

*What do I know already? (Do not look up the answers to these questions. The purpose is to assess your current level of knowledge on these topics.)*

A. Among the three measures of the center we studied (mean, median, and mode), which is generally preferred? Provide reasons why the other two measures in the list might be used instead of the preferred option at times. The mean is the preferred measure. The median should be used when extreme values might make the mean unrepresentative of what is typical in the data set. The mode is best for data that is qualitative.

B. What is the difference between a relative frequency table and relative frequency distribution? What is the difference between class limits and class boundaries? Why do we organize raw data into frequency distributions and then display the results in graphs like a histogram? Frequency tables are for qualitative data. Frequency distributions are for quantitative data. Class limits have gaps between the upper limit of a class and the lower limit of the class that follows it, the class boundaries do not have the gaps. We organize raw data and graph it because we want to understand how the data is distributed on the number line.

C. Describe what needs to be done (in the appropriate order) to evaluate the following expression  $\sum_{i=1}^n (x_i - \bar{x})^2$  for a given set of  $n$  measurements. step 1. subtract the sample mean from each value, step 2. square each of those results, step 3. sum the list of resulting values.

D. The median is sometimes referred to as a robust measure of the center. What is meant by this and explain how this quality of robustness is both a strength and a weakness. The median is more resistant to the effect of extreme values than the mean. That is useful when you want to describe the center of a data set that has extreme values on one end of the distribution; however, the same quality in general is a weakness because the median is insensitive to differences between data sets that just happen to have the same middle number.

**Learning Objectives:** (Click the learning objectives below for a short clip on the topic.)

Define Relative Frequency (1)

Calculate the Relative Frequencies for a Set of Classes (3)

Form Class Boundaries from a Given Set of Classes (3)

Know the left-endpoint-convention (1)

Calculate an Appropriate Class Width Given a Set of Raw Data (3)

Form Classes from a Set of Raw Data (3)

Know the best practices for drawing a histogram (1)

Use Summation Notation to simplify an Expression (3)

List Four Important Properties of the Arithmetic Mean (1)

Distinguish between Situations where the Mean, Median, or Mode is Most Appropriate (4)

**Exercises:**

- Some barbers, hairdressers, and cosmetologists earn very large annual salaries due to celebrity clients, product lines, and/or very successful salons. However, workers in the industry do not typically earn such large salaries. If you worked for the Bureau of Labor Statistics, what measure of the center would you recommend to describe the typical earning for barbers, hairdressers, and cosmetologists?

A. Mean

B. Median

C. Mode

D. Range (this is not a measure of the center)

E. Mean Absolute Deviation (this is not a measure of the center)

- Use the provided values  $\{-4, 2, 9, 1, -3, 5, 4\}$  to find the following sum:  $\sum_{i=1}^7 (x_i - 1)^2$

131

- The relative frequency for a class is defined as the class frequency divided by the sum of the frequencies.
- Which of the following are properties of the arithmetic mean? (select all that apply)
  - Every data value is included in the calculation of the arithmetic mean.
  - The sum of the deviations from the mean is always zero. In other words,
 
$$\sum_{i=1}^n (x_i - \bar{x}) = 0.$$
  - The arithmetic mean is robust, which means that it is not heavily influenced by extreme values.
  - The mean is the "center of mass" or centroid of the data set.

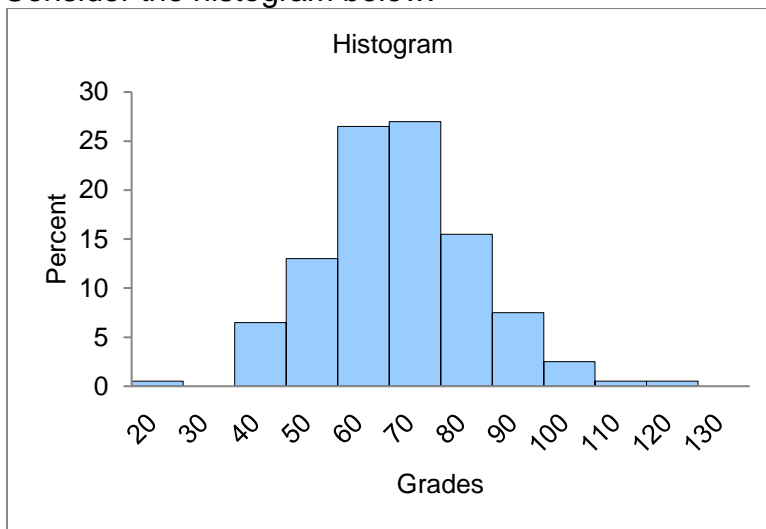
5. Use the provided values {3, 2, -5, 1, 0, 3} to find the following sum:  $\sum_{i=1}^6 x_i^2 = 48$
6. True or False: In a histogram, the proportion of the total area of a histogram for a given rectangle should be equal to the relative frequency for the class the given rectangle represents. In other words, if a rectangle in a histogram contains 2% of the total area of the histogram, that rectangle should represent a class in the frequency table that has a relative frequency equal to 0.02.
7. The following frequency distribution is for grades in a course that has a maximum of 130 available points. Find the relative frequency for the class that includes grades between 40 and 50 points.  $13/200 = 0.065$

