**Oral Exam:**

Question Bank:

You will answer two questions, one you choose and one the professors choose. For the one the professor chooses, you will be asked either the exact question or a question very similar (i.e., Different parameter estimate table, plot, contrast etc.).

You may also be asked a follow up question(s) based on your responses.

As you study for this exam, try and pay careful attention to each word in your delivery. Little words can make a big difference.

Have fun! ☺.

**Easy:**

1. Define the terms population and sample. How do they differ?

Population is the entire set of data points, every possible observation in your study.

Sample is a subset of data used for the actual study that should represent the population it pulls from well.

2. What is the importance of a representative sample in a study?

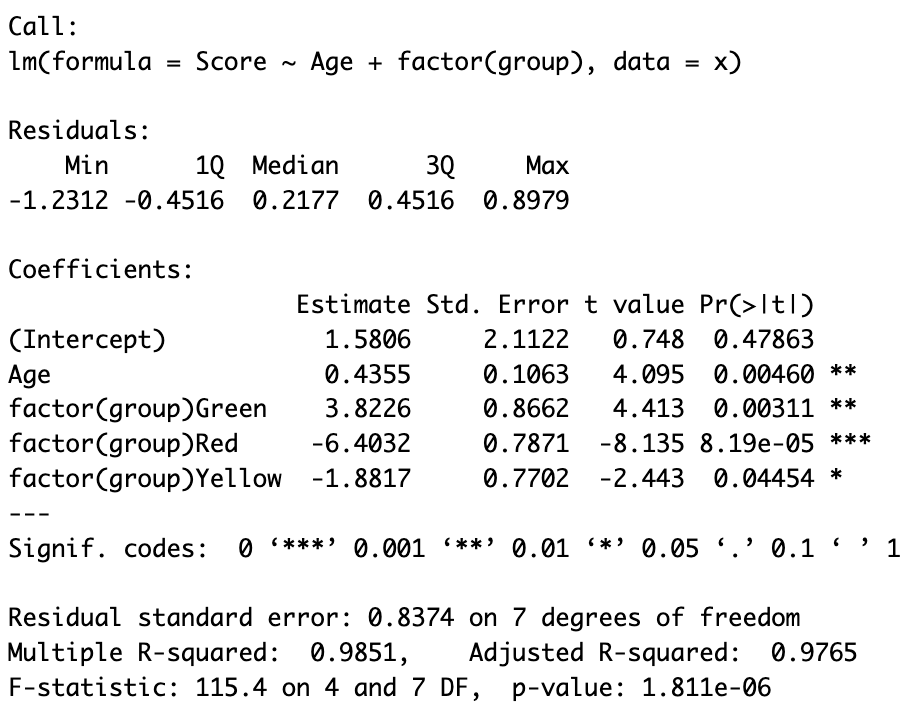
3. Discuss the differences between mean, median, and mode. In which scenarios might one be more appropriate to use than the others?

4. Can you explain what a standard deviation is and what it indicates about a data set?

10. What is a correlation? How does it differ from causation?

19. Assume you want to test the mean score on an exam between males and females. If there is strong evidence to suggest the standard deviations of the distributions of the scores for males and females are different, what should we do? What is the difference?

20. Take a second and review the following model and parameter estimate table. Interpret the slope of the Age variable from the following regression model:



Follow up… interpret .

