Homework:
Bias & Variability

MATH 150

Due: Jan 26, 2024

Sepehr Akbari

Problem 1

In each of the following cases, identify both the parameter and statistic. Do you think the difference between the two is more likely attributable to bias or variability, or might it be impossible to tell? Briefly explain your choice.

(a) The average SAT math score in 2022 was 521. In a simple random sample of 25 test-takers from that year, the average score was 595.

(b) On a certain day, 150 reported gas prices at their local gas stations to a tracking website. The mean price was \$5.53/gallon, while the national average was 4.12.

(c) In ten flips of a particular coin, seven were heads

Answer

(a) The gap between the sample average (595) and population average (521) is likely due to natural variability in the sample. Assuming correct sampling, bias is less probable in a simple random sample.

- Parameter: 521
- Statistic: 595

(b) The difference between the parameter and the statistic is likely due to bias in this case. The sample of 150 individuals reporting gas prices may not represent the entire nation, introducing potential regional differences and non-random sampling, which can lead to biased statistics.

- Parameter: \$4.12
- Statistic: \$5.53

(c) The difference between the parameter and the statistic is likely due to variability in this case. While the probability of getting heads in a single coin flip is 0.5, over ten flips, natural variability can occur, and obtaining seven heads is not unexpected. Bias is less likely given the assumption of a fair coin, and the observed outcome falls within the expected range of variability.

- Parameter: $\frac{5}{10}$
- Statistic: $\frac{7}{10}$

Problem 2

In an opinion poll, researchers asked 200 random customers at Chicago-area McDonalds if they identified as republicans. Of those surveyed, 31% answered "yes."

(a) Identify at least two possible sources of bias in this study.

(b) Suppose the actual percentage of republicans in the Chicago-area was in fact 31%. Does this mean that the study was not biased? Briefly explain.

Answer

(a) Possible sources of bias in this study:

- **Sampling Bias:** If the researchers only surveyed customers at McDonald's in Chicago, the sample may not be representative of the entire population. For example what if McDonald's is specifically more popular for democrats?
- **Non-response Bias:** The statement does not specify how many of the people surveyed actually responded, therefore articulating inaccurate data. For example, what if the actual amount of republicans in this sample is %35, with %4 not responding?
- **Social Desirability:** McDonald is not a good choice of location to ask political questions like such, and may lead to people not giving honest opinions for various reasons. Especially those not identifying with neither democrats or republicans.

(b) Even if the actual percentage of Republicans in the Chicago-area is 31%, **it doesn't necessarily mean the study is unbiased**. The key issue lies in whether the sample is truly representative of the entire population. If the sample is biased, for example, if it disproportionately includes more politically active individuals or McDonald's customers, then the study may still be biased, even if the observed percentage matches the true population percentage. The accuracy of the study depends on the representation of the sample, not just the alignment of the observed percentage with the population percentage.

Problem 3

Why is the following a poor question to ask in order to learn how much people exercise? Suggest better wording.

How much do you exercise most weeks?

Answer

There are a couple of issues with the question:

1. Exercise is not defined.
2. "How much" should be represented by a unit.
3. "most weeks" is not fixed, a fixed window should be used to analyze data within. Furthermore, it is good to inquire only the average or median of that week for better analysis.

On daily average, how many minutes of physical activity have you had in the past 5 weeks?

Problem 4

Give an example of a statistic with high variability but low bias. Briefly explain your answer.

Answer

The heights of 10 year old, in a randomly selected school in Chicago.

- **High Variability:** The heights of 10 year old can vary significantly due to genetic factors, nutritional differences, and other individual characteristics. This natural variability can result in a wide range of height values within the population.
- **Low Bias:** A truly random and representative sample would result in a low-bias measurement of height, even with the inherent variability in heights among the 10 year old students.