

GLM Interpretation 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 format, are available to you on Blackboard at **Materials from teachers > Additional Materials > Data > GSS2002.dta**.

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 the 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. How do they compare (both statistically and substantively)? Repeat all of the analyses below for logit and probit.
8. Divide the estimated coefficient on each variable from the probit model by the corresponding coefficient estimate from the logit model. How do these values compare?
9. Use `eststo` and `esttab` to display the results of both models side-by-side.
10. Repeat the process of displaying the logit and probit results, but this time create an output table that displays the average marginal effects of each variable rather than coefficient estimates. To do this, estimate the logit model, call `margins, dydx(*) post` where `dydx(*)` requests marginal effects and `post` request that those results overwrite the stored coefficients, then call `eststo`, repeat for the probit specification, and finally call `esttab`. How do the results compare?
11. 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`.
12. 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`.*
13. 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.
14. 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))`.
15. If you don't like the look of the resulting `marginsplot` figure, play around with the display options. You may want to use something like `marginsplot, recast(line) recastci(rarea)` to obtain a more attractive result.
16. Compare the empirical probabilities (i.e., the sample proportions) reported in Table 1 to predicted probabilities from the logit and probit specifications. What is the predicted probability of government and non-governmental employees engaging in each of the three types of voluntary behavior?
17. A common measure of model fit (albeit a problematic one) is the "percent correctly classified," which compares the proportion of observations observed as 1 to the proportion of observations predicted to be 1 (i.e., where $Pr(Y = 1) > 0.5$). Calculate this statistic for the various models.

18. 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)`.
19. 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?
20. Compare the average marginal effect to the marginal effect at the means of the other covariates: `margins, dydx(pubemp) atmeans`. How do they differ? Why?
21. 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:

```
quietly margins pubemp, at(age=(20 (5) 80))
marginsplot
```

How large does the marginal effect appear to be at each age level? If you add the `, yline(0)` option to `marginsplot`, you will see a horizontal line where the marginal effect is zero.
22. 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? Do you understand the distinction between the syntax used in the last two `margins` commands?
23. Introduce an interaction between `pubemp` and `educ` into the regression model. Now re-estimate the predicted probabilities and the marginal effect of `pubemp` across levels of education. Plot the results. What interpretations do you take away?
24. Practice obtaining predicted probabilities and marginal effects at representative values. Here, we will use the *stat* form of the `at()` option. Try obtaining predicted probabilities at various percentiles of age:

```
margins, at((p25) age)
margins, at((median) age)
margins, at((p75) age)
```

Then do the same for marginal effects of `pubemp`. You can also combine the *stat* form with the `var=(values)` form from above to specify interesting combinations of multiple variables. See `help margins` for more details.
25. Use `margins` to calculate various marginal effects at means (MEM), marginal effects at representative values (MER), and average marginal effects (AME) for both logit and probit models. Use `marginsplot` to produce additional interpretations and consider including additional variables, interactions, and non-linear terms in your model specification in order to understand the full functionality of `margins` and `marginsplot`.