PLSC 503: "Multivariate Analysis for Political Research"

Exercise Eight

Part I

For the simulation part of this exercise, consider a data-generating process for a latent variable Y^* of the form:

$$Y_i^* = 0 + 2X_{1i} - 4X_{2i} + 3X_{3i} + u_i \tag{1}$$

with

$$Y_i = \begin{cases} 0 \text{ if } Y_i^* \leq 0 \\ 1 \text{ if } Y_i^* > 0, \end{cases}$$

$X_1, X_2, X_3 \sim$ i.i.d. $U(0,1)$, and $u_i \sim$ Logistic$(0,1)$.[1] Using simulations:

1. Estimate the fraction of cases where the OLS regression

$$Y_i = \gamma_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i} + \gamma_3 X_{3i} + e_i$$

   yields predictions $\hat{Y}_i$ which are either less than zero or greater than one. What can you say about this fraction? Why is it what it is?

2. Illustrate, describe, and discuss the relationship between the estimated logit coefficients

$$\Pr(Y_i = 1) = \Lambda(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i})$$

   and the regression coefficients in the data-generating process in (1).

3. Reexamine the relationship in question 2, changing the data-generating process to one where $u_i \sim$ Logistic$(0,5)$. How does this change the nature of the relationship? Why? Why does it matter (or not)?

Part II

In their 2018 book Pruis or Pickup?, Hetherington and Weiler describe how and why in recent years political valences have crept into people's opinions about many things, including those that are not at all obviously political in nature. We're going to explore one such phenomenon.

As part of Washington University's American Panel Study (TAPS), researchers in July 2016 asked roughly 1300 U.S. survey respondents a series of questions. These included both conventional and unconventional survey items, including one question which asked:

Have you ever taken the shampoo and conditioner bottles from a hotel or motel?

- Yes
- No
- Not sure

---

[1]Note that here the second term is the "scale parameter" of the logistic distribution; the standard logistic distribution has a scale parameter of 1.0.

We'll assess the determinants of respondents' answers to this burning political question. The data we use here (PLSC503-2022-ExerciseEight.csv) code the answer to the above question (StealShampoo) as "Yes" = 1 and "No" = 0, with "Not sure" and nonresponses coded as missing. We'll examine the association between answers to this question and several demographic and political variables:

- Political party identification indicators – binary variables for Democrat and GOP, with independents serving as the reference category,

- Ideology – a seven-point indicator variable, where higher values indicate greater political conservatism,

- Education – measured as a twelve-category ordinal variable with values ranging from 3 to 15,

- Income – a 15-category ordinal indicator,

- The respondent's Age in years, as of 2016,

- Female – a binary indicator of sex, naturally-coded, and

- Racial classifications – indicator variables for White, Black, and Asian identification (with "other" as the reference category),

Your assignment is simple: Estimate a binary-response regression model (or more than one) of your choosing, where the response variable is StealShampoo. Feel free to specify the model as you see fit. Interpret your findings, using the techniques discussed in class. This assignment is due (electronically, obviously) at 11:59 p.m. EST on Friday, April 15, 2022, and is worth the usual 50 points.