

City University of New York (CUNY)

CUNY Academic Works

Open Educational Resources

Queensborough Community College

2020

Clear-Sighted Statistics: Module 3: Where Do Data Come From?

Edward Volchok

CUNY Queensborough Community College

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/qb_oers/56

Discover additional works at: <https://academicworks.cuny.edu>

This work is made publicly available by the City University of New York (CUNY).

Contact: AcademicWorks@cuny.edu

Clear-Sighted Statistics: An OER Textbook

Module 3: Where Do Data Come From?

“The object of statistical methods is the reduction of data. A quantity of data, which usually by its mere bulk is incapable of entering the mind, is to be replaced by relatively few quantities which shall adequately represent the whole, or which...shall contain...the relevant information contained in the original data.”¹

-- Ronald A. Fisher

I. Introduction

Fisher’s comment raises many questions that we will deal with throughout *Clear-Sighted Statistics*. Given that the primary concern of statistics is the accurate analysis of data, the critical issue we will now address is: How do we acquire good data to help gain a better understanding of the [phenomena](#) we are investigating? We want data that will not only improve our [comprehension](#), but will also help us make informed decisions. We do not want to misuse data, as many decision-makers do when they “...use it [data] as a [drunkard](#) uses a lamp post, for support, rather than for illumination”², as advertising great David Ogilvy was fond of saying. As investigators—as skeptics—we need to understand that good data shed light on the issues we are examining. How we acquire good data is, therefore, critical.

In this module, we will review the basic issues of obtaining high-quality data.³ We will also deal with related issues of obtaining accurate measurement. High-quality data is better than anecdotal evidence—evidence that is supported with a mere handful of stories.



Figure 1: *Where Do Data Come From?*
Volchok©1990

Anecdotal evidence lacks scientific rigor. Unlike widely circulating conspiracy theories that invoke sinister schemes based on circular reasoning that resist falsification, data developed through the application of the scientific method provides the only basis for improving our understanding and decision-making. Falsification is an essential element of the scientific method. Scientific facts must be susceptible to falsification; that is, they must be testable and possibly proven false. Karl Popper, a twentieth century philosopher of science, writing on falsification declared, "...it must be possible for an empirical scientific system to be refuted by experience."

After completing this module, you will be able to:

- Understand the basic research process.
- Discuss the differences between *primary* and *secondary research* or data.
- Distinguish *basic* or pure research from *applied* research.
- Identify the different types of research: *Exploratory*, *descriptive*, [*causal*](#), and *meta-analysis*.
- Understand what is meant by the term *big data* and discuss some of the opportunities and problems it presents.
- Understand the basic sampling techniques.
- Define *random sampling error* and discuss its importance.
- Identify the basic types of *systematic errors*.
- Understand the importance of *reliability* and *validity* when dealing with data obtained from surveys.
- Identify steps to take when reviewing research.

Some of these topics are usually introduced in research methodology courses. These topics, however, are central to statistical analysis. Students who successfully complete an

introductory statistics class, can construct *frequency distributions*, calculate the *mean* and *standard deviation*, explain *confidence intervals*, perform *significance tests*, and understand *correlation* and *regression*. Statistics students will miss important aspects of statistical literacy if they are not aware of the topics mentioned above. The goal of this module is to inform you about these important research issues.

II. The Research Process

There are two broad categories of research: 1) *applied research* and 2) *basic or pure research*. Applied research seeks [pragmatic](#) solutions to a pressing problem. In essence, it is research that will help decision-makers make more informed decisions by reducing the uncertainty they face. The objective of basic research, on the other hand, is to advance our knowledge, rather than solve a pressing problem. When distinguishing between practical and basic research, we should see them as poles on a [continuum](#) and not an either/or, or [binary](#) choice.

Research is a multi-stage process that involves several basic steps regardless of whether it is applied or basic research. We engage in research because we are faced with a problem. Facing a problem will cause the researcher to start asking questions. Here are examples of typical questions a researcher might have:

- How do Facebook users feel about the privacy of their data?
- What candidates for President of the United States get increased Google searches after a presidential debate and who is searching?
- What behavioral, biological, [pharmacological](#), and treatment factors contribute to better blood glucose control for Type I diabetics?
- What factors have contributed to the decline of Cable TV subscriptions?

- What factors are associated with a person not voting in local, state, and federal elections?
- What factors are associated with higher graduation rates at community colleges and four-year colleges?

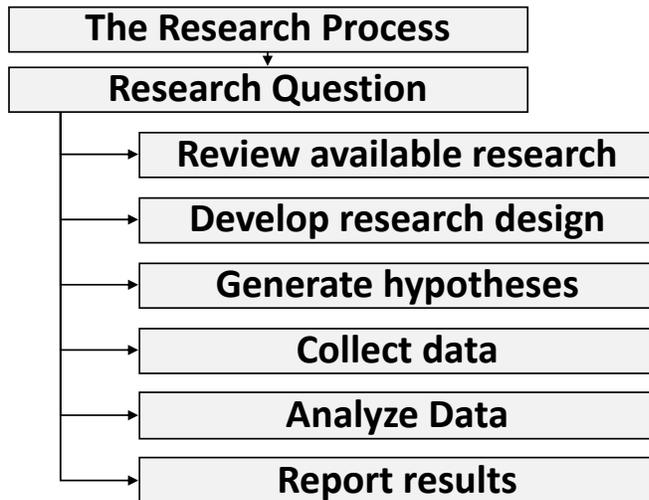


Figure 2: The Research Process

Once researchers articulate the basic questions, they will review available research or data. Two kinds of data and research that will be examined: *secondary data or research* and *primary data or research*. We will start with secondary data first. Secondary data have been collected by others and this information is generally available to the public. You can find secondary data in government reports, a variety of publications, published research reports, among many government sources like the Bureau of Labor Statistics, the U.S. Census Bureau, the National Center of Education Statistics, and the National Institute of Health. Reference librarians are extremely helpful for finding secondary data. Primary data is generally not available to the public. It is data generated by researchers who have processed internal databases and conducted various kinds of research. They examine secondary research or data first because it has been collected previously and, therefore, is readily available. Primary data becomes secondary data once it is published.

Secondary data are often useful because they:

- 1) Provide background information to help researchers refine their questions.
- 2) Might actually answer the researchers' question.
- 3) Might alert researchers about problems they need to avoid.
- 4) Might help researchers decide on the data they need and the most appropriate research methods to employ.
- 5) Might provide information that will help researchers with their sampling.

Secondary data, however, have some limitations:

- 1) Appropriate secondary data or research may not be available.
- 2) May be out-of-date.
- 3) May lack relevance for the problem being investigated.
- 4) May be inaccurate.
- 5) While useful, they may not adequately address the problem under investigation.

Whenever we use secondary data, we should be skeptical. We should be concerned about any biases embedded in the data. Bias is anything that distorts the accuracy of the data. When reviewing secondary data, we should consider the following questions:

- 1) Who gathered the data/conducted the research?
- 2) For what purpose was the data collected/research conducted?
- 3) How was the data collected/research conducted?
- 4) When was the data collected/research conducted?
- 5) What is included in the data/research and what is not?
- 6) Is it consistent with other secondary and primary data/research?

Once the secondary data or research has been reviewed, the researchers will determine whether additional research—primary research—is necessary. If so, the researchers will develop the research design. First, they will state the research objectives or goals. Like all objectives, these should be specific, measurable, achievable with the available budget and timetable, relevant to the questions being posed, and time-constrained, which is to say, that there should be a completion date for the research. To remember these requirements, some people use the following [mnemonic](#): **SMART** (**S**pecific, **M**easurable, **A**chievable, **R**elevant, and **T**imed).

