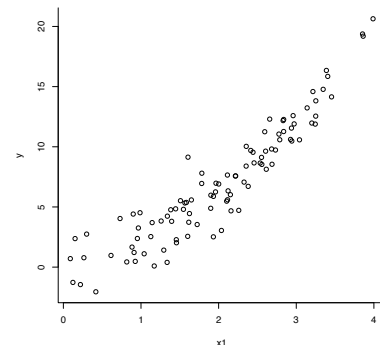


This is meant to gauge your understanding coming into BIFX 553. Full credit will be given for participation.

- 1) Define in your own words what a p-value is.
- 2) Write a short R script to add all numbers from 1 to 100 using a for loop.
- 3) Repeat problem 2 using only one line of code (i.e. without using a for loop).
- 4) Given the data.frame, *dat*, with the variables $\{x1, x2, x3, y\}$, write the R command to create a linear model for *y* using *x1*, *x2* and *x3* as predictors.
- 5) Write an R command to plot the relationship between *x1* and *y*.
- 6) Given the relationship between *x1* and *y* in this figure, update the R command in question 4 with a more appropriate model.
- 7) A collaborator presents a data analysis to you. The p-value they give you is $p = 0.012$. Is this statistically significant?
- 8) After discussing the results further, you discover that this p-value represents just one of 20 tests performed in the analysis. What is the Bonferroni threshold for significance if you want the family-wise error rate to be $\alpha = 0.05$? Is the result statistically significant?
- 9) What assumptions are made in a simple linear regression? What R command(s) can you use to get a quick look at whether those assumptions have been violated?

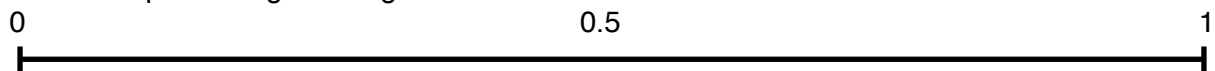


10) Explain the relationship between confidence intervals and p-values. Why are confidence intervals more useful than p-values?

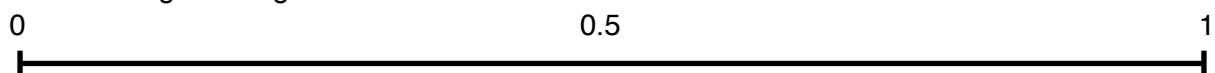
11) Rate your knowledge of the following topics by placing an 'X' on the scales below, where

- 0 = *"I know nothing"*
- 0.5 = *"I am proficient when I can have access to Google"*
- 1 = *"Given a complex problem, I could write a flawless, elegant solution on the back of an envelope in indelible ink without breaking a sweat!"*

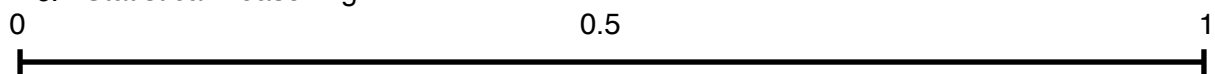
a. Computer Programming



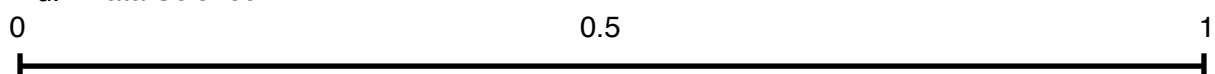
b. R Programming



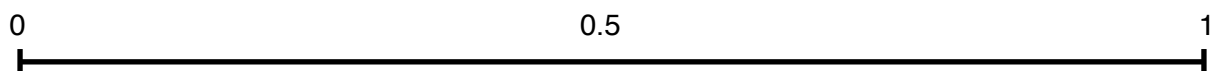
c. Statistical Reasoning



d. Data Science



e. Genetics



f. Biomedical Science

