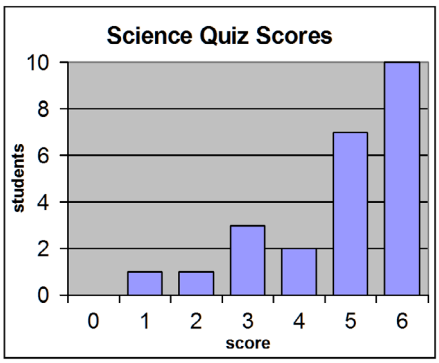
Using Mean, Median, and Mode

- The *mode* can be used with any type of data.
- The *median* can only be used if the data can be put in order.
- The *mean* can be used only if the data is numerical.

Whether you use mean, median, or mode depends both on

- the **type of data** and
- the **shape of the distribution.**

Example. This distribution of science quiz scores is heavily skewed (asymmetrical), and its "peak" is at 6. Clearly, most students did very well on the quiz.



Which of the three measures of center — mean, median, or mode — would best describe this distribution?

Mode: We can see from the graph that the mode is 6.

Median: There are 24 students. The students' actual scores can be read from the graph. They are 1, 2, 3, 3, 3, 4, 4, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6.

The median is the average of the 12th and 13th scores, which is 5.

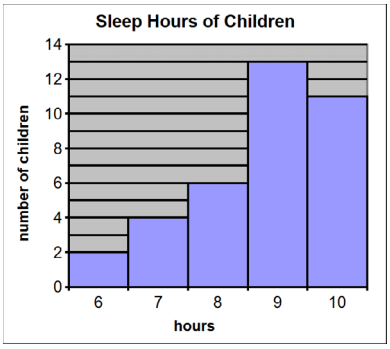
The mean is $\dfrac{1+2+3\cdot 3+2\cdot 4+7\cdot 5+10\cdot 6}{24} = 4.79167 \approx 4.79.$

Notice that the mean is less than 5, but the two highest bars on the graph are at 5 and 6. In this case, the mean does *not* describe the peak of the distribution very well because it actually falls outside the peak!

The median describes the peak reasonably well, but the mode is actually the best in this situation.

1. Fill in.

a. Is the original data numerical? _____



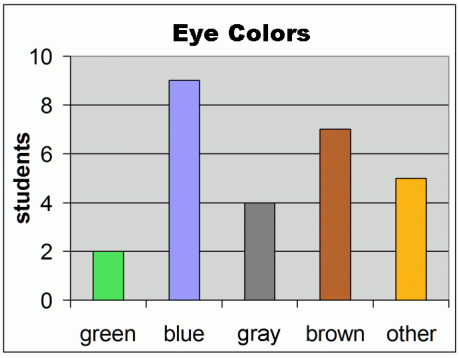
Calculate those measures of center that are possible.

The mode: _____ The median: _____

The mean: _____

Which measure(s) of center describe the peak of the distribution well?

b. Is the original data numerical? _____



Calculate those measures of center that are possible.

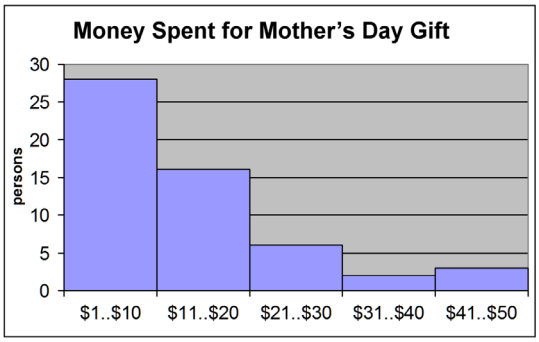
The mode: _____ The median: _____

The mean: _____

Which measure(s) of center describe the peak of the distribution well?

- Mean works best if the distribution is fairly close to a bell shape and does not have outliers. An outlier easily throws off the mean, and then the mean does not accurately describe the center of the data.
- If the distribution is very skewed or has outliers, it is better to use median than mean. Median is not sensitive to outliers — it doesn't get thrown off by an outlier (neither does the mode).

2. Judith asked 55 teenagers about how much money they spent to purchase a Mother's Day gift.

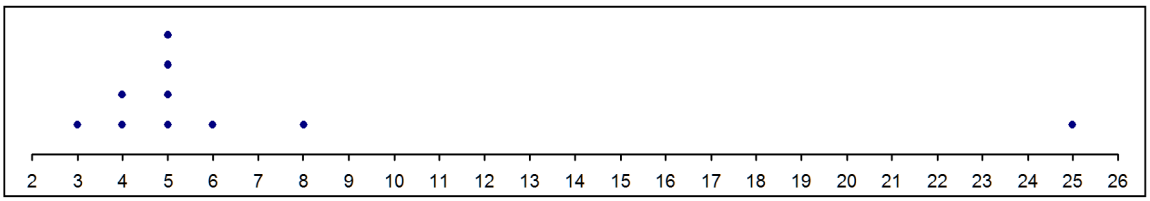


a. The mean is \$11 and the median is \$9. Which of the two better describes this data? Also, explain how your choice relates to the shape of the distribution.

b. *Approximately* what percentage of these teenagers spent \$10 or less on a Mother's Day gift?

