

Guide to statistics in the Eurocodes for timber engineers

Mike Bather, April 2021

University of Liverpool

Contents

1	Introduction	1
2	Finding the mean of a sample	2
3	Finding the confidence interval around the mean of a sample	2
4	Characteristic value of MoE based on visual grading	3
5	Characteristic value of density based on visual grading.....	6
6	Characteristic value of MoR based on visual grading.....	8
7	References	9

1 Introduction

The Eurocodes use statistical approaches firstly to determine realistic values for loads or actions that are applied to structures and secondly to determine the values of the material properties of structural elements. This brief guide describes the statistical background needed to understand the determination of the characteristic values of the material properties of timber based on sample statistics. The key properties of modulus of elasticity (MoE), bending strength or modulus of rupture (MoR) and density are considered. Following this, the statistical approaches for the determination of material properties, used in the Eurocodes are explained.

EN 1990 (CEN, 2005) provides the basis of all of the structural Eurocodes, covering principles and requirements for reliability and safety and it covers the basis of structural design. A reliability based approach to failure is used which, for normal buildings requires the probability of ultimate limit state failure, during an assumed 50 year life, to be less than 10^{-5} . The probability of a serviceability limit state failure over the same period is much greater, being less than 10^{-2} and much closer to 10^{-1} .

To achieve these levels of reliability, a harmonised set of standards for testing of materials and products has been devised which is used with a harmonised set of standards for the design of new structures and buildings. Characteristic properties of

materials are determined, making use of the mean and 0.05 quantile. Characteristic actions or loads are determined, making use of the 0.95 quantile. In design, partial safety factors are applied to both materials and actions, and these factors change according to the limit state in question and the material in question.

Typical partial factors for permanent and variable actions are 1.35 and 1.5 respectively. A typical partial factor for the material properties of timber is 1.3.

In this way, the reliability of the entire structure under design is controlled. The above description is a useful basic introduction, but as is shown in this chapter, the determination of the characteristic material properties of timber is a little more complicated than this. Firstly, the final value of the material characteristic property is based on the lower bound of a confidence band drawn around the mean or the 0.05 quantile and secondly, some further reductions are made where low sample numbers and size are used in the testing programme for the determination of the material properties.

2 Finding the mean of a sample

Finding the mean of a sample is elementary, and so a sample of size n has a mean value \bar{y}

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (2.1)$$

This is useful to describe a sample, but in considering the mean of a population, it is useful to determine the confidence interval around the sample mean statistic, within which the population mean parameter will lie with a specified degree of probability.

3 Finding the confidence interval around the mean of a sample

For a normally distributed sample, to determine a confidence interval, it is necessary to determine the variance, the standard deviation, the standard error of the mean and an appropriate t-multiplier value. In the Eurocodes, it is assumed that MoE and density are distributed normally and that MoR is distributed lognormally (CEN, 2016a). So, for MoR, a normal distribution is assumed to be created by taking the natural logarithm of all MoR data points. In any case, it is worthwhile checking the distribution assumed for any particular dataset.

In the Eurocodes, the standard deviation of the sample is denoted s_y and is the square root of the variance. It gives an indication of the scatter of the data points. The formulae for variance s^2 and standard deviation s_y are found in statistical textbooks

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1} \quad (3.1)$$

$$s_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}} \quad (3.2)$$

Mean values vary from sample to sample and the standard deviation of a sampling distribution is termed the standard error of the mean (SEM). SEM is an indication of the uncertainty of the estimate of the mean and is given by

$$SEM = \frac{s_y}{\sqrt{n}} \quad (3.3)$$

A two sided confidence interval with a confidence level of say $\alpha = 95\%$ indicates that there is a 95% probability that a population parameter lies within the upper and lower limits of the interval. This accounts for the variation between the means of the samples and the mean of the population.

Its formula can be written, first in words

'sample mean value \pm (t – multiplier) \times standard error of the mean (SEM)'

$$MoE_{mean} = \bar{y} \pm t_{\alpha, n-1} \times \frac{\sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}}}{\sqrt{n}} \quad (3.4)$$

$t_{\alpha, n-1}$ is the α percentile of a central t-distribution with n-1 degrees of freedom and can be found from tables or software. The use of a central t-distribution is appropriate for estimates of the mean.

4 Characteristic value of MoE based on visual grading

Following the visual grading of a sample of timber joists, the characteristic values of the key mechanical and physical properties of each group of graded joists needs to be determined. This calculation follows three stages and the group of similarly graded joists is simply termed the 'sample' in the Eurocodes and represents that grade of timber in the population:

- (i) Calculate the sample mean value \bar{y}
- (ii) Calculate MoE_{mean} , which is the lower confidence limit of the mean value \bar{y} for $\alpha = 75\%$
- (iii) Adjust this lower confidence limit MoE_{mean} (according to the number and size of sub-samples) to obtain the characteristic modulus of elasticity $E_{0,mean}$

It is necessary to refer to three Eurocodes to accomplish the above, and the notation within each code is not perfectly consistent with the others. So, a notation is adopted here (based on the codes) which most clearly describes the process of obtaining the characteristic value of MoE needed for the strength classes in EN338 (CEN, 2016b).

