University of Canterbury | Professor Bob Reed
ECON 213 | Second Semester 2009

WEEK #11 ASSIGNMENT

***DUE*: Monday, October 5th**

***NOTE*: Be sure to read the note at the end of the questions.**

Do the following exercises:

1A. Read Chapter 12 through Section 12-1 in the textbook, pages 257-270.

1B. Answer Questions #1-9.

2A. Download the Excel spreadsheet file "Voting" from the course website: "*http://www.econ.canterbury.ac.nz/personal_pages/bob_reed/Courses/ECON_213/Week11/index.html*" and save it into your WEEK11 folder. Confirm that the data match those in TABLE 12-A on pages 262-263 in the textbook.

Create a new workfile, calling it "Voting", and import the Excel data into it.

Estimate the following regression model using OLS (cf. Equation 2-6 on page 261 in the textbook):

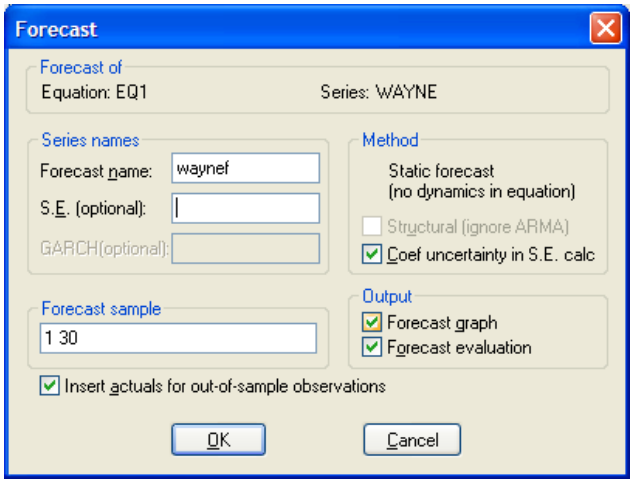
$WAYNE = B_0 + B_1\ INCOME + B_2\ AGE + B_3\ MALE + e$.

Compare your regression results with those in TABLE 12-B on page 263 of the textbook. They should be identical. Name this equation "EQ1".

2B. Answer Questions #10-13.

The following set of exercises demonstrates how to use EViews to obtain "fitted" (predicted) values for all the observations.

3A. Open up the "Equation" window for EQ1. Click on the "Forecast" button. The following "Forecast" window will appear:



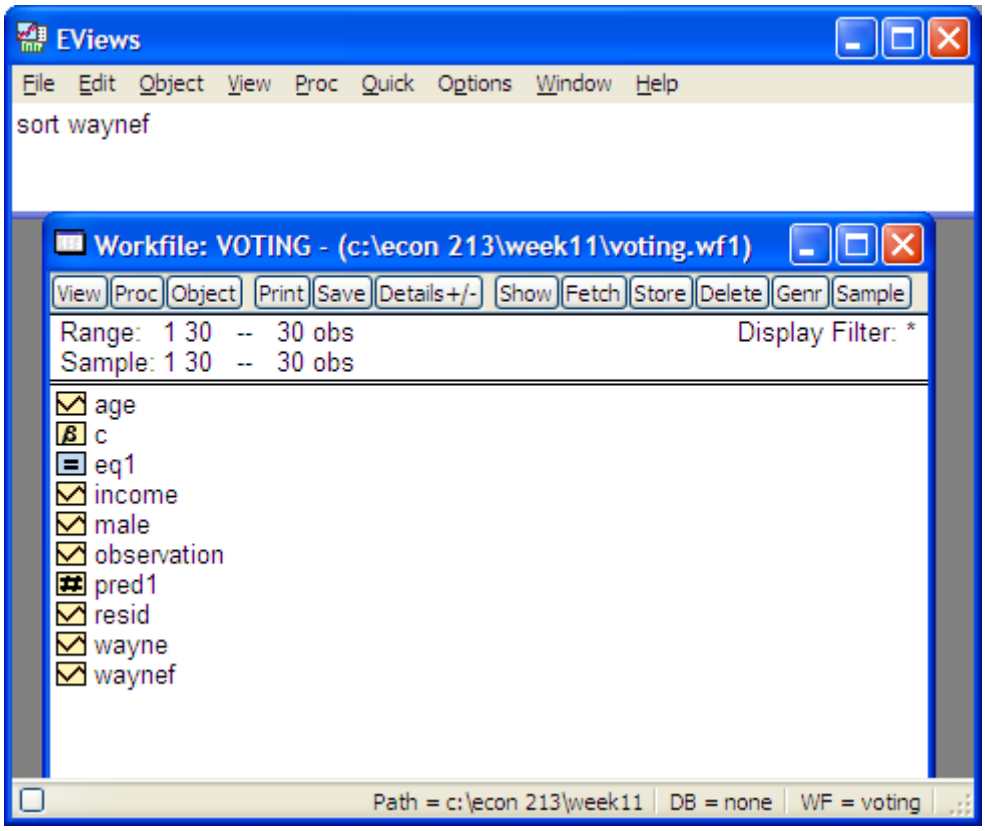
Note that the "Forecast sample" lists observations 1 through 30; in other words, it lists the entire sample used for the regression. Click "OK."

