

Statistical Terms to Guide the Review of Published Research

NR714 Analytic Methods II

Week 4 and Week 5

Absolute value: The positive numeric value of a number (the minus sign in front of the number is disregarded).

Abstract: A short description of a completed research project. Typically, it is between 150 and 300 words and contains information on the purpose, methods, results, and conclusions of the study.

Alpha level: The p -value defined by the researcher as being statistically significant. It is the chance that the researcher is willing to take of committing a type I error. The most commonly used alpha levels are .05; .01, and .10.

Alternative hypothesis (H_a): The hypothesis that states a statistically significant relationship exists between the variables. It is the hypothesis opposite to the null hypothesis. It is also referred to as the “acting” hypothesis or the research hypothesis.

Analysis of variance (ANOVA): A **parametric** statistical technique used to compare the means of three or more groups as defined by one or more factors.

Baseline: Measures taken at the start of a study before any interventions; sometimes referred to as the pretest.

Bell-shaped: A graphical shape, typical of the normal distribution.

Beta Level: The probability of making a type II error.

Between-group variance: The extent to which the group means deviate from the grand mean.

Bonferroni t-test: A post-hoc test used in ANOVA to help determine which of the means are significantly different from each other after an ANOVA has been found to be statistically significant. It is a type of t-test that adjusts the alpha level for multiple comparisons that are being made.

Causal relationship: A relationship in which one or more variables is presumed to cause changes in another variable.

Chi-square: A **nonparametric** test used to assess whether a statistically significant association exists between the rows and columns in a cross-tabulation (contingency) table. The chi-square test statistic is used to compare the proportion of participants having a given characteristic among different groups.

Confidence interval: The range of values within which the population parameter is estimated to fall within.

Continuous variable: A variable that can take on any possible value within a range. For example, weight is a continuous variable because a weight of 152.5 lb makes sense. In contrast, number of children is a discrete variable because it can take on only certain values (0, 1, 2, and so on). A value of 1.2 for children does not make any conceptual sense.

Control group: The group that is used for comparison in an experimental or quasi-experimental study.

Correlation coefficient: A measure of the extent to which the variation in one variable is related to the variation in another variable.

Critical value: The value or values that determine the rejection region or threshold for a specific computed statistic to be defined as statistically significant (e.g., that will result in the researcher's rejecting the null hypothesis). The critical value of each statistic is based on the distribution of that particular statistic. Depending on the particular statistical test, the computed value must either exceed or be less than the critical value for the test to be defined as statistically significant.

Cross-tabulation: A way of presenting the relationship between two variables in a table format; the rows and columns of the table are labeled with the values of the variables.

Data set: Collection of different values of all the variables used to measure the characteristics of the sample or population.

Degrees of freedom: The freedom of a variable's values to vary given what is already known about the other values and the sum of the values.

Dependent variable: The variable that measures the effect of some other variable (e.g., the variable whose values are expected to be predicted by the independent variable). Also referred to as the outcome variable or the response variable.

Descriptive statistics: Statistics used to summarize and describe data.

Dichotomous variable: A nominal variable having only two categories.

Dummy variables: A set of dichotomous variables used to represent a nominal variable with three or more categories in regression models.

Effect size: The extent to which the independent variable influences the value of the dependent variable.

Empirical study: Study based on observation or experience.

Estimation: The procedure for testing a model in which a sample of data is used to make an estimate of the values for the whole population.

Experiment: A research study with the following characteristics: an intervention that the investigator controls and that only some groups receive, random selection of participants into the study, and random assignment of participants to intervention and control groups.

External validity: The extent to which the results of a study can be generalized to other populations or settings than the sample that was studied.

Extraneous variable: A variable that confounds the relationship between the dependent variable and the independent variable.

Factor(s): In an ANOVA model, the variable(s) that defines the groups whose means are being compared.

Fisher's exact test: An alternative to the chi-square test that can be used when the expected frequencies in each cell are too small.

Frequency distribution: A systematic array of data together with a count of the raw frequency that each value occurs, the relative frequency with which it occurs, and the cumulative frequency with which it occurs.

Friedman's ANOVA by rank: A **nonparametric** analogue to the repeated-measures ANOVA. It is often used to compare rankings and preferences.