After the research objectives are stated, the researchers will determine how to acquire the data, what sampling method to use, how large a sample they need, and how they will analyze the data. How they acquire the data, depends on budgetary constraints, time considerations, and the type of research. There are four basic types of research:

- 1) Exploratory
- 2) Descriptive
- 3) Causal
- 4) Meta-analyses

Exploratory research is preliminary research designed to get a deeper understanding of the research problem. It does not seek to test or confirm hypotheses. A hypothesis is a tentative answer to a research question. Exploratory research cannot prove causal relationships. Exploratory research includes focus groups and in-depth interviews. Focus groups are guided discussions of a group of six-to-ten people designed to elicit the respondents' perceptions on a particular topic. In-depth interviews are similar to focus groups except the discussion is with only one respondent. Exploratory research is qualitative, not quantitative. The techniques used to analyze exploratory research are typically not covered in an introductory statistics course.

Descriptive research addresses the questions of who, what, where, when, and how. The goal is to provide detailed descriptions of the studied phenomena. Descriptive research can be both qualitative and quantitative. The three major kinds of descriptive research are: 1) observational, 2) case studies, and 3) surveys.

With the observational approach, the subjects of interest are observed in a natural or laboratory setting. There are two types of observational studies: 1) *Prospective* and 2) *Retrospective*. Prospective studies identify subjects of interest and then collect data as events occur. With retrospective studies, data are collected after the events have happened.

The case study method involves detailed investigations of one or two examples of an issue in hopes of finding lessons for all similar cases. Quantitative and qualitative data are collected from a variety of sources so that generalized conclusions may be made about all cases.

With surveys, respondents provide data by answering a questionnaire. With survey research, issues regarding sampling technique, wording of the questions, and the *reliability* and *validity* of the data become very important. Reliability and validity will be explained later in this module.

Causal or experimental research uses controlled experiments to determine cause and effect relationships, although in Module 18, Linear Correlation and Regression, we shall see that observational research is able to establish causal links. With causal research, as [Turing award](#) winner Judea Pearl points out in *The Book of Why: The New Science of Cause and Effect*⁴, we hope to answer the *why* question by establishing causal relationships. The cause must precede the effect. If the cause is not present, we should not be able to observe any effect. We should also observe a [concomitant](#) variation between cause and effect; that is to

say, if we adjust the variable that causes the response, we should be able to measure the size of the effect. Here are several examples of causal statements:

- When we reduce the price of Diet Coke by 20 percent, dollar sales increase by 25 percent.
- Switching to an insulin pump will contribute to reducing Type I diabetics A1C score to below 7 percent.
- Smoking causes 7 out of 10 lung cancer cases in the United Kingdom.

Causality is barely discussed in introductory statistics textbooks. As Judah Pearl and his co-authors of *Causal Inference in Statistics* point out, most introductory statistics books do not even include the words “cause” or “causation” in the index. At most, textbooks emphasize that “*correlation does not imply causation.*”⁵ With cause and effect relationships, the cause or causes are called *independent* or *predictor variables* while the effect is called the *dependent* or *response variable*. There are also *confounding variables*, or variables that obscure the causal relationship if one exists. There is also a concern about *spurious* or *false correlations*. A spurious correlation occurs when two or more variables are linked mathematically due to the presence of unseen variables or mere coincidence, but have no causal relationship. We will discuss the issue of causation in Module 18.

The fourth type of research is a statistical method called *meta-analysis*. Studies using traditional Null Hypothesis Significance Testing (NHST) often yield contradictory results. Meta-analysis is a set of powerful statistical techniques that are used to develop quantitative analyses of multiple studies on the same or similar subjects. Meta-analysis helps researchers reconcile contradictory results among different studies. According to the Australian statistician, Geoff Cumming, “At its simplest, it [meta-analysis] gives a point estimate that is a weighted average of the separate study means.”⁶ Do not be concerned

that you do not yet know what weighted averages or point estimates are. We will define weighed averages in Module 5 and point estimates when we introduce confidence intervals in Module 11. This introductory textbook, however, will not explain how to conduct meta-analyses.

III. The Opportunity and Threat of Big Data

One of today's hottest topics is big data. Big data deals with huge amounts of data having three features called the three Vs:

- 1) **Volume:** Huge volumes of data are collected from a variety of sources including business transactions, social media platforms like Facebook or [TikTok](#), large websites like Google or Amazon, as well as sensors or machines.
- 2) **Velocity:** The data streams in with unprecedented speed and must be dealt with in a timely manner.
- 3) **Variety:** The data is available in all types of formats from structured numerical data to unstructured text, video, graphic, email, audio, stock market transactions, and even mouse clicks.

In his book called *Everyone Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Are*, data scientist and former Google employee, Seth Stephens-Davidowitz argues that big data has four advantages over traditional data sources:

- 1) It provides new types of data.
- 2) It is honest data. Unlike traditional surveys in which people may lie, big data tracks what people actually do, not what they say they do.
- 3) It offers the means to run large-scale randomized controlled experiments or A/B tests, which are usually extremely laborious and expensive, at almost no cost, and in this way uncovers causal links in addition to mere correlations.
- 4) The sheer volume of data allows researchers to “zoom in” on very small subsegments of the population.

One of Stephens-Davidowitz's central ideas is that traditional survey research has a major limitation: *People do not tell the truth*. To make his point, he reports survey results on the use of condoms in heterosexual sex. "Women say they have sex, on average, fifty-five times per year, using a condom 16 percent of the time. This adds up to about 1.1 billion condoms per year. But heterosexual men say they use 1.6 billion condoms every year." Are people lying or are they unable to remember the truth? Either way, the problem is that the survey findings are wrong. "According to Nielsen, the global information and measurement company that tracks consumer behavior, fewer than 600 million condoms are sold every year."⁷ Apparently both men and women are consciously or unconsciously exaggerating how much sex they have because their claimed use of condoms far exceeds the total number of condoms sold in the United States.

Stephens-Davidowitz argues that data derived from Google searches provide better clues to what people actually do than surveys. Google searches provide information that is missing from surveys. "More than half of citizens who don't vote tell surveys immediately before an election that they intend to," writes Stephens-Davidowitz, "skewing our estimation of turnout, whereas Google searches for 'how to vote' or 'where to vote' weeks before an election can accurately predict which parts of the country are going to have a big showing at the polls."⁸ During the 2016 presidential election, African Americans told pollsters that they would turn out in large numbers to vote against Donald Trump. Google searches on Hillary Clinton, Donald Trump, and the 2016 election in heavily African American neighborhoods, however, were "way down." One reason why Secretary Clinton lost the presidential election was low turnout by African American voters.⁹

Another great power of big data is that it makes randomized experiments “...much, much easier to conduct—anytime, more or less anywhere, as long as you’re online. In the era of Big Data all the world’s a lab.”¹⁰ Companies like Google and Facebook can use the data collected from their websites to rapidly conduct continuous causal research to tweak their websites. “Facebook now runs a thousand A/B tests per day, which means that a small number of engineers at Facebook start more randomized, controlled experiments in a given day than the entire pharmaceutical industry starts in a year,” writes Stephens-Davidowitz.¹¹

Big data, however, has a dark side, as data scientist Cathy O’Neil shows in her book *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. O’Neil bases her argument on the fact that predictive models based on big data often have “many poisonous assumptions [that] are camouflaged by math and go largely untested and unquestioned.”¹² Models are often “[inscrutable black boxes](#)” that are considered intellectual property by the developers who created them. O’Neil draws examples from these big data black box models in the areas of higher education, online advertising, the job market, acquiring credit and insurance, and civic life. She concludes that these models—these “weapons of math destruction” as she calls them—promise efficiency and fairness but “distort higher education, drive up debt, spur mass incarceration, pummel the poor at nearly every juncture, and undermine democracy.”¹³

IV. Non-Probability and Probability Sampling Methods

While we occasionally use data collected from a [census](#)—some of the data available through the United States Bureau of the Census or data that include *all* transactions from a business, we typically use data obtained through samples. There are two

broad categories of samples: *non-probability samples* and *probability samples*. A probability sample is one in which every variable in a population has a known, “non-zero chance” of being included in the sample. A “zero chance” means that this variable, this population element, would *never* be included in the sample. With non-probability sampling, the chance that an element of the population will be selected is unknown and cannot be calculated. Non-probability sampling may be appropriate for qualitative research, but this sampling technique is avoided when conducting statistical analysis.

