Comparing Mean and Median

Learning Objective: Relate measures of center and spread to the shape of the distribution, and choose the appropriate measures in different contexts.

Comparing the Mean and the Median

As we have seen, mean and the median, two of the common measures of center, each describe the center of a distribution of values in a different way. The mean describes the center as an average value, in which the *actual values* of the data points play an important role. The median, on the other hand, locates the middle value as the center, and the *order* of the data is the key to finding it.

To get a deeper understanding of the differences between these two measures of center, consider the following example.

Here are two datasets:

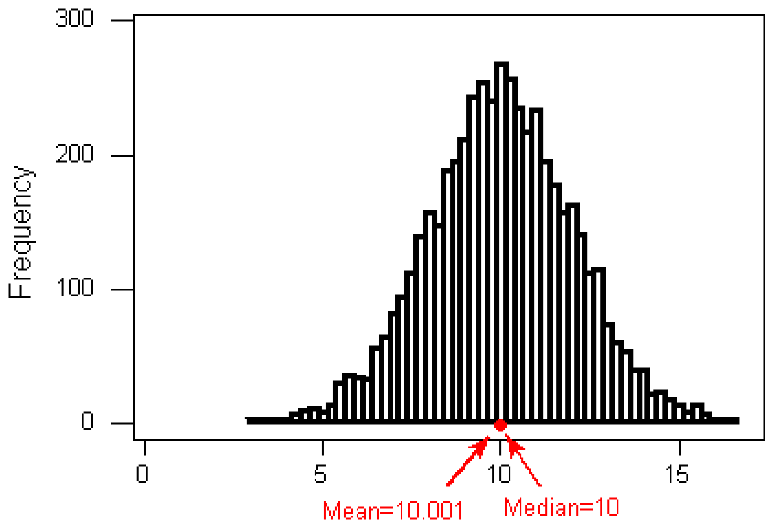
- Data set A → 64 65 66 68 70 71 **73**
- Data set B → 64 65 66 68 70 71**730**

For dataset A, the mean is 68.1, and the median is 68. Looking at dataset B, notice that all of the observations except the last one are close together. The observation 730 is very large, and is certainly an outlier. In this case, the median is still 68, but the mean will be influenced by the high outlier, and shifted up to 162. The message that we should take from this example is:

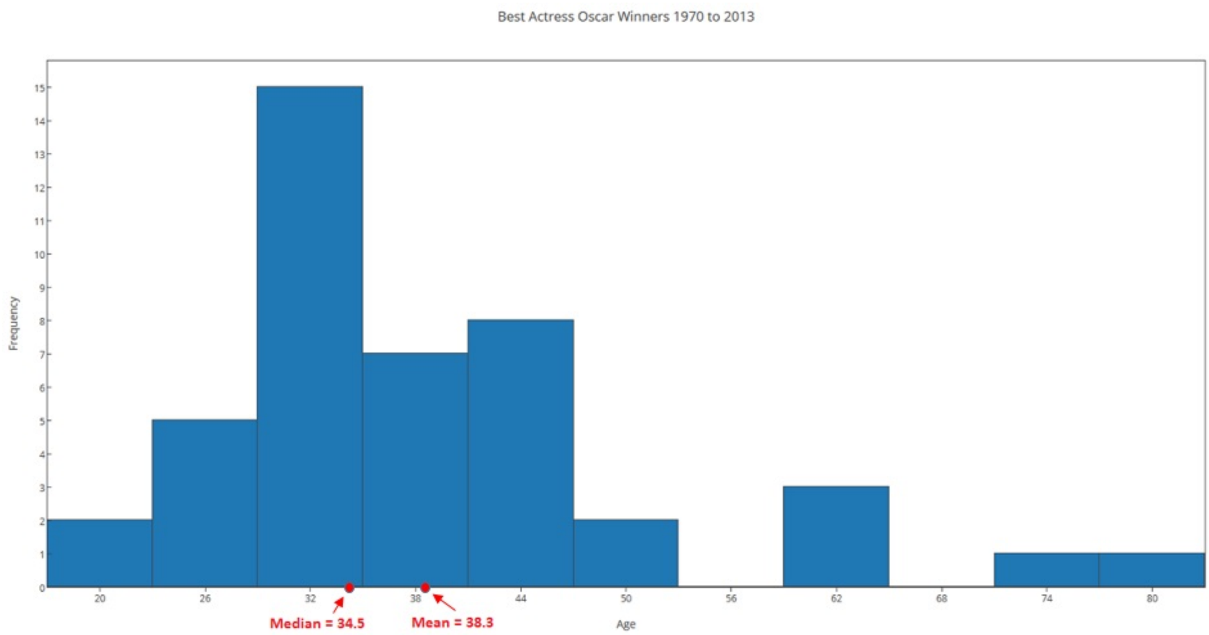
The mean is very sensitive to outliers (because it factors in their magnitude), while the median is resistant to outliers.