F-statistic: A statistic computed by measuring the ratio of two variances to each other.

Generalizability: The extent to which the research findings can be applied to situations beyond those of the immediate group that was studied. The extent to which the findings can be used to make inferences about the population that the sample came from.

Heterogeneity of variance: Significant differences in the variance of the values of the dependent variable within two or more groups that are being compared with each other.

Histogram: A way of graphically displaying ordinal, interval, and ratio-level data. It shows the shape of the distribution.

Homogeneity of variance: When there are no significant differences in the variance of the values of the dependent variable within two or more groups that are being compared with each other.

Hypothesis: A formal statement of the expected relationship between two or more variables.

Independent random sample: A sample in which the value of the variables for each subject are not related to the value of the variables for the other subjects and in which each and every subject has an equal chance of being selected to be in the study.

Independent samples: Groups or samples in which the measurements of the members of one sample have no preexisting relationship to the measurements of members of the other sample(s).

Independent t-test: A **parametric** test used to determine if the means of two independent groups are significantly different from each other.

Independent variable: A variable that is expected to influence the value of another variable.

Inference: A conclusion about a population drawn from results based on a sample of data from that population.

Institutional review board (IRB): An established body of individuals that convenes to review proposed and ongoing research studies to protect the welfare of human subjects recruited to participate in biomedical or behavioral research. IRBs are typically located on college campuses, medical facilities, and governmental institutions.

Internal validity: The extent to which the findings of a study truly and accurately represent the relationship between the independent variable(s) and the dependent variable.

Interquartile range: The range of values from the 25th to the 75th percentile; the middle 50% of the data.

Interval-level measurement: A rank order scale with equal intervals between units but no true zero. IQ scores, SAT scores, and GRE scores are all examples of interval level data.

Kruskal-Wallis H-test: A **nonparametric** alternative to ANOVA. It is used to determine the difference in the medians between three or more groups.

Likert scale: A measurement scale that asks respondents to register the level to which they agree or disagree with a set of statements. There are typically five to seven response categories that range from strongly disagree to strongly agree. Other descriptors, such as level of satisfaction, may be used instead of level of agreement.

Mann-Whitney U-test: The **nonparametric** analogue of the independent t-test. It is used to determine the statistical significance of the difference in the medians of two independent groups.

Matched-pairs design: Test design in which study participants are matched based on characteristics that might influence their treatment reactions.

McNemar test: A **nonparametric** test used to examine differences in paired dichotomous measures.

Mean: A measure of central tendency. It is the arithmetic average of a set of data.

Measurement: The assignment of numbers to objects or events.

Median: A measure of central tendency. It is the central point or middle value of an ordered set of data.

Missing values: Values that are missing from a variable for some participants. These values may be missing because the research subject refused to answer certain questions or because certain questions do not apply to the subject (e.g., the question, "Are you pregnant?" would be missing for male study subjects).

Mode: A measure of central tendency. It is the most frequently occurring score in a population or data set.

Mutually exclusive and exhaustive categories: Each subject (e.g., item, event) can fit into one and only one category, and each subject (e.g., item, event) fits into a category.

***N*:** The total number of subjects in a research study across all groups.

***n*:** The number of subjects in specific sub-groups.

Negative relationship: As the value of one variable increases, the value of the other decreases. Also called an inverse relationship.

Negatively skewed distribution: Asymmetric distribution that has a disproportionate number of cases with high values and a tail that points to the left. Also called a left-skewed distribution.

Nominal measure: A measurement scale in which the numbers have no intrinsic meaning but are merely used to label different categories. Ethnic identity, religion, and health insurance status (e.g., none, Medicaid, Medicare, private) are all examples of nominal level data.

Nonparametric tests: Statistical tests designed to be used when the data fail to meet one or more of the assumptions required of **parametric** tests. These tests are “distribution free” but usually have less power than **parametric** tests.

Normal distribution: A theoretical probability distribution in which the horizontal axis represents all possible values of a variable and the vertical axis represents the probability that those values will occur. Normal distributions are unimodal (mean, median, and mode are the same), symmetrical about the mean, and have a shape commonly described as a bell-shaped curve.

Null hypothesis: The hypothesis that states that two or more variables being compared will not be related to each other (i.e., no significant relationship between the variables will be found).

