**Week 5 Lab Handout- Limited Dependent Variables**

**PA 5033 – Multivariate Techniques**

April 16th, 2021

Wenchen Wang

**PART A: LINEAR PROBABLITY MODEL (LPM) ~10min**

**PART B: LOGIT MODEL ~15min**

**PART C: DUMMY VARIABLE MODEL ~10min**

**PART D: INSTRUMENTAL VARIABLE METHOD ~15min**

**Data Set:** We will be using the Problem Set 2, Data Set D. It is ‘**AgeGrp4\_20data.dta**’ on the Canvas site.

**Part A: Linear Probability Model:** **Fired = f (age, performance)**

Run an OLS regression of ***fired*** as a function of ***age*** and ***performance*.**



* *To interpret the predictions and calculate (average of the percentage of ones and zeros predicted correctly) for the linear probability model (lpm) use the following four steps:*

1. Calculate the predicted values for each individual (observation) being fired and recode the predicted values into 0 (if < 0.5) or 1 (if ≥ 0.5).



1. Count the number of individuals who were actually fired (fired=1) and our model predicted would be fired (problem=1). Additionally, count the number of individuals who were not fired (fired=0) and our model predicted would not be fired (fired=0).



1. Count the total number individuals who were actually fired (fired=1) and were actually not fired (fired=0).



1. Calculate :

For our “LPM” model plug in the data generated in steps (2) and (3) to calculate

**Part B: Logit Model:** **Fired = f (age, performance)**

****

**Marginal effects** are partial derivatives of the regression equation with respect to each variable in the model for each unit in the data; average marginal effects are simply the mean of these unit-specific partial derivatives over some sample. In ordinary least squares regression with no interactions or higher-order term, the estimated slope coefficients are marginal effects. In other cases and for generalized linear models, the coefficients are not marginal effects at least not on the scale of the response variable. **margins** therefore provides ways of calculating the marginal effects of variables to make these models more interpretable.

To analyze the impact of each coefficient, we need to calculate the **marginal effect on the average (Method 1)** and the **average marginal effect (Method 4)**.

1. Marginal effect on the average

***margins, dydx (age perf) atmeans***

**

*How do we interpret the impacts of age and performance on the probability of being fired with marginal effects at the average?*

1. Average Marginal Effect

***margins, dydx (age perf)***

**

* *How do you interpret the impacts of age and performance on the probability of being fired using average marginal effects?*
* *Calculate  for this logit model by using the following two steps:*

1. Generate a model prediction table for this model using the **“lstat”** command.



1. Calculate using the guidelines to interpreting the “lstat table” on the following page.

**Interpreting an “lstat” table:**

|  |  |  |
| --- | --- | --- |
| Classified | -------True---------------------------  D ~D | Total |
| +  - | [A] [B]  [C] [D] | [E]  [F] |
| Total | [G] [H] | [I] |

**A: Predicted fired *and* actually fired (i.e., Predicted 1 and actual observation is 1)**

B: Predicted fired *but not* actually fired (i.e., Predicted 1 but actual observation is 0)

C: Predicted not fired *but* actually fired (i.e., Predicted 0 but actual observation is 1)

**D: Predicted not fired *and* actually not fired (i.e., Predicted 0 and actual observation is 0)**

E: Total predicted fired (i.e., Total predicted 1s by the model)

F: Total predicted not fired (i.e., Total predicted 0s by the model)

**G: Total actually fired (i.e, Total actual 1s in the dataset)**

**H: Total actually not fired (i.e, Total actual 0s in the dataset)**

I: Total sample size

Therefore, we calculate the following replacement for R^2:

* *Generate a graphical representation for your findings:*

1. Calculate the predicted values for each individual (observation) being fired and recode the predicted values into zero (if < 0.5) or one (if ≥ 0.5).



1. Create a scatterplot of how well our model predicted being fired and not fired by age and performance

**twoway (scatter age perf) , by(problogitr fired)**



* *How do your logit results compare to the LPM estimates?*

**Part C: Over or Under 50 Age Dummy Variable**:

Now change the ***age*** variable to be consistent with the current practical legal requirements, that real “discrimination begins at fifty” in the following way.



Rerun the logit equation from part 2 using the new variable “***newage***”: **Fired = f (newage, performance)**



To analyze the impact of each coefficient, calculate the **marginal effect on the average** (Method 1). Note: you could also calculate the **average marginal effects** (Method 4).

Marginal effect on the average

***margins, dydx (age perf) atmeans***



* *How do you interpret the impacts of newage and performance on the probability of being fired using marginal effects on the average?*
* *Generate a model prediction table using* *the* ***“lstat”*** *command and calculate using the same approach as in the previous model*



* *What do your results show?*

**Part D: Instrumental Variable of *Performance Cleansed of Age*:**

To address possible simultaneity issues between ***age*** and ***performance***, we will create a new instrumental variable. Run the ***linear*** regression **Performance = f (age)** and save the unstandardized residuals. The residual (res\_1) functions as a performance variable cleansed of age.



Run the following ***logistic*** regressions:

1. **Fired = f (age, res\_1)**

****

* To analyze the impact of age and performance (cleansed of age), calculate the **marginal effect on the average** (Method 1).

Marginal effect on the average

***margins, dydx (age res\_1) atmeans***

**

* *How do you interpret the impacts of age and res\_1 (performance cleansed of age) on the probability of being fired using marginal effects on the average?*
* *Generate a model prediction table using* *the* ***“lstat”*** *command and calculateusing the same approach as in the previous model*



* *How does this model compare with the LPM and the previous logit model with age and performance as independent variables?*

1. **Fired = f (newage, res\_1)**

****

* *Generate a model prediction table using* *the* ***“lstat”*** *command and calculateusing the same approach as in the previous model.*

****

* *What do these results show and what do they mean?*