Statistics

An Outline for Beginning Researchers

Wayne Briner, PhD

Bad Wolf Laboratories

Table of Contents

Statistics, What They Are, What They Are Not

Nearly every student in the natural or social sciences encounters and needs to understand statistics. Yet few of us are actually mathematicians, and statistics can take on a near-magical quality. They occupy that unique place for most of us where a few numbers add credibility to our study and approval from our colleagues and advisors. Yet, in reality, they evade our complete understanding. Here we will outline some of the fundamentals of statistical analysis.

First and foremost, statistics allow us to objectively determine if there is a pattern in the data that we have. This is very useful when you have a large data set and the sheer size of the data set is beyond easy comprehension. They also remove simple opinion when looking at the data we have gathered. For example, did our new teaching method really work? As a researcher you are probably very invested in the study you just conducted. You *want* things to work. Statistics allow us to state, in an objective fashion that yes, it did, or no, it did not.

Secondly, but connected to the first idea, no data set comes out looking neat and tidy. The numbers tend to spread out for all of the things you have measured. How do we know that any differences, or relationships, that we see are "real"? That is where the idea of statistical significance comes into play. All statistical significance means is "I am 95% certain that what we have here is not simply random". It does not mean your difference is big, nor does it mean it is important. Those questions are related more to your research question than they are to statistics.

The Types of Statistics and What They Tell Us

Statistics come in four basic flavors:

1. Descriptive Statistics: These do just as the name implies, tell us what a group of data looks like. The most common is the mean (average), that we are all taught in the 6th grade. Others include the median and the mode, and they have their uses. The mean, median and mode all tell us where the "center" of the data is at. Typically these are fairly close to each other, other times they are not and can represent a problem with the data, which can be resolved in a number of ways. Other descriptive stats include those that tell us how much spread (variance) there is to the data. These include the range (distance from the highest to lowest score) and the standard deviation (the typical spread of the scores). Nearly all data presentations include the mean and standard deviation (ex: 100 $\pm$ 15, the mean and standard deviation for IQ). Other, more sophisticated descriptives include the standard error of the mean and confidence intervals, which describe not the sample we have studied, but about the population that the sample comes from.

2. Differences Between Groups: Many studies try to determine if two or more groups differ from each other on some type of measure. For example, does a new team building approach improve productivity when compared to the typical method? There are any number of methods to do this. The simplest method that many people are familiar with is the t-test. It will compare the means of two groups and, taking the spread of the scores into account, determine if the means have enough distance between them to be due to actual differences between the groups and not just from some random effect. If your design includes more than one group it is then typical to use an ANOVA (Analysis of Variance), which can look at several groups at once. ANOVAs can become very sophisticated. If you are looking at several groups and looking at them in terms of sub-groups (ex; boys and girls in three different classrooms) you can use a factorial ANOVA. If you are looking at several performance measures in those classrooms you can use a MANOVA (multivariate ANOVA). There are additional ways of using ANOVA when you want to measure from the same group several times (factorial ANOVA) and take some measures into account while comparing on different ones (analysis of covariance, ANCOVA). You can also combine these in a number of ways. ANOVA, and all of its types, are extremely powerful tools. However, they come with a catch; they can be difficult to interpret. Careful research design and a very clear hypothesis will make the interpretation much easier.

3. Relationships Between Variables: This group of statistics asks the basic question: when one thing changes, does another change with it? For example, as hours in the classroom increase does test performance also increase? These are correlational studies and very common in all fields. The simplest statistics is the Pearson's correlation and gives you a number between 0 and 1 that tells you how strong the relationship is and, if significant, that the relationship is "real" (not random or due to chance). You can also put a line through the data (linear regression) to help illustrate the relationship. One word of caution, however: correlation is *not* causation. That two variables move together does not mean that one *causes* the other to move, other issues you have not looked at may be at play. Take the example of increasing classroom hours. Perhaps scores are increased not because of sheer hours in the class, but that the teacher has been freed up from other responsibilities and can give students more attention. Other correlational methods include factor analysis (FA) which determines how several measures relate to each other and which ones are the strongest at predicting some type of outcome (which measures are most important to your question). FA is extremely popular but there are a couple of catches to FA. Most importantly, you need a large sample size. Many statisticians would say that if you do not have at least 150 subjects with *all* the measures present (no missing data) you should not attempt FA. The other problem is that you need to make sure your measures do not overlap too closely. For example, if you are interested in weight loss and measure weight with 15 different scales they are all telling you essentially the same thing and FA will give you spurious results.

