



AP[®] Psychology:
**Teaching Statistics
and Research Methodology**

2008
Curriculum Module

AP[®] Psychology Curriculum Module: Teaching Statistics and Research Methodology

Table of Contents

Editor's Introduction	1
<i>Chris Hakala</i> <i>Western New England College</i> <i>Springfield, MA</i>	
The Transition from Descriptive Statistics to Inferential Statistics	2
<i>Elliott Hammer</i> <i>Xavier University of Louisiana</i> <i>New Orleans, LA</i>	
A Lesson on Correlation	6
<i>Amy C. Fineburg</i> <i>Spain Park High School</i> <i>Hoover, AL</i>	
Teaching Statistics Concepts	11
<i>Jonathan Stadler</i> <i>Fisk University</i> <i>Nashville, TN</i>	
Contributors	18

Teaching Statistics and Research Methodology

Editor's Introduction

Chris Hakala
Western New England College
Springfield, Massachusetts

The resources that follow provide both content and teaching tips designed to assist teachers with statistics and research methodology instruction. They have been developed to assist with material that has traditionally been considered difficult to present to high school students. Too often, the goals of using statistics and methodology are lost on the introductory psychology student. Many students operate under the assumption that the theories they are studying appeared out of thin air.

Understanding the concepts of statistics and methodology helps students recognize the fact that psychologists do research, discover things from that research, and are sometimes incorrect in their assumptions. It also helps students understand the critical thinking process within the context of human behavior. It is my hope that by delivering these topics to students in a unified conceptual framework you will be more successful in your AP Psychology teaching.

The Transition from Descriptive Statistics to Inferential Statistics

Elliott Hammer

Xavier University of Louisiana

New Orleans, Louisiana

Overview

One of the reasons that other sciences sometimes disparage psychology is that we often admit that we can't "prove" support for our theories. Proof is virtually impossible for psychology researchers to attain because controlling enough variables to isolate a relationship definitively is, well, virtually impossible. In other sciences, researchers can regulate conditions for their experiment to a greater degree than can researchers studying behavior. Chemicals rarely "have a bad day," and the structure of a cell is pretty much the same for one person as for any other. People, however, (and nonhuman animals, for that matter) operate much less in a vacuum than do the subjects of lab research for biologists, chemists, and physicists. As a result, psychologists are subject to issues of mood, relationships, traffic, and the like that affect the performance of our participants.

Assessing Dependability

In the absence of proof, the claims that psychologists make must rely on assessment of the likelihood that one's results are dependable. That likelihood never reaches 100 percent, but it can get very close. How close is close enough to count on a finding is a matter of some debate, but most researchers agree to a reasonable degree on 95 percent; such findings would mean that I would have only a 5 percent chance of being wrong in a claim that I might make about a relationship between a couple of variables. I'll discuss where this number comes from in a bit, but keeping that 5 percent value in mind can be helpful. First, let's consider a bit further why proof is unattainable and how we can assess the dependability of our results.

Consider a researcher who wants to know if a new way of teaching math can help students learn math more effectively. The first thing that researcher might want to do is to assess the math proficiency of the general population. Even this step presents a bit of a problem in that it would be impossible to get every member of a population to take a math test. So, what would the researcher do? First, he or she would gather a sample from the population that he or she hopes to make a statement about, and that sample would serve to represent the population for the purposes of the study. This sample would take the test and provide something of a baseline. Does that baseline perfectly match the population? Probably not, but it's the closest we can get, considering the circumstances. Obviously, the bigger and more random our sample is, the more closely it resembles the population, and the more dependable that sample's mean score is going to be.

This effort to gather a sample to represent the population is indicative of our goals in conducting such research. We typically don't care about our sample in and of itself. Instead, what we care about is the population that that sample represents. It's one thing to make a statement about a group of 50 or 1,000 people, but it's another to be able to

extrapolate what we learn about them to the greater population, who are those people who didn't participate as part of our sample. This is what inferential statistics is all about. We *infer* what the population is like, based on what we know about our sample.

Note that we actually can prove something about our sample. If we find that our sample averaged 47.5 correct answers on the test, we have proof about the sample mean. We can be sure how many questions each test-taker got right in the sample (as long as we calculated correctly). What we can't know for sure is what this says about everyone else—the people who weren't in the sample. So we take what we know (sample statistics) and infer what we don't know about the population (inferential statistics). What we're doing is estimating what the population's values would be, based on what the sample is like. This process is called *parameter estimation* because a parameter is a measurement for the entire population, as opposed to statistics, which refer only to samples. Because we can rarely know what the population is like, we estimate, through our inferential statistics.

