2020

Clear-Sighted Statistics: Module 1: An Introduction to Statistics

Edward Volchok
CUNY Queensborough Community College

Clear-Sighted Statistics: An OER Textbook

Module 1: An Introduction to Statistics

> To consult with the statistician after the experiment is finished is often merely to ask him to conduct a *post mortem* examination. He can perhaps say what the experiment died of.[1]
>
> -- Ronald A. Fisher

Introduction

The goal of this eBook is to help you understand and use elementary statistical techniques effectively…to help you clearly **SEE**:

- **S**elect the appropriate technique.
- **E**xecute that technique properly.
- **E**xplain the results of the analysis clearly and accurately.

This eBook:

1. Offers step-by-step instructions to help you avoid common mistakes, reduce your anxiety about studying statistics, and help you effectively communicate the findings of a statistical analysis.
2. Provides practical examples of common statistical analyses to show how you can perform the calculations using a paper and a pencil, a hand-held calculator, and Microsoft Excel.
3. Offers examples of how statistics is intentionally and inadvertently misused.

Throughout *Clear-Sighted Statistics*, new vocabulary will be introduced. Definitions of many words that may be new to you are provided with links to dictionary.com. These words will be underlined in blue. Clicking on the underlined word will take you to the word's definition in dictionary.com.

After completing this module, you will better understand:

- What the discipline of statistics is and some of the common misconceptions about it.
- The origins of the science of statistics.
- How the study of statistics should lead you to healthy skepticism and not suffocating, misanthropic cynicism.
- Why learning statistics is important for your future.
- What you need to succeed in an introductory statistics class.

I. What is (are) Statistics? Is it Math?

The science of statistics (statistics as a singular noun) is commonly referred to as the science involved with the collection, organization, analysis, interpretation, and presentation of data. This discipline is applied widely by people seeking solutions to scientific, social, psychological, economic, legal, and political problems. Some people go so far as to call this discipline "mathematical statistics." One reason people do this is to distinguish the discipline of statistics from another meaning of this word when it is used as a plural noun: The numerical data we collect from samples are called statistics (plural noun). **Please note:** Data is also a plural noun, a single piece of information is a datum.

I remember reading this basic definition of statistics during the first week of my sophomore year in college when I was taking my required statistics class. Quite frankly, the mention of mathematics terrified me. I was, like many of my readers, math phobic. I had hated math since seventh grade when I was introduced to algebra. While I enjoyed my elementary school arithmetic classes, especially when we solved word problems, I found the abstraction of "x" and "y" in my algebra class confusing. I was bored. I spent most of my time in my junior high school, high school, and college math classes daydreaming. Needless to say, my daydreams had nothing to do with solving quadratic equations.

I want to ease your math anxiety as you start your introductory statistics class. To quote John Wilder Tukey, the great statistician from Princeton University and Bell Labs, "Statistics is a science, not a branch of mathematics, but [it] uses mathematical models as an essential tool."[2] Elsewhere Tukey adds, "Statistics is a science in my opinion, and it is no more a branch of mathematics than are physics, chemistry, and economics; for if its methods fail the test of experience—not the test of logic—they are discarded."[3]

Unlike statistics, academic mathematics deals with idealized abstractions that nearly always have absolute solutions. It is about logical reasoning, patterns, and optimization. It is the science of deducing complex consequences from relatively simple axioms. Mathematics is a deductive system that involves four things:

1. A set of primitive undefined terms.
2. Definitions evolved from these terms.
3. Axioms or postulates.
4. Theorems and their proofs.[4]

Statistics, on the other hand, is inductive. Statistics does not deal with idealized abstractions. Real-world observations are viewed as evidence for the veracity—the accuracy—of a conclusion. *Clear-Sighted Statistics*, therefore, will not present statistical problems with sleep-inducing mathematical abstractions.

Statistics can be defined as the *science of making educated guesses* about how observable phenomena work. Unlike a mathematician dealing with theorems, users of statistical techniques are never 100 percent certain about the accuracy of their conclusions. We use statistical methods to address practical, real-world problems. Our goal is often to help decision-makers make better decisions. We are always aware that new evidence may overturn our assumptions and our conclusions. Whenever evidence shows that our

hypotheses are wrong, we reject them. We will review how this is done in the modules on Null Hypothesis Significance Testing (NHST).

To be sure, the foundation of statistics is built on very sophisticated mathematics. But the mathematics you need to succeed in an introductory statistics class is mostly basic arithmetic, stuff that you were introduced to by the time you completed sixth or seventh grade.

To succeed in an introductory statistics class, you need:

1. A "numbers sense" or an understanding of basic concepts of quantities (more and less, larger and smaller), an understanding of the order of numbers, the ability to make number comparisons, and knowledge of the symbols used to represent numbers). This "numbers sense" has been called an "'at homeness' with numbers and an ability to make use of mathematical skills which enables an individual to cope with the practical demands of everyday life."[5]
2. Proficiency with the basic arithmetic operations of addition, subtraction, multiplication, and division.
3. Competence using the mathematical order operations (PEMDAS).
4. A facility dealing with fractions, decimals, and percentages.
5. The ability to perform calculations involving negative numbers.
6. The ability to round numbers.
7. A grasp of exponents and square roots.

You should have learned all these skills by the time you complete middle school if not sooner. Do not be alarmed if your skills are a bit rusty or even under-developed. If you need to brush up, please review Appendix 1: Basic Math Review.

