The Magic of Pretest/Posttest & Longitudinal Data Analysis 6 Week Webinar Class

CLASS 1:

INTRODUCTION

Presented by

William M. Bannon, PhD

StatsWhisperer, Inc.

www.statswhisperer.com

Welcome to our 6 week webinar class *The Magic of Pretest Posttest & Longitudinal Data Analysis!* Doubtlessly, you have come to this class with specific goals in mind. Our primary goal is to make sure your goals are met. Therefore, please be sure to present all questions you have regarding the materials (email: wb@statswhisperer.com), as well as let us know how we can facilitate your goals being met.

The objectives of this webinar is to review and implement the fundamentals of: 1) Pretest/Posttest Data Analysis; and 2) Longitudinal Data Analysis. These objectives will be met through a six-week **Webinar Agenda** of:

Week 1: Introduction

Week 2: Pretest/Posttest Analysis

Week 3: Longitudinal Data Analysis

Week 4: Attrition/Missing Data

Week 5: Special Considerations in Pretest/Posttest & Longitudinal Data Analysis

Week 6: Write-up and Report

The **primary goal** of this webinar will not be just to present, review, and discuss information, but to illustrate how to apply materials to produce original research. In short, the focus of this program will be on making all the information practical and usable. The most important three words you can say after this webinar:

YES, I CAN!

Defining Pretest/Posttest & Longitudinal Data Analysis

Defining Pretest/Posttest Data Analysis

Before engaging in the actual statistical analysis, it would be wise to define the methods being employed. Subsequently, for our purposes we might define a *Pretest/Posttest Study* as:

1) A study with a dependent (outcome) variable measurement at two timepoints (e.g., pretest/posttest)
2) A study where there is most often an intervention or stimulus between the timepoints
3) A study involving ≥1 groups (e.g., control/experimental)

There are many other considerations that might be involved in the design of such a study, including if subjects are randomized and so on. However, for the purposes of statistical analysis, the three points mentioned above are the most important study characteristics. Specifically, the data analytic methods will largely progress the same based solely upon if the study incorporates a 2-timpoint measurement, along with the number of groups involved in the analysis, aside from the other characteristics of the study.

Next, let's consider an example of a Pretest/Posttest quantitative study. **Figure 1** presents Pretest/Posttest study examining the efficacy of *The Happiness Intervention*.

Figure 1.

Pretest/Posttest Quantitative Analysis Examining *The Happiness Intervention*

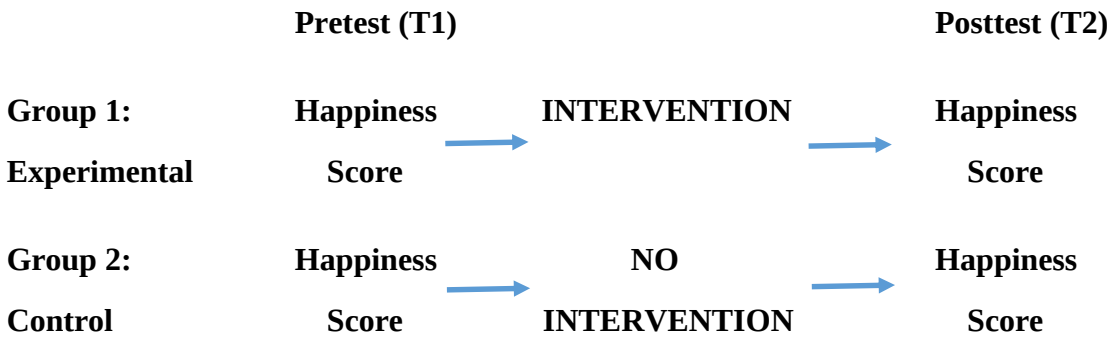


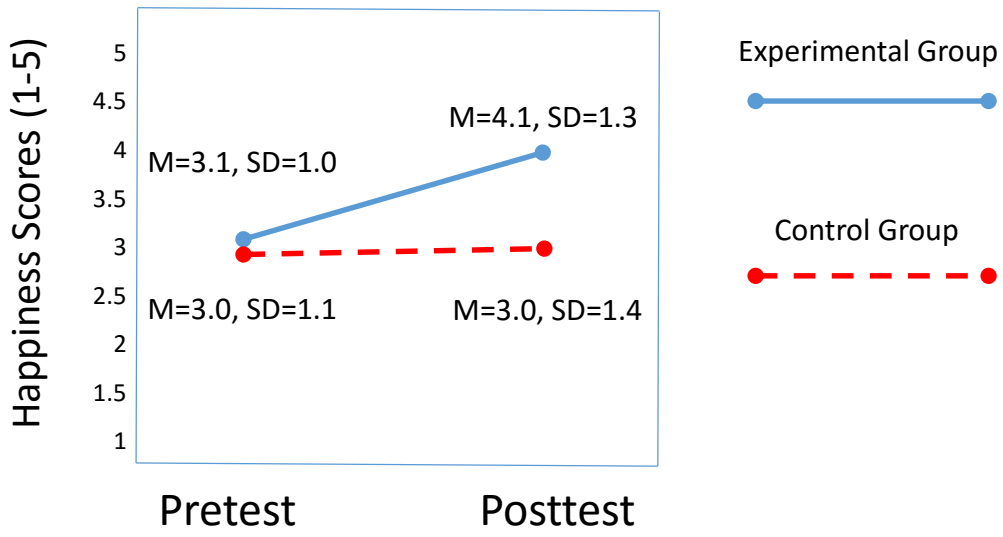
Figure 1 describes that the study incorporates two study groups. Each study group reports a *Happiness Score* at pretest (T1=Timepoint 1) and at posttest (T2=Timepoint 2). The first group is the experimental group, which will receive *The Happiness Intervention*. The second group is the control group, which will receive no intervention.

Figure 2 presents the intervention outcome that would be expected if *The Happiness Intervention* was effective. Keep in mind, the objective of *The Happiness Intervention* is to increase the level of happiness for the intervention participant. Figure 2 describes that at pretest (T1) the intervention and control group evidenced similar levels of happiness. Specifically, on a 1-5 scale measuring happiness (5=Highest level of happiness), the experimental group evidenced a mean score of 3.1 (SD=1.0) and the control group evidenced a mean score of 3.0 (SD=1.1).

However, at posttest (T2), i.e., after *The Happiness Intervention* was delivered to the experimental group, the scores between the two groups diverged. Specifically, the mean happiness score for the experimental group increased to 4.1 (SD=1.3), while the mean score for the control group was virtually unchanged (M=3.0, SD=1.4).

Figure 2

Expectation of the Pretest/Posttest Study: *The Happiness Intervention*



These data reflect the classic pretest/posttest intervention effect. For example, there is no reason to believe the level of happiness for the control group should increase or decrease as there has been no new stimulus (e.g., an intervention) introduced into their environment. Subsequently, in Figure 2 the red dashed line reflecting the control group is flat from pretest to posttest, which indicates no change in happiness scores.

However, the solid blue line reflecting the experimental group takes an upward sloping direction, which indicates an increase in happiness scores from pretest to posttest. Recall, the purpose on *The Happiness Intervention* is to increase the level of happiness for the intervention participant. Therefore, if the intervention is effective the line reflecting the intervention group should make an upward slope, which indicates an increase in scores from pretest to posttest.

The big question: After we observe the general pattern of scores from pretest to posttest between the experimental and control groups, the big question becomes ***are the differences between the mean score pretest to posttest changes significantly different?***

The question of significant pretest to posttest change can be examined in two contexts:

Within group changes: Within group changes would refer the mean score changes within a single group, such as the experimental group. Specifically, we might ask is the pretest to posttest change from 3.1 (SD=1.0) to 4.1 (SD=1.3) significant? This is being asked independent of considering the scores of the other study group (i.e., the control group).

Between group changes: Between group changes would refer to the comparison of the mean score changes between the study groups (e.g., the experimental and control groups) from pretest to posttest. Specifically, we might ask is there significant difference between pretest to posttest group scores when one group evidenced a change from 3.1 (SD=1.0) to 4.1 (SD=1.3) and the second group evidences a change from 3.0 (SD=1.1) to 3.0 (SD=1.4).