Using Inferential Statistics

Once we have a mean for our sample, that mean can serve as our best guess of what our population is like; we have no reason to assume that our sample mean over- or underestimates the population mean. So, if our sample mean is 47.5, our best guess of the population mean (parameter estimation again) is 47.5. Why do we care? Well, this estimated mean might be helpful, but when it really comes in handy is when we're trying to do some research, for example, to see if the new way of teaching math is any good. We would predict that the sample who goes through the treatment (the new technique, say, teaching math through pictures instead of numbers) will have a higher mean than the assumed population mean of 47.5. But how much higher would the mean have to be in order to be meaningful? Consider that even if the treatment does absolutely nothing, the treatment group will still almost certainly not score exactly 47.5 because of a bit of chance fluctuation or mild irregularities in the sample (called *sampling error*), such as some especially lucky guesses. In fact, if we are reasonably certain that the group's mean will not be 47.5 exactly, there's a 50 percent chance that it'll be higher, and a 50 percent chance that it'll be lower. So we can't just look at the mean and say, "Oh, the treatment sample's mean is 47.8, so the treatment works." Most of us would not be inclined to intuit that such a difference was important, so we can accept that. However, we also can't look and say, "Oh, the treatment sample's mean is 85, so the treatment works." Regardless of how big the difference is, we need to see how likely that sample's mean is to have occurred even if the treatment did nothing.

So how do we assess the likelihood that the score from our sample would have occurred simply by chance and not because the treatment has an effect? Well, first we have to establish our *null hypothesis*, which is our assumption that the treatment has no effect. We hope to reject this null hypothesis; doing so would support our research (or, alternative) hypothesis, which is our real prediction in the study. Our statistical calculations of the probability of rejecting the null hypothesis usually depend upon several factors. The one that we typically have the most control over is the sample size, symbolized with N . Most statistical formulas adjust so that the larger the N is, the larger the statistic we are calculating; larger statistical values are more likely to be beyond our *critical value* (so called because it's the value that our statistic must reach in order to be considered significant), leading us to reject our null hypothesis. We therefore would like

Curriculum Module: Teaching Statistics and Research Methodology

to have a fairly large sample, despite obvious limits on how many people we can get into our sample. We can also enhance our ability to reject the null hypothesis by having a fairly strong manipulation and fairly sensitive measures.

Now let's get back to the .05 value discussed at the beginning. In research, the number commonly refers to our alpha level, which the researcher sets; most researchers are willing to set alpha as high as .05, but typically not higher. We can think of the value as the probability of rejecting the null hypothesis when we shouldn't; to do so would be committing a Type I error (or alpha error, for obvious reasons). Clearly, we would like this probability to be as low as possible because we would like to avoid committing a Type I error. So why not set it even lower? We usually don't set it lower at the outset because there is a major trade-off. If, for example, we set alpha very low (say, .0001), we obviously have a very low probability of committing a Type I error. That's good. But, we also make it extremely difficult to reject the null hypothesis at all because the critical value in such a situation is so large; our calculated value (r, t, F, etc.) would have to be untenably high in order to be statistically significant. It's so high and difficult to obtain that we may miss some real effects, and we don't reject the null hypothesis even if we should. Missing those real effects is called committing a Type II error (or beta error). By setting alpha at .05, we are liberal enough that we are confident about finding an effect if one is really there, but we're conservative enough that we'll be wrong only 1 out of every 20 times that we reject the null hypothesis. Given the trade-offs, most researchers abide by this convention.

The table below summarizes how our decisions to reject or retain the null hypothesis can be correct or incorrect, given the true situation that we are trying to infer.

		In Reality	
		Null Hypothesis is True (treatment doesn't work)	Null Hypothesis is False (treatment does work)
<i>Decision Based on Statistical Information</i>	Reject Null Hypothesis	Incorrect Decision; Type I Error (probability = alpha)	Correct Decision
	Retain Null Hypothesis	Correct Decision	Incorrect Decision; Type II Error

What does all this have to do with the proof problem that elicits disrespect from some researchers in other scientific disciplines? In reality, the null hypothesis is either true or false. That is, the treatment really does have an effect, or it doesn't. Two variables are either really related to one another, or they aren't. How do we know? We don't, because the null hypothesis pertains to the population, which we'll never know for sure. As a consequence, when we make our decision to reject the null hypothesis or not, we can't be sure we're making the right decision. The right decision for a false null hypothesis would be to reject it. However, we make this decision based on our sample, not the population, so rejecting the null hypothesis is a bit of an educated guess (based on those probabilities) of what the population is like. It's not the same as knowing what the population is like. Similarly, when we retain the null hypothesis, we don't know for sure if we've made the correct decision either. Why? Once again, because we don't know whether the null hypothesis is, in fact, true, which is our assumption when we retain the null hypothesis.

Conclusion

In reality, we may not be able to prove our predictions, but we can determine how reliable our results are. If a researcher set a reasonably low alpha level and conducted the study appropriately, then critics would not have a very strong case in countering the findings. So even in the absence of proof, we can support the conclusions to which our data lead us, and we can determine the likelihood of our being wrong. This absence of proof also serves as encouragement for replication, which further strengthens our claims. Just as a large sample can be more convincing than a small one, a lot of studies can be much more convincing than one.

