M2-02: Descriptive Statistics

Part of the "Exploratory Data Analysis" Learning Badge

Video Walkthrough: *https://discovery.cs.illinois.edu/m2-02/*

Discovering Properties of Means and Medians

Let's explore a few basic properties of center and spread using our "Party Dataset", where we explored your perception of people at a party:

	Several	A Few	Many	A Couple	Dozens	Some	Scores Of
0	5	5	10	4	12	15	25
1	5	3	12	2	24	4	35
...							

Dataset URL: https://waf.cs.illinois.edu/discovery/party.csv

Property #1: Adding or Subtracting a Constant to All Observations

To discover the first property, we will find the mean and median of the dataset, then add (or subtract) the same number from every row in the data, and find the new mean and median.

Step 1: Find the mean and median before modifying the dataset for the variable _____:

	Python	Dataset Values	
		Original Data	**After Modification**
μ (Mean):			
Median:			

Step 2: Modify the dataset by adding a constant to each observation:

Python:	
Description:	Adds a constant value of _____ to each observation of _____.

Step 3: Find the mean and median with the modified dataset (completing Step 1).

Analysis: How did adding/subtracting the same number to all observations change the mean and median values?

Property #2: Multiplying or Dividing a Constant to All Observations

Similar to our first property, what happens if all of our data is multiplied or divided by the same number?

Step 1: Find the mean and median before modifying the dataset for the variable _____:

	Python	Dataset Values	
		Original Data	After Modification
μ (Mean):			
Median:			

Step 2: Modify the dataset by multiplying a constant to each observation:

Python:	
Description:	Adds a constant value of _____ to each observation of _____.

Step 3: Find the mean and median with the modified dataset (completing Step 1).

Analysis: How did multiplying/dividing the same number to all observations change the mean and median values?

Q: Does this work with negative numbers?

Property #3: Sensitivity to Outliers

Thinking back to the example where ten people with a salary of $100,000 were in a room and a trillionaire walked in. You learned that:

- The median value does not change, **but**
- The mean value changes significantly!

Exploring the "Party Dataset" again, what happens if we **add a new observation** where the respondent says that a "a couple of people were at the party" meant that 1,000,000 people were there! What happens?

Step 1: Find the mean and median before modifying the dataset for the variable _____:

	Python	Dataset Values	
		Original Data	After Modification
μ (Mean):			
Median:			

Step 2: Modify the dataset by adding a new record:

Python:	
Description:	Adds a new record with the variable "A Couple" observed to be 1,000,000.

Step 3: Find the mean and median with the modified dataset (completing Step 1).

Analysis: How did adding the record to "A Couple" change the mean and median?