MULTIVARIATE MODELLING OF COGNITIVE FUNCTION AND BRAIN STRUCTURAL DATA

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by:

CHRISTOPHER PAUL CHEYNE

OCTOBER 2011

ABSTRACT

Multivariate modelling of cognitive function and brain structural data

Christopher P. Cheyne

Previous studies have investigated links between cognitive ability and a number of factors including age, gender, handedness, musical ability as well as the volume and surface area of certain brain structures. However, in these studies either the explanatory variables are analysed independently of each other, or the investigation is based on a separate analysis for individual cognitive outcomes (e.g. language, visuospatial, etc.) The main objectives of this thesis are (1) to develop general multivariate models, which include mixed-effects terms, to account for the correlation in the data, (2) to explore the possible associations in children and adults between multiple cognitive ability test scores and the range of factors mentioned above by simultaneously applying the multivariate models designed in (1), and (3) to investigate the possible effects of missing data on the results.

To meet these objectives, a range of statistical and stereological methods was employed. Multivariate linear and linear mixed models were developed and fitted to multiple datasets. The approach used took into account the correlation of clustered data, the correlation between outcomes as well as the association between explanatory variables and a linear combination of the outcomes. Stereological methods were used to estimate the volume and surface area of a region of the brain called Broca's area, using magnetic resonance images. Also, the latest formulae in error prediction for these stereological estimates were described and applied to the data.

Results from the fitted multivariate linear mixed model to a dataset of 11-year old children (n=11843) showed that children whose writing hand has less hand skill than the opposite hand performed worse, on average, in both reading and maths scores, than those children whose writing hand had more hand skill than the opposite hand. A multivariate linear model fitted to a dataset of adults (n=142) revealed that the gender difference found in the non-musician groups for the vocabulary and arithmetic scores was not present in the musician group. Multivariate linear models were subsequently fitted to a subset of this cohort containing volume and surface area estimates of Broca's area (n=39). Musicians were associated with Broca's area being less convoluted in the right hemisphere than non-musicians. Other associations investigated were not found to be statistically significant.

Inverse probability weighting was then used to take the missing data into account for each of the analyses (aim (3)). The results and interpretations determined from the fitted multivariate models were consistent with the analyses when the missing data were accounted for. Only those results for the children dataset changed slightly, but not enough to alter the interpretations of the results. This adds weight to the belief that the results of the multivariate analyses gave a reasonably accurate description of the variability that exists within the children and adult datasets.

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank a number of people without whose advice and motivation would have made the experience of writing this thesis less fulfilling than it was. Firstly, I'd like to give recognition to my primary supervisor, Dr. Marta Garcia-Fiñana for many informative discussions and expert advice which have proven to be invaluable during the course of the entire PhD. Secondly, I'd like to thank my two secondary supervisors, Dr. Vanessa Sluming and Prof. Neil Roberts for their assistance in understanding about MR images, regions of the brain and for use of the musician-control dataset. Thanks also to my third supervisor, Prof. Kenneth Wilson for his comments about the common theme of the thesis.

Thanks must also go to Dr. Steven Lane for reading through the thesis in its entirety and giving feedback. Other people who have been involved in discussions related to the work involved in this PhD project included Prof. Tim Crow, Dr. Stuart Leask, Dr. Patricia Cowell, Joanne Powell and Dr. Ihssan Abdul-Kareem and my thanks to them all.

I would also like to acknowledge the funders of this project, the Medical Research Council, as well as the Centre for Longitudinal Studies for allowing myself to use the National Child Development Study dataset in this project.

Some of the deepest gratitude should certainly be reserved for my parents, Margaret and Stan, who have continually provided a strong, stable platform from which to continue my academic career. Without their love, care and attention over the years, I would not have reached as far as I have. Along with my parents' motivation and never-say-die spirit, other family members have also endeavoured to encourage me towards the end goal of the PhD, so my thanks also to them.

In terms of sorting out administrative issues, I couldn't have asked for a better admin team than the one in the Department of Biostatistics. So my thanks also to Karen Barnes, Kath Forrest and Jo Hollett for their assistance as well as sometimes lending an ear over a cup of tea when the stress of the thesis was greatest!

I also wish to express my sincere thanks to all members of the Department of Biostatistics, some of whom have left during the course of the PhD. You all made it a pleasure to come into university each day. Even my office colleagues, Dr. Jamie Kirkham and Dr. Steven Lane, contributed to the *PhD experience*!

Finally, I wish to thank all my friends for their patience and understanding in why I haven't always been available, especially in the past year or so, but still have been hanging in there and keeping in touch.

CONTENTS

LIST OF TABLES

Table # *Description*

CHAPTER 1 – Introduction

CHAPTER 2 – Methodology

CHAPTER 3 – Investigation of the associations between handedness and cognitive function in children

CHAPTER 4 – Investigation of the associations between musical ability and cognitive function in adults

CHAPTER 5 – Investigation of the associations between Broca's area and cognitive function in adults

CHAPTER 6 – Missing Data

R codes for some of the results in these tables can be seen in Appendix A.

LIST OF FIGURES

CHAPTER 3 – Investigation of the associations between handedness and cognitive function in children

CHAPTER 4 – Investigation of the associations between musical ability and cognitive function in adults

CHAPTER 5 – Investigation of the associations between Broca's area and cognitive function in adults

R codes for some of these figures can be seen in Appendix A.

LIST OF ABBREVIATIONS

Abbreviation	Definition
2D	Two-dimensional
3DMR	Three-dimensional mental rotation
3MS	Modified Mini-Mental State Examination
AC	Anterior commisure
AC-PC	Anterior commisure – Posterior commisure
AIC	Akaike's Information Criterion
BA44	Brodmann's area 44 (Pars Opercularis)
BA45	Brodmann's area 45 (Pars Triangularis)
BASA	Surface area of Broca's area (Broca's surface area)
BAV	Volume of Broca's area (Broca's volume)
BIC	Bayesian Information Criterion
CC	Chris Cheyne (observer)
CE	Coefficient of error
CSF	Cerebrospinal fluid
CVLT-II	California Verbal Learning Test, 2nd Edition
DAS	Differential Ability Scales
fMRI	Functional magnetic resonance imaging
GCA	General cognitive ability
GRO	General Registry Office
IFG	Inferior frontal gyrus
IPW	Inverse probability weighting
JOL	Benton judgment of line orientation
LH	Left hemisphere
LMM	Linear mixed model
MAR	Missing at random
MARIARC	Magnetic Resonance Imaging and Analysis Research Centre
MC	Musician-control
MCAR	Missing completely at random
MCAT	Medical College Admission Test
MI	Multiple imputation
ML	Maximum likelihood
MLM	Multivariate linear model
MLMM	Multivariate linear mixed model
MR	Magnetic resonance
MRI	Magnetic resonance imaging
NCDS	National Child Development Study
NMAR	Not missing at random
PC	Posterior commisure
PO	Pars Opercularis
PT	Pars Triangularis
RAVLT	Rey Auditory Verbal Learning Test
RelBASA	Broca's surface area relative to Broca's volume (relative Broca's surface area)

RelBAV	Broca's volume relative to total brain volume (relative Broca's volume)
RelHS	Relative hand skill
REML	Restricted maximum likelihood
RH	Right hemisphere
RHPO	Pars Opercularis in the right hemisphere
RIGLS	Restricted iteratively generalized least squares
RoI	Region of interest
RVDLT	Rey Visual Design Learning Test
SH	Superior hand
TBV	Total brain volume
ULM	Univariate linear model
VBM	Voxel-based morphometry
WAIS	Wechsler Adult Intelligence Scale
WH	Writing hand

CHAPTER 1

Introduction

This thesis encompasses a number of specific multivariate analyses to cover the aim of providing evidence, or identifying a lack of evidence, for associations between multiple cognitive abilities, handedness, musical ability and measurement estimates of a region of the brain called Broca's area. There are a number of abilities which are defined as cognitive (e.g. language, spatial awareness, etc.) which may have differing links to factors such as gender or musical ability. By taking multiple cognitive abilities and factors into account in individual models, better estimates of these links can be found.

There have been several studies where the connection between cognitive ability and a range of factors including handedness, age, musical ability and the activation of Broca's area was investigated (Schubotz & von Cramen 2002a,b; Salthouse, 2006; Thilers *et al.*, 2007; Jakobsen *et al.*, 2008; Nicholls *et al.*, 2010). However, the results from these studies are far from conclusive (more details about them can be seen in Section 1.2). It is interesting to note that most studies, while taking multiple factors into account, have focussed mainly on the effects of individual factors on cognitive ability without considering interactions between the factors.

Either the general cognitive ability (GCA) value, which is comprised of a number of different cognitive ability test scores (e.g. language, mathematical, memory, etc.), or individual cognitive abilities have been investigated as outcome variables in independent linear regression models. Furthermore, measurements of hand skill have been obtained by various methods (e.g. Purdue's Pegboard test, Edinburgh test, etc.) and links considered with cognitive ability. However, the links between cognitive ability and the handedness of a particular task (e.g. a component of the Edinburgh test such as writing, brushing teeth, picking up an object, etc.) have not been reported along with an estimate of hand skill in the same study.

As with most datasets, missing data may be a problem. If there are individuals with missing data, either outcomes or explanatory variables, and if there isn't an even spread of these individuals across groups of interest within the dataset, then this could lead to biased results. For example, let there be a dataset with equal numbers of people from three socio-economic groups (unemployed, manual employment and skilled employment) and 25% of these people have missing data. If the majority of those individuals with missing data are from any one particular group, then the influence on the results of any analyses performed will be greater from people in the other two socio-economic groups than those in the group with the majority of the missing data. However, if the missing data are spread evenly across all socio-economic groups then, given that the missingness of the outcomes is not dependent upon any other characteristics of interest, the missing data should not influence the results. Therefore, missing data and their effects on the original analyses' results should also be investigated.

The objectives of this thesis are presented and described in detail in the following subsection. A general introduction into the background of the research area follows. We then introduce the datasets that will be used in the multivariate analyses. Finally, the structure of the thesis is the final component of this introductory chapter.

1.1 Objectives

The objectives of this thesis and the breakdown by chapter can be seen in Table 1.1. Section 1.4 gives a more detailed description of the thesis structure. There are six objectives for this thesis. They are:

i. To construct a number of statistical models which:
    a. allow for the investigation of associations between explanatory variables and a linear combination of multiple outcomes.
    b. allow for the investigation of associations between explanatory variables and multiple outcomes individually.
    c. take into account correlations among clusters of data.
    d. take into account correlation between outcomes.

which when applied to specific datasets can be used to meet the objectives below. *The methodology for multivariate linear mixed and linear models can be seen in Chapter 2. Individual models and their structures are explained within Chapters 3-5.*

ii. To investigate the associations between multiple cognitive ability scores considered simultaneously in one model, and:

   a. handedness
   b. gender

   in children, whilst at the same time taking into account any other possible factors that may have some association with the cognitive ability test scores (e.g. UK region). *The analyses for this objective can be seen in Chapter 3, with writing hand, superior hand and relative hand skill (see Section 1.3.1 for definitions) all linked to handedness and cognitive ability scores represented by reading and mathematics scores.*

iii. To look into the possible links between multiple cognitive ability scores considered simultaneously in one model, and:

   a. musical ability
   b. gender
   c. age

   in adults. *The analyses for this objective can be seen in Chapter 4, with cognitive ability scores represented by vocabulary, arithmetic and visuospatial scores.*

iv. To examine the possible relationships between volume estimates of Broca's area (Broca's volume), and:

   a. musical ability
   b. gender
   c. age
   d. cognitive ability scores:
      i. vocabulary scores
      ii. arithmetic scores
      iii. visuospatial scores

in adults. *The analyses for this objective can be seen in Chapter 5. To account for differences in the size of Broca's area due to total brain size, we investigated the links described above with Broca's volume estimates relative to total brain volume (relative Broca's volume) as opposed to raw Broca's volume estimates.*

v. To explore the possible associations between surface area estimates of Broca's area (Broca's surface area), and:
   a. musical ability
   b. gender
   c. age
   d. cognitive ability scores:
      i. vocabulary scores
      ii. arithmetic scores
      iii. visuospatial scores

   in adults. *The analyses for this objective can also be seen in Chapter 5. To account for differences in Broca's surface area due to Broca's volume, we investigated the associations described above with Broca's surface area estimates relative to Broca's volume (relative Broca's surface area) as opposed to raw Broca's surface area estimates.*

vi. To investigate the effect of missing data on the analyses in Chapters 3 and 4 (see Table 1.1). *Both of the datasets used in the analyses in those chapters include participants that have missing data (either outcomes or explanatory variables). So, in Chapter 6, a method is deployed that adjusts the analyses' results for the missing data.*

Ch. #	Associations Investigated/ Objectives	Outcome Variables	Explanatory Variables	Statistical Analyses
3	Handedness, gender and UK region by cognitive ability	(1) Reading scores (2) Mathematics scores	Writing hand (L/R) Superior Hand (L/R) Relative Hand Skill Gender (M/F) UK region (NE&M, SE, WAL, SCOT)	Multivariate linear mixed model (MLMM)
4	Musical ability, gender and age by cognitive ability	(1) Vocabulary scores (2) Arithmetic scores (3) Visuospatial scores	Musical Ability (M/C) Gender (M/F) Age	Multivariate linear model (MLM)
5	Musical ability, gender, age, cognitive ability test scores by Broca's volume	(1) Left relative Broca's volume (2) Right relative Broca's volume	Musical Ability (M/C) Gender (M/F) Age Vocabulary scores Arithmetic scores Visuospatial scores	MLM
	Musical ability, gender, age, cognitive ability test scores by Broca's surface area	(1) Left relative Broca's surface area (2) Right relative Broca's surface area	Musical Ability (M/C) Gender (M/F) Age Vocabulary scores Arithmetic scores Visuospatial scores	MLM
6	Handedness, gender and UK region by cognitive ability taking missing data into account	(1) Reading scores (2) Mathematics scores	Writing hand (L/R) Superior Hand (L/R) Relative Hand Skill Gender (M/F) UK region (NE&M, SE, WAL, SCOT)	Inverse probability weighting (IPW) IPW weighted MLMM
	Musical ability, gender and age by cognitive ability adjusting for missing data	(1) Vocabulary scores (2) Arithmetic scores (3) Visuospatial scores	Musical Ability (M/C) Gender (M/F) Age	IPW IPW weighted MLM

Table 1.1: *A summary of the objectives of the thesis, variables involved and analyses by chapter (Ch.).*

1.2 Background

There have been a great many investigations aiming to understand the reasons behind the varying cognitive ability levels that people have. Cognitive ability was defined by C. Spearman in 1904, as comprising of two types of factors (Zhang *et al.*, 2010):

i. A general factor
ii. A number of specific factors

Each specific factor (individual ability; e.g. language, mathematical, etc.) that improves a person's performance for a given cognitive ability task can be summarised by the general factor. The general factor can be assessed by the use of psychometric battery tests, is comparable between everyone, and is comprised of a number of sub-tests (specific factors). There are a number of different cognitive ability battery tests such as the Wechsler Adult Intelligence Scale (WAIS) and the Differential Ability Scales (DAS). Both of these tests have a number of sub-tests, with these scores combined giving an estimate of IQ (general factor). The WAIS is split into two main areas which are verbal and performance abilities (Aitken, 1996). These in turn are made up of many sub-tests including vocabulary, arithmetic, digit span, block design and symbol search tests. The DAS similarly comprise a number of sub-tests including verbal, performance (non-verbal) and spatial ability tests (Elliott, 1990).

A number of possible factors have been investigated in terms of their association with cognitive ability and reported upon, such as handedness, gender, age and musical ability (Salthouse, 2006; Thilers *et al.*, 2007; Jakobsen *et al.*, 2008; Nicholls *et al.*, 2010). Results from some of the previous studies into these associations are detailed below starting with the link between cognitive ability and handedness.

1.2.1 Handedness and cognitive ability

An early investigation into the links between handedness and cognitive ability predicted that reading disability and issues surrounding speech were more common for participants with inconsistent cerebral asymmetry (i.e. where the participants' brains did not conform to having a larger, more dominant left hemisphere than right hemisphere) (Orton, 1937). This study also reported that inconsistent cerebral asymmetry was associated with left handedness and ambidextrousness (i.e. those people with no hand preference between their left and right hands). One issue that has been considered after Orton's publication is how to measure handedness. The simplest way to determine whether an individual is left- or right-handed is to simply ask them directly. An alternative method is to ask the participant to perform a task, such as write their name, or pick up a pencil, and the hand with which they perform the task is assigned to be the hand they use most often. However, these methods do not record how often individuals use their preferred hand. Some people may use a combination of their two hands, depending upon the type of task, or even depending upon the task itself. For example, if the pencil was to the left of the person, they may be more likely to pick it up with their left hand, even though they would pick up the pencil with their right hand if it was near their right hand. Some methods for being able to differentiate the level of skill attributed to each of a participant's hands have been constructed. Three of these methods are:

- Box ticking method
- Edinburgh Handedness Inventory test
- Purdue Pegboard test

An individual having their handedness measured by the box ticking method had to tick as many boxes in one minute with each hand in turn (Leask & Crow, 2006). The two scores for each hand could then be compared as a measure of hand skill. The Edinburgh Handedness Inventory test (or otherwise known as simply the Edinburgh test) involves the participant being asked twenty questions about which hand they use to perform specified tasks (Oldfield, 1971). Questions include which hand the individual uses to brush their teeth, use a hammer and swing a golf club. The Purdue

Pegboard test was originally designed to assist industry in selecting potential employees (Tiffin & Asher, 1948). A modified version of this test consists of the singular timing of both hands moving pegs between positions on the board (Annett, 2002). The peg-moving tests give a continuous outcome for each hand (time), as opposed to discrete outcomes for the other two tests (since there are only a finite number of possible outcomes).

Comparing left-handed people to right-handed people in terms of cognitive ability has produced some interesting results. Some studies have suggested that left-handed individuals show a cognitive advantage over right-handed individuals (McManus, 2002). However, conflicting studies have said this supposed cognitive advantage of left-handed participants does not exist (Resch *et al.*, 1997). Indeed, others have obtained results which suggest that left-handedness is associated with lower scores across a number of cognitive ability tests (Johnston *et al.*, 2009).

General cognitive ability (GCA) was reported to have a quadratic relationship with hand skill (Nicholls *et al.*, 2010). Hand skill was estimated by counting the number of times each participant tapped on a circle on a touch screen in 30 seconds with each hand. GCA was estimated using the cognition battery of tests as seen in Gordon, (2003). This result implies that people in the study with extreme handedness (i.e. those individuals that use their left or right hand much more than their opposite hand) were associated with lower levels of cognitive ability than those participants with moderate handedness (i.e. those individuals that have one hand only moderately more skilled than the other hand). A quadratic relationship was also reported between reading test scores and the absolute difference in skill scores between partakers' left and right hands (Annett & Manning, 1990). These results agree to a certain extent with Crow *et al.*, (1998) who stated that some moderate deficits in cognitive ability (including verbal, non-verbal, reading and mathematical skills) in a group of 11-year old children were evident for those children with extreme hand skill. However, where they differ, is that Crow *et al.*, (1998) suggest that lower cognitive abilities are associated with children that are at the "point of hemispheric indecision" (i.e. they are ambidextrous). This "point of indecision" was defined as the zero point of a laterality index variable (see McManus, 1985, for laterality index definition). This was also found when 1,355 participants of varying ages between 18 and 60+ years old from a New Zealand TV show were tested for IQ, with mixed-handedness associated with lower cognitive IQ sub-test scores than left- and right-handedness (Corballis *et al.*, 2008).

In 2005, a BBC online Internet survey was conducted where 255,100 participants were given a mental rotation task to test their spatial performance, as well as a fluency/reasoning task (Peters *et al.*, 2006). The fluency/reasoning task involved each subject listing as many words in 2 minutes that could be linked to something grey. The mental rotation task, a redrawn Vandenberg and Kuse, (1978) mental rotation task, involved 24 questions consisting of 5 images of 3-dimensional objects in different rotations and subjects had to identify the two images which were of identical objects (Peters *et al.*, 1995). The majority of those participants who performed worst in the spatial performance task appeared to be those with no preference for handedness. Also, participants who were either extremely left- or right-handed tended to show deficits when compared to those who were mostly left- or right-handed for the spatial task. The extreme handedness group performed worst in the fluency/reasoning task, followed by those with no preference for handedness.

1.2.2 Gender and cognitive ability

Differences between the sexes in a variety of cognitive abilities have long been commented on. Girls are generally better at articulation and sentence production when they are very young than boys and this may have further consequences later in childhood with girls also associated with larger vocabularies, making better use of grammar as well as better reading skills (Sommer, 2010). Females have also been linked with higher scores than males in memory and verbal tests (Thilers *et al.*, 2007). However, for some non-language skills, males have been found to have an advantage over females. For example, a meta-analysis found that males outperformed females in various spatial ability tests (Voyer *et al.*, 1995). One further interesting result is that Thilers *et al.*, (2007) suggested that the magnitude of the gender differences in these cognitive ability tests was greater for right-handed individuals than for left-handed individuals. This could indicate that future analyses should include an interaction term between gender and handedness so that this possible gender by handedness variability within cognitive ability can be adjusted for.

1.2.3 Age and cognitive ability

In terms of the effect of age on cognitive ability, it has been reported that older people score lower on cognitive ability tests than younger people (Gunstad *et al.*, 2006; Salthouse, 2006; Mitnitski & Rockwood, 2008; Mitnitski *et al.*, 2010). When an individual is between 20 and 40 years of age, there starts a general decline in episodic memory, perceptual reasoning and perceptual speed ability test scores with age (Salthouse, 2006). However, no statistically significant reduction in the scores from a vocabulary test occurred as age increased. Other studies have focussed on general cognitive ability both as a summary measure of different cognitive abilities as well as individual cognitive abilities, and when comparing middle-aged and younger-aged participants to old-age participants, the old-age participants have shown to have significantly lower cognitive function scores than the other age groups (Gunstad *et al.*, 2006; Mitnitski & Rockwood, 2008; Mitnitski *et al.*, 2010).

1.2.4 Musical ability and cognitive ability

Understanding the role of music on a variety of different cognitive abilities, such as language and spatial awareness, is another area where there has been much debate and interest in recent years. A number of studies have attempted to identify differences in cognitive functions, as well as in brain structures, between musicians and non-musicians. Differences have been found in a number of tasks, such as, verbal and visual memory, visuospatial and auditory tasks (Sluming *et al.*, 2002; Sluming *et al.*, 2007; Franklin *et al.*, 2008; Jakobsen *et al.*, 2008, Parbery-Clark *et al.*, 2009). In particular, Jakobsen *et al.*, 2008, reported that musicians have superior verbal and visual memory than controls. Verbal memory in this study was measured by the California Verbal Learning Test, 2nd edition (CVLT-II) whereas visual memory was measured by the Rey Visual Design Learning Test (RVDLT). The CVLT-II involved the immediate and delayed recall of word lists and the RVDLT is based around the learning, delayed recall and delayed recognition of visual designs. Even after adjusting for the IQ of each individual, there were still differences between musicians and controls in the delayed recall of both words and visual designs.

Similarly, Franklin *et al.*, (2008) highlighted that musicians appear to have an improvement in long-term verbal memory and a greater verbal working memory span when compared to controls. In this case verbal memory was assessed using the Rey Auditory Verbal Learning test (RAVLT) which is a very similar test to the CVLT-II. The RAVLT also involved the delayed recall of words from a list, although the list comprised 15 unrelated words (see Schmidt, (1996) for details) whereas the list of words of the CVLT-II consisted of 16 words with groups of 4 words related to each other (see Delis *et al.*, (2000) for more details). The verbal working memory tests administered involved sentences and mathematical equations which had to be read aloud before responded to as being correct or false. The capitalized letter or word which followed each equation was then meant to be repeated in the correct serial order.

In a study by Sluming *et al.*, (2007), the Benton judgement of line orientation (JOL) test was used to evaluate spatial ability and it was found that musicians achieved higher scores in this test than controls. The JOL test involved a two-dimensional (2D) line being rotated such that the participants had to estimate the degree of rotation. In addition to this result, a comparison was carried out between musicians and controls using a three-dimensional mental rotation (3DMR) task (Sluming *et al.*, 2007). This 3DMR task consisted of ten pairs of cubes arranged in chiral patterns, with one of these rotated about a given axis and participants have to decide whether the two objects were identical or not. Once again, as for the 2D spatial ability test, musicians attained significantly better scores than controls in the 3DMR task.

1.2.5 Broca's area and cognitive ability

Broca's area, a region of the brain named after Paul Broca who is credited with its identification, was linked to the processing of language after those with speech difficulties had lesions in the area now known as Broca's area (Broca, 1861). The two subjects of Broca's study, Leborgne and Lelong, each had only a very limited vocabulary and both had lesions in the frontal lobe of the left hemisphere (i.e. Broca's area) (Dronkers *et al.*, 2007). Located in the inferior frontal gyrus, Broca's area comprises of Pars Opercularis (PO) and Pars Triangularis (PT), which are two distinct parts. PO and PT are denoted in Brodmann's cytoarchitectonic map as areas 44 and 45, respectively (see Figure 1.1; Brodmann, 1909).

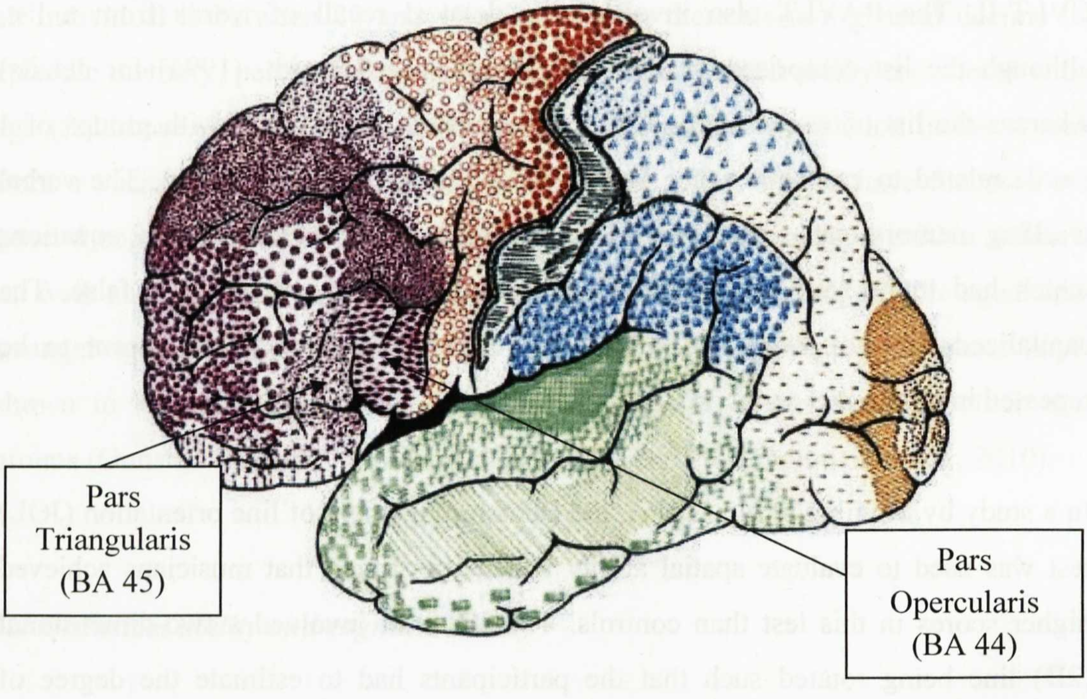


Figure 1.1: *Brodmann's 1909 cytoarchitectonic map with Pars Opercularis (BA 44) and Pars Triangularis (BA 45) highlighted (modified from Zilles & Amunts, 2010).*

Further investigations have confirmed that language, or more specifically speech production, is strongly linked with the inferior frontal gyrus (which Broca's area is a part of) in the left hemisphere (Stephan *et al.*, 2003, Toga & Thompson, 2003; Costafreda *et al.*, 2006). More specific language related skills, such as syntactic processing which is linked to the structure of sentences, have also been associated with Broca's area (Caplan, 2006).

Furthermore, Broca's area is involved in the processing of non-linguistic information such as music, sequencing, action recognition and visuospatial cognition according to other studies (Koelsch *et al.*, 2002; Schubotz *et al.*, 2002a,b; Hamzei *et al.*, 2003; Sluming *et al.*, 2007). Links between general cognitive ability scores, such as IQ, have been found to be correlated with certain regions of the brain (Haier *et al.*, 2005). Interestingly, Haier *et al.*, (2005) found these links to be different between males and females with males showing stronger correlations in the volumes of frontal and parietal lobes, whereas females had correlations that were strongest in the volumes of the frontal lobe and Broca's area. This result suggests that there may be, not only a gender difference in the volume of Broca's area, but also an interaction between gender and cognitive ability. When looking at possible gender differences in the volume of Broca's area, it would be useful to also consider the fact that females have smaller brain volumes than males due to their generally smaller stature (Cosgrove *et al.*, 2007). Therefore, a way of adjusting the volumes of Broca's area for this overall gender variation should be considered.

Following on from the studies linking musicians with higher cognitive ability scores than non-musicians (in Section 1.2.4), differences in the activation of a variety of areas of the brain have been observed between musicians and non-musicians (Schlaug *et al.*, 1995; Amunts *et al.*, 1997; Sluming *et al.*, 2002; Gaser & Schlaug, 2003; Lee *et al.*, 2003; Bengtsson *et al.*, 2005; Schneider *et al.*, 2005; Bangert & Schlaug, 2006; Stewart, 2008). Some areas have been highlighted as being active when music is being processed (Stewart *et al.*, 2003; Stewart, 2005; Sluming *et al.*, 2007). This can be seen in Broca's area where musicians have been found to have greater activation in Broca's area as well as having larger grey matter volume in the left hemisphere only, than non-musicians (Sluming *et al.*, 2002; Sluming *et al.*, 2007). This result along with the fact that language skills are mainly associated with the left hemisphere suggests that the volume estimates of Broca's area in different hemispheres should be considered as two separate variables.

After reviewing the previous results in the areas of the objectives of this thesis, it was clear that more research needed to be undertaken on these topics. We wanted to add to this knowledge-base and to do this, we needed to analyse a number of different datasets. One of the reasons for requiring multiple datasets is that we wanted to analyse associations between cognitive ability and the factors listed above for both children and adults where possible. We were, of course, limited to being able to test for links between cognitive ability and factors which were included at the data collection stage as well. These datasets are introduced in the following subsection.

1.3 Introduction to the datasets

This thesis includes analyses of two different datasets, as well as the analysis of a subset of one of the datasets. The first dataset came from the National Child Development Study and will be described in detail in Section 1.3.1. The second dataset was collected at the University of Liverpool in 2000 and includes data from musicians and non-musicians (controls). This dataset is described in detail in Section 1.3.2 and will be referred to throughout the thesis as the musician-control (MC) dataset. For the final set of analyses, a subset of the MC dataset was taken based on near-identical group sizes of male and female, musicians and controls with the measurement estimates of Broca's area added. An overview of this subset can be seen in Section 1.3.3.

1.3.1 The National Child Development Study dataset

Originally the National Child Development Study (NCDS) was set up to obtain data from all births on the UK mainland within a period of one week (3-9 March, 1958) as a study of perinatal mortality (Shepherd, 1995; Leask & Crow, 1997; Leask & Crow, 2006; Power & Elliott, 2006). Subsequently, a decision was taken to extend the scope of the study to investigate the main factors that affect human development. To do this the cohort was followed-up at various ages including at 7, 11 and 16 years of age. The NCDS is still collecting data with follow-ups on this cohort to the present day and beyond. These data are stored by and can be obtained from the Centre for Longitudinal Studies. A number of different types of variables were collected for the dataset, with different variables obtained at different sweeps. Some sweeps, such as the one at age 44-45 in 2002/03, were specifically targeted to gain data which would answer certain questions about the cohort, and this particular sweep was classed as a biomedical survey (Power & Elliott, 2006). Usually a range of factors such as socio-economic, medical and cognitive factors were included in various sweeps. One of the objectives of this thesis was to examine possible associations between multiple cognitive ability test scores, handedness and gender, in children, using a single model. Therefore, the NCDS data collected in 1969, consisting of 18,558, 11-year old children, was considered.

Cognitive ability in this dataset is comprised of two sets of test scores:

- *Reading scores*
- *Mathematics scores*

The reading test was comprised of 35 comprehension questions which involved the children having to complete a sentence by choosing one of five given words (Crow *et al.*, 1998). An example of this was:

A bird lays its eggs in a (pond, stream, cloud, house, nest)

with the correct answer being 'nest'. Children were given twenty minutes to complete the test. The mathematics test included questions based on three areas of mathematics (arithmetic, logic and geometry). A total of 40 questions were asked with the children having 40 minutes to answer them all.

The methods of statistical analysis that were used to address the research questions posed in this thesis comprised of multivariate linear and linear mixed models (see Table 1.1). These methods are explained in detail in the next chapter. One of the assumptions of both of these types of models is that the error terms are normally distributed and an indicator of this is the normality of the outcome variables (see Section 2.1 for more details). In the case of these analyses, the reading and maths scores will be the outcomes. Initial observations show that the reading scores were approximately normally distributed but the maths scores look highly skewed (i.e. not normally distributed). This lack of normality of the maths scores is confirmed by the Anderson-Darling test (p<0.001).

Variables which are not normally distributed can be transformed into new variables that follow a normal distribution more closely. By transforming the maths scores and using the transformed scores as an outcome in the statistical analyses the corresponding error terms are expected to show a distribution closer to the normal distribution. A variety of possible transformations of the maths scores were considered including a log transformation, an exponential transformation and a number of power transformations. The log and exponential transformations made the variable more skewed. Increasing or decreasing the power transformation also resulted in increasingly skewed variables. The transformation which produced a variable which most closely followed a normal distribution was a Box-Cox transformation. This transformation can be expressed as:

$$y^* = \left((y+3)^\lambda - 1\right)/\lambda \tag{1.3.1}$$

where y^* is the new transformed variable, y is the original, untransformed variable and λ is the transformation parameter (Box & Cox, 1964). The transformation parameter is chosen such that the transformed variable y^* is closer to a normal distribution by using maximum likelihood estimation, and in this instance $\lambda = 0.512$. The distribution of the newly transformed mathematics score variable looks closer to the normal distribution when compared to the distribution of the original mathematics scores (see histogram in Figure 1.2).

The mean of the transformed scores was 11.01 (sd=4.36) with median equal to 11.04. The fact that the mean and median are close together is a good indicator of a symmetric distribution. Although the Anderson-Darling test still suggests that the hypothesis of normality is rejected (p<0.001), when datasets are very large, as is the case here with a sample size equal to 18,558, normality tests will highlight and magnify small differences in the data from a normal distribution. To justify how good the final fitted model was to the actual data, the distribution of the standardized residuals derived from the model fitted were analysed to check whether the assumption of normality held (see Section 3.4.3).

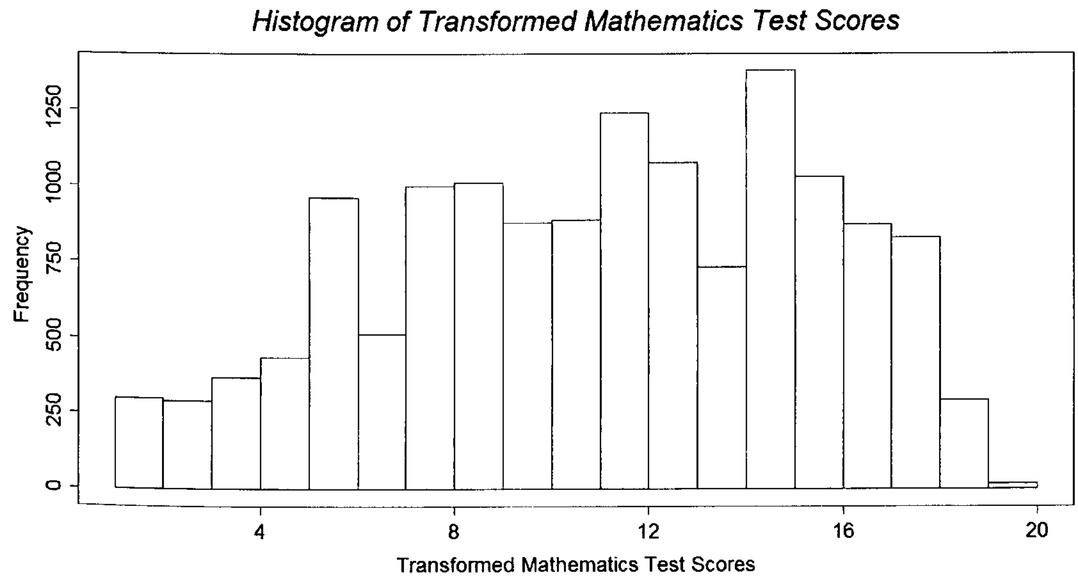


Figure 1.2: *Histogram of the transformed NCDS mathematics scores using Equation (1.3.1) with transformation parameter* $\lambda = 0.512$.

The main factors with links to cognitive ability being investigated according to the objective for this section are gender and handedness. *Gender* is a binary variable, with males set equal to 0 and 1 representing females. There were two variables included in the NCDS dataset that could be representative of handedness. Handedness was recorded both as a binary variable, in terms of which hand the child writes with, and a continuous variable, based on the number of boxes ticked in 1 minute with either hand (i.e. a measure of hand skill). The hand skill measure obtained from the box-ticking task is more informative than simply stating whether a person is left- or right-handed. As opposed to considering hand skill individually for every child's left and right hand it was decided that it would be more interesting to look at the relative hand skill between the left and right hands. Hence, two new variables were created from these two box-ticking measures. The first is called *relative hand skill* which was calculated as:

$$\text{relative hand skill} = \left|\frac{n(RH)-n(LH)}{n(RH)+n(LH)} \times 100\right| \quad (1.3.2)$$

where n(RH) and n(LH) are the number of boxes ticked in 1 minute with the right and left hands, respectively (McManus, 1985). Relative hand skill can be interpreted as the relative superiority of the dominant hand over the non-dominant hand. This variable will always be positive therefore, an indicator variable is required to give information as to whether or not the left or right hand had ticked the greatest number of boxes (i.e. had greater hand skill) for each child. The second new variable, *superior hand* (SH) was created to denote which hand is the dominant one (i.e. had the greatest hand skill). SH is set equal to 0 if the dominant hand is the right hand (i.e. where the number of boxes ticked in 1 minute is greater for the right hand than for the left hand) and equal to 1 if the dominant hand is the left hand (i.e. more boxes ticked in 1 minute by the left hand than by the right). In addition to these two handedness variables, it was also deemed appropriate to include the other handedness variable from the NCDS dataset itself, *writing hand* (WH). WH is a binary variable which takes values 0 if the child writes with their right hand, and is equal to 1 if the child is a left-hand writer.

As mentioned in Table 1.1, a further variable representing the region of the UK in which the children attended school is also included in the NCDS analyses. In the original dataset a region variable exists in which each child is recorded as attending a school in one of eleven designated regions of the UK mainland (such as North-West England, North-East England, South-West England, etc.). However, in the interest of reducing the complexity of the analyses while attempting to still explain as much variability as possible, it was decided to reduce the number of regions to four:

- Northern England and the Midlands
- Southern England
- Wales
- Scotland

This new variable, *UK region*, was then used in the analyses in Chapter 3 as a factor with Northern England and the Midlands set as the reference region. Additionally, another variable, *local authority*, indicated which of 185 local authorities in the UK the school attended by the child was located in. *Local authority* was included in the statistical analyses as a clustering variable.

Table 1.2 shows descriptive statistics (i.e. population mean and standard deviation) for each binary and categorical explanatory variable by each outcome variable. We considered using both parametric tests (e.g. t-tests) and non-parametric tests (e.g. Mann-Whitney U tests), however, because of the large sample size and the fact that our outcome variables (reading and transformed maths scores) were approximately normally distributed, we give the *p*-values for the parametric tests in Table 1.2. When we compared the results from the t-tests to those from Mann-Whitney U tests, the differences between the *p*-values for the two tests were small, and all resulting interpretations (i.e. whether comparisons between groups were statistically significant or not) remained the same.

We used Welch's t-test to compare the reading and transformed maths scores between the two groups of gender, writing hand and superior hand. Welch's t-test was used as opposed to Student's t-test because we could not assume that the variances of the scores in each group were equal (due to the differing group sizes). In addition, we used basic linear regression models to test the hypothesis that there were no differences in test scores between the children who attended school in the different regions of the UK mainland (see Table 1.2 for *p*-values). Confidence intervals were constructed for the differences between the means of each group and are shown in Table 1.2 These confidence intervals are unadjusted because they do not take any other variables into account. Confidence intervals constructed after statistical analyses are performed will take any variables in the model into account so they are called adjusted confidence intervals.

There were no significant differences between males and females in either test score nor between left- and right-hand writers in the transformed maths scores (p=0.3 and p=0.2 for gender in the reading and transformed maths scores, respectively; p=0.05 for writing hand; Table 1.2). However, we can see that there was a difference between left- and right-hand writers in the reading scores such that right-hand writers had 1 extra average mark (p=0.05, Table 1.2). The t-tests for superior hand also show that children who had left superior hand obtained lower scores on average, in both tests than children with right superior hand (both $p<0.001$, see Table 1.2).

In Table 1.2, we can see that there was some justification for including UK region in the analyses from the fact that differences exist in the mean test scores between pupils in different regions of the UK according to the t-tests from basic linear regression models. For example, those children from Southern England had a statistically significantly higher mean reading score than those children in Northern England and the Midlands (the reference region), and children from both Southern England and Scotland were associated with a higher mean transformed maths score than children from Northern England and the Midlands ($p<0.001$, $p=0.008$ and $p<0.001$, respectively, Table 1.2). The descriptive statistics for the only continuous variable, relative hand skill, are shown in Table 1.3. The Pearson correlation coefficients in Table 1.3 suggest that there is very little correlation between relative hand skill and both reading and transformed maths scores. However, if the associations between these variables are non-linear as reported in previous studies (e.g. a quadratic relationship, see Section 1.2.1) then the Pearson correlation coefficients will not identify this.

These initial observations and hypotheses tests give a flavour into possible associations between each explanatory variable and outcome. However, they are unadjusted statistics, which do not account for other factors. The statistical analyses in *Chapter 3* take all the variables into account simultaneously, and therefore give stronger evidence for any associations that are statistically significant, than from direct observations from the dataset.

Variable	Group	n	Reading Scores			Transformed Mathematics Scores		
			Mean (sd)	*p*-value	95% CI	Mean (sd)	*p*-value	95% CI
Gender	Male	9596	45.5 (18.7)	0.3	(-0.9, 0.3)	11.1 (4.5)	0.2	(-0.06, 0.23)
	Female	8959	45.8 (17.2)			10.97 (4.3)		
Writing Hand	Left	1553	45.0 (18.5)	0.05	(0.02, 2.0)	10.90 (4.4)	0.05	(-0.01, 0.48)
	Right	12161	46.0 (17.9)			11.14 (4.3)		
Superior Hand	Left	1587	44.5 (18.4)	<0.001	(1.0, 3.0)	10.71 (4.4)	<0.001	(0.29, 0.77)
	Right	11177	46.4 (17.8)			11.24 (4.3)		
UK Region	Northern England & the Midlands	8238	45.2 (17.8)	*Reference region*		10.87 (4.4)	*Reference region*	
	Southern England	4718	46.9 (18.6)	<0.001	(1.0, 2.4)	11.09 (4.4)	0.008	(0.06, 0.39)
	Wales	817	43.9 (18.3)	0.05	(-2.7, 0.02)	11.08 (4.4)	0.2	(-0.12, 0.53)
	Scotland	1584	45.2 (16.5)	0.9	(-1.0, 1.0)	11.43 (4.0)	<0.001	(0.32, 0.81)
Overall		18558	45.7 (18.0)			11.01 (4.4)		

Table 1.2: *Population mean, standard deviation, p-values of t-tests and unadjusted 95% CIs of the difference between the means for the reading and transformed mathematics scores with a breakdown by each discrete variable grouping in the NCDS dataset.*

Variable	Mean (sd)	Reading Scores		Transformed Maths Scores	
		Correlation	*p*-value	Correlation	*p*-value
Relative Hand Skill	16.32 (8.7)	0.05	<0.001	0.07	<0.001

Table 1.3: *Population mean, standard deviation and Pearson correlation coefficients related to cognitive ability, with p-values, for relative hand skill.*

1.3.2 The musician-control dataset

The musician-control (MC) dataset was collected at the Magnetic Resonance Image and Analysis Research Centre (MARIARC), University of Liverpool, in 2000 for the purpose of investigating differences in brain activation between musicians and non-musicians (controls). The musicians were from the Royal Liverpool Philharmonic Orchestra while the controls were staff and students from the University of Liverpool. In the full dataset there are 149 participants; 40 musicians and 109 controls. Analyses of this dataset were conducted to investigate the possible associations between musical ability, gender and age and cognitive ability in adults (see Table 1.1). Originally, handedness was also considered as a possible factor in the analyses since, in this dataset, handedness was recorded using the Edinburgh Handedness Inventory test. However, concerns were raised when the breakdown of left- and right-handed individuals into male and female, musicians and controls were tabulated (see Table 1.4). It can clearly be seen that since there is only one left-handed control participant and only 5 left-handed participants altogether compared to 142 right-handers (2 individuals have missing handedness values) there is little reward to be gained from including left-handers in the analyses. Therefore, a decision was taken to include just the 142 right-handed individuals in the dataset. This reduced dataset, referred to from now on as the MC dataset, consisted of 106 non-musicians (55 male and 51 female), and 36 musicians (27 male and 9 female). The age of participants involved ranged from 19 to 94 years.

		Musician	Control	Total
Gender	Male	1	1	2
	Female	3	0	3
	Total	4	1	5

Table 1.4: *The distribution of left-handed participants by gender and musician/control groups.*

The outcome variables, representing 'cognitive function' were taken from the Wechsler Adult Intelligence Scale (WAIS) test (Wechsler, 1981). These were the

WAIS vocabulary, arithmetic and block design scores, and represented vocabulary, mathematical and visuospatial ability, respectively. The WAIS vocabulary test involved the participant being shown a word on a card, such as 'assemble', and then asked what the word means. Individuals were awarded up to 2 points depending on how good their understanding of the word was. For example, 'assemble' could be defined as 'to put together' or 'to join' which would be awarded 2 points (good understanding), or defined as 'to make something' or 'to build' which would be awarded 1 point (not incorrect answer but showing poverty of content), or defined as 'people' or 'work in a factory' which would be awarded no points (wrong responses). There were 35 words, which meant that the maximum possible mark was 70.

The mathematical aspect of 'cognitive function' was represented by the WAIS arithmetic scores. The arithmetic test consisted of the participants being asked questions, starting off easy with 'how much is 5 pounds plus 4 pounds?' and getting progressively more difficult, with points being awarded for correct answers. There were 20 questions, with a maximum possible mark of 22 (the final two questions were worth 2 marks each, with all others being worth 1 mark).

The third score (visuospatial score) was obtained for each individual using the WAIS block design test. This test involved the participant having to replicate a two-dimensional (2D) pattern using a range of cubes (2 faces red, 2 faces white and 2 faces half red, half white, see Figure 1.3). There were a total of nine 2D patterns with the first two patterns worth up to 2 marks, the proceeding two worth up to 6 marks, and the rest up to 7 marks, for a maximum mark of 51.

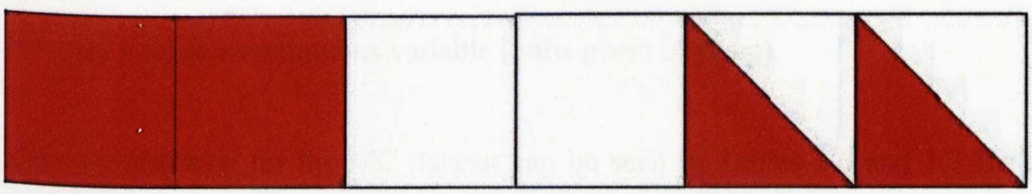


Figure 1.3: *The six sides of a cube used in the WAIS block design test.*

The WAIS test scores are combined in such a way that the resultant outcome is an estimate for a participant's IQ. To do this each individual WAIS component test score (e.g. vocabulary, arithmetic, etc.) is scaled non-linearly and weighted such that the lower scaled scores contain many more actual scores than the upper scaled scores. For example, from the vocabulary test, the scaled scores 1 to 9 (47.4% of all possible scaled scores) correspond to the actual scores 0 to 46 (66.2% of all possible unscaled scores), and the scaled scores 10 to 19 (52.6% of all possible scaled scores) represent the actual scores 47 to 70 (33.8% of all possible unscaled scores), as seen in Table 1.5. Also, an increase in one scaled mark in the vocabulary scores could be as large as 16 actual marks (an increase from the scaled score 6 to 7, could represent an actual score increase from 20 to 36) or as low as 1 actual mark (an increase from the scaled score 18 to 19, would represent an actual score increase from 69 to 70). These non-linear transformations are applied to the raw scores so that the scaled scores are reasonably normally distributed with a mean score roughly equal to 10 and standard deviation approximately equal to 3 (Kaufman & Lichtenberger, 1999). The original non-scaled test scores are not normally distributed, and yet these raw scores would be more useful in terms of interpretation of the statistical models. The MC dataset includes only the scaled WAIS sub-test scores with the actual test scores being unavailable. It is, therefore, most appropriate to use the scaled scores as the outcomes of the analyses in *Chapter 4*. After scaling according to WAIS, the possible range of scores was 0 to 17 for the arithmetic test and 0 to 19 for both the vocabulary and visuospatial tests.

Vocabulary Test		Arithmetic Test		Visuospatial Test	
Scaled Test Score	**Actual Test Scores**	**Scaled Test Score**	**Actual Test Scores**	**Scaled Test Score**	**Actual Test Scores**
1	0-5	1	0	1	0
2	6-8	2	1-2	2	1
3	9-10	3	3	3	2
4	11-13	4	4	4	3-7
5	14-19	5	5	5	8-13
6	20-28	6	6-7	6	14-19
7	29-36	7	8-9	7	20-22
8	37-42	8	10	8	23-26
9	43-46	9	11	9	27-30
10	47-51	10	12	10	31-34
11	52-54	11	13-14	11	35-37
12	55-59	12	15	12	38-41
13	60-62	13	16	13	42-43
14	63-64	14	17	14	44-46
15	65	15	18	15	47-48
16	66-67	16	-	16	49
17	68	17	19	17	50
18	69			18	-
19	70			19	51

Table 1.5: *The scaled and equivalent actual scores for the vocabulary, arithmetic and visuospatial tests from the WAIS.*

The explanatory variables to be tested for associations with the three cognitive ability tests according to Table 1.1 were:

- *Gender* - a binary variable with 0 = 'male' and 1 = 'female'.
- *Musician* - defined as a binary variable indicating whether a participant was a non-musician (control, factor musician = 0) or a musician (factor musician = 1).
- *Age* - set as a continuous variable (units given in years).

Descriptive statistics for the MC dataset can be seen in Tables 1.6 and 1.7 for the discrete and continuous variables, respectively.

Similarly as for the exploratory analyses of the NCDS dataset, we decided in favour of parametric tests over non-parametric tests. This was because the three cognitive ability score variables were all approximately normally distributed. Again, to check our results we performed Mann-Whitney U tests as well as t-tests and the resultant *p*-values were similar enough that we obtained the same interpretations as for the t-tests in Table 1.6. In terms of initial observations from Table 1.6, although there was no difference between males and females in both vocabulary and visuospatial scores according to Welch's t-test, this may not be accurate once multiple factors are adjusted for (p=0.6 and p=0.2, respectively). However, the difference between males and females in the arithmetic scores (males have a mean score of 11.1 compared to 9.6 for females) was statistically significant since the *p*-value given by Welch's t-test was less than 0.001.

In all three sets of test scores there does look to be a difference between musicians and controls with musicians having much higher mean test scores (14.4, 12.1 and 12.8 for musicians in the three tests, respectively compared to 10.3, 9.8 and 10.1 for the controls, Table 1.6). The t-tests agree with our assertions with differences found between musicians and non-musicians in each test score ($p<0.001$ for all three test scores; Table 1.6).

The mean age of these 142 subjects is approximately 48 years old (Table 1.7). We found that each set of cognitive test scores were negatively correlated with age and the Pearson correlation coefficients were significantly non-zero between age and both the arithmetic and visuospatial scores (both $p<0.001$, Table 1.7). This implied that there was fairly weak negative correlation (-0.3) between arithmetic scores and age and fairly strong, negative correlation (-0.7) between visuospatial scores and age.

Variable	Group	n	Vocabulary Test Scores			Arithmetic Test Scores			Visuospatial Test Scores		
			Mean (sd)	p-value	95% CI	Mean (sd)	p-value	95% CI	Mean (sd)	p-value	95% CI
Gender	Male	82	11.6 (2.9)	0.6	(-0.7, 1.2)	11.1 (2.5)	<0.001	(0.6, 2.4)	11.2 (2.5)	0.2	(-0.3, 1.6)
	Female	60	11.3 (2.2)			9.6 (2.4)			10.5 (2.8)		
Musician	Musician	36	14.4 (2.0)	<0.001	(-4.9, -3.4)	12.1 (2.4)	<0.001	(-3.2, -1.4)	12.8 (2.1)	<0.001	(-3.6, -1.9)
	Control	106	10.3 (1.8)			9.8 (2.3)			10.1 (2.5)		
Overall		142	11.5 (2.6)			10.4 (2.5)			10.9 (2.7)		

Table 1.6: *Population mean, standard deviation, p-values of t-tests and unadjusted 95% CIs of the difference between the means for the WAIS vocabulary, arithmetic and visuospatial scores with a breakdown by each discrete variable grouping in the MC dataset.*

Variable	Mean (sd)	Vocabulary Scores		Arithmetic Scores		Visuospatial Scores	
		Correlation	p-value	Correlation	p-value	Correlation	p-value
Age (in years)	48.4 (18.0)	-0.1	0.1	-0.3	<0.001	-0.7	<0.001

Table 1.7: *Population mean, standard deviation and Pearson correlation coefficients, with p-values, related to cognitive ability for age.*

The mean, standard deviation and test statistics for each explanatory variable and outcome are taken directly from the sample population, and so do not take any other factors into account. The statistical analyses applied in *Chapter 4* give coefficients for each factor as well as standard errors and confidence intervals which are adjusted for other factors in the model. This means that further associations which are not able to be directly observed from the data may be found. Also, added weight is given to any previously observed links between cognitive ability scores and gender or musical ability already mentioned above if they are observable in the statistical model.

1.3.3 Broca's area subset

As an extension to the analyses of the MC dataset, we decided that the relationships between Broca's area and all factors previously mentioned (age, gender, musical ability as well as the three cognitive ability scores) should be investigated by using a subset of the MC dataset itself. Broca's area has historically been linked to cognitive ability, but differences in Broca's area between males and females, musicians and non-musicians are less certain. Measurement estimates (volume and surface area) of Broca's area were added to this subset using stereological methodology which is introduced in the next chapter. It was decided that the subset should be balanced with 10 individuals in each of the male and female, musician and non-musician groups. There were more than 10 male and female non-musicians as well as male musicians in the MC dataset. Therefore, to obtain 10 individuals, a random sample was taken from each separate group. However, since there were only 9 female musicians in the dataset, it was not possible to include 10 female musicians in the subset. The characteristics of individuals in the subset were compared to those in the MC dataset, and were found to be similar (i.e. not statistically significantly different) in terms of cognitive scores and age.

The outcomes for this part of the thesis are related to the volume and surface area estimates of Broca's area. The estimates of both volume and surface area of Broca's area have been obtained from magnetic resonance imaging (MRI) scans, which were acquired from a 1.5 Tesla General Electric (Milwaukee, WI) system. A proprietary quadrature head coil and a 3D spoiled gradient echo sequence were used to obtain a series of 124 coronal T_1-weighted images. These images comprised tissue slices of thickness 1.6mm with the following parameters: TE, 9ms; TR, 34 ms; flip angle, 30°; with a field of view of 20cm which contained a 256 x 256 pixel matrix. These scans were then processed using a variety of programs including BrainVoyager, MRICro, ImageJ and EasyMeasure as described in *Chapter 2*. BrainVoyager was used to obtain surface models that are accurate geometrical reconstructions of the cortical surface of the brain (Kriegeskorte & Goebel, 2001). MRICro and ImageJ were subsequently used to rotate and demarcate the region of interest, in this case Broca's area, from the cortical surface reconstructions. Finally, EasyMeasure was used to obtain the estimates of surface area and volume of the region of interest.

The outcome variables considered were defined as a function of the volume and surface area estimates of Broca's area, with the estimate from each hemisphere recorded separately. It was decided to also take into account the total brain volume for the volume estimates of Broca's area so that differences in size between participants' brains were corrected for. This was achieved by dividing the volume of Broca's area (BAV) by the total brain volume (TBV):

$$\text{RelBAV} = (\text{BAV}/\text{TBV}) \times 100 \qquad (1.3.3)$$

where RelBAV represents the volume of Broca's area relative to the total brain volume. It is multiplied by 100 so that the relative volume estimates of Broca's area are representative of the percentage of the total brain volume that Broca's area encapsulates. From this point onwards the volume of Broca's area is referred to as Broca's volume.

For surface area, it was decided to correct for subjects having different size of Broca's area since the interest was in the tortuosity of the structure. Let us assume two participants have identical shape of Broca's area, but one has a larger volume, then the participant with the larger volume will have a larger surface area, but it is the convolution of the area that is of primary interest in this study. Therefore, the surface area estimates of Broca's area (BASA) were adjusted to take into account Broca's volume by dividing the surface area estimates for the left and right hemispheres by their respective left and right Broca's volume estimates. Since surface area and volume are in different units (cm^2 and cm^3, respectively), the volume estimates are adjusted by a power of 2/3 such that:

$$\text{RelBASA} = \text{BASA}/(\text{BAV})^{\frac{2}{3}} \qquad (1.3.4)$$

where the surface area of Broca's area relative to Broca's volume is represented by RelBASA. This variable is regarded as a dimensionless shape factor. From this point onwards the surface area of Broca's area is referred to as Broca's surface area.

The explanatory variables considered for the analysis of the MC dataset (see Section 1.3.2) were also considered for this investigation. Additionally, the WAIS vocabulary, arithmetic and block design (visuospatial) scores were also considered as explanatory variables in the analysis of the Broca's area subset. Hence, the full list of explanatory variables for this study is:

- *Gender* (0 = male, 1 = female)
- *Age* (in years)
- *Musician* (0 = non-musician, 1 = musician)
- *WAIS vocabulary scores*
- *WAIS arithmetic scores*
- *WAIS block design (visuospatial) scores*

The descriptive statistics can be seen in Tables 1.8 and 1.9. Although *p*-values for the parametric hypothesis tests (i.e. t-tests) are given in Table 1.8, we performed Mann-Whitney U tests (i.e. non-parametric tests) to compare the results. The Mann-Whitney U tests gave *p*-values which were close to those that we obtained from the-tests and the interpretations were identical.

Table 1.8 suggests that there are little discernible differences between either musicians and controls or males and females in either the RelBAV or RelBASA variables in either hemisphere. However, we can see in Table 1.8 that there is an indication that musicians have smaller RelBASA estimates in the right hemisphere than controls (p=0.03). The only way to test whether these differences are true is to conduct the statistical analyses. The results from these analyses give adjusted results with other factors taken into consideration and so give a better indication as to whether associations exist and, if so, whether they are statistically significant or not.

Table 1.9 shows that the mean age of individuals in this dataset was 38 years, with mean scores of 12, 11.5 and 12.5 in the vocabulary, arithmetic and visuospatial tests, respectively. The Pearson correlation coefficients were between -0.3 and 0.3 suggesting that there was, at best, weak positive or negative correlation between the continuous variables (the three cognitive ability scores and age) and the measurement estimates of Broca's area. Indeed, the two-sided t-tests suggested that the only correlation that was significantly different to zero was that between vocabulary scores and RelBASA in the right hemisphere (p=0.03, Table 1.9). This implied that participants who obtained higher vocabulary scores had a smaller Broca's surface area, on average, than participants who performed lower scores on the vocabulary test. However, more information about, and evidence of, associations will only be obtained through analysis of the data. The results for the statistical analyses of this dataset can be viewed in *Chapter 5*.

Variable	Group	n	RelBAV (cm^3)						RelBASA					
			Left Hemisphere			Right Hemisphere			Left Hemisphere			Right Hemisphere		
			Mean (sd)	p-value	95% CI	Mean (sd)	p-value	95% CI	Mean (sd)	p-value	95% CI	Mean (sd)	p-value	95% CI
Gender	Male	20	1.09 (0.3)	0.3	(-0.08, 0.24)	1.28 (0.3)	0.05	(-0.00, 0.32)	4.0 (0.6)	0.05	(-0.002, 0.77)	4.1 (0.7)	0.2	(-0.18, 0.72)
	Female	19	1.01 (0.2)			1.11 (0.2)			3.6 (0.6)			3.8 (0.7)		
Music	Musician	19	1.06 (0.2)	0.8	(-0.19, 0.14)	1.23 (0.3)	0.5	(-0.23, 0.11)	3.6 (0.7)	0.06	(-0.02, 0.77)	3.7 (0.7)	0.03	(0.05, 0.90)
	Control	20	1.04 (0.3)			1.17 (0.2)			4.0 (0.5)			4.2 (0.7)		
Overall		39	1.05 (0.2)			1.20 (0.3)			3.8 (0.6)			4.0 (0.7)		

Table 1.8: *Population mean, standard deviation, p-values of t-tests and unadjusted 95% CIs of the difference between the means for the* RelBAV *and* RelBASA *estimates in the left and right hemispheres with a breakdown by each discrete variable grouping in the Broca's area subset.*

Variable	Mean (sd)	RelBAV (cm^3)				RelBASA			
		Left Hemisphere		Right Hemisphere		Left Hemisphere		Right Hemisphere	
		Correlation	p-value	Correlation	p-value	Correlation	p-value	Correlation	p-value
Age (in years)	38.4 (10.3)	-0.02	0.9	0.1	0.5	0.3	0.07	0.3	0.1
Vocabulary Test Scores	12.4 (2.9)	0.1	0.5	0.1	0.4	-0.3	0.07	-0.3	0.03
Arithmetic Test Scores	11.5 (2.6)	-0.006	0.9	-0.06	0.7	-0.3	0.08	-0.3	0.1
Visuospatial Test Scores	12.5 (2.6)	-0.3	0.1	-0.09	0.6	-0.1	0.4	-0.1	0.5

Table 1.9: *Population mean, standard deviation and Pearson correlation coefficients, with p-values, related to left and right* RelBAV *and* RelBASA *estimates for the continuous variables in the Broca's area subset.*

1.4 Structure of the thesis

This section expands on the thesis structure as seen in Table 1.1. The various different methodologies that were applied in the thesis are introduced in *Chapter 2*. There are two main areas of methodology:

- Statistical methodology
- Stereological methodology

In the statistical methodology section, a number of different statistical models are defined. Univariate linear regression models are first described (i.e. linear models with one outcome). An extension to this type of model, involving the inclusion of multiple outcome variables in a multivariate linear model, is then discussed. These models can be made even more complex by including clustering variables associated with the outcome variables. These clustering variables are classed as random effects and the methodology based around models which include them can be seen in the sections on univariate and multivariate linear mixed models. The stereological methodology section includes a detailed explanation of how to estimate the volume and surface area of 3-dimensional objects. These methods were used to estimate the volume and surface area of Broca's area, which were incorporated into the analyses within *Chapter 5*. The methods of calculating the coefficient of error of the volume and surface area estimators are also provided along with worked examples. The coefficients of error for the estimators for Broca's area were applied to intra-rater studies which can also be seen in *Chapter 5*. Finally, there is a description of how the volume and surface area estimation methods were applied to Broca's area using different software packages and various screenshots are shown.

The investigation into the association between cognitive ability and handedness in children is reported in *Chapter 3*. Cognitive ability was defined as reading and mathematics scores and the explanatory variables included gender, writing hand and UK region. With the cognitive scores as the outcomes, a multivariate linear mixed model was fitted to the dataset. The results from this model were then compared to two independent univariate models.

Chapter 4 moves on to look into effects that musical ability had on cognitive ability in adults. Cognitive ability was set as the outcome comprising of three types of test scores; vocabulary, arithmetic and visuospatial. Similarly to the analyses in *Chapter 3*, gender was included as an explanatory variable. However, age was also included as an explanatory variable since the dataset analysed in this chapter included participants of varying ages, as well as musical ability. A multivariate linear model was fitted to the dataset and then compared to three separate univariate linear models.

The analyses in *Chapter 5* focussed on the possible links between the volume and surface area estimates of Broca's area and cognitive ability in adults. Also considered were the links between Broca's area and other factors such as age and musical ability. This was an extension from the investigation in *Chapter 4* albeit with a subset of the data including volume and surface area estimates of Broca's area. The volume estimates of Broca's area relative to the total brain volume in the left and right hemispheres were considered as the bivariate outcome in the multivariate model. As for previous analyses, the results were compared to the univariate models. The surface area estimates of Broca's area relative to the volume estimates of Broca's area for the right and left hemispheres were then set as the outcomes in a multivariate model fitted to the data, and compared to univariate models. The final section in this chapter aimed to calculate the within-observer variability for the volume and the surface area estimates.

In *Chapter 6*, the types of missing data, their effects on the results and ways of dealing with them were discussed. The dataset used in *Chapter 3* has quite a substantial amount of missing data (roughly 36%) along with approximately 14% of the participants in the musician-control dataset used in *Chapter 4*. Two techniques which adjust for the missing data, namely multiple imputation and inverse probability weighting, were considered. Inverse probability weighting was used to adjust for the missing data, and the multivariate linear mixed and linear models were re-fitted to the datasets. The results are given and the limitations of the method are discussed.

Chapter 7 brings the conclusions from the analyses in each chapter together. Possible issues that arose are discussed and further work is suggested.

CHAPTER 2

Methodology

To address the research questions of this thesis, a number of different statistical and estimation methodologies will be employed. This methodology can be found in a number of textbooks. However, there is not a definitive book which explains each of the different types of statistical models applied in this thesis, using consistent notation. This section will firstly introduce the statistical methodologies to be used with consistent notation, followed in turn by the stereological estimation methods used to estimate the volume and surface area of Broca's area. For the statistical methodologies we shall consider the univariate and the multivariate cases separately.

2.1 Statistical methodology

In this section, we introduce the statistical methodology behind linear regression models and linear mixed models for both the univariate and multivariate cases. Firstly, we describe the simplest form of the linear regression model and gradually explain the methodology behind more complex statistical modelling up to the multivariate linear mixed-effects model.

2.1.1 Linear regression model (the univariate case)

Univariate analyses are those that have only one outcome (dependent) variable. However, they can still include any number of explanatory (independent) variables. The first model type to be introduced will be the linear regression model.

Regression analyses are primarily used to model associations between variables (Ugarte *et al.*, 2008). At least one variable is defined as the response variable and other variables are called the explanatory variables. A regression analysis is then used to model the relationship between the explanatory variables and the response variable(s). An example of this is the relationship between a child's IQ and the age of the child's mother (maternal age). In this case the response variable would be IQ and the explanatory variable, maternal age. The simplest form of regression model is the linear regression model. This model is called 'linear' because the relationships between the response variable(s) and the parameters of the explanatory variables are assumed to be linear. Higher order relationships can however be considered. For example as en extension to the example used previously, we could include maternal age^2 as another explanatory variable. The simplest expression of the linear regression model comprises one outcome variable, y, an intercept, a, an explanatory variable, x, and an explanatory coefficient, c (Crawley, 2005; Crawley, 2007). This takes the form:

$$y = a + cx \tag{2.1.1}$$

The gradient of the slope of the linear regression line is represented by c. In practice the relationship between the dependent and independent variables will not be perfectly linear, and therefore, some deviation will exist between the outcomes predicted from the model and the observed outcomes. This deviation is then incorporated into the model as an error term, usually denoted by ε. The linear regression model, with n subjects in the dataset, can then be re-written as:

$$y_i = a + cx_i + \varepsilon_i \tag{2.1.2}$$

where y_i is the observed outcome for the i^{th} subject, a and c are the intercept and explanatory coefficient, respectively (as in Equation (2.1.1)), x_i is the value of the explanatory variable for the i^{th} subject, ε_i is the error term for the i^{th} subject such that $\varepsilon_i \sim N(0, \sigma^2)$, and $1 \leq i \leq n$. This can also be written in vector form:

$$\underline{Y} = a + c\underline{X} + \underline{\varepsilon} \tag{2.1.3}$$

where $\underline{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, $\underline{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, $\underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$ such that $\underline{\varepsilon} \sim N_n(0, \sigma^2 \underline{I})$, which is an n-dimensional multivariate normal distribution, and a and c are the intercept and explanatory coefficient, respectively.

Following the earlier example, let it be assumed that there are 100 subjects and the objective is to predict child IQ (the outcome variable) using maternal age as the predictor (explanatory) variable. A simple linear regression model could be written in the form of that in Equation (2.1.2) with y_i and x_i representing the IQ and maternal age for the i^{th} child, respectively. This could also be written in the form of Equation (2.1.3) with $n = 100$.

There are a number of assumptions which must be made for the model to give reliable information about the relationship between the dependent and independent variables. These assumptions are:

(1) There are no errors in the measurements of the independent and dependent variables in the model.

(2) The relationships between the independent and dependent variables in the model are linear.

(3) The error terms are independent and identically distributed.

(4) The error terms follow a Normal distribution with mean 0.

(5) The error terms are homoscedastic (have a constant variance) (Seber, 1977; Crawley, 2007).

If the outcome variable is normally distributed then this is an indicator that the error terms are more likely to be normally distributed. If the outcome variable is not normally distributed then this suggests that the error terms may not be normally distributed. This is because it is easier for a linear model to fit normally distributed data as opposed to non-normal data. These assumptions can be assessed further by checking the distribution of the residuals of the model after analysis. The residual, or error term, for the i^{th} subject, ε_i, is calculated as follows:

$$\varepsilon_i = y_i - \hat{y}_i \tag{2.1.4}$$

where y_i and $\hat{y}_i$ are the observed and expected (as predicted from the model) outcome, respectively, for the i^{th} subject. If the assumptions made above hold, then the residuals should be randomly scattered about the mean (zero). If this assumption does not hold, then there will be a noticeable pattern in the plot of the residuals (i.e. they are not randomly scattered). Figure 2.1 shows an example, using synthetic data for illustration purposes only, of both a standardized residual plot which meets assumptions (3)-(5) and a standardized residual plot for which assumptions (3)-(5) do not hold. These standardized residual plots come from an example model with child IQ as the outcome (these are the fitted values from the model itself). It can quite clearly be seen in the right hand plot of Figure 2.1 that these standardized residuals are mostly above zero (i.e. do not have mean zero) and the deviation appears to decrease with higher IQ's (i.e. evidence of funnelling) suggesting that the residuals are not homoscedastic. This is in stark contrast to the plot on the left hand side of Figure 2.1 where the standardized residuals appear to have mean 0 and constant variance across the range of all child IQ's.

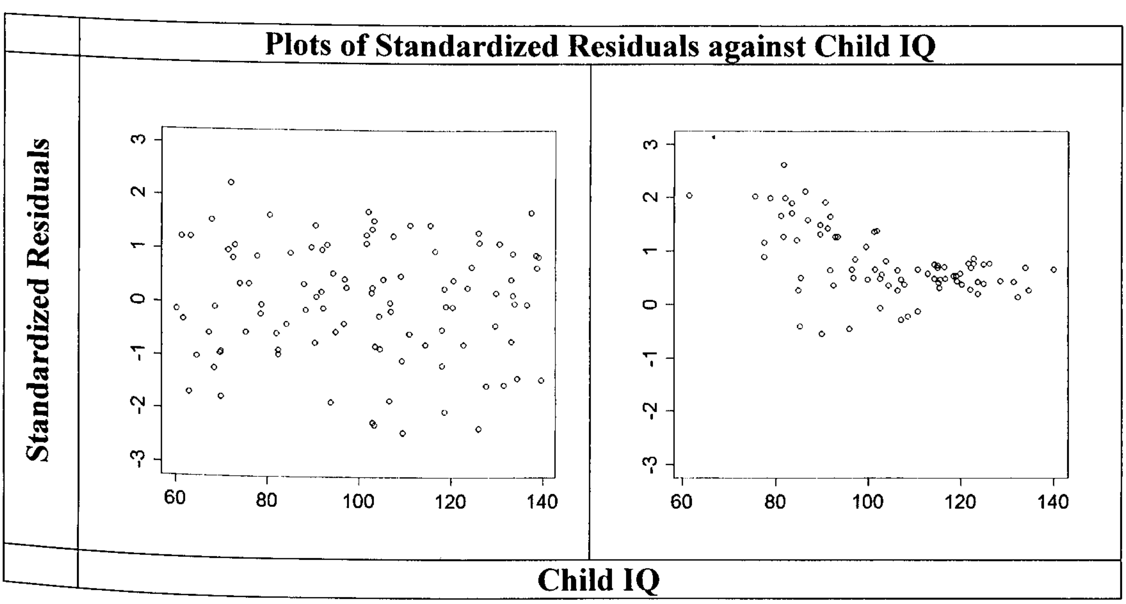


Figure 2.1: Example plots of fitted child IQ values against standardized residuals such that assumptions (3)-(5) are met for the plot on the left hand side whereas this is not the case for the standardized residuals in the right hand plot.

A simple linear regression model is useful if the dataset only has two variables of interest with no further variables which could affect the relationship (which is assumed to be linear) between them. However, this is not normally the case and there will be other variables which could also have a statistically significant effect on the outcome variable. In this case, there could be confounding relationships between these explanatory variables and the outcome. For example, two explanatory variables when investigated separately using a simple linear regression model with the same outcome variable may be significant, suggesting that both have an effect on the outcome. However, if the two explanatory variables are correlated then they may be explaining the same variability in the outcome. This can only be fully revealed by including all biologically and statistically significant variables in one single model.

From the example mentioned earlier, let it be assumed that is now desired to include both gestational age and maternal IQ as further explanatory variables with child IQ as the outcome. The simple linear regression modelling approach could be repeated for each individual explanatory variable, but this would not help to explain which of the three explanatory variables is most strongly correlated with child IQ, or their combined effect. To take these extra variables into account, the model in Equation (2.1.2) must be expanded to:

$$y_i = a + c_1 x_{i1} + c_2 x_{i2} + c_3 x_{i3} + \varepsilon_i$$

where y_i represents the outcome (child IQ) of the i^{th} subject $(1 \leq i \leq 100)$, a is the intercept, the explanatory variables maternal age, gestational age and maternal IQ for the i^{th} subject are represented by x_{i1}, x_{i2}, x_{i3}, respectively with c_1, c_2, c_3 being the respective regression coefficients and with ε_i being the error terms for the i^{th} child. Therefore, the effect of a change of 1 unit in maternal age on child IQ after gestational age and maternal IQ are taken into account is explained by c_1, c_2 explains the effect of gestational age on child IQ after maternal age and maternal IQ are accounted for, and c_3 explains the effect of maternal IQ on child IQ when the other two explanatory variables are accounted for.

The model in Equation (2.1.2) can, in fact, be expanded to include $(p-1)$ explanatory variables (p if we consider the intercept term to be an explanatory variable) where $p \in \{1,2,3,\dots\}$. The expanded form of the linear regression model, also called the multiple regression model, can be written as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{i(p-1)} + \varepsilon_i \tag{2.1.5}$$

where y_i is the outcome for the i^{th} subject, β_0 is the intercept, $\beta_1, \dots, \beta_{p-1}$ are the regression coefficients for the $(p-1)$ explanatory variables, $x_{i1}, \dots, x_{i(p-1)}$ are the values of the $(p-1)$ explanatory variables for the i^{th} subject and ε_i is the error term for the i^{th} subject such that $\varepsilon_i \sim N(0, \sigma^2)$ with $1 \le i \le n$. In matrix form this can be expressed as:

$$\underline{Y} = \underline{X\beta} + \underline{\varepsilon} \tag{2.1.6}$$

where $\underline{Y}$ is the $(n \times 1)$ outcome vector such that $\underline{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, $\underline{X}$ is an $(n \times p)$ design matrix such that $\underline{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1(p-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{n(p-1)} \end{bmatrix}$, $\underline{\beta}$ is a $(p \times 1)$ vector of coefficients such that $\underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$, $\underline{\varepsilon}$ is the $(n \times 1)$ error vector such that $\underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$ and $\underline{\varepsilon} \sim N_n(0, \sigma^2 \underline{I})$ as in Equation (2.1.3) (see for example, Ugarte *et al.*, 2008). Assumptions (1) - (5) mentioned above for the simple linear regression model (Equation (2.1.2)), are still applicable here. With this model we can now explain the variability in the dataset which can be attributed to the relationship between the explanatory variables and the outcome variable. The error terms give an indication of the amount of variability in the data that cannot be explained by the explanatory variables. Obviously, the smaller the error term is, the better the model fits the data.

The models described up to this point measure the influence that the explanatory variables have on the average of the outcome and there has only been one outcome measurement for each subject. This means that the variability in the outcome variable that can be explained by the fitted models is associated with the explanatory variables and the error term left for each subject contains all the unexplained variability associated with individual subjects. What has not been modelled here is the between-subject variability, that is, the variation across subjects and to differentiate this from the within-subject variability (all other unexplained variation associated with each individual subject). However, it is quite plausible to consider an example where for each subject there are multiple measurements for the outcome. One example of this is longitudinal data where a number of measurements of the same variable are taken, for each subject, over a period of time. This type of data allows the between-subject variability to be modelled with an extension to the linear regression model. This more complex model is called a linear mixed model.

2.1.2 Linear mixed model (the univariate case)

Taking the example used previously with child IQ as the outcome variable and maternal age, gestational age and maternal IQ as explanatory variables, let it now be assumed that the data are longitudinal. That is, child IQ was measured at two time points; at ages 6 and 12. These data can be analysed over both time points in one model using a linear mixed modelling approach. By analysing the data with a linear mixed model, the between-subject variability can be differentiated from the within-subject variability. Since it is expected that child IQ is correlated across the two time-points then it would be useful to apply a linear mixed model to the dataset such that the unexplained variability for each subject (the within-subject variability) is reduced.

Eisenhart, (1947) introduced a concept of two types of variable: fixed effects and random effects. Fixed effects, as with the explanatory variables in standard linear regression models, only have an influence on the mean of the outcome variable. Random effects, however, have an influence on the variance of the outcome variable, but not on the mean (Crawley, 2007). The linear mixed model incorporates both these type of variable (fixed and random effects). This model is used primarily when there is correlation within the data. That is, when there is some underlying association between the data as a whole, or between groups of data. This correlation can be viewed in the structure of the variance-covariance of the random effects and random errors (Davis, 2003). Examples of correlated data are clustered data (association within clusters) and longitudinal data (association within time points). The linear mixed model allows identification of the between-subject variability (from the random effects), separately to the within-subject variability (the random errors). The simplest form of linear mixed model is called the random intercept model (Zuur *et al.*, 2009), where the model contains one variance term representing the between-subject variability.

In the example with child IQ data measured at two time points, the linear mixed model can be written as:

$$y_{i1} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + b_i + \varepsilon_{i1}$$

$$y_{i2} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + b_i + \varepsilon_{i2}$$

where y_{i1}, y_{i2} represent child IQ at ages 6 and 12 (time points 1 and 2), respectively, for the i^{th} child $(1 \leq i \leq 100)$, β_0 is the intercept and x_{i1}, x_{i2}, x_{i3} are the explanatory variables (also called fixed effects) maternal age, gestational age and maternal IQ for the i^{th} subject, respectively, with $\beta_1, \beta_2, \beta_3$ being the respective fixed effect coefficients. The additional term b_i is called the random effect term and it takes into account the variability which inherently belongs to the individual subject, i, and is common across both time points 1 and 2 (ages 6 and 12). In the example, the quantity b_i is constant for each individual since it does not depend on the time point and it is called the random intercept. The final error terms, now called random error terms, $\varepsilon_{i1}, \varepsilon_{i2}$ include all unexplained variation within child IQ for the i^{th} child at time point 1 and 2, respectively.

