Exploratory Data Analysis

Section 3-4

Jacob Ayers

MAT 110

Lesson #8

Objectives

- Find the five-number-summary of a data set
- Construct boxplots

Exploratory Data Analysis

Purpose of analysis: confirm various conjectures about data

Exploratory Data Analysis

Purpose of analysis: confirm various conjectures about data

In exploratory data analysis (EDA), we organize data using stem-and-leaf plots.

- Measure of central tendency: median
- Measure of variation: IQR
- Graphical representation: Boxplot (aka box-and-whisker plot)

Exploratory Data Analysis

Purpose of analysis: confirm various conjectures about data

In exploratory data analysis (EDA), we organize data using stem-and-leaf plots.

- Measure of central tendency: median
- Measure of variation: IQR
- Graphical representation: Boxplot (aka box-and-whisker plot)

Purpose of EDA: find out what information can be discovered about the data (e.g. center and spread)

The Five-Number Summary

The *five-number summary* of a data set includes the following values:

- The lowest value (minimum)

The Five-Number Summary

The *five-number summary* of a data set includes the following values:

- The lowest value (minimum)
- Q_1

The Five-Number Summary

The *five-number summary* of a data set includes the following values:

- The lowest value (minimum)
- Q_1
- The median

The Five-Number Summary

The *five-number summary* of a data set includes the following values:

- The lowest value (minimum)
- Q_1
- The median
- Q_3

The Five-Number Summary

The *five-number summary* of a data set includes the following values:

- The lowest value (minimum)
- Q_1
- The median
- Q_3
- The highest value (maximum)

The Five-Number Summary

The *five-number summary* of a data set includes the following values:

- The lowest value (minimum)
- Q_1
- The median
- Q_3
- The highest value (maximum)

I'll be demonstrating finding the five-number summary using a calculator, but it can also be done by hand using methods we've seen in previous videos.

The Five-Number Summary

Identify the five-number summary and find the interquartile range for the data.

$362, 589, 437, 316, 192, 188$

The Five-Number Summary

Identify the five-number summary and find the interquartile range for the data.

$362, 589, 437, 316, 192, 188$

Using a graphing calculator:

Min $= 188$

$Q_1 = 192$

Median $= 339$

$Q_3 = 437$

Max $= 589$

IQR $= 437 - 192 = 245$

Boxplots

Once we have the five-number summary, we can draw a boxplot.

Boxplots

Once we have the five-number summary, we can draw a boxplot.

A **boxplot** is a graph of a data set obtained by drawing a horizontal line from the minimum data value to Q_1, drawing a horizontal line from Q_3 to the maximum data value, and drawing a box whose vertical sides pass through Q_1 and Q_3 with a vertical line inside the box passing through the median or Q_2.

Boxplots

Once we have the five-number summary, we can draw a boxplot.

A **boxplot** is a graph of a data set obtained by drawing a horizontal line from the minimum data value to Q_1, drawing a horizontal line from Q_3 to the maximum data value, and drawing a box whose vertical sides pass through Q_1 and Q_3 with a vertical line inside the box passing through the median or Q_2.

We don't have a great method of constructing a boxplot on a calculator or in Sheets, so we'll have to draw them by hand.

Boxplots

The number of meteorites found in 10 states of the United States is 89, 47, 164, 296, 30, 215, 138, 78, 48, 39. Construct a boxplot for the data.

Boxplots

The number of meteorites found in 10 states of the United States is 89, 47, 164, 296, 30, 215, 138, 78, 48, 39. Construct a boxplot for the data.

First, find the five-number summary (30, 47, 83.5, 164, 296).

Boxplots

The number of meteorites found in 10 states of the United States is 89, 47, 164, 296, 30, 215, 138, 78, 48, 39. Construct a boxplot for the data.

First, find the five-number summary (30, 47, 83.5, 164, 296).

Now, we can draw the boxplot.

Boxplots

Construct a boxplot for the data below.

79, 82, 77, 84, 80, 89, 60, 79, 91, 93, 88

First, find the find-number summary (60, 79, 82, 89, 93).

Boxplots

Construct a boxplot for the data below.

79, 82, 77, 84, 80, 89, 60, 79, 91, 93, 88

First, find the find-number summary (60, 79, 82, 89, 93).

Now, we can draw the boxplot.

Next Steps

- Prepare for Midterm 1
- Take Midterm 1
- Begin Module #6
  - Read 4-1
  - Watch Video Lesson #9

Thanks for watching!