E

TGP/14/1 Draft 4
SECTION 3: Statistical Terms
ORIGINAL: English only
DATE: December 12, 2007

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

Associated Document
to the
General Introduction to the Examination
of Distinctness, Uniformity and Stability and the
Development of Harmonized Descriptions of New Varieties of Plants (document TG/1/3)

DOCUMENT TGP/14

“GLOSSARY OF TECHNICAL, BOTANICAL AND STATISTICAL TERMS

USED IN UPOV DOCUMENTS”

Section 3: Statistical Terms

Document prepared by the Office of the Union

to be considered by the

Enlarged Editorial Committee at its meeting
to be held in Geneva, Switzerland, on January 8, 2008

Note for Draft version
Footnotes will be retained in published document
Endnotes are background information to assist in the consideration of this draft and will not appear in the final, published document

n:\orgupov\shared\document\tc_tgps\tgp-14\upov drafts\tgp_14_draft_4_section_3(stats).doc

Glossary of Statistical Terms

Acceptance probability:

Additivity: Effects, for example in an analysis of variance, are said to be additive if there is no interaction between them.

Alpha (α): Statisticians use the Greek letter alpha to indicate the probability of rejecting the statistical hypothesis tested when in fact, that hypothesis is true. α is called the significance level of a test. Before conducting any statistical test, it is important to set a value for alpha. For establishing distinctness, alpha is sometimes set at 0.01. This is the equivalent of asserting that one will reject the hypothesis tested 1 out of 100 times if the obtained test statistic is among those that would occur from random samples drawn from a population in which the hypothesis is true. If the obtained statistic leads to rejection of the tested hypothesis, it is not because the obtained statistic could not have occurred by chance, but because the odds of obtaining the statistic by chance are sufficiently low (one out of hundred), and so it is reasonable to conclude that the results are not due to chance.

Alpha-design: Alpha designs are a very flexible class of resolvable incomplete block designs. Such designs are particularly useful when there are many treatments to be examined, the variability of the experimental units is such that the block size needs to be kept small, and blocks can be combined into full replicates.

Alternative Hypothesis: In hypothesis testing, the null hypothesis and an alternative hypothesis are put forward. If the data support sufficiently strongly rejection of the null hypothesis, then the null hypothesis is rejected in favor of an alternative hypothesis. For instance, if the null hypothesis were that $\mu_1 = \mu_2$ then the alternative hypotheses would be $\mu_1 \neq \mu_2$ (two-sided), or $\mu_1 < \mu_2$ or $\mu_1 > \mu_2$ (one-sided).

ANOVA: This term is an acronym for a procedure entitled Analysis of Variance. This procedure employs the statistic (F) to test the statistical significance of the differences among the obtained means of two or more random samples from a given population. When there are one or two factors in the experiment, the analysis is called a one-way or a two-way analysis of variance respectively. See also factorial design.

Assumptions: see model assumptions

(Balanced) Complete Block Design / Randomized complete block design : An experimental lay-out where all treatments are present once in every block. Blocking is done to make the experimental units more homogeneous within each group. All treatments are randomly assigned within each block to minimize the confounding effect of the heterogeneous experimental units. This is a common design for field trials of agricultural crops.

Balanced Incomplete Block Design: This differs from a balanced complete block design in that the block size is less than the total number of treatments. Each treatment is replicated equally and the assignment of the treatments over the blocks is such that the SED of each pair of treatment means has the same value.

Bar graph: A bar graph is much like a histogram, differing in that the columns are separated from each other by a small distance. Bar graphs are commonly used for qualitative variables.

Beta (β): Statisticians use the Greek letter beta to indicate the probability of failing to reject the null hypothesis when it is false and a specific alternative hypothesis is true. For a given test, the value of beta is determined by the value of alpha, features of the statistic that is being calculated (particularly the sample size) and the specific alternative hypothesis that is being entertained. While it is possible to carry out a statistical test without defining a specific alternative hypothesis, neither beta nor power can be calculated. It is relevant to note here that power (the probability that the test will reject the hypothesis tested when a specific alternative hypothesis is true) is equal to one minus beta (i.e. power = 1 - beta). See Power

Between plot standard deviation: When speaking about variance components this term is commonly used for the variability between experimental units, like plots.

Bias: Bias is the difference between the true value of the parameter and the expected value of the estimator. An estimator is biased if the expected value of the estimator doesn't equal the parameter it is estimating.

Binomial Distribution: When a coin is flipped, the outcome is either a head or a tail. In this example, the event has two mutually exclusive possible outcomes. For convenience, one of the outcomes can be labeled "success" and the other outcome "failure." If an event occurs N times (for example, a coin is flipped N times), then the binomial distribution can be used to determine the probability of obtaining exactly r successes in the N outcomes. The binomial probability for obtaining r successes in N trials is:

$$P(r) = \binom{N}{r} \pi^{r} (1-\pi)^{N-r}, \quad r = 0,1\ldots N$$

where P(r) is the probability of exactly r successes, N is the number of events, and π is the probability of success on any one trial. This formula assumes that the events:

(a) are dichotomous (fall into only two categories)
(b) are mutually exclusive
(c) are independent and
(d) are randomly selected

Bivariate Normality: A particular form of distribution of two variables that has the traditional 'bell' shape (but not all bell-shaped distributions are normal). If plotted in three-dimensional space, with the vertical axis showing the number of cases, the shape would be that of a three-dimensional bell (if the variances on both variables were equal) or a ~~'fireman's hat'~~ flattened three-dimensional bell (if the variances were unequal). When perfect bivariate normality obtains, the distribution of one variable is normal for each and every value of the other variable. See also Normal Distribution.

