

Soc 73994, Homework #9

Models for Binary Outcomes II & III

Richard Williams, University of Notre Dame, <https://www3.nd.edu/~rwilliam/>

Last revised March 23, 2020

All answers should be typed and mailed to the TA. Be sure your response includes your name, the date, and a clear title, e.g. Homework # 9. If there is a huge amount of output for any analyses you run yourself, you may want to be selective in what you copy and paste into your assignment (but make sure you include enough so it is clear what commands you executed, e.g. you might show all the commands but only parts of the output).

This assignment focuses on standardized coefficients and comparing logit and probit coefficients across nested models. First you will interpret the results from a prepared problem, and you will then do a similar analysis using a data set of your choice. Ideally you will use your own data for the latter but you can use something else. Note that, if you normally use the `svy: prefix`, it may not work with all the commands you are asked to use so you may have to drop the prefix for this assignment.

1. Prepared problem (put together by Chris Quiroz). Run the following commands:

```
clear all
use https://www3.nd.edu/~rwilliam/statafiles/anes_codeddata, clear

*****
*Descriptives

des pollstation yearseduc income age
sum pollstation yearseduc income age

*****
*Logistic Regression Model nested

quietly logit pollstation yearseduc income age

logit pollstation yearseduc if e(sample)==1, nolog
est store m1
listcoef, std help

logit pollstation yearseduc age income, nolog
est store m2
listcoef, std help

esttab m1 m2

knb logit pollstation yearseduc || age income
```

Discuss

- a. How the coefficient for the variable in the baseline model changes as more variables are added.

- b. How the y-standardized coefficient changes as more variables are added.
 - c. Briefly explain why y^* gets rescaled with each equation. Then discuss how and why your comparison of coefficients across models may be affected by the rescaling. Remember that `listcoef` gives you the standard deviation of y^* , so you will want to mention that in your discussion.
 - d. What does the `khb` command tell you about the direct and indirect effects of that variable? In your case, do you think the `khb` command is helping you to make (at least slightly) more correct statements than you would have made had you only looked at how the coefficients changed across models? Remember that you have help files and citations for `khb` if you think you need to know more about it.
-

For the remaining problems use a data set of your choice.

2. Run a series of at least three nested logistic regression models. The first model should be very simple, maybe even only one variable (e.g. race, gender). The 2nd model should add a few variables while the third adds even more. There should be some sort of logical sequencing to the order in which you add variables. Ideally the added variables will be statistically significant and noticeably increase pseudo R^2 . Further these variables should be expected to account for at least some of the effects of the variables in your baseline model, e.g. the direct effect of race should decline as education is controlled for. Make sure the sample used is the same for each model. If you are not sure how to do this see the first few pages of the Missing Data Part 1 handout. Use something like `esttab`, `outreg2`, or `estimates table` to display your models in side by side columns.
3. After each model, run the command `listcoef, std help`. Then discuss
 - a. How the coefficients for the variable or variables in your baseline model change as more variables are added.
 - b. How the y-standardized coefficients change as more variables are added.
 - c. Briefly explain why y^* gets rescaled with each equation. Then discuss how and why your comparison of coefficients across models may be affected by the rescaling. Remember that `listcoef` gives you the standard deviation of y^* , so you will want to mention that in your discussion.
4. Now use the `khb` command. Choose a variable (or the variable) from your first model. What does the `khb` command tell you about the direct and indirect effects of that variable? In your case, do you think the `khb` command is helping you to make (at least slightly) more correct statements than you would have made had you only looked at how the coefficients changed across models? Remember that you have help files and citations for `khb` if you think you need to know more about it. (Don't feel too bad if the results don't seem too dramatic! That is often the case. But if you do have a reasonable example where KHB seems to make a major difference in how you interpret the findings, I would like to see it.)
5. Discuss anything else you think is useful from your analyses, e.g. do the x-standardized coefficients from your final model tell you much?