This guide contains a summary of the statistical terms and procedures. This guide can be used as a reference for course work and the dissertation process. However, it is recommended that you refer to statistical texts for an in-depth discussion of each of these topics.

Statistics Guide

Prepared by: Amanda J. Rockinson- Szapkiw, Ed.D.

Contents

Foundational Terms and Constructs

Introduction to Statistics: Two Major Branches

Broadly speaking, educational researchers use two research methods:

Quantitative research is characterized by objective measurement, usually numbers.

Qualitative research emphasizes the in depth understanding of the human experience, usually using words.

Statistical procedures are sometimes used in qualitative research; however, they are predominantly used to analyze and interpret the data in quantitative research. Accordingly, **statistics** can be defined as a numerical summary of data. There are two major branches of statistics:

Descriptive Statistics, the most basic form of statistics, are procedures for summarizing a group of scores to make them more comprehendible. We use descriptive statistics to describe a phenomenon.

Inferential Statistics are procedures that help researchers draw conclusions based on informational data gathered. Researchers use inferential statistics when they want to make inferences that extend beyond the immediate data alone.

Before you can understand these two major branches of statistics and the different type of procedures that fall under each, you first need to understand a few basics about variables.

Variables

A **variable** is any entity that can take on different values. We can distinguish between two major types of variables:

1. **Quantitative Variables** – numeric (e.g., test score, height, etc.)
2. **Qualitative Variables** – nonnumeric or categorical (e.g., sex, political party affiliation, etc)

Variables are typically measured on one of four levels of measurement: (a) nominal, (b) ordinal, (c) interval or (d) ratio. The level of measurement of a variable has important implications for the type of statistical procedure used.

Nominal Scale: Nominal variables are variables that can be qualitatively classified into discrete, separate categories. The categories cannot be ordered. Religion (Christian, Jewish, Islam, Other),

intervention or treatment groups (Group A, Group B), affiliations, and outcome (success, failure) are examples of nominal level variables.

Ordinal Scale: Ordinal variables are variables that can be qualitatively classified into discrete, separate categories (like nominal variables) and rank ordered (unlike nominal variables). Thus, with ordinal variables, we can talk in terms of which variable has less and which has more. The socioeconomic (SES) status of families (high, low, medium), height when we do not have an exact measurement tool (tall, short), military rank (private, corporal, sergeant), school rank (sophomore, junior, senior) and ratings (excellent, good, average, below average, poor; or agree, neutral, disagree) are all examples of ordinal variables. Although there is some disagreement, IQ is also often considered an ordinal scale variable. Likert-type scale data is also ordinal data; however, in social science research, we treat it as ratio or interval level data.

Interval Scale: Interval variables are variables that can be qualitatively classified into discrete, separate categories (like nominal and ordinal variables) and rank ordered (like ordinal variables), and measured. Interval variables have an arbitrary zero rather than a true zero. Temperature, as measured in degrees Fahrenheit or Celsius, is a common example of an interval scale. For example, we can say that a temperature of 40 degrees is higher than a temperature of 30 degrees (rank). Additionally, we can say that an increase from 20 to 40 degrees is twice as much as an increase from 30 to 40 degrees (measurement). However, since there is no arbitrary zero, we cannot say that it is twice as high.

Ratio Scale: Ratio variables have all the characteristics of the preceding variables in addition to a true (or absolute) zero point on a measurement scale. Weight (in pounds), height (in inches), distance (in yards), and speed (in miles per hour) are examples of ratio variables.

While we are discussing variables, let's also review a few additional terms and classifications. When conducting research, it is important to identify or classify variables in the following manner:

Independent variables (IV), also known as predictor variables in regression studies, are variables we expect to affect other variables. This is usually the variable we, as the researcher, plan to manipulate. It is important to recognize that the IV are not always manipulated by the researcher. For example, when using an *ex post facto* design, the IV is not manipulated by the researcher (i.e., the researcher came along after the fact). IV used in an *ex post facto* design may be are gender (male, female) or smoking condition (smoker, nonsmoker). Clearly, the researcher

did not manipulate gender or smoking condition. However, the IV is the variable that is manipulated or was affected.

Dependent variables (DV), also known as criterion variables in regression studies, are variables we expect to be affected by other variables. This is usually the variable that is measured.

In this research question, what is the IV and DV?

Do at-risk high school seniors who participate in a study skills program have a higher graduation rate than at-risk high school seniors who do not participate in a study skills program?

IV = participation in study skills program
DV = graduation rate

You can visualize the IV and DV as follows:

IV (cause; possible cause) $\longrightarrow$ DV (effect; outcome)

Additional types of variables to be aware of include:

- **Mediator variables** (or intervening) – variables that are responsible for the relationship between two variables (i.e., help to explain why the relationship exists).
- **Moderating variables** – a variable that influences the strength or direction of the relationship that exists between variables (i.e., specify conditions when the relation exists).
- **Confounding variables** are variables that influence the dependent variable and are often used as control variables.
- **Variables of Interest** – variables that are being studied in a correlational study, when it is arbitrary to use the labels of IV and DV.

Descriptive Statistics

Remember that with descriptive statistics, you are simply describing a sample. With inferential statistics, you are trying to infer something about a population by measuring a sample taken from the population.

Frequency Distributions

In its simplest form, a **distribution** is just a list of the scores taken on some particular variable. For example, the following is a distribution of 10 students' scores on a math test, arranged in order from lowest to highest:

69, 77, 77, 77, 84, 85, 85, 87, 92, 98

The **frequency** (*f*) of a particular data set is the number of times a particular observation occurs in the data. And the **frequency distribution** is the pattern of frequencies of the observations or listing of case counts by category. Frequency distributions can show either the actual number of observations falling in each range or the percentage of observations (the proportion per category). They can be shown using as frequency tables or graphics.

Frequency Table- is a chart presenting statistical data that categorizes the values along with the number of times each value appears in the data set. See the example below.

Table 1
Distribution of Students' Test Scores

	F	%
60-69	1	10%
70-79	3	30%
80-89	4	40%
90-99	2	20%

Frequency distributions also may be represented graphically; here are a few examples:

A **bar graph** is a chart with rectangular bars. The length of each bar is proportional to the value that it represents. Bar graphs are used for nominal data to indicate the frequency of distribution.

A **histogram** is a graph that consists of a series of columns; each column represents an interval having a category of a variable. The frequency of occurrence is represented by the column's height. A histogram is useful in graphically displaying interval and ratio data.

A **pie chart** is a circular chart that provides a visual representation of the data (100% = 360 degrees). The pie is divided into sections that correspond to a category of the variable (i.e. age, height, etc.). The size of the section is proportional to the percentage of the corresponding category. Pie charts are especially useful for summarizing nominal variables.

Central Tendency

Measures of central tendency represent the "typical" attributes of the data. There are three measures of central tendency.

The **mean (*M*)** is the arithmetic average of a group of scores or sum of scores divided by the number of scores. For example, in our distribution of 10 test scores,

$$69 + 77 + 77 + 77 + 84 + 85 + 85 + 87 + 92 + 98/ 10 = 83.1$$

The mean is used when working with interval and ratio data and is the base for other statistics such as standard deviation and t- tests. Since the mean is sensitive to extreme scores, it is the best measure for describing normal, unimodal distributions. The mean is not appropriate in describing a highly skewed distribution.

The **median (*Mdn*)** is the middle score of all the scores in a distribution arranged from highest to lowest. It is the midpoint of distribution when the distribution has an odd number of scores. It is the number halfway between the two middle scores when the distribution has an even number of scores. The median is useful to describe a skewed distribution. For example, in our distribution of 10 test scores, 84.5 is the median.

<div align="center">

69, 77, 77, 77, 84,   **<u>84.5</u>**   85, 85, 87, 92, 98

</div>

The median is useful when working with ordinal, ratio, and interval data. Of the three measures of central tendency, it is the least effected by extreme values. So, it is useful to describe a skewed distribution.

The **mode (*Md*)** is the value with the greatest frequency in the distribution. For example, in our distribution of 10 test scores, 77 is the mode because it is observed most frequently. The mode is useful when working with nominal, ordinal, ratio, or interval data.

