Fitting of an ordinary linear least squared (OLS) regression is rather straight forward via an analytical approach, see *Eq. (5)*. Whereby first *β1* and later *β0* are calculated and the R-squared can be derived, *Eq. (6).*

For OLS the best fitting line is found by reducing the sum of squares. Pearson’s r and consequently R-squared can be derived from *Eq.* *(7)*

Another option would be fit use our derived model on the data.

The numerator in *Eq. (8)* also directly gives us our squared residuals of the model. This can also be notated as , where is the predicted value of the model at the ith value (simply ), often pronounced as yhat.

Assuming the raw residuals approximate the normal distribution, with limited heteroscedasticity (equal variance along *x*) and is centred at 0 and denoted a*s*  (or for the t-distribution). Then denotes the standard deviation and the variance of the population (*s* and *s2* for a sample assuming a t-distribution). In most programs the t-distribution is used where the *SE* of the mean is calculated the sample standard deviation where *s* is the sample standard deviation and *s2* the variance of the sample. The difference with population standard deviation and variance is that there is not divided by *n* in the denominator but by . The latter part has however been criticized as in small sample sizes it overestimates the intervals and if we have a reasonable sample size it might be perfectly reasonable to move to the z instead of t distribution (Huang, 2018).

If these assumptions are reasonably met, the residual standard deviation (sigma) can be calculated, see *Eq. 8*. That sigma is not directly given as the numerator in *Eq. 8* is to highlight that sigma represents the leftover noise (hence residuals) of the fitted model to the data.

The denominator is the sum of absolute differences from the mean of *y*. Then it is possible (and useful) to calculate the SE of (*Eq. (9)*) Correspondingly the confidence intervals for the t-distribution be calculated at 95% by multiplying the SE by two and subtracting or adding by (*Eq. (10 and 11)*).

If the upper and lower confidence interval do not cross ~0 we know that our data is at least compatible at <5% with the null or more extreme, under the assumption our hypothetical null is true. Otherwise, if the t-score is > ~2 we can suggest the same by applying the t-tests, see *Eq. (12)*.

The t-score (*t*) might sound confusing, but far from it. For we divide the coefficient by the *SE* and what we have is a signal to noise ratio, where is the signal and *SE* is the noise.

To calculate the *SE* of the intercept we repeat a similar process. However, the intercept is in normal conditions compared to null (literally) the origin where the x-axis and y-axis cross (at 0). The SE of the intercept can be calculated by dividing the squared mean of *x* by the variance of *x* (plus 1/n) and multiplied by sigma.

Often the term normal or Gaussian distribution is used. In 99% of the cases this is in relation to the residuals of a model (note that there are +1000 other distributions to select from). Given we believe the distribution of the expected value approximates a normal distribution which can be relatively fast given the central limit theorem (CLT). Shortly, the CLT describes that if we have enough samples the distribution of the error will approximate a normal distribution.