[Note: to add 3 dimensional graphic of bivariate distribution]

Blocking: A method in the design of experiments used to reduce the variability of residuals. Types of designs that use this method are generally called block designs. A great number of types exist but only a few one are considered in this document. See also Block Design.

Block Design: see Balanced Complete Block Design, (Balanced) Incomplete Block Design, Randomized Complete Block Design, Alpha Design.

Box plot – also called box-and-whisker diagram: A schematic plot to display the distribution of a variable. The box spans the interquartile range of the values in the variable, so that the middle 50% of the data lie within the box, with a line indicating the median. Whiskers can extend beyond the ends of the box as far as the minimum and maximum values.

Central Limit Theorem: The Central Limit Theorem is a statement about the characteristics of the sampling distribution of means of random samples from a given population. That is, it describes the characteristics of the distribution of values we would obtain if we were able to draw an infinite number of random samples of a given size from a given population and we calculated the mean of each sample.

The Central Limit Theorem consists of three statements:

1. The mean of the sampling distribution of means is equal to the mean of the population from which the samples were drawn.

2. The variance of the sampling distribution of means is equal to the variance of the population from which the samples were drawn divided by the size of the samples.

3. If the original population is distributed normally (i.e. it is bell shaped), the sampling distribution of means will also be normal. If the original population is not normally distributed, the sampling distribution of means will increasingly approximate a normal distribution as sample size increases. (i.e. when increasingly large samples are drawn)

Chi-Square: The statistic X^2 (Chi-Square) is what statisticians call an enumeration statistic. Rather than measuring the value of each of a set of items, a calculated value of Chi-Square compares the frequencies of various kinds (or categories) of items in a random sample to the frequencies that are expected if the population frequencies are as hypothesized by the investigator. Chi-square is often used to assess the "goodness of fit" between an obtained set of frequencies in a random sample and what is expected under a given statistical hypothesis. For example, Chi-Square can be used to determine if there is reason to reject the statistical hypothesis that the frequencies in a random sample are as expected when the items are from a normal distribution.

Chi-squared (χ^2) distribution: distribution of the sum of squared independent standard normal variables. Used to do significance tests on chi-squared statistics.

Coefficient: A coefficient is a constant used to multiply another value. In the linear transformation Y = 3X + 7, the coefficient "3" is multiplied by the variable X. In the linear combination of means $L = (2)M_1 + (-1)M_2 + (-1)M_3$ the three numbers in parentheses are coefficients.

Completely Randomised Design: An experimental lay-out where the experimental units are homogenous and the treatments are randomly assigned to the uniform experimental units without any constraint. It is the simplest experimental design, which is used in the testing of many horticultural and ornamental crops under greenhouse condition where the experimenter has more control over the experimental units.

Confidence Interval: A confidence interval is a range of values that has a specified probability of containing the parameter being estimated. The 95% and 99% confidence intervals, which have 0.95 and 0.99 probabilities of containing the parameter respectively are most commonly used. If the parameter being estimated were μ, the 95% confidence interval might look like the following:

$$12.5 \leq \mu \leq 30.2$$

What this means is that the interval between 12.5 and 30.2 has a 0.95 probability of containing μ.

Confounding: Two factors are confounded if they vary together in such a way that it is impossible to determine which factor is responsible for an observed effect. For example, consider an experiment in which two fungicides treatments for foliar disease control were compared. Treatment one was given to the one variety and treatment two was given to another variety. If a difference between treatments were found, it would be impossible to tell if one treatment were more effective than the other or if treatments for disease control are more effective for one variety than the other. In this case, varieties and treatment are confounded. Sometimes, confounding is much more subtle. An experimenter may accidentally manipulate a factor in addition to the factor of interest.

Consistency: An estimator is consistent if the estimator tends to get closer to the parameter it is estimating as the sample size increases.

Contingency Table: A contingency table is a table showing the responses of subjects to one factor as a function of another factor. For instance, the following contingency table shows disease resistance as a function of different varieties (the data are hypothetical). The entries show the number of plants for each variety under different level of disease resistance. The Chi-square test of independence is used to test the relationship between rows (varieties) and columns (disease resistance) for significance.

~~Disease Resistance~~	~~Resistant~~	~~Moderately Resistant~~	~~Susceptible~~
~~Variety~~			
~~Candidate~~	~~18~~	~~20~~	~~2~~
~~Comparator 1~~	~~3~~	~~10~~	~~27~~
~~Comparator 2~~	~~6~~	~~24~~	~~10~~

[Note: example to be changed]

Continuous Variable: A continuous variable is one for which, within the limits the variable's range, any value is possible. For example, the variable 'plant height' is continuous since it may be 1.21m, 1.25m or even 1.30m etc to measure plant heights. The variable 'Number of lobed leaves' is not a continuous variable since it is not possible to get 54.12 lobed leaves from 100 leaves counted. It must be an integer. See also 'discrete variable'

Correlation (Pearson): Given a pair of related measures (X and Y) on each of a set of items, the correlation coefficient (r) provides an index of the degree to which the paired measures co-vary in a linear fashion. In general r will be positive when items with large values of X also tend to have large values of Y whereas items with small values of X tend to have small values of Y. Correspondingly, r will be negative when items with large values of X tend to have small values of Y whereas items with small values of X tend to have large values of Y. Numerically, r can assume any value between -1 and +1 depending upon the degree of the relationship. Plus and minus one indicate perfect positive and negative relationships whereas zero indicates that the X and Y values do not co-vary in any linear fashion. See Measures of association.

