UNIT 4: QUANTITATIVE METHODS

QUESTION #4.1: How does quantitative research "prove" hypotheses?

SHORT ANSWER: by calculating the probability of the results occurring by pure chance

Social scientists don't like to speak of their research as "proving" their hypotheses. The preferred term is "confirming" the hypotheses. The way that this is done is rather backward: we calculate (or estimate) the probability of the observed results occurring by random variation (i.e., pure chance, luck) and then if that probability is sufficiently low enough, we say that some alternate hypothesis is a better explanation for the results.

http://www.youtube.com/watch?v=G60Hp_iFW5I

This approach to **statistical significance** is called null hypothesis testing. A **null hypothesis** is a statement that that we did *not* really prove anything because the observed results could be explained by random variation. In order to prove something, I have to come up with an initial hypothesis that I will test. I do this by stating my hypothesis as an alternative to the null. The only way I can confirm my hypothesis is to reject the null. The only way I can reject the null is to show that it is a very improbable explanation of the results.

Many times, the null hypothesis is obviously the best explanation for the observed results. For example, suppose I say that the majority of the shoppers at a certain grocery store on Tuesday afternoons are women. (In other words, my alternate hypothesis is that most of the observed shoppers will be female.) So, I stand outside the store and observe the first three customers to exit the store. Number one is a woman in her thirties, pushing a basket with two little children inside; number two appears to be an older woman, alone, looks like she just finished some activity at the senior center across the street; number three is a younger women in her late teens or twenties, looks like she just got off of her office job. So, I'm three for three; does that mean that I confirmed my alternate hypothesis that most shoppers are women? If you are thinking like a scientist, your reply would be to stick with the null hypothesis as plausible explanation. Women are half of the population in the city where I did my observations, so the odds of observing three woman would be ½ times ½ times ½ which would be 0.125. Most scientists would not reject the null at that probability of random variation explaining the results.

Statistical significance is expressed in terms of a **p** value, which stands for probability (of the null hypothesis). P values range from 0.00 (indicating that something is impossible) to 1.00 (indicating that something is certain). The more improbable that the null hypothesis is, the more likely our alternate hypothesis is. Scientists usually accept the following cut off points for rejecting the null hypothesis: p below .05, reject the null with fair confidence; p below .01, reject the null with good confidence; p less than .001 reject the null with excellent confidence.

http://www.youtube.com/watch?v=G60Hp_iFW5I

STATISTICAL SIGNIFICANCE (probability of the null hypothesis)

```
p = 1.00  - - - - - - - - - - - - - - - - - - - - - - - - - - - - (certainty)

                p > .10        not significant ACCEPT THE NULL

p = .10   - - - - - - - - - - - - - - - - - - - - - - - - -

                p < .10        marginal        ACCEPT THE NULL

p = .05   - - - - - - - - - - - - - - - - - - - - - - - - -

                p < .05        fair            REJECT NULL

p = .01   - - - - - - - - - - - - - - - - - - - - - - - - -

                p < .01        good            REJECT NULL

p = .001  - - - - - - - - - - - - - - - - - - - - - - - - -

                p < .001       excellent       REJECT NULL

p = 0.00  - - - - - - - - - - - - - - - - - - - - - - - - - - - (impossibility)
```

In the case of the above example, the probability of getting three women out of three observations was p = .0125. We look at the above chart and find where that would place is: we are still in the area that is not significant. We must *accept the null hypothesis* and admit that our (alternate) initial hypothesis was not confirmed.

This process of calculating (or estimating) the probability of the null hypothesis is known as **inferential statistics**. They give us the p value that tells us if we have confirmed our hypothesis or whether we have to accept the null (and admit that we have proved nothing).

- -

GRAMMAR LESSON: Do not use the word *significant* unless

you mean p < .05.

- -

If we reject the null hypothesis when there really isn't an underlying difference or correlation, we have committed a Type I error. We can reduce the number of **Type I** errors by gathering our data carefully, selecting the right statistical tests or even by imposing a more stringent standard for the rejection of the null (e.g., the .01 level). However, we would not want to increase our vulnerability to **Type II** error (which means failure to reject the null even though real differences do exist.

QUESTION #4.2: How important is sample size?

SHORT ANSWER: sample size is very important in quantitative research

In doing qualitative research, it is less important how many are in our sample, and more important how much we get from them. In quantitative research, sample size is very important because the larger the sample size, the easier it is for an observed trend to be statistically significant. In the above example, I observed a definite trend: all shoppers leaving the store were women, but my sample size was small (n = 3). Suppose I had observed four customers leaving the store: all women. Now, the probability of that occurring would by half of 0.125 (p = .0625) but the above chart shows that although we are marginal (getting close to statistical significance) we would still have to accept the null. However, if my observation was five our of five, that is even less likely to occur by pure chance (p = .03125) we could then reject the null with fair confidence.

Let us be clear, you cannot just stand out in front of the store and wait for four women in a row to exit, ignoring all the men who came before and after. That would be as bad as flipping a coin, ignoring the tails, and claiming that you got a lot of heads. You should include all observations in your sample. However, with a sample size of a thousand, if you observed 55% females exiting the store (compared to only 45% males) those results might be statistically significant. Notice that most polling companies and marketing research companies usually have a sample size of at least a thousand. However, for student research, a sample size of fifty might be more realistic, but you might need to observe almost 40 women in order to reject the null hypothesis.

QUESTION #4.3: What other kinds of statistics are used in research on consumer behavior?

SHORT ANSWER: descriptive statistics such as mean, median, mode, percent, standard deviation

Most (alternate) hypotheses are phrased in terms of measures of central tendency about the dependent variable: mean, median, mode, percent. Some hypotheses are phrased in terms of the dispersion (e.g., standard deviation, variance, range). Which of these measures is appropriate depends upon the quantitative scale in which the dependent variable is measured. Statistics textbooks explain the distinctions between ratio and interval scales (and whether they are continuous or discrete). Most of that doesn't really matter in selecting the best measure to describe the central tendency of the dependent variable. What matters most is whether the distribution of the variable is close to symmetrical (i.e., normal, standard, Gaussian) such that most of the scores are close to the middle of the range. In that case, we can use the arithmetic **mean** as the average.

http://www.youtube.com/watch?v=16hUQrX8akI

If there are a few extremely high (or low) outliers, we say that the distribution is **skewed**, and it would be wiser to use the **median**, which would be the score attained by a person in the middle of the distribution.

http://www.youtube.com/watch?v=FwsImyWiqjY

Income distribution is a good example of a skewed variable. Suppose a men's group at a church has ten members. Nine of them are small business owners or professionals and make close to $100,000 annually. The tenth member is the C.E.O. of a large corporation, and his income last year was over ten million dollars. This would produce an extreme right skew in our distribution. If we did the calculation required by the arithmetic mean, we would show an average income of over a million dollars, yet only one man in that group made over that amount. If we used the median as our measure of central tendency, we would get a more realistic average of $100,000.

When we have an ordinal scale, comprised of various levels, we could describe the median as the level that a person in the middle distribution would occupy, or we can simply give the percents in each level.

Customer rating of service	Excellent	Good (median)	Fair	Poor
Percent of customers	23%	37%	25%	15%

When we have a nominal scale, comprised of categories, we could describe the **mode** as the category occupied by a plurality of the subjects, or we can simply give the percents in each category.

http://www.youtube.com/watch?v=t59YiuRTkr0

Suppose we started a new company early in 2011. We want to see if our 2012 customers were repeat customers.

Customer's previous purchase was from	Our company (mode)	A competitor	No previous purchase of this product
Percent of customers	40%	35%	25%

If you know how to do percents, you know how to do the most important calculation in marketing research. A **percent** is a part / whole relationship. It is like a proportion, except that we multiply by 100 at the end. The important part in correctly calculating a percent is to properly identify the relevant part and the relevant whole.

http://www.youtube.com/watch?v=TwQDnH3vvKg

For example, the above calculation uses as a whole the number of customers our company had during 2012 (n = 500). Within those, we identified the part (n = 200) that we had previously sold to. So we simply take the 200, divide by 500, then multiply the quotient by 100 to get 40%.

Perhaps a better question to ask would be what happened to all of the customers we had during 2011? How many of them came back in 2012? Suppose that had fewer customers in 2011 (n = 400). We know that 200 came back in 2012. The numerator (part) is the same as in our previous calculation, but now the denominator (whole) is different: so we take 200 divided by 400 and then multiply by 100. The answer is now 50%. We get a different percent as an answer because we are asking a different question. The first time we asked, "How many of our 2012 customers were customers from before?" and we got 40% as the answer. Now, we asked "How many of our 2011 customers came back in 2012?" and we got 50%.

Don't confuse these simple part / whole percents with the percentage change. The simple percents using the above formula can never be greater than 100% or less than 0% (and they can never be negative). These limits do not apply to percent change calculations.

http://www.youtube.com/watch?v=oxgt4cLYo1M

At times, you may be asked to provide a variable's measure of dispersion. Most of the time, simply providing the minimum and maximum scores should be sufficient, but a more precise measure would be the standard deviation.

