General Education Statistics

Class Notes

The Five-Number Summary and Boxplots (Section 3.5)
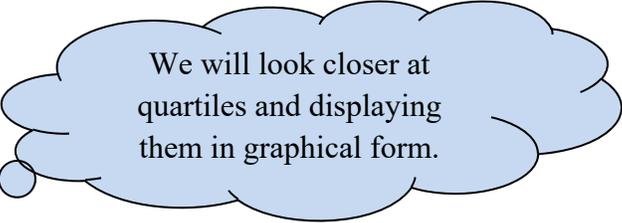
We saw quartiles in the previous section. Five-number summaries use those to define a useful way to explore the data. A boxplot is a graphic that shows these numbers off.

Definition: The **five-number summary** of a set of data consists of the smallest data value (or minimum), Q_1, the median, Q_3, and the largest data value (or maximum). We organize the five-number summary in increasing order.

Minimum Q_1 Median Q_3 Maximum
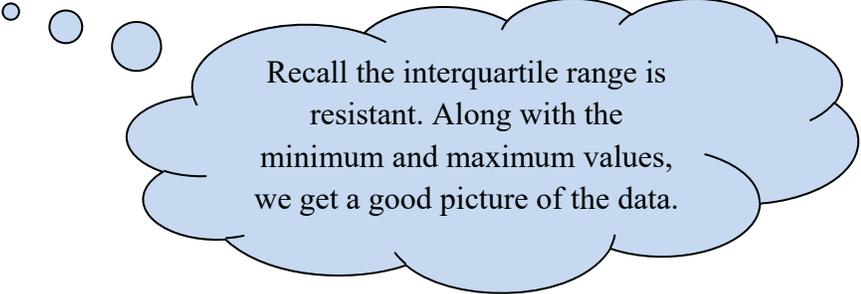
Worksheet: Effect of an Outlier:

This worksheet will give us practice in finding means, medians, standard deviations, quartiles, and five-number summaries. We will see how an outlier affects some of these measures more than others. Recall, this is the idea of resistance.

expl 1: A company pays its employees the following salaries. (Notice, I wrote the data in increasing order.) Find the five-number summary of this data.

$25,000
$26,000
$26,000
$26,000
$27,000
$28,000
$30,000
$30,000
$35,000
$36,000
$45,000
$100,000
$200,000

expl 2: Every six months, the United States Federal Reserve Board conducts a survey of credit card plans in the U.S. The following data are the interest rates charged by ten credit card issuers randomly selected for the July 2005 survey. Determine the five-number summary of the data.

Institution	Rate
Pulaski Bank and Trust Company	6.5%
Bank of Louisiana	9.9%
Rainier Pacific Savings Bank	12.0%
Infibank	13.0%
United Bank, Inc.	13.3%
First National Bank of The Mid-Cities	13.9%
Lafayette Ambassador Bank	14.3%
Wells Fargo Bank NA	14.4%
Firstbank of Colorado	14.4%
Bar Harbor Bank and Trust Company	14.5%

Definition: Boxplot: A **boxplot** (or **box-and-whiskers plot**) is a graphic that shows the five-number summary, with any outliers clearly marked. You will see a scale along the bottom to give meaning to the numbers.

The steps of creating a boxplot are below.

Step 1 Determine the lower and upper fences (where $IQR = Q_3 - Q_1$):

Lower Fence $= Q_1 - 1.5(IQR)$

Upper Fence $= Q_3 + 1.5(IQR)$

Step 2 Draw a number line long enough to include the maximum and minimum values. Above the number line, draw vertical lines at Q_1, M, and Q_3. Use these vertical lines to draw a rectangular box.

Step 3 Temporarily label the lower and upper fences.

Step 4 Draw a line from Q_1 to the smallest data value that is larger than the lower fence. Draw a line from Q_3 to the largest data value that is smaller than the upper fence. (Basically, do not draw all the way to outliers. They are dealt with next.) These lines are called **whiskers**.
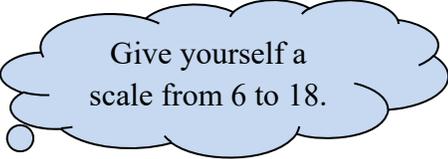
Step 5 Any data values less than the lower fence or greater than the upper fence are outliers and are marked with an asterisk (*). Erase the upper and lower fences (from step 3).