A Lesson on Correlation

Amy C. Fineburg
Spain Park High School
Hoover, Alabama

Overview

Correlation is a misunderstood topic. This misunderstanding stems mainly from our early learning about the scientific method. We are taught in elementary science courses that the results of studies “prove” relationships between variables, and that if one variable leads to another variable occurring, then one naturally “causes” the other. This misunderstanding is most prevalent in popular press reporting of scientific findings. Journalists want to report the most salient portions of longer (and often dryer) scientific articles, and they will often discuss correlational research as being causal for the sake of brevity, clarity, or even expediency. Scientists themselves will misrepresent their own correlational findings as causal for the same reasons. Reporting correlational research requires hedged and qualified statements rather than direct and concise statements, which, quite frankly, aren’t as much fun to report.

Making sure students understand what correlational research is and what it is not is vitally important to helping them become good consumers of scientific information.

Learning Objectives:

Students must understand the following about correlational research:

- Correlation does not mean causation.
- Correlated variables may influence each other in both directions.
- Intervening variables may explain the correlation’s existence.

Without an understanding of these three concepts, students will not be able to parse the real from the opinion when correlational research is presented to them. They could then be easily lead to believe that oatmeal causes cancer or that television watching causes obesity. Both of which are not true statements, by the way.

Beginning the Lesson

I begin teaching correlation by having my students recite the mantra “Correlation does not mean causation” numerous times. We say it aloud together, often with a dramatic flair. My students often end up laughing at me for having them do this, but the phrase begins to stick with them. After we recite the mantra, I begin the explanation of what that mantra means.

First, I discuss how correlational data is obtained. We discuss survey research and quasi-experimental research. When I discuss these types of methods, I try to relate them to their own experiences. Most, if not all, students have taken a survey in their lives. Many have informally compared data collected at different times. I also give some fictional, but concrete examples:

Curriculum Module: Teaching Statistics and Research Methodology

A teacher wants to know whether grade point average (GPA) is related to test scores.

A store owner wants to know whether people who have the store credit card spend more money on clothes than those who don't.

At this point, we also discuss independent variables that cannot actually be independent, like gender, age, height, weight, etc. These variables are more likely to be used in correlational research because they cannot be manipulated. They can be used with other types of statistical analysis, but correlational research lends itself to working with these types of variables.

Second, we discuss correlation coefficients. Correlation coefficients reveal the strength and direction of the relationship between variables. I present first that correlation coefficients fall between -1 and $+1$, with 0 indicating no relationship between variables. A negative correlation is an indirect (or inverse) relationship, with one variable going up while the other goes down. For example, as TV watching goes up, GPA goes down. A positive correlation indicates a direct relationship, with each variable going in the same direction as the other. For example, as hours of studying goes up, so does GPA, or as number of brain cells goes down, so does GPA. This concept is often difficult for students to grasp, since they are trained to think that "negative" means "bad" and "positive" means "good." They often mistake a negative correlation as a weak relationship (a "bad" one) and a positive correlation as a strong relationship (a "good" one). To combat this, I go through several possible test questions about correlation coefficients:

Which of the following correlation coefficients presents the strongest relationship between the variables?

- a. .02
- b. -.67
- c. .55
- d. -.14

The correct answer is *b*. I then pose the question of which of the above shows the weakest relationship, which is answer *a*. If students are not getting the concept, I will put more on the board until I feel they have grasped this concept.

In addition to this discussion of correlation coefficients, I also show how correlational data is plotted on a graph called a scatterplot. I show students how positive correlation scatterplots yield a line of best fit that has a positive slope while negative correlation scatterplots yield a line of best fit with a negative slope. Zero correlation scatterplots yield a line of best fit with no slope (or better yet, don't yield a line of best fit!). Not only does this reinforce the idea that positive and negative do not mean "good" and "bad," but it also connects with knowledge they gained in math classes, especially geometry and algebra. I then give out a worksheet in which students must plot a set of data and determine the type of scatterplot the data yields. The handout I give is included at the end of this article.

Using Correlational Studies

Third, we discuss directionality. Correlation coefficients do not indicate the direction of the relationship between variables, just that a relationship exists. Does variable A lead to variable B, or vice versa? To demonstrate this problem, I present the following correlational studies that show how directionality cannot be inferred:

Curriculum Module: Teaching Statistics and Research Methodology

Researchers have long known about the correlation between eye-movement patterns and reading ability: Poorer readers have more erratic patterns (moving the eyes from right to left and making more stops) per line of text. In the past, however, some educators concluded that “deficient oculomotor skills” caused reading problems and so developed “eye-movement training” programs as a corrective. Many school districts may still have “eye-movement trainers,” representing thousands of dollars of equipment, gathering dust in their storage basements. Careful research has indicated that the eye movement/reading ability correlation reflects a causal relationship that runs in the opposite direction. Slow word recognition and comprehension difficulty lead to erratic eye movements. When children are taught to recognize words efficiently and to comprehend better, their eye movements become smoother. Training children’s eye movements does nothing to enhance their reading comprehension (Stanovich, 1998, as cited in Fineburg, 2003).

