The Ideal Qualitative Study with a Quantitative Component
6 Week Webinar Class

CLASS 1:

INTRODUCTION & QUANTITATIVE 1

Presented by
William M. Bannon, PhD
StatsWhisperer, Inc.
www.statswhisperer.com

FIRST SECTION:

INTRODUCTION

Welcome to our 6 week webinar class *The Ideal Qualitative Study with a Quantitative Component*. 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@williambannonassociates.org), 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) Qualitative Data Analysis; 2) Quantitative Data Analysis; and 3) Mixed Methods Data Analysis. These objectives will be met through a six-week **Webinar Agenda** of:

Week 1: Intro. & Quantitative Analysis Part 1

Week 2: Quantitative Analysis Part 2

Week 3: Essentials of Qualitative Analysis

Week 4: Essentials of Mixed Methods Research

Week 5: Statistical Analysis with Small Samples

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!

The Traditional Roles of Quantitative & Qualitative Research Methods

Traditionally, Quantitative methods (research primarily involving measurement via numbers and statistical tests) and Qualitative methods (research primarily involving measurement via the written word and narratives) have had a somewhat linear relationship in the research process. For example, **Figure 1** presents a general representation of the inductive reasoning process associated with traditional qualitative research.

These steps are largely seen as the part of the research process where the researchers attempt to identify *What's going on?* Essentially, the researcher **observes** some phenomenon (often understudied or not yet studied) among a **small number** of study participants (e.g., a researcher observes that small number of people pray while skydiving). Next, using qualitative research methods, the researcher detects a **pattern**(s) that seems to be associated with how the phenomenon under study operates (e.g., a researcher further notes the higher the skydiving jump the more intense the prayers).

Next, the data is then used to begin to form a **tentative hypothesis**(es) and/or **inform theory** that begins to explain how the factors or variables within this phenomenon relate to one another (e.g., the researcher may hypothesize the level of risk posed by skydiving is associated with a more intense interest in one's well-being). *This is regarded as the beginning of the study of a certain topic or phenomenon.*

Figure 1.

<u>The Traditional Inductive Reasoning Process of Qualitative Research</u>
Inductive Reasoning asks: *What's going on?* within the **research process:**

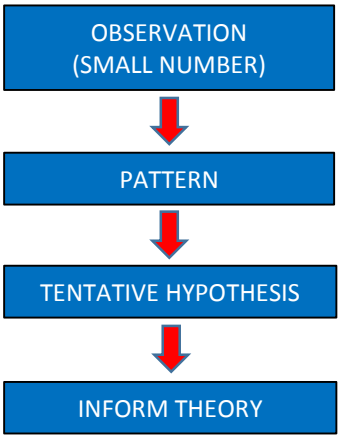


Figure 2 presents a general representation of the deductive reasoning process associated with traditional quantitative research. These steps are largely seen as the part of the research process where the researchers attempt to identify *Is this what's going on?* Essentially, the researcher **tests** the preliminary **hypothesis**(es) and **theory** identified within the qualitative study using more rigorous research methods, including **a large number of observations/a larger sample size**.

Additionally, the researcher would move beyond simple observation and may measure the level of prayer with a survey research scale that numerically gauges the gradation of prayer, as well as the actual height of the jump in approximate feet. Statistical tests are then used to gather evidence that will either support or not support (i.e., **confirm**) the study theory and/or hypothesis regarding the relation of prayer to the height of the skydive jump.

This is regarded as the second half of the research process where the relation of the factors comprising the topic or phenomenon are supported or not supported.

Figure 2

<u>The Traditional Deductive Reasoning Process of Quantitative Research</u>
Deductive Reasoning asks: *Is this what's going on?* within the **research process:**

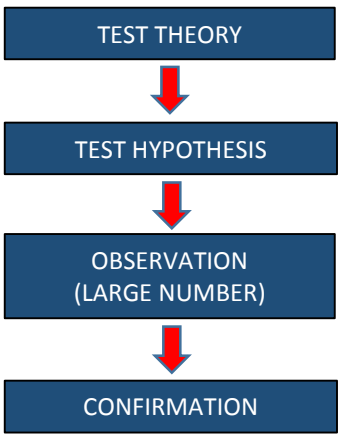


Figure 3 presents the traditional representation of conducting research as an indicative and deductive process, using qualitative and quantitative methods, respectively. We can see that in this classic progression the qualitative and quantitative research methods form **two distinct phases of the research process. Briefly, the methods comprise a useful crescendo** that may identify and tests hypotheses and theories in the name of understanding how the world around us works.

