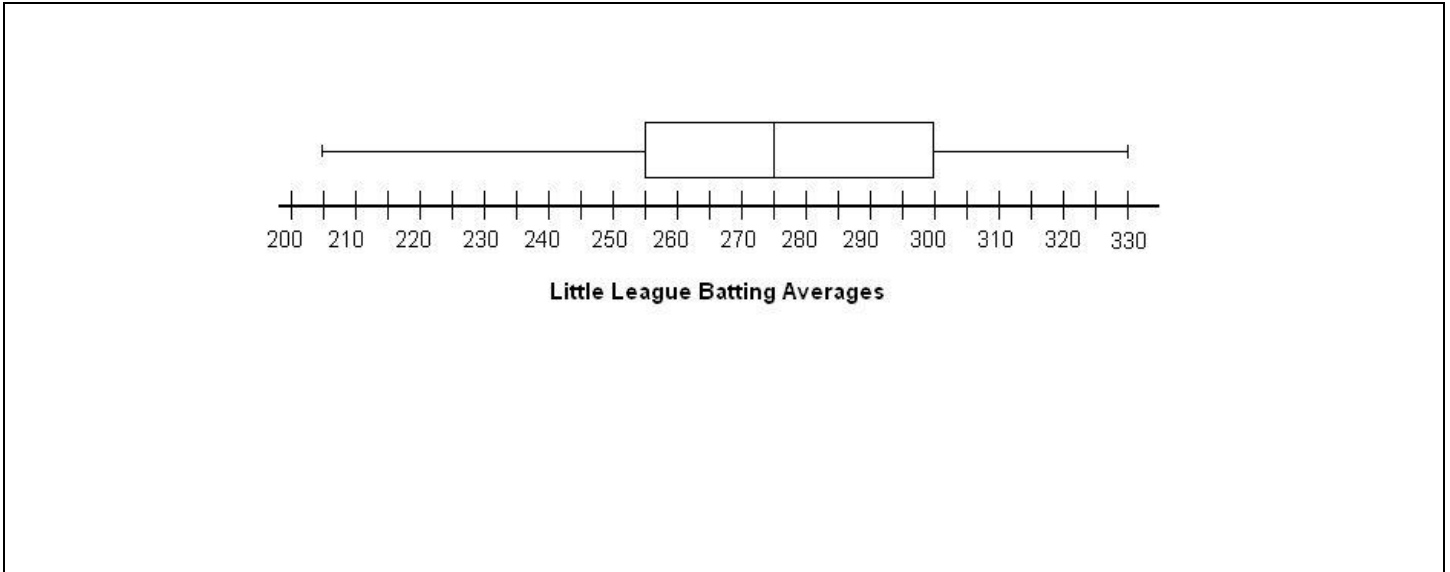


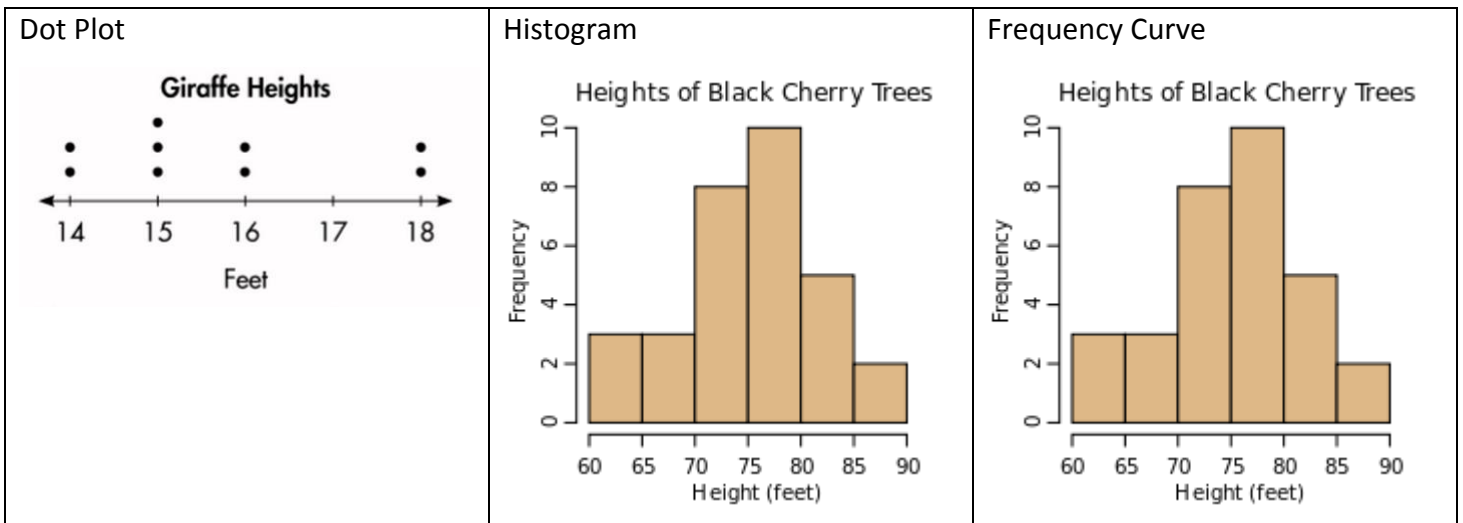
Unit 0 - Statistics

“Univariate” – Only one variable is measured

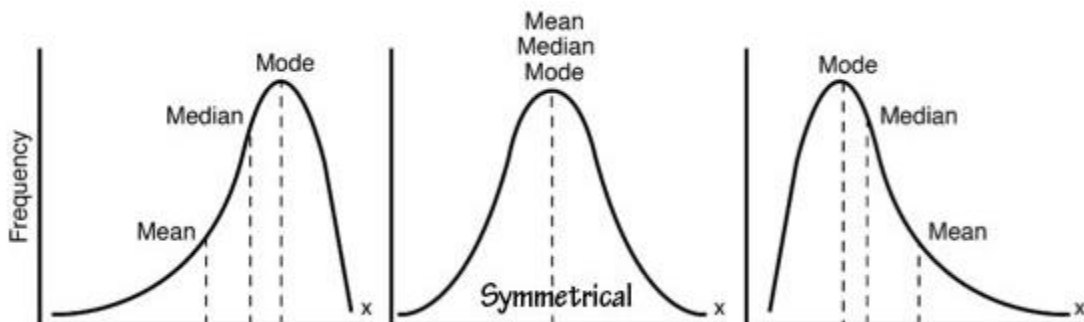
A. Box Plot



B. Frequency Plots:



C. Skew:

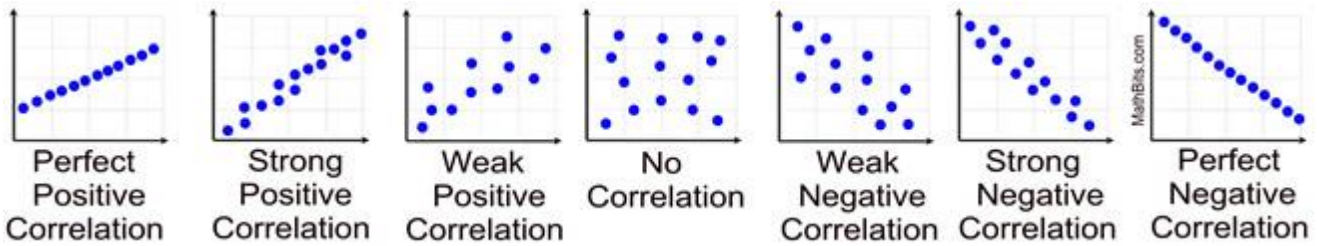


“Bivariate” – Relationship between two variables, x and y.

A. Correlation Coefficient

Tells you:

Example: Estimate the correlation coefficient for the following.



B. Residuals

Tells you:

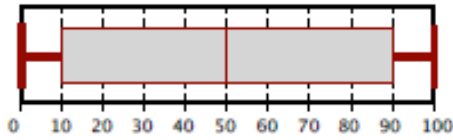
Residual plot for good fit:

Residual plots for bad fit:

C. The linear regression equation

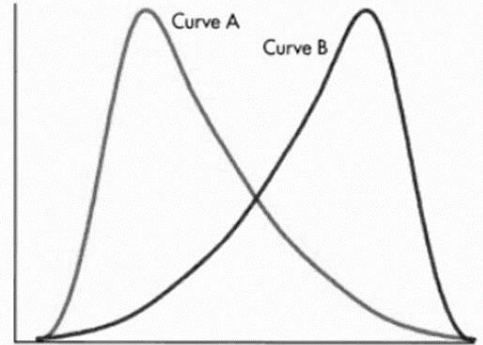
Unit 0 Practice:

1. The box plot below shows test scores for a class of 40 ninth graders.



- A) How many students got an "A"?
- B) How many students scored between 10 and 90 points?
- C) How many students scored below 50 points?
- D) How many students scored between 50 and 90 points?
- E) Is the data skewed right, left, or symmetric?

2. Answer the following:



- A) Which curve has the higher median?
- B) Which curve has the higher mode?

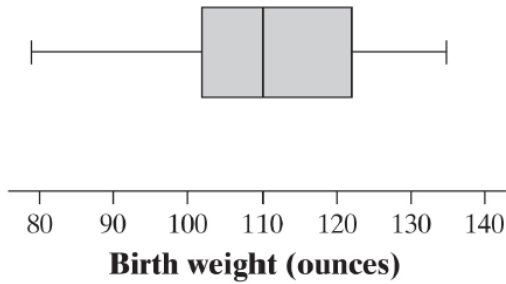