Table 2
Levels of Measurement and the Best Measure of Central Tendency

Measurement	Measure of Central Tendency
Nominal	Mode
Ordinal	Median
Interval	Symmetrical data: Mean Skewed data: Median
Ratio	Symmetrical data: Mean Skewed data: Median

Measures of Dispersion

Without knowing something about how data is dispersed, measures of central tendency may be misleading. For example, a group of 20 students may come from families in which the mean income is $200,000 with little variation from the mean. However, this would be very different than a group of 20 students who come from families in which the mean income is $200,000 and 3 of students' parents make a combined income of $1 million and the other 17 students' parents

make a combined income of $60,000. The mean is $200,000 in both distributions; however, how they are spread out is very different. In the second scenario, the mean is affected by the extremes. So, to understand the distribution it is important to understand its dispersion. Measures of dispersion provide a more complete picture of the data set. Dispersion measures include the range, variance, and standard deviation.

The **range** is the distance between the minimum and the maximum. It is calculated by taking the difference between the maximum and minimum values in the data set $(X_{max} - X_{min.})$. The range only provides information about the maximum and minimum values, but it does not say anything about the values between. For example, in our distribution of 10 test score, the range would be (98-69) = 29.

The **variance** of a data set is calculated by taking the arithmetic mean of the squared differences between each value and the mean value. It is symbolized by the Greek letter sigma squared for a population and s^2 for a sample. It tells about the variation of the data and is foundational to other measures such as standard deviation. In our example of 10 scores, the variance is 70.54.

The **standard deviation**, a commonly used measure of dispersion, is the square root of the **variance**. It measures the dispersion of scores around the mean. The larger the standard deviation is, the larger the spread of scores around the mean and each other. In our example of 10 scores, the standard deviation is the square root of 70.54, which is equal to 8.39. For normally distributed data values, approximately 68% of the distribution falls within ± 1 *SD* of the mean, 95% of the distribution falls within ± 2 *SD*s of the mean, and 99.7% of the distribution falls within ± 3 *SD*s of the mean. If our 10 scores were evenly distributed, with a *M* of 83.1 and *SD* of 8.39, this mean 68% of the scores would fall between 91.49 and 74.71; 95% of the scores would range between 99.88 and 66.32; 99.7% would fall between 108.27- and 57.93.
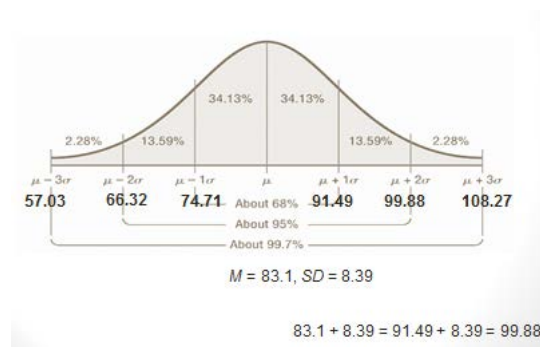
Normal Distribution

Throughout the overview of the descriptive statistics, the **normal distribution** (also called the **Gaussian distribution**, or bell- shaped curve) was discussed. It's important that we understand what this is. The normal distribution is the statistical distribution central to inferential statistics.

Due to its importance, every statistics course spends lots of time discussing it, and textbooks devote long chapters to it. One reason that it is so important is that many statistical procedures, specifically parametric procedures, assume that the distribution is normal. Additionally, normal distributions are important because they make data easy to work with. They also it make it easy to convert back and forth from raw scores to percentiles (a concept that is defined below).

The **normal distribution** is defined as a frequency distribution that follows a normal curve. Mathematically defined, a bell cure frequency distribution is symmetrical and unimodal. For a normal distribution, the following is true:

- o 34.13% of the scores will fall between M and +1 SD.
- o 13.59% of the scores will fall between 1 SD & 2 SD.
- o 2.28% of the scores will fall between 2 SD & 3 SD.
- o Mean = median = mode = Q_2 = $P_{50.}$

Another way to look at this is how we looked at it above. For normally distributed data values, approximately 68% of the distribution falls within $\pm$ 1 SD of the mean, 95% of the distribution falls within $\pm$ 2 SDs of the mean, and 99.7% of the distribution falls within $\pm$ 3 SDs of the mean.

To further understand the description of the normal bell curve, it may be helpful to define the shapes of a distribution:

- **Symmetry**. When it is graphed, a symmetric distribution can be divided at the center so that each half is a mirror image of the other.

- **Uimodal or bimodal peaks**. Distributions can have few or many peaks. Distributions with one clear peak are called **unimodal**, and distributions with two clear peaks are called **bimodal**. When a symmetric distribution has a single peak at the center, it is referred to as **bell-shaped**.

- **Skewness**. When they are displayed graphically, some distributions have many more observations on one side of the graph than the other. Distributions with most of their observations on the left (toward lower values) are said to be **skewed right**; and distributions with most of their observations on the right (toward higher values) are said to be **skewed left**.

- **Uniform**. When the observations in a set of data are equally spread across the range of the distribution, the distribution is called a **uniform distribution**. A uniform distribution has no clear peaks.

See graphic below.
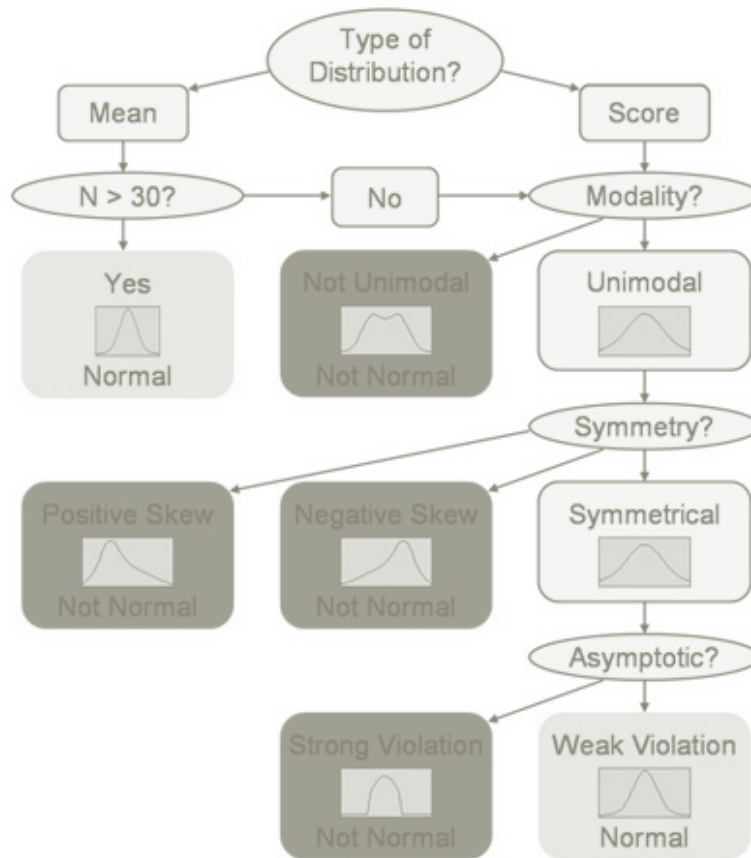
In our discussion of normal distribution, it is important to note that extreme values influence distributions. **Outliers** are extreme values. As we discuss, percentiles and quartiles below, it may be helpful to know that an extreme outlier is any data values that lay more than 3.0 times the interquartile range below the first quartile or above the third quartile. Mild outliers are any data values that lay between 1.5 times and 3.0 times the interquartile range below the first quartile or above the third quartile.

Measure of Position

In statistics, when we talk about the position of a value, we are talking about it relative to other values in a data set. The most common measures of position are percentiles, quartiles, and standard scores (*z*-scores and *t*-scores).

Percentiles are values that divide a set of observations into 100 equal parts or points in a distribution below which a given P of the cases lay. For example, 50% of scores are below P_{50}. For example, if $33 = P_{50}$, what can we say about the score of 33? That 50% of individuals in the distribution or data set scored below 33.

Quartiles divide a rank-ordered data set into four equal parts. The values that divide each part are called the first, second, and third quartiles; and they are denoted by Q_1, Q_2, and Q_3, respectively. In terms of percentiles, quartiles can be defined as follows: $Q_1 = P_{25}$, $Q_2 = P_{50} = Mdn$, $Q_3 = P_{75}$. For example, $33 = P_{50} = Q_2$.

Standard scores, in general, indicate how many standard deviations a case or score is from the mean. Standard scores are important for reasons such as comparability and ease of interpretation. A commonly used standard score is the *z*-score.

A ***z*-score** specifies the precise location of each X value within a distribution. The sign of the *z*-score (+ or -) signifies whether the score is above the mean or below the mean.

