**112 – Fall 25 -Stat Assignment one**

**Problem one**

1. Ask ChatGPT to tell you the difference between multiple linear regression and  
multivariate regression.

2. Ask ChatGPT to give you an example of a research question that requires using  
multivariate regression:  
a) Education  
b) Psychiatry  
c) Finance  
d) Data science  
e) Economics  
f) Entrepreneurship  
g) Film/cinema/music  
h) Etc.

3. Ask ChatGPT tp tell you the major difference between multivariate analysis of variance  
and multivariate regression.

4. Summarize what you learned from the above exercise in a few lines.

**Problem two**  
Ask ChatGPT to ...

1. Create a data set of N = 500 observations with no missing data for predicting academic  
self-confidence (0-100), happiness with campus climate (0-100), and respect for diversity  
(0-100) from major (engineering, stem, non-stem), and the combined effect (interaction)  
of stress (0-100) with being a transfer student (yes/no). Make interaction statistically  
significant.

2. Give you the correlation matrix between the three outcomes. What do you notice about  
this correlation table?

3. Give you the MANOVA (Multivariate analysis of variance) summary table and interpret  
it for you.

4. Give you the OLS (ordinary least square) summary table for each of the three outcomes.

5. Draw the relevant interaction plots for the three OLS models and interpret them for you  
within context.

6. Comment on the practical significance of the findings.

**Problem three**

Ask Chat Gpt to create a data set of N=500 for you (no missing data) that would include all the variable you used to create the research questions that you created in the field of your interest in problem two part one.

1. Perform the relevant EDA including creation of the correlation matrix for all the  
   numerical variables.
2. Use the following example command to run the Manava model on your data.

*manova\_model <- manova(cbind(academic\_confidence, happiness\_climate, respect\_diversity) ~  
major + stress + transfer + stress:transfer,data = df)*

1. *Interpret the results of the MANOVA model.*
2. Now run univariate regression for each outcome. Interpret the findings within context.
3. Create the relevant interaction plots and interpret them.
4. Do you see a problem of multicollinearity in the OLS models? Yes or no? Support your  
   answer by providing the right criteria for measuring multicollinearity.
5. Suppose the coefficient of correlation between the three outcome variables ranges from  
   0.10 to 0.25, would you still use multivariate regression? Yes or no and explain why.
6. Could you use MANOVA to answer the research question given in this problem? Yes  
   and no and why?
7. Write a five-to-seven-line abstract of your findings within context.

**Problem four.** Go through the final report posted in in week two and answer the following questions about the report you looked through.

1. What did you learn from this report?

2. What are the strengths of this report

3. What errors if any do you find in this report?

4. What are the weaknesses of this report?

5. What additional information would you include if it were your

project?

6. What other statistical method would you consider if it were your

Project?

7. Do you see any use for multivariate regression? If yes how and

Under what conditions? Please explain.