This model can be generalised and expanded to include g time points $(g \geq 1)$ and $(p-1)$ fixed effects where $p \in \{1,2,3,\dots\}$. Hence, the model takes the form:

$$y_{ij} = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{i(p-1)} + b_i + \varepsilon_{ij} \qquad (2.1.7)$$

where y_{ij} is the outcome at the j^{th} time point of the i^{th} subject $(1 \leq j \leq g)$, β_0 is the intercept, $\beta_1, \dots, \beta_{p-1}$ are the coefficients for the $(p-1)$ fixed effects, $x_{i1}, \dots, x_{i(p-1)}$ are the values of the $(p-1)$ fixed effects for the i^{th} subject, b_i is the random intercept for the i^{th} subject such that $b_i \sim N(0, \sigma_b^2)$ and ε_{ij} is the random error term for the j^{th} time point of the i^{th} subject such that $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. Equation (2.1.7) can be written in matrix form as:

$$\underline{\tilde{Y}} = \underline{\tilde{X}}\underline{\beta} + \underline{z}\,\underline{b} + \underline{\tilde{\varepsilon}} \qquad (2.1.8)$$

where $\underline{\tilde{Y}} = \begin{bmatrix} \underline{\tilde{Y}_1} \\ \vdots \\ \underline{\tilde{Y}_n} \end{bmatrix}$ is the $(gn \times 1)$ outcome vector such that $\underline{\tilde{Y}_i} = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{ig} \end{bmatrix}$, $\underline{\tilde{X}} = \begin{bmatrix} \underline{\tilde{X}_1} \\ \vdots \\ \underline{\tilde{X}_n} \end{bmatrix}$ is a $(gn \times p)$ design matrix such that $\underline{\tilde{X}_i} = \begin{bmatrix} 1 & x_{i1} & \cdots & x_{i(p-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & \cdots & x_{i(p-1)} \end{bmatrix}$ is a $(g \times p)$ matrix, $\underline{\beta}$ is the $(p \times 1)$ vector of coefficients (as defined for Equation (2.1.6)), $\underline{z}$ is a $(gn \times n)$ matrix such that $\underline{z} = \begin{bmatrix} \underline{1}^g & & \underline{0} \\ & \ddots & \\ \underline{0} & & \underline{1}^g \end{bmatrix}$ where $\underline{1}^g = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ is a $(g \times 1)$ vector, $\underline{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$ is the $(n \times 1)$ random effect vector such that $\underline{b} \sim N_n(\underline{0}, \underline{G})$ and $\underline{G} = \begin{bmatrix} \sigma_b^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_b^2 \end{bmatrix}$ is an $(n \times n)$ diagonal matrix, $\underline{\tilde{\varepsilon}}$ is the $(gn \times 1)$ random error vector such that $\underline{\tilde{\varepsilon}} = \begin{bmatrix} \underline{\tilde{\varepsilon}_1} \\ \vdots \\ \underline{\tilde{\varepsilon}_n} \end{bmatrix}$ where $\underline{\tilde{\varepsilon}_i} = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{ig} \end{bmatrix}$ and $\underline{\tilde{\varepsilon}} \sim N_{gn}(\underline{0}, \underline{R})$ where $\underline{R} = \begin{bmatrix} \sigma_\varepsilon^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_\varepsilon^2 \end{bmatrix}$ is a diagonal $(gn \times gn)$ matrix. This model assumes that the random effects, b_i, are independent and identically distributed and the same applies to the random errors, ε_{ij} (hence, the zeroes in the off-diagonal elements of the variance-covariance matrices $\underline{G}$ and $\underline{R}$). A further assumption is that the random effects and random errors are independent from each other. This model can be further expanded by including further random effects but for the work involved in this thesis, this is not required (for a description of the general, expanded linear mixed model see Laid & Ware, 1982). Similarly to the linear regression model, the linear mixed model has the 5 assumptions, (1) to (5), named earlier in Section 2.1.1. Furthermore, the standardized residual plots can also be plotted for these models to assess whether or not these assumptions hold.

Previously it has been considered that there is only one outcome. However, it is quite plausible that there could be multiple outcomes. There are two possible ways to analyse these multivariate data. Multivariate data consisting of d outcomes and $p-1$ explanatory variables could be analysed by simply fitting d univariate models. However, univariate analysis of multivariate data only gives information relating to associations between the explanatory variables and each outcome individually. We want to be able to obtain conclusions about the overall relationships between all of the explanatory and outcome variables simultaneously. Separate univariate analyses of multivariate data are unsatisfactory for two main reasons (Krzanowski, 2003):

- Any dependencies or correlation between the outcomes are ignored thus leading to possible major inference errors from the hypothesis testing of the data.
- The use of multiple univariate analyses is equivalent to having multiple comparisons and therefore requires an adjustment of the significance level, α (e.g. using the Bonferroni correction method).

In either of the situations mentioned previously, the possibility of obtaining false positives or false negatives from hypothesis tests based on these univariate analyses is greater than for a multivariate analysis. A multivariate model gives information about the associations between outcomes and explanatory variables individually (in the models based on the marginal distributions for each outcome) as well as those links between the explanatory variables and a linear combination of the outcomes. The linear combination of outcomes can be written as:

$$\underline{\tilde{Y}}^{(1)} + \underline{\tilde{Y}}^{(2)} + \cdots + \underline{\tilde{Y}}^{(d)} \qquad (2.1.9)$$

and this is the multivariate outcome (Hair, Jr., *et al.*, 2010). Hypothesis tests for the multivariate outcome in Equation (2.1.9) do not need to be adjusted using the Bonferroni correction method and dependencies between outcomes are already accounted for in the multivariate model itself.

2.1.3 *Linear regression model (the multivariate case)*

Let it be assumed that there are now 2 sub-scores of the child IQ score of interest, verbal IQ and non-verbal IQ and that the dataset consists of only 1 measurement for each outcome of each individual (as in Section 2.1.1). The effects of gestational age, maternal age and maternal IQ on the two sub-score IQs of the child are now of interest. Firstly, the correlation between the two outcomes is 0.6, suggesting that there is reasonably strong positive correlation between verbal and non-verbal child IQ. That is, that a child with a higher verbal IQ than another will, more than likely have a higher non-verbal IQ score, and vice versa. A multivariate linear regression model is then chosen as the analysis method, such that this correlation can be accounted for in the model. This model then would take the form:

$$y_i^{(1)} = \beta_0^{(1)} + \beta_1^{(1)}x_{i1} + \beta_2^{(1)}x_{i2} + \beta_3^{(1)}x_{i3} + \varepsilon_i^{(1)}$$

$$y_i^{(2)} = \beta_0^{(2)} + \beta_1^{(2)}x_{i1} + \beta_2^{(2)}x_{i2} + \beta_3^{(2)}x_{i3} + \varepsilon_i^{(2)}$$

where $y_i^{(1)}, y_i^{(2)}$ represents verbal IQ and non-verbal IQ, respectively, for the i^{th} child $(1 \leq i \leq 100)$, $\beta_0^{(1)}, \beta_0^{(2)}$ are the intercepts for the verbal and non-verbal IQ scores, respectively, the explanatory variables, maternal age, gestational age and maternal IQ for the i^{th} subject are represented by x_{i1}, x_{i2}, x_{i3}, respectively with $\beta_1^{(1)}, \beta_2^{(1)}, \beta_3^{(1)}$ being the respective coefficients for the verbal IQ scores, and $\beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)}$ representing the coefficients for the non-verbal IQ scores. The error terms $\varepsilon_i^{(1)}, \varepsilon_i^{(2)}$ are those for the verbal and non-verbal IQ scores, respectively for the i^{th} child, and the correlation between the two IQ scores can be found by writing the model in another way, which is to be discussed later in this subsection.

Generalising, the univariate linear regression model in Equation (2.1.5) can be expanded to include d outcome variables $(d \geq 1)$. Then the multivariate linear regression model can be written as:

$$y_i^{(k)} = \beta_0^{(k)} + \beta_1^{(k)}x_{i1} + \cdots + \beta_{p-1}^{(k)}x_{i(p-1)} + \varepsilon_i^{(k)} \qquad (2.1.10)$$

where $y_i^{(k)}$ is the k^{th} outcome for the i^{th} subject $(1 \le k \le d)$, $\beta_0^{(k)}$ is the intercept for the k^{th} outcome, $\beta_1^{(k)}, \dots, \beta_{p-1}^{(k)}$ are the coefficients of the $(p-1)$ explanatory variables for the k^{th} outcome, $x_{i1}, \dots, x_{i(p-1)}$ are the values of the $(p-1)$ explanatory variables for the i^{th} subject, $\varepsilon_i^{(k)}$ is the error term for the k^{th} outcome of the i^{th} subject such that $\varepsilon_i^{(k)} \sim N(0, \sigma_k^2)$, $1 \le i \le n$. In matrix form this can be re-written as:

$$\underline{Y}^* = \underline{X}^* \underline{\beta}^* + \underline{\varepsilon}^* \qquad (2.1.11)$$

where $\underline{Y}^*$ is the $(dn \times 1)$ outcome vector such that $\underline{Y}^* = \begin{bmatrix} \underline{Y}^{(1)} \\ \vdots \\ \underline{Y}^{(d)} \end{bmatrix}$ and $\underline{Y}^{(k)} = \begin{bmatrix} y_1^{(k)} \\ \vdots \\ y_n^{(k)} \end{bmatrix}$, $\underline{X}^*$ is a $(dn \times dp)$ design matrix such that $\underline{X}^* = \begin{bmatrix} \underline{X} & & \underline{0} \\ & \ddots & \\ \underline{0} & & \underline{X} \end{bmatrix}$, $\underline{\beta}^*$ is a $(dp \times 1)$ vector of coefficients such that $\underline{\beta}^* = \begin{bmatrix} \underline{\beta}^{(1)} \\ \vdots \\ \underline{\beta}^{(d)} \end{bmatrix}$ and $\underline{\beta}^{(k)} = \begin{bmatrix} \beta_0^{(k)} \\ \beta_1^{(k)} \\ \vdots \\ \beta_{p-1}^{(k)} \end{bmatrix}$, $\underline{\varepsilon}^*$ is the $(dn \times 1)$ error vector such that $\underline{\varepsilon}^* = \begin{bmatrix} \underline{\varepsilon}^{(1)} \\ \vdots \\ \underline{\varepsilon}^{(d)} \end{bmatrix}$ and $\underline{\varepsilon}^{(k)} = \begin{bmatrix} \varepsilon_1^{(k)} \\ \vdots \\ \varepsilon_n^{(k)} \end{bmatrix}$ such that $\underline{\varepsilon}^* \sim N_{nd}(\underline{0}, \underline{\Sigma})$ where $\underline{\Sigma} = \begin{bmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} & \cdots & \underline{\Sigma}_{1d} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} & \cdots & \underline{\Sigma}_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \underline{\Sigma}_{d1} & \underline{\Sigma}_{d2} & \cdots & \underline{\Sigma}_{dd} \end{bmatrix}$ is a $(dn \times dn)$ matrix with $(n \times n)$ matrices $\underline{\Sigma}_{kk} = \begin{bmatrix} \sigma_k^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_k^2 \end{bmatrix}$ and $\underline{\Sigma}_{k_1k_2} = \begin{bmatrix} \sigma_{k_1k_2} & & 0 \\ & \ddots & \\ 0 & & \sigma_{k_1k_2} \end{bmatrix}$, $(1 \le k_1, k_2 \le d;\ k_1 \ne k_2)$. It is useful to note that all off-diagonal elements in all the $\underline{\Sigma}_{kk}$ and $\underline{\Sigma}_{k_1k_2}$ matrices are zero. This is because there is no correlation between the outcome measurements from two different individual subjects (i.e. they are independent). The only correlation is between outcomes within the same subject. The structure of the error variance-covariance matrix, $\underline{\Sigma}$, in the multivariate regression model (Equations (2.1.10) & (2.1.11)) allows for the correlation between outcomes to be measured $(\sigma_{k_1k_2})$. For example if $\sigma_{k_1k_2} = 0$, then this means that the k_1^{th} and k_2^{th} outcome variables are independent and the model obtained from this method would be identical to fitting two separate univariate linear regression models to the data. In general, the correlation between the k_1th and k_2th outcomes, $\sigma_{k_1 k_2}$, can be different for each different pairing of outcomes 1 to d (called a general structure). There are also a number of different possible correlation structures in between (including autoregressive (AR($\tilde{p}$)), moving average (MA($\tilde{q}$)), and mixed autoregressive-moving average ARMA($\tilde{p}$,$\tilde{q}$) structures where $\tilde{p}$ is the order of the AR model and $\tilde{q}$ is the number of noise terms in the moving average model (Pinheiro & Bates, 2000). In the case with two outcome variables, only the independent and general structures can be considered. Once again, as with the univariate linear regression models, it is assumed that the outcome variables are normally distributed and that the error terms for the kth outcome, $\varepsilon_i^{(k)}$, are independent and identically normally distributed with mean 0 and variance σ_k^2. This means that we can again analyse the standardized residual plots for normality to check the goodness of fit of the model to the data.

The multivariate version of the linear regression model can be expanded similar to the univariate case of the linear mixed model in Section 2.1.2. This is relevant when there are multiple measurements taken for each outcome variable (e.g. clustered or longitudinal data). Now, however there are also multiple outcomes to consider, which add complexity to the equations. This model is called the multivariate linear mixed model.

2.1.4 Linear mixed model (the multivariate case)

From the running example, the two outcomes from Section 2.1.3 (verbal IQ and non-verbal IQ) are considered and the longitudinal component described in Section 2.1.2 is added, such that the two outcomes are measured at two time points (age 6 and age 12). Advantages of this model are that the correlation can be taken into account between outcomes, whilst at the same time, the between-subject variability can be differentiated from the within-subject variability. Let it be assumed that the explanatory variables (or fixed effects) remain the same (maternal age, gestational age and maternal IQ). The model used to analyse this data would then take the following form:

$$y_{i1}^{(1)} = \beta_0^{(1)} + \beta_1^{(1)}x_{i1} + \beta_2^{(1)}x_{i2} + \beta_3^{(1)}x_{i3} + b_i^{(1)} + \varepsilon_{i1}^{(1)}$$

$$y_{i2}^{(1)} = \beta_0^{(1)} + \beta_1^{(1)}x_{i1} + \beta_2^{(1)}x_{i2} + \beta_3^{(1)}x_{i3} + b_i^{(1)} + \varepsilon_{i2}^{(1)}$$

$$y_{i1}^{(2)} = \beta_0^{(2)} + \beta_1^{(2)}x_{i1} + \beta_2^{(2)}x_{i2} + \beta_3^{(2)}x_{i3} + b_i^{(2)} + \varepsilon_{i1}^{(2)}$$

$$y_{i2}^{(2)} = \beta_0^{(2)} + \beta_1^{(2)}x_{i1} + \beta_2^{(2)}x_{i2} + \beta_3^{(2)}x_{i3} + b_i^{(2)} + \varepsilon_{i2}^{(2)}$$

where $y_{i1}^{(1)}, y_{i2}^{(1)}$ represent verbal IQ at ages 6 and 12 (time points 1 and 2), respectively, and $y_{i1}^{(2)}, y_{i2}^{(2)}$ indicate non-verbal IQ at ages 6 and 12 (time points 1 and 2), respectively, for the i^{th} child ($1 \leq i \leq 100$). The intercepts are $\beta_0^{(1)}, \beta_0^{(2)}$ for the verbal and non-verbal IQ scores, respectively. The fixed effects, maternal age, gestational age and maternal IQ for the i^{th} subject are represented by x_{i1}, x_{i2}, x_{i3}, respectively with $\beta_1^{(1)}, \beta_2^{(1)}, \beta_3^{(1)}$ being the respective coefficients for the verbal IQ scores, and $\beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)}$ representing the coefficients for the non-verbal IQ scores. The random intercept term for the verbal IQ and non-verbal IQ for the i^{th} subject, take the form of $b_i^{(1)}$ and $b_i^{(2)}$, respectively. The error terms, $\varepsilon_{i1}^{(1)}, \varepsilon_{i2}^{(1)}$, are those for verbal IQ of the i^{th} subject at the two time points (ages 6 and 12), respectively with $\varepsilon_{i1}^{(2)}, \varepsilon_{i2}^{(2)}$ being the error terms corresponding to the non-verbal IQ of the i^{th} subject at the same two time points. Similarly as described previously in Section 2.1.3, the correlation is not shown at this level of the model, but will be explained in the matrix version of the general form of the multivariate linear mixed model later in this section.

The multivariate linear mixed model, with random intercept only from the example above, can be extended to include any number of time points for each subject. The general form of a multivariate linear mixed model with random intercept only which has g time points ($g \geq 1$) for each subject i ($1 \leq i \leq n; n \geq 2$) and d outcome variables ($d \geq 2$) can be written as:

$$y_{ij}^{(k)} = \beta_0^{(k)} + \beta_1^{(k)} x_{i1} + \cdots + \beta_{p-1}^{(k)} x_{i(p-1)} + b_i^{(k)} + \varepsilon_{ij}^{(k)} \tag{2.1.12}$$

where $y_{ij}^{(k)}$ is the k^{th} outcome at the j^{th} time point for the i^{th} subject ($1 \leq k \leq d; 1 \leq j \leq g$), $\beta_0^{(k)}, \dots, \beta_{p-1}^{(k)}$ are the p fixed effect coefficients (including the intercept) for the k^{th} outcome, $x_{i1}, \dots, x_{i(p-1)}$ are the $(p-1)$ explanatory variables for the i^{th} subject, $b_i^{(k)}$ is the random intercept for the k^{th} outcome of the i^{th} subject, such that $b_i^{(k)} \sim N\left(0, \sigma_{b_k}^2\right)$ and $\varepsilon_{ij}^{(k)}$ is the error term for the k^{th} outcome at the j^{th} time point for the i^{th} subject such that $\varepsilon_{ij}^{(k)} \sim N\left(0, \sigma_{\varepsilon_k}^2\right)$. This model can be written in matrix form as:

$$\underline{\bar{\bar{Y}}} = \underline{\bar{\bar{X}}}\underline{\beta}^* + \underline{z}^* \underline{b}^* + \underline{\bar{\bar{\varepsilon}}} \tag{2.1.13}$$

where $\underline{\bar{\bar{Y}}} = \begin{bmatrix} \underline{\tilde{Y}}^{(1)} \\ \vdots \\ \underline{\tilde{Y}}^{(d)} \end{bmatrix}$ is the $(dgn \times 1)$ outcome vector with $\underline{\tilde{Y}}^{(k)} = \begin{bmatrix} \underline{\tilde{Y}}_1^{(k)} \\ \vdots \\ \underline{\tilde{Y}}_n^{(k)} \end{bmatrix}$ and $\underline{\tilde{Y}}_i^{(k)} = \begin{bmatrix} y_{i1}^{(k)} \\ \vdots \\ y_{ig}^{(k)} \end{bmatrix}$, $\underline{\bar{\bar{X}}}$ is the $(dgn \times dp)$ fixed effect design matrix such that $\underline{\bar{\bar{X}}} = \begin{bmatrix} \underline{\tilde{X}} & & \underline{0} \\ & \ddots & \\ \underline{0} & & \underline{\tilde{X}} \end{bmatrix}$ with $\underline{\tilde{X}}$ defined as in Equation (2.1.8), $\underline{\beta}^*$ is the $(dp \times 1)$ fixed effect coefficient vector as defined for Equation (2.1.11), $\underline{z}^* = \begin{bmatrix} \underline{z} & & \underline{0} \\ & \ddots & \\ \underline{0} & & \underline{z} \end{bmatrix}$ is a $(dgn \times dn)$ matrix, $\underline{b}^*$ is the is the $(dn \times 1)$ random effect vector such that $\underline{b}^* = \begin{bmatrix} \underline{b}^{(1)} \\ \vdots \\ \underline{b}^{(d)} \end{bmatrix}$ where $\underline{b}^{(k)} = \begin{bmatrix} b_1^{(k)} \\ \vdots \\ b_g^{(k)} \end{bmatrix}$ with $\underline{b}^* \sim N_{dn}\left(\underline{0}, \underline{G}^*\right)$ where $\underline{G}^* = \begin{bmatrix} \underline{G}_{11}^* & \underline{G}_{12}^* & \cdots & \underline{G}_{1d}^* \\ \underline{G}_{21}^* & \underline{G}_{22}^* & \cdots & \underline{G}_{2d}^* \\ \vdots & \vdots & \ddots & \vdots \\ \underline{G}_{d1}^* & \underline{G}_{d2}^* & \cdots & \underline{G}_{dd}^* \end{bmatrix}$ (a $(dn \times dn)$ matrix) with $\underline{G}_{kk}^* = \begin{bmatrix} \sigma_{b_k}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{b_k}^2 \end{bmatrix}$ and $\underline{G}_{k_1 k_2}^* = \begin{bmatrix} \sigma_{b_{k_1} b_{k_2}} & & 0 \\ & \ddots & \\ 0 & & \sigma_{b_{k_1} b_{k_2}} \end{bmatrix}$ (both

$(n \times n)$ matrices). The error terms are represented by the $(dgn \times 1)$ matrix $\bar{\underline{\varepsilon}} = \begin{bmatrix} \underline{\tilde{\varepsilon}}^{(1)} \\ \vdots \\ \underline{\tilde{\varepsilon}}^{(d)} \end{bmatrix}$ with $\underline{\tilde{\varepsilon}}^{(k)} = \begin{bmatrix} \underline{\tilde{\varepsilon}}_1^{(k)} \\ \vdots \\ \underline{\tilde{\varepsilon}}_n^{(k)} \end{bmatrix}$ and $\underline{\tilde{\varepsilon}}_i^{(k)} = \begin{bmatrix} \varepsilon_{i1}^{(k)} \\ \vdots \\ \varepsilon_{ig}^{(k)} \end{bmatrix}$ such that $\bar{\underline{\varepsilon}} \sim N_{dgn}(\underline{0}, \underline{R}^*)$ where $\underline{R}^* = \begin{bmatrix} \underline{R}_{11}^* & \underline{R}_{12}^* & \cdots & \underline{R}_{1d}^* \\ \underline{R}_{21}^* & \underline{R}_{22}^* & \cdots & \underline{R}_{2d}^* \\ \vdots & \vdots & \ddots & \vdots \\ \underline{R}_{d1}^* & \underline{R}_{d2}^* & \cdots & \underline{R}_{dd}^* \end{bmatrix}$ is a $(dgn \times dgn)$ matrix with $\underline{R}_{kk}^* = \begin{bmatrix} \sigma_{\varepsilon_k}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{\varepsilon_k}^2 \end{bmatrix}$ and $\underline{R}_{k_1 k_2}^* = \begin{bmatrix} \sigma_{\varepsilon_{k_1}\varepsilon_{k_2}} & & 0 \\ & \ddots & \\ 0 & & \sigma_{\varepsilon_{k_1}\varepsilon_{k_2}} \end{bmatrix}$ (both $(gn \times gn)$ matrices). One can notice that the only correlation considered here is the correlation between outcomes for the same subject. All other correlations are set to zero since the outcome of each individual is assumed to be independent of each other. This is also the case for the random effects where time points for different subjects are assumed to not be correlated with each other, only time points within each individual are therefore correlated and have correlation elements which can be non-zero. This model can be extended to include further random effect terms. However, for the work involved in this thesis, such generalisation is not necessary.

2.2 Stereological methods

2.2.1 Introduction

To obtain estimates of geometrical properties of an object (such as volume, surface area, number of particles inside the object, etc.) a number of possible methods exist. In the 19th century, geology was one of the main areas which was in need of fast, reliable estimates for volume of 3-dimensional objects. In this field, it was the mineral content of rocks that was the parameter of interest. An old method to estimate volume consisted of the crushing of rocks into a fine powder and then separating the minerals from each other using physical and/or chemical techniques. However, the crushing of rocks, apart from being totally destructive and inefficient, becomes difficult for larger rocks. Another method was to simply infer the volume of mineral in the rock by multiplying a physical property of the rock (weight or mass, for example) by the known estimate of proportion in a rock of its type. In 1847, A.E. Delesse, a French geologist and mineralogist, began the investigation into the use of metallography, the technique of viewing under reflected light cut and polished flat surfaces of rocks and metals, quantitatively, later to be known as Stereology. Delesse reasoned that a plane section of a rock, once smoothed and polished, could be representative of the whole rock (see e.g. Howard & Reed, (2005); Baddeley & Jensen, (2005) and references therein). Under this assumption, a mineral's percentage of the area of the plane section is then considered to be equal to the mineral's percentage of the volume in the entire rock. That is:

$$V_p = A_p \tag{2.2.1}$$

where V_p is the proportion of mineral volume in the rock, and A_p is the proportion of mineral area in a plane section of the rock. For Equation (2.2.1) to hold it is assumed that the rock is homogeneous. That is, that the distribution of minerals throughout the entire rock is the same. In the late 19th century another geologist, A. Rosiwal, further simplified Delesse's theory. By superimposing a grid of equally-spaced parallel lines onto the plane, it was shown that the proportion of the lines which lie on the mineral in question can be approximated to the proportion of the volume of the mineral in the entire rock. Therefore, assuming equality, the relationship can be expressed as:

$$V_p = L_p \tag{2.2.2}$$

where L_p is the proportion of length of the lines that are superimposed on top of the mineral. Furthermore, this sampling strategy can be extended by superimposing a grid of equi-distant points. The proportion of points lying within the region of interest (RoI) (the mineral), is then equal to the proportion of the volume of the rock that consists of the mineral. That is:

$$V_p = P_p \tag{2.2.3}$$

where P_p is the proportion of points, superimposed onto the plane, which lie on top of the RoI (or the mineral in this case) (Cruz-Orive, 1997). Examples of volume and surface area estimates obtained using stereological methods can be seen in a range of recent studies (see e.g. Reed *et al.*, 2001; Dhaliwal *et al.*, 2002; Kannekens *et al.*, 2006; Sneddon *et al.*, 2006).

2.2.2 Unbiased estimation

The name 'Stereology' was proposed by a group of scientists who were holding discussions about common problems when attempting to estimate 3-D objects from 2-D sections (Weibel, 1987). This discussion group, formed in 1961, became the International Society of Stereology. The term 'Stereology' is adapted from the Greek word 'stereos' which is translated as 'solid'. The first definition of Stereology was "the spatial interpretation of sections" (Elias, 1963). More recently, Stereology has been defined as "a body of mathematical methods relating three-dimensional parameters defining the structure to two-dimensional measurements obtainable on sections of the structure" (Weibel, 1979). Note that the aim of Stereology includes the estimation of any parameters which can describe the structure of the 3-D object (e.g. volume and surface area). This newfound interest in the subject can be explained by the advancement of technology in microscopy and the variety of uses of Stereology has increased over the years. As well as in geology, there are now applications of stereological methods in life sciences, materials science, engineering, industry and clinical medicine (for references to a range of applied studies see Howard & Reed, 2005; Baddeley & Jensen, 2005).

In stereology, a parameter (e.g. length, volume or surface area) is to be estimated. The estimator proposed has to have certain properties to allow the estimates obtained to be acceptably close to the true parameter values. One of these properties is that the estimator must be accurate. Accuracy, in this context, can be thought of as a measure of how close (or far apart) from the true value the estimates are. The difference between the true value and the estimator mean is called the bias. Therefore, an estimator is deemed biased if the difference between its expectation and the true value that it is estimating, is non-zero. For example, if the volume, V, of an object was to be estimated by applying the estimator $\hat{V}$, then the bias of the estimator would take the form:

$$\mathrm{Bias}(\hat{V}) = \mathrm{E}(\hat{V}) - V \tag{2.2.4}$$

where $\mathrm{E}(\hat{V})$ indicates the average of the estimator $\hat{V}$ and V is the true value. An estimator is said to be unbiased when its bias is zero. Therefore, if the estimator $\hat{V}$ was unbiased then this would mean that the expected mean of the estimator $\hat{V}$ is equal to V, which reads:

$$\mathrm{Bias}(\hat{V}) = \mathrm{E}(\hat{V}) - V = 0$$

$$\therefore \mathrm{E}(\hat{V}) = V \tag{2.2.5}$$

The terms precision and accuracy, although related, refer to different concepts. Precision relates to the spread of estimates about their mean. For example, the estimator, $\hat{V}$, defined earlier, can be applied to obtain a number of estimates of the volume V. These estimates will always gravitate towards the expected mean of the estimator, $E(\hat{V})$. If $\hat{V}$ has low precision, then these estimates could be spread out, quite far away from $E(\hat{V})$. However, the more precise $\hat{V}$ is, the closer the estimates will be in relation to $E(\hat{V})$. A measure of the precision of an estimator is its variance $\mathrm{Var}(\hat{V})$. The measure of the accuracy of an estimator is its mean square error (MSE) and this is calculated by summing the estimator's variance with the square of its bias:

$$\mathrm{MSE}(\hat{V}) = \mathrm{Var}(\hat{V}) + \left(\mathrm{Bias}(\hat{V})\right)^2 \tag{2.2.6}$$

This means that the variance and MSE of an estimator are equal if and only if it is unbiased (i.e. $\mathrm{Bias}(\hat{V}) = 0$). This distinction between accuracy (unbiasedness) and precision (variance) is illustrated in Figure 2.2. A precise estimator may still be biased (inaccurate), as can be seen in the bottom left corner, where the mean of the volume estimates (blue spot) is not equal to the true parameter mean (green spot). The most desirable estimator is shown in the top left hand corner, which is an accurate estimator with all estimates close to the true parameter value.

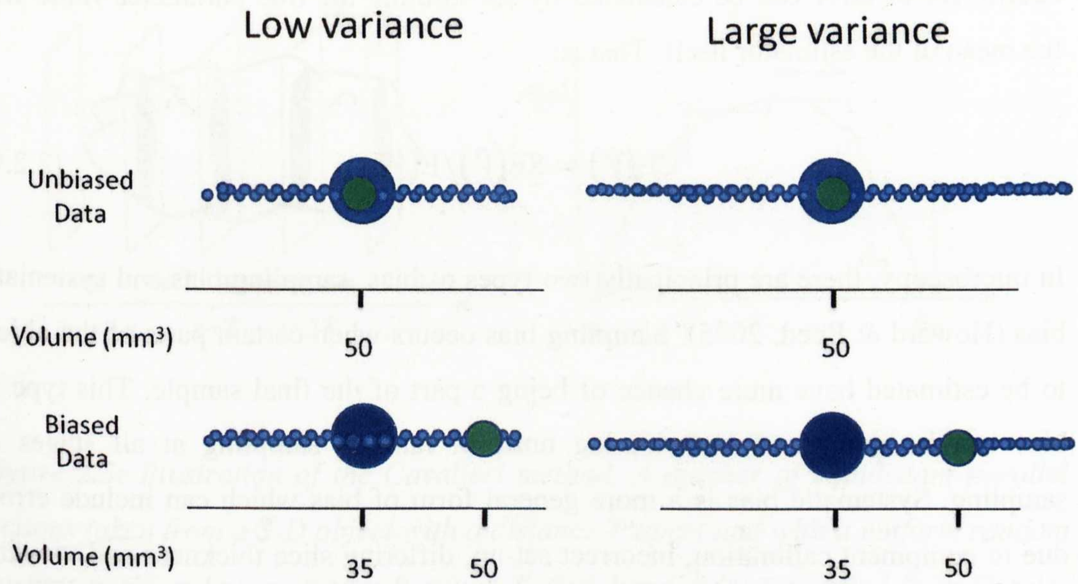


Figure 2.2: *Illustration of the difference between precision (variance) and bias for synthetic 3-D estimates of a 3-D object with true value equal to 50mm^3. For the biased estimator, a systematic measurement error had occurred which resulted in a mean value of the volume estimator of 35mm^3.*

A further useful measure for an unbiased estimator $\hat{V}$ is given by its standard error (SE) and is calculated as the square root of its variance:

$$\mathrm{SE}(\hat{V}) = \sqrt{\mathrm{Var}(\hat{V})} \tag{2.2.7}$$

A commonly used measure of an unbiased estimator in Stereology is neither the estimator's variance nor standard error but its coefficient of error (CE). CE is defined as the variance of the estimator divided by the true parameter value and it can be expressed as follows:

$$\mathrm{CE}(\hat{V}) = \mathrm{SE}(\hat{V})/V \tag{2.2.8}$$

In practice, however, the true parameter value, V in the example above, would be unknown, since this is what is attempting to be estimated. An estimate of the coefficient of error can be calculated by substituting the true parameter value with the mean of the estimator itself. That is:

$$\mathrm{CE}(\hat{V}) = \mathrm{SE}(\hat{V})/\mathrm{E}(\hat{V}) \tag{2.2.9}$$

In microscopy, there are principally two types of bias; sampling bias and systematic bias (Howard & Reed, 2005). Sampling bias occurs when certain parts of the object to be estimated have more chance of being a part of the final sample. This type of bias can be minimised by adopting uniform random sampling at all stages of sampling. Systematic bias is a more general form of bias which can include errors due to equipment calibration, incorrect set-up, differing slice thickness and viewing distortions to name but a few. This bias is very difficult to completely eradicate, but it should be minimised wherever possible. It should be noted here that the methods that are discussed in the next few sections are mathematically unbiased. That is, the estimation methods are unbiased if the data are unbiased. It should also be noted that for the volume and surface area estimators in the following sections that for the reader's benefit, referring to an estimator would be the same as referring to the realization of an estimator (an estimate).

2.2.3 Volume estimator

The volume of a 3-dimensional object, V, can be obtained by integrating the area of the intersection between the object and a plane along a sampling axis. That is:

$$V = \int_{a_*}^{b_*} A(x)\, dx \tag{2.2.10}$$

where $A(x)$ is the area of the section at the abscissa x (see Figure 2.3). As indicated earlier, the volume of a given structure could be, in principle, estimated from Equations (2.2.1)-(2.2.3) where the structure under study (e.g. a mineral) is homogeneously contained within a reference structure (e.g. rock). For most biological structures however, this is not the case.

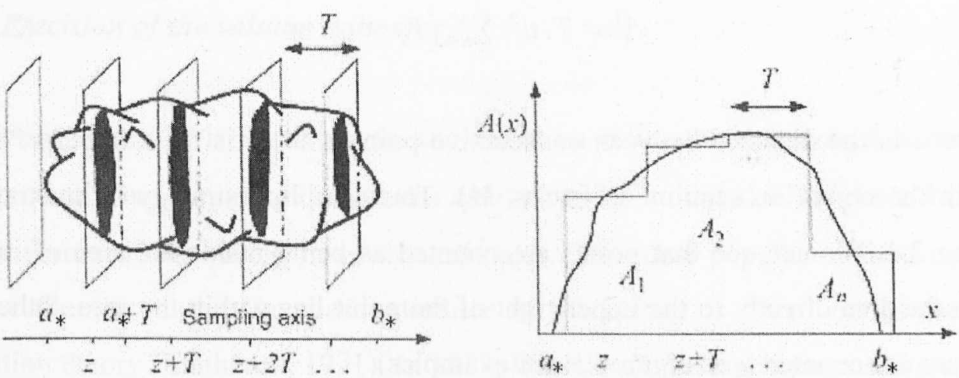


Figure 2.3: *Illustration of the Cavalieri method. A number of equidistant parallel sections taken from a 3-D object with a distance T apart and with a uniform random starting position* (z) *such that* $0 \leq z \leq T$ *(left hand side) (modified from* Garcia-Fiñana *et al.*, 2003).

One of the most common stereological methods for estimating volume is the Cavalieri method (see e.g., Gundersen & Jensen, 1987). This method was named in honour of the mathematician and student of Galileo in the 17th Century, Bonaventura Cavalieri, who was the first to consider the volume estimation of 3-D objects from a number of arbitrary sections. The Cavalieri estimator is mathematically designed to be unbiased. This estimator involves taking M sections with a distance T apart from each other (see Figure 2.3). The area of each section is then calculated, and the volume estimate is obtained by summing these section areas multiplied by the distance T. The Cavalieri volume estimator can be written as:

$$\hat{V} = (T.A_1) + \cdots + (T.A_M) = T.\sum_{s=1}^{M} A_s \qquad (2.2.11)$$

where A_s is the area of the object in the s^{th} section $(s = 1,2,\ldots,M)$. For this method to be unbiased, the distance of the first section from the front of the object should be randomly chosen from the uniform distribution between 0 and T. The area of interest can be estimated for each section by applying point counting (see e.g. Mathieu *et al.*, 1981). This involves the superimposition of a grid of points, randomly translated on each parallel section such that the volume estimator from Equation (2.2.11) takes the form:

$$\tilde{V} = T.u^2.\sum_{s=1}^{M} P_s \qquad (2.2.12)$$

where u is the distance between consecutive points and P_s is the number of points within the object on section $(1 \leq s \leq M)$. This sampling strategy is illustrated in Figure 2.4. We can see that points are counted as being inside the area of interest when the area directly to the upper right of the point lies within the area. Otherwise they are not counted (see Figure 2.4 for examples).

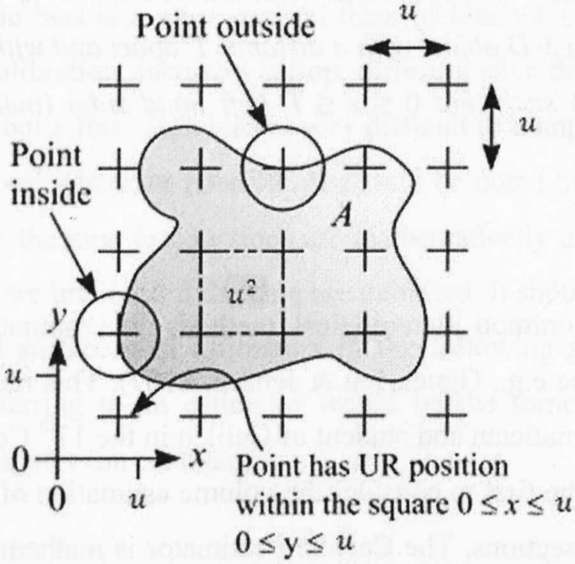


Figure 2.4: *Illustration of a grid of points applied to a Cavalieri section with area A and distance between points u such that the area per point is* u^2 *(figure modified from* http://www.mikepuddephat.com/Page/1599/Single-object-stereology-(part-1) (downloaded in February 2011)).

However, although the Cavalieri method is designed to be unbiased, it is prone to systematic bias called 'overprojection' or 'voluming' if the distance between sections (i.e. T) is large (Howard & Reed, 2005). If overprojection occurs then the volume of the object will be overestimated using the Cavalieri method. As long as the distance between Cavalieri sections, T, is set such that the number of sections, M, is equal to 10 or more, then this bias becomes negligible.

2.2.4 Precision of the volume estimator

In applied studies it is important to know that an estimator is unbiased as well as how precise it is. Assuming the volume estimator $\hat{V}$ is unbiased, $\mathrm{CE}(\hat{V})$ can be used as a measure of the precision of the estimator. The most popular method used to predict the CE of the Cavalieri estimator in Stereology is based on Matheron's transition theory (Matheron, 1971). In 1987, Gundersen & Jensen proposed a way to estimate the CE of the volume estimator described in Equation (2.2.11). The following calculation steps are required for the prediction of the CE:

$$\sum A = \sum_{s=1}^{M} A_s \tag{2.2.13}$$

$$D_1 = \sum_{s=1}^{M} (A_s)^2 \tag{2.2.14}$$

$$D_2 = \sum_{s=1}^{M-1} A_s . A_{s+1} \tag{2.2.15}$$

$$D_3 = \sum_{s=1}^{M-2} A_s . A_{s+2} \tag{2.2.16}$$

Using Equations (2.2.13)-(2.2.16) the coefficient of error of the volume estimator $\hat{V}$ can be estimated as (see Garcia-Fiñana & Cruz-Orive, 2004 and references therein):

$$CE(\hat{V}) = \frac{1}{\sum A} \cdot \left(\alpha(q).(3D_1 + D_3 - 4D_2)\right)^{1/2} \tag{2.2.17}$$

where $\alpha(q)$ is as defined in Equation (2.2.22). The CE of $\tilde{V}$, the volume estimator obtained by the Cavalieri method in combination with the point counting technique (see Equation (2.2.12)) can be written as:

$$\mathrm{CE}(\tilde{V}) = \sqrt{\mathrm{CE}^2(\hat{V}) + \mathrm{CE}^2_{\mathrm{PC}}(\tilde{V})} \tag{2.2.18}$$

where $\mathrm{CE}^2(\hat{V})$ is the contribution of the variability across sections and $\mathrm{CE}^2_{\mathrm{PC}}(\tilde{V})$ is the contribution of the mean variability due to point counting within all sections

(Cruz-Orive 1989; Cruz-Orive, 1999; Gundersen, *et al.*, 1999; Kiêu, *et al.*, 1999; Garcia-Fiñana & Cruz-Orive, 2000a,b; Garcia-Fiñana *et al.*, 2003).

However, since the CE's are obtained from observed data, the formula in Equation (2.2.19) is denoted in lower case:

$$ce(\tilde{V}) = \sqrt{ce^2(\hat{V}) + ce_{PC}^2(\tilde{V})} \qquad (2.2.19)$$

where the contribution of the variability across sections is defined as:

$$ce^2(\hat{V}) = \alpha(q).\frac{1}{(\sum P)^2}(3(D_0^* - \hat{v}) + D_2^* - 4D_1^*) \qquad (2.2.20)$$

such that:

$$\sum P = \sum_{s=1}^{M} P_s \qquad (2.2.21)$$

is the sum of points counted across all M sections. The numerical coefficient $\alpha(q)$ depends upon the fractional smoothness constant of the area function, q, as described in Garcia-Fiñana & Cruz-Orive (2000b), and can be expressed as:

$$\alpha(q) = \frac{\Gamma(2q+2).\zeta(2q+2).\cos(\pi q)}{(2\pi)^{2q+2}.(1-2^{2q-1})} \qquad (2.2.22)$$

where $\Gamma(x)$ is the gamma distribution:

$$\Gamma(x) = \int_0^\infty z^{x-1}e^{-t}dz \qquad x \geq 1 \qquad (2.2.23)$$

(Abramowitz & Stegun, 1972) and $\zeta(x)$ is the Riemann-zeta function:

$$\zeta(x) = \sum_{n=1}^{\infty}\frac{1}{n^x} \qquad x \geq 1 \qquad (2.2.24)$$

(Ivic, 1985; Karatsuba & Voronin, 1992). The value of q is dependent on both the geometrical properties of the structure and the direction in which sections are taken and it has been shown that q may also be dependent on the distance between sections, T (Gundersen, *et al.*, 1999; Garcia-Fiñana & Cruz-Orive, 2000a). The estimator of q can be applied when there are at least 5 sections and is defined as:

$$\hat{q} = \max\left\{0, \frac{1}{2\log(2)}.\log\left(\frac{3(D_0^*-\hat{v})+D_4^*-4D_2^*}{3(D_0^*-\hat{v})+D_2^*-4D_1^*}\right) - \frac{1}{2}\right\} \qquad (2.2.25)$$

(Kiêu, 1997, Kiêu, *et al.*, 1999). The estimate of q has been rounded off to the nearest integer in the past, but $\hat{q}$ is in fact a rational number which takes values between 0 and 1. As a rule of thumb, when the number of sections is less than 5 then $\hat{q}$ can be set equal to 0 if the shape of the object is irregular and if the shape of the object is regular then $\hat{q} = 1$. In terms of the numerical coefficient $\alpha(q)$, when $\hat{q} = 0$, $\alpha(\hat{q}) = 1/12$, when $\hat{q} = 0.5$, $\alpha(\hat{q}) = 1/45$ and when $\hat{q} = 1$, $\alpha(\hat{q}) = 1/240$. In Equations (2.2.20) and (2.2.25) the quantities $D_0^*, \ldots, D_4^*$ can be calculated by the following equations:

$$D_0^* = \sum_{s=1}^{M}(P_s)^2 \qquad (2.2.26)$$

$$D_1^* = \sum_{s=1}^{M-1} P_s.P_{s+1} \qquad (2.2.27)$$

$$D_2^* = \sum_{s=1}^{M-2} P_s.P_{s+2} \qquad (2.2.28)$$

$$D_4^* = \sum_{s=1}^{M-4} P_s.P_{s+4} \qquad (2.2.29)$$

The final unknown quantity from Equation (2.2.20) yet to be defined is $\hat{v}$ which is an estimator of the point counting variance, also known as 'nugget' variance:

$$\hat{v} = 0.0724.\left(\overline{F_2}/\sqrt{\overline{F_1}}\right).(M\sum P)^{1/2} \qquad (2.2.30)$$

The quantities $\overline{F_1}$ and $\overline{F_2}$ in Equation (2.2.30) are related to the shape of the object of interest. Take a section of two objects, one a rounded ellipsoid and the other an irregular shape, with the same total area. Each section has a uniform random grid of points superimposed on both. The irregular shaped section will have a greater variance in number of points lying within it than the rounded ellipsoid, due to the positioning of the point-grid and the irregularity of the shape of the object itself. For a more precise estimate of section area (i.e. the variance of number of points counted is reduced), the irregular shaped section would need a more dense grid of points than the rounded ellipsoid would. The dimensionless shape coefficient $\overline{F_2}/\sqrt{\overline{F_1}}$ from Equation (2.2.30) takes the variability of the shape of the object of interest into account where $\overline{F_1}$ is the mean area of the sections (also known as the mean transect area) and the mean boundary length is given by $\overline{F_2}$. The boundary length for section $(1 \leq s \leq M)$, F_{2s}, can be calculated, after superimposing a grid of parallel lines on the section with a uniform random position and isotropic orientation, as:

$$F_{2s} = (\pi/2).(a/l).I_s \tag{2.2.31}$$

where a is the test area around one point on an isotropic line grid, l is the length of the test lines per point and I_s is the number of intersections between the boundary of the shape of interest and the line grid. In the case of the volume estimation method in Section 2.2.3, the grid of points in Figure 2.4 has distance between points equal to u. This means that $a = u^2$ and assuming that $l = 2u$, Equation (2.2.31) becomes:

$$F_{2s} = (\pi/4).u.I_s \tag{2.2.32}$$

The area of section s, F_{1s}, can be estimated using the point counting technique as:

$$\hat{F}_{1s} = u^2.P_s \tag{2.2.33}$$

The shape coefficient for one section can therefore be calculated by combining Equations (2.2.32) and (2.2.33) as follows:

$$F_{2s}/\sqrt{\hat{F}_{1s}} = (\pi.u.I_s)/\left(4\sqrt{u^2P_s}\right) = \tfrac{\pi}{4}.I_s.\sqrt{P_s} \tag{2.2.34}$$

The mean of $\hat{F}_{1s}$ and F_{2s} can be calculated as $\overline{F_1} = \left(\sum_{s=1}^{M} \hat{F}_{1s}\right)/M$ and $\overline{F_2} = \left(\sum_{s=1}^{M} F_{2s}\right)/M$, respectively, and so the shape coefficient from Equation (2.2.30) can be written as:

$$\overline{F_2}/\sqrt{\overline{F_1}} = \frac{\pi}{4}.\frac{\sum I}{\sqrt{\sum P}}.\frac{1}{\sqrt{M}} \tag{2.2.35}$$

where $\sum I = \sum_{s=1}^{M} I_s$.

Finally, the second term on the right hand side of Equation (2.2.19), the mean variability due to point counting within all sections, can be written as:

$$ce_{PC}^2(\tilde{V}) = \frac{\hat{v}}{(\sum P)^2} \tag{2.2.36}$$

2.2.5 Worked example for the precision of the volume estimator

The coefficient of error (CE) of the volume estimator for Pars Opercularis in the right hemisphere (RHPO) grey matter of an example participant (FM032) can be calculated using the formulae in the previous subsection (Section 2.2.4). The distance between Cavalieri sections (T) was 1mm and the size of the square grid test system (u) was 3mm. In this particular example, the number of Cavalieri sections (M) was equal to 17. The aim of this worked example is fourfold:

- To obtain an estimate of the volume for the RHPO grey matter of the example subject using Equation (2.2.12).
- To estimate the coefficient of error that is representative of the variability across sections in relation to the total volume estimate (see Equation (2.2.20)).
- To estimate the coefficient of error due to point counting using Equation (2.2.36).
- To estimate the total coefficient of error present in the volume estimate (see Equation (2.2.19)).

Table 2.1 shows the number of points within the RHPO grey matter for the example participant, along with the calculations of the quantities $D_0^*, \ldots, D_4^*$ as described in Equations (2.2.26)-(2.2.29), respectively. These quantities are needed for some of the coefficient of error calculations. The sum of the points across sections will be used in the estimation of the volume of the region of interest itself.

The volume estimate, $\tilde{V}$, can now be estimated from Equation (2.2.12) using the points from each section as recorded in Table 2.1:

$$\tilde{V} = 1 \cdot (3)^2 \cdot (19 + 22 + 26 + 37 + 34 + 40 + 37 + 39 + 41 + 34 + 35 + 27 + 23 + 20 + 11 + 4 + 2) = 4059\text{mm}^3 \tag{2.2.37}$$

The nugget variance $\hat{\nu}$ is needed for the calculation of both the coefficient of error due to Cavalieri sections and point counting. From Equation (2.2.30), this can be written as:

$$\hat{\nu} = 0.0724 \times 7.7 \times \sqrt{17 \times 451} \approx 48.81 \tag{2.2.38}$$

where the estimate of the shape coefficient $\overline{F_2}/\sqrt{\overline{F_1}}$ used for Pars Opercularis is 7.7.

The fractional smoothness constant of the area function, $\hat{q}$ in Equation (2.2.25), is then calculated as:

$$\hat{q} = \max\left\{0, \frac{1}{2\log(2)} \times \log\left(\frac{3 \cdot (14397 - 48.81) - 4 \cdot 13386 + 11195}{3 \cdot (14397 - 48.81) - 4 \cdot 13974 + 13386}\right) - \frac{1}{2}\right\}$$

$$\hat{q} = \max\{0, -0.31\} = 0 \tag{2.2.39}$$

As was stated in Section 2.2.4, when $\hat{q} = 0$:

$$\alpha(\hat{q}) = 1/12 \tag{2.2.40}$$

Using these results, and by square rooting Equation (2.2.20), the coefficient of error due to Cavalieri sections becomes:

$$\mathrm{ce}(\hat{V}) = \sqrt{\frac{1}{12} \cdot (3(14397 - 48.81) - 4 \cdot 13974 + 13386) \cdot \frac{1}{451^2}} \approx 0.0148$$

(2.2.41)

This means that the contribution to the total coefficient of error of the volume estimator that is due to the Cavalieri sampling is approximately equal to 1.5%. The coefficient of error due to point counting can be obtained by square rooting Equation (2.2.36) such that:

$$\mathrm{ce}_{\mathrm{PC}}(\tilde{V}) = \sqrt{48.81 \cdot \frac{1}{451^2}} \approx 0.0155 \qquad (2.2.42)$$

Section s	P_s	P_s^2	$P_s \cdot P_{s+1}$	$P_s \cdot P_{s+2}$	$P_s \cdot P_{s+4}$
1	19	361	418	494	646
2	22	484	572	814	880
3	26	676	962	884	962
4	37	1369	1258	1480	1443
5	34	1156	1360	1258	1394
6	40	1600	1480	1560	1360
7	37	1369	1443	1517	1295
8	39	1521	1599	1326	1053
9	41	1681	1394	1435	943
10	34	1156	1190	918	680
11	35	1225	945	805	385
12	27	729	621	540	108
13	23	529	460	253	46
14	20	400	220	80	-
15	11	121	44	22	-
16	4	16	8	-	-
17	2	4	-	-	-
Total	451	14397	13974	13386	11195
		$\boldsymbol{D_0^*}$	$\boldsymbol{D_1^*}$	$\boldsymbol{D_2^*}$	$\boldsymbol{D_4^*}$

Table 2.1: *Table of points within the RoI (RHPO grey matter) for each Cavalieri section, together with the calculation of D_0^*, D_1^*, D_2^* and D_4^* using Equations (2.2.26)-(2.2.29), respectively.*

The contribution to CE of the volume estimator that is due to point counting can be estimated to be 1.6%. Finally the total coefficient of error for the volume estimator in Equation (2.2.19) can now be written as:

$$ce(\tilde{V}) = \sqrt{0.0148^2 + 0.0155^2} \approx 0.0214 \qquad (2.2.43)$$

meaning that the estimate of total coefficient of error of volume estimator due to point counting and Cavalieri sections combined is equal to 2.1%.

2.2.6 Surface area estimator

One approach to estimate surface area involves the uniform random superimposition of a grid of parallel equidistant lines in an isotropic orientation over uniform random parallel equidistant sections in an isotropic orientation throughout a 3-D object (e.g. Elias & Schwartz, 1969) (see Figure 2.5). The estimator of the surface area, $\hat{S}$, can be then expressed as a function of the intersections between the lines and the surface boundary of the object as follows:

$$\hat{S} = 2.I.\tilde{h}.T \qquad (2.2.44)$$

where I is the total number of intersections across all sections, $\tilde{h}$ is the distance between lines and T is the distance between sections.

An alternative method for the estimation of surface area was later developed in 1986 by Baddeley *et al.* (see also Cruz-Orive, 2006). To guarantee unbiasedness three requirements must be met during the sampling. These three requirements are: (i) the object to be estimated must have an identifiable directional axis or a direction should be generated by the experimenter (referred to as the 'vertical axis'); (ii) the vertical sections (sections parallel to the vertical axis or perpendicular to the horizontal plane) must be randomly orientated in respect of the horizontal plane (see Figure 2.6); (iii) the test line to be superimposed on the vertical sections must be given a weight proportional to $\sin(\theta)$ where θ is the angle between the vertical direction and the test line.

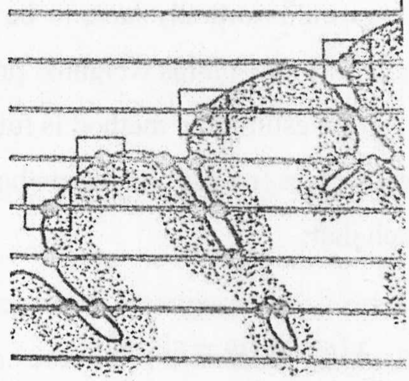


Figure 2.5: *Illustration of a grid of parallel equidistant lines superimposed onto a section of the cerebral cortex of a mammal obtained with isotropic orientation (with intersections marked with dots)* (Elias & Schwartz, 1969).

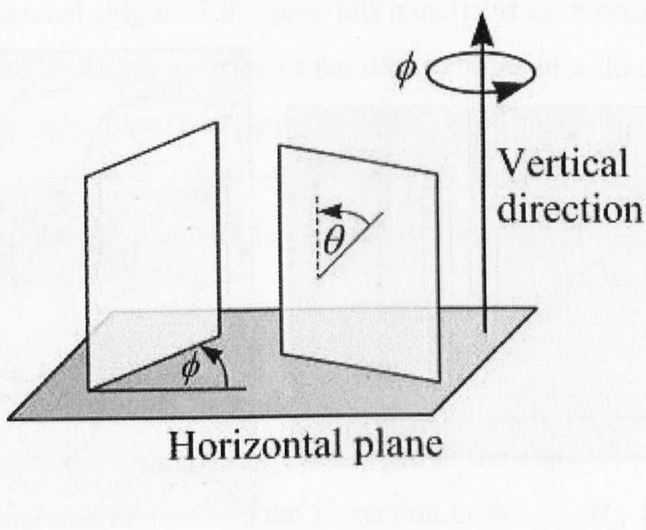


Figure 2.6: *A vertical axis is chosen perpendicular to a horizontal plane* (http://www.mikepuddephat.com/Page/1600/Single-object-stereology-(part-2) (downloaded in February 2011)).

By defining the vertical axis, anisotropy is introduced into the sampling design. Anisotropy is a property of being directionally dependent, which in this case is in the direction of the vertical axis. However, the sampling design should be isotropic, meaning that all directions should have identical properties. To compensate for this, a numerical weighting factor would normally have to be included in the estimation method. However, by using a grid of cycloids weighted proportional to $\sin(\theta)$ as the test lines, requirement (iii) for the estimation method is fulfilled and so the numerical weighting factor is not necessary in the estimation method (Baddeley *et al.*, 1986). The cycloids have an arc such that:

$$x(\varphi) = (\varphi - \sin\varphi)r \tag{2.2.45}$$

$$y(\varphi) = (1 - \cos\varphi)r \tag{2.2.46}$$

where $0 < \varphi < \pi$ and r is a variable linked to the cycloid size such that $r = l/16$ with l equal to the cycloid length (see Figure 2.7).

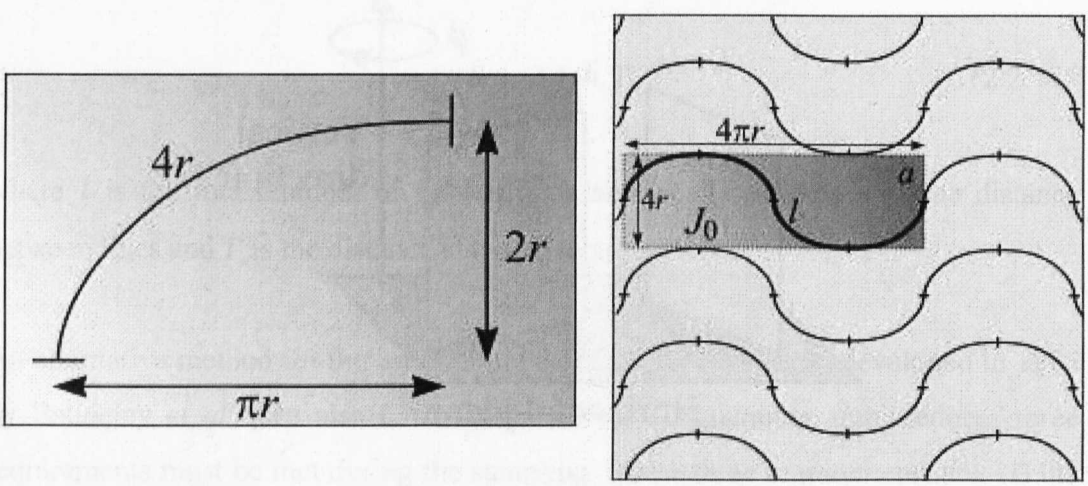


Figure 2.7: *Figure of the mathematical curve of the cycloid (left panel) and the cycloid grid (right panel)* (http://www.mikepuddephat.com/Page/1600/Single-object-stereology-(part-2) (downloaded in February 2011)).

For requirement (iii) to hold when this grid of cycloids is used as the test lines, the minor axis of the cycloid must be parallel to the defined vertical axis. This means that φ is the angle between the vertical direction and the test lines. The true surface area, S, can be thought of as the integral of a function f over all angles of orientation φ such that:

$$S = \int_0^{\pi} f(\varphi)\, d\varphi \qquad (2.2.47)$$

(see Cruz-Orive, 2006 for more details about the function f).

The estimation procedure itself involves three levels of sampling. Firstly, a horizontal plane is chosen. Secondly, a random angle, ϕ_0, is generated between 0 and π/w $(w \geq 1)$, on this horizontal plane. A systematic set of w orientations are then generated, $(\phi_0, \dots, \phi_{(w-1)})$, such that:

$$\phi_i = \phi_0 + (\pi/w)(i-1) \qquad (i = 1, \dots, w) \qquad (2.2.48)$$

An example of this can be seen in Figure 2.8 with $w = 3$ and $\theta = \phi_0$ being the random angle generated (Figure 2.8, upper left panel). At each orientation, Cavalieri sampling is applied to obtain a series of parallel sections at a distance T apart (see Figure 2.8, upper right panel). A grid of cycloids is then superimposed with a uniform, random position on each individual section and for each orientation such that the minor axis of the cycloids is parallel to the vertical axis (see Figure 2.8, lower panel).

The properties of the cycloids were defined earlier in this section. The number of intersections between the cycloids and the surface of the object are then counted. Let I_{ij} be the number of intersections on the j^{th} section $(j = 1, \dots, U_i;\ U_i \geq 1)$ for the i^{th} orientation $(i = 1, \dots, w)$. The angle between orientations is set at $h = \frac{\pi}{w}$ (i.e. the systematic sampling period). Then the surface area estimator with w orientations, $\tilde{S}_w$, can be estimated from the following equation:

$$\tilde{S}_w = h \sum_{i=1}^{w} \hat{f}_i \qquad (2.2.49)$$

The function $\hat{f}_i$ can be calculated as:

$$\hat{f}_i = \frac{2}{\pi}.\frac{a}{l}.T.\sum_{j=1}^{U_i} I_{ij} \tag{2.2.50}$$

where a is the cycloid test unit area, l is the cycloid test line length (see Figure 2.7) and T is the distance between the Cavalieri sections. Combining Equations (2.2.49) and (2.2.50), the surface area estimator $\tilde{S}_w$ becomes:

$$\tilde{S}_w = \frac{2}{w}.\frac{a}{l}.T.\sum_{i=1}^{w} I_i \tag{2.2.51}$$

where I_i is the total number of intersections on the i^{th} orientation (see Cruz-Orive, 2006).

An alternative method called the 'vertical spatial grid' method is similar to the method described above, except that the cycloids are set to specific vertical shifts on each Cavalieri section, as opposed to a random, uniform position (Cruz-Orive & Howard, 1995). The vertical shifts form cycloids in the orientation direction meaning that two systematic arrangements of cycloid chains are then contained in two mutually perpendicular systems of parallel vertical planes. Although this method is not used in this thesis, a comparison of these two methods on our images may be useful to examine the differences between them.

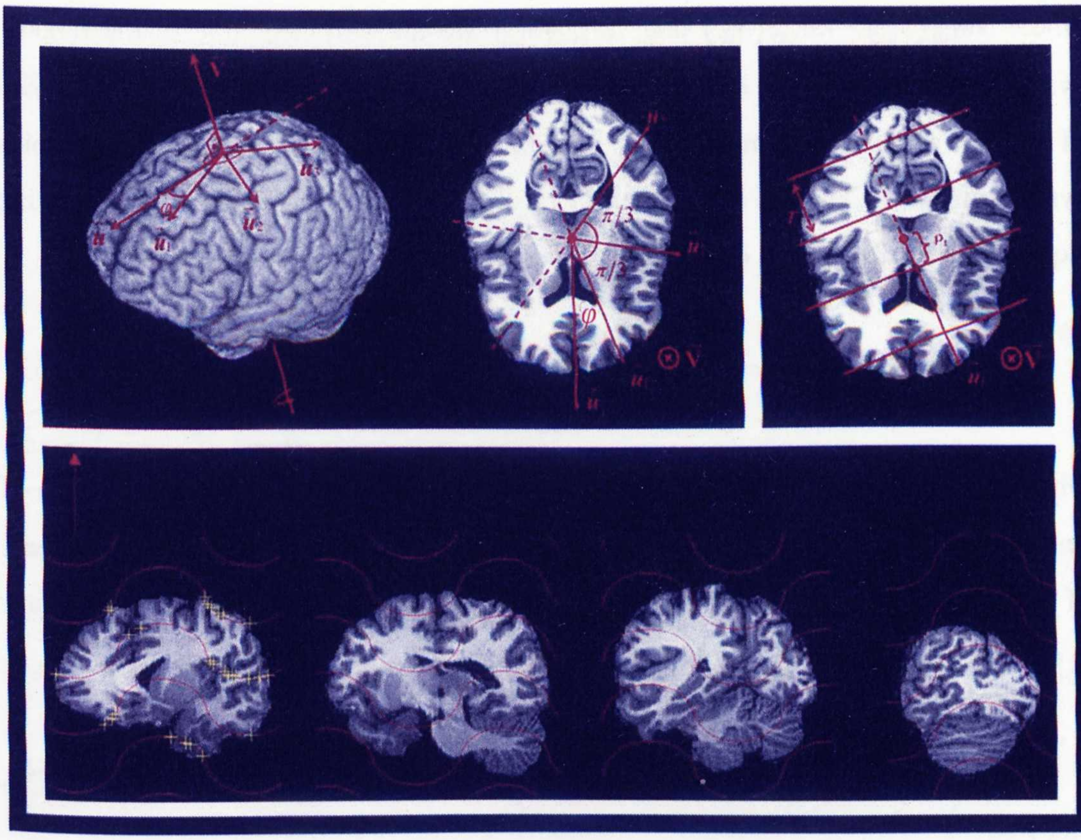


Figure 2.8: *In this example, 3 orientations are obtained using systematic sampling with φ being randomly generated such that $0 < \varphi < \pi/3$, Cavalieri sections are obtained at a distance T apart, and the cycloid grid is randomly 'thrown' onto each section* (modified from van Aarde, 2006).

2.2.7 *Precision of the surface area estimator*

To estimate the precision of the surface area estimator is a more complex task than to estimate the precision of the volume estimator. This additional complexity arises because of an additional level of sampling involved in the estimation of surface area (the volume estimation has two levels of sampling, whereas the surface area estimation involves three levels). The variance of the surface area estimator, $\mathrm{Var}(\tilde{S}_w)$, is equal to the sum of the variance due to each of the three levels of sampling:

$$\mathrm{Var}(\tilde{S}_w) = \mathrm{Var}_{\phi}(\tilde{S}_w) + \mathrm{Var}_{\mathrm{cav}}(\tilde{S}_w) + \mathrm{Var}_{\mathrm{cyc}}(\tilde{S}_w) \qquad (2.2.52)$$

where $\text{Var}_\phi(\tilde{S}_w)$, $\text{Var}_{\text{cav}}(\tilde{S}_w)$ and $\text{Var}_{\text{cyc}}(\tilde{S}_w)$ are the variances due to the random systematic sampling of w orientations (first level of sampling), the w Cavalieri series of vertical sections (second level of sampling) and the sampling using the cycloids system on each Cavalieri section (third level of sampling), respectively. Following Equation (2.2.52), the CE of the surface area estimator can be expressed as:

$$\text{CE}^2(\tilde{S}_w) = \text{CE}^2_\phi(\tilde{S}_w) + \text{CE}^2_{\text{cav}}(\tilde{S}_w) + \text{CE}^2_{\text{cyc}}(\tilde{S}_w) \qquad (2.2.53)$$

where $\text{CE}^2_\phi(\tilde{S}_w)$, $\text{CE}^2_{\text{cav}}(\tilde{S}_w)$ and $\text{CE}^2_{\text{cyc}}(\tilde{S}_w)$ are the square of the coefficients of error due to the three levels of sampling as described above for the three components of the variance. The estimator of the coefficient of error of the surface area estimator is denoted using lowercase symbols as follows:

$$\text{ce}^2(\tilde{S}_w) = \text{ce}^2_\phi(\tilde{S}_w) + \text{ce}^2_{\text{cav}}(\tilde{S}_w) + \text{ce}^2_{\text{cyc}}(\tilde{S}_w) \qquad (2.2.54)$$

The idea is to be able to distinguish between the exact expression of $\text{CE}^2(\tilde{S}_w)$ in Equation (2.2.53) and the estimator based on the observed data in Equation (2.2.54).

(i) Estimation of $\text{Var}_\phi(\tilde{S}_w)$

In order to estimate these three components of the coefficient of error, the variances from Equation (2.2.52) due to each level of sampling must be estimated. From this point on, the estimator of the variance of the surface area estimator (in other words, the estimator of $\text{Var}(\tilde{S}_w)$), will be referred to as $\text{var}_m(\tilde{S}_w)$ to maintain consistency with Cruz-Orive & Gual-Arnau (2002). The estimator $\text{var}_m(\tilde{S}_w)$ can be written as:

$$\text{var}_m(\tilde{S}_w) = \text{var}_m(\hat{S}_w) + h^2\hat{v}_w \qquad (2.2.55)$$

where $\text{var}_m(\hat{S}_w)$ is the variance due to the systematic sampling on a half circle and is based on a model that takes into account the circular sampling nature (see Cruz-Orive & Gual-Arnau (2002) for details). The parameter m can take the form of 0 or 1 and also comes from this 'global model'. The first term in the right hand side of Equation (2.2.55), $\mathrm{var}_m(\hat{S}_w)$ can be calculated from:

$$\mathrm{var}_m(\hat{S}_w)$$
$$= \max\left\{0, \left(\pi^2 . B_{2m+2} . \left(\hat{C}_0 - \hat{C}_1 - \hat{v}_w\right)\right) / \left(\left(B_{2m+2} - B_{2m+2}(1/w)\right) . w^{2m+3}\right)\right\} \tag{2.2.56}$$

where $\hat{C}_0$ and $\hat{C}_1$, can be calculated as:

$$\hat{C}_H = \sum_{i=1}^{w} \hat{f}_i \hat{f}_{i+H} \tag{2.2.57}$$

Note that $\hat{f}_{i+w} = \hat{f}_i$, due to the periodicity of circular systematic sampling, and when $H = 0,1, \dots, [w/2]$, with $[w/2]$ defined as the integer part of a real number $w/2$, it can then also be stated that:

$$\hat{C}_{w-H} = \hat{C}_H \tag{2.2.58}$$

The quantity B_{2m+2} is a Bernoulli polynomial as defined in Abramowitz & Stegun (1972) and m, as stated above, can be either 0 or 1. To determine whether m should be set at 0 or 1, the following should be adhered to:

$$m = \begin{cases} 0 & \text{if } |\hat{D}_0| \leq |\hat{D}_1| \\ 1 & \text{otherwise} \end{cases} \tag{2.2.59}$$

where $\hat{D}_0$ and $\hat{D}_1$ take the form:

$$\hat{D}_0 = \left(\hat{C}_0 - \hat{C}_2 - \hat{v}_w\right) / \left(\hat{C}_0 - \hat{C}_1 - \hat{v}_w\right) - \left(2(w-2)\right)/(w-1) \tag{2.2.60}$$

$$\hat{D}_1 = \left(\hat{C}_0 - \hat{C}_2 - \hat{v}_w\right) / \left(\hat{C}_0 - \hat{C}_1 - \hat{v}_w\right) - 4(w-2)^2/(w-1)^2 \tag{2.2.61}$$

(Cruz-Orive & Gual-Arnau, 2002). Note that the equations for $\hat{D}_0$ and $\hat{D}_1$ only hold if $w \geq 4$. That is, there has to be a minimum of 4 orientations so that there are enough data for the calculations to take place. From the definitions of the Bernoulli polynomials and Equation (2.2.56), $\mathrm{var}_m(\hat{S}_w)$ can now be written as two equations depending upon whether the value m equals 0 or 1, respectively:

$$\mathrm{var}_0(\hat{S}_w) = \max\{0, \pi h(\hat{C}_0 - \hat{C}_1 - \hat{v}_w)/6(w-1)\} \tag{2.2.62}$$

$$\mathrm{var}_1(\hat{S}_w) = \max\{0, \pi h(\hat{C}_0 - \hat{C}_1 - \hat{v}_w)/30(w-1)^2\} \tag{2.2.63}$$

Also note that Equations (2.2.62) and (2.2.63) hold only if $w \geq 2$. It is also useful to note that the term $\hat{v}_w$ is an estimator of the variance of the local Cavalieri sections based on the w orientations.

(ii) Estimation of $Var_{cav}(\tilde{S}_w)$

The second term in Equation (2.2.52), the variance due to the random systematic sampling of the w Cavalieri series of vertical sections, $\mathrm{Var}_{\mathrm{cav}}(\tilde{S}_w)$, can be estimated as the product of the square of the angle between the orientations ($h = \pi/w$) and the estimator $\hat{v}_w$ (the second term on the right side of Equation (2.2.55)). That is:

$$\mathrm{var}_{\mathrm{cav}}(\tilde{S}_w) = h^2 \hat{v}_w \tag{2.2.64}$$

such that

$$\hat{v}_w = \sum_{i=1}^{w} \hat{\sigma}_i^2 \tag{2.2.65}$$

where

$$\hat{\sigma}_i^2 = ((2/\pi).(a/l))^2.T^2.\alpha(q_i).(3(\hat{C}_{0i} - \hat{v}_{w_i}) - 4\hat{C}_{1i} + \hat{C}_{2i}) + ((2/\pi).(a/l))^2.T^2.\hat{v}_{w_i} \tag{2.2.66}$$

The values $\hat{C}_{Hi}$ ($H = 0,1,\dots,U_i - 1$) can be calculated from:

$$\hat{C}_{Hi} = \sum_{j=1}^{U_i - H} I_{ij}\, I_{i,j+H} \tag{2.2.67}$$

with I_{ij} being the number of intersections on the j^{th} section of the i^{th} orientation. The estimator $\hat{v}_{w_i}$ is defined as the second stage nugget component (an estimator of the variance due to the test system (i.e. the cycloid grid)) (Cruz-Orive, 2006), and is defined in Equation (2.2.71). Equation (2.2.65) only holds when $w \geq 3$. By substituting the estimator of the variance due to Cavalieri sampling for each orientation i in Equation (2.2.66) into the previous Equation (2.2.65), $\mathrm{var_{cav}}(\tilde{S}_w)$ from Equation (2.2.64) becomes:

$$\mathrm{var_{cav}}(\tilde{S}_w) = h^2 \sum_{i=1}^{w} \left(\left((2/\pi).(a/l)\right)^2.T^2.\alpha(q_i).\left(3\left(\hat{C}_{0i} - \hat{v}_{w_i}\right) - 4\hat{C}_{1i} + \hat{C}_{2i}\right)\right) + h^2 \sum_{i=1}^{w} \left(\left((2/\pi).(a/l)\right)^2.T^2.\hat{v}_{w_i}\right) \tag{2.2.68}$$

The first right hand term of Equation (2.2.68) is the estimator of the variance due to the U Cavalieri sections ($U = \sum_{i=1}^{w} U_i$) for all w orientations. The estimator of the variance due to the cycloid grid test system for all U Cavalieri sections takes the form of the second right hand term of Equation (2.2.68). The term $\alpha(q_i)$ represents the numerical coefficient that is dependent upon the fractional smoothness constant of the area function similar to that defined in Equation (2.2.22), such that:

$$\alpha(q_i) = [\Gamma(2q_i + 2).\zeta(2q_i + 2).\cos(\pi q_i)]/[(2\pi)^{2q_i+2}.(1 - 2^{2q_i-1})] \tag{2.2.69}$$

where $\Gamma(x)$ is the gamma distribution as defined in Equation (2.2.23), $\zeta(x)$ is the Riemann-zeta function as defined in Equation (2.2.24) and $q_i \in [0,1]$. An estimate of q_i can be obtained using Equation (2.2.70), only if the number of vertical sections, U_i, in each orientation i is greater than or equal to 5.

$$\hat{q}_i = \max\left\{0, \frac{1}{2\log(2)}.\log\left(3\left(\hat{C}_{0i} - \hat{v}_{w_i}\right) + \hat{C}_{4i} - 4\hat{C}_{2i}/3\left(\hat{C}_{0i} - \hat{v}_{w_i}\right) + \hat{C}_{2i} - 4\hat{C}_{1i}\right) - \frac{1}{2}\right\} \tag{2.2.70}$$

(Kiêu *et al.*, 1999). The estimator $\hat{v}_{w_i}$ can be written as:

$$\hat{v}_{w_i} = \sum_{j=1}^{U_i} \hat{\sigma}_{ij}^2 \tag{2.2.71}$$

with $i = 1, \dots, w$ and $\hat{\sigma}_{ij}^2$ being an estimator of the variance of the number of intersections between the boundary of the object and cycloid grid, on the j^{th} Cavalieri section of orientation i. Since this is a rolling cycloid probe test system, the formula relating to the coefficient of error estimator from a single intersection count I_{ij} cannot be directly estimated. However, the total variance estimate due to sampling with cycloids can be calculated using the Poisson model. It can be shown that the standard deviation of this data is equal to the square root of the mean (Cruz-Orive & Gual-Arnau, 2002). Hence, the variance estimator, $\hat{\sigma}_{ij}^2$, is equal to the number of intersections on the j^{th} section of the i^{th} orientation:

$$\hat{\sigma}_{ij}^2 = I_{ij} \tag{2.2.72}$$

(iii) Estimation of $var_{cyc}(\tilde{S}_w)$

The second term on the right hand side of Equation (2.2.68) provides the estimator of variance of the third level of sampling from Equation (2.2.52) (sampling using the cycloids system on each Cavalieri section), which can therefore be written as:

$$\mathrm{var_{cyc}}(\tilde{S}_w) = h^2 \textstyle\sum_{i=1}^{w} \left((2/\pi).(a/l)\right)^2 . T^2 . \hat{v}_{w_i} \tag{2.2.73}$$

2.2.8 Worked example for the precision of the surface area estimator

A worked example for the CE of the surface area estimator for right hemisphere Pars Opercularis (RHPO) is given below for a single subject (FM032). The number of random, systematic orientations (w) is 4 and the angle between orientations, h, can therefore be written as $h = \pi/4$. The distance between Cavalieri sections on each orientation (T) was equal to 1mm, and the ratio of test area to the cycloid test line length (a/l) was set to 3mm. The objectives of this worked example are:

- To estimate the surface area, $\tilde{S}_w$, for the RHPO of a single subject using Equation (2.2.51).

- To estimate the coefficient of error of the surface area estimator that is due to the third level of sampling using the cycloid grid test system, $\mathrm{ce}_{\mathrm{cyc}}(\tilde{S}_w)$ (third term on the right hand side of Equation (2.2.54)).
- To estimate the coefficient of error of the surface area estimator that is directly related to the Cavalieri sections, $\mathrm{ce}_{\mathrm{cav}}(\tilde{S}_w)$ (second term on the right hand side of Equation (2.2.54).
- To estimate the coefficient of error of the surface area estimator due to the first level of sampling (systematic sampling on a semi-circle), $\mathrm{ce}_{\phi}(\tilde{S}_w)$ (first term on the right hand side of Equation (2.2.54)).
- To calculate the total coefficient of error of the surface area estimator (due to the contribution of the three levels of sampling mentioned above), $\mathrm{ce}(\tilde{S}_w)$ in Equation (2.2.54).

Surface area estimate

Table 2.2 shows the intersection counts for each Cavalieri section j of each orientation i, along with the total sum of intersections for each orientation. These are needed for the surface area estimation.

The surface area of right hemisphere PO for this subject can then be estimated by using the intersection counts from Table 2.2, as well as some of the parameters specified earlier (such as T and h) and substituting them into Equation (2.2.51):

$$\tilde{S}_w = 2/4 \times 3 \times 1 \times (47 + 61 + 57 + 42) \approx 310.5\mathrm{mm}^2 \quad (2.2.74)$$

Orientation	Section	I_{ij}						$\sum I_{ij}$
i	j	1	2	3	4	5	6	
1	5	13	10	9	11	4	0	47
2	6	16	14	13	6	9	3	61
3	6	18	13	11	10	4	1	57
4	4	2	14	15	11	0	0	42

Table 2.2: *The intersection counts for each Cavalieri section of each orientation for the RHPO of subject FM032.*

Coefficient of error due to the third level of sampling

By using the Poisson model to calculate the total variation due to the sampling with cycloid grids, the variance of the variable number of intersections between the boundary of the object and cycloid grid on the j^{th} section of the i^{th} orientation is simply equal to the number of intersections (see Equation (2.2.72)). This means that $\hat{v}_{W_i}$ can be calculated by applying Equation (2.2.71) to the data in Table 2.2 such that:

$$\hat{v}_{W_1} = 47 \quad (2.2.75)$$

$$\hat{v}_{W_2} = 61 \quad (2.2.76)$$

$$\hat{v}_{W_3} = 57 \quad (2.2.77)$$

$$\hat{v}_{W_4} = 42 \quad (2.2.78)$$

$$\hat{v}_{W_i} = 47 + 61 + 57 + 42 = 207 \quad (2.2.79)$$

This can now be substituted into Equation (2.2.73) to calculate the variance due to cycloid grid sampling:

$$\mathrm{var_{cyc}}(\tilde{S}_W) = (\pi/4)^2 \times \left((2/\pi) \times 3\right)^2 \times 1^2 \times 207$$

$$\mathrm{var}_{\mathrm{cyc}}(\tilde{S}_w) = 465.8\mathrm{mm}^4 \quad (2.2.80)$$

The coefficient of error due to cycloids sampling can now be predicted by the application of Equation (2.2.9) to Equations (2.2.80) and (2.2.74):

$$\mathrm{ce}_{\mathrm{cyc}}(\tilde{S}_w) = \left(\sqrt{465.8}/310.5\right) \times 100 = 7.0\% \quad (2.2.81)$$

Coefficient of error due to the second level of sampling

To be able to estimate the coefficient of error of the surface area estimator due to Cavalieri sampling, $\hat{v}_w$ has to be calculated (see Equation (2.2.65)). To do this, $\hat{\sigma}_i^2$ has to be calculated for each orientation i (Equation (2.2.66)), which in turn needs $\hat{C}_{0i}, \ldots, \hat{C}_{4i}$, $\hat{q}_i$ and $\alpha(\hat{q}_i)$ for each orientation i (see Equations (2.2.67), (2.2.70) and (2.2.69), respectively). The values of $\hat{C}_{0i}, \ldots, \hat{C}_{4i}$ are products of the number of intersections between the cycloid grid and the region of interest on Cavalieri sections, from the same orientation (i) (see Equation (2.2.67) for details). These products can be seen in Table 2.3 and a summary of $\hat{C}_{0i}, \ldots, \hat{C}_{4i}$ for each orientation i can be seen in Table 2.4.