Defining Longitudinal Data Analysis

We might define a *Longitudinal Study* as:

1) A study with a dependent (outcome) variable measurement at more than two timepoints (e.g., pretest/posttest/follow-up)
2) A study where there is most often an intervention or stimulus between at least one set of timepoints (e.g., pretest/posttest)
3) A study involving $\geq$1 groups (e.g., control/experimental)

Next, let's consider an example of a Longitudinal quantitative study. **Figure 3** presents a Longitudinal study examining the efficacy of *The Happiness Intervention*. The figure describes that the study incorporates two study groups. Each study group reports a *Happiness Score* at pretest (T1=Timepoint 1), posttest (T2=Timepoint 2), and follow-up (T3=Timepoint 3). The first group is the experimental group, which will receive *The Happiness Intervention*. The second group is the control group, which will receive no intervention.

Figure 3.

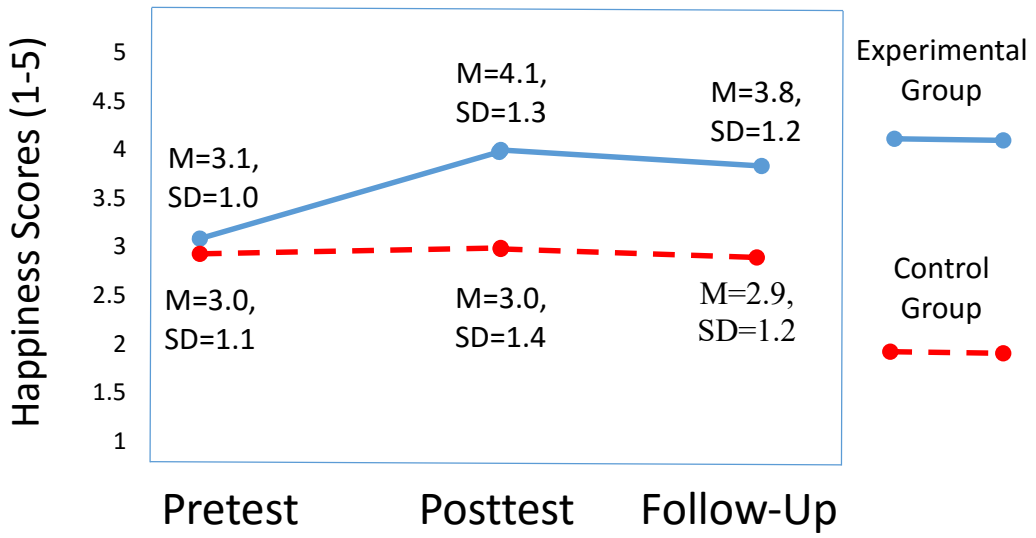
Longitudinal Quantitative Analysis Examining *The Happiness Intervention*

	Pretest (T1)		Posttest (T2)	Follow-Up (T3)
Group 1: Experimental	**Happiness Score** →	**INTERV-ENTION** →	**Happiness Score**	**Happiness Score**
Group 2: Control	**Happiness Score** →	**NO INTERV-ENTION** →	**Happiness Score**	**Happiness Score**

Figure 4 presents the intervention outcome that would be expected if *The Happiness Intervention* was effective. Again, keep in mind, the objective of *The Happiness Intervention* is to increase the level of happiness for the intervention participant. Figure 4 describes that at pretest (T1) the intervention and control group evidenced similar levels of happiness. Specifically, on a 1-5 scale measuring happiness (5=Highest level of happiness), the experimental group evidenced a mean score of 3.1 (SD=1.0) and the control group evidenced a mean score of 3.0 (SD=1.1).

Figure 4

Expectation of the Pretest/Posttest Study: *The Happiness Intervention*



However, at posttest (T2), i.e., after *The Happiness Intervention* was delivered to the experimental group, the scores between the two groups diverged. Specifically, the mean happiness score for the experimental group increased to 4.1 (SD=1.3), while the mean score for the control group was virtually unchanged (M=3.0, SD=1.4). Lastly, at follow-up (T3), the mean happiness score for the experimental group decreased to 3.9 (SD=1.2), while the mean score for the control group was again virtually unchanged (M=2.9, SD=1.2).

These data reflect the classic pretest/posttest/follow-up intervention effect. For example, there is no reason to believe the level of happiness for the control group should increase or decrease as there has been no new stimulus (e.g., an intervention) introduced into their environment. Subsequently, in Figure 4 the red dashed line reflecting the control group is almost entirely flat from pretest, to posttest, to follow-up, which indicates no change in happiness scores.

However, the solid blue line reflecting the experimental group initially takes an upward sloping direction, which indicates an increase in happiness scores from pretest to posttest. Recall, the purpose on *The Happiness Intervention* is to increase the level of happiness for the intervention participant. Therefore, if the intervention is effective the line reflecting the intervention group should make an upward slope, which indicates an increase in scores from pretest to posttest.

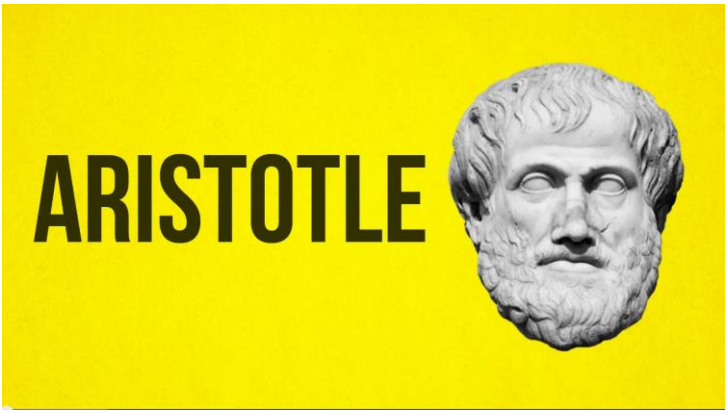
Next, from posttest to follow-up the solid blue line reflecting the experimental group makes a slight decrease from M=4.1 (SD=1.3) to M=3.8 (SD=1.2). This is also classically seen in a pretest/posttest/follow-up intervention. Specifically, there is expected to be an immediate increase in scores for the experimental group at posttest because the intervention will have just been delivered. However, as time goes on the effects of the intervention often fade, which produces a slight decrease in scores at follow-up.

The big question: After we observe the general pattern of scores from pretest to posttest to follow-up between the experimental and control groups, the big question becomes ***are the changes in mean scores between pretest, posttest, and follow-up significantly different?*** Once more, are the changes in mean scores between the three timepoints significantly different *Within group* and/or *Between group?*

This 6-week webinar will prepare us to answer these big questions by instructing on how to use statistical tests to examine if changes between timepoints are statistically significant (in the context of within and between groups) using SPSS software.

Webinar Sample Study

Aristotle once stated “For the things we have to learn before we can do them, we learn by doing them.”



Within this webinar we will illustrate and implement the fundamentals of Pretest/Posttest and Longitudinal data analysis through a sample study.

Specifically, we will test an *intervention to increase Happiness* first using a Pretest/Posttest study design, then using a Longitudinal (Pre/Post/Follow-up) study design. Each study will incorporate two study groups (i.e., an Experimental and Control group).

Overall, in the context of each study we will:

- Prepare data for analysis
- Analyze the data (Pre/Post & Longitudinal)
- Write-up findings

However, before we can approach these quantitative data analysis studies, we must first review the “must know” core concepts needed to effectively implement a quantitative analysis study. The following section will present this information.

The "Must Know" Core data analysis concepts for this webinar

This section will address 7 "must know" core concepts needed to implement a Pre/Post and/or Longitudinal quantitative study. These facts are the foundational learning needed to implement our sample studies. Some of these facts are rather simple, but knowledge of each is absolutely necessary to successfully conduct quantitative analysis.