In EN 14358, it is assumed that the MoE test values are normally distributed and that the sample of size n has a mean value \bar{y} similar to Equation (2.1)

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (4.1)$$

In EN14358, the standard deviation of the sample is denoted s_y and is the square root of the variance, similar to Equation (3.1)

$$s_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}} \quad (4.2)$$

The standard error of the mean (SEM) is the same as Equation (3.3) and is given by

$$SEM = \frac{s_y}{\sqrt{n}} \quad (4.3)$$

In EN14358, the value of the standard deviation, s_y , must not be taken as lower than a limit of $0.05 \bar{y}$. This is to force the grader to keep above a lower limit for the size of variance of timber in a sample.

The lower confidence limit MoE_{mean} is determined at “...a confidence level of $\alpha = 75\%$, where the confidence level α is defined as the probability of which the characteristic value is greater than the estimator on the characteristic value” (CEN, 2016a). This accounts for the variation between the means of the samples and the mean of the population. The single sided confidence limit of $\alpha = 75\%$ is the same as the lower confidence limit of a two sided confidence band with a level of $\alpha = 50\%$. From Section 3, the formula for the former can be written, first in words and then using the equation in Section 3. It can then be manipulated to create a middle term, which is how this formula is presented in the code

'sample mean value – (t – multiplier) × standard error of the mean (SEM)'

$$MoE_{mean} = \bar{y} - t_{\alpha, n-1} \times \frac{\sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}}}{\sqrt{n}} \quad (4.4)$$

$$MoE_{mean} = \bar{y} - \frac{t_{\alpha, n-1}}{\sqrt{n}} \times \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}} \quad (4.5)$$

In this instance, working with a lower 75% confidence limit, $t_{\alpha, n-1}$ is the α percentile in a central t-distribution with n-1 degrees of freedom and can be found from tables or software. Rather than present the statistics given above, EN14358 provides a simplified approach, substituting the middle term of Equation (4.5) with a new term $k_s(n)$ and giving both tabulated values and a simple formula for its calculation.

$$k_s(n) = \frac{t_{\alpha, n-1}}{\sqrt{n}} \quad (4.6)$$

$$MoE_{mean} = \bar{y} - k_s(n) \times s_y \quad (4.7)$$

Thus, three ways are given for calculating MoE_{mean} .

Finally, Equation (12) of EN384 (CEN, 2018, p. 12) requires one further adjustment to obtain $E_{0,mean}$ for each of the visual grades. This adjustment accounts for the variations between the means of samples and the mean of a population. It is this value that is compared with the tabulated stiffness values in Table 1 of EN338 (CEN, 2016a) to obtain the Strength Class.

$$E_{0,mean} = \min \left[1.1 \bar{E}_{i,\min}, \frac{\sum_{i=1}^{n_s} n_i \bar{E}_i}{n} \right] \times \frac{k_n}{0.95} \quad (4.8)$$

\bar{E}_i is the mean value of modulus of elasticity of sub-sample i

$\bar{E}_{i,\min}$ is the lowest mean value of modulus of elasticity of all i – sub-samples

n_s is the number of sub-samples

n_i is the number of specimens in sub-sample i

n is the total number of specimens

k_n is the factor to adjust for the number of sub-samples and this is taken from a table presented in EN384 (Table 1).

Table 4.1. Extracted from EN384, Table 1: Factor to adjust test results to the number of sub-samples

Number of sub-samples	1	2	3	4	5 or more
k_n for modulus of elasticity and density	0.88	0.91	0.94	0.97	1.00
k_n for strengths parallel to the grain	0.70	0.80	0.90	0.95	1.00

Thus, for modulus of elasticity and for a single sample of $n_s = 1$, $k_n = 0.88$, the first term in brackets, $\bar{E}_{i,\min}$, is the same as the sample mean and the second term, $\frac{\sum_{i=1}^{n_s} n_i \bar{E}_i}{n}$, is also the same as the sample mean and so, as the first term is multiplied by 1.1, the minimum of the two must be simply the sample mean with no adjustment. Therefore

$$E_{0,mean} = MoE_{mean} \times \frac{0.88}{0.95} = MoE_{mean} \times 0.926$$

The final adjustment to determine $E_{0,mean}$ is a direct outcome of a consideration of sampling and the homogeneity of a population of timber joists. How populations and sub-samples are defined is open to wide interpretation. For a characteristic value determined from multiple sub-samples, two approaches could have been adopted by the code:

- i. Use a lower confidence limit $\frac{\sum_{i=1}^{n_s} n_i \bar{E}_i}{n}$ based on all the test pieces together from all the sub-samples

- ii. Use the lowest of the separate lower confidence limits $\bar{E}_{i,\min}$ of each of the sub-samples

The code combines these two approaches and adds 10% to the lowest lower confidence limit $\bar{E}_{i,\min}$ of the separate sub-samples. The choice of how to apply this adjustment is based on judgement rather than statistical analysis (Ravenshorst, 2015). In a study in which graded sub-samples are not used primarily due to, for instance, small sample size, then the greater potential precision of outcomes is outweighed by the degree of uncertainty introduced into the analysis.

The final adjustment to determine $E_{0,mean}$, for a single sample, with no sub-sampling, introduces a 7% reduction in stiffness. It is questionable whether this reduction is appropriate for a single large sample, however, it is clearly a requirement of the code and so is required. An explanation of the origins of this factor can be found elsewhere (Fewell and Glos, 1988).

5 Characteristic value of density based on visual grading

The calculations for MoR and density are similar to each other; the chief difference being that it is assumed that density values follow a normal distribution but that MoR values follow a lognormal distribution, necessitating the taking of natural logs prior to calculating the sample mean \bar{y} and sample standard deviation s_y . Subsequently, for MoR, the exponential of these values (based on natural logs) is needed to determine the 5-percentile values. For ease of understanding, density is considered before MoR. The calculations described in EN14358 and EN384 follow three stages (and once again, due to a lack of consistency in terminology between the codes, and to improve clarity, an amended notation is used):

- i. Calculate the values of the sample mean \bar{y} , standard deviation s_y and 5-percentile ρ_{05}
- ii. Calculate $\rho_{05,LCL}$, which is the lower confidence limit of the 5-percentile value ρ_{05} for $\alpha = 75\%$
- iii. Adjust this lower confidence limit $\rho_{05,LCL}$ (according to the number and size of sub-samples) to obtain the characteristic 5-percentile value ρ_k

As noted above, it is assumed that the density test values are normally distributed. The mean value \bar{y} , the standard deviation s_y , and the standard error of the mean (SEM) are derived in the way described in the previous section (for MoE).

Once again, in EN14358, the value of the standard deviation must not be taken as lower than a limit of $0.05 \bar{y}$. For large data sets, the lower limit is not likely to be required, due to the size of the sample variance being adequately large.

The lower confidence limit, $\rho_{05,LCL}$, is determined at a confidence level of $\alpha = 75\%$, and the formula for this is determined in accordance with the principles explained in Section 3.

EN14358 follows the same approach and also provides a simplified approach, substituting the t-multiplier, $\frac{t_{(1-\alpha/2; n-1; \delta)}}{\sqrt{n}}$ with a new term $k_s(n)$ and giving both tabulated values and a simple formula for its calculation

$$\rho_{05,LCL} = \bar{y} - k_s(n) \times s_y \quad (5.1)$$

$t_{(\gamma; n-1; \delta)}$ is the γ quantile of a noncentral t-distribution with $n-1$ degrees of freedom and δ is the noncentrality parameter. The use of a noncentral t-distribution is appropriate for estimates of percentiles or quantiles (Hahn, Meeker and Escobar, 2017).

Thus, three ways are given for calculating $\rho_{05,LCL}$.

Finally, EN384 requires one further adjustment to obtain ρ_k for each of the visual grades. This adjustment accounts for the variations between the means of samples and the mean of a population. It is this value that is compared with the tabulated density values in Table 1 of EN338 (CEN, 2016a) to obtain the Strength Class.

$$\rho_k = \min \left[1.1 \rho_{05,LCL,min} , \frac{\sum_{i=1}^{n_s} n_i \rho_{05,LCL,i}}{n} \right] \times k_n \quad (5.2)$$

$\rho_{05,LCL,i}$ is the 5-percentile density of sub-sample i

$\rho_{05,LCL,min}$ is the lowest 5-percentile density of all i – sub-samples

n_s is the number of sub-samples

n_i is the number of specimens in sub-sample i

n is the total number of specimens

k_n is the factor to adjust for the number of sub-samples and this is taken from Table 1, presented in EN384.

Thus, for a single sample of $n_s = 1$, $k_n = 0.88$, the first term in brackets, $\rho_{05,LCL,min}$ is the same as the $\rho_{05,LCL,i}$ the 5-percentile density of sub-sample i and so the second term, $\frac{\sum_{i=1}^{n_s} n_i \rho_{05,LCL,i}}{n}$, is also the same as $\rho_{05,LCL}$ the 5-percentile density of the sample and so, as the first term is multiplied by 1.1, the minimum of the two must be simply the 5-percentile density of the sample with no adjustment. Therefore

$$\rho_k = \rho_{05,LCL} \times 0.88$$

This 12% reduction factor would reduce to zero if five or more sub-samples are used, in place of a single large sample.

