



Quiz: Measures of Centrality and Variability

The **mean** of a list of numbers x_1, x_2, \dots, x_n is

$$\text{mean} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

The **median** is a number m for which half the data values are less than or equal to m . To find the median, sort the data then count to the center. The **population standard deviation** is

$$\text{SD} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \text{mean}_x)^2}$$

If the data values are only a sample from the population, then the **sample standard deviation** is

$$\text{sd} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \text{mean}_x)^2}$$

1. Find the mean, median and sd for each list. Explain how lists b)–d) are related to list a) and what impact that has on the three statistics.

a) 1, 3, 4, 4, 5, 7

b) 11, 13, 14, 14, 15, 17

c) 2, 6, 8, 8, 10, 14

d) -3, -9, -12, -12, -15, -7

2. Two histograms are shown below. Indicate the approximate locations of the mean and median for each.



3. True or false and explain: The mean and the median can't be negative.
4. True or false and explain: The mean and the median must be equal.
5. True or false: the mean is always greater than the sd? Find an example to justify your answer.
6. True or false and explain: SD and sd can't be negative.
7. What are the SD and sd of the list 5, 5, 5, 5, 5?
8. Make a rough sketch of a histogram of family incomes in the U.S. Which is greater, the mean income or the median income? Why?
9. Find a value x to make the mean of the list 1, 2, x equal to 2. Justify your answer.
10. Find a value x to make the mean of the list 1, 2, x equal to 100. Justify your answer.



Quiz (Continued): Percentiles and Quantiles

The mean and standard deviation can be used to summarize a data set. Frequency tables and histograms give more details about the distribution of the data. Percentiles (including quantiles) are another tool for summarizing the data distribution. The median is an example of a percentile. It is the 50th percentile. Half the data is less than the median and half is greater.

Example. Here is a small data set:

6 2 17 5 5 3 20 19 10 12 19

To compute the median, we first need to sort the data.

2 3 5 5 6 10 12 17 19 19 20

Here are a few examples of percentiles:

- The minimum is the 0% since none of the data is smaller. The 0% is 2.
- The median is 10.
- The maximum is the 100% since none of the data is greater. The 100% is 20.
- The 25% is a value v for which 25% of the data is at most v . It is the median of the first half of the data: 2, 3, 5, 5, 6, 10. So, the 25% is 5.
- The 75% is the median of the second half of the data: 10, 12, 17, 19, 19, 20. So, the 75% is any value between 17 and 19. Most statistical software will take the average and output 18.

Here is how to do this in R.

```
> x = c(6, 2, 17, 5, 5, 3, 20, 19, 10, 12, 19)
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   2.00   5.00   10.00   10.73   18.00   20.00
> fivenum(x)
[1]  2  5 10 18 20
> range(x)
[1]  2 20
```

11. Explain in words what the `summary`, `fivenum` and `range` commands in R do.

12. For the following data, find the 0%, 25%, 50%, 75% and 100% by hand then check your answers with R.

6 10 20 19 1 1 11 13 13 17 2

.

13. Repeat #2 using the following data.

98 83 65 56 58 13 97 4 100 63

.