## Frequency Distribution - Grades

| <b>Grades</b> |   |              |                  |
|---------------|---|--------------|------------------|
| <i>lower</i>  |   | <i>upper</i> | <i>frequency</i> |
| 20            | < | 30           | 1                |
| 30            | < | 40           | 0                |
| 40            | < | 50           | 13               |
| 50            | < | 60           | 26               |
| 60            | < | 70           | 53               |
| 70            | < | 80           | 54               |
| 80            | < | 90           | 31               |
| 90            | < | 100          | 15               |
| 100           | < | 110          | 5                |
| 110           | < | 120          | 1                |
| 120           | < | 130          | 1                |
|               |   |              | 200              |

8. I am interested in the average fine issued for speeding tickets in Miami. I reviewed a sample of the tickets issued. The average I calculated from the sample of data is an example of a statistic or a parameter? Why? The average describes an attribute of a sample of data
9. The heights of the athletes playing on the men's basketball team are recorded each year. This set of measurements is an example of: (select all that apply)
- A. continuous data
  - B. discrete data
  - C. qualitative data
  - D. ordinal level data
  - E. quantitative data

10. Consider the histogram below:



What should the relative frequency be (approximately) for the rectangle covering the grade interval from 80 to 90?

- A. 0.25  
 B. 0.85  
 C. 0.90  
 D. 0.80  
 E. 0.15
11. A women's volleyball coach would like to report the typical height of athletes that compete in her sport. Which measure of the center would be the best choice in this situation?  
 A. Median  
 B. Mean (Since human height has physical limitations, it is unlikely to have extreme values that will unduly influence the mean.)  
 C. Mode  
 D. None of the above
12. Use the provided values {3, 2, -5, 1, 0, 3} to evaluate the following expression:  

$$\left( \sum_{i=1}^6 x_i \right)^2 = ?$$
 16
13. The largest value in a set of data is 80. The smallest value in the set is 19. If we wish to create a frequency distribution for the data that has 5 classes, what class width should we use among the choices below? Assume that your first class interval will start with the number 16 and that you will use class boundaries to define the classes.  
 A. 12      B. 12.2      C. 13      D. 61      E. 10

14. In the following frequency distribution in which class interval would a grade of 30 belong? The left-end-point convention says that the lower boundary values are included in the intervals, but the upper boundary values are not in the interval. [30, 40)

## Frequency Distribution - Grades

| <i>Grades</i> |   |              |                  |
|---------------|---|--------------|------------------|
| <i>lower</i>  |   | <i>upper</i> | <i>frequency</i> |
| 20            | - | 30           | 1                |
| 30            | - | 40           | 0                |
| 40            | - | 50           | 13               |
| 50            | - | 60           | 26               |
| 60            | - | 70           | 53               |
| 70            | - | 80           | 54               |
| 80            | - | 90           | 31               |
| 90            | - | 100          | 15               |
| 100           | - | 110          | 5                |
| 110           | - | 120          | 1                |
| 120           | - | 130          | 1                |
|               |   |              | 200              |

- A. 20 - 30  
B. 30 - 40  
C. 40 - 50  
D. 50 - 60  
E. None of these
15. True or false: When creating a histogram with a very large amount of data, it is possible to use more classes than when creating a histogram for a relatively small amount of data. Also, a histogram created from a large data set that has been organized into many classes will tend to appear like an almost smooth curve.
16. The NHTSA (National Highway Traffic Safety Administration) conducted a study of accidents involving motorcycles. The study included observations of helmet color. If the researchers want to describe the typical color of helmet involved in a motorcycle crash, which measure of the center is most appropriate?  
A. Mean  
B. Median  
C. Mode  
D. Range (this is not a measure of the center)  
E. Mean Absolute Deviation (this is not a measure of the center)

17. Organize the following data set into a frequency distribution with 5 classes, a starting lower class limit (or boundary) of 5, and a uniform class width.

|    |    |    |    |    |
|----|----|----|----|----|
| 7  | 8  | 9  | 9  | 10 |
| 11 | 11 | 12 | 15 | 19 |
| 20 | 21 | 21 | 22 | 23 |
| 24 | 27 | 29 | 30 | 31 |
| 33 | 34 | 34 | 34 | 37 |

| Classes | Frequency |
|---------|-----------|
| 5 - 11  | 7         |
| 12 - 18 | 2         |
| 19 - 25 | 7         |
| 26 - 32 | 4         |
| 33 - 39 | 5         |

or (using boundaries)

| Classes | Frequency |
|---------|-----------|
| 5 - 12  | 7         |
| 12 - 19 | 2         |
| 19 - 26 | 7         |
| 26 - 33 | 4         |
| 33 - 40 | 5         |

18. Convert the given set of class limits below into a suitable set of class boundaries so that the boundaries can be used to draw a histogram for the data. Note, you do not need to create the histogram.

| Heights (in) | Frequency | Boundaries              |
|--------------|-----------|-------------------------|
| 60 - 62      | 1         | 59.5 - 62.5 or [60, 63) |
| 63 - 65      | 2         | 62.5 - 65.5 or [63, 66) |
| 66 - 68      | 22        | 65.5 - 68.5 or [66, 69) |
| 69 - 71      | 25        | 68.5 - 71.5 or [69, 72) |
| 72 - 74      | 7         | 71.5 - 74.5 or [72, 75) |
| 75 - 77      | 3         | 74.5 - 77.5 or [75, 78) |

19. The number of friends each Facebook user has is an example of: (select all that apply)
- A. continuous data
  - B. discrete data
  - C. qualitative data
  - D. nominal level data
  - E. quantitative data