A new variable, "waynef", will appear in your workfile. This variable consists of the "fitted" values for the dependent variable "WAYNE." Inspect the variable "waynef" and confirm that these "fitted" values are the same as those reported in TABLE 12-C on pages 264-265 of the text. Then close the "Equation" window.

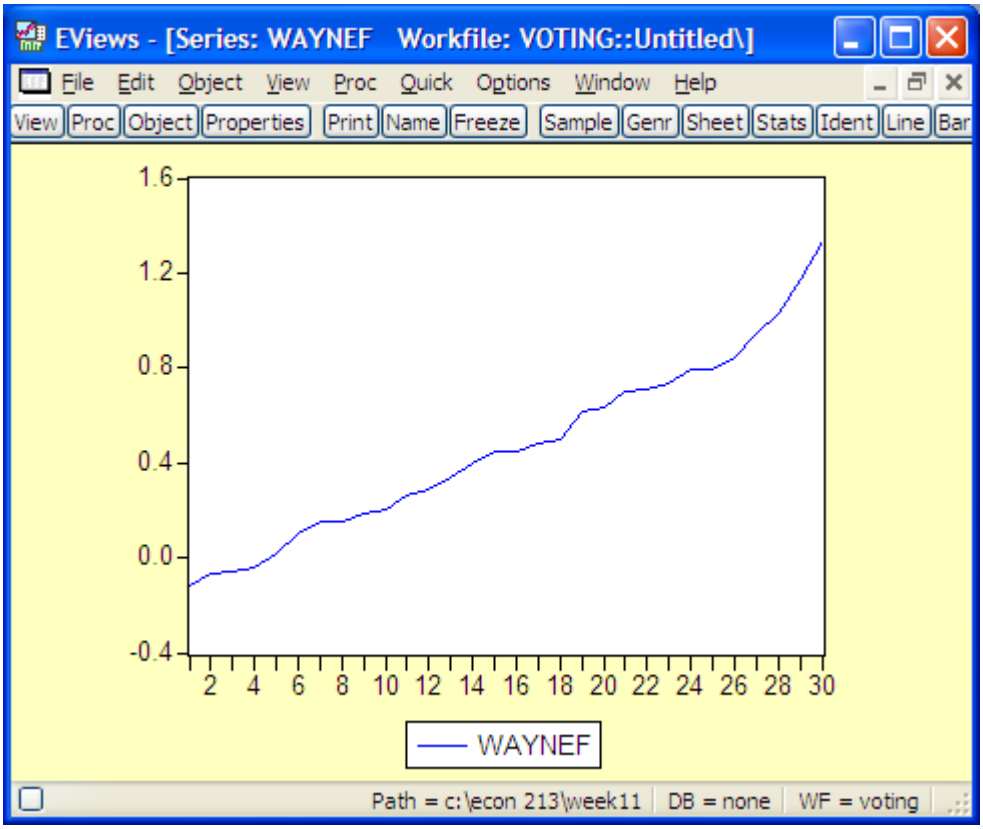
3B. In this exercise we graph the "fitted" values of WAYNE. For reasons that will become clear, we first sort the data in ascending order of the variable "waynef". To do that, go to the white work area below the main EViews taskbar and type the following command:

```
sort waynef
```

(See below.) Then hit the "Enter" key.