Orientation	Section	Intersections				
i	j	I_{ij}	$I_{ij}.I_{ij}$	$I_{ij}.I_{i,j+1}$	$I_{ij}.I_{i,j+2}$	$I_{ij}.I_{i,j+4}$
1	1	13	169	130	117	52
	2	10	100	90	110	0
	3	9	81	99	36	0
	4	11	121	44	0	0
	5	4	16	0	0	0
Total		47	487	363	263	52
2	1	16	256	224	208	144
	2	14	196	182	84	42
	3	13	169	78	117	0
	4	6	36	54	18	0
	5	9	81	27	0	0
	6	3	9	0	0	0
Total		61	747	565	427	186
3	1	18	324	234	198	72
	2	13	169	143	130	13
	3	11	121	110	44	0
	4	10	100	40	10	0
	5	4	16	4	0	0
	6	1	1	0	0	0
Total		57	731	531	382	85
4	1	2	4	28	30	0
	2	14	196	210	154	0
	3	15	225	165	0	0
	4	11	121	0	0	0
Total		42	546	403	184	0

Table 2.3: *Number of intersections, and their products, for each orientation i.*

Orientation (i)	$\sum I_{ij}$	$\widehat{C}_{0i}$	$\widehat{C}_{1i}$	$\widehat{C}_{2i}$	$\widehat{C}_{4i}$
1	47	487	363	263	52
2	61	747	565	427	186
3	57	731	531	382	85
4	42	546	403	184	0

Table 2.4: *Sum of intersections and $\hat{C}_{0i}, \ldots, \hat{C}_{4i}$ for each orientation i.*

The results in Table 2.4 can then be used along with Equations (2.2.75)–(2.2.78) to calculate $\hat{q}_i$, from Equation (2.2.70), for each orientation i, such that:

$$\hat{q}_1 = \max\left\{0, \frac{1}{2\log(2)} \times \log\left(\frac{3\cdot(487-47) - 4\cdot 263 + 52}{3\cdot(487-47) - 4\cdot 363 + 263}\right) - \frac{1}{2}\right\}$$

$$\hat{q}_1 = \max\{0, 0.144\} = 0.144 \qquad (2.2.82)$$

$$\hat{q}_2 = \max\left\{0, \frac{1}{2\log(2)} \times \log\left(\frac{3\cdot(747-61) - 4\cdot 427 + 186}{3\cdot(747-61) - 4\cdot 565 + 427}\right) - \frac{1}{2}\right\}$$

$$\hat{q}_2 = \max\{0, 0.126\} = 0.126 \qquad (2.2.83)$$

$$\hat{q}_3 = \max\left\{0, \frac{1}{2\log(2)} \times \log\left(\frac{3\cdot(731-57) - 4\cdot 382 + 85}{3\cdot(731-57) - 4\cdot 531 + 382}\right) - \frac{1}{2}\right\}$$

$$\hat{q}_3 = \max\{0, 0.024\} = 0.024 \qquad (2.2.84)$$

$$\hat{q}_4 = \max\left\{0, \frac{1}{2\log(2)} \times \log\left(\frac{3\cdot(546-42) - 4\cdot 184 + 0}{3\cdot(546-42) - 4\cdot 403 + 184}\right) - \frac{1}{2}\right\}$$

$$\hat{q}_4 = \max\{0, 1.104\} = 1.104 \qquad (2.2.85)$$

One of the stipulations for this methodology is that $\hat{q}_i \epsilon [0,1]$ therefore, for this reason, the result in Equation (2.2.85) should be modified so that:

$$\hat{q}_4 = \min\{1, 1.104\} = 1 \qquad (2.2.86)$$

A number of other components are necessary to be able to calculate $\alpha(\hat{q}_i)$, including $\Gamma(2\hat{q}_i + 2)$ and $\zeta(2\hat{q}_i + 2)$ where $\Gamma(x)$ is the gamma distribution as defined in Equation (2.2.23) and $\zeta(x)$ is the Riemann-zeta function as defined in Equation (2.2.24). These components can be seen in Table 2.5.

Orientation (i)	$\hat{q}_i$	$\Gamma(2\hat{q}_i+2)$	$\zeta(2\hat{q}_i+2)$	$\cos(\pi\hat{q}_i)$
1	0.144	1.158	1.439	0.899
2	0.126	1.134	1.459	0.923
3	0.024	1.021	1.602	0.997
4	1	6.000	1.082	-1.000

Table 2.5: *Summary of the components necessary to calculate $\alpha(\hat{q}_i)$ for each orientation i.*

Using the results in Table 2.5 and substituting them into Equation (2.2.69) gives:

$$\alpha(\hat{q}_1) = [1.158 \times 1.439 \times 0.899]/[(2\pi)^{2.288}.(1 - 2^{-0.712})]$$

$$\alpha(\hat{q}_1) = 0.0574 \tag{2.2.87}$$

$$\alpha(\hat{q}_2) = [1.134 \times 1.459 \times 0.923]/[(2\pi)^{2.252}.(1 - 2^{-0.748})]$$

$$\alpha(\hat{q}_2) = 0.0602 \tag{2.2.88}$$

$$\alpha(\hat{q}_3) = [1.021 \times 1.602 \times 0.997]/[(2\pi)^{2.048}.(1 - 2^{-0.952})]$$

$$\alpha(\hat{q}_3) = 0.0783 \tag{2.2.89}$$

$$\alpha(\hat{q}_4) = [6 \times 1.082 \times -1]/[(2\pi)^{4}.(1 - 2^{1})]$$

$$\alpha(\hat{q}_4) = 0.0042 \tag{2.2.90}$$

Now, $\hat{\sigma}_i^2$ can be calculated for each orientation i from Equation (2.2.66):

$$\hat{\sigma}_1^2 = \left(3(2/\pi)\right)^2 \times 1^2 \times 0.0574 \times (3(487 - 47) - (4 \times 363) + 263) + \left(\left(3(2/\pi)\right)^2 \times 1^2 \times 47\right)$$

$$\hat{\sigma}_1^2 = 198.9\text{mm}^4 \tag{2.2.91}$$

$$\hat{\sigma}_2^2 = \left(3(2/\pi)\right)^2 \times 1^2 \times 0.0602 \times (3(747 - 61) - (4 \times 565) + 427) + \left(\left(3(2/\pi)\right)^2 \times 1^2 \times 61\right)$$

$$\hat{\sigma}_2^2 = 271.9\text{mm}^4 \tag{2.2.92}$$

$$\hat{\sigma}_3^2 = \left(3(2/\pi)\right)^2 \times 1^2 \times 0.0783 \times (3(731 - 57) - (4 \times 531) + 382) + \left(\left(3(2/\pi)\right)^2 \times 1^2 \times 57\right)$$

$$\hat{\sigma}_3^2 = 287.9\text{mm}^4 \tag{2.2.93}$$

$$\hat{\sigma}_4^2 = \left(3(2/\pi)\right)^2 \times 1^2 \times 0.0042 \times (3(546 - 42) - (4 \times 403) + 184) + \left(\left(3(2/\pi)\right)^2 \times 1^2 \times 42\right)$$

$$\hat{\sigma}_4^2 = 154.5\text{mm}^4 \tag{2.2.94}$$

Thus $\hat{v}_w$ can be calculated as in Equation (2.2.65):

$$\hat{v}_w = 198.9 + 271.9 + 287.9 + 154.5 = 913.1\text{mm}^4 \tag{2.2.95}$$

and hence, the variance due to Cavalieri sampling can be estimated from Equation (2.2.64):

$$\text{var}_{\text{cav}}(\tilde{S}_w) = (\pi/4)^2 \times 913.1 = 563.3\text{mm}^4 \tag{2.2.96}$$

Using both the variance due to Cavalieri sections in Equation (2.2.96) and the surface area estimate in Equation (2.2.74), the coefficient of error due to Cavalieri sections can be calculated from Equation (2.2.9):

$$\text{ce}_{\text{cav}}(\tilde{S}_w) = (\sqrt{563.3}/310.5) \times 100 = 7.6\% \tag{2.2.97}$$

Coefficient of error due to the first level of sampling

To calculate the coefficient of error due to the first level of sampling, the variance component of the total variance of the surface area estimator, $\text{var}_m(\hat{S}_w)$ from Equation (2.2.56) needs to be calculated. To do this $\hat{C}_0$, $\hat{C}_1$ and $\hat{C}_2$ are also to be calculated, but first the function $\hat{f}_i$, for each orientation i (see Equation (2.2.50)) are demanded:

$$\hat{f}_1 = \frac{2}{\pi} \times 3 \times 1 \times 47 = 89.76 \tag{2.2.98}$$

$$\hat{f}_2 = \frac{2}{\pi} \times 3 \times 1 \times 61 = 116.50 \tag{2.2.99}$$

$$\hat{f}_3 = \frac{2}{\pi} \times 3 \times 1 \times 57 = 108.86 \tag{2.2.100}$$

$$\hat{f}_4 = \frac{2}{\pi} \times 3 \times 1 \times 42 = 80.21 \tag{2.2.101}$$

Applying Equations (2.2.98)-(2.2.101) to Equation (2.2.57) gives:

$$\hat{C}_0 = \sum_{i=1}^{w} \hat{f}_i \hat{f}_i = 89.76^2 + 116.5^2 + 108.86^2 + 80.21^2$$

$$\hat{C}_0 = 39913 \tag{2.2.102}$$

$$\hat{C}_1 = \sum_{i=1}^{w} \hat{f}_i \hat{f}_{i+1}$$

$$= (89.76 \times 116.5) + (116.5 \times 108.86) + (108.86 \times 80.21) + (80.21 \times 89.76)$$

$$\hat{C}_1 = 39071 \tag{2.2.103}$$

$$\hat{C}_2 = \sum_{i=1}^{w} \hat{f}_i \hat{f}_{i+2}$$

$$= (89.76 \times 108.86) + (116.5 \times 80.21) + (108.86 \times 89.76) + (80.21 \times 116.5)$$

$$\hat{C}_2 = 38231 \qquad (2.2.104)$$

Following this, the value of m is to be determined from $\widehat{D}_0$ and $\widehat{D}_1$ (see Equations (2.2.60) and (2.2.61), respectively):

$$\widehat{D}_0 = (39913 - 38231 - 913.1)/(39913 - 39071 - 913.1) - 4/3$$

$$\widehat{D}_0 = -12.1 \qquad (2.2.105)$$

$$\widehat{D}_1 = (39913 - 38231 - 913.1)/(39913 - 39071 - 913.1) - 16/9$$

$$\widehat{D}_1 = -12.6 \qquad (2.2.106)$$

Since $|\widehat{D}_0| \leq |\widehat{D}_1|$ then $m = 0$. Therefore, the variance due to the orientations from Equation (2.2.62) can be written as:

$$\mathrm{var}_0(\hat{S}_w) = \max\left\{0, \frac{\pi(\pi/4)}{18}(39913 - 39071 - 913.1)\right\}$$

$$\mathrm{var}_0(\hat{S}_w) = \max\{0, -9.75\} = 0 \qquad (2.2.107)$$

Hence, the coefficient of error due to orientations can be stated to be (again from Equation (2.2.9)):

$$\mathrm{ce}_\phi(\tilde{S}_w) = (0/310.5) \times 100 = 0\% \qquad (2.2.108)$$

This value shows the limitations of the error predictor formulae when applied to particular cases, and the need for further improvement. The total variance within the surface area estimate can now be estimated from Equation (2.2.52):

$$\mathrm{var}(\tilde{S}_w) = 0 + 563.3 + 465.8 = 1029.1\mathrm{mm}^4 \qquad (2.2.109)$$

Finally the total coefficient of error due to the three levels of sampling (orientations, Cavalieri sections and cycloid grid sampling simultaneously) can be estimated by substituting Equation (2.2.109) and (2.2.74) into Equation (2.2.9) giving:

$$\mathrm{ce}(\tilde{S}_w) = \left(\sqrt{1029.1}/310.5\right) \times 100 = 10.3\% \qquad (2.2.110)$$

2.3 Image analysis

The volume and surface area estimation methods described in Section 2.2, can be applied to magnetic resonance (MR) images (e.g. Roberts *et al.*, 2000; Barta & Dazzan, 2003; Ronan *et al.*, 2006; Sharpe *et al.*, 2009; Acer *et al.*, 2010). In this next section, the application of the stereological estimation methods to magnetic resonance imaging (MRI) is described and the programs used are stated at each step. This section focuses on Broca's area and starts with the pre-processing of the images, followed by volume and then surface area estimation methods.

2.3.1 Preparation of the images

The purpose of the pre-processing of the images is to transform the images into a format that is compatible with the program used to perform the volume and surface area estimation. The information of the MR images taken from the scanner are stored in two files: an image file (.img) and a header file (.hdr). The images themselves are stored in the image file, whereas the role of the header file is to give information to a program attempting to accessing the image data, about how to read the image file correctly.

The first program used is called BrainVoyager QX (version 1.9). The main uses of this program are to transform the images into the correct file format, voxel size and orientation. The first step is to open BrainVoyager and then start a 'new project'. The new project will be in the form of an anatomical 3-D dataset and will be saved as a .vmr file. The data type is also set to 'analyze'. BrainVoyager can open the header and image files of the raw MR image data when we set the file type to 'analyze'. The contrast and brightness are then adjusted such that not only are the grey and white matter distinguishable from each other, but that the boundary between the brain and the cerebrospinal fluid (CSF) is also easily distinguishable. Since this varies from brain to brain, and requires individual adjustment, possible bias may be introduced. This contrast and brightness bias may add to the potential bias of overprojection in the stereological methodology mentioned in Section 2.2.3. The bias due to overprojection would be affected because if it is difficult to observe the edges of the region of interest (e.g. due to poor contrast/brightness levels) then it would also be difficult to choose a value of Cavalieri section thickness, T, which would guarantee that the number of sections within the region of interest, M, is at least equal to 10.

Following this step will lead to being asked whether one wishes to iso-voxel the data or not. The answer should always be 'yes' if the dimensions of the voxels in the images obtained from the scanner are not 1mm x 1mm x 1mm. This is because by having voxels with differing size dimensions, the output sections and images would be skewed and would introduce bias to estimations in favour of the dimension with the largest length. At present, to obtain high quality MR images, the maximum true resolution should be no more than $1mm^3$ (see BrainVoyager QX User's Guide - Version 0.9.8 at http://www.brainvoyager.com/bvqx/doc/UsersGuide/WebHelp/BrainVoyagerQXUsersGuide.htm).

Some tools in BrainVoyager also need a resolution of $1mm^3$ to function without further extensive editing of the MR images. Smaller cubed voxel sizes than 1mm x 1mm x 1mm could be used, but in this case voxels of size $1mm^3$ are used as the standard. The 'target voxel size' should be set to 1.0 (X), 1.0 (Y) and 1.0 (Z), to represent this standard voxel size

The iso-voxel transformation should be completed using sinc interpolation. Sinc interpolation for signal reconstruction is defined as:

$$x(\tilde{t}) = \sum_{k=-\infty}^{\infty} x_{\tilde{n}} \cdot \text{sinc}\left(\left(\pi/\tilde{T}\right)\left(\tilde{t} - \tilde{n}\tilde{T}\right)\right) \qquad (2.3.1)$$

where $x(\tilde{t})$ is the reconstructed signal, $\tilde{T}$ is the sampling period used to identify the sample $x_{\tilde{n}}$ from the original signal and $\mathrm{sinc}\left((\pi/\tilde{T})(\tilde{t}-\tilde{n}\tilde{T})\right)$ can be written as:

$$\mathrm{sinc}\left((\pi/\tilde{T})(\tilde{t}-\tilde{n}\tilde{T})\right) = \sin\left((\pi/\tilde{T})(\tilde{t}-\tilde{n}\tilde{T})\right)/(\pi/\tilde{T})(\tilde{t}-\tilde{n}\tilde{T}) \qquad (2.3.2)$$

(Oppenheim & Schafer, 1975). This type of interpolation is used because it is useful for retaining a higher proportion of data quality than the trilinear interpolation and is faster due to requiring less working memory than the cubic spline interpolation. The newly transformed image file should then be saved. At this stage the window shown by BrainVoyager should look like Figure 2.9. It is useful to note that in Figure 2.9 the standard orientation is always in the upper left hand corner of the brain views in BrainVoyager. This is the sagittal view through the images. In this case the view is upside down, so this needs to be transformed such that the neck and spinal column are located at the bottom of the screen. The view in both the coronal and transverse screens also needs to be transformed. Currently, in Figure 2.9, these can be seen to be posterior to anterior in direction, from left to right, when it should be viewed as anterior to posterior. To perform these spatial transformations, the dialogue box should be opened (by pressing on the 'Full Dialog>>' button in the '3D Volume Tools' window). From this box the spatial transformation tab (called 'Spatial Transf') should be opened and beneath the 'Standardize' menu, the 'To Sag...' button is then pressed. This brings up the 'transform to standard orientation' window. Any transformations referred to in this window are described for the sagittal view, but will affect both the coronal and transversal views simultaneously. To get the desired transformations, the images must be rotated by 90 degrees anti-clockwise in the y-axis (+90), twice, followed by a rotation of 90 degrees clockwise in the z-axis (-90). The resultant transformed images can be seen in Figure 2.10.

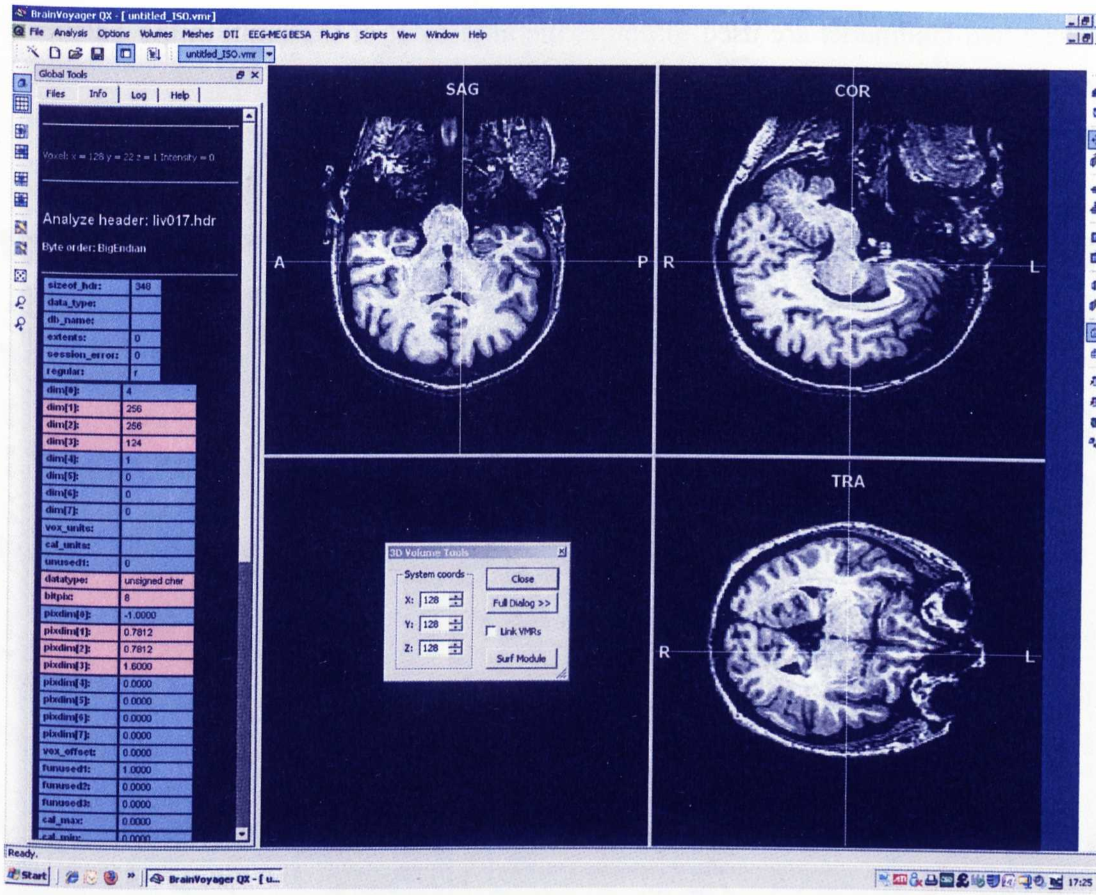


Figure 2.9: BrainVoyager screen after iso-voxel transforming the images to 1mm × 1mm × 1mm.

The images in Figure 2.10 are still not ready to be used in the programs to be used later, including the Stereology application program. This is because they have not been anterior commissure-posterior commissure (AC-PC) corrected. That is, they have not been corrected for the positioning of the head in the scanning itself. Tilting or slight rotating of the head, in the vertical or horizontal planes may have occurred. In order for identification of regions of interest to be easier and for correct use of other programs, this must be corrected to guarantee that there is a standard positioning of all of the individual brains. The anterior commissure (AC) is a grouping of nerve fibres that bridges between the two hemispheres just in front of the columns of the fornix. The posterior commissure (PC) is a band of white fibres crossing between the two hemispheres at the upper end of the cerebral aqueduct.

These two landmarks are used to define the alignment of the brain in the Talairach Atlas (Talairach & Tournoux, 1988).

__Figure 2.10:__ BrainVoyager screen after spatial transformation of the images.

An illustration of the AC-PC line can be seen in Figure 2.11. In BrainVoyager, the AC is needed to be identified. This is done by altering the co-ordinate values for the three planes (x-axis, y-axis and z-axis) until the co-ordinate which corresponds to the AC point. Figure 2.12 shows BrainVoyager with the co-ordinates of the AC point for the example brain. The AC point is the point in the centre of the crosses in each of the three views (sagittal, coronal and transversal).

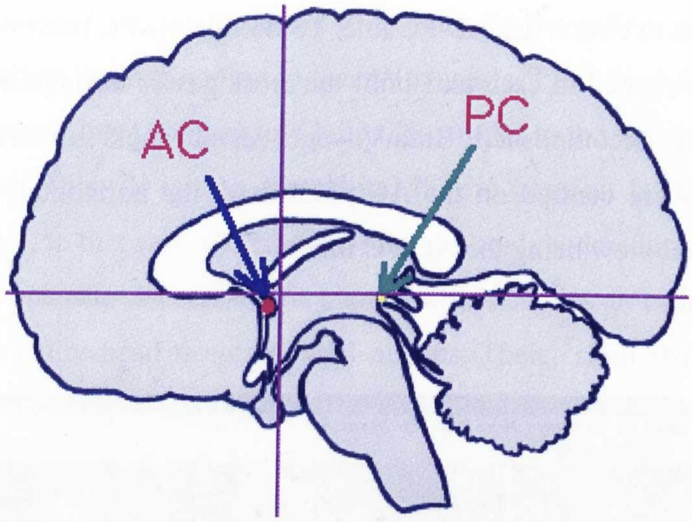


Figure 2.11: *Illustration of the AC-PC line* (http://airto.ccn.ucla.edu/BMCweb/HowTo/AC-PC.html (downloaded in February 2011)).

Figure 2.12: *BrainVoyager screen after the AC point has been identified and highlighted.*

The next stage is to obtain the AC-PC line. To do this the PC must be identified and then the images rotated in each axis until the cross passes through both the AC and PC. When this is accomplished, BrainVoyager should look as in Figure 2.13. The green crosshairs are centred on the AC point, with the horizontal, green crosshair line in the sagittal view being the AC-PC line.

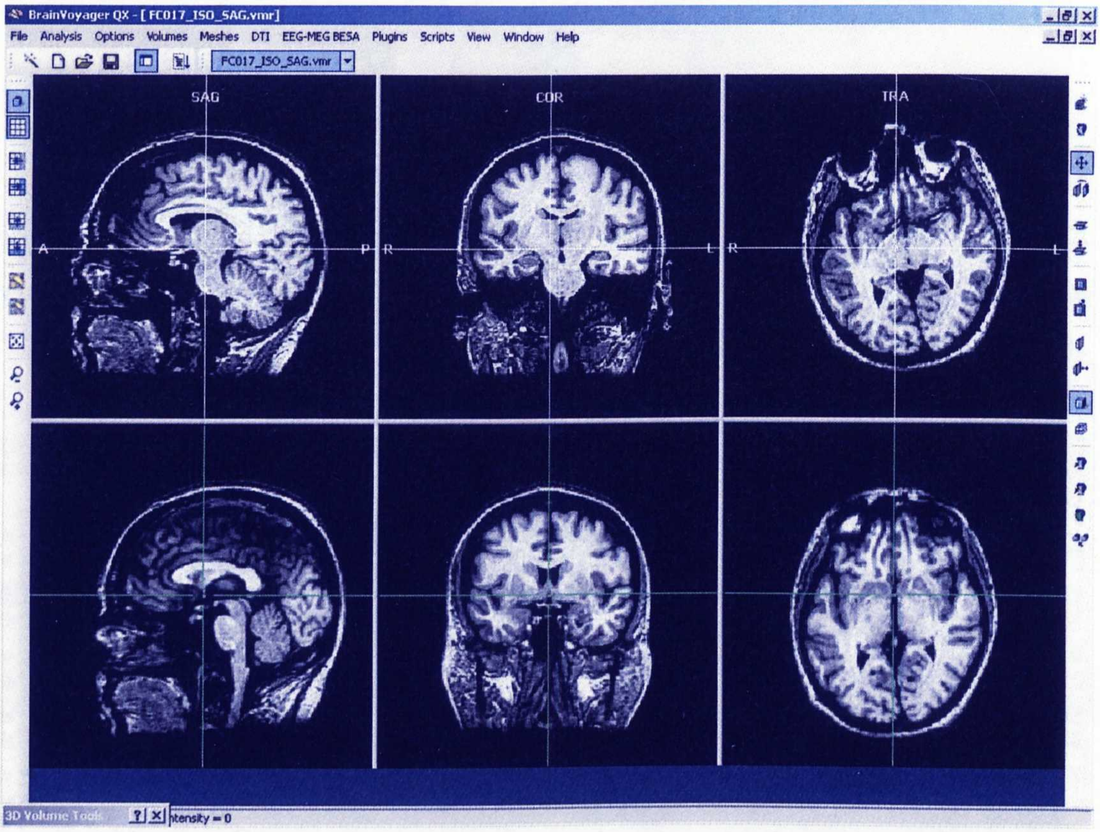


Figure 2.13: *BrainVoyager screen after the AC-PC line has been identified and highlighted (green, horizontal crosshair line, bottom row, left hand panel) with the original uncorrected views on the top row and the AC-PC corrected views on the bottom row.*

At this point the original image files need to be transformed with the AC-PC corrected rotations. Similarly as described above, a sinc interpolation transformation is used and the new image file is saved (as .vmr format). Following this step the AC-PC corrected views are visible in BrainVoyager as can be seen in Figure 2.14.

Subsequently, the .vmr file is exported to the file format called analyze using the 'Export To Analyze Format' option in BrainVoyager's 'File' drop-down menu. This generates an image (.img) and header (.hdr) file, similar to the raw MRI data derived from the scanner. This file is then opened in MRICro (version 1.40, build 1), which is a program used, in this case, to create a 3-dimensional rendered image of the brain. Once the file is opened, '3-D render' is chosen which creates an estimation of the 3-D image of the entire head from the MRI images. Then, 'skull strip' is completed which removes anything which MRICro identifies as being skin and bone, leaving nothing but the 3-D rendered brain itself. This can then be rotated around to highlight, for example, the left and right hemisphere as shown in Figure 2.15. These 3-D rendered images are then used to identify the region of interest. It is useful to note here, that the AC-PC corrected images that are obtained from BrainVoyager, and subsequently the 3-D rendered images that are given in MRICro, are mirrored images. That is, what appears to be the left hemisphere in BrainVoyager and MRICro is, in fact, the right hemisphere, and vice versa.

Now the region of interest can be demarcated and estimated. The next subsection covers the identification of the region of interest, which in this case is Broca's area.

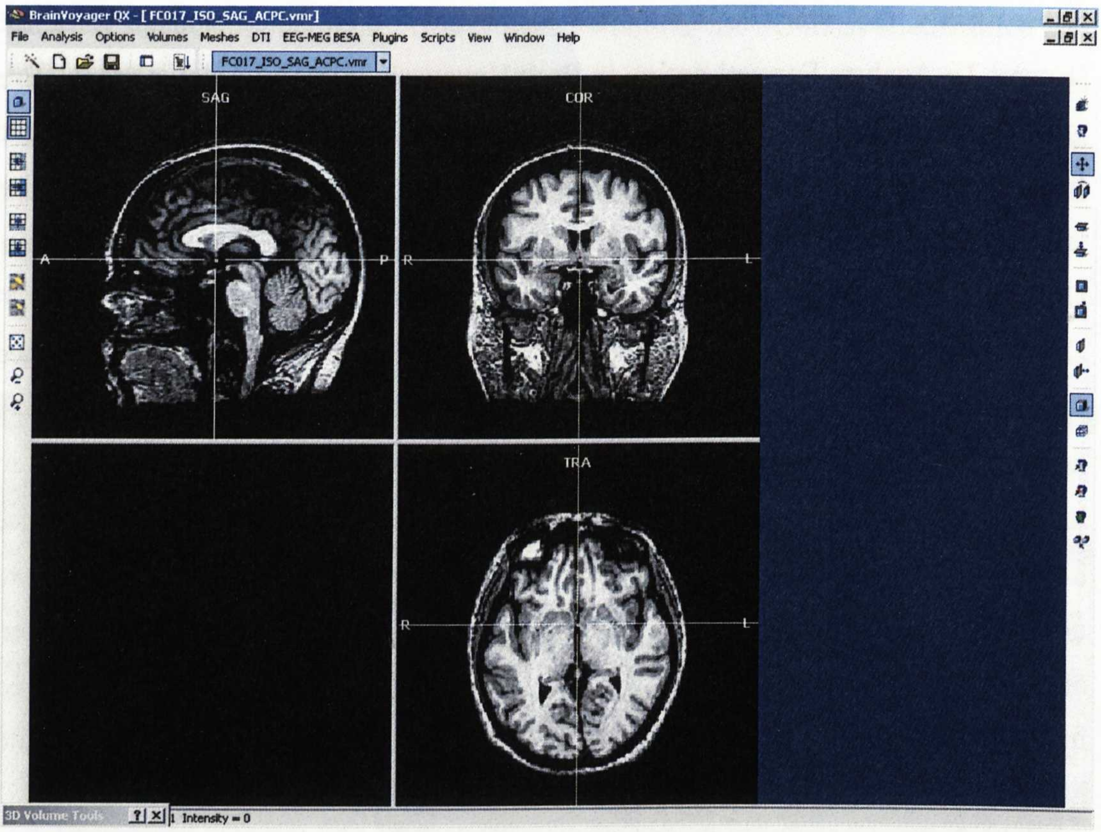
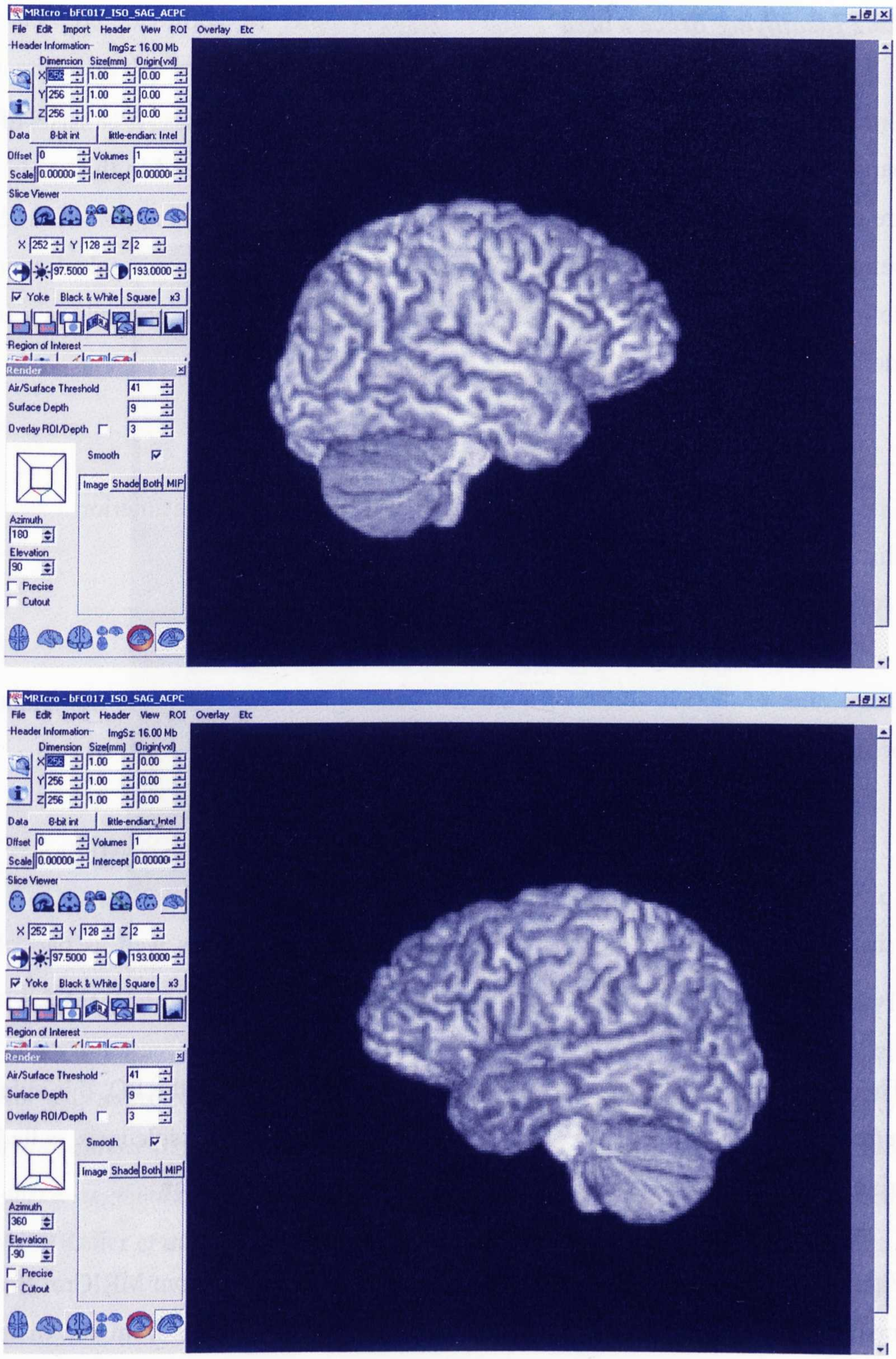


Figure 2.14: *BrainVoyager screen showing the AC-PC corrected sagittal, coronal and transversal views.*



(i)

(ii)

Figure 2.15: *MRICro screens showing 3-D rendered images of the (i) left hemisphere and (ii) right hemisphere of the example brain.*

2.3.2 Identifying Broca's area

Broca's area is a region of the brain located in the cerebral cortex. It consists of two parts: the pars opercularis (PO) and the pars triangularis (PT) and in this study they are estimated separately. Paul Broca was the first to identify this region of the brain being associated with language (Dronkers *et al.*, 2007). The major sulcal contours defining PO and PT can be seen in Figure 2.16. The pars opercularis can be identified since it is bounded by the following sulcal contours:

i. The inferior prefrontal sulcus to the posterior.
ii. The anterior ascending ramus of the Sylvian fissure to the anterior.
iii. The inferior frontal sulcus to the superior.
iv. The Sylvian fissure itself to the inferior.

(see Keller *et al.*, (2007) for a more detailed description). The pars triangularis is bounded by:

i. The anterior ascending ramus of the Sylvian fissure to the posterior.
ii. The inferior frontal sulcus to the superior and anterior.
iii. The anterior horizontal ramus of the Sylvian fissure as well as part of the Sylvian fissure itself to the inferior.

There may exist an extra sulcus within both the PO and PT. In the PO, this sulcus is referred to as the diagonal sulcus, whereas in the PT, this sulcus is called the triangular sulcus. However, this is not always found in every brain.

Figure 2.16 shows an image similar to that which is obtained from MRICro. Broca's area can be identified more clearly in this view than directly from the magnetic resonance (MR) images in BrainVoyager. However, this viewpoint only allows the surface of Broca's area to be identified. It is not possible to tell how deep into the brain the region of interest (RoI) actually goes, nor does it say whether the shape of the RoI continues exactly the same as it does at the surface, throughout the entire depth of the region. BrainVoyager must be used then to answer these queries. Using the MRICro 3-D rendered images and the area identified as Broca's area, this area can also be identified in BrainVoyager. Figure 2.17 shows the Broca's area demarcated in the brain (one hemisphere only) and three views from BrainVoyager (one sagittal and two coronal) as well as the MRICro view.

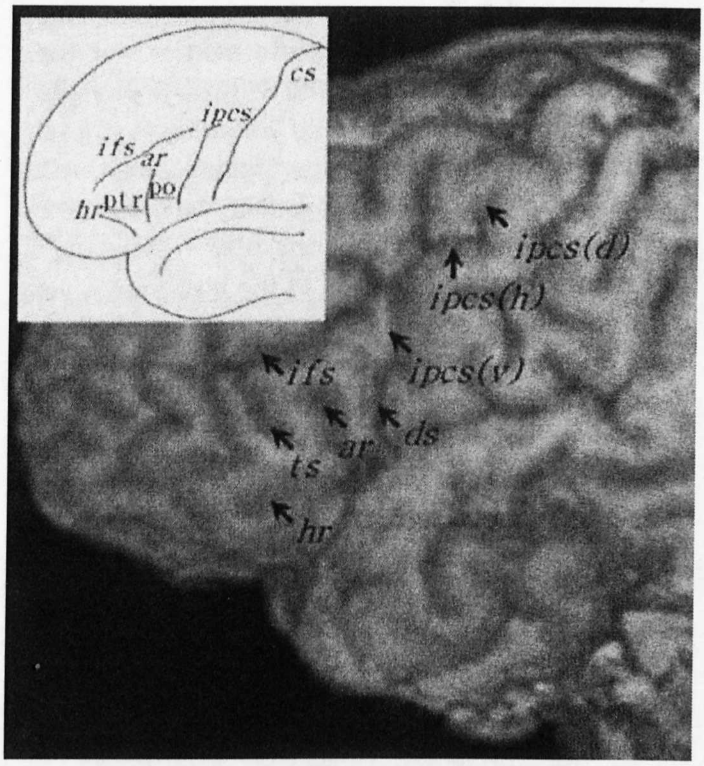


Figure 2.16: *The major sulcul contours defining Broca's area: Pars Opercularis (po); Pars Triangularis (ptr) with the horizontal ramus of the Sylvian fissure (hr), anterior ascending ramus of the Sylvian fissure (ar), the inferior frontal sulcus (ifs), triangular sulcus (ts), diagonal sulcus (ds) and the inferior precentral sulcus split into three segments: ventral (ipcs(v)); horizontal (icps(h)); and dorsal vertical (icps(d))* (Keller *et al.*, 2007).

Now that Broca's area has been identified the next step is to demarcate this region (the RoI). This will be explained in the next two subsections along with the estimation methods for volume and surface area (the demarcation needed is different for each method).

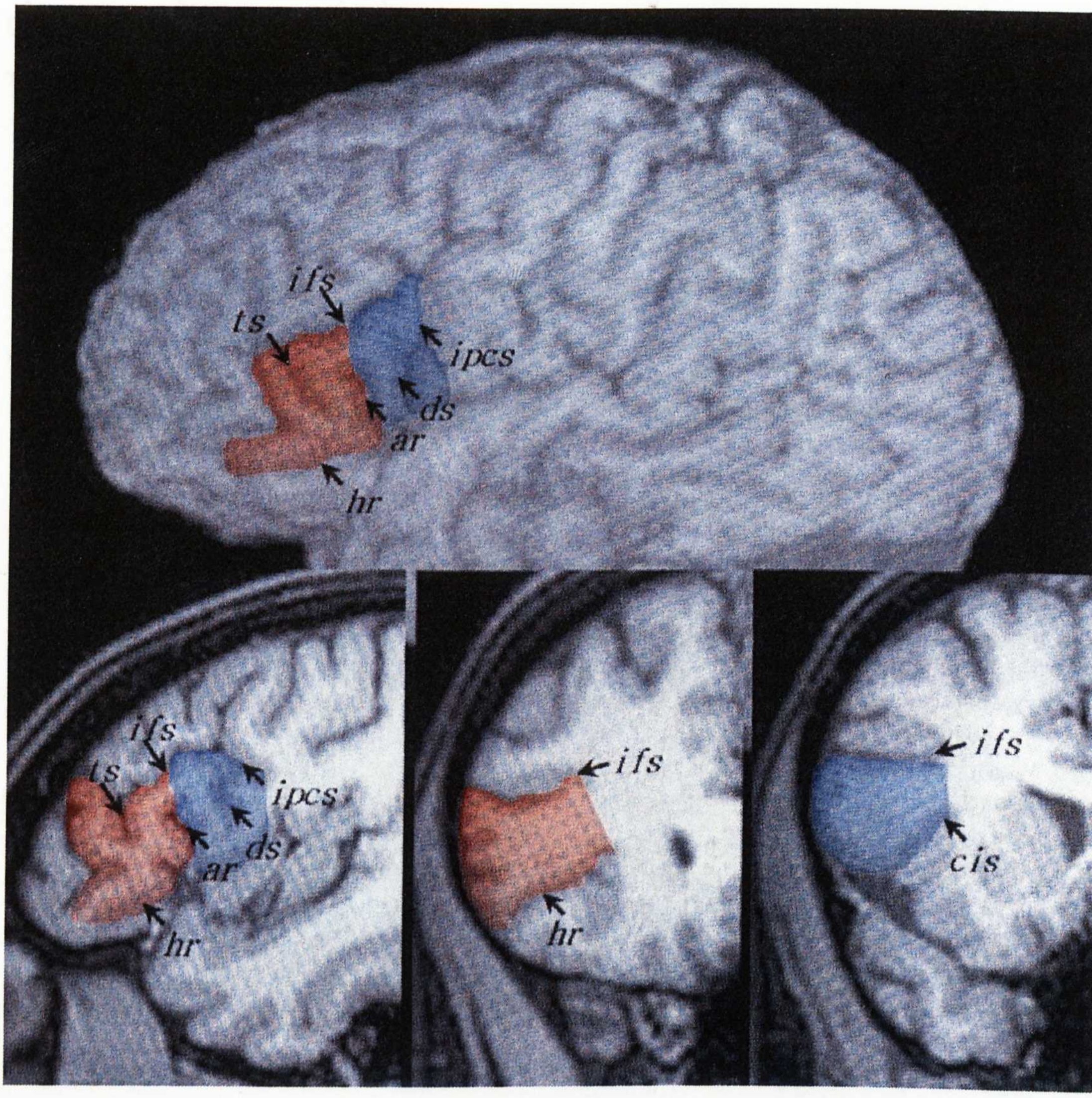


Figure 2.17: *Broca's area demarcated (Pars Opercularis (blue) and Pars Triangularis (red)) in the MRICro 3-D rendered image (top) and in three 2-D views from BrainVoyager (a sagittal view (bottom left), and two coronal views (bottom middle and right). The anatomy is here defined as follows: inferior precentral sulcus (ipcs); inferior frontal sulcus (ifs); diagonal sulcus (dr); triangular sulcus (tr); horizontal ramus of the Sylvian fissure (hr); anterior ascending ramus of the Sylvian fissure (ar); and circular insular sulcus (cis)* (Keller *et al.*, 2007).

2.3.3 Estimation of the volume of Broca's area

For stereological estimation it is necessary to be able to clearly distinguish the region under study. In this case the RoI, the Broca's area, needs to be demarcated such that it can be clearly identified during the sampling process. Since one of our main aims when estimating volume and surface area is to minimise the amount of variability due to errors, then a precise demarcation process is needed. Broca's area was demarcated using BrainVoyager and allowing the outlines to lie outside the region by at least one or two voxels.

The first task is to demarcate the anterior and posterior ends of Pars Opercularis. Pars Opercularis and Pars Triangularis are demarcated and estimated separately as has been previously mentioned. In the segmentation option within BrainVoyager there is an option to draw with the mouse on the images. A very large size is needed here for the line thickness and a size of around 30mm is useful to demarcate the end points of PO. The reason for this is so that the entire hemisphere, on one section, can be demarcated with one click of the left-hand mouse button.

After scrolling through the images, the coronal view should be set such that the particular coronal slice is at least one or two pixels outside of PO at the posterior end, throughout the entire depth of the region. Then a mark is made in BrainVoyager such that it encompasses the whole hemisphere. This is repeated for the anterior end of PO. Now the size of the mark is reduced to around size 2mm from 30mm, and manually, in the sagittal view on one section, PO is drawn around keeping at least two voxels outside of the region on the superior side. The reason for this reduction in line thickness is that since for volume estimation the region can clearly be highlighted when demarcated on only one sagittal section, and this smaller line thickness allows the region to be drawn around more precisely.

This is repeated for the inferior side of the region. The views from BrainVoyager should then be as in Figure 2.18. This process is then repeated for the pars triangularis, followed by the PO and PT of the opposite hemisphere. The output file from BrainVoyager contains the images as a series of sagittal slices. These images are also in the wrong orientation, with the lower part of the head at the top of the screen and the upper part of the head at the bottom. This can be seen when opening the image file in ImageJ (see Figure 2.19).

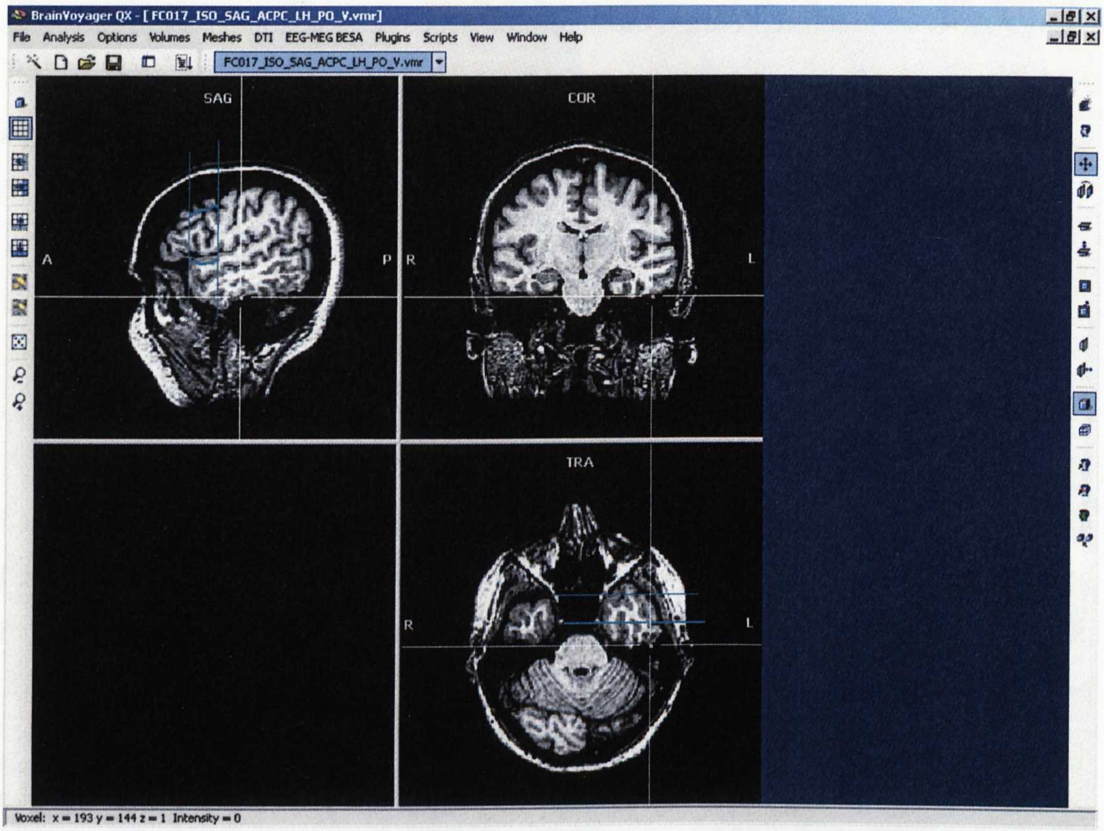


Figure 2.18: *BrainVoyager screen showing the demarcated example brain.*

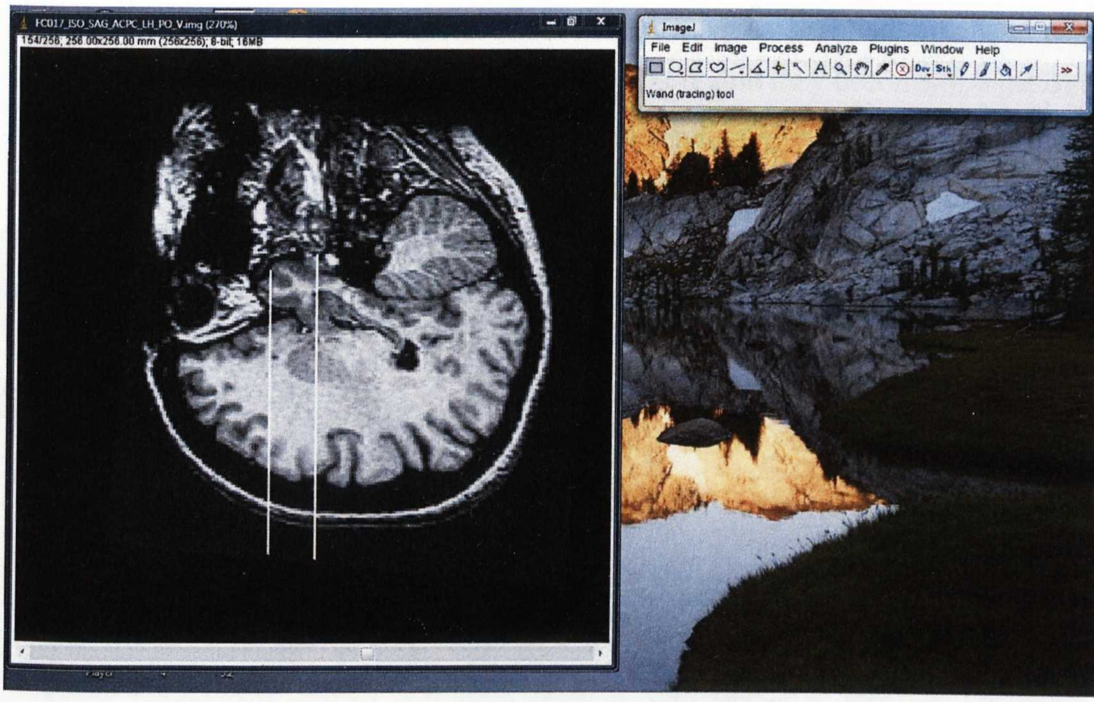


Figure 2.19: *ImageJ view of the demarcated example brain once opened directly from BrainVoyager.*

ImageJ allows for the output image file of the demarcated brains from BrainVoyager to be changed. Using the add-on plug-in ‘Align Stacks’, the image file can be changed to coronal sections from sagittal sections. As it was noted earlier the image files in BrainVoyager are mirror images of reality in terms of hemispheres. That is, the left hemisphere in BrainVoyager is actually the subject’s right hemisphere, and vice versa. This can be corrected in ImageJ by choosing the ‘sagittal (right-to-left)’ option in ‘To CoronalTP’. While also performing the conversion from sagittal to coronal slices, it also flips the images such that the left and right hemispheres are on the left and right hand side of the images, respectively. Finally, in ImageJ by selecting ‘Flip Vertical’ the images are flipped vertically so that the orientation is correct with the upper part of the head at the top of the viewing screen and the lower part of the head at the bottom. The ImageJ screen should now look as that in Figure 2.20. The file should be saved in analyze 7.5 format (this is the .hdr and .img file combination). These steps in ImageJ are needed to be repeated for the image files with demarcated PO and PT for each hemisphere. The output files from ImageJ are now ready to be used to perform the estimations in the stereological program EasyMeasure.

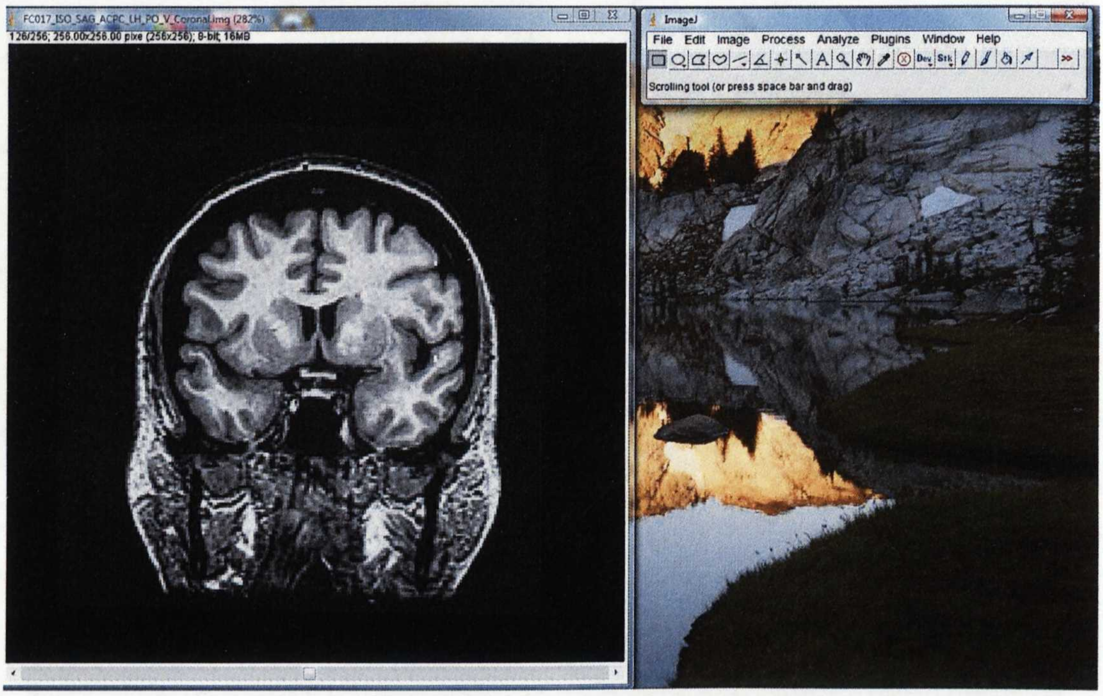


Figure 2.20: *ImageJ view of the demarcated example brain after editing in ImageJ.*

EasyMeasure is the program used to perform the stereological estimation methods needed in order to obtain both volume and surface area estimates. To estimate volume, the image file obtained from ImageJ is opened and 'Volume' is chosen from the 'Run Stereology' menu. It is then possible to change the grid size and the distance between consecutive sections (specified, respectively by u and T in Equation (2.2.12)). A random starting section between 1 and T is then generated in EasyMeasure when 'Run' is pressed and the Cavalieri set of images is selected. A random positioning of the grid is then generated on each section. Figure 2.21 shows an example of a section with random grid positioning in EasyMeasure. The crosses, where the upper right hand corner lies entirely within the region of interest are then removed by clicking on them. This is repeated for each structure as defined earlier (e.g. white matter and grey matter). It is useful to note that once points have been removed for one region (e.g. white matter), the points show up when a subsequent region is chosen as different coloured points. Normally the points are red, but after they have been removed for one region, for the next region they become orange. This is so that experimenters do not include the same points in multiple regions and is especially useful for points which are difficult to identify as to which region they lie in. This can also be seen in Figure 2.21 where the grey matter points have been removed and the white matter points (already removed in the white matter view) show as orange crosses. This point removal is continued for every slice that is within the demarcated area. The number of points that have been removed for each region on each section is automatically recorded in a Microsoft Excel compatible file, as well as the total number removed for each region. The estimations of volume and coefficient of error are automatically generated for each region. This process is repeated for each PO and PT of each hemisphere for each subject.

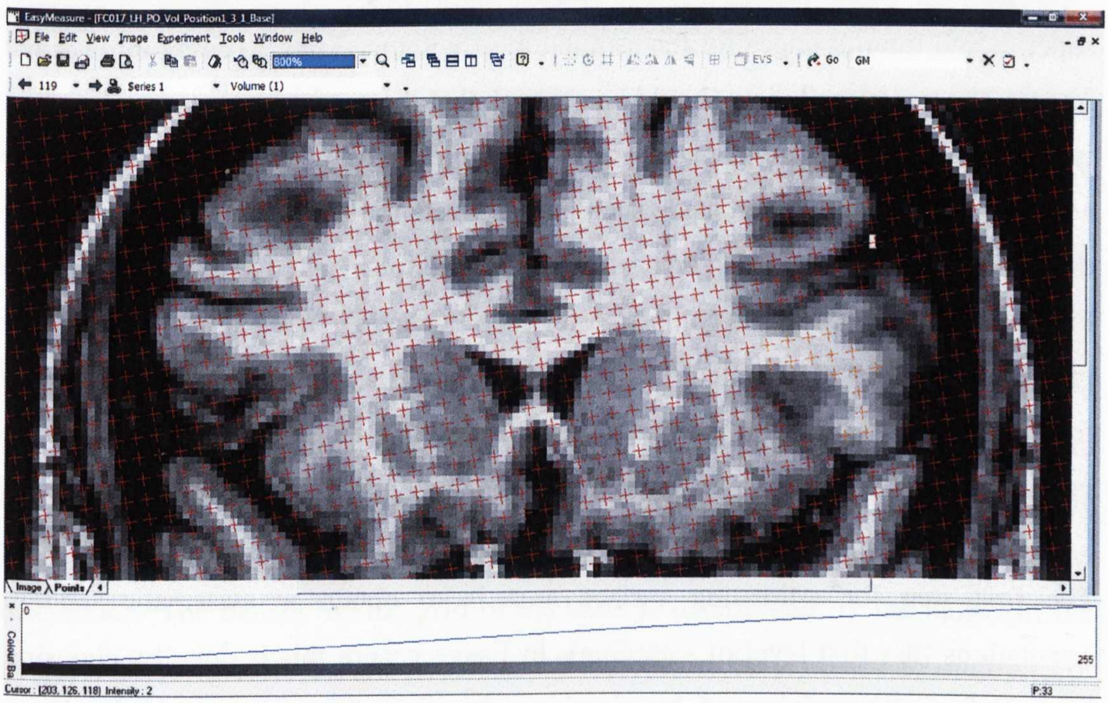


Figure 2.21: *One slice from a demarcated brain in EasyMeasure after volume stereology has been run with grid size=3 and section interval=1. The points lying within the grey matter of Broca's area have been removed and the points lying within the white matter are shown as orange crosses.*

2.3.4 Estimation of the surface area of Broca's area

Following on from the volume estimation, the first thing that needs to be done for surface area estimation is the demarcation of the region. The difference here between the demarcation for the volume estimation and that for surface area estimation is that after the posterior and anterior ends of the region of interest have been demarcated, the mark size is kept at 30mm for the demarcation of the superior and inferior ends of the RoI. The reasons for the superior and inferior sides of the region requiring the thicker line thickness is that with surface area estimation, the orientations obtained will be between the sagittal and coronal views and due to this, the resolution for sections taken from these orientations will not be as high quality as the coronal or sagittal images. This makes it far more difficult to identify the RoI and therefore, to make the demarcation more noticeable, one large thick line just outside the edge of all four sides of the RoI is preferable than the small, more precise outline used on the superior and inferior edges for the estimation of volume. After scrolling through the images, the transversal view should be set such that the particular transversal section is at least one or two pixels outside the superior-most edge of the RoI, throughout its entire depth. A mark is then made such that it covers the entire hemisphere in which the region of interest is. This is repeated for the inferior-most edge of the region (see Figure 2.22). This process is completed for each PO and PT of each hemisphere. The images then need to be rotated, flipped and converted to a coronal view in ImageJ, following exactly the same procedure as described in the previous subsection for volume estimation. Then they are ready to be opened in EasyMeasure.

The first step for estimating surface area is to obtain w systematic random orientations (the first level of sampling). In EasyMeasure this is done by choosing 'Run Reformatting' followed by changing the view from a coronal view to a transversal one. Then exhaustive vertical sections are thrown onto the image. The number of orientations, w, can be chosen, and it is called the 'number of series' in EasyMeasure (this is the number of orientations between 0 and π).

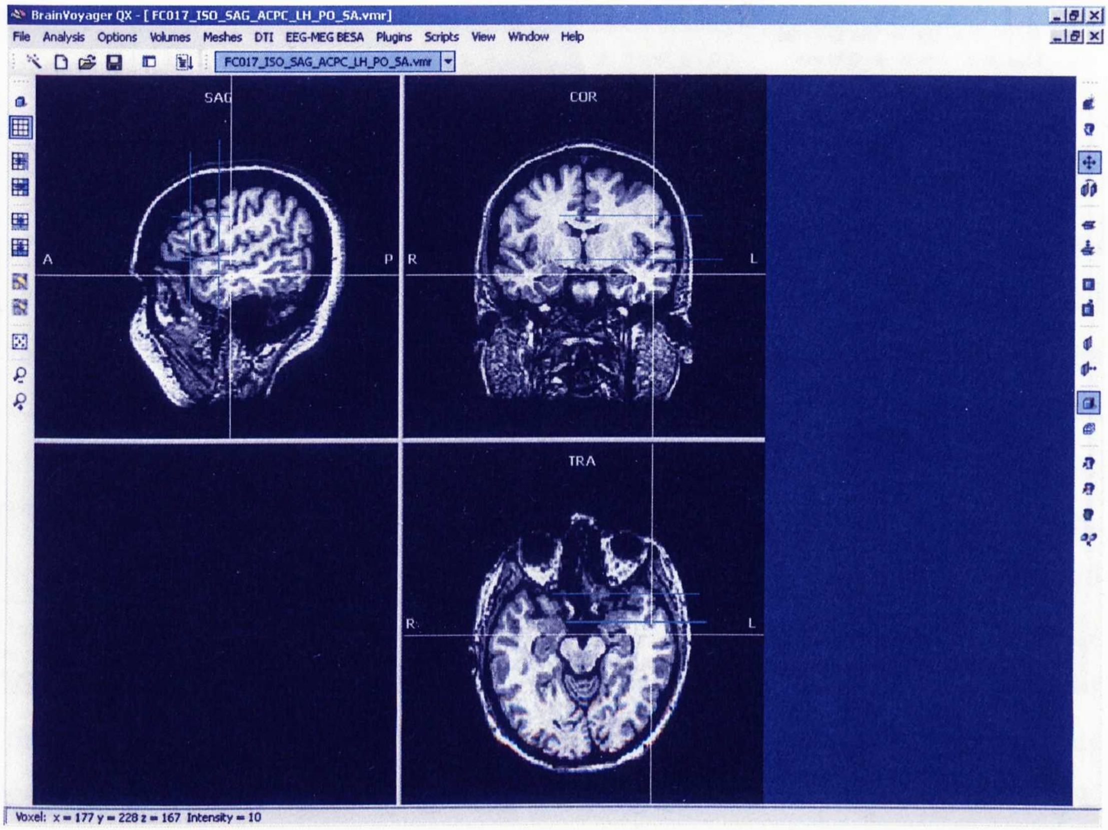


Figure 2.22: *BrainVoyager screen showing PO demarcated for surface area estimation.*

In the second level of sampling, the Cavalieri sections are generated and the distance between sections, T, can be chosen. Once these have been randomly 'thrown' onto the image, the transversal view can be seen in Figure 2.23. For the third sampling level, a grid of cycloids is randomly positioned on each section and for each orientation. The density of the grid of cycloids is determined by the radius of the cycloids (see Figure 2.7, right panel) and this needs to be provided in multiples of the distance between sections, T, specified earlier.

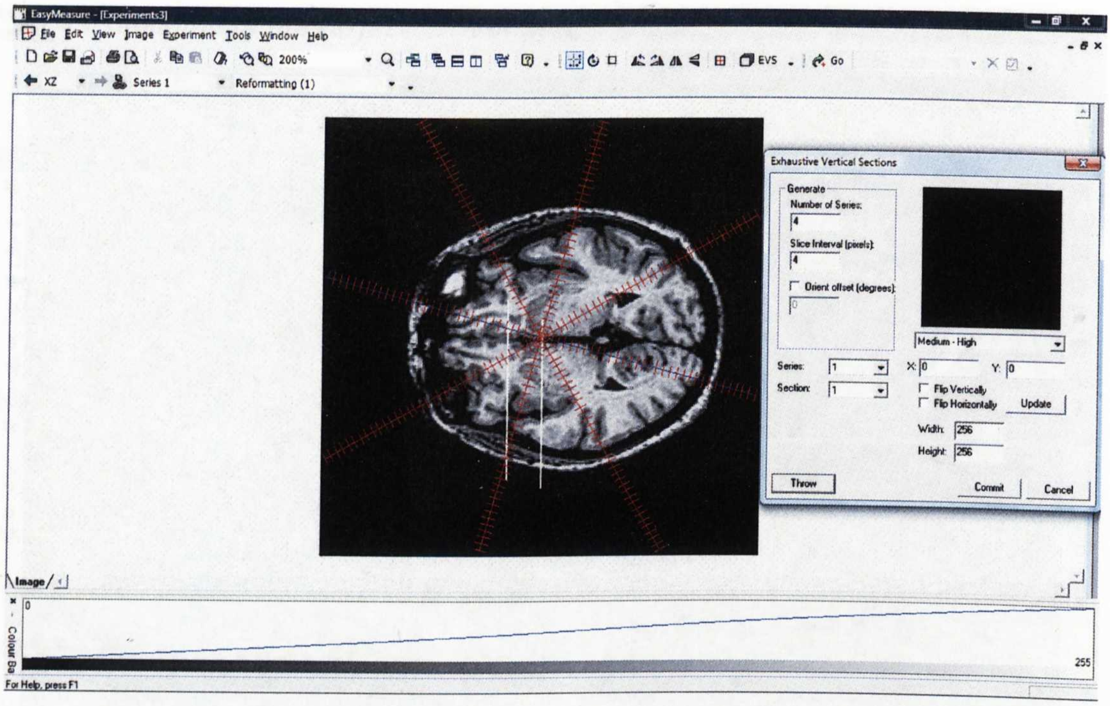


Figure 2.23: *EasyMeasure showing the two levels of systematic random sampling; 4 orientations with U_i $(i = 1,2,3,4)$ Cavalieri sections on each one.*

In order to estimate the surface area, crosses are placed on the intersections between the region of interest and the cycloids, by clicking on those points. An example of a slice with crosses on intersections is given in Figure 2.24. This is completed for all slices within the demarcated region for orientation 1. This is then repeated for orientations 2,…, w. The entire process is then repeated for PO and PT for both the left and right hemispheres. The results obtained from EasyMeasure are the total number of intersections for each orientation. From the known variables, and these results, the surface area can be estimated manually.

One of the issues surrounding this technique is that there is a loss of image resolution as a result of taking virtual sections which are not parallel to the voxel sides. Each voxel, as mentioned earlier, can be seen to be equivalent to 1mm x 1mm x 1mm cube. However, for surface area estimation, we can see that the systematically random sampled orientations, as seen in Figure 2.23, are not exactly perpendicular to either the sagittal or coronal axes throughout the RoI. The result of this is that the voxels appear to be stretched (they are no longer represented in 2D as a 1mm x 1mm square, but by a rectangular shape) which makes identifying the RoI boundaries on each Cavalieri section (e.g. in Figure 2.24) more difficult.

One possible way of overcoming this problem would be to improve the quality of the images obtained during the scanning process (e.g. by using a scanner which gives higher resolution images). This would mean that voxel sizes could be reduced thereby giving a much more precise image of the contours and boundaries of the RoI. Another possible method for reducing this virtual section image resolution effect would be to demarcate the region in three dimensions. However, to do this the RoI would have to be demarcated on every slice in every view in BrainVoyager, which could mean as many as 20 slices per RoI. With 4 RoI's per brain, this would be a very time consuming, and therefore, impractical approach using the currently available programs. Although these could be considered for future studies, neither of these methods were available, given the dataset and the numbers of individuals within it.

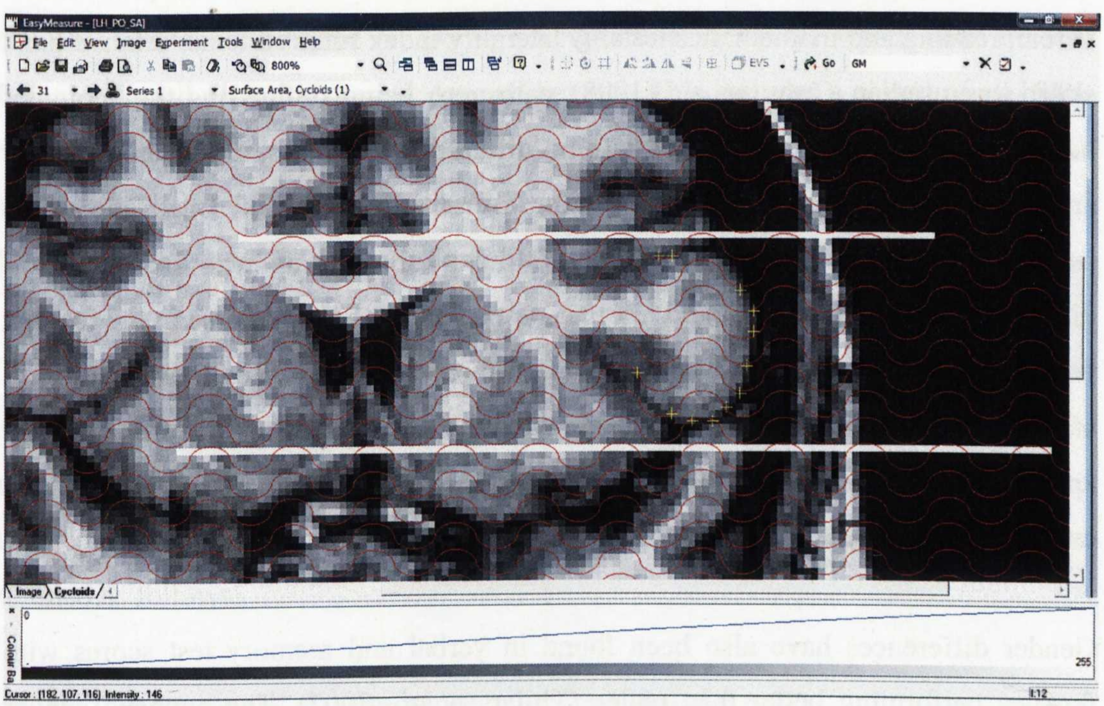


Figure 2.24: *EasyMeasure showing the cycloid grid with radius=3, with intersections (yellow crosses) when pars opercularis is demarcated.*

CHAPTER 3

Investigation of the associations between handedness and cognitive function in children

3.1 Introduction

Over the past two decades there has been much research into the links between cognitive function and certain associated variables. In particular, handedness and hand skill have been considered as explanatory variables in a number of studies of cognitive ability (e.g., Annett & Manning, 1990; Nettle, 2003; Faurie *et al.*, 2006; Peters *et al.*, 2006; Denny, 2008).

A possible association between hand skill and academic ability was investigated using data from a group of 11-year old children (Crow *et al.*, 1998). In this study, academic ability was measured by the number of marks obtained from verbal, non-verbal, reading and mathematics tests. A laterality index referred to as 'relative hand skill' was used in Crow *et al.*, (1998) with zero being defined as the 'point of hemispheric indecision' (i.e. there was no difference in hand skill between the left and right hand; see McManus, (1985) for laterality index definition). The results suggested that children with a laterality index of zero or with extreme skill differences between the right and left hand appeared to perform worse in the reading test. In the maths test, it was highlighted that those children with relative hand skill at the point of hemispheric indecision obtained the lowest scores. However, there did appear to be a substantial decrease in maths scores as children became more extremely right-handed.

Gender differences have also been found in verbal and memory test scores with females performing better than males (Thilers *et al.*, 2007). The opposite gender effect, with males linked with higher scores in spatial ability tests has also been reported (Voyer *et al.*, 1995). Results of previous studies are discussed in more detail in *Chapter 1.*

The investigations discussed above gave an insight into the connection of factors, such as gender and handedness on differing cognitive abilities. However, these factors had not been jointly considered in a single statistical model where their interactions were also taken into account. The aim of this chapter was to investigate the effect of gender, relative hand skill (decomposed into two variables related to its absolute value and direction), writing hand and region of the UK mainland on reading and mathematics scores in children, by applying a multivariate linear mixed-effects model to the National Child Development Study dataset (age 11). In particular, the correlation between reading and maths test scores, and potential interactions between the factors mentioned above, were taken into account in the analysis. The following two subsections contain a description of the variables considered along with the statistical methodology used.

3.2 Introduction to the data

In this chapter, a dataset collected for the National Child Development Study (NCDS) was analysed. The NCDS is a longitudinal study for which data are still being collected up to the present day and beyond. This includes a cohort of all births within a one week period (3-9 March, 1958) on the United Kingdom mainland. Follow-ups have occurred at various time points during the course of the individuals' lives. For a more detailed description of this dataset, see Section 1.3.1.

The variables of interest for this study were reading and transformed mathematics scores (which were considered as the outcome variables), gender, writing hand, UK region, relative hand skill and superior hand. The reading and transformed mathematics scores are as defined in Section 1.3.1, and the explanatory variables are defined as follows:

- *Gender*: 'male' = 0 (reference value) and 'female' = 1.
- *Writing hand* (WH): 'right-hand writers' = 0 (reference value) and 'left-hand writers' = 1.

- *Region* was created based on the UK region variable defined earlier in Section 1.3.1. Region was defined as a categorical variable with 4 levels:
  - Northern England and the Midlands = 0 (Reference region)
  - Southern England = 1.
  - Wales = 2
  - Scotland = 3
- Relative hand skill was decomposed into two variables:
  - *Relative hand skill* which was calculated as:

$$\text{relative hand skill} = \left|\frac{\text{n(RH)}-\text{n(LH)}}{\text{n(RH)}+\text{n(LH)}} \times 100\right| \tag{3.2.1}$$

    where n(RH) and n(LH) are the number of boxes ticked in 1 minute with the right and left hands, respectively (McManus, 1985). *Relative hand skill* can be interpreted as the relative superiority of the dominant hand over the non-dominant hand.
  - *Superior hand* (SH) was created to denote which hand was the dominant one. *Superior hand* is 0 if the dominant hand was the right hand (i.e. where the number of boxes ticked in 1 minute was greater for the right hand than for the left hand) and equal to 1 if the dominant hand was the left hand (i.e. more boxes ticked in 1 minute by the left hand than by the right).

A more specific location than simply the region of the UK mainland in which a child attended school was given by the variable *local authority*. This variable recorded the local authority which contained the school in which each child was enrolled. There were 185 local authorities in total and we used *local authority* as a clustering variable in the analyses undertaken in this chapter.

3.3 Statistical models

The statistical models here for the NCDS analysis were the univariate and multivariate linear mixed models (LMMs) described by Equations (2.1.8) and

(2.1.13), respectively. The intercept random effect is often used to take into account when repeated measurements are taken over time per individual (i.e. for longitudinal data) by assigning a specific value of a normally distributed random variable to each individual (i.e. the same value for every time point of a given individual). In this section, the intercept random effect terms take into account that children are clustered by local authority. Therefore, a value of a normally distributed random variable is assigned to each local authority in order to take into account the local authority effect (see Section 1.3.1, for a description of the local authority variable). The univariate linear mixed model adopts the following form:

$$\underline{\tilde{Y}} = \underline{\tilde{X}}\underline{\beta} + \underline{z}\,\underline{b} + \underline{\tilde{\varepsilon}} \tag{3.3.1}$$

as defined in Equation (2.1.8). However, there is a slight deviation of notation for the analyses in this chapter from that in *Chapter 2*. For the univariate and multivariate LMMs, local authority will be denoted by i such that $1 \leq i \leq g$ (where $g = 185$ is the number of local authorities), with individual j in each local authority ($1 \leq j \leq n_i$), where n_i denotes the number of children in the i^{th} local authority. The total number of children in the dataset ($n = 18558$) can be expressed as $n = \sum_{i=1}^{g} n_i$. From Equation (3.3.1), $\underline{\tilde{Y}}$ is the $(n \times 1)$ vector of test scores (reading or transformed mathematics scores) for the n children. The matrix $\underline{\tilde{X}}$ is the $(n \times p)$ fixed effect design matrix with $x_{ij1}, \dots, x_{ij(p-1)}$ representing the $(p-1)$ fixed effects for each child. In particular, there are 9 fixed effects which correspond to relative hand skill (x_{ij1}), (relative hand skill)^2 (x_{ij2}), superior hand (x_{ij3}), writing hand (x_{ij4}), superior hand and writing hand interaction $(x_{ij5} = x_{ij3} \times x_{ij4})$, gender (x_{ij6}), and region (4-level factor: x_{ij7}, x_{ij8}, x_{ij9}) such that:

$$\underline{\tilde{X}} = \begin{bmatrix} \underline{\tilde{X}_1} \\ \vdots \\ \underline{\tilde{X}_g} \end{bmatrix} \tag{3.3.2}$$

is an $(n \times 10)$ matrix where $\underline{\tilde{X}_i}$ is the $(n_i \times 10)$ matrix which can be written as:

$$\underline{\tilde{X}_i} = \begin{bmatrix} 1 & x_{i,1,1} & \cdots & x_{i,1,9} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i,n_i,1} & \cdots & x_{i,n_i,9} \end{bmatrix} \tag{3.3.3}$$

The vector $\underline{\beta}$ denotes the fixed effect coefficients to be estimated from the dataset and which can be expressed as:

$$\underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_9 \end{bmatrix} \tag{3.3.4}$$

The random effects term $\underline{z}\,\underline{b}$ allows the variability across local authorities to be estimated. Any further variables which are dependent upon or correlated with local authority could be included here and therefore, the number of random effect terms would increase. However, in the NCDS dataset there were no other variables that were expected to be dependent upon the local authority that a child is attached to, and hence, a random intercept model was deemed to be the most suitable structure of mixed model in this case. In a random intercept model, the random effects matrix $\underline{z}$ is an $(n \times g)$ matrix such that:

$$\underline{z} = \begin{pmatrix} \underline{\tilde{z}_1} & & \underline{0} \\ & \ddots & \\ \underline{0} & & \underline{\tilde{z}_g} \end{pmatrix} \tag{3.3.5}$$

with $\underline{\tilde{z}_i}$ being the $(n_i \times 1)$ constant vector of the form:

$$\underline{\tilde{z}_i} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \tag{3.3.6}$$

The $(g \times 1)$ random effects coefficient vector $\underline{b}$ explains the variability between children due to differences between local authorities:

$$\underline{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_g \end{bmatrix} \tag{3.3.7}$$

The random effects coefficient vector $\underline{b}$ also follows a g-dimensional multivariate normal distribution such that $\underline{b} \sim N_g(\underline{0}, \underline{G})$ and:

$$\underline{G} = \begin{bmatrix} \sigma_b^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_b^2 \end{bmatrix} \tag{3.3.8}$$

is a $(g \times g)$ diagonal matrix with σ_b^2 representing the random intercept variance term for the i^{th} local authority. This model could be made more complex, by allowing not only the random effect intercept variances to differ between different local authorities, but also the covariances between authorities (replacing the zeroes with non-zeroes in the off-diagonal elements of Equation (3.3.8)) However, in the case of the NCDS analyses, each local authority was assumed to be independent of each other (hence, the zero off diagonal elements in the matrices in Equation (3.3.8)). Lastly, the random error terms, represented by the $(n \times 1)$ vector $\underline{\tilde{\varepsilon}}$, are defined exactly as by Equation (2.1.8). This means that both the random effects and random errors are independent and identically distributed between subjects.

By applying the multivariate linear mixed model with random intercept only (see Equation (2.1.13)) to the data, we get a model of the form:

$$\underline{\bar{\bar{Y}}} = \underline{\bar{\bar{X}}\beta^*} + \underline{z}^*\underline{b}^* + \underline{\bar{\bar{\varepsilon}}} \tag{3.3.9}$$

where the $(2n \times 1)$ outcome vector $\underline{\bar{\bar{Y}}} = \begin{bmatrix} \underline{\tilde{Y}}^{(1)} \\ \underline{\tilde{Y}}^{(2)} \end{bmatrix}$ with $\underline{\tilde{Y}}^{(k)}$ ($k = 1,2$ is an indicator variable for the reading and transformed mathematics scores, respectively) as defined for Equation (2.1.13). The fixed effect design matrix $\underline{\bar{\bar{X}}} = \begin{bmatrix} \underline{\tilde{X}} & \underline{0} \\ \underline{0} & \underline{\tilde{X}} \end{bmatrix}$ is a $(2n \times 20)$ matrix (note that $\underline{\tilde{X}}$ is a $(n \times 10)$ matrix as defined in Equation (3.3.2)). The vector $\underline{\beta}^* = \begin{bmatrix} \underline{\beta}^{(1)} \\ \underline{\beta}^{(2)} \end{bmatrix}$ is the (20×1) vector of coefficients where $\underline{\beta}^{(k)}$ ($k = 1,2$) is the (10×1) vector defined as:

$$\underline{\beta}^{(k)} = \begin{bmatrix} \beta_0^{(k)} \\ \beta_1^{(k)} \\ \vdots \\ \beta_9^{(k)} \end{bmatrix} \tag{3.3.10}$$

$\underline{\beta}^{(k)}$ $(k = 1,2)$ represents the coefficients of the 9 fixed effects, as defined with the univariate model, for the k^{th} set of test scores. The $(2n \times 2g)$ random effects design matrix $\underline{z}^* = \begin{bmatrix} \underline{z} & \underline{0} \\ \underline{0} & \underline{z} \end{bmatrix}$ where $\underline{z}$ is as defined as in Equation (3.3.5) and $\underline{b}^* = \begin{bmatrix} \underline{b}^{(1)} \\ \underline{b}^{(2)} \end{bmatrix}$ is the $(2g \times 1)$ random effects vector such that:

$$\underline{b}^{(k)} = \begin{bmatrix} b_1^{(k)} \\ \vdots \\ b_g^{(k)} \end{bmatrix} \tag{3.3.11}$$

is the $(g \times 1)$ random effects vector for the k^{th} set of test scores. The random effects follow the $2g$-dimensional multivariate normal distribution such that $\underline{b}^* \sim N_{2g}(\underline{0}, \underline{G}^*)$ and where $\underline{G}^* = \begin{bmatrix} \underline{G}_{11}^* & \underline{G}_{12}^* \\ \underline{G}_{21}^* & \underline{G}_{22}^* \end{bmatrix}$ is a $(2g \times 2g)$ matrix with $\underline{G}_{kk}^* = \begin{bmatrix} \sigma_{b_k}^2 & & \underline{0} \\ & \ddots & \\ \underline{0} & & \sigma_{b_k}^2 \end{bmatrix}$.