Must Know Core Concept 1: There are 7 Steps in Quantitative Data Analysis

Must Know Core Concept 2: Parametric and Non-Parametric Statistics

Must Know Core Concept 3: Descriptive and Inferential Statistics

Must Know Core Concept 4: Level of Variable Measurement

Must Know Core Concept 5: Study Variable Type

Must Know Core Concept 6: Three Dimensions of a Relationship

Must Know Core Concept 7: Haste or Paste: Save the Syntax

Must Know Core Concept 1:
There are 7 Steps in Quantitative Data Analysis

Bannon (2013) has listed 7 steps that must be followed in order to produce an effective and legitimate quantitative study. These steps are:

Step 1: Study Map

Step 2: Data Entry

Step 3: Check Data Integrity

Step 4: Univariate Analysis

Step 5: Bivariate Analysis

Step 6: Multivariate Analysis

Step 7: Write-up and Report

These steps form a logical crescendo toward completing a quantitative research study. **Step 1: Study Map**, involves illustrating the relationships being examined within the quantitative study via a diagram, as well as presenting the study independent, dependent, and covariate variables, as well as research question(s) and hypothesis(es) in text. **Step 2:Data Entry**, involves structuring a software database (e.g., SPSS), entering data into the database for analysis, and "cleaning" the data to be sure the data have been entered correctly.

Step 3: Check Data Integrity, involves making sure the data are appropriate for the various statistical tests used in analysis via checks of test assumptions, statistical power, the validity/reliability of study measurement instruments, and analysis of missing data. This is often the most intricate and important step within the statistical analysis, as if the data are not appropriate for analysis the results of the study are likely to be distorted.

Steps 4 through 6 are three steps that involve statistical analysis. **Step 4: Univariate Analysis**, involves the analysis of one variable at a time. Variable are examined individually (via the *Frequencies* procedure) and the characteristics of each are presented (e.g., *Mean, Standard Deviation, Number/Percentage by category*).

Step 5: Bivariate Analysis, involves examining the relationship between 2 study variables (usually the Independent Variable with the Dependent Variable or a Covariate Variable with the Dependent Variable) at a time using various statistical tests (e.g., *chi-square, ANOVA, Correlation, t-test*).

Step 6: Multivariate Analysis, involves using statistical tests (e.g., *multiple linear, logistic regression*) to examine the influence of multiple predictor variables (*e.g., independent and covariate variables*) on a dependent variable. Lastly, **Step 7: Write-up and Report**, involves reporting the results of the quantitative study in text and tables in a professional format (*e.g., AMA, APA*).

Since this section of the webinar illustrates how to perform quantitative analysis, we will focus on **Step 1: Study Map**, as well as **Step 4: Univariate Analysis, Step 5: Bivariate Analysis, and Step 6: Multivariate Analysis**, which are the steps involving statistical analysis. A knowledge of the 7 "must know" core concepts we will be instrumental in successfully implementing these steps.

Must Know Core Concept 2:

Parametric and Non-Parametric Statistics

Statistical tests are often categorized as *parametric* and *non-parametric*. Parametric statistical tests are the most widely applied family of tests in research. However, non-parametric tests are often useful resources to keep in mind. Parametric statistics are a group of tests that make certain *assumptions* (e.g., a normal distribution of scores) about the data used in analysis. Parametric tests used in our study include the t-test, correlation, ANOVA, and linear (multiple) regression. If assumptions are met, these tests can produce accurate and precise estimates. However, if these assumptions are not met, results are likely to be misleading.

Non-Parametric statistics are a group of statistical tests that examine (typically through rank) the relationship between variables without making as many assumptions about the parameters and qualities of the data. A non-parametric test presented in this text is the chi-square analysis. Non-parametric tests are of use in many instances, such as when the dependent variable is an ordinal variable or a rank, as well as when the assumptions (e.g., no undue effect of outlier scores) of parametric testing are not met.

Must Know Core Concept 3:
Descriptive and Inferential Statistics

Statistical procedures are often categorized as descriptive or inferential. Before discussing the differences between descriptive and inferential statistics, it is helpful to review the concepts related to the *population* and *sample* as each pertains to quantitative research. A *population* is the total set of individuals, groups, objects, or events that the researcher is studying. For example, if a study focused upon the employment patterns of recent U.S. college graduates, the population might be defined as every college student who graduated within the past one year from any college across the United States.

A *sample* is a relatively small subset of people, objects, groups, or events selected from the population. This subset taken from the population is intended to *represent* the population in statistical analysis. For example, instead of surveying every recent college graduate in the United States (which would be impractical in term of cost and time constraints), a sample of this population would be surveyed and have their data analyzed. An objective of such an analysis would then be to *generalize* the study findings generated from the study sample, to the larger population. Of course, there are a great many factors that impact how *representative* the sample is of the population, which effects how *generalizable* the findings are from the sample to the population. However, these factors often relate more to study design (e.g., a probability vs. non-probability sample) rather than statistical analysis. Therefore, since this text is focused upon applied data analysis rather than study design and related concepts, these concepts, as well as other important topics such as *probability distributions* will not be explored at length.

Inferential Statistics are the statistical procedures used to make predictions or inferences about a population based upon the observations made through the sample. In other words, the results of the *inferential* statistical analysis are used *to make inferences* about the larger population based upon the quantitative findings involving the sample. Within our study, inferential statistics are largely presented in **Step 5: Bivariate Analysis** (t-test, ANOVA, correlation, chi-square) and **Step 6: Multivariate Analysis** (multiple linear regression analysis and binary logistic regression analysis). *Descriptive statistics* (also described as *univariate statistics* in this text) are a group of statistical procedures used to describe the characteristics of a sample or population. Within this text, descriptive statistics are largely presented in **Step 4: Univariate Analysis**.

Must Know Core Concept 4:
Level of Variable Measurement

Many of the essential decisions made in quantitative analysis are based upon considering the level of measurement of a study variable(s) (also known as the *structure* of a variable). There are generally four levels of measurement regarding a study variable, which are *Nominal*, *Ordinal*, *Interval*, and *Ratio*. Within this section, we will review these four levels of variable measurement in detail, as well as present them in the context of composing categorical or continuous study variables. Specifically, in real world application, variable structure is often referred to as categorical (*Nominal* and *Ordinal* level measurement) or continuous (*Interval* and *Ratio* level measurement). Subsequently, in preparation of real world experience, in this section we will describe variable measurement as such.

Additionally, this section is also the first place we use the term *construct*. It is important to realize while this term is important in research, it is also a bit difficult to define. The word *construct* refers to the *idea(s)* associated with the label given to a variable or concept. It is as if the variable or concept is composed or *constructed* of several ideas and dimensions. For example, the variable *depression* is a construct that is associated with the dimensions of being unhappy, unmotivated, and so on. These dimensions are often measured through the identification of *indicators* that reflect the presence of the construct. For example, an interruption in sleep patterns is often an *indicator* of the *construct* depression.

Categorical study variables

Categorical variables place study participants (or any other *unit of observation*) into a *category*. This section describes the two general types of categorical variables, nominal and ordinal, as well as the definition of a dichotomous variable (a variable with a two category response). Although, the concept of a nominal or ordinal level of measurement is important, this detail is not reflected in the remainder of the text. Specifically, these variables are simply referred to as categorical and/or dichotomous.

Nominal variable. A nominal variable has two or more categories, *but cannot be ranked into an intrinsic order*, such as the variables gender (male/female) and racial identity (Caucasian/White, African American/Black, etc.).

Ordinal variable. An ordinal variable has two or more categories that can be ranked into a meaningful order. For example, if you were to rate a meal you might rank the order of categories as *poor*, *good*, *very good*, and *excellent*. However, there is *not an equal distance between categorical measurement points*. For example, there would not be a reason to believe there is an equal measurement distance between *good* to *very good*, in relation to *good* to *excellent*. In some research, *Ordinal* level variables are treated as continuous, but in this text, these variables are largely treated as categorical.

Dichotomous variable. Many times in research you will hear the term *dichotomous variable*. Therefore, it is important to realize that a dichotomous variable is a categorical variable (either nominal or ordinal level) that has a two category response, such as Yes or No, or good or bad.

Continuous study variables

Continuous variables reflect a greater or lesser degree of a *construct* as scores increase and decrease. This section presents the two general types of continuous variables, interval and ratio. However, this detail is not reflected in the remainder of the text, as these variables are simply referred to as continuous.

Interval variable. An interval variable measures the presence of a construct along a continuum with equal distances between measurement points, *but does not incorporate a meaningful true zero*. For example, an *IQ test* is an interval level measure of *general intelligence*. As IQ test scores rise, a greater presence of *general intelligence* is indicated. Additionally, there is an equal distance between measurement points where the distance between a score of 100 to 110 is 10 points, just like the difference between 120 and 130 is 10 points. However, there is not a meaningful true zero where a score of zero reflects a complete absence of general intelligence.