Hippocrates’ delightful *Good News Survey* (GNS) was designed to illustrate errors that can be hidden in seemingly sound scientific studies. The survey found that people who often ate Frosted Flakes as children had half the cancer rate of those who never ate the cereal. Conversely, those who often ate oatmeal as children were four times more likely to develop cancer than those who did not. Does this mean that Frosted Flakes prevents cancer while oatmeal causes it? Ask your students to suggest explanations for these correlations. The answer? Cancer tends to be a disease of later life. Those who ate Frosted Flakes are younger. In fact, the cereal was not around when older respondents were children, and so they are much more likely to have eaten oatmeal. The GNS finding that children who took vitamins were more than twice as likely to go on to use marijuana and cocaine was also likely due to these respondents being younger than average. Finally, the GNS revealed that people who had had routine physicals in the previous three years were twice as likely to report high blood pressure and cholesterol levels. Do physical exams cause health problems? No, the survey researchers suggest that the unmasking bias is probably operating, with those having had physicals simply more likely to know they have a problem (Tierney, 1987, as cited in Fineburg, 2003).

Scientists have linked television watching with childhood obesity. In fact, the degree of obesity rises 2 percent for each hour of television viewed per week by those ages 12 to 17, according to a study in *Pediatrics*, the Official Journal of the American Academy of Pediatrics. One explanation is that TV watching results in less exercise and more snacking (often on the high-calorie, low-nutrition foods pitched in commercials). Is that conclusion justified? What are some alternative explanations for the correlation? The causal relationship may be reversed. Obesity may lead children to prefer more sedentary activities, such as TV viewing. Or, some third factor may explain the relationship. For example, parents having little formal education may not emphasize good nutrition or good use of leisure time (Tierney, 1987, as cited in Fineburg, 2003).

Fourth, I discuss the problem of third (or extraneous) variables. Simply because two variables are correlated does not mean they are related only to each other. Often, other variables may explain the relationship between the variables. I again present the following stories to demonstrate this issue:

For instance, a positive correlation between milk consumption and incidence of cancer in various societies is probably explained by the relative wealth of these societies,

bringing about both increased milk consumption and more cancer as a function of greater longevity (Paulos, 1989, as cited in Fineburg, 2003).

In the New Hebrides islands, body lice were at one time thought to produce good health. When people became ill, their temperatures rose and caused the body lice to seek more hospitable abodes. Both the lice and good health departed with the onset of the fever. Similarly, the unexpected positive correlation between the quality of a state's day care programs and the reported rate of child abuse is not causal but merely indicates that better supervision results in more consistent reporting of incidents (Paulos, 1989, as cited in Fineburg, 2003).

This is my favorite correlation story! In the early twentieth century, thousands of Americans in the South died from pellagra, a disease marked by dizziness, lethargy, running sores, and vomiting. Finding that families struck with the disease often had poor plumbing and sewage, many physicians concluded that pellagra was transmitted by poor sanitary conditions. In contrast, Public Health Service doctor Joseph Goldberger thought that the illness was caused by an inadequate diet. He felt that the correlation between sewage conditions and pellagra did not reflect a causal relationship, but that the correlation arose because the economically disadvantaged were likely to have poor diets as well as poor plumbing. How was the controversy resolved? The answer demonstrates the importance of the experimental method. Finally he selected two patients—one with scaling sores and the other with diarrhea. He scraped the scales from the sores, mixed the scales with four cubic centimeters of urine from the same patients, added an equal amount of liquid feces, and rolled the mixture into little dough balls by the addition of four pinches of flour. The pills were taken voluntarily by him, by his assistants and by his wife. None of them came down with pellagra. To further make his case, Goldberger asked two groups from a Mississippi state prison farm to volunteer for an experiment. One group was given the high-carbohydrate, low-protein diet that Goldberger suspected to be the culprit, while the other group received a balanced diet. Within months, the former group was ravaged by pellagra, while the latter showed no signs of the disease (Stanovich, 1998, as cited in Fineburg, 2003).

Students usually find the last story particularly interesting (and gross!), but it clearly demonstrates that directionality cannot be inferred from a correlational study.