4. Determining What Are Important Contributors: Related to correlational studies and factor analysis are a variety of other methods that look at predictors of outcome. These are, in many ways, derivatives of factor analysis. These tests include principle component analysis, structural equation modeling, logistic regression, discriminant analysis, survival analysis and others. Use of these methods should be determined ahead of time, and not after the study has been designed and data collected. They can be difficult to interpret and can be misleading, especially if the sample sizes are small, or the measures not well thought out.

Some Things to be Mindful Of

There are three things that researches need to keep in mind when looking at statistical results.

1. **What does "*p*" mean?** The thing we are all looking for in statistics is statistical significance, typically $p < .05$. Many computer programs will give you an output with exact probability (ex; $p = .003$). As long as it is less than .05, it is significant, but the cutoff is $p < .05$. What this means is that we are 95% certain that what we found is not due to some random effect from the way data likes to spread around. That is *all* it means. A $p = .001$ is not better than $p = .04$. It only indicates that there is greater certainty that the results are not random, but this can be due to a number of things. Things that go into a smaller (better) *p* value include the nature of the sample, the nature of the measure you take, and the size of the sample. Note that it has no bearing on how "important" the finding is, that is determined by your research question, not the statistics. For example, studying a new contraceptive in nuns would be "highly" significant, but not important. However, a $p = .049$ for cancer treatment would get everyone's attention.

2. **Effect Size:** If statistics has a measure of importance, it would be effect size. Effect size gives us an idea of how far apart groups are, or how strong relationships are. It does play into statistical significance in that larger effect sizes are more often statistically significant. There are many measures of effect size and they all have advantages and disadvantages. One that is commonly used (and calculated by SPSS) is "partial eta-squared". Partial eta-squared ranges from 0-1, with a larger number indicating a stronger effect. For correlational research r^2 or R^2 is common and follows the same 'larger is stronger' rule.

Effect sizes can be tricky to interpret. Certainly a large effect size is easy to make sense of. However, small effect sizes can also be very important when considering large groups, such as populations. A small effect size (say .05) for an improvement in cardiovascular fitness in the population of the US would mean a massive improvement in terms of sheer numbers. Importance is, again, derived from the research question and design; the statistics are secondary in that regard.

3. **Power:** Everybody wants more power; that includes statisticians. Simply put, power is the ability to find a relationship, or differences between groups. Keep in mind that there must be a relationship or difference to find. If there is nothing to find, it doesn't matter how much power you have, you will not find anything. Power is derived principally from your sample size, with larger samples giving you more power. Keep in mind that subtle differences or relationships will require more power to find. In other words, small effect sizes need larger samples to separate out. Other things go into how large a sample (or, how much power) you will need. Those other things largely amount to "noise"; other factors that play into what you are interested in, or how cleanly you can measure what you are after.

Power can be used to calculate the sample size you will need before you run a study. There are several programs out there that can help you do this; search for G-Power on Google. Knowing how many subjects you need before you run a study can save you a fair amount of grief and frustration later on. Many graduate colleges will require you to calculate anticipated sample sizes before they will approve your study.

Many statistics programs (such as SPSS) will give you an “observed power” calculation after you run some stats. Beware of these calculations, they often amount to the equivalent of statistical circular arguments with statistically significant findings having a lot of power simply because there was a difference found.

Important Aspect of Research Design

Statistics is just a tool. Much like a hammer is needed to build a house; stats can put your research together for you. However, the quality of the house is dependent on the design of the house and the quality of the materials, not the hammer. Similarly, the outcome of your research is dependent on your research design, not the statistics you use. Below, are outlined some key ideas to keep in mind that will make your research stronger and easier:

1. Have a clear, unambiguous hypothesis/question. The simpler, the better. “Adding an additional hour to school each day will improve test scores by 5%” is better than “Additional classroom time of 30-90 minutes at least once a week will improve at least one test score for at least one subgroup of children that come from specific neighborhoods”.