Ratio variable. A ratio variable measures the presence of a construct along a continuum with equal distances between measurement points and does *incorporate a meaningful true zero*. For example, measuring *weight* in pounds would be considered a ratio level measurement. As weight increases, a greater presence of the construct *weight* is indicated. Additionally, there is an equal distance between measurement points. For example, the distance between 100 and 110 is 10 pounds, just like the difference between 120 and 130 is 10 pounds. There is also a meaningful true zero score, as a weight of zero reflects a complete absence of the construct weight.

Must Know Core Concept 5: Study Variable Type

In any quantitative study it is essential to identify the study variable type. For example, in the sample study presented in this text, the study variable type includes independent, dependent, and covariate variables. Before statistical analysis can be approached, it is necessary to understand the roles of each type of variable in a study, as well as which specific study variables have been designated each role.

Independent variable

The **independent variable** (at times referred to as a **predictor variable** in this text) is the *variable of interest that explains the dependent variable* within the data analysis study. Specifically, the independent variable is tested toward identifying if the factor significantly predicts, causes, or impacts the dependent variable. For example, in each of our sample studies, *study group* (experimental/control) is examined as an independent variable predicting the dependent variable *changes in levels of happiness* between timepoints.

Dependent variable

The **dependent variable** (also referred to as the **outcome variable** in this text) is the *variable being explained through the data analysis study*. The dependent variable is explained by the **predictor variables** (the independent and covariate study variables). For example, in our Pretest/Posttest sample study the dependent variable is the *change in happiness scores* from between pretest to posttest.

Covariate variable

The **covariate variable** (at times referred to as a **predictor variable** in this text and **control variables** in the field) is a *variable in a dataset that may not be the focus of the study, but must be included in analysis because the factor has a significant impact on the dependent or independent variables*. The covariate variable is frequently included in the data analysis study in an attempt to minimize or *control for* the effect of that variable on the relationship of interest between the independent and dependent variables. For example, in our sample study the factors *Income Level* is included as covariate variable because we posit that personal income may have a significant influence on changes in levels of happiness between timepoints. Through including this

covariate variables in analysis, the relationship between the independent variable and dependent variable can be seen more clearly, as the influence of the covariate variable upon the dependent variable is minimized or *controlled for*.

While the covariate variables may not be the focus of the study, these variables are essential to the study. These important factors are instrumental in accounting for the complex set of influences on the other study variables, which is vital to revealing the true relationship between the variables that are the focus of the study.

Must Know Core Concept 6:

Three Dimensions of a Relationship

The goal of most data analysis studies is to examine the relationship between study variables. However, before such a relationship can be examined, a researcher must be aware of the three dimensions of any variable relationship that must be considered, which are *significance, directionality*, and *magnitude*.

Aspect 1: The significance of the relationship

In testing the relationship between variables, significance, or more specifically, statistical significance is expressed through the level of probability (a.k.a., *p*-value), which reflects how likely the relationship presented is a result of chance. Although the level of probability may vary, the typical scientific standard is that probability must be less than 5 times in 100 (i.e., $p<.05$) to achieve statistical significance.

In statistical testing if a relationship between variables is not statistically significant, one would not consider the other two aspects (i.e., directionality, magnitude) of the relationship between variables. Specifically, if the relationship between variables does not achieve statistical significance it is assumed to be by chance, and therefore any other aspects of the relationship are not credible.

Thus, the indicators of significance in statistical testing are:

$p>.05$ = Not a Significant Relationship

vs.

$p<.05$ = Significant Relationship

It is important to realize, that reporting the level of probability, i.e., is probability greater than .05 ($p>.05$) or equal to/less than .05 ($p<.05$), is one of the most focused upon activities in inferential testing. Many times people realize that *p* (probability)

must be at or below .05 for a finding to be statistically significant, but are not sure what that means. As mentioned above, setting the probability level at .05 means that 5 times out of 100 a significant finding would be by chance, but what does that mean?

Essentially, when performing a statistical test, we are asking the statistical procedure to tell us if a significant relationship exists between variables (e.g., Does a significant relationship exist between levels of *Class Attendance* and *Assigned Class Time*). However, at the same time, we realize that sometimes the statistical procedure will in a way lie to us, where the test will indicate a significant relationship exists, when it actually does not. Therefore, we set probability to specify how many times we are willing to be told a significant relationship exists when actually the finding is a result of chance, with the standard being 5 times out of 100 ($p<.05$).

It is rather like, if you knew someone that gave you tips for race horses to bet on. Suppose this person was known to provide you with both winning tips, as well as losing tips. Well, these winning and losing tips could be expressed as a percentage reflecting how many times the tips were accurate, such as 50% of the time the tips given by this person were accurate winners. Realistically, this person would need to give you accurate information a high percentage of the time before you would consider believing the information provided.

In essence, there would be a **cutoff point** for accurate information that must be met before you would listen to him. For example, if he gave you accurate tips that turned out to be winners only 40% of the time (i.e., inaccurate 60% of the time), you might not act on his tips since over half of the time he was leading you astray. However, if 80% of his tips were accurate (i.e., 20% were inaccurate), you might be more willing to listen to him. When we set the probability level at .05, it is as if we are saying that we will only believe these race horse tips if they are accurate 95% of the time (inaccurate 5% of the time). We are again **selecting a cutoff point** for the rate for which we will accept accurate versus inaccurate information.

However, regarding statistics, we are not referring to horse race tips, but how often we are willing to accept inaccurate information regarding variable relationships from a statistical test. Our **cutoff point** is .05 ($p<.05$), which means 95% of the time the statistical test will supply *accurate* information about relationships between variables (the equivalent of accurate race horse tips), but 5% (i.e., .05) of the time we will be told *inaccurate* information about relationships between variables (the equivalent of inaccurate race horse tips) where findings appear significant by *chance*.

That may lead you to ask, what does it mean when significant relationships occur by *chance*? You could think of a *chance* occurrence as a development that gives the appearance of a certain situation, but is actually not that way and is something else (see any episode of the sitcom *Three's Company* as a reference). For example, let's say you did not like pizza. However, one day at work a client gave you his slice of pizza to hold for him as he signed a million dollar contract for you. Right at that point a co-worker saw you holding that slice of pizza and asked you "Oh, you like pizza?" And you say "No, I just happen to be holding this slice *by chance* when you saw me." Note, it appeared that there was a significant relationship (between you and pizza). However, there was not a significant relationship, it only *appeared* that way out of a chance occurrence (you happened to be holding a slice of pizza).

In statistics, a chance occurrence does not look like you and a slice of pizza. The chance occurrence would more likely be a statistically significant relationship between variables on the computer that is not really significant in the real world. For example, consider our research question: Does a significant relationship exist between *Assigned Class Time* and *Class Attendance*? Suppose the findings indicated that residents assigned to classes in the *evening* on average attended classes at a rate twice that of residents who were assigned to classes in the *morning*. This could be expressed as:

1) The average number of days/week for those assigned to evening classes: **4.0**.
2) The average number of days/week for those assigned to morning classes: **2.0**.
3) The statistical test indicated the relationship between *Class Attendance* and *Assigned Class Time* (i.e., morning or evening) is statistically **significant** at the **.05 (p<.05)** level.

In terms of probability being set at .05, we are saying that 95% of the time the above finding will be an accurate significant relationship that actually exists between variables. However, 5% of the time this finding will not be an accurate significant relationship that exists between variables, but will be the result of *chance*, where things sort of fall in place in a way that makes a relationship look significant when it is really not.

Notice, the **statistical findings stay absolutely the same**, but conceptually in our minds we acknowledge that 5% of the time they will not be accurate, but will occur by chance. Thus, it is important to keep in mind when a statistical test indicates that a relationship between variables is significant, **we should not accept that this is absolutely the case**. In explaining the significant relationship between variables,

always remember that the results may have nothing to do with the variables, but might be a function of the *chance* occurrence inherent in statistical testing.

Keep in mind, when probability is set at .05, 5 times out of 100 significant findings will occur by chance, which could also be expressed as 1 out of 20 significant findings will occur by chance. Therefore, if you conduct 20 statistical tests (or around that number) and identify one statistically significant finding from that pool of tests, it is likely that this one significant finding is the finding that has occurred by chance. In statistics, you might hear the term *fishing expedition*, which often refers to a researcher conducting many statistical tests to *fish* for a statistically significant finding. This often implies that the statistically significant finding(s) *fished for* is likely to be the chance finding that will occur when enough statistical tests are conducted.