As result of watching some introductory statistics students struggle while others succeed the following suggestion, is offered: When we visualize data, it is much easier to bring to light the meaning of the data. Throughout *Clear-Sighted Statistics*, data

visualization is used to help you clearly see how to analyze data and make the results of your analyses visible.

Some good news: After teaching introductory statistics classes for fourteen years to over 2,000 students in over sixty classes, I know that with a little effort, students can do well in their statistics class.

The study of statistics is not beyond the abilities of people who have not completed or even started high school. Fostering statistical literacy among junior high school students is, in fact, a stated goal of the Ministry of Education of the People's Republic of China. At a 2004 international conference on statistics education, Li Jun of the East China Normal University, commenting on teaching statistics to junior high school students, declared: "The MOE [Chinese Ministry of Education] emphasized the following three components of statistical literacy [for junior high school students]:

1. Familiarity with using statistical thinking to deal with problems containing data.
2. Appreciating the role statistics plays in decision making by going through the process of collecting, displaying, and analyzing data.
3. Being able to read data resources, data analyses, and summarized information critically."[6]

If junior high school students can develop statistical literacy, you can too.

There are two main branches of the science of statistics: Descriptive Statistics and Inferential Statistics.

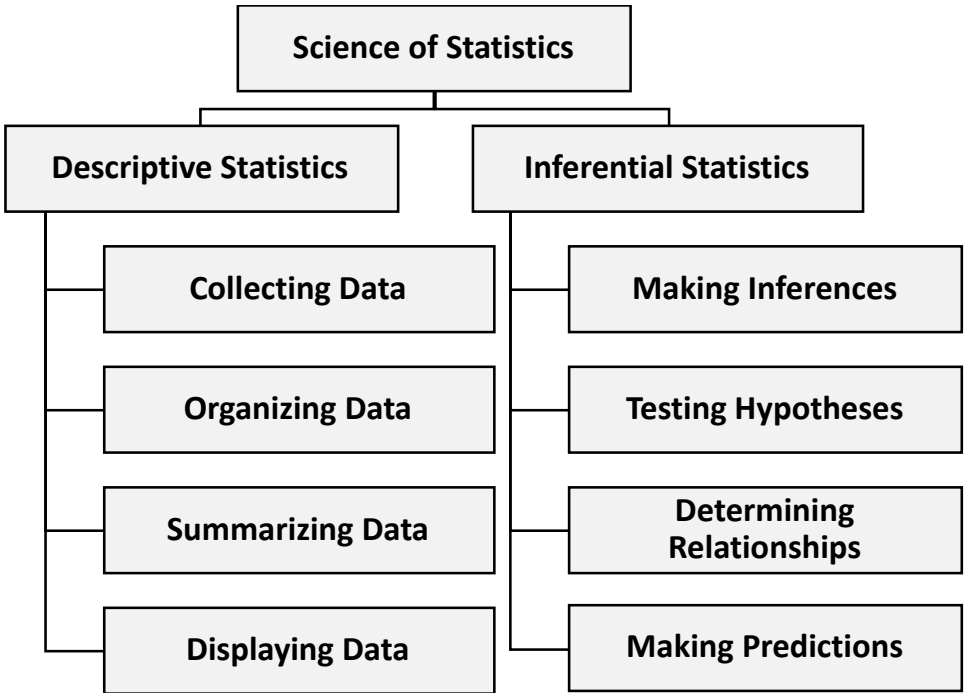


Figure 1: Descriptive and Inferential Statistics

1. **Descriptive statistics** deals with the collection, organization, and summation of data in tables and charts. Descriptive statistics will be covered in Modules 2, 4, 5, and 6.
2. **Inferential statistics** deals with making inferences or estimates about a population based on data derived from samples: Confidence Intervals, Null Hypothesis Significance Testing (NHST), Determining relationships among different data sets, and making predictions. The foundation of inferential statistics is probability theory, which we will cover in Module 7. We will turn to inferential statistics starting in Module 8. An important issue that we will cover is sampling error. Sampling error occurs when the sample measure differs from the usually unknown population measure. As we shall see, sampling error is not the result of a human mistake. It is part of the sampling process.

Populations and Samples: A population represents the total number of persons or things we are investigating. A sample is a small part of a population. As noted above, data derived from samples are called *statistics*. Data taken from the entire population are called *parameters*. We will discuss sampling in Module 2: Statistics and the Types of Data in Module 10: Sampling and Sample Errors. Inferential statistics is based on probability, or the likelihood that something will happen.

II. A Quick History of Statistics

We can trace its origins to ancient times, but the modern science of statistics grew out of developments in the eighteenth and nineteenth centuries.[7] Three historical factors gave rise to the discipline of statistics[8]:

1. The codification of probability theory in the seventeenth century.
2. The rise of the modern nation-state.
3. Government administrators' growing interest in basing policy decisions on accurate information of economic, political, and social phenomena.

The link between statistics and the nation-state can be found in the etymology (the origin) of the word statistics. In Latin, the word *statisticum* means "of the state." In Italian, *statistica* means "statesman." It is not unheard of for people to use the word statesman when they deliver a eulogy for a great statistician.[9] Statistical analysis in the seventeenth and eighteenth centuries was often called "political arithmetic."[10]

Statistics and social science got a boost in the early nineteenth century from the French philosopher and proto-sociologist, Auguste Comte, and the Belgian astronomer, Adolphe Quételet. Both men used statistics to study social phenomena. They called their techniques "social physics."[11] Quételet was the first to apply the bell-shaped curve, or the

normal curve, to social matters. Comte disagreed with how Quételet collected his statistics. He also started to call his discipline sociology. We will study the normal curve in several modules of *Clear-Sighted Statistics*.

