

Nothing Up My Sleeve: Unveiling the Magic of Statistics

*Essential Statistics for Graduate Students in Education*

Julie Alonzo

University of Oregon

<b>PREFACE</b> .....	<b>4</b>
<b>BEFORE YOU BEGIN</b> .....	<b>5</b>
<b>CHAPTER 1: A BASIC INTRODUCTION TO VOCABULARY AND IDEAS IN STATISTICS</b> .....	<b>6</b>
SYMBOLS.....	6
ASSIGNMENT #1 .....	7
<b>CHAPTER 2: WORKING WITH NUMBERS</b> .....	<b>8</b>
<i>INTRODUCTION</i> .....	8
<i>THE SUMMATION OPERATOR</i> .....	8
<i>WORKING WITH NUMBERS</i> .....	8
<i>MEASURES OF VARIATION</i> .....	10
<i>THE NORMAL CURVE</i> .....	11
ASSIGNMENT #2 .....	13
<b>CHAPTER 3: APPROACHING ANALYSES</b> .....	<b>14</b>
<i>MEASUREMENT SCALES</i> .....	14
<i>TYPES OF STATISTICAL ANALYSES</i> .....	15
ASSIGNMENT #3 .....	15
GUIDE TO BASIC STATISTICAL ANALYSES.....	16
<b>CHAPTER 4: CORRELATIONS</b> .....	<b>17</b>
BASIC DEFINITION .....	17
WHAT ASSUMPTIONS MUST BE MET IN ORDER TO RUN A CORRELATION ANALYSIS?.....	17
HOW DO YOU REPORT THE RESULTS OF A CORRELATION?.....	18
ASSIGNMENT # 4.....	19
<b>CHAPTER 5: REGRESSION</b> .....	<b>20</b>
<i>BASIC DEFINITION</i> .....	20
<i>WHAT ASSUMPTIONS MUST BE MET IN ORDER TO RUN A REGRESSION ANALYSIS?</i> .....	21
<i>YOUR OPTIONS</i> .....	22
<i>Linear</i> .....	22
<i>Multivariate Linear</i> .....	22
<i>Logistic and Multivariate Logistic</i> .....	22
<i>HOW DO YOU REPORT THE RESULTS OF A REGRESSION ANALYSIS?</i> .....	22
<i>Regression Summary for Grade 6 Statewide Reading Assessment</i> .....	23
ASSIGNMENT # 5.....	24
<b>CHAPTER 6: CHI-SQUARE</b> .....	<b>25</b>
<i>BASIC DEFINITION</i> .....	25
<i>WHEN WOULD YOU WANT TO RUN CHI-SQUARE ANALYSIS?</i> .....	25
<i>WHAT ASSUMPTIONS MUST BE MET IN ORDER TO RUN A CHI-SQUARE ANALYSIS?</i> .....	25
<i>HOW DO YOU REPORT THE RESULTS OF A CHI-SQUARE ANALYSIS?</i> .....	26
ASSIGNMENT #6 .....	26
<b>CHAPTER 7: T TESTS</b> .....	<b>27</b>
<i>BASIC DEFINITION</i> .....	27
<i>WHAT ASSUMPTIONS MUST BE MET IN ORDER TO RUN A T TEST?</i> .....	27
<i>YOUR OPTIONS</i> .....	27
<i>Independent Observations t Test</i> .....	27
<i>Dependent Observations t Test</i> .....	28
<i>One-tailed or Two-tailed t Tests</i> .....	28

<i>HOW DO YOU REPORT THE RESULTS OF A T-TEST?</i> .....	29
ASSIGNMENT #7 .....	30
<b>CHAPTER 8: ANALYSIS OF VARIANCE</b> .....	<b>31</b>
<i>BASIC DEFINITION</i> .....	31
<i>WHAT ASSUMPTIONS MUST BE MET IN ORDER TO RUN AN ANOVA?</i> .....	32
<i>YOUR OPTIONS</i> .....	32
<i>One-Way ANOVA</i> .....	32
<i>Multivariate ANOVA</i> .....	33
<i>HOW DO YOU REPORT THE RESULTS OF AN ANOVA?</i> .....	33
<i>Example 1</i> .....	33
<i>Example 2</i> .....	34
<i>Example 3</i> .....	34
<i>Descriptive Statistics for Grade 6 District ORF Test</i> .....	35
<i>Analysis of Variance Summary Table for Grade 6 District ORF Test</i> .....	36
ASSIGNMENT #8 .....	36
<b>REFERENCES</b> .....	<b>37</b>

## Preface

If you're reading this text, it is likely that you did not major in mathematics in college. In fact, if you DID major in math, you probably ought to seek a different path into the world of statistics, for you may find this introduction too non-mathematical for your liking.

If, on the other hand, you are one of the many folks out there who believes it is no coincidence that 'math' is a four letter word, then perhaps you are in the right place. This text was written with one purpose clearly in mind:

*To provide people with enough rudimentary knowledge of statistics to enable them to make appropriate decisions about the types of statistical analyses that might be appropriate for the research they read about or conduct.*

Specifically, it was written for a target audience of graduate students in education programs. Having learned the material in this text, students should be able to read and critique empirical journal articles. Equally as important, they also should be conversant enough in the common analyses used in educational research to be able to design, conduct, analyze, and report their own quantitative empirical study completed for their doctoral dissertation.

True statisticians and hard-core researchers might justifiably criticize this text for being too superficial. My use of *CliffsNotes* and the SPSS manual as guiding texts is unusual, to say the least. The decision to use these sources was not made lightly, however. This text is designed to give relative newcomers to the world of research a working knowledge of basic statistics as quickly and painlessly as possible. It is written with the notion that although the topics we research and the methods we use to conduct that research are extremely serious, a little humor might make the subject more palatable.

To that end, this book is dedicated to Nancy Heapes, an educational leader whose integrity and earnest desire to learn material she was sure would be a challenge compelled me to write a text that would actually teach her what she needed to know. I hope that others also benefit from the work.

## Before You Begin

The lessons in statistics through which I will be guiding you in this document will borrow heavily from two books:

Green, S. B., & Salkind, N. J., (2003). *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*, 3<sup>rd</sup> ed. Upper Saddle River, NJ: Pearson.

Voelker, D. H., Orton, P. Z., & Adams, S. V. (2001). *CliffsQuickReview™ Statistics*. New York, NY: Hungry Minds.

You should purchase these books prior to beginning your study of statistics, and I expect you will return to them often as you move into your own analyses in the future. You may find that you are left wanting additional guidance as you become more comfortable with the subject, ready—perhaps—to delve more deeply into the magic.

Additional recommended texts:

Kerlinger, F. N. & Lee, H. B. (2000). *Foundations of Behavioral Research*, 4<sup>th</sup> ed. Thomson Learning.

Gives really solid background on quantitative research design as well as covering a wide variety of statistical analyses.

Salkind, N. J. (2000). *Statistics for People Who [Think They] Hate Statistics*. Thousand Oaks, CA: Sage.

Written by the same author as the SPSS manual, this book gives common-sense explanations about many of the same topics covered in this document.

Salsburg, D. (2001). *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*. New York, NY: Henry Holt.

Provides a delightful history of some of the ‘greats’ of the statistical world and a good introduction for the ways in which statistics have shaped much of our modern world-view.

## Chapter 1: A Basic Introduction to Vocabulary and Ideas in Statistics

We use statistics (math magic) to make inferences (logical guesses) about populations (big groups) based on the responses of a sample (small group drawn from the big group).

These inferences are *more likely to be true* when we have met certain assumptions on which the psychometricians (math magicians) based their calculations. We're never *sure* that our inferences are true; we NEVER *prove* something with statistics, merely present evidence that supports or refutes our hypotheses. Because it's all math magic, we must use exact and proper terminology when talking about the analyses we're conducting (I liken this to the horrific effect adding frog's legs rather than hog's eggs might have on a potion to decrease one's leg hair; before you know it, you might be green and slimy).

