PSYC 4330 – Seminar in Statistics

Assignment #2

For this assignment, you will be using the *sat.act* dataset available within the `psych` package in R.

For your analyses, be sure that your code and output are both visible. You can use screenshots from RStudio to make things easier.

Here is some information regarding the variables in the dataset:

1) Gender: 1 = males; 2 = females
2) Education: 0 = less than high school; 1 = high school; 2 = post-secondary diploma program; 3 = some university/college; 4 = completed university/college; 5 = graduate work.
3) Age = age in years
4) ACT = American College Test (ranges from 0 to 36)
5) SATQ = Standard Aptitude Test – Quantitative (ranges from 200-800)
6) SATV = Standard Aptitude Test – Verbal (ranges from 200-800)

For the questions below, use α = .10 for any null hypothesis significance tests.

Part A (12 marks)

Explore how the interaction between gender and age affects scores on the SATQ. More specifically, summarize the findings related to this effect, including the following:

i)   Set-up the variables/model and print the results (3 marks)

ii)   Summarize the traditional null hypothesis significance testing results (1 mark)

iii)   Plot the interaction, with a summary of visualization (2 marks)

iv)   Conduct a negligible effect test on the interaction (using a negligible effect interval of *b* = -2 to *b* = 2), and summarize the results (3 marks)

-    Note that specifying the formula with the following formats will not work in neg.reg: y ~ x1*x2 or y ~ x1 + x2 + x1:x2. There are two ways around this: a) compute the interaction term before running the model (e.g., int = x1*x2, using a numeric version of gender), and then use neg.reg (formula = y ~ x1 + x2 + int), or b) compute the confidence interval for the interaction separately using *confint*(model).

v)   Provide an interpretation of the magnitude of the interaction using both standardized and unstandardized units, as well as an interpretation of magnitude using the proportional distance (specific to the negligible effect test). (3 marks)

Part B (12 marks)

Explore the effect of level of education (treat education as a categorical variable/factor) on ACT scores, controlling for SATQ scores. More specifically, summarize the findings related to this effect, including the following:

i) Set-up the variables/model and print the model results (2 marks)

ii) Conduct all pairwise comparisons of the means of the levels of education, using each of: 1) no multiplicity control; 2) Bonferroni familywise error control; 3) Benjamini-Hochberg false discovery rate control. Use a table to summarize which pairwise comparisons were statistically significant using each method. Also, provide a brief (couple of sentences) summary of the null hypothesis significance testing results for each method. (8 marks)

iii) Summarize which of the pairwise comparisons resulted in meaningful differences between the levels of education (where meaningful is operationally defined as a difference in magnitude on ACT, controlling for SATQ, of at least 2 points) (2 marks)

iv) State, in one sentence, whether you believe multiplicity control has any value within the field of Psychology. (0 marks, but you need to complete this question to get any marks for Part B 😊)

Part C (3 marks)

For this question, interest is in how age affects SATV scores. However, we suspect that the state in which individuals live might impact the results (we don't have this variable in the dataset, but imagine that it is available). More specifically, we would like to allow the intercepts to vary across the states.

Write out the model (for each level, and for the composite model) for predicting SATV scores from age, while allowing the intercepts to vary across states. Don't use X and Y in your model, use the actual variable names (e.g., SATV, AGE).