Non-Probability & Probability Sampling

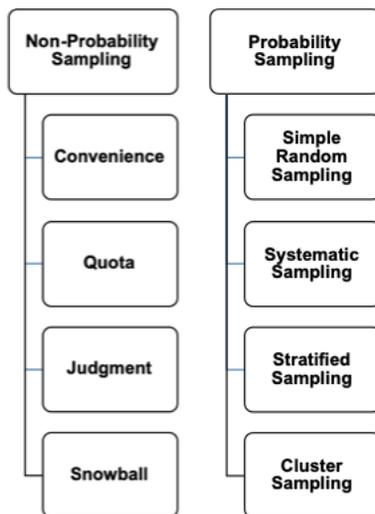


Figure 3: Non-Probability and Probability Samples

There are four kinds of non-probability samples: 1) Convenience or Haphazard Samples, 2) Judgment Samples, 3) Quota Samples, and 4) Snowball or Network Samples. There are also four kinds of probability samples: 1) Simple Random Samples, 2) Systematic Samples, 3) Stratified Samples, and 4) Cluster Samples.

A. Non-Probability Samples

Convenience or haphazard samples are based on using people or items that are readily available. This is a non-probability sample because it is impossible to determine the probability that an individual member of the population will be included in the sample. An example of a convenience sample would be your statistics professor using students in your class as a sample for all students in your college. She would do this because accessing you and your classmates is convenient, but she would not know the probability of all students enrolled at the college being included in her sample.

Quota samples are based on using [demographic](#), [psychographic](#), or other classification segments of the population. The sample is selected based on the proportion of the population each segment represents. The quota sample is not a probability sample because each element selected for the sample is not chosen at random; which is to say, by chance. The quotas used in a quota sample are based on the researcher's judgment and the best available information, which are often flawed.

Judgment or expert samples are based on the judgment of an expert about how well a sample represents the population being investigated. These samples are often biased because "expert" opinions about a population may be difficult to verify.

Snowball or network samples are used when a researcher wants to sample people with unusual characteristics and a *sampling frame* is not available. A sampling frame is a list of all the elements in the population. A researcher might use a snowball sample when it is difficult to obtain subjects who are members of small, hidden, or [clandestine](#) groups. Such groups could be musicians who play the oboe,

sex workers, or users of illegal drugs. Snowball samples are built by finding a small group of respondents who have the sought-after characteristics. Then the researcher obtains referrals from these respondents to increase the number of respondents to the desired sample size. This method reduces the cost of recruiting respondents, but may also increase the probability that the sample will be biased.

B. Probability Samples

The four kinds of probability samples are: 1) Simple Random Samples, 2) Systematic Samples, 3) Stratified Samples, and 4) Cluster Samples.

Simple random sampling is the simplest form of probability samples. Both the [Mega Millions](#) and [Powerball](#) lotteries select their winning numbers based on simple random sampling. Each element of the population has a known, non-zero chance of being selected. **Please note:** We will calculate the odds of winning these lotteries in Module 7: Basic Concepts of Probability. The probability of an element of the population being selected is based on the following formula:

- Probability of a Single Element Being Selected = Sample Size/Population Size

During the First World War, the United States drafted men into the armed forces using simple random sampling. The names of eligible draftees were written on paper tags which were then placed in a large fish bowl. The paper tags were mixed, and then names of men drawn at random until the draft board had the required number of draftees.

Suppose your statistics instructor wanted a sample of 50 students from the 15,000 students at your school. The probability of any student being selected for the sample is 0.0067 or 0.67%, found by:

$$\text{Probability of Being Selected} = \frac{50}{15,000} = 0.0067 = 0.67\%$$

Equation 1: The Probability of Being Selected

To conduct a simple random sample, you need a *sampling frame*, which you will recall, is a list of all the elements in the population. Your statistics instructor could get a sampling frame by requesting that the registrar provide a list of all 15,000 students enrolled at the college.

Once the registrar provides this list, your instructor would assign each student on the list a number, 1 through 15,000. Then using a *table of random numbers* or assigning each student a random number using Microsoft Excel, your instructor would select a number at some [arbitrary](#) point. Then he or she would randomly move up and down (and left and right if a Random Numbers Table is used) until fifty students are selected. Figure 4 shows a section of a random numbers table.

07833	02410	07329	04459	08473	05359	08302	09722	01365
00190	05509	08282	04442	09609	06485	06751	04749	06025
00885	03010	04549	02151	05379	08860	08294	03874	09901
00427	06833	08026	08634	08386	02997	03189	06357	04561
09129	07787	08631	04837	07263	03858	01870	04064	06466
05603	03722	02765	08479	06314	00878	05422	05561	09731
01371	08746	08304	05497	05261	00118	03352	08981	09774
06946	02811	06314	05605	09560	03338	00738	02817	05313
07300	02687	00844	07176	00891	06843	07622	02747	07195
01381	03491	09947	01460	09809	02508	06993	04025	06123
02614	07682	02401	01329	06736	09005	00137	06555	07323
00926	06058	04282	02517	02545	09584	08803	03446	07284
09172	06502	03031	02560	03336	08374	09641	09486	03017
01245	05609	04418	08989	06790	04986	02751	03946	08600
08821	06982	09257	00168	02102	06043	00509	07491	08170
03393	01320	05759	00801	07450	09448	01547	07640	02823
03977	05361	04732	03865	01859	04615	03322	01011	05822
04846	01612	08258	08852	04989	07031	01908	07933	06908
01218	02998	07893	03479	01877	06754	04517	04001	01684
06530	04076	07443	06263	05338	00531	09759	01423	00588
04609	04620	08813	01487	01641	09656	04542	08758	08820
00271	02975	03832	09517	09931	05072	03399	03459	08436
00172	05111	07628	05151	03641	00351	07021	07305	04281
06359	06270	01421	02607	05261	02607	09999	03393	00409
02906	08715	03299	05755	08171	06970	04030	00090	01648

Figure 4: Partial Random Numbers Table

The simple random sample has two major advantages over the four non-probability samples: 1) it is easy to execute, 2) every member of the population has

an equal, non-zero chance of being included in the sample. Not having a sampling frame, however, would present a problem.

Simple random sampling has disadvantages. Sampling frames are required and these may not exist with large populations. There are ways to create sampling frames, but these methods require time and money. They may also increase the probability of *sample selection bias*, which is a real concern. Sample selection bias occurs when the sampling frame is missing elements of the population.

Systematic sampling is another probability sampling method. It is often used as a substitute for simple random sampling.

Here is how systematic sampling works. The researcher selects a random number upon which she can base her sample. In our sample of 50 students from a population of 15,000, the researcher could use a random numbers table or generate a random number from 1 to 15,000 using Microsoft Excel. Let's say, student number 2,396 was selected at random. The researcher then selects a *skip interval* or space between the next selected student: 300 would be a reasonable skip interval, found by $15,000/50$. The next student selected would be number 2,696 ($300 + 2,396$). She repeats the selection interval until all 50 students have been selected. Here is how the skip interval is calculated:

$$\text{Skip Interval} = \text{Population Size} / \text{Sample Size}$$

$$\text{Skip Interval} = 15,000 / 50 = 300$$

Because of this ease in constructing, comparing, and interpreting, systematic sampling is used when researchers are operating under tight budgets. But this method has limitations. It assumes that the size of the population can be reasonably

estimated. The ordering of the population, however, may also introduce errors. It is faster and less expensive than simple random sampling. But there is also a small risk that patterns in the population may introduce sample selection bias. For example, a directory of telephone numbers is organized in alphabetical order of last names. Last names are linked to people's ethnicity, which could bias a sample. One way to limit this bias is to shuffle the population's order by assigning a random number to every element of the population and then sorting the population using the random numbers.

Strata sampling is another probability sampling method. The first step requires that the population be divided into two or more *mutually exclusive* and *collectively exhaustive* strata. Mutually exclusive means that each variable in the population fits in only one [stratum](#). Collectively exhaustive means that all variables are included in one of the strata. Traditional binary gender classes—male and female—are often used. Facebook's 51 gender identities are not mutually exclusive and therefore would not be a good starting point for strata sampling. The second step is to conduct random samples for each of the strata.

