Mr Murphy
AP Statistics
4.1 Correlation
HW Pg. 205 #5.1, 5.2, 5.5, 5.11, 5.12, 5.15

#Goals: 1. Calculate the correlation coefficient.
2. Know how to interpret ***r*** in context.

- The **correlation** measures the strength and direction of the linear relationship between two quantitative variables. Correlation is usually written as ***r***. The correlation r between x and y is

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

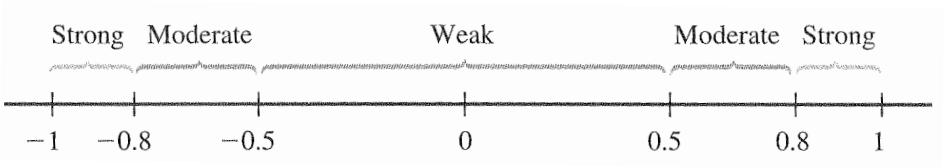
Ex1 Are more expensive bike helmets safer than less expensive ones? The accompanying data on x = price and y = quality rating for 12 different brands of bike helmets is given below. Quality rating was a number from 0 (worst possible rating) to 100, and was determined based on factors that included how well the helmet absorbed the force of an impact, the strength of the helmet, ventilation, and ease of use.

Price	Quality Rating
35	65
20	61
30	60
40	55
50	54
23	47
30	47
18	43
40	42
28	41
20	40
25	32

Let's enter this in our calculators and find *r:*

Put data in L_1 (x) and L_2 Stat -> Calc -> 8. LinReg(a + bx) -> L_1, L_2, Y_1
Note: Make sure your diagnostics are on: 2nd -> Catalog -> Scroll down to Diagnostics On -> Enter
Y_1: Vars -> Y-Vars -> 1. Function -> Y_1

- **Properties of *r***
  - The value of r does not depend on the unit of measurement for either variable.
  - The value of r does not depend on which of the variables is considered x.
  - The value of r is between -1 and +1.



  - The correlation coefficient r = 1 only when all the points in a scatterplot of the data lie exactly on a straight line that slopes upward. Similarly, r = -1 only when all the points lie exactly on a downward-sloping line.
  - The value of r is a measure of the extent to which x and y are linearly related.
  - When interpreting r there are three qualifiers you must always mention
    - Positive or negative
    - Linear
    - Strength

Let's check out this wicked cool web-site…
http://istics.net/stat/Correlations/

Ex1 (cont'd) Interpret r.

Ex2 The following data on the average finishing time by age group for female participants in the New York City marathon is given below. Find and interpret r.

Age Group	Representative Age	Average Finish Time
10 - 19	15	302.38
20 - 29	25	193.63
30 - 39	35	185.46
40 - 49	45	198.49
50 - 59	55	224.3
60 - 69	65	288.71

- The sample correlation r measures how strongly x and y values in a SAMPLE of pairs are linearly related to one another. ρ is the population correlation coefficient.
- **CAUTION: Correlation DOES NOT imply causation.**
    - Among all elementary school children, the relationship between the number of cavities in a child's teeth and the size of his or her vocabulary is strong and positive. Is it true that eating more foods that lead to cavities will increase vocabulary size?

Checkpoint:
Multiple Choice

1. If there is a very strong correlation between two variables, then the correlation coefficient should be

(a) close to +1
(b) close to -1
(c) close to -1 or +1
(d) close to zero
(e) There is not way to determine the correlation coefficient.

