Overview

27 December 2021 08:42

Lesson Overview

In this lesson, we will continue to cover more topics related to analyzing quantitative variables and you will learn to use **measures of spread.** Measures of spread are used to provide us an idea of how spread-out our data are from one another.

In this lesson you will:

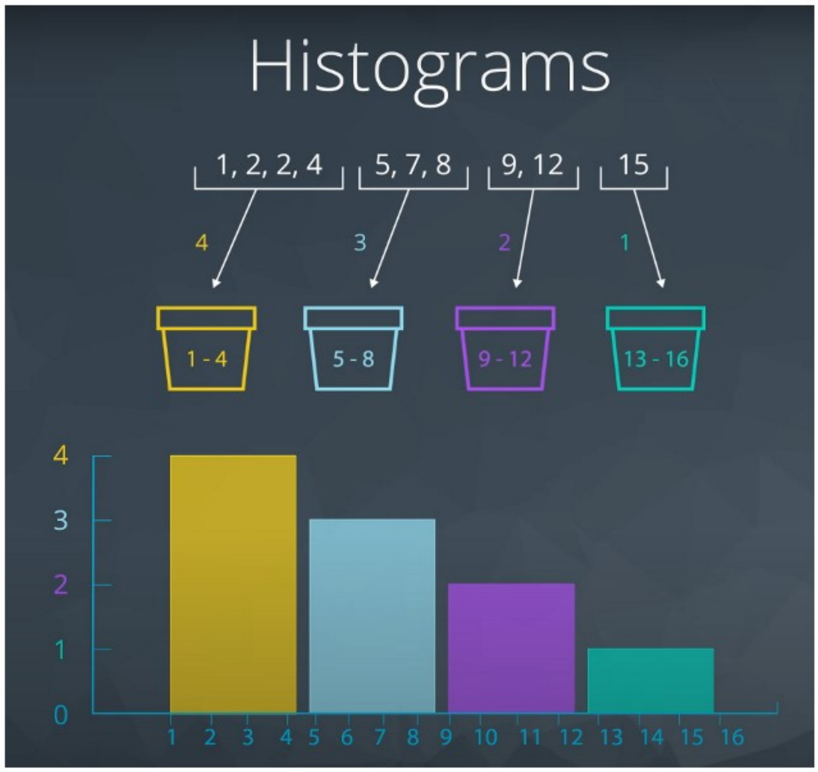
- Evaluate measures of spread
- **Range**
- **Interquartile Range (IQR)**
- **Standard Deviation**
- **Variance**
- Analyze outliers
- Evaluate descriptive and inferential statistics

Throughout this lesson, you will learn how to calculate these, as well as why we would use one measure of spread over another.

Histograms

27 December 2021 08:49

Histograms



- Super useful for understand'g different aspect of Data and they are most common visuals for quantitative data.
- Helps in understand'g the four aspects regarding a quantitative variable:
  - Centre
  - Shape
  - Spread
  - Outliers

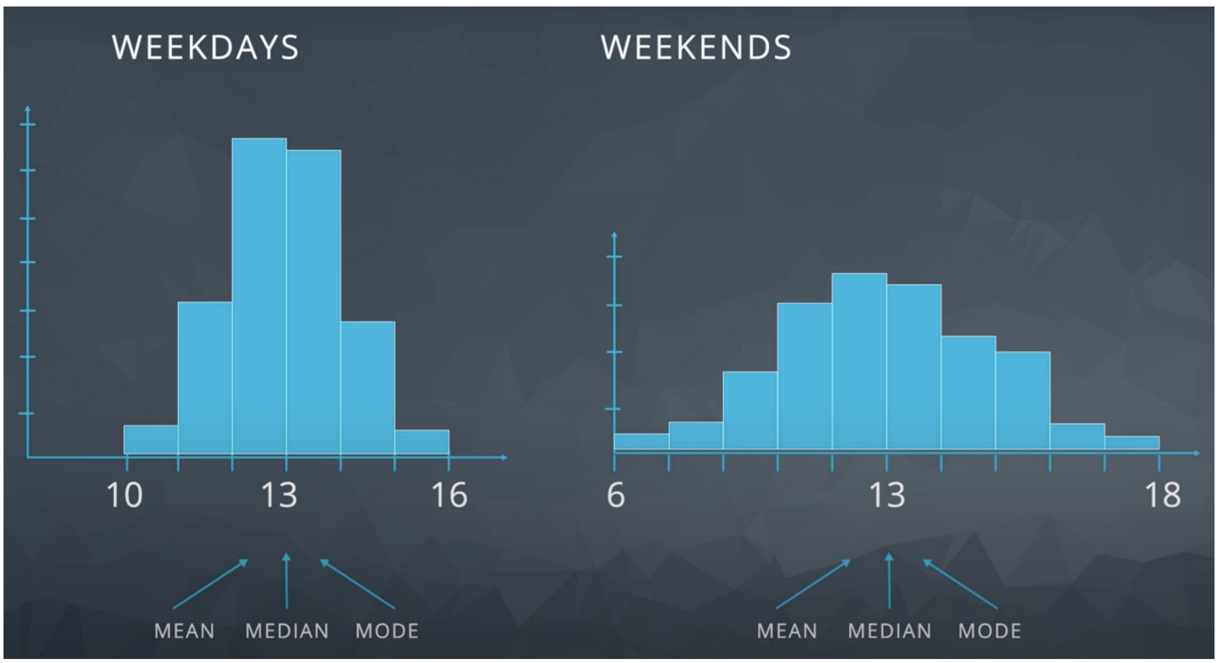
How are Histograms are constructed?

First, we need to bin our data. Each **bin** represents a range of values in a dataset. The number of values that fall in the range of each bin determines the height of each histogram bar. As shown in the picture above, changing the range of our bins can result in slightly different visuals. However, there is no right or wrong answer in choosing how to bin, and in most cases, the software you use will choose the appropriate bins for you.

