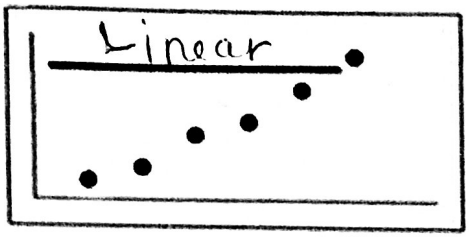
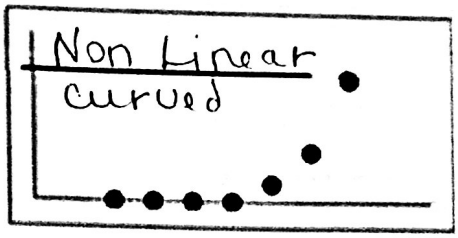
Statistical Reasoning
Scatterplots and Correlation

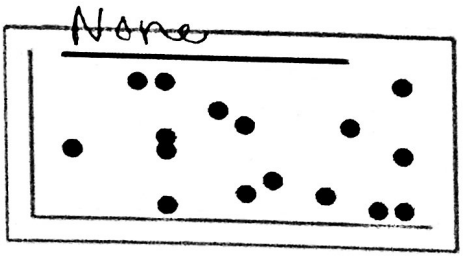
Name: _____ Date: _____ Class: _____

CHARACHTERISTICS OF SCATTERPLOTS

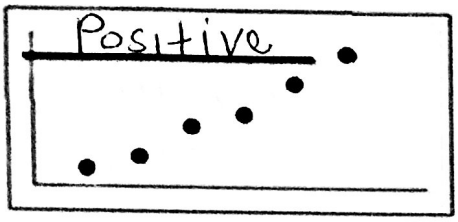
SHAPE

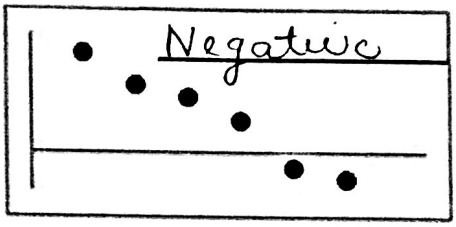


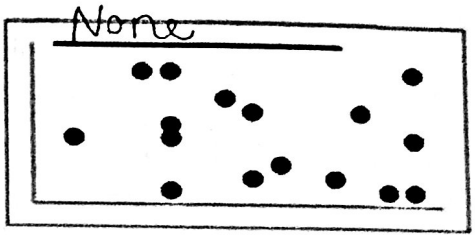




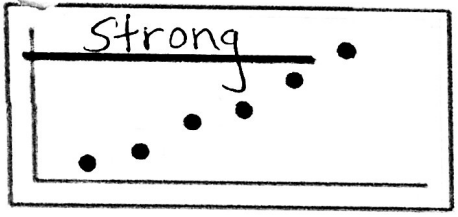
DIRECTION

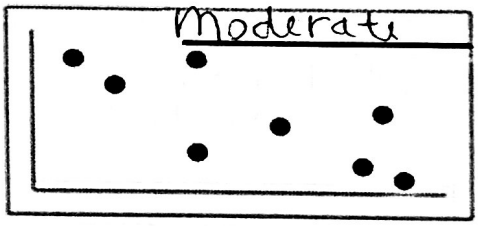


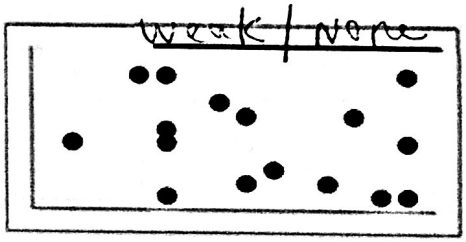




STRENGTH







Two Quantitative Variables: Correlation (STRENGTH)