Aspect 2: The directionality of the relationship

When we are testing the relationship between variables, once we have identified a relationship is significant, we would then want to observe the second relationship aspect, which is *directionality*. Directionality indicates if a negative or positive relationship exists between variables. A **positive relationship** between two variables in statistics typically means as the first variable increases in score, the second variable increases in score. For example, a positive relationship between the two variables Pretest to Posttest *Change in Happiness Scores* and *Income Level* would suggest as scores reflecting *Income Level* increase, Pretest to Posttest *Changes in Happiness Scores* also increase.

A **negative relationship** between two variables in statistics could mean as the first variable increases in score, the second variable decreases in score. For example, a negative relationship between the two variables Pretest to Posttest *Change in Happiness Scores* and *Income Level* would suggest as scores reflecting the *Income Level* increase, Pretest to Posttest *Changes in Happiness Scores* decrease.

Typically, in inferential statistical testing, a negative relationship is indicated by the presence of a minus (-) sign in front of the statistical coefficient. A positive relationship is indicated if there is not a minus sign (no -) in front of the statistical coefficient. The statistical coefficient is the statistic produced in each statistical test, such as the *Pearson's r* coefficient produced in a bivariate correlation. This will be reviewed in greater detail, but as a short introductory example, if the correlation yields a correlation coefficient of $r = -.35$ (the number .35 selected here is arbitrary) the relationship is negative (notice the minus sign), but a correlation coefficient of $r = .35$ (no minus

sign), indicates a positive relationship. Thus, in statistical testing the indicators of directionality are:

Minus (-) Sign = Negative Relationship
vs.
No Minus (no -) Sign = Positive Relationship

Aspect 3: The magnitude of the relationship

Finally, once you find out if a relationship is significant, as well as positive or negative, you might then wonder how strong the relationship is, which is a question of magnitude. In inferential statistical procedures the magnitude of a relationship can be indicated several ways, including the calculation of an effect size.

Essentially, an effect size reflects how impactful one variable is upon another, thereby estimating the magnitude of a relationship between variables. For example, if we consider what factors impact a variable such as Pretest to Posttest *Changes in Happiness Scores*, we could imagine that there are things that might be very impactful such as concerns over one's emotional well-being, as well as things that are less impactful, such as good/bad weather or having a good/bad hair day.

The Small, Medium, and Large Effect Size. In statistical testing, the magnitude of an effect size is categorized as being small, medium, or large. Before we discuss this concept statistically, it may be useful to examine how the concept of small, medium, and large effect sizes might relate to the outcome of our study Pretest to Posttest *Changes in Happiness Scores*.

For example, let's say there are three factors that significantly influence the outcome variable Pretest to Posttest *Changes in Happiness Scores*, which are:

1) **emotional well-being** for that day (how a study participant feels that day)

2) if the **weather** is good or bad (dark sky = dark mood?)

3) a good or bad **hair day** (feel good about look)

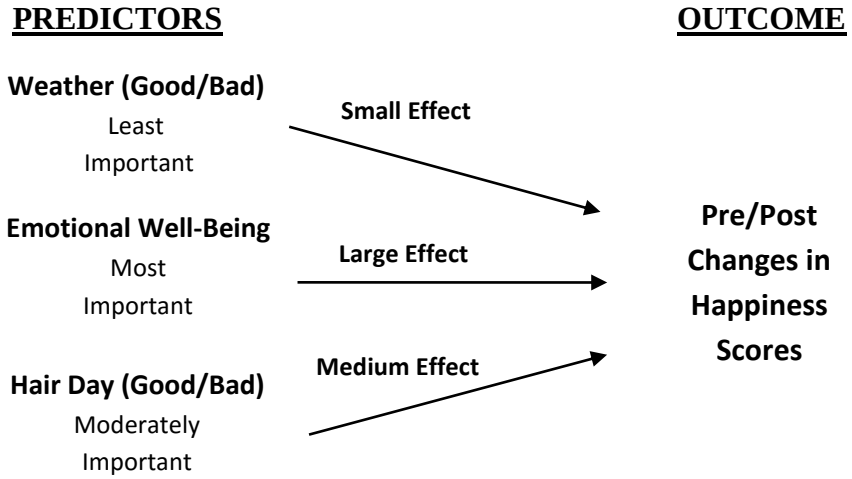
Figure 8 illustrates how the concept of a small, medium, and large effect size might be evidenced by each of these three factors upon the outcome Pretest to Posttest *Changes in Happiness Scores*. For example, you may be able to conceive of how the weather might influence class attendance. However, although significant, it is really your *least important* consideration, so it has a *small effect* on Pretest to Posttest *Changes in*

Happiness Scores. Having a good or bad hair day also has a significant effect and is more important to you than the weather. But in all it is *moderately* important, so it has *a medium effect* on Pretest to Posttest *Changes in Happiness Scores*.

The real driver of Pretest to Posttest *Changes in Happiness Scores* is emotional well-being that day. In other words, if you are already feeling happy the day you take the intervention and report your scores, the pretest to posttest changes in level of happiness will be most enhanced. **Again, all three factors may be *significantly* related to the outcome, but each might play a much different role in analysis due to the *magnitude* of the effect each has upon the outcome**.

Figure 8

Small, medium, and large effect sizes among significant predictors of an outcome



How Can We Indicate a Small, Medium, and Large Effect Size Statistically? Now that we have discussed small, medium, and large effect sizes, we might wonder how each can be reflected statistically. As indicated in **Table 1**, these categories of effect size are reflected by clear numerical values. Please note, the actual raw numerical values reflecting a small, medium, and large effect size often varies by the statistical test used (listed in the left hand column). However, all effect size numbers can be categorized as small, medium, or large. These will be reviewed later as these tests are employed in statistical analysis.

Table 1

Small, medium, and large effect sizes by statistical tests

Effect Size	Small	Medium	Large	A Test Where You Will See this Effect Size Used
R	0.10	0.30	0.50	Pearson *r* Correlation
r^2	0.01	0.09	0.25	Pearson *r* Correlation
Eta-squared	0.01	0.06	0.14	ANOVA
R^2	0.01	0.06	0.14	Linear/Multiple Regression
Cohen's *d*	0.20	0.50	0.80	T-Test
Phi	0.10	0.30	0.50	Chi Square (2X2)
Cramer's V	0.10	0.30	0.50	Chi Square >(2X2)
Cohen's f^2	0.02	0.15	0.35	Linear/Multiple Regression
Odds Ratio	1.44	2.47	4.25	Logistic Regression

Thus, an indicator of the magnitude of the relationship between variables could be described as:

Small Effect Size,

Medium Effect Size,

or

Large Effect Size

In summation. Whenever you conduct an inferential statistical test, you should look at the statistical output for indicators of the three aspects of a variable relationship, which are significance, directionality, and magnitude. As a mnemonic to remember these three aspects, try a term that reflects something like in a *relationship* you must *Savor Daily Miracles* (SDM). An example of a daily miracle in a relationship would be when someone close to you pays you a spontaneous compliment. When things like this happen, it is important to savor these daily miracles. Therefore, when you think of a relationship, think of SDM, S=Significance, D=Directionality, and M=Magnitude.

Must Know Core Concept 7:
Haste or Paste: Save the Syntax

Throughout this text, the instructions for each statistical test and procedure end with "Click **OK** or **Paste**." If you choose to click **OK** the statistical test or procedure will be implemented immediately. If you choose to click **Paste**, the SPSS software will produce the syntax necessary to implement the procedure, which can then be saved as a syntax file. Once the syntax is in this file, all you need to do is *highlight the syntax and run the test by clicking the run icon*, which is the large green button in the top center of the screen. It may be a bit more expedient to click **OK**, but I would urge you to click **Paste**, so you can produce and save your syntax. This is useful on a number of levels, including producing an electronic paper trail verifying the procedures were conducted properly, as well as the fact that you can repeat the procedures through opening and running the syntax instead of repeating the entire analysis.

Approaching The 7 Steps of Data Analysis

After reviewing the *Must Know Core Concepts* we can approach *The 7 Steps of Data Analysis* toward completing our sample studies. In the remainder of this text we will present **Step 1: Study Map**, as well as part of **Step 3: Checks of Data Integrity**, which will be *checks of Statistical Power*.