**Medium:**

6. Can you explain the null hypothesis and alternative hypothesis in hypothesis testing?

7. What is a confidence interval and how does it relate to hypothesis testing?

8. What is a Type I error and a Type II error in the context of hypothesis testing?

9. Can you explain the concept of the p-value and its significance in hypothesis testing?

13. What is a box plot, and what information can it tell us about a data set?

14. What is the Central Limit Theorem, and why is it important in statistics?

21. What does it mean to have a high Cook’s D? Which point on the scatter plot below has the highest Cook’s D?

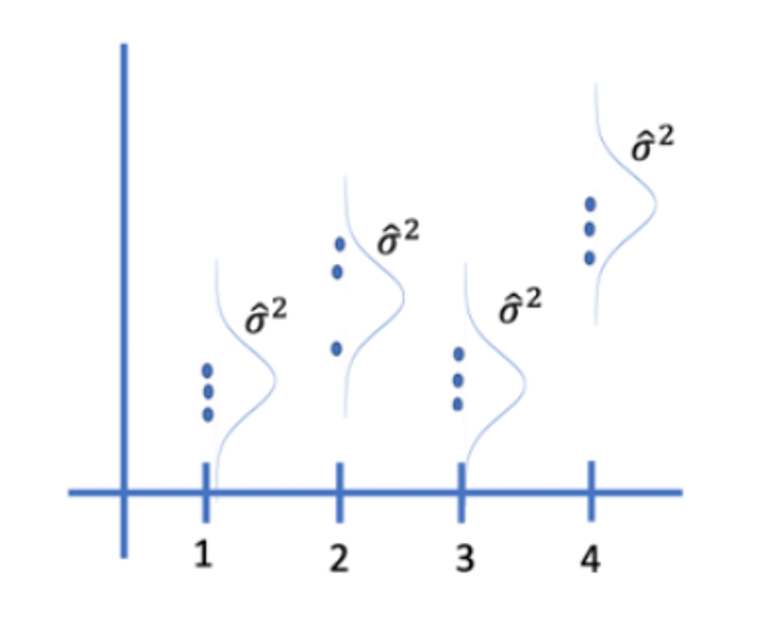
23. We want to test for an effect of a test prep program on student scores on a test. To test for this effect, we randomly select 100 students give those students a form of the test and record their scores. At the same time, we randomly select 100 different students from the same population and give them the test prep program. Upon completion of the program, we have the second group of students take the same test as the first group and record their score. What statistical test would you recommend to test for this effect?

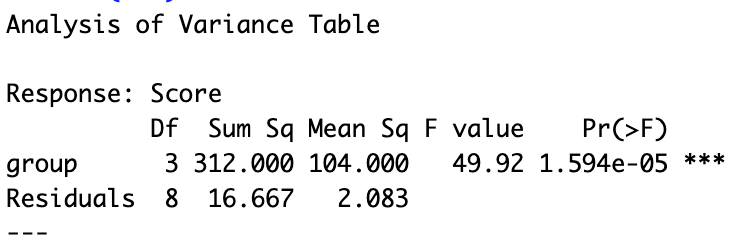
**More Involved:**

5. Please explain the need for a correction when we have multiple tests / comparisons. Name a method for making this correction. Follow up: Create a situation where a particular test would be appropriate and ask which method is best applied in this situation?

11. Explain the difference between parametric and non-parametric statistical methods and when you would use one method over another. Provide examples in your explanation.

12. Consider the plot and the corresponding ANOVA table below. What number represents the variances of the assumed normal distributions of Y for each group?





15. Can you explain the concept of statistical power and how sample size may affect it? Follow up: What is alpha and beta in the context of statistical power?

16. Write a null hypothesis that describes a contrast in which the average of the means of the A and D groups are tested for equality of the C and E groups? Simply hold it up to the camera when you are finished and we will take a snap shot.

17. With respect to the last question, what set of contrast weights reflects this contrast (assume they are ordered alphabetically)?

18. Assume we want to test the mean of group A and D for equality. Would it be better to filter the data to get only the As and Ds and then do a two sample t-test (assuming the assumptions are met) or include all the data including groups B,C and E in the analysis? Why?

22. We want to test for an effect of a test prep program on student scores on a test. To test for this effect we give the students a form of the test and record their scores. During the next week, we give them the test prep program. Immediately after completing the program, we have them take the test again and record their score again. What statistical test would you recommend to test for this effect?

24. An analyst suggested that a SLR model would not be appropriate here because both the response and predictor variables are not normally distributed.  Do you agree with the analyst or do you have something else to say about it?