For this reason, we'll start with some basic vocabulary. Learn the words as well as the symbols. As I used to tell my high school students, "Language is power; wield it like a master!"

Because we don't really care about the sample (we really care about the population from which it was drawn), you'll also see the funny notations used to refer to population statistics in articles and books [there are times when you can say, "I am confident enough in the results that I got from my sample that I can now make statements about the population." We'll talk more about this later.].

Usually, the symbols used to refer to population statistics come from the Greek alphabet while those used to describe the sample come from the Roman alphabet we're much more accustomed to seeing. (Perhaps this has something to do with the fact that populations, like letters from the Greek alphabet, are kinda hard to pin down... we have to work a lot harder to get to them).

### Symbols

Sample	What it's called	Population	What it's called
$n$	# in your sample	$N$	# in the population
$x$	Sample mean	$\mu$	Population mean
$s$	Sample standard deviation	$\sigma$	Population standard deviation
$s^2$	Sample variance	$\sigma^2$	Population variance

$\alpha$  = alpha This is a symbol that you use to indicate the % of times you are willing to risk being wrong in your inferences...trust me, you'll understand this soon!

$\eta^2$  = eta squared. This symbol represents the proportion of the variance in one variable that can be predicted from another variable.

Note: we seldom know much about the population (if we did, we wouldn't need to keep spending so much time and energy studying samples!)... all the same, our goal is to make reasonable statements about the population as a whole based on the degree to which our sample indicates that our data supports such statements. Because we are trying to generalize from a

small sample to a large population, it is always really important to describe the sample in as much detail as possible.

What sort of information do you normally use to describe the sample? Whatever info you have! Typically, you would at minimum report the sample size ( $n$ ), mean ( $\bar{x}$ ), and standard deviation (which is known as either 's' or "SD", depending on the text you're reading).

### **Assignment #1**

1. Read and understand these notes. Memorize the 9 symbols listed above.
2. Read Chapters 1 & 2 from *CliffsQuickReview™ Statistics*.
3. Read pgs. 15-21 again.

## Chapter 2: Working with Numbers

### *Introduction*

Now that you have a basic handle on some of the vocabulary used by statisticians and educational researchers like yourself (folks who differ from you, really, only in their knowledge of the stuff you're currently learning here about math magic), it's time to move into some additional knowledge you need to understand before you can begin to apply your new spells.

First, you need to learn one more symbol that intimidates a lot of folks just because it looks like it must mean something really profound. It doesn't.

### *The Summation Operator*

The symbol is known in English as 'the summation operator,' but you'll usually see it looking like  $\Sigma$ . Now, most folks' reaction when they see this jagged-toothed nasty-looking E is to regress to their worst memories of feeling dumb in math class and give up. Trust me, its bark is MUCH worse than its bite.

When you see  $\Sigma$ , all it means is that you're about to add a whole bunch of things together. It's actually your friend because it can be used to save you a whole bunch of typing.

For example,  $\Sigma x$  simply tells you to add up all the individual values ( $x$ ). Let's say that you're teaching a leadership class and you give a pretest to all 20 students in your class. They get the following scores: 15, 18, 11, 14, 15, 19, 9, 14, 11, 20, 9, 5, 12, 1, 16, 20, 12, 13, 18, 9. You decide to have your GTF figure out the class mean, so rather than write out the directions in great detail, you simply write out, "Dear Jessica, please  $\Sigma x$  and  $/ n$ ." And she, because you have trained her in basic statistics, gets out her trusty calculator and plugs in:  $15 + 18 + 11 + 14 + 15 + 19 + 9 + 14 + 11 + 20 + 9 + 5 + 12 + 1 + 16 + 20 + 12 + 13 + 18 + 9$  and then divides the total by 20 to get your class's mean score on the pretest.

The moral of the story is: Try not to hyperventilate when you see  $\Sigma$ . Just think, 'add a whole bunch of things up.'

### *Working with Numbers*

When statisticians talk about the results of their study, they pretty much always present you with basic information about the ways in which the sample performed on whatever measure they're using before they get into the more complicated analyses of whether or not the scores are statistically significant in any way. This basic information is generally referred to as *descriptive statistics*, and it includes both *measures of central tendency* and *measures of variation*.

Yes, yes, I know, fancy words again. But, as usual, the fancy terms don't really refer to anything too complicated.

*Measures of central tendency* is simply a fancy way of approaching where the ‘center’ of your scores lie. There are three types of measures of central tendency. *Mode* refers to the score that most frequently appears in your sample. Look back at the example from your leadership class pretest listed above. What would the mode be? <sup>1</sup>

One interesting thing about modes is that there may not always be just one of them in a given set of scores. If, for example, you had *two* scores that appeared the same number of times in your set of scores, and both appeared more than any other score, you would say that your sample is *bimodal*. In other words, it has two modes.

*Median* refers to the number that appears in the exact center of your set of scores if you line them all up in numerical order. It’s a little tedious, but do it here for the experience:

1. Arrange the 20 pretest scores above in ascending numerical order (I started the number line for you):

1 5 \_\_\_\_\_ 20 20

2. Now circle the number that is in the exact center.
3. Yes, I know, there are an even number of scores, so you just circled two numbers. Unlike a mode, you can only have one median. What do you think you’ll do?
4. If you answered, “average the two numbers in the exact center,” then you’d be completely correct. If that wasn’t your answer, then please learn it.

*Mean* is what is normally meant by ‘average.’ To find the mean, simply  $\sum x$  and  $/ n$ . Go ahead and do that for your pretest scores.

Compare the three measures of central tendency you got for your sample.

Mode = <sup>2</sup>  
 Median = <sup>3</sup>  
 Mean = <sup>4</sup>

There are a few very rare times when it will make more sense to use mode or median rather than mean for your research. Let’s think about it.

Say you’re administering a survey on leadership preferences to a bunch of administrators in training. If you want to come up with a description of “Jane Average Administrator” in the program, you might decide to take the modal response for each item on your survey and say that *that* provides the best information for your purpose.

---

<sup>1</sup> 9

<sup>2</sup> Still 9!

<sup>3</sup> 13.50

<sup>4</sup> 13.05

But maybe that's not how you want to use your data. Maybe instead you want to find out the degree to which the average administrator in training is more inclined to believe that managerial skills are more important to be an effective administrator than leadership skills. To find this out, you could use either the median or the mean. Which do you think would give you the better information if you have one person who is REALLY massively different from the rest of the group?

Do you have an answer yet? If not, think some more.

Which would it be? Why?<sup>5</sup>

### *Measures of Variation*

OK, so you are able to understand where the center of your distribution of scores lies, but that's not the only information that is really useful when trying to describe scores. Another basic concept is that *how* your scores *vary* is also quite informative.

If you're trying to describe a group, it's good to talk about what the 'average' person is like, but it's also important not to forget that not everyone is exactly in the middle. Sometimes, the variation within a group can tell you a great deal.

Say, for example, you are a professor applying for a tenure track job at a Research I university. You have a solid background in statistics and are looking forward to a challenging year working with students who will push your own knowledge of stats. You hear that the mean score on the final exam in the SPSS class last year was 92%, and you think you understand the level of knowledge students will bring to your class. Unfortunately, unless you also seek information about the variation in class performance, you might be sorely disappointed.

You might be expecting a class with students whose grades range from 88% to 95%, making a 92%. But what if the class was small, and students' abilities were widely scattered. The previous year's class happened to have one extremely capable student who scored 100% on the final AND earned an additional 25% worth of extra credit... (in stats-speak, this student would be known as a 'severe outlier') other students in the class ranged from 60% to 81%, but that one severe outlier would make it look as though the class performed at a much higher level overall than they actually did.

Beware the trap of ignoring measures of variation!