The strata used in the sampling may be based on demographics, values, attitudes, lifestyles, usage patterns, and purchase behaviors. A key question arises: How do we define the strata? Strata are selected based on the research question. Binary gender identities, for example, could be appropriate in political polls when we think gender plays a role in voting behavior.

Stratified samples have less sampling error than simple random samples.

Remember: Sampling error is defined as the difference between the population

parameter and the sample statistic. In addition, when using stratified samples, we can use smaller samples.

Stratified sampling has its drawbacks: 1) The information required to properly stratify a population may not be available. You may not know much about the strata in a population. Each stratum must have distinctive features that do not overlap with other strata. When you cannot identify these features, you cannot use strata sampling. 2) When distinctive strata cannot be identified, it may not be worth the time and expense to identify them.

Three steps are used to stratify a population properly:

- 1) **Identify the important stratification factors.** These factors should be related to the question being investigated. Stratification factors could be:
 - a. Demographics
 - b. Psychographics or values, attitudes, and lifestyles
 - c. Usage or purchase behaviors

- 2) **Determine the proportion of the population that each stratum represents** with each stratum being mutually exclusive.
 - a. With *proportional allocation*, the number of items selected from each stratum are proportional to the proportion of the stratum to the population
 - b. With *disproportional or optimal allocation*, the number of items selected from each stratum is based on weighting the proportion of each stratum to the population and the variability of the characteristic under consideration

- 3) **Select a random sample from each stratum**

Cluster sampling involves [partitioning](#) or dividing a population into separate groups called clusters. Unlike the strata of stratified sampling, clusters are [heterogeneous](#) groups. Clusters are often based on geographic areas to reduce sampling costs. There are two critical steps in cluster sampling: 1) The population of

interest is divided into mutually exclusive and collectively exhaustive clusters; 2) A random sample of the clusters is conducted.

There are two types of cluster samples: *One-Stage Cluster Samples* and *Two-Stage Cluster Samples*. In the former, a simple random sample is conducted in all of the clusters. In the latter, a sub-set of clusters are selected through random samples and then the sample is selected by conducting a random sample of the selected clusters.

Cluster samples are considered probability samples because of the random selection of clusters and the random selection of elements within the clusters. The assumption is that the characteristics of the clusters are as heterogeneous as the population. If this is not the case, bias will be introduced.

The advantage of two-stage cluster sampling is that the time to complete the sampling and the cost of sampling are reduced because sampling is restricted to only a few clusters, assuming a two-stage cluster sample is used. But the trade-off, the disadvantage, is that the rate of sample error is higher than other methods.

V. Random Sampling Errors

As discussed, all samples are liable to have random sampling error. Simply stated, random sampling error occurs whenever the sample statistic is not equal to the population parameter. In Module 10, Sampling and Sample Errors, we shall see that these errors are not the result of mistakes made by people involved with the sampling. Sampling errors occur whenever samples are selected.

VI. Systematic Errors

Systematic errors result from errors in the research design and execution either because of actions taken by researchers or respondents. There are two broad categories of systematic errors: *sample design errors* and *measurement errors*.

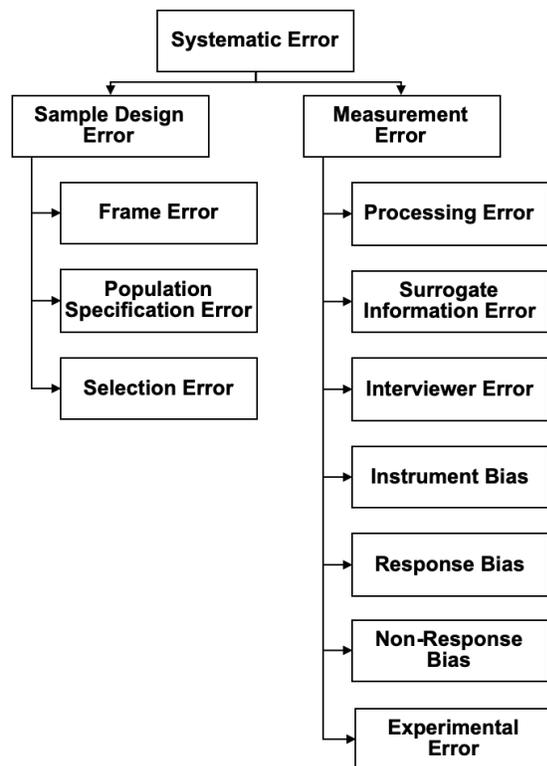


Figure 5: Systematic Errors

A. Sample Design Errors

Sample Design Errors result from problems with the sample design or how the sampling was conducted. There are three kinds of sample design errors: 1) frame error, 2) population specification error, and 3) selection error.

1) Frame Error: You will recall that a sample frame is a list of all the elements in a population from which the researchers will draw their sample. If researchers draw a sample from an incomplete or inaccurate sample frame, they will have frame error. Here is an example of frame error. A marketing researcher plans to test an email campaign and will need at least two groups of people for the

test. The first group is called the *treatment group*. This group will receive the new type of email that the marketer is considering using. The second group, the *control group*, will receive the email the marketer is currently using. Through this research, the marketing researcher will determine whether the treatment group yields better results—higher response rates—than the control group. Frame error will occur if the email lists for these two groups—the sample frames—are inaccurate, which may bias the results.

2) Population Specification Error: This type of error results from an incorrect definition of the population. Suppose a researcher wants to sample people with Type II Diabetes, which is sometimes still referred to as Adult Onset Diabetes. The researcher specifies the population as adults 18-years-old or older. Later, it is determined that obese children as young as 8-years-old are developing Type II Diabetes. If these children are excluded from this population, there will be population specification error and the samples derived from this inaccurately defined population will be biased.

3) Selection Error: Even when researchers have an accurate sample frame and the population is properly defined, there can be selection error. Selection errors occur when the sampling procedures are not properly followed or when the procedures themselves are improper or inaccurate. One way selection error occurs is when an interviewer decides to avoid interviewing certain types of people.

B. Measurement Errors

Measurement Errors are human errors made when the data are collected. Unlike random sampling error and sample design error, measurement error can happen

with censuses or samples. There are seven basic kinds of measurement errors: 1) processing error, 2) surrogate information error, 3) interviewer error or bias, 4) instrument or questionnaire bias, 5) response bias, 6) non-response bias, and 7) experimental error.

1) **Processing Error:** Processing error includes a wide range of errors that occur after the data have been collected. Processing error includes errors in coding, transcribing, assigning weights to the data as well as the use of inappropriate statistical techniques.

2) **Surrogate Information Error:** Surrogate information error occurs when there is an inconsistency between the information sought and the information needed to solve a problem. This error is usually caused by the researchers' lack of understanding of how respondents view the questions they being asked. The following survey question will generate surrogate information error: What is your favorite breakfast beverage? a) Coffee, b) Tea, c) Milk. With this question, the researcher is not measuring respondents' favorite breakfast beverage, rather he or she is measuring the preference among these three beverages. The researcher should have formulated an open-ended question—a question without fixed answers—to get a wide range of breakfast beverages: Hot chocolate, orange juice, tomato juice, grapefruit juice, prune juice, water, gin and tonic, etc.

3) **Interviewer Error or Bias:** Interviewer error occurs when the interviewer consciously or unconsciously influences respondents' answers. How respondents react to the interviewers' age, gender, race, body language, attire, accent, and tone of voice can also cause this type of bias. Interview bias can arise

when interviewers record inaccurately and interpret observation data incorrectly. Another type of interviewer error occurs when the interviewer commits fraud. An interviewer who does not conduct the survey and covers up this fact by completing the questionnaires without the input of respondents has committed fraud. An interviewer would also be guilty of fraud if he or she changed respondents' answers.

4) **Instrument or Questionnaire Bias:** This type of error is due to poorly worded or confusing questionnaires. (**Please note:** Researchers often call questionnaires *instruments*.) These errors result from unskilled writing or deliberate attempts to get results that support a [foregone](#) conclusion rather than shedding light on a problem. Asking *leading*, *loaded*, and *double-barrel* questions causes instrument bias.

