Appendix 1.

Pearson's correlation

Regression parameters for a straight line model (Y = a + bx) are calculated by the least squares method (minimisation of the sum of squares of deviations from a straight line). This differentiates to the following formula for the slope (b) and the Y intercept (a) of the line:

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$a = \bar{Y} - b\bar{x}$$

A residual for a Y point is the difference between the observed and fitted value for that point, i.e. it is the distance of the point from the fitted regression line.

Pearson's product moment correlation coefficient (r) is given as a measure of linear association between the two variables:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

r^2 is the proportion of the total variance (s^2) of Y that can be explained by the linear regression of Y on x. $1-r^2$ is the proportion that is not explained by the regression. Thus $1-r^2 = s^2xY / s^2Y$. Confidence limits are constructed for r using Fisher's z transformation. The null hypothesis that r = 0 (i.e. no association) is evaluated using a modified t test [9].