Figure 3

The traditional representation of conducting research as an indicative and deductive process, using qualitative and quantitative methods, respectively

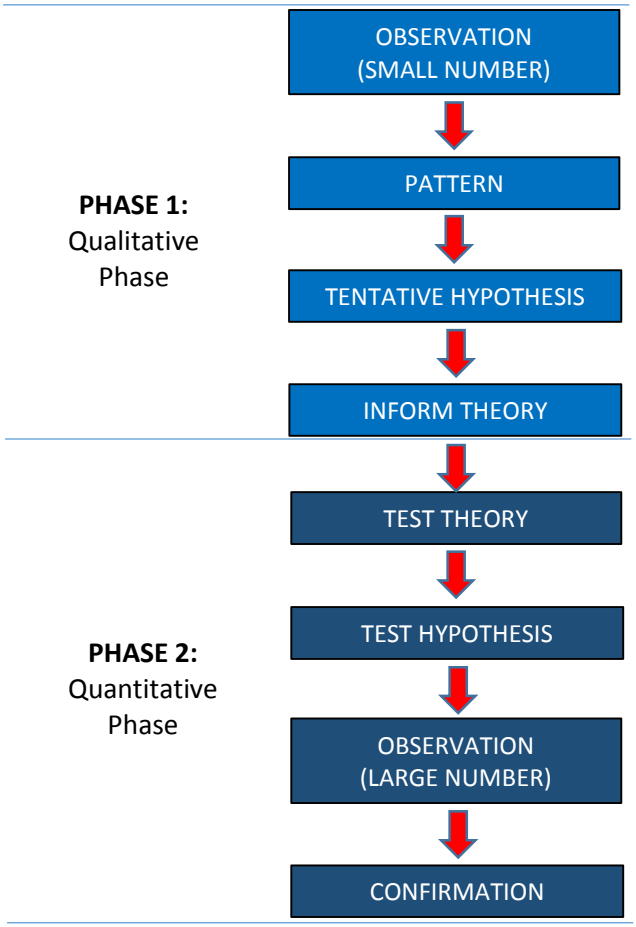


Figure 4 suggests another way of utilizing qualitative and quantitative methods to produce sound and informative research. Specifically, this figure suggests that instead of a linear relationship, the worlds of qualitative and quantitative research methods, or words and numbers (respectively) might be employed in a parallel manner that might simultaneously inform a single research study. However, now the questions begs: *Why is this necessary?*

Figure 4

These Methods Can Also Be Two Parallel Worlds Informing One Study

It is quite easy to see why the combination of qualitative and quantitative methods are beneficial within a single research study when we consider the potential shortcomings of a purely qualitative or quantitative research study.

Before we review this topic, consider this subject from a practical standpoint. A qualitative research study has many considerable strengths, such as examining contextual factors, as well as presenting more in-depth and detailed responses. However, qualitative studies often incorporate limitations such as not incorporating statistical tests to identify association and presenting findings that are limited in terms of generalizability.

However, if a qualitative research study was enhanced with a quantitative component, this study might incorporate the strengths of a qualitative study, without the limitations traditionally associated with qualitative research (i.e., no statistical tests to identify association and limited generalizability). Thus, we might ponder, might the combination of qualitative and quantitative methods create a **Super Study?**

The Potential Shortcomings of a Purely Quantitative Research Study

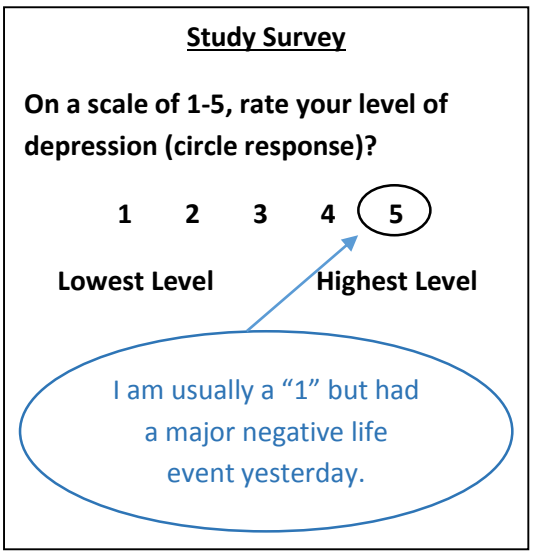
To illustrate the potential shortcomings of a purely quantitative research study, suppose we wanted to conduct a study where we quantified the typical general level of chronic depression within a community. Specifically, suppose we administered a single-item measure to examine the level of chronic depression (among the first 100 community members that pass by a random street corner in the community), which read:

On a scale of 1-5, how would you rate your level of depression?
(with 5 reflecting the highest level of depression).

Almost invariably, we would get at least one survey back that had a commentary inserted similar to that presented within **Figure 5**. In essence, the respondent wished to voice a comment, but there was no place on the quantitative study survey to enter the text. Subsequently, the respondent inserted his comment (see the circled blue text) in a random section of the survey. This is a common occurrence in quantitative research. Typically, the written text is simply discarded, as there is no place to enter the data within the database as it is unanticipated.

