

Sharon Jepkosgei

Homework 1

Due Date: 2/13/23

Question One: Why do we use statistics

We encounter large numbers in about every field and aspect of life. These large numbers are characterized by errors, deviation, and difference. Therefore, statistics are employed to understand, explore, summarize and draw conclusions about large numbers. Essentially, it enables effective decision-making and planning by various entities. It also helps us understand trends and patterns in huge quantities of data.

Question Two: Difference between a sample and a population

A population refers to all subjects (items or individuals) that a researcher wants to study. On the other hand, a sample is a part or subset of a population under review. It is used to make generalizations about a population. For example, all UTD students can be the population while some students (500) can be respondents in a study.

Question Three Why do we use samples instead of populations?

First, using samples is cheaper since it is too costly to collect data from the entire population and researchers often have very limited funds. Second, a sample saves time because collecting data from a population is time-consuming. Lastly, in many cases, it is impractical or not possible to access every individual or item in a population. For example, it can be impossible to collect data on all whales in an ocean.

Question Four: Defining the “Mystery of Statistics”

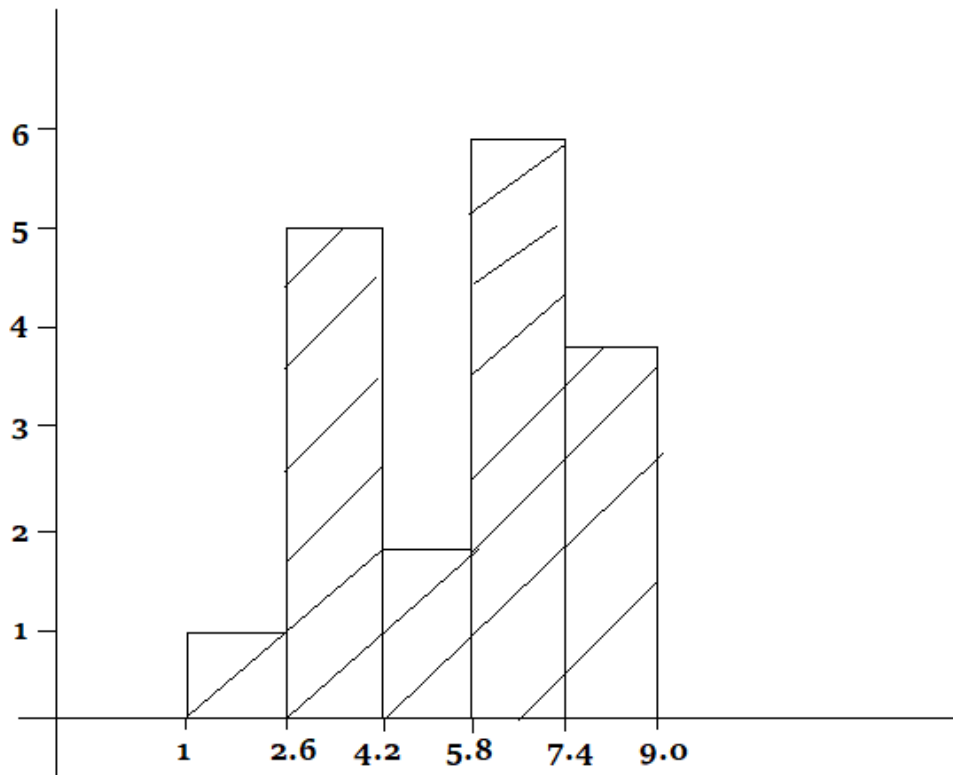
As stated above collecting data from an entire population is expensive hence; samples are used in research. The ‘mystery’ therefore is uncovering how resulting data from a few individuals or items of a population can be used to make a conclusion about an entire population.

Question Five

Dataset: 7 5 8 3 6 6 2 4 9 1 5 4 9 8 6 6 2 7

- i. Order data: 1 2 2 3 4 5 5 6 6 6 6 7 7 8 8 9 9
- ii. Range: 1-9
- iii. Width: $9 - 1 = 8$
- iv. Boxes: 5
- v. Size of boxes: $8/5 = 1.6$

a.