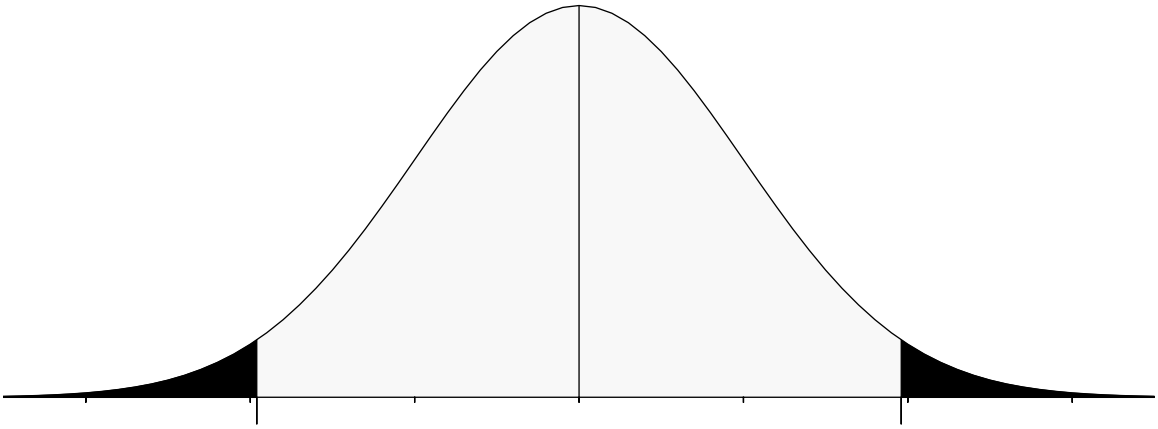


Figure 2: The Normal or Bell-Shaped Curve

Florence Nightingale, another great nineteenth century pioneer of statistics, wrote in the margin of her copy of Quételet's *Physique Sociale*, "[Statistics is] the most important science in the whole world: for upon it depends the practical application of every other [science]."[12] Similarly, Karl Pearson, the influential English statistician who developed many of the techniques that are always covered in an introductory statistics class, published a book in 1892 entitled *The Grammar of Science*. In this work, Pearson argues that "...the Theory of Measurement, Theory of Errors, Theory of Probability, Theory of Statistics [essentially the modern science of statistics]…" were developed to enable scientific measurements.[13]

III. Statistics: Some Common Associations and Misconceptions

The acclaimed American novelist and humorist, Mark Twain made an eminently memorable comment about statistics in his lengthy three-volume autobiography.

> "Figures often beguile [mislead] me," Twain declared, "particularly when I have the arranging of them myself; in which case the remark attributed to Disraeli [Benjamin Disraeli was a prime minister of the United Kingdom in 1868 and again 1874 to 1880] would often apply with justice and force: There are three kinds of lies: lies, damned lies, and statistics."[14]

Twain's comment underscores the well-known link between statistics and lies. There are over 25 Statistics books on Amazon with "lies" in the title. This tally does not include multiple editions of the same book.[15] I should point out that Disraeli most likely never said "lies, damned lies, and statistics."[16] We call this a spurious quotation.

The second common thought about statistics embedded in Twain's opinion is that statistics is difficult. Advanced statistics, in fact, is difficult. Fortunately, introductory statistics, as I have said, is relatively simple. Handheld calculators and computer programs like Microsoft Excel make the statistical calculations relatively easy.

We will review both perceptions of statistics starting with its association with lies. We will then move onto the question of statistics being difficult and how this can undermine our attempts to arrive at accurate conclusions from our analyses.

A) Statistics and Lies

A common perception of statistics is that it is associated with lies. The Austrian scientist and statistician, Richard von Mises echoed Twain's comments about statistics and lies in his famous lecture entitled "The Definition of Probability":

> …may I quote a remark I once heard: "There are three kinds of lies: white lies, which are justifiable; common lies—these have no justification; and statistics." …a great deal of meaningless and unfounded talk is presented to the public in the name of statistics....my purpose is to show that, starting from statistical observations and applying to them a clear and precise concept of probability it is possible to arrive at conclusions that are as reliable and 'truthful' as those obtained in any other exact science.[17]

It is not surprising that devious and dishonest people lie and mislead with statistics. Numbers are powerful. Numbers can confound us. In his book *Proofiness*, New York University Professor Charles Seife writes, "Numbers, figures, and graphs all have the aura of perfection." "If you want to get people to believe something really, really stupid," Seife adds, "just stick a number on it."[18] We see such lies in business and politics all of the time. Authoritarian leaders lie constantly. Demagogues corrode the stability of democratic governments when they resort to feeding citizens a steady diet of lies. After including numbers in a lie, the tip from the dictator's handbook is to repeat the lie *ad nauseam*. "[A]ll effective propaganda must be ... persistently repeated until the very last individual has come to grasp the idea that has been put forward."[19]

It is not a great leap to argue that a democratic society demands statistically literate citizens if self-government is to survive in our data-drenched world. Robert Hooke, author of *How to Tell the Liars from the Statisticians*, makes this argument:

> The science of statistics has made great progress in this [20th] century, but *progress has been accompanied by a corresponding increase in the misuse of statistics* [italics added]. The public, whether it gets its information from television, newspapers, or news magazines, is not well prepared to defend itself against those who would manipulate it with statistical arguments. Many people either believe everything they hear or come to believe nothing statistical, which is even worse.[20]

Hooke's argument is not unique. Many have made it before and since. In fact, Karl Pearson drew this link in 1892 in the first few pages of *The Grammar of Science*, "…it cannot fail to be suggestive as to methods of eliminating individual bias; it [the scientific method and statistics] ought to be one of the best training grounds for citizenship."[21]

Data can be used to reveal or conceal facts. Professor Aaron Levenstein once remarked, ""Statistics are like bikinis. What they reveal is suggestive, but what they conceal

is vital."[22] When dealing with statistical analysis, we need to know what is being concealed and what is being revealed.