COYD: Abbreviation of Combined-Over-Years Distinctness criterion. Statistical method to test distinctness in DUS testing. See TGP/9

COYU: Abbreviation of Combined-Over-Years Uniformity criterion. Statistical method to test uniformity in DUS testing. See TGP/10

Critical Value: A critical value (which depends on the level of significance, alpha) is used in significance testing. It is the value that a test statistic must exceed in order for the null hypothesis to be rejected. For example, the critical value of t (with 12 degrees of freedom in a two-sided test using the alpha=0.05 significance level) is 2.18. This means that for the probability to be less than or equal to 0.05, the absolute value of the t statistic must be 2.18 or greater.

Degrees of Freedom: Statisticians use the terms ‘degrees of freedom’ to describe the number of values in the final calculation of a statistic that are free to vary. Consider, for example the statistic s^2, the estimated variance of a sample. To calculate the estimated variance of a random sample, we must first calculate the mean of that sample and then compute the sum of the several squared deviations from that mean. While there will be n such squared deviations only (n - 1) of them are, in fact, free to assume any value whatsoever. This is because the final squared deviation from the mean must include the one value of X such that the sum of all the Xs divided by n will equal the obtained mean of the sample. All of the other (n - 1) squared deviations from the mean can, theoretically, have any values whatsoever. For these reasons, the statistic s^2, the estimated variance of a sample, is said to have only (n - 1) degrees of freedom.

Dependent Variable: A variable which the analyst is trying to explain in terms of one or more independent variables. The distinction between dependent and independent variables is typically made on theoretical grounds-in terms of a particular causal model or to test a particular hypothesis. This is often called the Y-variable.

Design of experiment: see Experimental design

Discrete Variable: A discrete variable is one that cannot take on all values within the limits of the variable. For example, responses to a five-point rating scale can only take on the values 1, 2, 3, 4, and 5. The variable cannot have the value 1.7. A variable such as a plant height can take on any value. Variables that can take on any value and therefore are not discrete are called continuous. Statistics computed from discrete variables can be continuous. The mean on a five-point scale could be 3.117 even though 3.117 is not possible for an individual score.

Dispersion: Synonyms are variation, variability or spread. A variable’s dispersion is the degree to which scores on the variable differ from each other. If every score on the variable were about equal, the variable would have very little dispersion. There are many measures of dispersion, eg. variance, standard deviation, range, interquartile range etc.

Distribution: Form of a function that describes the possible outcomes of a variable. The (probability) distribution of a (random) variable specifies the chance that the variable takes a value in any subset of the real numbers. See Binomial Distribution, Chi-squared distribution, Continuous Distribution, Discrete Distribution, F-Distribution, Frequency Distribution, Normal Distribution, Relative Frequency Distribution, Standard Normal Distribution, Symmetric Distribution, Student's t-Distribution, t-Distribution, Z-Distribution

Effect: see Main Effect

Efficiency: The efficiency of a statistic is the degree to which the statistic is stable from sample to sample. That is, the less subject to sampling fluctuation a statistic is, the more efficient it is. The efficiency of a statistic is measured relative to the efficiency of other statistics and is therefore often called the relative efficiency. If statistic A has a smaller standard error than statistic B, then statistic A is more efficient than statistic B. The relative efficiency of two statistics may depend on the distribution involved. For instance, the mean is more efficient than the median for normal distributions but not for many types of skewed distributions. The efficiency of a statistic can also be thought of as the precision of the estimate: the more efficient the statistic, the more precise the statistic is as an estimator of the parameter.

Estimation: The process of using a statistic to estimate a parameter of a distribution.

Estimator: An estimator is used to estimate a parameter. Normally a statistic is used as an estimator. Three important characteristics of estimators are: bias, consistency, and relative efficiency.

Expected Value: A theoretical average value of a statistic over an infinite number of samples from the same population.

Experimental Design: The lay-out of an experiment. See Completely Randomised Design, Balanced Complete Block Design, Incomplete Block Design, Alpha Design, Factorial Design.

Experimental Unit: An experimental unit is the smallest subdivision of the experiment (trial) to which the varieties are randomized. If there are more than one plant within a plot, the observations of a certain characteristic on each plant are used for estimating the plant-to-plant variability of the variety. The mean (or other function) of the observations can be considered as the plot measurement for that characteristic. Usually the experimental unit in a field trial is a plot.

F Distribution: The F distribution is the distribution of the ratio of two chi-squared variables, e.g. ratio of two estimates of variance. It is used to compute probability values in the analysis of variance. The F distribution has two parameters: degrees of freedom numerator (dfn) and degrees of freedom denominator (dfd). The dfn is the number of degrees of freedom of the numerator, and dfd is the number of degrees of freedom of the denominator. The dfd is often called the degrees of freedom for error or dfe. In the simplest case of a one-factor between-subjects ANOVA, dfn = a-1

dfd = N-a

where "a" is the number of groups and "N" is the total number of subjects in the experiment. The shape of the F distribution depends on dfn and dfd. The lower the degrees of freedom, the larger the value of F needed to be significant. For instance, if dfn = 4 and dfd = 12, then an F of 3.26 would be needed to be significant at the 0.05 level. If the dfn were 10 and the dfd were 100, then an F of 1.93 would suffice.

Factor: Each basic treatment will be called a factor. If an experiment is testing the effect of fertiliser dosage, then 'fertiliser' is a factor. Some experiments have more than one factor. For example, if the effect of fertiliser dosage and irrigation water were both manipulated in the same experiment, then these two variables would be factors. The experiment would then be called a two-factor experiment.

Factor Level: The possible forms of a factor are called the levels of that factor. The levels of factor 'variety' for example are the different varieties in an experiment.