Motivation for spread of data

27 December 2021 08:59

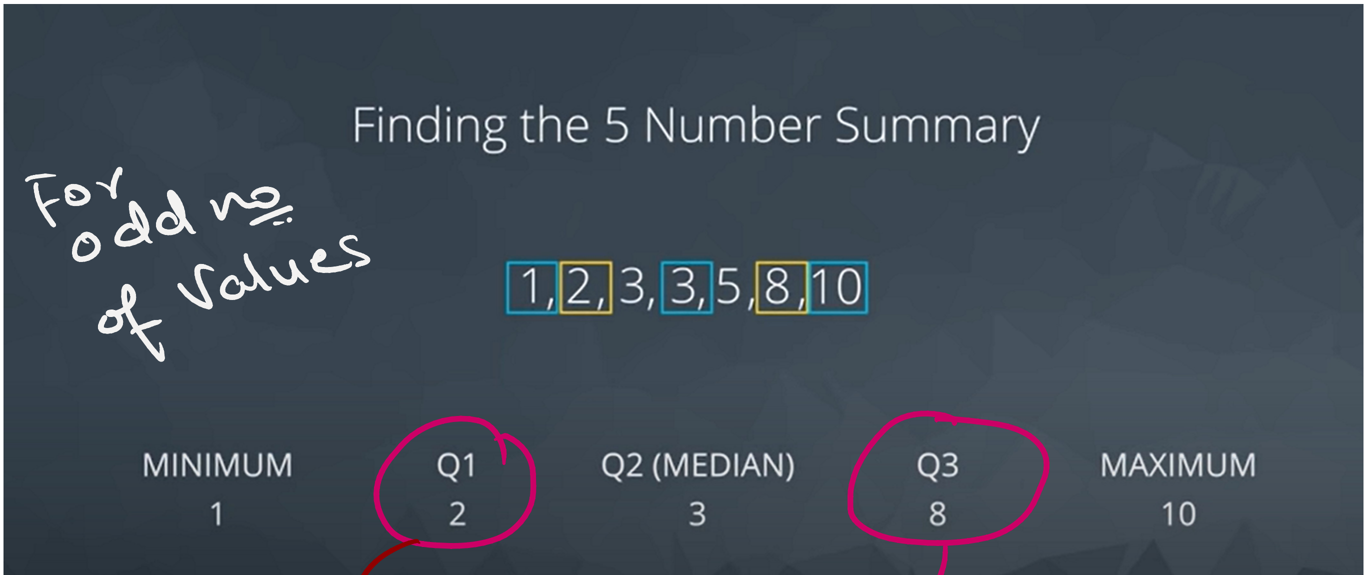
The two histograms below illustrate the number of dogs Josh saw on weekdays versus weekends. The measures of center for both histograms (**mean, median, mode**) are basically the same and centered about the highest bin for both histograms, 13.



Visually, the difference between the histograms is the range or spread of dogs Josh sees during each time period. In the upcoming lessons, we will discuss the most common ways to **measure the spread** of our data.