2. Make sure you have clear, independent and dependent variables. They should follow your hypothesis. For example, if your hypothesis is linked to leadership style, do not make the corporation your independent variable and hours worked your dependent variable.

3. Make sure the measurements you take (the instruments you use) actually measure what you are after. If you are interested in leadership style, is that best measured by surveying the leaders, or the group members?

4. Avoid missing data points. Big gaps in data add a lot of problems, and not just for the statistical analysis. Gaps in data call into question the instruments you use to collect data, as well as the nature of the group you were sampling.

5. Be sure you have an adequate sample size. Be sure the sample meshes well with your hypothesis. If you are measuring classroom involvement you probably want to study teachers, not administrators.

6. Do you need a control group? Depending on what your design looks like, you may need some type of control or reference group. If you want to study the effects of team building, you probably need some type of group that does not get a team building session, or at least gets some type of dummy activity. You will need this to bring out subtle effects. In some cases, before and after measures will work.

7. Include demographic data. Telling the reader about your group helps round out the story. How many, how old, how well educated, gender, race, income, location? Which bits of demographic data you collect will depend on the type of study you are conducting. It allows the reader to picture the sample in your study and decide what groups in the real world it pertains to. It also shows that you are thorough and diligent.

8. When gathering data keep your data grouped by individual, if possible. In other words, you need to be able to track down that subject #20 had certain responses and scores. This is absolutely required for some types of analysis, in particular if you are going to examine the relationships between variables, as in correlations and factor analysis.

A STATISTICS DECISION TREE

Which statistics should I use? This is the question every researcher asks, no matter how experienced they are. The answer depends, of course, on what you want to know. It also depends on where you are at in your analysis. We typically recommend a layered approach for analyzing data where you start with the simplest question your research needs to answer, and then move to other analyses in an effort to explain your findings. That being the case, you may use several statistical approaches before you are done. Having said that, below is a step by step question approach for aiding the decision making process.

A. Are you looking for differences between groups? If yes then use:
- One-Way ANOVA (t-test if only two groups)
- If you measure the outcome more than once: Within subjects ANOVA
- If you have more than one independent variable: Factorial ANOVA
- If you have more than one IV and measure more than once: Mixed ANOVA
- If you are looking at more than one dependent variable: MANOVA

B. Are you looking for a relationship between two variables (typically dependent variables)? If yes, then use:
- Person's correlation
- Are there several variables where you want to determine the relationship?
  - Then use Factor Analysis/Principle Component Analysis/Structural Equation Modeling

C. Do you want to predict an outcome? If yes, then use:
- Survival Analysis/Time-series Analysis/Logistic Regression

Please note that this is not an exhaustive list and subtleties in your design may preclude the use of some approaches. For a detailed discussion of the approaches see: *Using Multivariate Statistics* by Tabachnick & Fidell, now in its 6^{th} edition.

PRESENTATION OF RESULTS

Once you have all that data and performed your analysis, you then need to tell the reader about it. It is typically better to tell a story and build your argument rather than jump to the big finale. Generally, you will want to start with some descriptive statistics directed at the demographic characteristics of your sample. Tables are often used here giving the means and standard deviations and other important aspects of your group (how many men/women, ages, income, education, etc. (see Table 1 for an example). For something important you want to really emphasize, and sometimes just to make the paper more interesting and readable, a histogram bar graph is good to use (Figure 1). You will use the means as the reference point. It is typically a good idea to include error bars (the standard deviation). Error bars can both emphasize that differences you see are “real” or explain why what seems to be a difference is not actually there (see Figure 1).

TABLE 1. DESCRIPTIVE STATISTICS OF TEST SCORES FOR FOUR HYPOTHETICAL CLASSROOMS

Classroom	*n*	*Mean*	*SD*	*SEM*	*95% CI*
A	10	13.80	4.78	1.51	±3.41
B	10	18.80	6.48	2.05	±4.63
C	10	24.40	1.51	.48	±1.08
D	10	21.05	1.99	.63	±1.64

FIGURE 1. MEAN TEST SCORE FOR FOUR HYPOTHETICAL CLASSROOMS.