Factorial Design: When an experimenter is interested in the effects of two or more factors, it is usually more efficient to combine these factors in one experiment than to run a separate experiment for each factor. Moreover, only in experiments with more than one factor is it possible to test for interactions between factors. Consider a hypothetical experiment on the effects of the factor nitrogen on grain yield in a cereal crop. There were three levels of nitrogen dosage: 50kg, 100kg and 150kg per hectare. A second factor, water level, was also manipulated. There were two levels of irrigation water on the field: 5cm and 10cm. The grain yield data (t/ha) for each condition (often called treatment) in the experiment is shown below:

Water	5cm	10cm
Dosage		
50 kg/ha	1.5	1.8
100 kg/ha	2.5	2.2
150 kg/ha	2.8	1.9

The number of combinations (six) is therefore the product of the number of levels of dosage (three) and levels of water (two). Also see: Main Effect.

Fitted Values of dependent variable: Explained part of observed values of the dependent variable. These values are calculated by using the estimated parameters in a model.

Fitted Constants: Special type of an (non-orthogonal) analysis of variance model assuming additivity of the factors.

Fixed term/Fixed factor: see mixed model

F Ratio: Ratio (quotient) of two variances that is F-distributed. It is used for example in ANOVA's to test the effect of factors and their interactions.

Frequency Distribution: A frequency distribution shows the number of observations falling into each of several intervals of values. Frequency distributions are portrayed as frequency tables, histograms, or polygons. Frequency distributions can show either the actual number of observations falling in each interval or the percentage of observations. In the latter instance, the distribution is called a relative frequency distribution.

Frequency Table: A frequency table is constructed by allocating the scores on a variable into intervals and counting the number of scores in each interval. The actual number of scores is displayed as well as the percentage of scores in each interval.

Heteroscedasticity: The absence of homogeneity of variance. See Homogeneity of Variance

Heterogeneity: The absence of homogeneity of variance. See Homogeneity of Variance

Hierarchical Analysis: In the context of multidimensional contingency table analysis, a hierarchical analysis is one in which inclusion of a higher order interaction term implies the inclusion of all lower order terms. For example, if the interaction of two factors is included in an explanatory model, then the main effects for both of those factors are also included in the model.

Histogram: A histogram is constructed from a frequency table. The intervals are shown on the X-axis and the number of scores in each interval is represented by the area of a rectangle located above the interval, which, if the intervals are of equal width, is equivalent to the rectangle's height.

Homogeneity of Variance: The assumption of homogeneity of variance (or homoscedasticity of variance) is that the variance within each of the populations is equal. This is an assumption of analysis of variance (ANOVA). ANOVA works well even when this assumption is violated except in the case where there are unequal numbers of subjects in the various groups. If the variances are not homogeneous, they are said to be heterogeneous or heteroscedastic.

Homoscedasticity: See Homogeneity of Variance

Hypothesis Testing: Hypothesis testing is a method of inferential statistics. An experimenter starts with a hypothesis about a population parameter called the null hypothesis. Data are then collected and the viability of the null hypothesis is determined in light of the data. If the data are very different from what would be expected under the assumption that the null hypothesis is true, then the null hypothesis is rejected. If the data are not greatly at variance with what would be expected under the assumption that the null hypothesis is true, then the null hypothesis is not rejected. Failure to reject the null hypothesis is not the same thing as accepting the null hypothesis.

Incomplete Block Design: Block design where the number of plots within each block is smaller than the number of treatments.

Independence: Observations on one plot are called independent if they are not influenced by varieties on other plots. For example if tall varieties are planted next to a small one there could be a negative influence of the big ones on the small one. In such a case a row of plants on both sides of the plot can be planted in order to avoid dependency. See also Statistical Independence.

Independent Variable: Two variables are independent if knowledge of the value of one variable provides no information about the value of another variable. For example, if you measured the terminal leaf length and the degree of fragrance in a rose variety, then these two variables would in all likelihood be independent. Knowing that leaf length would not effect the fragrance of rose. However, if the variables were leaf length and leaf width, then there may be a high degree of dependence. When two variables are independent then the correlation between them is 0.

Interaction: A situation in which the direction and/or magnitude of the relationship between two factors depends on (i.e., differs according to) the value of one or more other factors. When interaction is present, simple additive techniques are inappropriate; hence, interaction is sometimes thought of as the absence of additivity. Synonyms: non-additivity, conditioning effect, moderating effect, contingency effect.

Interquartile Range: The interquartile range is a measure of spread or dispersion. It is computed as the difference between the 75th percentile [often called (Q3)] and the 25th percentile (Q1). The formula for interquartile range is therefore: Q3-Q1. Since half the scores in a distribution lie between Q3 and Q1, the interquartile range is the distance needed to cover 1/2 the scores. The interquartile range is little affected by extreme scores, so it is a good measure of spread for skewed distributions. However, it is more subject to sampling fluctuation in normal distributions than is the standard deviation and therefore not often used for data that are approximately normally distributed.

Interval Scale: A scale consisting of equal-sized units. On an interval scale the distance between any two positions is of known size. Results from analytic techniques appropriate for interval scales will be affected by any non-linear transformation of the scale values. See also Scale of Measurement

Intervening Variable: A variable which is postulated to be a predictor of one or more dependent variables, and simultaneously predicted by one or more independent variables. Synonym: mediating variable.

Kurtosis: Kurtosis indicates the extent to which a distribution is more peaked or flat-topped than a normal distribution.

Least Significant Difference (LSD): A commonly used mean separation procedure. For example, the difference between two means (based on the same number of observations) is declared significant at any desired level of significance if it exceed the value derived from the following formula:

$$LSD = t\,\sqrt{(2S^2/n)},$$

where t is the tabulated two-tailed t-value at the required probability and degrees freedom. S is the pooled standard deviation of the observations and n is the number of observations per mean.