Throughout *Clear-Sighted Statistics*, we will discuss the liars and cheats who willfully distort data to suit their own ends and the suckers they prey upon. We will refer to this two broad groups of people as knaves and fools. We do not see the word "knave" used a lot today. A knave is an unprincipled, untrustworthy, or dishonest person. There are a host of synonyms for this word: scoundrel, blackguard, fraud, scallywag, lowlife, and rapscallion to name just a few. The jacks on playing cards are sometimes called knaves.

Figure 3: The Four Jacks (Knaves)

There are also many terms for fools: few: Suckers, victims, numskulls, schlemiels, lamebrains, nitwits, stooges, dupes, and easy marks.

There have always been knaves. David Hume, one of the most important philosophers of the eighteenth century, warned his readers about the ubiquity of knaves in his essay, "On the Independence of Parliament." Hume declared, "Political writers have established it as a maxim that…every man ought to be supposed a knave and to have no other end, in all his actions, than private interest."[23] This sentiment was shared by Edmund Burke, the Anglo-Irish political philosopher and member of parliament, who wrote, "The credulity [the capacity of being easily deceived or cheated] of dupes is as inexhaustible as the invention [false statements] of knaves."[24] While Burke, and to a lesser extent Hume, are considered conservatives, liberals also bemoan the presence of knaves and fools. One such contemporary liberal is Robert B. Reich, Chancellor's Professor of Public Policy at the

University of California at Berkeley and Senior Fellow at the Blum Center for Developing Economies. He served as Secretary of Labor under President Clinton. In his blog, Reich wrote, "There are two kinds of liars—fools and knaves. Fools lie because they don't know the truth. Knaves lie because they intend to mislead."[25]

Harry G. Frankfurt, a retired professor of philosophy from Princeton University, puts an interesting spin on the pervasive use of misinformation in his little book, *On Bullshit*. "One of the most salient features of our culture," Professor Frankfurt declares, "is that there is so much bullshit."[26] Frankfurt even distinguishes a liar from a bullshitter. The bullshitter's statements may be true or false; but, unlike the liar, the bullshitter has little interest in the truth.[27] Because of this, the bullshitter is more dangerous than the liar. "The bullshitter ignores these demands [the demands of truth] altogether. He does not reject the authority of the truth, as the liar does, and opposes himself to it. He pays no attention to it at all. By virtue of this, bullshit is a greater enemy of the truth than lies are."[28]

While we need not trouble ourselves with the subtle differences between the liar and the bullshitter, we may wonder which is the lessor of two evils: Being a knave or a fool. Fools fall prey to knaves and get cheated. Fools also do damage to society by believing and spreading nonsense about how things work in order social, cultural, economic, and political lives. Knaves, on the other hand, often rise to the highest levels of power, prestige, and wealth. Sometimes they also end up in prison. I suspect, indeed hope, that you want to avoid the traps of being either a knave or fool.

Throughout *Clear-Sighted Statistics* we shall play a game called "Knave or Fool." In this game, I will present statistical analyses. You will determine whether the authors are knaves, fools, or honest people who have conducted the analysis properly.

B) Statistics is Hard

Twain's second point is that statistics is hard, especially when we do it ourselves. I have already stated that introductory statistics is not hard. Advanced statistics, however, can be difficult. Advanced statistical methods like multiple regression, probit and logistic regression, factor analysis, and structural equation modeling are difficult to use properly even with the assistance of powerful software like Stata, SPSS, or SAS. Statisticians, in fact, often disagree on the proper use of advanced methods. This led John Wilder Tukey to make the following joke about contentious disagreements among statisticians. The collective noun for a group of birds is called a flock, a group of fish is called a school, and a group of wolves is called a pack. **Question:** What is the collective noun for a group of statisticians?" Tukey's answer: A quarrel.[29] Other commentators, who downplay the contentious nature of disputes among statisticians but recognize that statisticians often disagree, have suggested the word *variance* to describe a collection of statisticians. Variance, as you will learn, is a measure of how far data are dispersed.

It is very easy to find evidence supporting the notion that advanced statistics is very challenging. All we have to do is look at the numerous studies on the high proportion of academic studies published in leading peer-reviewed journals that contain serious statistical errors. In 2014, Dutch researcher Coosje L. S. Veldkamp and four of her colleagues wrote an article entitled "Statistical Reporting Errors and Collaboration on Statistical Analysis in Psychological Science." After reviewing 697 articles published in six leading psychological journals, they found that 63 percent had statistical errors.[30] Writing in the *Journal of Thoracic Disease*, Matthew S. Thiese, Skyler Walker, and Jenna Lindsey found that 78 percent of the 91 published comparative surgical papers contained

potentially meaningful errors in the application of analytical statistics.[31] Spanish researchers Emili García-Berthou and Carles Alcara found that 38 percent of the papers published in *Nature* and 25 percent published in the *British Medical Journal* had serious statistical errors.[32] Ironically, two years after this article appeared, Monwhea Jeng, a physicist at Syracuse University, published a response in the same journal critical of the statistics presented by García-Berthou and Alcara. Jeng argued that the Spaniards used the wrong statistical analysis and, as a result, their findings were wrong.[33]

The misuse of statistics and methodological shortcomings has lead John Ioannidis, a leading medical researcher and professor of medicine and statistics at Stanford University, to write, "There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims."[34]

At this point you might feel like throwing up your hands and screaming, "Statistics! It is all…malarkey! If scholars with their advanced degrees who publish in the world's leading peer-reviewed journals cannot get their statistics right, how can I, a mere undergraduate, be expected to learn this stuff?"

