ICPSRCDA17 Assignment 4: testing and fit

Your Name: Points received: ____ out of 110

This assignment gives you practice with statistical tests, assessing models using the BIC and AIC statistics, and examining residuals.

Using the *science4* data (e.g., *usecda cda_science4*), create a binary dependent variable (Y) based on a measure of publications or citations. At least 10% of the cases must be in the smallest category (to avoid convergence problems). In this assignment you will develop four competing models. Each model must contain at least one continuous (C) and one binary independent variable (B). These models can have completely different right-hand-sides, or the models can be nested (e.g., model 1 includes age and race as an independent variables; model 2 includes age, race, and education; model 3 includes age, race, education, and income; etc.). **This means you will need to choose no fewer than 5 independent variables.**

For questions marked with *'s (and only for these questions), do the following:

a) State the hypothesis mathematically (e.g., $H_0$: $\beta_{PHD}=0$).

b) State the hypothesis in prose (e.g., Ph.D. prestige has no effect on scientific productivity).

c) Interpret the results of the test as though it were a sentence lifted from a substantive paper.

When typing your answers, use alt-Insert-Symbol for adding symbols (or the Symbol font), the subscript font for subscripts, and the superscript font for superscripts. At the start of each question, before you begin your answer, include in a fixed font the output used for your answer.

1. ___ of 5: After opening your data, keep only the variables you will be using and drop all missing cases using `misschk , gen()` (listwise deletion). Demonstrate that the data are clean by including the output from the following commands:

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

2.* ___ of 10: Pick three of your independent variables, including on binary (B) and one continuous (C). Estimate a logit model of Y on B, C, and X. Use the z-statistic from the logit output to test if B significantly impacts productivity.

3. ___ of 10:

a) Use `test` to test the same hypothesis using a Wald test.

b) How is the specific value of the Wald test related to the z-test in question 1?

4. ___ of 5: Test the same hypothesis using an LR test with `lrtest`. Show the appropriate output and write a sentence indicating your conclusion based on the LR test.

5.* ___ of 10: Test that the effects of B and C are simultaneously equal to zero using a Wald test using `test`.

6. ___ of 5: 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.

7.* ___ of 10: Test the hypothesis that all coefficients are simultaneously equal to zero using a Wald test.

8. ___ of 5: 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.

9.* ___ of 10: Now test a more complicated hypothesis, for example, that the effect of C is twice the effect of X, or the effect of C is equivalent to the effect of X. Use a Wald test.

10. ___ of 15: Estimate your four logit models.

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, what does this suggest?

11. ___ of 15: Using one of the models from Question 10:

a) Estimate the logit model.

b) Compute the standardized Pearson residuals.

c) List the 10 most negative and 10 most positive residuals (not observations with residuals!) along with 1) the dependent variable, and 2) the independent variables. Use the commands `sort` and `list` to do this. Note: you may also wish to include the predicted probability for y=1 (obtained through the `predict` command) to help you in evaluating these residuals.

d) What do these residuals suggest about how the model or sample should be modified?

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