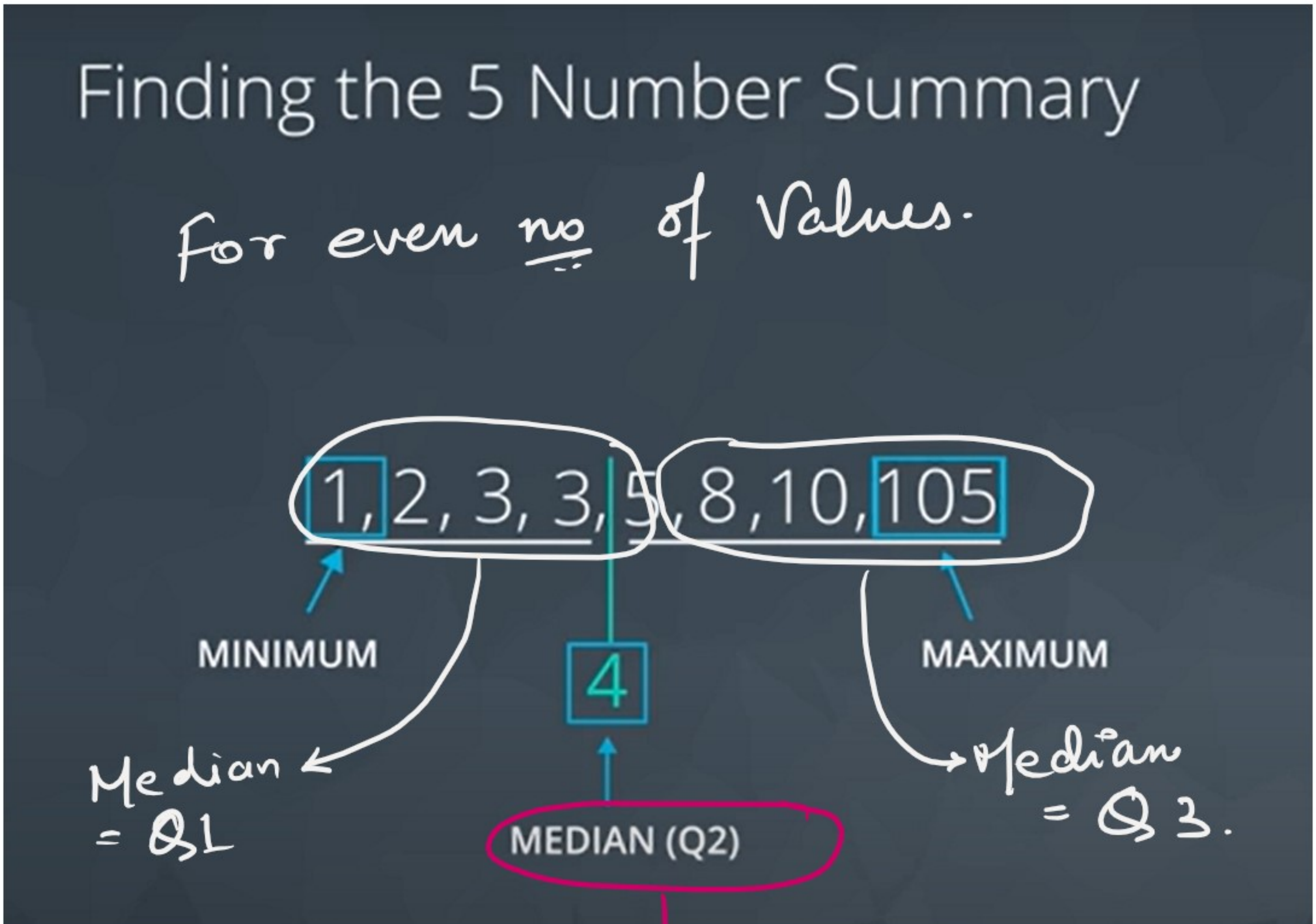
5 number Summary

27 December 2021 09:02

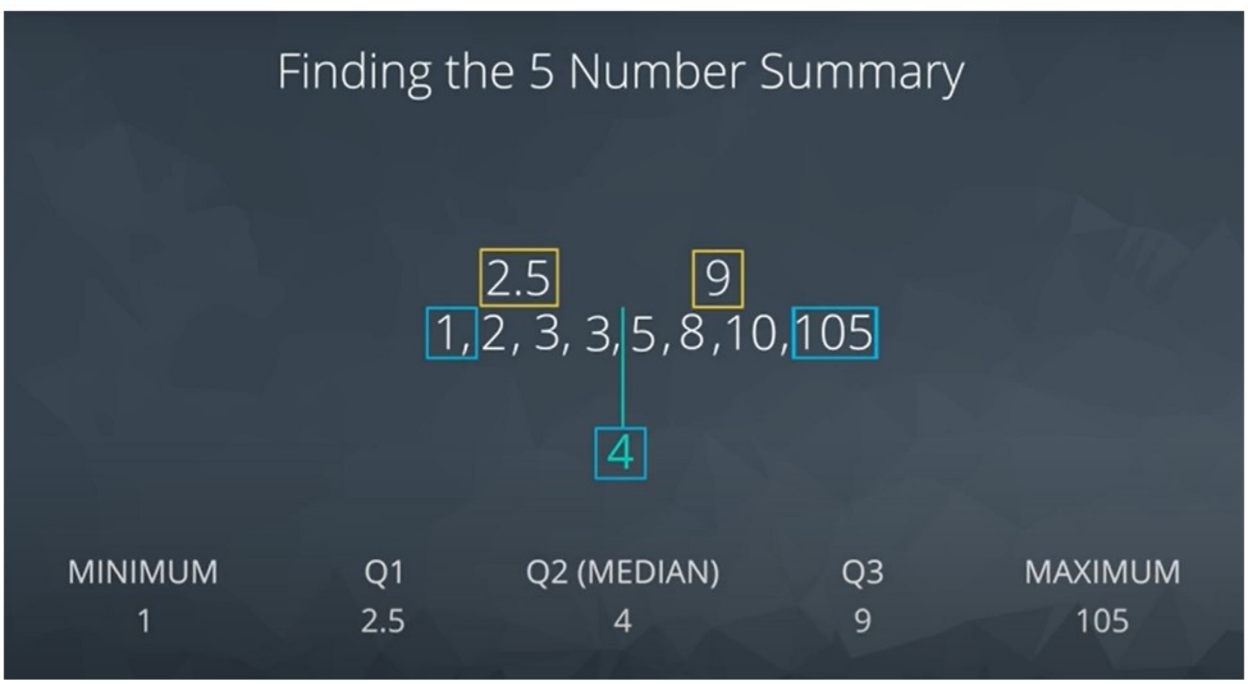


Median of numbers left of Q2.

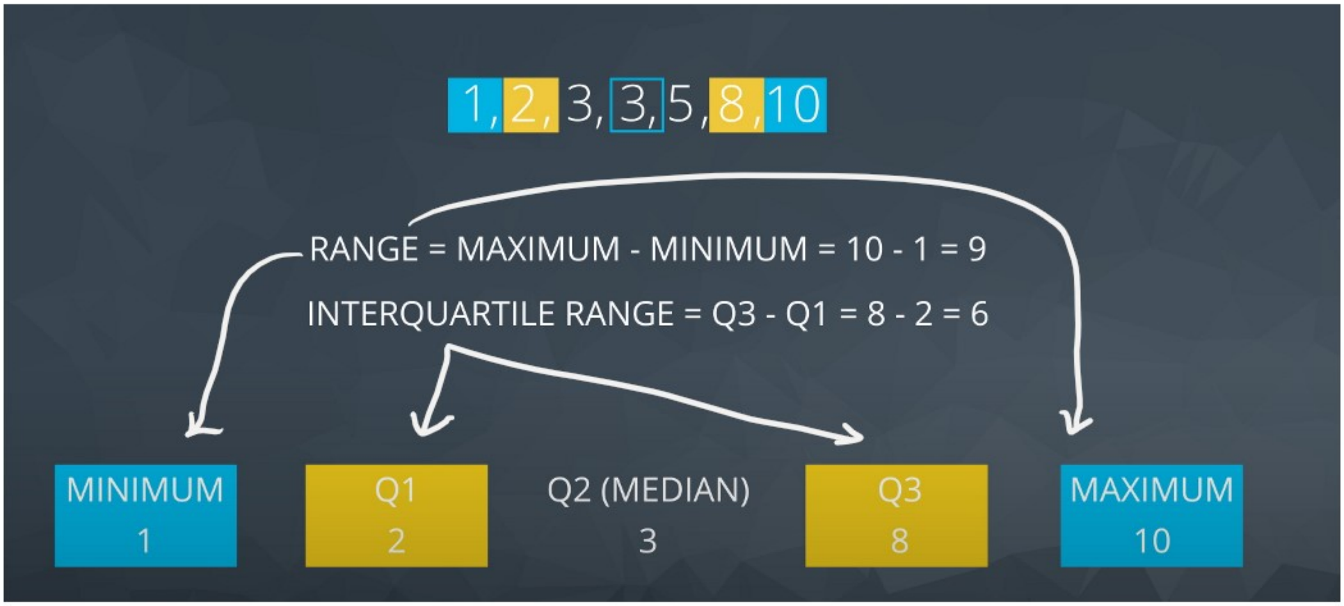
median of numbers ri8. of Q2


Finding the 5 Number Summary
For even no of Values.
1, 2, 3, 3, 5, 8, 10, 105
MINIMUM
4
MAXIMUM
Median = Q1
MEDIAN (Q2)
Median = Q3.

Find Q2.


Finding the 5 Number Summary
2.5
9
1, 2, 3, 3, 5, 8, 10, 105
4
MINIMUM
1
Q1
2.5
Q2 (MEDIAN)
4
Q3
9
MAXIMUM
105

Range and IQR(Inter Quartile Range)



Calculating the 5 Number Summary

The five-number summary consist of 5 values:

1. **Minimum:** The smallest number in the dataset.
2. **Q1**: The value such that 25% of the data fall below.
3. **Q2**: The value such that 50% of the data fall below.
4. **Q3**: The value such that 75% of the data fall below.
5. **Maximum:** The largest value in the dataset.
   In the above video, we saw that calculating each of these values was essentially just finding the median of a bunch of different datasets. Because we are essentially calculating a bunch of medians, the calculation depends on whether we have an odd or even number of values.

Range

The **range** is then calculated as the difference between the **maximum** and the **minimum**.

IQR

The **interquartile range** is calculated as the difference between **Q3** and **Q1**.