Relax. You are not expected to develop into a professional statistician after completing an introductory statistics class. I shall offer reasons why you should learn basic statistics. But first, a brief lesson in classical Greek philosophy is warranted to help you adopt the appropriate "worldview" or a comprehensive concept of the world when using statistics. To impress your friends, you can even use the German word for worldview, *weltanschauung*.

IV. Cynicism and Skepticism

Western philosophy was founded in Ancient Greece. We are going to quickly highlight the worldview of two schools of Greek philosophy: Cynicism and Skepticism. The Cynicism is to be avoided while Skepticism is to be embraced.

A) Cynicism and Diogenes of Sinope (c. 412 to c. 323 BCE)

Diogenes is often called the dog philosopher. The word "cynic" comes from the ancient Greek word for dog-like. Many of Diogenes' contemporaries thought him mad. Diogenes was born in the town of Sinope, now part of Turkey. His father, it is said, earned his living minting coins. Diogenes was exiled from Sinope for defacing those coins.

A statue of Diogenes was erected in his hometown in 2006 depicting him standing on a barrel while holding a lantern. A dog accompanies him. The barrel is shown because the philosopher eschewed creature comforts. It is said that he lived in an abandoned barrel down by the river. The dog is shown because Diogenes contemporaries thought his manners were no better than a stray dog's. He lacked respect for public decorum and had poor personal hygiene. He habitually relieved himself in public. The lantern is shown because he used to walk the streets of Athens in broad daylight holding a lantern. When asked why, he would snap back, "I'm looking for an honest man." This, of course, was a joke because Diogenes thought that there were no honest men.

Here are the central tenets of Diogenes' cynical worldview:

- People are dishonest and greedy.
- Society is incompatible with true happiness.
- Following social conventions and striving for wealth and physical comfort lead to misery.
- There is no truth.

A famous incident regarding the elderly Diogenes meeting the young Alexander the Great has been told by Plutarch in the *Life of Alexander* and Diogenes Laërtius in the *Lives of Eminent Philosophers, Book VI*. This story underscores Diogenes' contempt for authority and convention. The meeting took place in Corinth, the city where the elderly Diogenes had been residing. At barely twenty-years-old, Alexander was the brash young king of Macedonia. His late father, Philip II of Macedonia had recently conquered Sparta, Athens, and the rest of the Greek city-states, which were located to the south of Macedonia. Aristotle, who was arguably the greatest of all ancient Greek philosophers, was Alexander's tutor.

The young king, who was planning his successful invasion of the mighty Persian Empire, went to Diogenes' home; yes, the barrel along the riverbank. When the two men met, Diogenes was reclining in the sun while leaning against his barrel. The young king, the most powerful man in Greece, asked the philosopher if there was anything he could do for him. Diogenes snarled, "Yes, move! You're blocking my sunlight." Alexander's entourage wanted to punish Diogenes for this public display of contempt for their king, but Alexander was said to be amused by Diogenes' insolence saying something along the lines of: "If I were not Alexander, I would want to be Diogenes."[35]

Diogenes' view that people are inherently dishonest and there is no such thing as truth is not uncommon today. It is said that we live in a post-truth era.[36] The rhetoric used in political discourse has become divorced from the truth. Unfounded conspiracy theories are given widespread credence by highly paid personalities in the media and people holding high political office. Politicians say anything, without regard for the truth, in hopes of appealing to people's biases, fears, and emotions.[37] These leaders will do or say

whatever they feel will get them what they want. If a politician does not like what a journalist writes, he declares the news fake. Eager followers cheer him on.[38]

This cynicism leads to disaster because it undermines the public's trust in our institutions. It is my hope that by studying subjects like statistics, you will be protected from what philosopher and scholar of totalitarianism, Hannah Arendt called the "gullibility and cynicism" that leads the masses to "believe everything and nothing."[39]

B) Skepticism

Skeptic philosophers in ancient Greece include: Xenophanes (c. 570 – c. 472 BCE), Democritus (c. 460 – c. 370 BCE), Gorgias (c. 483 – c. 375 BCE), and Pyrrho of Elis (c. 360 – 270 BCE). The word skepticism is derived from the ancient Greek word for investigation. The common thread among skeptics is that they are investigators. They generally believe that truth while uncertain can be uncovered by carefully applying reason.

Statistics should lead us to healthy skepticism where truth is considered uncertain, but with the application of appropriate techniques we can uncover the ways our world actually works. We can do this by collecting evidence to answer the questions:

- Who?
- What?
- When?
- Where?
- Why?

The skeptic's worldview is embedded in Richard von Mises' comment: "Starting from statistical observations and applying to them a clear and precise concept of probability it is possible to arrive at conclusions that are as reliable and 'truthful' as those

obtained in any other exact science." We also see the attitude of a skeptic in a statement made by physicist Richard P. Feynman in his book, *QED: The Strange Theory of Light*. Feynman writes, "Nature permits us to calculate only probabilities. Yet science has not collapsed."[40] As we shall discover in later modules, probabilities are not precise and our understanding of our world needs repeated verification.

