**Week04 and week05 Assignment Due Oct 02 (30 points)**

**Test assumptions, experimental units, p-values**

Note: This one is a bit longer than previous assignments as it covers 2 weeks of lectures.

Q1 – Q12. Let’s revisit Figure 2b of the SilvestreRoig paper. In the week02 assignment, you plotted frequency histograms and made some guesses about symmetry and skew. You will now do a more detailed exploration of the data distribution of the 4 groups in the figure. Although the adherent and polarized groups are shown on the same graph, the authors really wanted to compare the adherent control to the adherent PDGF-BB and the polarized groups to each other.

Q1. Complete the table below. (3 pts)

N mean SD median skewness

Ctrl Adherent 39 40.9 19.1 37 0.61

PDGF-BB Adherent 35

Ctrl Polarized 26

PDGF-BB Polarized 28

Q2. From the table above, based on mean and median and skewness, which variable is most likely to have a symmetrical distribution and why? (1 pt)

Q3. Create individual box plots for each variable. Do the adherent groups side-by-side. So the same for the polarized groups. Be sure to label each graph and the Y-axis. Copy and paste below. (2 pts)

Q4. Based on the box plots, which group looks to be the least symmetrical and why? (1 pt)

Q5. Create a Q-Q plot for each group. Copy the graphs below. (2 pts)

Q6. Which of the Q-Q plots demonstrates the least symmetry? (1 pt)

Q7. Complete the table below. (2 pts)

Anderson-Darling D’Agostino-Pearson Shapiro-Wilk

Ctrl Adherent Normal Normal Normal

PDGF-BB Adherent

Ctrl Polarized Not normal

PDGF-BB Polarized

Q8. Do you think that doing a normality test is appropriate on the data and why or why not? (1 pt)

Q9. What type of data transformation would you do on each of the groups? (3 pts)

Group Type of transformation Why

Ctrl Adherent

PDGF-BB Adherent

Ctrl Polarized Power Continuous data with Left skew

PDGF-BB Polarized

Q10. You want to determine if the number of adherent neutrophils are different between the Ctrl Adherent and PDGF-BB Adherent groups. You would like to do a parametric independent (unpaired) t-test that has the assumptions the data are: normally distributed (symmetrical), homoscedastic, and independent. I will tell that from the study design, the data are independent. Do you think the data are symmetrical (in both groups) and are homoscedastic (between groups)? Why or why not? (2 pts)

Q11. Even if you concluded the data do not meet the assumptions of an independent t-test in Q10, for this example, we will assume that that they do. State the null hypothesis for the t-test and two-sided alternate hypothesis. (2 pts)

Q12. You perform the t-test and find that the p=0.001. What do you conclude about the differences between the number of neutrophils between the two adherent groups? (1 pt)

Q13-Q15. In the Week03 assignment, you did box plots of the Macrophage vs. SMC groups. The data for each group is highly skewed, but there is no evidence of heteroscedasticity as the standard deviations for both groups are almost the same. The mean, median, SD, and skewness values are shown below:

Macrophage SMC

Mean 12.97 9.48

Median 9.87 5.75

SD 10.44 10.92

Skewness 1.56 1.92

Q13. Apply an appropriate data transformation and complete the table below (Data to two decimal points; 2 pts)

Macrophage SMC

Mean

Median

SD

Skewness

Transformation used \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Q14. Plot the transformed data using box plots. Copy the box plots of the transformed data below. You can show them on the same plot. (1 pt)

Q15. Do you think your transformation worked (i.e., the data are more symmetrical)? State all the evidence for yes or no. (2 pts)

Q16a-d. An experiment was done where cells from the same breast cancer cell line were injected into two mammary pads (one on the right side, one on the left side) of each of 46 mice. The mice were randomized into either placebo or treatment group. The mice were studied for four weeks. At the end of the 4 week period, the volume for each tumor in each animal was measured using calipers. Total tumor burden was calculated as the sum of volumes of the tumors on the right and left sides.

Q16a. What is the total n for this experiment? (1 pt)

Q16b. What is the experimental unit? (1 pt)

Q16c. What is the observational unit? (1 pt)

Q16d. What is the statistical unit? (1 pt)