The off diagonal elements of $\underline{G}^*$ are $\underline{G}_{k_1k_2}^* = \begin{bmatrix} \sigma_{b_{k_1}b_{k_2}} & & 0 \\ & \ddots & \\ 0 & & \sigma_{b_{k_1}b_{k_2}} \end{bmatrix}$ $(k_1, k_2 = 1,2;\ k_1 \neq k_2)$. The variability in the k^{th} outcome variable due to the i^{th} local authority is equal to $\sigma_{b_k}^2$ with the covariance between the two sets of test scores in the i^{th} local authority being equivalent to $\sigma_{b_{k_1}b_{k_2}}$. The multivariate LMM can be extended so that the random intercept term differs amongst local authorities by including the covariance between individual authorities (replacing the zeroes with non-zeroes in the off-diagonal elements of $\underline{G}_{kk}^*$). However, as stated earlier for the univariate case, we assume that each local authority is independent of each other as well as sharing a single random intercept value. The reported random effect variances and covariance, in the analyses of this chapter, are representative of the overall variance due to local authority across all subjects for each outcome, and the overall covariance between the outcomes, at the local authority level.

Finally, the $(2n \times 1)$ random error term vector, $\underline{\bar{\bar{\varepsilon}}}$, can be expressed as:

$$\underline{\bar{\bar{\varepsilon}}} = \begin{bmatrix} \underline{\bar{\varepsilon}}^{(1)} \\ \underline{\bar{\varepsilon}}^{(2)} \end{bmatrix} \qquad (3.3.12)$$

where

$$\underline{\bar{\varepsilon}}^{(k)} = \begin{bmatrix} \varepsilon_1^{(k)} \\ \vdots \\ \varepsilon_n^{(k)} \end{bmatrix} \qquad (3.3.13)$$

such that $\underline{\bar{\bar{\varepsilon}}} \sim N_{2n}\left(\underline{0}, \underline{R}^*\right)$ where $\underline{R}^* = \begin{bmatrix} \underline{R}_{11}^* & \underline{R}_{12}^* \\ \underline{R}_{21}^* & \underline{R}_{22}^* \end{bmatrix}$ with $\underline{R}_{kk}^*$ and $\underline{R}_{k_1k_2}^*$ set as $(n \times n)$ matrices as defined in Equation (2.1.13). The variability within subjects for the k^{th} set of test scores can then be written as $\sigma_{\varepsilon_k}^2$ $(k = 1,2)$ with the covariance between the two sets of test scores at the subject level represented by $\sigma_{\varepsilon_{k_1}\varepsilon_{k_2}}$. The analyses and plots for the both the univariate and multivariate LMMs were conducted using a combination of MLwiN version 2.16 and R version 2.12.1.

3.4 Statistical analyses

Random intercept linear mixed models were initially applied to the National Child Development Study (NCDS) dataset with reading and transformed maths scores as the outcome variables. The random intercept takes into account the variability between local authorities and the random error term explains the variation in the data that cannot be explained by the fitted model. The final fitted model was constructed using a stepwise model selection process. That is, at each stage of the model selection process, not only were new variables checked to see if their inclusion benefitted the model (p-value <0.05), but also current variables in the model were checked to see if they could be removed (i.e. if they became insignificant as more variables were added; Fahrmeir & Tutz, 1994). The p-value for stepwise removal used here was 0.1. While this process was being completed restricted maximum likelihood (REML) was used as the estimation method for univariate LMMs. REML was introduced by Patterson & Thompson (1971) in the context of incomplete block designs for estimating variance parameters. The reason that REML was used as opposed to maximum likelihood (ML) is that REML takes into account the loss of degrees of freedom resulting from the estimation of the fixed effects when estimating the variance components (Harville, 1977). REML also produces estimation equations for the variance parameters that are unbiased (Smyth & Verbyla, 1996). For the multivariate LMMs, restricted iteratively generalized least squares (RIGLS) was used as the estimation method. For multivariate linear models (and multivariate LMMs) that are assumed to follow a multivariate normal distribution, the parameter estimates obtained using a RIGLS method is equivalent to those obtained using a REML method (Goldstein, 1989).

Issues regarding missing data are covered in detail in *Chapter 6*, however, in the case of the analyses contained within this chapter, any subjects with missing data in any of the variables entered into the model were automatically omitted. The missing data for the three analyses (univariate and multivariate) can be seen in Table 3.1. Taking all variables into account it can be seen that 36.1% of the subjects have at least one missing variable value when considering reading scores as the only outcome, whereas when considering the mathematics scores either independently or in conjunction with the reading scores, the percentage of those subjects with at least one variable value missing is 36.2%.

When fitting the LMMs, it was noticed that three children were associated with much lower fitted values than the others. When investigated, these children had values of relative hand skill, much higher than other children in the dataset. Two of the children had relative hand skill equal to 100, and one had a relative hand skill of 96. This meant that each child had ticked a number of boxes with one hand, but had zero or just one or two ticked with the other hand. This could be due to a number of reasons including recording error, the reluctance of the child to co-operate or the child only having one hand. Considering these possible reasons, the three children with extreme values of relative hand skill were excluded from the analyses.

Variable	*Missing (%)*	*Non-missing (%)*
Reading Scores	4,427 (23.9%)	14,131 (76.1%)
Mathematics Scores	4,431 (23.9%)	14,127 (76.1%)
Gender	3 (0.02%)	18,555 (99.98%)
Local Authority	3,226 (17.4%)	15,332 (82.6%)
Relative Hand Skill	5,792 (31.2%)	12,766 (68.8%)
Superior Hand (SH)	5,788 (31.2%)	12,770 (68.8%)
Writing Hand (WH)	4,838 (26.1%)	13,720 (73.9%)
UK Region	3,201 (17.2%)	15,357 (82.8%)
Total*	6,708[a] (36.1%) 6,711[b] (36.2%) 6,712[c] (36.2%)	11,850[a] (63.9%) 11,847[b] (63.8%) 11,846[c] (63.8%)

*A number of subjects have multiple missing variable values and hence, are counted as 'missing' only once.

[a] These figures correspond to the univariate model with reading scores as the outcome.

[b] These figures correspond to the univariate model with transformed mathematics scores as the outcome.

[c] These figures correspond to the multivariate model with both reading and transformed mathematics scores as outcomes.

Table 3.1: *Missing and non-missing data for the outcome, explanatory and random effect variables of the NCDS dataset*

Further evidence for this comes from plotting mean reading and transformed mathematics scores against relative hand skill by SH and WH. Figure 3.1 shows the mean reading and transformed mathematics scores against relative hand skill both for all children in the full dataset and for all children in the reduced dataset with the outliers removed. Of the three children with extreme relative hand skill values, one had left WH and SH and the other two had right SH and WH. Hence, the only groups that were affected in Figure 3.1 were the two consistent WH and SH groups which are associated with the blue and red lines (children with right SH-right WH and left

SH-left WH, respectively). Those children with inconsistent WH and SH took the form of the purple and green lines (right SH-left WH and left SH-right WH, respectively) and since these groups contain no children with extreme values of relative hand skill, there were no differences in the distribution of reading or maths scores across relative hand skill in the inconsistent handedness groups between the two datasets (with and without outliers, see Figure 3.1).

For children in the full dataset with relative hand skill greater than 40, it can be seen in Figure 3.1, by observing the almost parallel blue and red lines, that the mean reading scores decrease at an almost constant rate for both consistent WH and SH groups, as relative hand skill increases. However, once the three outliers are removed, the left WH-left SH group are associated with an increased rate of decline in terms of mean reading scores by relative hand skill, whereas the rate of decline of mean reading scores by relative hand skill for the right WH-right SH group has decreased (see Figure 3.1). Changes in the rate of decline of transformed maths marks by relative hand skill also occurred once the outliers were removed for children in both the consistent WH and SH groups with relative hand skill values greater than 40 (see Figure 3.1).

This change in the rate of decline of transformed maths marks by relative hand skill is most noticeable for the left WH-left SH group where the decrease in maths scores as relative hand skill increases between values of 40 and 100 in Figure 3.1 is small before the outliers are removed (a reduction in maths marks of around 1.75 marks), compared to a larger reduction of around 2.25 maths marks between the relative hand skill values of 40 to 57. Since there were only three individuals with relative hand skill greater than 64, any confidence intervals for the curves would be extremely large for these areas and this is the reason why the changes that occurred in Figure 3.1 between the two datasets were so noticeable. The information that could therefore, be obtained about children with relative hand skill between 64 and 100 was limited even if there was no uncertainty surrounding the measurement and recording of the children's results. Furthermore, these three individuals represented only 0.03% of the total data after missing data was removed. After discussion and much thought, we decided to exclude these three subjects from further analyses.

Following this decision, the LMMs with reading and transformed mathematics test scores as outcomes were re-fitted and the results can be seen in Section 3.4.1.

In Figure 3.1, a loess smoothing function was applied to the raw data in order to obtain the curves. The loess smoothing function was introduced by Cleveland (1979) and is a regression method which, similar to time series analyses, takes account of data which are in close proximity to each individual data-point. In this type of regression method, for example, each particular datum could be slightly modified (weighted) to take into account a particular number of data nearest to the data-point, or to take into account data within a certain range. The loess smoothing function uses a weighting function, W, to account for data in close proximity to other data. This is a function of u, where u is the proportion of the dataset closest to each data-point. For example, if $u = 0.5$, then the nearest 50% of the data would be accounted for by the weights for each data-point. The extreme values are $u = 0$ (i.e. when each data-point is totally independent of each other) and $u = 1$ (i.e. when every data-point is dependent upon every other datapoint in the entire dataset). The weighting function itself, $W(u)$, can be calculated by:

$$W(u) = (1 - u^3)^3 \qquad (3.4.1)$$

such that $0 \leq u < 1$ (Cleveland & Devlin, 1988). To alter the smoothing of the regression lines the proportion of data which the weighting function is applied to can be modified. In Figure 3.1 u was set such that the regression lines were relatively smooth and yet still gave a reasonable indication of the relationship between relative hand skill and the two test score variables for each SH-WH group. In this case, u was set equal to 0.75.

Plots of relative hand skill against mean reading and transformed maths scores are given in Figures 3.2(a) and 3.2(b), respectively. In these plots, the term 'relative hand skill' is defined as:

$$\text{relative hand skill} = \frac{\text{n(RH)} - \text{n(LH)}}{\text{n(RH)} + \text{n(LH)}} \times 100 \qquad (3.4.2)$$

where n(RH) and n(LH) are the number of boxes ticked in 1 minute with the right and left hands, respectively. Note that the formula in Equation (3.4.2) is the same as Equation (3.2.1) except for the modulus sign.

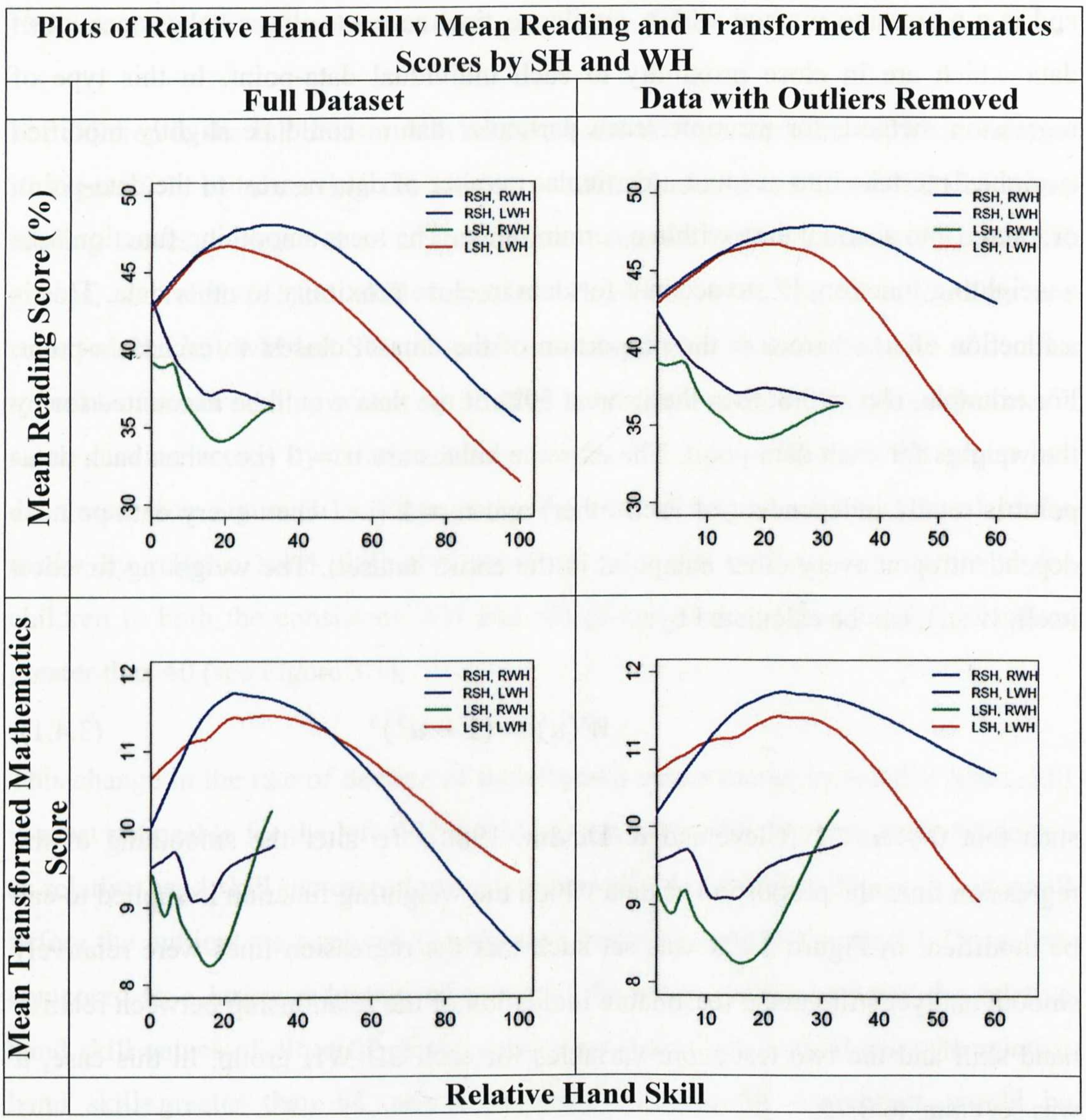
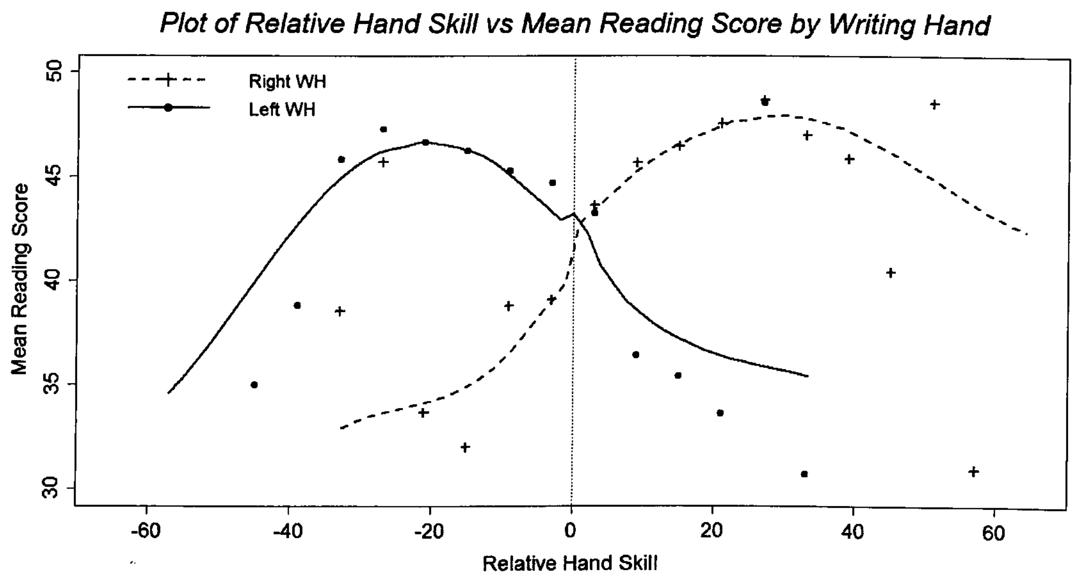


Figure 3.1: *Plots of relative hand skill against mean reading test scores and mean transformed mathematics scores by left and right superior hand (LSH and RSH, respectively) and left and right writing hand (LWH and RWH, respectively) for both the full dataset (left hand column) and the dataset with the outliers removed (right hand column). The curves were obtained by using the 'Loess' smoothing function with smoothing parameter equal to 0.75.*

The crosses and dots in Figure 3.2 represent mean values of the actual data for equidistant intervals of relative hand skill for right-hand and left-hand writers, respectively. Some of these empirical mean values differ from the values of the loess curve (which is weighted across the nearest three-quarters of the data to each data-point (i.e. $u = 0.75$)), and this is caused by the small sample sizes that are available for certain groups (e.g. for right-hand writers, there were 10,313 pupils with right superior hand compared to 313 with left superior hand, and for left-hand writers there were 1,252 pupils with left superior hand and only 103 with right superior hand). It is interesting to note that although there are more right-hand writers with left superior hand than left-hand writers with right superior hand, this reverses when expressed as percentages of the total sample of right- and left-hand writers, with 2.3% of right-hand writers having left superior hand compared to 7.7% of left-hand writers having right superior hand.

Scores for pupils that have a left superior hand and are right-hand writers are represented by the dashed curve where relative hand skill is less than zero. Scores for left-hand writers with right superior hand can be identified by the solid curve where relative hand skill is greater than zero. Scores for pupils with consistent writing hand and superior hand are included in the solid curve when relative hand skill is less than zero (i.e. left superior hand) for left-hand writers and in the dashed curve when relative hand skill is greater than zero (i.e. right superior hand) for right-hand writers. It can also be seen that there appears to be a quadratic relationship for left- and right-hand writers between both reading and transformed maths scores, and relative hand skill.

(a)



(b)

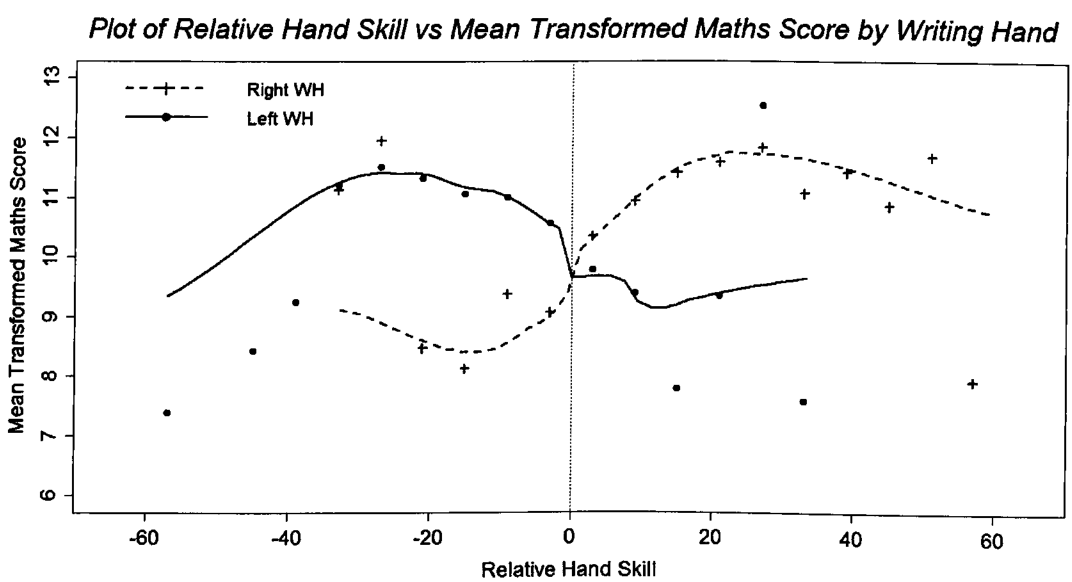


Figure 3.2: *Plot of relative hand skill against (a) mean reading score and (b) mean transformed mathematics score by writing hand where pupils with left SH appear on the left hand side of the graphic with negative relative hand skill values. The curves were obtained by using the "Loess" smoothing function. The dots and crosses represent the mean values for 20 equidistant intervals across the range of relative hand skill (dots for left-hand writers, and crosses for right-hand writers).*

3.4.1 Multivariate model

Following a stepwise model selection process as described earlier in Section 3.4, the results of the final, fitted multivariate LMM can be seen in Tables 3.2-3.5. This model was applied to the NCDS dataset with 11,843 children included (after taking into account those with missing outcomes and covariates and the three with extreme handedness values, see Table 3.1 and Section 3.4 for details). The multivariate models in this chapter were fitted using MLwiN which uses Wald tests to determine whether the multivariate parameters equal zero or not. The Wald statistic follows a chi-square distribution with degrees of freedom equal to the number of outcomes (d) (Goldstein, 2003). Therefore, the p-values presented in Table 3.2 came from the chi-square distribution with 2 degrees of freedom ($d = 2$). Furthermore, hypothesis tests were also conducted using a Wald test with a statistic following a chi-square distribution with degrees of freedom equal to the number of parameters being tested (Dobson, 1999).

Multivariate models give information about both the associations between the explanatory variables and a linear combination of the outcomes as well as those links between the explanatory variables and each outcome individually. In this case, the overall coefficients, standard errors, confidence intervals and p-values for the explanatory variables with reading and transformed maths scores combined as the outcome (i.e. $\underline{\tilde{Y}}^{(1)} + \underline{\tilde{Y}}^{(2)}$ from Equation (3.3.9)) can be seen in Table 3.2.

We found that relative hand skill and the square of relative hand skill both had significant coefficients of 0.6 and -0.01, respectively (both $p<0.001$, Table 3.2). This implies that when both sets of scores were considered simultaneously, there was a quadratic relationship between relative hand skill and the cognitive ability test scores (i.e. the reading and transformed maths scores). Since the coefficients for relative hand skill and (relative hand skill)^2 were positive and negative, respectively, then this implies that the quadratic relationship, if plotted would form an inverted u-shape similar to those seen in Figure 3.2. We calculated the optimum relative hand skill (the level of relative hand skill associated with the maximum average combined reading and transformed maths scores from the model) by differentiating and setting the quadratic function equal to zero after coefficients from the model were substituted in as follows:

$$y = 0.6x - 0.01x^2 + c$$

$$\mathrm{d}y/\mathrm{d}x = 0.6 - 2(0.01x)$$

$$0 = 0.6 - 0.02x$$

$$x = -0.6/-0.02 = 30$$

where y and x represent the multivariate outcome (combined reading and transformed mathematics scores) and optimum relative hand skill, respectively. Therefore, the level of relative hand skill associated with the maximised multivariate outcome was equal to 30.

Superior hand (SH), writing hand (WH) and the interaction term between both SH and WH were all significant ($p<0.001$, $p=0.003$, $p<0.001$, respectively, Table 3.2). The interaction term between SH and WH being included in the model implied that children with different combinations of SH and WH could be tested for equality or non-equality of the sum of reading and transformed maths scores together. We conducted these multiple tests and the results can be seen in Table 3.3. The overall conclusions from these hypothesis tests were that those children with inconsistent WH and SH (i.e. those children that wrote with the hand which had least hand skill) performed worse, on average, than those individuals with consistent WH and SH (i.e. those that write with their hand that has the most hand skill). For right-hand writers, we found that those with right SH performed better, on average, across both tests simultaneously than those with left SH ($p<0.001$, Table 3.3). For left-hand writers, those with left SH obtained higher test scores on average, across both test combined, than those with right SH ($p=0.01$, Table 3.3). Children with left SH-left WH achieved higher scores across both tests on average, than those with left SH-right WH, and children with right SH-right WH obtained higher scores across both tests on average that those with right SH-left WH ($p<0.001$ and $p=0.003$, respectively,

Table 3.3). Also, when we compared individuals with right SH-right WH to individuals with left SH-left WH in terms of the combined test scores, no difference was found (p=0.4, Table 3.3).

Fixed Effect	Coefficient	St. Error	95% CI	p-value
Intercept	52.2	0.7	(50.8, 53.7)	<0.001
Relative hand skill	0.6	0.07	(0.4, 0.7)	<0.001
(Relative hand skill)^2	-0.01	0.002	(-0.014, -0.008)	<0.001
Superior hand (SH) (0: Right, 1: Left)	-7.8	1.4	(-10.5, -5.0)	<0.001
Writing hand (WH) (0:Right, 1:Left)	-6.3	2.1	(-10.5, -2.2)	0.003
SH x WH	13.2	2.6	(8.1, 18.4)	<0.001
Gender *(0: Male, 1: Female)*	-0.02	0.4	(-0.8, 0.7)	0.9
Northern England & Midlands	*Reference Region*			
Southern England	1.8	0.5	(0.9, 2.7)	<0.001
Wales	-0.4	0.9	(-2.3, 1.4)	0.6
Scotland	0.2	0.7	(-1.2, 1.5)	0.8

Table 3.2: *Estimates of the multivariate fixed effect coefficients for the multivariate linear mixed model with both the reading and mathematics scores combined (including the standard error of the corresponding estimators, the 95% confidence intervals for the coefficients and the p-values).*

WH	SH	Estimate	Standard Error	95% CI	p-value
Right	Left vs Right	-7.8	1.4	(-10.5, -5.0)	<0.001
Left	Left vs Right	5.5	2.2	(1.2, 9.8)	0.01
Left vs Right	Right	-6.3	2.1	(-10.5, -2.2)	0.003
Left vs Right	Left	6.9	1.5	(4.0, 9.9)	<0.001
Left vs Right	Left vs Right	-0.8	0.8	(-2.5, 0.8)	0.4

Table 3.3: *Results of the hypothesis tests for differences between right and left superior handed children, and between right- and left-hand writers, in both the reading and transformed mathematics scores combined.*

Table 3.2 suggests that there are no differences in the combined reading and transformed maths scores between boys and girls (p=0.9). Children that attended schools in Southern England outperformed children that attended schools in Northern England and the Midlands when both sets of test scores were combined (p<0.001, Table 3.2). However, we found no differences in the overall test scores between pupils of Scottish and Welsh schools and those pupils from schools in Northern England and the Midlands (p=0.8 and p=0.6, respectively, Table 3.2).

Information about the associations between the explanatory variables and each individual outcome (in this case the reading and transformed mathematics scores) was obtained from the models based on the marginal distributions of the MLMM. These results can be seen in Tables 3.4 and 3.5. We note here that due to the similarity of multiple models from a multivariate model to multiple comparisons, as mentioned in *Chapter 2,* a p-value adjustment was needed for these marginal model results. We used the Bonferroni correction to adjust the α-value (the critical level at which a p-value becomes statistically significant) to account for the fact that we have two outcomes in our multivariate model such that:

$$\alpha^* = \alpha/2 \tag{3.4.4}$$

where α^* is the adjusted critical value (Hair, Jr., *et al.*, 2010). In our case we wanted a 5% significance level (i.e. $\alpha = 0.05$) hence, α^* was set equal to 0.025.

Using the Bonferroni-adjusted significance level for this model of 0.025 we can see that relative hand skill had a significant effect on both the reading and mathematics scores and that this effect was quadratic (p<0.001 for both relative hand skill and (relative hand skill)^2 in the models based on the marginal distribution of each set of test scores, Table 3.4). In Figures 3.2 (a) and (b) this quadratic effect of relative hand skill could be seen in those children with consistent SH and WH, obtained directly from the raw data. In particular, the increase in reading test scores due to a 1 unit increase in relative hand skill was 0.4%. However, this increase was eroded by the quadratic term until the value of relative hand skill reached an optimum point. After this optimum point, further increases in relative hand skill were associated with decreases in reading scores, in much the same way as for the multivariate outcome. This quadratic effect of relative hand skill was very similar for the transformed mathematics scores. We calculated the optimum relative hand skill from the model for the multivariate outcome to be equal to 30. The optimum relative hand skill for each individual set of test scores, according to the model, was calculated by differentiating and setting equal to zero the following quadratic functions:

$$y_1 = 0.4x - 0.009x^2 + c \qquad\qquad y_2 = 0.1x - 0.002x^2 + c$$

$$\mathrm{d}y_1/\mathrm{d}x = 0.4 - 2(0.009x) \qquad\qquad \mathrm{d}y_2/\mathrm{d}x = 0.1 - 2(0.002x)$$

$$0 = 0.4 - 0.018x \qquad\qquad 0 = 0.1 - 0.004x$$

$$x = -0.4/-0.018 = 22.2 \qquad\qquad x = -0.1/-0.004 = 25$$

where y_1 and y_2 represent the reading and transformed mathematics scores, respectively, and in each case x is the optimum relative hand skill. Therefore, it can be seen that the optimum relative hand skill according to our model was 22.2 for the reading scores and 25 for the transformed mathematics scores.

We found that superior hand, writing hand and their interaction term were all strongly significant, even with the Bonferroni correction, for both sets of test scores (p<0.001 for SH and WH x SH for both outcomes and p=0.005 and p=0.002 for WH in the reading and transformed maths scores, respectively, Table 3.4). The coefficients of SH and WH were negative for both sets of test scores and the interaction term between SH and WH was positive. The negative coefficients of SH suggest that for right-hand writers (reference group) those with left superior hand perform worse in both tests than those with right superior hand. In particular, the mean difference was 6% for the reading scores. Furthermore, for left-hand writers, those with left superior hand performed, on average, significantly better in both tests than those with right superior hand. To confirm this last statement, we conducted the corresponding hypothesis test for each of the two outcomes and this difference was found to be 5%, on average, for the reading scores (p=0.01 and p=0.006, for both tests, respectively, Table 3.5). Writing hand had a significant effect for both reading and mathematics scores (negative coefficients; p=0.005 and p=0.002, respectively, Table 3.4), which suggests that for pupils with right superior hand (reference group), left-hand writers perform worse than right-hand writers (and this difference was 5%, on average, for the reading scores, see Table 3.4). Finally, for pupils with left superior hand, left-hand writers achieved higher scores on both tests, on average, than right-hand writers (Table 3.5, both p=<0.001). Specifically, this difference was equal to 5.5%, on average, for the reading scores (Table 3.5).

The above results reveal that there is a discrepancy in test scores between pupils with inconsistent writing hand and superior hand and pupils with consistent writing hand and superior hand. Alternatively, our analysis reveals that left-hand writers with left superior hand had, on average, no differences in either the reading or transformed maths scores than right-hand writers with right superior hand (see Table 3.5, p=0.2 and 0.2, respectively). Although the data did not provide evidence to suggest a difference in performance between consistent left and consistent right handed pupils, what should be borne in mind is that the sample size of left handed pupils is relatively small for large values of relative hand skill (for example, there are only 23 left handed pupils with absolute relative hand skill greater than 35).

No evidence was found to suggest that there was a difference between boys and girls in either the reading or the transformed mathematics scores (p=0.8 and p=0.1, respectively, Table 3.4). Furthermore, there was a significant difference in both reading and maths scores between pupils attending schools in Southern England and those pupils enrolled in schools in Northern England and the Midlands (see Table 3.4, p<0.001 and p=0.02, respectively). In the reading tests, we observed that children based in Southern England performed, on average, 2% better than their counterparts from Northern England and the Midlands (Table 3.4).

Further analysis based on this fitted model showed that pupils in Southern England also outperformed pupils in Wales and Scotland by 2%, on average, in the reading test (p=0.003 and p=0.001, respectively). A further difference was also found in the transformed mathematics scores between children that attended schools in Southern England and Scotland and those children that attended schools in Northern England

and the Midlands (p=0.02 and p<0.001, respectively, Table 3.4) such that those in Northern England and the Midlands performed worse than the others.

In our multivariate linear mixed model, we were able to estimate the variance in the reading and transformed maths scores, as well as the covariance and correlation between the two sets of scores which were due to the local authority in which each child's school was located (in the random effects, as seen in Table 3.4). The correlation coefficient, r, between the reading and transformed maths scores at the local authority level was 0.97 (p<0.001, Table 3.4). Therefore, a given local authority had similar proportions of pupils performing well on both tests, and conversely, of pupils doing badly on both tests.

Similarly, the random error variances which represent the random variation at pupil level in the reading and transformed maths scores, along with the covariance and correlation between the two sets of test scores at the pupil level are given in Table 3.4. The correlation was r=0.74, which suggests that there is also a strong positive correlation between the two sets of test scores (p<0.001, Table 3.4). Precisely, there is enough evidence to suggest that those children that performed well on one of the tests, also tended to do well on the other test, and conversely, those children that did not perform well on one of the tests did not perform well on the other test.

Fixed Effect	Coeff.	St. Error	95% CI	*p*-value
Reading Test Scores				
Intercept	42.2	0.6	(41.0, 43.4)	<0.001
Relative hand skill	0.4	0.06	(0.3, 0.5)	<0.001
(Relative hand skill)^2	-0.009	0.001	(-0.01, -0.006)	<0.001
Superior hand (SH) *(0: Right, 1: Left)*	-6.2	1.2	(-8.5, -3.9)	<0.001
Writing hand (WH) *(0:Right, 1:Left)*	-5.0	1.8	(-8.5, -1.5)	0.005
SH x WH	10.5	2.2	(6.2, 14.9)	<0.001
Gender *(0: Male, 1: Female)*	0.1	0.3	(-0.5, 0.7)	0.8
Northern England & Midlands	*Reference Region*			
Southern England	1.6	0.4	(0.9, 2.4)	<0.001
Wales	-0.8	0.8	(-2.3, 0.8)	0.3
Scotland	-0.4	0.6	(-1.5, 0.8)	0.5
Transformed Mathematics Test Scores				
Intercept	10.0	0.1	(9.8, 10.3)	<0.001
Relative hand skill	0.1	0.01	(0.09, 0.2)	<0.001
(Relative hand skill)^2	-0.002	<0.0005	(-0.003, -0.001)	<0.001
Superior hand (SH) *(0: Right, 1: Left)*	-1.6	0.3	(-2.1, -1.0)	<0.001
Writing hand (WH) *(0:Right, 1:Left)*	-1.3	0.4	(-2.2, -0.5)	0.002
SH x WH	2.7	0.5	(1.7, 3.7)	<0.001
Gender *(0: Male, 1: Female)*	-0.1	0.08	(-0.3, 0.03)	0.1
Northern England & Midlands	*Reference Region*			
Southern England	0.2	0.09	(0.03, 0.4)	0.02
Wales	0.3	0.2	(-0.03, 0.7)	0.07
Scotland	0.5	0.1	(0.2, 0.8)	<0.001
Random Effect	**Est.**	**St. Error**	**95% CI**	***p*-value**
Variance (Reading Scores)	23.5	3.8	(16.1, 30.9)	<0.001
Variance (Transformed Maths Scores)	1.1	0.2	(0.6, 1.5)	<0.001
Covariance [Correlation] (Reading/Trans. Maths Scores)	4.8 [0.97]	0.8 [0.2]	(3.3, 6.4) [0.66, 1.00]	<0.001
Random Error	**Est.**	**St. Error**	**95% CI**	***p*-value**
Variance (Reading Scores)	293.2	5.1	(283.2, 303.2)	<0.001
Variance (Trans. Maths Scores)	17.3	0.3	(16.7, 17.9)	<0.001
Covariance [Correlation] (Reading/Trans. Maths Scores)	52.4 [0.74]	1.1 [0.02]	(50.2, 54.5) [0.71, 0.77]	<0.001

Table 3.4: *Estimates of the fixed effects coefficients and the random effect and random error variance, covariance and correlation terms for the multivariate linear mixed model (including the standard error of the corresponding estimators, the 95% confidence intervals for the coefficients/terms and the p-values; NCDS dataset).*

WH	SH	Test	Estimate	St. Error	95% CI	*p*-value
Right	Left vs Right	Reading	-6.2	1.2	(-8.5, -3.9)	<0.001
		Transformed Mathematics	-1.6	0.3	(-2.1, -1.0)	<0.001
Left	Left vs Right	Reading	4.3	1.8	(0.7, 7.9)	0.02
		Transformed Mathematics	1.1	0.3	(0.3, 2.0)	0.009
Left vs Right	Right	Reading	-5.0	1.8	(-8.5, -1.5)	0.005
		Transformed Mathematics	-1.3	0.4	(-2.2, -0.5)	0.002
Left vs Right	Left	Reading	5.5	1.3	(3.0, 8.0)	<0.001
		Transformed Mathematics	1.4	0.3	(0.8, 2.0)	<0.001
Left vs Right	Left vs Right	Reading	-0.7	0.5	(-1.7, 0.4)	0.2
		Transformed Mathematics	-0.2	0.1	(-0.4, 0.08)	0.2

Table 3.5: *Results of the hypothesis test for differences between right and left superior handers, and between right and left-hand writers, in both the reading and transformed mathematics scores.*

The Pearson correlation coefficient, ρ, between reading and transformed mathematics scores can be obtained directly from the dataset. Figure 3.3 shows the reading scores plotted against the transformed mathematics scores along with a line of best fit. The correlation was estimated to be $\rho=0.75$ ($p<0.001$).

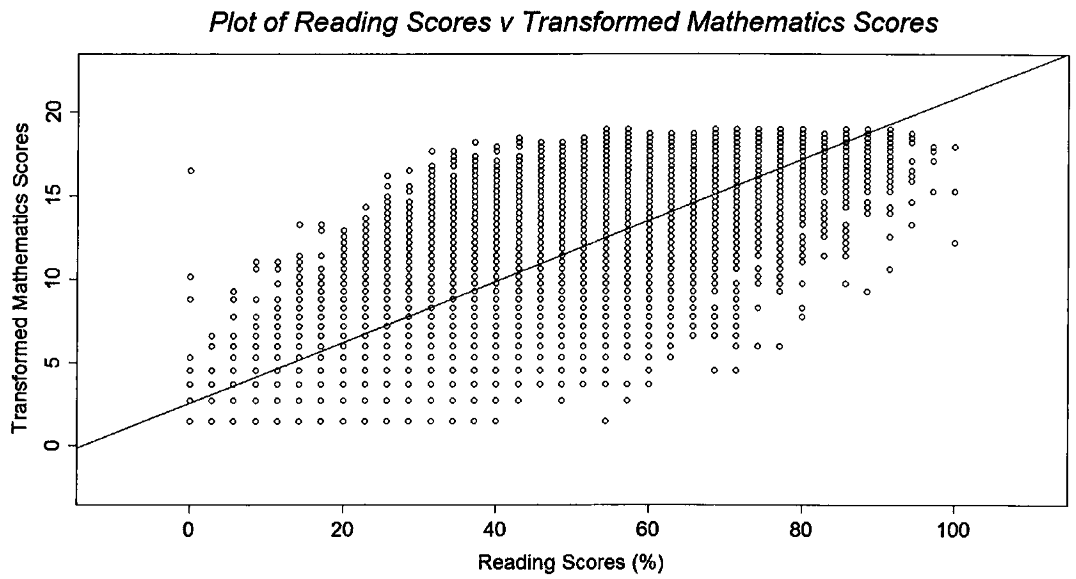


Figure 3.3: *Plot of reading scores against transformed mathematics scores from the raw dataset with a line of best fit.*

3.4.2 Comparison between multivariate and univariate linear mixed models

As expected, the univariate models with reading and transformed maths scores as the outcome variables had similar results to the models based on the marginal distributions from the multivariate linear mixed model (MLMM) in Table 3.4. The results were not identical due to the fact that there were differences in the number of children included for each model (see Table 3.1). These slight differences meant that the coefficients and *p*-values were slightly different. However, the coefficients and *p*-values did not change sufficiently to alter the interpretation from that of the MLMM models based on the marginal distributions.

In this subsection, the question of whether the MLMM in Section 3.4.1 provided a better representation of the associations within the NCDS dataset than two univariate LMMs is addressed. In other words, the question is whether the complexity added by the MLMM with the calculation of the extra parameters (the covariance and correlation in the random effects and random errors, along with the explanatory coefficients associated with the multivariate outcome) provides a better fit when compared to two fitted univariate LMMs. We already stated two main reasons for fitting a multivariate model to multivariate data as opposed to multiple univariate models in *Chapter 2* (correlation between outcomes is ignored and multiple comparisons mean that the critical level, α, has to be adjusted). As we have seen, there is strong evidence to suggest that the reading and transformed maths scores are correlated. This correlation was taken into account in the MLMM but not by the univariate LMMs as the two sets of test scores were treated as if they were independent.

Since the fitted univariate models are almost identical to the models based on the marginal distributions from the MLMM, then the deviance from an MLMM with both the random effect and random error covariances set equal to zero (i.e. with both outcomes set as being independent of each other) is a reasonable approximation to the combined deviance from the two univariate LMMs. Since the MLMM with independent outcomes (model 1) is nested within the multivariate model with correlated outcomes (model 2) (i.e. model 1's parameters are all included in model 2) then the deviance test can be used to compare the models (Krzanowski, 2003). The deviance, D, of a model is a goodness-of-fit statistic that can be calculated for generalized linear models (including linear mixed-effects models) and is defined as:

$$D = -2[l(y; y) - l(\mu; y)] \tag{3.4.5}$$

where $l(y; y)$ is the log-likelihood of the model if all the fitted values exactly equalled the observed values from the dataset and $l(\mu; y)$ is the log-likelihood of the current fitted model (Galwey, 2006). The deviance statistic is asymptotically distributed as a χ^2 statistic with v degrees of freedom where v is equal to the total number of data multiplied by the number of outcomes (i.e. dn) less the total number of independent parameters in the model (Krzanowski, 2003). The difference in deviance between two models (one of which is a sub-model of the other) can be regarded as a χ^2 statistic, with degrees of freedom equal to the difference in degrees of freedom between the two models (McCullagh & Nelder, 1989). The deviance values for the MLMMs with independent and correlated outcomes are shown in Table 3.6.

#	Model	Deviance	df
1	Independent MLMM	169858.7	23662
2	Correlated MLMM	160070.3	23660

Table 3.6: *Table showing the deviances and degrees of freedom of the two multivariate linear mixed models with reading and transformed maths scores as outcomes, one with independent outcomes (independent MLMM) and the other with correlated outcomes (correlated MLMM).*

The deviance statistic between the multivariate model with independent outcomes (independent MLMM) and the multivariate model with correlated outcomes (correlated MLMM), D^*, can be written as:

$$D^* = |D_1^* - D_2^*| = |169858.7 - 160070.3| = 9788.4 \qquad (3.4.6)$$

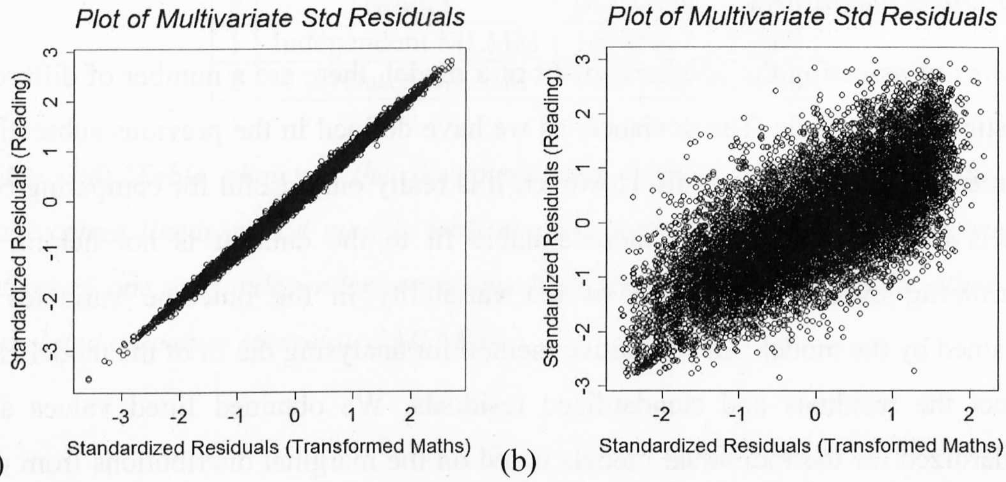
where D_1^* represents the deviance of the independent MLMM within the 'full' model (correlated MLMM) which has deviance D_2^*. D^* has degrees of freedom (df) equal to the difference in degrees of freedom between the two models (i.e. df $= 23662 - 23660 = 2$). Thus the test statistic (9788.4) with df equal to 2, corresponds to the chi-square distribution with p-value <0.001. Therefore, there is strong evidence at the 1% level of significance to suggest that there is a difference between the goodness-of-fit of the models to the dataset. The fact that the correlated MLMM has a smaller deviance implies that it is a better fit to the data than the independent MLMM.

We have seen the evidence in favour of fitting a multivariate model as opposed to multiple univariate models from the MLMM itself (the statistically significant correlation between test scores), the information in *Chapter 2* and that the deviance statistic of the MLMM with correlated outcomes was significantly different to the deviance for the MLMM with independent outcomes (approximate to two univariate LMMs). Given these observations and results, the correlated MLMM (i.e. our fitted multivariate linear mixed model) was therefore, regarded as the better fitting model (i.e. the model of choice).

3.4.3 Model diagnostics

In terms of assessing the goodness-of-fit of a model, there are a number of different statistics and methods. The deviance, as we have defined in the previous subsection is a measure of goodness-of-fit. However, it is really only useful for comparing two models to check which is a more suitable fit to the data. It is not helpful in determining exactly how much of the variability in the outcome variables is explained by the model. An alternative method for analysing the fit of the model is to inspect the residuals and standardized residuals. We obtained fitted values and standardized for the individual models based on the marginal distributions from our multivariate model. Firstly, the standardized residuals for each model based on the marginal distributions of the MLMM were plotted against each other. This is an indicator of how correlated the two outcomes are. Figure 3.4 shows two such plots, one for the random effects and one for the random errors.

We can see in Figure 3.4(a) that there is little variability in the standardized residuals between the two outcomes at the local authority level. We saw that the estimated correlation in the random effects from the model was 0.97 and this very strong, positive correlation is also visible in the standardized residuals plot. Figure 3.4 (b) shows more variability at the pupil level (random errors) and strong, positive correlation, albeit not as strong as for the random effects. We have seen that our model suggested that this correlation was equal to 0.74.



(a) (b)

Figure 3.4: *Marginal distribution standardized residual plots for the (a) random effects and (b) random errors of the multivariate linear model with reading and transformed maths scores as outcomes.*

The next step was to plot a variety of diagnostic plots for the models based on the marginal distributions of the MLMM with respect to the random error standardized residuals (see Figure 3.5). For reading scores, the standardized residuals approximately follow a normal distribution (Figures 3.5(c) and (e)). However, the standardized residuals for the transformed maths scores look less normally distributed (Figures 3.5(d) and (f)). The mean of the unstandardized residuals was approximately zero for each set of test scores. The plots of fitted values against standardized residuals show that the standardized residuals are reasonably homoscedastic for both the reading and mathematics scores (Figures 3.5(a) and (b), respectively).

In addition to these plots, similar figures were constructed for the random effect standardized residuals. However, these were not included as they closely followed the main assumptions (i.e. close to being normally distributed and looking more homoscedastic for both outcomes than those for the random error terms). In conclusion, it can be said that the overall model diagnostics point toward a reasonably good fit to the data with a better fit to the reading scores than for the transformed mathematics scores.

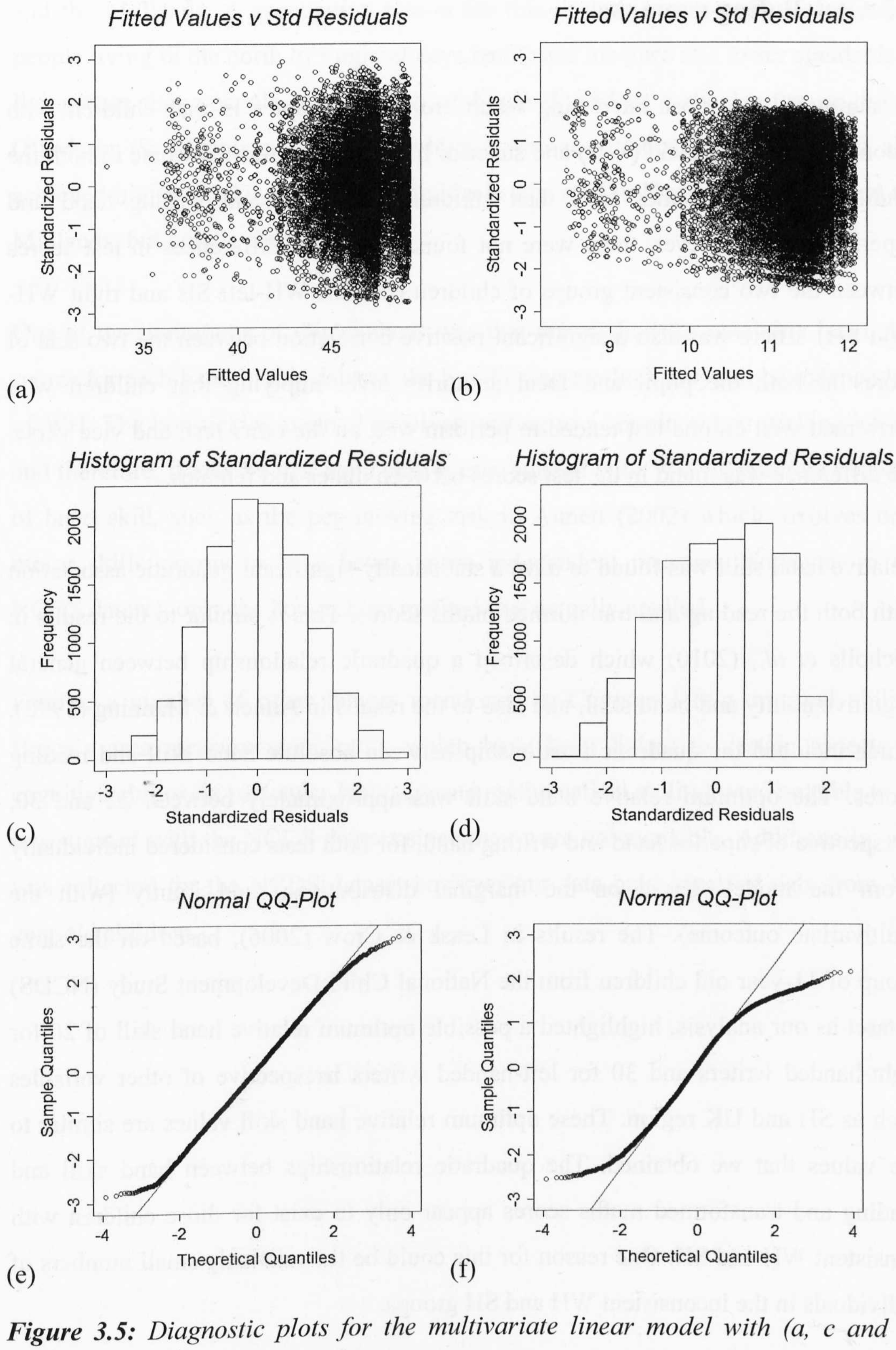


Figure 3.5: *Diagnostic plots for the multivariate linear model with (a, c and e) reading and (b, d and f) transformed mathematics scores as the outcome variables.*

3.5 Concluding remarks

In summary, the most interesting result from this chapter is that children with inconsistent writing hand (WH) and superior hand (SH) performed worse in both the reading and mathematics tests than children with consistent writing hand and superior hand. However, there were not found to be any differences in test scores between the two consistent groups of children (i.e. left WH-left SH and right WH-right SH). There was also a significant positive correlation between the two sets of scores at both the pupil and local authority level implying that children who performed well on one test tended to perform well on the other test, and vice versa. No difference was found in the test scores between males and females.

Relative hand skill was found to have a statistically significant quadratic association with both the reading and transformed maths scores. This is similar to the results in Nicholls *et al.*, (2010) which described a quadratic relationship between general cognitive ability and hand skill, and also to the results in Annett & Manning (1990), which included the quadratic relationship between absolute hand skill and reading scores. The optimum relative hand skill was approximately between 22 and 30, irrespective of superior hand and writing hand, for both tests considered individually (from the models based on the marginal distributions) and jointly (with the multivariate outcome). The results in Leask & Crow (2006), based on the same group of 11-year old children from the National Child Development Study (NCDS) dataset as our analysis, highlighted a possible optimum relative hand skill of 20 for right-handed writers and 30 for left-handed writers irrespective of other variables such as SH and UK region. These optimum relative hand skill values are similar to the values that we obtained. The quadratic relationships between hand skill and reading and transformed maths scores appear only to exist for those children with consistent WH and SH. One reason for this could be the relatively small numbers of individuals in the inconsistent WH and SH groups.

A difference between children who attended school in Southern England and their counterparts in Northern England and the Midlands was detected in both sets of test scores, with those in Southern England outperforming those in Northern England

and the Midlands. A speculative reason for this is that, over a period of decades, people living in the north of England have had lower incomes and lower standards of living than those people in the south of England, and in particular the south-east (Blackaby & Manning, 1990). Individuals who were educated in Scotland also achieved higher marks than those individuals educated in Northern England and the Midlands, but only in the maths test.

One of the limitations of this analysis was that the method for obtaining hand skill scores for each hand in this dataset, the box-ticking method, may not be independent of WH. The box-ticking method involves writing to a certain extent (ticking boxes) and therefore, it may be that hand skill is confounded with WH. A different measure of hand skill, such as the peg-moving task in Annett (2002) which involves only motor skills, seems to be a better, more independent measure. However, in the NCDS dataset, only the box-ticking method was actually applied.

Finally, a number of other factors mentioned in *Chapter 1* (e.g. musical ability, Broca's area measurements, etc.), which have been linked to certain aspects of cognitive ability (in particular language and mathematical ability) were not able to be investigated with the NCDS dataset since they were not available. Additionally, age was collected for the NCDS dataset however our data only involved data from 11-year old children.

CHAPTER 4

Investigation of the associations between musical ability and cognitive function in adults

The associations between handedness and cognitive ability in children (namely, reading and mathematics scores) were investigated in *Chapter 3*. In this chapter, we consider the links between cognitive ability and factors such as age, gender and musical ability in adults. Prior to giving the results from our statistical analyses, we introduce some past results in this particular area of research and then explain how we constructed our statistical models (referring to methodology from *Chapter 2*) and revisit the dataset breakdown from *Chapter 1*.

4.1 Introduction

In adults, differences between musicians and non-musicians in a number of cognitive function tests, including verbal, verbal memory and visuospatial tests have been observed (Sluming *et al.*, 2002; Sluming *et al.*, 2007; Franklin *et al.*, 2008; Jakobsen *et al.*, 2008). When musicians were compared to non-musicians in two different verbal memory tests (California Verbal Learning test, 2nd edition (CVLT-II) and Rey Auditory Verbal Learning test (RAVLT)), the same result occurred with musicians performing better (Franklin *et al.*, 2008; Jakobsen *et al.*, 2008). Furthermore, Jakobsen *et al.*, (2008) reported that in a visual memory test (Rey Visual Design Learning test (RVDLT)) given to participants, musicians obtained higher scores, on average, than non-musicians (controls). The Benton judgement of line orientation (JOL) test was used to evaluate spatial ability in Sluming *et al.* (2007) and it was seen that musicians achieved higher scores than controls. Evidence for this greater spatial ability in musicians was suggested earlier in Sluming *et al.* (2002) after comparing the two groups' results from a three-dimensional mental rotation (3DMR) task. A description of the CVLT-II, RAVLT, RVDLT, JOL and 3DMR tests can be seen in Section 1.2.4.

In addition to the associations between musical ability and cognitive ability, there are many other factors which are known to have an effect on cognitive ability, such as age (see e.g. Gunstad et al., 2006; Salthouse, 2006; Mitnitski & Rockwood, 2008; Mitnitski et al., 2010). For example, Salthouse (2006) suggested that there is a general decline in cognitive function with age starting when subjects are between 20 and 40 years of age. This decline occurred in three cognitive function tests: episodic memory, perceptual reasoning and perceptual speed. However, no statistically significant reduction in the scores from a vocabulary test was observed as age increased. When general cognitive ability scores were considered (the Modified Mini-Mental State Examination (3MS) (for more details, see Teng & Chui, 1987) and the Cognitive Abilities Screening Instrument), older-aged participants obtained lower marks, on average, (i.e. more errors) than younger participants (Mitnitski & Rockwood, 2008; Mitnitski *et al.*, 2010). In the study by Gunstad *et al.*, (2006) individuals were split into three age groups (young-age, middle-age and old-age) and scores across a range of cognitive ability tests (including attention, memory, executive functioning, language and motor skills) were shown to be lower, in general, for the older participants than for the young and middle-aged participants. Further discussion of previous results can be read in *Chapter 1*.

The objective of this chapter was to investigate the links between multiple cognitive ability scores and factors, such as musical ability, gender and age, simultaneously in a single statistical model. To do this we fitted multivariate linear models to a dataset of adults, with three cognitive ability scores, representing vocabulary, arithmetic and visuospatial abilities, as outcome variables. Interaction terms were also considered between the explanatory variables (musical ability, gender and age) as well as correlation between the outcomes. It is the increase in complexity of the applied statistical model (multivariate linear model) which allows for more refined decomposition of the variability within the dataset, and which is novel in comparison to the past studies.

4.2 The dataset

The dataset used in this chapter was collected at the Magnetic Resonance Image and Analysis Research Centre (MARIARC), University of Liverpool, in 2000. From now on we will refer to this dataset as the musician-control (MC) dataset. The MC dataset was obtained for the purpose of investigating differences between musicians and controls, in terms of the activation of various regions of the brain, mainly using voxel-based morphometry (VBM) and functional magnetic resonance imaging (fMRI) (Sluming *et al.*, 2002; Sluming *et al.*, 2007). The full MC dataset consisted of 149 participants (detailed information on the dataset can be found in Section 1.3.2). The outcome variables considered here were the vocabulary, arithmetic and block design test scores, all from the Wechsler Adult Intelligence Scale (WAIS). The WAIS block design test can be described as a visuospatial test and so, from this point onwards, the WAIS block design scores will be referred to as visuospatial scores.

Since the outcome variables (vocabulary, arithmetic and visuospatial scores) are taken directly from the WAIS, it was noted that these scores are scaled. However they are not scaled linearly, and therefore, the meaning of a one scaled unit increase in the actual test is difficult to interpret. This is explained in more detail in Section 1.3.2. The covariates considered were:

- *Gender* (0 = 'male', 1 = 'female')
- *Age* (in years)
- *Musician* (0 = 'non-musician' or 'control', 1 = 'musician')

Originally, we wanted to include handedness in this analysis as an explanatory variable but, due to the very small number of left-handers in the dataset, only right-handed individuals were considered. The missing data for the forthcoming analyses can be seen in Table 4.1. When all the variables were taken into account, the proportion of participants with at least one missing variable value was equal to 13.9% (see Table 4.1). This was also the case when only arithmetic and visuospatial scores were taken into account. However, when we had vocabulary score as the only outcome, then the number of individuals with at least one missing variable value was slightly fewer (19 or 13.2%, Table 4.1). For the analyses in this chapter, individuals with missing data in the variables included in the statistical models were omitted. Further information about missing data, along with the implications that omitting participants with missing variable values has on the results of the analyses in this chapter, can be seen in *Chapter 6*.

Variable	Missing (%)	Non-missing (%)
Vocabulary Test Scores	19 (13.2%)	125 (86.8%)
Arithmetic Test Scores	20 (13.9%)	124 (86.1%)
Visuospatial Test Scores	20 (13.9%)	124 (86.1%)
Musician	2 (1.4%)	142 (98.6%)
Gender	0 (0%)	144 (100%)
Age	0 (0%)	144 (100%)
Total*	19[a] (13.2%) 20[b] (13.9%)	125[a] (86.8%) 124[b] (86.1%)

*A number of subjects have multiple missing variable values and so are counted as 'missing' only once.
[a] These figures correspond to the univariate model with vocabulary scores as the outcome.
[b] These figures correspond to the univariate models with arithmetic and visuospatial scores as the outcome, as well as the multivariate model with all three cognitive scores as outcomes.

Table 4.1: *Missing and non-missing data for the outcome and explanatory variables of the MC dataset.*

4.3 Statistical models

In this chapter, we fitted a number of multivariate linear models, as described in Section 2.1.3, to the MC dataset. Our final, fitted multivariate model followed the structure of that in Equation (2.1.11) with the total number of participants, $n = 124$

after the omission of those individuals with missing data, the number of outcomes $d = 3$ (vocabulary, arithmetic and visuospatial scores) and 4 explanatory variables (i.e. $p - 1 = 4$; *musician* (we henceforth denote the factor *musician* in italics to differentiate it from a musician as a member of the group of individuals with musical ability), gender, *musician* and gender interaction and age). From Equation (2.1.11) we have that $\underline{Y}^* = \begin{bmatrix} \underline{Y}^{(1)} \\ \underline{Y}^{(2)} \\ \underline{Y}^{(3)} \end{bmatrix}$ is the $(3n \times 1)$ outcome vector with $\underline{Y}^{(k)}$ as previously defined, where $k = 1,2,3$ corresponds to the vocabulary, arithmetic and visuospatial scores, respectively. The design matrix $\underline{X}^* = \begin{bmatrix} \underline{X} & & \underline{0} \\ & \ddots & \\ \underline{0} & & \underline{X} \end{bmatrix}$ is a $(3n \times 3p)$ matrix with $\underline{X}$ as defined for Equation (2.1.6) where the explanatory variables were *musician* (x_{i1}), gender (x_{i2}), *musician* and gender interaction $(x_{i3} = x_{i1} \times x_{i2})$ and age (x_{i4}). We denoted the $(3p \times 1)$ vector of coefficients as $\underline{\beta}^* = \begin{bmatrix} \underline{\beta}^{(1)} \\ \underline{\beta}^{(2)} \\ \underline{\beta}^{(3)} \end{bmatrix}$ where $\underline{\beta}^{(k)}$ $(k = 1,2,3)$ is as defined for Equation (2.1.11). Finally, the random error vector $\underline{\varepsilon}^* = \begin{bmatrix} \underline{\varepsilon}^{(1)} \\ \underline{\varepsilon}^{(2)} \\ \underline{\varepsilon}^{(3)} \end{bmatrix}$ is a $(3n \times 1)$ vector with $\underline{\varepsilon}^{(k)}$ $(k = 1,2,3)$ as defined for Equation (2.1.11) and such that $\underline{\varepsilon}^* \sim N_{3n}(\underline{0}, \underline{\Sigma})$ where $\underline{\Sigma} = \begin{bmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} & \underline{\Sigma}_{13} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} & \underline{\Sigma}_{23} \\ \underline{\Sigma}_{31} & \underline{\Sigma}_{32} & \underline{\Sigma}_{33} \end{bmatrix}$ is a $(3n \times 3n)$ matrix with $(n \times n)$ matrices $\underline{\Sigma}_{kk}$ and $\underline{\Sigma}_{k_1k_2}$, $(1 \leq k_1, k_2 \leq d;\ k_1 \neq k_2)$ as previously defined (see Equation (2.1.11)).

From *Chapter 2*, it can also be seen that the covariance, and therefore correlation, can be estimated between each set of cognitive ability test scores. The analysis and plots were conducted and constructed, respectively using MLwiN version 2.16 and R version 2.11.1. The analyses were completed using Restricted Iteratively Generalised Least Squares (RIGLS) which is equivalent to restricted maximum likelihood estimation (REML) as described in Section 3.4.

4.4 Statistical analyses

We fitted a multivariate linear model (MLM) to the Musician-Control (MC) dataset using a similar approach to that for the analyses in *Chapter 3* (i.e. using a stepwise model selection method as defined in Section 3.4). The Wald test was again used to test the null hypotheses that the multivariate parameters equal zero. As mentioned in Section 3.4.1, the Wald test statistic follows a χ^2_d distribution where d is the number of outcomes (Goldstein, 2003). In this instance, $d = 3$. The results from the final, fitted multivariate model with the three cognitive ability scores as the outcome variables and following the omission of the individuals with missing data (124 participants left in the dataset, see Table 4.1) can be seen in Tables 4.2-4.5.

The main difference between a MLM and a univariate linear model (ULM) is that the MLM gives additional information about the associations between the explanatory variables and a linear combination of the outcomes. From the MLM we can also obtain results for the links between the explanatory variables and the individual outcomes from the models based on the marginal distributions as well as take the correlation between each outcome into account. The overall coefficients, standard errors, confidence intervals and p-values for the multivariate outcome (i.e. a linear combination of the vocabulary, arithmetic and visuospatial test scores, as shown in Equation (2.1.9)) are given in Table 4.2.

We found a positive association between *musician* and the combined cognitive ability test scores (p<0.001, Table 4.2). The coefficient for *musician* was positive which suggested that across all test scores combined, musicians outperform non-musicians (controls). Also, there was an overall link between age and the three cognitive ability scores combined and the coefficient was negative (p<0.001, Table 4.2). This can be interpreted as the older a participant was the lower their combined cognitive ability scores were.

Explanatory Variable	Coeff.	St. Error	95% CI	*p*-value
Intercept	37.0	1.3	(34.5, 39.4)	<0.001
Musician (0: Control, 1: Musician)	7.4	1.0	(5.5, 9.3)	<0.001
Gender (0: Male, 1: Female)	-1.0	0.6	(-2.6, 0.6)	0.2
Musician × *Gender*	1.2	1.7	(-2.1, 4.6)	0.5
Age	-0.1	0.02	(-0.17, -0.08)	<0.001

Table 4.2: *Estimates of the multivariate explanatory variable coefficients for the multivariate linear model with the vocabulary, arithmetic and visuospatial scores combined (including the standard error of the corresponding estimators, the 95% confidence intervals for the coefficients and the p-values).*

Although the coefficients for gender and the interaction term between gender and *musician*, were not significant (p=0.2 and p=0.5, respectively, Table 4.2), the contingency table of comparisons between the groups male and female, musicians and non-musicians was constructed, with the results shown in Table 4.3. There was not any overall difference in the combined cognitive ability test scores between males and females, in either the musician or control groups (p=0.9 and p=0.2, Table 4.3, respectively). However, there is a clear difference between the male and female musician groups and both the male and female control groups, with the musicians having higher cognitive ability scores than the non-musicians (all comparisons have p<0.001, Table 4.3). These results suggest that being able to play a musical instrument is strongly linked to the level of cognitive ability across the three defined cognitive ability scores (vocabulary, arithmetic and visuospatial).

		Musicians		Controls	
		Male	**Female**	**Male**	**Female**
Musicians	**Male**	X			
	Female	0.2 (0.9)	X		
Controls	**Male**	-7.4 (**<0.001**)	-7.6 (**<0.001**)	X	
	Female	-8.4 (**<0.001**)	-8.6 (**<0.001**)	-1.0 (0.2)	X

Table 4.3: *Cross-tabulated results (coefficients and p-values) for the overall effects of gender and* musician *on vocabulary, arithmetic and visuospatial scores combined (taken from the multivariate linear model).*

From the models based on the marginal distributions of each individual outcome, we obtained results for the associations between each cognitive ability test score and the explanatory variables. These results can be seen in Table 4.4.

As has been discussed in *Chapters 2* and *3*, due to the fact that the models based on the marginal distributions of the MLM, or equivalently ULMs, can be considered to be similar to multiple comparisons, then adjustments must be made to the p-values of these results (i.e. to the results in Table 4.4). Once again, the Bonferroni correction was used to modify the critical level at which a p-value becomes statistically significant (i.e. the level of α) to allow for the fact that we have three outcomes in the model. This can be shown to be:

$$\alpha^* = \alpha/3 \qquad (4.4.1)$$

where α^* is the adjusted critical value (Hair, Jr. *et al.*, 2010). We wanted to have a significance level of 5% (i.e. $\alpha = 0.05$), so α^* was set equal to 0.017 (to 2sf).

Interpreting the results from Table 4.4 using the Bonferroni corrected significance level of 0.017, we can see that *musician* had a strong association with all three cognitive ability scores individually (p<0.001, p=0.01 and p=0.004, for the vocabulary, arithmetic and visuospatial scores, respectively). The coefficient of gender was also significant for the vocabulary and arithmetic scores (p=0.01 and p=0.005, respectively, Table 4.4). However, we found no gender effect at the 5% level of significance for the visuospatial scores (p=0.08, Table 4.4).

Without interaction terms in the model, these results would be interpretable directly from Table 4.4. For example, with all three coefficients of *musician* being positive, this would suggest that musicians outperform non-musicians in all three tests regardless of gender. However, since the interaction term between *musician* and gender is included in the model, this interpretation is not strictly correct and needs further comparisons to explain the associations between musician, gender and the three cognitive ability test scores. Given the Bonferroni adjusted α-level of 0.017, we can also see that the interaction term between *musician* and gender, was not statistically significant for any of the three test scores, with the closest p-value to significance being for the visuospatial scores (p=0.06, Table 4.4). To interpret the results of the associations for each test score in relation to *musician* and gender, we compared the test scores for male and female, musicians and non-musicians using the models based on the marginal distributions of the MLM and multiple hypothesis tests (Wald tests). Table 4.5 shows the cross-tabulation of these hypotheses tests by cognitive ability and gives coefficients and p-values for each one.

Explanatory Variable	Coefficient	St. Error	95% CI	*p*-value
Vocabulary Test Scores				
Intercept	9.3	0.6	(8.2, 10.5)	<0.001
Musician (0: Control, 1: Musician)	4.7	0.4	(3.8, 5.6)	<0.001
Gender (0: Male, 1: Female)	1.0	0.4	(0.2, 1.7)	0.01
Musician × Gender	-0.8	0.8	(-2.4, 0.7)	0.3
Age	0.01	0.01	(-0.01, 0.03)	0.4
Arithmetic Test Scores				
Intercept	12.4	0.7	(11.0, 13.8)	<0.001
Musician (0: Control, 1: Musician)	1.4	0.5	(0.3, 2.5)	0.01
Gender (0: Male, 1: Female)	-1.3	0.5	(-2.2, -0.4)	0.005
Musician × Gender	0.6	1.0	(-1.3, 2.5)	0.5
Age	-0.04	0.01	(-0.06, -0.01)	0.002
Visuospatial Test Scores				
Intercept	15.3	0.6	(14.2, 16.4)	<0.001
Musician (0: Control, 1: Musician)	1.3	0.4	(0.4, 2.1)	0.004
Gender (0: Male, 1: Female)	-0.7	0.4	(-1.4, 0.07)	0.08
Musician × Gender	1.5	0.8	(-0.07, 3.0)	0.06
Age	-0.1	0.01	(-0.12, -0.08)	<0.001
Random Error	**Estimate**	**St. Error**	**95% CI**	***p*-value**
Variance (*Vocabulary*)	3.3	0.4	(2.4, 4.1)	<0.001
Variance (*Arithmetic*)	4.8	0.6	(3.6, 6.0)	<0.001
Variance (*Visuospatial*)	3.1	0.4	(2.3, 3.8)	<0.001
Covariance [Correlation] (*Vocabulary/Arithmetic*)	0.6 [0.2]	0.4 [0.1]	(-0.06, 1.4) [-0.02, 0.4]	0.07
Covariance [Correlation] (*Vocabulary/Visuospatial*)	0.2 [0.1]	0.3 [0.1]	(-0.3, 0.8) [-0.09, 0.3]	0.4
Covariance [Correlation] (*Arithmetic/Visuospatial*)	0.9 [0.2]	0.4 [0.09]	(0.2, 1.6) [0.05, 0.4]	0.01

Table 4.4: *Estimates of the explanatory coefficients and random error variance, covariance and correlation terms for the multivariate linear model with vocabulary, arithmetic and visuospatial scores as outcomes (including the standard error of the corresponding estimators, the 95% confidence intervals for the coefficients/terms and the p-values).*

Our cross-tabulated hypotheses tests gave results which suggested that females obtain a higher vocabulary score than males in the control group by 1 scaled mark on average (p=0.01, Table 4.5). A gender difference was again found in the arithmetic scores of the control group, although it was males that outperformed females by 1.3 scaled marks on average (p=0.005, Table 4.5). However, we found no difference between male and female non-musicians in the visuospatial scores (p=0.08, Table 4.5). Table 4.5 does show that in all three sets of test scores, there was no difference between males and females in the musicians group (p=0.9, p=0.4 and p=0.2, for the vocabulary, arithmetic and visuospatial scores, respectively). If we had not kept the musician and gender interaction term in the model, then our results from Table 4.4 would have been interpreted such that there was a gender difference in vocabulary and arithmetic scores irrespective of whether participants were musicians or non-musicians. However, we can clearly see from Table 4.5 that we have a gender difference in the control group, but not in the musician group. These gender differences between musician and non-musician groups would not have been identifiable without the inclusion of the interaction term in the model.

In the vocabulary scores, male musicians achieved higher scores, by 4.7 and 3.8 scaled marks on average, than male and female controls, respectively (p<0.001 for both comparisons, Table 4.5). This result was also replicated with female musicians, since they had higher vocabulary scores than male and female controls by 4.8 and 3.9 scaled marks, respectively (both p<0.001, Table 4.5). Male musicians were also found to have higher arithmetic scores on average than male and female controls by 1.4 and 2.7 scaled marks, respectively (p=0.01 and p<0.001, Table 4.5). Female musicians, however, had no significantly different arithmetic scores, on average, to female and male controls (p=0.02 and p=0.4, Table 4.5, respectively). These non-significant results would not have been identified without the interaction term, *musician* × gender, being a covariate in the MLM. We found that in the visuospatial scores, both male and female musicians outperformed male controls, by 1.3 and 2.1 scaled marks, respectively (p=0.004 and p=0.002, Table 4.5), and female controls by 1.9 and 2.7 scaled marks, respectively (p<0.001 in both cases, Table 4.5). Some of these significant comparisons can be easily seen in Figure 4.1 which illustrates the results when the 95% confidence intervals of the mean test scores of different groups of gender and musician do not overlap for the three test scores individually

		Vocabulary Test Scores				Arithmetic Test Scores				Visuospatial Test Scores			
		Musicians		Controls		Musicians		Controls		Musicians		Controls	
		Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Musicians	Male	X				X				X			
Musicians	Female	0.1 (0.9)	X			-0.7 (0.4)	X			0.8 (0.2)	X		
Controls	Male	-4.7 (**<0.001**)	-4.8 (**<0.001**)	X		-1.4 (**0.01**)	-0.7 (0.4)	X		-1.3 (**0.004**)	-2.1 (**0.002**)	X	
Controls	Female	-3.8 (**<0.001**)	-3.9 (**<0.001**)	1.0 (**0.01**)	X	-2.7 (**<0.001**)	-2.0 (0.02)	-1.3 (**0.005**)	X	-1.9 (**<0.001**)	-2.7 (**<0.001**)	-0.7 (0.08)	X

Table 4.5: *Cross-tabulated results (coefficients and p-values) for the effects of gender and* musician *on vocabulary, arithmetic and visuospatial scores (taken from the multivariate linear model).*

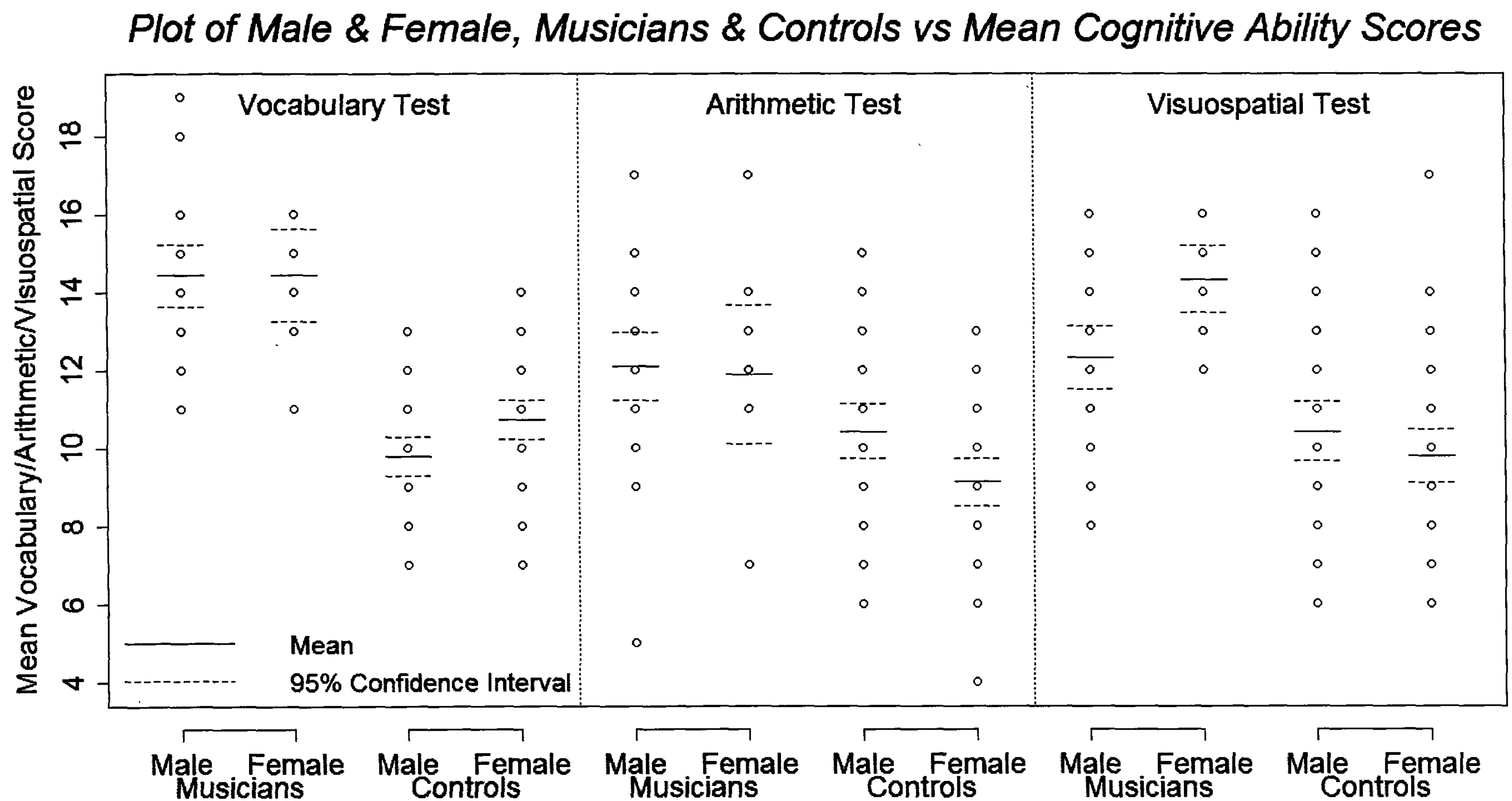


Figure 4.1: Plot of male/female, musicians/non-musicians (controls) by mean vocabulary, arithmetic and visuospatial scores.

From Table 4.4 we can also see that there is a negative association between age and both arithmetic and visuospatial scores (p=0.002 and p<0.001, respectively). For the arithmetic scores this is interpreted such that two participants with a ten-year age difference have a difference in arithmetic score of 0.4 scaled marks, on average, with the older person scoring lower than the younger person. A similar interpretation can be made for the visuospatial scores, with the older individual, by ten years, scoring, on average, 1 scaled mark less than the younger person.

From our multivariate linear model, we were able to observe the estimated variance in the three cognitive ability scores as well as the covariance and correlation between the test scores. The correlation, r, between the arithmetic and visuospatial scores was estimated by the model to be equal to 0.2 (p=0.01, Table 4.4). Although the p-value is statistically significant at the Bonferroni-corrected 5% level of significance, we can see that this means that there is evidence to suggest that the correlation does not equal 0, and does not mean that the correlation is biologically significant. It is clear that a correlation of 0.2 is a very weak, positive correlation. The other two correlation estimates in the model, between the vocabulary and arithmetic scores and between the vocabulary and visuospatial scores, were not significant, suggesting that there was no evidence at the Bonferroni-corrected 5% level of significance to reject the null hypothesis that the correlation was equal to zero (p=0.07 and p=0.4, Table 4.4, respectively).

We can compare the correlation estimates from the model to correlation estimates obtained directly from the dataset. To do this comparison, the Pearson correlation coefficients from the data were obtained and plots constructed of each test score against each other (see Figure 4.2). In each plot, we also plotted a line of best fit which gives a more visual indication of the correlation. The Pearson correlation coefficient, ρ, between the arithmetic and visuospatial scores, obtained directly from the data, was estimated to be ρ=0.5 (p<0.001). This is a higher estimate than we obtained from the model (r=0.2, Table 4.4).

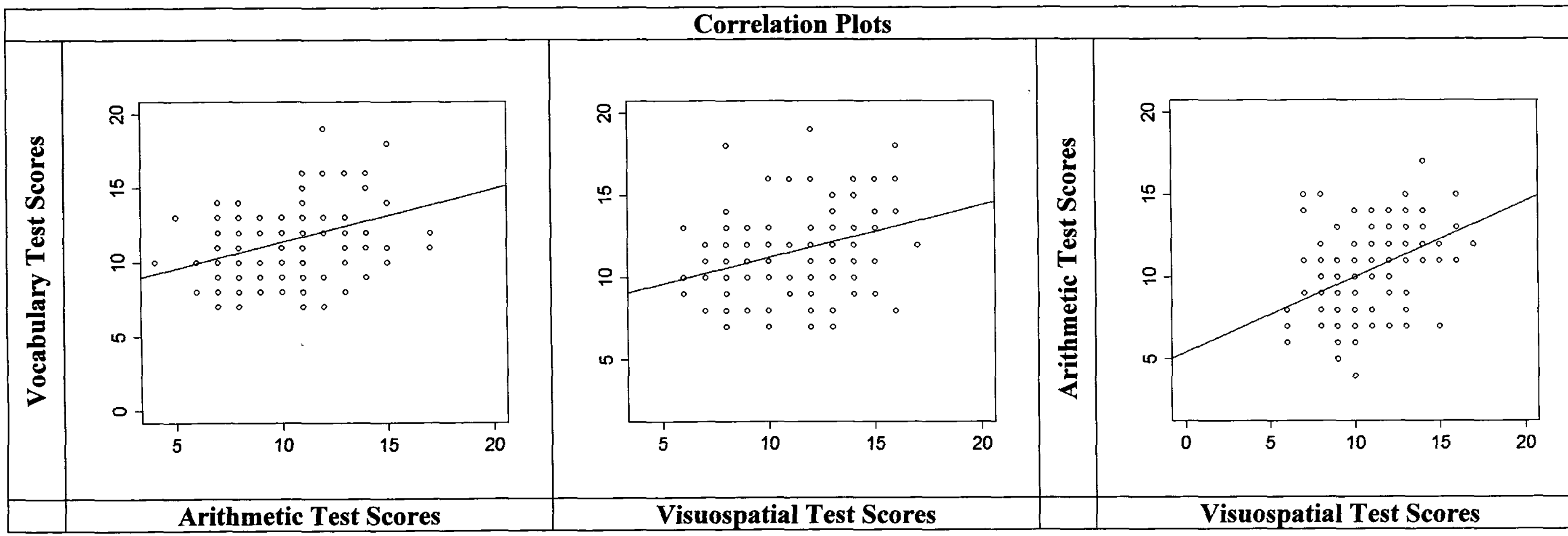


Figure 4.2: Plots of vocabulary scores against arithmetic and visuospatial scores, arithmetic scores against visuospatial scores, along with lines of best fit in each case.

