

Measuring Variability

Sections 2.4, 2.5, 2.6

Lecture 6

Robb T. Koether

Hampden-Sydney College

Thu, Jan 21, 2016

Outline

- 1 Variation
- 2 The Five-Number Summary
- 3 Displaying Variability – Boxplots
- 4 Outliers and Modified Boxplots
- 5 Assignment

Outline

- 1 Variation
- 2 The Five-Number Summary
- 3 Displaying Variability – Boxplots
- 4 Outliers and Modified Boxplots
- 5 Assignment

Variation

- In addition to locating the “center” of a distribution, we are also interesting knowing how spread out the data are, or equivalently, how variable the data are.
- In reference to the coming snow storm, one report said that snow-fall amounts would vary considerably over small area.

Percentiles and Quartiles

Definition (p^{th} Percentile)

The p^{th} percentile of a set of numbers is a number that divides the lower $p\%$ of the numbers from the rest.

Definition (1st Quartile)

The 1st quartile, denoted Q_1 , of a set of numbers is the 25th percentile.

Definition (3rd Quartile)

The 3rd quartile, denoted Q_3 , of a set of numbers is the 75th percentile.

Percentiles and Quartiles

Definition (p^{th} Percentile)

The p^{th} percentile of a set of numbers is a number that divides the lower $p\%$ of the numbers from the rest.

Definition (1st Quartile)

The 1st quartile, denoted Q_1 , of a set of numbers is the 25th percentile.

Definition (3rd Quartile)

The 3rd quartile, denoted Q_3 , of a set of numbers is the 75th percentile.

- The 2nd quartile is the median (50th percentile).

Net Worth Percentiles

- Try out the Net Worth Percentile Calculator: [click here](#).

Finding Quartiles

- First, the data must be arranged in order, smallest to largest.
- To find the quartiles,
 - Find the position of the median.
 - The 1st quartile is the median of all the numbers that are *below* that position.
 - The 3rd quartile is the median of all the numbers that are *above* that position.

Example

Example (Median Rainfall)

- Rainfall data for August in Richmond, VA (1986 - 2015).

6.74	1.24	4.04	4.90	5.72	2.88
6.91	5.58	2.52	8.42	4.44	1.41
1.84	2.00	2.79	2.30	3.15	3.59
16.02	2.56	5.99	6.81	5.73	4.04
3.92	7.10	3.50	7.64	3.61	2.77

- First, arrange the data in order.
- Then find the median and the quartiles.

Example

Example (Median Rainfall)

- Rainfall data for August in Richmond, VA (1986 - 2015).

1.24	1.41	1.84	2.00	2.30	2.52
2.56	2.77	2.79	2.88	3.15	3.50
3.59	3.61	3.92	4.04	4.04	4.44
4.90	5.58	5.72	5.73	5.99	6.74
6.81	6.91	7.10	7.64	8.42	16.02

- First, arrange the data in order.
- Then find the median and the quartiles.

Example

Example (Median Rainfall)

- Rainfall data for August in Richmond, VA (1986 - 2015).

1.24	1.41	1.84	2.00	2.30	2.52
2.56	2.77	2.79	2.88	3.15	3.50
3.59	3.61	3.92	4.04	4.04	4.44
4.90	5.58	5.72	5.73	5.99	6.74
6.81	6.91	7.10	7.64	8.42	16.02

- First, arrange the data in order.
- Then find the median and the quartiles.

Example (Gasoline Prices)

- Gasoline prices for 13 Sheetz stores in Virginia.

1.53	1.69	1.69	1.62	1.65
1.61	1.50	1.69	1.65	1.55
1.67	1.65	1.59		

- Find the median and the quartiles.

Outline

- 1 Variation
- 2 The Five-Number Summary**
- 3 Displaying Variability – Boxplots
- 4 Outliers and Modified Boxplots
- 5 Assignment

Definition (Five-Number Summary)

The **five-number summary** of a set of numbers consists of the five quantities

- Minimum
- Q_1
- Median
- Q_3
- Maximum

Example (Gasoline Prices)

- Gasoline prices for 13 Sheetz stores in Virginia.

