

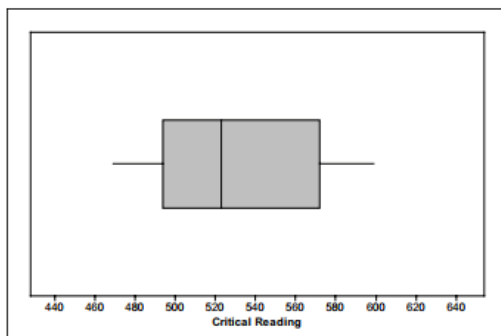
Homework Answers:

1. a. 469, 494, 523, 572, 599

b. California's average SAT Critical Reading score does not fall in the top half of the states' average Critical Reading scores because 499 is below the median score of 523. It does fall above the bottom quarter because 499 is greater than the first quartile, which is 494.

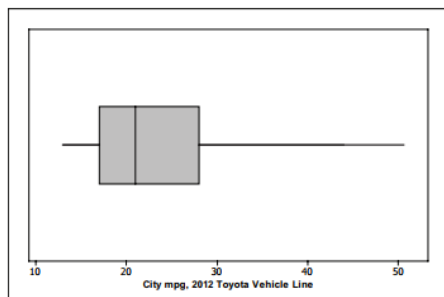
c. About 25% of the states; 12 states. That is because Wyoming's score is the third quartile.

d. In the boxplot below, the third quarter of the data appears to be more spread out than the other quarters of the data.



3. a. 13, 17, 21, 28, 51

b.



c. There appear to be three potential outliers: 43, 44, 51.

d. Calculations for fences:

$$\text{IQR} = 28 - 17 = 11; \text{step} = (1.5)(11) = 16.5$$

Lower inner fence = $17 - 16.5 = 0.5$; no data values lie below this fence.

$$\text{Upper inner fence} = 28 + 16.5 = 44.5$$


$$\text{Upper outer fence} = 28 + 2(16.5) = 61$$

Only one data value, 51, is an outlier and it is a mild outlier because it falls between the two upper fences (see below).



Measures of Center

Video #4

 <http://www.learner.org/courses/againstallodds/unitpages/unit04.html>

Fill in the Video Guide

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.

1. What *variable* is examined in comparing men and women workers at the beginning of the video?

1. The variable is the weekly wages for Americans, separated by gender.

2. Would you describe the shape of the distribution of men's weekly wages as symmetric, skewed to the left or skewed to the right?

2. The men's distribution is skewed to the right.

3. What is the most important difference between the distributions of weekly wages for men and for women?

3. The medians of the two distributions differ. Median measures the 50-50 point. The median for men's wages was larger than the median for women's wages.

4. Would a few very large incomes pull the mean of a group of incomes up, down, or leave the mean unaffected?

4. A few very large incomes inflate the mean of a group of incomes. Hence, these very large incomes would pull the mean up.

5. Would a few very large incomes pull the median of a group of incomes up, down, or leave the median unaffected?

5. A few very large incomes have no effect on the median.

Summarizing Symmetric Distributions -- The Mean

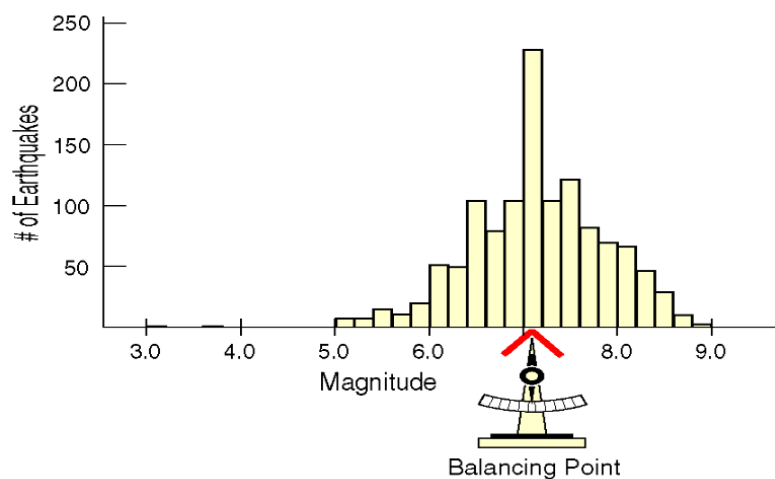
- When we have symmetric data, there is an alternative other than the median.
- If we want to *calculate* a number, we can *average* the data.
- We use the Greek letter sigma to mean “sum” and write:

$$\bar{y} = \frac{\text{Total}}{n} = \frac{\sum y}{n}$$

The formula says that to find the mean, we add up all the values of the variable and divide by the number of data values, n .

Summarizing Symmetric Distributions -- The Mean (cont.)

- The **mean** feels like the center because it is the point where the histogram balances:



Mean or Median?

- Because the median considers only the order of values, it is **resistant** to values that are extraordinarily large or small; it simply notes that they are one of the “big ones” or “small ones” and ignores their distance from center.
- To choose between the mean and median, start by looking at the data. If the histogram is symmetric and there are no outliers, use the mean.
- However, if the histogram is skewed or with outliers, you are better off with the median.

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

Stat - Edit - 1 Put data
into L₁

Stat - SortA(L₁)

a. Make a graph to describe the distribution and write a brief description of its important features.

Stat - Calc - 1: 1Var Stats

b. Find the median salary. Med = 61
Median Salary = \$61,000

c. Find the mean salary. $\bar{x} = 62.05$
Mean salary = \$62,050

d. Find the mode of the salaries.

Mode = 50 = \$50,000

e. Is the mean about the same as the median or not? What feature of the distribution explains the difference between the mean and the median? Is the mode a good measure of the center for these data?

Not equal b/c the outlier (\$90,000) is pulling the mean higher.
Median is a better measure of center here b/c of the outlier

4		89
5		00134569
6		345689
7		678
8		
9		0

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

$$\text{mean}(\bar{x}) = 26.1 \quad \text{mean w/out 61} (\bar{x}) = 22.22$$

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

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

Removing outlier make mean go down.

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

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

$$\text{med} = 24.5 \quad \text{Median w/out 61} = 23 \quad \text{not much change}$$

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

Median better because more "typical" and not distorted by outlier.

Homework: Review Exercises #1-3