Therefore:

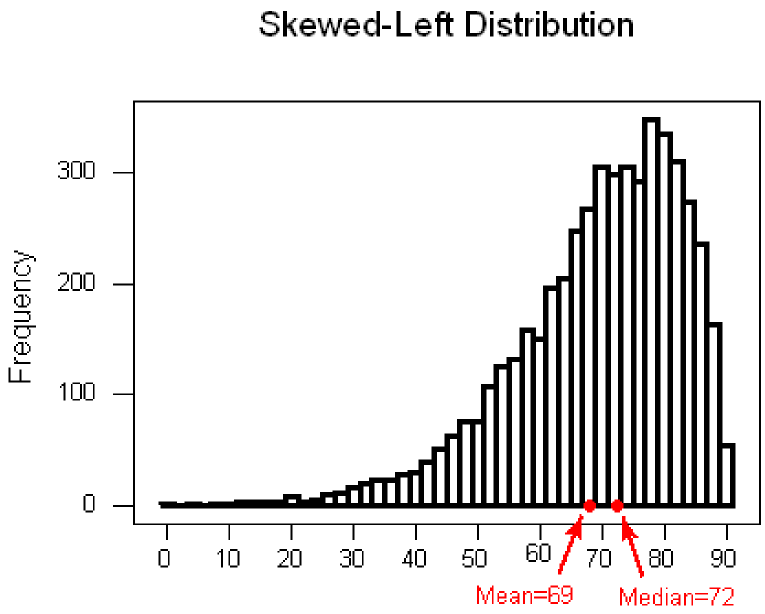
- For symmetric distributions with no outliers: $\overline{x}$ is approximately equal to M.



- For skewed right distributions and/or datasets with high outliers: $\overline{x} > M$



- For skewed left distributions and/or datasets with low outliers: $\overline{x} < M$

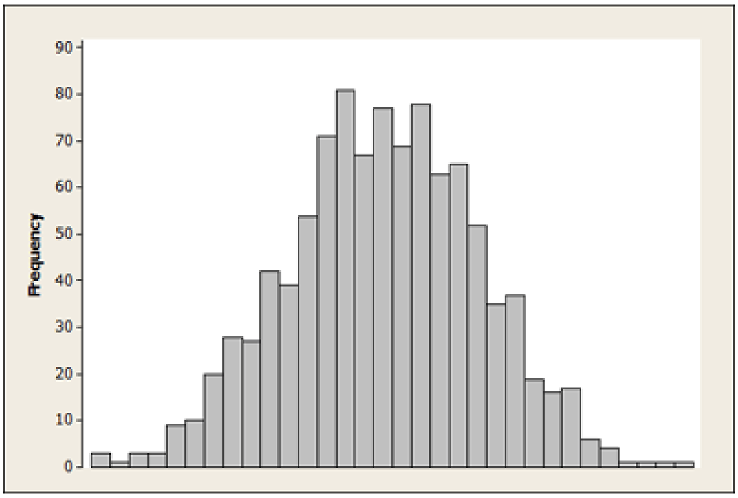


We will therefore use $\overline{x}$ as a measure of center for symmetric distributions with no outliers. Otherwise, the median will be a more appropriate measure of the center of our data.

Did I Get This

1/1 point (graded)

In the histogram below, what will be the relationship between the mean and median of the collected data?



- The mean will be bigger than the median.
- The mean will be smaller than the median.

