**Week03 Assignment Due SEP 11 (21 points total)**

Descriptive Stats and Graphing Data

For this assignment, we will be using data from the same Nature paper we used last week (you can find it in the Assignment folder): Silvestre-Roig, C. et al. Externalized histone H4 orchestrates chronic inflammation by inducing lytic cell death. *Nature* 2019; 569:236-264.

Q1a. Using the data from Figure 1K in the paper, calculate the mean, median, standard deviation, variance, range (both ways), interquartile range (either way), and skewness. Enter the missing data in the table below (to 2 decimal points only). (3 pts)

Macrophage SMC

Mean 12.97 9.48

Median 9.87 5.75

SD 10.44 10.92

Variance 108.99 119.25

Range1 0.00 to 61.80 0.00 to 66.40

Range2 61.80 66.40

IQR (either way) 5.83 to 16.22 1.73 to 12.91 or 10.39 11.18

Skewness 1.56 1.92

TAs each entry is worth 0.25 of a point.

Q1b. Based on the table above, would it be better to report the mean or the median as the measure of central tendency and why? (1 pt)

Median (because of skew)

Q1c. From the table above, what is the evidence that there is a right skew to the data in both groups? (give all evidence; 1 pt)

Mean larger than the median, skewness is positive and >1

Q1d. Reproduce the violin plot in Figure 1K. Make everything exactly the same including the axes and labels except for the following: 1) plot SMC without using any color but with the quartile and median lines visible; 2) no need to reproduce the box plots within the violin plots (If someone can figure out how they did this, let me know and we can share it with the class). Now do the violin plot showing all data points. Copy and paste the figures below. (2 pts)



Q1e. Now, using the same data for Figure 1k, create three new graphs (no need to use any different colors, but be certain your axes are correct and you label the graphs). The first is a box and whiskers plot of the data with the whiskers showing the min and max values. The second is a box plot with Tukey error bars. The third is a bar chart with standard deviation. Copy all graphs below. (4 pts)





Q1f. Vote for the graph you think best displays the data (Check one).

\_\_\_\_ Violin plot \_\_\_\_ Box plot with Tukey Bars

\_\_\_\_ Violin plot with all data points \_\_\_\_ Bar graph with SD

\_\_\_\_ Box plot with min and max \_\_\_\_ Bar graph with SD and all data points

TA’s no right answer. Give me a tally of the votes. No points

Q1g. Now create 3 graphs that show the 1) mean with SD, 2)mean with SEM; and 3) mean with 95% CI. Copy all three graphs below. Be sure to label each graph. (3 pts)



Q1h. Do you think p<0.05 based on the 95%CI graph above? (1 pt)

Yes

Q2a. Reproduce Figure 2e (minus the p-value; we will calculate that later). (1 pt)



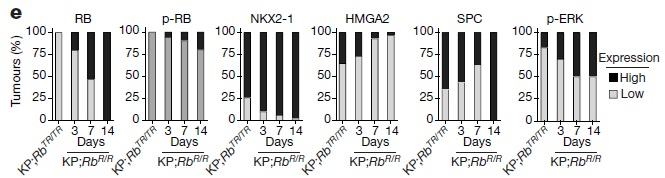
Q2b. What error bars did the authors use? (1 pt)

Standard deviation

Q3a. Reproduce the graphs for figures 2f and 2j. Do not worry about the lines through the graph or the r or p-values on the graph. We will revisit these when we do correlation and simple linear regression. (2 pts)



Q4. Figure 4e below Is from Walter RM *et al.* RB constrains lineage fidelity and multiple stages of tumour progression and metastasis. *Nature* 2019; 569:423-427. The data for p-ERK is below the figure.



|  |  |  |
| --- | --- | --- |
| Days Post Restoration | ERK(P) Low | ERK(P) High |
| Day 0 | 29 | 6 |
| Day 3 | 21 | 10 |
| Day 7 | 15 | 15 |
| Day 14 | 17 | 17 |

Q4a. Using the data above, create pie charts for Low and High groups for p-ERK. Use different colors to indicate the day (same color for the same time point in each graph). See if you can place the charts side-by-side with only one legend between them. (2 pts)