There are several types of measures of variation. The *range* refers to the difference between the highest and lowest scores in your sample. The range in the pretest example above is  $20 - 1 = 19$ . The *interquartile range* refers to the range of numbers that would be in the middle two blocks if you divided all your scores numerically into four quartiles (that's merely a fancy way of saying

---

<sup>5</sup> You would want to use the median because the mean would be heavily influenced by that one person who was so very different from the rest of the group.

‘fourths’). Thus, the interquartile range refers to the numbers between the 25<sup>th</sup> and 75<sup>th</sup> percentile in your sample.

The *deviation* refers to how far away a particular score is from the mean. You hardly ever hear this term used except as an intermediate step to other information about the larger sample, such as *variance* and *standard deviation*. *Variance* is a term that describes your sample, not an individual score. To figure out what the sample variance is, you (a) first figure out all the deviations of individual scores, (b) then you square each of those numbers, (c) add them all together, and then (d) divide by  $(n - 1)$ . Again, this is more an intermediate step to other information than truly useful itself. Why you do this particular set of steps is part of the mystery of the universe. Some things you don’t question, you merely accept.

NOTE: You should have a passing familiarity with these two terms—enough to recognize them when you meet them in the hallway, but don’t bother inviting them over for dinner; they don’t need to be that important in your life.

The variance, however, leads to the standard deviation, a very important number in statistics.

*Standard deviation* tells you how much, on average, the scores in your sample vary from the mean. It is THE measure of variation that you need to know. Mathematically speaking, the standard deviation is the square root of the variance. (So if your variance is 49, your standard deviation is 7. If your variance is 81, your standard deviation is 9). It is one of the basic pieces of information that you always report when you are describing how your sample performed on any measure you administered to them. Fortunately, SPSS calculates it for you automatically.

When you are comparing two different groups to determine the degree to which they are comparable, you generally want to check both the mean and standard deviation to be sure they are similar. (If the means are about the same, but one group has a much larger SD, then you can be pretty sure that the two groups are NOT comparable). Similarity in standard deviation is normally referred to as *homogeneity of variance*. Learn that term because you’ll see it in articles AND it’s one of the basic assumptions you’ll need to check before you run some of the fun statistical analyses you’ll be learning soon.

### *The Normal Curve*

Now we start to approach one of the coolest things about statistics, something so amazing that some argue it supports the existence of God. Honest.

Pretty much everything having to do with populations and the frequency with which certain things appear in populations follows what is called either a Gaussian distribution or normal curve (the terms are interchangeable). The normal curve has some rather useful properties.

1. Approximately 68% of all the measurements—on whatever you’re measuring!—occur within one standard deviation above or below the mean.

2. Approximately 95% of all the measurements occur within two standard deviations above or below the mean.
3. Approximately 99.7% of all the measurements occur within three standard deviations above or below the mean.

Figure 1 shows a drawing of a normal curve with the mean set to 0 and the SD set to 1. The numbers in the colored sections of the curve indicate the proportion of people in your sample in those areas of the distribution.

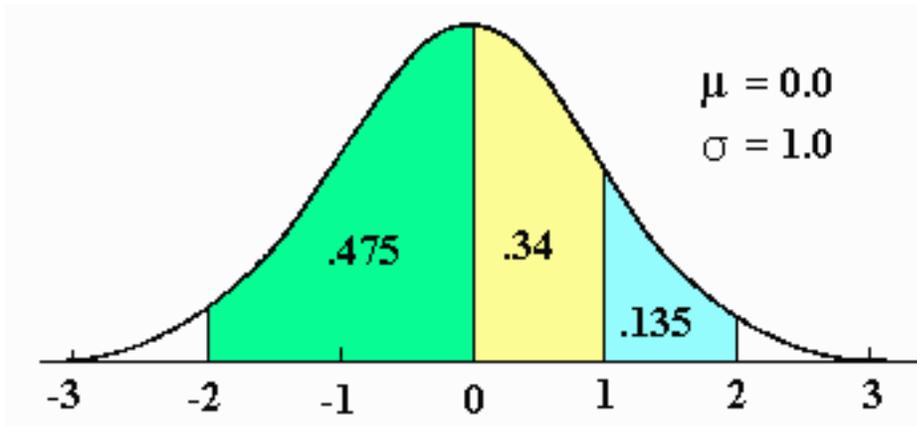


Figure 1: *The Normal Curve*

OK, so what does THAT tell us? Why do researchers get so gleeful over this fact?

It actually [get ready for this!] has practical implications that don't involve any further calculations to figure out! Did you hear that? Math-free useful stats!

Here's how it works. Say you want to find out whether you should be particularly worried – or impressed by—an individual's performance on a test (survey, 10K race, etc.). You collect scores on some dependent measure and calculate the mean and standard deviation simply because that's what you always do when you give information about descriptive statistics. Then you look at the scores that individual people received.

To figure out how an individual person compares with others in the population, you take the mean score from your sample and subtract and add the standard deviation to it. This will give you two numbers: one score that is one SD below the mean, and another that is one SD above the mean. 68% of all people in the population will have scores in this range. Do the same thing, but this time subtract and add the standard deviation to the values you got in the first step. This will give you a score that is two SDs below the mean and another that is two SDs above the mean. 95% of all people will have scores in this range.

Now when you look back at individual people's scores, you'll know how they compare to others in the population. If their score falls more than two SDs below the mean, for example, you'll know that they are in the bottom 2.5% of the population (you have to split the 5% that remain

from the 95% that are in the range of two SDs into 2.5% below two SDs and 2.5% above two SDs).

If this score was a measure of your risk factors for a heart attack, you would have reason to celebrate. If, however, it was a measure of your charm, it might indicate a need to learn some social skills.

Most of the statistical procedures I'll be teaching you in later chapters start with the basic assumption that your sample has a normal distribution. Because the normal curve is so prevalent in nature, a lack of normalcy in a distribution is cause to question the adequacy of the sample.

### **Assignment #2**

1. Read and understand these notes.
2. Answer all the questions that are embedded in these notes.
3. Read pgs. 23 – 33 and 50 - 60 from *CliffsQuickReview™ Statistics*.
4. Memorize the % of population under each part of the normal curve (be able to mark out up to 3 SDs from the mean and indicate what % of folks will lie in the various parts).

### Chapter 3: Approaching Analyses

Before I go over the different kinds of statistical procedures you can run, you need to be able to identify the four ways in which numbers can be used to represent reality. These four ways are also referred to as *measurement scales*.

#### *Measurement Scales*

*Nominal Scales* use numbers to indicate membership in a group, but the numbers don't really have any real numerical significance. For example, if on your survey you had people circle their ethnicity from six choices, and you coded the choices: 1 = Native American, 2 = Latino, 3 = Asian, 4 = White, 5 = African-American, 6 = Other, the numbers themselves don't actually mean anything. They *stand* for something rather than having their own intrinsic value. This is an example of using numbers in a nominal scale. Sometimes, folks refer to variables coded in this way as qualitative variables, because the numbers represent differences in the basic quality of the item under study rather than differences in quantity.

*Ordinal Scales* use numbers to indicate rank order, but you can not tell from the numbers whether or not each increase in number marks a similar increase in whatever it is you are measuring. For instance, you could develop a survey using an ordinal scale in which you ask people to respond to the following question:

How important is the ability to collaborate with others?

1 = not at all important, 2 = somewhat important, 3 = important, 4 = very important

You can't be sure that the difference between "not at all important" and "somewhat important" is the same numerical value as the difference between "somewhat important" and "important" even though in both cases, the numerical difference is 1. Ordinal scales indicate relative rankings rather than absolutes because they are always relative to the other items in the particular question rather than a true number with value of its own.

