Chapter 3 Describing, Exploring, and Comparing Data
Measures of Center (3.1)

Objectives

In this lesson, we will learn how to

- describe the meaning of a measure of center;
- compute and interpret the mean, median, mode, and midrange;
- decide when a measure of center is resistant;
- apply round-off rules for measures of center;
- compute an approximate mean from a frequency distribution;
- compute a weighted mean, such as a grade-point average.

1. Measures of Center

A measure of center is a value at the center or middle of a data set.

Common measures of center include

mean, median, mode, midrange.

Big Idea

A measure of center gives us a single value that represents a typical or central value in a data set. However, different measures of center can give different answers.

2. Mean

The mean, or arithmetic mean, is found by adding all data values and dividing by the number of data values.

Mean

For a sample,

$$\bar{x} = \frac{\sum x}{n}.$$

For a population,

$$\mu = \frac{\sum x}{N}.$$

Notation

- $\sum$ means "the sum of."
- x represents an individual data value.
- n is the number of values in a sample.
- N is the number of values in a population.
- $\bar{x}$, pronounced "x-bar," is the sample mean.
- μ, pronounced "mu," is the population mean.

Important Properties of the Mean

- The mean uses every data value.
- Sample means from the same population tend to vary less than other measures of center.
- The mean can be greatly affected by an extreme value, or outlier.

A statistic is resistant if extreme values do not cause it to change very much. The mean is not resistant.

Example 1: Mean: Find the mean of the first eleven wait times, in minutes, for Space Mountain at 10 AM:

$$50, 25, 75, 35, 50, 25, 30, 50, 45, 25, 20.$$

Solution

Add all values and divide by the number of values.

$$\bar{x} = \frac{50 + 25 + 75 + 35 + 50 + 25 + 30 + 50 + 45 + 25 + 20}{11}$$

$$= \frac{430}{11} = 39.0909\ldots$$

Rounded to one more decimal place than the original data,

$$\bar{x} = 39.1.$$

The mean wait time is 39.1 minutes.

3. Median

The median is the middle value when the original data values are arranged in order.

Finding the Median

First sort the data values.

- If the number of values is odd, the median is the exact middle value.
- If the number of values is even, the median is the mean of the two middle values.

Important Properties of the Median

- The median is resistant because it is not affected much by a few extreme values.
- The median does not directly use every data value.

Example 2: Median with an Odd Number of Values: Find the median of the first eleven wait times, in minutes, for Space Mountain at 10 AM:

$$50, 25, 75, 35, 50, 25, 30, 50, 45, 25, 20.$$

Solution

First sort the values:

$$20, 25, 25, 25, 30, 35, 45, 50, 50, 50, 75.$$

There are 11 values, so the median is the 6th value:

$$\text{median} = 35.0.$$

The median wait time is 35.0 minutes.

Example 3: Median with an Even Number of Values: Find the median after including the twelfth wait time:

$$50, 25, 75, 35, 50, 25, 30, 50, 45, 25, 20, 50.$$

Solution

First sort the values:

$$20, 25, 25, 25, 30, 35, 45, 50, 50, 50, 50, 75.$$

There are 12 values, so use the mean of the two middle values, 35 and 45:

$$\text{median} = \frac{35 + 45}{2} = 40.0.$$

The median wait time is 40.0 minutes.

4. Mode

The mode is the value or values that occur with the greatest frequency.

Finding the Mode

- A data set can have one mode.
- A data set can have more than one mode.
- A data set can have no mode.
- The mode can be used with qualitative data.

When two values occur with the same greatest frequency, the data set is called bimodal. When more than two values occur with the same greatest frequency, the data set is called multimodal.

Example 4: Mode: Find the mode of the first eleven wait times, in minutes, for Tower of Terror at 10 AM:

$$35, 35, 20, 50, 95, 75, 45, 50, 30, 35, 30.$$

Solution

Sort the list:

$$20, 30, 30, 35, 35, 35, 45, 50, 50, 75, 95.$$

The value 35 occurs three times, more than any other value.

$$\text{mode} = 35.$$

The mode is 35 minutes.

Other Mode Examples:

- The wait times $30, 30, 50, 50, 75$ have two modes: 30 and 50.

