

Definition**Point Estimator**

A **point estimator** is a single-valued estimate, calculated from the sample data, which is intended to be close to the true population value.

Lava Lamps, Randomness, and Your Bank Account

True randomness is something computers are not currently very good at creating. To generate “random” numbers, computers use an algorithm that produces pseudo-random numbers. A simple example of a pseudo-random number generator would be to take the current day of the month and add the current time in minutes to it, then find the digit of π that corresponds to that “seed” number. For example, if the date is 7/4/2020 and the time is 10:48 AM, then the seed number would be 52 and the 52nd digit of π is 5. Therefore, 5 would be the “random” number that this simple pseudo-random number generator returns.

Every digital transaction that occurs relies on encryption to protect your money. Encryption keys are a random string of characters. Encryption relies on non-predictable encryption keys. The problem with a pseudo-random number generator for encryption purposes is that if a hacker knows how the generator works, then they may be able to “predict” the encryption key and thus destroy the security that the encryption provides. If encryption keys are “predictable,” all wire transfer and other online financial transactions would be vulnerable to fraud. For encryption purposes the security of the encryption key is directly associated with the true randomness of its key.

If true randomness is required, the best place to find randomness is in nature.

Cloudflare is an internet security company that uses a wall of lava lamps to generate truly random strings for encryption. The chaotic nature of the fluid dynamics (the bubbles) in the lava lamps makes it impossible to predict what they will look like at any given moment. This enables Cloudflare to take billions of pictures of their wall of lava lamps and generate billions of truly random encryption strings from these pictures.

close to their population counterparts. In other words, the sample mean ought to be close to the population mean, the sample proportion ought to be close to the population proportion, and the sample standard deviation should be close to the population standard deviation. Since sample statistics will be used as the basis of the statistical inference, we must know how those statistics vary from one sample to another. Once the variability of the sample statistic is understood, we will be able to make probability statements regarding our inferences.

Why Calculate the Sample Mean?

When analyzing ratio data, the first piece of summary information that an analyst wants to determine is the mean. For most populations, performing a census to determine the population mean is impractical. The only alternative is to use sample information. It seems reasonable that the sample mean, \bar{x} , would contain an enormous amount of information about the population mean, μ , and would thus be a sensible estimate of the population mean. Generally, if you wish to estimate a population value—be it the mean, standard deviation, or proportion—the corresponding sample value will be a good **point estimator**.

Can you be sure that the sample mean will always be close to the population mean? When dealing with random variables, nothing is certain, but there are methods of reducing the probable error. To understand how this is achieved, we must examine how the sample mean varies. The next section will introduce the distribution of the sample mean and how this distribution can be used to make statistical inferences.

8.1 Exercises**Basic Concepts**

1. Why is the quality of sample data so important?
2. Why is randomness useful in sampling?
3. What is wrong with a voluntary survey?
4. What is a biased sample?
5. What is a sampling frame? Why is this concept important?
6. Discuss how you would draw a simple random sample of the students at your college.
7. What makes drawing a simple random sample from a geographic area a difficult task?
8. Is the sample mean always close to the population mean?
9. Under what conditions is the sample mean considered a random variable?
10. What is the sampling distribution of the sample mean?
11. Describe how statistics as random variables are crucial to statistical inference.
12. What is a point estimator? Give an example.

Exercises

13. A magazine reported the results of a survey in which readers were asked to send in their responses to several questions regarding good eating. Consider the reported results to the question, *How often do you eat chocolate?*

Survey Responses	
Category	% of Responses
Frequently	13
Occasionally	45
Seldom	37
Never	5

- a. Were the responses to this survey obtained using voluntary sampling techniques? Explain your answer.
 - b. What types of biases may be present in the responses?
 - c. Is 13% a reasonable estimate of the proportion of all Americans who eat chocolate frequently? Explain.
14. A magazine reported the results of a survey in which readers were asked to send in their responses to several questions regarding anger. Consider the reported results to the question, *How long do you usually stay angry?*

Survey Responses	
Category	% of Responses
A few hours or less	48
A day	12
Several days	9
A month	1
I hold a grudge indefinitely	22
It depends on the situation	8

- a. Were the responses to this survey obtained using voluntary sampling techniques? Explain your answer.
 - b. What types of biases may be present in the responses?
 - c. Is 22% a reasonable estimate of the proportion of all Americans who hold a grudge indefinitely? Explain.
15. Students in a marketing class have been asked to conduct a survey to determine whether or not there is a demand for an insurance program at a local college. The students decide to randomly select students from the local college and mail them a questionnaire regarding the insurance program. Of the 150 surveys that were mailed, 50 students responded to the following survey item: *Pick the category which best describes your interest in an insurance program.*

Survey Responses	
Category	% of Responses
Very Interested	50
Somewhat Interested	15
Interested	10
Not Very Interested	5
Not At All Interested	20

- a. What types of biases may be present in the responses?
- b. Is 50% a reasonable estimate of the proportion of all students who would be very interested in an insurance program at the local college? Explain.
- c. Is 50% a reasonable estimate of the proportion of all business majors who would be very interested in an insurance program at the local college? Explain.
- d. What strategies do you think the marketing students could have used to get a less biased response to their survey?
- e. Suppose the program was created and only a few people registered. How could the survey question have been reworded to better predict actual enrollment?

16. Television news programs often conduct opinion surveys by announcing some question on the air and advising viewers to call different numbers for a *yes* or *no* response. National television programs do the same thing except they use 900 numbers and the respondent must pay for the call. Suppose that a national news program asks its viewers to phone in a response to the following: *Women should be permitted to assume combat roles in the military*. The results of the particular survey were 34% *yes* and 66% *no*. Is it reasonable to believe that the results of the survey reflect the attitudes of the nation on this issue? What biases exist in this sampling method?
17. A local politician wants to know what the residents of his community think about an increase in the local property tax to pay for improvements to the highway. He decides to conduct a survey.
- What is the population of interest to the politician?
 - Can you think of any good sources for a sampling frame?
 - What are the shortcomings (if any) of the sources you picked for the sampling frame?

8.2 The Distribution of the Sample Mean and the Central Limit Theorem

Sample means vary because sample data vary from sample to sample. As an illustration, suppose that an automobile manufacturer wished to determine the average miles per gallon (mpg) of a specific vehicle model that it manufactures. Since determining the mpg of each vehicle is very time consuming, the manufacturer has decided to select two vehicles from a batch of six. Suppose that the actual mpg of the six vehicles are given in Table 8.2.1.

Car	MPG
A	25
B	27
C	40
D	29
E	28
F	30
Mean	29.8
Variance	23.14
Standard Deviation	4.81

It is important to realize that we are assuming the above set of data constitutes a population. The mean mpg rating of the population in Table 8.2.1 is approximately 29.8 and the population standard deviation is approximately 4.81.

$$\mu \approx 29.8$$

$$\sigma \approx 4.81$$

Both of these measures are considered population parameters. The mpg ratings given in Table 8.2.1 are not known by the manufacturer when the shipment arrives. The manufacturer's job is to estimate the population mean using a sample estimate, in this case using the sample mean from a sample of size two.

How many different samples of size two can be drawn? Assuming no replacement, there would be 15 possible samples of size two if order does not matter. A list of all possible samples and the resulting sample means is given in Table 8.2.2.