

ICPSRCA18 Assignment 7: count models

Your Name:

Points received: \_\_\_\_ out of 130

⇒ **Email to Shawna (shawnana@umich.edu) by 10am Friday, August 17.**

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`, or their R equivalents.
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: Regress Y on C, B and X using both poisson and negative binomial models. Note that C and B must be statistically significant in BOTH models (*Hint: which model should you run first??*).
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. Include the output.
9. \_\_\_\_ of 10: Interpret the standardized factor change coefficient for C and the unstandardized 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: Calculate the discrete change for B holding the other variables at some location you find interesting or useful.
11. \_\_\_\_ of 15: Calculate the 95% confidence interval for the discrete change for B calculated in question 10. (If you computed 10 using `mchange` in Stata, be sure to highlight the appropriate coefficient in the table). Interpret these results.
12. \_\_\_\_ of 5: Estimate a zero-inflated poisson model and a zero-inflated negative binomial model. Think carefully about which IVs you will "inflate," but include at least one of C or B in both portions of your model. 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. Interpret the unstandardized factor change coefficients for B and C in **both portions** of the model (i.e.,

interpret both the factor change coefficients relating to the factor change in the count and the factor change in the odds of being “always zero”). This should read as though it were part of a journal article and should include information on statistical significance.

15. \_\_\_ of 10: 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 output that supports your conclusion (e.g., graphs of residuals, test statistics, information on under/overprediction, etc.)
16. \_\_\_ of 10: My assessment of the overall effectiveness of your answers.