The skepticism of the scientist was best expressed by the poet Walt Whitman. Whitman's statement as recorded by his friend Horace Traubel. Here is the great poet's remark:

> I like the scientific spirit—the holding off, the being sure but not too sure, the willingness to surrender ideas when the evidence is against them: this is ultimately fine—it always keeps the way beyond open—always gives life, thought, affection, the whole man, a chance to try over again after a mistake—after a wrong guess.[41]

V. Why You Need to Learn Statistics

There are many reasons why you need to develop your quantitative, information, and statistical literacies. Advocates of democratic education going back as far as Thomas Jefferson have pushed for the education of citizens.[42] With the rise of authoritarian populists in many democratic societies, these literacies are more important than ever. You will have to navigate political, economic, and personal life by making important decisions about data. A person who has developed these literacies will understand such economic concerns as payday lending, credit card debt, servicing student loans, or deciding on which career to pursue. He or she will understand the effect of gerrymandered districts on the kinds of laws enacted and will comprehend problems involving health, education, and family life. Thus, statistically literate citizens are more likely to see through demagogues' lies, and may also be less susceptible to resentments demagogues play upon.[43]

Becoming comfortable with data and statistics will help you avoid becoming someone's dupe. It will help you achieve your personal and financial goals. Here is what John Pullinger, the past president of the United Kingdom's Royal Statistical Society, said about the importance of learning statistics: "In the data rich world…those nations, governments, businesses and individuals who use the power of numbers will prosper. Those who 'get statistics will get on. Those who do not, will get left behind." Do not get left behind. Learn statistics!

We are going to conclude Module 1 with a little exercise about which careers offer students more money. I have used this exercise during the first class of the semester. I found it in the January 2006 issue of the journal *Teaching Sociology*.[44] The article, written by Pamela Paxton, is entitled "Dollars and Sense." A word of caution about the starting salaries presented: The data were collected from 1996 to 2003. It is reasonable to expect that these salaries have increased over time. We will not address the question of whether these salaries have kept pace with inflation.

Figure 3 shows a table of twelve job titles listed in the article with the starting salaries and whether a knowledge of statistics is required.

	Job Title	Statistics Needed	Starting Salary
1	Community Organizer	No	$14,954
2	Corrections Officer	No	$27,000
3	Criminologist (Counseling)	No	$23,963
4	Criminologist (Research)	Yes	$31,500
5	Human Resources	No	$27,040
6	Marketing Research Analyst	Yes	$32,000
7	Marketing Research Interviewer	No	$12,000
8	Private Detective	No	$26,750
9	Probation Officer	No	$23,594
10	Public Opinion Researcher (w/Statistics)	Yes	$27,101
11	Public Opinion Researcher (No Statistics)	No	$16,500
12	Public Relations Specialist	No	$21,325
13	Residential Counselor	No	$19,680
14	Social Worker	No	$23,147

Source: *Chronicle Guidence Publications*, The Bureau of Labor Statistics and the Nation Association of Colleges and Employers. Salary are from 1996 to 2003. Pamela Paxton, "Dollars and Sense," *Teaching Sociology*, Vol. 34. January 2006, pp. 65-70.

Figure 4: Starting Sales list by Job Title in Alphabetical Order

As you can see, the job titles are listed in alphabetical order. This makes it difficult to determine whether the jobs that require statistics pay higher salaries than those that do not. We should reorder the table based on salary:

	Job Title	Statistics Needed	Starting Salary
1	Marketing Research Analyst	Yes	$32,000
2	Criminologist (Research)	Yes	$31,500
3	Public Opinion Researcher (w/Statistics)	Yes	$27,101
4	Human Resources	No	$27,040
5	Corrections Officer	No	$27,000
6	Private Detective	No	$26,750
7	Criminologist (Counseling)	No	$23,963
8	Probation Officer	No	$23,594
9	Social Worker	No	$23,147
10	Public Relations Specialist	No	$21,325
11	Residential Counselor	No	$19,680
12	Public Opinion Researcher (No Statistics)	No	$16,500
13	Community Organizer	No	$14,954
14	Marketing Research Interviewer	No	$12,000

Source: *Chronicle Guidence Publications*, The Bureau of Labor Statistics and the Nation Association of Colleges and Employers. Salary are from 1996 to 2003. Pamela Paxton, "Dollars and Sense," *Teaching Sociology*, Vol. 34. January 2006, pp. 65-70.

Figure 5: Starting Sales Listing in Descending Order of Salary

By sorting the table on salary, the fact becomes very clear that the starting salaries for jobs that require statistics are higher than those that do not. In fact, the average salary for a job that demands statistics is $30,200 versus $21,450 for those jobs that do not require statistics. The difference is $8,750. Jobs requiring knowledge of statistics pay nearly 41%

more than non-statistics jobs. Although this quick comparison lacks the rigor that we will demand when we conduct Null Hypothesis Significance Tests, the implication is clear: If you want to earn more money, learn statistics.

V. A Final Thought

I have repeatedly said that introductory statistics is relatively easy. This is true only if you put in the necessary effort. You need to do more than just passively listen in class, casually read this book the night before the exam, and take the exam without having practiced of the techniques covered in class. Here is a suggestion on how to do well in all of your classes. It comes from an old Chinese proverb:

> Not hearing is not as good as hearing.
> Hearing is not as good as seeing.
> Seeing is not as good as knowing.
> Knowing is not as good as acting.
> True learning continues until it is put into action.

