

Stat 503 Assignment 7: count models

Your Name:

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

⇒ **OPTIONAL ASSIGNMENT**

⇒ **Please email write-up and log file to [sns3@indiana.edu](mailto:sns3@indiana.edu) by 10am Friday, August 12**

1. \_\_\_\_ of 10: Using one of the class datasets, choose a substantively reasonable dependent count variable Y. [Note: Count variables with a limited range often don't work well as dependent variables. Please choose a DV that has at least 10 counts 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. Please keep in mind that you will be estimating zero-inflated models, so pick at least one IV that you feel is reasonable to include in the "inflation" portion of the model. Also note that C and B must be statistically significant in the PRM and NBRM and significant in either count or binary portion of the zero-inflated models. 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 5. Present a simple table describing your variables. Include information on the distribution of the dependent variable. If your outcome has many values, you might want to group the counts for higher values (e.g., present the percent for 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.
4. \_\_\_\_ of 5: Using Excel or another method, construct a table comparing the effects and significance of C and B from the PRM and the NBRM. Column1=Name of variable; Col2=PRM unstandardized coefficient; Col3=NBRM unstandardized coefficient; Col4=ratio of Col2 and Col3; Col5=PRM z-value; Col6=NBRM z-value; Col7=ratio of Col5 and Col6.
5. \_\_\_\_ of 5: Why do you expect the ratio in Col4 to be close to 1.0?
6. \_\_\_\_ of 10: What do you expect for Col7?
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 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: Using `prchange`, 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 `prchange` output, highlight these numbers, and reproduce them using `prvalue`, `save` and `prvalue`, `dif`. Interpret these results. It is not necessary to interpret the discrete change for all values of the dependent variable (although you should interpret more than just one). Rather, paint an overall picture of the effect. Include information on statistical significance.
12. \_\_\_\_ of 5: Use `zip` and `zinb` to estimate a zero-inflated poisson model and a zero-inflated negative binomial 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 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 `prgen` and the `graph` command to plot the  $\Pr(y=0)$  across the range of C while holding the other variables at some location you find interesting or useful. This should look similar to the graph on slide 59 of the Lecture Notes. Describe, in a general way, what you find.
16. \_\_\_ 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.
17. \_\_\_ of 10: My assessment of the overall effectiveness of your answers.