Level of a factor: See Factor Level

Level of significance: See Significance Level

Linear: The form of a relationship among variables such that when any two variables are plotted, a straight line results. A relationship is linear if the effect on a dependent variable of a change of one unit in an independent variable is the same for all possible such changes.

Linear Regression: Linear regression is the prediction of one variable from another variable when the relationship between the variables is assumed to be linear ($Y=aX+b$).

Linear Transformation: A linear transformation of a variable involves multiplying each value of the variable by one number and then adding a second number. For example, consider the variable X with the following three values: 2, 3, and 7. One linear transformation of the variable would be to multiply each value by 2 and then to add 5. If the transformed variable is called Y, then $Y = 2X+5$. The values of Y are: 9, 11 and 19.

LSD: See Least Significant Difference

Main Effect: The main effect of a factor is the effect of the factor averaging over all levels of other factors in the experiment. The main effect of irrigation water given in Factorial Design example could be assessed by computing the mean for the two levels of water averaging across all three levels of nitrogen dosage. The mean for the 5cm water is: $(1.5 + 2.5 + 2.8)/3 = 2.27$ and the mean for the 10cm water is: $(1.8 + 2.2 + 1.9)/3 = 1.97$. The main effect of water, therefore, involves a comparison of the mean of the 5cm water (2.27) with the mean of the 10cm water (1.97). Analysis of variance provides a significance test for the main effect of each factor in the design.

Mean: The arithmetic mean is what is commonly called the average. When the word "mean" is used without a modifier, it can be assumed that it refers to the arithmetic mean. The mean is the sum of all the scores divided by the number of scores. The formula in summation notation is: $\mu = \sum X/N$, where μ is the population mean and N is the number of scores. If the scores are from a sample, then the symbol M refers to the mean and N refers to the sample size. The formula for M is the same as the formula for μ. The mean is a good measure of central tendency for roughly symmetric distributions but can be misleading in skewed distributions since it can be greatly influenced by extreme scores. Therefore, other statistics such as the median may be more informative for distributions such as reaction time or family income that are frequently very skewed. The sum of squared deviations of scores from their mean is lower than their squared deviations from any other number. For normal distributions, the mean is the most efficient and therefore the least subject to sample fluctuations of all measures of central tendency.

Mean Square Error: The mean square error (MSE) is an estimate of the population variance in the analysis of variance. The mean square error is the denominator of the F ratio.

Measure of Association: A number (a statistic) whose magnitude indicates the degree of correspondence i.e. strength of relationship between two variables. An example is the Pearson product-moment correlation coefficient. Measures of association are different from statistical tests of association (e.g. Pearson chi-square, F-test) whose primary purpose is to assess the probability that the strength of a relationship is different from some pre-selected value (usually zero). See also Statistical Measure, Statistical Test

Median: The median is the middle of a distribution: half the scores are above the median and half are below the median. The median is less sensitive to extreme scores than the mean and this makes it a better measure than the mean for highly skewed distributions.

Missing Data: Information that is not available for a particular case for which at least some other information is available.

Mixed model: [definition to be provided]

[Mode: The mode is the most frequently occurring score in a distribution and is used as a measure of central tendency. The advantage of the mode as a measure of central tendency is that its meaning is obvious. Further, it is the only measure of central tendency that can be used with nominal data. The mode is greatly subject to sample fluctuations and is therefore not recommended to be used as the only measure of central tendency. A further disadvantage of the mode is that many distributions have more than one mode. These distributions are called "multimodal." In a normal distribution, the mean, median, and mode are identical.]

[Note: to reconsider whether the term "Mode" should be in the glossary]

Model: see statistical model

Model assumptions: With all statistical models assumptions are assumed. For example, with ANOVA two assumptions are: the residuals are normally distributed and have homogeneity of variance.

Modified Joint Regression Analysis: A statistical method used to adjust for when marked differences between years in the range of expression of a characteristic can occur. For example, in a late spring, the heading dates of grass varieties can converge. The method involves fitting a model to the variety-by-year table of means for the characteristic such that the model allows for a proportionately larger or smaller variety response depending on the year the data was observed in. For greater detail see TGP/8

Multiple Comparison Test: See Range Test

Multivariate Normality: The form of a distribution involving more than two variables in which the distribution of one variable is normal for each and every combination of categories of all other variables. See also Normal Distribution

Mutually Exclusive Events: Two events are mutually exclusive if it is not possible for both of them to occur at once. For example, if a dice is rolled, the event "getting a 1" and the event "getting a 2" are mutually exclusive since it is not possible for the dice to be both a one and a two on the same roll. The occurrence of one event "excludes" the possibility of the other event.

Nominal Scale: A classification of cases which defines their equivalence and non-equivalence, but implies no quantitative relationships or ordering among them. Analytic techniques appropriate for nominally scaled variables are not affected by any one-to-one transformation of the numbers assigned to the classes. See also Scale of Measurement

Non-additive: Not additive. See Interaction

Normal Distribution: A particular form for the distribution of a variable which, when plotted, produces a ‘bell’ shaped curve- symmetrical, rising smoothly from a small number of cases at both extremes to a large number of cases in the middle. Not all symmetrical bell-shaped distributions meet the definition of normality.

Normality: See Normal Distribution

Normal Probability Plot: Gives a visual indication of whether the distribution of a set of data is approximately normal. The data are ranked and the percentile of each data value is obtained. The data value is then plotted against the normal equivalent deviate of the data value’s percentile. If the distribution is close to normal, the plotted points will lie close to a straight line.

