ICPSRCDA18: Binary Regression Model (Part 1)

Your Name: Points received: ____ out of 85

In what follows, I refer to the dependent variable as Y, the binary variable you interpret as B, the continuous variable you interpret as C, and all other variables in your model as X. Remember that B and C must be statistically significant at the .05-level.

1. ___ of 5: Choose one of the data sets available for class. Pick or construct a binary dependent variable and three independent variables, including your B and C variables (see above). Keep only the variables you will be using and drop all missing cases (listwise deletion). Save the data under a new name (e.g. *addhealth-brm-sns.dta*). (Note that in this and all future assignments, constructed variables should be named and labeled in accordance with the "rules" outlined in the handout: *Workflow for CDA*.)
2. ___ of 5: Demonstrate that the data are clean by including the output from the following Stata commands or their R equivalents:

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

3. ___ of 5: Estimate a **logit** model of Y on B, C, and X. Be sure to use `i.` for indicator variables, as appropriate.
4. ___ of 10: Compute the factor change coefficients. Present a table with the logit coefficients, the appropriate standardized and unstandardized factor change coefficients, and the z-values.
5. ___ of 10: Interpret the appropriate standardized and unstandardized factor change coefficient(s) for C; include statistical significance in your interpretation.
6. ___ of 10: Interpret the appropriate standardized and unstandardized factor change coefficient(s) for B; include statistical significance in your interpretation.
7. ___ of 15:
    a) Compute Pr(y=1) when B is equal to 1 and the other variables are held at their mean.
    b) Compute Pr(y=1) when B is equal to 0 and the other variables are held at their mean.
    c) Calculate the factor change coefficient for B using these values (you can do this by hand or by using the `display` command in Stata or the R equivalent).
    d) Repeat steps (a)-(c) holding the other variables at some value other than their mean.
    e) What do you find?
8. __ of 5: Estimate the **probit** model corresponding to the logit model estimated above. Create a table that displays (1) the unstandardized coefficients from both the logit and probit models and the ratio between the two; (2) the y-standardized coefficients from both models and the ratio between the two; and (3) the z-values from both models and the ratio between the two. NOTE: Use the full number of digits for computing ratios, but only include the appropriate number of decimal digits in the table. Excel is an easy way to do this.
9. ___ of 10: In the following questions, compare logit and probit estimates. Each question requires only one or two sentences to answer.
    a) How different are the logit and probit unstandardized coefficients?
    b) Why are they so different?
    c) Why are the y-standardized coefficients similar but not exactly the same?
10. ___ of 10: My assessment of the overall effectiveness of your answers.