STATISTICAL THEORY I (STA 4321): Homework 1

Evan Waldmann

August 30, 2018

Problem 1.1. For each of the following situations, identify the population of interest, the inferential objective, and how you might go about collecting a sample.

Solution.
a) The National Highway Safety Council wants to estimate the proportion of automobile tires with unsafe tread among all tires manufactured by a specific company during the current production year.

- Population: All tires manufactured by the specific company during the current production year

- Objective: To determine the proportion of the population with unsafe tread

- Sampling Method: Take an independent random sample of the population

b) A political scientist wants to determine whether a majority of adult residents of a state favor a unicameral legislature.

- Population: Adult residents of a state

- Objective: To determine if there is a majority of the population that favor a unicameral legislature

- Sampling Method: Take an independent random sample of the population

c) A medical scientist wants to estimate the average length of time until the recurrence of a certain disease.

- Population: People infected with the certain disease

- Objective: To estimate average length of time until recurrence of the certain disease in the population

- Sampling Method: Take an independent random sample of the population

d) An electrical engineer wants to determine whether the average length of life of transistors of a certain type is greater than 500 hours.

- Population: Transistors of a certain type

- Objective: To determine whether the average life span of the population is greater than 500 hours

- Sampling Method: Take an independent random sample of the population

e) A university researcher wants to estimate the proportion of U.S. citizens from "Generation X" who are interested in starting their own businesses.

- Population: Generation X

- Objective: To estimate the proportion of the population who are interested in starting their own businesses

- Sampling Method: Take an independent random sample of the population

f) For more than a century, normal body temperature for humans has been accepted to be 98.6 Fahrenheit. Is it really? Researchers want to estimate the average temperature of healthy adults in the United States.

- Population: Healthy adults in the United States

- Objective: To estimate the average temperature of the population (and I am assuming that the overall goal is see if that estimate differs from 98.6)

- Sampling Method: Take an independent random sample of the population

g) A city engineer wants to estimate the average weekly water consumption for single-family dwelling units in the city.

- Population: All single-family dwelling units in the city

- Objective: To estimate the average weekly water consumption of the population

- Sampling Method: Take an independent random sample of the population

Problem 1.10. It has been projected that the average and the standard deviation of the amount of time spent online using the Internet are, respectively, 14 and 17 hours per person per year (many do not use the internet at all!)

Solution.
a) 17-14 = -3
b) 1-.5-.34 = .16
c)No. The for the data to be normally distributed, the data has to follow the Empirical Rule, which requires that 68% of the data falls within the first standard deviation from the mean, 95% of the data falls withing two standard deviations from the mean, and 99.7% of the data falls within three standard deviations from the mean. If the data were normally distributed, we would expect to find about 16% of the data to be below -3, which is impossible because you cannot have a negative number of hours of time spent online per person per year.

Problem 1.11. Use (a), (b), and (c) to show that

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n}y_i\right)^2\right]$$

Solution.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$= \left(\frac{1}{n-1}\right) \sum_{i=1}^{n} \left(y_i^2 - 2y_i\bar{y} + \bar{y}^2\right)$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} y_i^2 - 2\bar{y} \sum_{i=1}^{n} y_i + n(\bar{y})^2 \right]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} y_i^2 - \frac{2}{n}\left(\sum_{i=1}^{n} y_i\right)^2 + \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 \right]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 \right]$$

Problem 1.22. Prove that the sum of the deviations of a set of measurements about their mean is equal to zero; that is,

$$\sum_{i=1}^{n} (y_i - \bar{y}) = 0$$

Proof.

$$\sum_{i=1}^{n} (y_i - \bar{y})$$

$$= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \bar{y}$$

$$= \sum_{i=1}^{n} y_i - n\bar{y}$$

$$= \sum_{i=1}^{n} y_i - n\frac{1}{n}\sum_{i=1}^{n} y_i$$

$$= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} y_i$$

$$= 0$$