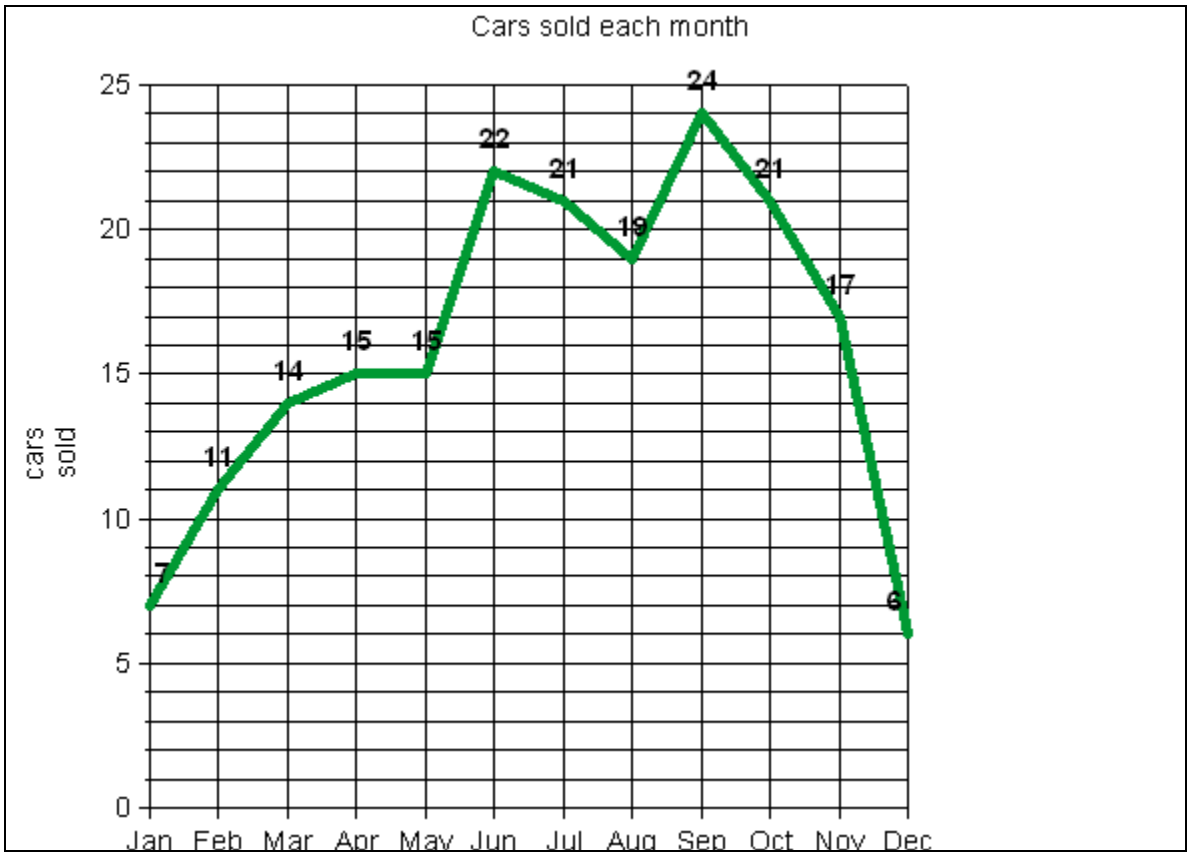
http://www.youtube.com/watch?v=NQa7fheN7vk

QUESTION #4.4: What kinds of graphs can be used for consumer behavior research?

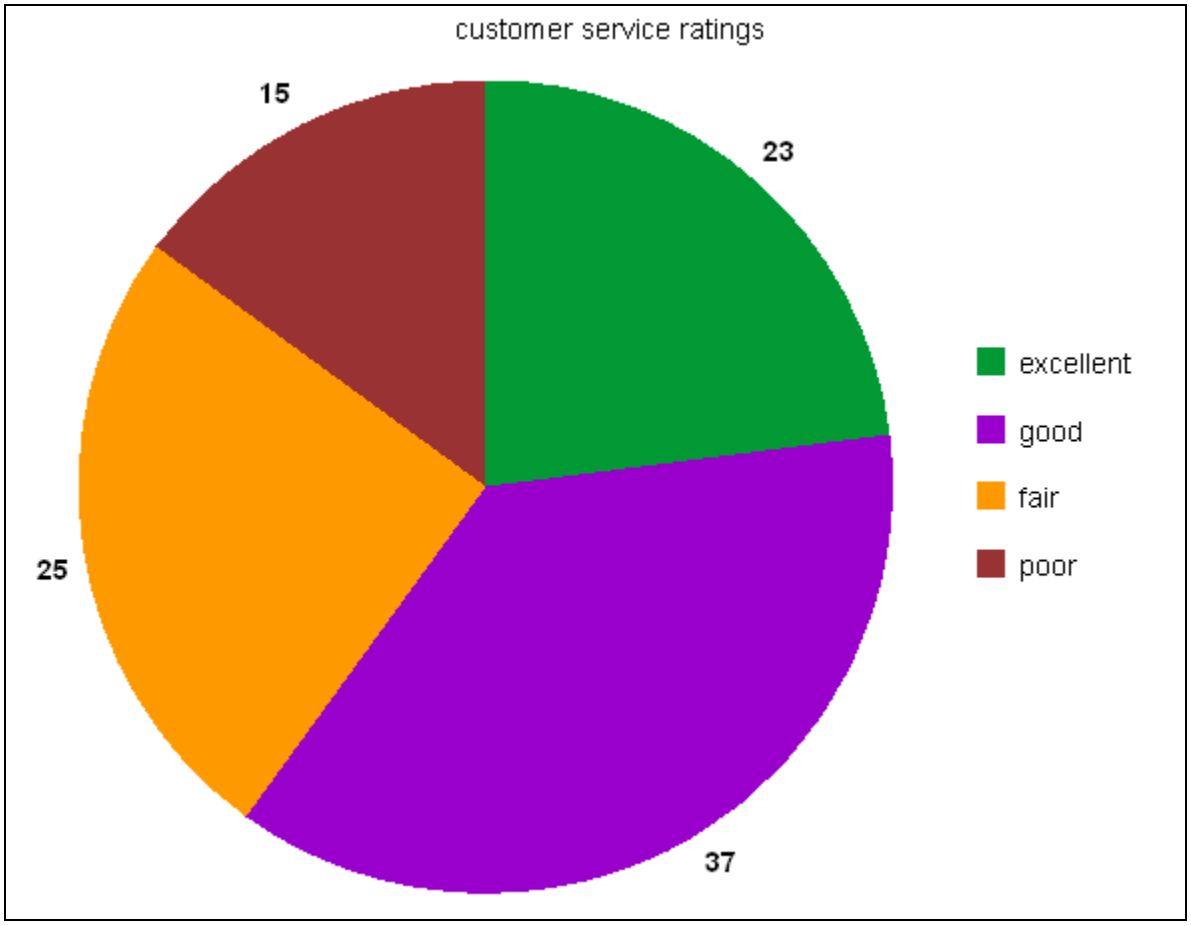
SHORT ANSWER: line graphs, bar charts, pie charts, and scatterplots

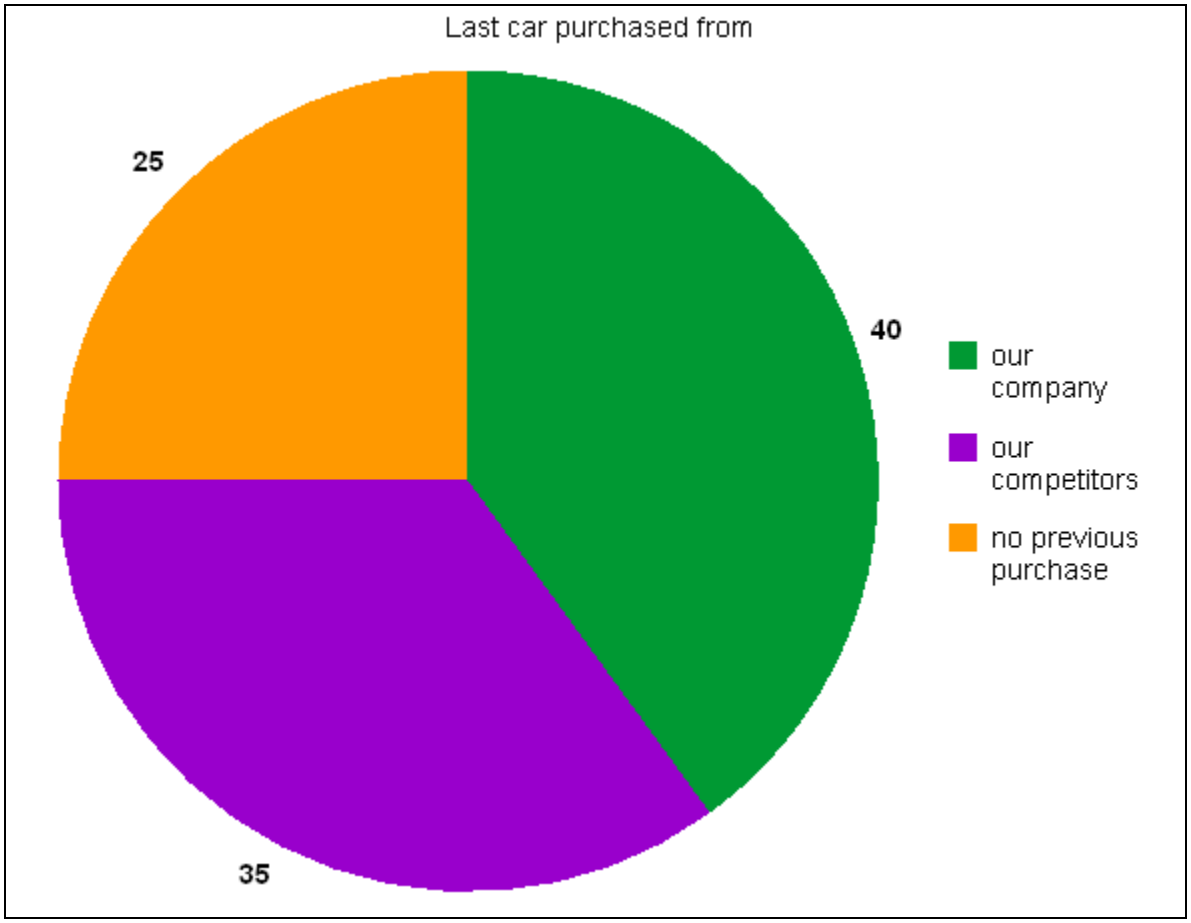
http://www.youtube.com/watch?v=CH_cn4SbWcE

In many areas of business (e.g., finance, marketing, production, human resources) **line graphs** are used to show the change in a variable over time. The time unit may be years, quarters, months, weeks, or even shifts. The line can represent trends in absolute numbers of a variable (e.g., the number of accidents, gross sales) or means (e.g., production) or percents or even some measure of dispersion (e.g., plotting standard deviation as a measure of quality control). However, consumer behavior research has less use for line graphs.