You interpret *z*-scores as follows:

- A *z*-score equal to 0 represents a factor equal to the mean.
- A *z*-score less than 0 represents a factor less than the mean.
- A *z*-score greater than 0 represents a factor greater than the mean.
- A *z*-score equal to 1 represents a factor 1 standard deviation greater than the mean; a *z*-score equal to -1 represents a factor 1 standard deviation less than the mean. A *z*-score equal to 2, 2 standard deviations greater than the mean.

For example, a *z*-score of +1.5 means that the score of interest is 1.5 standard deviations above the mean. A *z*-score can be calculated using the following formula:

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

$X =$ value, $\mu =$ mean of the population,, $\sigma =$ standard deviation.

For example, 5th graders take a national achievement test annually. The test has a mean score of 100 and a standard deviation of 15. If Bob's score is 118, then his *z*-score is $(118 - 100) / 15 = 1.20$. That means that Bob scores 1.20 *SD* above the average in the population. A *z*-score can also be converted into a raw score. For example, Bob's score on the test is $X = (z * \sigma) + 100 = (1.20 * 15) + 100 = 18 + 100 = 118$.

SPSS will compute *z*-scores for any continuous variable and save the *z*-scores to your dataset as a new variable.

- SPSS: Analyze > Descriptive Statistics > Descriptives
    - Select variable(s) and click *Save Standardized values as variables* check box.

Inferential Statistics

With inferential statistics, you are trying to infer something about a population by measuring a sample taken from the population.

Hypothesis Testing

A **hypothesis test** (also known as the statistical significance test) is an inferential statistical procedure in which the researcher seeks to determine how likely it is that the results of a study are due to chance or whether or not to attribute results observed to sampling error. **Sampling error** is chance or the possibility that chance affected the co-variation among variables in sample statistics or the relationship between the variables being studied. In hypothesis testing, the researcher decides whether to reject the null hypothesis or fail to reject the null hypothesis.

Research and Null Hypothesis

There are two types of hypotheses:

First, there is the **research hypothesis.** The research hypothesis (also referred to as the alternative hypothesis) is a tentative prediction about how change in a variable(s) will explain or cause changes in another variable(s). **Example:** H_1: There will be a statistically significant difference in high school dropout rates of students who use drugs and students who do not use drugs.

Second, the **null hypothesis** is the antithesis to the research hypothesis and postulates that there is no difference or relationship between the variables under study. That is, results are due to chance or sampling error. **Example:** H_0: There will be no statistically significant difference in high school dropout rates of students who use drugs and students who do not use drugs.

How to Develop a Research Question

The formulation and selection of a research question can be overwhelming. Even experienced researchers usually find it necessary to revise their research questions numerous times until they arrive at one that is acceptable. Sometimes the first question formulated is unfeasible, not practical, or not timely.

A well formulated research question, according to Bartos (1992):
- ask about the relationship between two or more variables
- be stated clearly and in the form of a question
- be testable (i.e. possible to collect data to answer the question)
- not pose an ethical or moral problem for implementation
- be specific and restricted in scope (your aim is not to solve the world's problems)
- identify exactly what is to be solved.

Example of a Poorly Formulated Research Question: What is the effectiveness of parent education for parents with problem children? (Note this is not specific or clear. Thus, it is not testable. What is meant by effectiveness, parent education, and parents with problem children? Since it is not clear it would be hard to test.)

Example of a Well Formulated Research Questions: What is the effect of the STEP parenting program on the ability of a parents to use natural, logical consequences (as opposed to punishment) with their child who has been diagnosed with Bipolar disorder? Or, Is the difference in the number of times that parents use natural, logical consequences (as opposed to punishment) with their child who has been diagnosed with Bipolar disorder who participate in the STEP parenting program as opposed to parents who participate in a parent support group? (These questions are clearer and more focused. The researcher would now be better able to design an experimental study to compare the number of natural, logical consequences used by parent who have participated in the STEP program and those who have not.)

12

After your research question is formulated, you will write your research hypotheses.

A good hypothesis has the following characteristics:

- the hypothesis stated the expected relationship between variables
- the hypothesis is testable
- the hypothesis is stated simply and concisely as possible
- the hypothesis is founded in the problem statement and supported by research (Bartos, 1992).

Example Research Hypotheses: There will be a statistically significant difference in the number of times that parents use natural, logical consequences (as opposed to punishment) with their child who has been diagnosed with Bipolar disorder who participate in the STEP parenting program as opposed to parents who participate in a parent support group as measured by the Parent Behavior Rating Sale. (Note: This is non-directional; **Directional hypotheses** specify the direction of the expected and indicate a one-tailed test will be conducted; whereas, **nondirectional hypotheses** indicate a difference is expected with no specification of the direction. In the latter, a two-tailed test will be conducted.

Example Null Hypothesis: There will be no statistically significant difference in the number of times that parents use natural, logical consequences (as opposed to punishment) with their child who has been diagnosed with Bipolar disorder who participate in the STEP parenting program as opposed to parents who participate in a parent support group as measured by the Parent Behavior Rating Sale.

A **hypothesis test** is then an inferential statistical procedure in which the researcher seeks to determine how likely it is that the results of a study are due to chance or sampling error.

 There are four steps to hypothesis testing:

1. State the null and research hypothesis as just discussed.
2. Choose a statistical significance level or alpha.
   a. An alpha is always set ahead of time. It is usually set to either .05 or .01 in social science research. The choice of .05 as the criterion for an acceptable risk of Type I error is a tradition or convention, based on suggestions made by Sir Ronald Fisher. A significance level of .05 means that if we reject the null hypothesis, we are willing to accept that there is no more than 5% chance that we made the

wrong conclusion. In other words, with a .05 significance level, we want to be at least 95% confident that if we reject the null hypothesis we have made the correct decision.

3. Choose and carry out the appropriate statistical test.
4. Make a decision regarding hypothesis (i.e. reject or fail to reject the null hypothesis).
   a. If the *p*-value for the analysis is equal to or lower than the significance level established prior to conducting the test, the null hypothesis is rejected. If the *p*-value for the analysis is more than the significance level established prior to conducting the test, the null hypothesis is not rejected. Let's say we set our alpha level at .05. If our *p*-value is less than .05 (*p* = .02), we can consider the results to be statistically significant, and we reject the null hypothesis.

Type I and II Error

The purpose of an inferential statistical procedure is to test a hypothesis (es). It is important to note that there is always the possibility that you could reach the wrong conclusions in your analysis. After you have determined the results of your statistical tests, you will decide to reject or fail to reject the null hypothesis. If your results indicate no statistically significant difference based on the chosen significance level (e.g. your results indicate that *p* = .07, when you set the significance level at .05), you will fail to reject the null hypothesis. If your results indicate statistically significant difference (e.g. your results indicate that *p* = .02, when you set the significance level at .05), you reject the null hypothesis. Sometimes researchers make error in rejecting or failing to reject the null hypothesis. Reaching the wrong conclusion refers to as an error, a Type I error or Type II error.

Type I Error

If the researcher says that there is statistically significant difference and rejects the null hypothesis when the null hypothesis is true (i.e. there really is there really is statistical significance), the researcher makes a Type I error.

Example: A researcher concludes that the STEP parenting program is more effective than the parent support group. Based on these results, a counseling center revises their parenting program and spends a significant amount of money for the STEP program. After the expenditure of time and money, the parents do not appear to show improvement in their implementation of logical, natural consequences. Subsequent research does not reveal results demonstrated in the original study. Although the ultimate truth about the null hypothesis is not known, evidence supports that a Type I error likely occurred in the original study.

Type II Error
If the researcher says that there is no statistically significant difference and fails to reject the null hypothesis when the null hypothesis is false (i.e. there really is statistical significance), the researcher makes a Type II error.

Example: A researcher concludes that there is no statistically significant difference in the number of times that parents use natural, logical consequences (as opposed to punishment) with their child who participates in the STEP parenting program as opposed to parents who participate in the parent support group. That is, the programs are approximately equivalent in their effect. Future research however demonstrates that the STEP program is more effective; thus, it is likely that the researcher made a Type II error. Type II errors may often be a result of lack of statistical power; that is, you may not obtain statically significant results because the power is too low (a concept reviewed below).

Note: A Type I error is often called alpha. The Type II error is often called beta. The power of the test = 100% - beta. Ideally the statistical procedure should correctly identify if a difference or relationship exists between variables, an appropriate level of power is needed (ideally .80 or above).