Figure 5

The Purely Quantitative Study Survey



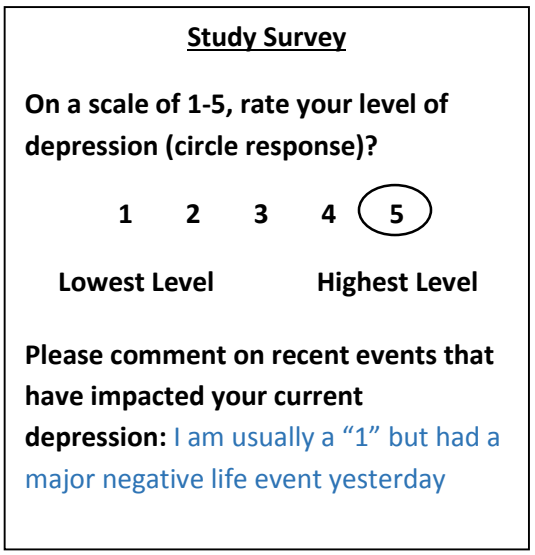
When we read the insertion, we see written: I am usually a "1" but had a major negative life event yesterday. Please recall, our study is intended to measure the level of *chronic* depression within the community. What the respondent has indicated is that he is usually a "1," which reflects the lowest level of depression. Although, as per his response he will be coded in the database as a "5," which reflects the highest level of depression. However, this number is misleading because his depression is not chronic, but circumstantial. Thus, the level of chronic depression in the community is misrepresented (inflated) because there was no space for respondents to indicate whether this rating reflected their typical level of depression or not.

In essence, we can see quite easily that the purely quantitative study might easily misrepresent the level of chronic depression within the community.

The limitation posed by the purely quantitative study survey can be quite easily addressed with the addition of a qualitative element. For example, **Figure 6** on the presents the original quantitative study survey with room for commentary (a qualitative response) at the bottom of the study survey if needed. The addition allows for a fuller expression of the *perspective* of the study participant, which is key information in a research study.

Figure 6

The Quantitative Study Survey with an Added Qualitative Element



The Potential Shortcomings of a Purely Quantitative Research Study

In our former example, the inability to note the qualitative aspects of a respondent might be termed the *blindside* of quantitative research. In research, one might consider the *blindside* of a study to be an aspect of the data not visible to the researcher, which might make the study vulnerable to distorted or inaccurate findings (e.g., the inflated level of chronic depression posed in our example). It is important to note that a purely qualitative study might also have blindside.

Consider we are instead conducting a purely qualitative study in the same community that asks participants to list factors related to their feeling depressed. Again, assume we will interview 100 study participants in the same fashion (i.e., the first 100 passersby on a random street corner).

The *blindside* of this purely qualitative study may relate to the common practice in qualitative research to not report the *number of study participants* that evidenced a certain theme. For example, a purely qualitative study may report "*A theme emerged linking feelings of depression to worries about paying monthly household bills*."

The purely qualitative study would not mention the number or percentage of study participants evidencing this theme. Computing the number and percentage of participants that evidence a theme is a quantitative method known as a *frequency count*.

Thus, we are left to wonder what number and percentage of these 100 respondents provided qualitative data that reflected this theme. Did 70% (70 of 100) of respondents evidence qualitative data reflecting this theme? Did 85% (85 of 100) of respondents evidence qualitative data reflecting this theme?

This is problematic as perhaps the number and percentage of respondents that must evidence certain qualitative data to constitute a *theme* might be different than the researcher reporting the findings. For example, I believe that 80% of respondents must evidence data on a certain topic for it to be considered a theme.

However, suppose the researcher believes only 70% of respondents must evidence like comments to constitute a theme. Due to the fact the researcher has not reported the number and percentage of respondents that have made like comments, I cannot see what he or she has termed comments as a theme that do not meet my criteria as a researcher.

Lastly, the number and percentage of respondents that mention a topic makes a big difference in terms of impact, prevalence, and so on. Thus, one can easily see how the purely quantitative study may incorporate important limitations.

However, one can also easily see how adding even a small quantitative component such as a *frequencies count* regarding the number and percentage of respondents that mention certain qualitative data that constitutes a theme.

The Power of Combining Dichotomies

The possibility of creating a superior product through combining complementing elements is not new. Consider the theme presented by the 1980s Rees's Peanut Butter Cup commercial where an unintentional collision resulted in a combined chocolate & peanut butter product.

Image 1

1980s Rees's Peanut Butter Cup commercial

After the collision, the person originally holding the peanut butter would state "*Your chocolate is in my peanut butter*," which was promptly retorted with the person originally holding the chocolate stating "*Your peanut butter is in my chocolate*." However, after the collision each person sampled the combined food. Next, each person stated the word "*Delicious*." The tone suggested the combination of foods was unexpectedly preferable to each food individually.