Here is an example of a leading question:

- How dumb was President Obama's policy on North Korea?

This is a leading question because it contains the embedded—and biased—idea that President Obama's policy on North Korea was dumb.

Here is an example of a loaded question:

- Don't you think that the liberal media push fake news to undermine President Trump?

This is a loaded question because it contains unjustified assumptions in the hope of skewing the respondents' answers.

While leading and loaded questions are asked by researchers who hope to get responses that will confirm a desired result, double-barreled questions are the result of poor questionnaire design.

Here is an example of a double-barreled question:

- Would you vote for a presidential candidate who supports cutting spending on education and health care?

This is a double-barreled question because it asks two separate questions: 1) cutting spending on education and 2) cutting spending on health care in one. Respondents may be confused if their answers for the two questions differ. They may feel trapped because they want to answer affirmatively to one question and not to the other.

Good surveys never pose double-barreled questions. We see double-barreled questions posed all the time by cable news hosts when they interview politicians. This enables the politician to avoid the thrust of the interviewer's questions and insert whatever talking point he or she wants to push or to answer only one of the questions while avoiding the other.

5) **Response Bias:** Response bias occurs whenever a respondent gives false or misleading answers to a survey question. There are two types of response bias: *deliberate falsification* and *unconscious falsification*. Deliberate falsification is a nice way of saying that a respondent lied. Respondents lie to appear more intelligent or successful, to conceal confidential information, or to avoid embarrassment. We call this *social desirability bias*.

Unconscious falsification occurs when the respondent does not understand the question, is unable to recall details, thinks the events that he or she is discussing happened more recently or less recently than they did.

Other types of response biases are:

- **Acquiescence bias:** [Acquiescence](#) bias is the tendency of some respondents to agree with all questions or to answer all question with positive [connotations](#).
- **Extremity bias:** Extremity bias is the tendency of some respondents to use extremely positive or negative responses. Extremity bias is associated with non-response bias, or the bias caused by people not completing a questionnaire. Non-responders are sometimes thought to have no strongly held views and therefore lack motivation to complete the questionnaire.
- **Auspices bias:** Auspices bias is the tendency of some respondents to be influenced by the organization sponsoring the study.

6) **Non-Response Bias:** Non-response bias occurs when respondents do not completely answer the questions on a survey. Or, when they simply refuse to participate. Non-response bias or error is the difference between a “perfect” survey in which every person contacted completes the questionnaire and the results derived from those who completed the questionnaire. Researchers should be concerned with **self-selection bias** that occurs when respondents are more likely to have more deeply held positive or negative opinions than those who do not respond.

Usually researchers seek ways to reduce non-response bias. But remember, some dishonest executives are like the drunkard who seeks support rather than illumination of the problems. We do, however, find executives who want non-response bias. President Trump’s repeated insistence on including a question on citizenship in the country’s 2020 census is an attempt to undercount immigrants. The *Wall Street Journal* reports that the Harvard University’s Shorenstein Center on Media, Politics, and Public Policy estimated that adding a citizenship question to the census would have reduced the number of Hispanic respondents in the 2010 census by 12 percent or 6.07 million people. Undercounting Hispanics will lower their

representation in Congress, as well as shift allocation of financial resources to demographic groups that are not undercounted.¹⁴

7) **Experimental Error:** Experimental errors occur when conducting research with human respondents. These experimental errors are often called *reactive effects* because respondents react to the fact that they are being observed by altering their usual behavior. These errors distort the study's findings. Here are three common experimental errors:

a) Hawthorne Effect or Observer Effect: The Hawthorne effect was discovered as a result of studies conducted between 1924 to 1932 at the Hawthorne Works near Chicago. Executives at the factory commissioned a study to determine whether workers would be more productive in higher or lower levels of light. The researchers found that workers' productivity increased whenever any change in lighting was made and then slumped after the study. This *reactive effect* occurred because workers did not behave normally knowing that they were in a study. This, of course, distorted the results. Researchers, therefore, hope to keep research participants unaware of the study. If that is not possible, participants should not be aware of the desired outcome of the study.

b) Placebo Effect: The [placebo](#) effect is a major concern of medical researchers. It occurs when a study's participants show positive results regardless of whether they are in the group receiving the experimental treatment or in a control group receiving a placebo, which is an inert substance that has no therapeutic value. The placebo effect can distort the results of a study. A 2015 study of clinical tests on neuropathic pain conducted from 1990 to 2013 found that placebo responses have increased.¹⁵

c) John Henry Effect: The John Henry effect occurs when members of the control group become aware that they are in a control group, and then they respond by working harder to overcome this disadvantage. John Henry, you may recall, is a folk hero. He was the “steel driving man” who raced against the railroad’s new steam-powered rock drilling machine using his hammer and his own strength. According to the lyrics of the popular folk song written by [Pete Seeger](#), “The man that invented the steam drill thought he was mighty fine. But John Henry made fifteen feet, the steam drill only made nine.”¹⁶ Hooray for John Henry! But the researchers’ experiment is biased due to the John Henry effect. By the way, the steel driving man met a sad end. After the race, he “laid down his hammer and died.”

VI. Measurement and the Problem of Reliability and Validity

Whenever we deal with data, we are concerned with measurement. Measurement is the assignment of numbers to variables. Sometimes measuring variables, or concepts which the variables represent, is very easy. Sometimes it is not. It is easy, for example, to measure sales for Coca-Cola. Two major marketing research companies have subscription services that monitor retail sales data for consumer packaged goods: Information Resources Incorporated (IRI) and Nielsen. Consumer packaged goods is a term used for products that consumers use and repurchase frequently. Both IRI and Nielsen measure sales as recorded by check-out scanners. There are a number of ways in which these sales are measured.

When using sales data, we would ask the following questions:

- What time period was the data collected?
- What geographic areas are included in the data?

- What types of sales are being measured? “Retail sales” (sales made to people who buy Coca-Cola for their own enjoyment) or “factory sales” (sales made to companies that sell Coca-Cola to resell to other people).
- Does the data include dollar sales, unit sales, or both? If unit sales are included, how are they measured?

Dollar sales versus unit sales: Obviously sales can be measured in how much money was spent in any currency we choose. The type of currency used depends upon where the sales were made. Unit sales, however, are a bit more complex. Sales recorded by retailers’ check-out scanners will record the total number of packages scanned. Each package, whether, it is a single 7.5 ounce can or a box of 24 16.9 ounce plastic bottles counts as one unit even though the prices are very different. Given the fact that Coca-Cola is sold in a wide variety of package sizes, we may seek a more refined measure of unit sales. This measure is called *equivalized units* or EQ for short. The EQ for Coca-Cola and other carbonated soft drinks is usually considered 24 8-ounce servings or 192-ounces.

Other easy-to-measure concepts include weight, height, blood pressure, time spent completing a task, or distance travelled.

When the concepts we want to measure are abstract, our task is not so easy. How do we measure someone’s personality using the 16 [Myer-Briggs](#) personality types? How do we measure the extent to which a person has an authoritarian personality using a version of Theodor W. Adorno’s [F Test](#)?

A marketing researcher may seek to measure abstract concepts like brand loyalty, a sociologist may seek to measure social deviance, and an educator may seek to measure quantitative literacy. We call these abstract concepts *constructs*. Measuring constructs that deal with people’s psychological traits are called [psychometrics](#).

When trying to measure constructs we develop two types of definitions: a *constitutive definition*, which is often called a *theoretical* or *conceptual definition*, and an *operational definition*. A constitutive definition states the essential meaning of the construct. An operational definition, on the other hand, is a description of the observable features of a construct that will be measured. Brand loyalty is a difficult construct to measure. Marketing researchers have been arguing about the definition of brand loyalty for over fifty years.¹⁷ The constitutive definition for brand loyalty could be the positive associations felt by a consumer for a particular brand versus competitive brands that contribute to an inclination to buy that brand. Brand loyalty is generally considered a *multi-dimensional* construct. An operational definition would focus on different *dimensions* or aspects of how brand loyalty would be measured. The dimensions might include: The number of times out of ten purchase occasions the consumer buys the brand (*brand switching*), how often the consumer buys the brand (*frequency*), the amount of money spent on the brand (*monetary value*), the time since the last purchase (*recency*) and the likelihood that the respondent would recommend the brand to an acquaintance (*brand recommendations*). The metrics used to create an operational definition would depend on the product category. Brand loyalty would be measured—*operationalized*—differently for carbonated soft drinks, cars, cigarettes, or condoms because the frequency of purchase—the *purchase cycle*—varies widely for these product categories.

