**FORMULAS:**

 OR 

(Formula Difference is in denominator – sq, roots then mult, vs. mult. then sq. root)

  

**STAT ESSENTIALS:** Be able to: 1) go through the steps leading to correlation and regression; 2) Conducting a correlation by hand and via SPSS; 3) Scatter Plot – building and interpreting; 4) Basics of Regression. [NOTE: see Orange sheets for additional problems and resources.]

**PROBLEMS:**

**1)** Identify steps that lead to regression:

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
4. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
5. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
6. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2)** When should you not conduct regression?

**3)** If a correlation is not statistically significant, what statistic represents your “best guess” estimate of a y variable for a given x variable?

**4)** What the … What type of relationship is this? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (one word will do here)



**5)** Located below are data collected for three paired variables. For which pairs would it appear appropriate to obtain a correlation coefficient? How did you figure this out?



**6)** What is the value of “n” for the relationship between Exercise & Milk Consumption in the prior problem?

**7)** Assume that there is a statistically significant correlation (r = .89) between the length of new born female infants and their birth weight. Using length as the X variable, determine the coefficient of determination AND explain what the heck this means in this relationship. (the internet may help here)

**8)** To use a least squares equation to estimate values outside the range of data values is called exsanguination. Wait a minute, I know from countless CSI programs that that isn’t right. So what is it called?

**9)** Age & Exercise







**10)** A medical researcher wishes to describe the relationship between the prescription cost of a brand name drug and its generic equivalent. The data (in dollars) are shown.



**11)** Given a line of best fit equation for the number of fires and acres burned is

Is the correlation between these two variables positive or negative? How do you know?

**12)** Hours Online The number of hours 12 students spent online during the weekend and the scores of each student who took a test the following Monday are shown below.





**13)** Data on tornadoes and deaths are shown below.



