

Sunday, October 14, 2018
5:00 PM

Key Terms:

The **median** gives the midpoint of a set of data – it separates the upper half of the data from the lower half.

- To calculate the median, order the data from **smallest to largest** and count up $(n+1)/2$ places in the ordered list.

The **mean** is the **arithmetic average** or **balance point** of a set of data.

- To calculate the mean, sum the data and divide by the number of data: $\bar{x} = \frac{\sum x}{n}$

The **mode** is the **data value that occurs most frequently**.

A **resistant measure** of some aspect of a distribution (such as its center) is

relatively unaffected by a small subset of extreme data values.

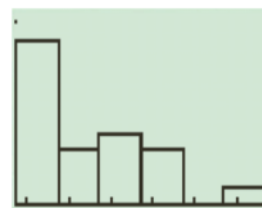
Classwork Exercises

1. Here are the starting salaries, in thousands of dollars, offered to the 20 students who earned degrees in computer science in 2011 at a university.

63 56 66 77 50 53 78 55 90 65
64 69 59 76 48 54 49 68 51 50

- a.** Make a graph **of your choice** to describe the distribution. Be sure to include a title and labels!

Either a histogram or a stemplot would be a good choice of graphic display.



Starting salaries offered to 20 Students with Computer Science degrees from 2011

4	89
5	00134569
6	345689
7	678
8	
9	0

Key: 4 | 8 = \$ 48,000 starting salary

- b. Write a brief description of its important features.

The distribution is skewed to the right. There is a gap in the 80 thousands, which makes the \$90,000 salary appear to be an outlier.

- c. Find the median salary. $\frac{n+1}{2} = \frac{21}{2} = 10.5$ th entry
 $(59 + 63)/2 = 61$; median starting salary is \$61,000.

- d. Find the mean salary. $\bar{x} = \frac{\sum x_i}{n}$ ← 2nd STAT SUM LI
 $(1241/20) = 62.05$; mean starting salary is \$62,050.

- e. Find the mode of the salaries. 50 thousand

- f. Is the mean about the same as the median or not?

The mean starting salary is higher than the median.

- g. What feature of the distribution explains the difference between the mean and the median?

The feature of the distribution that explains the difference between the mean and the median is largely due to the \$90,000 outlier but also due to the shape of the data, which is skewed to the right.

- h.** Is the mode a good measure of the center for these data? Why or why not?

The mode does not do a good job in measuring the center or location of these data. Normally, the mode is used for categorical data where we wish to know which is the most common category.

- 2.** Each month, the Commerce Department reports the “average” price of new single-family homes. For August 2012, the two “averages” reported were \$256,900 and \$295,300. The distribution is skewed to the right. Which of these numbers was the mean price and which was the median price? Explain your answer.

The median was \$256,900 and the mean is \$295,300. The mean is inflated because of a few extremely expensive houses, houses with prices in the millions.

- 3.** In 1961 New York Yankee outfielder Roger Maris held the major league record for home runs in a single season, with 61 home runs. That record held for 37 years. Here are Maris’s home run totals for his 10 years in the American League.

13, 23, 26, 16, 33, 61, 28, 39, 14, 8

- a.** Find the mean number of home runs that Maris hit in a year, both with and without his record 61.

Mean for all 10 years: $\bar{x} = \frac{\sum x_i}{n}$
 $(13 + 23 + 26 + 16 + 33 + 61 + 28 + 39 + 14 + 8)/10 = 261/10$
 $= 26.1.$

Mean (excluding 61):
 $(13 + 23 + 26 + 16 + 33 + 28 + 39 + 14 + 8)/9 = 200/9$
 $\approx 22.22.$

- b. How does removing the record number of home runs affect his mean number of runs?

Omitting his record year lowers the mean by 3.88.

- c. Find the median number of home runs that Maris hit in a year, both with and without his record 61.

All 10 years: ordered data

8 13 14 16 23 | 26 28 33 39 61

median = $(23 + 26)/2 = 24.5$.

$$\frac{n+1}{2} = \frac{10+1}{2} = \frac{11}{2} = 5.5^{\text{th}} \text{ value}$$

Excluding 61: ordered data

8 13 14 16 23 26 28 33 39

$$\frac{n+1}{2} = \frac{9+1}{2} = \frac{10}{2} = 5^{\text{th}}$$

median = 23.

- d. How does removing the record number of home runs affect his median number of runs?

Omitting his record year lowers the median by 1.5. Hence, the median is less affected by the record number of home runs than the mean.

- e. If you had to choose between the mean and median to describe Maris's home run hitting pattern, which would you use?

The mean overstates Maris's usual performance because of the influence of the outlier. But the median doesn't point to the great achievement of his career. Perhaps we should say "Maris hit 61 home runs in 1961, and averaged about 22 home runs a year in his other 9 years in the American League."