

Lesson 8: Five Number Summary, Box-plot, Outliers

Daily Data Collection

Each student will record the following information on the board:

Head Circumference & Hand span (thumb to pinky spread out)

Distribution Part 3: Spread [you learned Range earlier]

Using a mean or a median by itself can be irresponsible. A measure of center alone can sometimes be misleading. We need to know how spread out the data is, or the variability of the observations. For that reason, when we describe a distribution, we must also address its spread.

- Q_1 is the first quartile. It is the median of the bottom half of the observations.
25% of observations are less than this number and 75% are above this number.
- Q_2 is the median.
50% of observations are less than this number and 50% are above this number.
- Q_3 is the third quartile. It is the median of the top half of the observations.
75% of observations are less than this number and 25% are above this number.

IQR The Interquartile Range (IQR) is the range of the middle 50% of the data: $Q_3 - Q_1$

Finding the quartiles and IQR by hand:

1. Find the median. This splits the data into a lower half and an upper half. The median is the second Quartile.
2. Perform the steps for finding a median on the lower half of the data. This is the First Quartile.
3. Perform the steps for finding a median on the upper half of the data. This is the Third Quartile.
4. To find the IQR, subtract $Q_3 - Q_1$

Five Number summary describes: **Min** **Q1** **Median** **Q3** **Max**

Class Data:

Find the five number summary and the IQR for head circumference and hand span.

Distribution Part 4: Outliers

Outliers are observations that are so far above or below the majority of the data that they do not fit. Here is the process to find the cutoff values.

1. find the IQR
2. multiple the IQR by 1.5
3. Subtract the new number from Q_1 Any data value below this is considered an outlier (too low)
4. Add the value from step 2 to Q_3 Any data value above this is considered an outlier (too high)

Cutoff Value Formulas: $Q_1 - 1.5(IQR)$ $Q_3 + 1.5(IQR)$

Class Data:

Find the cutoff values for head circumference and hand span. Then identify any outliers in the data.

BoxPlots:

A boxplot is very useful because the graph shows the center, shape and spread of a particular data set. To create a boxplot, you need the five number summary.

Boxplot Info:

- Create a horizontal axis that evenly counts from the min to the max.
- Above this line, place 3 vertical lines to show the location of Q_1, Q_2, Q_3
- Connect the tops and bottoms of the vertical lines to make a box.
- This box shows the size of the middle 50% (IQR).
- Draw a horizontal line from the left side of the box to the location of the min.
- Draw a horizontal line from the right side of the box to the location of the max.
- The graph shows the range from the dot at the min to the dot at the max.
- The boxplot can be used to see if data is symmetric (approximately normal) or if it is skewed.

Class Data:

Create a box-plot for head circumference and hand span. Then describe if the distribution is symmetric or skewed.

Guided Practice:

Home Run King

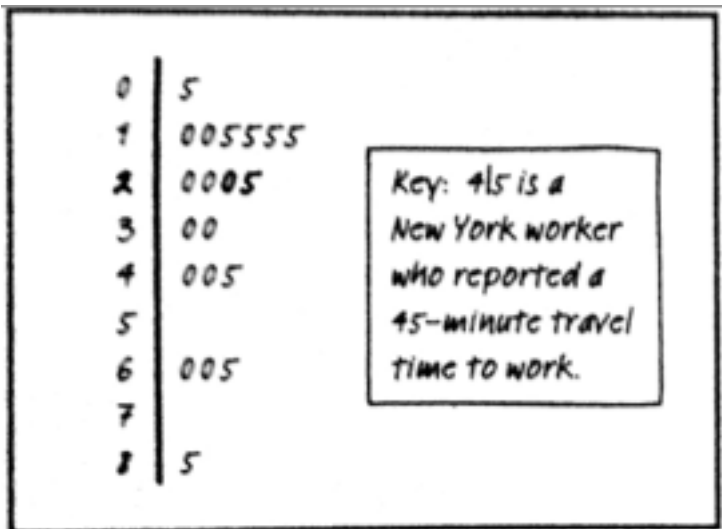
Making a boxplot

Barry Bonds set the major league record by hitting 73 home runs in a single season in 2001. On August 7, 2007, Bonds hit his 756th career home run, which broke Hank Aaron's longstanding record of 755. By the end of the 2007 season when Bonds retired, he had increased the total to 762. Here are data on the number of home runs that Bonds hit in each of his 21 complete seasons:

16 25 24 19 33 25 34 46 37 33 42 40 37 34 49 73 46 45 45 26 28

PROBLEM: We want to make a boxplot for these data.

Identify any outliers:



Describe the distribution of the boxplot above.