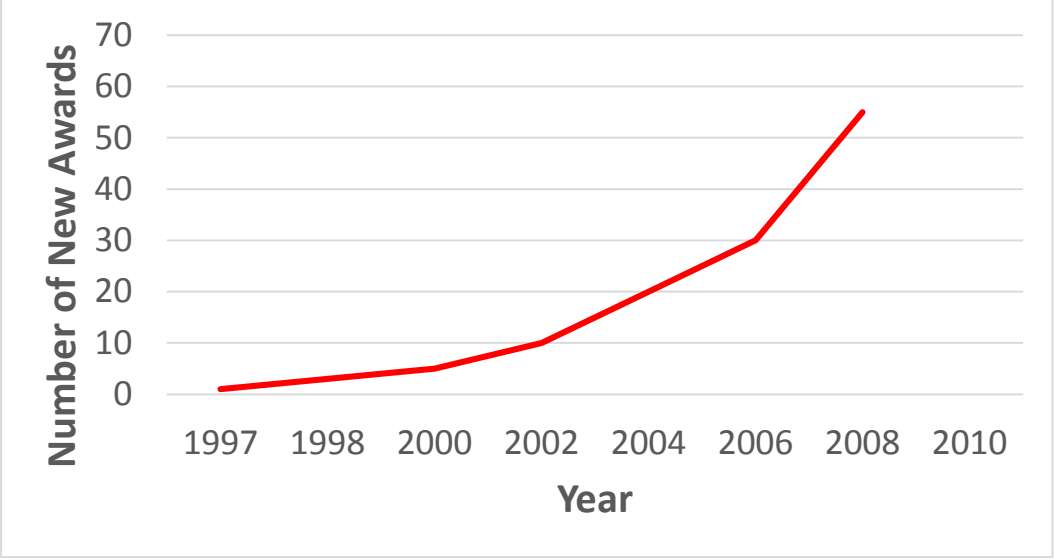
This scenario is actually very pertinent to combining qualitative and quantitative research methods. Specifically, the tone of the commercial suggested that chocolate and peanut butter were fine as separate products. Indeed, there was a palpable initial resistance on behalf of each person regarding combining the two foods. Likewise, in the field of research there may be a hesitation to combine qualitative and quantitative methods. However, just as in the commercial, the combination might result in a rich product not possible using one method alone.

Combing quantitative and qualitative methods is commonly referred to as **Mixed Methods research**. Seemingly, an increasing number of researchers have realized how combining qualitative and quantitative methods in mixed methods research produces sound and effective research. For example, **Figure 7** presents the increasing number of NIH Funding Awards that Self-Identify as Mixed Methods over the finding period from

1997 to 2008. The red line reflects a steep increase in the number of funded applications employing mixed method research over this time period.

Figure 7

NIH Funding Awards that Self-Identify as Mixed Methods from 1997-2008 (*N* = 226)



SOURCE CRISP DATABASE; PLANO CLARK (2010)

Table 1 presents the number of dissertations and theses containing the term *Mixed Methods* in the citation or abstract from 1980 to 2009. It is clear that the number of dissertations with the term mixed methods in the title from 2005-2009 is considerably higher than the previous 25 years combined.

Table 1

This number represents the number of dissertations and theses containing the term *Mixed Methods* in the citation or abstract

Year Range	Number
2005-2009	2524
2000-2004	532
1995-1999	100
1985-1989	17
1980-1984	3

Note: The source of these numbers is the "proquest" search engine (Proquest Search Engine, 2009). Haines, C., 2011 unpublished dissertation, University of Nebraska – Lincoln.

Clearly, the value of mixed methods research seems to be incrementally apparent in the field of research. However, performing mixed methods research presents a special set of challenges that a purely qualitative or quantitative researcher may not face. Specifically, the purely qualitative or quantitative researcher may only need to have knowledge about his or her specific research method, where the mixed methods researcher must have a working knowledge of the essentials of:

1) Producing a Qualitative Study

2) Producing a Quantitative Study

3) Mixed Methods research (how to combing the two forms of methods)

Note this is a 3 step process. In other words, one can't just jump into doing mixed-methods research. Certainly, there must be a foundational knowledge of both qualitative and quantitative methods before a mixed methods study is appropriate for a researcher.

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 quantitative, quantitative, and mixed methods research through a sample study.

Within this study, the administrators of the retirement community *Lotsa Noise Pines* would like to encourage residents to attend meditation classes. The administrators have been offering morning and evening classes for about one year and have noticed that the evening classes seem to be packed, while the morning classes have uneven attendance.

The process of offering morning or evening classes to residents includes randomly giving meditation class daily passes to all residents on the first day of the month. There are an equal number of classes for morning and evening classes. Again, these daily passes are assigned randomly to community residents via a delivery person who places these passes in each of the physical mailboxes of each community resident.

Each pass specifies if the resident is assigned for a morning or evening class for a particular day. Each resident is assigned approximately 50% morning class passes and 50% evening class passes in a given month.

Thus, it seems when a community resident is assigned to an evening class in a given day, he or she seems much more likely to actually attend the meditation class.

The administrators have recorded the attendance level of each community resident over the past month. The administrators want to know the average weekly attendance rate for each community resident. These data will be used as the outcome variable (i.e., the dependent variable) for our study.

The administrators are considering moving both classes to the evening to encourage attendance. **Their initial thought is that residents attend classes more often when they are assigned an evening class because evening classes are more convenient than morning classes for their schedules.** However, the administrators are fearful this might be the wrong move as the literature states meditation is generally preferred in the morning among this population.

In essence, the administrators believe the predictor of interest regarding attending classes (the study *dependent variable*) is assigned class time in the morning or evening (the study *independent variable*). The administrators also believe and want to consider in analysis other factors that influence class attendance, which are age, doctor

prescription status (i.e., did the doctor prescribe meditation classes?), and level of class enjoyment (these are the study *covariate variables*). A detailed definition of the roles of the study covariate, independent, and dependent variables will be presented in the next section.

