Practice quiz

Learning outcomes:

LO1: Define common principles of data reduction.

LO2: Calculate components of data reduction.

1. Any random sample selected from a population always produces the same results as a different random sample selected from the same population.

a. True

> Incorrect. Sample size and sample selection are important statistical decisions that set the foundation for further analysis. Each selected random sample differs from the next. If the sample is an inaccurate representation of the population, then all analyses after that point will be inaccurate, resulting in incorrect inferences being drawn.

b. False

> Correct, well done. Sample size and sample selection are important statistical decisions that set the foundation for further analysis. Each selected random sample differs from the next. If the sample is an inaccurate representation of the population, then all analyses after that point will be inaccurate, resulting in incorrect inferences being drawn.

---

2. For each of the following examples of quantitative variables, select the type of variable they represent. Select from the types "Measurable", "Categorical (ordinal)", or "Categorical (nominal)".

a. A customer survey where the possible responses on the survey are "Likely, neutral, and unlikely"

> Categorical (ordinal)

b. The average weight of products manufactured in a factory

> Measurable

c. An opinion poll asking voters which political party they are planning on voting for in an upcoming election

> Categorical (nominal)

d. An opinion poll asking participants what their religion is

> Categorical (nominal)

e. The ages of employees in a large corporation

> Measurable

f. A survey asking consumers how they would rate the service of a restaurant on a scale of 1 to 10

> Categorical (ordinal)

General feedback:

Measurable variables are those that are formulated using standardized methods of measurement, such as metric units and the imperial system. Examples include weight, time, and age. Categorical variables are those for which no generally accepted method of measurement exists. Ordinal categorical variables can be ranked (such as scale-ranked surveys), but nominal categorical variables cannot be ranked (such as religions and political parties).

---

3. Match each of the following descriptions with the data characteristic it most closely describes.

a. The middle value of an ordered dataset

> Median

b. The most frequently occurring value in a dataset

> Mode

c. The preferred measure of central tendency or simply the average of the data

> Mean

d. The largest value in a dataset minus the smallest

> Range

e. The average squared distance between each value and the mean of the dataset

> Variance

General feedback:

The measures of location include the preferred measure of central tendency (the mean), the middle value in an ordered set (the median), and the most frequently occurring value (the mode). The measures of dispersion are the range (the difference between the largest and smallest values) and the variance (describing the average squared distance between each value and the mean of the dataset).

---

4. For the dataset containing the values 1, 1, 3, 4, 5, 8, 9, 10, 11, 12, and 13, calculate the mean, median, and mode.

a. Mean = 77, median = 8, mode = 1

> Incorrect, please try again. Keep in mind that the mean should be divided by the total number of observations in the dataset.

b. Mean = 1, median = 8, mode = 7

Incorrect, please try again. Revisit the definitions of mean and mode and recall that the mean is the preferred measure of central tendency, while the mode is the most frequently occurring value.

c. Mean = 7, median = 1, mode = 8

Incorrect, please try again. Revisit the definitions of median and mode and recall that the median is the middle value of an organized dataset, while the mode is the most frequently occurring value.

d. Mean = 7, median = 8, mode = 1

Correct, well done. Dividing the sum of all the values, 77, with the number of values, 11, results in a mean of 7. The middle value, or median, of the set is 8, and the most frequently occurring value, the mode, is 1.

---

5. For the graph shown below, select whether it is positively skewed, negatively skewed, or symmetrical.
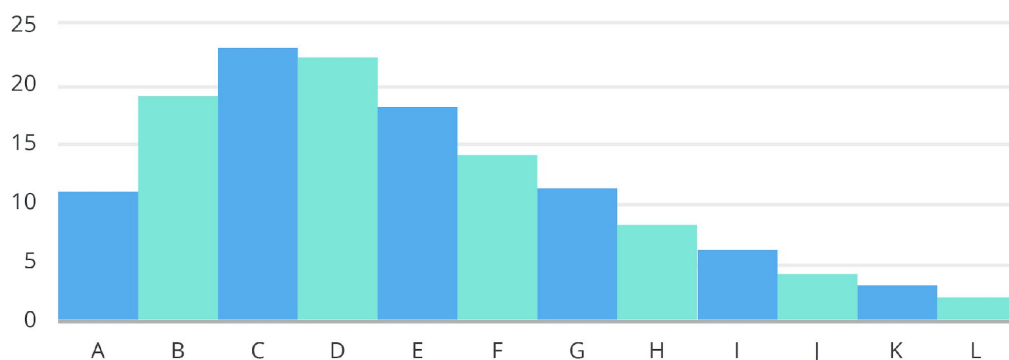
a. Positively skewed

