Stats: Regression, Part 1: Correlation and Regression (based on TF Chapter 3, pp. 56-58)

Name ____

1. Under what circumstances might you want to use a correlation instead of a t-test or ANOVA? Give an example of each with developmental variables.

Use this sample of data for the next few questions:

X	Y						
3	2						
5	4						
2	5						
6	4						
8	7						
4	5						
9	8						
3	4						
5	6						
6	7						

2. Calculate the correlation (r) between X and Y. Show your work. (You may want to use some of the blank columns above.) You can use the formula in TF, or one of the formulas below.

$$r = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

You may also need:

(sample) variance of X: $s_x^2 = \dfrac{\sum(X_i - \bar{X})^2}{N-1}$

$$(sample)\ covariance\ (X,Y) = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{N-1}$$

What is r?

ANSWER:

3. What does the value of r tell you?

ANSWER:

4. What does the sign of r tell you?

ANSWER:

5. Now compute the regression coefficient (B) for regressing Y on X using either of these algebraically equivalent formulas. (Note that you may be able to reuse a variance or covariance if you calculated those above.)

$$B = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} = \frac{N \sum XY - (\sum X)(\sum Y)}{N \sum X^2 - (\sum X)^2}$$

6. What is the major difference between the formula for B and the one for *r*? What is the significance of that difference?

ANSWER:

7. Ultimately we want to build a linear equation of the form Y = A + BX. (Also written as $\hat{Y} = b_0 + b_1 X$.) We computed B above. Now compute A, the Y-intercept, with the formula below. Show your work.

$$A = \overline{Y} - B\overline{X}$$