Subsequently, to generate a well-rounded picture of the situation, the administrators have asked us to:

- Analyze survey data regarding class attendance (Quantitative Analysis)
- Conduct/Analyze interviews regarding class attendance (Qualitative Analysis)
- Merge/Interpret results regarding class attendance (Mixed-Methods Analysis)

The remainder of this webinar will illustrate the foundations of these research methods through carrying out these analyses. The first analysis that will be conducted is the quantitative portion, as the administrators have collected surveys from the 100 residents that describe their class attendance over the past month, as well as their age, status regarding if their doctor prescribed meditation, their level of enjoyment regarding meditation class, and if they attend class in the morning or evening. These survey data have already been entered in SPSS for us by the administrators.

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

SECOND SECTION:

QUANTITATIVE ANALYSIS 1

This section on *Quantitative Analysis 1*, will discuss the 8 "must know" core concepts needed to implement a quantitative study, as well how to draw up the Study Map (Step 1 within the 7 Steps of Data Analysis). These activities will serve as our foundational learning before we implement quantitative data analysis. The actual quantitative data analysis for our study will be performed in the next section on *Quantitative Analysis 2*.

The "Must Know" Core data analysis concepts

This section will review several "must know" core concepts that you must understand before you approach a quantitative analysis study. We will review 8 core concepts. Many of these concepts 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: The Clarity of the Study Hypothesis

Must Know Core Concept 7: Three Dimensions of a Relationship

Must Know Core Concept 8: 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 8 "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 the sample study the administrators believe the predictor of interest is *Assigned Class Time* (morning or evening) is examined as an independent variable predicting the dependent variable *Class Attendance*.

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 sample study one presented in this text, the dependent variable *Class Attendance* is examined in relation to the independent variable *Assigned Class Time*, as well as the covariate variables *Age, Doctor Prescription Status,* and *Level of Class Enjoyment*.

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 *Age, Doctor Prescription Status,* and *Level of Class Enjoyment* are included as covariate variables because the administrators of the retirement community believe each has a significant influence on the dependent variable. Through including these covariate variables in analysis, the relationship between the independent variable and dependent variable can be seen more clearly, as the influence of each 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:
The Clarity of the Study Hypothesis

A hypothesis is a predictive statement regarding the relationship between study variables. The hypothesis is the central figure in a quantitative study. Specifically, the study hypothesis is like the sun within the universe known as the data analysis study, around which the planets known as the steps of data analysis revolve. Quite appropriately, the essentials of the study hypothesis are regularly presented in statistics textbooks. For example, the hypothesis(es) is typically presented in terms of:

The Alternative and Null Hypothesis. The alternative hypothesis (also called the research hypothesis) and null hypothesis are the two rival hypotheses, one of which will be supported through statistical testing. The **alternative hypothesis** (**research hypothesis**) is the relationship between variables that the researcher believes exists and is testing for (e.g., *There **is** a significant relationship between Assigned Class Time* and *Class Attendance*). The **null hypothesis** is the default position that states that there is no relationship between the variables under examination (e.g., *There **is not** a significant relationship between Assigned Class Time* and *Class Attendance*).

Statistical testing ultimately results in *rejecting* either the alternative (research) or null hypothesis based upon the results of statistical testing. For example, if the statistical testing indicated that *Assigned Class time* and *Class attendance* are significantly related ($p<.05$), the null hypothesis will be rejected and the alternative hypothesis (research hypothesis) would be supported. However, if the statistical testing indicated that *Assigned Class Time* and *Class Attendance* are not significantly related ($p>.05$), the null hypothesis will be supported and the alternative hypothesis (research hypothesis) would be rejected.

Type I Error and Type II Error. Also conceptually relevant to the hypothesis is the concept of **type I error** and **type II error**. **Type I error** involves a rejection of a true null hypothesis. Specifically, type I error leads one to conclude that a significant relationship between variables exists, when in reality it does not. For example, a type I error would occur when analysis indicates a significant relationship exists between *Assigned Class time* and *Class attendance*, when the relationship does not exist. A **type II error** is the failure to reject a false null hypothesis. Type II error would occur

when analysis indicates a significant relationship does not exist between *Assigned Class time* and *Class attendance*, when the relationship does exist.

Directional and Non-Directional. The hypothesis is also described as being directional or non-directional. A directional hypothesis posits how variables will be related, such as the hypothesis stating *Those who are assigned class in the evening will have higher class attendance levels relative to those assigned class in the morning*. A non-directional hypothesis posits a relationship exists, but does not specify how the variables will be related, such as the hypothesis stating *There is a significantly different level of class attendance between people assigned class in the evening and morning*.

The concept of the directional hypothesis relates to the **one-tailed test** and **two-tailed test**, which is available in certain statistical procedures (e.g., correlation, t-test). Specifically, some tests provide an option to select a one-tailed test or two-tailed test, which are alternative ways of computing the statistical significance ($p<.05$). The **one-tailed test** is said to be appropriate when a directional hypothesis is considered, while a **two-tailed test** is said to be appropriate when a hypothesis is non-directional.