Statistical Power

Power may be defined as a number or percentage that indicates the probability a study will obtain statistically significant results or how much confidence you can have in rejecting the null hypothesis. For example, a power of 40% or 0.4 indicates that if the study was conducted 10 times it is likely to produce results (i.e. statistically significant) 4 times. Another way to say this is that the researcher can say with 40% certainty that his or her conclusion about the null hypothesis was correct.

Power is influenced by 3 factors:

- *Sample size*
- *Effect size*
- *Significance level*

Assuming that all terms in an analysis remain the same, as effect size increases, statistical power increases. As an alpha level is made smaller, for example, if we change the significance level from .05 to .01, statistical power decreases.

There are multiple methods for calculating power prior to a study and after a study: (1) Cohen's (1988) charts; (b) open source or paid software (G*Power); SPSS. Statistical power is important

for both planning a study to determine the needed sample size and interpreting results of a study. If the power is below .80, then one needs to be very cautious in interpreting results. You, as the researcher, should inform the reader in the results or discussion section that a Type II error was possible due to power when power is low.

Note: Power for nonparametric tests is less straightforward. One way to calculate it is to use the Monte Carlo simulation methods (Mumby, 2002).

Effect Size

Although you may identify the difference between the groups you are studying as statistically significant (which you, like most researchers, will find exciting), there is more than just obtaining statistical significance.

The **effect size** tells us the strength of the relationship, giving us some practical and theoretical ideas about the significance of our results. In terms of statistics, effect size can be defined as a statistic that is used to depict the relationship magnitude between the means.

The three most common effect size statistics are:

- **Partial eta squared** indicates the proportion of variance in the DV that is explained by the IV. The values range from 0 to 1 and are interpreted as the chart below indicates.
- **Cohen's *d*** refers to the difference in groups based on standard deviations.
- **Pearson's *r*** refers to the strength and direction of the relationship between variables.

Cohen (1988) sets forth guidelines for interpreting effect size.

Table 3.
Guidelines for Interpreting Effect Size for Group Comparisons.

Effect size	Cohen's *d*	Partial Eta squared	Pearson's *r*
Small	.2	.01	.10- .29
Medium	.5	.06	.30- .49
Large	.8	.138	Over .50

Non Parametric and Parametric

There are two types of data analysis, parametric and nonparametric. Parametric tests are said to be more powerful and precise, for nonparametric tests do not require the number of assumptions that parametric tests do. Although the type of analysis chosen depends on a number of factors, nonparametric tests usually used under two conditions: 1) the dependent variable or variable of interest is measured at the nominal or ordinal level and 2) the data does not satisfy assumptions

required by parametric tests. Here is a list of the parametric and nonparametric statistical procedures:

Table 4

Parametric vs. Nonparametric Procedures

	Parametric	Nonparametric
Assumed Distribution	Normal	Any
Assumed Variance	Homogeneous	Any
Level of Measurement	Ratio and Interval	Any; Ordinal and Nominal
Central Tendency Measure	Mean	Median, Mode
	Statistical Procedures	
	Independent samples *t* test	Mann – Whitney test
	Paired Sample *t* test	Wilcoxon
	One way, between group ANOVA	Kruskal- Wallis
	One way, repeated measures ANOVA	Friedman Test
	Factorial ANOVA	None
	MANOVA	None
	Pearson	Spearman, Kendall Tau, Chi Square
	Bivariate Regression	None

Table derived from Pallant.J. (2007). SPSS survival manual.

Assumption Testing

There are some assumptions that apply to all parametric statistical procedures. They are listed here. There are additional assumptions for some specific procedures. Texts such as Warner (2013) discuss these in detail.

Level of Measurement: The dependent variable measured should be measured on the interval or ratio level.

Random sampling: It is assumed that the sample is a random sample from the population.

Independent Observations: The observations within each variable must be independent; that is, that the measurements did not influence one another.

Normality: This assumption assumes that the population distributions are normal. Univariate normality is examined by creating histograms or by conducting a normality test, such as the Shapiro-Wilk (if sample size is smaller than 50) and Kolmogorov-
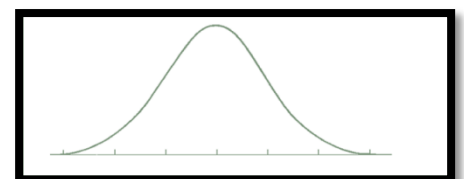
Smirnov (if sample size is larger than 50) tests. On the histogram, normality is assumed when there is a symmetrical, bell shaped curve present. For the normality tests, non-significant results (a significance level more than .05) indicate tenability of the assumption. That is, normality can be assumed.

Equal Variances (homogeneity of variance): This assumption assumes that the population distributions have the same variances. If this assumption is violated, the averaging of the 2 variances is futile. This assumption is evaluated using Levene's Test for Equality of Variance for both the ANOVA and t-test. A significance level larger than .05 indicates that equal variance can be assumed. A significance level less than .05 means that variance cannot be assumed; that is, the assumption is not tenable. Bartlett's test is also an alternative test to the Levene's test. Scatterplots and Box's M are used to test this assumption with correlational procedures and multivariate procedures such as a MANOVA.

Note: Some of these are discussed in more detail below as they apply to specific statistical procedures. Although non-parametric tests have less rigorous assumptions, they still require a random sample and independent observations.

Parametric Statistical Procedures
t-tests
Independent *t* test

Description: Independent *t*-tests (also known as independent sample *t*-tests) are used when comparing the mean scores of two different groups. The procedure tells you if two groups statistically significantly differ in terms of their mean scores.

Variables:

- Independent variable, nominal
- Dependent variable, ratio or interval

Assumptions: (All above listed for parametric procedures). Using data set, examine:

1) ***Normality:*** This assumption assumes that the population distributions are normal. The *t*-test is robust over moderate violations of this assumption, especially if a two-tailed test is used and the sample size is not small (30+). Check for normality by creating a histogram or by conducting a normality test, such as the Shapiro-Wilk and Kolmogorov-Smirnov tests. On the histogram, normality is assumed when there is a symmetrical, bell shaped curve. For the normality tests, non-significant results (a significance level more than .05) indicate tenability of the assumption. That is, normality can be assumed.

2) **Equal Variances**: This assumption assumes that the population distributions have the same variances. If this assumption is violated, the averaging of the 2 variances is futile. If it is violated, use modified statistical procedure (in SPSS, this the alternative t- value on the second line of the t-test table, which says *equal variance not assumed*). Evaluate variance using Levene's Test for equality of variance. A significance level larger than .05 indicates that equal variance can be assumed. A significance level less than .05 means that variance cannot be assumed; that is, the assumption is not tenable.
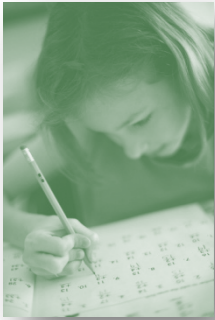
Example Research Questions and Hypotheses:
Example #1

RQ: Is there a significant difference between male and female math achievement? (Non directional/ two-tailed)

H$_0$: There is no statistically significant difference between male and female math achievement.

Example #2

RQ: Do at-risk high school seniors who participate in a study skills program have a higher graduation rate than at-risk high school seniors who do not participate in a study skills program? (Directional/one-tailed)

H$_0$: At-risk high school seniors who participate in a study skills program do not have a statistically significant higher graduation rate than at-risk high school seniors who do not participate in a study skills program.

Non-parametric alternative: Mann-Whitney U Test

Reporting Example: t (93) = -.67, p = .41, d = -.10. Males (M = 31.54, SD = 5.16, n = 29) on average do not statistically significantly differ from females (M = 32.46, SD = 4.96, n = 65) in terms of math achievement. The observed power was .30, indicating the likelihood of a Type II error (Note that numbers and studies are not real).

Items to Report:

- Assumption testing
- Descriptive statistics (*M, SD*)
- Number (*N*)
- Number per cell (*n*)
- Degrees of freedom (df)
- t value (*t*)
- Significance level (*p*)
- Effect size and power

Paired Sample *t* test

Description Paired sample *t*-tests (also known as the repeated measures *t*-tests or dependent *t*-tests) are used when comparing the mean scores of one group at two different times. Pretest/posttest designs are an example of the type of situation in which you may choose to use this procedure. Another time that this procedure may be used is when you examine the same person in terms of his or her response to two questions (i.e. level of stress and health). This procedure is also used with matched pairs (e.g. twins, husbands and wives).