6 Characteristic value of MoR based on visual grading

Reference should be made to Section 5 which deals with the determination of the characteristic value of density (as MoR follows a similar approach). Strength parameters are assumed to be logarithmically normally distributed and so the mean, the variance and the standard deviation of the sample are based on the natural logs of the MoR values of the data points.

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n \ln y_i \quad (6.1)$$

$$s_y = \sqrt{\frac{\sum_{i=1}^n (\ln y_i - \bar{y})^2}{n-1}} \quad (6.2)$$

As for MoE and density, EN14358 requires that a minimum value for variance is used, thus

$$s_y \geq \sqrt{\ln(1 + 0.05^2)} \approx 0.05 \quad (6.3)$$

The lower limit may not be required if the sample size is large enough and the sample variance is adequately large. Once again, the lower confidence limit $f_{05,LCL}$, is determined at a confidence level of $\alpha = 75\%$, and the formula for this is determined in accordance with the principles explained in Section 3 and Section 5.

$$f_{05,LCL} = \exp\left(\bar{y} - t_{(1-\alpha/2; n-1; \delta)} \times \frac{s_y}{\sqrt{n}}\right) \quad (6.4)$$

The calculations described in EN14358 and EN384 follow three stages, and again, an amended notation is used:

- i. Calculate the values of the sample mean \bar{y} , standard deviation s_y and 5-percentile f_{05}
- ii. Calculate $f_{05,LCL}$, which is the lower confidence limit of the 5-percentile value f_{05} for $\alpha = 75\%$
- iii. Adjust this lower confidence limit $f_{05,LCL}$ (according to the number of sub-samples) to obtain the characteristic 5-percentile value f_k

The lower confidence limit $f_{05,LCL}$, is determined at a confidence level of $\alpha = 75\%$, and the method is the same as for density

$$f_{05,LCL} = \bar{y} - t_{(1-\alpha/2; n-1; \delta)} \times \frac{s_y}{\sqrt{n}} \quad (6.5)$$

The term $k_s(n)$ from EN14358 is applicable for MoR (in just the same was as for density), substituting for $\frac{t_{(1-\alpha/2; n-1; \delta)}}{\sqrt{n}}$

$$f_{05,LCL} = \bar{y} - k_s(n) \times s_y \quad (6.6)$$

The final adjustment required by EN384 differs slightly from that for density

$$f_k = \min \left[1.2 f_{05,LCL,min} , \frac{\sum_{i=1}^{n_s} n_i f_{05,LCL,i}}{n} \right] \times k_n \quad (6.7)$$

Also, the factors k_n for MoR are lower than those for MoE and density. Thus, for a single sample of $n_s = 1$, $k_n = 0.70$, the first term in brackets, $f_{05,LCL,min}$ is the same as the sample 5-percentile and the second term, $\frac{\sum_{i=1}^{n_s} n_i f_{05,LCL,i}}{n}$, is also the same as the sample 5-percentile and so, as the first term is multiplied by 1.2, the minimum of the two must be simply the sample 5-percentile with no adjustment. Therefore

$$f_k = f_{05,LCL} \times 0.70$$

That some reduction of the characteristic 5-percentile value is needed is clear and it should be noted that the size of this reduction (30%) is significant.

7 References

CEN (2005) 'EN 1990:2002+A1:2005 Incorporating corrigendum December 2008 Eurocode - Basis of structural design'. Brussels: European Committee for Standardization.

CEN (2016a) 'EN 14358:2016 Timber structures — Calculation and verification of characteristic values'. Brussels: European Committee for Standardization.

CEN (2016b) 'EN 338:2016 Structural timber — Strength classes'. Brussels: European Committee for Standardization.

CEN (2018) 'EN384:2016+A1:2018 Structural timber — Determination of characteristic values of mechanical properties and density'. Brussels: European Committee for Standardization.

Fewell, A. and Glos, P. (1988) 'The determination of characteristic strength values for stress grades of structural timber. Part 1', in *CIB W18 proceedings paper 21-6-2*. Parkesville, Canada.

Hahn, G. J., Meeker, W. Q. and Escobar, L. A. (2017) *Statistical Intervals: A Guide for Practitioners and Researchers*. Second Edi. Hoboken, New Jersey: John Wiley & Sons Inc. doi: 10.2307/2290749.

Ravenshorst, G. (2015) *Species Independent Strength Modeling of Structural*. Technische Universiteit Delft.