Exercise

27 December 2021 09:11

QUESTION 1 OF 2

Identify the following for this dataset:

1, 5, 10, 3, 8, 12, 4, 1, 2, 8

QUESTION 1 OF 2

Identify the following for this dataset:

- Dataset: 1, 5, 10, 3, 8, 12, 4, 1, 2, 8
  - Range
  - 1st quartile
  - Median
  - 3rd Quartile

[37]
```
nums = [1, 5, 10, 3, 8, 12, 4, 1, 2, 8]
sorted_nums = sorted(nums)
print(sorted_nums)
```

```
[1, 1, 2, 3, 4, 5, 8, 8, 10, 12]
```

[73]
```
median = statistics.median(sorted_nums)
Q_1 = statistics.median(sorted_nums[:math.floor(median)+1])
Q_3 = statistics.median(sorted_nums[math.ceil(median):])
print(f"Sorted Numbers = {sorted_nums}")
print(f"{'Minimum:': <16} {min(sorted_nums)}")
print(f"{'Q1:': <16} {Q_1}")
print(f"{'Median or Q2:':<16} {median}")
print(f"{'Q3:': <16} {Q_3}")
print(f"{'Maximum:':<16} {max(sorted_nums)}")
print(f"{'Range:':<16} {max(sorted_nums) - min(sorted_nums)}")
print(f"{'IQR:':<16} {Q_3 - Q_1}")
```

```
Sorted Numbers = [1, 2, 3, 4, 5, 8, 8, 10, 12]
Minimum:         1
Q1:              3.5
Median or Q2:    5
Q3:              9.0
Maximum:         12
Range:           11
IQR:             5.5
```

QUESTION 2 OF 2

Identify the following for this dataset:

- Dataset: 5, 10, 3, 8, 12, 4, 1, 2, 8
    - Range
    - 1st quartile
    - Median
    - 3rd Quartile

```
"""Odd number of Values"""

nums = [5, 10, 3, 8, 12, 4, 1, 2, 8]
sorted_nums = sorted(nums)
print(f"Sorted Numbers = {sorted_nums}")
median = statistics.median(sorted_nums)
median_index = sorted_nums.index(median)
Q_1 = statistics.median(sorted_nums[:median_index])
Q_3 = statistics.median(sorted_nums[median_index+1:])
print(f"{'Minimum:': <16} {min(sorted_nums)}")
print(f"{'Q1:': <16} {Q_1}")
print(f"{'Median or Q2:':<16} {median}")
print(f"{'Q3:': <16} {Q_3}")
print(f"{'Maximum:':<16} {max(sorted_nums)}")
print(f"{'Range:':<16} {max(sorted_nums) - min(sorted_nums)}")
print(f"{'IQR:':<16} {Q_3 - Q_1}")
```

```
Sorted Numbers = [1, 2, 3, 4, 5, 8, 8, 10, 12]
Minimum:         1
Q1:              2.5
Median or Q2:    5
Q3:              9.0
Maximum:         12
Range:           11
IQR:             6.5
```

Box Plots

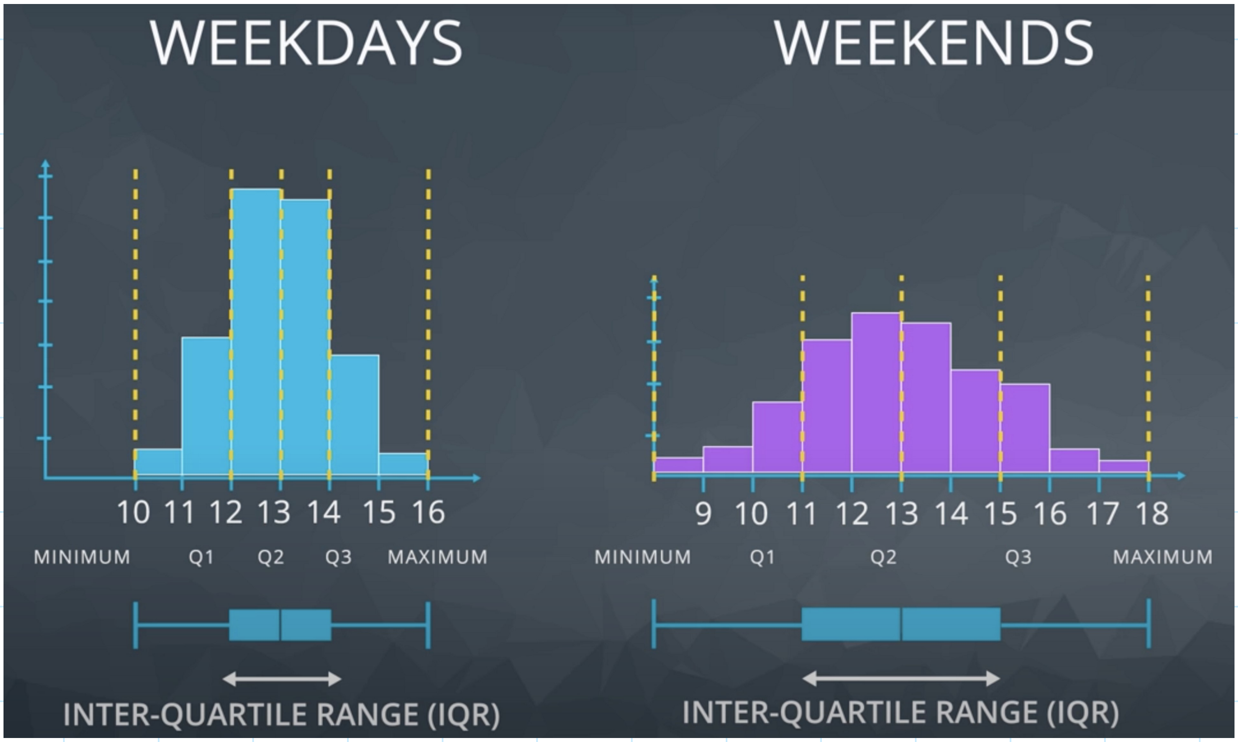
27 December 2021 10:27

Box Plots.

Box plots are useful for quickly comparing the spread of two datasets across some key metrics, like quartiles, maximum, & minimum.

How do we create the box plot?

1. The beginning of the line to the left of the box and the end of the line to the right of the box represent the minimum and maximum values in a dataset.
2. The visual distance between these markings is an indication of the range of the values.
3. The box itself represents the IQR. The box begins at the Q1 value, ends at the Q3 value, and Q2, or the median, is represented by a line within the box.