Variables:

- Independent variable, nominal
- Dependent variable, interval or ratio

Assumptions: (All above listed for parametric procedures). Using data set, examine:

Normality: This assumption assumes that the population distributions are normal. The *t*-test is robust over moderate violations of this assumption. It is especially robust if a two-tailed test is used and if the sample sizes are not small (30+). Check for normality by creating a histogram or by conducting a normality test, such as the Shapiro-Wilk and Kolmogorov-Smirnov tests. On the histogram, normality is assumed when there is a symmetrical, bell shaped curve. For the normality tests, non-significant results (a significance level more than .05) indicate tenability of the assumption. That is, normality can be assumed. It is important to examine this assumption in each group or grouping variable.

Equality of variance does not apply here, as two populations are not being examined.

Example Research Questions and Hypotheses:

Example #1

RQ: Is there a significant change in participants' tolerance scores as measured by the Tolerance in Diversity Scale after participating in diversity training? (Non directional/ two-tailed)

H$_0$: There is no statistically significant difference in participants' tolerance scores as measured by the Tolerance in Diversity Scale after participating in diversity training.

Example #2

RQ: Do students perform better on the analytical portion of the SAT than on the verbal portion? (directional/ one-tailed)

H$_0$: Students do not score statistically significantly better on the analytical portion of the SAT than on the verbal portion.

Non-parametric alternative: Wilcoxon Signed Rank Test

Reporting Example: t (28) = -.67, p = .41, d = -.10. Participants on average do not statistically significantly score better on the analytical portion of the SAT (M = 31.54, SD = 5.16) than on the verbal portion (M = 32.46, SD = 4.96, N = 29). The observed power was .45, indicating a Type II error may be possible. (Note that numbers and studies are not real)

Items to Report:

- Assumption testing
- Descriptive statistics (*M, SD*)
- Number (*N*)
- Degrees of freedom (df)
- t value (*t*)
- Significance level (*p*)
- Effect size and power

Correlation Procedures

Bivariate Correlation

Description: A bivariate correlation assists in examining the strength and direction of the linear relationship between two variables. The Pearson product moment coefficient is used with interval or ratio data. The Pearson product moment coefficient ranges from +1 to -1. A plus indicates a positive relationship (as one variable increases, so does the other), whereas a negative sign indicates a negative relationship (as one variable increases, the other decreases). The value indicates the strength of the relationship. 0 indicates no relationship; .10 to. 29 = a small relationship; .30 to .49 = a medium relationship; .50 to 1.0 = large relationship.

Variables:

- two variables, ratio or interval or
- one variable (i.e. test score), ratio or interval, and one variable, ordinal or nominal

Spearman rank order correlation is used with ordinal data or when data does not meet assumptions for the Pearson product moment coefficient.
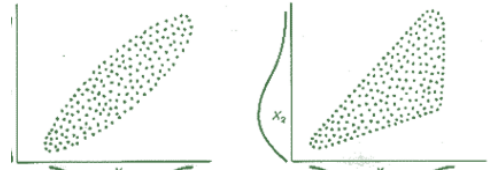
Assumptions:

1) *Normality:* This assumption assumes that the population distributions are normal. Check for normality by creating histograms or by conducting normality tests, such as the Shapiro-Wilk and

Kolmogorov-Smirnov tests. On the histogram, normality is assumed when there is a symmetrical, bell shaped curve. For the normality tests, non-significant results (a significance level more than .05) indicate tenability of the assumption. That is, normality can be assumed.

2) *Independent Observations:* The observations within each variable must be independent.

3) *Linearity*: This assumption assumes the relationship between the two variables is linear. Check for linearity using a scatterplot; a roughly straight line (no curve) indicates that the assumption is tenable.

4) *Homoscedasticity*: This assumption assumes the variability in scores in both variables should be similar. Check for homoscedasticity using a scatterplot; a cigar shape indicates that the assumption is tenable.

Example Research Questions and Hypotheses:

RQ: Is there a significant relationship between second grade students' math achievement and level of math anxiety? (Non directional/ two-tailed)

H$_0$: There is no statistically significant relationship between second grade students' math achievement and level of math anxiety.

Non-parametric alternative: Spearman rank order correlation; chi- square

Reporting Example: The two variables were strongly, negatively related, r (90) = -.67, p = .02. As students' math anxiety increased ($M = 41.54$, $SD = 5.16$) their math achievement decreased ($M = 32.46$, $SD = 4.96$, $N = 91$). The observed power was .80. (Note that numbers and studies are not real).

Items to Report:

- Assumption testing
- Descriptive statistics (*M, SD*)
- Number (*N*)
- Degrees of freedom (df)
- Observed *r* value (*r*)
- Significance level (*p*)
- Power

Partial Correlation

Description: A partial correlation assists you in examining the strength and direction of the linear relationship between two variables, while controlling for another variable (i.e. confounding variable; a variable you suspect influences the other two variables).

Variables:

- two variables, ratio or interval or
- one variable (i.e. test score), ratio or interval, and one variable, ordinal or nominal
- AND one variable you wish to control

Assumptions:

Same as bivariate correlation.

Example Research Questions and Hypotheses:

RQ: After controlling for practice time, is there a relationship between the number of games a high school basketball team wins and the average number of points scored per game?

H$_0$: While controlling for practice time, there is no significant relationship between the number of games a high school basketball team wins and the average number of points scored per game.

Reporting Example: While controlling for practice variable, the two variables were strongly, positively correlated, r (90) = + .67, p = .02, N = 92. As the average number of points scored per game increased ($M = 15$, $SD = 6$), the number of games a high school basketball team won increased ($M = 30$, $SD = 4$). (Note that numbers and studies are not real). An inspection of the zero order correlation, r (91) = + .69, p = .02, suggested that controlling for practice variable ($M = 750$, $SD = 45$) had little effect on the strength of the relationship between the two variables. The observed power was .90 for both analyses.

Items to Report:

- Assumption testing
- Descriptive statistics (M, SD)
- Number (N)
- Observed r value for the zero order and partial analysis (r)
- Degrees of freedom (df)
- Significance level for the zero order and partial analysis (p)
- Power

Bivariate Linear Regression

Description: A bivariate correlation assists you in examining the ability of the independent or predictor variable to predict the dependent or criterion variable.

Variables:
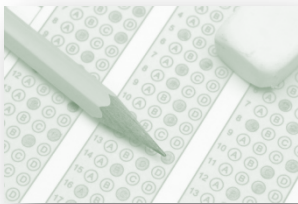
- two variables, ratio or interval or

Assumptions:

1) *Normality:* This assumption assumes that the population distributions are normal. Check for normality by creating histograms or by conducting normality tests, such as the Shapiro-Wilk and Kolmogorov-Smirnov tests. On the histogram, normality is assumed when there is a symmetrical, bell shaped curve. For the normality tests, non-significant results (a significance level more than .05) indicate tenability of the assumption. That is, normality can be assumed.

2) *Independent Observations:* The observations within each variable must be independent.

3) *Linearity*: This assumption assumes the relationship between the two variables is linear. Check for linearity using a scatterplot; a roughly straight line (no curve) indicates that the assumption is tenable.

4) *Homoscedasticity*: This assumption assumes the variability in scores in both variables should be similar. Check for homoscedasticity using a scatterplot; a cigar shape indicates that the assumption is tenable.

Example Research Questions and Hypotheses:

RQ: How well does the amount of time college students study predict their test scores? Or, How much variance in test scores can be explained by the amount of time a college student studies for the test?

H$_0$: The amount of time college students study does not significantly predict their test scores.

Reporting Example: The regression equation for predicting the test score is, Y = 1.22X $_{GPA}$ + 47.54. The 95% confidence interval for the slope was .44 to 2.00. There was significant evidence to reject the null hypothesis and conclude that the amount of time studying (*M* = 6.99, *SD* = .91) significantly predicted the test score (*M* = 87.13, *SD* = 1.83), *F*(1, 18) = 10.62, *p* <.01. Table 1 provides a summary of the regression analysis for the variable predicting test scores. Accuracy in predicting test scores is moderate. Approximately 38% of the variance in the final exam was accounted for by its linear relationship with GPA.

Table1
Summary of Regression Analysis for Variable Predicting Final Exam (*N* = 20)

Variable	B	SE B	β
GPA	1.22	.37	.61*

Note. *r* = .61*; *r*^{2} = .38*; *p* < .05

Items to Report:

- Assumption testing
- Descriptive statistics (*M, SD*)
- Number (*N*)
- Degrees of freedom (df)
- *r* and r^2
- *F* value (*F*)
- Significance level (*p*)
- B, beta, and SE B
- Regression equation
- Power

Here it is important to note that authors such as Warner (2013) and the APA manual suggest that reporting a single bivariate correlation or regression analysis is usually not sufficient for a dissertation, thesis, or publishable paper. However, when significance tests are reported for a large number of regression or correlational procedures, there is an inflated risk of Type I error, that is, finding significance when there is not.