Null Hypothesis: The null hypothesis is an hypothesis about a population parameter. The purpose of hypothesis testing is to test the viability of the null hypothesis in the light of experimental data. Depending on the data, the null hypothesis either will or will not be rejected as a viable possibility. Consider a researcher interested in whether the Variety 1 is taller than Variety 2. The null hypothesis is that $\mu_1 - \mu_2 = 0$ where μ_1 is the mean height of Variety 1 and μ_2 is the mean height of Variety 2. Thus, the null hypothesis concerns the parameter $\mu_1 - \mu_2$ and the null hypothesis is that the parameter equals zero. The null hypothesis is often the reverse of what the experimenter actually believes; it is put forward to allow the data to contradict it. In the experiment, the experimenter probably expects that Variety 1 is taller than Variety 2. If the experimental data show that Variety 1 has a sufficiently higher plant height, then the null hypothesis that there is no difference in plant height can be rejected.

Ordinal Scale: A classification of cases into a set of ordered classes such that each case is considered equal to, greater than, or less than every other case. Analytic techniques appropriate for ordinally scaled variables are not affected by any monotonic transformation of the numbers assigned to the classes. See also Scale of Measurement

Outlier: See Outlying Case

Outlying Case (Outlier): A case whose score on a variable deviates substantially from the mean (or other measure of central tendency). Such cases can have disproportionately strong effects on statistics.

Paired t-Test:

Parameter: A parameter is a numerical quantity measuring some aspect of a population of scores. For example, the mean is a measure of central tendency. Greek letters are used to designate parameters. Following are some examples of parameters of great importance in statistical analyses and the Greek symbol that represents each one. Parameters are rarely known and are usually estimated by statistics computed in samples. To the right of each Greek symbol is the symbol for the associated statistic used to estimate it from a sample.

Quantity	***Parameter***	**Statistic**
Mean	μ	M
Standard deviation	σ	S
Proportion	π	P
Correlation	ρ	R

Pattern Variable: A nominally scaled variable whose categories identify particular combinations (patterns) of scores on two or more other variables.

Pooled Standard Deviation: Square root of pooled variance

Pooled Variance: Weighted average of a number of variances.

Population: A population consists of an entire set of objects, observations, or scores that have something in common. The distribution of a population can be described by several parameters such as the mean and standard deviation. Estimates of these parameters taken from a sample are called statistics.

Population standard:

Power: Power is the probability of correctly rejecting a false null hypothesis. Power is therefore defined as: 1 - β where β is the Type II error probability. If the power of an experiment is low, then there is a good chance that the experiment will be inconclusive. That is why it is so important to consider power in the design of experiments. There are methods for estimating the power of an experiment before the experiment is conducted. If the power is too low, then the experiment can be redesigned by changing one of the factors that determine power.

Precision:

Predicted Values:

Prediction:

Probability Value: In hypothesis testing, the probability value is the probability of obtaining a statistic as different from or more different from the parameter specified in the null hypothesis as the statistic obtained in the experiment. The probability value is computed assuming the null hypothesis is true. If the probability value is below the significance level then the null hypothesis is rejected. The probability value is also known as the significance probability.

P-Value: See Probability Value

Qualitative Variable: see Variable

Quantitative Variable: see Variable

Random Sampling: In random sampling, each item or element of the population has an equal chance of being chosen at each draw. A sample is random if the method for obtaining the sample meets the criterion of randomness (each element having an equal chance at each draw). The actual composition of the sample itself does not determine whether or not it was a random sample.

Random Term / Random Factor: see Mixed models

Random Variable:

Randomized complete block design: See (Balanced) complete block design

Randomisation: In designing an experiment to compare a number of varieties with each other it is important to randomize the varieties over the plots.

Range: The range is the simplest measure of spread or dispersion. It is equal to the difference between the largest and the smallest values. The range can be a useful measure of spread because it is so easily understood. However, it is very sensitive to extreme scores since it is based on only two values. The range should almost never be used as the only measure of spread, but can be informative if used as a supplement to other measures of spread such as the standard deviation or semi-interquartile range; e.g. the range of the numbers 1, 2, 4, 6, 12, 15, 19, 26 is 25 (=26 – 1).

Range Test: Range tests are used to compare each mean in an experiment with every other mean; they are based on the studentized range distribution. The most commonly used range tests are: Duncan's Multiple range Test, Student-Newman-Keul's Test, Tukey's Test.

Ranks: The expression of a particular characteristic (e.g., plant height) relative to other cases on a defined scale-as in 'Short,' 'Medium,' 'Tall' etc. Note that when the actual values of the numbers designating the relative positions (the ranks) are used in analysis they are being treated as an interval scale, not an ordinal scale. See also Interval Scale, Ordinal Scale

Ratio Scale: Ratio scales are like interval scales except they have true zero points. A good example is the Kelvin scale of temperature. This scale has an absolute zero. Thus, a temperature of 300 Kelvin is twice as high as a temperature of 150 Kelvin.

Regression Line: A regression line is a line drawn through a scatter-plot of two variables, one is the independent variable (Y) and the other is the dependent variable. The line is chosen so that it comes as close to the points as possible. In linear regression, Y values are obtained from several populations, each population being determined by a corresponding X value. The randomness of Y is essential and it is assumed that the Y populations are normally distributed and have a common variance.

Relative Frequency Distribution: See Frequency Distribution

REML: Restricted Maximum Likelihood method used to analyse a non-orthogonal ANOVA with more than one type of experimental unit.

Residual: Unexplained part of an observation. Remains after fitting a model. It is the difference of the observation and the prediction from the model.

Replication: In order to know whether a difference between a new variety and another variety exists, replicates are needed of the varieties. This is in order to know whether the difference is a real difference between the varieties or a difference due to random fluctuations.

Resolvable Design: A resolvable design is one in which each block contains only a selection of the treatments, but the blocks can be grouped together into subsets in which each treatment is replicated once. The groupings of blocks thus form replicates.

