GSERMCDA17: Binary Regression Model + Testing & Fit

Your Name: Points received: ____ out of 150

In what follows, I refer to the dependent variable as Y, the binary variable you interpret as B, the continuous variable you interpret as C, and all other variables in your model as X. B and C must be statistically significant at the .05-level.

1. ___ of 5: Choose one of the data sets available for class. Pick or construct a binary dependent variable and three independent variables. Keep only the variables you will be using and drop all missing cases (listwise deletion). Save the data under a new name (e.g. *addhealth-brm-sns.dta*). (Note that in this and all future assignments, constructed variables should be named and labeled in accordance with the "rules" outlined in the handout: *Workflow for CDA*.)
2. ___ of 5: Demonstrate that the data are clean by including the output from the following commands:

```
. codebook, compact
. sum
```

3. ___ of 5: Estimate a **logit** model of Y on B, C, and X. Be sure to use `i.` for indicator variables, as appropriate.
4. ___ of 5: Use `listcoef` to compute the factor change coefficients.
5. ___ of 10: Interpret the appropriate standardized and unstandardized factor change coefficient(s) for C. Use the z-statistic from the logit output to test if C significantly impacts Y.
6. ___ of 5:

   a) Use `test` to test the significance of C on Y using a Wald test.

   b) How is the specific value of the Wald test related to the z-test in question 5?
7. ___ of 10: Interpret the appropriate standardized and unstandardized factor change coefficient(s) for B. Use the z-statistic from the logit output to test if B significantly impacts Y.
8. ___ of 5: Use an LR test (`lrtest`) to test if B significantly impacts Y. Show the appropriate output and write a sentence indicating your conclusion based on the LR test.
9. __ of 5: Estimate the **probit** model corresponding to the logit model estimated above. Create a table that displays (1) the unstandardized coefficients from both the logit and probit models and the ratio between the two; and (2) the z-values from both models and the ratio between the two. NOTE: Use the full number of digits for computing ratios, but only include the appropriate number of decimal digits in the table. Excel is an easy way to do this.
10. ___ of 10: Answer the following questions (in 2 sentences or less):
    a) How different are the logit and probit unstandardized coefficients?
    b) Why are they so different?
    c) Why are the z-scores similar but not exactly the same?
11. ___ of 10:
    a) Test that the effects of B and C are simultaneously equal to zero using a Wald test using `test`.
    b) Test the same hypothesis using an LR test using `lrtest`. Show the appropriate output and write a sentence indicating your conclusion based on the LR test.
12. ___of 10: Re-estimate your **logit** model from (1). Compute the predicted probabilities for all cases using the `predict` command and include a `dotplot` of these probabilities (ensure your y-axis runs from 0 to 1). What substantive insights do you gain from this graph?
13. ___ of 5: Compute the discrete change coefficients for C and B using `mchange`. Hold other variables at values you find interesting or useful.

14. ___ of 10: Choose an appropriate discrete change coefficient for B from your `mchange` output and report the associated 95% confidence interval. Make sure the results match your output in 3 above. Interpret this discrete change coefficient, including the confidence interval.

15. ___ of 15: Use `mgen` and the `graph` commands to generate and plot predicted probabilities over the range of variable C for both values of B. Hold the other variables at the same location you chose in question 3 above. Present the plot. Write a paragraph telling the **story** of your results. This should read as though it were part of a journal article. Incorporate the magnitude of the effects and the associated confidence intervals as needed (Hint: Use `mchange` to calculate these*).* Also make sure to indicate the levels of any other variables in your model.

16. ___ of 10: Looking back on your work in this assignment, which method(s) of interpretation did you find most useful (factor change, discrete change, plotting, some combination)? Why?

17. __ of 15: Using the variables you've selected (& maybe a few more) estimate four logit models. The models can be as simple or complex as you like, and they do not have to make substantive sense. Try to make at least one non-nested in the others.

    a) Store and present the estimates of each model using `estimates store` and `estimates table` or `esttab`. Include the BIC and AIC statistics.

    b) Based on the BIC statistic, which model is preferred and how strong is the evidence?

    c) How does your answer to 10b correspond to your substantive evaluation of the models?

    d) Does AIC give you the same conclusion? If not, why might this be?

18. ___ of 10: My overall evaluation of your work.