A similar result for the multivariate model estimating lower correlations between test marks than the data itself is also observed between the vocabulary and arithmetic scores (ρ=0.4 (p<0.001) from the data compared to r=0.2 from the model, Table 4.4) as well as between the vocabulary and visuospatial scores (ρ=0.3 (p<0.001) from the data compared to r=0.1 from the model, Table 4.4). Therefore, once the multiple factors have been adjusted for in the model, the values of r were estimated to be between 50% and 67% less than the values of ρ which were observed from the data.

4.4.1 Comparison between multivariate and univariate linear models

The construction of our fitted univariate linear models (ULMs) followed a stepwise model selection process as did the earlier fitting of the multivariate linear model (MLM). However, there were differences between the number of explanatory variables in the ULMs and those in the models based on the marginal distributions of the MLM. Some explanatory variables had non-significant coefficients for individual outcomes in the MLM (e.g. age had a p-value of 0.4 in the vocabulary scores) and these variables were not included in the ULMs. Thus, the three ULMs did not give identical coefficients or p-values to the equivalent models based on the marginal distributions from the MLM. In addition, the ULM with vocabulary score as the outcome had one less participant in the sample for the MLM because of missing data (see Table 4.1). These slight differences in coefficients and p-values were not sufficient to change the interpretation from those given from the MLM marginal distributions in Section 4.4.

We subsequently addressed the question of whether the three ULMs or the MLM gave a better fit to the data. In Section 2.1.3 we described how the MLM is more complex than a ULM. This is because, the MLM also considers the correlation between outcomes whereas ULMs do not. In ULMs, the outcomes are considered to be independent of other outcomes. However, we can construct a MLM with independent outcomes by setting:

$$\sigma_{1,2} = \sigma_{1,3} = \sigma_{2,3} = 0 \qquad (4.4.2)$$

in Equation (2.1.11) (i.e. setting the correlation between the three outcomes equal to zero). Since the ULMs did not include all the explanatory variables that were in the MLM (the non-significant explanatory variables in the marginal distribution results for the MLM in Table 4.4 were not included in the ULMs) we also had to set coefficients for those particular explanatory variables equal to zero. The resulting MLM with independent outcomes (independent MLM) is therefore, a reasonable approximation to a combination of three ULMs. It was this independent MLM which we then compared with the correlated MLM (i.e. the MLM with correlated outcomes in Tables 4.2 and 4.4).

It is clear to understand that the MLM with independent outcomes (model 1) is nested within the MLM with correlated outcomes (model 2) and, as discussed in Section 3.4.2, this means that we were able to use a deviance test to compare the two models. The deviance of a model is defined in Equation (3.4.5) and is a goodness-of-fit statistic. As seen in *Chapter 3*, the deviance test statistic is simply the difference between two models' deviances. In this case, one model is representative of the three univariate models combined, and the other is the multivariate model. The deviance test statistic also has degrees of freedom equal to the difference in the degrees of freedom of the two models being tested. The deviances and the degrees of freedom for the correlated and independent MLMs can be seen in Table 4.6.

The deviance statistic between the independent and correlated MLMs, D^*, can then be written as:

$$D^* = |D_1^* - D_2^*| = |1527.261 - 1513.831| = 13.43 \qquad (4.4.3)$$

where D_1^* is the deviance of the nested model (independent MLM) within the 'full' model (correlated MLM) which has deviance D_2^*. D^* has degrees of freedom (df) equal to the difference in degree of freedom between the two models (i.e. $\text{df} = 357 - 351 = 6$). The test statistic (13.43 with df=6) can be regarded as a χ^2 statistic and corresponds to the χ^2 distribution with p-value 0.04. Therefore, there is evidence at the 5% significance level to suggest a difference between the goodness-of-fit of the models. Since the correlated MLM has the smallest deviance, it is the better fitting model. When accumulating the evidence, including that there is some correlation between the outcomes (see Figure 4.2), it can be concluded that there is some improvement in the fit of the correlated MLM over the independent MLM and thus also the ULMs. Therefore, the concluding remark here is that the MLM would be the model of choice to fit to this dataset.

#	Model	Deviance	df
1	Independent MLM	1527.261	357
2	Correlated MLM	1513.831	351

Table 4.6: *Table showing the deviances and degrees of freedom of the independent MLM and the correlated MLM, with vocabulary, arithmetic and visuospatial scores as outcomes.*

4.4.2 Model diagnostics

To assess the goodness-of-fit of a model a range of statistics can be used (e.g. R^2, adjusted R^2, Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC)). The deviance can also be classed as a goodness-of-fit statistic, but it is really only useful for comparing two models to determine which is a more suitable fit to the data. The AIC and BIC are based around the calculation of the log-likelihood of the model (as is the deviance) and are useful for comparing models, but not for determining a model's fit to the data (for details of AIC see Jobson, 1992; for details of BIC see Pinheiro & Bates, 2000). The adjusted R^2 statistic is also useful for comparing models as it takes into account the number of parameters in a model, but is not particularly helpful with the goodness-of-fit of a model to the data (see Ugarte *et al.*, 2008 for more details).

R^2, otherwise known as the coefficient of determination, can be calculated as:

$$R^2 = 1 - SS_{res}/SS_{tot} \qquad (4.4.4)$$

where SS_{res} and SS_{tot} are the residual and total sum of squares, respectively (Ugarte *et al.*, 2008). The statistic R^2 takes values between 0 and 1, and when multiplied by 100 represents the proportion of variability in the outcome variable which is explained by the explanatory variables in the model, as a percentage. That is, the closer an R^2 value is to 1, the better the fit of the model to the data. The R^2 statistic was unobtainable for the MLM, but it was calculable for the three ULMs with each cognitive ability score as the outcome. Since we already know that the MLM is a better fit to the model, we can say that the MLM will have a higher R^2 value than the ULMs. Therefore, we can say that the proportion of variability explained by the MLM was greater than that explained by the ULMs. The values of R^2 obtained for each of the three ULMs were:

- $R^2 = 0.545$ (ULM with vocabulary scores as the outcome).
- $R^2 = 0.276$ (ULM with arithmetic scores as the outcome).
- $R^2 = 0.585$ (ULM with visuospatial scores as the outcome).

These R^2 values indicate that the univariate models explained 54.5%, 27.6% and 58.5% of the total variability in the vocabulary, arithmetic and visuospatial scores, respectively. The ULM with arithmetic score as the outcome is the poorest fitting ULM to the data. Even the ULMs with vocabulary and visuospatial scores are not that good at explaining the variability in the outcomes. However, what we can say is that the MLM must explain at least the same proportion of variability in the outcome if not a higher proportion, than we have seen with the three ULMs.

In addition to the R^2 statistic, we can analyse the fit of a model by inspecting the standardized residuals. Using the standardized residuals extracted from the models based on the marginal distributions of the MLM in Table 4.4, we constructed plots of the standardized residuals for each outcome against each other (see Figure 4.3). These plots give an indication of the correlation between the outcomes as estimated by the model. We can see that there is little correlation in all the plots of Figure 4.3. However, this is understandable given the fact that the correlation estimates from the MLM between the outcomes were very low (0.1 or 0.2 in all three cases).

Figure 4.4 shows the diagnostic plots for the models based on the marginal distributions of the MLM. Overall, the conclusion of the diagnostics is that the fit of the model seems reasonable for the visuospatial scores, but not as good a fit to the vocabulary and arithmetic scores. The standardized residuals look to be normally distributed for the visuospatial scores (Figures 4.4 (f) and (h)). However, even though the histograms of standardized residuals look to be reasonably normally distributed, the normal-QQ plots suggest that it is less likely that the standardized residuals for both the vocabulary and arithmetic scores follow a normal distribution (see Figures 4.4 (d), (e), (g) and (h) for the histograms and normal-QQ plots for the vocabulary and arithmetic scores, respectively). When we used the Anderson-Darling normality test to check the normality of the standardized residuals, there was no evidence to reject the hypothesis that they were normally distributed at the 5% level of significance (p=0.06, p=0.1 and p=0.5 for the vocabulary, arithmetic and visuospatial scores, respectively).

Even though there may be a slight heteroscedascity of the standardized residuals of the arithmetic and vocabulary marks, overall, the standardized residuals do appear to be reasonably homoscedastic for all three outcomes (Figures 4.4(a), (b) and (c)). Also, the means of the unstandardized residuals were approximately zero for each outcome. In conclusion, it can be said that the R^2 statistics and the model diagnostics show that this model is reasonable, but not particularly good at explaining a large proportion of the variability within the three cognitive ability scores.

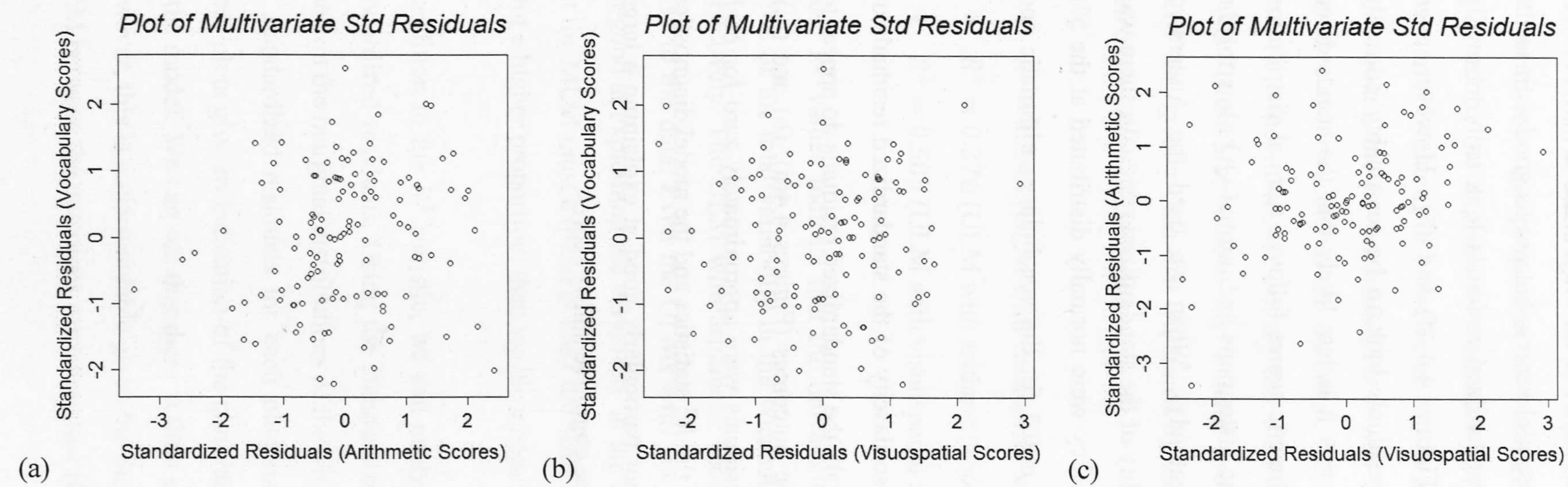


Figure 4.3: *Marginal distribution standardized residual plots for the (a) arithmetic and vocabulary (b) visuospatial and vocabulary and (c) visuospatial and arithmetic scores from the multivariate linear model.*

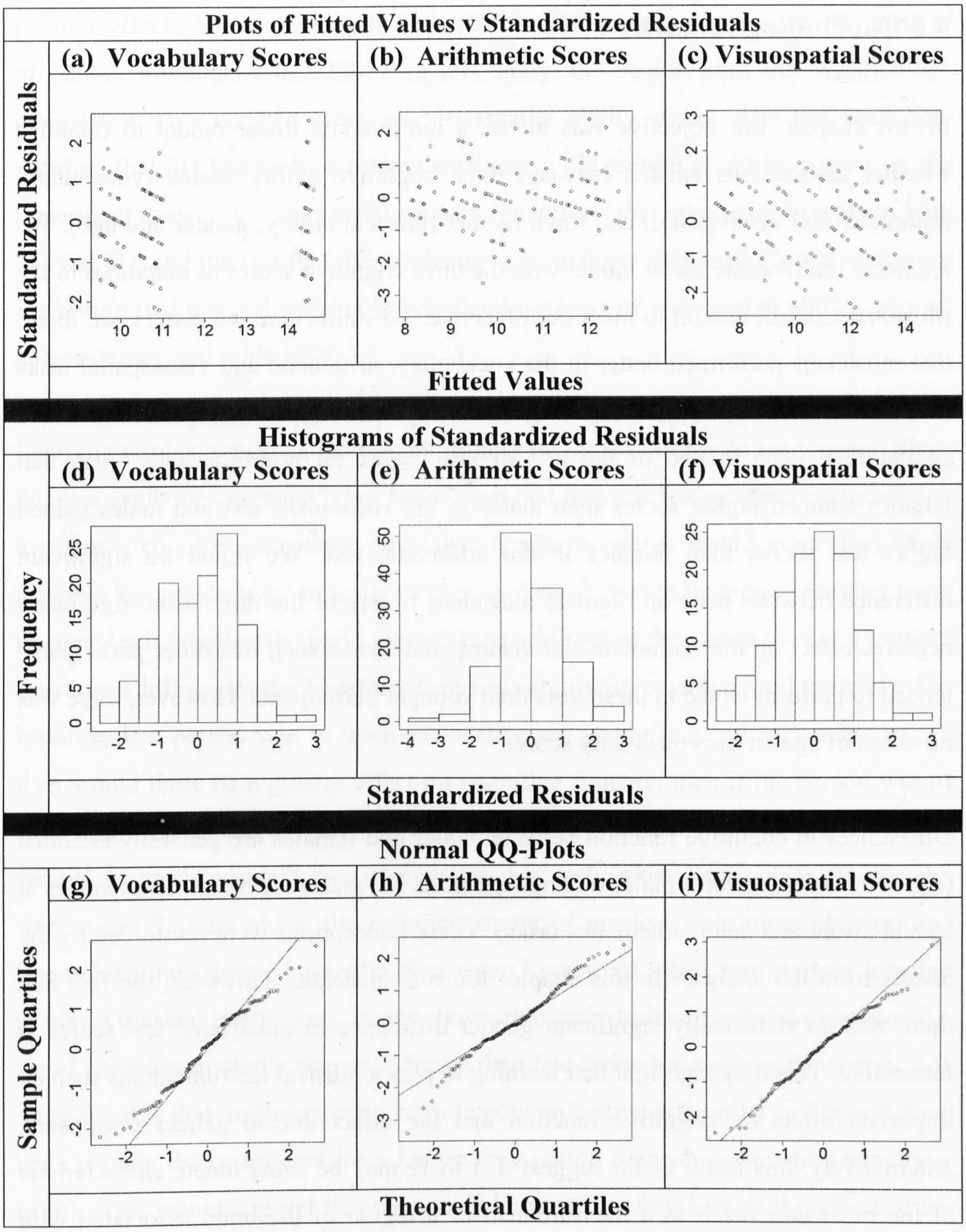


Figure 4.4: Diagnostic plots for the multivariate linear model with (a, d and g) vocabulary, (b, e and h) arithmetic and (c, f and i) visuospatial scores as the outcome variables.

4.5 Concluding remarks

In this chapter, our objective was to use a multivariate linear model to establish whether associations existed between three cognitive ability scores (vocabulary, arithmetic and visuospatial) and three factors (musical ability, gender and age). We applied a multivariate linear model with the three cognitive scores as outcomes to the Musician-Control dataset to meet our objective. To summarise the results, we found that musicians performed better in the vocabulary, arithmetic and visuospatial tasks than non-musicians, irrespective of gender. Furthermore, our analysis indicated a gender difference in two of the sets of test scores for non-musicians, such that females attained higher scores than males in the vocabulary test and males gained higher test scores than females in the arithmetic test. We found no significant difference between male and female musicians in any of the three tests. Age had a negative effect on the arithmetic and visuospatial scores such that older participants tended to perform worse in these tests than younger participants. However, there was no effect of age on the vocabulary scores.

Differences in cognitive function between males and females are generally assumed (e.g. females outperform males in language skills but males are better than females at spatial awareness and mathematics tasks). These conceptions were reinforced by the results from the analyses in this chapter for non-musicians. However, the fact that there was no statistically significant gender difference in musicians' test scores is interesting. This may highlight that learning to play a musical instrument has such an important effect on cognitive function that the effect due to gender disappears. Alternatively, this result could suggest that there may be some innate characteristic of the musicians (such as a measurement of a region of the brain associated with both cognitive and musical ability (e.g. Broca's area)) which has to be present, irrespective of gender. This would mean that an individual would have a characteristic which is linked to both a higher cognitive and musical ability, overriding the gender effect.

Musicians outperformed non-musicians in all three tests irrespective of gender. We previously noted that the effect of learning to play a musical instrument had a positive effect on spatial ability in both a 2D and 3D mental rotation task (Sluming *et al.*, 2002; Sluming *et al.*, 2007). In this study, the scores from the WAIS block design test were used to represent visuospatial ability scores. The test used here required that 3D blocks be arranged such that a 2D pattern could be copied on the uppermost faces of the rearranged squares. Therefore, this involved elements of both 2D and 3D, and the fact that the results agree with those previously found reinforces the notion that musical training is beneficial in terms of visuospatial ability, as well as vocabulary and arithmetic.

Learning to play a musical instrument usually involves learning to read music. Since this is a symbolic language it has been suggested that this has an effect on increasing a person's cognitive functions other than language, which would more than likely improve if a person were to learn another spoken or written language. Another point to consider is whether the period in a person's life when they learn to play a musical instrument influences the benefits of musical ability on cognitive ability or not. For instance, if a person was to learn while they are young (e.g. primary school age), then would there be a greater effect on cognitive function than if the person was of adult age (e.g. over 18 years old)? The study of Watanabe *et al.*, (2007) split a sample of musicians into two distinct groups: early trained musicians (those who were under 7 years old at the commencement of musical instrument playing) and late-trained musicians (those who were over 7 years old when they starting learning to play a musical instrument). It was shown that early-trained musicians performed a novel complex rhythmic sequence better than late-trained musicians. However, it is quite feasible that musicians who learn to play an instrument earlier in life will have been playing the instrument for a longer time period than someone who learns later in life. This would suggest that there may be some confounding of the effect of the number of years of experience of playing a musical instrument and the age at which learning to play the instrument commenced.

Furthermore, we considered that the effect of playing different types of instruments may also have a bearing on the benefits of being a musician in terms of cognitive ability. This is because different types of instrument require different skills. Ideally we would have liked to investigate these effects in our study, but due to limitations of the dataset (e.g. the number of participants, the lack of an age at the commencement of musical instrument training, etc.), we did not do so.

The negative effect of age on the arithmetic and visuospatial scores follows the generally accepted view that there is an age-related decline in memory, perceptual reasoning and perceptual speed (Jacobs *et al.*, 2001; Salthouse, 2006). However, in our sample (age range of 19-94 years old) we found no evidence for an age-related change in vocabulary scores, which matches with the findings of Salthouse (2006). In the study of Salthouse (2006) a large group of participants ranging in age (in years) from early-20's up to mid-80's, no differences were found in the average WAIS vocabulary score between older individuals and younger individuals. Differences between the effects of age on vocabulary scores to those for arithmetic and visuospatial scores may be explained by the different mechanisms that govern linguistic and non-linguistic functions. Linguistic functions appear to rely on previous knowledge that has been gained and used comprehensively, over a long period of time thereby becoming almost intuitive (i.e. using areas of long-term memory) rather than rely on short-term memory. However, non-linguistic functions require information to be processed at the time of assessment, as well as being reliant upon both short- and long-term memory for previous methodological information.

We did not find *musician* by age interaction to have a statistically significant coefficient for any of the three cognitive ability scores. Therefore, we did not have any evidence to suggest that the age-related effects on the three scores were any different for musicians to non-musicians. However, our group of musicians consisted of just 36 individuals (age range 24-65 years) compared to 106 non-musicians (age range 19-94 years). Thus, in a future study, it would be advisable to include more musicians with a larger range of ages to have a more balanced dataset (between musicians and non-musicians).

CHAPTER 5

Investigation of the associations between Broca's area and cognitive function in adults

In the previous two chapters we have considered associations between cognitive ability and a range of factors, in both children and adults. In addition to handedness, musical ability and the other factors we considered, there are certain regions of the brain which are linked with cognitive ability. One such region is Broca's area which is historically linked with language. In this chapter, we consider the links between measurements of Broca's area (volume and surface area) and factors such as cognitive ability, age, gender and musical ability in adults. We begin this chapter by describing some results from previous research in this area and then follow this by explaining the methodology used on the data defined. After the results are given, a further subsection explains the within observer variability from our volume and surface area estimates of Broca's area.

5.1 Introduction

Broca's area is a region of the brain associated with various language functions (Broca, 1861; Caplan, 2006; Dronkers *et al.*, 2007; see also Section 2.3.2). Broca's area was first associated with language function by Paul Broca in 1861 following his observations from the post-mortem brains of two men, named Leborgne and Lelong, who both had a very limited vocabulary (Broca, 1861; Dronkers *et al.*, 2007). Other language related functions such as speech production and syntactic processing have also been strongly linked with either the inferior frontal gyrus (in which Broca's area resides) or Broca's area itself (Stephan *et al.*, 2003, Toga & Thompson, 2003; Caplan, 2006; Costafreda *et al.*, 2006).

A variety of areas of the brain have been identified where differences have been observed between musicians and non-musicians (Schlaug *et al.*, 1995; Amunts *et al.*, 1997; Sluming *et al.*, 2002; Gaser & Schlaug, 2003; Lee *et al.*, 2003; Bengtsson *et* *al.*, 2005; Schneider *et al.*, 2005; Bangert & Schlaug, 2006; Stewart, 2008). Some areas have been highlighted as being active when music is being processed (Stewart *et al.*, 2003; Stewart, 2005; Sluming *et al.*, 2007). Broca's area has also been identified as one such region that has some interaction with music processing (Sluming *et al.*, 2007). Furthermore, it has also been suggested that Broca's area is involved in the processing of non-linguistic information including sequencing, action recognition and visuospatial cognition, as well as music (Koelsch *et al.*, 2002; Schubotz *et al.*, 2002a,b; Hamzei *et al.*, 2003; Sluming *et al.*, 2007). In terms of comparing functionality of Broca's area between musicians and controls, musicians have been found to have greater activation in Broca's area than non-musicians (Sluming *et al.*, 2002; Sluming *et al.*, 2007). More detailed descriptions of previous results can be seen in *Chapter 1.*

The aim of this chapter was to examine the relationships between measurements of Broca's area (more specifically, volume and surface area) and an array of different variables such as cognitive ability, age, gender and musical ability, using more complex statistical methods which take into account both the correlation between Broca's area in each hemisphere as well as interaction terms between the explanatory variables, where necessary. To complete this aim, we used multivariate linear models fitted to a subset of data from the larger Musician-Control dataset used in *Chapter 4.* Our outcome variables were modifications of the volume and surface area estimates of Broca's area, as defined in *Chapter 1*, which took the total brain volume and the volume of Broca's area into account, respectively. It is the complex nature of the analysis, and the fact that interaction terms and correlation between outcomes were considered which add novelty to our study.

To investigate the relationship between Broca's area and cognitive ability, we could have simply extended the analysis conducted in the previous chapter considering measurements of Broca's area (more specifically, volume and surface area) as additional explanatory variables. However, if no significant association between the Broca's area measurements and the cognitive ability scores are found, then this is all the information that can be gained from that extension of analysis. By setting the Broca's area measurements as the outcomes, then we were not only able to explore the possible links between Broca's area and cognitive ability, but also possible relationships between Broca's area and gender, age and musical ability.

5.2 The dataset

We used a subset of the Musician-Control (MC) dataset in the analyses within this chapter. The full MC dataset was not used due to volume and surface area estimates of Broca's area being obtained for the subset only. This subset comprised of 19 musicians (9 females and 10 males) and 20 non-musicians (10 females and 10 males) (i.e. 39 individuals in total). The full dataset contained data for just 9 female musicians and hence, it was not possible for us to obtain a balanced dataset in terms of male and female, musicians and non-musicians. It is useful here to point out that there are no missing values in this dataset and further, more detailed information about this dataset can be seen in Section 1.3.3. The covariates considered were:

- *Gender* (0 = male, 1 = female)
- *Age* (in years)
- *Musician* (0 = non-musician, 1 = musician)
- *WAIS vocabulary scores*
- *WAIS arithmetic scores*
- *WAIS block design (visuospatial) scores*

which were the outcome and explanatory variables considered in the analyses within *Chapter 4*. Note that similarly to *Chapter 4*, we denote the factor *musician* in italics so that it is easier to identify when we refer to the factor as opposed to a musician in general.

The outcome variables considered were adjusted values of the volume and surface area estimates of Broca's area. The volume estimates of Broca's area (henceforth denoted by Broca's volume) were modified by dividing them by the total brain volume of each participant and then multiplying by 100 (see Equation (1.3.3)). The Broca's volume estimates relative to total brain volume (RelBAV) can be interpreted as expressing the volume of Broca's area as a percentage of the total brain volume. We performed this adjustment because each individual's brain size may be associated with the size of Broca's area (i.e. a scaling effect may occur). Broca's area exists in each hemisphere of a person's brain, hence, estimates were obtained and adjusted for each hemisphere of every participant included in the subset. To investigate whether associations between Broca's area and the explanatory variables existed across both hemispheres or just in one, both left and right hemisphere RelBAV estimates were included as the two outcomes in our statistical models.

The surface area estimates of Broca's area (henceforth referred to as Broca's surface area) were adjusted by dividing through by Broca's volume to the power of 2/3 (see Equation (1.3.4)). We divided through by Broca's volume to the power of 2/3 to correct for the fact that volume is a 3-dimensional measurement (in units of cm^3), whereas surface area is only 2-dimensional (in units of cm^2). This new adjusted variable, Broca's surface area relative to Broca's volume (RelBASA), was of interest because it provides information about the convoluted nature of Broca's area, and therefore, it can be regarded as a type of dimensionless shape factor. For similar reasons as for including the RelBAV estimates from the left and right hemispheres, we included the left and right RelBASA estimates as two outcomes in our models.

5.3 Statistical models

In this chapter we firstly fit a multivariate linear model to the dataset with left and right RelBAV estimates as the outcomes. In Section 5.4 our fitted multivariate model had a similar structure to that in Equation (2.1.11). In this multivariate model we had $n = 39$ (i.e. the number of participants), $d = 2$ (i.e. the number of outcomes; left and right RelBAV) and 1 explanatory variable (i.e. $p - 1 = 1$; gender). From Equation (2.1.11) we have that:

$$\underline{Y}^* = \begin{bmatrix} \underline{Y}^{(1)} \\ \underline{Y}^{(2)} \end{bmatrix} \tag{5.3.1}$$

is the $(2n \times 1)$ outcome vector with $\underline{Y}^{(k)}$ as previously defined, where $k = 1,2$ corresponds to the left and right RelBAV estimates, respectively. The design matrix $\underline{X}^*$ is a $(2n \times 2p)$ matrix such that:

$$\underline{X}^* = \begin{bmatrix} \underline{X} & & \underline{0} \\ & \ddots & \\ \underline{0} & & \underline{X} \end{bmatrix} \tag{5.3.2}$$

with $\underline{X}$ as defined for Equation (2.1.6) where the explanatory variable was gender (x_{i1}). We denoted the $(2p \times 1)$ vector of coefficients as:

$$\underline{\beta}^* = \begin{bmatrix} \underline{\beta}^{(1)} \\ \underline{\beta}^{(2)} \end{bmatrix} \tag{5.3.3}$$

where $\underline{\beta}^{(k)}$ $(k = 1,2)$ is as defined for Equation (2.1.11). We defined the $(2n \times 1)$ random error vector, $\underline{\varepsilon}^*$, to be:

$$\underline{\varepsilon}^* = \begin{bmatrix} \underline{\varepsilon}^{(1)} \\ \underline{\varepsilon}^{(2)} \end{bmatrix} \tag{5.3.4}$$

with $\underline{\varepsilon}^{(k)}$ $(k = 1,2)$ as defined for Equation (2.1.11) such that $\underline{\varepsilon}^* \sim N_{2n}(\underline{0}, \underline{\Sigma})$ where $\underline{\Sigma} = \begin{bmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{bmatrix}$ is a $(2n \times 2n)$ matrix with $(n \times n)$ matrices $\underline{\Sigma}_{kk}$ and $\underline{\Sigma}_{k_1 k_2}$, $(1 \leq k_1, k_2 \leq d;\ k_1 \neq k_2)$ also as previously defined (see Equation (2.1.11)). The variance for the random errors of the left and right RelBAV estimates can be written as σ_1^2 and σ_2^2, respectively, with covariance between the two outcomes equal to $\sigma_{1,2}$.

In Section 5.5 we then focus on the associations between the RelBASA estimates and the explanatory variables as named in Section 5.2. Our approach to this investigation was, similarly to that for RelBAV as outcomes, to fit multivariate linear models to the data structured as in Equation (2.1.11). The final, fitted multivariate model included 2 explanatory variables (i.e. $p - 1 = 2$; gender and *musician*), 2 outcomes (i.e. $d = 2$; left and right RelBASA estimates) and $n = 39$. This model followed the structure in Equations (5.3.1)-(5.3.4) apart from the following differences:

- $k = 1,2$ corresponded to the left and right RelBASA estimates, respectively.
- In Equation (5.3.2), $\underline{X}$ was defined as for Equation (2.1.6) but with explanatory variables: gender (x_{i1}) and *musician*(x_{i2}).
- σ_1^2 and σ_2^2 represented the left and right RelBASA random error variances, respectively.
- $\sigma_{1,2}$ was equivalent to the covariance between the two outcomes (left and right RelBASA).

MLwiN version 2.16 was used to fit both multivariate liner models. R version 2.11.1 was the program of choice for constructing the plots for these models. Similarly to the analyses of *Chapter 4*, Restricted Iteratively Generalised Least Squares was used as the estimation method in fitting the models as described in Section 3.4.

5.4 Broca's volume estimates relative to total brain volume

In a similar approach to the investigation in *Chapter 4* (i.e. using a stepwise method of model selection), we fitted a multivariate linear model (MLM) to the subset of the Musician-Control dataset containing Broca's area measurement estimates, as discussed in Section 5.2. In this section, we focus on the associations between the left and right volume estimates of Broca's area relative to the total brain volume (RelBAV) and the explanatory variables described in Section 5.2. We used Wald tests to determine whether the multivariate parameters were non-zero. In this section the Wald test statistic followed a chi-square distribution with 2 degrees of freedom (the number of outcomes, $d = 2$). The results from the final, fitted MLM can be seen in Tables 5.1 and 5.2.

We can obtain information about the overall associations between explanatory variables and the outcomes combined from a MLM which is not available from multiple univariate linear models (ULMs). Furthermore, results from the models based on the marginal distributions for each outcome as well as the covariances and correlations between the outcomes were also obtained from the MLM. The overall multivariate coefficients, standard errors, confidence intervals and p-values for our fitted model can be seen in Table 5.1.

No effect of gender was found, across both hemispheres combined, at the 5% level of significance (p=0.06, Table 5.1). The p-value for gender was very close to α (i.e. the critical significance level, 0.05) which means that at that significance level we found no difference between males and females across both left and right RelBAV estimates combined, on average. However, we would have found a significant difference between males and females in RelBAV estimates if we had considered the 10% significance level as the level of α.

Fixed Effect	Coefficient	St. Error	95% CI	p-value
Intercept	2.4	0.09	(2.2, 2.5)	<0.001
Gender	-0.2	0.1	(-0.5, 0.005)	0.06

Table 5.1: *Estimates of the multivariate explanatory variable coefficients for the multivariate linear model with left and right* RelBAV *estimates combined (including the standard error of the corresponding estimators, the 95% confidence intervals for the coefficients and the p-values).*

The results from the models based on the marginal distribution of each outcome, which allowed us to scrutinize possible associations between each RelBAV estimate and explanatory variable individually, can be viewed in Table 5.2. Since we have two outcomes, we needed to refer again to the rule of multiple comparisons from *Chapter 2*, and also to obtain a Bonferroni-adjusted α-value. We used the same modification as in *Chapter 3* in Equation (3.4.4) and therefore, the new adjusted α-value, α^*, was set equal to 0.025.

Using the new Bonferroni-corrected critical significance value, we can see that the coefficient for gender was not significant in either hemisphere (p=0.3 and p=0.04 for the left and right hemispheres, respectively, Table 5.2). However, the gender coefficient in the right hemisphere would have been statistically significant if we had set the original significance level to 10%, even after Bonferroni correction (i.e. $\alpha^* = 0.1/2 = 0.05$). If we consider this gender difference between the RelBAV estimates in the right hemisphere to be present in the dataset, then the model suggests that females have a smaller Broca's volume relative to total brain volume and in particular, a reduction by 0.2% on average. Although this may appear small, the mean RelBAV for the right hemisphere was 1.2% (of the total brain volume) and therefore, the difference in Broca's volume estimates between males and females was around 17%.

Explanatory Variable	**Coefficient**	**St. Error**	**95% CI**	***p*-value**
		Left Hemisphere		
Intercept	1.1	0.06	(1.0, 1.2)	<0.001
Gender	-0.08	0.08	(-0.2, 0.08)	0.3
		Right Hemisphere		
Intercept	1.3	0.06	(1.2, 1.4)	<0.001
Gender	-0.2	0.08	(-0.3, -0.004)	0.04
Random Error	**Estimate**	**St. Error**	**95% CI**	***p*-value**
Variance (Left Hemisphere)	0.06	0.01	(0.04, 0.09)	<0.001
Variance (Right Hemisphere)	0.06	0.01	(0.03, 0.09)	<0.001
Covariance [Correlation] (Left/Right Hemispheres)	0.01 [0.2]	0.01 [0.2]	(-0.006, 0.03) [-0.1, 0.5]	0.2

Table 5.2: *Estimates of the explanatory coefficients and random error variance, covariance and correlation terms for the multivariate linear model with left and right relative Broca's volume estimates as outcomes (including the standard error of the corresponding estimators, the 95% confidence intervals for the coefficients/terms and the p-values).*

Insufficient evidence was found to suggest that the correlation between the left and right RelBAV estimates was non-zero at the Bonferroni-adjusted 5% level of significance (p=0.2, Table 5.2). The Pearson correlation coefficient between the left and right RelBAV estimates were obtained directly from the raw data (see the line of best fit in Figure 5.1). The Pearson correlation coefficient obtained from the raw data was equal to 0.25 (p=0.1). Once again, this indicated that there was insufficient evidence to suggest that the correlation between the left and right RelBAV estimates was non-zero.

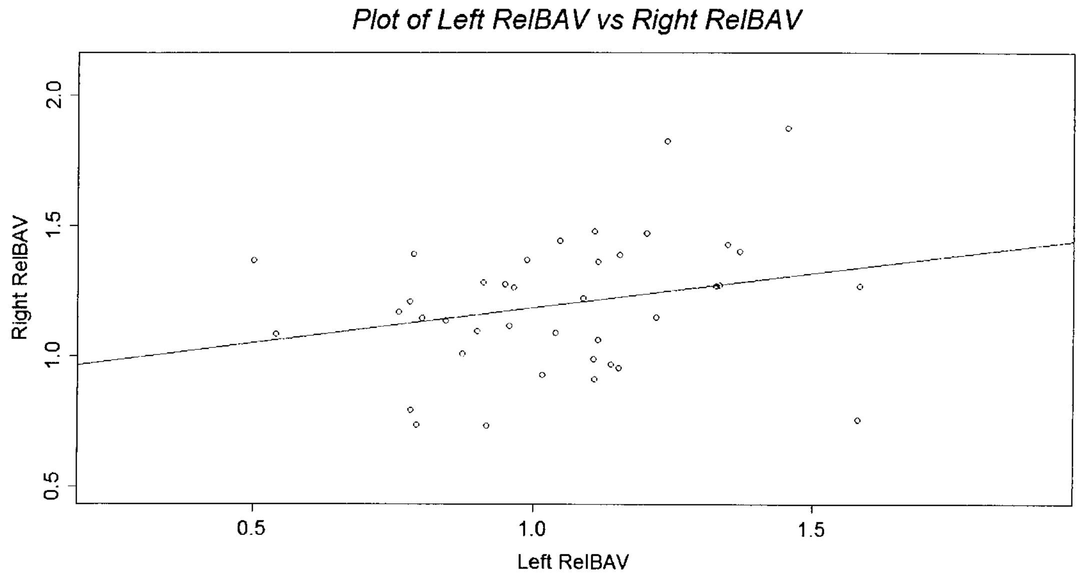


Figure 5.1: *Plot of left* RelBAV *against right* RelBAV *with a line of best fit.*

5.4.1 Comparison between multivariate and univariate linear models

Since there were no significantly non-zero multivariate coefficients for any of the explanatory variables in Table 5.1 and the correlation between the two outcomes was also non-significant in Table 5.2, this indicates that two univariate linear models (ULMs) with the same outcomes as the multivariate linear model (MLM) may be sufficient to explain the variability within left and right RelBAV estimates. Table 5.2 shows the results of the models based on the marginal distributions for each hemisphere independently and these results were almost identical to those obtained from the ULMs. However, during the stepwise model selection process for the ULMs, as described in Section 3.4, the coefficient for gender was not significant for the left hemisphere and was therefore, not included in the final model for left RelBAV.

We discussed in both Sections 3.4.2 and 4.4.1 the approximation of a MLM with independent outcomes to multiple ULMs. In our case we approximated the two ULMs with left and right RelBAV estimates as outcomes by a MLM with the covariance between the outcomes set equal to zero (i.e. $\sigma_{1,2} = 0$ from Equation (2.1.11)) and with the coefficient for gender set equal to zero in the left hemisphere. This MLM with independent outcomes (independent MLM) was nested within the MLM we fitted earlier in Section 5.4 (see Table 5.1 and 5.2) which had correlated outcomes (correlated MLM), which meant that we were able to use a deviance test to compare the fit of the models.

We have already stated that the deviance of a model is a goodness-of-fit statistic in Section 3.4.2 and we defined it in Equation (3.4.5). For a full description of the deviance test see Section 3.4.2. The deviances for the independent MLM (model 1) and the correlated MLM (model 2) can be seen in Table 5.3, along with their respective degrees of freedom.

Using the deviances for models 1 and 2 from Table 5.3, the deviance test statistic, D^*, can be written as:

$$D^* = |D_1^* - D_2^*| = |1.786 - (-1.161)| = 2.947 \qquad (5.4.1)$$

with D_1^* equal to the deviance for the independent MLM and D_2^* representing the deviance of the correlated MLM. The D^* value of 2.947 with 2 degrees of freedom (i.e. $73 - 71$) corresponds to the χ^2 distribution with p=0.23. This is interpreted as there being no evidence to suggest that the MLM explains more variability in the data than the two ULMs, which fits well with our observations about the multivariate explanatory coefficients and correlation mentioned earlier.

#	Model	Deviance	df
1	Independent MLM	1.786	73
2	Correlated MLM	-1.161	71

Table 5.3: *Table showing the deviances and degrees of freedom of the independent MLM and correlated MLM with left and right* RelBAV *as outcomes.*

5.4.2 Model diagnostics

We discussed a variety of statistics to assess the goodness-of-fit of a MLM in Section 4.4.2. The R^2 statistic was deemed to be the only one of the suggested statistics which could give an indication of how well a model fits the data as opposed to simply being able to compare the fit of two models. However, R^2 is unobtainable for a multivariate model. In Section 5.4.1 we showed that there was no difference between two ULMs and our fitted MLM. Therefore, we can assess the fit of the MLM by obtaining values of R^2 for the ULMs. The R^2 values for the ULMs was found to be:

- $R^2 = 0.025$ (ULM with left RelBAV as the outcome).
- $R^2 = 0.099$ (ULM with right RelBAV as the outcome).

The ULMs explain just 2.5% and 9.9% of the total variability in the left and right RelBAV estimates, respectively, according to the R^2 statistics. This means that these two ULMs are inadequate and poorly fitting to the data, and consequently the MLM may also be a poor fit to the data, but all we can state is that the R^2 statistic for the MLM would be greater than or equal to the R^2 values given above for the ULMs.

The standardized residuals can also be inspected to assess the goodness-of-fit of a model. Firstly, we plotted the standardized residuals for each of the marginal distributions from the MLM against each other (see Figure 5.2). This plot shows the correlation between the outcomes as shown by the model. Our MLM in Table 5.2 gives us a non-statistically significant estimate of the correlation between the outcomes and this is reflected in Figure 5.2 with little correlation indicated.

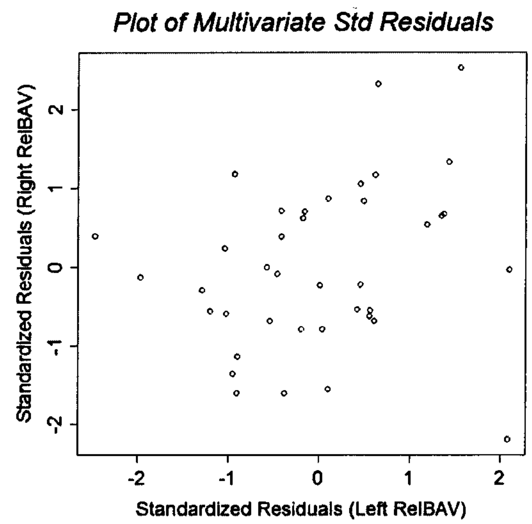


Figure 5.2: *Marginal distribution standardized residual plot for the left and right RelBAV estimates from the multivariate linear model.*

The model diagnostics for the models based on the marginal distributions of the MLM are shown in the plots of Figure 5.3. The overall conclusion from the model diagnostics was that there was not enough evidence to suggest that the model assumptions stated in Section 2.1.1 were not met. The standardized residuals for the left RelBAV estimates approximately follow a normal distribution (see Figures 5.3(c) and (e)). Although there could a hint of heteroscedascity in both Figures 5.3(a) and (b) for the left and right RelBAV estimates, respectively, due to the small sample size there is not enough evidence to clear state that the two sets of standardized residuals did not meet the homoscedascity assumption. The normal-QQ plot in Figure 5.3(f) shows that the standardized residuals for right RelBAV look approximately normally distributed, but the histogram in Figure 5.3(d) looks bimodal (i.e. not normally distributed). The indication of the standardized residuals approximately following a normal distribution is also highlighted by the Anderson-Darling normality test (for both outcomes, p=0.7) suggesting that there was insufficient evidence to reject the assumption of normality.

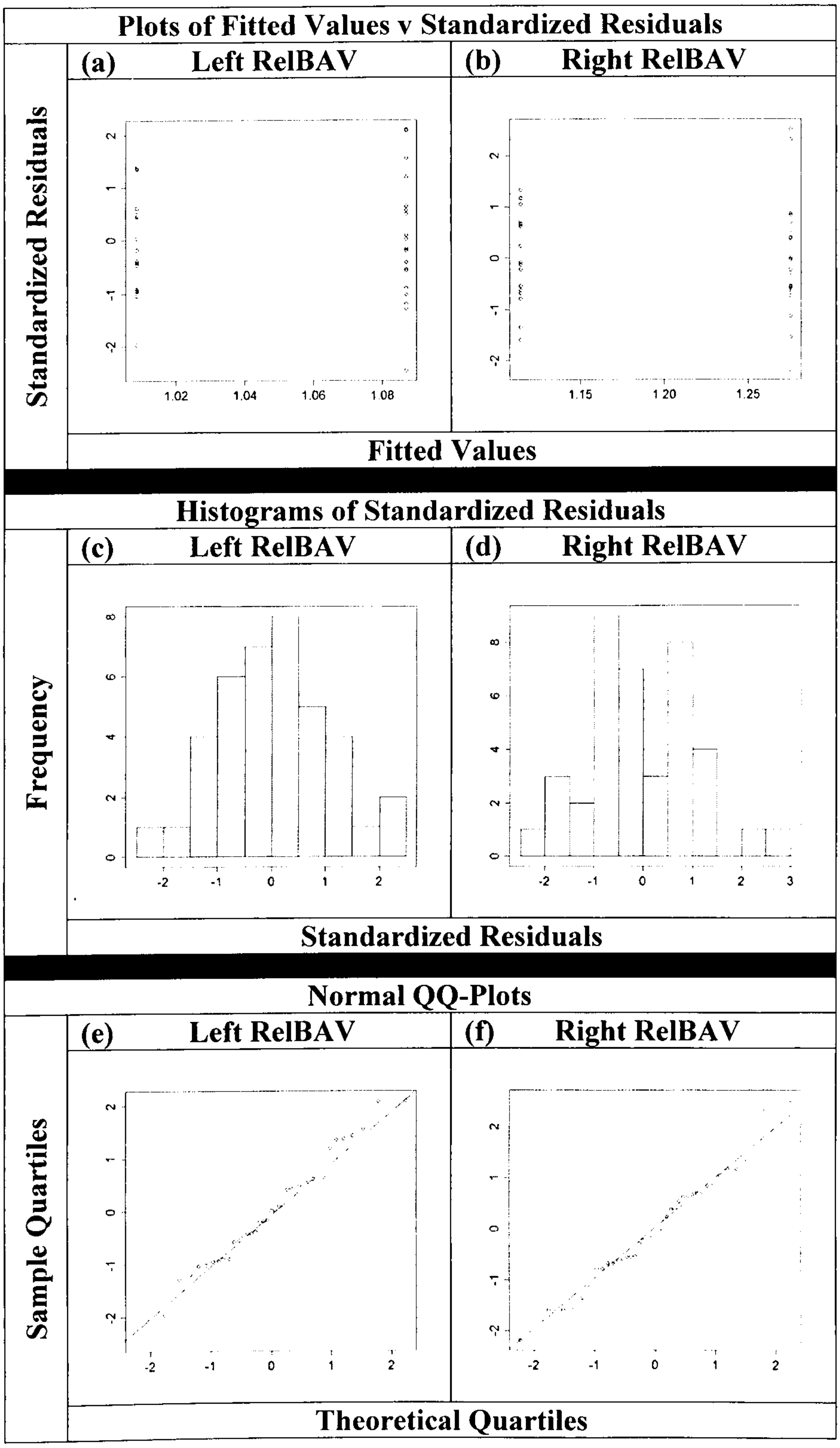


Figure 5.3: *Diagnostic plots for the multivariate linear model with (a, c and e) left and (b, d and f) right* RelBAV *as the outcomes.*

5.5 Broca's surface area estimates relative to Broca's volume

Following on from the investigation into the associations between the explanatory variables and Broca's volume, one question which remains to be answered is whether associations exist between the explanatory variables and Broca's surface area. Our outcome variables for this section were the Broca's surface area estimates relative to Broca's volume (RelBASA or relative Broca's surface area) for the left and right hemispheres (see Section 5.2 for details). The methodology used was identical to that for the volume estimates in Section 5.4. We used a stepwise model selection process (see Section 3.4) and Wald tests (as defined in Section 5.4) to check whether the multivariate parameters were non-zero. The fitted model can be viewed in Tables 5.4 and 5.5.

The multivariate linear model allows the overall association between explanatory variables and RelBASA across both hemispheres to be described. The results show that there was a difference in RelBASA estimates between musicians and non-musicians when both hemispheres are considered simultaneously (p=0.02, Table 5.4). Since the explanatory coefficient for *musician* in Table 5.4 was negative, then this suggests that musicians have a smaller relative Broca's surface area (RelBASA) than non-musicians, across both hemispheres combined. Not enough evidence was found to reject the null hypothesis that there was no difference in the RelBASA of both hemispheres combined between males and females (p=0.06, Table 5.4).

Fixed Effect	Coefficient	St. Error	95% CI	*p*-value
Intercept	8.5	0.3	(7.9, 9.1)	<0.001
Gender	-0.7	0.4	(-1.4, 0.03)	0.06
Musician	-0.9	0.4	(-1.6, -0.2)	0.02

Table 5.4: *Estimates of the multivariate explanatory variable coefficients for the multivariate linear model with left and right* RelBASA *estimates combined (including the standard error of the corresponding estimators, the 95% confidence intervals for the coefficients and the p-values).*

From the models based on the marginal distributions of the RelBASA estimates for each hemisphere, we obtained results for the associations between each outcome and the explanatory variables (see results in Table 5.5). We have two outcomes in our model, hence, we need to adjust the α-value in the same way as in Section 5.4. Using the Bonferroni correction, we obtained a new adjusted α-value, $\alpha^* = 0.5/2 = 0.025$ (see Equation (3.4.4)). The results show that there was a difference in the right RelBASA between musicians and controls, but not in the left hemisphere (p=0.02 and p=0.04, respectively, Table 5.5). The coefficient for *musician* in the right hemisphere was negative, suggesting that musicians have a smaller relative Broca's surface area (RelBASA) than non-musicians in the right hemisphere. The coefficient for gender was found to be significant, but only at the 10% Bonferroni-corrected α-level in the left hemisphere (p=0.03, Table 5.5). There was not enough evidence at the 5% level of significance to reject the null hypothesis that the coefficient for gender in the RelBASA estimates of the left hemisphere was equal to zero. We found no gender difference in the RelBASA estimates of the right hemisphere (p=0.2, Table 5.5).

Explanatory Variable	**Coefficient**	**St. Error**	**95% CI**	***p*-value**
Left Hemisphere				
Intercept	4.2	0.2	(3.9, 4.5)	<0.001
Gender	-0.4	0.2	(-0.8, -0.04)	0.03
Musician	-0.4	0.2	(-0.7, -0.03)	0.04
Right Hemisphere				
Intercept	4.4	0.2	(4.0, 4.7)	<0.001
Gender	-0.3	0.2	(-0.7, 0.1)	0.2
Musician	-0.5	0.2	(-0.9, -0.07)	0.02
Random Error	**Estimate**	**St. Error**	**95% CI**	***p*-value**
Variance (Left Hemisphere)	0.3	0.07	(0.2, 0.5)	<0.001
Variance (Right Hemisphere)	0.4	0.1	(0.2, 0.6)	<0.001
Covariance [Correlation] (Left/Right Hemispheres)	0.3 [0.7]	0.07 [0.2]	(0.1, 0.4) [0.3, 1.0]	<0.001

Table 5.5: *Estimates of the explanatory coefficients and random error variance, covariance and correlation terms for the multivariate linear model with left and right relative Broca's surface area estimates as outcomes (including the standard error of the corresponding estimators, the 95% confidence intervals for the coefficients/terms and the p-values).*

We saw the correlation (r) between the left and right RelBASA estimates was estimated by the model to be equal to 0.7 in Table 5.5 ($p<0.001$). This means that there is strong evidence to reject the hypothesis that the correlation between the two outcomes equals zero. A value of $r = 0.7$ is fairly strong, positive correlation. As a comparison, we plotted the left and right RelBASA estimates against each other with a line of best fit (see Figure 5.4). The Pearson correlation coefficient, ρ, can be examined directly from the raw data and can be seen by the gradient of the line of best fit in Figure 5.4. This plot shows that there is strong, positive correlation between the left and right relative Broca's surface area estimates. Indeed, ρ was estimated directly from the raw data and was also equal to 0.7 ($p<0.001$).

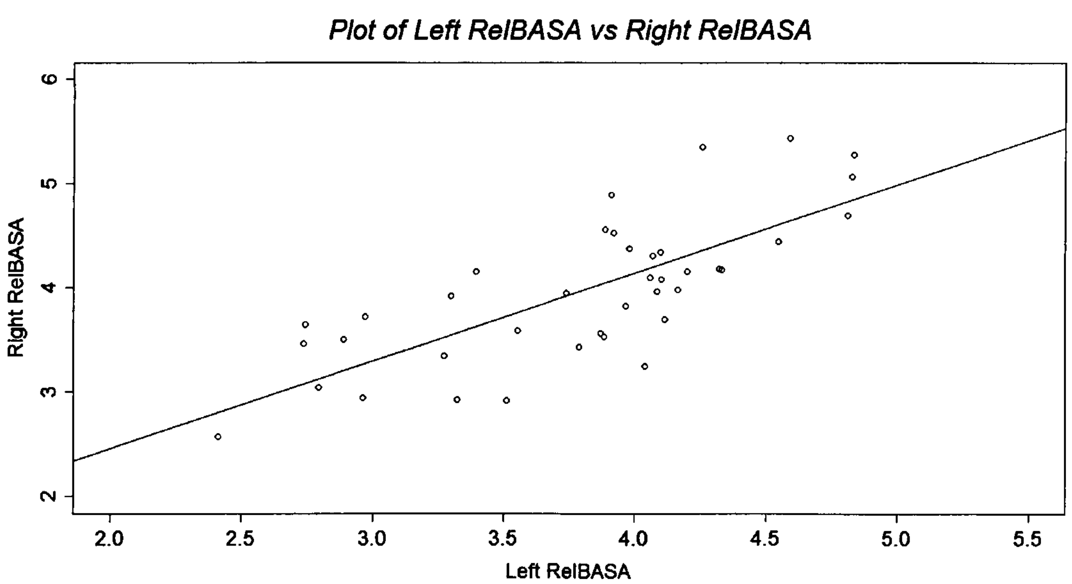


Figure 5.4: *Plot of left against right, relative Broca's surface area with a line of best fit.*

5.5.1 Comparison between multivariate and univariate linear models

At this point in the investigation, the question to be answered was whether or not the MLM is a better fit to the data than two ULMs. From the results in the previous section, we saw that *musician* was statistically significant for the multivariate outcome (i.e. across both left and right RelBASA estimates combined; see Table 5.4) and we had a non-zero correlation between the outcomes in Table 5.5. These two observations suggest that a MLM may be more suitable than two ULMs since these observations would not be able to be made from univariate models.

We discussed reasons for differences between ULMs and the models based on the marginal distributions from the MLM in Section 4.4.1. Each ULM was fitted independently using a stepwise model selection process (see Section 3.4). Explanatory variables which were included in the MLM were not necessarily included in the ULM for each outcome individually. This can be explained by the fact that each explanatory variable in the MLM does not need to have a significantly non-zero coefficient for each outcome. For example, it is possible for an explanatory variable to be strongly associated with both the multivariate outcome, and one individual outcome, but not associated with any other outcomes. In cases where these variable coefficients were not significant for the models based on the marginal distributions of the MLM they were not included in the ULMs (e.g. the coefficient of gender had a p-value of 0.2 in the model based on the marginal distribution for right RelBASA (see Table 5.5) and hence, this explanatory variable was not included in the respective ULM).

What we expected to see was a univariate model for the right RelBASA estimates including the intercept and *musician* as the only significant explanatory variables. However, in our case, the univariate model obtained has the intercept and vocabulary score as the statistically significant explanatory variables. The results for this univariate model can be seen in Table 5.6. The negative coefficient of vocabulary score meant that those participants that achieved higher vocabulary marks had smaller estimated right relative Broca's surface area than individuals with lower vocabulary scores (p=0.02, Table 5.6).

Explanatory Variable	Coefficient	St. error	95% CI	*p*-value
Intercept	5.0	0.5	(4.1, 6.0)	<0.001
Vocabulary Test Score	-0.08	0.04	(-0.2, -0.01)	0.02

Table 5.6: *Estimates of the explanatory coefficients for the univariate linear model with right* RelBASA *estimates as the outcome (including the standard error of the corresponding estimators, the 95% confidence intervals for the coefficients and the p-values).*

As described in Sections 3.4.2, 4.4.1 and 5.4.1, a way of testing whether or not the ULMs significantly differ from the MLM is to approximate the ULMs to a MLM with independent outcomes and then use a deviance test. However, no direct comparison can be made here because the multivariate model does not include the right RelBASA univariate model from Table 5.6 as a sub-model. One of the requirements of the deviance test is that the models are nested (McCullagh & Nelder, 1989). Consequently, a deviance test cannot be performed. However, it was noticed that during the fitting of the ULM, both vocabulary score and *musician* had significantly non-zero coefficients, but vocabulary score had the smaller *p*-value (the difference in *p*-values between vocabulary score and *musician* was equal to 0.0007).

Following the inclusion of vocabulary score into the model, the factor *musician* then became non-significant. However, in the multivariate model, the effect of *musician* across both hemispheres was more significant than vocabulary score was. This can be explained by *musician* also being significant in the left hemisphere whereas vocabulary score was not. Thus, once *musician* was included in the ULM with right RelBASA as the outcome, vocabulary score became non-significant. This suggests that the variables *musician* and vocabulary score appear to be explaining some, if not all, of the same variability within the data.

Figure 5.5 shows the vocabulary scores plotted against the right RelBASA estimates with musician and control groups identified independently. We observed that musicians obtained, on average, higher vocabulary scores than non-musicians, and also that participants with higher vocabulary scores were associated with smaller right RelBASA than those participants with lower vocabulary scores. The combination of these results suggest that the variability due to musicians and vocabulary test scores in the right RelBASA estimates are confounded and are therefore, explaining similar variability in the dataset. The almost parallel linear regression lines for the musician and non-musician groups in Figure 5.5 shows that there is little evidence of a significant interaction between *musician* and vocabulary score in relation to right RelBASA.

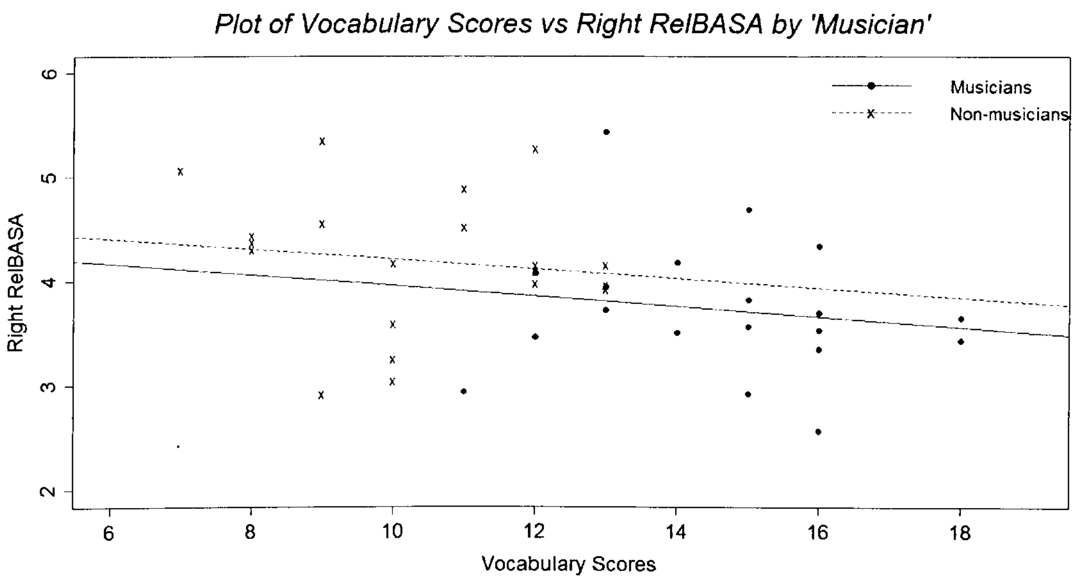


Figure 5.5: *Plot of WAIS vocabulary scores against right hemisphere Broca's surface area relative to Broca's volume by musician with a line of best fit for musicians and non-musicians.*

In Sections 3.4.2 and 4.4.1 we discussed the approximation of a MLM with independent outcomes (independent MLM) to a number of ULMs. In the case of the ULMs in this section, we used the approximation to them of a MLM with correlation between left and right RelBASA fixed as equal to zero (i.e. $\sigma_{1,2} = 0$ from Equation (2.1.11)) and with the coefficient of gender in the right hemisphere also set equal to zero. We have explained that there is relatively little difference between the ULM with vocabulary score as an explanatory variable in Table 5.6 and the ULM with *musician* as an explanatory variable, therefore, we included *musician* as an explanatory variable for the right hemisphere in the independent MLM, and not vocabulary score. This made it possible for us to perform a deviance test, as described in Section 3.4.2 where the deviance of a model is defined as in Equation (3.4.5) to check whether the MLM with correlated outcomes from Tables 5.4 and 5.5 (correlated MLM) was a better fit to the data than the independent MLM (i.e. the ULMs). The deviances for these models, along with their respective degrees of freedom can be seen in Table 5.7.

#	Model	Deviance	df
1	Independent MLM	139.960	71
2	Correlated MLM	110.952	69

Table 5.7: *Table showing the deviances and degrees of freedom of the independent MLM and correlated MLM with left and right relative Broca's surface area as outcomes.*

Using the results in Table 5.7, the deviance test statistic, D^*, can therefore, be calculated as:

$$D^* = |D_1^* - D_2^*| = |139.960 - 110.952| = 29.008 \qquad (5.5.1)$$

This test statistic (29.008) which has 2 degrees of freedom (the difference between the 71 and 69 degrees of freedom of the two models), corresponds to a chi-squared distribution with $p<0.001$. Therefore, there is evidence to suggest that the correlated MLM is different to the independent MLM. Since the correlated MLM has the smaller deviance then we can say that the correlated MLM is a better fitting model than the independent MLM and thus also better than two ULMs.

5.5.2 Model diagnostics

For the reasons given in Section 4.4.2 we were able to assess the goodness-of-fit of the MLM by scrutinizing the R^2 statistic for the two fitted ULMs with left and right RelBASA as the respective outcomes. The value of R^2 for each of the ULMs was:

- $R^2 = 0.198$ (ULM with left RelBASA as the outcome).
- $R^2 = 0.162$ (ULM with right RelBASA as the outcome).

These R^2 are interpreted as proportions of the total variability which are explained by the model. Therefore, we can see that 19.8% and 16.2% of the total variability in the left and right RelBASA estimates are explained by the respective ULMs. In Section 5.5.1 we showed that the MLM was a better fit to the data than the two ULMs. Therefore, we know that the total variability in the outcomes explained by the MLM will be greater than that obtained from the ULMs (i.e. greater than 19.8% and 16.2% of the variability of left and right RelBASA, respectively, was explained by the MLM).

We can further assess the goodness-of-fit of the MLM by examining the standardized residuals to see if the model assumptions were met (see Section 2.1.1 for assumptions). Standardized residuals and fitted values for each of the models based on the marginal distributions from the MLM were obtained. As an indicator of the correlation between the two outcomes, we plotted the standardized residuals for each outcome against each other (see Figure 5.6). it can clearly be seen in Figure 5.6 that the standardized residuals of each model based on the marginal distributions from the MLM are positively correlated. The model estimates this correlation to be equal to 0.7 (see Table 5.5).

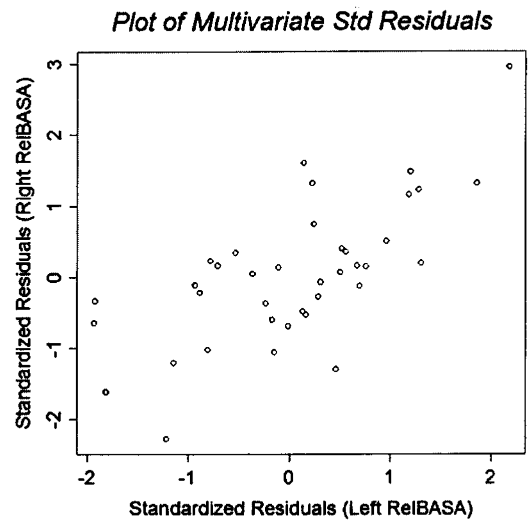


Figure 5.6: *Marginal distribution standardized residual plot for the left and right RelBASA estimates from the multivariate linear model.*

A range of diagnostic plots were then constructed for the two models based on the marginal distributions of the MLM (see Figure 5.7). The standardized residuals, for both left and right RelBASA, approximately follow a normal distribution according the histograms (see Figures 5.7(c) and (d) and the normal-QQ plots (see Figures 5.7(e) and (f)). The Anderson-Darling test for normality agrees with this observation finding that there is no evidence to reject the null hypothesis that the standardized residuals are normally distributed (p=0.7 and p=0.3 for left and right RelBASA, respectively). Observing the standardized residuals in Figures 5.7(a) and (b), we can see that they look to be slightly heteroscedastic. However, due to the small sample size, this is not sufficient to state with conviction that the model does not meet the assumption of homoscedasticity. Overall, the diagnostic plots suggest that the model is a reasonably good fit to the data, and although the R^2 statistics are poor, we can only tell that the variability explained by the MLM is greater than those figures showing the fit of the ULMs.

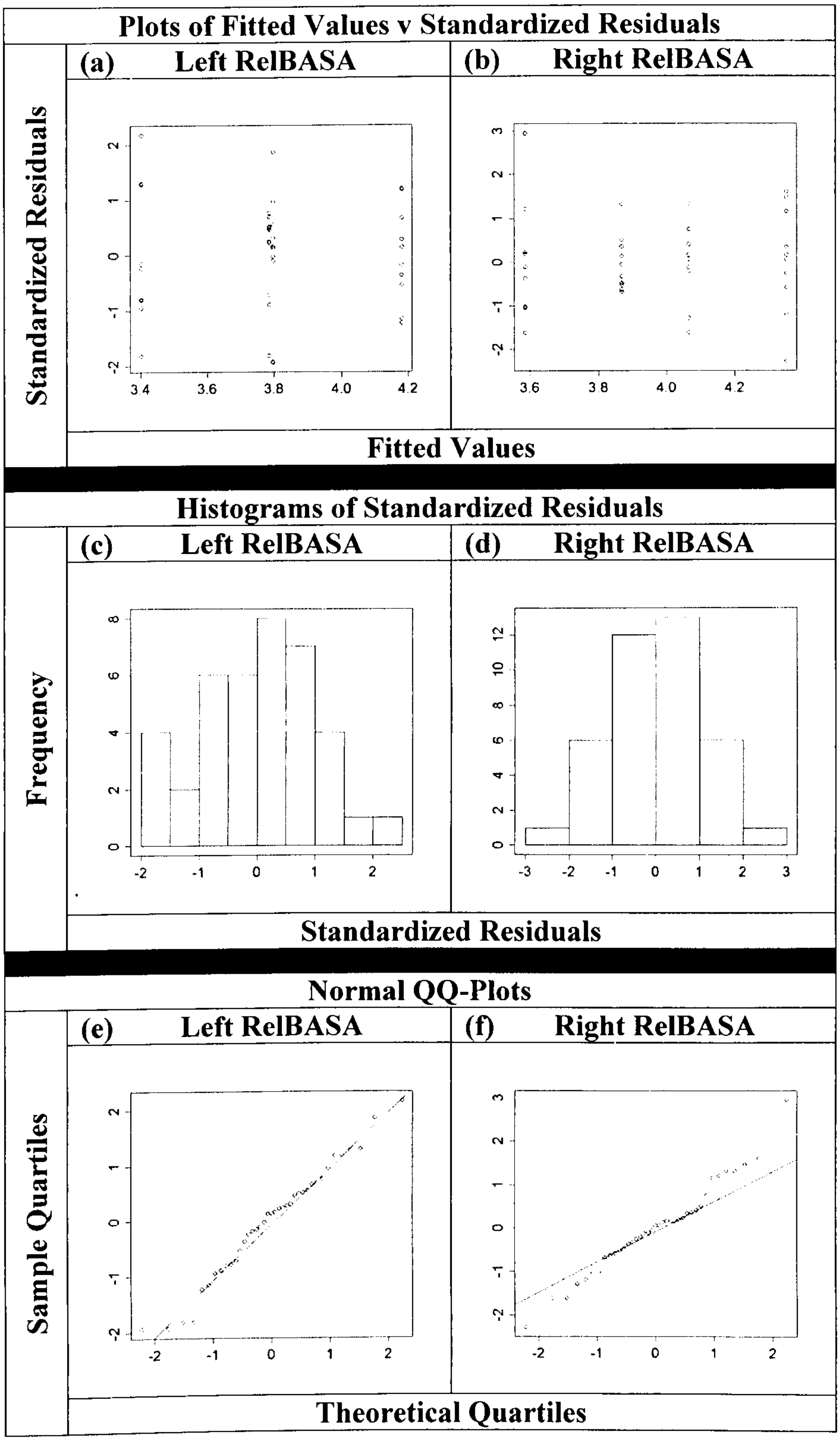


Figure 5.7: *Diagnostic plots for the multivariate linear model with (a, c and e) left and (b, d and f) right relative Broca's surface area as the outcomes.*

5.6 Within observer variability

The process by which Broca's area can be identified was explained in Section 2.3.2, and the stereological methodology used to obtain estimates for the volume and surface area of Broca's area was detailed in Sections 2.2.3 and 2.2.6, respectively. The precision of these estimators, including a worked example for each case, was also investigated in *Chapter 2*. Estimates of the variance of the volume estimator applying Cavalieri sectioning and point counting were obtained for one of the sub-structures of Broca's area in one hemisphere along with estimates for the corresponding coefficients of error. Similarly for the surface area estimator, estimates of the variance and coefficient of error due to the three levels of sampling (systematic sampling of orientations, Cavalieri sectioning and cycloid grid positioning) were obtained for a right hemisphere Pars Opercularis. Other possible sources of variability among the volume and surface area estimates could include variability due to:

- Observer variability
    - within observer variability
    - between observer variability
- Biological variation
- Other causes (measurement errors, magnetic resonance imaging (MRI) limitations such as limited resolution, intensity inhomogeneities, partial voluming, etc)

Some of these sources of variation could also lead to bias within the estimates. Examples of this could be if the region of interest (RoI) is continually wrongly demarcated to be larger than the actual RoI, or the observer continually counts too many or too few points within the RoI. It would be possible to estimate any potential bias by comparing estimates between observers. However, some of the other causes such as measurement error and the MRI limitations could also lead to bias which would be confounded in the actual magnetic resonance (MR) images and therefore, extremely difficult to estimate. Multiple MR scanners and additional equipment would be required to obtain estimates for some of these additional errors.

In this subsection, the variability that will be extracted from the data will be the within observer variability. Unfortunately, it was not possible to extract estimates for the between observer and demarcation variation in this thesis. However, these are areas for further work in an extension to the work reported in this thesis. The other sources of variability, biological variation, measurement error and MRI limitations are very difficult to estimate individually and for the purposes of this study they will be grouped together as unexplained variability. To estimate the within observer variability, $var_{w.obs}$, two intra-rater studies were constructed.

The first intra-rater study involved the estimation of the volume of Pars Opercularis (PO) and Pars Triangularis (PT) for both hemispheres of two participants by observer CC (Chris Cheyne), repeated ten times. For each of the ten repeated estimations, the same demarcations of the RoI and Cavalieri sectioning were used. Only the positioning of the grid of points differed between each of the ten repeated estimations. In each case, the grid of points was randomly placed on each Cavalieri section. The result of this study meant that the variability between the volume estimates, consisting of the within observer variability and the variability due to point counting, was obtained.

For the second intra-rater study, PO and PT surface area of both the left and right hemispheres were estimated for the same two participants investigated in the first intra-rater study. This estimation was again obtained by observer CC, and repeated ten times, with the same RoI demarcation, orientations and Cavalieri sectioning on each orientation. In this study, only the positioning of the cycloid grid on each Cavalieri section was different (randomly placed) for each of the ten repeated observations. From this study we obtained the variability between the surface area estimates which consisted of the within observer variability and the variability due to cycloid grid sampling. The results for each of these two intra-rater studies can be seen below.

5.6.1 Intra-rater study 1 - Volume

We obtained ten volume estimates (in cm^3) of the two sub-structures of Broca's area PO and PT in the left and right hemispheres of two participants which can be viewed in Table 5.8.

		Participant							
		1				2			
		LH PO	LH PT	RH PO	RH PT	LH PO	LH PT	RH PO	RH PT
Observation	1	6.54	7.04	5.92	7.54	3.23	4.18	6.23	2.63
	2	6.43	6.52	5.95	7.36	3.16	4.37	6.09	2.68
	3	6.63	6.71	5.90	7.87	3.05	4.45	6.08	2.75
	4	6.45	7.20	5.86	8.03	3.00	4.37	6.28	2.66
	5	6.62	7.07	5.99	8.01	2.84	4.23	5.73	2.92
	6	6.54	7.07	5.94	7.34	3.00	4.30	5.74	3.04
	7	6.79	6.71	5.78	7.76	3.05	4.36	5.54	2.65
	8	6.71	6.82	5.99	7.16	2.94	4.41	5.65	2.84
	9	6.36	6.89	5.93	7.51	3.04	4.57	5.93	2.50
	10	6.34	6.80	5.94	7.63	3.11	4.75	5.74	2.90

Table 5.8: *Ten estimates from two participants of the volume (in cm^3) of Pars Opercularis and Pars Triangularis in both the left and right hemispheres.*

As mentioned earlier, the variation of the volume estimates obtained from this study consisted of two parts:

- Variation due to point counting
- Within observer variation

Therefore, the total coefficient of error of the corresponding volume estimator can be approximated as the square root of the sum of the square of the coefficient of error due to point counting and the square of the within observer coefficient of error. Denoting the estimator of the coefficient of error in lower case, ce, this reads:

$$\mathrm{ce}(\tilde{V}) = \sqrt{\mathrm{ce}_{\mathrm{PC}}^2(\tilde{V}) + \mathrm{ce}_{\mathrm{w.obs}}^2(\tilde{V})} \quad (5.6.1)$$

where $\mathrm{ce}_{\mathrm{PC}}(\tilde{V})$ represents the estimator of the coefficient of error due to point counting, and $\mathrm{ce}_{\mathrm{w.obs}}(\tilde{V})$ is equivalent to the within observer coefficient of error estimator. The coefficient of error due to the within observer variability can then be calculated by rearranging Equation (5.6.1) as long as the other two coefficients of error are known. The unit size of the grid of points, u, used in this study is the same as the one used in the worked example of Section 2.2.5 (i.e. u=3mm). Moreover, the shape coefficients of PO and PT which, as discussed in Section 2.2.4, are required for the estimation of the coefficient of error of the volume estimator due to point counting and allow the variability of the shape of the regions of interest to be accounted for (i.e. taking into account the irregularity of the RoI on each Cavalieri section), were expected to be similar across individuals. Consequently, it was assumed that the coefficient of error due to point counting will be similar to that in the worked example ($ce_{PC}(\tilde{V}) = 0.0155$; see Equation (2.2.42)) and therefore, this estimate was used in our calculation of the within observer coefficient of error. To find the total coefficient of error, two components had to be derived from the observations from the two brains. The first component was the mean of the mean volumes across each brain:

$$\mathrm{E}[(\bar{V}_1, \bar{V}_2)] = \frac{1}{2}(\bar{V}_1 + \bar{V}_2) \quad (5.6.2)$$

where $\bar{V}_1 = \sum_{i=1}^{10} \tilde{V}_{1i}/10$ and $\bar{V}_2 = \sum_{i=1}^{10} \tilde{V}_{2i}/10$ are the means for the two individual brains for a particular sub-structure with $\tilde{V}_{1i}$ representing the volume estimate of the i^{th} observation of the 1st brain $(1 \leq i \leq 10)$ and $\tilde{V}_{2i}$ representing the volume estimate of the i^{th} observation of the 2nd brain. We calculated the mean of mean volumes for PO and PT of each hemisphere by using Equation (5.6.2). The second component is the mean variance of the volume estimator across the two brains:

$$\mathrm{var}(\tilde{V}) = \frac{1}{2}\left(\mathrm{var}(\tilde{V}_1) + \mathrm{var}(\tilde{V}_2)\right) \quad (5.6.3)$$

where $\mathrm{var}(\tilde{V}_1)$ is the variance of the volume estimator for the 1st brain, and $\mathrm{var}(\tilde{V}_2)$ is the variance of the volume estimator for the 2nd brain. By applying Equations (5.6.2) and (5.6.3) to Equation (2.2.9) the total coefficient of error of the volume estimator can be written as:

$$\mathrm{ce}(\tilde{V}) = \sqrt{\mathrm{var}(\tilde{V})}/\mathrm{E}[(\tilde{V}_1, \tilde{V}_2)] \quad (5.6.4)$$

For each of the four sub-structures of Broca's area, the total mean, variance and coefficient of error of the volume estimator can be seen in Table 5.9.

Broca's Area Sub-structure	$\mathrm{E}[(\hat{V}_1, \hat{V}_2)]$	$\mathrm{var}(\hat{V})$	$\mathrm{ce}(\hat{V})$
LH PO	4.79	0.017	0.027
LH PT	5.64	0.036	0.034
RH PO	5.91	0.035	0.032
RH PT	5.19	0.056	0.046

Table 5.9: *Mean (in cm³), variance (in cm⁶) and coefficient of error for the volume estimator of each sub-structure of Broca's area in the left and right hemispheres. The within-observer and point counting components are accounted for by both the variance and coefficient of error. LH: Left hemisphere; RH: Right hemisphere; PO: Pars Opercularis; PT: Pars Triangularis.*

By rearranging Equation (5.6.1), the within observer coefficient of error for each sub-structure of Broca's area in each hemisphere can be calculated by using:

$$\mathrm{ce}_{\mathrm{w.obs}}(\tilde{V}) = \sqrt{\mathrm{ce}^2(\tilde{V}) - \mathrm{ce}^2_{\mathrm{PC}}(\tilde{V})} \quad (5.6.5)$$

Given the results in Table 5.9 and the estimate of $ce_{PC}(\tilde{V}) = 0.0155$, from Section 2.2.5, then the coefficient of error due to the within observer variability for each sub-structure of Broca's area were:

LHPO: $$ce_{w.obs}(\tilde{V}) = \sqrt{0.027^2 - 0.0155^2} = 0.022 \quad (5.6.6)$$

LHPT: $$ce_{w.obs}(\tilde{V}) = \sqrt{0.034^2 - 0.0155^2} = 0.030 \quad (5.6.7)$$

RHPO: $$ce_{w.obs}(\tilde{V}) = \sqrt{0.032^2 - 0.0155^2} = 0.028 \quad (5.6.8)$$

RHPT: $$ce_{w.obs}(\tilde{V}) = \sqrt{0.046^2 - 0.0155^2} = 0.043 \quad (5.6.9)$$

The within observer coefficient of error suggests that the contribution to the within observer variability is greater in Pars Triangularis than in Pars Opercularis in both hemispheres (3.0% and 4.3% for left and right PT, respectively, compared to 2.2% and 2.8% for left and right PO, respectively). Equations (5.6.6)-(5.6.9) also suggest that, given the set parameters of u and T from Equation (2.2.12) (i.e. the distance between points on the point grid and the distance between Cavalieri sections), the CE that is due to within observer variability is greater than the CE due to Cavalieri sectioning and point counting combined (2.1% in Equation (2.2.43)), irrespective of structure and hemisphere.

5.6.2 Intra-rater study 2 – Surface area

Similarly to Section 5.6.1, we obtained ten estimates, this time of the surface area (in cm^2) of the two sub-structures of Broca's area PO and PT in the left and right hemispheres of two participants. Table 5.10 show the surface area estimates for these two participants.