Odds ratio: The ratio of the probability of occurrence over the probability of non-occurrence of an event.

Ordinal scale: A measurement scale that ranks subjects on some variable. The interval between the ranks does not necessarily have to be equal. Examples of ordinal variables are scale items that measure any subjective state (e.g., happiness: very happy, somewhat happy, somewhat unhappy, very unhappy; attitude: strongly agree, somewhat agree, somewhat disagree, strongly disagree; and military rank: general, colonel, sergeant, private).

Outliers: Extreme values of a variable that are at the tail end of the distribution of the data.

***p*-value:** The actual probability of getting the obtained results or results even more extreme. The smaller the *p*-value, the more statistically significant (i.e., the less likely the result is due to chance).

Paired t-test: Test used to determine if a relationship exists between two correlated means.

Parameter: A characteristic of a population.

Parametric tests: Tests based on the assumption that the data are normally distributed.

Pearson correlation coefficient: A **parametric** test used to determine if a linear association exists between two measures of interval or ratio measurement scale. The variables need to be normally distributed.

Percentile: The score that is at or below a certain percentage of the scores.

Population: The entire group having some characteristic (e.g., all people with depression, all residents of the United States). Often a sample is taken of the population and then the results are generalized to that population.

Post-hoc tests: Tests of paired comparisons made when an overall test, such as an ANOVA, is statistically significant. Post-hoc tests are used to control for the problems caused by multiple comparisons. Many post-hoc tests, including the Scheffe test, Tukey's honestly significant difference test, and Bonferroni post-hoc test, are available.

Positive relationship: Also called a "direct" relationship. The values of x and y increase or decrease together. As x increases, y also increases.

Positively skewed distribution: An asymmetric distribution with a disproportionate number of cases with a low value. The tail of this distribution points to the right. Also known as a right-skewed distribution.

Power: The probability that the null hypothesis will be correctly accepted by the test.

Probability: The likelihood that an event will occur; it is expressed as a number from 0 to 1.

Quartile: The four "quarters" of the data distribution. The first quartile is the 25th percentile, the second quartile is the 50th percentile, the third quartile is the 75th percentile, and the fourth quartile is the 100th percentile.

Randomization: The assignment of study subjects to groups by chance in which each and every subject has an identical chance of selection.

Range: The difference between the largest and the smallest value in the dataset.

Ratio scale: A measurement scale in which there are both equal intervals between units and a true zero. Most biologic measures (e.g., weight, pulse rate) are ratio-level variables.

Rejection region: The region of a statistical distribution not under the null hypothesis. The null hypothesis is rejected if the computed value of the sample statistic falls into the rejection region. These regions are typically located at the extreme tail(s) of a sampling distribution.

Reliability: The degree of consistency with which an instrument measures what it purports to measure. Reliability can be broken down into test-retest reliability, interrater reliability, and internal consistency.

Repeated ANOVA: A **parametric** technique used to compare the means of three or more related groups.

Sample: A group selected from the population in the hope that the smaller group will be representative of the entire population.

Scatter diagram: A diagram that graphically represents the relationship of two ordinal, interval, or ratio level variables to each other. The diagram is typically presented with correlation coefficients.

Significance level: The p -value; the probability of rejecting the null hypothesis when it is true.

Spearman correlation coefficient: A **nonparametric** correlation coefficient.

Standard deviation: The average distance of each point from the mean.

Theory: A well-tested and systematic explanation about the relationships that make up a phenomenon.

Two-tailed test: A test of statistical significance in which values at both extremes (tails) of a distribution are considered in determining the statistical significance of an analysis.

Type I error: Rejecting the null hypothesis when it is true.

Type II error: Accepting the null hypothesis when it is false.

Valid percent: The percent of subjects in each category after the missing data are excluded.

Validity: The extent to which an instrument measures what it intends to measure; the extent to which the measurements are "true."

Variable: Any characteristic that can be expressed with more than one value.

Variance: A measure of the spread or dispersion of scores in a distribution of interval or ratio level data. The square root of the variance of a variable is the standard deviation.

Wilcoxon matched-pairs test: A **nonparametric** test analogous to the paired t -test. It determines whether the medians of two related groups are equal.

z-scores: A standardized score that represents the values in a data set as the number of standard deviations from the mean.