

CHAPTER 3

COMMONLY USED STATISTICAL TERMS

There are many statistics used in social science research and evaluation. The two main areas of statistics are descriptive and inferential. The third class of statistics is design and experimental statistics. *Descriptive* statistics involve the tabulating, depicting, and describing of collections of data. These data may be either quantitative or qualitative. They provide a picture or description of the properties of data collected in order to summarize them into manageable form. *Inferential* statistics are a formalized body of techniques that infer the properties of a larger collection of data from the inspection of that collection. They build on these statistics as they infer the properties of samples to various populations. *Design and analysis* statistics were developed for the discovery and confirmation of causal relationships among variables in social science experiments. They use a variety of statistical tests related to aspects such as prediction and hypothesis testing. *Experimental analysis* is related to

As in the previous glossary chapter, if you are looking for a term and it is not here, please send it to my e-mail address, mholosko@uga.edu, with the subject line: "New Statistical Terms Needed." It will then be added in the next edition.

There are many statistics used in social science research and evaluation. The two main areas of statistics are descriptive and inferential. The third class of statistics is design and experimental statistics. *Descriptive* statistics involve the tabulating, depicting, and describing of collections of data. These data may be either quantitative or qualitative. They provide a picture or description of the properties of data collected in order to summarize them into manageable form. *Inferential* statistics are a formalized body of techniques that infer the properties of a larger collection of data from the inspection of that collection. They build on these statistics as they infer the properties of samples to various populations. *Design and analysis* statistics were developed for the discovery and confirmation of causal relationships among variables in social science experiments. They use a variety of statistical tests related to aspects such as prediction and hypothesis testing. *Experimental analysis* is related to

As in the previous glossary chapter, if you are looking for a term and it is not here, please send it to my e-mail address, mholosko@uga.edu, with the subject line: "New Statistical Terms Needed." It will then be added in the next edition.

comparisons, variance, and ultimately testing whether variables are significant between each other. The latter two types of statistics are usually either parametric or nonparametric. The importance of statistics in the research process is sometimes exaggerated. Thus, a highly sophisticated statistical analysis rarely, if ever, compensates for a poorly conceived project, a poorly constructed research design, or an inaccurate data collection instrument. Thus, statistics certainly may aid the researcher but are never a substitute for good, sound thinking and attention to the scientific method and research process. For researchers, then, statistics are simply a tool to help them study the phenomena they are interested in.

DESCRIPTIVE STATISTICS

Measures of Central Tendency

Mean, arithmetic mean (\bar{X} or M): The sum of the scores in a distribution divided by the number of scores in the distribution. It is the most commonly used measure of central tendency. It is often reported with its companion statistic, the standard deviation, which shows how far things vary from the average.

Median (Mdn): The midpoint or number in a distribution having 50% of the scores above it and 50% of the scores below it. If there are an odd number of scores, the median is the middle score.

Mode (Mo): The number that occurs most frequently in a distribution of scores or numbers. In some fields, notably education, sample data are often called scores, and the sample mode is known as the modal score.

Measures of Variability

Interquartile range (IQR): A measure of statistical dispersion being equal to the difference between the third and first quartiles. The first quartile (designated Q_1) is the lower and cuts off the lowest 25% of data (the 25th percentile); the second quartile (Q_2), or the median, cuts the data set in half (the 50th percentile); and the third quartile (Q_3) cuts off highest 25% of data, or the lowest 75% (the 75th percentile).

Range (Ra): The difference between the highest and lowest scores in a distribution; a measure of variability.

Standard deviation (SD): The most stable measure of variability, it takes into account each and every score in a normal distribution. This descriptive statistic assesses how far individual scores vary in standard unit lengths from its midpoint of 0. For all normal distributions, 95% of the area is within 1.96 standard deviations of the mean.

Variance (SD²): A measure of the dispersion of a set of data points around their mean value. It is a mathematical expectation of the average squared deviations from the mean.

Inferential Statistical Tests

Tests concerned with using selected sample data compared with population data in a variety of ways are called *inferential statistical tests*. There are two main bodies of these tests. The first and most frequently used are called *parametric statistical tests*. The second are called *nonparametric tests*. For each parametric test, there may be a comparable nonparametric test, sometimes even two or three.

Parametric tests are tests of significance appropriate when the data represent an interval or ratio scale of measurement

and other specific assumptions have been met, specifically, that the sample statistics relate to the population parameters, that the variance of the sample relates to the variance of the population, that the population has normality, and that the data are statistically independent.

Nonparametric tests are statistical tests used when the data represent a nominal or ordinal level scale or when assumptions required for parametric tests cannot be met, specifically, small sample sizes, biased samples, an inability to determine the relationship between sample and population, and unequal variances between the sample and population. These are a class of tests that do not hold the assumptions of normality.

In the list of statistical terms below, when the test is a parametric test, the designation of *PT will be used at the end of the definition. Conversely, when the test is a nonparametric test, the designation of *NPT will be used at the end of the definition.