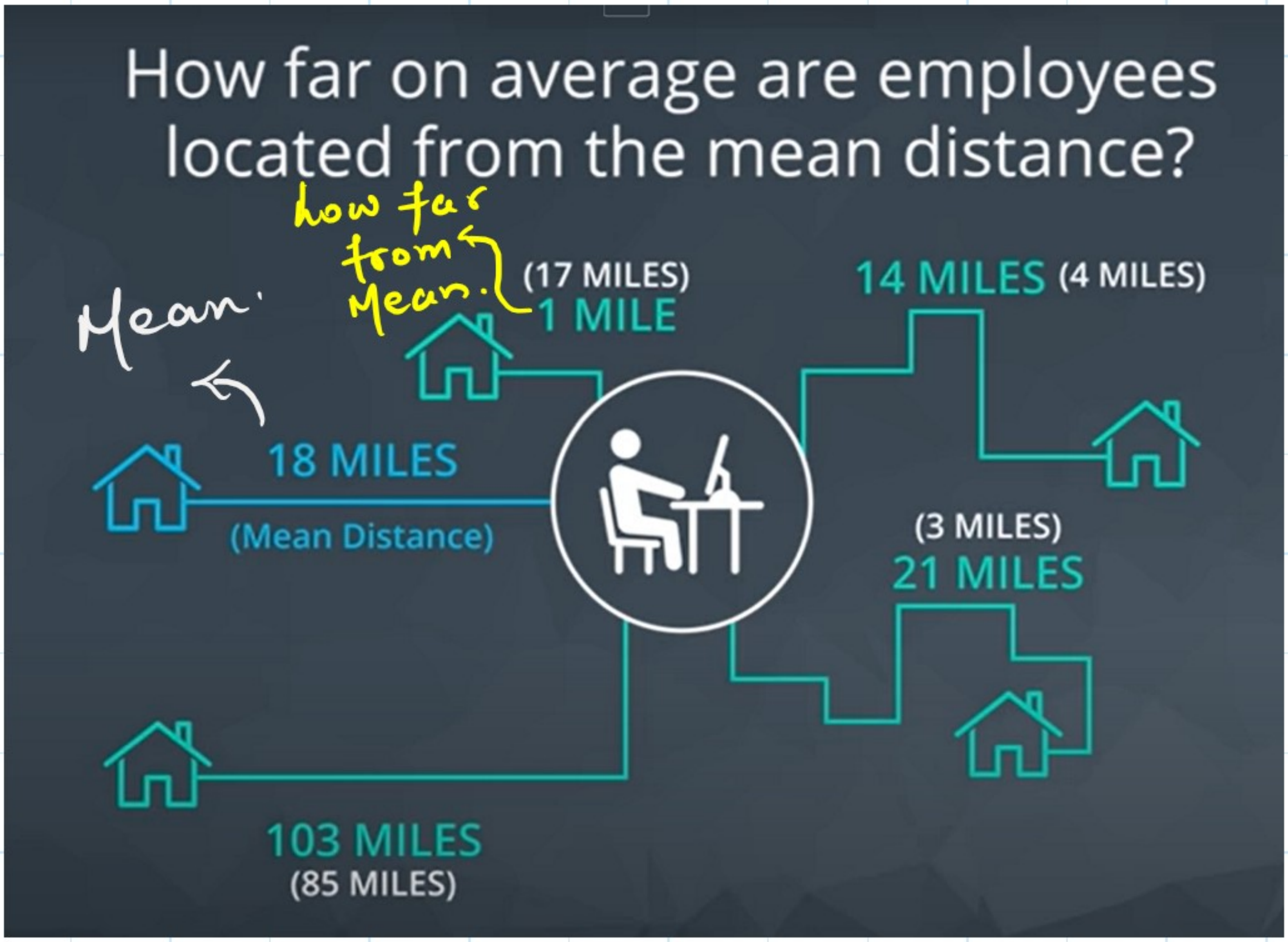
From both the histograms and box plots, we can see that the number of dogs seen on weekends varies much more than on weekdays.

Insted of depending on a visual of 5 number summary to compare our data, we may use single value to compare the two distribution spreads - Standard Deviation

Standard Deviation

27 December 2021 10:37

Standard Deviation



Standard Deviation and Variance

The **standard deviation** is one of the most common measures for talking about the spread of data. It is defined as **the average distance of each observation from the mean**.

In the above picture, we saw this as how far individuals were from the average distance from work (the example distances shown are examples from the full data set, the mean of just those 4 numbers is 38.5. The mean of 18 shown later in the picture is the mean of the full data set.

Standard Deviation Calculation

27 December 2021 10:44

How to Calculate Standard Deviation

Dataset = 10, 14, 10, 6

1. Calculate the mean $(\sum_{i=1}^{4} x_i)/n$ = 40/4 = 10
2. Calculate the distance of each observation from the mean and square the value

$(x_i - \overline{x})^2$	=
10-10	0
14-10	16
10-10	0
6-10	16

3. Calculate the **variance**, the average squared difference of each observation from the mean

$\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$	=
(0+16+0+16)/4	8

4. Calculate the **standard deviation**, the square root of the variance

$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$	=
$\sqrt{8}$	2.83

is on average, how far each point in our dataset is from the mean.

Exercise

27 December 2021 10:59

Standard Deviation Question 2 OF 2

For the following set of data match the correct values.

- Variance and Standard Deviation
- Remember to find the variance we first find the mean average of the values.
- Then subtract the mean from each value
- Then square each of these values, then add them up
- Then divide by the number of values.
- (Round your answer to two decimal places at the end of your calculation - don't round along the way.)

DataSet: 1, 5, 10, 3, 8, 12, 4

```
[92] nums = [1, 5, 10, 3, 8, 12, 4]
     mean_data = statistics.mean(nums)
     print(f"{'Mean of DataSet:': <33} {mean_data}")
     mean_subt_each_val = [(val - mean_data) for val in nums]
     print(f"{'Mean Subtracted from each value:': <33} {mean_subt_each_val}")
     mean_subt_each_val_squared = [val **2 for val in mean_subt_each_val]
     print(f"{'Each value Squared:': <33} {mean_subt_each_val_squared}")
     each_val_squared_added = sum(mean_subt_each_val_squared)
     print(f"{'Each value Squared and Added:': <33} {each_val_squared_added}")
     final_division_by_num_values = round(each_val_squared_added / len(nums),2)
     print(f"{'Variance of Data:': <33} {final_division_by_num_values}")
     print(f"{'Standard Deviation of Data:': <33} {round(final_division_by_num_values ** 0.5, 2)}")
```

```
Mean of DataSet:                  6.142857142857143
Mean Subtracted from each value:  [-5.142857142857143, -1.1428571428571432, 3.8571428571428568, -3.1428571428571432, 1.8571428571428568, 5.857142857142857, -2.1428571428571432]
Each value Squared:               [26.44897959183674, 1.3061224489795926, 14.877551020408161, 9.877551020408166, 3.4489795918367334, 34.306122448979586, 4.591836734693879]
Each value Squared and Added:     94.85714285714286
Variance of Data:                 13.55
Standard Deviation of Data:       3.68
```

More on Standard Deviation

27 December 2021 11:00

Other Measures of Spread

5 Number Summary

In the previous sections, we have seen how to calculate the values associated with the **five-number summary** (**min**, Q_1, Q_2, Q_3, **max**), as well as the measures of spread associated with these values (**range** and **IQR**).

For datasets that are **not symmetric**, the five-number summary and a corresponding box plot are a great way to get started with understanding the spread of your data. **Although I still prefer a histogram in most cases, box plots can be easier to compare two or more groups.** You will see this in the quizzes towards the end of this lesson.