References

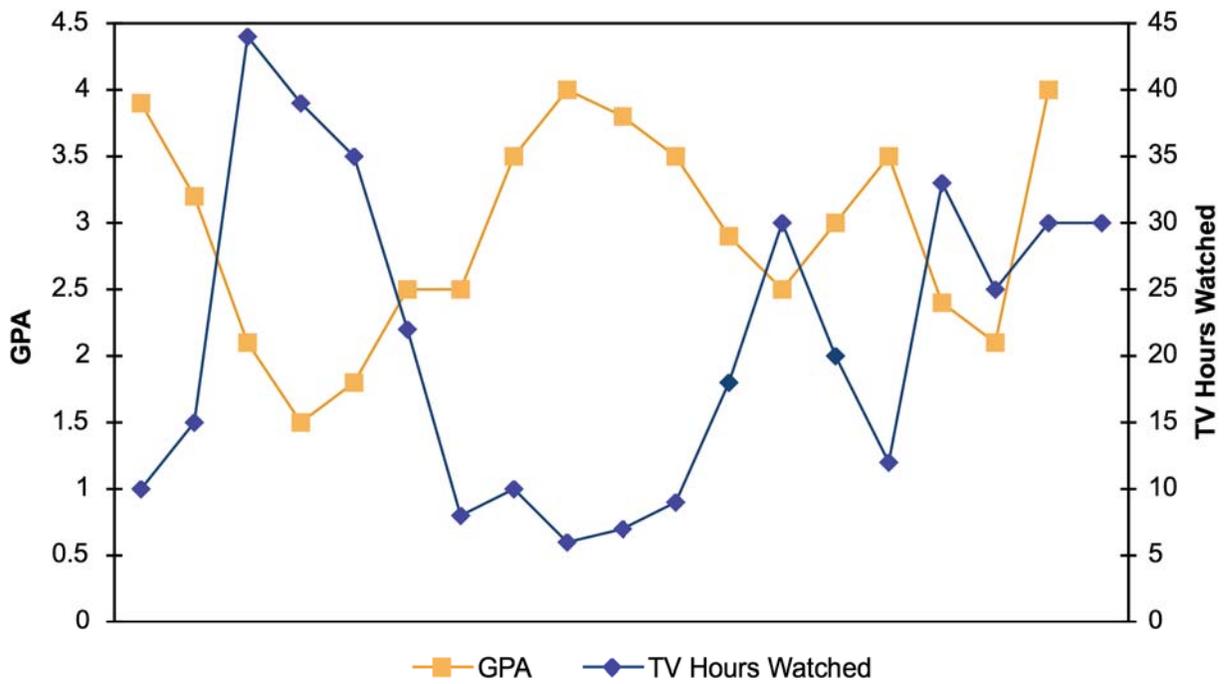
- Paulos, John Allen. 1989. *Innumeracy: Mathematical Illiteracy and Its Consequences*. New York: Hill and Wang.
- Stanovich, Keith E. 1998. *How to Think Straight About Psychology* (5th ed.). New York: Longman.
- Tierney, John. 1987, September/October. *Good News! Better Health Linked to Sin, Sloth*. *Hippocrates*, pp. 30–35.

Student Handout: Charting Correlations on Scatterplots

Directions: Chart the following data pairs on the graph below. Indicate whether the data resemble a positive correlation, negative correlation, or zero correlation.

GPA	Hours Watching TV per week
3.9	10
3.2	15
2.1	44
1.5	39
1.8	35
2.5	22
2.5	8
3.5	10
4.0	6
3.8	7
3.5	9
2.9	18
2.5	30
3.0	20
3.5	12
2.4	33
2.1	25
4.0	30

Source: Instructor's resources to accompany Blair-Broeker & Ernst. (Year) *Thinking About Psychology*. New York: Worth Publishers.



Teaching Statistics Concepts

Jonathan Stadler
Fisk University
Nashville, Tennessee

Overview

Teaching statistics concepts to psychology students is often a challenge because students often find the material boring, think that the concepts are irrelevant to their lives, or suffer from math anxiety. Although students in AP Psychology classes are usually quite capable in approaching most material, these same students may not appreciate the degree to which mathematical concepts, especially statistical concepts, are an integral part of the scientific study of psychology. I have found that a certain percentage of students have taken psychology specifically because they thought that they would avoid math as a result; to their chagrin, they are quickly disabused of this notion. In an effort to communicate important statistical concepts in a way that is active and engaging, the following exercises are presented for some of the basic concepts that students should understand.

Number Scales

Before addressing the specific concepts, I find it useful to describe the ways in which psychological data can be collected and measured numerically. I inform them that all data that can be measured and reported as numbers are categorized by one of four number scales. These scales are Nominal, Ordinal, Interval, and Ratio. A helpful mnemonic device is the acronym NOIR, which allows me to go on a brief tangent about my love of film and the genre of film noir. Following this introduction, I then give the students an example of measures on each scale and the characteristics of the scale, as well as any shortcomings of using the scale to measure psychological phenomena. (I also try to communicate that the type of statistics that you use depends on the number scale of your data, but this point may be more detailed than you need to worry about.)

Nominal: These numbers are purely for categorizing data into groups. As such, they have no quantitative properties. As an example, I describe collecting data about people's political views, where "1" designates being a Democrat, "2" designates being a Republican, and "3" designates being an Independent. The numbers are used solely to group participants into the appropriate categories. I try to illustrate this by joking that by adding a Democrat and a Republican together, you don't get an Independent, rather you get an argument. I then ask the students to come up with other information that we might desire to collect using a nominal measure.