Once we have constitutive and operational definitions, the researchers will develop a questionnaire to measure each of the dimensions of the construct. Typically, researchers will write many more questions than they will ultimately use. One survey researcher who I used to work with said “we start fat and work to get thin.” By that he meant, the

questionnaire would be tested and refined to make certain that only the bare minimum number of questions are used to measure the construct, and that the questionnaire actually measures what we hope it will measure. Whenever we measure constructs we want our measurements and questionnaires to be *reliable* and *valid*. *Reliability* measures consistency while *validity* a measures accuracy. As the graphic shown in Figure 6, there are three alternatives for a measurement. A measurement can have poor validity and good reliability, poor validity and poor reliability, or good validity and good reliability.



Figure 6: The Relationship Between Reliability and Validity

A) Reliability

Reliability refers to the consistency or stability of a research instrument. Measurements that provide consistent results over time are reliable. **But reliability is not accuracy.** Let's say that we have a yardstick. Like all yardsticks, this yardstick shows 36 inches. Every time we measure something with it, the yardstick shows the same 36 inches. The yardstick would be reliable even if the inches on the yardstick were actually just over 1.11 inches long, and what we thought was 36 inches was actually 40 inches.

As shown in Figure 7, there are four methods used to determine whether a measurement scale or questionnaire is reliable: a) test-retest, b) equivalent form, c) internal consistency, and d) inter-rater consistency.

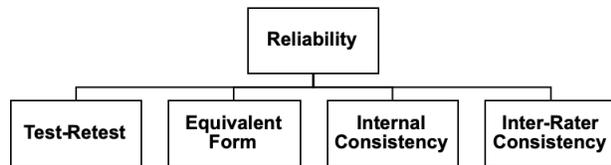


Figure 7: The Four Kinds of Reliability

a) Test-Retest Reliability:

With the test-retest method, respondents are given the survey at two different times under similar conditions. When there are very few differences between the results of the first and second surveys, the instrument is said to be stable and we conclude that it is reliable.

There are some problems with this technique:

- a) It may be difficult or impossible to get respondents to complete the questionnaire twice.
- b) The administration of the first survey may cause a respondent's answers to change.
- c) [Extraneous](#) factors in the respondents' environment may cause the measurements to change.

b) Equivalent Form Reliability:

Equivalent form reliability uses similar research instruments—forms that are essentially equivalent—to determine whether they yield similar results. Equivalent form reliability attempts to overcome the limitations of test-retest reliability by measuring the strength of the consistency or [correlation](#) between two similar research instruments. When the equivalent forms yield similar results, the instruments are considered reliable.

c) Internal Consistency Reliability:

Internal consistency reliability is the ability of a research instrument to return similar measurements when given to different samples during the same time period. There are two ways to test for internal consistency reliability. The first is called the *split ballot technique*.

Questions that explore the same construct are divided into two groups of equal numbers of questions. The groups of questions are given to different samples of respondents. The problem with the split ballot technique is that it depends entirely on how the items were split. A better technique is based on using a statistical technique called *Chronbach's Alpha*, which calculates the mean reliability scores for all possible ways of splitting the questions. Here is how Chronbach's Alpha scores are generally interpreted:

Table 1: Interpreting Chronbach's alpha Scores

Chronbach's alpha	Internal Consistency
$0.9 \geq \alpha$	Excellent
$0.8 \geq \alpha < 0.9$	Good
$0.7 \geq \alpha < 0.8$	Acceptable
$0.6 \geq \alpha < 0.7$	Questionable
$0.5 \geq \alpha < 0.6$	Poor
$0.5 < \alpha$	Unacceptable

Because *Clear-Sighted Statistics* is an introductory textbook, we will not cover the calculation of Chronbach's alpha.

d) Inter-Rater Reliability:

Inter-rater consistency reliability refers to how independent researchers compare their assessments when they observe the same behavior. An example of inter-rater consistency is when the quantitative scores of Olympic judges of a gymnastics competition are consistent.

B) Validity

Validity refers to the accuracy of measurements, how the research plan is being executed, whether the research "instrument" is accurately measuring the constructs under investigation. With validity we want to make certain that our 36" yardstick accurately measures 36". By the way, no measurement instrument is perfect. They all have some level

of inaccuracy. Physicists at the National Institute of Standards and Technology have developed the most accurate clock ever made. The estimates are that this clock will gain or lose a second in 33 billion years or two-and-a-half times the age of the universe.¹⁸ Not perfect. The clock on your smart phone is far less accurate than the NIST's new atomic clock, but for most purposes, it is a valid instrument for telling time.

Figure 8 shows the basic types of validity:

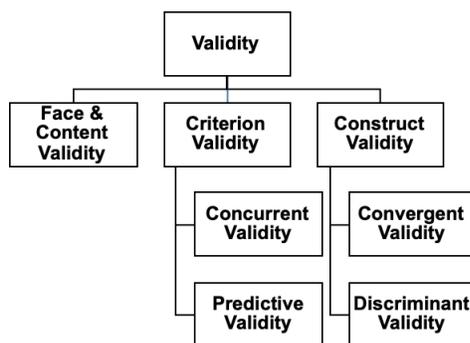


Figure 8: Types of Validity

Face Validity: Face validity is a non-statistical form of validity. It is the degree to which the measurement is *judged* to accurately measure what it is supposed to measure. The people making this judgment are not considered experts in testing methodology. It is, therefore, the lowest level of validity.

Content Validity is another non-statistical form of validity. It is the extent to which the measurement of the content actually represents the content as determined by the judgment of experts in the field.

Both face validity and content validity are considered starting points for establishing validity. They are necessary but not sufficient for establishing the validity of an instrument. But whether an instrument has face or content validity without *criterion* or *construct validity*, the instrument is not considered valid.

Criterion Validity: The degree to which the research instrument can predict the designated criterion. There are two forms of *criterion validity*: *concurrent validity* and *predictive validity*.

- **Concurrent Validity:** The degree to which one variable measured can be predicted by the instrument concurrently—at the same time—with another variable of interest. If we have a predictive relationship, we say that these variables [concur](#). Suppose you develop a new and less expensive test to determine if a person has lung cancer. If your test and the established test with proven validity are given to patients at the same time, your test would have concurrent validity if its results were similar to those of the established test.
- **Predictive Validity:** The degree to which the future level of the criterion variable can be predicted by the instrument. SAT scores would have predictive validity if these scores are predictive of students' future academic performance in college.

Construct Validity: The degree to which the instrument accurately measures the construct of interest. There are two forms of *construct validity*: *convergent validity* and *discriminant validity*.

- **Convergent Validity:** The strength of the association among different questions or instruments that purport to measure the same construct. When the measures for these dimensions are similar—when they converge— you have convergent validity.
- **Discriminant Validity:** When test results show that two measures that are supposed to be dissimilar are, in fact, dissimilar, you have discriminant validity.

VII. A Skeptic's Guide to Reviewing Research

When you examine research, adopt the attitude of a skeptic, not a cynic. **Be aware:** Any skeptic whose standards of proof are so strict that he cannot be convinced by strong evidence, is not really a skeptic. He is, in fact, a cynic who believes that there is no truth. In addition, you should avoid the pitfalls of conspiracy theorists. A conspiracy theorist may

declare himself a skeptic. He is not. He attempts to explain random events as the result of a sinister and secret plot by powerful people who aim to control the world. A conspiracy theorist may declare himself skeptical about the [efficacy](#) of vaccinating children against measles. A recent study of 5,323 participants in 24 countries found evidence that the belief in anti-vaccine conspiracies is associated with beliefs in other conspiracies like: Princess Diana was murdered, the American government had advanced knowledge of the 9/11 attacks and chose to let them happen, a clandestine group of conspirators is plotting a new world order, and so forth.¹⁹

Sadly, people who believe in conspiracy theories are highly resistant to evidence-based arguments. Conspiracy theories typically suffer from two kinds of bias: *confirmation bias* and *hindsight bias*. Confirmation bias is the tendency to look for evidence that confirms what you already believe while ignoring evidence that undermines those beliefs. Hindsight bias occurs when someone perceives historical events as having been more predictable than they were before the event occurred.