Correct, well done. The high number of frequencies to the left of the distribution, and the tail pointing toward the right, indicate that this distribution is positively skewed.

b. Negatively skewed

Incorrect, please try again. For a negatively skewed distribution, a higher number of frequencies occur to the right of the distribution, with the tail pointing toward the left.

c. Symmetrical

Incorrect, please try again. A symmetric distribution has a higher number of frequencies in the middle of the distribution and the tails to the left and the right look similar.

6. From the given box plot, calculate the interquartile range, or IQR, of the dataset.
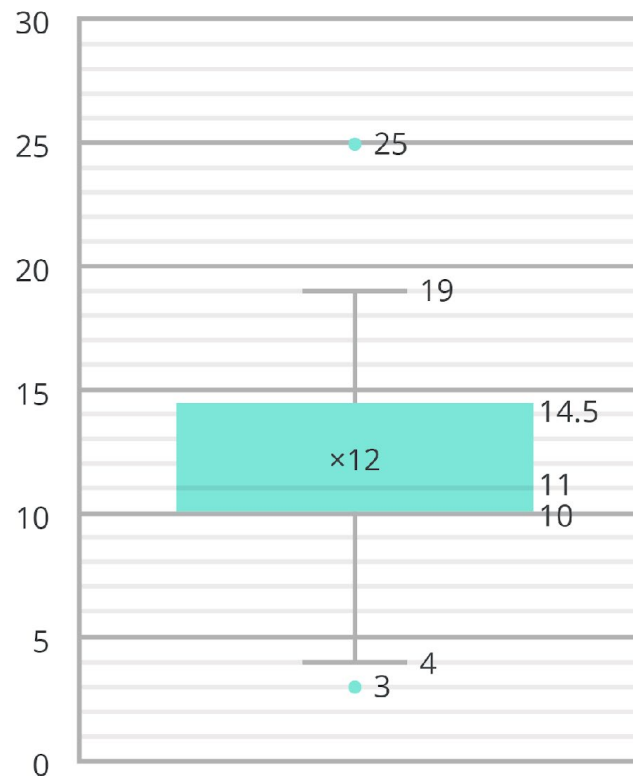
a. IQR = 3.5

  Incorrect, please try again. The lower-most value in the box is 10, and not the median of 11. The IQR is the difference between the upper quartile and the lower quartile.

b. IQR = 4.5

  Correct, well done. The IQR is the difference between the upper quartile of 14.5 and the lower quartile of 10.

c. IQR = 11.5

  Incorrect, please try again. The lower-most value in the box is 10, and not the bottom-most value of 3. The IQR is the difference between the upper quartile and the lower quartile.

d. IQR = 14.5

  Incorrect, please try again. This is the upper-most value of the box. The IQR is the difference between the upper quartile and the lower quartile.

e. IQR = 15

Incorrect, please try again. The upper-most value of the box is 14.5 and not the upper-most outlier of 25. The IQR is the difference between the upper quartile and the lower quartile.

f. IQR = 22

Incorrect, please try again. The range of this dataset is 22. The IQR is the difference between the upper quartile and the lower quartile.

---

7. A dataset contains the values 1, 3, 5, 8, and 10. These values represent the whole population. Calculate the standard deviation.

a. 10.64

Incorrect, please try again. The variance is 10.64, and you can use this value to calculate the standard deviation.

b. 6.31

Incorrect, please try again. Remember to subtract the mean from each value before it is squared.

c. 3.65

Incorrect, please try again. This standard deviation would be 3.65 if calculated for a sample dataset. If calculated for the entire population, do not subtract 1 from the total number of values.

d. 3.26

Correct, well done. Review the formula to make sure the correct method was used to get to this result.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{5} (x_i - \mu)^2}{N}} = \sqrt{\frac{19.36 + 5.76 + 0.16 + 6.76 + 21.16}{5}} = 3.26$$