Next, obtain a graph of the series "waynef" using the same procedure you used in the Week #9 Assignment. Your graph should look like the one below.



To sort the data back in the original order, type

```
sort observation
```

and then hit the "Enter" key.

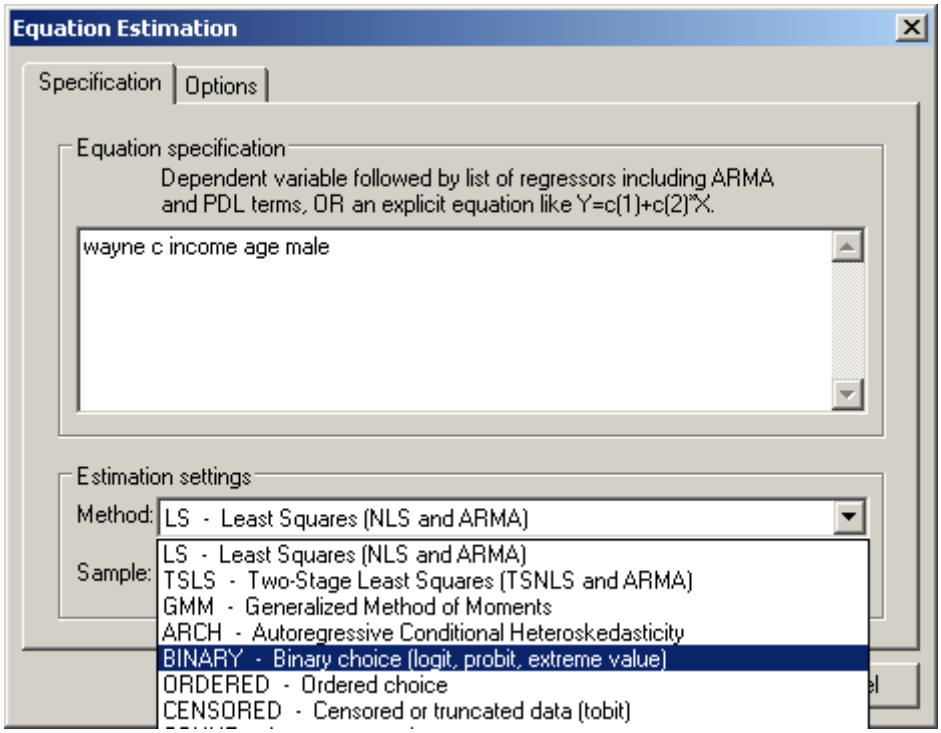
3C. Answer Question #14.

The next set of exercises demonstrates how to use EViews to estimate a Probit model (cf. pages 266-268 in the text).