However, you may notice within the quantitative research literature (as well as the statistical analysis conducted in this text) the two tailed test is applied consistently, even when a hypothesis is directional. This is likely related to the fact that the two-tailed test is a more rigorous test, which may suggest a more stable relationship between variables. In other words, a one-tailed test indicates a statistically significant relationship more easily, so we might favor the more stringent two-tailed test. Thus, even though the hypotheses are directional in sample studies one and two, analysis will only employ two-tailed tests.

However, beyond the fundamental concepts related to the hypothesis, lies the essential consideration of hypothesis *clarity*. *Clarity* is an indispensable virtue related to the hypothesis. Therefore, we will now discuss several elements related to hypothesis clarity, as well as a test to examine if the hypothesis is clear in terms of being succinct.

Succinct

A potentially useful test that can be applied to examine if the hypothesis is succinct is the *first half/second half test*. I noticed when I began writing hypotheses they were so long that by the time I read the second half of my hypothesis, I had forgotten what the first half said. I realized that was a bad sign, so a regular test of mine became *reading the hypothesis aloud and observing if the hypothesis was held in my mind as one thought without rereading it*. Chances are if you cannot do this, your hypothesis can be shortened and made more succinct.

Clarity

The hypothesis is to an empirical study, what the bull's-eye in the center of a target is to an archer. The archer needs that bull's-eye to focus and guide his or her actions to achieve profound success. All actions, thoughts, and considerations revolve around hitting that bull's-eye. The more clearly the archer can see the bull's-eye, the better are his or her chances of success. If the archer's view of the bull's-eye is cloudy, the better the chances the target will be missed.

Likewise, for a quantitative researcher, all the actions, thoughts, and considerations involved in the data analysis aspect of a study revolve around effectively answering the study hypothesis(es). Furthermore, the more clearly the researcher can see the hypothesis, the more likely he or she will be able to effectively indicate if the hypothesis is supported or not supported. If the researcher's view of the hypothesis is unclear, the better the chances the hypothesis will not be answered properly. Thus, please note the following three points regarding hypothesis clarity to maximize the chances of study success.

Point 1: Present one relationship at a time. The first step in making the hypothesis clear is to first identify the constructs (e.g., independent and dependent variables) to be included within the hypothesis and be sure only <u>one relationship among variables</u> is being hypothesized. Many times you will see several variable relationships being hypothesized and then presented as a single hypothesis, such as:

People who attend class in the evening are <u>happier</u>, <u>healthier</u>, and <u>more optimistic</u> than people who attend class in the morning (we will call this the original longer hypothesis).

This hypothesis is actually three shorter hypotheses, namely:

Hypothesis 1: People who attend class in the evening are <u>happier</u> than people who attend class in the morning.

Hypothesis 2: People who attend class in the evening are <u>healthier</u> than people who attend class in the morning.

Hypothesis 3: People who attend class in the evening are <u>more optimistic</u> than people who attend class in the morning.

(we will call these the three shorter hypotheses)

One of the reasons the original longer hypothesis needs to be broken up into the three shorter hypotheses is because of the need to definitively state if a hypothesis is supported or not supported. If several relationships are presented in a single hypothesis, then some may be supported and some may not be supported by the data, which would

preclude our ability to state if the hypothesis is definitively supported or not. Using the original longer hypothesis as an example, the findings could have indicated that:

*Hypothesis 1: People who attend class in the evening **ARE** happier than people who attend class in the morning.*

*Hypothesis 2: People who attend class in the evening **ARE** healthier than people who attend class in the morning.*

*Hypothesis 3: People who attend class in the evening **ARE NOT** more optimistic than people who attend class in the morning.*

Subsequently, you can't definitively state if the original longer hypothesis was supported or not supported as parts of the hypothesis were supported, while other parts were not. However, if the longer original hypothesis was broken into the three shorter hypotheses you could definitively state if each was or was not supported using the same findings as such:

*Hypothesis 1: People who attend class in the evening **ARE** happier than people who attend class in the morning.*

Hypothesis 1 was supported by the data.

*Hypothesis 2: People who attend class in the evening **ARE** healthier than people who attend class in the morning.*

Hypothesis 2 was supported by the data.

*Hypothesis 3: People who attend class in the evening **ARE NOT** more optimistic than people who attend class in the morning.*

Hypothesis 3 was not supported by the data.

To recap, the first suggestion regarding making your hypothesis clear is to present *only* one relationship at a time.

Point 2: Making the cause and effect relationship clear. In sample study one we are hypothesizing a direct cause and effect relationship. Specifically, we are hypothesizing that *Class attendance* is directly impacted by the variable *Class time*. When you are presenting a hypothesis that addresses a direct relationship, it is often useful to see if it passes the *"If A then B"* test to be sure you are clearly identifying the cause and effect relationship. For example, regarding sample study one, our hypothesis would state:

***If** a person is assigned class in the evening, **then** he or she will attend a significantly higher number of days per week relative to those who attend classes in the morning.*

Of course the hypothesis does not need to be presented in this format. However, it is usually a good idea to check that the hypothesis *can* be phrased in this way as a means of examining if a clear cause and effect relationship is being presented.

