



## Empiricism at Work on Mars

The planet Mars contains numerous populations that scientists have wanted to learn about since the planet was first discovered. Scientists have been trying to estimate all sorts of things about the planet, including the planet's climate, geology, and history. NASA has sent 14 successful missions to sample various aspects of this planet: four Mariner missions between 1965-1971, two Viking Landers in 1976, Mars Pathfinder and Mars Surveyor in 1997, the Mars Odyssey in 2001, the Mars Exploration Rovers in 2004, the Mars Reconnaissance Orbiter in 2006, Phoenix in 2008, Curiosity in 2012, and MAVEN in 2014. Three "rover" missions (Curiosity in 2012, Insight in 2018, and Perseverance in 2021) have landed on Mars. If you are interested in seeing some of the images or looking at some of the sample data collected from these missions, go to [mars.jpl.nasa.gov](https://mars.jpl.nasa.gov).

term "empirical" (from the Greek word *empeiria*, meaning experience) to the English language. This work and the inductive method were highly influential on the development of science and the evolution of the scientific method.

Induction embraces the philosophical principle "what is true of the many is true of the whole" and can be thought of as bottom-up thinking. All applications of statistical inference use inductive reasoning.

Let's consider an example. Down syndrome (DS) is a relatively common disorder in humans. Geneticists tell us that individuals with DS have three copies of chromosome 21 instead of two. How did they reach this conclusion? A French geneticist in 1959 analyzed chromosomes of individuals with what was believed to be Down syndrome and observed an extra copy of chromosome 21 in each of them.<sup>4</sup> This discovery was later confirmed by further studies using more advanced techniques. Is this absolute proof that an extra copy of chromosome 21 is the only cause of DS? At some time in the future, could a person exhibit DS and not have an extra copy of chromosome 21?

Unlike deductive reasoning, induction is not guaranteed to produce absolutely true conclusions. Thus, inductive inference must be associated with *probable* conclusions and some degree of uncertainty. As the data increases, the more probable the inductive conclusion. Probability is the language of uncertainty and will be used to define the degree of belief we have in our conclusions. Being able to express uncertainty in a precise manner is one of the reasons we study probability in a statistics course before statistical inference.

Inductive reasoning is one of the standards used to determine whether someone's beliefs about the world are "justified." As the statistician R. A. Fisher said, "inductive inference is the only process known to us by which essentially new knowledge comes into the world... (and was a) ... contribution to the intellectual development of mankind".<sup>5</sup> As the philosopher Robert Audi put it, induction is empiricism's "role ... in grounding rationality."<sup>6</sup>

## 1.1 Exercises

### Basic Concepts

1. Complete the sentence: An empirical claim would be one that ....
2. What is the difference between knowledge and belief?
3. Is the idea that three copies of chromosome 21 causes Down Syndrome closer to a belief or to knowledge?
4. Do a Google search to determine the difference between anecdotal evidence and empirical evidence.
5. Restate the sentences to have the same meaning but without using the word "empirical":
  - a. "Only empirical research can decide whether bankruptcy law helps or hurts entrepreneurs."
  - b. "The book reviews the new theory carefully and in language accessible to the general reader, and then subject it to a detailed empirical examination."
  - c. "It was established as an empirical matter, that when average incomes rise, the average incomes of the poorest fifth of society rise proportionately."
  - d. "And empirical estimates are being replaced by mathematical exactness."

