ICPSRCDA17 Assignment 7: count models

Your Name: Points received: ____ out of 130

⇒ **Hand in to TA by 10am Friday, August 18.**

1. ___ of 10: Using one of the class datasets, choose a substantively reasonable dependent count variable Y. **Note:** Counts such as # of children that have limited range often cause problems and should not be used for this assignment. Choose or create three substantively meaningful independent variables. I'll refer to the binary independent variable as B, the continuous as C, and the other as X. Keep only the variables you will be using and drop all missing cases (listwise deletion). Demonstrate that the data are clean by including output from `codebook, compact` and `sum`.

2. ___ of 10. Present a simple table describing your variables. Include information on the distribution of the dependent variable. If your outcome variable has many categories, you may want to group the counts for higher values (e.g., 0, 1, 2, 3, 4, 5-9,10-15, etc.) for purposes of presentation.

3. ___ of 5: Use `poisson` and `nbreg` to regress Y on C, B and X. Note that C and B must be statistically significant in BOTH models.

4. ___ of 10: Using Excel or another method, construct a table comparing the effects and significance of C and B from the PRM and the NBRM. Column 1=Name of variable; Column 2=PRM unstandardized coefficient; Column 3=NBRM unstandardized coefficient; Column 4=ratio of Col2 and Col3; Column 5=PRM z-value; Column 6=NBRM z-value; Column 7=ratio of Col5 and Col6.

5. ___ of 5: Why do you expect the ratio in Column 4 to be close to 1.0? Relate this to the mean structure of the two models.

6. ___ of 10: What do you expect for Column 7? Relate what you find to the results of question 4 and the effects of ignoring unobserved heterogeneity.

7. ___ of 5: Test the NBRM against the alternative of the PRM. Write up the result as though it were part of a research paper. Include information on statistical significance.

8. ___ of 5: Using the preferred model from above, compute the factor change coefficients using `listcoef, help`. Include the output.

9. ___ of 10: Interpret the <u>standardized</u> factor change coefficient for C and the <u>unstandardized</u> factor change coefficients for C and B. Include statistical significance. This should read as though it were part of a published article.

10. ___ of 5: Using `mchange`, calculate the discrete change for B holding the other variables at some location you find interesting or useful.

11. ___ of 15: Choose an appropriate form of the discrete change from your `mchange` output, highlight these numbers, and reproduce them using `margins` and `mlincom`. Interpret these results.

12. ___ of 5: Use `zip` and `zinb` to estimate a zero-inflated poisson model and a zero-inflated negative binomial model. Think carefully about which IVs you will "inflate." Note that C and B must be statistically significant in either the count OR the binary portion of the model.

13. ___ of 5: Test the ZINB against the alternative of the ZIP. Write up the result as though it were part of a research paper. Include information on statistical significance.

14. ___ of 10: Using the preferred model from question 13, calculate the factor change coefficients using the command `listcoef, help`. Interpret the unstandardized factor change coefficients for B and C (and X if B or C is not included in both portions of the model). This should read as though it were part of a journal article and should include information on statistical significance.

15. ___ of 10: Use `countfit` to compare PRM, NBRM, ZIP and ZINB. Based on these results and your substantive understanding of the process being studied, which model would you use and why? Include the graph from `countfit` and any other output that supports your conclusion.

16. ___ of 10: My assessment of the overall effectiveness of your answers.