Point 3: Specify the variables being examined. Perhaps the most important factor in making your hypothesis clear is to specify the variables being examined as clearly as possible. For example, the study hypothesis for sample study one would be much less clear if we were more general in stating the study variables as:

Class scheduling time will be significantly related to class attendance.

Notice, although it may seem like a clear hypothesis is being presented, the statement is so general that it does not describe an exact testable relationship. However, if we replace the term *class scheduling time* with the more specific *Being assigned to either a morning or evening meditation class* the relationship we are examining is much clearer:

Assignment to either a morning or evening meditation class will significantly impact the average number days per week a community resident attends class.

Thus, the clearer the terms we use to describe the variables being studied, the more powerful the hypothesis will be.

Must Know Core Concept 7:
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 *Level of Class Enjoyment* and *Class Attendance* would suggest as scores reflecting *Level of Class Enjoyment* increase, *Class Attendance* 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 *Level of Class Enjoyment* and *Class Attendance* would suggest as scores reflecting the *Level of Class Enjoyment* increase, *Class Attendance* 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 meditation *Class Attendance*, 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 *Class Attendance*.

For example, let's say you are a resident assigned to a meditation class. You have three acknowledge three factors that significantly influence your class attendance, which are:

1) Your **emotional well-being** for that day (do you feel you need class that day)

2) If the **weather** is good or bad (as you walk to class and enjoy the sun)

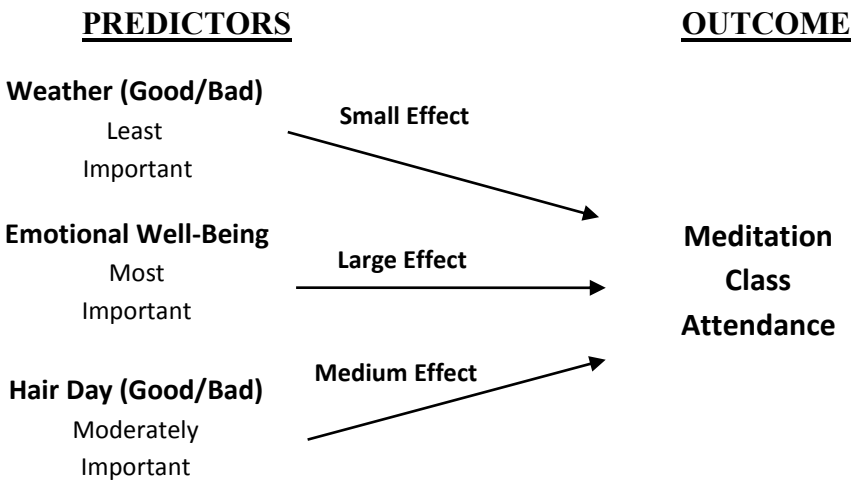
3) Weather you are having a good or bad **hair day** (you like to look presentable)

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 *Class Attendance*. 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 your class attendance. 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 your class attendance.

The real driver of your meditation class attendance is your emotional well-being that day. In other words, if you feel that you really need to sit down and meditate to serve your emotional state, you will! Thus, this is the *most important* factor in deciding if you will attend class, so it has a *large effect* on class attendance. **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 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 2**, 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.

For example, the raw numbers for an r^2 effect size describe that a small, medium, and large effect size would be indicated by the values 0.01, 0.09, and 0.25, respectively. However, the raw numbers reflecting a small, medium, and large effect for the eta squared effect size are 0.01, 0.06, and 0.14, respectively. Thus, keep in mind while the raw numbers of effect sizes may not be comparable, the categories of small, medium, and large effect sizes will be. As we conduct the statistical procedures listed in the column to the far right in this Table, we will review how to produce and interpret many of these effect sizes.

Table 2

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 8:
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.

Applying Step 1: Study Map to the Sample Study

As stated earlier, our sample study will involve explaining the factors that influence attendance at meditation classes among residents of a retirement community. 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. Before we create our study map. Let's consider the input of the administrators of the retirement community, as well as the goals they have for this study. Specifically, let's assume the **administrators noticed:**

1) The **evening classes are usually full**, so it seems assigned class time (morning/evening) is the big predictor of attendance (this is our predictor variable of interest, termed the **Independent Variable**)

2) Also suspect age, doctor prescribing meditation class, and level of class enjoyment are factors (these predictors may be related to our dependent variable, but are not the predictor variable of interest, so these are our **Covariate Variables**)

3) Considering moving **all class to the evening** to encourage attendance

4) Administrators want research findings to **support** or **dispute** this decision

Lastly, the administrators want two things explained. What predicts:

5) The number of days classes were attended per week. This is our **Dependent Variable**, measured as a continuous variable (measured as: 1-7), which we will call **Dependent Variable 1a**.

6) Attending class at least three days per week. This is also our **Dependent Variable**, but measured as a categorical variable (measured as: Yes/No), which we will call **Dependent Variable 1b**.