*Interval Scales* use equal intervals to divide responses into a continuum that makes precise numerical sense (but for which *zero* is an arbitrary number rather than an absolute lack of whatever you are measuring). For example, when reporting temperature, the difference between 35 degrees and 55 degrees is the same, numerically speaking, as the difference between 45 and 65 degrees. And, the zero doesn't mean that there is an absolute lack of temperature; it's just a number on the continuum.

*Ratio Scales* are the same thing as interval scales with one addition: Zero has a true meaning. For instance, the difference between 5 feet 6 inches and 5 feet 10 inches is exactly the same as the difference between 7 feet 6 inches and 7 feet 10 inches, and zero inches on this scale indicates a complete absence of length.

### *Types of Statistical Analyses*

Different statistical procedures are appropriate for use with different measurement scales. This is one reason it is important to be clear about what type of measurement scale you are using prior to running analyses. In SPSS, interval and ratio scales are grouped together as *interval scales*.

Once you know what type of measurement scale you are using, you are ready to decide what type of statistical procedure to run. The chart on the next page can be used as a quick reference to guide you in your selection. Don't worry: each of these analyses is explained in later chapters of this booklet.

### *Reporting Results*

Statisticians are sticklers for details. I guess that's not too surprising when you think about the precision with which they must operate in order to work the magic they do. Still, we lesser beings sometimes feel a little overwhelmed by the structure expected of us when reporting results. Each of the following chapters will introduce a type of statistical analysis, explain when you would use it, the assumptions your data must meet for its use to be appropriate, and the exact way in which you are expected to report the results of your analysis.

For now, suffice it to say that there are usually two ways to report your results: in text and in table. Generally, you use text when you have very little to report and tables when you have multiple results to report and saying the same thing over and over and over and over and over with words would be irritating—not to mention a waste of paper. You'll see examples of how to report results, tailored to the specific analysis being run, at the end of each chapter.

Writing up your results is NOT an area where creativity on your part will be appreciated. Instead, refer back to these pages, your APA or SPSS manual, or previously published research, and model your own presentation of the results on the standard way of doing it.

### **Assignment #3**

1. Read and understand these notes. Examine the chart below, but don't spend much time on it.
2. Memorize the different types of measurement scales.
3. Read pgs. 34 – 35 from *CliffsQuickReview™ Statistics*.
4. Read pgs. 24 – 39 of *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*.

**Guide to Basic Statistical Analyses**

Dependent Variable(s)	Independent Variable(s)				
	Ordinal, Interval, or Ratio		Nominal		Any Type
	1	2 or more	1	2 or more	2 or more
1 ordinal, interval, or ratio variable	<ul style="list-style-type: none"> <li>• Regression</li> <li>• Correlation</li> </ul>	<ul style="list-style-type: none"> <li>• Multiple Regression</li> </ul>	<ul style="list-style-type: none"> <li>• <i>t</i>-Test (either Independent Observations or Dependent Observations) if there are only 2 levels.</li> <li>• ANOVA if there are more than 2 levels.</li> </ul>	<ul style="list-style-type: none"> <li>• Two-Way (or more) ANOVA (Between Subjects, Within Subjects, or Mixed Effects)</li> </ul>	<ul style="list-style-type: none"> <li>• ANCOVA (Between Subjects Analysis of Covariance)</li> </ul>
2 or more ordinal, interval or ratio variables	<ul style="list-style-type: none"> <li>• Multivariate Regression</li> </ul>	<ul style="list-style-type: none"> <li>• Multivariate Multiple Regression</li> </ul>	<ul style="list-style-type: none"> <li>• Multivariate ANOVA (Between Subjects or Within Subjects)</li> </ul>	<ul style="list-style-type: none"> <li>• Multivariate Two-Way (or more) ANOVA (Between Subjects, Within Subjects, or Mixed Effects)</li> </ul>	<ul style="list-style-type: none"> <li>• Multivariate ANCOVA (Between Subjects)</li> </ul>
1 nominal variable	<ul style="list-style-type: none"> <li>• Logistic Regression</li> <li>• Linear Discriminant Function</li> <li>• <i>t</i>-test</li> <li>• ANOVA</li> </ul>	<ul style="list-style-type: none"> <li>• Logistic Regression</li> <li>• Linear Discriminant Function</li> <li>• Multivariate ANOVA (Between Subjects or Within Subjects)</li> </ul>	<ul style="list-style-type: none"> <li>• Chi-Square (can be either Independent Observations or Dependent Observations).</li> </ul>	<ul style="list-style-type: none"> <li>• Chi-Square (can be either Independent Observations or Dependent Observations).</li> </ul>	
2 or more nominal variables	<ul style="list-style-type: none"> <li>• Two-Way (or more) ANOVA (Between Subjects, Within Subjects, or Mixed Effects)</li> </ul>	<ul style="list-style-type: none"> <li>• Multivariate Two-Way (or more) ANOVA (Between Subjects, Within Subjects, or Mixed Effects)</li> </ul>	<ul style="list-style-type: none"> <li>• Chi-Square (can be either Independent Observations or Dependent Observations).</li> </ul>	<ul style="list-style-type: none"> <li>• Chi-Square (can be either Independent Observations or Dependent Observations).</li> </ul>	

## Chapter 4: Correlations

### Basic Definition

A correlation analyzes the degree to which two variables measured on interval or ratio scales are linearly related to one another. The most common correlation is actually referring to what is more properly called the Pearson product-moment correlation coefficient, and the statistical symbol for this statistic is a lower-case, italicized  $r$ .

When you test the statistical significance of  $r$ , you are testing to see if the relationship is strong enough in your sample that you can confidently assert that the relationship also exists in the population from which your sample was drawn. Although you can test the correlation between many, many variables at once using computers, keep in mind that what you are actually doing is comparing two variables at a time, regardless of how many such pair-wise comparisons you make.

If your variables have been measured on an *ordinal* scale but you still want to determine the degree to which they are linearly related, you would use SPSS or another statistical software package to calculate Kendall's tau-b rather than Pearson's  $r$ . (Gotta love these commonsensical stats names!) This distinction is something to keep in mind when critiquing research articles.

In addition,  $r$  can be squared (multiplied by itself) to indicate effect size—the proportion of variance of one variable that is accounted for by its linear relationship with the other variable. Thus, a correlation of .70 would indicate that  $.70^2 = .49 = 49\%$  of the variance on one variable can be predicted by the other.

### What Assumptions Must Be Met in Order to Run a Correlation Analysis?

Two assumptions must be met in order for you to trust the results of testing for a significant relationship using the Pearson correlation.

1. The variables must be bivariate normally distributed. This means that not only is each variable normally distributed on its own, it is also normally distributed at all levels of the other variable. This assumption is usually checked through a visual inspection of a scatterplot of the data. When you look at the scatterplot, you should not see a *widening or narrowing* of the points of data as you move from left to right. Check out pg. 108 in *CliffsNotes* for a fairly clear illustration of this concept.
2. The sample must be randomly drawn from the population and the scores for each case need to be independent of other cases. (A violation of this assumption might occur if, for example, one of your variables involved responses on a survey, and individual people completing the survey discussed the questions with each other before responding.)

If you are computing correlations between multiple variables, you need to adjust your alpha value accordingly (otherwise, with each new correlation that you run, you increase the risk of committing a Type I error and mistakenly reporting a relationship as 'significant' when in

actuality it was just a chance occurrence.) This is simple enough to do using the Bonferroni correction (yes, you guessed it, Bonferroni was the guy who developed this technique).

If you're using SPSS, you can simply tell the computer to use the Bonferroni correction. You should know, though, that what you are basically doing is dividing your pre-set alpha by the number of statistical tests you are going to run. For example, if you originally set your alpha at .05 and you want to analyze 10 correlations, you simply divide .05 by 10, yielding a new adjusted alpha of .005.

There are other approaches that achieve the same results using slight variations, but the main thing to keep in mind is that a researcher who is conducting multiple statistical analyses using the same data set should use *some* means of correcting for family-wise Type I error inflation.