## Exercises

6. If a doctor uses an empirical therapy on your illness, what exactly would that mean?
7. Sherlock Holmes pleaded to his friend Watson “Data! Data! Data! I can’t make bricks without clay”. What were the “bricks” and “clay” that Holmes was referring to?
8. Retailers often have customer loyalty or rewards programs in which a person supplies some basic contact and demographic information in exchange for discounts. It probably does not surprise you to learn that the company is collecting data from your shopping preferences and habits. Give examples of the information that might be gathered. How is the “story” that the data tells potentially useful to the retail company?
9. Suppose there were two bags of marbles. In one bag there are 500 marbles and the other bag contains 100 marbles. You reach in and take 50 marbles out of each bag. If all 50 marbles selected from each bag are red, could you reasonably state a hypothesis that every marble in each bag is red? With respect to the conclusion, do you feel there is a difference in the likelihood of the conclusion based on the number of marbles in each bag?
10. Research shows the London taxicab drivers who have to memorize a map of London to get their taxicab license have larger brain regions devoted to spatial, or mapping, memories. Was this empirical research? Was the research deductive or inductive?
11. For the past 10 years geese have come to our pond in May, therefore geese will come every year in May. Is this an example of deduction or induction?
12. Research indicated that the frontal lobes of the brain involve higher cognitive functions such as conscious thought and problem-solving. Do you think these conclusions were reached primarily by induction or deduction?
13. Determine if this is an example of deduction or induction: All second-degree polynomial equations can be solved using the quadratic formula. The solutions to  $3x^2 + 11x - 4 = 0$  are calculated as  $x = \frac{-11 \pm \sqrt{(-11)^2 - 4(3)(-4)}}{(2)(3)} = \frac{-11 \pm 13}{6}$ ,  
 $x = -4$  or  $x = \frac{1}{3}$ .
14. Consider a decision that you recently made. What kind of information did you use to help you to come to a conclusion? Did you use an inductive or a deductive thinking process?

Exercises 15-17 give examples of valid syllogisms which use deductive reasoning. Fill in the missing premise or conclusion.

15. **Major premise:** All birds have two legs.  
**Minor premise:** Pigeons are birds.  
**Conclusion:** \_\_\_\_\_



### René Descartes

René Descartes, a 17<sup>th</sup>-century French philosopher, mathematician, and scientist, is known as the father of modern philosophy. He pioneered the use of skepticism and doubt as a method of inquiry. He famously doubted everything except his own existence in his famous statement “Cogito, ergo sum” (I think, therefore I am), and made significant contributions to mathematics, notably in developing Cartesian geometry.

16. **Major premise:** All plants are green.  
**Minor premise:** \_\_\_\_\_  
**Conclusion:** A cactus is green.
17. **Major premise:** \_\_\_\_\_  
**Minor premise:** You were stung by a bee.  
**Conclusion:** You had an allergic reaction.
18. René Descartes (1596-1650), a French philosopher and mathematician, summarized his belief in the distinction between the mind and body in the quote, “Cogito, ergo sum,” translated to English as “I think, therefore I am.” Complete a valid syllogism related to his statement given that the major premise is, “The act of thinking requires a conscious self.”

## 1.2 Basic Statistical Concepts

Statistics has its roots in empiricism. The practice of empiricism requires a focus—what do you want to learn more about? This focus defines a fundamental concept in statistics, the **population**.

### Population

A **population** is the total set of subjects or things we are interested in studying.

**DEFINITION**

The notion of a population is a very general concept. Populations are defined by what a researcher is studying and can come in all shapes and sizes. If you are researching Hank Aaron’s major league batting performance, then the population would consist of all 12,364 of Aaron’s major league at bats.<sup>7</sup> If someone is studying toucans in Brazil, then all the toucans in Brazil would constitute the population. If you are studying students at your college, then all the students attending your college represent a population.

A list of all members of a population is called a population **frame**.

### Frame

A list containing all members of the population is referred to as a **frame**.

**DEFINITION**

According to the Census Bureau in 2020 there are about 332 million people in the United States.<sup>8</sup> The frame for the population of the United States would be a rather long list containing about 332 million names. Although a previous census would be a good start in developing a frame for the US population, it is doubtful that an exact frame could ever be developed at a given point in time since there is one new birth every 8 seconds, and one death every 12 seconds. There are just too many people being born, dying, and immigrating over a 10-year period to get an exact frame for the US population. But for problems that deal with smaller populations, frames are easily developed. For example, if your statistics class were the population you were studying, the final class roster would be the frame for the population.