Warner (2013) suggests several ways to reduce risk of inflated Type I error:

- Report significance tests for a limited number of analyses (i.e. don't correlate every variable with every other variable; let theory guide the analyses)
- Use Bonferroni corrected a levels to test each individual Regression
- Use cross validation within the sample
- Replicate analyses across new samples

Thompson (1991) in *A Primer on the Logic and Use of Canonical Regression Analysis*, suggested: *when seeking to understand the relationship between sets of multiple variables, canonical analysis limits the probability of committing Type I errors, finding a statistically significant result when it does not exist, because instead of using separate statistical significant tests, a canonical can assess these relationships between the two set of variables (independent and dependent) in a single relationship rather than using separate relationships for each dependent variable*. Thompson discusses this further, so this article is well worth reading if you are planning to conduct a correlation analysis. In summary, Thompson is recommending that a multivariate analysis may be more appropriate when the aim is to analyze the relationship between multiple variables.

Two commonly used multivariate analyses include (Note: there are many analyses not discussed in this guide):

- **Multiple Regression** (Standard, Hierarchal, Stepwise) is used to explore how multiple predictor variables contribute to the prediction model for a criterion variable and the relative contribution of each of the variables that make up the model. You need criterion variables measured at the interval or ratio level and two or more predictor variables, which can be measured at any of the four levels of measurement.

- **Canonical Correlations** are used to analyze the relationship between two sets of variables. You need two sets of variables.

Analysis or Variance Procedures
One-way analysis of variance (ANOVA)

Description: The one way between group ANOVA involves the examination of an independent variable with three or more levels or groups and one dependent variable; thus, it tells you when there is a difference in the mean scores of the dependent variable across three or more groups. If statistical significance is found, post hoc tests need to be run to determine between which groups differences lie.

Variables:

- Independent variable with three of more categories, nominal (i.e. young, middle age, old; Poor, Middle Class, Upper Class)
- Dependent variable, ratio or interval

Assumptions: (All above listed for parametric procedures). With data set, examine:

1) *Normality:* This assumption assumes that the population distributions are normal. The *ANOVA* is robust with moderate violations of this assumption when the sample size is large. Check for normality by creating histograms or by conducting normality tests, such as the Shapiro-Wilk and Kolmogorov-Smirnov tests. On the histogram, normality is assumed when there is a symmetrical, bell shaped curve. For the normality tests, a non-significant result (a significance level more than .05) indicates tenability of the assumption. That is, normality can be assumed. It is important to examine this assumption in each group or grouping variable.

2) *Equal Variances*: This assumption assumes that the population distributions have the same variances. If this assumption is violated, the averaging of the 2 variances is futile. If it is violated, use modified statistical procedure (In SPSS, this alternative can be found under the Robust Test of Equability of Means output; Welsh or Brown-Forsythe). Evaluate variance using Levene's Test for Equality of Variance. A significance level larger than .05 indicates that equal variance can be assumed. A significance level less than .05 means that variance cannot be assumed; that is, the assumption is not tenable.

Example Research Questions and Hypotheses:

RQ: Is there a significant difference in sense of community for university students using three different delivery mediums for their courses (residential, blended, online)?

H$_0$: There is no statistically significant difference in sense of community for university students based on the delivery medium of their courses.
Non-parametric alternative: Kruskal- Wallis

Reporting Example: An analysis of variance demonstrated that the effect of delivery system was significant, $F(3,27) = 5.94$, $p = .007$, $N = 30$. Post hoc analyses using the Scheffé post hoc criterion for significance indicated that the average number of errors was significantly lower in the online condition ($M = 12.4$, $SD = 2.26$, $n = 10$) than in the blended condition ($M = 13.62$, $SD = 5.56$, $n = 10$) and the residential condition ($M = 14.65$, $SD = 7.56$, $n = 10$). The observed power was .76.No other comparisons reached significance. (Note that numbers and studies are not real).

Items to Report:

- Assumption testing
- Descriptive statistics (*M, SD*)
- Number (*N*)
- Number per cell (*n*)
- Degrees of freedom (df within/ df between)
- Observed *F* value (*F*)
- Significance level (*p*)
- Post hoc or planned comparisons
- Effect size and power

Factorial (or two-way) analysis of variance (ANOVA)

Description: The two way between group ANOVA involves the simultaneous examination of two independent variables and one dependent variable; thus, it allows you to test for an interaction effect as well as the main effect of each independent variable. If a significant interaction effect is found, additional analysis will be needed.

Variables:

- Two independent variables, nominal (i.e. age, gender, type of treatment)
- Dependent variable, ratio or interval

Assumptions: (All above listed for parametric procedures)

1) *Normality:* This assumption assumes that the population distributions are normal. The *ANOVA* is robust over moderate violations of this assumption, Check for normality by creating a histogram or by conducting a normality test, such as the Shapiro-Wilk and Kolmogorov-Smirnov tests. On the histogram, normality is assumed when there is a symmetrical, bell shaped curve. For the normality tests, non-significant results (a significance level more than .05) indicate tenability of the assumption. That is, normality can be assumed. It is important to examine this assumption in each group or grouping variable for each independent variable.

2) *Equal Variances*: This assumption assumes that the population distributions have the same variances. If this assumption is violated, the averaging of the 2 variances is futile. If it is violated, use modified statistical procedure (In SPSS, this alternative can be found under the Robust Test of Equability of Means output; Welsh or Brown-Forsythe). Evaluate variance using Levene's

Test for Equality of Variance. A significance level larger than .05 indicates that equal variance can be assumed. A significance level less than .05 means that variance cannot be assumed; that is, the assumption is not tenable.

Sample Research Questions and Hypotheses:

RQ: What is the influence of type of delivery medium for university courses (residential, blended, online) and age (young, middle age, older) on sense of community?

H_{01}: There is no statistically significant difference in sense of community for university students based on the delivery medium of their courses and their age.

H_{02}: There is no statistically significant difference in sense of community for university students based on the delivery medium of their courses.

H_{03}: There is no statistically significant difference in sense of community for university students based on their age.

Non-parametric alternative: None

Reporting Example: A 3 x 3 analysis of variance demonstrated that the interaction effect of delivery system and age was not significant, $F(2,27) = 1.94$, $p = .70$, observed power = .56. There was a statistically significant main effect for delivery system, $F(2, 27) = 5.94$, $p = .007$, $\eta^2 = .02$, observed power = .87; however, the effect size was small. Post hoc comparisons to evaluate pairwise differences among group means were conducted with the use of Tukey HSD test since equal variances were tenable. Tests revealed significant pairwise differences between online ($M = 13.62$, $SD = 5.56$, $n = 10$) and residential ($M = 14.62$, $SD = 6.56$, $n = 10$), and blended ($M = 15.62$, $SD = 4.56$ $n = 10$), $p < .05$. The main effect for age did not reach statistical significance, $F(2, 27) = 3.94$, $p = .43$, $\eta^2 = .001$, observed power = .64; indicating that age did not influence students' sense of community (Note that numbers and studies are not real).

Items to Report:

- Assumption testing
- Descriptive statistics (*M, SD*)
- Number (*N*)
- Number per cell (*n*)
- Degrees of freedom (df within/ df between)
- Observed *F* value (*F*)
- Significance level (*p*)
- Effect size and power

One-way repeated measures analysis of variance (ANOVA)

Description: The one way repeated measures ANOVA is used when you want to measure participants on the same dependent variable three or more times. You may also use it to measure participants exposed to three different conditions or to measure subjects' responses to two or more different questions on a scale. This ANOVA tells you if differences exist within a group among the sets of scores. Follow up analyses are conducted to identify where differences exist.

Variables: (One group)

- Independent variable, nominal (i.e. time 1, time 2, time 3; pre intervention, prior to intervention; 3 month following intervention)
- Dependent variable, ration or interval
- One group measured on the same scale (DV) three or more times (IV) or each person in a group measured on three different questions using the same scale.