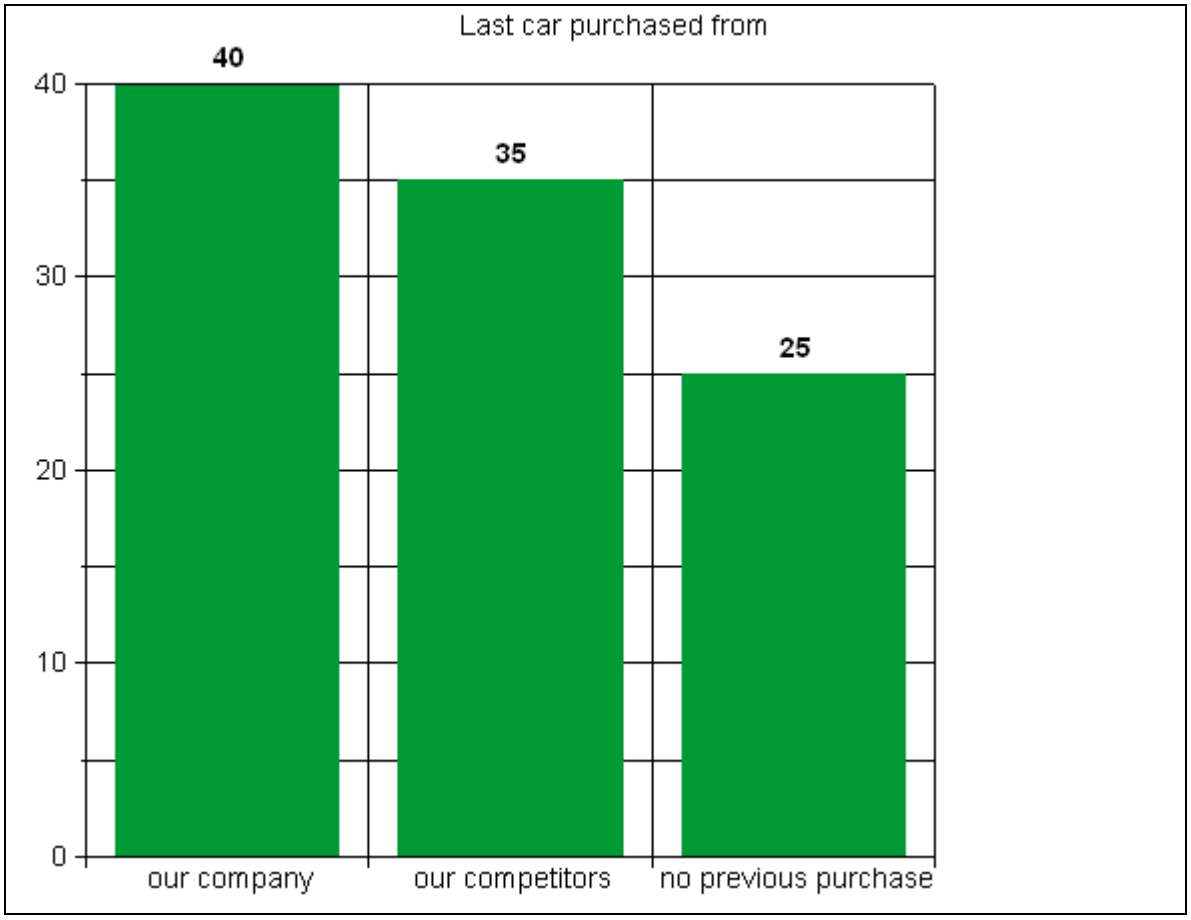
Pie charts are circular diagrams representing the proportionate distribution of a variable. Pie charts are good when we have a variable measured in simple percents (whether the distribution is on an ordinal or nominal scale). The size of each slice is proportionate to the percentage: the larger the percentage, the larger the slice. Here is what the previous examples of percentage distribution would look like as pie charts.


customer service ratings
15
23
25
37
excellent
good
fair
poor

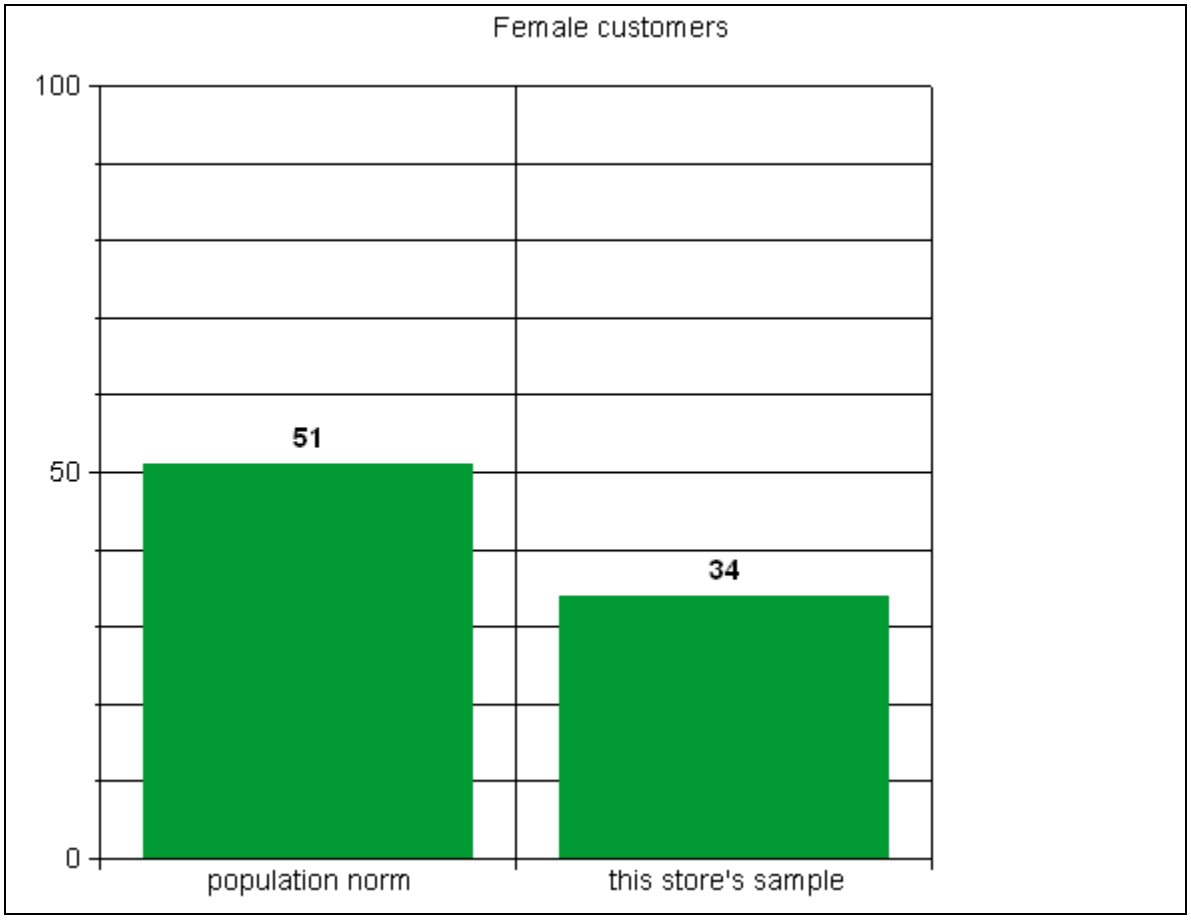

Last car purchased from
25
40
35
our company
our competitors
no previous purchase

Bar charts use horizontal or vertical bars to show a comparison of categories of a variable, or between a sample and a population, or between different groups, or between different time periods. Anything that a line graph can do, a bar chart can do: have each time period be represented by a different bar such that the length of the bar depicts the level of the variable for that time period. That could be the absolute number of the variable, a mean, median, percent, range or standard deviation. Our above example of a line graph could be portrayed as a bar chart, just make each month a different column and have the height of the column be the number of cars sold.

Anything that a pie chart can do, a bar chart can do. Each slice of the pie can be represented by a bar. The bigger the slice of pie, the longer the bar. Here is a bar graph depicting was the two previous examples of pie charts say.

