

Technical Glossary

1.21 Technical Glossary

The aim of this glossary of technical terms is to be a resource when researchers need brief definitions of statistical terms for reports etc. It is not intended to be exhaustive, or to define these terms for statisticians.

Age-standardised scores	Test scores adjusted to give values which are evenly distributed about a fixed value (often 100), and which take account of the age of the individuals taking the test.
Analysis of variance (ANOVA)	Analysis of variance is a technique for assessing whether the distribution of the values of a particular measurement varies from one group to another. The measurement could be the heights of a set of people, or pupils' scores on a test. The factors are usually categorical, such as gender, season of birth, or entitlement to free school meals. Analysis of covariance (ANCOVA) is analysis of variance on the residuals of a measurement, after allowing for a set of related numerical variables.
Average	An 'average' value purports to represent or to summarise the relevant features of a set of values. There are three types of average mean, median and mode for their definitions please refer to the <i>mean, median and mode</i> sections.
Bias (of samples)	In general, this refers to any situation in which there is an unwanted influence acting on the data collected, in such a way that it no longer properly represents the desired population. For example, if we sampled only from grammar schools to represent the population of all secondary school pupils, then it is very likely that the mean ability of the sample would be higher than the mean ability of the population, and any conclusions we drew from the sample would not necessarily be valid for the wider population and it would be regarded as a biased sample
Bias (of tests or items)	A test is biased if performance on the test for different groups of individuals is affected by factors unrelated to the characteristic being tested. Items are biased if, controlling for overall performance in the characteristic being tested, individuals from different groups perform differently on the item. (See differential item functioning).
Biserial Correlation	A correlation coefficient computed between a dichotomous and a continuous variable.

Bootstrap Methods	Methods for estimating uncertainties in complex situations, by repeated sampling of the given dataset.
Chi-Square Test	A test statistic for categorical data. Chi-squared as a test statistic is used to test independence, but the chi-square test also is used as a goodness-of-fit test. The chi-square test statistic can be converted into one of several measures of association, including phi coefficient, the contingency coefficient and Cramer's V .
Cluster analysis	A set of techniques for combining a large number of individuals or objects into a smaller set of clusters, such that individuals in the same cluster are similar to each other and less similar to those in other clusters. In some cases, different techniques can give rise to quite different sets of clusters.
Coefficient	In <i>regression analysis</i> , the estimated relationship between the outcome measure and one of the background variables, expressed as the change in the value of the outcome associated with one unit change in the background variable.
Cohen's Kappa	A measure of interrater reliability for categorical data. This percentage-of-agreement measure corrects for chance or random agreement. Kapp is 1 when agreement is perfect and it is 0 when agreement is no better than would be expected by chance.
Confidence interval	No statistical estimate is ever totally accurate, but the degree of accuracy will depend on a number of factors, including the amount of data on which it is based. If we estimate something, say the overall national mean score on a test based on a sample of individuals, we may compute a confidence interval which is a range of values, enclosing the best estimate, which has a specified probability of containing the true value. For example, if we say that a 95% confidence interval for the mean score is 15.7 to 16.1, this implies that if we repeated the whole exercise lots of times with different samples, then 95 times out of 100 the true national mean score would lie inside the confidence interval we give. Obviously, the higher the confidence level we want, the wider the interval we must specify.
Correlation	A measure of association between two measurements, e.g. between size of school and the mean number of GCSE passes at grades A, B & C obtained by each pupil. A positive correlation would occur if the number of passes increased with the size of the school. If the number of passes decreased with size of school there would be a negative correlation. Correlations range from -1 to +1 (perfect negative to perfect positive correlations); a value of zero indicates no linear association between the two measures
Cramer's V	A measure of association appropriate for measuring the strength of a relationship between two variables, one or both of which have more than two categories.

