**Stat 141XP -Spring 25**

**Review Exercise on survey design**

**Professor Esfandiari**

Download stemq2.csv to R, create all of the outputs using the given R codes. Once you have created the relevant outputs, answer the questions under parts I, II, and III.

**(You will find the lecture on survey design useful)**

The following is part of a survey that I conducted to learn more about the attitude of STEM majors toward their life at UCLA. This data set can be found in the Wednesday folder for week five on Bruinlearn. The data is labeled stemq2.csv.

SA = Strongly agree

A = Agree

NS = Not Sure

D = Disagree

SD = Strongly Disagree

Some items are written in a positive way meaning agreeing with them indicates satisfaction with our campus. Agreeing with negative items indicates lack of satisfaction. For positive items SA is assigned score of “5” and “SD” is assigned score of “1”. This is reversed for negative item.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Statement | SA | A | NS | D | SD |
| Q1. My family believes STE majors make more money. (+) |  |  |  |  |  |
| Q2. My high school prepared me well for STEM classes. (+) |  |  |  |  |  |
| Q3.Myfamily believes that STEM is more appropriate for males. (-) |  |  |  |  |  |
| Q4. My family struggled financially when I was growing up. (-) |  |  |  |  |  |
| Q5. I need to work to pay for my expenses. (-) |  |  |  |  |  |
| Q6. I need to graduate and get a job as soon as possible. (-) |  |  |  |  |  |
| Q7. Most of my friends are from own major. (+) |  |  |  |  |  |
| Q8. The workload in my major is overwhelming. (-) |  |  |  |  |  |
| Q9. The workload in my major does not leave enough room for a social life.(-) |  |  |  |  |  |
| Q10. I experience anxiety about academics. (-) |  |  |  |  |  |
| Q11. I experience anxiety about getting along with my roommates. (-) |  |  |  |  |  |
| Q12. I experience anxiety with my dating life. (-) |  |  |  |  |  |
| Q13. I have learned how to cope with the stress in my academics. (+) |  |  |  |  |  |
| Q14. I seek help when I have an academic problem. (+) |  |  |  |  |  |
| Q15. I have a general idea about what I want to do after graduation. (+) |  |  |  |  |  |
| Q16. I feel excluded at UCLA. (-) |  |  |  |  |  |
| Q17. There is not enough students like me in my major. (-) |  |  |  |  |  |
| Q18. I wish there were more faculty I could identify with. (-) |  |  |  |  |  |
| Q19. I feel I belong at UCLA. (+) |  |  |  |  |  |
| Q20. My parents understand what is takes to get into UCLA. (+) |  |  |  |  |  |
| Q21. My parents understand the value of a college education. (+) |  |  |  |  |  |
| Q22. I am happy at UCLA. (+) |  |  |  |  |  |
| Q23. I am confident in my ability to succeed at UCLA. (+) |  |  |  |  |  |
| Q24. My experiences at UCLA have helped me to interact with diverse groups (+) |  |  |  |  |  |

Given the above answer the following questions.

Part I

1. Estimate the reliability of the above survey by calculating Cronbach’s alpha.
2. Estimate the reliability after removing each items.
3. Given the guidelines in the lecture, is the reliability acceptable.

Part II

1. Carry out the relevant calculation to show which items work in the same direction as the overall survey and which ones do not.
2. Remove the items that do not work in the same direction of the test and recalculate Cronbach’s alpha. Use a correlation of 0.4 as the cutoff. Did anything change? The items that loaded to factor one are colored in red and those that loaded to factor two are colored blue.

Part III

1. Conduct parallel analysis, and create the plot of parallel analysis, identify the number of underlying factors, and interpret, the resulting lines.
2. Assuming two factors in the survey, conduct exploratory factor analysis using the number of factors that you decided underlies the data.
3. What percentage of the variance is explained by each component?
4. What do the items marked in red show?
5. Go back read the questions marked in red and decide what

Clusters in the survey the identify?

1. Create the plot of EFA and determine which items loads to which factors. Use two factors
2. Read the given questions and see if you can define the resulting factors. For example, if most of the questions related to academics load on one of the factors, you can label that factor “academics”.
3. Calculate the coefficient of correlation between the total score and the 24 items on the survey and identify the factors that did not load into any of the two factors. What could the reasoning be?