Variance and Standard Deviation

Two additional **measures of spread** that are used all the time are the **variance** and **standard deviation**. At first glance, the variance and standard deviation can seem overwhelming. If you do not understand the expressions below, don't panic! In this section, I just want to give you an overview of what the next sections will cover. We will walk through each of these parts thoroughly in the next few sections, but the big picture goal is to generally understand the following:

1. How the mean, variance, and standard deviation are calculated.
2. Why the measures of variance and standard deviation make sense to capture the spread of our data.
3. Fields, where you might see these values used.
4. Why we might use the standard deviation or variance as opposed to the values associated with the 5 number summary for a particular dataset.

Calculation

We calculate the variance in the following way:

$$\frac{1}{n}\sum\limits_{i=1}^n(x_i - \bar{x})^2$$

The variance is **the average squared difference of each observation from the mean**.

To calculate the variance of a set of 10 values in a spreadsheet application, with our 10 data points in column A, we would create a new column B by typing in something like **=A1-AVERAGE(A$1:A$10)** and copying this down for all 10 rows. This would find us the difference between each data point and the mean average of all the data. Then we create a new column C having the square of these differences, using the formula **=B1^2** in cell C1, and copying that down for all rows. Then in the cell below this new column, cell C11, type in **=SUM(C1:C10)**. This adds up all these values in column C. Finally in cell C12, we divide this sum by the number of data points we have, in this case, ten: **=C11/10**. This cell C12 now contains the variance for our 10 data points.

More detailed guidance on using spreadsheets like this may be included in a future lesson in your program.

The standard deviation is the square root of the variance. Therefore, the formula for the standard deviation is the following:

$$\sqrt{\frac{1}{n}\sum\limits_{i=1}^n(x_i - \bar{x})^2}$$

In the same spreadsheet as above, to find the standard deviation of our same set of 10 data values, we would use another cell like C13 to take the square root of our variance measure, by typing in **=sqrt(C12)**.

The standard deviation is a measurement that has the same units as our original data, while the units of the variance are the square of the units in our original data. For example, if the units in our original data were dollars, then units of the standard deviation would also be dollars, while the units of the variance would be dollars squared.

Again, **this section is designed as background knowledge for the following sections**. If it doesn't make sense on this first pass, do not worry. You will be guided in future sections in performing these calculations, and building your intuition, as you work through an example using the salary data. Then we will provide context about why these calculations are important, and where you might see them!

Why Standard Deviation

27 December 2021 11:05

Standard deviation is a common metric used to compare the spread of two datasets. The benefits of using a single metric instead of the 5 number summary are:

- It simplifies the amount of information needed to give a measure of spread
- It is useful for inferential statistics

Some important points on Std Dev

27 December 2021 11:07

Important Final Points

1. The variance is used to compare the spread of two different groups. A set of data with higher variance is more spread out than a dataset with lower variance. Be careful though, there might just be an outlier (or outliers) that is increasing the variance when most of the data are actually very close.
2. When comparing the spread between two datasets, the units of each must be the same.
3. When data are related to money or the economy, higher variance (or standard deviation) is associated with higher risk.
4. The standard deviation is used more often in practice than the variance because it shares the units of the original dataset.

Use in the World

The standard deviation is associated with risk in finance, assists in determining the significance of drugs in medical studies, and measures the error of our results for predicting anything from the amount of rainfall we can expect tomorrow to your predicted commute time tomorrow.

These applications are beyond the scope of this lesson as they pertain to specific fields, but know that understanding the spread of a particular set of data is extremely important to many areas. In this lesson, you mastered the calculation of the most common measures of spread.

Final Quiz - Measures Spread

27 December 2021 11:10

Final quiz Measures Spread 1 and 2

For the following dataset, match each value to the appropriate label:

- DataSet: 15, 4, 3, 8, 15, 22, 7, 9, 2, 3, 3, 12, 6

```python
nums = [15, 4, 3, 8, 15, 22, 7, 9, 2, 3, 3, 12, 6]
sorted_nums = sorted(nums)
print(f"Sorted Numbers = {sorted_nums}")
print(f"Number of values = {len(nums)}")
median = statistics.median(sorted_nums)
median_index = sorted_nums.index(median)
Q_1 = statistics.median(sorted_nums[:median_index])
Q_3 = statistics.median(sorted_nums[median_index+1:])
print(f"{'Minimum:': <16} {min(sorted_nums)}")
print(f"{'Q1:': <16} {Q_1}")
print(f"{'Median or Q2:':<16} {median}")
print(f"{'Q3:': <16} {Q_3}")
print(f"{'Maximum:':<16} {max(sorted_nums)}")
print(f"{'Range:':<16} {max(sorted_nums) - min(sorted_nums)}")
print(f"{'IQR:':<16} {Q_3 - Q_1}")
print(f"{'Mean:':<16} {statistics.mean(nums)}")
```

```
Sorted Numbers = [2, 3, 3, 3, 4, 6, 7, 8, 9, 12, 15, 15, 22]
Number of values = 13
Minimum:         2
Q1:              3.0
Median or Q2:    7
Q3:              13.5
Maximum:         22
Range:           20
IQR:             10.5
Mean:            8.384615384615385
```

```
Mean of DataSet:                 8.384615384615385
Mean Subtracted from each value: [6.615384615384615, -4.384615384615385,
Each value Squared:              [43.76331360946745, 19.22485207100592,
Each value Squared and Added:    441.0769230769231
Variance of Data:                33.93
Standard Deviation of Data:      5.82
```

Recap

27 December 2021 11:50

Variable Types

We have covered a lot up to this point! We started with identifying data types as either categorical or quantitative. We then learned we could identify quantitative variables as either continuous or discrete. We also found we could identify categorical variables as either ordinal or nominal.

Categorical Variables

When analyzing categorical variables, we commonly just look at the count or percent of a group that falls into each **level** of a category. For example, if we had two **levels** of a dog category: lab and not lab. We might say, 32% of the dogs were lab (percent), or we might say 32 of the 100 dogs I saw were labs (count).

However, the 4 aspects associated with describing quantitative variables are not used to describe categorical variables.

Quantitative Variables

Then we learned there are four main aspects used to describe quantitative variables:

1. Measures of **Center**
2. Measures of **Spread**
3. **Shape** of the Distribution
4. **Outliers**

We looked at calculating measures of Center

1. **Means**
2. **Medians**
3. **Modes**

We also looked at calculating measures of Spread

1. **Range**
2. **Interquartile Range**
3. **Standard Deviation**
4. **Variance**

Calculating Variance

We saw that we could calculate the variance as:

$$\boldsymbol{\frac{1}{n}\sum\limits_{i=1}^n(x_i - \bar{x})^2}$$

You will also see:

$$\boldsymbol{\frac{1}{n-1}\sum\limits_{i=1}^n(x_i - \bar{x})^2}$$

The reason for this is beyond the scope of what we have covered thus far, but you can find an explanation here.

You can commonly find answers to your questions with a quick Google search. Now is a great time to get started with this practice! This answer should make more sense at the completion of this lesson.

