

## Practice Assignment: Interpretation of Mean and Median

---

1) In which of the following situations is it likely that the mean will be **lower** than the median?

- a) The dataset has many low and high outliers, which are symmetrically distributed.
- b) The dataset has 500 data values.
- c) The dataset has right skew and several high outliers.
- d) The dataset is mostly symmetric but contains several extremely low outliers and no high outliers.

Answer: d

2) When is it most appropriate to use the mean as a measure of center?

- a) When the distribution has no outliers
- b) When the distribution is symmetric
- c) When the distribution is skewed
- d) When the distribution has outliers

Answer: b

Note: Some students might choose (a), but a distribution without outliers may still have skew.

3) Determine whether this statement is true or false: It doesn't make sense to calculate a mean or median when describing categorical data.

Answer: True

Questions 4–8: These questions reference the *FiveThirtyEight* hate\_crimes dataset,<sup>1</sup> which describes hate crime rates within different states.

The following table displays selected variables from the full dataset for the first 10 states (including DC).

---

<sup>1</sup> *Hate crimes*. (2018, February 9). FiveThirtyEight.  
<https://github.com/fivethirtyeight/data/tree/master/hate-crimes>

state_abbrev	median_house_inc	hate_crimes_per_100k_spl c	avg_hatecrimes_per_100k_fb i
AL	42278	0.125838926	1.806410489
AK	67629	0.143740118	1.656700109
AZ	49254	0.225319954	3.413927994
AR	44922	0.069060773	0.869208872
CA	60487	0.255805361	2.397985899
CO	60940	0.390523301	2.804688765
CT	70161	0.335392269	3.772701469
DE	57522	0.322754169	1.469979563
DC	68277	1.52230172	10.95347971

The following are descriptions of each variable in the table:

*state\_abbrev*: State abbreviation

*median\_house\_inc*: Median household income, 2016

*hate\_crimes\_per\_100k\_splc*: Hate crimes per 100,000 population, Southern Poverty Law Center, Nov. 9–18, 2016

*avg\_hatecrimes\_per\_100k\_fbi*: Average annual hate crimes per 100,000 population, FBI, 2010–2015

- 4) Why do you believe the dataset authors chose to report the median household income within each state rather than the mean?

Answers will vary.

Sample answer: Income data tend to be severely right skewed. Most households have earnings within a common range (e.g., \$0–\$80,000), but some households tend to have very high incomes (e.g., \$250,00+). Also, the asymmetric nature of incomes—that they can climb very high but cannot get below \$0—often results in high outliers. So, the mean will provide a relatively high measure of center due to undue influence from larger incomes. The median, which is more resistant to skew and outliers, will be more representative of the “typical” household income within a state.

- 5) Go to the *Describing and Exploring Quantitative Variables* tool at [https://lumen-learning.shinyapps.io/eda\\_quantitative/](https://lumen-learning.shinyapps.io/eda_quantitative/). Select the **Single Group** tab, locate the drop-down menu under “Dataset,” and select “Hate Crimes (SPLC).” This dataset includes the data from the variable *hate\_crimes\_per\_100k\_splc*.

Part A: Find the **mean** hate crimes committed per 100,000 people across the states and DC. Note that four states with missing values were excluded.

Answer: mean  $\approx 0.304$  hate crimes per 100,000 people

Part B: Find the **median** hate crimes committed per 100,000 people across the states and DC.

Answer: median  $\approx 0.226$  hate crimes per 100,000 people

Part C: Use the histogram displayed on the online tool to visualize the distribution of hate crimes committed per 100,000 people across the states and DC. Comment on the shape of the distribution and how it relates to the relationship between the mean and median you calculated previously.

Answers will vary.

Sample answer: The distribution is skewed right, with one particularly high outlier. The mean tends to follow skew and outliers, so it is higher than the median (which is more resistant to skew and outliers).

6) There is one data value that has a particularly powerful influence on the previous mean calculation. Use the previous abbreviated table to identify which U.S. state (including DC) is associated with that data value and comment on whether it should be removed before calculating the mean hate crimes per 100,000 people.

Answers will vary.

Sample answer: The high outlier of Washington DC, with 1.52 hate crimes per 100,000 people, has a powerful influence on the mean. Including DC in the calculation inflates the mean value.

One could argue that the value should be removed because DC is a city, and lumping together the crime rate of one city with the crime rates of full states (which are composed of cities, suburban areas, and rural areas) may lead to strange results. However, one may argue that failing to include DC would be unjust exclusion.

Note: Other strategies (outside the scope of this lesson) may include:

1) Using the median

2) Finding the national total hate crimes committed and the national population size; then, dividing these numbers to get the true hate crime rate nationally