Assumptions: (All above listed for parametric procedures). Use data to examine:

1) *Normality:* This assumption assumes that the population distributions are normal. The *ANOVA* is robust over moderate violations of this assumption, Check for normality by creating histograms or by conducting normality tests, such as the Shapiro-Wilk and Kolmogorov-Smirnov tests. On the histogram, normality is assumed when there is a symmetrical, bell shaped curve. For the normality tests, non-significant results (a significance level more than .05) indicate tenability of the assumption. That is, normality can be assumed. This assumption needs to be examined for each level of the within group factors (e.g. Time 1, Time 2, Time 3). Multivariate normality also needs to be examined.

2) *Sphericity or Homogeneity-of-variance-of difference*: This assumption is only meaningful when there are more than two levels of the within group factor.

Example Research Questions and Hypotheses:

Example #1

RQ: Is there a significant difference in university students' sense of community based on the type of technology (1, 2, 3) they use for their course assignments?

H$_0$: There is no statistically significant difference in university students' sense of community based on the type of technology they use for their education course assignments.

Example #2

RQ: Is there a significant difference in mothers' perceived level of communication intimacy with her father, her son, and her husband?

H₀: There is no statistically significant difference in a mothers' perceived level of communication intimacy with her father, her son, and her husband.

Non-parametric alternative: Friedman Test

Reporting Example: A one way repeated measures ANOVA was conducted to compare university students' (*N*= 30) sense of community scores at the completion of assignment 1 (*M* = 13.62, *SD* = 5.56), assignment 2 (*M* = 12.46, *SD* = 2.26), and assignment 3 (*M* = 14.4, *SD* = 3.26). A significant effect was found, Wilks' Lambda = .25, *F* (3,27) = 5.94, *p* = .007, partial eta squared =.70, observed power = .82. Follow-up comparisons indicated that students had significantly higher sense of community while participating in Assignment 1 as compared to Assignments 2 and 3, *p* = .002. No other comparisons reached significance (Note that numbers and studies are not real).

Items to Report:

- Assumption testing
- Descriptive statistics (*M, SD*)
- Number (*N*)
- Degrees of freedom (df within/ df between)
- Observed *F* value (*F*)
- Significance level (*p*)
- Follow up comparisons
- Effect size and power

Multivariate analysis of variance (MANOVA)

Description: The MANOVA compares two or more groups in terms of their means on a group of related dependent variables. It tests the null hypothesis that the population means on a set of dependent variables do not vary across different levels of a factor or grouping variable.

Variables:

▸ One independent variable, nominal (e.g. treatment and control group); and

▸ Two or more underlined (i.e. shown in the literature to be significantly related) dependent variables, ratio or interval

MANOVAs can also be extended to two-way and higher-order designs involving two or more categorical, independent variables.

Assumptions: (All above listed for parametric procedures). Use data to examine:

1) ***Extreme outliers***. Use box plots.

2) *Univariate normality*. This assumption assumes that the population distributions are normal. The *ANOVA* is robust over moderate violations of this assumption, Check for normality by creating histograms or by conducting normality tests, such as the Shapiro-Wilk and Kolmogorov-Smirnov tests. On the histogram, normality is assumed when there is a symmetrical, bell shaped curve. For the normality tests, non-significant results (a significance level more than .05) indicate tenability of the assumption. That is, normality can be assumed. This is checked for each grouping variable. The MANOVA is reasonably robust to modest violations of normality when the sample size is at least 20 in each cell (Tabacknick & Fidell, 2007, p.251). The exception to this is when normality is affected by outliers.

3) *Multivariate normality*. This is examined using Mahalanobis distance. The data's Mahalanobis distance value is compared against the critical value outlined in a chi-square critical value chart found in statistical texts. If it exceeds the value in the chart, this assumption is not tenable.

4) *Multicolinearity and singularity.* These assumptions are checked by creating correlation matrixes. A MANOVA is most robust when the dependent variables are moderately correlated. When correlations are low, separate univariate analyses need to be run. Multicolinearty is an issue when correlation coefficient values are above .8 or .9. When multicolinearty exists, it is usually preferable to collapse the variables into a single measure.

5) *Linearity.* This assumption assumes that the relationship among variables is linear. This is examined using scatter plots. The presence of a straight line indicates linearity. A curvilinear line would indicate that the assumption is not tenable.

6) *Equal Variances*: This assumption assumes that the population distributions have the same variances. If this assumption is violated, the averaging of the 2 variances is futile. If it is violated, use modified statistical procedure (In SPSS, this alternative can be found under the Robust Test of Equability of Means output; Welsh or Brown-Forsythe). Evaluate variance using Levene's Test for Equality of Variance. A significance level larger than .05 indicates that equal variance can be assumed. A significance level less than .05 means that variance cannot be assumed; that is, the assumption is not tenable.

7) *Homogeneity of variance-covariance*. Box's M is checked to examine the tenability of this assumptions. In SPSS, this is part of the MANOVA output. A significance level larger than .05 indicates that equal variance can be assumed. A significance level less than .05 means that variance-covariance cannot be assumed; that is, the assumption is not tenable. It's reported as follows: Box's test $M = X$, $F (X,X) = X$, $p = .X$.

Example Research Questions and Hypotheses:

Example #1

RQ: Is there a statistically significant difference in university students' linear combinations of spirit, trust, interaction, and learning based on the type of program they are enrolled in for their course of study?

H_{01}: Residential and distance students do not statistically significantly differ in terms of the linear combinations of spirit, trust, interaction, and learning.
H_{02}: Residential and distance students do not statistically significantly differ in their feelings of sprit.
H_{03}: Residential and distance students do not statistically significantly differ in their feelings of trust.
H_{04}: Residential and distance students do not statistically significantly differ in their interaction.
H_{05}: Residential and distance students do not statistically significantly differ in their learning.
Non-parametric alternative: None

Reporting Example: A one-way multivariate analysis of variance (MANOVA) was conducted to investigate the effect of course differences (i.e. residential, distance) in the combination of feelings of spirit, feelings of trust, feelings of interaction, and learning. The pooled means and standard deviations for feeling of spirit, trust, and interaction are $M= 26.57$ $(SD = 6.16)$, $M= 25.78$ $(SD = 6.64)$, $M= 27.06$ $(SD = 5.16)$, respectively. The pooled means and standard deviations for learning is $M= 32.16$ $(SD = 5.74)$. Descriptive statistics disaggregated by traditional $(n= 19)$ and distance $(n = 101)$ are shown in Table 1.

Table 1

Descriptive Statistics on Dependent Variables disaggregated by types of course ($N=120$)

Variable	Traditional		Distance	
	M	*SD*	*M*	*SD*
Spirit	34.47	2.72	26.16	6.73
Trust	33.73	4.40	26.65	5.57
Learn	35.02	3.10	28.78	5.73
Interaction	31.16	3.63	25.06	5.36

Preliminary assumption testing was conducted. One-Sample Kolmogorov-Smirnov test for normality with a Lilliefor's correction, Shapiro-Wilk test for normality, and a Mahalanobis distances. Normality for the distance education group on the dependent variables of learning was not found tenable at the .05 alpha level. Analysis of the Mahalanobis distance values compared to a critical value yielded one outlier. As it did not appear to significantly influence the data, it remained as part of the data set. Normality on all additional variables was found tenable. Several of the dependent variables were correlated (see Table 2, not shown here); however, the assumption of no multicollinearity was tenable as none of the coefficient were over .80. The assumption of the homogeneity of variance-covariance was tenable based on the results of the Box's test, $M = 36.29$, F (10,4528) = 5.67, p = .23. The results of Levene's test of equality of error provided evidence that the assumption of homogeneity of variance across groups was not tenable for the spirit, trust, learn, and interaction posttest, $F (1,118) = 11.82$, $p < .01$; $F (1,118) = 4.34$, $p =.04$; $F (1,118) = 7.05$, $p < .01$; and $F (1,118) = 4.40$, $p =.04$, receptively. Consequently,

Pillai's Trace was used, instead of the Wilks' Λ, because it is a more robust test when assumptions are violated. There was a statistically significant difference between residential and distance students on the combined dependent variables, Pillai's Trace = .26, F (4,115) = 10.16, p $< .001$, partial η^2 = .26, observed power = ,78. Univariate ANOVAs on the dependent variables were conducted as follow-up tests to the MANOVA. Using the Bonferroni method, each ANOVA was tested at a .013 (.05/ 4)alpha level. These results showed significant differences between the distance and residential students for feelings of spirit, F (1,118) = 48.52, $p < .001$, partial η^2 = .27, observed power = .84; for feelings of trust, F (1,118) = 31.78, $p < .001$, partial η^2 = .24, observed power = .82; for feelings of interaction, F (1,118) = 24.23, $p < .001$, partial η^2 = .17, observed power = .85; and for learning, F (1,118) = 15.81, $p < .001$, partial η^2 = .15, observed power = .88. The residential students scored higher than the distance students on each dependent variable (see Table 1). Based on Cohen's (1988) threshold of .01 for small, .06 for medium, and .14 for large, the effect sizes were large for all the dependent variables, except for learning. The effect size for learning was medium. (Note that numbers and studies are not real).