You must act. The act you must perform is solving statistical problems on your own. Do the examples in this book. Do the problems your instructor covered in class.

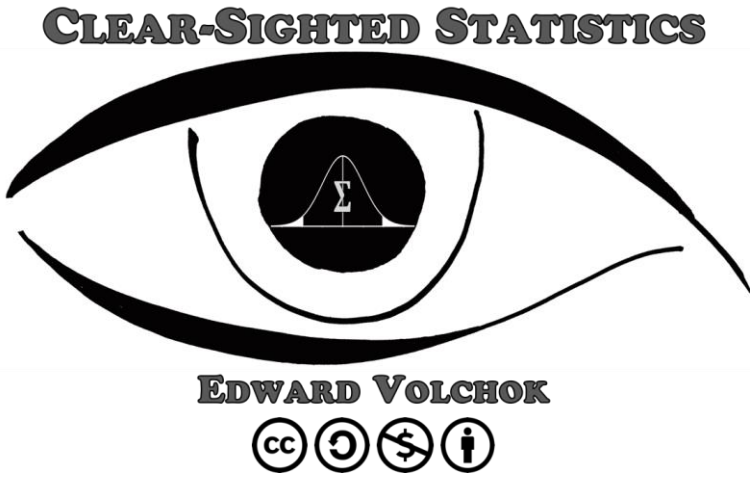
VI. Exercises

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

1. **What does the word "statistics" mean?**
2. **Is Statistics a branch of Mathematics?**
3. **How much math do I need to pass an introductory statistics class?**
4. **What is the difference between a Population and a Sample?**
5. **What is the difference between Descriptive and Inferential Statistics?**
6. **Why does Statistics lead to healthy skepticism?**
7. **Why should I want to learn Statistics?**

8. **True or False: Do statisticians tend to argue among themselves?**

* * *



* * *

---

[1] Ronald A. Fisher, "Presidential Address," *Sankhyā: The Indian Journal of Statistics*, Volume 4, No. 1, 1938, p. 17.

[2] J. W. Tukey quoted by David R. Brillinger in "John Wilder Tukey (2015 – 2000)." https://www.stat.berkeley.edu/~brill/Papers/boas.pdf

[3] L. V. Jones (ed.) The Collected Works of John W. Tukey: Philosophy and Principles of Data Analysis 1949-1964, Volume III. (New York: Chapman & Hall, 1986), p. 66.

[4] Liaqat Aki Khan, "What is Mathematics – An Overview," *International Journal if Mathematics and Computational Science*. Vol. 1, No. 3, 2015. p. 98. Link to "What is Mathematics – An Overview".

[5] W. H. Cockcroft. *Mathematics Counts: Report of the Committee of Inquiry into the Teaching of Mathematics in Schools Under the Chairmanship of Dr. WH Cockcroft*. (London: Her Majesty's Stationery Office, 1982), p. 11. Link to Mathematics Counts.

[6] Li Jun, "Statistics Education for Junior High School Students." In *Curricular Development in Statistics Education: International Association for Statistical Education(IASE) Roundtable*, edited by Gail Burrill and Mike Camden. 2004, p. 221. https://iase-web.org/documents/papers/rt2004/5.2_Li.pdf.

[7] https://wol-prod-cdn.literatumonline.com/cms/attachment/8c216a3d-882b-4324-83ce-2ecdc84c89cd/sign707-gra-0001-m.jpg

[8] Helen M. Walker, *Studies in the History of Statistical Method: With Special Reference to Certain Educational Problems*. (Philadelphia: Williams & Wilkins, 1929).

[9] Frederick Mosteller. "Samuel S. Wilks: Statesman of Statistics." *The American Statistician*, Vol. 18, No. 2, April 1964, p. 15.

[10] Ted McCormick, "Political Arithmetic and Sacred History: Population Thought in the English Enlightenment, 1660-1750." *Journal of British Studies*, Vol. 52, No. 4. October 2013. Pp. 829-857. JSTOR, www.jstor.org/stabke/2400883.

[11] Adolphe Quételet, *Sur l'homme et le développement de ses facultés, ou Essai de physique sociale*. (Paris: Bachelier, Imprimeur-Libraire, 1835). Auguste Comte. Social Physics: From the Positive Philosophy of Auguste Comte. (Miami, FL: HardPressNet, 2015).

[12] Nightingale's annotations to Adolphe Quételet's Physique Sociale cited in "Florence Nightingale," I. Bernard Cohen, *Scientific American*, 250(3): p. 136, March 1984.

[13] Karl Pearson, *The Grammar of Science, 2nd edition*, (London: Adam and Charles Black, 1900), p. 516. Kindle location 9123.

[14] Mark Twain, *The Project Gutenberg eBook of Chapters from My Autobiography*. Chapter 20. 2016. http://www.gutenberg.org/files/19987/19987-h/19987-h.htm. 2016. Location 141460.

[15] Edward Volchok, "Social Justice, Numeracy, and Teaching Statistics at a Community College," *Numeracy*. Volume 1, Issue 1, 2019. p. 7. https://doi.org/10.5038/1936-4660.12.1.6.

[16] https://www.york.ac.uk/depts/maths/histstat/lies.htm

[17] Richard von Mises. *Probability, Statistics and* Truth. (New York: Dover Publications, 1981), p. 1.

[18] Charles Seife. p. 7.

[19] Adolf Hitler. *Mein Kampf*. Translated by James Murphy. (London, UK: Hurst and Blackett, Ltd., 1939), p. 148. *https://mk.christogenea.org/_files/Adolf%20Hitler%20-%20Mein%20Kampf%20english%20translation%20unexpurgated%201939.pdf.*

[20] Robert Hooke, *How to Tell the Liars from the Statisticians*. (New York: Marcel Dekker, 1983), p. vii.