Cronbach's Alpha	A measure of internal reliability or consistency of the items in a test. It is a widely used form of Kuder-Richardson formula 20 (KR 20), it can be used for test items that have more than two answers which KR 20 cannot be. Like other reliability coefficients cronbach's alpha ranges from 0 to 1. Scores towards the high end of the range suggest that the items in a test are measuring the same thing.
Crosstabulation	A table of two or more variables cross-classified, consisting of cells showing the number of cases in each combination of categories from the variables. Also known as a contingency table.
Degrees of Freedom (<i>df</i>)	The number of values free to vary when computing a statistic; the number of pieces of information that can vary independently from on another. The <i>df</i> tells you how much data was used to calculate a particular statistic and is usually one less than the number of variables.
Differential item functioning (dif)	A statistical method for testing for the apparent existence of <i>item bias</i> , looking at different performance on an item by two groups of individuals, controlling for their overall performance on the test. One of the most commonly used procedures is the Mantel-Haenszel procedure.
Discriminant Analysis	A form of regression analysis designed for classification. It allows two or more continuous independent variables to be used to place individuals or cases into the categories of a categorical dependent variable.
Effect Size	A statistic, often abbreviated to D or delta, indicating the difference in outcome for the average subject who received a treatment from the average subject who did not. This statistic is often used in meta-analysis.
Eigenvalue	A statistic used in factor analysis to indicate how much of the variation in the original group of variables is accounted for by a particular factor.
Facility	The facility of an item or a test question is the overall percentage of candidates within the group in question who give the correct answer. For multiple-mark items, the facility may be defined as the average mark on the item expressed as a percentage of the maximum mark.

Factor analysis	Factor analysis is a technique which is widely used in dealing with large numbers of measurements made on different individuals or objects, when many of the measurements may be strongly correlated with each other. In factor analysis we attempt to define a smaller set of underlying factors which are related to the variables measured, and which explain or represent most of the correlation structure of the data. The set of factors we define in this way is not unique, and the final set used can be chosen according to a number of criteria. The process of finding the 'best' or simplest factor solution is known as factor rotation.
Function	<p>A function is a mathematical way of expressing certain kinds of relationships between two measures. A linear function implies a relationship which is steadily increasing or decreasing – plotted as a graph it give a straight line sloping upwards or downwards. A quadratic function is slightly more complex (it contains a squared element) and graphically can be represented either as a curve which rises, reaches a maximum and then falls, or alternatively one which falls, reaches a minimum, and then rises again. A cubic function is the next most complex, as it gives a graphical plot with both a maximum and a minimum point (i.e. rise-fall-rise or fall-rise-fall).</p> <p>Other common mathematical functions used are sine, log, square root etc.</p>
Goodness of Fit	How well a model, a theoretical distribution or an equation matches actual data.
Interaction	It is sometimes the case in regression models that the relationship between one of the variables and the outcome measure is different for different groups – for example the relationship between achievement and prior attainment may be different for boys and girls. This is modelled using an <i>interaction term</i> , which takes account of this possibility. If statistically significant, it implies that the strength of the underlying relationship is not the same for all groups.
Interquartile Range (IQR)	One way of looking at the spread of responses to a question is to look at the IAR. The three values which split the distribution of answers into four equal portions are known as quartiles, and the difference between the upper and lower quartile is the IQR. This is the range of the central 50% of the responses and median is the middle quartile.
IRT	Item Response Theory. The branch of psychometrics that is concerned with the probability of success when someone attempts to answer a test item (widely used in tests of intelligence, aptitude, ability, achievement and knowledge). Different types of model are used, with one, two or three parameters. The one-parameter model is sometimes know as the Rasch model.

Likelihood	The conditional probability of observed values given the expected values.
Likelihood Ratio	A ratio of two likelihoods. It is widely used as a test statistic, especially for relations among categorical variables displayed in contingency tables. The smaller the likelihood ratio the stronger the relationship.
Logistic regression	A form of regression in which the outcome of interest is binary, i.e. just takes two values – for example: passing an exam or not; going into further education or not; achieving 5+ A* to C grades at GCSE etc. A set of background variables can be used to predict the probabilities of the binary outcome, as in conventional regression analysis, but the interpretation of the coefficients is less straightforward.
Mean	The conventional way of calculating the ‘average’ of a set of data values, by adding them up and dividing by the number of data values. Can be seriously affected by a few extreme data values (see <i>median</i>).
Measurement Error	Inaccuracy resulting from flaws in a measuring instrument.
Median	The central value in a set of data, such that half the cases lie below and half above that value. It is less affected by extreme values than the <i>mean</i> as a measure of the ‘average’ of a dataset.
Mode	The most common response in a set of data.
Multilevel modelling	Multilevel modelling is a recent development of linear <i>regression</i> which takes account of data which is grouped into similar clusters at different levels. For example, individual pupils are grouped into year groups or cohorts, and those cohorts are grouped within schools. There may be more in common between pupils within the same cohort than with other cohorts, and there may be elements of similarity between different cohorts in the same school. Multilevel modelling allows us to take account of this hierarchical structure of the data and produce more accurate predictions, as well as estimates of the differences between pupils, between cohorts, and between schools. (Multilevel modelling is also known as hierarchical linear modelling).
Phi coefficient	A measure of association appropriate for measuring the strength of a relationship between two dichotomous variables.