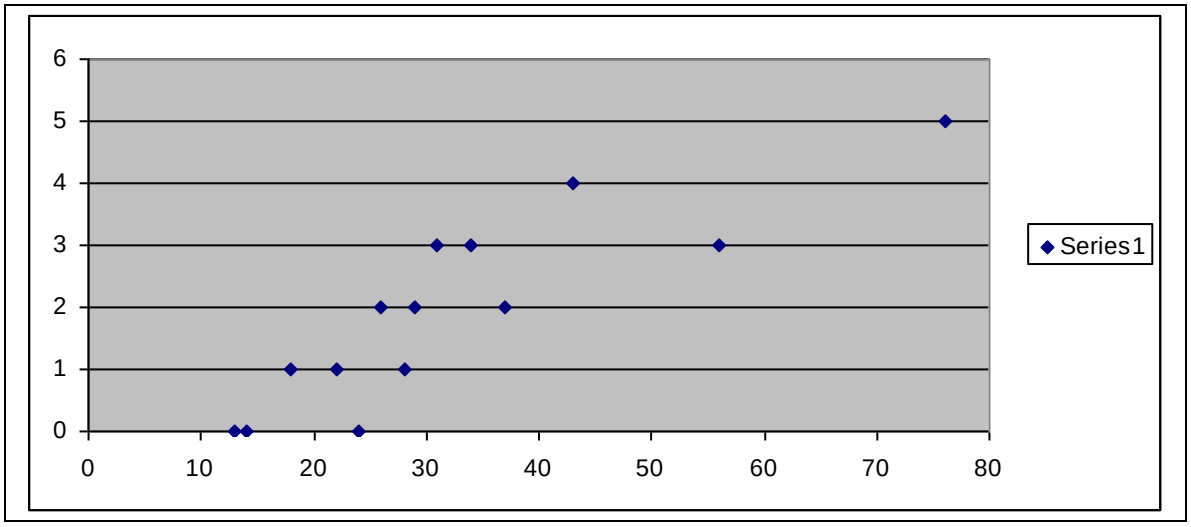


A further advantage of the bar chart over the pie chart is that one bar graph may compare several pie charts representing different groups, time periods, or a sample vs. a population.



A bivariate **scatterplot** depicts the relationship between two variables, such as customer income (variable X) and the amount spent on vacations (variable Y).

http://www.youtube.com/watch?v=CJX6TegkQe8



To make graphs for your variables, you can use different spread sheet programs, such as Excel (where vertical bar graphs are called column graphs) or you can use websites such as

http://nces.ed.gov/nceskids/createagraph/

QUESTION #4.5: Are correlations used in consumer behavior research?

SHORT ANSWER: Yes, very often.

Correlation describes the association between two variables (or two measures of the same variable). The direction of a correlation may be positive, negative, or zero.

http://www.youtube.com/watch?v=0AUjt_MA72U

- -

GRAMMAR LESSON: In this class, do not use "positive" to mean good or "negative" to mean bad.

- -

A **positive correlation is a direct relationship** between the two variables. (We never use the words *good* or *bad* to describe a correlation.) A direct relationship means that if a given subject scores high on one variable, he is likely to score high on the other variable; and if a subject scores low on one variable he is likely to score low on the other variable. An example would be the one about consumers' income level and the amount they spend on vacations. The *higher* the income the *more* spent on vacations; the *lower* the income, the *less* spent on vacations. Notice that as we would go from left to right, the regression line slopes up (a positive slope).

The **contingency table** is another diagram for showing correlation, especially in nominally scales variables (e.g., male / female, yes/no, Brand A / Brand B). Here is an example of the direct correlation between whether consumers saw an advertisement (independent variable) and whether or not they expressed an intention to purchase the product (dependent variable).

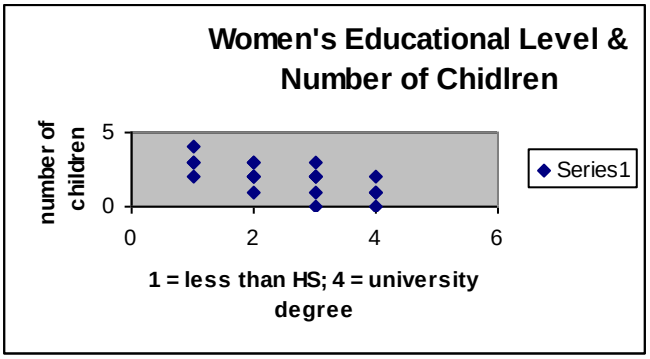
http://www.youtube.com/watch?v=zEjA2n-3LFk

Grouping	Intention to purchase	No intention to purchase	Totals
Saw advertisement	15	35	25
Did not see advertisement	5	45	25

Totals	20	80	50 = n

What this direct correlation means is that those who saw the advertisement were *more* likely to express an intention to purchase.

Negative correlations occur when subjects who score high on one variable tend to score low on the other variable. Another name for this relationship is **inverse**. (We never use the words "good" or "bad" to describe a correlation.) For example, there is a negative correlation between a woman's education level and the number of children that she will bare. The term “negative” does not imply that education is bad, or that having children is bad, it merely means that the *more* education a woman has, the *fewer* children she tends to give birth to. A scatterplot of that relationship might look something like this for a hypothetical sample of 40 year old women.



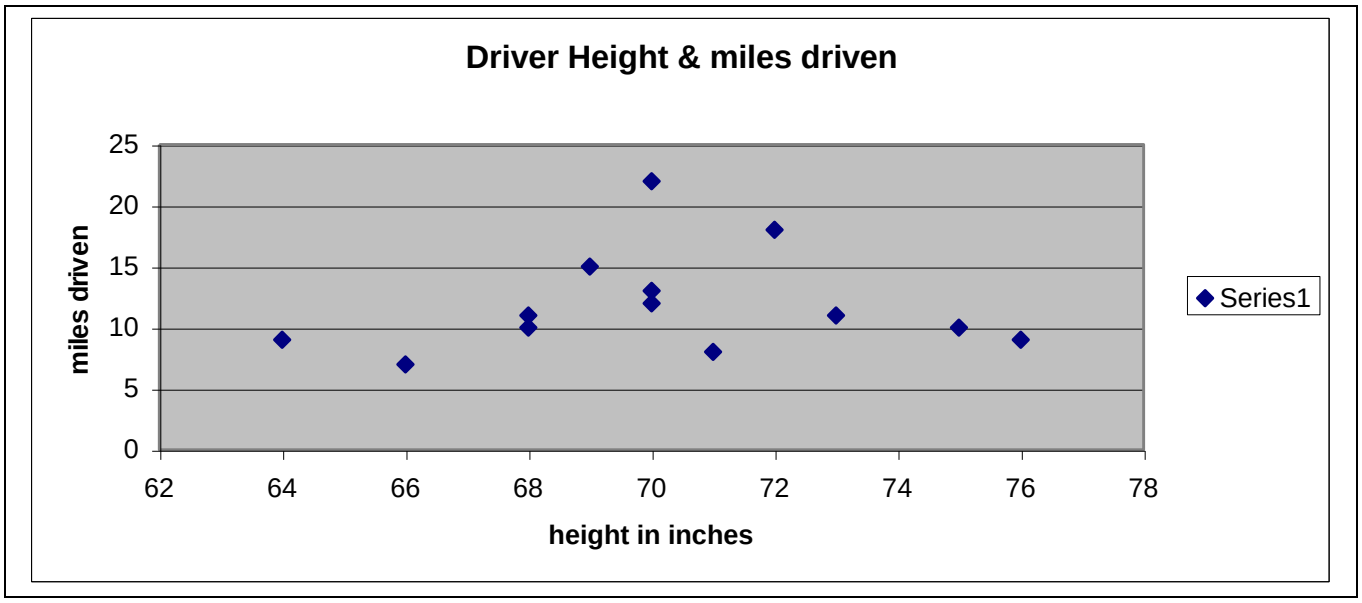
Notice that if we were to draw a regression line representing the shape of the data points, it would move down from left to right (a negative slope).

Here is how a contingency table might look for those data.

Educational level	No children	One child	Two children	More than two children
Less than high school	5%	10%	25%	60%
Stopped after High School	10%	15%	50%	25%
Some college or technical training	15%	25%	45%	15%
University degree	25%	35%	30%	10%

A **zero correlation** is one in which there is **no relationship** between the variables observed. If you choose any two variables at random (say, the price of beans at the local market, and whether or not Brazil will win the World Cup) there will most likely be no relationship between them, a zero correlation. When the correlation is zero, there is no way that we could use a knowledge about a subject's score on one variable to predict what kind of score that subject would have on the other variable.

Here is the way that a zero correlation looks on a bivariate scatterplot: no trend between how tall a driver is and how many miles he drives per year. So, there's no way I could like at the height of a driver and estimate how many miles he drives. Let's look at a scatterplot for hypothetical male drivers between the ages of 25 and 40. Notice that there is no real upward or downward slope to the regression line.



Here is the way that same relationship would be represented with a contingency table.

Height of driver	Drives over 10,000 miles per year	Drives less than 10,000 miles per year
Over 6 feet	28%	72%
Under 6 feet	28%	72%

Notice that the percentages are the same for each group of drivers: it makes no difference whether the driver is tall or short.

QUESTION #4.6: What is the strength of a correlation and how is it measured?

SHORT ANSWER: correlation coefficients range from -1.00 to +1.00, the closer to zero, the weaker they are

The strength of a correlation is expressed by a decimal number ranging from 0.00 to 1.00. A **weak** correlation is a relationship that approximates the zero correlation discussed above. A weak correlation means that there are many exceptions to any observed trend about the relationship between variables. A perfect correlation would have a coefficient of 1.00, meaning that there are no exceptions to the trend: every data point in the sample would be right on the regression line. Correlations close to 1.00 are **strong** because there are few exceptions to the trend.