Items to Report:

- Assumption testing
- Descriptive statistics (*M, SD*)
- Number (*N*)
- Number per cell (*n*)
- Degrees of freedom
- Observed *F* value (*F*)
- Significance level (*p*)
- Follow up ANOVAs
- Effect size and power

There are a number of other related analyses that are note within the scope of the discussion in the guide. However, you want to refer to statistical texts, as they may be helpful. For example, other related analyses may include:

- **Mixed with-in and between groups ANOVA.** This is used to determine if there is a main effect for each IV and whether there is an interaction between the 2 variables. (e.g. Which intervention is more effective in reducing the participants fear score over three periods of time (pre- intervention, during, post intervention)?
- **Analysis of Covariance (ANCOVA).** This is used when you want to statistically control for the possible effects of a confounding variable (covariate) in situations where the criteria is met to use an ANOVA or MANOVA; it statistically removes the effect of the covariate. They can be one-way, two-way, or multivariate (MANCOVA).

Test Selection

To select the appropriate statistical test, it is helpful to ask these questions (derived from Pallant (2007) and Rovai, et al (2013)):

- What are the variables under study? Label them (e.g. independent, dependent, etc.) and identify their scales of measurement (i.e., ratio, interval, ordinal, or nominal).
- Does the research question imply a correlation, a prediction, or a difference test? Or, is the hypothesis an (a) hypothesis of difference or (b) hypothesis of association?
  - If the hypothesis is a hypothesis of association, is the focus a relationship between variables or a predictive relationship between variables? How many variables are being studied? Explain the relationship being examined.
  - If the hypothesis is a hypothesis of difference, is the data independent (e.g., separate groups) or dependent (e.g., pretest/ posttest)? How many categories, levels, or groups are you dealing with (e.g., one, two, or more)? How many dependent variables?
- What assumptions are violated?

Once you have answered these questions, the table below can be useful in selecting the most appropriate analysis. However, it is recommended that you always identify the purpose of the analysis for the study and ask, what statistical procedure aligns with this purpose? For example, "the purpose of the study is to examine the difference between one independent variable with two groups on one dependent variable". What analysis of differences examines one independent variable with two groups on one dependent variable? An independent t test.

Test Selection Chart

Test of Difference				
Independent Variable(s)/ Measurement	Dependent Variable(s)/ Measurement	Type of data (independent/ dependent)/ number of groups	Significance Test	Non-parametric Alternative
1 nominal (2 groups)	1 interval/ ratio	2 independent groups	Independent Samples t-Test	Wilcoxon Test
1 nominal (2 levels)	1 interval/ ratio	1 group measured two times/ dependent	Paired Samples t- Test/ Dependent t- Test	Mann- Whitney U Test
1 nominal	1 interval/ ratio	1 group (compared to a population)	One Sample t- Test	
1 nominal (3 or more groups)	1 interval/ ratio	3 or more independent groups	One way between groups ANOVA	Kruskal- Wallis

1 nominal (3 or more levels)	1 interval/ ratio	1 group measured 3 or more times/ dependent	One way repeated measures ANOVA	Friedman Test
2 nominal (2 or more groups)	1 interval/ ratio	2 or more independent groups for each independent variable; at least two IVs	Two way between groups ANOVA	None
2 nominal (2 or more groups)	1 interval/ ratio	2 or more independent groups for one independent variable and 1 group measured 3 or more times/ dependent; at least two IVs	Within- between groups ANOVA	None
1 or more nominal (2 or more groups)	2 or more interval/ ratio (associated)	2 or more independent groups	MANOVA	None
1 or more nominal (2 or more groups)	1 interval/ ratio and 1 interval/ ratio covariate	2 or more independent groups or 2 or more independent groups for each independent variable; at least two IVs	(one way or factorial) ANCOVA (Note: If data is dependent, then this can be performed as a within group analysis)	None
1 or more nominal (2 or more groups)	2 or more interval/ ratio (associated) and 1 interval/ ratio covariate	2 or more independent groups	MANCOVA (Note: If data is dependent, then this can be performed as a within group analysis)	None
Test of Association				
Variables of Interest/ Predictor Variables/ Control Variables	**Criterion Variables**	**Design**	**Significance Test**	**Non-parametric**
2 ratio/ interval		Relationship	Pearson Product-moment	Spearman's Rank Order
2 ratio/ interval and 1ratio/ interval covariate		Relationship	Partial Correlation	Spearman's Rank Order

1 ratio/ interval (could be ordinal or nominal)	1ratio/ interval	Predictive Relationship	Bivariate Linear Regression	None
More than 1 ratio/ interval	1ratio/ interval	Predictive Relationship	Multiple Regression	Logistic regression
More than 1 ratio/ interval; two set of them		Relationship	Canonical Correlation	

*** Here analysis is presented in isolation; however, in social science research, it cannot be separated from the literature and research design. Think about a dissertation. Remember that the results of a quantitative dissertation are usually in Chapter 3 or 4. The researcher uses the preceding chapters to present and justify the research based on the professional literature. The results chapter needs to logically connect with the preceding chapters and set the stage for the proceeding chapter – the discussion. In the discussion chapter, the researcher discusses or explains the results, provides connections with the professional literature, describes the implications of the research and recommends future research ideas.

References Used to Create this Guide

Anastasi, A., & Urbina, S. (1997). Measuring interests and attitudes. In *Psychological testing* (7th Edition). Upper Saddle River, NJ: Prentice-Hall.

Anderson, C. W. and Loynes, R. M. (1987). *The Teaching of Practical Statistics*. John Wiley and Sons, New York, U.S.A.

Bancroft, T. A. and Huntsberger, D. V. (1961). *Some Papers Concerning the Teaching of Statistics*. Iowa State University Press, Ames, Iowa, U.S.A.

Barnett, V. (1985). *Advanced Level Studies; Statistics, Statistical Background*. Center for Statistical Education, University of Sheffield, Sheffield, U.K.

Baron, R. M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Social and Personality Psychology, 51*(6), 1173-1182.

Campbell, D. T. & Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Corston, R. & Colman, A. (2000). *A crash course in SPSS for Windows*. Blackwell, Oxford.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change"- or should we? *Psychological Bulletin, 74*(1), 68-80.

Diekhoff, G. (1992). *Statistics for the social and behavioral sciences: Univariate, bivariate, multivariate*. Dubuque, IA: Wm. C. Brown Publishers.

Field, A. P. *Discovering statistics using SPSS*. London: Sage.

Fisher, R. A. (1959). *Statistical Methods & Scientific Inference*. New York: Hafner Publishing.

Green, S. B., & Salkind, N. J. (2003). *Using SPSS for Windows and Macintosh: Analyzing and understanding data*, (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Gregory, R. J. (2000). Validity and test development. In *Psychological testing: History, principles, and applications*. Needham Heights, MA: Allyn and Bacon.

Howell, D. C. (2008). *Fundamental statistics for the behavioral sciences* (6th ed.). Belmont, CA: Wadsworth.

Kazdin, A. E. *Research design in clinical psychology.* Boston, MA: Pearson.

Morgan, S. E., Reichert, T., & Harrison, T. R. (2002). *From numbers to words: Reporting statistical results for the social sciences.* Boston: Allyn and Bacon.

Pallant, J. (2007). *SPSS: Survival manua*l. New York: McGraw-Hill.

Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods, 1*, 331-340.

Rovai, A.; Baker, J.D; Ponton, M.K. (2013) *Social science research design and statistics*. Chesapeake, VA: Watertee Press.

Schoeninger, D.W. and Insko, C.A. (1971). *Introductory statistics for the behavioral sciences*. Boston: Allyn and Bacon, Inc.

Sprinthall, R.C. (1994). *Basic statistical analysis*. Boston: Allyn and Bacon.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences. Mahwah, NJ: Lawrence Erlbaum Associates.*

Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics.* Boston: Allyn and Bacon.

Warner, R. M. (2012). *Applied statistics: From bivariate through multivariate techniques (2ed).* Thousand Oaks, CA: Sage Publications.

Wolfe, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: Sage Publications.