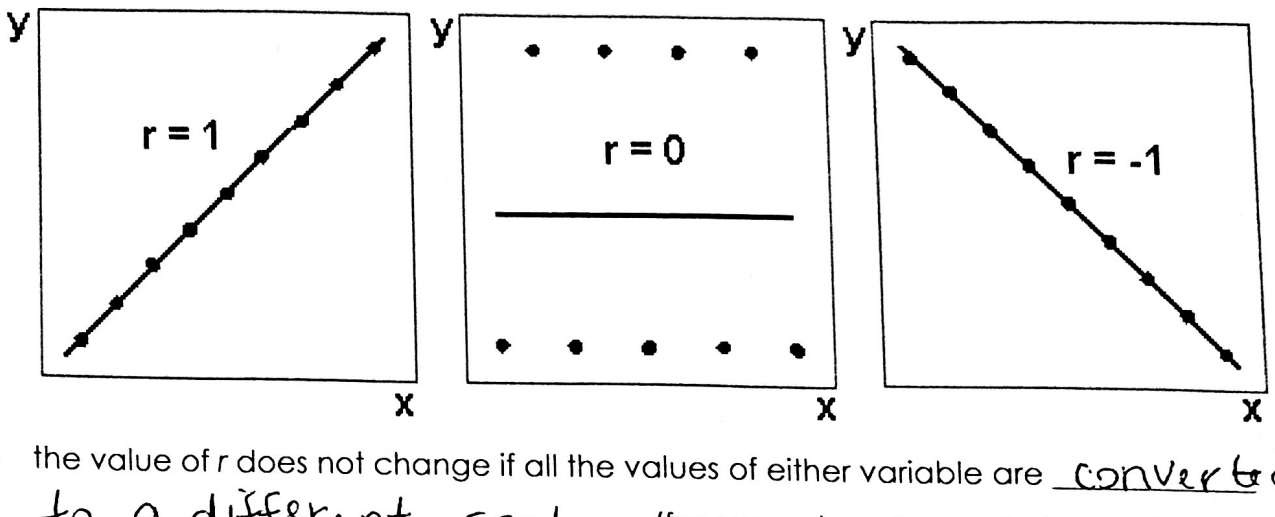
correlation analysis — a measure of the direction and strength of the linear relationship between the two variables

regression analysis — provides a method for drawing a straight line through the data points to summarize this linear structure

The correlation coefficient measures the strength of the linear association between two quantitative variables—called ***r***.

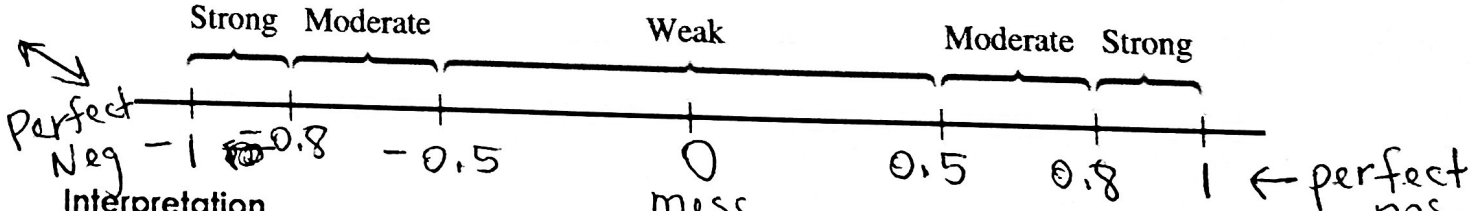
The correlation coefficient, *r* —

- Values of r range from +1 (perfect positive correlation), through 0 (no correlation), to −1 (perfect negative correlation):



- the value of *r* does not change if all the values of either variable are converted to a different scale (for example: r does not change if you change from °C to °F)
- *r* has no units
- the value of ***r*** is not affected by the choice of **x** or **y** — they are not necessarily explanatory / response variables, although they might be
- ***r*** measures the strength of linear relationship — you can't really guess *r* from visual inspection

A correlation greater than 0.8 is generally described as strong, whereas a correlation less than 0.5 is generally described as weak. These values can vary based upon the "type" of data being examined and the context. A study utilizing scientific data may require a stronger correlation than a study using social science data.



Interpretation

r = 1 perfect + correlation

r = −1 perfect − correlation

r = 0 (or close to 0) indicates that there is no linear association —look at the scatterplot to confirm (could be a non-linear relationship)

*** Always plot data 1st to make sure it looks linear and has no outliers.