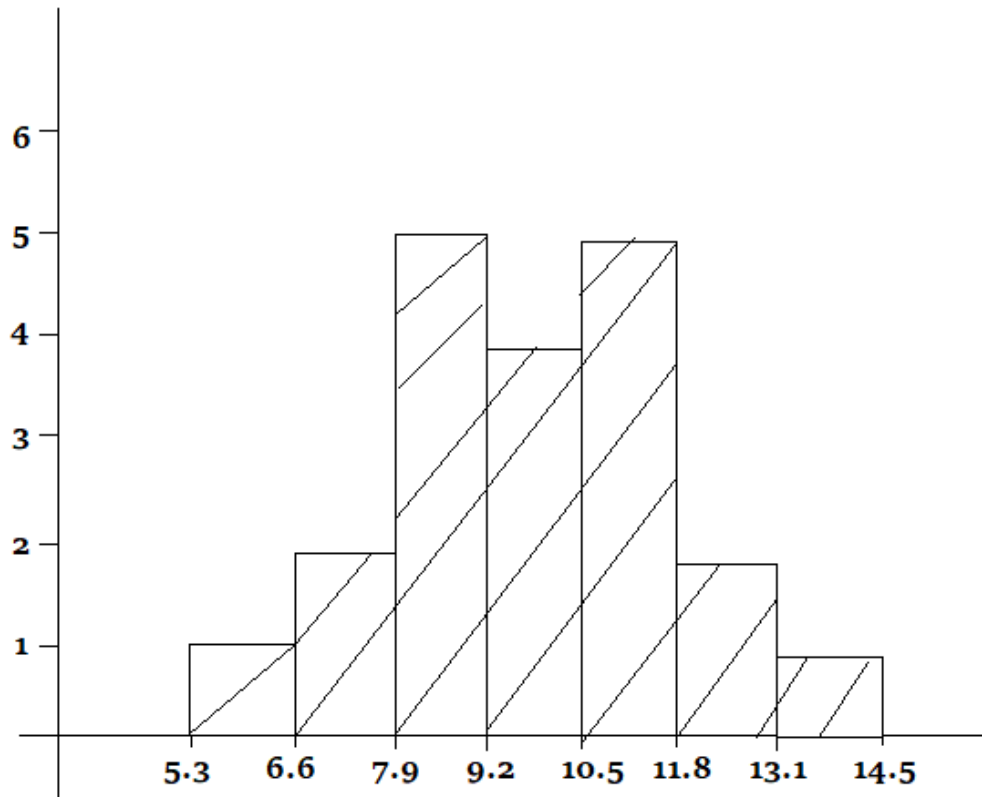


- b. The histogram indicates that there are no apparent outliers. However, it does not show a clear measure of central tendency. You cannot easily tell the mean or median using the chart. The histogram is bell-shaped indicating a normal distribution.

Question Six

Dataset: 10.3 11.1 9.6 9.0 14.5 13.0 6.7 11.0 8.4 10.3 8.0 11.8 8.7 10.6 9.5

- i. Order data: 5.3 6.7 7.3 8.0 8.0 8.4 8.7 9.0 9.5 9.6 10.3 10.3 10.6 11.0 11.1 11.2
11.8 12.5 13.0 14.5
- ii. Range: 5.3-14.5
- iii. Width: $14.5 - 5.3 = 9.2$
- iv. Boxes: 7
- v. Size of the boxes: $9.2/7 = 1.3$
 - a.



- b. The histogram is symmetrical. It shows that the dataset has no apparent outliers.

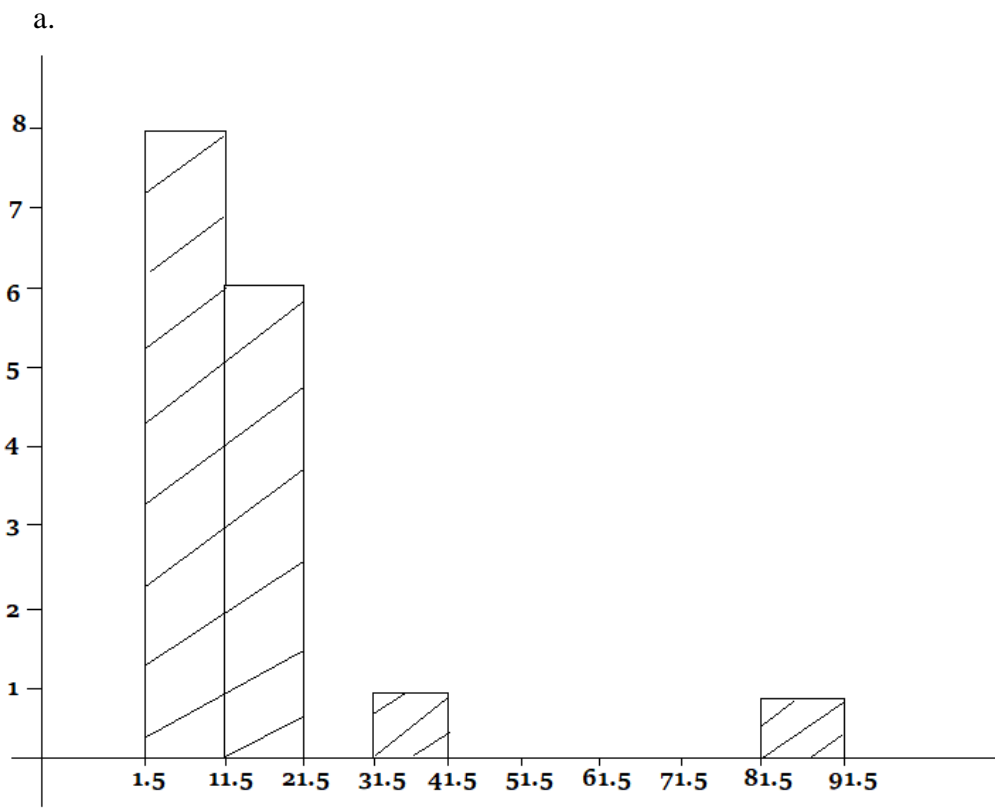
The measures of central tendency that can be obtained from the chart are mean and median which are approximately 9.85. The histogram is bell-shaped indicating a normal distribution.

Question Seven

Dataset: 3.2 9.9 17.7 1.5 11.5 4.6 11.8 32.5 8.9 11.7 14.0 7.1 13.8 91.6 9.0 14.0

- i. Order data: 1.5 3.2 4.6 7.1 8.9 9.0 9.9 11.5 11.7 11.8 13.8 14.0 14.0 17.7 32.5 91.6
- ii. Range: 1.5-91.6
- iii. Width: $91.6 - 1.5 = 90.1$
- iv. Boxes: 9

v. Size of boxes: $90.1/9 = 10$



b. The histogram shows that there are outliers in the data. It also indicates that the data values are skewed to the right. You cannot easily obtain a measure of central tendency by looking at the chart. The histogram has a right tail.