Sample: A sample is a subset of a population. Since it is usually impractical to test every member of a population, a sample from the population is typically the best approach available. Inferential statistics generally require that sampling be random although some types of sampling seek to make the sample as representative of the population as possible by choosing the sample to resemble the population on the most important characteristics.

Sample Size: The sample size is very simply the size of the sample. If there is only one sample, the letter "N" is often used to designate the sample size. If samples are taken from each of "a" populations, then the small letter "n" is often used to designate size of the sample from each population. When there are samples from more than one population, N is used to indicate the total number of subjects sampled and is equal to (a)*(n). If the sample sizes from the various populations are different, then n_1 would indicate the sample size from the first population, n_2 from the second, etc. The total number of subjects sampled would still be indicated by N. When correlations are computed, the sample size (N) refers to the number of subjects and thus the number of pairs of scores rather than to the total number of scores. The symbol N also refers to the number of subjects in the formulas for testing differences between dependent means. Again, it is the number of subjects, not the number of scores.

Sampling Fluctuation: Sampling fluctuation refers to the extent to which a statistic takes on different values with different samples. That is, it refers to how much the statistic's value fluctuates from sample to sample. A statistic whose value fluctuates greatly from sample to sample is highly subject to sampling fluctuation.

Scale of Measurement: Scale of measurement refers to the nature of the assumptions one makes about the properties of a variable; in particular, whether that variable meets the definition of nominal, ordinal, interval or ratio measurement. See also Nominal Scale, Ordinal Scale, Interval Scale, Ratio Scale

SED: Abbreviation of Standard Error of Difference of two means.

SEM: Abbreviation of Standard Error of Mean. See Standard Error of Mean

Semi-Interquartile Range: The semi-interquartile range is a measure of spread or dispersion. It is computed as one half the difference between the 75th percentile [often called (Q3)] and the 25th percentile (Q1). The formula for semi-interquartile range is therefore: (Q3-Q1)/2. Since half the scores in a distribution lie between Q3 and Q1, the semi-interquartile range is 1/2 the distance needed to cover 1/2 the scores. In a symmetric distribution, an interval stretching from one semi-interquartile range below the median to one semi-interquartile above the median will contain 1/2 of the scores. This will not be true for a skewed distribution, however. The semi-interquartile range is little affected by extreme scores, so it is a good measure of spread for skewed distributions. However, it is more subject to sampling fluctuation in normal distributions than is the standard deviation and therefore not often used for data that are approximately normally distributed.

Significance Level: In hypothesis testing, the significance level is the probability threshold used for rejecting the null hypothesis. The significance level is used in hypothesis testing as follows: First, the results of the experiment are compared with the results that would be expected if the null hypothesis were true. Then, assuming the null hypothesis is true, the probability of observing as or more extreme results is computed. Finally, this probability is compared to the significance level. If the probability is less than or equal to the significance level, then the null hypothesis is rejected and the outcome is said to be statistically significant. Traditionally, experimenters have used either the 0.05 level (sometimes called the 5% level) or the 0.01 level (1% level), although the choice of levels is largely subjective. The lower the significance level, the more the data must diverge from the null hypothesis to be significant. Therefore, the 0.01 level is more conservative than the 0.05 level. The Greek letter alpha (α) is used to indicate the significance level.

Significance Test: A significance test is performed to determine if an observed value of a statistic differs enough from a hypothesized value of a parameter to draw the inference that the hypothesized value of the parameter is not the true value. The hypothesized value of the parameter is called the "null hypothesis". A significance test consists of calculating the probability of obtaining a statistic as or more extreme than the statistic obtained in the sample assuming that the null hypothesis is correct. If this probability is sufficiently low, then the difference between the parameter and the statistic is said to be "statistically significant". Just how low is sufficiently low? The choice is somewhat arbitrary but by convention levels of 0.05 and 0.01 are most commonly used. For instance, in Plant Breeder's Rights varietal distinctness based on measured characteristics are often tested at 0.01 level.

Significant: A test is said to be significant if the test statistic supersedes a predetermined threshold.

Simple Effect: A simple effect of a factor is the effect at a single level of another factor. Often simple effects are computed following a significant interaction.

Size of Test: Synonym of Significance Level

Skewness: A measure of lack of symmetry of a distribution.

Spread: See Dispersion

Standard Deviation: It is the square root of the average squared deviation of each observation from the arithmetic mean. In other words it is the square root of variance. See Variance

Standard Error: The standard error of a statistic is the standard deviation of the sampling distribution of that statistic. Standard errors are important because they reflect how much sampling fluctuation a statistic will show. The inferential statistics involved in the construction of confidence intervals and significance testing are based on standard errors. The standard error of a statistic depends on the sample size. In general, the larger the sample size the smaller the standard error. The standard error of a statistic is usually designated by the Greek letter sigma (σ) with a subscript indicating the statistic. For instance, the standard error of the mean is indicated by the symbol: σ_M.

Standard Error of Mean: The standard error of the mean is designated as: σ_M. It is the standard deviation of the sampling distribution of the mean. The formula for the standard error of the mean is: $\sigma_M = \sigma/\sqrt{N}$, where σ is the standard deviation of the original distribution and N is the sample size (the number of scores each mean is based upon). This formula does not assume a normal distribution. However, many of the uses of the formula do assume a normal distribution. The formula shows that the larger the sample size, the smaller the standard error of the mean. More specifically, the size of the standard error of the mean is inversely proportional to the square root of the sample size.