- The wait times $20, 30, 35, 50, 75$ have no mode because no value is repeated.

5. Midrange

The midrange is the value halfway between the maximum and minimum values.

Midrange

$$\text{midrange} = \frac{\text{maximum value} + \text{minimum value}}{2}.$$

The midrange is very easy to compute, but it uses only the maximum and minimum values, so it is not resistant.

Example 5: Midrange: Find the midrange of the first eleven wait times, in minutes, for Space Mountain at 10 AM:

$$50, 25, 75, 35, 50, 25, 30, 50, 45, 25, 20.$$

Solution

The minimum value is 20 and the maximum value is 75.

$$\text{midrange} = \frac{75 + 20}{2} = \frac{95}{2} = 47.5.$$

The midrange is 47.5 minutes.

6. Round-Off Rules

Round-Off Rules for Measures of Center

- For the mean, median, and midrange, round to one more decimal place than the original data.
- For the mode, leave the value as is, because the mode is one of the original data values.

7. Critical Thinking with Measures of Center

We can calculate measures of center from many sets of numbers, but we should ask whether the calculation is meaningful.

Be Careful

Mean and median may not be meaningful when the numbers are really labels, ranks, jersey numbers, or values from a biased list.

Example 6: When Measures of Center Are Not Meaningful: For each situation, decide why the mean or median may not be meaningful.

(a) Zip codes of several famous locations.

(b) Ranks of selected medical schools.

(c) Jersey numbers of football players.

(d) Top 5 CEO compensation values.

(e) The 50 state mean ages, one from each state.

Solution

(a) Zip codes are labels, not measurements.

(b) Ranks show order, but they do not measure an actual amount.

(c) Jersey numbers are labels, not numerical measurements.

(d) A top 5 list is not representative of the larger population.

(e) State populations differ, so the state means should be weighted by population size.

8. Mean from a Frequency Distribution

When data are summarized in a frequency distribution, we can approximate the mean by using class midpoints.

Mean from a Frequency Distribution

$$\bar{x} = \frac{\sum(f \cdot x)}{\sum f},$$

where f is the frequency and x is the class midpoint.

Example 7: Mean from a Frequency Distribution: Use the frequency distribution below to approximate the mean.

Time (seconds)	f	x	$f \cdot x$
$0 - 14$	6	7.0	42.0
$15 - 29$	18	22.0	396.0
$30 - 44$	14	37.0	518.0
$45 - 59$	5	52.0	260.0
$60 - 74$	5	67.0	335.0
$75 - 89$	1	82.0	82.0
$90 - 104$	1	97.0	97.0
Totals	50		1730.0

Solution

Use

$$\bar{x} = \frac{\sum(f \cdot x)}{\sum f}.$$

Substitute the totals:

$$\bar{x} = \frac{1730.0}{50} = 34.6.$$

The approximate mean is 34.6 seconds.

9. Weighted Mean

A weighted mean is used when different data values have different weights.

Weighted Mean

$$\bar{x} = \frac{\sum(w \cdot x)}{\sum w},$$

where w represents the weight assigned to each data value.

Example 8: Computing Grade-Point Average: A student takes five courses with grades and credits as follows:

Grade	Quality Points x	Credits w
A	4	3
A	4	4
B	3	3
C	2	3
F	0	1

Compute the student's GPA.

Solution

Use the credits as weights.

$$\bar{x} = \frac{\sum(w \cdot x)}{\sum w}$$

$$= \frac{3(4) + 4(4) + 3(3) + 3(2) + 1(0)}{3 + 4 + 3 + 3 + 1}$$

$$= \frac{12 + 16 + 9 + 6 + 0}{14} = \frac{43}{14} = 3.0714\ldots$$

The GPA is approximately

$$3.07.$$

Summary

Key Ideas

- The mean uses every data value, but it is not resistant.
- The median is the middle value and is resistant.
- The mode is the most frequent value and can be used with qualitative data.
- The midrange is easy to compute, but it is not resistant.
- Always think about whether a measure of center is meaningful for the data.
- For grouped data, the mean from a frequency distribution is an approximation.
- A weighted mean is used when some values count more than others.