

IND 503 M2.2 The Scientific Method

The scientific method is used to determine fact and truth. Doing so is more difficult than might be immediately evident.

The scientific method is used to determine fact and truth. Doing so is more difficult than might be immediately evident as what we observe is very often influenced by factors we *don't* see. Those observations, even if true, might not have broader applicability.

Evidence collected in an unstructured, uncontrolled manner, such as what we experience in our daily lives, is anecdotal. Anecdotal evidence is quite weak.

Evidence collected in a structured, controlled manner, such as what we collect in a formal research study, is called empirical. Empirical evidence is far stronger yet still can be erroneous if the data are not collected in a sound manner.

The lesson is to not believe what we read until we assess the soundness of the methods used to collect data, determine that the data were analyzed in a credible manner, and are convinced of the logic of the interpretation.

Quantitative vs. Qualitative

The important concepts for now are *qualitative* and *quantitative*, *inductive* and *deductive*.

Research methods are generally categorized as qualitative and quantitative. Combined, that process reflects the scientific method. Let's explore...

Qualitative

Qualitative methods are exploratory and use [inductive reasoning](#). We make observations, describe them in words, identify patterns, and develop a theory.

[Deductive vs. Inductive Reasoning](#)

Quantitative

Video

[How Sherlock Changed the World - Inductive Reasoning](#)

Time: 0:59 mins

(Click for [closed-captioning](#))

Quantitative methods are confirmatory and use [deductive reasoning](#).

Summary

Knowledge is developed through a cycle in which qualitative methods use inductive reasoning to develop theory. From there, quantitative methods use deductive reasoning to confirm theory.

More simply, we make observations, describe them in words, identify patterns, and develop a theory. That's qualitative and inductive. Later, we test theory by collecting relevant data and subjecting it to statistical analysis.

Hypotheses

What are hypotheses?

The first step in the scientific method is to create a specific statement that you are trying to prove. These statements are called hypotheses and set the bar for what we are trying to accomplish with our data collection. Let's say that you are a human resource manager for a large firm which has been experiencing a high volume of sexual harassment complaints in the last year. You contract with an outside company to conduct your annual sexual harassment training, and now need to evaluate if the training was effective.

Begin with your variable of interest, also known as the independent variable. In this scenario, we are interested in the effectiveness of the sexual harassment training. Now, what are we hoping the sexual harassment training will accomplish? A reduction in the number of sexual harassment complaints would be the goal. The number of sexual harassment complaints is the variable we are measuring, which we hope will be impacted by the independent variable. The variable being measured is known as the dependent variable (DV), as changes in the variable should be "dependent" upon our independent variable (IV).

So we now have the three pieces we need for hypothesis creation, our IV, DV, and in this case a direction. Let's take a stab at the hypothesis we are trying to prove:

The new sexual harassment training significantly reduced the number of sexual harassment complaints.

Note that we have all three components in our hypothesis. The hypothesis we are trying to prove is known as the research or alternative hypothesis. We can either say our hypothesis is true, or that it is not true. The next step is to create our null hypothesis. This hypothesis states that our IV had NO impact upon our DV. By creating our research hypothesis first, it is easy to negate to create our null:

The new sexual harassment training did not significantly reduce the number of sexual harassment complaints.

Although null hypotheses might have little function in the real world, in the scientific method they are integral. In statistical terms, we always either reject or fail to reject the null hypothesis. If we fail to reject the null hypothesis, this means our IV was ineffective. In this case, if the number of complaints was not significantly reduced, this means our training was not effective. If the null hypothesis is rejected, this means our training was effective in significantly reducing the DV, or number of harassment complaints. Let's take a look at how hypotheses play a role in testing:

Video

[Understanding Hypothesis Testing, P-Value, T-Test](#)

Time: 7:37 mins

Closed Captioned

Validity & Reliability

Race and Crime?

In the early to mid 1900s, sociologists and criminologists correlated race and crime, noting a strong correlation between minorities and crime. They concluded that race is a *predictor* of criminal behavior. However, later researchers determined that if they held

socio-economic levels constant, the relationship between race and crime disappeared. Instead, the true relationship was between socio-economic level and crime, not between race and crime.

Poorer people tend to commit more crime than wealthier people and minorities tend to be poorer. Poor people commit more crime, irrespective of race. That's quite an error. The problem is called spuriousness, which is produced by a confounding variable - a variable that influences our research but of which we are unaware.

Validity

Have you ever completed a survey question and asked yourself "What was the purpose of that question"? Survey questions and other types of information-gathering instruments must be carefully constructed in order to measure what they SAY they are measuring. Let's take an example. Many of us have the dreaded "bathroom scale" in our homes and measure our weight on a regular basis. Let's say that you weigh yourself in the morning and the scale says you weigh 160 pounds. That afternoon you visit your doctor, and the scale at the office says you weigh 150 pounds. Because your scale is not accurate, it is not a valid measure of your actual weight.

Validity pertains to truthfulness. Does the test measure what it purports to measure? Many threats to validity exist and researchers must design studies with sufficient control to ensure that variables or influences not under study do not affect the results.

Be aware that validity is not so often clear cut. Let's say that you want to administer a personality test to your employees. How exactly do you create questions that will specifically measure personality traits such as extraversion and agreeableness? Valid instruments are created and assessed using complex methods and data analysis, typically on samples of individuals.

It is important to recognize that the operationalization (or precise definition) of variables we measure with instruments must be considered. Whenever possible, it is advised to utilize questions that have been proven valid (and reliable) in the past. Most research articles will give measures of validity and reliability for the data collection instruments they utilize.

Reliability

Reliability is the measure of how stable, dependable, trustworthy, and consistent a test is in measuring the *same thing each time* (Worthen et al., 1993) or the degree of consistency between two measures of the same thing (Mehrens and Lehman, 1987). For us to consider data credible, they must be collected in a valid and reliable manner.

Reliability is a term that can be thought of as consistency. Let's return to our "bathroom scale" example. Let's say that you weigh yourself at 9:00 am using your scale and you

weigh 160 pounds. Later in the day, say after lunch, you weigh yourself again and you weigh 170 pounds. Perplexed, you wait until the next morning and weigh yourself again and the scale reads 150 pounds. The variability of your weights is not possible within that short of a time frame and it appears your scale is unreliable. If you were to weigh yourself at the different times and the scale gave you the same readings, then it would be reliable. Given the fact that we have established that our scale is 10 pounds off, it is not valid, BUT if the weights are consistent, then it is reliable.

Reliability is another imperative factor in research. Many factors can influence the reliability of an instrument, and again, the process for establishing reliability is lengthy and complex. Most scientific research articles include measures of reliability for their instruments. It is recommended that an instrument with proven reliability be utilized for data collection.

Operationalization

When is crime a crime?

Statutory rape is defined in many different ways around the country. Rape in one state is not rape in another, even though the precipitating acts might be identical.

Beyond statutes, police department policies can produce significant variances in crime reporting.

Let us say that City A has an annual residential burglary rate of 1000 per 100,000 households and City B has a residential burglary rate of 2000 per 100,000 households. On the surface, City B appears to have a significantly higher residential burglary rate than City A. But the City A police department policy is to report a home garage burglary as a *petty theft* while City B reports according to the statute, home garage burglaries are *residential burglaries*.

Such differences are commonplace and the ethics debatable, but a comparison of the two cities would yield erroneous interpretations if *operationalization* is not considered.

Operationalization pertains to the measurement of variables and, in particular, the way variables are defined to make them measurable. This matter is far from trivial.