Statistical Terms

Alpha coefficient (a): See *Cronbach's alpha coefficient*.

Analysis of covariance (ANCOVA): A statistical technique for equating groups on one or more variables when testing for statistical significance using the *F*-test statistic. It adjusts scores on a dependent variable for initial differences on other variables, such as pretest performance or IQ. *PT

Analysis of variance (ANOVA): A statistical technique for determining the statistical significance of differences among means; it can be used with two or more groups and uses the *F*-test statistic. *PT

Autoregressive integrated moving average (ARIMA): This statistic is a Box-Jenkins approach to time series analysis. It tests for changes in the data patterns pre- and

postintervention within the context of analyzing the outcomes of a time series design.

Binomial test: An exact test of the statistical significances of derivations from a theoretically expected distribution of observations into two categories. *NPT

Chi-square (χ^2): A nonparametric test of statistical significance appropriate when the data are in the form of frequency counts; it compares frequencies actually observed in a study with expected frequencies to see if they are significantly different. *NPT

Cochran's Q: Used to evaluate the relation between two variables that are measured on a nominal scale. One of the variables may even be dichotomous or consisting of only two possible values. *NPT

Coefficient of determination (r^2): The square of the correlation coefficient (r), it indicates the degree of relationship strength by potentially explained variance between two variables.

Cohen's d : A standardized way of measuring the effect size or difference by comparing two means by a simple math formula. It can be used to accompany the reporting of a t -test or ANOVA result and is often used in meta-analysis. The conventional benchmark scores for the magnitude of effect sizes are as follows: small, $d = 0.2$; medium, $d = 0.5$; large, $d = 0.8$. *NPT

Cohen's kappa (κ): A statistical measure of interrater agreement for qualitative (categorical) items. Scores range from -1.0 to 1.0 . *NPT

Confidence interval (CI): Quantifies the uncertainty in measurement. It is usually reported as a 95% CI, which is the range of values within which it can be 95% certain that the true value for the whole population lies. For example,

for a number needed to treat (NNT) of 10 with a 95% CI of 5 to 15, there would be 95% confidence that the true NNT value lies between 5 and 15.

Correlation coefficient (r): A decimal number between 0.00 and ± 1.00 that indicates the degree to which two quantitative variables are related. The most common one used is the Pearson Product Moment correlation coefficient or just the Pearson coefficient.

Cronbach's alpha coefficient (α): A coefficient of consistency that measures how well a set of variables or items measures a single, unidimensional, latent construct in a scale or inventory. Alpha scores are conventionally interpreted as follows: high, “.90; medium, 0.70 to 0.89; and low, 0.55 to 0.69.

Cumulative frequency distribution: A graphic depiction of how many times groups of scores appear in a sample.

Dependent t -test: A data analysis procedure that assesses whether the means of two related groups are statistically different from each other, for example, one group's mean score (time one) compared with the same group's mean score (time two). It is also called the *paired samples t-test*. *PT

Effect size (δ): Any measure of the strength of a relationship between two variables. Effect size statistics are used to assess comparisons between correlations, percentages, mean differences, probabilities, and so on.

Eta (μ): An index that indicates the degree of a curvilinear relationship.

F -test (F): A parametric statistical test of the equality of the means of two or more samples. It compares the means and variances between and within groups over time. It is also called *analysis of variance (ANOVA)*. *PT

Factor analysis: A statistical method for reducing a set of variables to a smaller number of factors or basic

components in a scale or instrument being analyzed. Two main forms are *exploratory (EFA)* and *confirmatory factor analysis (CFA)*. *PT

Fisher's exact test: A nonparametric statistical significance test used in the analysis of contingency tables where sample sizes are small. The test is useful for categorical data that result from classifying objects in two different ways; it is used to examine the significance of the association (contingency) between two kinds of classifications. *NPT

Friedman two-way analysis of variance: A nonparametric inferential statistic used to compare two or more groups by ranks that are not independent. *NPT

G²: This is a more conservative goodness-of-fit statistic than the χ^2 and is used when comparing hierarchical models in a categorical contingency (two-by-two) table.

Independent t-test: A statistical procedure for comparing measurements of mean scores in two different groups or samples. It is also called the *independent samples t-test*. *PT

Kendall's tau (c): A nonparametric statistic used to measure the degree of correspondence between two rankings and to assess the significance of the correspondence. *NPT

Kolmogorav-Smirnov (K-S) test: A nonparametric goodness-of-fit test used to decide if a sample comes from a population with a specific distribution. The test is based on the empirical distribution function (ECDF). *NPT

Kruskal-Wallis one-way analysis of variance: A nonparametric inferential statistic used to compare two or more independent groups for statistical significance of differences. *NPT

Mann-Whitney U-test (U): A nonparametric inferential statistic used to determine whether two uncorrelated groups differ significantly. *NPT

McNemar's test: A nonparametric method used on nominal data to determine whether the row and column marginal frequencies are equal. *NPT

Median test: A nonparametric test that tests the null hypothesis that the medians of the populations from which two samples are drawn are identical. *NPT

Multiple correlation (R): A numerical index describing the relationship between predicted and actual scores using multiple regression. The correlation between a criterion and the best combination of predictors.