2

expl 3: Draw a boxplot for the interest rates data in the last example. Follow the steps below.

a.) (Step 1) Determine the lower and upper fences (where $IQR = Q_3 - Q_1$):

$$\text{Lower Fence} = Q_1 - 1.5(IQR)$$

$$\text{Upper Fence} = Q_3 + 1.5(IQR)$$

> Give yourself a scale from 6 to 18.

b.) (Step 2) Draw a number line long enough to include the maximum and minimum values, leaving enough space above it to draw the plot. Above the number line, draw vertical lines at Q_1, M, and Q_3. Use these vertical lines to draw a rectangular box.

(Step 3) Temporarily label the lower and upper fences.

(Step 4) Draw a line from Q_1 to the smallest data value that is larger than the lower fence. Draw a line from Q_3 to the largest data value that is smaller than the upper fence. (Basically, do not draw all the way to outliers. They are dealt with next.) These lines are called **whiskers**.

(Step 5) Any data values less than the lower fence or greater than the upper fence are outliers and are marked with an asterisk (*). You can now erase the upper and lower fences.

You should also label the horizontal axis and title your plot.

c.) Do you think this distribution is symmetric, skewed left, or skewed right? Explain.

3

expl 4: Here we see the boxplot as generated by StatCrunch for the interest rate data.

Let's suppose another sample was taken of ten banks and whose interest rates are summarized in the second boxplot. The medians are marked with a red, vertical line.

Let's explore the comparisons we can make. Notice the scales are roughly lined up.

a.) How would you compare the spread of the two data sets as measured by the interquartile range?

Random Sample of Interest Rates
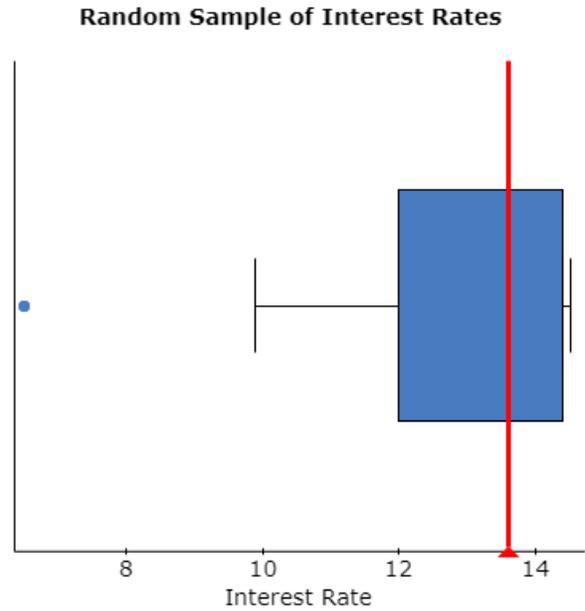
Interest Rate

b.) How would you compare the centers of the two data sets?

Random Sample of Interest Rates
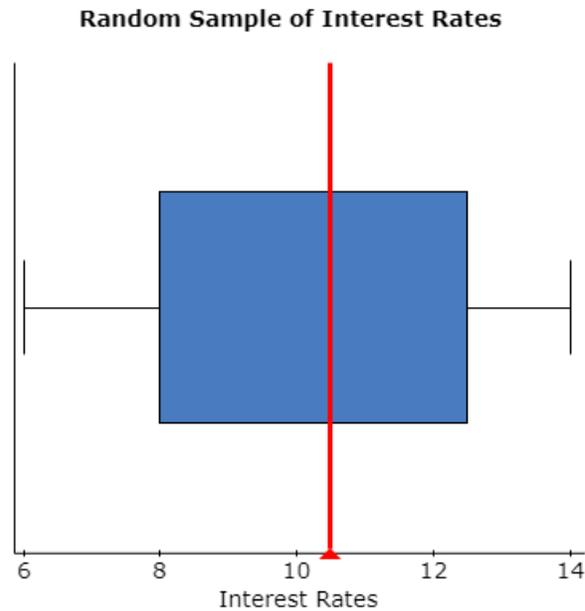
Interest Rates

c.) How would you compare the overall ranges (maximum minus minimum values) of the two data sets?