		Participant							
		1				2			
		LH PO	LH PT	RH PO	RH PT	LH PO	LH PT	RH PO	RH PT
Observation	1	3.50	3.56	3.11	3.95	1.85	2.15	3.11	1.71
	2	3.35	4.10	3.12	3.87	1.92	2.64	3.14	1.94
	3	3.53	3.99	3.56	4.35	1.85	2.58	3.30	2.07
	4	3.56	3.95	3.32	3.83	2.00	2.30	3.23	1.92
	5	3.32	3.68	3.06	3.87	2.16	2.49	3.00	1.79
	6	3.41	3.26	3.14	3.80	1.95	2.40	3.14	2.13
	7	3.60	3.57	3.41	3.56	2.12	2.66	3.23	2.10
	8	3.63	3.78	3.45	3.33	2.18	2.72	3.30	1.86
	9	3.20	3.77	3.44	3.74	2.22	2.66	3.33	2.03
	10	3.48	4.25	3.32	3.77	2.09	2.69	3.17	2.03

Table 5.10: *Ten estimates from two participants of the surface area (in cm^2) of Pars Opercularis and Pars Triangularis in both the left and right hemispheres.*

The variation of the surface area estimates obtained from this study comprised of two components:

- Variation due to cycloid grid positioning
- Within observer variation

Therefore, the estimator of the total coefficient of error, $\mathrm{ce}(\tilde{S}_w)$, can also be decomposed into two parts:

$$\mathrm{ce}(\tilde{S}_w) = \sqrt{\mathrm{ce}^2_{\mathrm{cyc}}(\tilde{S}_w) + \mathrm{ce}^2_{\mathrm{w.obs}}(\tilde{S}_w)} \qquad (5.6.10)$$

where $\mathrm{ce}_{\mathrm{cyc}}(\tilde{S}_w)$ is the estimator of the coefficient of error of the surface area estimator due to cycloid grid positioning, and the coefficient of error due to the within observer variability is denoted by $\mathrm{ce_{w.obs}}(\tilde{S}_w)$. Furthermore, the calculations necessary to obtain the estimate of the total coefficient of error for the surface area estimator are similar to those for volume estimates in Equations (5.6.2) and (5.6.3). Namely, the mean and variance:

$$\mathrm{E}[(\bar{S}_{w1}, \bar{S}_{w2})] = \frac{1}{2}(\bar{S}_{w1} + \bar{S}_{w2}) \tag{5.6.11}$$

$$\mathrm{var}(\tilde{S}_w) = \frac{1}{2}(\mathrm{var}(\bar{S}_{w1}) + \mathrm{var}(\bar{S}_{w2})) \tag{5.6.12}$$

where $\bar{S}_{w1} = \sum_{i=1}^{10} \tilde{S}_{w1i}/10$ and $\bar{S}_{w2} = \sum_{i=1}^{10} \tilde{S}_{w2i}/10$ are the means for a Broca's area sub-structure surface area in brain 1 and 2, respectively, with $\tilde{S}_{w1i}$ representing the surface area estimate of the i^{th} observation of the 1st brain $(1 \leq i \leq 10)$ and $\tilde{S}_{w2i}$ representing the surface area estimate of the i^{th} observation of the 2nd brain, $\mathrm{var}(\bar{S}_{w1})$ is the variance of the surface area estimator for the 1st brain and $\mathrm{var}(\bar{S}_{w2})$ is the variance of the surface area estimator for the 2nd brain. The application of Equations (5.6.11) and (5.6.12) to Equation (2.2.9) gives the total coefficient of error for the surface area estimator:

$$\mathrm{ce}(\tilde{S}_w) = \sqrt{\mathrm{var}(\tilde{S}_w)}/\mathrm{E}[(\bar{S}_{w1}, \bar{S}_{w2})] \tag{5.6.13}$$

Table 5.11 shows the mean, variance and total coefficient of error across both brains, for each sub-structure of Broca's area in each hemisphere. If $\mathrm{ce}(\tilde{S}_w)$ and $\mathrm{ce_{cyc}}(\tilde{S}_w)$ are known then the coefficient of error due to the within observer variability can be calculated by rearranging Equation (5.6.10).

Region	$E[(\tilde{S}_{w1}, \tilde{S}_{w2})]$	$\text{var}(\tilde{S}_w)$	$\text{ce}(\tilde{S}_w)$
LH PO	2.74	0.019	0.050
LH PT	3.16	0.061	0.078
RH PO	3.24	0.020	0.044
RH PT	2.88	0.044	0.073

Table 5.11: *Mean (in cm^2), variance (in cm^4) and coefficient of error for the surface area estimator of each sub-structure of Broca's area in the left and right hemispheres. The within-observer and cycloid grid positioning components are accounted for by both the variance and coefficient of error. LH: Left hemisphere; RH: Right hemisphere; PO: Pars Opercularis; PT: Pars Triangularis.*

The ratio of test area to the cycloid test line length (a/l), used in this intra-rater study was the same as that used in the worked example (i.e. a/l = 3mm). Also, since the total number of intersections across Cavalieri sections and orientations are expected to be similar for all sub-structures of Broca's area across the hemispheres of both individuals, the estimator $\hat{v}_{w_i}$ (an estimator of the variance due to the cycloid grid test system; see Equation (2.2.71)) can be assumed to be constant across all subjects. This means that it is assumed that the coefficient of error due to cycloid grid positioning will be similar to that in the worked example and therefore, the value of $\text{ce}_{\text{cyc}}(\tilde{S}_w) = 0.070$ was used (see Equation (2.2.81)). Rearranging Equation (5.6.10) means that the coefficient of error due to the observer can be calculated to be:

$$\text{ce}_{\text{w.obs}}(\tilde{S}_w) = \sqrt{\text{ce}^2(\tilde{S}_w) - \text{ce}^2_{\text{cyc}}(\tilde{S}_w)} \qquad (5.6.14)$$

Thus by using the values from Table 5.11 and the value of $\text{ce}_{\text{cyc}}(\tilde{S}_w)$ obtained in Equation (2.2.81), the estimates of the within observer coefficient of error for each of the left and right Broca's area sub-structures' surface area estimates were:

LHPO: $$\text{ce}_{\text{w.obs}}(\tilde{S}_w) = \sqrt{0.050^2 - 0.070^2} = NaN \qquad (5.6.15)$$

LHPT: $$\text{ce}_{\text{w.obs}}(\tilde{S}_w) = \sqrt{0.078^2 - 0.070^2} = 0.021 \qquad (5.6.16)$$

$$\text{RHPO:} \qquad \mathrm{ce_{w.obs}}(\tilde{S}_w) = \sqrt{0.044^2 - 0.070^2} = NaN \qquad (5.6.17)$$

$$\text{RHPT:} \qquad \mathrm{ce_{w.obs}}(\tilde{S}_w) = \sqrt{0.073^2 - 0.070^2} = 0.034 \qquad (5.6.18)$$

We can clearly see that the within observer coefficient of error was 2.1% for PT in the left hemisphere, and 3.4% for PT in the right hemisphere. These figures are similar to those we obtained for volume (3.0% and 4.3%, for left and right PT, respectively) suggesting that our estimates of within observer CE for surface area were reasonable. However, we could not obtain estimates for the within observer CE for right and left PO. This was because the estimate of CE due to cycloid sampling was greater than the estimate of total coefficient of error. It was mentioned earlier that the value of the CE due to cycloid sampling that we used came from the worked example in Section 2.2.8. To estimate the variance of the cycloids system on each Cavalieri section the Poisson model was used as described in Section 2.2.7. This method tends to produce conservative estimates which, while reasonable as upper limits, are not appropriate for extracting further CEs as in this intra-rater study. Other examples of the overestimation of CEs using the Poisson approach can be seen in Cruz-Orive & Gual-Arnau, (2002).

5.7 Concluding remarks

In this chapter we had two main objectives. The first objective was to investigate possible links between Broca's volume estimates relative to total brain volume (relative Broca's volume or RelBAV) and an assortment of factors (musical ability, cognitive ability, age and gender). Our second objective was to repeat the investigation but using Broca's surface area estimates relative to Broca's volume (relative Broca's surface area or RelBASA) instead of volume estimates. To do this we fitted two multivariate linear models with left and right hemisphere RelBAV and RelBASA as outcomes, respectively. We also performed an intra-rater study to obtain estimates of the within observer variability from our volume and surface area estimates.

From the multivariate model with left and right relative Broca's volume as the outcomes we saw that none of the explanatory coefficients had an association to either outcome. However, the closest explanatory coefficient to being statistically significant was the coefficient of gender and particularly for the right hemisphere. This suggests that there was some evidence (at the Bonferroni-adjusted 10% significance level) that females had a smaller relative Broca's volume than males but only in the right hemisphere. It is known from previous studies that the absolute brain volume, for both white and grey matter, is larger for men than for women (Smith *et al.*, 2007). Although, since relative Broca's volume measurements are adjusted for brain size differences between males and females, the associations detected here between gender and relative Broca's volume must be influenced by other factors.

Broca's area in the left hemisphere has previously been found to have no effect on non-verbal tasks, but is strongly linked to the processing of language output functions (Schaffler *et al.*, 1993). However, other studies suggest that Broca's area is involved in non-linguistic skills such as the processing of music, sequencing, action recognition and visuospatial cognition (Koelsch *et al.*, 2002; Schubotz *et al.*, 2002a,b; Hamzei *et al.*, 2003; Sluming *et al.*, 2007). Arithmetic processing requires a combination of different types of knowledge apart from simple number processing and calculation (e.g. reading and writing numbers in both digits and words; Zamarian *et al.*, 2009). Therefore, arithmetic processing comprises a number of non-linguistic skills, which could be linked to Broca's area, specifically, in the right hemisphere. There exists a general perception that females perform better linguistically compared to males, and conversely that males perform better at spatial awareness and mathematical tasks than females. If this perception is true, and if the non-linguistic skills linked with Broca's area are associated more in the right hemisphere, then this could be a reason as to why females have a smaller relative Broca's volume in the right hemisphere, but not in the left hemisphere.

Our fitted multivariate linear model with left and right relative Broca's surface area as outcomes detected that musicians had a smaller relative Broca's surface area, in comparison to non-musicians, in the right hemisphere. Since there isn't any difference in volume of Broca's area between musicians and non-musicians, then this indicates that musicians have less convoluted Broca's area in the right hemisphere, in comparison to controls.

In a similar result to that for the model with relative Broca's volume as the outcome, we found that the coefficient for gender was not significant for RelBASA in either hemisphere at the Bonferroni-corrected 5% significance level. However, we did find the gender coefficient to be statistically significant at the Bonferroni-adjusted 10% level of significance for left hemisphere relative Broca's surface area. The negative coefficient suggested that females have a smaller relative Broca's surface area in the left hemisphere than males. Since Broca's volume is already adjusted for then this result indicates that Broca's area in females is less convoluted than in males, for the left hemisphere only. In *Chapter 4* it has already been pointed out that females achieved better scores than males in the vocabulary test, and since language skills are prominently used in the left hemispheric Broca's area, this reduction in relative Broca's surface area might be associated with the increased ability in the cognitive functions for which Broca's area is used.

From the intra-rater study for Broca's volume estimates, the coefficient of error due to within observer variability was greater than both the estimated coefficient of error due to Cavalieri sectioning and point counting combined. This result suggests that the contribution due to the observer had a greater effect on the variability of the volume estimates than the volume estimation techniques given the sampling parameters in this study. We have identified that the demarcation problems may be causing some additional within observer variability, as could a lack of experience at identifying Broca's area from magnetic resonance images, which can only be built up over time. Additional variability could be estimated from a future study with multiple observers or it could be reduced by increasing the precision of the demarcations of the Broca's area sub-structures, Pars Opercularis and Pars Triangularis, for each participant.

The results from the intra-rater study for Broca's volume were not observed in the intra-rater study for Broca's surface area estimates. In the surface area intra-rater study, the within observer coefficient of error was observed to be 2.1% and 3.4% for the left and right hemisphere PT estimates, respectively. However, for Pars

Opercularis it was not possible to obtain an estimate due to the overestimation of the coefficient of error due to cycloid sampling (see Cruz-Orive, 2002, for examples). We can therefore, conclude that further work is necessary in the estimation method of the coefficient of error of the surface area estimator due to cycloid sampling such that less conservative values are given. This would then allow us to obtain more reliable estimates for the within observer coefficient of error for our intra-rater and inter-rater studies of future applications.

CHAPTER 6

Missing Data

In the analyses described in *Chapters 3* and *4*, there were a number of children and adults that were not actually included in the final results due to them having a number of missing covariate or outcome data. In this chapter, we examine methods for accounting for individuals with missing outcome data in statistical analyses. One of these methods, Inverse Probability Weighting is then used to fit weighted multivariate linear mixed and linear models to the National Child Development Study dataset and the Musician-Control dataset used in *Chapters 3* and *4*, respectively. We compared the results of the weighted models to the unweighted models and conclude whether or not there was any evidence to suggest that missing data had an effect on the conclusions drawn from the results in previous chapters.

6.1 Introduction

In two of our datasets we saw that a number of participants had missing covariates, missing outcomes or both missing covariates and missing outcomes. The National Child Development Study (NCDS) dataset contained information on 6,712 children (36.2% of the total number of children) who had missing data (see Table 3.1). In Table 4.1 we saw that the Musician-Control (MC) dataset included 20 adults (13.9% of the total number of adults in the dataset) who had at least one covariate or outcome value missing. In general, there are three categories of missing data that can occur in a dataset (Tsiatis, 2006):

i. Missing Completely At Random (MCAR)
ii. Missing At Random (MAR)
iii. Not Missing At Random (NMAR)

Missing data are classified as MCAR if the probability of missingness does not depend on any variables, irrespective of whether they are observed or unobserved variables (Everitt & Dunn, 2001). If missing data are categorised as MAR then the probability of those data being missing is dependent upon at least one variable which is included in the dataset, but is independent of all unobserved data (i.e. those variables not included in the dataset) (Tsiatis, 2006). NMAR missing data is defined as such when the probability of missingness is dependent upon some unobserved data which is not included as variables in the dataset (Wayman, 2003). In both of our datasets (the NCDS and MC datasets) we earlier assumed that the missing data was MCAR. However, it may be the case that the data are MAR as we cannot be one hundred percent sure that there are no underlying reasons connected to the outcome measures why data were not collected for every individual included in both datasets. We make the assumption that the missing data are not NMAR because, after discussions with the investigators responsible for the data collection of the MC dataset, and after searching the literature about the National Child Development Study, there were no indications that the probability of participants having missing outcomes was linked to any variables other than those already in the datasets. Also, models which adjust for NMAR data are unverifiable (due to their dependability on data which is unknown) and hence, their correctness is debatable (Tsiatis, 2006).

The statistical methods that we used in *Chapters 3-5*, multivariate linear mixed and linear models, assumed that missing data were MCAR. These methods are called deletion methods due to the fact that any individuals within the datasets that have at least one missing covariate or missing outcome value are omitted from the analyses. This type of method is reliable when the missing data are MCAR. That is, a deletion method is most useful when the remaining data (after deletion) are representative of the complete dataset (before deletion), otherwise bias will occur (Wayman, 2003).

If we assume the missing data to be MAR then there are a number of methods which would adjust for the probability of missingness, of which two methods are:

- Multiple Imputation (MI)
- Inverse Probability Weighting (IPW)

Imputation procedures involve the missing variable values being replaced by predicted values. There are a number of different types of imputation methods including replacement by the variable mean. We applied this approach (replacement of missing variable values by the variable mean) to both the NCDS and MC datasets. The results obtained by refitting the analyses from *Chapters 3* and *4* were very similar to those obtained when individuals with missing values were omitted. Indeed, all statistical significance and interpretations were the same, with only p-values, standard errors and coefficients changing slightly. An issue with basic imputation methods, such as the replacement by variable mean method, is that conservative variances within regression models can be produced. One imputation method that overcame this issue was introduced in 1978 by DB Rubin and was called Multiple Imputation (MI) (Schafer & Olsen, 1998). MI involves an iterative procedure to obtain imputed values for each of the missing data, which is repeated a number of times, resulting in a series of completed datasets (with the missing data imputed). The completed datasets are then used simultaneously to obtain pooled regression models. One of the advantages of MI over other imputation methods is that both the *between* and *within* imputation variability is taken into account in the pooled data (due to the multiple imputations and iterations obtained, respectively; Molenberghs & Kenward, 2007). MI proves to be useful, and is efficient even when there is a relatively high proportion (e.g. 30%) of missing data in the dataset (Schafer & Olsen, 1998). However, on closer inspection, the majority of the children with missing data (36.2% (6,712) of the total number of children with at least one missing outcome or covariate; Table 3.1) have multiple missing variables. This implied that the time needed to calculate MI estimates for these missing covariates and covariates was large.

After investigation using the Multiple Imputation by Chained Equations (MICE) package in R it became apparent that for the estimates to converge larger numbers of iterations and imputations were required. We also found that there was no clearly defined way of fitting linear mixed models and multivariate models to the pooled MI data. A method of exporting the pooled MI dataset to a text file for use in other programs such as MLwiN or SAS was also unclear. Finally, to run the MI analyses in R with the very large number of missing data (both outcomes and covariates) would have taken a number of days or weeks for a computer with the computational power available to us. Therefore, even though MI results in good estimates for the missing data, it was not feasible to explore the effect of MI using the Multiple Imputation by Chained Equations (MICE) package in R at this time.

The Inverse Probability Weighting (IPW) method involves each individual with non-missing outcomes in the dataset being given a specific weight. This weight is associated with the probability that the participant, given his/her characteristics, does not have a missing outcome. These weights are then included in regression analyses performed on this dataset. The method for accounting for missing outcomes using IPW does not use a time consuming iterative procedure like that needed for MI, although the IPW estimates for the missing data can be inefficient relative to likelihood-based analyses and can be sensitive to the model specification for the probability of response (Carpenter, *et al.*, 2006).

Given the impracticality of the MI method with the large amounts of missing data that we had, we chose the IPW method to re-analyse the NCDS and MC datasets by taking into account the missing outcomes.

6.2 Inverse probability weighting methodology

The methodology of inverse probability weighting (IPW) consists of several stages which can be associated with each constituent word (i.e. inverse, probability and weighting). The first stage, *probability*, involves the computation of the probability of a participant having a non-missing outcome by constructing a logistic regression model. In *Chapter 2*, the outcome variable in each model (univariate and multivariate, linear and linear mixed model) was represented in matrix form by the $(n \times 1)$ vector $\underline{Y}$. In Section 6.1, we recalled that those individuals with missing outcomes were omitted from the analyses reported in *Chapters 3* and *4*. When referring to the complete dataset (the dataset which included all the participants with and without missing outcomes) the outcome variable vector will be written as $\underline{\acute{Y}}$ such that:

$$\underline{\acute{Y}} = \begin{bmatrix} \underline{Y} \\ \underline{\breve{Y}} \end{bmatrix} \tag{6.2.1}$$

is an $(n_{tot} \times 1)$ vector with n_{tot} equal to the total number of subjects in the complete dataset and where $\underline{\check{Y}}$ is the $\left((n_{tot} - n) \times 1\right)$ vector of missing outcomes. Let r_i be a binary variable for the i^{th} subject $(1 \leq i \leq n_{tot})$ such that it is equal to 1 when the outcome, y_i, is known and equal to 0 when the outcome is missing. This binary variable can be written as an $(n_{tot} \times 1)$ vector, $\underline{R}$, where:

$$\underline{R} = \begin{bmatrix} r_1 \\ \vdots \\ r_{n_{tot}} \end{bmatrix} \tag{6.2.2}$$

The variable r_i, since it is binary, can be described in terms of a binomial distribution with probability of success for the i^{th} individual (i.e. a non-missing outcome, $r_i = 1$) equal to $\hat{p}_i$. Therefore, the probability of the i^{th} individual having a missing outcome can be written as $(1 - \hat{p}_i)$. When a regression model is to be used and the outcome variable is a binary variable (in this case, r_i) then the type of regression model used is a logistic regression model. Logistic regression does not assume that the relationships between the outcome variable and the explanatory variables are linear. The form of the linear model for the i^{th} subject can be written as:

$$ln\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{i(p-1)} \tag{6.2.3}$$

where $\beta_0, \beta_{1,\ldots,}\beta_{p-1}$ are the p explanatory coefficients and $x_{i1},\ldots,\ x_{i(p-1)}$ are the explanatory variables for the i^{th} individual. Let $z_i = ln\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right)$ then Equation (6.2.3) can be written in matrix form as:

$$\underline{Z} = X\underline{\beta} \tag{6.2.4}$$

where $\underline{Z}$ is the $(n^* \times 1)$ logistic outcome vector such that $\underline{Z} = \begin{bmatrix} z_1 \\ \vdots \\ z_{n^*} \end{bmatrix}$ and n^* is the number of participants from the complete dataset with no missing covariates, $\underline{X}$ is an

$(n^* \times p)$ explanatory design matrix such that $\underline{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1(p-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n^*1} & \cdots & x_{n^*(p-1)} \end{bmatrix}$ and $\underline{\beta}$ is a $(p \times 1)$ vector of explanatory coefficients such that $\underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$. By rearranging the logistic regression model in Equation (6.2.3), the probability of the i^{th} participant not having a missing outcome, $\hat{p}_i$, can be computed as:

$$\hat{p}_i = \frac{exp(\beta_0+\beta_1 x_{i1}+\cdots+\beta_{p-1}x_{i(p-1)})}{1+exp(\beta_0+\beta_1 x_{i1}+\cdots+\beta_{p-1}x_{i(p-1)})} \tag{6.2.5}$$

The word *inverse* is included in the terminology IPW because the probabilities computed in Equation (6.2.5) now have to be inversed. Since we know that $\hat{p}_i$ is the probability that an individual does not have a missing outcome then individuals who are more likely to have a non-missing outcome will have a probability $\hat{p}_i$ that is closer to 1. Individuals who are more likely to have a missing outcome will have probability $\hat{p}_i$ that is closer to 0. However, we want to weight the data such that participants who are more likely to have a missing outcome are given more weight (i.e. a larger weight value), and those participants who are less likely to have a missing value are given less weight. Therefore, the probability $\hat{p}_i$ is inversed such that the weight for the i^{th} individual, $\hat{w}_i$, can be calculated by:

$$\hat{w}_i = 1/\hat{p}_i \tag{6.2.6}$$

It is also useful to note that whereas in the logistic regression model of Equation (6.2.4) we had n^* participants, we only use the weights for the n participants which we have known outcomes for. This is because, as we will shortly show, we multiply the outcomes by the weights in the weighted linear regression model.

The *weighting* part of IPW is now relevant as the weights for those subjects with recorded outcomes, $\hat{w}_1$, ..., $\hat{w}_n$, can be added into the multivariate analyses from *Chapters 3* and *4*, such that they become weighted multivariate linear mixed and linear models, respectively. For the multivariate linear models from *Chapter 4*, *i* represents subjects without a missing outcome or covariate $(1 \leq i \leq n)$. The matrix form for the weighted multivariate linear model (an extension of Equation (2.1.11)) can be written as (Xu & Eckstein, 1995):

$$\underline{Y}^* = \underline{X}^* \underline{\beta}^* + \underline{\varepsilon}^* \qquad (6.2.7)$$

where $\underline{Y}^*$ is the $(dn \times 1)$ outcome vector such that $\underline{Y}^* = \begin{bmatrix} \underline{Y}^{(1)} \\ \vdots \\ \underline{Y}^{(d)} \end{bmatrix}$ and $\underline{Y}^{(k)} = \begin{bmatrix} \widehat{w}_1 y_1^{(k)} \\ \vdots \\ \widehat{w}_n y_n^{(k)} \end{bmatrix}$, $\underline{X}^*$ is a $(dn \times dp)$ design matrix such that $\underline{X}^* = \begin{bmatrix} \underline{X} & & \underline{0} \\ & \ddots & \\ \underline{0} & & \underline{X} \end{bmatrix}$ and $\underline{X} = \begin{bmatrix} \widehat{w}_1 & \widehat{w}_1 x_{11} & \cdots & \widehat{w}_1 x_{1(p-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{w}_n & \widehat{w}_n x_{n1} & \cdots & \widehat{w}_n x_{n(p-1)} \end{bmatrix}$, $\underline{\beta}^*$ is a $(dp \times 1)$ vector of coefficients such that $\underline{\beta}^* = \begin{bmatrix} \underline{\beta}^{(1)} \\ \vdots \\ \underline{\beta}^{(d)} \end{bmatrix}$ and $\underline{\beta}^{(k)} = \begin{bmatrix} \beta_0^{(k)} \\ \beta_1^{(k)} \\ \vdots \\ \beta_{p-1}^{(k)} \end{bmatrix}$, $\underline{\varepsilon}^*$ is the $(dn \times 1)$ error vector as defined in Equation (2.1.11).

Since we adapted the matrix form for the multivariate linear mixed model in Equation (2.1.13) for the analyses in *Chapter 3* (see Equation (3.3.9)) and because these analyses were again performed in this chapter, the weighted multivariate linear mixed model was expressed as an extension of Equation (3.3.9). Let i indicate the i^{th} local authority such that $1 \leq i \leq g$ $(g = 185)$ and j refers to the j^{th} child for the i^{th} local authority where $1 \leq j \leq n_i$ and $n = \sum_{i=1}^{g} n_i$. In this instance, the weights are denoted as $\widehat{w}_{ij}$ which can be interpreted as the weight corresponding to the j^{th} child of the i^{th} local authority such that:

$$\widehat{w}_{ij} = 1/\hat{p}_{ij} \qquad (6.2.8)$$

where $\hat{p}_{ij}$ represents the probability of the j^{th} child for the i^{th} local authority not having a missing outcome. Hence, the weighted bivariate linear mixed model can be written as:

$$\underline{\bar{\bar{Y}}} = \underline{\bar{\bar{X}}}\underline{\beta}^* + \underline{z}^*\underline{b}^* + \underline{\bar{\bar{\varepsilon}}} \tag{6.2.9}$$

where the $(2n \times 1)$ outcome vector $\underline{\bar{\bar{Y}}} = \begin{bmatrix} \underline{\tilde{Y}}^{(1)} \\ \underline{\tilde{Y}}^{(2)} \end{bmatrix}$ with $\underline{\tilde{Y}}^{(k)} = \begin{bmatrix} \underline{\tilde{Y}}_1^{(k)} \\ \vdots \\ \underline{\tilde{Y}}_g^{(k)} \end{bmatrix}$ and $\underline{\tilde{Y}}_i^{(k)} = \begin{bmatrix} \hat{w}_{i1} y_{i1}^{(k)} \\ \vdots \\ \hat{w}_{in_i} y_{in_i}^{(k)} \end{bmatrix}$ ($k = 1,2$ is an indicator variable for the reading and transformed mathematics test scores, respectively). The fixed effect design matrix $\underline{\bar{\bar{X}}} = \begin{bmatrix} \underline{\tilde{X}} & \underline{0} \\ \underline{0} & \underline{\tilde{X}} \end{bmatrix}$ is a $(2n \times 20)$ matrix (there are 10 explanatory variables in the model, including the intercept) such that $\underline{\tilde{X}} = \begin{bmatrix} \underline{\tilde{X}}_1 \\ \vdots \\ \underline{\tilde{X}}_g \end{bmatrix}$ is a $(n \times 10)$ design matrix such that $\underline{\tilde{X}}_i = \begin{bmatrix} \hat{w}_{i1} & \hat{w}_{i1}x_{i1} & \cdots & \hat{w}_{i1}x_{i(p-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{w}_{in_i} & \hat{w}_{in_i}x_{i1} & \cdots & \hat{w}_{in_i}x_{i(p-1)} \end{bmatrix}$ is an $(n_i \times 10)$ matrix. The vector $\underline{\beta}^* = \begin{bmatrix} \underline{\beta}^{(1)} \\ \underline{\beta}^{(2)} \end{bmatrix}$ is the (20×1) vector of coefficients as defined for Equation (3.3.9). The random effect terms $\underline{z}^*$ and $\underline{b}^*$ along with the random error term $\underline{\bar{\bar{\varepsilon}}}$ are also as defined for Equation (3.3.9) as the weighting does not affect these terms. The IPW methodology using different notation can be seen in Tsiatis, (2006) and Molenberghs & Kenward, (2007).

6.3 National Child Development Study dataset

We re-analysed the National Child Development Study (NCDS) data using inverse probability weighting (IPW) and the results are given in this subsection. Firstly, a logistic regression model was fitted to the data, with the additional variable R as the outcome. As stated in Section 6.1, 36.2% of the total number of children in the NCDS dataset had at least one missing outcome or covariate (6,712 out of 18,558) as seen in Table 3.1. The resulting probabilities, $\hat{p}_{ij}$, were then obtained for each child with complete data (i.e. no missing outcomes or covariates), using the associations between the explanatory variables in the logistic regression model and the binary variable R and a weighting (probability) variable $\hat{w}_{ij}$ was calculated. These weights were then applied to the linear regression models fitted to the data to obtain models, which may or may not differ to those seen in *Chapter 3*.

After comparing the model reported in *Chapter 3* to the model in this subsection, if they were found to be very similar, then this would suggest that the complete data (data with no missing outcomes or covariates) provide a reasonable representation of the total dataset (data including all subjects irrespective of whether they have missing outcomes, missing covariates or no missing data at all; i.e. the missing data can be assumed to be missing completely at random (MCAR)). If the models were different, then there would have been evidence to suggest that the individuals with missing data have similar characteristics that are not the same as those without missing data. Therefore, the missing data should be treated as missing at random (MAR) and require an adjustment to the standard linear regression procedure (e.g. IPW).

The logistic regression models were considered with the variable R as the outcome variable where:

$$r_{ij} = \begin{cases} 0 & \text{when at least one of reading and mathematics test scores are missing} \\ 1 & \text{when both reading and mathematics test scores have been recorded} \end{cases}$$

The explanatory variables considered were (see *Chapters 1* and *3* for definitions):

- Gender
- Writing Hand (WH)
- Region
- Relative Hand Skill (RelHS)
- Superior Hand (SH)

Using a stepwise approach to model selection (as described in Section 3.4), five possible logistic models were considered. While we were using the stepwise model selection approach, we found that some explanatory variables appeared to be closely related to each other in terms of the variability they were explaining in the variable R. These close relationships meant that as we removed one variable which became non-significant, another variable would become significant, which in turn would cause another to become non-significant, and another variable would become significant. This caused a loop in the stepwise selection process, which meant that we had to use a different approach to selecting which model, out of the five candidate models, best fitted the data. The explanatory variables which were significant for each model can be seen in Table 6.1.

Our simplest model (model 1) included just one explanatory variable: region. The most complex model (model 5) contained 11 explanatory variables including 5 first order interaction terms and 1 second order interaction term. We decided to compare these 5 models to investigate which model should be used to provide the weights for the IPW method, by considering the goodness-of-fit statistics for each model (as seen in Table 6.2). The model with the smallest AIC and BIC value is defined to be the better fitting model, which in Table 6.2 looks to be model 5 (5810 and 5944, respectively). However, the number of subjects omitted in models 1, 2-4 and 5 differ due to missing covariates. This means that the AIC and BIC cannot be directly compared.

Model #	Explanatory Variables	AIC	BIC	Residual Deviance
1	Region	8583	8614	8575
2	Region, RelHS, Region × RelHS	5900	5960	5884
3	Region, RelHS, Gender, Region × RelHS, Region × Gender	5902	5991	5878
4	Region, RelHS, Gender, SH, Region × RelHS, Region × Gender, RelHS × SH	5902	6006	5874
5	Region, RelHS, Gender, SH, WH, Region × RelHS, Region × Gender, RelHS × SH, RelHS × WH, SH × WH, RelHS × SH × WH	5810	5944	5774

Table 6.1: *Explanatory variables of five logistic models applied to the NCDS dataset with goodness-of-fit statistics (AIC, BIC and residual deviance) for each model.*

An analysis of deviance test (see Section 3.4.2 for more details) would normally be used to test for differences between the models, but this is not possible for models 1 and 5 due to the differences in number of children included (i.e. differing numbers of children with missing covariates). However, the deviance test can be performed for models 2 to 4, since they have the same number of children included. The results of these deviance tests can be seen in Table 6.2. No differences were found between the three models (p=0.2 and p=0.1 for the tests when comparing models 2 and 3 and models 3 and 4, respectively, Table 6.2). Therefore, we can see that almost identical information can be gathered about the logistic regression outcome variable R from model 2 as can be obtained from models 3 and 4. Models 3 and 4 are more complex and therefore, given the fact that they provide little extra information about R, we can deem this added model complexity to be unnecessary. The AIC an BIC values in Table 6.1 also add extra weight to this conclusion, since out of the three models, the model with the smallest AIC and BIC values was model 2 (5900 and 5960, respectively). Even though we cannot compare models 1, 2 and 5 directly we can observe a very simplistic comparison from Table 6.1. Model 1 is the most simplistic model, containing only one explanatory variable and although we cannot gauge a comparison using the deviance, AIC or BIC statistics we can suggest that this would not be a suitable model. The unsuitability of model 1 was due to it being too simplistic as well as model 2 being reasonable and simplistic enough to fit without any issues. The most complex model (model 5) would be more difficult to implement in the construction of the inverse probability weights due to the fact that it contains eleven explanatory variables including six interaction terms. Therefore, we decided to not use model 5, but instead opt for model 2. The results of this fitted logistic regression model (i.e. model 2) can be seen in Table 6.3.

Test	**Residual Deviance**	**df**	***p*-value**
Model 2 vs Model 3	5884 – 5878 = 6.6	4	0.2
Model 3 vs Model 4	5878 – 5874 = 3.9	2	0.1

Table 6.2: *The deviance test results between logistic regression models 2, 3 and 4.*

Explanatory Variable	**Coeff.**	**St. error**	**95% CI**	***p*-value**
Intercept	2.7	0.1	(2.5, 2.9)	<0.001
Relative Hand Skill	-0.006	0.005	(-0.02, 0.005)	0.3
Northern England & Midlands	*Reference region*			
Southern England	0.07	0.2	(-0.3, 0.4)	0.7
Wales	-0.4	0.4	(-1.2, 0.4)	0.3
Scotland	0.1	0.3	(-0.4, 0.7)	0.7
Northern England & Midlands × Relative Hand Skill	*Reference region*			
Southern England × Relative Hand Skill	0.006	0.01	(-0.01, 0.02)	0.5
Wales × Relative Hand Skill	0.06	0.02	(0.01, 0.1)	0.01
Scotland × Relative Hand Skill	0.02	0.02	(-0.009, 0.06)	0.2

Table 6.3: *Estimates of the explanatory coefficients for the logistic regression model (model 2) with R as the outcome (including the standard error of the corresponding estimators, the 95% confidence intervals for the coefficients and the p-values).*

Using the results from Table 6.3, the weights were calculated as in Equation (6.2.8). These weights were then applied in the linear mixed effects models fitted to the NCDS dataset. The stepwise model selection method was utilised in choosing the models, as described in *Chapter 3*. The resulting model from this process can be seen in Tables 6.4 and 6.5. The model takes the same structure as that in *Chapter 3* (see Tables 3.2 and 3.4, respectively). The explanatory variable coefficients, standard errors, 95% CIs and *p*-values are also very similar, with most values very close to those before the weighting was applied. Indeed, the interpretation of the model is unchanged with those variables significant in Section 3.4.1 also being significant with the IPW model in Tables 6.4 and 6.5. Associations between the linear combination of the two outcomes (reading and transformed maths scores; i.e. the multivariate outcome) can be seen in Table 6.4. To give an idea of how close the results were, from the IPW weighted multivariate model the coefficient for superior hand × writing hand interaction was 13.5 (95% CI: (8.3, 18.6), p<0.001; Table 6.4) and for the unweighted model it was 13.2 (95% CI: (8.1, 18.4), p<0.001; Table 3.2). That is, the change in coefficient was approximately equal to 2.3%, and made no difference to the interpretation of the results. Other comparisons between the coefficients, standard errors, 95% CIs and *p*-values of the two models have similar differences which make no difference to the overall interpretations from Section 3.4.1.

Similarly to the results of the multivariate outcome, the results from the individual models based on the marginal distributions of the weighted multivariate model were very similar to those for the multivariate model of Section 3.4.1, with the interpretations from *Chapter 3* remaining unchanged. For example, in Table 6.5 the coefficient for writing hand in the reading scores was -5.1 (95% CI: (-8.6, -1.6), p=0.004) whereas from the unweighted multivariate model we obtained a coefficient of -5.0 (95% CI: (-8.5, -1.5), p=0.005; Table 3.4). This change in coefficient represented an approximate decrease of 2%, and did not affect the interpretation of the result.

Fixed Effect	Coefficient	St. Error	95% CI	*p*-value
Intercept	52.3	0.7	(50.8, 53.7)	<0.001
Relative hand skill	0.6	0.07	(0.4, 0.7)	<0.001
(Relative hand skill)^2	-0.01	0.002	(-0.014, -0.008)	<0.001
Superior hand (SH) *(0: Right, 1: Left)*	-7.8	1.4	(-10.6, -5.1)	<0.001
Writing hand (WH) *(0:Right, 1:Left)*	-6.5	2.1	(-10.6, -2.3)	0.002
SH x WH	13.5	2.6	(8.3, 18.6)	<0.001
Gender *(0: Male, 1: Female)*	-0.02	0.4	(-0.8, 0.7)	0.9
Northern England & Midlands	*Reference Region*			
Southern England	1.8	0.5	(0.9, 2.8)	<0.001
Wales	-0.3	1.0	(-2.2, 1.6)	0.8
Scotland	0.1	0.7	(-1.2, 1.5)	0.8

Table 6.4: *Estimates of the multivariate fixed effect coefficients for the IPW weighted multivariate linear mixed model with both the reading and mathematics scores combined (including the standard error of the corresponding estimators, the 95% confidence intervals for the coefficients and the p-values).*

When comparing the random effect and random error, variances, covariances and correlations we also saw that the estimates were very similar between the two multivariate models (in Tables 3.4 and 6.5). One of the slight differences was that the correlation between the two outcomes at the local authority level (i.e. from the random effects) was equal to 0.95 for the weighted model compared to 0.97 for the uncorrelated model (a change of just 2.1%; both $p<0.001$; see Tables 6.5 and 3.4, respectively) Also, in Table 3.4 it was seen that 5.6% of the total variation not explained by the fixed effects in the reading scores and 4.5% of the total variation unexplained in the transformed mathematics scores were accounted by the variability across subjects due to local authority. For the inverse probability weighted model in Table 6.5, this variability due to local authority accounted for 5.4% and 4.8% of the total unexplained variation in the reading and transformed mathematics scores, respectively.

Fixed Effect	Coeff.	St. Error	95% CI	*p*-value
Reading Test Scores				
Intercept	42.2	0.6	(41.0, 43.4)	<0.001
Relative hand skill	0.4	0.06	(0.3, 0.5)	<0.001
(Relative hand skill)^2	-0.009	0.001	(-0.012, -0.006)	<0.001
Superior hand (SH) *(0: Right, 1: Left)*	-6.3	1.2	(-8.6, -4.0)	<0.001
Writing hand (WH) *(0:Right, 1:Left)*	-5.1	1.8	(-8.6, -1.6)	0.004
SH x WH	10.7	2.2	(6.4, 15.0)	<0.001
Gender *(0: Male, 1: Female)*	0.1	0.3	(-0.6, 0.7)	0.8
Northern England & Midlands	*Reference Region*			
Southern England	1.7	0.4	(0.9, 2.4)	<0.001
Wales	-0.7	0.8	(-2.3, 0.9)	0.4
Scotland	-0.4	0.6	(-1.5, 0.8)	0.6
Transformed Mathematics Test Scores				
Intercept	10.0	0.1	(9.8, 10.3)	-
Relative hand skill	0.1	0.01	(0.1, 0.2)	<0.001
(Relative hand skill)^2	-0.002	<0.0005	(-0.003, -0.001)	<0.001
Superior hand (SH) *(0: Right, 1: Left)*	-1.6	0.3	(-2.1, -1.0)	<0.001
Writing hand (WH) *(0:Right, 1:Left)*	-1.3	0.4	(-2.2, -0.5)	0.002
SH x WH	2.7	0.5	(1.7, 3.8)	<0.001
Gender *(0: Male, 1: Female)*	-0.1	0.08	(-0.3, 0.03)	0.1
Northern England & Midlands	*Reference Region*			
Southern England	0.2	0.09	(0.03, 0.4)	0.02
Wales	0.4	0.2	(-0.009, 0.8)	0.06
Scotland	0.5	0.1	(0.2, 0.8)	<0.001
Random Effect	**Est.**	**St. Error**	**95% CI**	***p*-value**
Variance (Reading Scores)	18.2	3.1	(12.1, 24.3)	<0.001
Variance (Trans. Maths Scores)	0.9	0.2	(0.6, 1.3)	<0.001
Covariance [Correlation] (Reading/Trans. Maths Scores)	3.9 [0.95]	0.7 [0.2]	(2.6, 5.2) [0.63, 1.00]	<0.001
Random Error	**Est.**	**St. Error**	**95% CI**	***p*-value**
Variance (Reading Scores)	318.3	5.0	(308.4, 328.2)	<0.001
Variance (Trans. Maths Scores)	18.6	0.3	(18.0, 19.1)	<0.001
Covariance [Correlation] (Reading/Trans. Maths Scores)	56.8 [0.74]	1.1 [0.01]	(54.7, 58.9) [0.71, 0.77]	<0.001

Table 6.5: *Estimates of the fixed effects coefficients and the random effect and random error variance, covariance and correlation terms for the IPW weighted multivariate linear mixed model with reading and transformed maths scores as outcomes (including the standard error of the corresponding estimators, the 95% confidence intervals for the coefficients/terms and the p-values).*

The fact that the model from Section 3.4.1, fitted using a deletion method which omits all individuals with missing outcomes or covariates gave such similar results to our model in this subsection, fitted using IPW which takes individuals with missing outcomes into account, suggests that the missing outcomes may be MCAR. That is, the missing outcomes do not depend on any variables either observed or unobserved. One way of checking this is by using Mann-Whitney tests and chi-square tests to test the hypothesis that there are no differences in means and numbers, respectively, between the group with missing outcomes and the group with non-missing outcomes. We used these non-parametric tests due to our outcome variable (missing/non-missing) being binary. The p-values for the Mann-Whitney and chi-squared tests can be seen in Table 6.6. Our results from these tests show that apart from UK region, there was no evidence to suggest that there was a difference between left and right superior handers, left- and right-hand writers, males and females or between the mean relative hand skill value of each group (p=0.8, p=0.9, p=0.08 and p=0.6, respectively; Table 6.6). However, there was a difference in the numbers of individuals from each UK region between the missing and non-missing outcome groups (p=0.005, Table 6.6).

Explanatory Variable	***p*-value**
Chi-square tests	
Superior Hand	0.8
Writing Hand	0.9
Gender	0.08
UK Region	0.005
Mann-Whitney test	
Relative Hand Skill	0.6

Table 6.6: *Estimated p-values for the Mann-Whitney and chi-square tests between the missing and non-missing outcome groups for each explanatory variable in the multivariate linear mixed model with reading and transformed maths scores as outcomes.*

Table 6.7 shows the numbers of individuals in each UK region by the missingness of outcomes. Although there was a slight difference in the percentage of the total groups with the largest being for the Northern England and Midlands region (53.3% of the non-missing group compared to 57.1% of the missing group; a difference of 3.8%; Table 6.7), they were not large. Therefore, an explanation as to why this difference is being highlighted by the chi-square test as being significant is down to the fact that the dataset is so large, and thus small differences are more likely to be found by hypotheses tests. Considering the small differences between UK regions and the lack of differences in the other explanatory variables, then this may be why the effect of missing outcomes in the IPW model is very small in comparison to the unweighted model.

Outcome	**UK Region**				***Total***
	Northern England & the Midlands	**Southern England**	**Wales**	**Scotland**	
Not Missing	7534 (53.3%)	4339 (30.7%)	767 (5.4%)	1483 (10.5%)	14123
Missing	704 (57.1%)	379 (30.7%)	50 (4.1%)	101 (8.2%)	1234
Total	8238	4718	817	1584	15357

Table 6.7: *Breakdown of the numbers in each UK region of individuals with missing and non-missing outcomes.*

We saw that the results from the IPW model were very similar to the unweighted model. The similarity between models meant that the diagnostic plots for the IPW multivariate model were also very similar to those in Figures 3.4 and 3.5. Therefore, we can see that the model is a reasonable fit to the data, and that the fit was better to the reading scores than for the transformed maths scores.

6.4 Musician-control dataset

In *Chapter 4*, the musician-control (MC) dataset was analysed without adjusting for missing data. The missing data for this dataset can be seen in Table 4.1. To re-analyse this data, taking the missing outcomes into account, the IPW method was utilised. A logistic regression analysis was conducted first. Similar to the definition in Section 6.2, the variable r_i was the outcome variable for the logistic regression analysis where:

$$r_i = \begin{cases} 0 & \text{when at least one verbal, arithmetic and visuospatial test score was missing} \\ 1 & \text{when scores for the verbal, arithmetic and visuospatial tests were recorded} \end{cases}$$

for the i^{th} subject. The explanatory variables to be considered not only included those considered for the unweighted analyses in *Chapter 4*, but also some other characteristic variables related to each subject. These variables were:

- *Gender* (0=male; 1=female)
- *Age* (in years)
- *Musician* (0=non-musician; 1=musician)
- *Height* (in cm)
- *Weight* (in kg)

Please note that from this point onwards we describe the factor *musician* in italics to distinguish it from a musician (i.e. someone who plays a musical instrument). We commenced the logistic regression analysis using a stepwise model selection method (as described in Section 3.4). However, after considering all variables, it was found that the only suitable logistic regression model was the model with the intercept only (i.e. none of the variables mentioned above were statistically significant). This can be interpreted as there being no significant associations between any of the observed explanatory variables within the dataset and the probability of missingness in terms of an individual's outcome. This implies that the missing outcomes should be regarded as MCAR. Therefore, the unweighted model in Section 4.4 was identical to the IPW model since all participants were given a constant weight.

We performed a similar check to the analyses in Section 6.3, by testing for differences between the missing and non-missing outcome groups in terms of the explanatory variables, using Mann-Whitney and Fisher exact tests. We used Fisher exact tests instead of chi-square tests because at least one combination of male or female, musician or non-musician for either the missing or non-missing groups, contained less than or equal to 5 individuals. The p-values from these tests can be seen in Table 6.8. The Mann-Whitney tests that we fitted for the continuous variables all suggested that there was no evidence to reject the null hypothesis which was equality of means between the missing and non-missing outcome groups (p=0.3, p=0.2 and p=0.7 for age, height and weight, respectively, Table 6.8). The Fisher exact tests suggested that there was no difference in the proportion of males and females between the missing and non-missing groups (p=0.2, Table 6.8). However, Table 6.8 did show a significant result for the Fisher exact test of *musician* (p=0.007). This implies that there was a difference in the proportion of musicians and non-musicians between the participants with missing outcomes and those participants without missing outcomes.

Explanatory Variable	***p*-value**
Fisher exact tests	
Musician	0.007
Gender	0.2
Mann-Whitney tests	
Age	0.3
Height	0.2
Weight	0.7

Table 6.8: *Estimated p-values for the Mann-Whitney and Fisher exact tests between the missing and non-missing outcome groups for each explanatory variable considered for the logistic regression model from the MC dataset.*

Table 6.9 shows the breakdown of *musician* by missingness groups and we can easily see that there are no musicians with missing outcomes. We stated previously that there were no explanatory coefficients which were significantly non-zero in the logistic regression model with R as the outcome and this included the coefficient for *musician*. The p-values for the coefficients of *musician* in the models based on the marginal distributions of the MLM ranged from less than 0.001 up to 0.01 for all outcomes. Since *musician* is strongly associated with the three outcomes, it suggests that, given the additional 18 non-musicians with missing outcomes (around 17% of the number of non-musicians with non-missing outcomes), the change in the p-

values for *musician* would, in all likelihood, not be enough to alter the final interpretations from the model.

Outcome	Musician (Factor)		Total
	Non-Musicians	Musicians	
Not Missing	88 (71.0%)	36 (29.0%)	124
Missing	18 (100%)	0 (0.0%)	18
Total	106	36	142

Table 6.9: *Breakdown of the numbers in the musician and non-musician groups of individuals with missing and non-missing outcomes.*

6.5 Concluding remarks

The aim of this chapter was to investigate the possible effect of missing data on the results from the analyses conducted in earlier chapters of the thesis. We considered multiple imputation (MI) and inverse probability weighting (IPW) as two methods for adjusting for missing outcomes. MI was deemed to be too time consuming given the amount of children with, not only missing outcomes, but missing covariates as well. The method of IPW included the fitting of a logistic regression model to explore which explanatory variables were associated with a binary variable stating whether an individual had a missing outcome or not. From this logistic regression model we obtain probabilities of missingness for each individual and weights are calculated which are then used to fit weighted multivariate linear mixed or linear models.

We firstly performed the IPW method on the National Child Development Study (NCDS) dataset. From our IPW multivariate linear mixed model we obtained coefficients, standard errors, 95% confidence intervals and p-values which were highly similar to the unweighted model in *Chapter 3*, for both the multivariate outcome and for each outcome based on the marginal distributions of the multivariate model. Such was the similarity that although changes were approximately within a range of 2%, all of the interpretations for each of the fixed effects remained the same as those for the analyses in *Chapter 3*. We then investigated whether we could find any characteristics from the dataset itself in terms of differences between the group of individuals with missing outcomes and the group who had no missing outcomes. We performed chi-squared and Mann-Whitney tests on the data and found that the only variable which differed between the two groups was UK region. However, even though this difference was statistically significant there was only a maximum difference of 3.8% in one of the regions, between the two groups (Northern England and the Midlands). This suggested that there was no evidence of differences between the missing and non-missing outcome groups for the other explanatory variables (relative hand skill, superior hand, writing hand and gender) and possibly not a biologically significant difference between the groups in term of UK region. Therefore, we suggested that the missing outcomes were missing completely at random (MCAR). MCAR missing data do not rely on any variables in the dataset or any unobserved data. We can also say that the individuals with missing outcomes had very little overall effect on the original analyses of this dataset from *Chapter 3*.

We then used the IPW method to consider the effect of individuals with missing outcomes on the analyses in *Chapter 4*, based on the musician-control (MC) dataset. Our fitted logistic regression model contained the intercept as the only explanatory variable. Therefore, the weights applied to the individuals in the data were constant, meaning that the IPW multivariate model was identical to the unweighted multivariate linear model fitted in *Chapter 4*. We investigated this result by performing Fisher exact and Mann-Whitney tests to test for differences between the groups with missing and non-missing outcomes for each explanatory variable (age, height, weight, *musician* and gender). The only result that was significant was for *musician* due to the fact that no musicians, out of 36, had missing outcomes. All 18 of the participants with missing outcomes were non-musicians. However, we concluded that since the number of controls with missing outcomes represented approximately 17% of the total number of controls, this is unlikely to have altered the strongly significant p-values in the MLM for *musician*. Therefore, the interpretation obtained from the MLM would not have changed. Thus, we can assume the missing outcomes to be MCAR and say that there was no evidence to suggest that individuals with missing outcomes had an effect on the results of the original analyses in *Chapter 4*.

CHAPTER 7

Discussion

In this chapter we firstly recap the pathway through the thesis before focussing on the results obtained in detail. We discuss the interpretation of the results from each chapter and suggest possible explanations for these results. Limitations of the analyses that we undertook are then discussed. We consider improvements which could have increased the reliability and increased the scope of our investigation. These improvements could be implemented in a future study. The final section of this chapter summarises the conclusions that we obtained from our investigations.

7.1 Introduction

We set out six main objectives in *Chapter 1*, which were to be investigated in this thesis. These objectives mainly revolved around associations between cognitive ability scores, volume and surface area estimates of Broca's area and a range of other factors in both children and adults. One thing that all the objectives had in common was that in each case, we wanted to determine if there were any benefits of applying multivariate linear and linear mixed models to the datasets, as opposed to univariate models, in obtaining results for the objectives of the study.

We assessed the possible links between two cognitive ability scores (reading and mathematics) and handedness, gender and UK region in a dataset of 11-year old children (see *Chapter 3*). In *Chapter 4*, multivariate models were fitted to a dataset containing adults in order to complete the objective relating to adults. The relationships between three cognitive ability test scores (verbal, arithmetic and visuospatial) and three explanatory variables (musical ability, gender and age) were considered.

Chapter 5 took a slightly different approach to the previous two chapters such that the cognitive ability scores were not included in the models as outcome variables, but as explanatory variables. In Section 5.4, we fitted a multivariate model with volume estimates of Broca's area relative to the total brain volume (relative Broca's volume) in both the left and right hemispheres as outcomes to identify possible associations between relative Broca's volume and a variety of explanatory variables (musical ability, gender, age and the three cognitive ability scores as defined in *Chapter 4*). We investigated similar associations in Section 5.5 between surface area estimates of Broca's area relative to the volume estimates of Broca's area (relative Broca's surface area) and the same six explanatory variables as in Section 5.4. The examination of the links including relative Broca's surface area involved multivariate models fitted to the dataset which had two outcomes corresponding to the left and right relative Broca's surface area estimates (i.e. an estimate for each hemisphere).

Missing data were taken into account in the analyses for both children and adults in *Chapters 3* and *4*, respectively (i.e. some individuals had missing outcomes, missing covariates or both). However, when we fitted multivariate linear mixed and linear models the individuals with missing data were omitted. In *Chapter 6*, we applied the inverse probability weighting method to both datasets. The use of this method allowed multivariate models to be fitted while adjusting for missing outcomes, although individuals with missing covariates were still omitted. Results obtained using the inverse probability weighting method were then compared to those stated earlier from the analyses in *Chapters 3* and *4*.

7.2 Results and interpretations

The results that we obtained for each objective were described and interpreted separately within each chapter. This section contains a number of subsections relating to the interpretation of results for analyses based around individual objectives. Although not explicitly stated as an objective, a theme that runs through all the chapters and objectives was one of testing whether a multivariate model gives a better fit to the data. We start with results from this comparison (multivariate models versus univariate models) in Section 7.2.1.

7.2.1 Multivariate versus univariate

Three different datasets were used in the analyses of *Chapters 3-6*. The main differences between these three datasets were that one was a dataset of children, one comprised of adults, and the third was a dataset of adults which included volume and surface area estimates of Broca's area of each participant (see *Chapter 1* for more details). In each case, multivariate linear mixed models (MLMMs) or multivariate linear models (MLMs) were applied.

Multivariate models are applicable where there are multiple dependent outcomes which are measured at the same time. When a dataset has multiple outcome variables there are two approaches to the statistical analyses. One approach is to perform multiple independent univariate analyses with each dependent variable as the outcome. However, it can be considered equivalent to performing multiple comparisons on the same data which reduces the power of the tests (Littell *et al.*, 2002). The second approach is to include all the outcomes in one multivariate model. All the information that can be gained from the univariate models can be obtained from the multivariate model, in addition to associations between the explanatory variables and the outcome variables combined, and the correlation between outcomes. However, the multivariate models are more complex, and therefore require more computational power. It has been reported that reasons for using univariate models as opposed to a single multivariate model when investigating multiple outcomes include (Huberty & Morris, 1989):

i. Low correlation between outcome variables.
ii. Small number of outcome variables.
iii. Small number of data.

In this thesis, the datasets used ranged in size from 39 participants in the dataset containing Broca's area volume and surface area estimates, to 18,558 children in the National Child Development Study dataset. Also, there were either 2 or 3 outcome variables for each set of analyses.

In *Chapter 3*, we found that there was strong evidence to suggest that the reading and mathematics test scores of 11-year old children from the National Child Development Study (NCDS) were positively correlated at both the individual and local authority levels ($r = 0.74$, $p < 0.001$ and $r = 0.97$, $p < 0.001$, respectively). The strong, positive correlation at the individual level implies there was evidence pointing toward children with high or low scores in one of the tests obtaining equivalently high or low scores on the other test. Similarly, the strong, positive correlation at the local authority level suggested that a given local authority had, in general, similar proportions of pupils performing well on both tests, and conversely, of pupils doing badly on both tests.

We approximated the two univariate models by a MLMM with independent outcomes (independent MLMM). A deviance test was then constructed to compare the independent MLMM against our fitted MLMM with correlated outcomes (correlated MLMM). The correlated MLMM was shown to be significantly different to the independent MLMM, and therefore, the univariate models as well ($p < 0.001$). Furthermore, we also saw that the deviance of the correlated MLMM was smaller than the deviance of the independent MLMM, which demonstrated that our fitted multivariate model was a better fit to the data than two independent univariate models.

The correlation between the vocabulary, arithmetic and visuospatial scores in *Chapter 4*, observed directly from the data were estimated to be equal to 0.5 between the arithmetic and visuospatial scores, 0.4 between the vocabulary and arithmetic scores and 0.3 between the vocabulary and visuospatial scores. The multivariate linear model fitted gave conservative estimates of the correlation between the three cognitive ability scores which were between 50%-67% less than those observed directly from the data. However, we constructed a deviance test which indicated that the multivariate model was different to the univariate models (again by comparing the MLM with correlated outcomes (correlated MLM) to a MLM with independent outcomes (independent MLM); $p = 0.04$). The fact that the deviance was smaller for the correlated MLM than for the independent MLM suggested that the multivariate model was a more suitable model in terms of describing the variability within the outcomes and their associations with the explanatory variables.

In *Chapter 5*, we saw that the estimated correlation between the surface area estimates of Broca's area relative to the volume estimates of Broca's area (relative Broca's surface area) of the left and right hemispheres, both from the raw data and the multivariate was equal to 0.7. This fairly strong, positive correlation suggests that adults who have a larger relative Broca's surface area in the left or right hemisphere are more likely to also have a larger relative Broca's surface area in the other hemisphere, and vice versa. The deviance test for the correlated MLM and independent MLM (equivalent to two univariate models) showed that the difference between the two models was statistically significant ($p < 0.001$). The deviance value for the correlated MLM was smaller which implied that it fitted the data better than the independent MLM (and consequently the two univariate models) did.

When we applied the deviance test to the correlated and independent MLMs with left and right volume estimates of Broca's area relative to total brain volume (relative Broca's volume) as the outcomes, also in *Chapter 5*, we found that there was no evidence to suggest that the two models were different ($p = 0.2$). The correlation between these two outcomes (relative Broca's volume in the left and right hemispheres) was low and there was not enough evidence to suggest that the correlation was non-zero ($r = 0.2, p = 0.1$). On this occasion, we could therefore, conclude that the two univariate models would have been sufficient to explain the variability in the outcomes without the added complexity of a multivariate model.

Our conclusion for this objective is that the multivariate model should still be the first step when multiple outcomes are considered from the same dataset, even where the multivariate model explains no further variability in the outcomes than multiple univariate models. This is due to the fact that all the information that can be obtained from univariate models is also included in the multivariate model (i.e. associations between the explanatory variables and individual outcomes can be assessed from the models based on the marginal distributions of a multivariate model). Hence, to achieve the objectives set out in *Chapter 1*, pertaining to associations between cognitive ability, handedness, gender, age and Broca's volume and surface area estimates, in both children and adults, our focus was on the results from multivariate linear and linear mixed models as opposed to multiple univariate linear and linear mixed models. Interpretations of, and possible explanations for the results from these analyses can be seen in the following subsections.

7.2.2 Associations between handedness and cognitive function in children

Investigating the associations between cognitive ability and handedness and gender in children was one of the objectives as described in *Chapter 1*. These associations were examined by performing multivariate linear mixed model analyses on a dataset of 11-year old children from the National Child Development Study. Cognitive ability were represented by reading and transformed mathematics scores and the explanatory variables considered were relative hand skill (defined as the modulus of the difference in hand skill between the left and right hands divided by the combined left and right hand skill), superior hand (SH) (a binary variable defined as the hand which had the greatest hand skill score), writing hand (WH), gender, and UK region (a 4-level factor) (see *Chapter 1* for further details on the outcome and explanatory variables). A further variable, local authority, which denoted which local authority each child attended school in was used as a clustering variable (a random effect), in the linear mixed models that were fitted.

After taking both WH and SH into account and since the interaction term between WH and SH was strongly significant (for both the models based on the marginal distributions and that with the multivariate outcome), in *Chapter 3* we saw that children with inconsistent WH and SH (i.e. those children that had a different WH to their SH) performed worse, on average, in both the reading and maths tests than those children with consistent WH and SH (i.e. those that write with their hand with the greatest level of hand skill), irrespective of whether they were left- or right-handed.

The rule of multiple models meant that we had to adjust the critical α significance level using a Bonferroni correction (α was reduced from 0.05 to 0.025). Even so, in the reading test, it was shown that those children with left WH and right SH performed worse than their peers that had consistently right and left WH and SH by 5% and 4%, respectively, on average ($p = 0.02$, and $p = 0.005$, respectively). Also, children with right WH and left SH were shown to have attained lower, on average, reading scores than those children with consistently right and left WH and SH by 6% and 5.5%, respectively (both $p < 0.001$). Since the mean reading score was 46% (from Table 1.2), we can see that these differences between children with consistent and inconsistent WH and SH ranged from 9% to 13% of the mean score.

For the transformed maths scores, we saw that these differences were identical in terms of being positive or negative to those for the reading scores, giving similar interpretations, as well as being significant ($p = 0.009$, and $p = 0.002$ for differences between those children with left WH, right SH and those with consistently right and left WH and SH, respectively; $p < 0.001$ for both differences between those children with right WH, left SH and those with consistently right and left WH and SH). So we can say that children with left WH and right SH obtained lower transformed maths scores than those that had consistently right and left WH and SH by 7% and 6% (12% and 10% relative to the mean score), respectively, on average with right WH and left SH children also associated with lower transformed maths scores, on average, than those children with consistently right and left WH and SH by 8% and 7% (14% and 12.5% relative to the mean score), respectively. It is useful to note that the percentage differences are shown here to give an idea of the effect size, but since the transformation of the original maths scores was non-linear these percentages do not translate to a linear relationship between the transformed scores and the covariates.

The percentage differences in both reading and maths scores between children with consistent WH and SH, and those with inconsistent WH and SH have been shown to be between 4% and 8% of the maximum score as well as between 9% and 14% of the mean score. Literacy and numeracy are both regarded as basic skills which are assessed through both English and Mathematics subjects in the education system (Bynner, 1997). Both reading and maths skills are related to literacy and numeracy, respectively, and are therefore important in examinations and tests throughout the course of a child's time at school. Some of the most important examinations which future employers will scrutinize are GCSEs, which children in the UK sit at age 16. Grade boundaries in these examinations for English language and mathematics (related to literacy and numeracy) are approximately 10% apart (Assessment and Qualifications Alliance (AQA), 2012). Thus, for biological interpretations, differences in the reading and transformed maths scores of 5% or more (i.e. close to 10%) should be deemed as being biologically significant, and possibly equivalent to an increase in grade in those subject areas. We can see that the differences we identified earlier are all biologically significant apart from that between children with left WH and right SH and those that had left WH and SH.

It can also be seen that there was no discernable difference in reading or transformed maths scores between the two consistent WH and SH groups (one left-handed and one right-handed) with $p = 0.2$ for both comparisons. These results were consistent with what can also be seen in the overall combination of the two cognitive ability scores (the multivariate outcome).

In a study of 4,942 4- and 5-year old children Johnston *et al.*, (2009) showed that there were no differences in skills linked to language (such as in vocabulary and writing test scores) between left- and right-handers which agrees with our result for reading scores. However, the same study also claims that left-handed children did appear to perform worse on non-verbal tests from the Australian Council for Education Research's 'Who Am I?' test than right-handed children. Other studies also claim to have noted significant differences between left- and right-handed individuals including that of Resch *et al.*, (1997) who found those people aged between 16 and 30-years old (from a dataset comprising of 545 individuals) that were left-handed obtained lower scores in both Cattell's Culture Fair intelligence test and a spelling test than right-handers. This overall left-handed cognitive disadvantage is still debatable (e.g. see McManus, (2002) offering the alternative view that left-handed individuals have a cognitive advantage over right-handers).

We found that there was a quadratic relationship between relative hand skill and the multivariate outcome such that there was an optimal level of relative hand skill which was linked to the highest combination of the two outcomes. Since the models based on the marginal distributions of the multivariate model for both outcomes included the square of relative hand skill ($p < 0.001$), it suggested that there is a quadratic relationship between both reading and transformed maths scores and relative hand skill. This implies that there was an optimum value of relative hand skill which was linked to the highest average score in the reading and maths tests, as well as for the multivariate outcome. In both tests, children with relative hand skill values increasingly less or greater than this optimum level of relative hand skill had greater disparity in average test score (lower scores) to those linked with the optimum relative hand skill. We found the optimum value of relative hand skill to be 22, 25 and 30 for both left- and right-handed children in both the reading, transformed maths and combined scores, respectively (i.e. the children with the highest reading, transformed maths and combined scores had ticked 57%, 67% and 86% more boxes with one hand than the other, respectively).

The optimum relative hand skill values that we obtained from our model tally up with the results in Leask & Crow (2006), where it was stated that a possible optimum relative hand skill for right-handed subjects was 20 and 30 for left-handed subjects irrespective of other variables, when using the definition of relative hand skill from this thesis. A quadratic relationship between reading scores and the absolute difference in hand skill of the left and right hands of subjects had also been reported by Annett & Manning (1990). Our study has shown that this quadratic relationship exists even after dividing the difference in left- and right-hand skill by the total hand skill in both hands. Overall cognitive ability was also reported to have a quadratic association with hand skill in Nicholls *et al.*, (2010).

Another interesting point of note is that three different ways of recording hand skill were used in the three previously highlighted studies. In the study by Leask & Crow, (2006) as well as in this thesis a timed box-ticking task was used to record hand skill, with children having to tick as many boxes as they could in a minute firstly with their right hand, and then with their left hand. Annett & Manning (1990), used the modified Purdue pegboard test as described in Annett (2002). The Purdue pegboard test consists of a subject having to move a set number of pegs around a board with each hand in turn. In this case, hand skill is measured in seconds with the time taken to complete the movement of the set number of pegs being measured. In Nicholls *et* *al.*, (2010), hand skill was measured using a circle tapping task. This task comprised a subject having to tap inside a circle as many times as they could in 30 seconds. Therefore, even with hand skill being measured in three different ways (tapping, ticking and moving pegs) a quadratic relationship was still present between cognitive ability scores and hand skill.

We also found that there was insufficient evidence to suggest that the coefficient of gender was non-zero for either the reading or maths test scores ($p = 0.8$, and $p = 0.1$, respectively). This implies that there was no statistically significant difference in either reading or maths test scores between 11-year old boys and girls. Our result contradicts the studies by Halpern *et al.*, (1998) and Thilers *et al.*, (2007). Analyses of Medical College Admission Test (MCAT) scores from 152,669 adults aged between 19 and 40 years old suggested that males outperformed females in both the biological and physical sciences components of the MCAT (Halpern *et al.*, 1998). However, similarly to our study there was little discernable difference found between males and females in the verbal reasoning or writing components of the MCAT. An older group of adults (2,802 aged between 35 and 90 years old) were analysed and females had significantly higher episodic memory and verbal fluency test marks, on average, than males (Thilers *et al.*, 2007). Notice that the studies in Halpern *et al.*, (1998) and Thilers *et al.*, (2007) were on datasets comprised of adults aged over 19 years old, whereas in this thesis the analyses were based on 11-year old children.

The intelligence of individuals, as measured by IQ, increases until about they reach their peak intelligence at an age between 18 and 21 years old, on average, which is followed by a gradual decline as they grow older (McArdle *et al.*, 2002). The age at which this peak intelligence occurs differs from person to person. Hence, even if no difference is found between children as they are still developing and gaining intelligence, the differences in peak intelligence between individuals could explain how gender differences of cognitive ability then occur in adult samples. Furthermore, the Flynn effect shows that from generation to generation there are IQ gains ranging from 5 to 25 points (Kanaya *et al.*, 2003). The Flynn effect is taken into account by the normalising of IQ test scores every 20-30 years, but this could still mean that there are generational differences with the average rise per year of IQ, obtained using the Wechsler Adult Intelligence Scale, being equal to 0.3 points per year. This effect could also add to some of the differences between the groups of adults which include individuals of varying ages (i.e. from different generations). This would not be the case for the NCDS dataset as it only included children of the same age. Differences could also be due to the type and level of test given. For example, the MCAT is given to prospective medical school students as an entrance examination and therefore set to challenge the most gifted individuals whereas the maths and reading tests given to children in the NCDS were far more general and set to encompass 11-year olds of all abilities.

The final explanatory variable that we found to be statistically significant was the 4-level factor, UK region. This variable had Northern England and the Midlands set as the reference region with the other regions being Southern England, Scotland and Wales. In the reading scores, children from Southern England outperformed those children from the rest of the UK mainland (Northern England, the Midlands, Wales and Scotland) by 2% (4% of the mean score) on average ($p < 0.001$, $p = 0.003$, and $p = 0.001$, respectively). Also, Southern England's children obtained higher scores than subjects from Northern England and the Midlands in the mathematics test by 1% (2% of the mean score) on average ($p = 0.02$). Considering both outcomes jointly, the result of children from Southern England outperforming children who were educated in Northern England and the Midlands is repeated ($p < 0.001$). One speculative reason for this finding is the perceived north-south divide in living standards and income. People living and working in the south of England are thought of as having higher incomes and a higher standard of living than their equivalents in the north of England (Blackaby & Manning, 1990). It is plausible that these differences in incomes and standards of living has been passed on in the form of education with people in the south having a greater amount of disposable income, and therefore able to pay for better quality schools, than those in the north.

Furthermore, children from Scotland also achieved higher marks, on average, in the maths test than children that attended schools in Northern England and the Midlands by 3% (4% of the mean score) on average ($p<0.001$). Differences existed between the education systems of England and Scotland in the 1960's at all levels of education (i.e. primary, secondary, higher) (Paterson & Iannelli, 2007). Whitfield, (1970-71), stated that not only did the organisation of schooling differ between Scotland and the rest of the UK, but so did the school curriculums and those who set them. These contrasting educational systems between England and Scotland could speculatively explain our result with regard to the transformed mathematics scores. However, both the differences observed in the reading and transformed maths scores with regard to UK region were not biologically significant (all differences associated with UK region were 4% or less).

This concluded the analyses on children but we wanted to know whether further links could be identified between cognitive ability scores and other variables such as musical ability and age. These factors, along with gender, were included in the analysis of the musician-control dataset in *Chapter 4* and the results are summarised in the next subsection.

7.2.3 Associations between musical ability and cognitive function in adults

Following on from the results of the previous investigation we wanted to examine the associations in adults between cognitive ability, musical ability, age and gender. In *Chapter 4*, we conducted a multivariate linear model analysis of the musician-control (MC) dataset which was introduced in *Chapter 1*. This data consisted of a group of musicians from the Royal Liverpool Philharmonic Orchestra along with a group of non-musicians (controls), including males and females of varying ages. Cognitive ability was represented by vocabulary, arithmetic and visuospatial scores from the Wechsler Adult Intelligence Scale (WAIS). Explanatory variables considered were gender, age and *musician* (a binary variable was used to denote whether a person was a musician or a non-musician and henceforth *musician* is highlighted in italics to differentiate the variable from an individual musician). More information about these outcome and explanatory variables can be viewed in *Chapter 1*. The associations between the outcomes and the explanatory variables were assessed by performing multivariate linear model (MLM) analyses.

For the multivariate outcome with all three cognitive test scores combined it was seen that the only two statistically significant explanatory variable coefficients were *musician* and age. Across all of the cognitive ability scores when taken together, musicians outperformed non-musicians irrespective of gender ($p < 0.001$). Also we found that there was a linear relationship between the cognitive ability scores and age with older people in the study associated with lower marks, on average, than younger people ($p < 0.001$). To further understand the results in terms of individual cognitive abilities, the models based on the marginal distributions of the MLM for each of the three cognitive ability scores were examined with the results described below. We had three outcomes and therefore needed to use the Bonferroni correction to account for multiple comparisons such that the significance level α was reduced from 0.05 to 0.017.

The most interesting result from the models based on the marginal distributions of the multivariate model revolved around the associations between the three sets of test scores, gender and musical ability. This is due to the interaction term between musical ability and gender being present in the model. Even though this interaction term was not statistically significant for the multivariate outcome, or for the models based on the marginal distributions of the individual outcomes, we included this in the model because it allowed us to construct hypothesis tests for cognitive ability score differences between male and female, musicians and non-musicians. The effect of the interaction term alongside *musician* and gender singularly can be seen in the interpretation of the results breakdown. Firstly, for the vocabulary test scores, we saw that musicians obtained an average of 5 marks (25%) more than male controls, as well as an average of 4 marks (20%) more than female controls irrespective of the gender of the musicians ($p < 0.001$ for all comparisons). In terms of comparisons to the mean score (11.5), these differences were around 41%-42% between musicians and male controls, and around 33%-34% between musicians and female controls. The percentage differences highlighted in this section should be used to give an idea of the effect size, but since the scores have been transformed using a non-linear transformation from the raw WAIS scores, they are not associated in terms of a linear relationship with the raw scores. In terms of biological significance, a difference of 5% or more should be considered as significant. Since the three sets of scores are sub-scores from an IQ assessment test, then these scores can again be linked to academic ability and therefore the comment mentioned earlier about grade boundaries being around 10% can justify our choice of biological significance level of 5% or more. Our results tend to agree with previous studies which have found that musicians have superior verbal memory, including the delayed recall of words (Franklin *et al.*, 2008; Jakobsen *et al.*, 2008).

In addition to the comparisons between musicians and non-musicians, we also considered differences between males and females in both the musician and control groups. The models based on the marginal distributions of the MLM suggested that female controls attained 1 extra vocabulary test mark (5% of the maximum score, 10% of the mean score) than male controls, on average, out of 19 marks ($p = 0.01$). Two other studies suggested that females have better scores in verbal tests, which support the result derived from the MC dataset analyses where females performed better than males in the vocabulary test (Halpern *et al.*, 1998; Thilers *et al.*, 2007). However, no difference was found between male musicians and female musicians in the vocabulary scores ($p = 0.9$). After considering the model interpretation that musicians outperformed controls as well as the gender differences then it therefore appears to be that the discrepancy between male and female controls' vocabulary scores is eliminated by gaining the ability to play a musical instrument.

Our multivariate model also showed that female musicians were associated with higher visuospatial scores than both female and male controls, by 3 and 2 marks (14% and 11%) on average, out of a possible 19 marks ($p < 0.001$ and $p = 0.002$, respectively). It was also suggested that male musicians achieve, on average, 2 marks (10%) more than female controls and 1 mark (7%) more than male controls on the visuospatial test ($p < 0.001$ and $p = 0.004$, respectively). These differences are even more noticeable when compared to the mean score of 10.9, with the differences ranging from 12% between male musicians and male controls, up to 25% between female musicians and male controls. Similarly to this result, the study of Sluming *et al.*, (2007) reported an increased level of visuospatial ability in musicians over non-musicians, from both the Benton judgement of line orientation test and a three-dimensional mental rotation task.

In contrast to the gender differences in the vocabulary scores, in the visuospatial test we noticed that there was no difference between the marks of males and females in either the musician or control groups ($p = 0.2$ and $p = 0.08$, respectively). The lack of a gender difference in the visuospatial results appears to contradict the overall results from the meta-analysis in Voyer *et al.*, (1995) which stated that males outperformed females in the a variety of different types of spatial ability tests. However, differences between males and females in the WAIS block design test (i.e. the visuospatial test used in our dataset) were only found in a few age groups, and not across the entire range of ages (Voyer *et al.*, 1995).

In terms of the arithmetic marks, our results suggested that male musicians attain a higher number of correct answers than both male and female non-musicians (controls) by 1 and 3 marks out of a maximum of 17 marks or 8% and 16%, ($p < 0.001$ and $p = 0.01$), respectively. However, no difference was found between the arithmetic scores of female musicians and both male and female controls ($p = 0.4$ and $p = 0.02$, respectively). When comparing males against females in the musician and control groups separately from each other, it was noted that there was a significant gender difference found in the control group only, with males linked to a 1 mark (8%) higher arithmetic score than females ($p = 0.005$ and $p = 0.4$ for the control and musician groups, respectively). The mean arithmetic score was 10.4 marks, so these differences are even more prominent with the lowest difference between male and female controls of 12.5%. These three statistically significant results can therefore be regarded as being biologically significant. The gender difference in the sample group of non-musicians is in line with the reported conclusions from Halpern *et al.*, (1998) which investigated the difference in mathematical ability tests between males and females. An interesting point of note is that this gender difference between males and females, similarly to the vocabulary results, appears to become negligible when a person has the ability to play a musical instrument. Indeed this improvement in mathematical ability can also be seen in that there was no significant difference between female musicians' and male non-musicians' arithmetic marks.

Furthermore, even though the gender comparison in the visuospatial scores of the musician and control group were both non-significant, Figure 4.1, showed there not only appeared to be a slight male advantage in the visuospatial scores of the control group, but a slight female advantage in the visuospatial scores of the musician group. We did not consider the type of instrument that musicians played. However, this partial reversal of visuospatial skills by gender may be linked to the type of instrument.

Age was statistically significant for the arithmetic and visuospatial scores such that in each case, older subjects had, on average, lower scores than younger subjects. The relationship between age and both arithmetic and visuospatial abilities was linear. The models based on the marginal distributions from our MLM suggested that if two subjects had a difference in age of 10 years, then the older person would have had 1 and 0.4 marks lower (5% and 2% of the maximum score; 9% and 4% of the mean score), on average, out of 19 and 17 marks, respectively, than the younger person on the visuospatial and arithmetic ability assessments ($p < 0.001$ and $p = 0.002$, respectively). The association of age with visuospatial scores may be biologically significant, but the association of age with arithmetic scores does not look to be so. This is similar to the results of previous studies which have reported that older subjects obtain lower scores on cognitive ability tests than younger subjects (Gunstad *et al.*, 2006; Salthouse, 2006; Mitnitski & Rockwood, 2008; Mitnitski *et al.*, 2010). However, no difference was found in the vocabulary scores which was related to age ($p = 0.4$). Salthouse, (2006) stated that there was no reduction in verbal ability between younger and older subjects, from a group with ages in years ranging from the mid-20's to the mid-80's, which agrees with our findings.

7.2.4 Associations between relative Broca's volume and cognitive function

Due to previous studies linking Broca's area with cognitive ability and following the results of the previous subsection we wanted to include some measurement estimates corresponding to Broca's area in a similar type of analysis. Therefore, by using the estimation methods described in *Chapter 2*, we obtained estimates for the surface area and volume of Broca's area for 39 subjects comprising 10 male musicians, 10 male controls, 10 female controls and 9 female musicians. Information from these individuals came from the MC dataset as defined in *Chapter 1* with the same explanatory variables as those involved in the analyses of the previous subsection in addition to the three cognitive ability scores (vocabulary, arithmetic and visuospatial). The first objective of this part of the study was to identify possible associations between the volume estimates of Broca's area relative to the total brain volume (relative Broca's volume) and the other factors (cognitive ability, musical ability, gender and age). We also wanted to take into account that Broca's area may have different functions in the left and right hemispheres and so they were considered separately. Therefore, in our analyses the left and right relative Broca's volume estimates were set as the outcome variables in a multivariate linear model.

We saw that when both left and right relative Broca's volume estimates were considered simultaneously as a multivariate outcome, gender was not significant ($p = 0.06$). Thus, our interpretation was that there was insufficient evidence to suggest that there was an overall difference in relative Broca's volume between males and females in total relative Broca's area across both hemispheres combined. Even considering the two hemisphere estimates individually from the models based on the marginal distributions, the coefficient of gender was not biologically relevant for either left or right relative Broca's volume at the Bonferroni-adjusted 2.5% level of significance ($p = 0.3$ and $p = 0.04$, respectively). However, if we considered statistical significance to be at the Bonferroni-corrected 5% level of significance, then there would have been evidence to suggest that the coefficient of gender in the right hemisphere was non-zero. The significant gender coefficient would have shown us that females were associated with a smaller relative Broca's volume in the right hemisphere than males by 0.2% on average. A difference in relative Broca's volume of 0.2% translates as being equal to a difference in actual Broca's area volume of around 17% due to the mean Broca's volume in the right hemisphere being equivalent to approximately 1.2% of the total brain volume.

The study of Amunts *et al.*, (1999) investigated the left and right hemisphere Pars Opercularis (PO) and Pars Triangularis (PT) volumes independently. In Section 1.2.5 we stated that Broca's area is comprised of the union of PO and PT. No differences were found in the volume of either PO or PT in either left or right hemispheres, between males and females (Amunts *et al.*, 1999). However other studies have found gender differences in Broca's area volume (e.g. Harasty *et al.*, (1997) found that females had larger inferior frontal gyrus (IFG) volume (of which Broca's area is a part of) than males). Gender disparity in the volume asymmetry of PO and PT across the hemispheres was reported in Uylings *et al.*, (1999) with males associated with larger asymmetry (in particular, larger volumes in the left than in the right hemisphere). These results show that there is much debate in this area, and this may in part be due to the small sample sizes in each study (10, 11 and 21 participants in Amunts *et al.*, (1999), Uylings *et al.*, (1999) and Harasty *et al.*, (1997), respectively).

In Section 5.6.1 we estimated the within observer coefficient of error (CE) for each sub-structure of Broca's area (PO and PT) in each hemisphere. Our method for obtaining the estimate involved making ten observations of the brains of two individuals. We found that the CEs for left and right PT were greater than those for left and right PO (3.0% and 4.3% for left and right PT compared to 2.2% and 2.8% for left and right PO, respectively). When comparing the within observer CE with the CEs due to Cavalieri sectioning and point counting, it was seen that the within observer CE was greater than the other CEs combined. This suggests that there was a greater error in the variability of the Broca's volume estimates than the error due to elements of the estimation method itself.

7.2.5 Associations between relative Broca's surface area and cognitive function

Using the same participants as in the previous subsection (39 subjects consisting of 20 non-musicians and 19 musicians) we wanted to investigate the associations between surface area estimates of Broca's area relative to Broca's volume (relative Broca's surface area) and the other factors identified previously (vocabulary, arithmetic and visuospatial scores, *musician*, gender and age). We adjusted the surface area estimates by Broca's volume estimates because it was the surface convolution that was of more interest than the total surface area. In our multivariate linear model analyses, the two variables set as the outcomes were the relative Broca's surface area estimates for the left and right hemispheres, with the three cognitive ability scores, *musician*, age and gender fixed as the explanatory variables.

When identifying associations across both hemispheres simultaneously the multivariate outcome was examined and it was found that musicians had a smaller relative Broca's surface area compared to controls across both hemispheres with musicians having a less convoluted surface area relative to volume than non-musicians ($p = 0.02$). However, this was not the case for gender which did not have a significant effect when both hemispheres were considered jointly ($p = 0.06$).