Applying Step 1: Study Map to the Sample Studies

As stated earlier, our sample studies will involve explaining the factors that influence changes in levels of happiness from Pretest to Posttest, as well as from pretest, posttest, and follow-up between two study groups. Also stated earlier, we mentioned that the process on quantitative analysis incorporates 7 steps, the first of which is **Step 1 (Study Map)**. At this stage of analysis the study variables, research question(s), and hypothesized variable relationship(s) are listed in both text and diagram form. This practice is great for clarity and organization, which is a good thing. The next two pages will present the text/tables study maps for the Pretest/Posttest study and Longitudinal study, respectively.

Study Map 1a (for the Pretest/Posttest study)

Study Map in Text

Research Question 1a: Is the *Happiness Intervention* associated with significantly higher happiness scores from pretest to posttest?

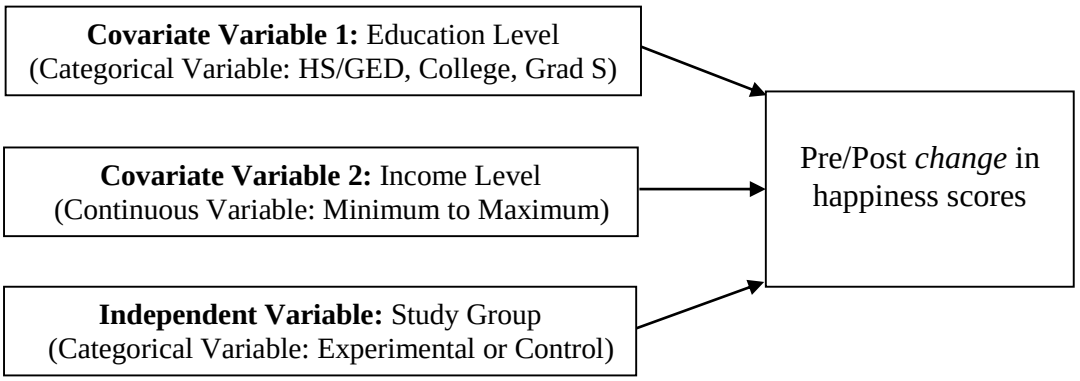
Hypothesis: Participants assigned to the *Happiness Intervention* will evidence significantly higher pre/post happiness scores, while those in the control will not.

Dependent Variable: Pre/Post *change* in happiness scores

Independent Variable: Study Group (Experimental/Control)

Covariate Variable: Education Level, Income Level

Study Map in a Diagram



Study Map 1b

(for the Longitudinal Pretest, Posttest, Follow-up study)

Study Map in Text

Research Question 1b: Is the *Happiness Intervention* associated with significantly higher happiness scores from pretest, posttest, to follow-up?

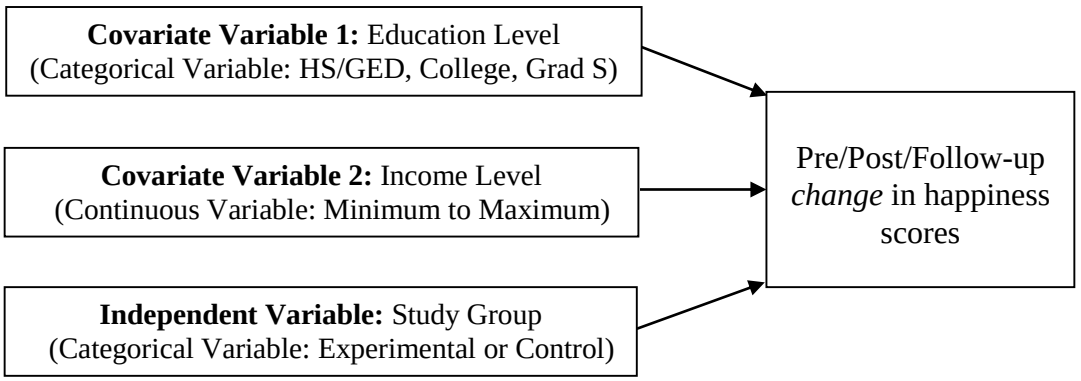
Hypothesis: Participants assigned to the *Happiness Intervention* will evidence significantly higher pre, post, follow-up scores, while those in the control will not.

Dependent Variable: Pre/Post/Follow-up *change* in happiness scores

Independent Variable: Study Group (Experimental/Control)

Covariate Variable: Education Level, Income Level

Study Map in a Diagram



Our two study maps (1a and 1b) have outlined the data analysis plans that will be implemented in our quantitative studies.

Step 3: Checks of Data Integrity

The last topic that will be addressed in this presentation is determining the needed sample size for a Pretest/Postest and/or Longitudinal data analysis study. This question addresses whether or not a study has sufficient statistical power to answer the research question/hypothesis. The consideration of *statistical power* is actually the first area of **Step 3** (**Checks of Data Integrity**) within *The 7 Steps of Data Analysis*. Specifically, there are four areas in which data integrity must be assessed:

1) Statistical Power
2) Missing Data
3) Test Assumptions
4) Study Instruments

This discussion will address the first, *statistical power*, while the remaining three checks will be explored in later presentations.

Power Analysis

The question of the needed sample size to generate sufficient statistical power is most often addressed through conducting the statistical procedure known as a power analysis. There are a great many statistical software programs that may be used to conduct a power analysis. We will consider the G*Power software program as it is adequate, effective, and available free online (http://www.gpower.hhu.de/en.html).

The power analysis may be conceived of as a **dish of 4 ingredients** (see below). When the first three are set (i.e., statistical power, level of probability, and effect size), the power analysis software will compute the final feature, the sample size. As described below statistical power should be set at .80 or above, probability is typically .05 (or below), and the default setting for the effect size is medium. Since we have no compelling reasons to do otherwise, we will use these conventions in our power analysis for the sample studies:

1. **Statistical Power** (Power=.80)
2. **Level of Probability** (ALPHA=.05)
3. **Effect Size** (=Medium)
4. **Sample Size** (needed sample size will be presented based on settings 1-3)

A different power analysis is needed for different statistical tests. Also, keep in mind the power analysis is most often done for the statistical text used in final multivariate analysis, not every test involved in the study. In this webinar series we will utilize two types of tests in final multivariate analysis. Subsequently, we will need a power analysis for each here.

First, we will need a power analysis with in study 1, the Prestest/Posttest Analysis, for a Multiple Linear Regression (MLR). This method will be used for final multivariate analysis when examining *Difference Scores*. We will now present a *Power Analysis* to determine the needed sample size for this analysis using the G*Power statistical software.

Figure 9

G*Power Software: MLR (Pre/Post Difference Scores)