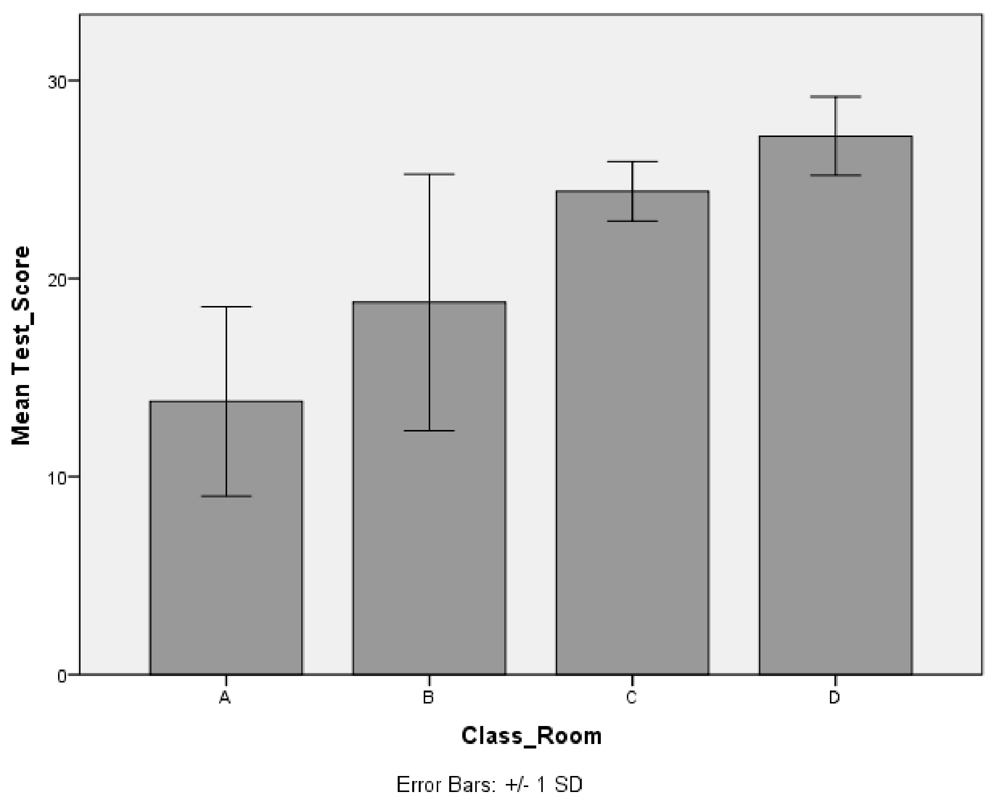

FIGURE 1. Note that classrooms A & B do not differ from each other statistically (p > .05). This is seen by inspecting the error bars which indicate a large amount of overlap between the groups. Both classrooms C & D differ from A & B ($p < .05$) and this can be seen by way of the much smaller error bars with little overlap.

After descriptive statistics, you will want to make your broadest point. Did your weight loss program actually produce a general weight loss? As study time increased, did grades also increase? Do employees notice differences in leadership style? Once you have made your broad points, then go to a finer level of analysis. Did your weight loss program effect one group more than another? Did study time increase only for certain students? Did differences in leadership style improve productivity? Typically, you continue to take apart your data layer by layer until you have thoroughly explored your question.

If you have a larger data set it is easy to go running off in all sorts of directions, exploring all sorts of things. Stay focused on your hypothesis, don't cloud the issue. Answer all the questions that pertain to it before heading off in other directions. Include other interesting side issues only if space permits and the side issue might lead to the next study that you will want to do.

Presentation of your results in graph form can add considerable clarity to your results and discussion. Figure 1, above, provides not only good descriptive data but explains why groups do and do not differ from each other. Often patterns emerge by graphing data that would not otherwise be seen. When presenting descriptive data, or demonstrating the differences between groups, bar graphs are typically used, although other formats (e.g., pie charts) may be acceptable.

When presenting the relationships between variables, such as correlations, a scattergram is often used. Each pair of data points is plotted on a grid and the pattern of the relationship can be seen (FIGURE 2).