### How Do You Report the Results of a Correlation?

Typically, you report the results of statistically significant correlational analyses in text and use a table to provide the reader with the full array of correlations (in the form of a correlation matrix). For example, If I were researching the variables that contribute to swimming speed, I might recruit 80 swimmers for my study and come up with an array of six variables to test: amount of time needed to swim one complete lap of the pool, quantity of body hair, % body fat, VO2 Max, length of arms, and length of legs. Once I ran the correlations, I could report my results like this:

Correlation coefficients were computed among the six variables in this study. Using the Bonferroni approach to control for Type I error, a  $p$  value of less than .008 (.05 / 6) was required for significance. The results of the correlational analyses (see Table 1) indicate that there is a significant strong positive correlation between amount of time needed to swim one complete lap of the pool and quantity of body hair,  $r(78) = .85, p < .001$ ; and a moderate positive correlation between amount of time needed to swim one complete lap of the pool and VO2 Max,  $r(78) = .48, p < .01$ . There was also a strong significant positive correlation between length of arms and length of legs,  $r(78) = .91, p < .001$ . All other correlations were not statistically significant.

Note: to interpret the  $r(78) = .85, p < .001$ , you should know that the  $r$  indicates that you are reporting a Pearson product moment correlation coefficient. The (78) indicates how many degrees of freedom the computer used in calculating significance (for simple bivariate correlations like this, the number in the parentheses will always be  $n - 2$ . Thus, in this case, the computer used 78 degrees of freedom in calculating significance because I had 80 swimmers in my sample). The number that comes after the equals sign indicates the actual computed value for the correlation. Finally, the  $p$  value is given, indicating how statistically significant the correlation is. Although the computer will tell you an actual  $p$  value, it is customary to report it in a more general sense: in relation to .05, .01, or .001, was the obtained  $p$  value  $<$  (less than your alpha level) or  $>$  (greater than your alpha level).

Table 1

*Correlations Among the Six Variables Analyzed in the Swim Speed Study (n = 80)*

	Speed	Hair	% Body Fat	VO2 Max	Arm Length
Hair	.85*				
% Body Fat	.22	.18			
VO2 Max	.48*	.14	.24		
Arms	.11	.08	.14	.15	
Legs	.11	.09	.13	.14	.91*

\* $p < .008$

#### Assignment # 4

1. Read pgs. 96 – 102 of *CliffsQuickReview™ Statistics*.
2. Read pgs. 237 – 355 of *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*.
3. Find a data set that you can use to practice using SPSS to run correlation analyses. What will your data set have to look like? Remember back to the start of this chapter, when I was explaining what a correlation really is? Hopefully, your answer is a resounding *yes*, and you'll guess [correctly] that in order to run a correlation, your data set will have to consist of a group of subjects for which you have data from two or more variables expressed as interval or ratio scales.
4. First, using SPSS, calculate the descriptive statistics ( $n$ ,  $M$ ,  $SD$ ) for your sample and present your results in a table, just as you would if you were reporting them in your dissertation.
5. Then, using SPSS, run a correlation analysis. Write up your results in a short paragraph and a table, just as you would if you were reporting them in your dissertation.

## Chapter 5: Regression

### *Basic Definition*

When you want to predict a score on one (ordinal, interval, or ratio) variable based on scores from one or more other (ordinal, interval, or ratio) variables, you can use regression to do so. Basically, what you are doing is using math magic to figure out where the *most likely* data point is for a given value, when you take into account the other information you have. Regression equations are not perfect—there is always some residual error that you should watch out for—but in many cases they do a decent job of prediction.

In regression analyses, it is common to discuss Independent Variables (IVs) and Dependent Variables (DVs). A Dependent Variable is the score you are trying to predict, while IVs are the data you are using to make that prediction.

One thing you should be aware of is that you are using data from your sample (*including* classifications based on the IV or IVs you are studying) to figure out how individuals in the population—for whom you have no data—might perform on your DV, given a particular set of IVs. To calculate a regression equation, you will need to have data on any IVs you want to include in the formula as well as on the DV you are trying to predict in the larger population. Statistics is strong magic, but it's not quite powerful enough to predict population performance without the initial data from your selected sample!

I use regression in my work with a local school district. They are interested in figuring out which of their students are likely to have trouble passing the state large-scale assessment in reading and which of their students are likely to pass with flying colors so they can allocate resources to those students who most need supplemental instruction. In my case, I use students' scores on three tests administered by the district (an oral reading fluency (ORF) measure, a vocabulary test, and a multiple choice reading test) to predict their performance on the state large-scale test of reading.

*Note:* Because I have gathered data from a previous sample in years past, I am now able to use the regression equation I obtained then to predict the current group of students' scores on the DV even though I do not have scores for them on that particular measure yet.

To figure out an individual's predicted score on the state test, I first use SPSS to run a regression analysis for all the scores in a given testing cycle. I then grab the values that the program provides and use them in an equation to calculate the student's predicted score. A regression equation always looks something like this:

DV (predicted score on the variable in which you are interested) =  $IV^{\#1} \times \text{a value} + IV^{\#2} \times \text{another value} + \text{error}$  (this is also known as the *residual*)

The math at this point is really quite simple: you just multiply the values for the Independent Variables you are using by the number SPSS tells you goes with those particular variables and then add the resulting figures plus the error value to come up with a final score.

*What Assumptions Must Be Met in Order to Run a Regression Analysis?*

There are 4 basic assumptions to be aware of when you are running regressions or evaluating others' research in which regression analyses are used.

1. The variables must be bivariate normally distributed. This means that not only is each variable normally distributed on its own, it is also normally distributed at all levels of the other variable. This assumption is usually checked through a visual inspection of a scatterplot of the data. When you look at the scatterplot, you should not see a *widening or narrowing* of the points of data as you move from left to right. Again, you can check out pg. 108 in *CliffsNotes* for a fairly clear illustration of this concept.
2. The variables should be independent of one another. In other words, one person's score should not influence another person's.
3. Outliers can *severely* skew the results of a regression equation. You should always check for outliers. Anytime regression is reported in an article or dissertation, the author should state whether or not there were any outliers in the data, and if there were, he/she should explain the steps taken to ensure that the resulting equation is, in fact, trustworthy. Typically, an outlier is considered any score that falls more than 2 standard deviations from the mean.
4. A mistake that beginning researchers might be prone to making is using the results of a regression analysis to predict beyond the scope of the data. Because the equation has been calculated using actual data points, the prediction is only reliable for scores that fall within the range of that data. Check to be sure that you (and others) are not predicting a score that falls outside your data range. If you are, then you should also state clearly that the values should be approached with caution.

For example, say you are doing a regression analysis in order to predict running speed (your DV) based on a combination of VO<sub>2</sub> max, body composition, and stride length (your 3 IVs). Your sample contains scores on all your variables of interest for 50 runners whose stride length varies from 24 to 48 inches. Running the regression will give you values to use to multiply each of your IVs by in order to predict the running speed of Jessie, somebody who was not in your original sample but whose VO<sub>2</sub> max, body composition, and stride length you know.

BUT you should be very wary of the number you come up with for predicted running speed if Jessie has a stride length of 67 inches. Because her score on the stride length IV falls out of the range of scores in the sample you used to develop your regression equation, there's a good chance that the equation will provide an inaccurate prediction of her running speed.

### *Your Options*

#### *Linear*

This is the most simple of the regression equations. A linear regression is simply the procedure you use to predict scores on one interval or ratio variable (considered your DV) from scores on another interval or ratio variable (considered your IV).

#### *Multivariate Linear*

A multivariate linear regression, as the name implies, is the procedure you use when you are working with more than one IV. In the example I gave above about how I use ORF, vocabulary, and multiple choice reading scores to predict how a student will perform on the state reading assessment, I used multivariate linear regression.