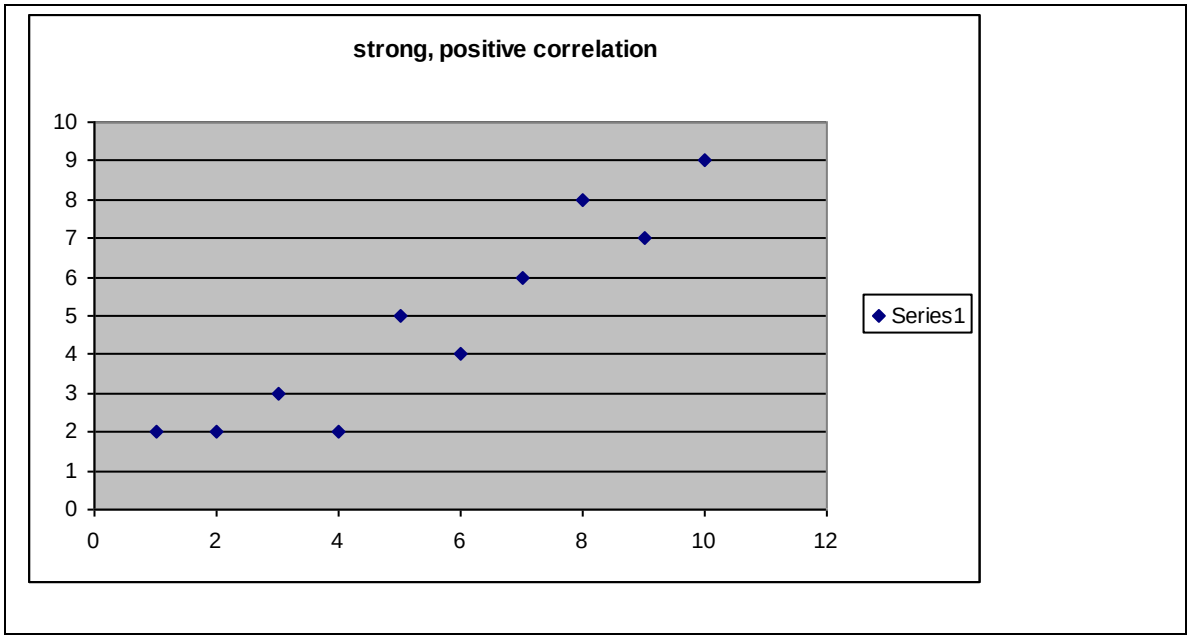
We also put positive and negative signs in front of correlation coefficients to indicate whether the relationship is direct (positive) or inverse (negative), so in that sense correlation coefficients range from -1.00 to +1.00. However, the strength of the correlation is determined by how close it is to zero. The further away from zero (toward either -1.00 or +1.00) the stronger the correlation. Therefore, a correlation of -.60 is stronger than a correlation of +.20. Therefore, don't look at the – or + sign in front of the decimal number to understand its strength, only to understand its direction.

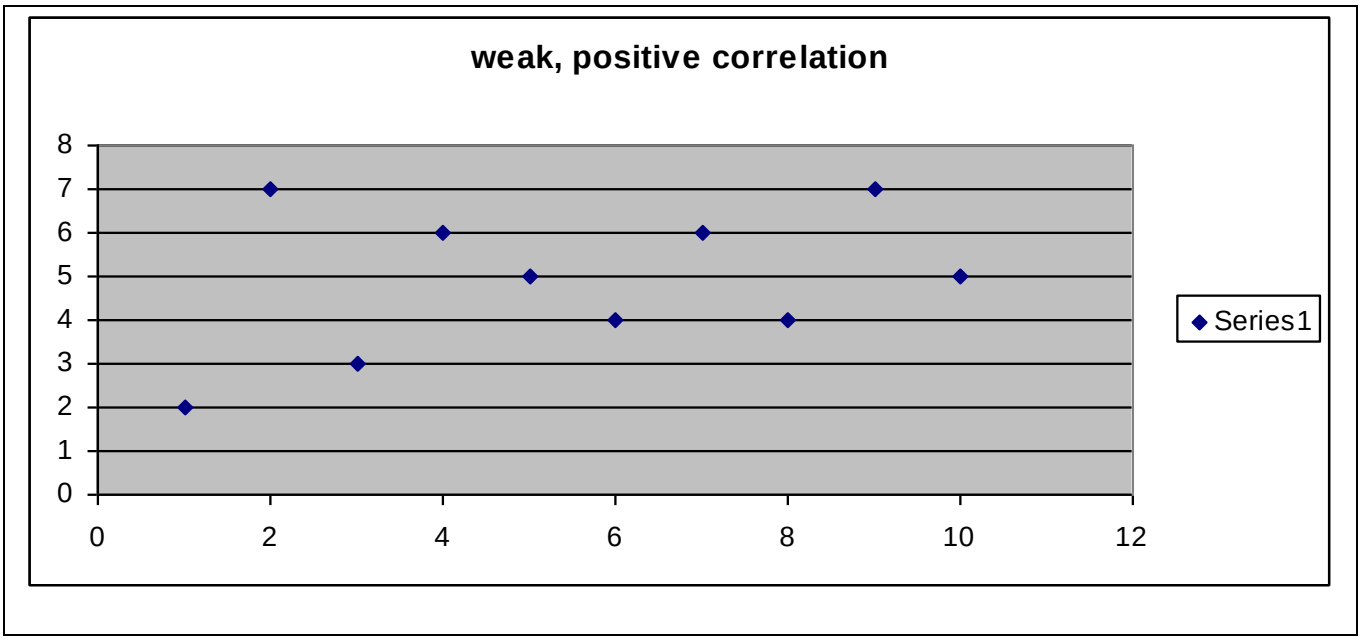
```
 ==============================================================

+1.00          perfect positive        no exceptions to trend


        high    strong positive         few exceptions to trend


+.60 ------------------------------------------------------


                 moderate positive       some exceptions to trend


+.20 ------------------------------------------------------


        low     weak positive           many exceptions to trend


0.00 ------------------------------------------------------
```

```
        low     weak negative           many exceptions to trend

-.20 ---------------------------------------------------------

                 moderate negative       some exceptions to trend

-.60 ---------------------------------------------------------

        high    strong negative         few exceptions to trend

-1.00           perfect negative        no exceptions to trend

==============================================================
```

These cut offs are not hard and fast. In experimental psychology, it is more common to find correlations above .6 (positive or negative) than it is in consumer behavior. This is because much of the research in experimental psychology takes place in the confines of the laboratory where the impact of extraneous independent variables can be controlled, while in the open environment of the workplace, there are many more influences on the subjects' choices, and this creates more exceptions to the trend, hence, weaker correlations.

When dealing with bivariate scatterplots, the strength of the correlation tells us how closely the individual data points approximate a theoretical regression line that expresses the general trend of the data points. (The exact slope and intercept of the regression line are separate calculations.)





In a contingency table, a strong correlation has few exceptions to the trend, and a weak correlation has many exceptions to the trend.

Grouping	Y variable high	Y variable low
X variable high	*Cell* ***A***	*Cell* ***B***
X variable low	*Cell* ***C***	*Cell* ***D***

If the correlation were positive, people who are high on X would be high on Y; and those low on X would be low on Y. We would see most of the subjects stack up in cells A and D. The exceptions would be found in cells B and C. With a strong correlation, there would be very few in B or C. The more subjects who would end up in cells B and C, the weaker the positive correlation.

If the correlation were negative, people who are high on X would be low on Y; and those low on X would be high on Y. We would see most of the subjects stack up in cells B and C. The exceptions would be found in cells A and D. With a strong correlation, there would be very few in A or D. The more subjects who would end up in cells A and D, the weaker the negative correlation.

QUESTION #4.7: What is reliability?

SHORT ANSWER: consistent measurement (do not confuse with validity)

http://www.youtube.com/watch?v=fmqKQBMgB4M

All measurement of data should strive for validity and reliability. **Reliability** means consistency of measurement. This is especially important in standardized psychological tests, but reliability is a criterion for any operational measure of a variable. Imagine a twelve inch ruler made out of elastic instead of wood. One carpenter might measure a board as being 5 inches, but another carpenter using the same ruler might stretch it a little less and determine that the board was 6 inches. This kind of inconsistency is not tolerable in science. When one marketing researcher reports that a consumer is "upper middle class" does that mean the same thing as another researcher might infer?

Correlation coefficients are a useful way to determine just how reliable measurements are. Reliable tests have high, positive correlation coefficients. Most of the subjects end up in cells A and D where there is agreement between the first measure of the test and the second measure, and there are very few disagreements in cells B and C.

	Measure #2 **high**	Measure #2 **low**
Measure #1 **high**	*Cell* ***A*** AGREE	*Cell* ***B*** DISAGREE
Measure #1 **low**	*Cell* ***C*** DISAGREE	*Cell* ***D*** AGREE

One form of reliability is **inter-rater**. Two different raters (e.g., judges, interviewers, diagnosticians, observers) evaluate the same subjects on the same variable. For example, the first measure might be the judgment of one sales professional watching a video of a sales presentation given to a customer, and then that rater would have to assess the customer's level of interest in a product. The second measure might be the judgment of another sales professional who is watching the same video. If both raters agree that the customer has a high level of interest in purchasing the product, we would categorize that subject in cell A; if both raters agree that there is little interest, then the subject goes in cell D; and if the raters disagree, then the subject is categorized in cell B or cell C. If we looked at enough videos, we would be able to express reliability as a percent agreement or as a correlation coefficient. Remember that percents would have to be a lot higher, because the percents could not be negative, and a truly random pattern of agreements equaling disagreements would show a 50% rate but a 0.00 correlation.