Study Map (for Dependent Variable 1a)

Study Map in Text

Research Question 1a: Is assigned class time the strongest predictor of the number of days classes were attended per week?

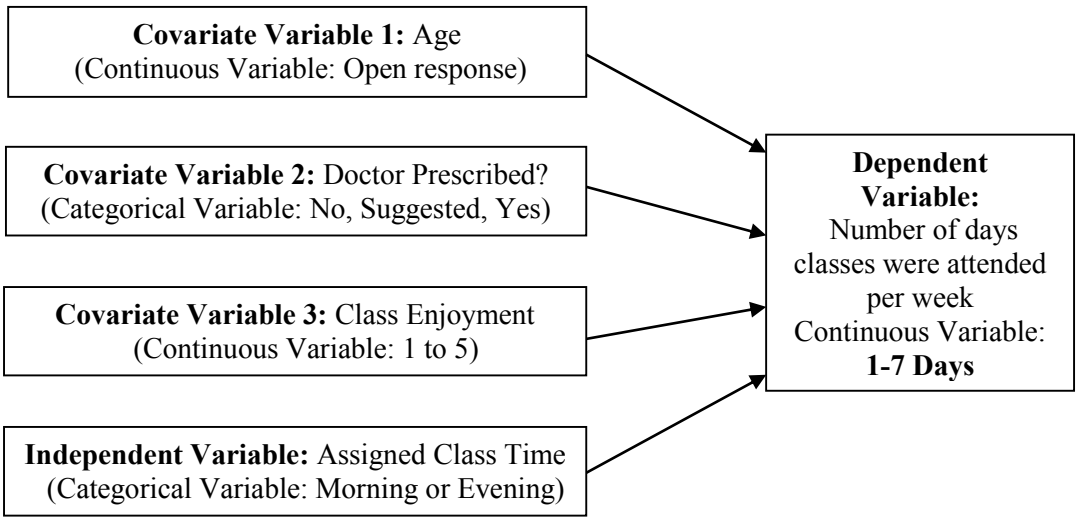
Hypothesis: Assigned class time is the strongest predictor of the number of days classes were attended per week.

Dependent Variable: Number of days classes were attended per week

Independent Variable: Assigned class time (morning/evening)

Covariate Variable: Age, Doctor Prescription Status, Class Enjoyment

Study Map in a Diagram



Study Map (for Dependent Variable 1b)

Study Map in Text

Research Question 1b: Is assigned class time the strongest predictor of attending at least 3 classes per week?

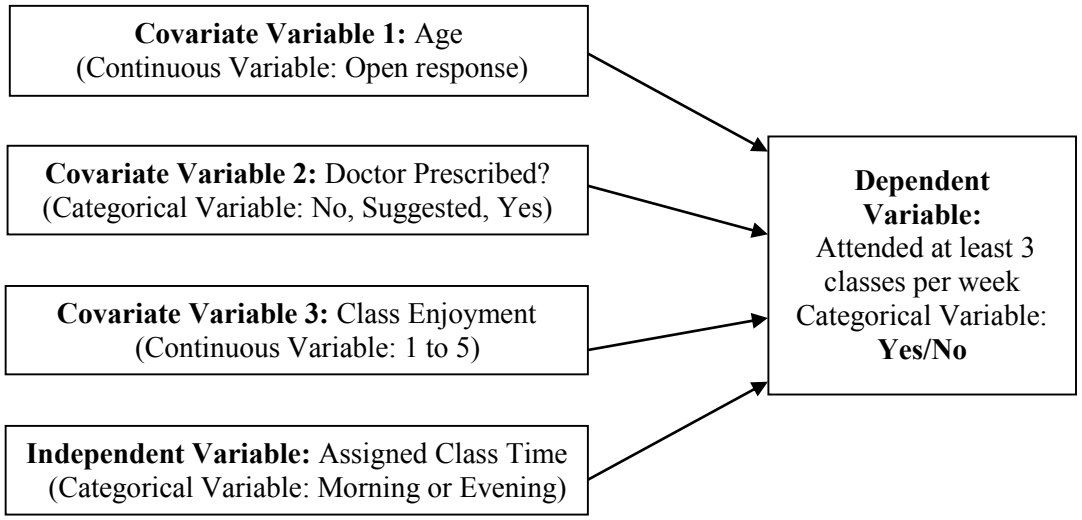
Hypothesis: Assigned class time is the strongest predictor of attending at least 3 classes per week.

Dependent Variable: Attending at least 3 classes per week

Independent Variable: Assigned Class time (morning/evening)

Covariate Variable: Age, Doctor Prescription Status, Class Enjoyment

Study Map in a Diagram



Our two study maps (1a and 1b) have outlined the data analysis plan that will be implemented in our quantitative study. You will soon see that having these two study maps, rather than just one, will be important to your leaning. Specifically, different types of analyses are applied when the dependent variable is continuous relative to when the dependent variable is categorical. Subsequently, we will use the continuous and categorical forms our dependent variable to illustrate how to conduct a quantitative analysis when you are faced with each scenario.