#### *Logistic and Multivariate Logistic*

The only difference, really, between linear and logistic regression is that logistic regression is used when the variables you are working with are ordinal rather than interval or ratio.

#### *How Do You Report the Results of a Regression Analysis?*

Because you are trying to predict score on the DV, it is fairly common to report your results in terms of what are called 95% Confidence Intervals. As the name implies, a 95% Confidence Interval provides a high and low score so you can easily determine the range of values within which you are 95% sure your predicted DV *will* fall.

Typically, when you run a regression analysis, you both report the equation that results *and* provide a table that shows the values that you'll use to multiply with each of the IVs, the error value (usually represented by  $\beta$ ), and the 95% Confidence Intervals.

As an example, on the next page I have provided an excerpt from an article some colleagues and I wrote for the American Educational Research Association annual conference, using the reading data I've already discussed (Alonzo, Duesbery, & Tindal, 2004).

*Regression Analysis of District Reading Assessments*

When combined, district ORF, Vocabulary, and Reading Tests provided a statistically significant prediction of student performance on the previous spring's statewide assessment in reading  $F(4, 164) = 38.57, p < .05$ . The district measures taken together accounted for 47% of the variability in state reading test performance. Table 14 presents the results of regression analyses.

Table 14

*Regression Summary for Grade 6 Statewide Reading Assessment*

Independent Variables	Unstandardized Coefficients		Standardized Coefficients		95% Confidence Interval for B	
	B	Std. Error	Beta	t	Lower Bound	Upper Bound
ORF	.001	0.02	.30	4.61	0.06	0.14
District Vocabulary	1.13	2.45	.03	0.46	-3.71	5.97
District Reading Test (SR)	32.45	4.72	.47	6.87	23.13	41.78
District Reading Test (CR)	1.48	2.56	.04	0.58	-3.56	6.53
Constant	189.09	3.32		56.99	182.54	195.64

**Assignment # 5**

- 1) Read pgs. 102 – 109 of *CliffsQuickReview™ Statistics*.
- 2) Read pgs. 257 – 379 of *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*.
- 3) Find a data set that you can use to practice using SPSS to run regression analyses. In order to run a regression, your data set will have to consist of a group of subjects for which you have data from two or more variables expressed as ordinal, interval, or ratio scales.
- 4) Decide which variable will be your DV (the score which you will be ‘predicting’) and which variable(s) will be your IV(s).
- 5) Now, using SPSS, calculate the descriptive statistics ( $n$ ,  $M$ ,  $SD$ ) for your sample on all of the IVs as well as the DV and present your results in a table, just as you would if you were reporting them in your dissertation.
- 6) Then, using SPSS, run a regression analysis. Write up your results in a short paragraph complete with an equation, and then prepare a table, just as you would if you were reporting these results in your dissertation.

## Chapter 6: Chi-Square

### *Basic Definition*

Unlike any of the other statistical procedures discussed in this booklet, Chi-Square tests (also written as  $\chi^2$ ) are *non-parametric*. This means that they are not based on sample/population parameters (mean, standard deviation, kurtosis, and skew). Instead, they simply compare the proportion of observed events to the proportion of events that one would predict would occur if the groups being compared did not, in fact, differ from each other.

### *When Would You Want to Run Chi-Square Analysis?*

Chi-Square analyses are used when the variables you are working with are nominal rather than ordinal, interval, or ratio.

For instance, if you were conducting a program evaluation of an administrator licensure program, you might want to compare the experiences of people of different racial/ethnic backgrounds on several different variables. You might want to investigate whether or not people from different racial/ethnic groups and different genders experience different rates of completion of the program, whether they have different rates of employment as administrators two to five years after they complete the program, and whether they rate the program as meaningful. Since all of these variables are nominal (they are black, white, latino, Native American, or other; typically, they will be female or male; they either completed the program or did not complete it; they are employed as administrators, or not, etc.), a Chi-Square analysis is the most appropriate choice.

In all of these cases, if there is no significant difference between the experiences of people from different racial/ethnic backgrounds in the program, you would expect to see equal proportions of them in each of the variables of interest. To the extent that the proportions of the groups vary from the expected “no difference,” you can assert that the experiences of people from different racial/ethnic backgrounds differ on the variables of interest.

### *What Assumptions Must Be Met in Order to Run a Chi-Square Analysis?*

1. Your sample must be random (this assumption can often be violated without much harm).
2. Each of your participants and their scores must be independent of each other (using scores from married couples, for instance, violates this assumption).
3. The sample size needs to be relatively large. At the very least, you should have 5 scores in each ‘cell’ that you are dividing your results into. (For this reason, I may need to ‘collapse’ one or more of the groups I have identified into the ‘other’ category. For instance, if only 3 Native Americans are participating in the administrator training program, then I would not have sufficient numbers to analyze their experiences as a separate sub-group. I could combine them with another group for which I have some logical reason to believe that they might be similar. Perhaps they have been reported in conjunction with the group in question

in other research I have read or share many of the traits of interest in this particular study. If that is not the case, I would most likely include the Native American subjects in the ‘other’ category for purposes of analysis.)

### *How Do You Report the Results of a Chi-Square Analysis?*

Typically, Chi-Square results are reported in text rather than table. For instance, I might report the results of one part of my program evaluation of the differential experiences of women and men in the administrator licensure program as follows:

A one-sample chi-square test was conducted to assess whether the rates of completion of the administrator licensure program differed for female vs male administrators in training. The results of the test were significant,  $\chi^2(1, N = 80) = 12.70, p < .01$ . The proportion of women who complete the program ( $P = .89$ ) was much greater than the proportion of men who complete the program ( $P = .34$ ). These results suggest that women in general complete the administrator licensure program at the University of Oregon at a higher rate than men.

### **Assignment #6**

1. Read pgs. 110 – 112 of *CliffsQuickReview™ Statistics*.
2. Read pgs. 341 – 349 of *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*.
3. Find a data set that you can use to practice using SPSS to run a Chi-Square analysis. In order to run a Chi-Square, you may need to re-categorize some of your variables as nominal rather than ordinal, interval, or ratio scale variables
  - a. For instance, if your data set consists of one nominal variable (sex) and a bunch of interval variables (such as scores on a survey of leadership vs managerial style preference), you may need to find a way to transform your interval variables into a nominal variable.
  - b. Perhaps it would make sense to add the total score on the “prefers managerial style” questions and compare it to the total score on the “prefers leadership style” questions. Folks who scored higher on the managerial preference could be coded as “0” while those who scored higher on the leadership preference could be coded as “1”.
  - c. You could then run a Chi-Square analysis to see if the proportion of men who preferred a managerial style differed significantly from the proportion of women with that preference.
4. Then, using SPSS, run a Chi-Square analysis. Write up your results in a short paragraph, just as you would if you were reporting these results in your dissertation.

## Chapter 7: *t* Tests

### *Basic Definition*

A *t* test is the most commonly used statistical procedure to test the significance of the difference between two sample means when the means represent scores from interval or ratio measurement scales. It is also sometimes referred to as Student's *t* test, after the man who developed the procedure.<sup>6</sup>

A statistically significant result on a *t* test indicates that there is sufficient evidence to suggest that a true difference between the groups on the variable being tested really does exist. For example, if you were attempting to determine whether Master's students in EDLD had significantly higher scores on the GRE than MBA students, you would enter all the scores from both groups into your computer program, ask the statistical wizard within to work its magic, and wait in anticipation for the resulting *p* value.

### *What Assumptions Must Be Met in Order to Run a *t* Test?*

For a *t* test to be appropriate for a given analysis, three assumptions must be met:

1. The scores being compared must come from interval or ratio measurement scales
2. Scores in the populations under study (or at least in the different groups in your sample) are normally distributed.

There IS a way around this assumption. When running your analyses, you can ask the computer to run a non-parametric analysis (such as the Mann-Whitney *U* Test) that uses the median rather than the mean to compare the groups.

