Scatter Plots and Correlation

Def'n$_1$: Causal Relationship: Where one variable directly affects another.
Note: Proving a causal relationship between two (or more) variables is the purpose of many studies.

Def'n$_2$: Scatter Plot: A graph in the Cartesian plane which displays the joint distribution of two variables in which each point represents a pair of variable values.

Def'n$_3$: Correlation: The relationship between two variables.

Def'n$_4$: Dependent Variable: Always placed on the vertical axis. It is dependent upon the independent variable. (ie. It changes when the independent variable is changed)

Def'n$_5$: Independent Variable: Always placed on the horizontal axis. It affects the dependent variable.

Remark$_1$: We say that a **Linear Correlation** is:

Strong if the two variables vary at similar rates
- Even without drawing a line of best fit, the scatter plot will have a very clear pattern to it
- Most of the points in the scatter plot will be quite close to a line of best fit. Visualize a tight oval.

Weak if the two variables vary at rates that are not similar
- The scatter plot will have a visible pattern to it but not as clear as with a strong correlation
- The points will roughly follow the line of best fit, but will be somewhat scattered. Think fat oval.

Positive if the slope of the line of best fit is positive. The two quantities increase together. ← positive

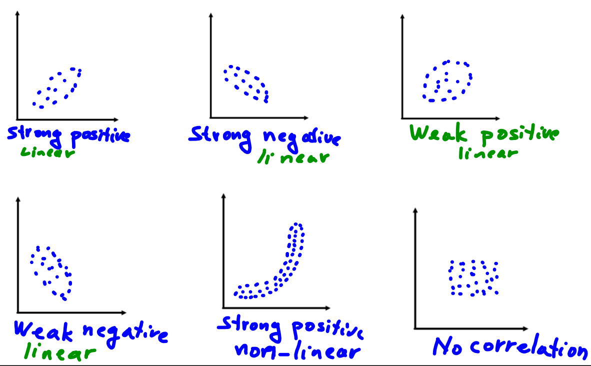
Negative if the slope of the line of best fit is negative. As x increases, y decreases. ← negative

If there is **no trend** or pattern with which to make predictions about future events:
- The points on the scatter plot will look completely random. Picture a square enclosing the points.
- If we construct a line of best fit, the points will not appear to follow it. The line itself will often be close to horizontal (zero slope)

Oct 2-8:24 AM

When you interpolate, you are making a prediction within the data. You are interpolating when the value you are finding is somewhere between the first point and the last point.

When you extrapolate, you are making a prediction beyond the data. Extend the linear model.

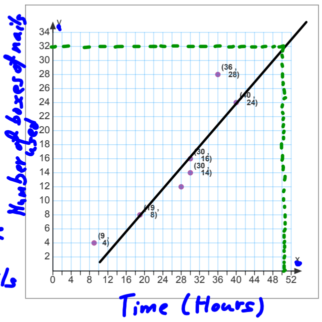


Oct 2-8:26 AM

Example 1: A roofer recorded the number of hours each employee worked in a week and the number of boxes of nails each employee used.

a) Draw a scatter plot

Time (Hours)	Number of Boxes of Nails Used
40	24
9	4
19	8
36	28
30	16
28	12
30	14



b) Observe the trend and identify the type of correlation.

There is a strong correlation between the hours worked and the number of boxes of nails used.

c) Estimate the number of hours Joe worked last week if he used 32 boxes of nails.

∴ Joe worked approximately 50 hours last week.

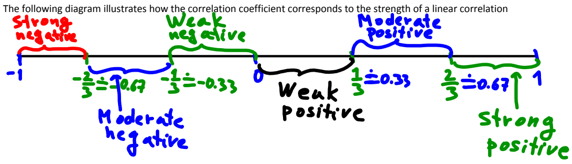
Oct 2-8:46 AM

The correlation coefficient, r is calculated using the following formula:
Note: You may use your calculator to determine values like $\sum y, \sum x, \sum xy$

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Σ - summation

The following diagram illustrates how the correlation coefficient corresponds to the strength of a linear correlation



r - correlation coefficient
R^2 - coefficient of determination

$r = \sqrt{R^2}$

Hwk: Pg. 142: #3, 8, 13, 16

Mar 8-12:51 PM