We examined the models based on the marginal distributions for the relative Broca's surface area estimates of each hemisphere separately to identify associations between the outcomes and explanatory variables. The critical significance level (α) was Bonferroni-corrected because we had two outcomes in the model simultaneously. The difference in the convolution of Broca's surface area that we saw across both hemispheres together applied for the right hemisphere only. That is, musicians have a less convoluted surface area of Broca's area relative to Broca's volume than controls, in just the right hemisphere ($p = 0.02$). Indeed, this difference was seen to be around 12.5% of the mean of the right hemisphere relative Broca's surface area estimates. Two explanatory coefficients were statistically significant only at the Bonferroni-adjusted 10% level of significance: gender and *musician*, both in the left hemisphere, and both at around 10.5% of the mean left hemisphere relative Broca's surface area estimates ($p = 0.03$ and $p = 0.04$, respectively). We found that both females and musicians were weakly linked to smaller relative Broca's surface area than males and non-musicians in the left hemisphere.

There is not a wealth of information about associations with the surface area of Broca's area. However, a recent study by Frye *et al.*, (2010) investigated links between reading-related skills and the surface area of a few individual regions of the brain including the inferior frontal gyrus (IFG) in two groups of participants (one group with dyslexia and the other without). In the group without dyslexia, the results suggested that there was a negative relationship between a sublexical decoding score (a measurement of how well a participant can break down words into individual sounds) and surface area of the IFG (i.e. individuals with higher scores on the sublexical decoding test were associated with smaller IFG surface area). However, this association with IFG surface area was not consistent across all test scores. For example, no links with IFG surface area were found for the 'alternative phonological awareness' (a test of the ability to perceive and manipulate sounds which create spoken words) scores in the group of participants without dyslexia.

Therefore, we can see that there was a link between some language related ability scores and smaller surface area in the region which encapsulates Broca's area. We have seen that greater cognitive abilities, including language skills, have been linked with having the ability to play a musical instrument (Sluming *et al.*, 2007, Franklin *et al.*, 2008, Jakobsen *et al.*, 2008). Also, females have been linked with having higher verbal test scores in Thilers *et al.*, (2007) along with the left hemisphere being more strongly linked with language skills (Schaffler *et al.*, 1993). The combination of musicians and females being linked to better language skills than non-musicians and males may explain the associations between these explanatory variables and the lower relative Broca's surface area estimates.

The within observer CE for each sub-structure of Broca's area (PO and PT) in each hemisphere was estimated in Section 5.6.2. The method for estimating the within observer CE for Broca's surface area, similarly to the method for estimating within observer CE for Broca's volume, involved ten observations being made of two individuals' brains. We saw that the within observer coefficient of error was observed to be 2.1% and 3.4% for the left and right hemisphere PT estimates. The estimates for within observer CE for left and right PO were not possible to obtain due to the coefficient of error due to cycloid sampling being overestimated by the Poisson model (see Cruz-Orive & Gual-Arnau, 2002, for examples). We concluded that further work is necessary in the estimation method of the coefficient of error of the surface area estimator due to cycloid sampling such that less conservative and more reliable estimates can be obtained in future studies.

7.2.6 The effect of missing data

The final objective of this thesis was to investigate the effect of missing data on the analyses that we had performed in the previous chapters to determine whether the results obtained can be relied upon to give a reasonably accurate description of the previously stated associations. The National Child Development Study (NCDS) dataset which had been used to analyse the links between cognitive ability and a range of factors including handedness in children (see *Chapter 3*) included a substantial amount of missing data. The number of children in the study with at least one missing outcome was 4,435 (23.9%). When the missing covariates are also taken into account then the number of children with at least one missing covariate or outcome was 6,712 which is equal to 36.2% of the total number of children. In the musician-control (MC) dataset from *Chapter 4*'s analyses the number of adults in the study with at least one missing outcome was equal to 20 which can be expressed as approximately 13.9% of the total number of subjects. This is equal to the number with missing outcomes and covariates combined.

For the original multivariate linear and linear mixed models in *Chapters 3* and *4* to be unbiased, we had to assume that the missing outcomes from both datasets were missing completely at random (MCAR). MCAR means that there are no underlying associations between any of the variables in the dataset, any unobservable parameters and the missingness of the outcome variables. We could not classify the missing outcomes as MCAR if those children or adults with missing outcomes shared some characteristics which were more common than the characteristics of those children or adults with non-missing outcomes.

After having studied the datasets it was decided that although there do not appear to be any confounding reasons as to why the subjects have missing outcomes, we cannot guarantee that they are MCAR. Methods for accounting for missing at random (MAR) data in multivariate analyses were discussed and described in *Chapter 6*. We decided to use the inverse probability weighting method to adjust the multivariate models to account for individuals with missing outcomes. This method involved the calculation of a weight value for each person which was obtained from a logarithmic model. These weights were then incorporated into the multivariate linear and linear mixed models in a similar approach to those previously performed.

The results for the weighted multivariate linear mixed model fitted to the NCDS dataset were very similar to the results we obtained in *Chapter 3* with changes to the coefficients of a approximate maximum of 2%. So similar were the results that the overall interpretations from the multivariate models were the same. Our weighted model therefore suggests that the effect of accounting for those children with missing outcomes was negligible and therefore that the missing outcomes were similar to what would be predicted given the children with complete data. We checked this by constructing hypothesis tests between the missing and non-missing outcome groups of children for each explanatory variable. Thus, using chi-squared tests and Mann-Whitney tests we were able to assess that there were no differences between the two groups in any of the explanatory variables except UK region. However, when we examined the numbers in each region by missing and non-missing outcomes, it was apparent that this difference wasn't large. Indeed, the statistical significance may be down to the large sample size and the chi-square test being over-sensitive. This added weight to the argument that the missing outcomes were MCAR and that the original multivariate analysis was reasonable.

The only suitable logarithmic model for the MC dataset included the intercept only. This meant that each adult in the study had a constant weight. Therefore, the weighted multivariate linear model fitted to the MC dataset was identical to that in *Chapter 4*. We again constructed hypothesis tests to test for possible differences between the missing and non-missing outcome groups of participants in terms of the explanatory variables. We saw that the only hypothesis test that gave a statistically significant result was the Fisher exact test for *musician*. The explanation for this result was that there were no musicians with missing outcomes. However, the number of non-musicians with missing outcomes totalled approximately 17% of the non-musicians with non-missing outcomes. Hence, due to these observations and the relatively small number of musicians compared to non-musicians, we explained that the strongly significant p-values for the coefficients of musician in the MLM would not have changed enough to alter the overall interpretations. This combination of factors all suggested that the missing outcomes can be assumed to be MCAR. We can conclude therefore that, in both cases, the missing outcomes have very little effect on the overall message that was obtained from the non-missing data.

7.3 Limitations

The limitations of the methodology that were used in this thesis can be associated with the two different types of methodologies used (i.e. statistical and stereological). The statistical limitations will be described first, followed by the stereological limitations.

7.3.1 Statistical limitations

One of the most basic limitations of studies such as the ones undertaken in this thesis is if the number of data is small. The NCDS dataset contained 11,843 children after outliers and those with missing data were omitted, which is still a very large number. However, the MC dataset, used in *Chapter 4*'s analyses, after those with missing covariates and outcomes were excluded comprised just 124 adults. The analyses based on the MC dataset would have been relatively straightforward to implement on a larger dataset. However, there are issues surrounding obtaining information on career musicians, since they are quite a small population of individuals. Although, we do not believe that a similar number of individuals are needed as in the NCDS dataset, a dataset of several hundred individuals with around 50% musicians and 50% non-musicians would have been preferable.

The lack of data really becomes quite an important issue for the dataset that includes the Broca's volume and Broca's surface area estimates. This dataset contains information about just 39 adults. As a comparison, the datasets of ten recently published studies based on stereological estimates of human organs ranged from just 5 up to 77 subjects (Acer *et al.*, 2010; Mechlenburg *et al.*, 2010; Abdul-Kareem *et al.*, 2011; Cadnapaphornchai *et al.*, 2011; Ertekin *et al.*, 2011; Hallahan *et al.*, 2011; Keller *et al.*, 2011; Mazonakis *et al.*, 2011; Mignani *et al.*, 2011; Powell *et al.*, 2012). One of the main reasons for studies involving stereological estimates having such small numbers of participants is the time consuming nature of both obtaining the images (from MRI techniques) and performing the stereological techniques. Ideally, we would like to see datasets including several hundred individuals, as mentioned for the MC dataset, but with the limitations specified for stereological studies, this would be relatively difficult to obtain given current technology and techniques.

The power of a test is defined as the probability of a type II error not occurring (i.e. the probability that a given null hypothesis is rejected when the null hypothesis is false). The size of a dataset is one of the factors that effects power with smaller datasets corresponding to smaller power values. As the power is reduced so the likelihood of a type II error increases (i.e. the probability of the null hypothesis not being rejected even though the null hypothesis is false). In our case when the dataset consists of 39 individuals, this implies that it is more likely that associations between Broca's volume, surface area and the other factors are not identified as being statistically significant when they possibly should be. Indeed, only relatively few explanatory variables had coefficients which were significant in the analyses with Broca's area measurement estimates included. It can be seen that as more covariates are considered and more complex statistical models are fitted to the data then associations which exist may not be fully recognised and identified by the model.

Further limitations occurred due to memory limitations of individual statistical programs. We used a variety of different statistical programs including R, MLwiN and SAS. The most flexible one of these is R and aside from being free to download, people are constantly developing new functions to perform different types of analysis. However, for the NCDS dataset we found that when the data is large, and as univariate linear (and linear mixed) models become more complex, more RAM is needed. Under Microsoft Windows, R can use a maximum of 2GB of RAM and despite our best efforts to increase this we could not. This 2GB limit became an artificial ceiling, in that we could only fit models up to a certain level of complexity before the limit was reached. A further problem with R was that it was not intuitive to fit a multivariate model without spending much time consuming effort on restructuring the dataset with dummy variables for each outcome. This meant that other programs had to be considered for fitting multivariate models (i.e. MLwiN and SAS).

Additionally, the program that we used to conduct the multivariate analyses, MLwiN, uses Wald tests to determine whether explanatory coefficients for the multivariate outcome are zero or non-zero. There are other possible tests which could have been used including the Hotelling's T^2 test. Hotelling's T^2 test has been suggested to outperform the Wald test (Poh, 1984). However, the Hotelling's T^2 test was not available as an option in MLwiN.

As it has previously been mentioned, MCAR could not be guaranteed for either the NCDS or MC datasets. Therefore, the missing outcomes should be taken into effect in some way. We used inverse probability weighting as the method of adjusting the multivariate analyses so that the missing outcomes were accounted for. In *Chapter 6*, multiple imputation was mentioned as an alternative method to inverse probability weighting. However, due to time constraints and the length of time needed for the implementation of MI when analysing large datasets with large amounts of missing data, it was deemed unsuitable to be included in this thesis. Yet, MI has been stated to be efficient even when there is a relatively high proportion (>30%) of missing data in the dataset (Schafer & Olsen, 1998). In the case of the NCDS there was a relatively high proportion of missing data (36%) and hence, this method would have also been appropriate.

A final limitation to the analyses from a statistical point of view is that only associations involving variables in the datasets can be investigated. Links between other variables which may be interesting cannot be examined if data were not collected and included in the datasets. One particular point of discussion was the measurement of relative hand skill in the NCDS. This was based on the box-ticking method as introduced in *Chapter 1*, which consisted of each subject trying to tick as many boxes in one minute with each hand (Leask & Crow, 2006). The main result of the analyses conducted in *Chapter 3*, involved subjects writing hand (WH) and superior hand (SH) (i.e. the hand which had the most skill as measured by the box-ticking method). Since box-ticking involves something akin to writing, there could be some confounding variability between these two variables (i.e. they may not be independent) and therefore, another method of estimating hand skill should be used in the future.

Suggestions have included the modified Purdue's pegboard test (as seen in Annett, 2002). In the pegboard test, individuals are asked to move a number of pegs on a board, with each hand, and they are timed in doing so. This test therefore relies on hand motor skills, but it is not directly related to writing skills. The Edinburgh test, as described by Oldfield, (1971) was also discussed. In this test a number of different tasks are listed, such as brushing teeth or combing hair, and subjects in the study are asked to identify whether they use their right hand, left hand or both hands to perform the task. However, the measure of handedness obtained from the Edinburgh test could be deemed to be discrete as there are only a limited number of values that it can take (with just three possible outcomes per task). Certainly, another measure of handedness would have been very useful in the NCDS dataset to be able to compare the results with the box-ticking measurement. Yet, the data collection for the NCDS data used in the analyses in this thesis occurred in 1969, and additional methods for assessing handedness were not considered necessary at the time. Thus, it is unfortunately not possible to incorporate a second measurement of handedness. A couple of reasons why the previous studies of the effects of handedness on academic ability mentioned in *Chapter 1* gave conflicting results could be that related to different measurements of handedness and academic ability along with differences in sample sizes in each study. A limitation of the MC dataset analyses in *Chapter 4* was that handedness could not be included as a factor due there being only 1 left-handed non-musician participant. Subsequently the Broca's area dataset analyses (*Chapter 5*) also did not include handedness as a variable because it was a subset of the MC dataset.

An additional variable that was collected but not included in the MC and Broca's area analyses was related to the years of experience playing an instrument (for the musicians). However, this was not included in the analyses due to the increasing complexity involved (an interaction term with the musical ability variable would have had to be included) given the reasonably small sample size. Another variable which was collected and not considered for this study was the type of instrument that a subject plays. Verbal and non-verbal tests were administered to the children in the National Child Development Study and although their scores were recorded, they were not included in this study. Similarly, each child's paternal social class, as defined by the General Registry Office (GRO) classification of occupations, was included in the NCDS dataset but not included in our study.

Other variables which were not considered in any of the analyses but which we would have liked to have examined included socio-economic status and a measure of IQ obtained using for example the Differential Ability Scales (DAS) with the various sub-tests used as multiple cognitive ability tests.

7.3.2 Stereological and imaging limitations

The limitations identified in terms of the stereological methods used in this thesis revolved around the estimation of the coefficient of error of the surface area estimator due to the systematic sampling on a semi-circle and the cycloid sampling. In the worked example in Chapter 2, we estimated that the contribution to the total coefficient of error of the surface area estimator due to systematic sampling on a semi-circle was 0%. It is highly unlikely for a level of sampling to provide 0% contribution to the total coefficient of error. Hence, this leads us to the conclusion that the method of obtaining the coefficients of error for the surface area estimator is incorrect and therefore requires further work to resolve this problem.

Our intra-rater study for the surface area estimates brought a further issue with the coefficients of error for the surface area estimator, to the forefront. We found that for Pars Opercularis, the coefficient of error of the surface area estimator due to cycloid sampling was greater than the total coefficient of error of the surface area estimator. Examples in Cruz-Orive & Gual-Arnau, (2002) suggest that the Poisson method used to estimate the cycloid grid sampling coefficient of error gives conservative values. Therefore, this overestimation implies that we could not acquire a reasonable estimate of the within observer coefficient of error.

The reliability and quality of the volume and surface area estimates themselves were limited by imaging techniques. The intra-rater study carried out on the Broca's volume estimates and documented in *Chapter 5*, showed that the estimated contribution to the error in the volume estimates due to the observer was greater than the contribution due to the estimation process. This was seen by the coefficient of error of the volume estimator due to within observer variability being larger in value than both the estimated coefficient of error of the volume estimator due to Cavalieri sectioning and point counting combined. To improve the estimates, both the within observer error and reliability of the estimates should be attempted to be reduced. However, a number of factors contributed to the overall error for both the volume and surface area estimates:

- Resolution of the magnetic resonance (MR) images
- Demarcation of the region of interest (RoI)
- Level of expertise of the observer
- Resolution of the monitor
- Observer error from the manual process

Firstly, the resolution of the MR images is correlated with both the length of time a subject is scanned for and how powerful the magnetic resonance imaging scanner is. In our case, the data came from a 1.5 Tesla scanner and live individuals. Obtaining MR images from a live subject means that the scanning process has to be relatively short (20-30 minutes) due to the exposure of the person to the magnetism in the MRI scanner. On the other hand, post-mortem studies have been known to run scanning processes for hours at a time, producing much better resolution of images. The resolution of MR images of brains from live individuals can also be affected by any movement of the head during the scanning process. This obviously does not affect the MR imaging scans of post-mortem brains. A more powerful 3 Tesla MR imaging scanner would have produced higher resolution images.

The demarcation of the RoI is crucial in obtaining reasonable estimates for both volume and surface area. Any mistakes in the demarcation process can result in parts of different regions, not belonging to the RoI, being included or parts of the RoI being excluded from the volume and surface area. An interesting point which we noted that can cause huge fluctuations in the volume and surface area estimates either with the same or different observers was that the demarcation is not complete in all three dimensions for each estimation method. We demarcate the extremities of the region in both the coronal and transversal planes. However, it is not possible to demarcate the innermost edge of the region in the sagittal view because there are no landmarks to be able to say where our region ends and where the next begins. The decision of where the sagittal extremity is, depends on each observer. There will also be some within observer variability even in the positioning of the endpoint. To be able to demarcate the edge correctly we would need to demarcate the innermost edge of each Cavalieri section in the coronal plane. When we consider that each RoI (pars Opercularis and Pars Triangularis) contains an average of between 10 and 15 Cavalieri sections and each brain contains two of each, then the total number of sections to be demarcated varies between 40 and 60 for each individual. Performing these demarcations would be a far too time consuming process.

The level of expertise of the observer will obviously have an effect. An experienced anatomist or neurologist would be far better at identifying and demarcating a region of the brain than someone with little experience in this field. In addition to this the resolution of the PC monitor also has an effect on the estimates. Since the manual method of volume and surface area estimation relies on visual accuracy, then it is obvious that a monitor with a higher resolution which can display more precise images would be better to use for the estimation process using EasyMeasure. There will always be a level of observer error due to the fact that this is a manual process.

7.4 Further work

We can extend the work that we conducted in this thesis by addressing some of the limitations that we found. We discussed these limitations in the previous section but in this section we reveal how each limitation could be resolved in a future study, in addition to extending the investigation where possible. Firstly, although the NCDS dataset was large (18,558 children), the other datasets were reasonably small (n=149 and n=39 for the MC dataset and Broca's subset, respectively). For the reasons discussed in Section 7.3.1, it is advisable that in a future study we increase the power of our hypotheses tests by increasing the sample size. As mentioned in Section 7.3.1, ideally increasing the sample size of the MC and Broca's dataset to one including several hundred individuals with a 50%-50% split between musicians and controls (and a further 50%-50% split in each musician and control group between males and females). However, due to the number of professional musicians required, along with the time consuming nature of obtaining the stereological estimates, this may not be an easy task.

In *Chapter 1* we stated that there were conflicting results from different studies investigating the links between handedness and academic ability. In Section 7.3.1, we suggested that possible reasons for this could have been due to differences between the studies in terms of sample size, methods of measuring handedness, and methods of estimating academic ability. A way of exploring the heterogeneity of these results would be to conduct a systematic review and meta-analysis of all previous studies of handedness. The variability across all studies could then be compared to check if there were significant differences between the reported studies which could be influencing the results, such that it appears that they are giving conflicting conclusions.

We observed memory limitations when using R, as well as having time constraints in fitting the multivariate models. However, we were able to fit and compare results between MLwiN and SAS. In a future study, it would be interesting to fit a multivariate model in R because of the program's flexibility and to compare the results between the three programs. If we have large datasets (e.g. the NCDS dataset) then we need to increase the amount of RAM that is available for R to use. This is not possible on a Windows PC but it is possible to increase the RAM that R uses by running the program on a Linux PC.

We discussed in Section 7.3.1 that the Hotelling's T^2 test gives more reliable information about whether multivariate coefficients are zero or non-zero, than the Wald test. Since the Wald test is the only available option in MLwiN (the program we used to fit our multivariate models), then an extension of the study would be to fit the models in an alternative program, and use the Hotelling's T^2 test instead of the Wald test. However, the Wald test has been stated to work well if used in conjunction with large datasets (Poh, 1984). Therefore, how much the Hotelling's T^2

test is an improvement over the Wald test for our analyses in *Chapter 3* (the NCDS dataset including 11,843 children) is debatable.

We used inverse probability weighting as a method of accounting for individuals with missing outcomes in our multivariate analyses. However, an alternative method was discussed called multiple imputation (MI). We did not use MI due to problems surrounding the implementation of the multiple imputation by chained equations (MICE) program in R. We found that our multivariate linear and linear mixed models were not able to be fitted to the pooled multiply imputed dataset. Also, due to the high proportion of missing data (both outcomes and covariates), to obtain reliable, stable imputed values required high computer processing power which we did not have available. Running a MI program on the NCDS would have required days or weeks to run correctly. It would have been desirable to have used MI because if two or more missing data methods agree with each other results-wise then it is a reasonable indicator that the initial method is taking the missing data into account correctly. MI is efficient even when there is a relatively high proportion (>30%) of missing data in the dataset and in the NCDS dataset this was proportion of children with missing data was 36%. Therefore, as an extension to the work in this thesis, it would be useful to use MI to impute values for the missing data (both outcomes and covariates) to compare the results obtained with the IPW multivariate models in this thesis. Thus the reliability of our results for both the NCDS and musician-control datasets could be reviewed.

Further extensions to the analyses could involve a new dataset being compiled including both children and adults. This would mean that we could investigate associations between cognitive ability and a range of factors for both children and adults in the same model. This dataset would need to be large with a number of additional variables to those in our datasets including:

- A measure of handedness which does not depend on handwriting (e.g. Purdue's pegboard test)
- Years of experience of playing a musical instrument
- Type of musical instrument played

- Other cognitive abilities (e.g. verbal and non-verbal test scores)
- Socio-economic status
- IQ (including sub-test scores; e.g. DAS)
- Volume and surface area of Pars Opercularis (PO) and Pars Triangularis (PT) (i.e. the two components of Broca's area)
- Volume of grey matter and white matter for Broca's area, PO and PT

Of course, these variables are either not available in the current datasets, or are available but cannot be included for reasons as explained in Section 7.3.1 (e.g. we could not investigate handedness in the MC dataset because there were only five left-handed participants). We did not include the breakdowns of volume by PO and PT or by grey and white matter due to the fact that our dataset was not large enough for us to have enough power to perform multivariate analyses. The lack of power also meant that we also could not investigate associations involving the breakdowns of PO and PT surface area estimates. However, these could be rectified if we obtained a larger dataset.

A number of improvements could be made in a future study by obtaining new, more reliable estimates for Broca's volume and surface area. Possible ways to make the estimates more reliable were discussed in the previous section. Firstly, if we were to obtain a new, larger dataset, then we could obtain MR images from a more advanced scanner. Our images were obtained from a 1.5 Tesla scanner, but currently 3 Tesla scanners are available. A more powerful scanner would mean that the MR images would have a better resolution meaning that more precise estimation would be possible. The resolution of the monitor viewed when obtaining the volume and surface area estimates also has an effect on the quality of the estimates. When applying stereological methods to obtain new estimates, we should use a relatively new monitor (to avoid dust build-up) with a good resolution. Larger monitors would also allow for more precise point and intersection marking in EasyMeasure (the program we used to implement the stereological methods from *Chapter 2*).

To improve the reliability of the estimates in a future study, we will require a reduction in the overall variability. We identified that the demarcation of the RoI, in our case Broca's area, was a contributor to this variability. This is due to the border of the region with other sections of the brain (i.e. not bordered by sulci or by the outer brain surface) being very difficult to identify, as well as the demarcation itself being time consuming. As an extension to this study, it would be interesting to get an estimation of the contribution to the total variability due to demarcation. This could be achieved by having an experienced anatomist identify and demarcate Broca's area for each participant and then the observer from this study repeating the estimations from the newly demarcated images.

An alternative, automated method for obtaining volume and surface area estimates for individual regions of the brain can be found in BrainVoyager (a program we used to demarcate Broca's area). This method independently identifies the region of choice by standardizing each individual brain to a template. Measurement estimates (e.g. volume and surface area) are then given. In a new study it would be very interesting to compare estimates from the automated method with those obtained from the manual method using EasyMeasure.

7.5 Conclusions

The associations between handedness and a variety of cognitive abilities have been reported upon in a large number of studies. However, the literature has been inconsistent in describing whether left- or right-handers have a cognitive advantage. Other investigations have focussed on individual links between cognitive abilities and factors such as age, gender and musical ability. Furthermore, Broca's area is a region which has been historically connected to language skills in the left hemisphere, although skills linked to Broca's area in the right hemisphere have not been fully explained. The objectives expressed in this thesis revolved around using multivariate analyses to investigate associations between different cognitive abilities and factors such as handedness, musical ability, age, gender and Broca's area measurement estimates (i.e. volume and surface area) in both children and adults where possible. Multivariate analyses were used so that all associations were adjusted for other factors, including correlations between the multiple outcomes.

Finally, we wanted to check whether the individuals with missing outcomes had an effect on the results of these multivariate analyses.

One of the main, novel results of this thesis was that in *Chapter 3*, from the National Child Development Study data, we concluded that those children with inconsistent writing and superior hand (i.e. the children who wrote with the opposite hand to their superior hand) performed worse in both a reading and maths test, than those children with consistent writing and superior hand. This result was found to be true irrespective of whether the children were left-or right-handed. Handedness was measured using a box-ticking method such that the hand skill for a hand (left or right) of a child was recorded by the number of boxes the child ticked in one minute with that hand. Superior hand was then defined as the hand which ticked the most number of boxes. This result was biologically significant even taking into account the correlation between outcomes, the region of the UK in which the children attended school, the quadratic relationships between relative hand skill and each cognitive ability and the gender of each child.

The second novel result that we found was that gender differences that were present in vocabulary and arithmetic tests in a group of non-musicians (controls) aged between 19 and 94 years old, were not present in a group of musicians. The literature contains many references to the perception that females perform better in language tasks and males perform better in mathematical tasks, which are both reflected by our results for the control group. Musicians were associated with higher scores in a vocabulary, arithmetic and visuospatial test than non-musicians which again agree with the literature. However, the combination of these results suggests that either the ability of being able to play a musical instrument cancels the gender effect for the vocabulary and arithmetic test scores or there is some innate characteristic within musicians (e.g. linked to Broca's area in the brain) which must be present in individuals to allow them to attain both a high cognitive ability and musical ability. Indeed, although no differences between males and females was found for the visuospatial scores in either the musician or control groups, we can see in Figure 4.1 that the slight male advantage in controls becomes a slight female advantage in musicians. We also discussed whether or not this may be related to the type of instruments that musicians play as opposed to just the effect of being a musician.

Our missing data analyses concluded that the results from the multivariate analyses stated above were consistent after accounting for individuals with missing outcomes. This added weight to our novel results.

To our knowledge, the study involving the links between relative Broca's surface area (Broca's surface area relative to Broca's volume) and musical ability and gender have not been previously reported. In the literature at least one study has discussed associations between the surface area of the inferior frontal gyrus, of which Broca's area is a component, and reading-related test scores, but not specifically Broca's area itself. We found that musicians were associated with a less convoluted Broca's surface than non-musicians across both hemispheres combined, but mainly in the right hemisphere. However, we did not find any significant links between relative Broca's volume (Broca's volume relative to total brain volume) and musical ability, gender, age or cognitive ability (defined as vocabulary, arithmetic and visuospatial scores).

The overall conclusions from this thesis can be related directly to the objectives in *Chapter 1*. We constructed statistical models which accounted for multiple outcomes as both a combination and individually (multivariate linear model) and models which also took correlation between clustered data into account (multivariate linear mixed models). We showed that it was beneficial to fit multivariate models when we have multiple outcomes if either we know the outcomes are correlated or we do not know whether they are correlated. If outcomes are not correlated then there is no advantage to fitting multivariate models over multiple univariate models.

We saw that there a number of different factors which are associated with different cognitive abilities. Some of these associations were seen to be more complex with interaction terms necessary to represent them in the statistical models. We used three different datasets and our results showed that possible associations between the explanatory variables and outcomes are less likely to be shown to be statistically significant when the sample size is small. This was especially true for the dataset containing Broca's volume and surface area estimates which included information about 39 participants. Previous studies have highlighted links between Broca's area and cognitive ability, but we did not find the relevant associations to be significant.

In future studies, we recommend that the sample size is increased with volume and surface area measurements of Broca's area. A larger study would allow us to fully understand whether the volume and surface area of Broca's area effects cognitive ability or whether it is another characteristic (e.g. the functional activation of Broca's area).

BIBLIOGRAPHY

Abdul-Kareem IA, Stancak A, Parkes LM & Sluming V (2011) Increased gray matter volume of left pars opercularis in male orchestral musicians correlate positively with years of musical performance. *Journal of Magnetic Resonance Imaging*, 33, pp. 24-32.

Abramowitz M & Stegun IA (1972) *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables (10th Edition).* Washington DC, USA: US Department of Commerce.

Acer N, Çankaya MN, İşçi O, Baş O, Çamurdanoğlu M & Turgut M (2010) Estimation of cerebral surface area using vertical sectioning and magnetic resonance imaging: A stereological study. *Brain Research*, 1310, pp. 29-36.

Aitken LR (1996) *Assessment of Intellectual Functioning: 2nd Edition.* New York, USA: Plenum Press.

Amunts K, Schlaug G, Jancke L, Steinmetz H, Schleicher A, Dabringhaus A & Zilles K (1997) Motor cortex and hand motor skills: Structural compliance in the human brain. *Human Brain Mapping,* 5, pp. 206-215.

Amunts K, Schleicher A, Bürgel U, Mohlberg H, Uylings HBM & Zilles K (1999) Broca's region revisited: Cytoarchitecture and intersubject variability. *Journal of Comparative Neurology*, 412, pp. 319-341.

Annett M (2002). *Handedness and Brain Asymmetry: The Right Shift Theory.* Hove, UK: Psychology Press.

Annett M, & Manning M (1990). Reading and a balanced polymorphism for laterality and ability. *Journal of Child Psychology and Psychiatry*, 31(4), pp. 511-529.

Assessment and Qualifications Alliance (AQA) (2012) Grade Boundaries – January 2012 Exams – GCSE. [pdf] Manchester, UK: AQA. Available at <http://store.aqa.org.uk/over/stat_pdf/AQA-GCSE-GDE-BOUND-JAN12.PDF> [Accessed 16 March 2012].

Baddeley A & Jensen EBV (2005) *Stereology for Statisticians*. Boca Raton, USA: CRC Press.

Baddeley AJ, Gundersen HJG & Cruz-Orive LM (1986) Estimation of surface area from vertical sections. *Journal of Microscopy*, 142(3), pp. 259-276.

Bangert M & Schlaug G (2006) Specialization of the specialized in features of external human brain morphology. *European Journal of Neuroscience,* 24, pp. 1832-1834.

Barta P & Dazzan P (2003) Hemispheric surface area: Sex, laterality and age effects. *Cerebral Cortex*, 13, pp. 364-370.

Bengtsson SL, Nagy Z, Skare S, Forsman L, Forssberg H & Ullen F (2005) Extensive piano practising has regionally specific effects on white matter development. *Nature Neuroscience,* 8(9), pp. 1148-1150.

Blackaby DH & Manning DN (1990). The north-south divide: Earnings, unemployment and cost of living differences in Great Britain. *Papers of the Regional Science Association,* 69, pp. 43-55.

Box GEP & Cox DR (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B (Methodological),* 26(2), pp. 211-252.

Broca P (1861) Nouvelle observation d'aphémie produite par une lesion de la troisième circonvolution frontale. *Bulletins de la Société d'anatomie (Paris), 2e serie,* 6, pp. 398-407.

Brodmann K (1909) *Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues.* Leipzig, Germany: Barth.

Bynner JM (1997) Basic skills in adolescents' occupational preparation. *The Career Development Quarterly*, 45, pp. 305-321.

Cadnapaphornchai MA, Masoumi A, Strain JD, McFann K & Schrier RW (2011) Magnetic resonance imaging of kidney and cyst volume in children with ADPKD. *Clinical Journal of the American Society of Nephrology*, 6, pp. 369-376.

Caplan D (2006) Why is Broca's area involved in syntax? *Cortex*, 42, pp. 469-471.

Carpenter JR, Kenward MG & Vansteelandt S (2006) A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society, Series A*, 169, pp. 571-84.

Cleveland WS (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association,* 74, pp. 829-836.

Cleveland WS & Devlin SJ (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association,* 83, pp. 596-610.

Corballis MC, Hattie J & Fletcher R (2008) Handedness and intellectual achievement: An even-handed look. *Neuropsychologia,* 26, pp. 374-378.

Cosgrove KP, Mazure CM & Staley JK (2007) Evolving knowledge of sex differences in brain structure, function, and chemistry. *Biological Psychiatry,* 62, pp. 847-855.

Costafreda SG, Fu CHY, Lee L, Everitt B, Brammer MJ & David AS (2006) A systematic review and quantitative appraisal of fMRI studies of verbal fluency: Role of the left inferior frontal gyrus. *Human Brain Mapping,* 27, pp. 799-810.

Crawley MJ (2005) *Statistics: An Introduction Using R.* Chichester, UK: Wiley

Crawley MJ (2007) *The R Book.* Chichester, UK: Wiley

Crow TJ, Crow LR, Done DJ, & Leask S (1998). Relative hand skill predicts academic ability: Global deficits at the point of hemispheric indecision. *Neuropsychologia*, 36(12), pp. 1275-1282.

Cruz-Orive LM (1989) On the precision of systematic sampling: a review of Matheron's transitive methods. *Journal of Microscopy*, 153, pp. 315-333.

Cruz-Orive LM (1997) Stereology of single objects. *Journal of Microscopy*, 186(2), pp. 93-107.

Cruz-Orive LM (1999) Precision of Cavalieri sections and slices with local errors. *Journal of Microscopy*, 193, pp. 182-198.

Cruz-Orive LM (2006) *Estimation of Surface Area from Systematic Vertical Sections.* Num 1/2006, Department of Mathematics, Statistics and Computation, University of Cantabria.

Cruz-Orive LM & Gual-Arnau X (2002) Precision of circular systematic sampling. *Journal of Microscopy*, 207(3), pp. 225-242.

Cruz-Orive LM & Howard CV (1995) Estimation of individual feature surface area with the vertical spatial grid. *Journal of Microscopy*, 178(2), pp. 146-151.

Davis CS (2003) *Statistical Methods for the Analysis of Repeated Measurements.* New York, USA: Springer.

Delis DC, Kramer JH, Kaplan E & Ober BA (2000) *California Verbal Learning Test, 2nd Edition: Adult Version: Manual.* San Antonio, USA: Psychological Corporation.

Denny K (2008). Cognitive ability and continuous measures of relative hand skill: A note. *Neuropsychologia*, 46(7), pp. 2091-2094.

Dhaliwal GS, Murray RD, Rees EM, Howard CV & Beech DJ (2002) Quantitative unbiased estimates of endometrial gland surface area and volume in cycling cows and heifers. *Research in Veterinary Science*, 73, pp. 259-265.

Dobson AJ (1999) *An Introduction to Generalized Linear Models.* Boca Raton, USA: Chapman & Hall/CRC Press.

Dronkers NF, Plaisant O, Iba-Zizen MT & Cabanis EA (2007) Paul Broca's historic cases: High resolution MR imaging of the brains of Leborgne and Lelong. *Brain,* 130(5), pp. 1432-1441.

Eisenhart C (1947) The assumptions underlying the analysis of variance. *Biometrics*, 3(1), pp. 1-21.

Elias H (1963) Address of the president. *Proceedings of the First International Congress for Stereology*, Ch 2, pp 1-3.

Elias H & Schwartz D (1969) Surface area of the cerebral cortex of mammals determined by stereological methods. *Science*, 166(3901), pp. 111-113.

Elliott C (1990) *Differential Ability Scales: Administration and Scoring.* San Antonio, USA: The Psychological Corporation.

Ertekin T, Acer N, Turgut AT, Aycan K, Özçelik Ö & Turgut M (2011) Comparison of three methods for the estimation of the pituitary gland volume using magnetic resonance imaging: A stereological study. *Pituitary*, 14, pp. 31-38.

Everitt BS & Dunn G (2001) *Applied Multivariate Data Analysis, 2nd Edition.* London, UK: Arnold.

Fahrmeir L & Tutz G (1994) *Multivariate Statistical Modelling Based on Generalized Linear Models.* New York, USA: Springer-Verlag.

Faurie C, Vianey-Liaud N, & Raymond M (2006). Do left-handed children have advantages regarding school performance and leadership skills? *Laterality*, 11 (1), pp. 57-70.

Franklin MS, Moore KS, Yip C, Jonides J, Rattray K & Moher J (2008) The effects of musical training on verbal memory. *Psychology of Music,* 36, pp. 353-365.

Frye RE, Liederman J, Malmberg B, McLean J, Strickland D & Beauchamp MS (2010) Surface area accounts for the relation of gray matter volume to reading-related skills and history of dyslexia. *Cerebral Cortex*, 20(11), pp. 2625-2635.

Galwey NW (2006). *Introduction to Mixed Modelling: Beyond Regression and Analysis of Variance*. Chichester, UK: Wiley.

Garcia-Fiñana M & Cruz-Orive LM (2000a) New approximations for the efficiency of Cavalieri sampling. *Journal of Microscopy*, 199, pp. 224-238.

Garcia-Fiñana M & Cruz-Orive LM (2000b) Fractional trend of the variance in Cavalieri sampling. *Image Analysis & Stereology*, 19, pp. 71-79.

Garcia-Fiñana M & Cruz-Orive LM (2004) Improved variance prediction for systematic sampling on R. *Statistics*, 38(3), pp. 243-272.

Garcia-Fiñana M, Cruz-Orive LM, Mackay CE, Pakkenberg B & Roberts N (2003) Comparison of MR imaging against physical sectioning to estimate the volume of human cerebral compartments. *NeuroImage*, 18, pp. 505-516.

Gaser C & Schlaug G (2003) Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience, 23(27),* pp. 9240-9245.

Goldstein H (1989). Restricted unbiased iterative generalized least-squares estimation. *Biometrika,* 76 (3), pp. 622-623.

Goldstein H (2003) *Multilevel Statistical Models, 3rd Edition.* London, UK: Arnold.

Gordon (2003) Integrative neuroscience and psychiatry. *Neuropsychopharmacology,* 28, pp. 2-8.

Gundersen HJG & Jensen EB (1987) The efficiency of systematic sampling in Stereology and its prediction. *Journal of Microscopy*, 147, pp. 229-263.

Gundersen HJG, Jensen EBV, Kiêu K & Nielsen J (1999) The efficiency of systematic sampling in Stereology – reconsidered. *Journal of Microscopy*, 193, pp. 199-211.

Gunstad J, Paul RH, Brickman AM, Cohen RA, Arns M, Roe D, Lawrence JJ & Gordon E (2006) Patterns of cognitive performance in middle-aged and older adults: A cluster analytic examination. *Journal of Geriatric Psychiatry and Neurology,* 19(2), pp. 59-64.

Haier RJ, Jung RE, Yeo RA, Head K & Alkire MT (2005) The neuroanatomy of general intelligence: Sex matters. *NeuroImage,* 25, pp. 320-327.

Hair, Jr JF, Black WC, Babin BJ & Anderson RE (2010) *Multivariate Data Analysis: A Global Perspective, 7th Edition.* Upper Saddle River, USA: Pearson.

Hallahan BP, Craig MC, Toal F, Daly EM, Moore CJ, Ambikapathy A, Robertson D, Murphy KC & Murphy DGM (2011) In vivo brain anatomy of adult males with Fragile X syndrome: An MRI study. *NeuroImage*, 54, pp. 16-24.

Halpern DF, Haviland MG & Killian CD (1998) Handedness and sex differences in intelligence: Evidence from the medical college admission test. *Brain and Cognition,* 38, pp. 87-101.

Hamzei F, Rijntjes M, Dettmers C, Glauche V, Weiller C & Buchel C (2003) The human action recognition system and its relationship to Broca's area: An fMRI study. *NeuroImage*, 19, pp. 637-644.

Harasty J, Double KL, Halliday GM, Kril JJ & McRitchie DA (1997) Language-associated cortical regions are proportionally larger in the female brain. *Archives of Neurology*, 54(2), pp. 171-176.

Harville DA (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72, pp. 320-338.

Howard CV & Reed MG (2005) *Unbiased Stereology: Three-Dimensional Measurement in Microscopy (2nd Edition)*. Oxon, UK: BIOS Scientific Publishers.

Huberty CJ & Morris JD (1989) Multivariate analysis versus multiple univariate analysis. *Psychological Bulletin*, 105(2), pp. 302-308.

Ivic A (1985) *The Riemann Zeta-Function: The Theory of the Riemann Zeta-Function with Applications*. New York, USA: Wiley.

Jacobs DM, Rakitin BC, Zubin NR, Ventura PR & Stern Y (2001) Cognitive correlates of mnemonics usage and verbal recall memory in old age. *Neuropsychiatry, Neuropsychology, and Behavioural Neurology*, 14(1), pp. 15-22.

Jakobsen LS, Lewycky ST, Kilgour AR & Stoesz BM (2008) Memory for verbal and visual material in highly trained musicians. *Music Perception*, 26(1), pp. 41-55.

Jobson JD (1992) *Applied Multivariate Data Analysis, Volume II: Categorical and Multivariate Methods*. New York, USA: Springer-Verlag.

Johnston DW, Nicholls MER, Shah M & Shields MA (2009) Nature's experiment? Handedness and early childhood development. *Demography*, 46, pp. 281-301.

Kanaya T, Scullin MH & Ceci SJ (2003) The Flynn effect and US policies: The impact of rising IQ scores on American society via mental retardation diagnoses. *American Psychologist*, 58(10), pp. 778-790.

Kannekens EM, Murray RD, Howard CV & Currie J (2006) A stereological method for estimating the feto-maternal exchange surface area in the bovine placentome at 135 days gestation. *Research in Veterinary Science*, 81, pp. 127-133.

Karatsuba AA & Vorinin SM (1992) *The Riemann Zeta-Function*. Berlin, Germany: Walter de Gruyter.

Kaufman AS & Lichtenberger EO (1999) *Essentials of WAIS-III Assessment*. New York, USA: Wiley.

Keller SS, Highley JR, Garcia-Finana M, Sluming V, Rezaie R & Roberts N (2007) Sulcal variability, stereological measurement and asymmetry of Broca's area on MR images. *Journal of Anatomy*, 211, pp. 534-555.

Keller SS, Roberts N, García-Fiñana M, Mohammadi S, Ringelstein EB, Knecht S & Deppe M (2011) Can the language-dominant hemisphere be predicted by brain anatomy? *Journal of Cognitive Neuroscience*, 23(8), pp. 2013-2029.

Kiêu K (1997) *Three Lectures on Systematic Geometric Sampling*. Memoirs. Department of Theoretical Statistics, University of Aarhus, Vol. 13.

Kiêu K, Souchet S & Istas J (1999) Precision of systematic sampling and transitive methods. *Journal of Statistical Planning and Inference*, 77, pp. 263-279.

Koelsch S, Gunter TC, von Cramon DY, Zysset S, Lohmann G & Friederici AD (2002) Bach speaks: A cortical "language-network" serves the processing of music. *NeuroImage*, 17, pp. 956-966.

Kriegeskorte N & Goebel R (2001) An efficient algorithm for topologically correct segmentation of the cortical sheet in anatomical MR volumes. *NeuroImage, 14,* pp. 329-346.

Krzanowski WJ (2003) *Principles of Multivariate Analysis: A User's Perspective, Revised Edition.* Oxford, UK: Oxford University Press.

Laird NM & Ware JH (1982) Random-effects models for longitudinal data. *Biometrics*, 38(4), pp. 963-974.

Leask SJ & Crow TJ (1997) How far does the brain lateralize?: An unbiased method for determining the optimum degree of hemispheric specialization. *Neuropsychologia,* 35(10), pp. 1381-1387.

Leask SJ & Crow TJ (2006). A single optimum degree of hemispheric specialisation in two tasks, in two UK national birth cohorts. *Brain and Cognition,* 62, pp. 221-227.

Lee DJ, Chen Y & Schlaug G (2003) Corpus callosum: Musician and gender effects. *Neuroreport*, 14, pp. 205-209.

Littell RC, Stroup WW & Freund RJ (2002) *SAS for Linear Models, 4th Edition.* Cary, USA: SAS Institute Inc.

Matheron G (1971) *The Theory of Regionalised Variables and its Applications.* Fontainebleau, France: École Nationale Supérieure des Mines de Paris.

Mathieu O, Cruz-Orive LM, Hoppeller H & Weibel E (1981) Measuring error and sampling variation in Stereology. *Journal of Microscopy*, 121, pp. 75-88.

Mazonakis M, Sahin B, Pagonidis K & Damilakis J (2011) Assessment of left ventricular function and mass by MR imaging: A stereological study based on the systematic slice sampling procedure. *Academic Radiology*, 18(6), pp. 738-744.

McArdle JJ, Ferrer-Caja E, Hamagami F & Woodcock RW (2002) Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, 38(1), pp. 115-142.

McCullagh P & Nelder NA (1989). *Generalized Linear Models, 2nd edition.* London, UK: Chapman & Hall.

McManus C (2002) *Right hand: Left hand.* London, UK: Weidenfeld & Nicholson.

McManus IP (1985). Right- and left-hand skill: Failure of the right shift model. *British Journal of Psychology*, 76, pp. 1-16.

Mechlenburg I, Nyengaard JR, Gelineck J, Soballe K & Troelsen A (2010) Cartilage thickness in the hip measured by MRI and stereology before and after periacetabular osteotomy. *Clinical Orthopaedics and Related Research*, 468, pp. 1884-1890.

Mignani R, Corsi C, De Marco M, Caiani EG, Santucci G, Cavagna E, Severi S & Cagnoli L (2011) Assessment of kidney volume in polycystic kidney disease using magnetic resonance imaging without contrast medium. *American Journal of Nephrology*, 33, pp. 176-184.

Mitnitski A & Rockwood K (2008) Transitions in cognitive test scores over 5 and 10 years in elderly people: Evidence for a model of age-related deficit accumulation. *BMC Geriatrics*, 8:3.

Mitnitski A, Fallah N, Wu Y, Rockwood K & Borenstein AR (2010) Changes in cognition during the course of eight years in elderly Japanese Americans: A multistate transition model. *Annals of Epidemiology,* 20(6), pp. 480-486.

Molenberghs G & Kenward MG (2007) *Missing Data in Clinical Studies.* Chichester, UK: Wiley.

Nettle D (2003). Hand laterality and cognitive ability: A multiple regression approach. *Brain and Cognition*, 52, pp. 390-398.

Nicholls MER, Chapman HL, Loetscher T & Grimshaw GM (2010). The relationship between hand preference, hand performance, and general cognitive ability. *Journal of the International Neuropsychological Society,* 16, pp. 585-592.

Oldfield RC (1971)The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia,* 9, pp. 97-113.

Oppenheim AV & Schafer RW (1975) *Digital Signal Processing.* Englewood Cliffs, NJ, USA: Prentice-Hall.

Orton SJ (1937) *Reading, Writing and Speech Problems in Children.* New York, USA: Norton.

Parbery-Clark A, Skoe E, Lam C & Kraus N (2009) Musician enhancement for speech-in-noise. *Ear & Hearing,* 30(6), pp. 653-661.

Paterson L & Iannelli C (2007) Social class and educational attainment: A comparative study of England, Wales, and Scotland. *Sociology of Education*, 80, pp. 330-358.

Patterson HD & Thompson R (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika,* 58 (3), pp. 545-554.

Peters M, Reimers S, & Manning JT (2006). Hand preference for writing and associations with selected demographic and behavioural variables in 255,100 subjects: The BBC internet study. *Brain and Cognition*, 62, pp. 177-189.

Peters M, Laeng B, Latham K, Jackson M, Zaiyouna R & Richardson C (1995) A redrawn Vandenberg and Kuse mental rotation test: Different versions and factors that affect performance. *Brain and Cognition,* 28, pp. 39-58.

Pinheiro JC & Bates DM (2000) *Mixed-Effects Models in S and S-Plus.* New York, USA: Springer-Verlag.

Poh K-K (1984) *The effect of two-stage sampling on non-parametric tests.* PhD Thesis, Department of Mathematics, University of Auckland, NZ.

Powell J, Lewis PA, Roberts N, García-Fiñana M & Dunbar RIM (2012) Orbital prefrontal cortex volume predicts social network size: An imaging study of individual differences in humans. *Proceedings of the Royal Society B*, [online]. Available at <http://rspb.royalsocietypublishing.org/content/early/2012/01/27/ rspb.2011.2574.long> [Accessed 16 March 2012].

Power C & Elliott J (2006) Cohort profile: 1958 British birth cohort (National Child Development Study). *International Journal of Epidemiology,* 35, pp. 34-41.

Reed MG, Shanks E, Beech DJ, Barlow L & Howard CV (2001) Stereological estimation of eye volume using the Pappus method. *Journal of Microscopy*, 202(3), pp. 473-479.

Resch F, Haffner J, Parzer P, Pfueller U, Strehlow U & Zerahn-Hartung C (1997) Testing the hypothesis of the relationships between laterality and ability according to Annett's right-shift theory: Findings in an epidemiological sample of young adults. *British Journal of Psychology,* 88, pp. 621-635.

Roberts N, Puddephat MJ & McNulty V (2000) The benefit of stereology for quantitative radiology. *British Journal of Radiology*, 73, pp. 679-697.

Ronan L, Doherty CP, Delanty N, Thornton J & Fitzsimons M (2006) Quantitative MRI: A reliable protocol for measurement of cerebral gyrification using stereology. *Magnetic Resonance Imaging*, 24, pp. 265-272.

Salthouse TA (2006) Mental exercise and mental aging: Evaluating the validity of the "use it or lose it" hypothesis. *Perspectives on Psychological Science*, 1(1), pp. 68-87.

Schafer JL & Olsen MK (1998) Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, 33, pp. 545-571.

Schaffler L, Luders HO, Dinner DS, Lesser RP & Chelune GJ (1993) Comprehension deficits elicited by electrical stimulation of Broca's area. *Brain*, 116, pp. 695-715.

Schlaug G, Jancke L, Huang Y, Staiger JF & Steinmetz H (1995) Increased corpus callosum size in musicians. *Neuropsychologia*, 33, pp. 1047-1055.

Schmidt M (1996) *Rey Auditory and Verbal Learning Test: A Handbook.* Los Angeles, USA: Western Psychological Services.

Schneider P, Sluming V, Roberts N, Scherg M, Goebel R, Specht HJ, Dosch HG, Bleeck S, Stippich C & Rupp A (2005) Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature Neuroscience*, 8(9), pp. 1241-1247.

Schubotz RI & von Cramon DY (2002a) A blueprint for target motion: fMRI reveals perceived sequential complexity to modulate premotor cortex. *NeuroImage*, 16, pp. 920-935.

Schubotz RI & von Cramon DY (2002b) Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: An fMRI study. *NeuroImage*, 15, pp. 787-796.

Seber GAF (1977) *Linear Regression Analysis*. New York, USA: Wiley.

Sharpe SA, Eschelbach E, Basaraba RJ, Gleeson F, Hall GA, McIntyre A, Williams A, Kraft SL, Clark S, Gooch K, Hatch G, Orme IM, Marsh PD & Dennis MJ (2009) Determination of lesion volume by MRI and stereology in a macaque model of tuberculosis. *Tuberculosis*, 89, pp. 405-416.

Shepherd P (1995) The National Child Development Study (NCDS): An introduction to the origins of the study and the methods of data collection. *NCDS User Support Group Working Paper 1 (Revised).*

Sluming V, Barrick T, Howard M, Cezayirli E, Mayes A and Roberts N (2002) Voxel-Based Morphometry Reveals Increased Gray Matter Density in Broca's Area in Male Symphony Orchestra Musicians. *NeuroImage,* 17(3), pp. 1613-1622.

Sluming V, Brooks J, Howard M, Downes J and Roberts N (2007) Broca's area supports enhanced visuospatial cognition in musicians. *Journal of Neuroscience,* 27(14), pp. 3799-3806.

Smith CD, Chebrolu H, Wekstein DR, Schmitt FA and Markesbery WR (2007) Age and gender effects on human brain anatomy: A voxel-based morphometric study in healthy elderly. *Neurobiology of Aging*, 28, pp. 1075-1087.

Smyth GK & Verbyla AP (1996). A conditional likelihood approach to residual maximum likelihood estimation in generalized linear models. *Journal of the Royal Statistical Society Series B (Methodological),* 58 (3), 565-572.

Sneddon JC, Boomker E & Howard CV (2006) Mucosal surface area and fermentation activity in the hind gut of hydrated and chronically dehydrated working donkeys. *Journal of Animal Science*, 84, pp. 119-124.

Sommer IEC (2010) Sex differences in handedness, brain asymmetry, and language lateralization. In: Hugdahl K & Westerhausen R, eds. *The Two Halves of the Brain: Information Processing in the Cerebral Hemispheres.* Cambridge, USA: MIT Press, pp. 287-312.

Stephan KE, Marshall JC, Friston KJ, Rowe JB, Ritzl A, Zilles K & Fink GR (2003) Lateralized cognitive processes and lateralized task control in the human brain. *Science,* 301(5631), pp. 384-386.

Stewart L (2005) Neurocognitive studies of musical literacy acquisition. *Musicae Scientae,* 9, pp. 223-227.

Stewart L (2008) Do musicians have different brains? *Clinical Medicine,* 8(3), pp. 304-308.

Stewart L, Henson R, Kampe K, Walsh V, Turner R & Frith U (2003) Brain changes after learning to read and play music. *NeuroImage,* 20, pp. 71-83.

Talairach J & Tournoux P (1988) *Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: An Approach to Cerebral Imaging.* Stuttgart, Germany: Thieme Medical.

Thilers PP, MacDonald SWS & Herlitz A (2007) Sex differences in cognition: The role of handedness. *Physiology & Behavior,* 92(1-2), pp. 105-109.

Teng E & Chui HC (1987) The Modified Mini-Mental State (3MS) examination. *Journal of Clinical Psychiatry,* 48, pp. 314-318.

Tiffin J & Asher EJ (1948) The Purdue Pegboard: Norms and studies of reliability and validity. *Journal of Applied Psychology,* 32(3), pp. 234-247.

Toga AW & Thompson PM (2003) Mapping brain asymmetry. *Nature Reviews Neuroscience,* 4, pp. 37-48.

Tsiatis AA (2006) *Semiparametric Theory and Missing Data.* New York, USA: Springer.

Ugarte MD, Militino AF & Arnholt AT (2008) *Probability and Statistics with R.* Boca Raton, USA: CRC Press.

Uylings HBM, Malofeeva LI, Bogolepova IN, Amunts K & Zilles K (1999) Broca's language area from a neuroanatomical and developmental perspective. In: Brown CM & Hagoort P, eds. *The Neurocognition of Language.* Oxford, UK: Oxford University Press, pp. 319-336.

van Aarde J (2006) *Stereological estimation of surface area in MR images.* PhD Thesis, Medical Vision Laboratory, Department of Engineering Science, University of Oxford, UK.

Vandenberg SG & Kuse AR (1978) Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills,* 47, pp. 599-604.

Voyer D, Voyer S & Bryden MP (1995) Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin,* 117(2), pp. 250-270.

Watanabe D, Savion-Lemieux T & Penhune VB (2007) The effect of early musical training on adult motor performance: Evidence for a sensitive period in motor learning. *Experimental Brain Research*, 176, pp. 332-340.

Wayman JC (2003) Multiple imputation for missing data: What is it and how can I use it? *Paper presented at the 2003 annual meeting of the American Educational Research Association*, Chicago, USA.

Wechsler D (1981) *Wechsler Adult Intelligence Scale-Revised.* San Antonio, USA: Psychological Corporation.

Weibel ER (1979) *Stereological Methods, Vol. 1: Practical Methods for Biological Morphometry.* London, UK: Academic Press.

Weibel ER (1987) Ideas and tools; The invention and development of Stereology. *Acta Stereologica*, 6[suppl.II], pp. 23-33.

Whitfield RC (1970-71) Curriculum development in Britain. *Paedagogica Europaea*, 6, pp. 107-118.

Xu M & Eckstein Y (1995) Use of weighted least-squares method in evaluation of the relationship between dispersivity and field scale. *Ground Water*, 33(6), pp. 905-908.

Zamarian L, Ischebeck A & Delazer M (2009) Neuroscience of learning arithmetic – Evidence from brain imaging studies. *Neuroscience and Behavioural Reviews*, 33, pp. 909-925.

Zhang K, Gao X, Qi H, Li J, Zheng Z & Zhang F (2010) Gender differences in cognitive ability associated with genetic variants of NLGN4. *Neuropsychobiology*, 62, pp. 221-228.

Zilles K & Amunts K (2010) Centenary of Brodmann's map – conception and fate. *Nature Reviews Neuroscience*, 11, pp. 139-145.

Zuur AF, Ieno EN, Walker NJ, Saveliev AA & Smith GM (2009) *Mixed effect models and extensions in ecology with R*. New York, USA: Springer.

APPENDIX A

R Codes

DATASETS

data1 – Full NCDS dataset (n=18551) excluding outliers
data1a – Full NCDS dataset (n=18558) including outliers
data1b – Reduced NCDS dataset (n=11843) (i.e. all children with missing outcomes or covariates omitted)
data2 – Full MC dataset (n=149)
data2a – Full MC dataset (n=142) right-handed participants only
data3 – Broca's subset (n=39)
intra1 – Intra-observer volume estimates for participant FM032
intra2 – Intra-observer volume estimates for participant FC017
intra3 – Intra-observer surface area estimates for participant FM032
intra4 – Intra-observer surface area estimates for participant FC017

OUTPUT FROM MLwiN

STANDARDIZED RESIDUALS & FITTED VALUES FOR THE MULTIVARIATE LINEAR MIXED MODEL IN CHAPTER 3

stdres1 – Standardized residuals for the reading scores (random errors)
stdres1a – Standardized residuals for the reading scores (random effects)
stdres2 – Standardized residuals for the transformed maths scores (random errors)
stdres2a – Standardized residuals for the transformed maths scores (random effects)
fitted1 – Fitted values for the reading test scores
fitted2 – Fitted values for the transformed maths test scores

STANDARDIZED RESIDUALS & FITTED VALUES FOR THE MULTIVARIATE LINEAR MODEL IN CHAPTER 4

stdres3 – Standardized residuals for the vocabulary scores
stdres4 – Standardized residuals for the arithmetic scores
stdres5 – Standardized residuals for the visuospatial scores
fitted3 – Fitted values for the vocabulary test scores
fitted4 – Fitted values for the arithmetic test scores
fitted5 – Fitted values for the visuospatial test scores

STANDARDIZED RESIDUALS & FITTED VALUES FOR THE MULTIVARIATE LINEAR MODELS IN CHAPTER 5

stdres6 – Standardized residuals for the left and right relative Broca's volume estimates
fitted6 – Fitted values for the left and right relative Broca's volume estimates
stdres7 – Standardized residuals for the left and right relative Broca's surface area estimates
fitted7 – Fitted values for the left and right relative Broca's surface area estimates

CHAPTER 1

NORMALITY TEST OF MATHS VARIABLE (NCDS DATASET)

```
library(nortest)
aa<-data1$maths_raw
ad.test(aa)
```

TRANSFORMATIONS

```
loga<-log(aa+1)          # log(x)
ad.test(loga)
hist(loga)

expa<-exp(aa)            # exp(x)
ad.test(expa)
hist(expa)

halfpa<-(aa)^0.5         # x^0.5
ad.test(halfpa)
hist(halfpa)

tpa<-(aa)^0.1            # x^0.1
ad.test(tpa)
hist(tpa)

sqpa<-(aa)^2             # x^2
ad.test(sqpa)
hist(sqpa)

halfpa<-(aa)^0.8         # x^0.8
ad.test(halfpa)
hist(halfpa)
```

BOX-COX TRANSFORMATION

```
library(MASS)
maths11<-(data1$maths_raw)+3
boxcox(maths11~1,lambda=seq(-2,2,0.1))
boxcox(maths11~1,lambda=seq(0,1,0.02))
boxcox(maths11~1,lambda=seq(0.4,0.6,0.01))
boxcox(maths11~1,lambda=seq(0.5,0.57,0.001))
library(car)
transfmaths11<-bcPower(maths11,0.512)
ad.test(transfmaths11)
```

FIGURE 1.2

```
hist(data1$maths_trans,cex.main=2,cex.lab=1.5,cex.axis=1.5,main=list("Histogram of Transformed
Mathematics Test Scores",font=3),xlab="Transformed Mathematics Test Scores",ylab="Frequency",
axes=FALSE)
axis(2,at=seq(0,1500,by=250),cex.axis=1.5)
axis(1,at=seq(0,20,by=4),cex.axis=1.5)
box(lty=1)
```

TABLE 1.2

T TESTS

SH

```
dataSHright<-subset(data1a,data1a$SH==0)
dataSHleft<-subset(data1a,data1a$SH==1)
t.test(dataSHright$read_pc,dataSHleft$read_pc,paired=FALSE,na.action=na.omit)
t.test(dataSHright$maths_trans,dataSHleft$ maths_trans,paired=FALSE,na.action=na.omit)
```

WH

```
dataWHright<-subset(data1a,data1a$WH==0)
dataWHleft<-subset(data1a,data1a$WH==1)
t.test(dataWHright$read_pc,dataWHleft$read_pc,paired=FALSE,na.action=na.omit)
t.test(dataWHright$maths_trans,dataWHleft$maths_trans,paired=FALSE,na.action=na.omit)
```

GENDER

```
dataMale<-subset(data1a,data1a$gender==0)
dataFemale<-subset(data1a,data1a$gender==1)
t.test(dataMale$read_pc,dataFemale$read_pc,paired=FALSE,na.action=na.omit)
t.test(dataMale$maths_trans,dataFemale$maths_trans,paired=FALSE,na.action=na.omit)
```

UK REGION

```
summary(lm(read_pc~factor(UK_region),data=data1a))
confint(lm(read_pc~factor(UK_region),data=data1a))
summary(lm(maths_trans~factor(UK_region),data=data1a))
```

```
confint(lm(maths_trans~factor(UK_region),data=data1a))
```

TABLE 1.3

PEARSON CORRELATION COEFFICIENTS

RELATIVE HAND SKILL

```
dataRelHScomp<-subset(data1a,is.na(data1a$DOH)==FALSE)
dataRelHSreadcomp<-subset(dataRelHScomp,is.na(dataRelHScomp$read_pc)==FALSE)
dataRelHSmathscomp<-subset(dataRelHScomp,is.na(dataRelHScomp$maths_trans)==FALSE)
cor.test(dataRelHSreadcomp$read_pc,dataRelHSreadcomp$DOH)
cor.test(dataRelHSmathscomp$maths_trans,dataRelHSmathscomp$DOH)
```

TABLE 1.6

WELCH'S T TESTS

MUSICIAN

```
dataControl<-subset(data2a,data2a$music==0)
dataMusician<-subset(data2a,data2a$music==1)
t.test(dataControl$vocab,dataMusician$vocab,paired=FALSE,na.action=na.omit)
t.test(dataControl$arith,dataMusician$arith,paired=FALSE,na.action=na.omit)
t.test(dataControl$block,dataMusician$block,paired=FALSE,na.action=na.omit)
```

GENDER

```
dataMale<-subset(data2a,data2a$gender==0)
dataFemale<-subset(data2a,data2a$gender==1)
t.test(dataMale$vocab,dataFemale$vocab,paired=FALSE,na.action=na.omit)
t.test(dataMale$arith,dataFemale$arith,paired=FALSE,na.action=na.omit)
t.test(dataMale$block,dataFemale$block,paired=FALSE,na.action=na.omit)
```

TABLE 1.7

PEARSON CORRELATION COEFFICIENTS

AGE

```
dataAgecomp<-subset(data2a,is.na(data2a$age)==FALSE)
dataAgevocabcomp<-subset(dataAgecomp,is.na(dataAgecomp$vocab)==FALSE)
dataAgearithcomp<-subset(dataAgecomp,is.na(dataAgecomp$arith)==FALSE)
dataAgeblockcomp<-subset(dataAgecomp,is.na(dataAgecomp$block)==FALSE)
cor.test(dataAgevocabcomp$vocab,dataAgevocabcomp$age)
cor.test(dataAgearithcomp$arith,dataAgearithcomp$age)
cor.test(dataAgeblockcomp$block,dataAgeblockcomp$age)
```

TABLE 1.8

WELCH'S T TESTS

MUSICIAN

```
dataControl<-subset(data3,data3$music==0)
dataMusician<-subset(data3,data3$music==1)
t.test(dataControl$leftrelbavolpc,dataMusician$leftrelbavolpc,paired=FALSE,na.action=na.omit)
t.test(dataControl$rightrelbavolpc,dataMusician$rightrelbavolpc,paired=FALSE,na.action=na.omit)
t.test(dataControl$lhbasarelbalhvol2o3,dataMusician$lhbasarelbalhvol2o3,paired=FALSE,na.action=
na.omit)
t.test(dataControl$rhbasarelbarhvol2o3,dataMusician$rhbasarelbarhvol2o3,paired=FALSE,na.action=
na.omit)
```

GENDER

```
dataMale<-subset(data3,data3$gender==0)
dataFemale<-subset(data3,data3$gender==1)
t.test(dataMale$leftrelbavolpc,dataFemale$leftrelbavolpc,paired=FALSE,na.action=na.omit)
t.test(dataMale$rightrelbavolpc,dataFemale$rightrelbavolpc,paired=FALSE,na.action=na.omit)
```

```
t.test(dataMale$lhbasarelbalhvol2o3,dataFemale$lhbasarelbalhvol2o3,paired=FALSE,na.action=na.o
mit)
t.test(dataMale$rhbasarelbarhvol2o3,dataFemale$rhbasarelbarhvol2o3,paired=FALSE,na.action=na.o
mit)
```

TABLE 1.9

PEARSON CORRELATION COEFFICIENTS

AGE

```
cor.test(data3$leftrelbavolpc,data3$age)
cor.test(data3$rightrelbavolpc,data3$age)
cor.test(data3$lhbasarelbalhvol2o3,data3$age)
cor.test(data3$rhbasarelbarhvol2o3,data3$age)
```

VOCABULARY SCORES

```
cor.test(data3$leftrelbavolpc,data3$vocab)
cor.test(data3$rightrelbavolpc,data3$vocab)
cor.test(data3$lhbasarelbalhvol2o3,data3$vocab)
cor.test(data3$rhbasarelbarhvol2o3,data3$vocab)
```

ARITHMETIC SCORES

```
cor.test(data3$leftrelbavolpc,data3$arith)
cor.test(data3$rightrelbavolpc,data3$arith)
cor.test(data3$lhbasarelbalhvol2o3,data3$arith)
cor.test(data3$rhbasarelbarhvol2o3,data3$arith)
```

VISUOSPATIAL SCORES

```
cor.test(data3$leftrelbavolpc,data3$block)
cor.test(data3$rightrelbavolpc,data3$block)
cor.test(data3$lhbasarelbalhvol2o3,data3$block)
cor.test(data3$rhbasarelbarhvol2o3,data3$block)
```

CHAPTER 2

FIGURE 2.1(a)

```
a<-rnorm(100,0,1)
b<-runif(100,60,140)
plot(b,a,cex.main=2,cex.lab=1.5,cex.axis=1.5,ylim=c(-3,3),main=list("",font=3),xlab="",ylab="")
```

FIGURE 2.1(b)

```
a<-rnorm(30,1,0.75)
c<-rnorm(20,0.5,0.5)
d<-rnorm(30,0.5,0.2)
b<-rnorm(30,85,10)
e<-rnorm(20,105,10)
f<-rnorm(30,120,10)
g1<-min(c(a,c,d))
g2<-max(c(a,c,d))
h1<-min(c(b,e,f))
h2<-max(c(b,e,f))
plot(b,a,cex.main=2,cex.lab=1.5,cex.axis=1.5,xlim=c(h1,h2),ylim=c(-
3,3),main=list("",font=3),xlab="",ylab="")
lines(e,c,type="p")
lines(f,d,type="p")
```

CHAPTER 3

FIGURE 3.1

```
hl<-subset(data1a, data1a$SH==1)          # Left SH
hm<-subset(data1a, data1a$SH==0)          # Right SH
hlg<-subset(hl,hl$WH==1)                  # Left WH + Left SH
hlh<-subset(hl,hl$WH==0)                  # Right WH + Left SH
hmg<-subset(hm,hm$WH==1)                  # Left WH + Right SH
hmh<-subset(hm,hm$WH==0)                  # Right WH + Right SH
plot(loess.smooth(hlg$DOH,hlg$read_pc,span=0.75),xlim=c(3.8,100),ylim=c(30,50),cex.main=1.8,ce
x.lab=1.5,cex.axis=1.5,type="l",lwd=3,col="red",main=list("",font=3),xlab="",ylab="")
lines(loess.smooth(hlh$DOH,hlh$read_pc,span=0.75),type="l",lwd=3,col="green")
lines(loess.smooth(hmg$DOH,hmg$read_pc,span=0.75),type="l",lwd=3,col="purple")
lines(loess.smooth(hmh$DOH,hmh$read_pc,span=0.75),type="l",lwd=3,col="blue")
legend(71,51,cex=1.1,c("RSH, RWH","RSH, LWH","LSH, RWH", "LSH,
LWH"),lty=c(1,1,1,1),lwd=c(3,3,3,3),col=c("blue","purple","green","red"),bty="n")

hl<-subset(data1a, data1a$SH==1)          # Left SH
hm<-subset(data1a, data1a$SH==0)          # Right SH
hlg<-subset(hl,hl$WH==1)                  # Left WH + Left SH
hlh<-subset(hl,hl$WH==0)                  # Right WH + Left SH
hmg<-subset(hm,hm$WH==1)                  # Left WH + Right SH
hmh<-subset(hm,hm$WH==0)                  # Right WH + Right SH
plot(loess.smooth(hlg$DOH,hlg$maths_trans,span=0.75),xlim=c(3.8,100),ylim=c(8,12),cex.main=1.8
,cex.lab=1.5,cex.axis=1.5,type="l",lwd=3,col="red",main=list("",font=3),xlab="",ylab="")
lines(loess.smooth(hlh$DOH,hlh$maths_trans,span=0.75),type="l",lwd=3,col="green")
lines(loess.smooth(hmg$DOH,hmg$maths_trans,span=0.75),type="l",lwd=3,col="purple")
lines(loess.smooth(hmh$DOH,hmh$maths_trans,span=0.75),type="l",lwd=3,col="blue")
legend(71,12.2,cex=1.1,c("RSH, RWH","RSH, LWH","LSH, RWH", "LSH,
LWH"),lty=c(1,1,1,1),lwd=c(3,3,3,3),col=c("blue","purple","green","red"),bty="n")

hl<-subset(data1b,data1b$SH==1)           # Left SH
hm<-subset(data1b, data1b$SH==0)          # Right SH
hlg<-subset(hl,hl$WH==1)                  # Left WH + Left SH
hlh<-subset(hl,hl$WH==0)                  # Right WH + Left SH
hmg<-subset(hm,hm$WH==1)                  # Left WH + Right SH
hmh<-subset(hm,hm$WH==0)                  # Right WH + Right SH
plot(loess.smooth(hlg$DOH,hlg$read_pc,span=0.75),xlim=c(3.8,65),ylim=c(30,50),
cex.main=1.8,cex.lab=1.5,cex.axis=1.5,type="l",lwd=3,col="red",main=list("",font=3),
xlab="",ylab="")
lines(loess.smooth(hlh$DOH,hlh$read_pc,span=0.75),type="l",lwd=3,col="green")
lines(loess.smooth(hmg$DOH,hmg$read_pc,span=0.75),type="l",lwd=3,col="purple")
lines(loess.smooth(hmh$DOH,hmh$read_pc,span=0.75),type="l",lwd=3,col="blue")
legend(47,51,cex=1.1,c("RSH, RWH","RSH, LWH","LSH, RWH","LSH,
LWH"),lty=c(1,1,1,1),lwd=c(3,3,3,3),col=c("blue","purple","green","red"),bty="n")

hl<-subset(data1b, data1b$SH==1)          # Left SH
hm<-subset(data1b, data1b$SH==0)          # Right SH
hlg<-subset(hl,hl$WH==1)                  # Left WH + Left SH
hlh<-subset(hl,hl$WH==0)                  # Right WH + Left SH
hmg<-subset(hm,hm$WH==1)                  # Left WH + Right SH
hmh<-subset(hm,hm$WH==0)                  # Right WH + Right SH
plot(loess.smooth(hlg$DOH,hlg$maths_trans,span=0.75),xlim=c(3.8,65),ylim=c(8,12),cex.main=1.8,
cex.lab=1.5,cex.axis=1.5,type="l",lwd=3,col="red",main=list("",font=3),xlab="",ylab="")
lines(loess.smooth(hlh$DOH,hlh$maths_trans,span=0.75),type="l",lwd=3,col="green")
lines(loess.smooth(hmg$DOH,hmg$maths_trans,span=0.75),type="l",lwd=3,col="purple")
lines(loess.smooth(hmh$DOH,hmh$maths_trans,span=0.75),type="l",lwd=3,col="blue")
legend(47,12.2,cex=1.1,c("RSH, RWH","RSH, LWH","LSH, RWH","LSH,
LWH"),lty=c(1,1,1,1),lwd=c(3,3,3,3),col=c("blue","purple","green","red"),bty="n")
```

FIGURE 3.2

```
paa<-subset(data1,data1$WH==1)          #LEFT HANDWRITERS
paa$doh<-((paa$RH_squares_1min-
paa$LH_squares_1min)/(paa$RH_squares_1min+paa$LH_squares_1min)*100)
pab<-subset(data1,data1$WH==0)          #RIGHT HANDWRITERS
pab$doh<-((pab$RH_squares_1min-
pab$LH_squares_1min)/(pab$RH_squares_1min+pab$LH_squares_1min)*100)
pac<-subset(paa,paa$SH==1)      #LEFT HW, LEFT HD
pad<-subset(paa,paa$SH==0)      #LEFT HW, RIGHT HD
pae<-subset(pab,pab$SH==1)      #RIGHT HW, LEFT HD
paf<-subset(pab,pab$SH==0)      #RIGHT HW, RIGHT HD
pag<-c(0,0)
pah<-c(0,100)
pas<-loess.smooth(paa$doh,paa$read_pc,span=0.75)[[1]]
pas<-subset(pas,pas<=0)
lengthpas<-length(pas)
pall<-loess.smooth(paa$doh,paa$read_pc,span=0.75)[[2]]
pall<-pall[1:lengthpas]
pat<-loess.smooth(paa$doh,paa$read_pc,span=0.075)[[1]]
pat<-subset(pat,pat>0)
lengthpat<-length(pat)
palr<-loess.smooth(paa$doh,paa$read_pc,span=0.075)[[2]]
aaa<-lengthpas+1
bbb<-lengthpas+lengthpat
palr<-palr[aaa:bbb]
pav<-loess.smooth(pab$doh,pab$read_pc,span=0.075)[[1]]
pav<-subset(pav,pav<=0)
lengthpav<-length(pav)
parl<-loess.smooth(pab$doh,pab$read_pc,span=0.075)[[2]]
parl<-parl[1:lengthpav]
paya<-loess.smooth(paa$doh,paa$read_pc,span=0.075)[[1]]
payb<-loess.smooth(pab$doh,pab$read_pc,span=0.075)[[1]]
pau<-loess.smooth(pab$doh,pab$read_pc,span=0.75)[[1]]
pau<-subset(pau,pau>0)
lengthpau<-length(pau)
parr<-loess.smooth(pab$doh,pab$read_pc,span=0.75)[[2]]
aab<-lengthpav+1
bba<-lengthpav+lengthpau
parr<-parr[aab:bba]
parhw<-c(parl,parr)
palhw<-c(pall,palr)
readpaa<-subset(paa,paa$read_pc>=0)
readpab<-subset(pab,pab$read_pc>=0)
mathtpaa<-subset(paa,paa$maths_trans>=0)
mathtpab<-subset(pab,pab$maths_trans>=0)
```

(a)

```
start<--60
end<-60
n<-20
space<-(end-start)/n
outputlx<-matrix(0,ncol=1,nrow=n)
outputrx<-matrix(0,ncol=1,nrow=n)
outputy<-matrix(0,ncol=1,nrow=n)
outputnumdataleft<-matrix(0,ncol=1,nrow=n)
outputnumdataright<-matrix(0,ncol=1,nrow=n)
for(i in 1:n){
set1<-0
set2<-0
```

```
set3<-0
set4<-0
a<-start+((i-1)*space)
b<-start+(i*space)
set1<-subset(readpaa,readpaa$doh>=a)
set2<-subset(set1,set1$doh<=b)
set3<-subset(readpab,readpab$doh>=a)
set4<-subset(set3,set3$doh<=b)
c<-length(set2$number)
d<-length(set4$number)
if(c>=1){
outputlx[i]<-mean(set2$read_pc)}
else{
outputlx[i]<-NA}
if(d>=1){
outputrx[i]<-mean(set4$read_pc)}
else{
outputrx[i]<-NA}
outputy[i]<-(b+a)/2
outputnumdataleft[i]<-c
outputnumdataright[i]<-d}
plot(paya,palhw,xlim=c(-65,65),ylim=c(30,50),type="l",lwd=2,col="black",cex.main=2,cex.lab=1.5,
cex.axis=1.5,main=list("Plot of Relative Hand Skill vs Mean Reading Score by Writing
Hand",font=3),
xlab="Relative Hand Skill",ylab="Mean Reading Score")
lines(payb,parhw,type="l",col="black",lty=2,lwd=2)
lines(pag,pah,type="l",col="black",lty=3)
lines(outputy,outputlx,type="p",pch=19,col="black",lwd=2)
lines(outputy,outputrx,type="p",pch=3,col="black",lwd=2)
legend(-65,51,cex=1.3,c("Right WH","Left WH"),lty=c(2,1),lwd=c(2,2),pch=c(3,19),merge=TRUE,
col=c("black","black"),bty="n")
```

(b)

```
start<--60
end<-60
n<-20
space<-(end-start)/n
outputlx<-matrix(0,ncol=1,nrow=n)
outputrx<-matrix(0,ncol=1,nrow=n)
outputy<-matrix(0,ncol=1,nrow=n)
outputnumdataleft<-matrix(0,ncol=1,nrow=n)
outputnumdataright<-matrix(0,ncol=1,nrow=n)
for(i in 1:n){
set1<-0
set2<-0
set3<-0
set4<-0
a<-start+((i-1)*space)
b<-start+(i*space)
set1<-subset(mathtpaa,mathtpaa$doh>=a)
set2<-subset(set1,set1$doh<=b)
set3<-subset(mathtpab,mathtpab$doh>=a)
set4<-subset(set3,set3$doh<=b)
c<-length(set2$number)
d<-length(set4$number)
if(c>=1){
outputlx[i]<-mean(set2$maths_trans)}
else{
outputlx[i]<-NA}
```

```
if(d>=1){
outputrx[i]<-mean(set4$maths_trans)}
else{
outputrx[i]<-NA}
outputy[i]<-(b+a)/2
outputnumdataleft[i]<-c
outputnumdataright[i]<-d}
paklx<-c(-57.14285714,-55.29640428,-53.44995141,-51.60349854,-49.75704568,-47.91059281,
-46.06413994,-44.21768707,-42.37123421,-40.52478134,-38.67832847,-36.83187561,
-34.98542274,-33.13896987,-31.29251701,-29.44606414,-27.59961127,-25.75315841,
-23.90670554,-22.06025267,-20.21379981,-18.36734694,-16.52089407,-14.67444121,
-12.82798834,-10.98153547,-9.13508260,-7.28862974,-5.44217687,-3.59572400,-1.74927114,
0.09718173,1.94363460,3.79008746,5.63654033,7.48299320,9.32944606,11.17589893,
13.02235180,14.86880466,16.71525753,18.56171040,20.40816327,22.25461613,24.10106900,
25.94752187,27.79397473,29.64042760,31.48688047,33.33333333)
pakly<-c(9.342010,9.466312,9.604976,9.755075,9.913683,10.077875,10.244723,10.411303,
10.574686,10.731948,10.880162,11.016402,11.137742,11.241255,11.324015,11.383096,11.415572,
11.418517,11.407586,11.421365,11.382263,11.308168,11.228681,11.164528,11.136784,11.116992,
11.015821,10.907313,10.754625,10.605019,10.480288,9.651971,9.661456,9.680489,9.660583,
9.571854,9.247934,9.146494,9.121150,9.186432,9.269206,9.314575,9.376069,9.424457,9.465378,
9.502656,9.538024,9.573220,9.609980,9.650039)
pakrx<-c(-32.7433628,-30.8506411,-28.9579195,-27.0651978,-25.1724761,-23.2797544,-
21.3870327,-19.4943110,-17.6015893,-15.7088676,-13.8161459,-11.9234242,-10.0307025,-
8.1379809,-6.2452592,-4.3525375,-2.4598158,-0.5670941,1.3256276,3.2183493,5.1110710,
7.0037927,8.8965144,10.7892360,12.6819577,14.5746794,16.4674011,18.3601228,20.2528445,
22.1455662,24.0382879,25.9310096,27.8237313,29.7164530,31.6091746,33.5018963,35.3946180,
37.2873397,39.1800614,41.0727831,42.9655048,44.8582265,46.7509482,48.6436699,50.5363915,
52.4291132,54.3218349,56.2145566,58.1072783,60.0000000)
pakry<-c(9.107922,9.062213,8.992268,8.905046,8.807505,8.706604,8.609302,8.522557,8.453327,
8.408572,8.395249,8.420317,8.490736,8.613462,8.793931,8.899995,9.077718,9.370907,10.137287,
10.339965,10.542102,10.745183,10.970011,11.158430,11.321409,11.460538,11.564133,11.637914,
11.697844,11.748119,11.741991,11.717695,11.717773,11.706137,11.682854,11.649083,11.605982,
11.554711,11.496428,11.432292,11.363462,11.291096,11.216355,11.140397,11.064380,10.989463,
10.916806,10.847567,10.782906,10.723980)
plot(paklx,pakly,xlim=c(-65,65),ylim=c(6,13),type="l",lwd=2,col="black",cex.main=2,cex.lab=1.5,
cex.axis=1.5,main=list("Plot of Relative Hand Skill vs Mean Transformed Maths Score by Writing
Hand",
font=3),xlab="Relative Hand Skill",ylab="Mean Transformed Maths Score")
lines(pakrx,pakry,type="l",col="black",lty=2,lwd=2)
lines(pag,pah,type="l",col="black",lty=3)
lines(outputy,outputlx,type="p",pch=19,col="black",lwd=2)
lines(outputy,outputrx,type="p",pch=3,col="black",lwd=2)
legend(-65,13.3,cex=1.3,c("Right WH","Left WH"),lty=c(2,1),lwd=c(2,2),pch=c(3,19),merge=TRUE,
col=c("black","black"),bty="n")
```