3. There is homogeneity of variance (score variances for the populations / samples under study are roughly equal).

There IS a way around this assumption. When running your analyses, you can simply ask the computer to run Levene's Test for Equality of Variances at the same time. If it comes back significant, you simply report the *t* value that is marked "Equal variances not assumed" and go on with it. If it comes back non-significant, then you report the *t* value that is marked "Equal variances assumed."

### *Your Options*

#### *Independent Observations *t* Test*

This is the most common kind of *t* test. You should select an Independent Observations *t* Test when the grouping variable you have used to divide up your sample is NOT linked closely to the

---

<sup>6</sup> Read *A Lady Tasting Tea* for the story behind the name!

variable you are using as your DV. For example, if you have 40 lifeguards who are also qualitative researchers and you divide them into two groups (those who have read Miles and Huberman, 1994, and those who have not) and then compare the time it takes members of each group to swim 12 laps of the pool, you would not expect that their score on the DV (time it takes to swim 12 laps) would vary systematically with group membership. In this case, using an Independent Observations *t* test would be appropriate.

### *Dependent Observations t Test*

When your DV is linked closely to the rationale you used to divide your group, it is more appropriate to use a Dependent Observations *t* test. For example, if you have 40 lifeguards and you divide them into two groups (those who have successfully qualified for the Olympic swimming trials and those who have not) and then compare the time it takes members of each group to swim 12 laps of the pool, you would expect that their score on the DV (time it takes to swim 12 laps) would vary systematically with group membership. In this case, using an Independent Observations *t* test would be inappropriate. A Dependent or Correlated Observations *t* test would give a much fairer judgement about the true similarity/ difference between the groups.

### *One-tailed or Two-tailed t Tests*

Old-time statisticians must have gotten a bit giddy at times when they had been adding up thousands of scores by hand and chugging through miles of equations because they sure did come up with some cute names to describe rather mundane reality.

If you remember back to the Normal Curve discussion from Chapter 3, you may recollect that the infamous curve is tallest in the middle and tapers down to almost nothing at both ends. These tapered ends are what statisticians refer to as the ‘tails’ of the normal curve. If you have a strong reason to believe that one group’s score *will be higher* than the other group’s score AND you make this hypothesis *a priori* (a funny way of saying ‘in advance of looking at your data’), you can use a one-tailed *t* test. When you ask the computer to run a one-tailed *t* test, what you’re basically doing is asking it to throw all its effort into one side of the curve. This means, for instance, that rather than dividing that .05 alpha into two equal parts of .025 on each side of the curve, you’re clumping it all at one end. It gets a little complicated, but the main thing to remember is that a one-tailed test is more powerful (in other words, it is more likely to call a difference between the means *significant*).

It’s important to be brutally honest with yourself on this one. If you do not specify *a priori* that one group’s mean will be higher than the other group’s, you should *always* use a two-tailed test. This is something to consider when reading published studies as well. If the authors report the results of a one-tailed *t* test but don’t mention that they went into the study with a clearly directional hypothesis about the results, you might want to question the appropriateness of their statistical analysis.

*How Do You Report the Results of a t-Test?*

When you run a *t* test, you generally report the results in text rather than a table. Usually, it goes something like this:

An independent-samples *t* test was conducted to evaluate the hypothesis that lifeguards who had read Miles and Huberman, 1994, and lifeguards who had not read Miles and Huberman, 1994, would perform differently on a timed trial of how long it would take them to swim 12 laps of the pool. Alpha was set at .05. The test was significant,  $t(38) = 2.48, p = .02$ . Lifeguards who had read Miles and Huberman, 1994, on the average swam at a faster pace than those who had not read the book. Table 1 presents descriptive statistics for the groups.

Table 1

*Descriptive Statistics for the Qualitative Difference in Lifeguard Study, Comparing Time Needed to Swim 12 Laps of the Pool by Group Membership*

Group	<i>n</i>	<i>M</i>	<i>SD</i>
Read Miles & Huberman, 1994	20	58.08	6.79
Did not read Miles & Huberman, 1994	20	81.03	6.85

Remember that information about statisticians being a tad detail-oriented? Well, here's where you need to be on your toes!

Please note (and remember!) that the 't' in *t* test is always italicized. as is the 'p' when you are reporting a *p* value. Take a look at the number inside the parentheses. Now look at the number of folks in your two groups. Think for a moment about how many groups you have in this particular analysis.

When you report a *t* test, you start off with the letter 't', then follow it [with NO space] with (n-2). This number, by the way, is known as *degrees of freedom* in stats-speak. You'll see it coming into play in a variety of different ways with the different forms of analyses, but it always has something to do with the number of subjects in your sample and the number of different ways you are dividing those folks up.

As for the table, pay careful attention to how it looks. Items within the table are single spaced, but they have 6pt spacing both before and after them (you highlight the table, then go under 'Format—Paragraph—Spacing' and indicate your 6pt preference. There are no vertical lines anywhere in the table. Horizontal lines divide the title from the column headers, the column headers from the data itself, and the last line of data from the rest of your document. Items within each cell should be centered and decimal aligned, except for items in the first column on the left margin, which should be aligned to the left. The title should be italicized, but the word Table should not be. Ah, the lovely details!

**Assignment #7**

1. Read pgs. 79 – 89 of *CliffsQuickReview™ Statistics*.
2. Read pgs. 138 – 158 in *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*.
3. Find a data set that you can use to practice using SPSS to run a  $t$  test. In order to run a  $t$  test, your data set will have to consist of a group of subjects for which you have data from one nominal variable and one variable that is on an ordinal, interval, or ratio scale.
4. Now, using SPSS, calculate the descriptive statistics ( $n$ ,  $M$ ,  $SD$ ) for your sample comparing the performance on your DV (the ordinal, interval, or ratio scale variable) of the two groups (as determined by the nominal variable you are using). Present your  $t$  test results in text and your descriptive statistics in a table, just as you would if you were reporting them in your dissertation.

## Chapter 8: Analysis of Variance

### *Basic Definition*

Reminiscent of a  $t$  test, analysis of variance (ANOVA or AOV) evaluates whether the means of the groups being studied vary significantly from one another. If it accomplishes basically the same thing, why do we even need an ANOVA?

Well, besides their cool name, ANOVAs allow you to analyze data in more interesting ways than mere  $t$  tests. While they can be used to compare differences on the DV between two groups, ANOVAs can also be used to compare differences on the DV between *more than two* groups, or between two or more levels within a group. The way that ANOVAs work is through a complex process of comparing the within-group and between-group variances to determine whether or not folks *within* a group vary more than they do with folks in *other* groups.

Try to understand the reasoning behind this; it may make it easier to keep things straight when you look at ANOVA results later. Any time you measure a DV, there is going to be the ‘true score’ as well as ‘unexplained variance’. This variance that you can’t explain is typically described as ‘error variance.’ When you run an analysis of variance, what you are doing is trying to figure out the different forces at work. If you are comparing three groups, you know going into it that you will NOT find identical scores on your DV, simply because there are forces operating beyond your control. Maybe someone will be having a bad day, maybe their allergies or medication is affecting their performance, maybe your instrument isn’t as reliable as you might want... for whatever reason, there will be some ‘error’ in your measurement.

In order to know if the OBSERVED difference in scores represents a TRUE difference in whatever it is you are trying to measure, you need to figure out whether the variance is due to error or due to differences between the groups. This is the foundation of ANOVA. Variance due to error is referred to as within-group variance. That which is due to real actual difference is referred to as between-group variance.

Like the  $t$  test, ANOVA results are always reported with degrees of freedom. And, for some strange reason, an ANOVA yields an  $F$  statistic, which is sometimes referred to as an  $F$  ratio (because it is calculated with a within-groups/ between-groups variance equation).

Please notice that the “ $F$ ” is both capitalized and italicized.