3. **a.** Find the mean, median, and mode of this data set: 3, 4, 4, 5, 5, 5, 5, 6, 8, 25. Note that the distribution has an outlier at 25.

mean _____ median _____ mode _____



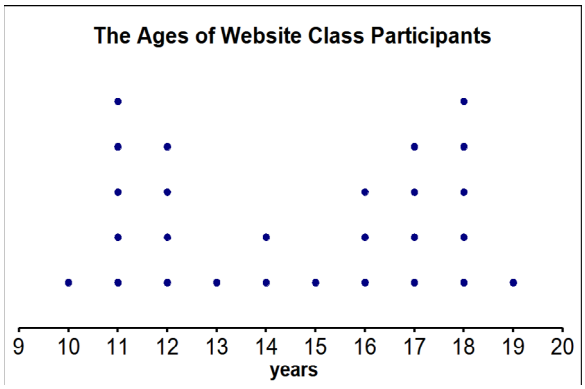
b. Which of the three, mean, median, or mode, best describes the center of this data?

Clearly, either the _____ or the _____ , but *not* the _____!

The _____ is off from the central peak of the distribution.

c. Calculate the mean again if the outlier 25 is omitted. After all, it is so different from the other data items, it could even be a typing error!

4. **a.** Describe the overall shape or pattern of this distribution.

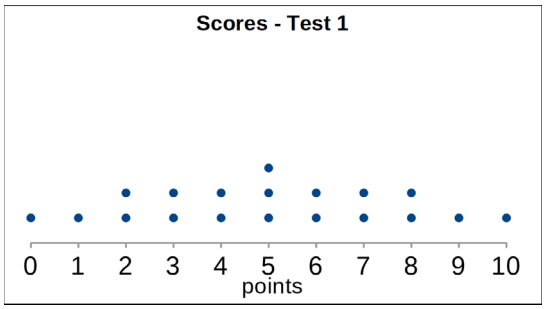


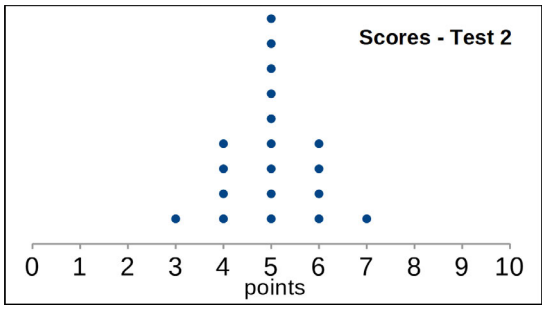
b. Find the mode and the median.

mode _____ median _____

c. Which of the two better describes this distribution, and why?

Range and Interquartile Range





Example 1. Look at the two graphs. The first gives the scores for test 1 and the second for test 2. *Both* sets of data have a mean of 5.0 and a median of 5. Yet the distributions are very different.

How? In test 1, the students got a wide range of different scores; the data is very scattered and **varies a lot**. In test 2, nearly all of the students got a score from 4 to 6. The data is concentrated, or *clustered*, around 5.

We have several ways of measuring the variation in a distribution. One way is to use **range**. Simply put, range is **the difference between the largest and smallest data items.**

For test 1, the smallest score is 0 and the largest is 10 so the range is 10. For test 2, the smallest score is 3 and the largest is 7 so the range is 4. Clearly, the range is much smaller for test 2, indicating the data is much more concentrated than in test 1.

Another measure of variation is the **interquartile range**.

To determine this measure, we first identify the **quartiles**, which are the numbers that divide the data into quarters. The **interquartile range is the difference between the first and third quartiles**. Since the quartiles divide the data into quarters, exactly half of it lies between the first and third quartiles—and it is the middlemost half of the data. The smaller this measure is, the more concentrated the data is.

Example 2. The scores for test 1 are: 0, 1, 2, 2, 3, 3, 4, 4, 5, 5, 5, 6, 6, 7, 7, 8, 8, 9, 10. Let's now find the interquartile range. For that, we need to divide the data into quarters.

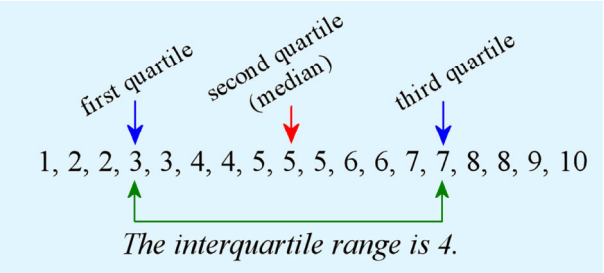
The median naturally divides the data into two halves: 0, 1, 2, 2, 3, 3, 4, 4, 5, **5** , 5, 6, 6, 7, 7, 8, 8, 9, 10.

Now we take the lower half of the data, *excluding the median*, and find *its* median: 0, 1, 2, 2, **3**, 3, 4, 4, 5. That is the **first quartile.**

Similarly, the median of the upper half of the data is the **third quartile**: 5, 6, 6, 7, **7**, 8, 8, 9, 10

The median itself is the **second quartile**.

Together, the three quartiles divide the data into quarters. The interquartile range is the difference between the third and first quartile, or in this case 7 − 3 = **4 points**.



So, exactly half of the test scores lie within 4 points (from 3 to 7 points) around the median of 5 points.

The scores for test 2 are: 3, 4, 4, 4, **4**, 5, 5, 5, 5, **5**, 5, 5, 5, 5, **6**, 6, 6, 6, 7. The quartiles are marked in bold.

The interquartile range is 6 − 4 = **2 points**. So, the middlemost half of the data lies within only 2 points of the median (5 points)—very close to the middle peak of the distribution. This is what we also see in the graph. Clearly, the interquartile ranges show us the same story: the data for test 1 varies much more than for test 2.