Here are some guidelines to follow when reviewing research that uses statistical analysis.

Step 1 - Have the study's goals been identified?

Detailed descriptions of a study's goals always appear in peer-reviewed journals. **Please note:** A [peer](#)-reviewed journal use a small panel of "peers" to evaluate scientific, academic, or professional work to filter out invalid or poor-quality research. The decision to publish an article is made based on the panel's recommendation. The goals of a study are explained in the article's introduction and in the abstract, which is a 100-to-200 word summary that accompanies published papers.

Step 2 – Has the research methodology been described?

Determine the type of study you are reviewing. Is it exploratory, descriptive, causal, or a meta-analysis? All trustworthy researchers explain the research methodology they followed. The researchers' conclusions must be compatible with the type of research they conducted. Exploratory research, for instance, cannot be used to make or confirm causal claims.

Step 3 – Who funded the research?

Good studies are unbiased. When reviewing research, you need to know the biases of the researchers as well those of the sponsors of the research. Researchers are obligated to disclose their funders of the research and whether they had input into the analysis. Readers must be alerted to potential conflicts of interest. When studies make claims that seem to support the sponsor's interest, the reader should determine whether other studies have replicated the study's results. (See Step 4.) You should be skeptical of a study that shows no relationship between the consumption of high calorie, sugar-sweetened beverages and obesity when Coca-Cola is a sponsor. The authors of one such study noted the potential conflict of interest and tried to allay concerns by declaring, "[This research] was funded by the Coca-Cola Company. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript."²⁰

Be aware: Research is not always published in peer-reviewed journals. Sometimes misleading, biased research findings are presented by pro-industry advocacy groups that do not divulge the sources of their funding. These [front groups](#) issue papers that downplay concerns raised by scientists and consumers about the safety of products like certain genetically modified foods, pesticides, herbicides, and high-sugar diets. One such group is

the [American Council on Science and Health](#). On October 9, 2018, the ACSH published a piece with the provocative title, “If You Accept Science, You Accept RoundUp Does Not Cause Cancer.”²¹ RoundUp is an herbicide made by Monsanto. It contains glyphosate. Some research suggests that glyphosate may cause Non-Hodgkin Lymphoma and other cancers.

According to [U.S. Right to Know](#), a non-profit investigative research group that focuses on the food and chemical industries, the ACSH is:

“...funded by chemical, pharmaceutical and tobacco companies, according to leaked internal documents that document how the group pitches its services to corporations for product-defense campaigns. Emails released via court proceedings show that Monsanto agreed to fund ASCH in 2015, and asked the group to write about the [International Agency for Research on Cancer] IRAC cancer report on glyphosate; ACSH later claimed the cancer report [which classified glyphosate as ‘probably [carcinogenic](#) to humans’ was a “scientific fraud.”²²

In April 2013, the United States Environmental Protection Agency reviewed the literature on glyphosate and cleared the chemical of any public health risks. But, the World Health Organization declared the herbicide a “probable” [carcinogen](#). Despite these mixed reviews, juries have ordered Monsanto to pay billions of dollars to people claiming their cancer was caused by RoundUp.²³

Product-defense campaigns are part of a marketing communication effort. As someone who spent nearly 30 years working for Fortune 500 companies in marketing communications, I feel that I must disclose my bias against this kind of public relations. I think it often crosses the line by favoring corporate financial concerns over the pursuit of the truth. Early in my career I worked in the flagship office of the advertising firm [Ogilvy & Mather](#). I am reminded of David Ogilvy’s comments on lying in advertising. In 1963, Mr. Ogilvy wrote:

“Never Write an Advertisement Which You Wouldn’t Want Your Family to Read. You wouldn’t tell a lie to your wife. Don’t tell them to mine. Do as you would be done by. If you tell lies about a product, you will be found out— either by the government, which will prosecute you, or by the consumer, who will punish you by not buying your product a second time.”²⁴

Given that Mr. Ogilvy wrote this statement over 57 years ago, we can forgive his [paternalism](#), and applaud his stance against fraudulent marketing communication masquerading as research.

Readers should check the sources of the research they are reviewing. A good resource for getting information on corporate public relations campaigns is [SourceWatch](#) published by the [Center for Media and Democracy](#).

Step 4 – Have the results of the research been replicated?

Good research findings can be reproduced by other researchers. If you are surprised by the research findings, ask yourself whether they have been replicated by other researchers.

Good researchers will place the results of their study in the context of the findings published by other researchers. You should be concerned when the findings are surprising and inconsistent with the findings of other researchers.

A famous case of published research that could not be replicated occurred in 1989.²⁵ A University of Utah chemist named Stanley Pons and British chemist Martin Fleischmann claimed to have discovered cold fusion. With cold fusion we would have an inexhaustible supply of clean, cheap, and reliable energy. This study created a lot of excitement because it would mean that we could say goodbye to polluting fuels that contribute to global warming. Carbon emissions would plunge. If Pons and Fleischmann were right, their discovery would be on par with those made by Newton, Galileo, Copernicus, and Einstein.

They would probably win the Nobel prize and become extremely wealthy as their technology would become the basis of the world's power generation.

What is fusion and why is cold fusion so exciting? Fusion is a process whereby two or more atomic nuclei fuse, releasing an enormous amount of energy. Unfortunately, fusion takes place in the core of a star at an extremely high temperature. The temperature of the core of our closest star—the sun—is estimated at 27,000⁰ F. Pons and Fleischmann claimed that they had fused atoms at room temperature. Before the history books and the physics books were rewritten, scientists tried to replicate cold fusion in their laboratories. In science, major discoveries must be verified with multiple sources of evidence. No scientist, however, was able to verify the findings of Pons and Fleischmann. As a result, Pons and Fleischmann became [infamous](#). They did not win a Nobel prize and no one ranks them among history's greatest scientists.

Step 5 – Has the sample size been justified?

The size of the sample must be stated in the researchers' statement on methodology. Most researchers do this. But merely mentioning sample size is not sufficient. Researchers must go further and address whether their sample size is sufficient to illuminate the problem under investigation.²⁶ To do this, they must address the issue of the *statistical power* of their analysis. Statistical power is the ability of the research to detect an effect, given the variability of the data and the size of the sample. We will examine the issues of statistical power and effect size when we review Null Hypothesis Significance Testing, NHST, starting in Module 13. Most statisticians suggest the minimum acceptable statistical power is 80 percent; which is to say, the research has an 80 percent chance of finding a statistically

significant result when there actually is one. Statistical significance will be discussed in detail when Null Hypothesis Significance Testing is covered.

Step 6 – Have the key variables been adequately defined and measured?

As part of the methodology discussion, the researchers must assign all the variables operational definitions. Then you have to make a judgment about whether the research properly measured these variables and identified any variables that may [confound](#) the study's findings.

Step 7 – If the researchers used a questionnaire, is it reliable and valid?

Poorly structured questionnaires are neither reliable nor valid. When reviewing survey research, it is important that researchers provide evidence that supports the instrument's reliability and validity. A copy of the questionnaire should be included in an appendix. The wording of each question should be reviewed to determine whether the researchers failed to mention any possible biases.

Step 8 – Have the researchers identified potential sources of systematic errors?

As you review the research results, consider sources of systematic errors and determine whether the researchers adequately addressed concerns about systematic errors.

Step 9 – Are the tables and charts misleading?

You should consider whether the tables and charts presented in the research are misleading. We will review how to describe data using tables and charts in Module 4: Picturing Data with Tables and Charts.