Say, for example, you didn’t want to limit your study of lifeguard prowess to two groups; instead, you wanted to examine the degree to which quantity of body hair affects swimming speed. You obtain a sample of 90 really hairy lifeguards and randomly assign them to one of three groups: those who will retain all their hair, those who will have exactly one half of the hair on their entire bodies removed through selective electrolysis, and those who will have all of the hair on their entire bodies removed through whole-body electrolysis. Once the hair removal procedure has been completed, you have all individuals from all three groups swim 12 laps of the pool in a controlled environment, and you track the time it takes each person to complete the swim.

When you are done, you have a data file with three columns, set up something like this:

Subject	Group	Time
Lifeguard 1	whole hair	67.90
Lifeguard 2	whole hair	71.89
Lifeguard 3	no hair	56.87
Lifeguard 4	half hair	59.76 ... etc.

This is the perfect situation for an ANOVA because you want to be able to tell not only whether the whole hairs outperform the no hairs, but also whether they outperform the half hairs—and how the performance of half hairs and no hairs compares.

#### *What Assumptions Must Be Met in Order to Run an ANOVA?*

There are 4 assumptions that must be met for an ANOVA to be an appropriate analytic choice:

1. The DV must be normally distributed for the populations (samples) within each of the levels in your analysis.
  - b. Generally, if you have a large enough sample size (at least 15 subjects per group), it's fairly safe to violate this assumption, but you should always check it out.
2. There is homogeneity of variance (score variances for the populations / samples under study are roughly equal).
3. The subjects were randomly selected from the population (or randomly assigned to experimental conditions).
4. The scores on the DV are independent of one another. (what one subject scores on the DV does not affect what another subject scores on it).

#### *Your Options*

ANOVAs offer tremendous flexibility, but with this flexibility comes the need to be precise about what variables you are actually interested in.

#### *One-Way ANOVA*

A one-way ANOVA is used to determine whether there is a statistically significant difference between two or more groups on one dependent variable of interest. When you have only two groups, it is nearly identical to a *t* test. With more than two groups, an ANOVA allows you to compare each group to all the other groups in your study.

### *Multivariate ANOVA*

A multivariate ANOVA is also sometimes referred to as a factorial analysis of variance. If there are only two comparisons being made, it is also sometimes called a two-way ANOVA. Multivariate ANOVA is particularly well-suited to studying complex inter-related variables—the sort educators and psychologists work with all the time.

A multivariate ANOVA uses the same basic approach as a one-way ANOVA, but it takes it another step. Think back to your hairy lifeguard study. Say you ran the study and discovered that whole hairs significantly outperformed both half hairs and no hairs, and that there was no statistically significant difference between the performance of half hairs and no hairs. You're feeling pretty good about your results, until someone asks you whether the swimmers' sex made a difference in the findings. Because you were only comparing ONE IV, you are unable to answer this question.

All is not lost, however, because you remember that a two-way ANOVA allows you to compare not only the main effect that amount of hair has on swimming speed, but also the main effect that the subjects' sex has on swimming speed, and —most wonderful of all— the amount of hair/ sex interaction effect.

### *How Do You Report the Results of an ANOVA?*

ANOVA results are frequently presented in a combination of text and table. Example 1, below, presents fictionalized results from the first hairy lifeguard study. Example 2 presents fictionalized results from a multivariate ANOVA when additional variables were added to the study design.

Please note that both Example 1 and 2 present the results *only* in text and would thus be incomplete in an article or dissertation.

Finally, Example 3 presents actual results from a technical report I wrote analyzing a school district's performance on a reading test. Example 3 shows how you would work the tables into the mix by referring to them in the appropriate places in the text.

#### *Example 1*

A one-way ANOVA was conducted to evaluate the effect of three states of hair coverage on swimming speed. The independent variable, hair coverage, included three levels: whole hair, half hair, and no hair. The dependent variable was the amount of time in seconds it took subjects to swim 12 laps of the pool. The strength of relationship between hair coverage and swimming speed, as assessed by  $\eta^2$  was strong, with hair coverage accounting for 35% of the variance in swimming speed,  $F(2, 87) = 4.84, p = .02$ .

Follow-up tests were conducted to evaluate pairwise differences among the means. Post hoc comparisons of the three groups indicate that swimmers with all their body hair intact swam significantly faster than swimmers with only half their body hair or no body hair at all. There

was no statistically significant difference between the swimming speeds of swimmers with half their body hair and those with no body hair at all.

*Example 2*

A 3 x 2 (amount of body hair x sex) ANOVA was conducted to evaluate the effect of amount of body hair and sex on speed in swimming 12 laps. The results for the ANOVA indicated a significant main effect for amount of body hair [ $F(2, 87) = 18.81, p < .01$ , partial  $\eta^2 = .48$ ], a nonsignificant effect for sex [ $F(1, 87) = 1.02, p = .56$ , partial  $\eta^2 = .01$ ], and a significant interaction effect between amount of body hair and sex [ $F(2, 87) = 11.45, p < .01$ , partial  $\eta^2 = .29$ ]. Hairy women significantly outperformed all other groups.

*Example 3*

Analysis of Variance (AOV) of the four ORF forms reveals a significant difference between the forms  $F(3, 259) = 7.26, p < .05$ . Levene's test of homogeneity of variances was not significant, so post-hoc analysis was conducted using Bonferroni's procedure. This analysis indicated that students performed significantly better on ORF Form C than they did on ORF forms D and A. There was no significant difference in student performance on ORF Form B compared to the other forms, or on ORF Form A compared to D.

Analysis of student performance by group (see Table 1) revealed that the only group of students who performed significantly differently on the ORF test were those designated as receiving Special Education services [ $F(1, 223) = 18.19, p < .05$ ]. No statistically significant differences were found between student performance on the ORF when blocked by gender, ethnicity, or ELL status (see Table 2).

Table 1

*Descriptive Statistics for Grade 6 District ORF Test*

	Group	<i>n</i>	<i>M</i>	<i>SD</i>
Gender	Male	131	143.36	36.03
	Female	132	145.03	39.12
Ethnicity	White	179	146.55	36.30
	Hispanic	17	136.47	38.12
	African American	5	128.20	46.24
	Asian	8	155.38	16.28
	Native American	5	127.00	64.61
	Other	11	155.91	41.55
	SPED		19	112.21
ELL		5	120.00	38.72
Total		263	144.20	37.55

Table 2

*Analysis of Variance Summary Table for Grade 6 District ORF Test*

Source	<i>df</i>	<i>F</i>	$\eta^2$	<i>p</i>
Gender	1	0.13	.00	.72
Error	261	(1414.61)		
Ethnicity	5	0.98	.02	.43
Error	219	(1379.53)		
SPED	1	18.19**	.08	.00
Error	223	(1280.59)		
ELL	1	2.47	.01	.12
Error	224	(1369.89)		

*Note.* Values enclosed in parentheses represent mean square errors.

\* $p < .05$ , \*\* $p < .01$ .

### Assignment #8

1. Read pgs. 160 – 190 in *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*.
2. Find a data set that you can use to practice using SPSS to run an ANOVA. In order to run an ANOVA, your data set will have to consist of a group of subjects for which you have data from one nominal variable and one or more variables on an ordinal, interval, or ratio scale.
3. Now, using SPSS, calculate the descriptive statistics ( $n$ ,  $M$ ,  $SD$ ) for your sample comparing the performance on your DV(s) of the two or more groups (as determined by the nominal variable(s) you are using). Present your ANOVA results in text and a table, just as you would if you were reporting them in your dissertation.

## References

- Green, S. B., & Salkind, N. J., (2003). *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data, 3<sup>rd</sup> ed.* Upper Saddle River, NJ: Pearson.
- Voelker, D. H., Orton, P. Z., & Adams, S. V. (2001). *CliffsQuickReview™ Statistics*. New York, NY: Hungry Minds.