1. **Ordinal:** These numbers contain some quantitative information, namely that of determining ranking: BETTER (as in National Collegiate Athletic Association (NCAA) basketball tournament rankings), FASTER (as in finishing positions in the 100 m dash), and SMARTER (as in class rankings). So, I describe the fact that the #1 seed in the NCAA tournament is better than the # 4 seed. However, I point out that beyond this general sense of ranking, the numbers don't provide much more information. For instance, I discuss that the interval in skill between the #1

- seed and the #4 seed isn't necessarily equal to the interval in skill between the #5 seed and the #8 seed, i.e., the intervals between ranks are not meaningful. I try to drive the point home by saying that the #1 seed is not four times better than the #4 seed; otherwise the #1 seed would never be upset by lower-seeded teams.
2. **Interval:** These numbers contain quite a bit of quantitative information, such that now adding and subtracting data measured on these scales is meaningful. The typical example used by nearly everyone is temperature, although perhaps a more relevant example to use for high school students is the SAT® or ACT score. In this case, the intervals between the numbers are equal in value, so that a 10 point change in the ACT means the same change whether the change is from 10 to 20 or from 20 to 30. Similarly, measures of intelligence (IQ), personality traits (Myers-Briggs Type Indicator®), and depression (Beck Depression Inventory®) are used on the assumption of an interval scale. It must be stressed to students, however, that interval scales do not have a true zero point, i.e., scoring a 0 on the extraversion scale of the Myers-Briggs test does not necessarily mean a lack of extraversion. Therefore, you cannot meaningfully discuss the RATIO between scores, e.g. you cannot say that a person with a Beck Depression Inventory score of 10 is half as depressed as a person with a score of 20.
 3. **Ratio:** These numbers contain the most amount of quantitative information of any of the scales because these scales have a true zero point, i.e., there is the possibility of having an absence of a measure. Typical examples are speed, time, distance, height, and weight, although you can also include grade point average (GPA). With these measures, you can meaningfully describe something as twice as long, half as fast, being four times as heavy, etc.

(Note: After introducing GPA as a ratio measure, I like to discuss the degree to which some of these measures are or are not truly measured on the scale that they are assumed to be. A good way to introduce the discussion might be to find an Internet survey that uses a Likert-type scale and ask students whether it represents a nominal, ordinal, or interval scale. You may be able to find a couple of examples where it could legitimately represent any of these scales. This then allows you to discuss the decisions that psychologists sometimes have to make concerning how data are collected and how the decisions aren't always straightforward.)

Descriptive Statistics

Once students are introduced to number scales, they are ready to tackle descriptive statistics. For the sake of students in my introductory psychology course, I discuss measures of central tendency and measures of dispersion. Students generally have an easy time understanding measures of central tendency but struggle to understand measures of dispersion. Following are some suggestions for ways to make these concepts more understandable to students.

For these exercises, some pre-class preparation is needed. Students often understand statistical concepts better if data sets are used that relate to them. Therefore, if you wish, you can construct a brief demographic sheet that you hand out to the students a week before you will cover these concepts. On these demographic sheets, gather data on rather innocuous characteristics, e.g. shoe size, height, number of visits to a local attraction (Walt Disney World, Sears Tower, etc.) in the previous year, number of

siblings, etc. Be creative, but be sure that none of the data are too personal. You can then use these data in place of the arbitrary numbers that will be provided below. Using the real data will require additional work, especially if you wish to make the computations less messy. If you prefer, the data provided will illustrate the points being made and result in straightforward answers.

For these demonstrations, you will need the following props. If you will be using the students' data as your data set, you will need 80 index cards. If you will be using the provided data sets, you will need either three decks of playing cards or 80 poker chips (preferably in more than one color—I prefer five colors). You will also need at least three differently colored pens or markers (five if you use all-white poker chips for your demonstrations).

Central Tendency

To illustrate the concept of central tendency, I begin with 17 poker chips. Five chips are green and each has a “5” written on them. Three chips are blue and each has a “10” written on them. One chip is red and has a “25” written on it. Eight chips are white and each has a “100” written on them. (Instead of poker chips, you can use index cards with student data in the same proportion or you can use the playing cards using number cards to represent the four groups: five “2”; three “3”; one “4”; eight “5”; and one “10” for the extreme score.)

I go around the classroom and have students pick a chip and go to the front of the room and order from lowest to highest score. I ask the students to hold out their chip in front and tell the class what their score is. After going down the line, I tell the entire class that we will be looking at different ways to characterize what is the typical behavior of this group, which is known as central tendency.

First, I ask them which chip (“score”) is the most numerous. Students answer “100” and I ask them the number of scores. I then tell them that we have determined the *mode* for the group. I ask them to think about whether this is the best way to characterize the group as a whole as we look at two other measures.

Second, I ask them which chip (“score”) is in the middle of the group. Based on the data, the student with the red chip (“25”) is in fact in the middle of the group (Score #9). I then tell them that we have determined the *median* for the group and have them take note of the fact that there are equal numbers of scores above and below the median. I also point out that the mode and median for this group are different. I then ask the student with a white “100” chip at the end of the line to briefly sit down. At this point I then discuss how the median is determined when there is an even number of scores. (In this case, you add 10 [Score #8] + 25 [Score #9] and divide by 2 to get 17.5) After this, I ask the 17th student to come back in line.