FIGURE 3.3

```
model1<-lm(maths_trans~read_pc,data=data1b)
summary(model1)
blank<-c(-50,150)
read1<-c(2.621226+0.183766*(-50),2.813972+0.181267*150)
plot(data1$read_pc,data1$maths_trans,cex.axis=1.5,cex.lab=1.5,cex.main=2,ylim=c(-
2.5,22.5),xlim=c(-10,110),
main=list("Plot of Reading Scores v Transformed Mathematics Scores",font=3),xlab="Reading
Scores (%)",
ylab="Transformed Mathematics Scores")
lines(blank,read1)
```

CORRELATION

```
cor.test(data1$read_pc,data1$maths_trans)
```

IMPORT STANDARDIZED RESIDUALS & FITTED VALUES FROM MLwiN

Random Errors

```
stdresread<-matrix(0,nrow=18551,ncol=1)
for(i in 1:18551){
stdresread[i,1]<- stdres1[i,1]}
for(i in 1:18551){
if(stdresread[i,1]==-9.9990e+29){
stdresread[i,1]<-NA}}

stdresmatht<-matrix(0,nrow=18551,ncol=1)
for(i in 1:18551){
stdresmatht[i,1]<- stdres2[i,1]}
for(i in 1:18551){
if(stdresmatht[i,1]==-9.9990e+29){
stdresmatht[i,1]<-NA}}
```

Random Effects

```
stdresreadla<-matrix(0,nrow=11166,ncol=1)
for(i in 1:11166){
stdresreadla[i,1]<-stdres1a[i,1]}
for(i in 1:11166){
if(stdresreadla[i,1]==-9.9990e+29){
stdresreadla[i,1]<-NA}}

stdresmathtla<-matrix(0,nrow=11166,ncol=1)
for(i in 1:11166){
stdresmathtla[i,1]<- stdres2a[i,1]}
for(i in 1:11166){
if(stdresmathtla[i,1]==-9.9990e+29){
stdresmathtla[i,1]<-NA}}
```

Fitted Values

```
fittedread<-matrix(0,nrow=18551,ncol=1)
count<-1
for(i in 1:18551){
fittedread[i,1]<-fitted1[count,1]
count<-count+2}
for(i in 1:18551){
if(fittedread[i,1]==-9.9990e+29){
fittedread[i,1]<-NA}}
```

```
fittedmatht<-matrix(0,nrow=18551,ncol=1)
count<-2
for(i in 1:18551){
fittedmatht[i,1]<- fitted2[count,1]
count<-count+2}
for(i in 1:18551){
if(fittedmatht[i,1]==-9.9990e+29){
fittedmatht[i,1]<-NA}}
```

FIGURE 3.4

(a)

```
plot(stdresmatht,stdresread,cex.axis=1.5,cex.lab=1.5,cex.main=2,main=list("Plot of Multivariate Std Residuals",font=3),xlab="Standardized Residuals (Transformed Maths)",ylab="Standardized Residuals (Reading)")
```

(b)

```
plot(stdresmathtla,stdresreadla,cex.axis=1.5,cex.lab=1.5,cex.main=2,main=list("Plot of Multivariate Std Residuals",font=3),xlab="Standardized Residuals (Transformed Maths)",ylab="Standardized Residuals (Reading)")
```

FIGURE 3.5

(a) and (b) FITTED VALUES v STANDARDIZED RESIDUALS

```
plot(fittedread,stdresread,cex.axis=1.5,cex.lab=1.5,cex.main=2,ylim=c(-3.1,3.1),main=list("Fitted Values v Std Residuals",font=3),xlab="Fitted Values",ylab="Standardized Residuals")

plot(fittedmatht,stdresmatht,cex.axis=1.5,cex.lab=1.5,cex.main=2,ylim=c(-3.1,3.1),main=list("Fitted Values v Std Residuals",font=3),xlab="Fitted Values",ylab="Standardized Residuals")
```

(c) and (d) HISTOGRAMS OF STANDARDIZED RESIDUALS

```
hist(stdresread,cex.axis=1.5,cex.lab=1.5,cex.main=2,ylim=c(0,2250),main=list("Histogram of Standardized Residuals",font=3),xlab="Standardized Residuals",ylab="Frequency")
box(lty=1)

hist(stdresmatht,cex.axis=1.5,cex.lab=1.5,cex.main=2,ylim=c(0,2250),main=list("Histogram of Standardized Residuals",font=3),xlab="Standardized Residuals",ylab="Frequency")
box(lty=1)
```

(e) and (f) QQ-PLOTS OF STANDARDIZED RESIDUALS

```
qqnorm(stdresread,cex.axis=1.5,cex.lab=1.5,cex.main=2,ylim=c(-3,3),main=list("Normal QQ-Plot",font=3),xlab="Theoretical Quantiles",ylab="Sample Quantiles")
qqline(stdresread)

qqnorm(stdresmatht,cex.axis=1.5,cex.lab=1.5,cex.main=2,ylim=c(-3,3),main=list("Normal QQ-Plot",font=3),xlab="Theoretical Quantiles",ylab="Sample Quantiles")
qqline(stdresmatht)
```

CHAPTER 4

FIGURE 4.1

```
hl<-subset(data2a, data2a$music==1)      # Musicians
hm<-subset(data2a, data2a$music==0)      # Controls
hlg<-subset(hl,hl$gender==1)             # Female Musicians
hlh<-subset(hl,hl$gender==0)             # Male Musicians
hmg<-subset(hm,hm$gender==1)             # Female Controls
hmh<-subset(hm,hm$gender==0)             # Male Controls
binary1<-c(5,10)
binary2<-c(15,20)
binary3<-c(25,30)
binary4<-c(35,40)
binary5<-c(45,50)
binary6<-c(55,60)
zz<-hlg$vocab
zz<-sort(zz)
zy<-hlh$vocab
zy<-sort(zy)
zx<-hmg$vocab
zx<-sort(zx)
zw<-hmh$vocab
zw<-sort(zw)
zv<-hlg$arith
zv<-sort(zv)
zu<-hlh$arith
zu<-sort(zu)
zt<-hmg$arith
zt<-sort(zt)
zs<-hmh$arith
zs<-sort(zs)
zr<-hlg$block
zr<-sort(zr)
zq<-hlh$block
zq<-sort(zq)
zp<-hmg$block
zp<-sort(zp)
zo<-hmh$block
zo<-sort(zo)
da<-mean(zz)
db<-mean(zy)
dc<-mean(zx)
dd<-mean(zw)
de<-mean(zv)
df<-mean(zu)
dg<-mean(zt)
dh<-mean(zs)
di<-mean(zr)
dj<-mean(zq)
```

```
dk<-mean(zp)
dl<-mean(zo)
dm<-length(zz)
dn<-length(zy)
do<-length(zx)
dp<-length(zw)
dq<-length(zv)
dr<-length(zu)
ds<-length(zt)
dt<-length(zs)
du<-length(zr)
dv<-length(zq)
dw<-length(zp)
dx<-length(zo)
malemusicianvocab<-matrix(5,ncol=1,nrow=27)
femalemusicianvocab<-matrix(10,ncol=1,nrow=9)
malecontrolvocab<-matrix(15,ncol=1,nrow=43)
femalecontrolvocab<-matrix(20,ncol=1,nrow=46)
malemusicianarith<-matrix(25,ncol=1,nrow=27)
femalemusicianarith<-matrix(30,ncol=1,nrow=9)
malecontrolarith<-matrix(35,ncol=1,nrow=42)
femalecontrolarith<-matrix(40,ncol=1,nrow=46)
malemusicianblock<-matrix(45,ncol=1,nrow=27)
femalemusicianblock<-matrix(50,ncol=1,nrow=9)
malecontrolblock<-matrix(55,ncol=1,nrow=42)
femalecontrolblock<-matrix(60,ncol=1,nrow=46)
ea<-sd(zy)
eb<-sd(zz)
ec<-sd(zw)
ed<-sd(zx)
ee<-sd(zu)
ef<-sd(zv)
eg<-sd(zs)
eh<-sd(zt)
ei<-sd(zq)
ej<-sd(zr)
ek<-sd(zo)
el<-sd(zp)
em<-sqrt(dm)
en<-sqrt(dn)
eo<-sqrt(do)
ep<-sqrt(dp)
eq<-sqrt(dq)
er<-sqrt(dr)
es<-sqrt(ds)
et<-sqrt(dt)
eu<-sqrt(du)
ev<-sqrt(dv)
ew<-sqrt(dw)
ex<-sqrt(dx)
fa<-db+(1.96*(ea/en))
fb<-db-(1.96*(ea/en))
fc<-da+(1.96*(eb/em))
fd<-da-(1.96*(eb/em))
fe<-dd+(1.96*(ec/ep))
ff<-dd-(1.96*(ec/ep))
fg<-dc+(1.96*(ed/eo))
fh<-dc-(1.96*(ed/eo))
fi<-df+(1.96*(ee/er))
fj<-df-(1.96*(ee/er))
```

```
fk<-de+(1.96*(ef/eq))
fl<-de-(1.96*(ef/eq))
fm<-dh+(1.96*(eg/et))
fn<-dh-(1.96*(eg/et))
fo<-dg+(1.96*(eh/es))
fp<-dg-(1.96*(eh/es))
fq<-dj+(1.96*(ei/ev))
fr<-dj-(1.96*(ei/ev))
fs<-di+(1.96*(ej/eu))
ft<-di-(1.96*(ej/eu))
fu<-dl+(1.96*(ek/ex))
fv<-dl-(1.96*(ek/ex))
fw<-dk+(1.96*(el/ew))
fx<-dk-(1.96*(el/ew))
ga<-c(fa,fa)
gb<-c(fb,fb)
gc<-c(fc,fc)
gd<-c(fd,fd)
ge<-c(fe,fe)
gf<-c(ff,ff)
gg<-c(fg,fg)
gh<-c(fh,fh)
gi<-c(fi,fi)
gj<-c(fj,fj)
gk<-c(fk,fk)
gl<-c(fl,fl)
gm<-c(fm,fm)
gn<-c(fn,fn)
go<-c(fo,fo)
gp<-c(fp,fp)
gq<-c(fq,fq)
gr<-c(fr,fr)
gs<-c(fs,fs)
gt<-c(ft,ft)
gu<-c(fu,fu)
gv<-c(fv,fv)
gw<-c(fw,fw)
gx<-c(fx,fx)
ha<-c(da,da)
hb<-c(db,db)
hc<-c(dc,dc)
hd<-c(dd,dd)
he<-c(de,de)
hf<-c(df,df)
hg<-c(dg,dg)
hh<-c(dh,dh)
hi<-c(di,di)
hj<-c(dj,dj)
hk<-c(dk,dk)
hl<-c(dl,dl)
binary11<-c(4,6)
binary12<-c(9,11)
binary21<-c(14,16)
binary22<-c(19,21)
binary31<-c(24,26)
binary32<-c(29,31)
binary41<-c(34,36)
binary42<-c(39,41)
binary51<-c(44,46)
binary52<-c(49,51)
```

```
binary61<-c(54,56)
binary62<-c(59,61)
plot(malemusicianvocab,zy,ylim=c(4,19),xlim=c(4,61),type="p",col="black",cex.main=2,cex.lab=1.5
,main=list("Plot of Male & Female, Musicians & Controls vs Mean Cognitive Ability
Scores",font=3),xlab="",ylab="Mean Vocabulary/Arithmetic/Visuospatial Score",axes=FALSE)
lines(femalemusicianvocab,zz,type="p",col="black")
lines(malecontrolvocab,zw,type="p",col="black")
lines(femalecontrolvocab,zx,type="p",col="black")
lines(malemusicianarith,zu,type="p",col="black")
lines(femalemusicianarith,zv,type="p",col="black")
lines(malecontrolarith,zs,type="p",col="black")
lines(femalecontrolarith,zt,type="p",col="black")
lines(malemusicianblock,zq,type="p",col="black")
lines(femalemusicianblock,zr,type="p",col="black")
lines(malecontrolblock,zo,type="p",col="black")
lines(femalecontrolblock,zp,type="p",col="black")
lines(binary11,hb,type="l",col="black")
lines(binary11,ga,type="l",lty=2,col="black")
lines(binary11,gb,type="l",lty=2,col="black")
lines(binary12,ha,type="l",col="black")
lines(binary12,gc,type="l",lty=2,col="black")
lines(binary12,gd,type="l",lty=2,col="black")
lines(binary21,hd,type="l",col="black")
lines(binary21,ge,type="l",lty=2,col="black")
lines(binary21,gf,type="l",lty=2,col="black")
lines(binary22,hc,type="l",col="black")
lines(binary22,gg,type="l",lty=2,col="black")
lines(binary22,gh,type="l",lty=2,col="black")
lines(binary31,hf,type="l",col="black")
lines(binary31,gi,type="l",lty=2,col="black")
lines(binary31,gj,type="l",lty=2,col="black")
lines(binary32,he,type="l",col="black")
lines(binary32,gk,type="l",lty=2,col="black")
lines(binary32,gl,type="l",lty=2,col="black")
lines(binary41,hh,type="l",col="black")
lines(binary41,gm,type="l",lty=2,col="black")
lines(binary41,gn,type="l",lty=2,col="black")
lines(binary42,hg,type="l",col="black")
lines(binary42,go,type="l",lty=2,col="black")
lines(binary42,gp,type="l",lty=2,col="black")
lines(binary51,hj,type="l",col="black")
lines(binary51,gq,type="l",lty=2,col="black")
lines(binary51,gr,type="l",lty=2,col="black")
lines(binary52,hi,type="l",col="black")
lines(binary52,gs,type="l",lty=2,col="black")
lines(binary52,gt,type="l",lty=2,col="black")
lines(binary61,hl,type="l",col="black")
lines(binary61,gu,type="l",lty=2,col="black")
lines(binary61,gv,type="l",lty=2,col="black")
lines(binary62,hk,type="l",col="black")
lines(binary62,gw,type="l",lty=2,col="black")
lines(binary62,gx,type="l",lty=2,col="black")
legend(2,20,cex=1.5,"Vocabulary Test",bty="n")
legend(22.6,20,cex=1.5,"Arithmetic Test",bty="n")
legend(42,20,cex=1.5,"Visuospatial Test",bty="n")
legend(0,6,cex=1.3,c("Mean","95% Confidence Interval"),lty=c(1,2),bty="n")
b1<-c(22.5,22.5)
c1<-c(0,25)
lines(b1,c1,type="l",lty=3,col="black")
b2<-c(42.5,42.5)
```

```
c2<-c(0,25)
lines(b2,c2,type="l",lty=3,col="black")
axis(1,at=c(5,10),cex.axis=1.5,labels=c("Male","Female"),line=1)
axis(1,at=c(15,20),cex.axis=1.5,labels=c("Male","Female"),line=1)
axis(1,at=c(25,30),cex.axis=1.5,labels=c("Male","Female"),line=1)
axis(1,at=c(35,40),cex.axis=1.5,labels=c("Male","Female"),line=1)
axis(1,at=c(45,50),cex.axis=1.5,labels=c("Male","Female"),line=1)
axis(1,at=c(55,60),cex.axis=1.5,labels=c("Male","Female"),line=1)
axis(1,at=c(7.5,17.5),cex.axis=1.5,labels=c("Musicians","Controls"),line=2,lty="blank")
axis(1,at=c(27.5,37.5),cex.axis=1.5,labels=c("Musicians","Controls"),line=2,lty="blank")
axis(1,at=c(47.5,57.5),cex.axis=1.5,labels=c("Musicians","Controls"),line=2,lty="blank")
axis(2,at=seq(4,20,by=2),cex.axis=1.5)
box(lty=1)
```

FIGURE 4.2

```
model1<-lm(vocab~arith,data=data2a)
summary(model1)
za<--10
zb<-100
blank<-c(za,zb)
vocab1<-c(7.76052+0.35812*za,7.76052+0.35812*zb)
plot(data2$arith,data2$vocab,cex.axis=1.5,ylim=c(0,20),xlim=c(4,20),main=list("",font=3),xlab="",
ylab="")
lines(blank,vocab1)

model2<-lm(vocab~block,data=data2a)
summary(model2)
za<--10
zb<-100
blank<-c(za,zb)
vocab2<-c(8.01511+0.31935*za,8.01511+0.31935*zb)
plot(data2$block,data2$vocab,cex.axis=1.5,ylim=c(2,20),xlim=c(4,20),main=list("",font=3),xlab="",
ylab="")
lines(blank,vocab2)

model3<-lm(arith~block,data= data2a)
summary(model3)
za<--10
zb<-100
blank<-c(za,zb)
arith2<-c(5.45206+0.45626*za,5.45206+0.45626*zb)
plot(data2$block,data2$arith,cex.axis=1.5,ylim=c(2,20),xlim=c(0,20),main=list("",font=3),xlab="",
ylab="")
lines(blank,arith2)
```

CORRELATION

```
cor.test(data2a$arith,data2a$vocab)
cor.test(data2a$block,data2a$vocab)
cor.test(data2a$block,data2a$arith)
```

CALCULATION OF R2 FOR EACH UNIVARIATE MODEL

```
model1<-lm(vocab~music*gender+age,data=data2a)
summary(model1)
model2<-lm(arith~music*gender+age,data= data2a)
summary(model2)
model3<-lm(block~music*gender+age,data= data2a)
summary(model3)
```

IMPORT STANDARDIZED RESIDUALS & FITTED VALUES FROM MLwiN

Fitted Values

```
multivocabfitted<-matrix(0,nrow=142,ncol=1)
count<-1
for(i in 1:142){
multivocabfitted[i,1]<-fitted3[count,1]
count<-count+3}
for(i in 1:142){
if(multivocabfitted[i,1]==-9.9990e+29){
multivocabfitted[i,1]<-NA}}

multiarithfitted<-matrix(0,nrow=142,ncol=1)
count<-2
for(i in 1:142){
multiarithfitted[i,1]<-fitted4[count,1]
count<-count+3}
for(i in 1:142){
if(multiarithfitted[i,1]==-9.9990e+29){
multiarithfitted[i,1]<-NA}}

multiblockfitted<-matrix(0,nrow=142,ncol=1)
count<-3
for(i in 1:142){
multiblockfitted[i,1]<-fitted5[count,1]
count<-count+3}
for(i in 1:142){
if(multiblockfitted[i,1]==-9.9990e+29){
multiblockfitted[i,1]<-NA}}
```

Standardized Residuals

```
multivocabstdres<-
matrix(0,nrow=142,ncol=1)
for(i in 1:142){
multivocabstdres[i,1]<-stdres3[i,1]}
for(i in 1:142){
if(multivocabstdres[i,1]==-9.9990e+29){
multivocabstdres[i,1]<-NA}}

multiarithstdres<-matrix(0,nrow=142,ncol=1)
for(i in 1:142){
multiarithstdres[i,1]<- stdres4[i,1]}
for(i in 1:142){
if(multiarithstdres[i,1]==-9.9990e+29){
multiarithstdres[i,1]<-NA}}

multiblockstdres<-
matrix(0,nrow=142,ncol=1)
for(i in 1:142){
multiblockstdres[i,1]<- stdres5[i,1]}
for(i in 1:142){
if(multiblockstdres[i,1]==-9.9990e+29){
multiblockstdres[i,1]<-NA}}
```

FIGURE 4.3

(a)

```
plot(multiarithstdres,multivocabstdres,cex.axis=1.5,cex.lab=1.5,cex.main=2,main=list("Plot of
Multivariate Std Residuals",font=3),xlab="Standardized Residuals (Arithmetic
Scores)",ylab="Standardized Residuals (Vocabulary Scores)")
```

(b)

```
plot(multiblockstdres,multivocabstdres,cex.axis=1.5,cex.lab=1.5,cex.main=2,main=list("Plot of
Multivariate Std Residuals",font=3),xlab="Standardized Residuals (Visuospatial
Scores)",ylab="Standardized Residuals (Vocabulary Scores)")
```

(c)

```
plot(multiblockstdres,multiarithstdres,cex.axis=1.5,cex.lab=1.5,cex.main=2,main=list("Plot of
Multivariate Std Residuals",font=3),xlab="Standardized Residuals (Visuospatial
Scores)",ylab="Standardized Residuals (Arithmetic Scores)")
```

FIGURE 4.4

(a), (b) and (c) FITTED VALUES v STANDARDIZED RESIDUALS

```
plot(multivocabfitted,multivocabstdres,cex.axis=2,main=list("",font=3),xlab="",ylab="")

plot(multiarithfitted,multiarithstdres,cex.axis=2,main=list("",font=3),xlab="",ylab="")

plot(multiblockfitted,multiblockstdres,cex.axis=2,main=list("",font=3),xlab="",ylab="")
```

(d), (e) and (f) HISTOGRAMS OF STANDARDIZED RESIDUALS

```
hist(multivocabstdres,cex.axis=2,main=list("",font=3),xlab="",ylab="")
box(lty=1)

hist(multiarithstdres,cex.axis=2,main=list("",font=3),xlab="",ylab="")
box(lty=1)

hist(multiblockstdres,cex.axis=2,main=list("",font=3),xlab="",ylab="")
box(lty=1)
```

(g), (h) and (i) QQ-PLOTS OF STANDARDIZED RESIDUALS

```
qqnorm(multivocabstdres,cex.axis=2,main=list("",font=3),xlab="",ylab="")
qqline(multivocabstdres)

qqnorm(multiarithstdres,cex.axis=2,main=list("",font=3),xlab="",ylab="")
qqline(multiarithstdres)

qqnorm(multiblockstdres,cex.axis=2,main=list("",font=3),xlab="",ylab="")
qqline(multiblockstdres)
```

CHAPTER 5

FIGURE 5.1

```
summary(lm(rightrelbavolpc~leftrelbavolpc,data3))
cc<--1000
cd<-1000000
ce<-0.9175+0.2665*cc
cf<-0.9175+0.2665*cd
ca<-c(cc,cd)
cb<-c(ce,cf)
plot(data3$leftrelbavolpc,data3$rightrelbavolpc,ylim=c(0.5,2.1),xlim=c(0.25,1.9),cex.main=2,
cex.axis=1.5,cex.lab=1.5,main=list("Plot of Left RelBAV vs Right RelBAV",font=3),xlab="Left
RelBAV",ylab="Right RelBAV")
lines(ca,cb)
```

CORRELATION

```
cor.test(data3$leftrelbavolpc,data3$rightrelbavolpc)
```

CALCULATION OF R^2 FOR UNIVARIATE MODELS

```
summary(lm(leftrelbavolpc~gender,data3))
summary(lm(rightrelbavolpc~gender,data3))
```

IMPORT STANDARDIZED RESIDUALS AND FITTED VALUES FROM MLwiN

Standardized Residuals

```
modeldiagnosticsleft<-
matrix(0,nrow=39,ncol=2)
modeldiagnosticsright<-
matrix(0,nrow=39,ncol=2)
modeldiagnosticsleft[,1]<-stdres6[,1]
modeldiagnosticsright[,1]<-stdres6 [,2]
```

Fitted Values

```
count<-1
count1<-2
for(i in 1:39){
modeldiagnosticsleft[i,2]<-fitted6[count,1]
modeldiagnosticsright[i,2]<-fitted6[count1,2]
count<-count+2
count1<-count1+2}
```

FIGURE 5.2

```
plot(modeldiagnosticsleft[,1],modeldiagnosticsright[,1],cex.axis=1.5,cex.lab=1.5,cex.main=2,
main=list("Plot of Multivariate Std Residuals",font=3),xlab="Standardized Residuals (Left
RelBAV)",ylab="Standardized Residuals (Right RelBAV)")
```

FIGURE 5.3

(a) and (b) FITTED VALUES v STANDARDIZED RESIDUALS

```
plot(modeldiagnosticsleft[,2],modeldiagnosticsleft[,1],cex.axis=1.5,main=list("",font=3),xlab="",
ylab="")

plot(modeldiagnosticsright[,2],modeldiagnosticsright[,1],cex.axis=1.5,main=list("",font=3),xlab="",
ylab="")
```

(c) and (d) HISTOGRAMS OF STANDARDIZED RESIDUALS

```
hist(modeldiagnosticsleft[,1],cex.axis=1.5,main=list("",font=3),xlab="",ylab="")
box(lty=1)

hist(modeldiagnosticsright[,1],cex.axis=1.5,main=list("",font=3),xlab="",ylab="")
box(lty=1)
```

(e) and (f) QQ-PLOTS OF STANDARDIZED RESIDUALS

```
qqnorm(modeldiagnosticsleft[,1],cex.axis=1.5,main=list("",font=3),xlab="",ylab="")
qqline(modeldiagnosticsleft[,1])

qqnorm(modeldiagnosticsright[,1],cex.axis=1.5,main=list("",font=3),xlab="",ylab="")
qqline(modeldiagnosticsright[,1])
```

FIGURE 5.4

```
summary(lm(rhbasarelbarhvol2o3~lhbasarelbalhvol2o3,data3))
cc<-0
cd<-1000000
ce<-0.7641+0.8453*cc
cf<-0.7641+0.8453*cd
ca<-c(cc,cd)
cb<-c(ce,cf)
plot(data3$lhbasarelbalhvol2o3,data3$rhbasarelbarhvol2o3,cex.main=2,cex.lab=1.5,cex.axis=1.5,
ylim=c(2,6),xlim=c(2,5.5),main=list("Plot of Left RelBASA vs Right RelBASA",font=3),xlab="Left
RelBASA",ylab="Right RelBASA")
lines(ca,cb)
```

CORRELATION

```
cor.test(data3$lhbasarelbalhvol2o3,data3$rhbasarelbarhvol2o3)
```

FIGURE 5.5

```
bamusic<-subset(data3, data3$music==1)
banm<-subset(data3, data3$music==0)
summary(lm(rhbasarelbarhvol2o3~vocab,bamusic))
summary(lm(rhbasarelbarhvol2o3~vocab,banm))
cc<-0
cd<-100
ce<-4.46968-0.05027*cc
cf<-4.46968-0.05027*cd
ca<-c(cc,cd)
cb<-c(ce,cf)
```

```
plot(bamusic$vocab,bamusic$rhbasarelbarhvol2o3,cex.main=2,cex.lab=1.5,cex.axis=1.5,ylim=c(2,6),
xlim=c(6,19),type="p",pch=19,col="black",cex.axis=1.5,cex.lab=1.5,cex.main=2,main=list("Plot of
Vocabulary Scores vs Right RelBASA by 'Musician'",font=3),xlab="Vocabulary Scores",ylab="Right
RelBASA")
points(banm$vocab,banm$rhbasarelbarhvol2o3,type="p",pch=120,col="black")
lines(ca,cb,type="l",col="black",lty=1)
cc<-0
cd<-100
ce<-4.69118-0.04672*cc
cf<-4.69118-0.04672*cd
ca<-c(cc,cd)
cb<-c(ce,cf)
lines(ca,cb,type="l",col="black",lty=2)
legend(16,6.15,cex=1.3,c("Musicians","Non-
musicians"),pch=c(19,120),lty=c(1,2),col=c("black","black"),bty="n")
```

CALCULATION OF R^2 FOR THE UNIVARIATE LINEAR MODELS

```
summary(lm(lhbasarelbalhvol2o3~gender+music,data3))
summary(lm(rhbasarelbarhvol2o3~gender+music,data3))
```

IMPORT STANDARDIZED RESIDUALS AND FITTED VALUES FROM MLwiN

Standardized Residuals

```
modeldiagnosticsleft<-
matrix(0,nrow=39,ncol=2)
modeldiagnosticsright<-
matrix(0,nrow=39,ncol=2)
modeldiagnosticsleft[,1]<-stdres7[,1]
modeldiagnosticsright[,1]<-stdres7[,2]
```

Fitted Values

```
count<-1
count1<-2
for(i in 1:39){
modeldiagnosticsleft[i,2]<-fitted7[count,1]
modeldiagnosticsright[i,2]<-fitted7[count1,2]
count<-count+2
count1<-count1+2}
```

FIGURE 5.6

```
plot(modeldiagnosticsleft[,1],modeldiagnosticsright[,1],cex.axis=1.5,cex.lab=1.5,cex.main=2,
main=list("Plot of Multivariate Std Residuals",font=3),xlab="Standardized Residuals (Left
RelBASA)",ylab="Standardized Residuals (Right RelBASA)")
```

FIGURE 5.7

(a) and (b) FITTED VALUES v STANDARDIZED RESIDUALS

```
plot(modeldiagnosticsleft[,2],modeldiagnosticsleft[,1],cex.axis=1.5,main=list("",font=3),xlab="",
ylab="")

plot(modeldiagnosticsright[,2],modeldiagnosticsright[,1],cex.axis=1.5,main=list("",font=3),xlab="",
ylab="")
```

(c) and (d) HISTOGRAMS OF STANDARDIZED RESIDUALS

```
hist(modeldiagnosticsleft[,1],cex.axis=1.5,main=list("",font=3),xlab="",ylab="")
box(lty=1)

hist(modeldiagnosticsright[,1],cex.axis=1.5,main=list("",font=3),xlab="",ylab="")
box(lty=1)
```

(e) and (f) QQ-PLOTS OF STANDARDIZED RESIDUALS

```
qqnorm(modeldiagnosticsleft[,1],cex.axis=1.5,main=list("",font=3),xlab="",ylab="")
qqline(modeldiagnosticsleft[,1])
```

```
qqnorm(modeldiagnosticsright[,1],cex.axis=1.5,main=list("",font=3),xlab="",ylab="")
qqline(modeldiagnosticsright[,1])
```

WITHIN OBSERVER CE CALCULATIONS - VOLUME

FM032

```
meanLHPObr1<-mean(intra1$LH_PO)
varLHPObr1<-var(intra1$LH_PO)
meanLHPTbr1<-mean(intra1$LH_PT)
varLHPTbr1<-var(intra1$LH_PT)
meanRHPObr1<-mean(intra1$RH_PO)
varRHPObr1<-var(intra1$RH_PO)
meanRHPTbr1<-mean(intra1$RH_PT)
varRHPTbr1<-var(intra1$RH_PT)
```

FC017

```
meanLHPObr2<-mean(intra2$LH_PO)
varLHPObr2<-var(intra2$LH_PO)
meanLHPTbr2<-mean(intra2$LH_PT)
varLHPTbr2<-var(intra2$LH_PT)
meanRHPObr2<-mean(intra2$RH_PO)
varRHPObr2<-var(intra2$RH_PO)
meanRHPTbr2<-mean(intra2$RH_PT)
varRHPTbr2<-var(intra2$RH_PT)
```

```
overallmeanLHPO2brVol<-(meanLHPObr1+meanLHPObr2)/2
overallmeanLHPO2brVol
varLHPO2brVol<-(varLHPObr1+varLHPObr2)/2
varLHPO2brVol
ceLHPO2brVol<-sqrt(varLHPO2brVol)/overallmeanLHPO2brVol
ceLHPO2brVol
overallmeanLHPT2brVol<-(meanLHPTbr1+meanLHPTbr2)/2
overallmeanLHPT2brVol
varLHPT2brVol<-(varLHPTbr1+varLHPTbr2)/2
varLHPT2brVol
ceLHPT2brVol<-sqrt(varLHPT2brVol)/overallmeanLHPT2brVol
ceLHPT2brVol
overallmeanRHPO2brVol<-(meanRHPObr1+meanRHPObr2)/2
overallmeanRHPO2brVol
varRHPO2brVol<-(varRHPObr1+varRHPObr2)/2
varRHPO2brVol
ceRHPO2brVol<-sqrt(varRHPO2brVol)/overallmeanRHPO2brVol
ceRHPO2brVol
overallmeanRHPT2brVol<-(meanRHPTbr1+meanRHPTbr2)/2
overallmeanRHPT2brVol
varRHPT2brVol<-(varRHPTbr1+varRHPTbr2)/2
varRHPT2brVol
ceRHPT2brVol<-sqrt(varRHPT2brVol)/overallmeanRHPT2brVol
ceRHPT2brVol
```

WITHIN OBSERVER CE CALCULATIONS – SURFACE AREA

FM032

```
meanLHPObr1<-mean(intra3$LH_PO)
varLHPObr1<-var(intra3$LH_PO)
meanLHPTbr1<-mean(intra3$LH_PT)
varLHPTbr1<-var(intra3$LH_PT)
meanRHPObr1<-mean(intra3$RH_PO)
varRHPObr1<-var(intra3$RH_PO)
meanRHPTbr1<-mean(intra3$RH_PT)
varRHPTbr1<-var(intra3$RH_PT)
```

FC017

```
meanLHPObr2<-mean(intra4$LH_PO)
varLHPObr2<-var(intra4$LH_PO)
meanLHPTbr2<-mean(intra4$LH_PT)
varLHPTbr2<-var(intra4$LH_PT)
meanRHPObr2<-mean(intra4$RH_PO)
varRHPObr2<-var(intra4$RH_PO)
meanRHPTbr2<-mean(intra4$RH_PT)
varRHPTbr2<-var(intra4$RH_PT)
```

```
overallmeanLHPO2brVol<-(meanLHPObr1+meanLHPObr2)/2
overallmeanLHPO2brVol
varLHPO2brVol<-(varLHPObr1+varLHPObr2)/2
varLHPO2brVol
ceLHPO2brVol<-sqrt(varLHPO2brVol)/overallmeanLHPO2brVol
ceLHPO2brVol
overallmeanLHPT2brVol<-(meanLHPTbr1+meanLHPTbr2)/2
overallmeanLHPT2brVol
```

```
varLHPT2brVol<-(varLHPTbr1+varLHPTbr2)/2
varLHPT2brVol
ceLHPT2brVol<-sqrt(varLHPT2brVol)/overallmeanLHPT2brVol
ceLHPT2brVol
overallmeanRHPO2brVol<-(meanRHPObr1+meanRHPObr2)/2
overallmeanRHPO2brVol
varRHPO2brVol<-(varRHPObr1+varRHPObr2)/2
varRHPO2brVol
ceRHPO2brVol<-sqrt(varRHPO2brVol)/overallmeanRHPO2brVol
ceRHPO2brVol
overallmeanRHPT2brVol<-(meanRHPTbr1+meanRHPTbr2)/2
overallmeanRHPT2brVol
varRHPT2brVol<-(varRHPTbr1+varRHPTbr2)/2
varRHPT2brVol
ceRHPT2brVol<-sqrt(varRHPT2brVol)/overallmeanRHPT2brVol
ceRHPT2brVol
```

CHAPTER 6

MISSING DATA ANALYSIS – NCDS DATASET

FIVE LOGISTIC REGRESSION MODELS

```
model1<-glm(Complete~factor(UK_region),data=data1,family=binomial)
summary(model1)
model2<-glm(Complete~factor(UK_region)*DOH,data=data1,family=binomial)
summary(model2)
model3<-glm(Complete~factor(UK_region)*DOH+gender+gender:factor(UK_region),data=data1,
family=binomial)
summary(model3)
model4<-glm(Complete~factor(UK_region)*DOH+gender+gender:factor(UK_region)+SH+
SH:DOH,data=data1,family=binomial)
summary(model4)
model5<-glm(Complete~factor(UK_region)*DOH+gender+gender:factor(UK_region)+SH+
SH:DOH+WH+WH:SH+WH:DOH+WH:SH:DOH,data=data1,family=binomial)
summary(model5)
anova(model2,model3)
anova(model2,model4)
anova(model3,model4)
AIC(model1,k=log(15357))
AIC(model2,k=log(12760))
AIC(model3,k=log(12760))
AIC(model4,k=log(12760))
AIC(model5,k=log(12617))
```

CALCULATING THE WEIGHTS FOR LOGISTIC REGRESSION MODEL 2

```
data1c<-subset(data1,is.na(data1$read_pc)==FALSE)
data1c<-subset(data1c,is.na(data1c$maths_pc)==FALSE)
data1nomissingUK_region1<-subset(data1b,data1b$UK_region==1)
data1nomissingUK_region2<-subset(data1b,data1b$UK_region==2)
data1nomissingUK_region3<-subset(data1b,data1b$UK_region==3)
data1nomissingUK_region4<-subset(data1b,data1b$UK_region==4)
length(data1nomissingUK_region1$number)
length(data1nomissingUK_region2$number)
length(data1nomissingUK_region3$number)
length(data1nomissingUK_region4$number)
outputUK_region1doh<-0
for(i in 1:6301){
ba<-data1nomissingUK_region1$DOH
a=exp(2.692401+(-0.005726*ba[i]))
```

```
b<-a/(a+1)
outputUK_region1doh[i]=1/b}
data1nomissingUK_region1$weightInc<-outputUK_region1doh
outputUK_region2doh<-0
for(i in 1:3578){
ba<-data1nomissingUK_region2$DOH
a=exp(2.692401+0.070651+(-0.005726*ba[i])+(0.006078*ba[i]))
b<-a/(a+1)
outputUK_region2doh[i]=1/b}
data1nomissingUK_region2$weightInc<-outputUK_region2doh
outputUK_region3doh<-0
for(i in 1:654){
ba<-data1nomissingUK_region3$DOH
a=exp(2.692401-0.443885+(-0.005726*ba[i])+(0.060624*ba[i]))
b<-a/(a+1)
outputUK_region3doh[i]=1/b}
data1nomissingUK_region3$weightInc<-outputUK_region3doh
outputUK_region4doh<-0
for(i in 1:1310){
ba<-data1nomissingUK_region4$DOH
a=exp(2.692401+0.135837+(-0.005726*ba[i])+(0.02382*ba[i]))
b<-a/(a+1)
outputUK_region4doh[i]=1/b}
data1nomissingUK_region4$weightInc<-outputUK_region4doh
nomissingdata1<-
rbind(data1nomissingUK_region1,data1nomissingUK_region2,data1nomissingUK_region3,
data1nomissingUK_region4)
data1missing1<-subset(data1c,is.na(data1c$UK_region)==TRUE)
data1missing2<-subset(data1c,is.na(data1c$DOH)==TRUE)
meanDOH<-0
for(i in 1:2154){
ca<-data1missing2$UK_region[i]
cb<-data1missing2$read_pc[i]
cc<-data1missing2$maths_pc[i]
da<-subset(nomissingdata1,nomissingdata1$UK_region==ca)
db<-subset(da,da$read_pc==cb)
dc<-subset(db,db$maths_pc==cc)
ea<-length(dc$number)
if(ea>=1){
meanDOH[i]<-mean(dc$DOH)}
else{
meanDOH[i]<-NA}}
meanDOH
data1missing2$DOH<-meanDOH
data1missing2NA<-subset(data1missing2,is.na(data1missing2$DOH)==TRUE)
meanDOHv2<-0
for(i in 1:104){
ca<-data1missing2NA$UK_region[i]
cb<-data1missing2NA$read_pc[i]
da<-subset(nomissingdata1,nomissingdata1$UK_region==ca)
db<-subset(da,da$read_pc==cb)
ea<-length(db$number)
if(ea>=1){
meanDOHv2[i]<-mean(db$DOH)}
else{
meanDOHv2[i]<-NA}}
meanDOHv2
data1missing2NA$DOH<-meanDOHv2
data1missing2NA2<-subset(data1missing2NA,is.na(data1missing2NA$DOH)==TRUE)
length(data1missing2NA2$number)
```

```
fa<-subset(data1missing2,data1missing2$UK_region==1)
fb<-subset(data1missing2,data1missing2$UK_region==2)
fc<-subset(data1missing2,data1missing2$UK_region==3)
fd<-subset(data1missing2,data1missing2$UK_region==4)
fe<-subset(data1missing2NA,data1missing2NA$UK_region==1)
ff<-subset(data1missing2NA,data1missing2NA$UK_region==2)
fg<-subset(data1missing2NA,data1missing2NA$UK_region==3)
fh<-subset(data1missing2NA,data1missing2NA$UK_region==4)
outputfa<-0
for(i in 1:1174){
ba<-fa$DOH
a=exp(2.692401+(-0.005726*ba[i]))
b<-a/(a+1)
outputfa[i]=1/b}
fa$weightInc<-outputfa
outputfb<-0
for(i in 1:705){
ba<-fb$DOH
a=exp(2.692401+0.070651+(-0.005726*ba[i])+(0.006078*ba[i]))
b<-a/(a+1)
outputfb[i]=1/b}
fb$weightInc<-outputfb
outputfc<-0
for(i in 1:107){
ba<-fc$DOH
a=exp(2.692401+0.070651+(-0.005726*ba[i])+(0.006078*ba[i]))
b<-a/(a+1)
outputfc[i]=1/b}
fc$weightInc<-outputfc
outputfd<-0
for(i in 1:168){
ba<-fd$DOH
a=exp(2.692401+0.135837+(-0.005726*ba[i])+(0.02382*ba[i]))
b<-a/(a+1)
outputfd[i]=1/b}
fd$weightInc<-outputfd
outputfe<-0
for(i in 1:18){
ba<-fe$DOH
a=exp(2.692401+(-0.005726*ba[i]))
b<-a/(a+1)
outputfe[i]=1/b}
fe$weightInc<-outputfe
outputff<-0
for(i in 1:26){
ba<-ff$DOH
a=exp(2.692401+0.070651+(-0.005726*ba[i])+(0.006078*ba[i]))
b<-a/(a+1)
outputff[i]=1/b}
ff$weightInc<-outputff
outputfg<-0
for(i in 1:36){
ba<-fg$DOH
a=exp(2.692401+0.070651+(-0.005726*ba[i])+(0.006078*ba[i]))
b<-a/(a+1)
outputfg[i]=1/b}
fg$weightInc<-outputfg
outputfh<-0
for(i in 1:24){
ba<-fh$DOH
```

```
a=exp(2.692401+0.135837+(-0.005726*ba[i])+(0.02382*ba[i]))
b<-a/(a+1)
outputfh[i]=1/b}
fh$weightInc<-outputfh
data1d<-rbind(nomissingdata1,fa,fb,fc,fd,fe,ff,fg,fh)
write.table(data3f,"M:\\NCDSModel2WeightedData.txt",sep="\t",col.names=T,row.names=F)
```

DIFFERENCES BETWEEN MISSING AND NON-MISSING GROUPS

```
datatotmiss<-subset(data1,data1$Complete==0)
datatotseen<-subset(data1,data1$Complete==1)
```

MEAN & STANDARD DEVIATION OF RELATIVE HAND SKILL

```
datatotmissDOHnoNA<-subset(datatotmiss,is.na(datatotmiss$DOH)==FALSE)
datatotseenDOHnoNA<-subset(datatotseen,is.na(datatotseen$DOH)==FALSE)
mean(datatotmissDOHnoNA$DOH)
sd(datatotmissDOHnoNA$DOH)
mean(datatotseenDOHnoNA$DOH)
sd(datatotseenDOHnoNA$DOH)
```

GROUP NUMBERS FOR DISCRETE VARIABLES

SH

```
datatotmiss1<-subset(datatotmiss,datatotmiss$SH==0)
datatotmiss2<-subset(datatotmiss,datatotmiss$SH==1)
length(datatotmiss1$number)
length(datatotmiss2$number)
datatotseen1<-subset(datatotseen,datatotseen$SH==0)
datatotseen2<-subset(datatotseen,datatotseen$SH==1)
length(datatotseen1$number)
length(datatotseen2$number)
```

WH

```
datatotmiss1<-subset(datatotmiss,datatotmiss$WH==0)
datatotmiss2<-subset(datatotmiss,datatotmiss$WH==1)
length(datatotmiss1$number)
length(datatotmiss2$number)
datatotseen1<-subset(datatotseen,datatotseen$WH==0)
datatotseen2<-subset(datatotseen,datatotseen$WH==1)
length(datatotseen1$number)
length(datatotseen2$number)
```

GENDER

```
datatotmiss1<-subset(datatotmiss,datatotmiss$gender==0)
datatotmiss2<-subset(datatotmiss,datatotmiss$gender==1)
length(datatotmiss1$number)
length(datatotmiss2$number)
datatotseen1<-subset(datatotseen,datatotseen$gender==0)
datatotseen2<-subset(datatotseen,datatotseen$gender==1)
length(datatotseen1$number)
length(datatotseen2$number)
```

UK REGION

```
datatotmiss1<-subset(datatotmiss,datatotmiss$UK_region==1)
datatotmiss2<-subset(datatotmiss,datatotmiss$UK_region==2)
datatotmiss3<-subset(datatotmiss,datatotmiss$UK_region==3)
datatotmiss4<-subset(datatotmiss,datatotmiss$UK_region==4)
length(datatotmiss1$number)
length(datatotmiss2$number)
length(datatotmiss3$number)
length(datatotmiss4$number)
```

```
datatotseen1<-subset(datatotseen,datatotseen$UK_region==1)
datatotseen2<-subset(datatotseen,datatotseen$UK_region==2)
datatotseen3<-subset(datatotseen,datatotseen$UK_region==3)
datatotseen4<-subset(datatotseen,datatotseen$UK_region==4)
length(datatotseen1$number)
length(datatotseen2$number)
length(datatotseen3$number)
length(datatotseen4$number)
```

CHI-SQUARE & MANN-WHITNEY U TESTS

RELATIVE HAND SKILL

```
wilcox.test(datatotseen$DOH,datatotmiss$DOH,paired=FALSE,na.action=na.omit)
```

SH

```
aa<-table(data1$Complete,data1$SH)
chisq.test(aa)
```

WH

```
aa<-table(data1$Complete,data1$WH)
chisq.test(aa)
```

GENDER

```
aa<-table(data1$Complete,data1$gender)
chisq.test(aa)
```

UK REGION

```
aa<-table(data1$Complete,data1$UK_region)
chisq.test(aa)
```

MISSING DATA ANALYSIS – MC DATASET

LOGISTIC REGRESSION MODELS

```
model1<-glm(complete~age,data=data2a,family=binomial)
summary(model1)
model2<-glm(complete~gender,data=data2a,family=binomial)
summary(model2)
model3<-glm(complete~music,data=data2a,family=binomial)
summary(model3)
model4<-glm(complete~handedness,data=data2a,family=binomial)
summary(model4)
model5<-glm(complete~height,data=data2a,family=binomial)
summary(model5)
model6<-glm(complete~weight,data=data2a,family=binomial)
summary(model6)
finalmodel<-glm(complete~1,data=data2a,family=binomial)
summary(finalmodel)
```

DIFFERENCES BETWEEN MISSING AND NON-MISSING GROUPS

```
datatotmiss<-subset(data2a,data2a$complete==0)
datatotseen<-subset(data2a,data2a$complete==1)
```

MEAN & STANDARD DEVIATION OF AGE

```
datatotmissagenoNA<-subset(datatotmiss,is.na(datatotmiss$age)==FALSE)
datatotseenagenoNA<-subset(datatotseen,is.na(datatotseen$age)==FALSE)
mean(datatotmissagenoNA$age)
sd(datatotmissagenoNA$age)
mean(datatotseenagenoNA$age)
sd(datatotseenagenoNA$age)
```

MEAN & STANDARD DEVIATION OF HEIGHT

```
datatotmissheightnoNA<-subset(datatotmiss,is.na(datatotmiss$height)==FALSE)
datatotseenheightnoNA<-subset(datatotseen,is.na(datatotseen$height)==FALSE)
mean(datatotmissheightnoNA$height)
sd(datatotmissheightnoNA$height)
mean(datatotseenheightnoNA$height)
sd(datatotseenheightnoNA$height)
```

MEAN & STANDARD DEVIATION OF WEIGHT

```
datatotmissweightnoNA<-subset(datatotmiss,is.na(datatotmiss$weight)==FALSE)
datatotseenweightnoNA<-subset(datatotseen,is.na(datatotseen$weight)==FALSE)
mean(datatotmissweightnoNA$weight)
sd(datatotmissweightnoNA$weight)
mean(datatotseenweightnoNA$weight)
sd(datatotseenweightnoNA$weight)
```

GROUP NUMBERS FOR DISCRETE VARIABLES

MUSICIAN

```
datatotmiss1<-subset(datatotmiss,datatotmiss$music==0)
datatotmiss2<-subset(datatotmiss,datatotmiss$music==1)
length(datatotmiss1$id)
length(datatotmiss2$id)
datatotseen1<-subset(datatotseen,datatotseen$music==0)
datatotseen2<-subset(datatotseen,datatotseen$music==1)
length(datatotseen1$id)
length(datatotseen2$id)
```

GENDER

```
datatotmiss1<-subset(datatotmiss,datatotmiss$gender==0)
datatotmiss2<-subset(datatotmiss,datatotmiss$gender==1)
length(datatotmiss1$id)
length(datatotmiss2$id)
datatotseen1<-subset(datatotseen,datatotseen$gender==0)
datatotseen2<-subset(datatotseen,datatotseen$gender==1)
length(datatotseen1$id)
length(datatotseen2$id)
```

FISHER EXACT & MANN-WHITNEY U TESTS

AGE

```
wilcox.test(datatotseen$age,datatotmiss$age,paired=FALSE,na.action=na.omit)
```

HEIGHT

```
wilcox.test(datatotseen$height,datatotmiss$height,paired=FALSE,na.action=na.omit)
```

WEIGHT

```
wilcox.test(datatotseen$weight,datatotmiss$weight,paired=FALSE,na.action=na.omit)
```

MUSICIAN

```
aa<-table(data2a$complete,data2a$music)
fisher.test(aa)
```

GENDER

```
aa<-table(data2a$complete,data2a$gender)
fisher.test(aa)
```

APPENDIX B

Publication

The effect of handedness on academic ability: A multivariate linear mixed model approach

Christopher P. Cheyne [a]; Neil Roberts [b]; Tim J. Crow [c]; Stuart J. Leask [d]; Marta García-Fiñana [a]

[a] University of Liverpool, UK; [b] University of Edinburgh, UK; [c] University of Oxford, UK;
[d] University of Nottingham, UK

LATERALITY, 2010, 15(4), pp. 451-464

LATERALITY, 2010, 15 (4), 451–464

Psychology Press
Taylor & Francis Group

The effect of handedness on academic ability: A multivariate linear mixed model approach

Christopher P. Cheyne
University of Liverpool, UK

Neil Roberts
University of Edinburgh, UK

Tim J. Crow
University of Oxford, UK

Stuart J. Leask
University of Nottingham, UK

Marta García-Fiñana
University of Liverpool, UK

In recent years questions have arisen about whether there are any links between handedness and academic abilities as well as other factors. In this study we investigate the effects of gender, writing hand, relative hand skill, and UK region on mathematics and reading test scores by applying a multivariate linear mixed-effects model. A data sample based on 11,847 11-year-old pupils across the UK from the *National Child Development Study* was considered for the analysis. Our results show that pupils who write with one hand while having better skill with their other hand (i.e., inconsistent writing hand and superior hand) obtained lower test scores in both reading and mathematics than pupils with consistent writing hand and superior hand. Furthermore, we confirm previous findings that degree of relative hand skill has a significant effect on both reading and maths scores and that this association is not linear. We also found higher scores of reading in children from the south of England, and of mathematics in children from the south of England and Scotland, when compared to other UK regions.

Address correspondence to: Dr Marta García-Fiñana, Centre for Medical Statistics and Health Evaluation, Shelley's Cottage, University of Liverpool, Liverpool L69 3GS, UK. E-mail: martaf@liv.ac.uk

The authors would like to thank Dr Vanessa Sluming and Dr Patty Cowell for earlier useful discussions. We also thank the MRC DTA and the Institute of Human Ageing of the University of Liverpool for funding this project, and the Centre For Longitudinal Studies for access to the NCDS dataset.

http://www.psypress.com/laterality DOI: 10.1080/13576500902976956

Keywords: Multivariate analysis; Handedness; Maths scores; Mixed-effects model; Reading scores; Writing hand.

The effect of handedness and hand skill on academic ability has been a subject of discussion in the last few years (Annett & Manning, 1990; Crow, Crow, Done, & Leask, 1998; Denny, 2008; Faurie, Vianey-Liaud, & Raymond, 2006; Nettle, 2003; Peters, Reimers, & Manning, 2006). For example, a possible association between hand skill and academic ability was investigated by analysing data from a group of 11-year-old children (Crow et al., 1998). A laterality index variable, referred to as *relative hand skill*, was used, and the value zero was defined as the "point of hemispheric indecision"; that is, the point at which there is no difference in hand skill between the right and left hand; (see McManus, 1985, for laterality index definition). The results suggested that pupils with laterality index zero performed worst in the reading test, with a negligible decrease from the peak score for right-hand extreme values (which involved 10% of pupils with the highest relative hand skill). Also, for maths scores, pupils with relative hand skill at the point of hemispheric indecision achieved the lowest test scores, and there was also a substantial decrease from the peak score for right-hand extreme values. Furthermore, the relationship between reading test scores and *hand skill* (defined as the absolute difference in skill scores between the right and left hand) has been described as being quadratic (Annett & Manning, 1990).

The significance of handedness and academic ability to predict mental illness has also been investigated (Crow, 2000; Crow, Done, & Sacker, 1995; Elias, Saucier, & Guylee, 2001; Grace, 1987; Klar, 1999; Merrin, 1984; Sommer, Ramsey, Kahn, Aleman, & Bouma, 2001). Crow (2000) suggests that pupils with relative hand skill close to zero (i.e., ambidextrous) are more likely to develop schizophrenia or psychosis at a later stage in their life, irrespective of whether they are left- or right-handed. In particular, of those pupils diagnosed with schizophrenia or psychosis later in their life, the percentage of ambidextrous pupils was approximately 22.4–25.8%, while in the control group the percentage was reduced to approximately 6.9%. Furthermore, pupils diagnosed with schizophrenia later in their life had performed worse in reading and mathematics tests at ages 7 and 11 than those pupils who were not diagnosed with schizophrenia (Crow et al., 1995). Also, pre-psychotic pupils performed worse than pupils from the control group, but only at age 7 for both reading and mathematics.

In 2005, a BBC online Internet survey was conducted where 255,100 participants were given a mental rotation task to test their spatial performance, as well as a fluency/reasoning task which involved listing a number of words linked to a given word (Peters et al., 2006). The majority of those participants who performed worst in the spatial performance task appeared to be those with no preference for handedness. Also, participants who were either extremely left- or right-handed tended to show deficits when compared to those who were mostly left- or right-handed for the spatial task. The extreme handedness group performed worst in the fluency/reasoning task, followed by those with no preference for handedness.

The studies mentioned above give an insight into the connection of factors, such as age, gender, and handedness on academic ability. However, these factors have not been jointly considered in a unique statistical model, but instead the effect of each factor has been investigated individually. In this paper we aim to analyse the effect of gender, relative hand skill, superior hand (defined as the most skilful hand), writing hand, and region of the UK mainland on reading and mathematics test scores by applying a multivariate linear mixed-effects model to the *National Child Development Study* dataset (age 11), the same dataset as used in the aforementioned studies. In particular, the correlation between reading and maths scores, and potential interactions between the factors mentioned above, are taken into account in the analysis.

METHOD

The *National Child Development Study* (NCDS) was set up to investigate the main factors that affect human development based on longitudinal data on children born on the UK mainland in one particular week in 1958 (3–9 March). The NCDS then involved following up the cohort at various ages over the course of their lifetime, including at ages 7, 11, and 16, and it is still continuing at present. These data are stored by and can be obtained from the Centre for Longitudinal Studies. In this study we considered data taken from 11,847 11-year-old children (i.e., the data were collected in 1969). In particular, the outcomes considered are reading and mathematics test scores. The reading test comprised 35 comprehension questions, which the child had to answer in 20 minutes (Crow et al., 1998). These questions consisted of choosing the correct word from five alternatives, which completed the sentence given. The mathematics test consisted of 40 questions involving geometry, arithmetic, and logic to be answered in 40 minutes. For ease of interpretation, both reading and maths scores were converted into percentages. Pupils' confidentiality is paramount in the study, so in the dataset each pupil had an ID number assigned and all identifiable data removed.

The mathematics test scores do not appear to be normally distributed. For the assumption of normality to hold in the model, the mathematics test scores were transformed. The chosen transformation was a Box-Cox transformation of the form:

$$y^* = \frac{((y+3)^\lambda - 1)}{\lambda} \tag{1}$$

(Box & Cox, 1964) where $\lambda = 0.512$ and y^* and y are the transformed and original mathematics test scores, respectively.

A multivariate linear mixed model (LMM) approach was applied to the dataset (for a detailed description of the theory behind this approach, see Laird & Ware, 1982; Sammel, Lin, & Ryan, 1999). In particular, the model chosen took the following form:

$$Y_k^{(i)} = X_k\beta^{(i)} + z_k b_k^{(i)} + \varepsilon_k^{(i)} \tag{2}$$

where i denotes the test type ($i = 1$ and $i = 2$ refer to the reading and maths scores, respectively). The array $Y_k^{(i)}$ represents the test scores for the k^{th} local authority of the i^{th} test, and it can be expressed as:

$$Y_k^{(i)} = \begin{bmatrix} Y_1^{(i)} \\ \vdots \\ Y_{n_k}^{(i)} \end{bmatrix} \tag{3}$$

where n_k represents the number of pupils within the k^{th} local authority (k = 1, 2, ..., 185).

The fixed effect design matrix for the k^{th} local authority is:

$$X_k = \begin{bmatrix} X_{1,1} & \cdots & X_{1,10} \\ \vdots & \ddots & \vdots \\ X_{n_k,1} & \cdots & X_{n_k,10} \end{bmatrix} \tag{4}$$

where $X_{l,j}$ represents the j^{th} fixed effect value for the l^{th} pupil (l = 1, 2, ..., n_k; j = 1, 2, ..., 10). In particular, the fixed effects refer to the intercept ($X_{l,1}$), relative hand skill ($X_{l,2}$), (relative hand skill)^2 ($X_{l,3}$), superior hand ($X_{l,4}$), writing hand ($X_{l,5}$), superior hand and writing hand interaction ($X_{l,6} = X_{l,4}*X_{l,5}$), gender ($X_{l,7}$), and region (4-level factor: $X_{l,8}$ $X_{l,9}$ $X_{l,10}$). The array $\beta^{(i)}$ represents the fixed effect coefficients for the i^{th} test, which will be estimated from the dataset ($i = 1,2$) and it can be expressed as:

$$\beta^{(i)} = \begin{bmatrix} \beta_1^{(i)} \\ \vdots \\ \beta_{10}^{(i)} \end{bmatrix} \tag{5}$$

Gender was a binary variable with value 0 for male and 1 for female. Writing hand (WH) was also binary with value 0 for right writing hand, and 1 for left writing hand. The region variable consisted of northern England and the Midlands (Region 1), southern England (Region 2), Wales (Region 3), and Scotland (Region 4). This region variable was entered as a factor with Region 1 as the reference region. Relative hand skill was calculated as:

$$\text{Relative Hand Skill} = \left| \frac{\text{n(RH)} - \text{n(LH)}}{\text{n(RH)} + \text{n(LH)}} \times 100 \right| \tag{6}$$

where n(RH) and n(LH) denote the number of boxes ticked with the right and left hand in a 1-minute marking task, respectively (McManus, 1985). Relative hand skill can be interpreted as the relative superiority of the dominant hand over the non-dominant hand. Superior hand (SH) is a binary variable calculated from the hand skill variable. If the dominant hand is the left hand (i.e., the number of boxes marked by the left hand is greater than the number of boxes marked by the right hand), SH is equal to 1, and otherwise, SH is equal to 0.

The random effects term z_k is the random intercept $(1,1)^T$. The random effects vector b_k is assumed to follow a bivariate normal distribution, which takes into account the variability between local authorities. We have:

$$b_k = \begin{pmatrix} b_k^{(1)} \\ b_k^{(2)} \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{r_{(1)}} & \sigma_{r_{(1)}r_{(2)}} \\ \sigma_{r_{(1)}r_{(2)}} & \sigma^2_{r_{(2)}} \end{pmatrix}\right) \tag{7}$$

where $\sigma^2_{r_{(i)}}$ is the random effect variance for the i^{th} test and $\sigma_{r_{(1)}r_{(2)}}$ is the random effect covariance between the two tests. Finally, the random errors are assumed to follow the bivariate normal distribution:

$$\begin{pmatrix} \varepsilon_k^{(1)} \\ \varepsilon_k^{(2)} \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{s_{(1)}} & \sigma_{s_{(1)}s_{(2)}} \\ \sigma_{s_{(1)}s_{(2)}} & \sigma^2_{s_{(2)}} \end{pmatrix}\right) \tag{8}$$

where $\sigma^2_{s(1)}$ is the random error variance of the i^{th} test and $\sigma_{s_{(1)}s_{(2)}}$ is the random error covariance between the two tests. The LMM analysis was conducted using MLwiN version 2.02 and the plots using R version 2.6.2.

RESULTS

Results of the multivariate analysis are given in Tables 1 and 2. We found no evidence to suggest that there is a difference between boys and girls in either the reading or the mathematics test scores ($p = .77$ and $p = .12$, respectively). Alternatively, relative hand skill has a statistically significant effect on both the reading and mathematics test scores ($p < .001$ in both cases). This effect is quadratic, as can be appreciated in Figure 1 and Table 1. The interaction terms between superior hand and writing hand are also significant ($p < .001$). The effect of superior hand is statistically significant for both reading and mathematics test scores (negative coefficients, $p < .001$), which suggests that for right-hand writers (reference group) those with left superior hand performed worse than those with right superior hand. In particular, the mean difference was 6.2% for the reading test scores. Furthermore, for left-hand writers, those with left superior hand performed on average significantly

TABLE 1
Results of the multivariate linear mixed model with fixed effects coefficients

Fixed effect	*Coefficient*	*St. error*	*Lower 95% CI*	*Upper 95% CI*	*p-value*
Reading test scores					
Relative hand skill	0.423	0.058	0.308	0.538	<.001
(Relative hand skill)^2	−0.009	0.001	−0.012	−0.006	<.001
Superior hand (SH) (0: Right, 1: Left)	−6.213	1.175	−8.516	−3.910	<.001
Writing hand (WH) (0: Right, 1: Left)	−5.011	1.773	−8.485	−1.537	.005
SH × WH	10.553	2.202	6.238	14.868	<.001
Gender *(0: Male, 1: Female)*	0.097	0.326	−0.542	0.736	.766
Northern England & Midlands			*Reference region*		
Southern England	1.620	0.389	0.857	2.383	<.001
Wales	−0.771	0.791	−2.322	0.780	.330
Scotland	−0.354	0.576	−1.482	0.774	.538
Transformed mathematics test scores					
Relative hand skill	0.123	0.014	0.095	0.151	<.001
(Relative hand skill)^2	−0.002	<0.0005	−0.003	−0.001	<.001
Superior hand (SH) (0: Right, 1: Left)	−1.556	0.283	−2.111	−1.001	<.001
Writing hand (WH) (0:Right, 1:Left)	−1.316	0.427	−2.153	−0.479	.002
SH × WH	2.703	0.531	1.663	3.743	<.001
Gender *(0: Male, 1: Female)*	−0.121	0.079	−0.275	0.033	.122
Northern England & Midlands			*Reference region*		
Southern England	0.210	0.093	0.028	0.392	.023
Wales	0.338	0.187	−0.029	0.705	.072
Scotland	0.502	0.137	0.234	0.770	.001

p-values refer to t-values that are based on the square roots of the F-values.

better than those with right superior hand. To confirm this last statement, the corresponding hypothesis test was conducted for each of the two test scores (see Table 3, $p = .018$ and $p = .009$, respectively).

Furthermore, the effect of writing hand is statistically significant for both reading and mathematics test scores (negative coefficients; $p = .005$ and $p = .002$, respectively), which suggests that for pupils with right superior hand (reference group) left-hand writers performed worse than right-hand writers (and this difference was 5.0%, on average, for the reading test scores; see Table 1). Finally, for pupils with left superior hand, left-hand writers achieved higher scores on average than right-hand writers (Table 3, both $p = < .001$). The above results reveal that there is a discrepancy in test scores

TABLE 2
Random effects and random error coefficients from the multivariate linear mixed model

	Estimate	*St. error*	*p-value*
Random effect			
Intercept (Reading Scores)	24.005	3.799	<.001
Intercept (Transformed Maths Scores)	1.078	0.211	<.001
Covariance (Reading/Transformed Maths Scores)	4.933	0.806	<.001
Random error			
Intercept (Reading Scores)	292.043	5.073	–
Intercept (Transformed Maths Scores)	17.241	0.295	–
Covariance (Reading/Transformed Maths Scores)	52.125	1.079	<.001

between pupils with inconsistent writing hand and superior hand and pupils with consistent writing hand and superior hand.

Alternatively, our analysis reveals that left-hand writers with left superior hand have mean test scores, in both reading and mathematics, that are no different from those of right-hand writers with right superior hand (see Table 4, $p = .21$ and .19, respectively). Although our data did not provide evidence to suggest a difference in performance between *consistent* left- and *consistent* right-handed pupils for large values of relative hand skill, we should bear in mind that the sample size of left-handed pupils is relatively small (for example, there are only 23 left-handed pupils with absolute relative hand skill greater than 35). Furthermore, there is a significant difference in both reading and maths scores between pupils in southern England and pupils in northern England and the Midlands (see Table 1; $p < .001$ and $p = .023$, respectively). In the reading tests pupils from southern England performed, on average, 1.6% better than their counterparts in northern England and the Midlands, and 2.4% and 2.0% better than pupils from Wales and Scotland, respectively. Furthermore it is interesting to note that pupils in Scotland and southern England achieved higher maths scores, on average, than those pupils in northern England and the Midlands (see Table 1; $p = .001$ and $p = .023$, respectively).

The random effect intercepts represent the random variation at local authority level in both the reading and transformed maths scores (see Table 2). The Pearson correlation coefficient between the reading and transformed maths scores at the local authority level was 0.971. Therefore local authorities, in general, had similar proportions of pupils performing well on both tests and, conversely, of pupils doing badly on both tests. Similarly, the random error intercepts, which are the random variation at pupil level in the reading and transformed maths scores, are given in Table 2. The Pearson correlation coefficient here was 0.735, which suggests that there is also a strong positive correlation between the two test scores. Precisely, there is

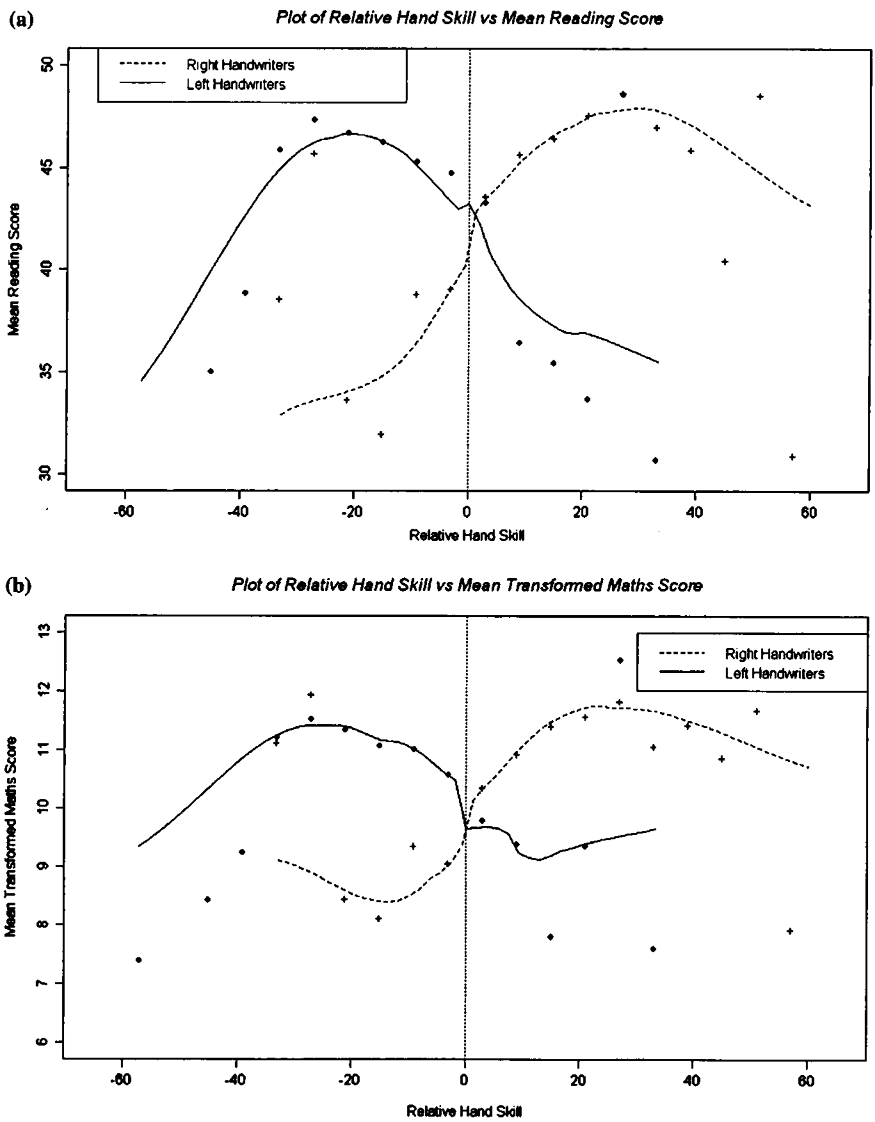


Figure 1. Plot of relative hand skill against (a) mean reading score and (b) mean transformed mathematics score by writing hand, where pupils with left SH appear on the left-hand side of the graphic with negative relative hand skill values. The curves were obtained by using the "Loess" smoothing function. The dots and crosses represent the mean values for 20 equidistant intervals across the range of relative hand skill (dots for left-hand writers, and crosses for right-hand writers).

evidence to suggest that pupils who performed well on one of the tests also tended to do well on the other test, and conversely, those pupils who did not perform well on one of the tests did not perform well on the other test. Plots

TABLE 3

Results of the hypothesis test for differences between right and left superior handers, and between right- and left-hand writers, in both the reading and transformed mathematics test scores

WH	*SH*	*Test*	*Estimate*	*Standard error*	*Lower 95% CI*	*Upper 95% CI*	*p-value*
Right	Right vs Left	Reading	−6.213	1.175	−8.516	−3.910	<.001
		Transformed Mathematics	−1.556	0.283	−2.111	−1.001	<.001
Left	Right vs Left	Reading	4.340	1.834	0.746	7.934	.018
		Transformed Mathematics	1.147	0.265	0.281	2.013	.009
Right vs Left	Right	Reading	−5.011	1.772	−8.485	−1.537	.005
		Transformed Mathematics	−1.316	0.427	−2.153	−0.479	.002
Right vs Left	Left	Reading	5.542	1.263	3.066	8.018	<.001
		Transformed Mathematics	1.387	0.305	0.790	1.984	<.001

TABLE 4
Results of the hypothesis test for differences between right superior, right-hand writers and left superior, left-hand writers, in both the reading and transformed mathematics test scores

Test scores	*Estimate*	*Standard error*	*Lower 95% CI*	*Upper 95% CI*	*p-value*
Reading	−0.671	0.536	−1.721	0.379	.211
Transformed Mathematics	−0.169	0.129	−0.422	0.084	.191

of relative hand skill against mean reading and transformed maths scores are given in Figure 1(a) and 1(b), respectively. The crosses and dots represent mean values for equidistant intervals of relative hand skill for right-hand and left-hand writers, respectively. Some of these empirical mean values differ from the values of the predicted curve, and this is caused by the small sample sizes that are available for those regions (e.g., for right-hand writers, there were 10,313 pupils with right superior hand compared to 313 with left superior hand, and for left-hand writers there were 1252 pupils with left superior hand and only 103 with right superior hand). Therefore, in the regions where there are more data, the curve provides a better fit to the data than in the regions where there are relatively few data. It is interesting to note here that although there are more right-hand writers with left superior hand than left-hand writers with right superior hand, this reverses when expressed as percentages of the total sample of right- and left-hand writers, with 2.3% of right-hand writers having left superior hand compared to 7.7% of left-hand writers having right superior hand. Scores for pupils who have a left superior hand but are right-hand writers are represented by the dashed curve where relative hand skill is less than zero. Scores for left-hand writers with right superior hand can be identified by the solid curve where relative hand skill is greater than zero. Scores for pupils with consistent writing hand and superior hand are included in the solid curve when relative hand skill is less than zero (i.e., left superior hand) for left-hand writers and in the dashed curve when relative hand skill is greater than zero (i.e., right superior hand) for right-hand writers.

We have observed that, for individuals with congruent hand performance and writing hand and with absolute values of relative hand skill greater than 10, the discrepancy in score between the hands seems to be caused by the skill of the non-writing hand. In particular the non-writing hand performed worse than the writing hand, and this difference becomes more pronounced as the relative hand skill increases, while the writing hand seems to maintain a skill level close to the average for the writing hand, irrespective of whether the pupil is right- or left-hand writer. Meanwhile, as relative hand skill decreases from 10 towards 0 for right-hand writers, the number of boxes ticked by the writing hand decreases so that it is below the average, whereas the non-writing hand increases slightly and it is above the average at all values of relative hand skill in this range. A similar behaviour is detected with left-hand writers although there are some fluctuations. In particular, for relative hand skill equal to zero, right-hand writers ticked on average 68.91 boxes per minute with each hand. This is slightly above the average for the non-writing hand (66.67), but much lower than the average for the writing hand (92.53). Left-hand writers with relative hand skill equal to zero ticked, on average, 61.31 with each hand, which is below the average for both the writing and non-writing hand (92.59 and 67.85 respectively).

DISCUSSION

In summary, our study suggests that pupils with inconsistent writing hand and superior hand performed worse in both reading and maths tests than pupils with consistent writing hand and superior hand. There is a strong, positive correlation between the two test scores, suggesting that pupils who achieved a high score or a low score in one test were more likely to achieve an equivalent high or low score, respectively, in the other test. We found no difference in test scores attributed to gender differences.

As the relative hand skill increases (decreases) from 0 (ambidexterity) towards a peak at around 25 to 35 (-25 to -35) in pupils with consistent writing and superior hand, both reading and mathematics mean test scores increase, but as the relative hand skill increases from about 35 to 60 (-35 to -60), both scores decrease (Figure 1). A previous study, where all pupils irrespective of writing hand were included, suggested that the optimum relative hand skill for verbal test scores was approximately 20 for right-handed pupils and approximately -30 for left-handed pupils (Leask & Crow, 2006). The fact that our results regarding the optimum hand skill for the consistent group are similar to these results can be explained since there are relatively few pupils with inconsistent writing hand and superior hand compared to those with consistent writing hand and superior hand. However, a similar trend cannot be identified for the pupils with inconsistent superior hand and writing hand in either test (see Figure 1). Further work is needed in this area to investigate whether there is some underlying factor that influences the distribution of these scores. Also, the biological and/or environmental predictors for which pupils in the consistent group with relative hand skill in the intervals (-60, -35), (-20, 20), and (35, 60) obtained lower scores on both tests than pupils with relative hand skill in the intervals (−35, −20) and (20, 35) remain unclear.

We found a significant difference between pupils from southern England and those from northern England and the Midlands in both reading and maths scores, with those from southern England performing better. Speculative reasons for this include the perceived north–south divide in the UK. For many decades people living in the north have had lower incomes and lower standards of living than those living in the south of England, particularly the south-east (Blackaby & Manning, 1990). This has been passed on in the form of education since people with greater disposable income could afford better quality schools. Pupils in Scotland also outperformed their counterparts from northern England and the Midlands in the mathematics test, and this could have been caused by subtle differences in the education system between England and Scotland in the 1960s. Nevertheless, the differences observed between regions, although statistically significant, might not be large enough to be regarded as important in academic terms.

Questions have arisen as to whether a task such as box-marking, which is related to writing, can be used as a sufficiently independent measure of hand skill. An alternative for this could have been the peg-moving task (Annett, 2002) or another such task, which does not involve the pupil writing in any way. Purdue's peg test consists of timing how quickly a person can move a quantity of pegs from one hole to another, and therefore, the variable *relative hand skill* used in this paper (and which is based on the box-marking task) is not strictly comparable to another investigation where hand skill is defined based on the pegboard task. Unfortunately, for this particular dataset no alternative tasks for the purpose of obtaining an individual pupil's hand skill were conducted in 1969.

Several studies have reported an association between handedness and brain asymmetry (Amunts et al., 1996; Annett, 2002). In particular, Broca's area, which is linked to language processing, may be affected or have an effect on handedness (Corballis, 2003; Liégeois et al., 2004; Tzourio, Crivello, Mellet, Nkanga-Ngila, & Mazoyer, 1998). It is also possible that this change in asymmetry is a function of gender (e.g., Annett, 2002). Further investigation based on a different cohort, where volume asymmetry of brain structures is added into the model, is currently being considered. A recent study has investigated the influence of family income and parent's education levels on the maths and reading test scores from the NCDS (Michael, 2008). These factors could be used to extend this model further.

Manuscript received 22 December 2008
Revised manuscript received 14 April 2009
First published online 17 June 2009

REFERENCES

Amunts, K., Schlaug, G., Schleicher, A., Steinmetz, H., Dabringhaus, A., Roland, P. E., et al. (1996). Asymmetry in the human motor cortex and handedness. *Neuroimage, 4*, 216–222.

Annett, M. (2002). *Handedness and brain asymmetry: The right shift theory*. Hove, UK: Psychology Press.

Annett, M., & Manning, M. (1990). Reading and a balanced polymorphism for laterality and ability. *Journal of Child Psychology and Psychiatry, 31*(4), 511–529.

Blackaby, D. H., & Manning, D. N. (1990). The north-south divide: Earnings, unemployment and cost of living differences in Great Britain. *Papers of the Regional Science Association, 69*, 43–55.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B (Methodological), 26*(2), 211–252.

Corballis, M. C. (2003). From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences, 26*(2), 199–208.

Crow, T. J. (2000). Schizophrenia as the price that homo sapiens pays for language: A resolution of the central paradox in the origin of the species. *Brain Research Reviews, 31*, 118–129.

Crow, T. J., Crow, L. R., Done, D. J., & Leask, S. (1998). Relative hand skill predicts academic ability: Global deficits at the point of hemispheric indecision. *Neuropsychologia, 36*(12), 1275–1282.

Crow, T. J., Done, D. J., & Sacker, A. (1995). Childhood precursors of psychosis as clues to its evolutionary origins. *European Archives of Psychiatry and Clinical Neuroscience, 245*, 61–69.

Denny, K. (2008). Cognitive ability and continuous measures of relative hand skill: A note. *Neuropsychologia, 46*(7), 2091–2094.

Elias, L. J., Saucier, D. M., & Guylee, M. J. (2001). Handedness and depression in university students: A sex by handedness interaction. *Brain and Cognition, 46*(1–2), 125–129.

Faurie, C., Vianey-Liaud, N., & Raymond, M. (2006). Do left-handed children have advantages regarding school performance and leadership skills? *Laterality, 11*(1), 57–70.

Grace, W. C. (1987). Strength of handedness as an indicant of delinquent's behaviour. *Journal of Clinical Psychology, 43*(1), 151–155.

Klar, A. J. (1999). Genetic models for handedness, brain lateralization, schizophrenia, and manic-depression. *Schizophrenia Research, 39*(3), 207–218.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*(4), 963–974.

Leask, S. J., & Crow, T. J. (2006). A single optimum degree of hemispheric specialisation in two tasks, in two UK national birth cohorts. *Brain and Cognition, 62*, 221–227.

Liégeois, F., Connelly, A., Cross, J. H., Boyd, S. G., Gadian, D. G., Vargha-Khadem, F., et al. (2004). Language reorganization in children with early-onset lesions of the left hemisphere: An fMRI study. *Brain, 127*(6), 1229–1236.

McManus, I. P. (1985). Right- and left-hand skill: Failure of the right shift model. *British Journal of Psychology, 76*, 1–16.

Merrin, E. L. (1984). Motor and sighting dominance in chronic schizophrenics. Relationship to social competence, age at first admission, and clinical course. *British Journal of Psychiatry, 145*, 401–406.

Michael, R. T. (2008). *Children's reading and math skills: The influence of family caring*. London: Centre for Longitudinal Studies.

Nettle, D. (2003). Hand laterality and cognitive ability: A multiple regression approach. *Brain and Cognition, 52*, 390–398.

Peters, M., Reimers, S., & Manning, J. T. (2006). Hand preference for writing and associations with selected demographic and behavioural variables in 255,100 subjects: The BBC Internet study. *Brain and Cognition, 62*, 177–189.

Sammel, M., Lin, X., & Ryan, L. (1999). Multivariate linear mixed models for multiple outcomes. *Statistics in Medicine*, *18*, 2479–2492.

Sommer, I., Ramsey, N., Kahn, R., Aleman, A., & Bouma, A. (2001). Handedness, language lateralisation and anatomical asymmetry in schizophrenia: Meta-analysis. *British Journal of Psychiatry*, *178*, 344–351.

Tzourio, N., Crivello, F., Mellet, E., Nkanga-Ngila, B., & Mazoyer, B. (1998). Functional anatomy of dominance for speech comprehension in left handers vs right handers. *Neuroimage*, *8*(1), 1–16.