[21] Karl Pearson, *The Grammar of Science, 2nd edition*, (London: Adam and Charles Black, 1900), p. 6. Kindle location 376.

[22] Cited in Alan Murray (2010), *The Wall Street Journal Essential Guide of Management*. (New York: HarperCollins, 2010). Chapter 9, location 1862.

[23] David Hume. "On the Independence of Parliament." 1742. https://ebooks.adelaide.edu.au/h/hume/david/of-the-independency-of-parliament/ also http://press-pubs.uchicago.edu/founders/documents/v1ch11s4.html.

[24] Edmund Burke. "A Letter From Mr. Edmund Burke, to A Member of the National Assembly in Answer to Some Objections to His Book on French Affair." *The Works of the Right Honourable Edmund Burke, Volume I*. (London: Holdsworth and Ball, 1836). p. 478. https://books.google.com/books?id=ezOUC_XUyzoC&pg=PA478&lpg=PA478&dq=The+credulity+of+dupes+is+as+inexhaustible+as+the+invention+of+knaves.+source&source=bl&ots=KodjjZbqSe&sig=ACfU3U1OVWFX39kKy0DDlMlEgF9jAXve1w&hl=en&sa=X&ved=2ahUKEwi3pMContriAhUSxVkKHdxuAGEQ6AEwAXoECAkQAQ#v=onepage&q=credulity&f=false

[25] Robert Reich, Blog. "Knaves and Fools," November 25, 2017. https://robertreich.org/post/167878608210

[26] Harry G. Frankfurt, *On Bullshit*. (Princeton, NJ: Princeton University Press, 2015), p. 1.

[27] Frankfurt, p. 55.

[28] Frankfurt, p. 61.

[29] Sharon Bertsch McGrayne, *The Theory that Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. (New Have, CT: Yale University Press, 2011). P. 51. The author footnotes the source as "Tukey, according to Brillinger email." Presumably, these would be John W. Tukey (1915-2000) and his student David R. Brillinger (1937-).

[30] Coosje L. S. Veldkamp, Michèle B. Nuijten, Linda Dominguez-Alvarez, Marcel A. L. M. van Assen, Jelte M. Wicherts, "Statistical Reporting Errors and Collaboration on Statistical Analyses in Psychological Science," *PLOS One*, December 10, 2014. https://doi.org/10.1371/journal.pone.0114876

[31] Matthew S. Thiese, Skyler Walker, Jenna Lindsey. "Truth, Lies, and Statistics." *Journal of Thoracic* Disease. Vol. 9, No. 10, October 2017, p. 4118. https://www.researchgate.net/publication/320749929_Truths_lies_and_statistics.

[32] Emili García-Berthou and Carles Alcaraz. "Incongruence between test statistics and P values in medical papers." *BMC Medical Research Methodology*. 2004, 4:13. May 28, 2004. https://bmcmedresmethodol.biomedcentral.com/track/pdf/10.1186/1471-2288-4-13.

[33] Monwhea Jeng, *BMC Medical Research Methodology*. 2006. 6: 45. January 25, 2006. https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-6-45

[34] Ioannidis, John P. A. "Why Most Published Research Findings Are False." *PLoS Medicine*. 2, 8, August 25, 2005. pp. 0696-0701. https://doi.org/10.1371/journal.pmed.0020124.

[35] Plutarch, "Alexander" in *Plutarch's Lives*, Translator: John Dryden and revised by Hugh Clough, (New York: Modern Library, 1975), p. 810. *Diogenes Laërtius, Lives of the Eminent Philosophers, Book VI*. Paragraph 31. https://en.wikisource.org/wiki/Lives_of_the_Eminent_Philosophers/Book_VI.

[36] Ralph Keyes, *The Post-Truth Era: Dishonesty and Deception in Contemporary Life*. (New York: St. Martin's Press, 2004).

[37] Christina Pazzanese, "Politics in a 'Post-Truth' Age." *The Harvard Gazette*. June 14, 2016. https://news.harvard.edu/gazette/story/2016/07/politics-in-a-post-truth-age/.

[38] Daniel A. Effron. "Why Trump Supporters Don't Mind His Lies." *The New York Times*. April 28, 2018. https://www.nytimes.com/2018/04/28/opinion/sunday/why-trump-supporters-dont-mind-his-lies.html.

[39] Hannah Arendt. *The Origins of Totalitarianism*. (New York, NY: Harcourt Brace & Co., 1979), p. 38.

[40] Richard P. Feynman. *QED: The Strange Theory of Light*. (Princeton, NJ, Princeton University Press, 1985), p. 19

[41] Horace Traubel. *With Walt Whitman in Camden, Vol. 1*. (Boston, MA: Small, Maynard & Company, 1906), p. 101. https://whitmanarchive.org/criticism/disciples/traubel/WWWiC/1/whole.html#p101.2.

[42] Johann Neem. "Is Jefferson a Founding Father of Democratic Education?" *Democracy & Education*, Vol. 21, No. 2. 2013.

[43] Edward Volchok, "Social Justice, Numeracy, and Teaching Statistics at a Community College," *Numeracy*. Volume 1, Issue 1, 2019. p. 1. https://doi.org/10.5038/1936-4660.12.1.6.

[44] Pamela Paxton. Dollars and Sense." *Teaching Sociology*. Vol. 34. January 2006, pp. 65-70.