Finally, I ask them to determine the average score for the group. I tell them that they need to add up all 17 scores and then divide the answer by 17 ($880/17 = 51.76$). I tell the students that the technical term for “average” is *mean*. Then I take a student who has a white chip, take back their chip, and hand them a black chip with “500” written on it. I then ask the student to go to the end of the line. Now I ask the students to recalculate the mode, median, and mean for the group with the new “score.” In this example, the mode and the median will remain the same, BUT the mean will increase dramatically ($1280/17 = 75.29$). With this recalculation, I am able to discuss how extreme scores affect the mean

while the mode and median are relatively unchanged by one extreme score. At this point I can ask the students which measure is the best way to characterize the typical behavior of this group. Taken together, an opportunity opens for discussing why the mean isn't always the best measure of central tendency, e.g. reporting median income versus mean income for a community.

Dispersion

For the presentation of dispersion measures, the poker chips are the best way to go. (If you would rather use playing cards, you simply need to draw out the cards that represent the numbers on the poker chips and sort them into the four groups. The personal demographic information becomes too unwieldy to construct a demonstration of dispersion that is easy to calculate.) This demonstration works well with class sizes of 20 and greater. If you have fewer than 20 students, you can do the demonstration sequentially with each group.

You will need to set up the chips beforehand and have four plastic sandwich bags to put the chips into. For this exercise you will need four sets of 10 chips (preferably of different colors, though not necessary) and three colored markers. For each group of chips, you will label the low score with a red marker, the high score with a blue marker, and the rest with a black marker. Here are the data for the four groups of 10 chips:

Curriculum Module: Teaching Statistics and Research Methodology

Group A: 4 5 5 6 6 6 6 7 7 8 (Mean = 6; Range = 4; Variance = 1.2; St. Dev. = 1.1)

Group B: 1 3 5 5 6 7 7 8 9 9 (Mean = 6; Range = 8; Variance = 6; St. Dev. = 2.4)

Group C: 3 6 7 7 7 7 8 9 9 (Mean = 7; Range = 6; Variance = 2.6; St. Dev. = 1.6)

Group D: 5 5 5 5 9 9 9 9 (Mean = 7; Range = 4; Variance = 4; St. Dev. = 2)

For this demonstration, you will first have students choose from Groups A and B (total of 20 students needed, or demonstrating Group A and then B separately). In the front of the classroom, have numbers 1 through 9 posted equidistant from each other (either on a whiteboard/chalkboard or on pieces of paper taped along the front wall). Have students from Group A line up with the number that corresponds to their score. Have the students with the red- and blue-labeled chips raise their hands. Have students in the class record the data (circling the score represented by the red-labeled chip and underlining the score represented by the blue-labeled chip) and put the data on the board or PowerPoint slide. Ask the students to note how closely grouped the scores are, and then have Group A sit down and have Group B line up in front. After following the same steps as with Group A, ask the students which group seemed more tightly packed. (They should answer Group A.) After Group B has sat down, have students calculate the mean for both groups. Students will notice that the means were the same. You can then discuss what dispersion means as it relates to the mean; that it is a measure of how spread out the scores are from the mean; and that in the first case you could visually see that the scores seemed to be more tightly clustered around the mean in Group A.

At this point, you can introduce the three measures of dispersion with the data still on the board:

First, there is the *range*, which is a measure of the absolute spread of the scores. The range is calculated by subtracting the lowest score (red score) from the highest score (blue score). Students should note that Group A has a smaller range than Group B, which confirms what they saw. Tell the students that using the range as a measure of dispersion is rather crude because one does not get a sense for the overall distribution of scores in the sample; in addition, one extreme score can radically alter the range and obscure the fact that the rest of the scores are rather tightly clustered.

Second, there is the *variance*. This is a numerical representation of the dispersion of scores around the mean. The smaller the number is, the less spread out the distribution of scores is around the mean. It is preferable to the range because variance takes into account all of the data, not just the highest and lowest score.

Calculating variance can be confusing for students. I find that it is best to break down the process into five steps. At each step, you can explain what the purpose is.

1. The first step is to calculate the mean. (The students have already done this for Groups A and B.) The mean is needed as an anchor point from which to evaluate how spread out the scores in the group are. In the mathematical formula, this step is represented by the symbol μ .
2. The second step is to subtract the mean from every individual score. We are calculating deviation scores, which represent how close each score is to the mean. (IMPORTANT: Always subtract the mean from the individual score [score – mean]. You should end up with positive AND negative numbers.) The larger the number is, the farther the score is from the mean. You should have as many

deviation scores as original scores. (In our example, there will be 10 deviation scores.) In the mathematical formula, this step is represented by $(x-\mu)$. If you add up the 10 deviation scores, the sum should be zero. This is a property of the true mean. If you get a number other than zero when adding up the deviation scores, this indicates that you have not calculated the mean correctly (or that there were rounding errors when there are decimal values). (This can be a good check that students didn't make any mistakes in their calculation of the mean.)