Multivariate analysis of covariance (MANCOVA): An extension of ANOVA that incorporates two or more dependent variables in the same analysis. It is an extension of MANOVA where artificial dependent variables (DVs) are initially adjusted for differences in one or more covariates. It computes the multivariate F statistic. *PT

Multivariate analysis of variance (MANOVA): It is an ANOVA with several dependent variables. *PT

Newman-Keuls test: A type of *post hoc* or *a posteriori* multiple comparison test of data that makes precise comparisons of group means after ANOVA has rejected the null hypothesis. *NPT

One-way analysis of variance (ANOVA): An extension of the independent group t -test where you have more than two groups. It computes the difference in means both between and within groups and compares variability between groups and variables. Its parametric test statistic is the F -test. *PT

Pearson correlation coefficient (r): This is a measure of the correlation or linear relationship between two variables x and y , giving a value between +1 and -1 inclusive. It is

widely used in the sciences as a measure of the strength of linear dependence between two variables. *PT

Pooled point estimate: An approximation of a point, usually a mean or variance, that combines information from two or more independent samples believed to have the same characteristics. It is used to assess the effects of treatment samples versus comparative samples.

Post hoc test: A *post hoc* test (or *post hoc* comparison test) is used at the second stage of the analysis of variance (ANOVA) or multiple analyses of variance (MANOVA) if the null hypothesis is rejected.

Runs test: Where measurements are made according to some well-defined ordering, in either time or space. A frequent question is whether or not the average value of the measurement is different at different points in the sequence. This nonparametric test provides a means for this. *NPT

Siegel-Tukey test: A nonparametric test named after Sidney Siegel and John Tukey, which tests for differences in scale between two groups. Data measured must at least be ordinal. *NPT

Sign test: A test that can be used whenever an experiment is conducted to compare a treatment with a control on a number of matched pairs, provided the two treatments are assigned to the members of each pair at random. *NPT

Spearman's rank order correlation (ρ): A nonparametric test used to measure the relationship between two rank-ordered scales. Data are in ordinal form. *NPT

Standard error of the mean (SEM): An estimate of the amount by which an obtained mean may be expected to differ by chance from the true mean. It is an indication of how well the mean of a sample estimates the mean of a population.

Statistical power: The capability of a test to detect a significant effect or how often a correct interpretation can be reached about the effect if it were possible to repeat the test many times.

Student-Newman-Keuls (SNK) test: A nonparametric post-ANOVA test, also called a *post hoc* test. It is used to analyze the differences found after the performed *F*-test (ANOVA) is found to be significant, for example, to locate where differences truly occur between means. *NPT

Student *t*-test (*t*): Any statistical hypothesis test in which the test statistic follows a Student's *t* distribution if the null hypothesis is true, for example, a *t*-test for paired or independent samples. *PT

***t*-distribution:** A statistical distribution describing the means of samples taken from a population with an unknown variance.

***T*-score:** A standard score derived from a *z*-score by multiplying the *z*-score by 10 and adding 50. It is useful in comparing various test scores to each other as it is a standard metric that reflects the cumulative frequency distribution of the raw scores.

***t*-test for correlated means:** A parametric test of statistical significance used to determine whether there is a statistically significant difference between the means of two matched, or nonindependent, samples. It is also used for pre-post comparisons. *PT

***t*-test for correlated proportions:** A parametric test of statistical significance used to determine whether there is a statistically significant difference between two proportions based on the same sample or otherwise nonindependent groups. *PT

***t*-test for independent means:** A parametric test of significance used to determine whether there is a statistically

significant difference between the means of two independent samples. *PT

***t*-test for independent proportions:** A parametric test of statistical significance used to determine whether there is a statistically significant difference between two independent proportions. *PT

Tukey's test of significance: A single-step multiple comparison procedure and statistical test generally used in conjunction with an ANOVA to find which means are significantly different from one another. Named after John Tukey, it compares all possible pairs of means and is based on a studentized range distribution q (this distribution is similar to the distribution of t from the *t*-test).

Wald-Wolfowitz test: A nonparametric statistical test used to test the hypothesis that a series of numbers is random. It is also known as the runs test for randomness. *NPT

Wilcoxon sign rank test (W_+): A nonparametric statistical hypothesis test for the case of two related samples or repeated measurements on a single sample. It can be used as an alternative to the paired Student's *t*-test when the population cannot be assumed to be normally distributed. *NPT

Wilks's lambda (h): A general test statistic used in multivariate tests of mean differences among more than two groups. It is the numeral index calculated when carrying out MANOVA or MANCOVA.

Z-score: A score expressed in units of standard deviations from the mean. It is also known as a *standard score*.

Z-test: A test of any of a number of hypotheses in inferential statistics that has validity if sample sizes are sufficiently large and the underlying data are normally distributed.