**\* include your name and section (4677/6001) on your assignment.**

**Assigned Wed., Nov.14; due Mon., Nov. 19th**

PURPOSE: This homework is designed to help you think through some of the concepts we introduced in class (or in your reading) relating to statistics with a bit more hands-on work and critical thinking. If you are struggling to find answers to some of the questions, consult the Cozby book (some questions were explicitly drawn from it.) As before, easy answers to part 2 of the homework can be found in the ebook on “Learning Stats with R,” provided in Canvas as a pdf file. However, as before, you are free to use any language or software you like to compute the answers.

INSTRUCTIONS: Write your responses under each question. Upload or insert a graph or table as necessary. Submit the homework on Canvas as a .docx or .pdf file. **I do not need to see any code this time.**  *Students in 4677 may work in pairs but will still each submit a homework assignment. In this case, you will indicate at the top not only your own name, but the name of the partner you worked with.*

**Part 1 (10 points)**

1. (2 points) Discuss the ways that a pregnancy test could give a false reading in terms of type I and type II errors.
2. (2 points) A study on musician’s personalities finds that 36% of the variance is shared between the factors ‘openness’ and ‘extraversion.’ What would the Pearson’s r value of the correlation between these two factors be?
3. In an experiment, one group of research participants is given 10 pages of material to proofread for errors. Another group proofreads the same material on a computer screen. The dependent variable is the number of errors detected in a 5-minute period. A .05 significance (alpha) level is used to evaluate the results.
4. (3 points) What statistical test would you use? (Choices are: Regression, t-test, F-test (ANOVA), Chi square test.) Explain your decision.
5. (3 points) When Professor Rodríguez conducted the proofreading study, the average number of errors detected in the print and computer conditions was 38.4 and 13.2, respectively; this difference was not statistically significant. When Professor Seuss conducted the same experiment, the means of the two groups were 21.1 and 14.7, but the difference was statistically significant. Explain how this could happen.

**Part 2 (7 points)**

1. Load the data from the provided csv file “parenthood2.” There are four columns of data. The first logs the number of hours the father (Dan) sleeps, the second logs the number of hours the baby sleeps, the third logs a “grumpiness” value on a scale of 1-100, and the last column records the days.
2. (2 points) Plot the father’s sleep and the baby’s sleep separately. Use whatever type of plot that makes the most sense to you to visualize and understand the distribution of the data. (Show plots)
3. (1 point) Calculate the correlation coefficient for the father’s sleep and the baby’s sleep.
4. (2 points) Construct a scatterplot showing the relationship between the father’s sleep and the baby’s sleep. Label all axes. (1 point Bonus: draw a regression line through the plot.)
5. (2 points) Compute a correlation matrix. (Show it). Which pair of variables show the strongest correlation?

Total: 17 points (will be converted to final score out of 10)