3. The third step converts the deviation scores to a form that allows us to add them up and not get zero. In this step, we take the square of the deviation scores so that we remove all negative values. For example, in Group A, the deviation score for the first score (4) is $(4-6)$, which is -2 . We then take the square of (-2) , which results in $(-2) \times (-2) = 4$. By following this through with all 10 scores, we have numbers that can be added together. In the mathematical formula, this step is represented by $(x-\mu)^2$.
4. The fourth step is to add the squared deviation scores together. The symbol for adding scores together is Σ , so the formula now looks like this: $\Sigma(x-\mu)^2$. Remind students that the larger this number is, the greater the dispersion is. (In our example, the sum of Group A is 12.)
5. The final step is to divide the sum by the number of scores in the group (N). This final step in calculating the variance allows you to get an estimate of the average distance that a score is away from the mean (this is a slightly oversimplified definition). The mathematical formula for variance in its final form is:

$$\Sigma(x-\mu)^2 / N$$

Often, students are told to divide by the number of scores minus one (N-1). There is good reason for doing so, but it relates to *inferential statistics*, which is outside this discussion. When using *descriptive statistics*, use N.

The final measure of dispersion is *standard deviation* (SD), which is related to variance. In fact, calculating standard deviation is relatively easy once students understand how to calculate variance because it involves adding one additional step: take the square root of variation. In the mathematical formula, it looks like:

$$\sqrt{(\Sigma(x-\mu)^2 / N)}$$

Like the variance, SD is an estimate of the amount of dispersion around the mean. More technically, it represents the extent of error that is being made by using the mean to represent the scores in the group. The smaller the SD, the more confident you can be that the mean is a good estimate of the group.

Now, walk through the demonstration again using Groups C and D. In this case, you can choose to have students simply come up and write their data on the board and then sit down rather than stand up in the front of the class. Have the students that got the low and high score in each group again circle or underline their data. Before you have students do any calculations, have them guess which group will have the larger variance and SD. Then have them calculate the range and ask them again. (This may lead to the answer that Group C will have the greater variance.) The data for Groups C and D were determined so that the range for Group C is greater, but the variance and SD for Group D

is greater. You can then reemphasize that the range is a crude measure of dispersion, whereas variance and SD take account of all of the group's data.

Once students feel comfortable calculating these values, you can have them calculate means, variances, and SDs for the demographic data that was collected earlier. It's likely that those data will result in more complicated calculations (e.g. involving decimals), so you'll want to make sure that students understand the concepts first.

An Optional Topic

As an optional topic, you can discuss which measures of central tendency and dispersion are appropriate for different measurement scales. Because the mean, variance, and SD provide the most information, they require measurement scales that provide the most information. Therefore, you can only calculate the mean, variance, and SD when you have data that are measured on an interval or ratio scale. The range and median require a measurement scale that provides rank order as a minimum. Therefore, you can calculate the range and median using an ordinal scale, as well as an interval or ratio scale. The mode can be calculated using all four measurement scales: nominal, ordinal, interval, and ratio. The mode is simply determining the number of scores, so the numbers do not actually have to represent quantitative properties. There is no dispersion measure for data recorded on a nominal scale.

Contributors

Amy C. Fineburg teaches AP Psychology at Spain Park High School in Hoover, Alabama. She is the author of the current edition of the Teacher's Guide for AP Psychology and has been a Reader since 2000. She has also authored a teacher's edition and compiled a resource manual for Thinking About Psychology, a non-AP high school psychology textbook. She has presented at several local, regional, and national conferences and served on the Teachers of Psychology in Secondary Schools (TOPSS) Executive Board as member-at-large. She is pursuing a Ph.D. in educational psychology at the University of Alabama.

Elliott Hammer is Associate Professor of Psychology and the John LaFarge Professor in Social Justice at Xavier University of Louisiana in New Orleans. He received his B.A. in Psychology from the University of Kansas and his M.S. and Ph.D. in Experimental Social Psychology from Tulane University. He teaches courses in research, statistics, and writing, Social Psychology, Cognitive Psychology, and a seminar in stereotyping and prejudice. His research focuses on the cognitive and affective components of stereotyping, particularly from the perspective of the target. He is currently writing a book on the psychology of disaster. He has been a Reader for AP since 2001.

Jonathan Stadler currently teaches at Fisk University. He received his B.A. from Cornell College in Mt. Vernon, Iowa, in psychology and biology. He earned his M.A. and Ph.D. in psychology, with a concentration in behavioral pharmacology from Vanderbilt University. His interests include the effects that stimulant drugs of abuse have on the brain; the role of withdrawal and tolerance on the perpetuation of drug abuse; and the societal issues surrounding the treatment and prevention of drug abuse. In addition, he has an interest in popular culture and its portrayal of psychological phenomena. Jonathan is a member of the Society for the Teaching of Psychology and has been a Reader for AP Psychology, as well as a consultant for online courses in psychology and biology.

Editor/College Board Advisor

Chris Hakala is Associate Professor of Psychology at Western New England College in Springfield, Massachusetts, where he has been teaching since 2002. Chris has a B.A. in psychology from Castleton State College and his M.A. and Ph.D. in experimental psychology from the University of New Hampshire. He is currently Associate Director for the Office of Teaching Resources in Psychology (OTRP) for the Division 2 of the American Psychological Association. He has read AP Exams since 2000 and is currently a Question Leader. Besides teaching Introduction to Psychology, he teaches courses in cognitive psychology, history of psychology, and research methodology.