Another form of reliability is **test-retest**. The subject is given the same test twice to see if he scores consistently. Suppose this is a test of personality that classifies subjects as introverts or extraverts. If the test is reliable, we should not see a subject looking like an introvert this week and looking like an extravert next week. This kind of reliability may be less important in consumer behavior, because we understand that situations change rapidly. A consumer may express a high interest in purchasing a new automobile today, but not on retest next week. Perhaps the diminished interest is due to losing his job, or even the fact that he has already purchased a new vehicle in the meantime.

Other forms of reliability include **alternate (parallel) forms** in which there might be two slightly different versions of the same test, and **internal** reliability in which we look at the different parts of a test and make sure that each part is really measuring the same thing as the other parts of the test. These are also very important in evaluating paper and pencil psychological tests with numerous items, but they are less pertinent to some of the variables measured in consumer behavior (e.g., purchase decisions and attitudes about products).

Establishing the reliability of a test	
Type of reliability	**Research involved**
Test-retest	Give the test twice to each subject; Correlate first administration to the second
Inter-rater	Have two judges evaluate each subject; Correlate the first ratings to the second

Alternate form	Give two versions of the test to each subject; Correlate the first version to the second
Internal	Give the entire test to each subject; Correlate one part of the test to the rest of it

QUESTION #4.7: What is validity?

SHORT ANSWER: measuring what we say we are measuring (not some other variable that was easier to measure)

http://www.youtube.com/watch?v=swwnbNurmTo

Validity means that a measurement actually measures the variable that it claims to measure. This is especially important in standardized psychological tests, but validity is a major criterion for any operational measure of a variable. Validity and reliability are both important for psychological measures, but they are not the same thing. Imagine that you need to weigh a brick, and someone brings out a ruler. That ruler may measure very reliably (consistently) but what it measures is distance, not what we need to measure now, which is weight. One of the biggest problems in psychological research is using the wrong tests to measure variables.

Here are examples of some questions used by marketing researchers, and some of the reasons why we might want to doubt the validity of those items.

QUESTION: How long have you been single?

VARIABLE RESEARCHER INTENDED TO MEASURE: Time since the subject ended the last serious, committed relationship.

VARIABLE THAT THE QUESTION COULD BE MEASURING: Time since a formal divorce was granted.

QUESTION: What is your monthly take home pay?

VARIABLE RESEARCHER INTENDED TO MEASURE: Income, wealth, or social class

VARIABLE THAT THE QUESTION COULD BE MEASURING: The size of the last paycheck, which may not be a typical measure of annual salary, other sources of income, or wealth from other sources (e.g., trust funds, pensions, settlements, investments)

QUESTION: What was the brand of the last automobile you purchased?

VARIABLE RESEARCHER INTENDED TO MEASURE: Whether the consumer drives a luxury car, sports car, or practical vehicle

VARIABLE THAT THE QUESTION COULD BE MEASURING: Just the last vehicle purchased, not what the subject leases or also owns (e.g., the last vehicle I purchased was a cheap little truck for my ranch, but I also own a classic luxury car, a classic sports car, and other more expensive trucks).

QUESTION: Are you Catholic?

VARIABLE RESEARCHER INTENDED TO MEASURE: Whether the subject grew up in a Catholic household

VARIABLE THAT THE QUESTION COULD BE MEASURING: Whether the subject converted to Catholicism when he married

The art of phrasing questions is beyond the scope of this course, but I refer you to my book

http://www.amazon.com/s/ref=nb_sb_noss?url=search-alias%3Dstripbooks&field-keywords=brink+questionnaires

Correlation coefficients are useful in describing how valid a test or question is. In order to validate, we must correlate it to some pre-established standard measure of that variable. (Just like if we wanted to see if a watch kept the correct time, we would have to compare it with the official government clock.)

In clinical psychology, there is an accepted standard for the diagnosis of each mental disorder, the ***Diagnostic and Statistical Manual*** (**DSM**) of the American Psychiatric Association. In consumer behavior research, there may be less agreement about absolute standards for the measurement of variables.

Even if you do not know the precise correlation coefficient describing the validity of your measure of a variable, try to get an idea what the pattern of its errors might be. Does it have more false positives or false negatives? Both of these reduce validity, but they have different implications for the researcher.

	Subject is actually **high** on that variable	Subject is actually **low** on that variable
Test shows subject scoring	*Cell* ***A*** **test is right**	*Cell* ***B*** False positive

high		
Test shows subject scoring **low**	*Cell* ***C*** False negative	*Cell* ***D*** **test is right**

False negatives are persons who score low on an assessment, but actually score high on the variable. **False positives** are persons who scored high on the assessment, even though they are low on the variable.

For example, suppose I am selling comprehensive automobile insurance. The variable I really need to measure is whether the subject had a need for such insurance. Support I ask, "How many vehicles do you own or lease?" Almost everyone who answers "none" can be placed in Cell D. There are probably no false negatives in cell C. Although many of the people who answer that they own at least one vehicle will be potential customers who need comprehensive insurance, others will be false positives: probably because they already have such insurance, or don't figure that the size of the risk matches the size of the premium.

In marketing here is the impact of false negatives and false positives. If you decide to target your advertising budget to those who score high on the assessment, you are betting on a high correlation between the assessment and the real variable (interest in purchasing the product). False negatives represent potential customers that you are missing. False positives represent advertising dollars wasted on people who will not become your customers. For example, I am frequently driving the freeways of southern California, where there are numerous billboards for lap band surgery. I'm not fat enough to consider something like that, so the billboard ads are wasted on me. I see the ads, but I don't need them. I'm a false positive.

- -

GRAMMAR LESSON: Do not use the words *accurate* or *accuracy* in this class. Figure out which of the following concepts you want to convey: precision, reliability or validity.

- -

QUESTION #4.8: How do we actually get the data for quantitative research?

SHORT ANSWER: direct observations, questionnaires, archives

The first technique envisioned by most new researchers is to come up with a questionnaire. The types of questions asked on questionnaires differ from those asked in interviews and focus groups (which are qualitative research techniques). The latter must use open-ended questions

evoking rich, narrative level data. The former must use questions, and specified response formats, that direct subjects to give quantifiable answers. Here are examples of such quantifiable response formats at each level of measurement.

BINARY NOMINAL (two categories)

Yes / No

Before / After

Pass / Fail

Male / Female

MULTIPLE NOMINAL (more than two categories)

Which product? Brand X / Brand Y / Brand Z

Which denomination? Catholic / Protestant / Muslim / Jewish

Which city? Mexico City / other city over 100,000 / small town / rural

Which career? Homemaker / student / farmer / factor / service / retail / professional

ORDINAL (ranks, or levels showing more or less of the variable)

How long? Under a month / 1 –12 months / 1 –3 years / over 3 years

How often? always / frequently / about half the time / rarely / never

How often? daily / weekly / monthly / once or twice a year / never

How intense? extremely / very / somewhat / slightly / not at all

Would you? definitely / probably / possibly / no way

Do you? strongly agree / mostly agree / mostly disagree / strongly disagree

Are you? very pleased / somewhat pleased / somewhat displeased / very displeased

How well? Excellent / good / fair / poor

How old? under 20, 20-29, 30-39, 40-49, 50-59, 60+

Social class? wealthy / financially secure / solid middle / working / poor

Questions can also be asked in hopes of getting a clear ratio response, but outside of a variable like age, or the number of children they have, where consumers do have the exact figure, it may be better to use an ordinal level estimate. For example, most Mexican consumers would not know exactly how many times they have been in an Oxxo convenience store last week. It would be better to use a response format such as

```
None / once or twice / three to seven times / at least 8 times
```

The quickest way to get data from subjects on variables is to observe those subjects engage in certain actions, where those actions have been agreed upon as operational definitions of the variable to be studied. This can be done in a laboratory situation or "the field" of the marketplace. If I want to find out the percentage of persons in my building who drive luxury cars, I walk around the parking lot and count. If I want to see if a given store attracts mostly female customers, I count the numbers of males and females going coming out. If I want to see if an ad campaign attracted more customers I count the number of customers these week and compare it to the number the previous week (before the ad campaign).

Now, with high tech scanning of the brain (e.g., functional magnetic resonant imagery) we can get inside the person's head and see if an advertisement for a project generates interest and intention to purchase.

Sometimes we cannot directly count the people, their actions, or their attitudes that really constitutes the dependent variable. So, we must look for a consequence of such actions as a **trace** from some other variable. For example, I did not see any birds atop my neighbor's fence, but I know he had quite a few last week, because I see the trace (the bird droppings on the fence posts). I did not observe any people at the football game shop at Startbucks before the game, but a found a trace that many had been there (the empty Starbucks cups left in the trashcans). I didn't ask the attendees of the museum's exhibits which one they liked the best, but I can infer that by looking at the trace of where they stopped and stood (the wear on the carpet in front of some of the exhibits).

The easiest way to get data from consumers is through **archives**. These are data collections that someone else (usually an institution) has already put together. All you have to do is gain access and make sure that you will not violate any ethical guidelines. Such files might include patient records, prison records, employment records, student records, military records, or job applications. For consumer behavior, the most valuable records will be product purchases, warrantees, credit ratings, payment records, complaints, service calls.