- [x] The mean and median will be about the same. ✔

Answer
Correct: Since the distribution is symmetric, the mean and median would be about the same.

Submit

Did I Get This

1/1 point (graded)
The Current Population Survey conducted by the Census Bureau records the incomes of a large sample of U.S. households each month. What will be the relationship between the mean and median of the collected data?

- [x] The mean will be bigger than the median. ✔
- [ ] The mean will be smaller than the median.
- [ ] The mean and the median will be about the same.

Answer
Correct: The distribution of incomes is skewed right, so the mean will be bigger than the median.

Submit

Did I Get This

1/1 point (graded)
The SAT Math scores of 1,000 future engineers and physicists are recorded. What will be the relationship between the mean and median of the collected data?

- [ ] The mean will be bigger than the median.
- [x] The mean will be smaller than the median. ✔
- [ ] The mean and median will be about the same.

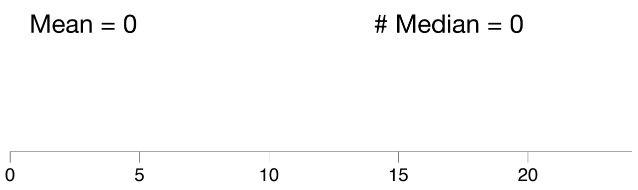
Answer
Correct:
Since the SAT Math scores for these students will be mostly high scores, the distribution will be skewed to the left. Thus, the few low scores (outliers) will make the mean smaller than the median.

Submit

Scenario: Mean and Median

A description of a distribution almost always includes a measure of its center or average. The two common measures of center are the **mean** and the **median**.

In this exercise, we will use an interactive simulation to explore the relationship between the mean and the median. Click below the line to add an observation. The red arrow marks the median. The green arrow marks the mean. When the median and mean are equal, a single yellow arrow is shown. You can use the mouse to drag a point along the line. Note that you can "stack" more than one point at the same location. Click on the trash to remove all points.



Clear

Learn By Doing (1/1 point)

Remove all points by clicking on the trash. Place three observations on the line by clicking below it, two close together near the center of the line, and one somewhat to the right of those two. Pull the single right-most observation out to the right. (Place the cursor on the point, hold down the mouse button, and drag the point.) How does the mean behave? How does the median behave? Make sure you understand why each measure acts as it does.

Your Answer:

median is pretty much in the middle if you don't touch the other lines. Because, everything else is stationary relative to wherever you move this single point. Even if it becomes a super outlier, the median won't really be affected. Meanm, however, will keep following and be affected by that single point.

Our Answer:
As we move the single right-most observation out to the right, the mean also moves to the right, while the median stays in its place. By moving the point to the extreme right we are creating an outlier, and outliers affect the mean and not the median. The mean is affected because it's a numerical average of all the values. The median is affected only by the order of the observations (which has not changed).

Resubmit Reset

Learn By Doing (1/1 point)

Remove all points by clicking on the 'Clear' button. Now place 5 observations on the line by clicking below it. Add one additional observation without changing the median. Where is your new point?

Your Answer:

Where the median is!

(I unfortunately didn't answer the question as it was stated at first)

Our Answer:
For the median to remain as it is, the new point has to be at exactly the same place where the median already is located.

Resubmit Reset

Learn By Doing (1/1 point)

Use the interactive simulation to convince yourself that when you add another observation to those already in place from the last exercise (there will now be 7 in total), the median does not change, regardless of where you put the 7th point. Do you understand why this is the case?

Your Answer:

Yes

Our Answer:

When we had six points, our median was the average of the two central points, which were equal in our case. Once we added one more point (now a total of 7 points), the median stays the same, because one of the two central points from the original dataset will now be among the upper half of the points or the lower half, depending on where you put the new point, but the other one remains as the "center" observation, i.e., the median.

Resubmit Reset

Let's Summarize

- The three main numerical measures for the center of a distribution are the mode, mean ($\overline{x}$), and the median (M). The mode is the most frequently occurring value. The mean is the average value, while the median is the middle value.
- The mean is very sensitive to outliers (as it factors in their magnitude), while the median is resistant to outliers.
- The mean is an appropriate measure of center only for symmetric distributions with no outliers. In all other cases, the median should be used to describe the center of the distribution.