MDM 4U

SCATTER PLOTS AND LINEAR CORRELATION

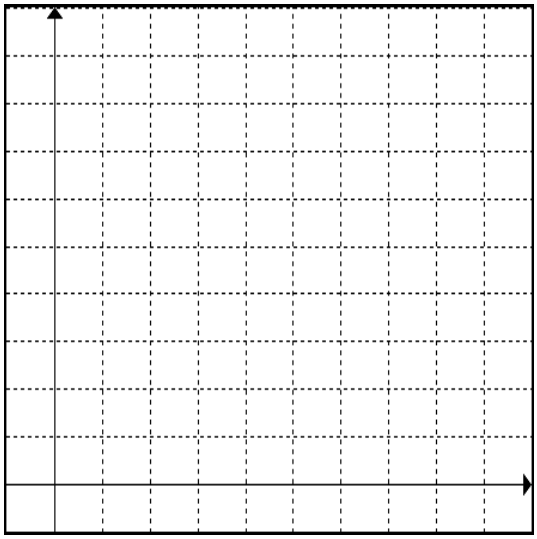
Two variable statistics provide methods for detecting relationships between variables and for developing mathematical models of these relationships.

A graphical model often reveals relationships between 2 variables.

Use the data below to sketch a scatter plot between the 2 variables.

- Prepare an appropriate scale to help spread out the data; label the axes.
- Plot the data.
- Sketch the line of best fit (straight line that passes as close as possible to all of the points on a scatter plot). Extend the line to the edges of the graph.

Hours of study time	Percentage on a test
2	62
4	80
1	52
6	92
4	75
8	95



2 variables have a ***linear correlation*** if changes in one variable tend to be proportional to changes in the other. The stronger the correlation, the more closely the data points cluster around the line of best fit.

1. Looking at the line of best fit (LOBF) you drew above, describe the line in terms of strength and slope.

2. Using the line, estimate the percentage on a test if the following numbers of hours studying were done.
   A) 0 hours B) 5 hours C) 30 minutes D) 12 hours

3. How many hours were spent studying if the following percentages were recorded?
   A) 50% B) 100% C) 35%

The **<u>STRENGTH</u> of a linear correlation** can be described by one of the following:

1. **Perfect Positive Linear Correlation** = all points lie on the line that slopes upwards from left to right.
2. **Strong Positive** OR **Moderate Positive** OR **Weak Positive**
3. **No Correlation** = there is little or no noticeable relationship between the variables.
4. **Strong Negative** OR **Moderate Negative** OR **Weak Negative**
5. **Perfect Negative Linear Correlation** = all points lie on the line that slopes downwards.

[Read Example 1 on pages 160-161]

A scatter plot often gives only a rough indication of the correlation between variables. The above descriptions are used to classify the correlation.

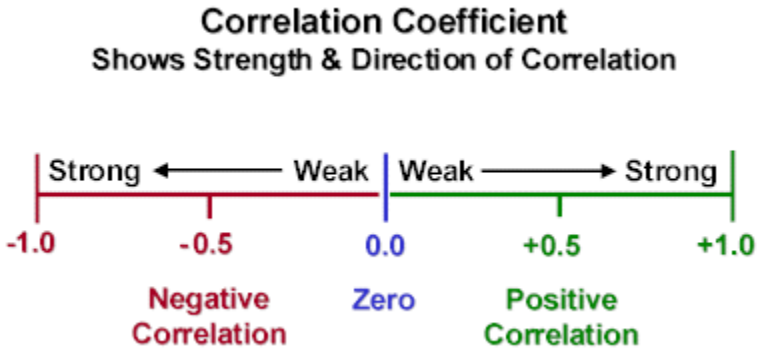
The strength of a correlation can be **measured using a formula** and **comparing the measure to a scale**.

The correlation is measured using the formula:

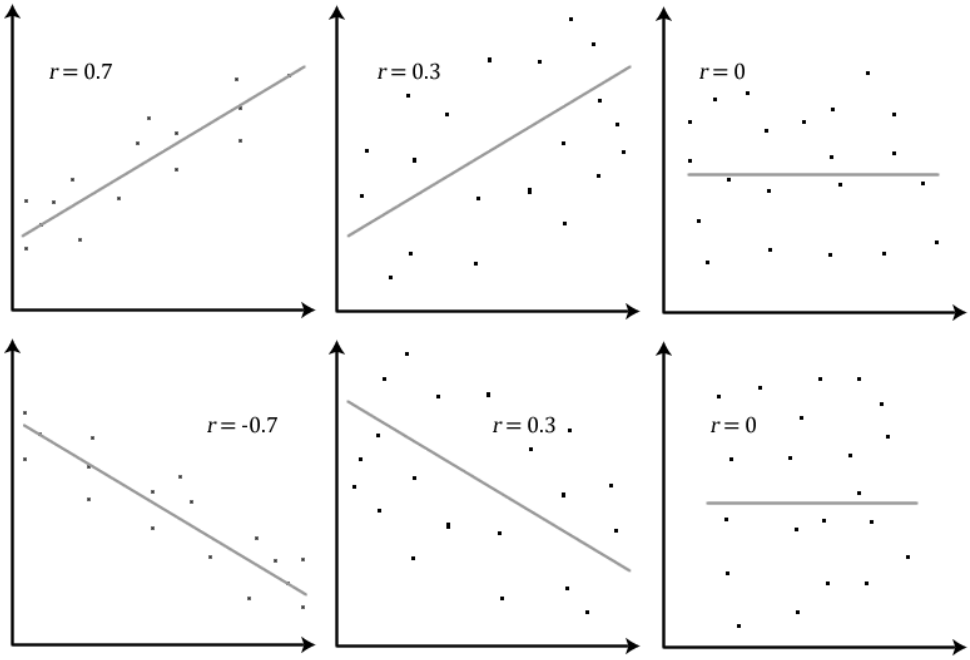
$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

where r = correlation coefficient
n = number of data points in sample
x = individual values of independent variable
y = individual values of dependent variable

Compare the "r" value to the strength of a linear correlation using the following diagram:



EXAMPLES OF SCATTERS PLOTS WITH CORRELATION COEFFICIENTS DISPLAYED:



For the data set below, complete the table, then use the formula provided to determine the correlation coefficient for the hours of study time and percentage on tests.

Hours of study time, x	Percentage on a test, y	x^2	y^2	xy
2	62			
4	80			
1	52			
6	92			
4	75			
8	95			
Σ=	Σ=	Σ=	Σ=	Σ=