Standard Deviation vs. Variance

The standard deviation is the square root of the variance. In practice, you usually use the standard deviation rather than the variance. The reason for this is because the standard deviation shares the same units with our original data, while the variance has squared units.

What Next?

In the next sections, we will be looking at the last two aspects of quantitative variables: **shape** and **outliers**. What we know about measures of center and measures of spread will assist in your understanding of these final two aspects.

Supporting Materials

- Calculating Variance

Shape

27 December 2021 11:53

Histograms

We learned how to build a **histogram** in this video, as this is the most popular visual for quantitative data.

Shape

From a histogram, we can quickly identify the shape of our data, which helps influence all of the measures we learned in the previous concepts. We learned that the distribution of our data is frequently associated with one of the three **shapes**:

1. Right-skewed

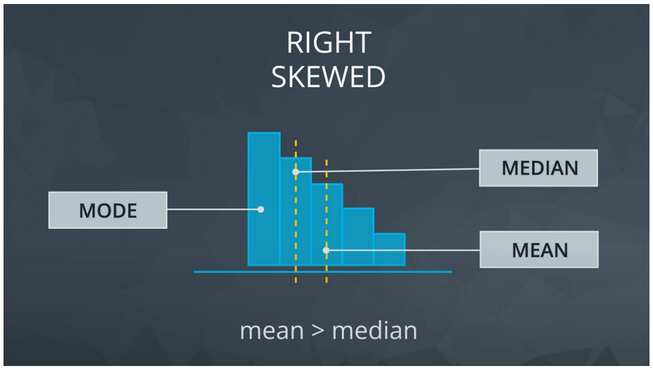
2. Left-skewed

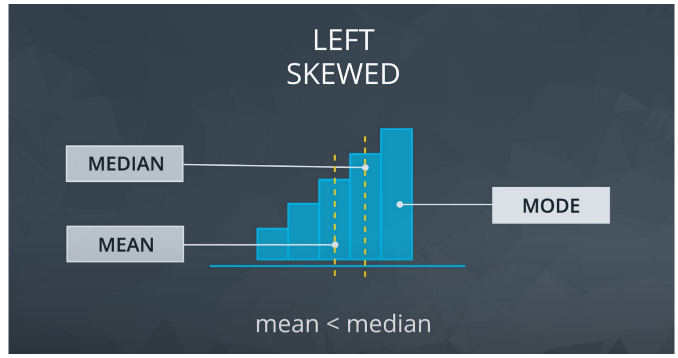
3. Symmetric (frequently normally distributed)

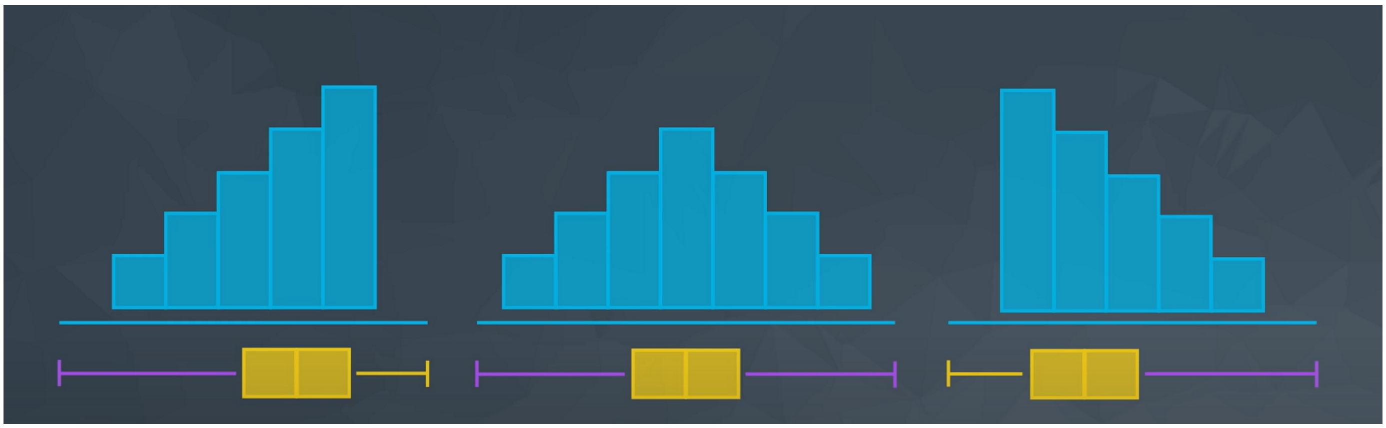
Summary

Shape	**Mean vs. Median**	**Real-World Applications**
Symmetric (Normal)	Mean equals Median	Height, Weight, Errors, Precipitation
Right-skewed	Mean greater than Median	Amount of drug remaining in a bloodstream, Time between phone calls at a call center, Time until light bulb dies
Left-skewed	Mean less than Median	Grades as a percentage in many universities, Age of death, Asset price changes

The mode of a distribution is essentially the tallest bar in a histogram. There may be multiple modes depending on the number of peaks in our histogram.


RIGHT
SKEWED
MODE
MEDIAN
MEAN
mean > median


LEFT
SKEWED
MEDIAN
MEAN
MODE
mean < median



Shape of Data in real world

27 December 2021 11:56

When working with data, building a **quick plot** lets you quickly see the shape of your data.

Distribution Shape	**Types of Data**
Bell Shaped	Heights, Weight, Scores
Left Skewed	GPA, Age of Death, Price
Right Skewed	Distribution of Wealth, Athletic Abilities

References

These are the references used to pull the applications of each shape.

- Quora
- University of Texas
- Stack Exchange

Supporting Materials

- Quora
- Stack Exchange

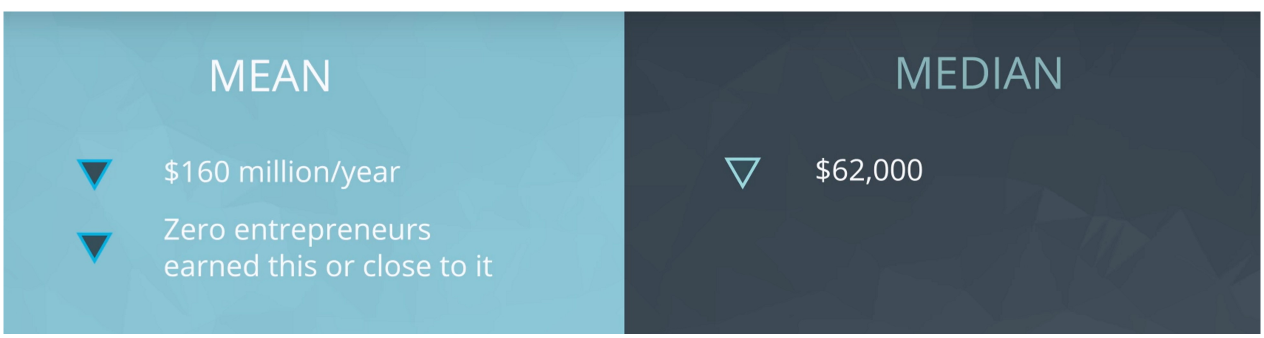
Outliers

27 December 2021 12:15

We learned that **outliers** are points that fall very far from the rest of our data points. This influences measures like the mean and standard deviation much more than measures associated with the five-number summary.