Standard Normal Distribution: The standard normal distribution is a normal distribution with a mean of 0 and a standard deviation of 1. Normal distributions can be transformed to standard normal distributions by the formula:

$$Z = (X- \mu)/ \sigma$$

where X is a score from the original normal distribution, μ is the mean of the original normal distribution, and σ is the standard deviation of original normal distribution. The standard normal distribution is sometimes called the Z-distribution.

Standard Scores: When a set of scores are converted to z-scores, the scores are said to be standardized and are referred to as standard scores. Standard scores have a mean of 0 and a standard deviation of 1.

Standardized Coefficient: When an analysis is performed on variables that have been standardized so that they have variances of 1.0, the estimates that result are known as standardized coefficients; for example, a regression run on original variables produces unstandardized regression coefficients known as b's, while a regression run on standardized variables produces standardized regression coefficients known as betas. (In practice, both types of coefficients can be estimated from the original variables.)

Standardized Variable: A variable that has been transformed by multiplication of all scores by a constant and/or by the addition of a constant to all scores. Often these constants are selected so that the transformed scores have a mean of zero and a variance (and standard deviation) of 1.0.

Statistical Independence: A complete lack of covariation between variables, a lack of association between variables. When used in analysis of variance or covariance, statistical independence between the independent variables is sometimes referred to as a balanced design.

Statistical Measure: A number (a statistic) whose size indicates the magnitude of some quantity of interest e.g., the strength of a relationship, the amount of variation, the size of a difference, the level of income, etc. Examples include means, variances, correlation coefficients, and many others. Statistical measures are different from statistical tests. See also Statistical Test.

Statistical Method: Examples include Analysis of Variance (ANOVA), Modified Joint Regression Analysis, COYD, COYU, and many others.

Statistical Model:

Statistical Significance: Significance tests are performed to see if the null hypothesis can be rejected. If the null hypothesis is rejected, then the effect found in a sample is said to be statistically significant. If the null hypothesis is not rejected, then the effect is not significant. The experimenter chooses a significance level before conducting the statistical analysis. The significance level chosen determines the probability of a Type I error.

Statistical Test: A statistical test can be used to assess the probability that a statistical measure deviates from some pre-selected value (often zero) by no more than would be expected due to the operation of chance if the cases studied were randomly selected from a larger population. Examples include Pearson chi-square, F test, t test, and many others. Statistical tests are different from statistical measures. See also Statistical Measure and Hypothesis Testing.

Statistic: Any numerical quantity (such as the mean) calculated from a sample. Such statistics are used to estimate parameters. The term “statistics” sometimes refers to calculated quantities regardless of whether or not they are from a sample.

Statistics: The word “statistics” is used in several different senses. In the broadest sense, “statistics” refers to a range of techniques and procedures for analyzing data, interpreting data, displaying data, and making decisions based on data. This is what courses in “statistics” generally cover. In a second usage, statistics is used as the plural of statistic.

Stochastic Variable:

Student’s t-Distribution:

Symmetric Distribution:

t-Distribution:

Test: See Statistical Test

Test Statistic: A numerical quantity calculated from the observations with which a test is performed.

~~Trimmed Mean:~~

Transformation: A change made to the scores of all cases on a variable by the application of the same mathematical operation(s) to each score. (Common operations include addition of a constant, multiplication by a constant, taking logarithms, arcsine, ranking, bracketing, etc.)

t-Test: A t-test is any of a number of tests based on the t distribution. The general formula for t is:

t = (statistic – hypothesised value) / estimated standard error of statistic

The most common t-test is a test for a difference between two means.

Two-Point Scale: If each case is classified into one of two categories (e.g., present/absent, tall/dwarf, dead/alive) the variable is a two-point scale. For analytic purposes, two-point scales can be treated as nominal scales, ordinal scales, or interval scales.

Type I and Type II Error: There are two kinds of errors that can be made in significance testing: (1) a true null hypothesis can be incorrectly rejected and (2) a false null hypothesis can fail to be rejected. The former error is called a Type I error and the latter error is called a Type II error. These two types of errors are defined in the following table. The probability of a Type I error is designated by the Greek letter alpha (α) and is called the Type I error rate; the probability of a Type II error (the Type II error rate) is designated by the Greek letter beta (β). A Type II error is only an error in the sense that an opportunity to reject the null hypothesis correctly was lost.

		Statistical Decision	
		Reject H_o	Do not Reject H_o
True situation	H_o True	Type I error	Correct
	H_o False	Correct	Type II error

Type of Characteristic: See TGP/8

Type of Expression: See TGP/8

Unbalanced Data: Observations not coming from a balanced design.

Variability: See Dispersion

Variable: A variable is any measured characteristic or attribute that differs for different subjects. For example, if the height of 30 plants were measured, then height would be a variable. Variables can be quantitative or qualitative. (Qualitative variables are sometimes called “categorical variables”). Quantitative variables are measured on an ordinal, interval, or ratio scale; qualitative variables are measured on a nominal scale.

Variance: The variance is a measure of how spread out a distribution is. It is computed as the average squared deviation of each observation from its arithmetic mean. Standard deviation is measured as the square root of variance. Both variance and standard deviation are measures of dispersion of data.

Variance Component: variance estimate of a random term in a mixed model.

Variation: See Dispersion

Weighted Data: Weights are applied when one wishes to adjust the impact of cases in the analysis, e.g., to take account of the number of population units that each case represents. In sample surveys weights are most likely to be used with data derived from sample designs having different selection rates or with data having markedly different subgroup response rates.

Within plot standard deviation: When speaking about variance components this term is commonly used for the variability within experimental units, e.g. within plots. For example, if observations are made on several plants on the same plot it is the standard deviation between these plants.

Z-Distribution: The standard normal distribution is sometimes called the Z-distribution. See Standard Normal Distribution

[End of Section 3]