Logistic Regression Lab

Please briefly read the following to get a sense of the article's main empirical analysis:

> Houston, David J. 2005. "'Walking the Walk' of Public Service Motivation: Public Employees and Charitable Gifts of Time, Blood, and Money." *Journal of Public Administration Research and Theory* 16: 67–86. doi:10.1093/jopart/mui028

The data for this article come from the General Social Survey, a nationally representative survey of the United States population. These data, in Stata (.dta) format, are available to you on Blackboard in the `Content > GSS` folder, along with the output of Stata's `codebook` command in a separate .txt file.

1. Load the data file into Stata.

2. Start a new do-file to contain the following analyses.

3. Create dummy variables for each of the three outcomes:

   - Recode `volchrty` as a binary variable called `volbin`
   - Recode `givblood` as a binary variable called `bloodbin`
   - Recode `givchrty` as a binary variable called `charbin`

4. Generate a new variable that recodes the key causal variable `wrkgovt` to be 1 for public sector employees and 0 for private sector employees. Call this new variable `pubemp`.

5. Recode or generate covariates as necessary based on th discussion on p.74 of Houston (2005). You will need versions of the following variables:

   - Female: `sex`
   - White: `race`
   - Education: `educ`
   - Income: `rincom98`
   - Occupational prestige: `prestg80`
   - Married: `marital`
   - Age: `age`
   - Children in household: a sum of `babies`, `preteen`, and `teens`
   - Community size (logged): `size`
   - Church attendance: `attend`

6. Replicate the analyses in Tables 3, 4, and 5 of Houston (2005). Focus on Model # 1 in each table, using the `logit` command to perform logistic regression. It is used in the same way as `reg`.

7. Re-estimate the same models using `probit` instead of `logit` to retrieve probit regression results.

8. Use `eststo` and `esttab` to display the results of both models side-by-side.

9. Use `tab pubemp volbin, row` to see the proportions of public and private sector employees who volunteered time in the sample. Repeat this for the other outcomes `bloodbin` and `charbin`.

10. Estimate the predicted probability of each outcome for public and private sector employees using `margins, at(pubemp=(0 1))`. *Remember to call* ***`margins`*** *immediately after* ***`logit`****.*

11. You can simplify the `margins` call by treating `pubemp` as a factor when you call `logit`. You can then just type `margins pubemp`. Try it. Manually calculate the average marginal effect by calculating the difference between the two predicted probabilities.

12. Estimate the predicted probability of the outcome for public and private sector employees across levels of education using: `quietly margins pubemp, at(educ=(0 (1) 20))`.

13. Plot the previous result using `marginsplot`. Because predicted probabilities range from 0 to 1, it is best to specify the scale of the y-axis. Try it again using `marginsplot, ylabel(0 (0.1) 1)`.

14. Estimate the average marginal effect of being a public employee using `margins, dydx(pubemp)`. How does this compare to the manually calculated change in predicted probability you calculated earlier?

15. Compare the average marginal effect to the marginal effect at the means of the other covariates using: `margins, dydx(pubemp) atmeans`. How do they differ? Why?

16. Estimate the predicted probabilities of the outcome across different age levels and plot the results. You can use `quietly` to suppress the printing of the `margins` command before you plot. Try the following code:
    ```
    margins pubemp, at(age=(20 (5) 80))
    marginsplot
    ```
    How large does the marginal effect appear to be at each age level?

17. Now let Stata figure out the size of the marginal effects for you and plot the result. Try the following code:
    ```
    margins, dydx(pubemp) at(age=(20 (5) 80))
    marginsplot
    ```
    How do the last two graphs compare?