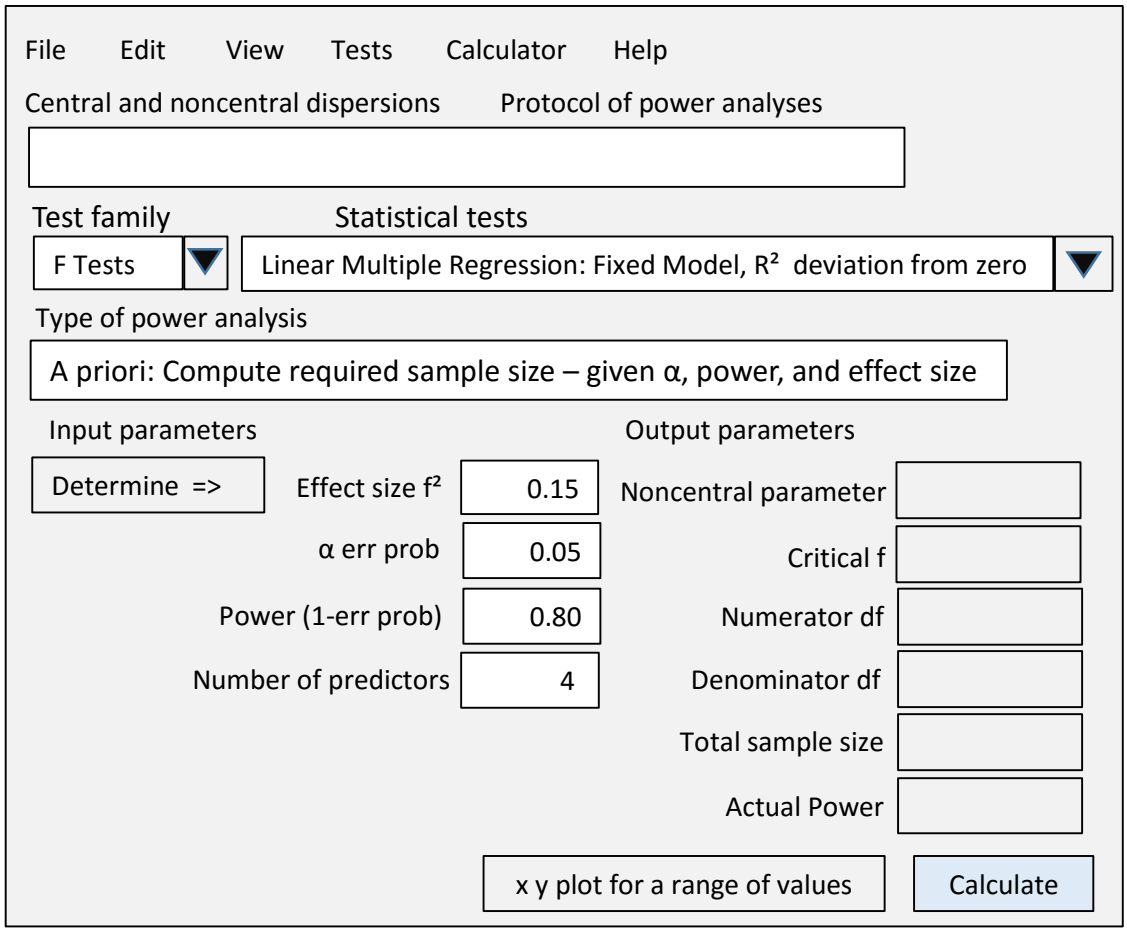
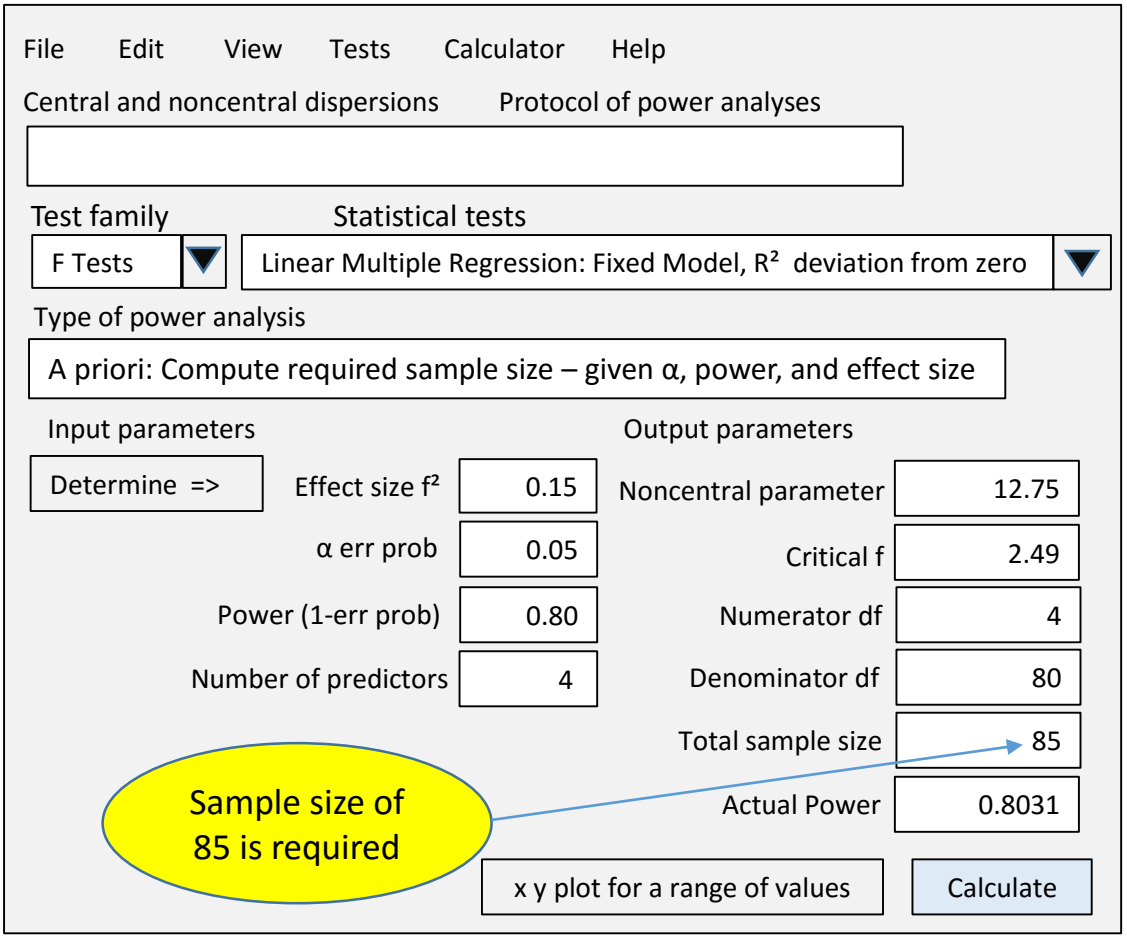


Figure 9 presents the set-up of the G*Power software dialogue box to perform a power analysis for a multiple linear regression. Notice, in the series of boxes in the lower left hand corner of the box we have an the effect size set at medium (Effect size f^2=0.15), alpha is set at .05 (α err prob=0.05), power is set at.80 (Power (1-err prob=0.80), and we have estimated 4 predictors in the multiple linear regression model.

Figure 10 presents the G*Power software dialogue box after the **Compute** button has been clicked (from Figure 9). In the lower right hand corner of Figure 10 the box labeled **Total Sample Size** lists that 85 study participants are needed to generate sufficient statistical power to answer the research question/hypothesis.

Figure 10

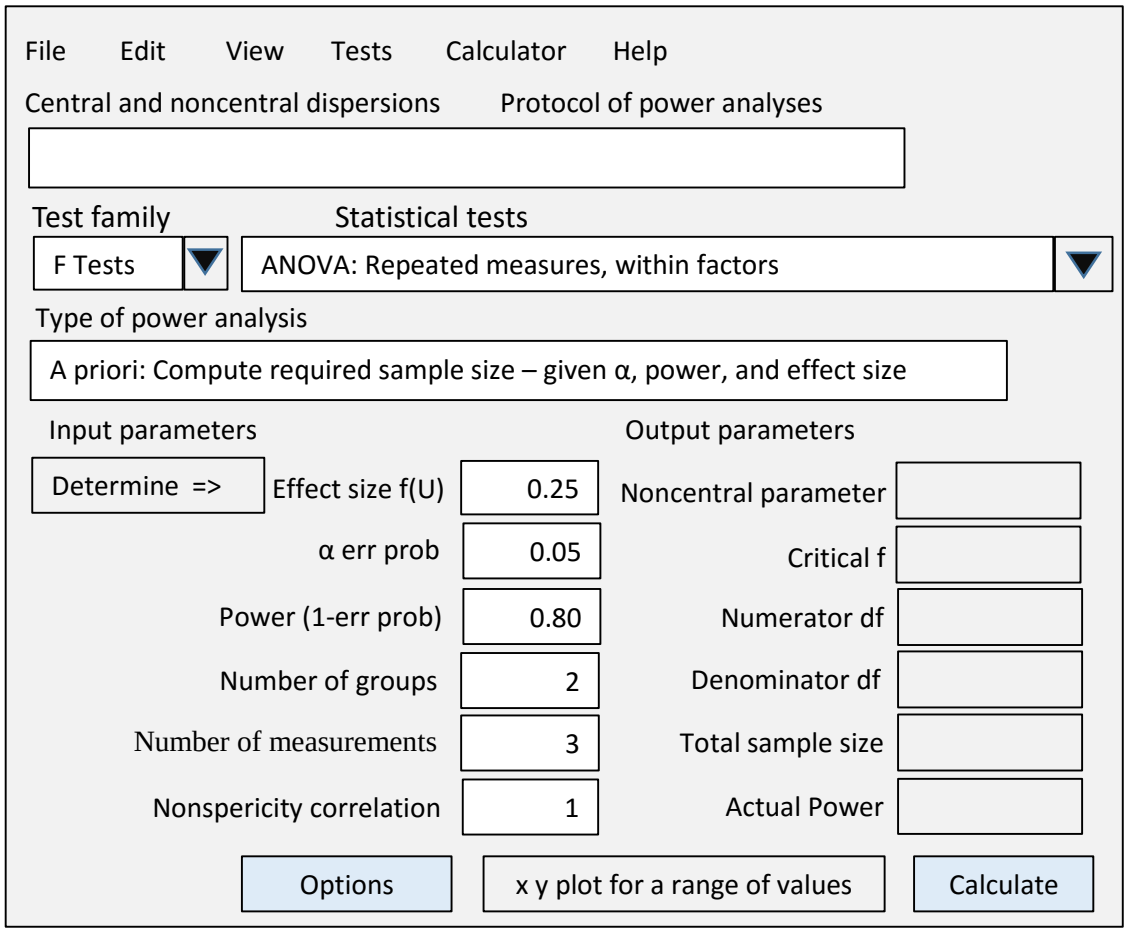
G*Power Software: MLR (Pre/Post Difference Scores)