2. You are given the following set of observations for variables x and y.

x	-3	-1	1	3
y	8	4	5	-1

The correlation coefficient is:
(a) -1.0 (b) -0.8971 (c) +1 (d) 0.8971 (e) .2349

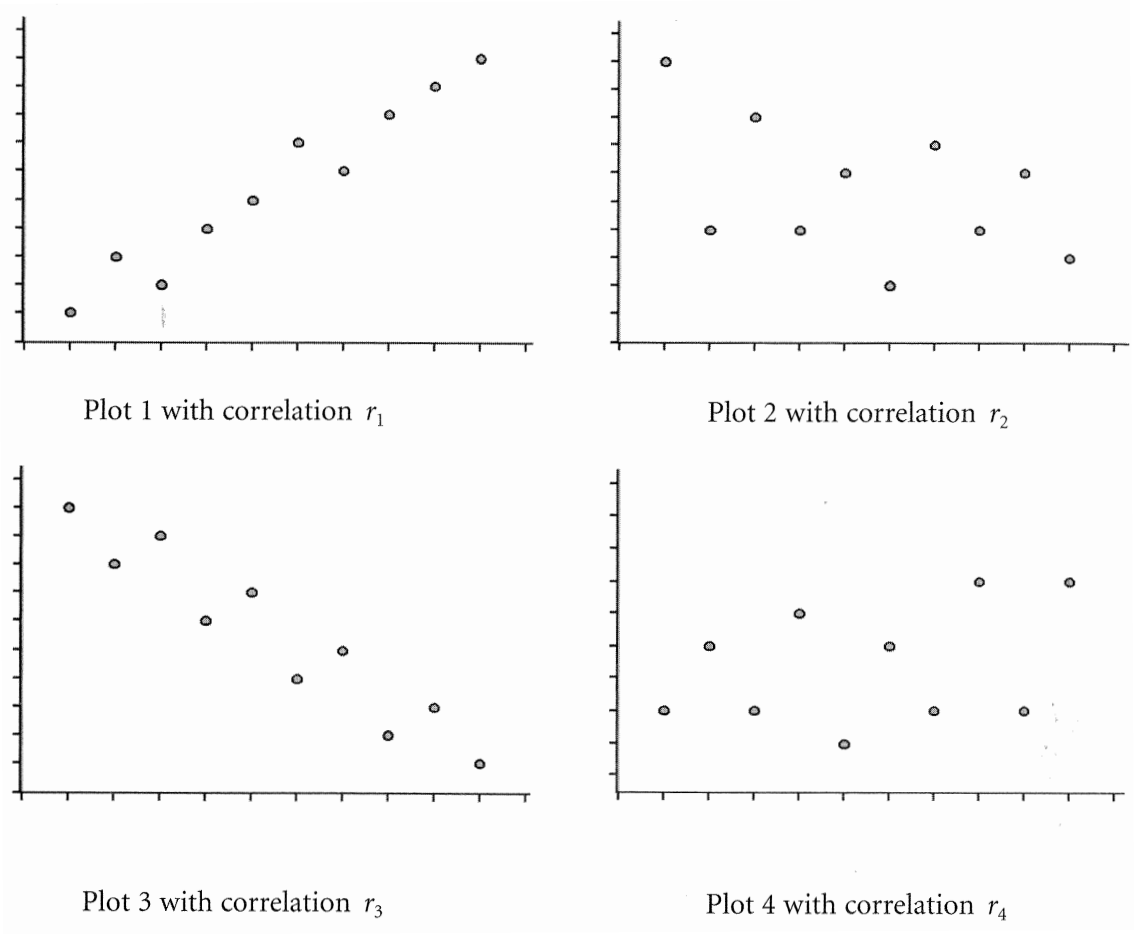
3. Pearson's correlation coefficient (r) is considered a symmetric measure because:

(a) its values range from 0 to 1.
(b) it indicates the causal relationship between two variables.
(c) the sign of r is the same as the sign of the slope.
(d) it will be the same regardless of which variable is the x and which is the y.
(e) None of the above.

4. Suppose the correlation between two variables is r = 0.23. What will be the new correlation if 0.14 is added to all values of the x variable, every value for the y variable is doubled, and the two variables are interchanged?

(a) 0.23 (b) 0.37 (c) 0.74 (d) -0.23 (e) -0.74

5. Order the correlation coefficients from least to greatest for the given scatterplots.



Plot 1 with correlation r_1

Plot 2 with correlation r_2

Plot 3 with correlation r_3

Plot 4 with correlation r_4

(a) $r_4 < r_3 < r_2 < r_1$

(b) $r_4 < r_2 < r_3 < r_1$

(c) $r_3 < r_2 < r_4 < r_1$

(d) $r_2 < r_3 < r_4 < r_1$

(e) $r_1 < r_2 < r_3 < r_4$