Supplemental Materials to Lab 2 of QM3

Shane Xinyang Xuan
ShaneXuan.com

Department of Political Science
University of California, San Diego

April 11, 2018

1 When not using a LPM?

The linear probability model (LPM) implies that a ceteris paribus unit increase in x_j always changes $\mathbf{Pr}(y = 1|\mathbf{X})$ by the *same* amount, regardless of the value of x_j, and this assumption might not be sensible in many situations. Moreover, LPM can produce fitted values that are outside the unit interval $[0, 1]$.

2 Building blocks for Binary Choice Model

The logit function is:

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) \tag{1}$$

The inverse logit is:

$$\text{logit}^{-1}(\pi) = \frac{e^{\pi}}{1+e^{\pi}} = \frac{1}{1+e^{-\pi}} \tag{2}$$

3 Choice of the Link Function

The systematic binary choice model is:

$$y_i \propto \text{Bern}(y_i|\pi_i) \tag{3}$$
$$\pi_i = g(X_i\beta) \tag{4}$$

The choice of $g(\cdot)$ is arbitrary, but a common choice is logit^{-1}:

$$\pi_i = \text{logit}^{-1}(X_i\beta) \Leftrightarrow \text{logit}(\pi_i) = X_i\beta \tag{5}$$
$$= \frac{1}{1+e^{-X_i\beta}} \tag{6}$$

We might also consider a probit model, which is the *inverse of the Normal CDF*. Consider that we have a latent variable y^*:

$$y_i^* = X_i\beta + \varepsilon,\ \varepsilon \sim \mathcal{N}(0,1), \tag{7}$$

and we observe y:

$$y_i = \begin{cases} 1, \; y_i^* > 0, \\ 0, \text{ otherwise} \end{cases} \tag{8}$$

Then, the probability of observing $y_i = 1$ given X_i is:

$$\begin{aligned} \mathbf{Pr}(y_i = 1|X_i) &= \mathbf{Pr}(y_i^* > 0) &(9)\\ &= \mathbf{Pr}(X_i\beta + \varepsilon > 0) &(10)\\ &= \mathbf{Pr}(\varepsilon < X_i\beta) &(11)\\ &= \Phi(X_i\beta) &(12)\\ &\equiv \text{probit}^{-1}(X_i\beta) &(13) \end{aligned}$$

Remark The probit is the inverse of the Normal CDF and vice versa.

4 Interpretation

4.1 Marginal Effect

Now, let's briefly discuss ways in which to analyze the regression outputs. For example, we can compute the derivative of the logit at the central value, and differentiate the function $\text{logit}^{-1}(X_i\beta)$ with respect to x_k:

$$\frac{\partial \pi_i}{\partial x_{ki}} = \beta_k \frac{-e^{-X_i\beta}}{(1 + e^{-X_i\beta})^2} \tag{14}$$

Note that Equation (14) indicates that the change in the predicted outcome induced by a change in x_k depends *not only* on β_k, *but also* on the value of x_k and the values of all the other covariates in the model. That is, it is not very intuitive to interpret the results from logit models in terms of marginal effects.

4.2 Direct Interpretation

The interpretation of coefficients in a logit model is log odds. To see this, consider Eq (6). With basic transformation:

$$\begin{aligned} \pi_i + \pi_i e^{-X_i\beta} &= 1 &(15)\\ \frac{1-\pi_i}{\pi_i} &= e^{-X_i\beta} &(16)\\ \log\left(\frac{\pi_i}{1-\pi_i}\right) &= X_i\beta &(17) \end{aligned}$$

where $\log \frac{\pi_i}{1-\pi_i}$ is the "log odds." Hence, we may interpret coefficients in the following way: a 1-unit increase in x_j increases the log odds of y by β_j, which is quite unintuitive as well.

Remark To illustrate this point, suppose we run the following model in Stata: `logit honor math`, and the coeffient of `math` is 0.156, with an intercept of -9.794. Then, we can write

$$\begin{aligned}
y_i &\propto \text{Bern}(y_i|\pi_i) \\
\pi_i &= \text{logit}^{-1}(\beta_1 math + \beta_0) \\
\text{logit}(\pi) &= 0.156 math - 9.794 \\
&= \log \frac{\pi}{1-\pi} \\
\text{logit}(\pi)_{math54} &= 0.156(54) - 9.794 \\
\text{logit}(\pi)_{math55} &= 0.156(55) - 9.794 \\
\text{logit}(\pi)_{math55} - \text{logit}(\pi)_{math54} &= 0.156
\end{aligned}$$

That is, for a one-unit increase in `math`, the expected change in log odds is 0.156.

Remark Note that we can convert the change in log odds to change in odds, by exponentiating the log odds:

$$\begin{aligned}
\exp\left(\log \frac{\pi}{1-\pi}^{(math55)} - \log \frac{\pi}{1-\pi}^{(math54)}\right) &= e^{0.156} \\
\frac{\exp\left(\log \frac{\pi}{1-\pi}^{(math55)}\right)}{\exp\left(\log \frac{\pi}{1-\pi}^{(math54)}\right)} &= e^{0.156} \\
\frac{\text{odds}^{(math55)}}{\text{odds}^{(math54)}} &= e^{0.156} \approx 116.9\%
\end{aligned}$$

Thus, we say for a one-unit increase in `math`, we expect to see about $(116.9 - 100) = 16.9\%$ increase in the odds of being in an honors class.

Remark To summarize, this is how we interpret coefficients from a logit regression:

1.) Exponentiate the coefficient (β_k) – this is the *odds*

2.) For every one-unit change in x_k, the *change in odds* is $[100 \times (e^{\beta_k} - 1)]\%$

4.3 Another way to interpret logit coefficients

Another way to interpret logit coefficients is given by Gelman & Hill (2007, 82). As a first-order approximation we can use the fact that the logistic curve is steepest in the middle. Since the slope of the inverse logit is 0.25 at that point, dividing $\hat{\beta}_k$ (from logit) by 4 gives an estimate of the *maximum* difference a one-unit change in x_k can induce in the probability of a success. In the example above, we can say that a one-unit increase in `math` increases the probability of being in an honors class by at most $\frac{0.156}{4} = 3.9\%$.