Point Biserial Correlation	A correlation to measure the association between two variables, one which is dichotomous and the other continuous. (See Biserial Correlation).
Population	All the members of a group of interest, e.g. all schools in England & Wales or the employers of graduates in Berkshire.
PPS	Probability proportional to size. Used in sampling, when the number of pupils per school is to be fixed but school sizes vary. If schools are drawn from the population with a fixed probability, pupils in larger schools have a reduced probability of being included. PPS sampling ensures all pupils have the same probability of inclusion in these circumstances.
Quartile	Numerical values which divide a dataset into four parts with equal numbers in each. The <i>first quartile</i> is such that 25% of the data lies below it and 75% above; the <i>third quartile</i> has 75% below and 25% above. The <i>second quartile</i> is equivalent to the <i>median</i> .
Range	The highest value of a set of numbers minus the lowest value - an indication of variation in the data.
Regression analysis (linear)	This is a technique for finding a straight-line relationship which allows us to predict the values of some measure of interest ('dependent variable') given the values of one or more related measures. For example, we may wish to predict schools' GCSE performance given some background factors, such as free school meals and school size (these are sometimes called 'independent variables'). When there are several background factors used, the technique is called multiple linear regression. If just a single background factor is used to predict, we have simple linear regression, and the results may be plotted as a straight line on a graph.
Residual	A residual is the difference between an actual observed value and the predicted value from a statistical model.
Sampling	The drawing of a fraction or sample of all the members of a population which is representative of that population. In a simple random sample all members, e.g. schools, have the same chance of being selected. If there is more interest in the pupils than in the schools then sampling in proportion to size of school will ensure that each pupil has an equal chance of being selected. Other considerations such as cost of school visits to any place in the country, may lead to the choice of a other strategies, e.g. a cluster sample. A proportion of LEAs might be selected first, followed by a simple random sample or sampling in proportion to size (see <i>PPS</i>).

Scatterplot	When relationships between two measures on the same set of objects are to be investigated, it is often useful to plot them on a graph with one measure on the X-axis and the other on the Y-axis. Each object or individual then appears as a point on the graph, and relationships can be seen quite easily. For example, a plot of schools' average GCSE results against percentage of pupils eligible for free school meals shows quite strikingly how the former decreases, in general, as the latter increases.
Scree Plot	A plot of eigenvalues in descending order. Names after its resemblance to a geological scree (an accumulation of stones) at the base of a steep hill or cliff. Used in Factor analysis to help determine the number of factors.
Standard deviation	Standard deviation is a measure of the spread of some quantity within a group of individuals. If the quantity is distributed approximately Normally, we would expect about 95% of the individuals to be within 2 standard deviations either side of the mean value.
Standard error	A measure of the uncertainty in the estimation of a statistical parameter. It is expressed as the <i>standard deviation</i> of the errors in the estimate, so that there is roughly a 95% chance that the 'true' value lies within 2 standard errors either side of the estimate.
Standard Error of Measurement (SEM)	A measure of the uncertainty in individuals' test scores due to factors unrelated to the purpose of the test.
Standardisation	The procedure to create standard scores (<i>z</i> scores) in order to compare variables measured in different units.
Statistic	A <i>statistic</i> is any numerical value computed from a set of data, whose aim is to represent the data in some way and from whose value a statistical judgement may be made. An example is the <i>mean</i> or arithmetical average of a set of numbers, often used as a measure of its 'central tendency'.
Statistical significance	We say that there is a statistically significant difference between two groups in some quantity if the probability of that difference arising by chance is less than a preset value (e.g. 5%). Similarly, we say that there is a significant relationship between two variables if the observed results have a low probability of arising by chance, that is by random fluctuations when the two variables are really unrelated.
Variable	A characteristic of members of the population, e.g. number of pupils in the school.

Variance	A measure of variability in data (the square of the <i>standard deviation</i>).
Variation	The way that a variable takes different values for different individuals; e.g. mean GCSE results differ from school to school. This is variation at the level of the school. Individual pupils also differ in their number of GCSEs: this is variation at the level of the pupil.
Weighting	If a sample is not fully representative of the population from which it is drawn and estimates of population distributions etc. are required, then weighting factors may be derived for different groups within the sample, such that weighted distributions of sample results give improved estimates of values in the underlying population.
Z score	Standardisation of a value by subtracting the variable's mean from the value and dividing by the variable's standard deviation. A distribution of <i>z</i> scores always has a mean of 0 and a standard deviation of 1.