The next power analysis needed is for a Repeated Measures General Linear Model (GLM), a.k.a., a Repeated Measures MANOVA, which is the procedure used to compare scores at different timepoints (e.g., pretest, posttest, and follow-up). This procedure may be used in our Pretest/Posttest study, as well as our Pretest/Posttest/Follow-up study. We will use our study with the 3 timepoints (i.e., pretest, posttest, and follow-up) in this power analysis.

Figure 11

G*Power Software: Repeated Measures General Linear Model (GLM)



Actual power

Figure 11 presents the set-up of the G*Power software dialogue box to perform a power analysis for a Repeated Measures General Linear Model (GLM). Notice, in the series of boxes in the lower left hand corner of the box we have an the effect size set at medium (Effect size f(U)=0.25), alpha is set at .05 (α err prob=0.05), and power is set at.80 (Power (1-err prob=0.80). We have entered 2 for Number of Groups (i.e., Experimental/Control) and 3 for Number of Measurements (i.e., Pretest/Posttest/Follow-up).

Next, click the button labeled **Options** in the lower left hand corner of the dialogue box. The new dialogue box in **Figure 12** will open. Within this box check "**As in SPSS**" then click **OK**.

Figure 12

G*Power Software: Options Box

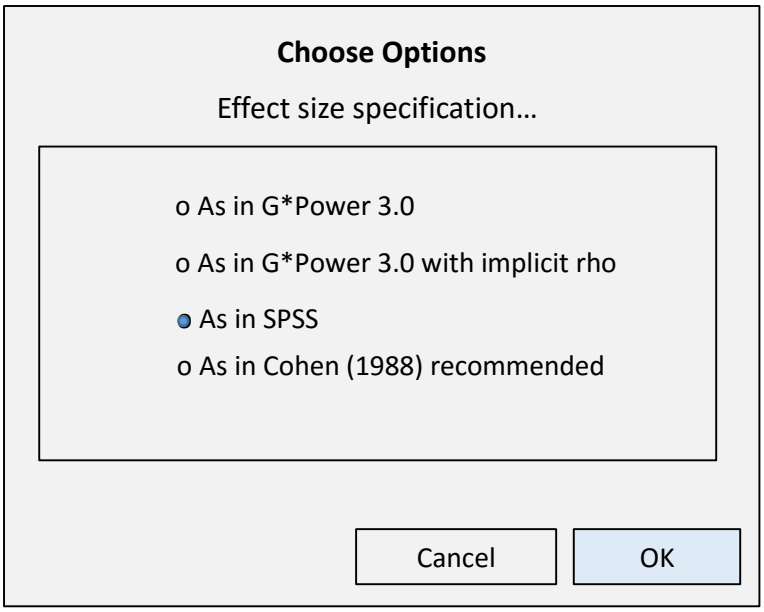
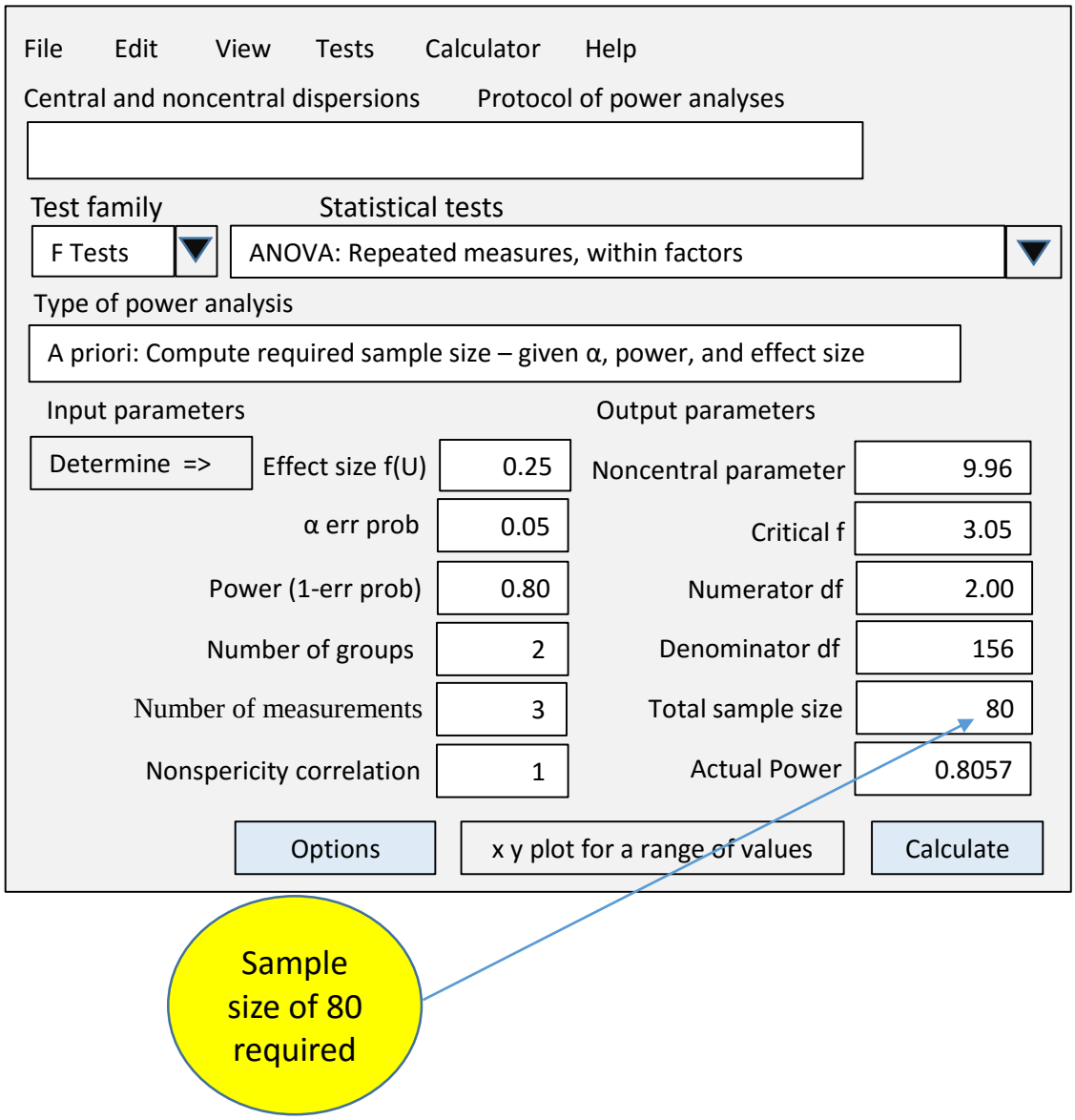


Figure 13 presents the G*Power software dialogue box after the **Compute** button has been clicked (from Figure 11). In the lower right hand corner of Figure 13 the box labeled **Total Sample Size** lists that 80 study participants are needed to generate sufficient statistical power to answer the research question/hypothesis.

Figure 13

G*Power Software: Repeated Measures General Linear Model (GLM)