Identifying Outliers

There are a number of different techniques for identifying outliers. A full paper on this topic is provided here. In general, I usually just look at a picture and see if something looks suspicious!



Working with Outliers

27 December 2021 12:19

Common Techniques

When outliers are present we should consider the following points.

1. Noting they exist and the impact on summary statistics.

2. If typo - remove or fix

3. Understanding why they exist, and the impact on questions we are trying to answer about our data.

4. Reporting the 5 number summary values is often a better indication than measures like the mean and standard deviation when we have outliers.

5. Be careful in reporting. Know how to ask the right questions.

Working with outliers (expert advice)

27 December 2021 12:21

Outliers Advice

Below are my guidelines for working with any column (random variable) in your dataset.

1. Plot your data to identify if you have outliers.

2. Handle outliers accordingly via the previous methods.

3. If no outliers and your data follow a normal distribution - use the mean and standard deviation to describe your dataset, and report that the data are normally distributed.

Side note

If you aren't sure if your data are normally distributed, there are plots called normal quantile plots and statistical methods like the Kolmogorov-Smirnov test that are aimed to help you understand whether or not your data are normally distributed. Implementing this test is beyond the scope of this class, but can be used as a fun fact.

4. If you have skewed data or outliers, use the five-number summary to summarize your data and report the outliers.

Supporting Materials

- Kolmogorov-Smirnov test
- Normal quantile plots

Descriptive Stats Summary

27 December 2021 14:49

Variable Types

We have covered a lot up to this point! We started with identifying data types as either categorical or quantitative. We then learned we could identify quantitative variables as either continuous or discrete. We also found we could identify categorical variables as either ordinal or nominal.

Categorical Variables

When analyzing categorical variables, we commonly just look at the count or percent of a group that falls into each **level** of a category. For example, if we had two **levels** of a dog category: lab and not lab. We might say, 32% of the dogs were lab (percent), or we might say 32 of the 100 dogs I saw were labs (count).

However, the 4 aspects associated with describing quantitative variables are not used to describe categorical variables.

Quantitative Variables

Then we learned there are four main aspects used to describe quantitative variables:

1. Measures of **Center**
2. Measures of **Spread**
3. **Shape** of the Distribution
4. **Outliers**

Measures of Center

We looked at calculating measures of Center

1. **Means**
2. **Medians**
3. **Modes**

Measures of Spread

We also looked at calculating measures of Spread

1. **Range**
2. **Interquartile Range**
3. **Standard Deviation**
4. **Variance**

Shape

We learned that the distribution of our data is frequently associated with one of the three **shapes**:

1. Right-skewed

2. Left-skewed

3. Symmetric (frequently normally distributed)

Depending on the shape associated with our dataset, certain measures of center or spread may be better for summarizing our dataset.

When we have data that follows a **normal** distribution, we can completely understand our dataset using the mean and standard deviation.

However, if our dataset is **skewed**, the 5 number summary (and measures of center associated with it) might be better to summarize our dataset.

Outliers

We learned that outliers have a larger influence on measures like the mean than on measures like the median. We learned that we should work with outliers on a situation by situation basis. Common techniques include:

1. At least note they exist and the impact on summary statistics.

2. If typo - remove or fix

3. Understand why they exist, and the impact on questions we are trying to answer about our data.

4. Reporting the 5 number summary values is often a better indication than measures like the mean and standard deviation when we have outliers.

5. Be careful in reporting. Know how to ask the right questions.

Histograms and Box Plots

We also looked at histograms and box plots to visualize our quantitative data. Identifying outliers and the shape associated with the distribution of our data are easier when using a visual as opposed to using summary statistics.

What Next?

Up to this point, we have only looked at **Descriptive Statistics**, because we are describing our collected data. In the final sections of this lesson, we will be looking at the difference between **Descriptive Statistics** and **Inferential Statistics**.

Descriptive vs Inferential

27 December 2021 14:52

In this section, we learned about how **Inferential Statistics** differs from **Descriptive Statistics**.

Descriptive Statistics

Descriptive statistics**is about describing our collected data**.

Inferential Statistics

Inferential Statistics**is about using our collected data to draw conclusions about a larger population**.

We looked at specific examples that allowed us to identify the

1. **Population -** our entire group of interest.
2. **Parameter -** numeric summary about a population
3. **Sample -** a subset of the population
4. **Statistic** numeric summary about a sample

Descriptive vs. Inferential Statistics

In this section, we learned about how **Inferential Statistics** differs from **Descriptive Statistics**.

Descriptive Statistics

Descriptive statistics **is about describing our collected data** using the measures discussed throughout this lesson: measures of center, measures of spread, the shape of our distribution, and outliers. We can also use plots of our data to gain a better understanding.

Inferential Statistics

Inferential Statistics **is about using our collected data to draw conclusions to a larger population**. Performing inferential statistics well requires that we take a sample that accurately represents our population of interest.

A common way to collect data is via a survey. However, surveys may be extremely biased depending on the types of questions that are asked, and the way the questions are asked. This is a topic you should think about when tackling the first project.

We looked at specific examples that allowed us to identify the

5. **Population -** our entire group of interest.

6. **Parameter -** numeric summary about a population
7. **Sample -** a subset of the population
8. **Statistic -** numeric summary about a sample

Looking Ahead

Though we will not be diving deep into inferential statistics within this course, you are now aware of the difference between these two branches of statistics. If you have ever conducted a hypothesis test or built a confidence interval, you have performed inferential statistics. The way we perform inferential statistics is changing as technology evolves. Many career paths involving **Machine Learning** and **Artificial Intelligence** are aimed at using collected data to draw conclusions about entire populations at an individual level. It is an exciting time to be a part of this space, and you are now well on your way to joining the other practitioners!

Lesson Review

Congratulations on completing this lesson on descriptive statistics. You learned some foundational metrics for understanding data, including how to :

- Evaluate measures of spread
- **Range**
- **Interquartile Range (IQR)**
- **Standard Deviation**
- **Variance**
- Analyze outliers
- Evaluate descriptive and inferential statistics

Looking Forward

In the next lessons, you'll switch over to the world of spreadsheets - powerful tools for performing analysis over data, including automated formulas for many of the statistics concepts you have seen so far!