Since the internet, we live in a world of **big data** in which we have automatic tracking of where a computer browser has been (traces of consumers' interests) and where cell phones have been (by pings on different towers) and how credit cards and club cards have been used for

purchases. The branch of research trying to come up with useful interpretations of such big data is known as **analytics**.

Wherever you are, you can be observing people, their traces, their archives or their electronic activity, and you can be doing consumer behavior research without a questionnaire. If you can count those variables on one of the scales we have mentioned, you are doing quantitative research.

QUESTION #4.9: How do we design the research in order to test our alternate hypotheses?

SHORT ANSWER: look for something to compare or correlate

http://www.youtube.com/watch?v=FpHee7l1cZg

There are four basic designs for quantitative research. One uses correlations and the other three use comparisons.

A **correlational** design starts with a null hypothesis saying that there is no correlation between two variables. The alternate hypothesis would state that there is a correlation (and should specify whether it is negative or positive). For example: the higher a consumer's income, the fewer the number of persons per room in the household: a negative correlation.

Once we know how each variable was measured, we know what correlation coefficient or inferential statistic to test the hypothesis. In general, we should use the most powerful test (i.e., the one most likely to show a significant difference). However, powerful tests are usually based upon **parametric** assumptions that the variables in this sample come from populations in which they are distributed like a normally distributed bell curve: a symmetrical distribution in which most of the cases are close to the mean. If these assumptions cannot be met, then the more robust (resistant to Type I error) should be used. These nonparametrics include percents, median, mode, and those tends highlighted in red in the table below.

	Binary nominal	Multiple nominal	Ordinal or skewed	Interval or ratio (normally distributed)
Binary nominal	Chi square or Fisher Exact	Chi square or Kolmogorov-Smirnov	Mann-Whitney or Kolmogorov-Smirnov	t-test for independent groups
Multiple nominal	Chi square or Kolmogorov-Smirnov	Chi square	Kruskal-Wallis	One way ANOVA
Ordinal or skewed	Mann-Whitney or Kolmogorov-	Kruskal-Wallis	Spearman	Spearman

	Smirnov			
Interval or ratio (normally distributed)	t-test for independent groups	One way ANOVA	Spearman	Pearson

A **sample vs. norms** design starts with a null hypothesis saying that there is no real difference between this particular sample and the norms (usually coming from a population or coming from pure chance variation). The alternate hypothesis would state that our sample should be higher (or lower) on a particular variable, compared to those norms. For example, I would predict a prosperous city like Toluca (my sample) would have more luxury car dealerships than the norm for all of Mexico. I could find out how many luxury car dealerships there are throughout the country of about a hundred million, and calculate how many should be expected for a city of a half million and see if Toluca significantly exceeds that figure.

Sample vs. norms designs are weak in a number of ways. One problem in conducting them is that we must have the population norms to begin with, otherwise we cannot do the study. Another is that our sample probably differs in several ways from the rest of the population. Toluca differs from most of Mexico, not just because it is more prosperous, but it is also closer to Mexico City, a state capital, has the highest elevation (almost three thousand meters) and is the coldest.

A sample vs, norms design based on a questionnaire could use a national poll for the variable, but the exact same wording of the question and response format would have to be used. I often give my students questions from Gallup polls. When my students' answers differ significantly from those national polls, is it because they are students? Because they live in southern California? Because the poll data are from back in 2009?

Binary nominal	Multiple nominal	Ordinal or skewed	Interval or ratio (normally distributed)
Chi square or binomial or test of proportions	Chi square or Kolmogorov-Smirnov	Kolmogorov-Smirnov	t-test for one sample vs. population

A **repeated measures** compares the entire sample to itself. This is sometimes called a dependent sample or matched pairs design. The null hypothesis would state that the first measure is similar to the subsequent measure(s). The alternate hypothesis would say that the first measure is going to be higher (or lower) than the subsequent measures. One example would be asking a sample of consumers for their ratings of three U.S. brands of automobiles: Ford, Chevrolet and Dodge. Each subject gives a rating of each kind of car. Another example would be asking smart phone users how confusing they find their phones to be: ask each new

owner during the first week with the smart phone, and then a month later. I would hypothesize that the level of confusion would go down. Another example would be to ask married couples in their thirties how often they watch reality programs. We get one answer from the wife and another answer from the husband, and see which gender admits to watching more of these programs.

One problem with repeated measures designs is that we have to record the data in such a way that we know which first measure and which subsequent measure belong to the same subject (or which husband's answer belongs to which wife's answer). As the name of the design implies, we have to match the pairs. Another problem is that the longer we wait between measures, the more than can happen to distort them. If we are measuring performance, that generally increases with practice (but in short time frames that could decrease with fatigue or boredom). If we are measuring a disease, we have to factor in the natural course of the disorder.

	Binary nominal	Multiple nominal	Ordinal or skewed	Interval or ratio (normally distributed)
Two measures	McNemar Chi square or binomial	Chi square	Wilcoxon	t-test for dependent groups
Three or more measures	Chi square	Chi square	Friedman	repeated ANOVA

Probably the best design for most purposes are **separate groups**. We separate our sample into two (or more) groups and compare the groups in terms of a dependent variable. The null hypothesis is that the two groups do not differ. The alternate hypothesis is that one group is much higher (or lower) than the other. An example would be to hypothesize that consumers who saw the advertisement would be more interested in purchasing the product, compared to the consumers who did not see the advertisement.

	Binary nominal	Multiple nominal	Ordinal or skewed	Interval or ratio (normally distributed)
Two groups	Chi square or Fisher Exact	Chi square or Kolmogorov-Smirnov	Mann-Whitney or Kolmogorov-Smirnov	t-test for independent groups
Three or more groups	Chi square or Kolmogorov-Smirnov	Chi square	Kruskal-Wallis	One way ANOVA

QUESTION #4.9: So how does this prove that the independent variable caused the dependent variable?

SHORT ANSWER: not necessarily

In quantitative research, there are four things that we can do with a variable: measure, control, randomize or manipulate. If we do not do one of those things with a potential individual variable, it can become a **confounding** variable and distort our understanding of what is really causing our dependent variable by creating a **spurious** correlation.

We can **measure** variables on one of the scales discussed above. Dependent variables must be measured, but independent variables can be measured, or they can be dealt with in one of the other ways.

We could **control** a variable by changing it into a constant. If we just sample men, we have controlled the variable of gender. If we just sample people with a college degree, we have controlled the variable of education. If we just sample people over age 65, we have controlled the variable of age. If we just sample people in our city, we have controlled the variable of geography. Each of these controls diminishes confounding variables in our study, but may make our results less generalizable to populations that don't have those controls.

We can **manipulate** an independent variable if we, the researchers, intentionally make it present or higher. This is the key feature of an experiment. Most quantitative research would not meet this definition. In order to be called an **experiment**, quantitative research must manipulate at least one independent variable. For example, if I want to demonstrate that a new pain reliever is effective, I would give some consumers a new medicine to see if they like it. On the other hand, if I just ask consumers which pain relievers they use, that is their choice, and so that would be a dependent variable, and such research would not be an experiment.

http://www.youtube.com/watch?v=hTEYaJPL-Zg

Few experiments use sample vs. norms or correlational designs. Some use repeated measures designs, but the methodological problems discussed above lead to questions about these before/after treatments. The most common kind of experiments are separate groups designs. If we told the participants that there would be two groups, one seeing this kind of advertisement and the other group seeing the other kind of advertisement, and let each subject choose his or her grouping, that would not be an experiment, because grouping would no longer be independent of the subjects' choices or preferences. Grouping would just be another variable determined by the subjects' choices.

The best way to group subjects in an experiment is through **random** assignment. Random does not imply haphazard or careless. When we select random samples, we are saying that each member of the population has an equal chance of being selected into the sample. When we then take a sample and randomly assign them to different groups, each subject has an equal chance

of winding up in the experimental group (compared to all the other subjects). Such random assignment allows us to assume that all kinds of background factors (e.g., age, gender, income, previous experiences, heredity) have been roughly equalized between the two groups. If our randomization process is well done, and if our sample size is large enough, this eliminates confounding variables.

A good experiment also works on controlling the variable of expectation. In clinical trials, for example, this is done by giving the control group a **placebo**, a fake pill or other treatment. In a double-blind placebo study neither the subjects nor the researchers who are recording the patients' progress initially know which subjects are receiving the placebo and which are getting the real treatment.

In practice, we cannot always randomly assign each individual subject to one group (experimental, getting the treatment) and control (not getting treatment). We may have to take two existing groups (e.g., two different classrooms of students) and call one the experimental group and the other control. This **quasi** experiment is not as good at randomizing all potentially confounding variables.

In order for us to come to a causal conclusion about our research (e.g., the presence of this independent variable causes the dependent variable to be higher) we must have three things.

First, we must have statistical significance. If we cannot reject the null, then we don't need a causal explanation, we have the null as an explanation: the pattern of results can be explained by pure chance.

Second, we must find the results in the direction predicted by the alternate causal hypothesis. If we claim that the new medication helps to reduce pain, then the subjects in the experimental group receiving the new medication must report significantly lower (not higher) levels of pain. Otherwise, we may have proved that the independent variable made the dependent variable worse.

Third, we have to be able to account for other possible factors. To the extent that an experiment (or other quantitative research) has effectively measured, controlled, or randomized the impact of other potential causes, then we can say that *this* independent variable is what is causing the dependent variable to change.