Step 10 – More Questions to ask.

Here are three more questions to address:

- 1) Did the study achieve its goals?

- 2) Are there alternative explanations? Remember Ockham's razor or the law of [parsimony](#). Ockham's razor, named after the medieval philosopher [William of Ockham](#), is a principle that states that when competing hypotheses make the same prediction, the simplest hypothesis—the most parsimonious—is *usually* correct.
- 3) Do the results have *statistical significance* and *practical significance*? Statistical significance means that the difference between a sample statistic and a population parameter cannot be explained by simple sampling error. Statistical significance does not imply practical or real-world significance. Practical significance, however, means that the findings have a practical application in the real world. Practical significance can be determined by measuring effect size (ES), which we will review in the context of null hypothesis significance testing.

VIII. Exercises

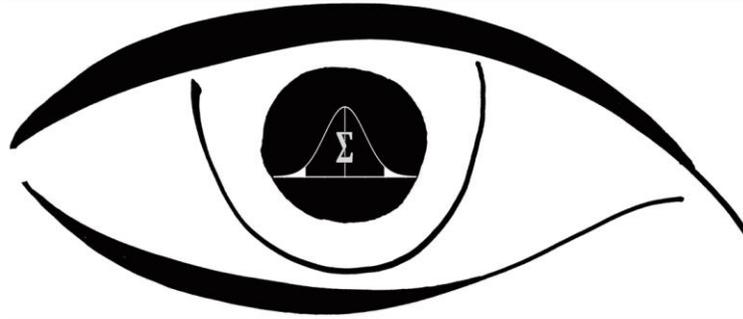
Answers to the following questions can be found by carefully reading this module.

1. **What is anecdotal evidence and why is it unreliable?**
2. **What is the difference between applied research and basic research?**
3. **What are secondary data?**
4. **What are the advantage of secondary research?**
5. **What questions should you ask when you review secondary research.**
6. **What are the limitation of secondary research?**
7. **What is exploratory research?**
8. **What is descriptive research?**
9. **What is observational research?**
10. **What are case studies?**

- 11. What are surveys?**
- 12. What is causal research?**
- 13. What is meta-analysis research?**
- 14. What are the advantages of “big data”?**
- 15. How do Non-Probability and Probability Samples differ?**
- 16. What are random sampling errors?**
- 17. What are systematic errors?**

* * *

CLEAR-SIGHTED STATISTICS



EDWARD VOLCHOK



Except where otherwise noted, *Clear-Sighted Statistics* is licensed under a [Creative Commons License](#). You are free to share derivatives of this work for non-commercial purposes only. Please attribute this work to Edward Volchok.

I would like to thank Ms. Leslie Ward, an Assistant Professor in the Library Department of Queensborough Community College. Ms. Ward reviewed a draft of this document and made numerous suggestions. Of course, I am solely responsible for any errors in this text.

* * *

¹ Ronald A. Fisher, "On the Mathematical Foundations of Theoretical Statistics." *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, Vol. 222, 1922, p. 311. *JSTOR*, www.jstor.org/stable/91208.

² David Ogilvy, *Confessions of an Advertising Man*, (New York: Ballantine Books, 1963), p. 10.

³ Karl R. Popper, *The Logical of Scientific Discover*, (New York: Basic Books, 1959), p. 41.

⁴ Judea Pearl and Dana Mackenzie, *The Book of Why: The New Science of Cause and Effect*, (New York: Basic Books, 2018).

⁵ Judea Pearl, Madelyn Glymour, and Nicholas P. Jewell, *Causal Inference in Statistics*, (West Sussex, UK, John Wiley & Sons, 2016), pp. xi-xii.

⁶ Geoff Cumming, *Understanding the New Statistics: Effect Size, Confidence Intervals, and Meta-Analysis*, (New York: Routledge, 2012), p. 5.

⁷ Stephens-Davidowitz, Seth. *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are*, (New York: HarperCollins, 2017), p. 5.

-
- ⁸ Stephens-Davidowitz, Seth. *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are*, (New York: HarperCollins, 2017), p. 9.
- ⁹ Stephens-Davidowitz, Seth. *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are*, (New York: HarperCollins, 2017), p. 9.
- ¹⁰ Stephens-Davidowitz, Seth. *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are*, (New York: HarperCollins, 2017), p. 11.
- ¹¹ Stephens-Davidowitz, Seth. *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are*, (New York: HarperCollins, 2017), p. 211.
- ¹² Cathy O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, (New York: Crown, 2016), p. 7.
- ¹³ Cathy O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, (New York: Crown, 2016), p. 199.
- ¹⁴ Jo Craven McGinty. “The Consequences of Asking the Citizenship Question,” *The Wall Street Journal*, May 31, 2019. <https://www.wsj.com/articles/could-a-census-citizenship-question-lead-to-an-undercount-11559304000>
- ¹⁵ Alexander Tuttle, Sarasa Tohyama, Tim Ramay, Jonathan Kimmelman, Petra Schweinhardt, Gary Bennett, and Jeffrey Mogil.S (December 2015). “Increasing Placebo Responses Over Time in U.S. Clinical Trials of Neuropathic Pain.” *Pain*. December 1, 2015. 156 (12), pp. 2616-26.
[doi:10.1097/j.pain.0000000000000333](https://doi.org/10.1097/j.pain.0000000000000333).
- ¹⁶ <https://www.lyrics.com/lyric/1057163/Pete+Seeger/John+Henry>
- ¹⁷ Yorick Odin, Nathalie Odin, and Pierre Valette-Florence, “Conceptual and Operational Aspects of Brand Loyalty: An Empirical Investigation. *Journal of Business Research*, Vol. 53, 2001, pp. 75-84.
- ¹⁸ Madeleine Gregory, “The Most Precise Atomic Clock Ever Could Change Our Understanding of Physics,” *Vice*, July 17, 2019. https://www.vice.com/en_us/article/bj9qzv/the-most-precise-atomic-clock-ever-could-change-our-understanding-of-physics.
- ¹⁹ Matthew J. Hornsey, Emily A. Harris, and Kelly S. Fielding. “Psychological Roots of Anti-Vaccination Attitudes: A 24-Nation Investigation.” *Health Psychology*, Vol. 37, No. 4, 2018, pp. 3047-315. <https://www.apa.org/pubs/journals/releases/hea-hea0000586.pdf>
- ²⁰ Peter T. Katzmarzyk, Staphanie T. Broyles, et al., “Relationship Between Soft Drink Consumption and Obesity in 9-11 Year Old Children in a Multi-National Study,” *Nutrients*, Vol. 8, No. 12, December 2016. <https://www.mdpi.com/2072-6643/8/12/770>.
- ²¹ ACSH Staff, “If You Accept Science, You Accept RoundUp Does Not Cause Cancer,” American Council on Science and Health, October 9, 2018. <https://www.acsh.org/news/2018/10/09/if-you-accept-science-you-accept-roundup-does-not-cause-cancer-13490>
- ²² Stacy Malkan, “Nina Fedoroff: Mobilizing The Authority of American Science to Back Monsanto,” *U.S. Right To Know*. Posted June 18, 2019. <https://usrtk.org/tag/acsh/>.
- ²³ Emily Moon, “Monsanto Was Ordered to Pay \$2 Billion in Cancer Lawsuit. There are 13,000 More Plaintiffs,” *PacificStandard*. <https://psmag.com/news/monsanto-ordered-to-pay-2-billion-in-cancer-lawsuit-13000-more-plaintiffs>.
- ²⁴ David Ogilvy, *Confessions of an Advertising Man*, (New York: Ballantine Books, 1963), p. 87.
- ²⁵ Martin Fleischmann and Stanley Pons, “Electrochemically Induced Nuclear Fusion of Deuterium,” *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*,” Vol. 261, Issue 2,

Part, 1, April 1989, pp. 301-308.

<https://www.sciencedirect.com/science/article/pii/0022072889800063?via%3Dihub>

²⁶ *Publication Manual of the American Psychological Association, Sixth Edition*. (Washington, DC: American Psychological Association, 2015), pp. 30-31.