d.) Overall, which data set would you say has higher interest rates?
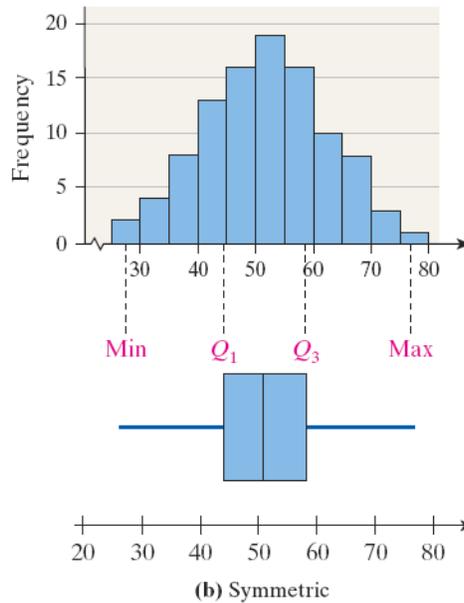
Distributions and the Boxplot:
A boxplot shows the distribution of a data set nicely. You can see similarities between a data set's histogram and its boxplot. Consider the pictures below.

Here, we see a symmetric distribution. The boxplot is nice and centered.
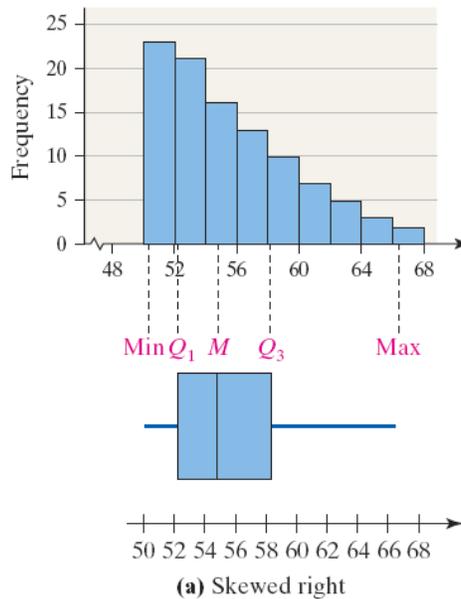
The median is centered in the box.

The whiskers are of equal lengths.

(b) Symmetric

Here, we see a distribution that is skewed right.

Notice how the median is off-center, just left of the box's center.

The right whisker mimics the tail we see on the histogram. It is longer than the left whisker.

(a) Skewed right

Here, we see the opposite of the last picture. We see a distribution that is skewed left.

Notice how the median is off-center, just right of the box's center.

The left whisker mimics the tail we see on the histogram. It is longer than the right whisker.

(c) Skewed left
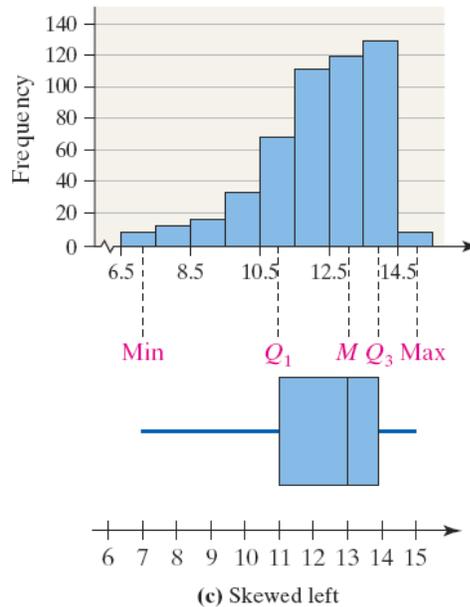
Making a Boxplot with Technology:
In Stat Crunch in MyStatLab (left-hand menu in MSL), follow these steps.

1. Enter the raw data if needed. Name the column variable.

2. Select **Graph** and highlight **Boxplot**.

3. Click on the variable whose boxplot you want to draw under "Select Column(s)". Check the boxes "Use fences to identify outliers" and "Draw boxes horizontally". Enter a label for the *x*-axis. (You'll have to scroll down.) Enter a title and click **Compute!**

Making boxplots with other technology such as your calculator is described in the book.