1.53	1.69	1.69	1.62	1.65
1.61	1.50	1.69	1.65	1.55
1.67	1.65	1.59		

- Give a five-number summary of these data.

Outline

- 1 Variation
- 2 The Five-Number Summary
- 3 Displaying Variability – Boxplots**
- 4 Outliers and Modified Boxplots
- 5 Assignment

Definition (Boxplot)

A **boxplot** is a graphical display of a five-number summary.

- Draw a horizontal (or vertical) scale that covers the range of data.
- Draw a rectangle over the scale that extends from Q_1 to Q_3 .
- Indicate the **median** by drawing a vertical line through the rectangle.
- Indicate the **maximum** and **minimum** by drawing “tails” or “whiskers” attached to the ends of the rectangle that extend to the minimum and the maximum.

Example

Example (Comparing Rainfall Data)

- Five-number summary of the rainfall data for August:

1.24 2.79 3.98 5.73 16.02

- Five-number summary of the rainfall data for April:

0.80 2.33 2.76 4.12 11.12

- Draw boxplots for both data sets and compare.

Outline

- 1 Variation
- 2 The Five-Number Summary
- 3 Displaying Variability – Boxplots
- 4 Outliers and Modified Boxplots**
- 5 Assignment

- Using percentiles, we can define what we mean by an outlier.
- This definition was developed by the statistician John Tukey in the late 1960s.
- We begin with the “interquartile range.”

The Interquartile Range

Definition (Interquartile range)

The **interquartile range**, denoted **IQR**, is

$$IQR = Q_3 - Q_1.$$

- The IQR is the width of the rectangle in the boxplot.

Interquartile Range

Example (Interquartile Range)

- Use the five-number summaries to find the *IQRs* of the August and April rainfall data.
- How do they compare?
- What does that tell us?

Modified Boxplots

Definition (Suspected Outlier)

A **suspected outlier** is any value that is more than $1.5 \times IQR$ below Q_1 or more than $1.5 \times IQR$ above Q_3 .

Definition (Outlier)

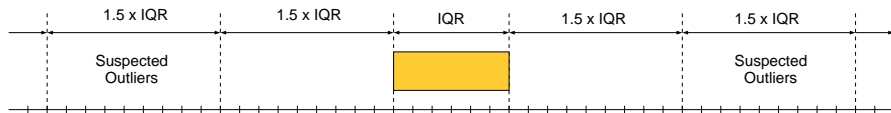
An **outlier** is any value that is more than $3 \times IQR$ below Q_1 or more than $3 \times IQR$ above Q_3 .

Definition (Modified boxplot)

A **modified boxplot** is a boxplot in which the outliers (and extreme outliers) are indicated by separate dots.

Modified Boxplots

- To draw a modified boxplot,
 - Draw the box part of the boxplot as usual.
 - Compute $1.5 \times IQR$.
 - The boundaries for suspected outliers are $Q_1 - 1.5 \times IQR$ and $Q_3 + 1.5 \times IQR$.
 - Extend the tails to the *last value that is not a suspected outlier*.
 - Draw a dot for each suspected outlier.



Example

Example (Rainfall Outliers)

- Do the August rainfall data contain any suspected outliers?
- Do the April rainfall data contain any suspected outliers?
- Do the gas price rainfall data contain any suspected outliers?

Outline

- 1 Variation
- 2 The Five-Number Summary
- 3 Displaying Variability – Boxplots
- 4 Outliers and Modified Boxplots
- 5 Assignment**

Assignment

Assignment

- Read Sections 2.4, 2.5, 2.6.
- Apply Your Knowledge: 2.6.
- Check Your Skills: 2.19.
- Exercises: 29, 30, 31, 32.