

QUIZ

You are creating a report on annual car sales from a dataset that contains information on every car sold at your dealership this year. One of the variables is the color of the car, which includes categories like “gray”, “white”, “red”, and “blue”.

To describe the variable for color, you want to make a table of measurements. What kind of measurements could go in your table for this kind of variable?

The frequency and proportion of each color in the dataset.



Correct! Color is a categorical variable, so we can record frequencies and proportions of its categories.

The median and interquartile range of the color variable.

The mean and standard deviation for the color variable.

True or False: We can always get precise summary information about the variables in a dataset just by viewing them in the dataframe.

False



Correct! We use summary statistics to measure and describe variables because it is usually difficult to understand anything precise just by viewing the dataset.

True

Fill in the blanks to make the statement about a correlation coefficient true.

A correlation coefficient of -0.97 tells us about the relationship between two variables.

The negative sign indicates that low values of the first variable are associated with values of the second, and high values of the first variable are associated with values of the second.

The distance of the coefficient from zero tells us the of the relationship, which is because -0.97 is very close to -1 .



You got it!

Fill in the blanks to make the following statements true.

Extreme values that are distant from the rest of the distribution are called **outliers**. These extreme values may heavily influence the **mean** more than the **median**. A measure that is not as heavily influenced by outliers and **skewness** is called a **robust** statistic.



You got it!

You are reading about a survey on time spent cleaning different areas of the home. Participants were asked how many hours per week they spent cleaning each room of their home. The mean and standard deviation times in hours are shown in the table. Which statement about the distributions of cleaning times is true?

Cleaning times in hours

ROOM	MEAN	STANDARD DEVIATION
Living room	1.2	0.9
Kitchen	3.8	1.9
Bathroom	1.5	0.6
Dining room	1.2	0.2

The center of the distribution for the kitchen is 1.9 hours.

The average distance from the center for values in the bathroom's distribution is 1.2 hours.

The distributions for the living room and dining room have the same center value, but the distribution for the living room has a wider spread.



Correct! The distributions for these rooms have the same mean, but different standard deviations.

You are reading an article about retirement ages. The article reports that the median retirement age is 65.4 years old with an interquartile range of 7.2 years. Which of the following statements about the distribution of retirement ages is true?

The second quartile is greater than 65.4.

Half of the ages in the distribution are less than 65.4 and the other half are greater than 65.4. The middle 50% of the data spans 7.2 years.



Correct! The median divides the data in half at the middle value, while the IQR gives us the range of the middle 50% of the data.

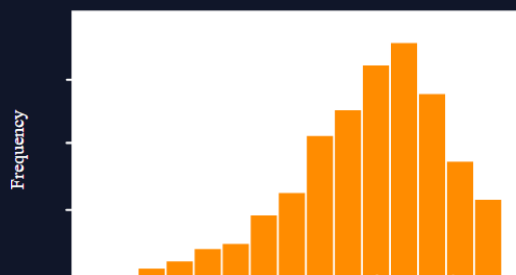
Fill in the blanks to make the statements true.

You want to explore the relationship between two numeric variables: age and height. To visually check the relationship, it would be best to make a ☒ **scatter plot** of the variables. To get a numeric summary of the linear relationship between the variables, you should also get their ☒ **correlation coefficient**.



You got it!

Which of the following is true about the distribution shown?

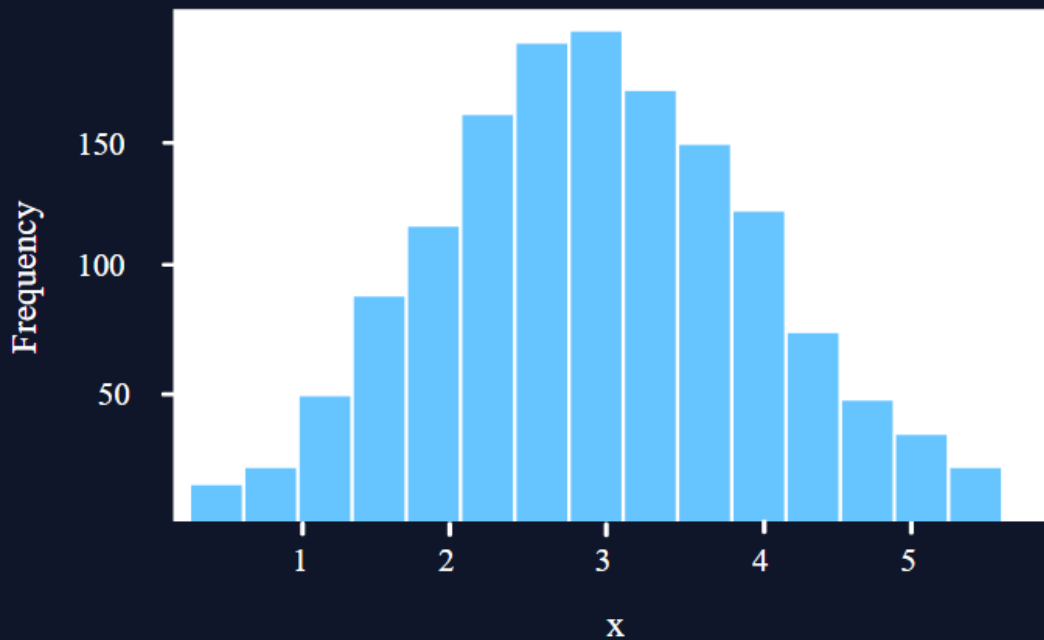


The distribution may be considered skewed because it is asymmetrical with a longer tail on the left side.



Correct! This distribution is negatively skewed.

Fill in the blanks to make the statements about the plot true.



This plot shows the ☒ distribution of the numeric variable x, which gives the possible ☒ values of a variable and their ☒ frequencies.



You got it!

True or False: Finding the mean number of cities for every region of a country is an example of aggregating data.

False

True



Correct! We can aggregate a numeric variable by summarizing it across every level of a categorical variable.

Fill in the blanks to make the statements true.

You want to explore the relationship between two numeric variables: age and height. To visually check the relationship, it would be best to make a of the variables. To get a numeric summary of the linear relationship between the variables, you should also get their .



You got it!

True or False: Finding the mean number of cities for every region of a country is an example of aggregating data.

True



Correct! We can aggregate a numeric variable by summarizing it across every level of a categorical variable.

False