UNIT 4: QUANTITATIVE METHODS

flashcards & matching2

http://www.quia.com/jg/2522647.html

jumbled words game2

http://www.quia.com/jw/469850.html

millionaire game2

http://www.quia.com/rr/938696.html

summary paragraph2

http://www.quia.com/cz/466066.html

UNIT 4 TERMS: quantitative research methods

ANALYTICS: (also known as "big data" and "data mining") involves accessing and analyzing digitally stored data from large populations and many variables

ARCHIVES: files where raw data have been stored; an archival study applies descriptive and inferential statistics to data coming from student records, employment records, patient records, job applications, customer registrations, records of customer complaints, etc.

BAR: a chart that can visually depict any scale of measurement; each bar represents a different group or measure, and the length of the bar can represent central tendency or absolute values

BIG DATA: (also known as "analytics" and "data mining") involves accessing and analyzing digitally stored data from large populations and many variables

CAUSATION: the inference of a cause and effect relation; best established by an experiment

CONFOUNDING: a variable which could potentially influence the dependent variable, but the researcher has not controlled, manipulated, measured, or randomized that variable; the present of confounding variables can lead to spurious correlations

CONTINGENCY: a table using rows and columns to depict cross tabulation of two variables; one variable becomes the rows, and the other variable becomes the columns

CONTROL: controlling a variable means to convert it to a constant within the research study

CONTROL GROUP: in an experiment, the group that does not receive the experimental treatment, but just serves as a comparison group; in clinical trials the control group usually receives a placebo; do not call it *controlled*

CORRELATION: a relationship between variables, usually symbolized by the letter r

DATA MINING: (also known as "big data" and "analytics") involves accessing and analyzing digitally stored data from large populations and many variables

DESCRIPTIVE: statistics describing the central tendency (e.g., mean, median, mode, percent) or dispersion (e.g., range, variance, standard deviation) of a variable

DIRECT: a positive correlation: when one variable is high, so is the other; when one variable is low, so is the other

E: a symbol in scientific notation telling us to move the decimal point, e.g., 4.6E-02 = .046

EXCEL: a Microsoft Office spreadsheet program which provides a good way to save quantitative data; most versions of Excel will perform some statistical analysis

EXPERIMENT: researcher manipulates an independent variable; best technique for identifying cause and effect

FALSE NEGATIVE: a subject who scores low on a test, when his actual measure on the variable is high

FALSE POSITIVE: a subject who scores high on a test, when his actual measure on the variable is low

FIELD COUNT: a (usually unobstrusive) survey in which the behavior of the subjects is simply observed and quantified in some way, without need for interaction such as a questionnaire

INDEPENDENT VARIABLE: causes or influences upon behavior; the variable manipulated in an experiment; background factors or external stimuli

INFERENTIAL STATISTICS: calculating or estimating the probability of the null hypothesis

INTER-RATER RELIABILITY: when subjects are evaluated by two different raters (e.g., interviewers, judges, diagnosticians) the ratings are said to have inter-rater reliability when there is agreement between the two examiners: the same subjects who are rated highly by one examiner are also rated highly by the other examiner; and subjects who are rated poorly by one examiner are rated poorly by the other examiner

INTERVAL: a scale of measurement in which each subject's score on a variable is represented by a number, such that the distance between the numbers is fixed (e.g., the difference between a 3 and a 4 is the same as the difference between a 5 and a 6); examples would be Celsius temperature and IQ test scores

INVERSE: a negative correlation: when variable is high, the other is low

LIKERT: an ordinal scale measuring the subject's level of agreement (e.g., completely agree / mostly agree / mostly disagree / completely disagree)

LINE: a graphical depiction, usually of one variable over time; the horizontal axis represents a time sequence, the vertical axis represents levels of the variable

MANIPULATE: when a researcher intentionally varies the level of an independent variable (usually by randomly assigning subjects to different groups, and then treating those groups differently)

MAXIMUM: the highest score in a data set

MEAN: usually this refers to a measure of central tendency (average) which is calculated by adding up all the scores in a data set and then dividing by the number of scores

MEASURE: quantitative measurement of a variable means using nominal, ordinal, interval or ratio scaling to represent the measure with a number

MEDIAN: if we arrange the cases of a data set, highest to lowest on a variable, the score attained by the middle case is the median; this is the best measure of central tendency for variables that are ordinally scaled or who have a skewed distribution

MINIMUM: the lowest score in a data set

MODE: the most frequent score in a data set

N: with a survey or experiment, the letter n indicates the number of subjects in a sample or group

NEGATIVE CORRELATION: an inverse relationship; when one variable goes up, the other goes down; do not say "negative" if you mean bad or unfavorable

NOMINAL: a scale of measuring a variable by categorizing each case into a specific category; binary nominal scales have only two categories (e.g., yes/no; pass/fail; male/female, experimental/control)

NONPARAMETRIC: statistical tests which make no assumptions about a variable's distribution on an interval or ratio scale; nonparametric tests include percent, mean, median, mode, binomial, Chi Square, Fisher Exact, Mann-Whitney, Kolmogorov-Smirnov, Wilcoxon, Kruskal-Wallis, Friedman; nonparametric tests resist Type I error

NOT SIGNIFICANT: $p > .05$, do not reject the null hypothesis, nothing was proved

NULL HYPOTHESIS: attributing the relationship between variables (or difference between groups) to random variation (pure chance, luck) rather than an underlying causal relationship (the null hypothesis should be rejected when $p < .05$)

ORDINAL: a quantitative scale of measurement using ranks or comparative levels (e.g., a Likert scale)

P: stands for probability (of the null hypothesis); $p < .05$ is fairly significant; $p < .01$ is good; $p < .001$ is excellent; $p > .05$ is not significant

PARAMETRIC: statistical tests which assume that a variable's distribution on an interval or ratio scale approximates that of the normal (i.e., Gauss, bell) curve, with most of the cases being close to the mean, and the distribution is symmetrical; examples of parametric tests would be mean, standard deviation, t test, ANOVA, Pearson coeffieicient

PARTICIPANTS: new term for subjects, the organisms participating in psychological research

PERCENT: a measure of a variable: 100 times part / whole; these simple percents are always positive and cannot exceed 100%.

PERCENT CHANGE: a measure of proportionate change: 100 times increase / start; increases have positive percent change; decreases have a negative percent change; negative percent change cannot exceed –100%, but there is no limit on positive percent change

PIE: a chart depicting a variable's distribution by having different slices for each category and the size of the slice indicates its percentage of the whole

PLACEBO: a fake treatment which patients believe in; often given to the control group in an experiment

PREDICTOR: when there is a strong correlation between two variables, the predictor variable is used to predict the level of the other (criterion) variable that may be outcome or performance; the predictor variable may be a background factor (IV) or other information about a consumer's behavior (DV)

QUASI: an "almost experiment" in which separate groups were compared, but grouping was not fully randomized

QUESTIONNAIRE: a series of questions in which the response patterns can be quantified on nominal, ordinal, interval or ratio scales

R: symbol for correlation coefficient (especially using Pearson's product moment formula)

RANDOM: in research, selection or assignment that is left to pure chance (such as a lottery)

RATIO: a scale of measurement in which the numbers represent proportionate differences, and there is a true zero point at which a subject has none of the variable

being studied; ratio scaling can be used with time, distance, area, volume, events (e.g., accidents), or units sold

RELIABILITY: when a test measures consistently from testing to retesting, item to item, examiner to examiner

REPEATED MEASURES: more than one measure of the same variable is taken from the same subject; e.g., matched pairs, dependent samples, before & after; vulnerable to some methodological weaknesses such as practice, fatigue, and the researcher's need to match the data

REPRESENTATIVE SAMPLE: a sample should be similar to the population on relevant background variables (e.g., age)

SAMPLE: all subjects actually observed in the research (e.g., all people filling out the questionnaire; all rats running the maze)

SAMPLE VS. NORMS: research design in which the entire sample is being compared with some external norm (e.g., the average in the population); vulnerable to some methodological weaknesses, especially the presence of many confounding variables

SCATTERPLOT: a bivariate graph for demonstrating correlations

SEPARATE GROUPS: research in which the sample is divided into separate groups which are then compared on a dependent variable; in a true experiment, the grouping is done by random assignment, and then the groups are treated differently (which constitutes the manipulation of the independent variable)

SIGNIFICANCE: when $p < .05$, reject the null hypothesis because the data are statistically significant

SKEW: when data are not distributed symmetrically about the mean, but there are some extremely high (or extremely low) outliers; if a distribution is skewed, nonparametric tests are more appropriate

SPURIOUS: a correlation between collateral effects; neither variable causes the other, but each can predict the other

STANDARD DEVIATION: a parametric measure of the dispersion in a distribution

STRONG: a high correlation close to +1.00 or -1.00; few cases if any are exceptions to a dominant trend

SUBJECT: the person or animal about whom we have data; the patient in the case study, the rats in the experiment

SURVEY: research that measures variables using a large sample (e.g., field count, questionnaire, archival data)

TYPE I: error in which we reject the null hypothesis prematurely, when in reality, there is no causal relationship between the variables

TYPE II: error in which we fail to reject the null hypothesis, even though there is in reality a causal relationship between the variables

VALIDITY: when a test measures what it says it measures, what it claims to measure, what it purports to measure

VARIABLE: something that varies and can be measured empirically; independent or dependent; opposite of a constant

WEAK CORRELATION: a low correlation close to 0.00; little or no relationship between the variables

X: horizontal coordinate on scatterplot diagram depicting the correlation between two variables; usually X represents the independent or predictor variable

Y: vertical coordinate on scatterplot diagram depicting the correlation between two variables; usually Y represents the dependent or criterion variable

ZERO CORRELATION: no relationship between two variables