3. Estimate the correlation coefficient of the scatter plot



4. Denise collects data on the number of siblings and hours of sleep per night from a group of 50 people at a local grocery store. She computes the correlation coefficient for the data to be -0.81 . Which of the following statements is true and why?

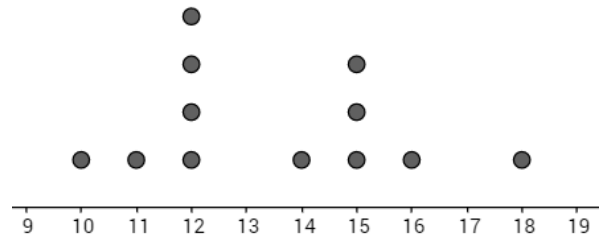
- A) People with more siblings got less sleep.
- B) People who had fewer siblings got less sleep.
- C) On a scatterplot, a line that models the data would have a negative slope.
- D) On a scatterplot, a line that models the data would have a positive slope.

5. The boxplot below shows the birth weights (in ounces) of a sample of 160 infants born in a local hospital.



Approximately how many children had birth weights between 102 and 122 ounces?

6. Find the mean, median, mode, and range of the data:



7. A survey was conducted where students were asked how many youtube videos they had watched that week. The frequency curve was skewed to the right. What does this tell you about the population?

8. Which of the following residual plots indicates a model/regression that is most appropriate for the data, and why?