FIGURE 2. RELATIONSHIP BETWEEN HYPOTHETICAL TEST SCORES AND STUDY HOURS FOR FOUR CLASSROOMS.

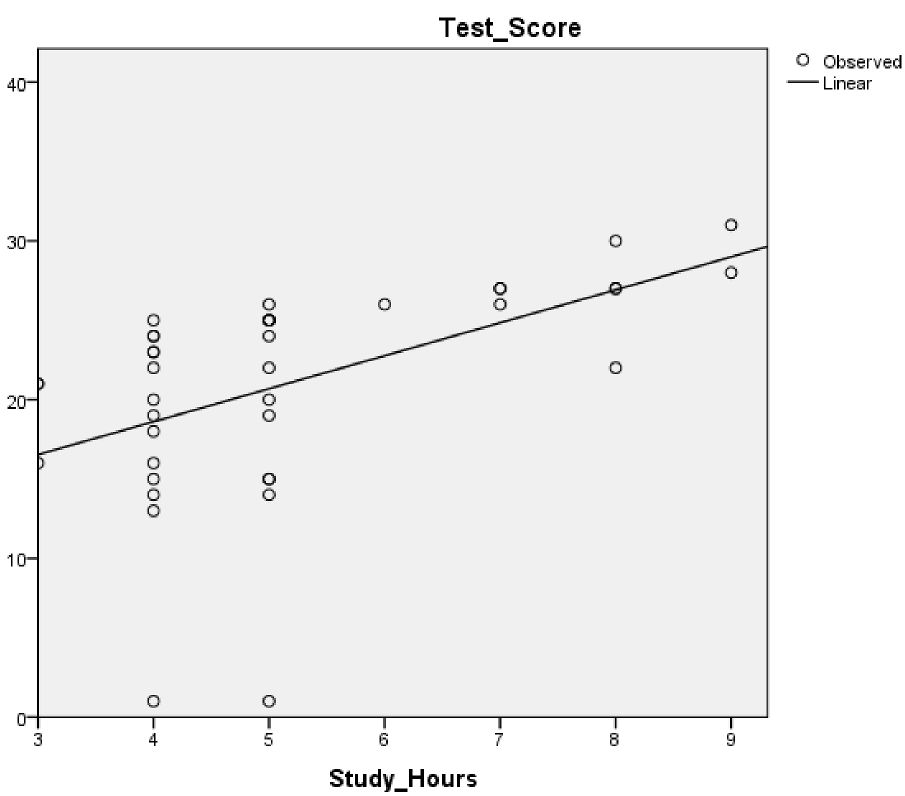

FIGURE 2. You can see the individual data points on the plot. This relationship is significant r (38) = .51, p < .05 and so a linear regression line has been included to emphasize the nature of the relationship. If the correlation is not significant a regression line is typically not included, if the data is presented at all.

When using graphs be sure they are well labeled across the top and have captions describing what the reader is seeing. This allows the reader to quickly grasp what is going on in the graph, and shows that you have been careful, thorough, and diligent.

Similarly, good writing, with a logical flow can make all the difference. Proofreading is an absolute must; grammatical errors and clumsy writing diminish the overall impact of your research. Be sure to follow the writing style your institution or journal requires very carefully.

IN CONCLUSION

Statistics are a powerful tool that can add considerable depth to your research. In many instances, good statistical analysis can “make” your research, but, you need to start with a solid research question and design, followed by careful data collection. The statistical approach you use will add the strength and depth you are looking for.

STATISTICS

AN OUTLINE FOR BEGINNING RESEARCHERS

IS PROVIDED AS A FREE SERVICE BY

BAD WOLF LABS

PUTTING TEETH INTO YOUR RESEARCH

WAYNE BRINER, PHD

DIRECTOR

TEAL BRINER

EDITING AND LAYOUT

Bad Wolf Laboratories provides statistical and research consulting for academics and industry.

Please visit us at *BadWolfLabs.com*, or email me at *wbriner@BadWolfLabs.com* or *info@BadWolfLabs.com* for your research and statistical needs.

© 2013, Wayne Briner.

Revision 1.

This booklet may be distributed freely in electronic or print form, provided it is *not altered in any way*.