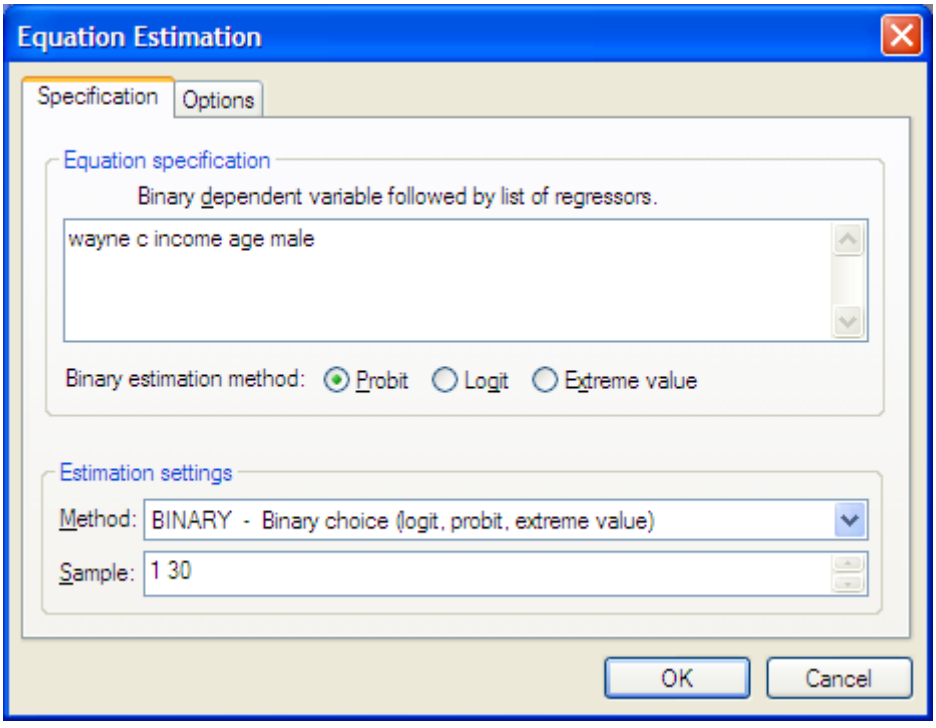
4A. Open up a new "Equation Estimation" window using the "**Quick / Estimate Equation…**" options from the main EViews taskbar. We will once again estimate the regression model

$$WAYNE = B_0 + B_1\ INCOME + B_2\ AGE + B_3\ MALE + e\ .$$

Type the appropriate set of variables into the "Equation specification" workspace. However, DO NOT click the "OK" button as you usually would in estimating an OLS regression. Instead, choose "BINARY – Binary choice (logit, probit, extreme value)" from the drop down menu by "Method" under "Estimation settings" (see below).



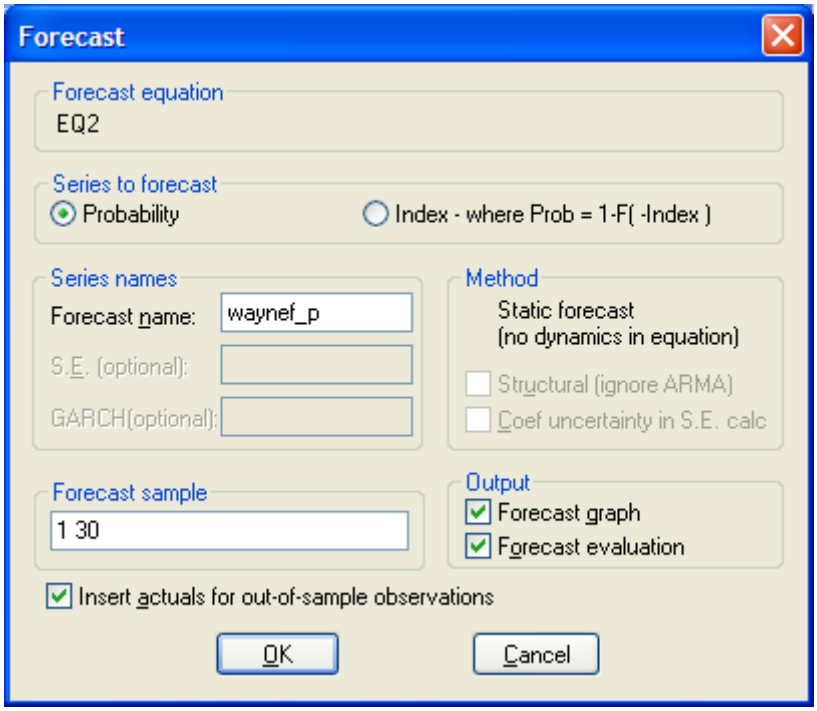
Then click "OK." The "Equation Estimation" window will change to look like the following:



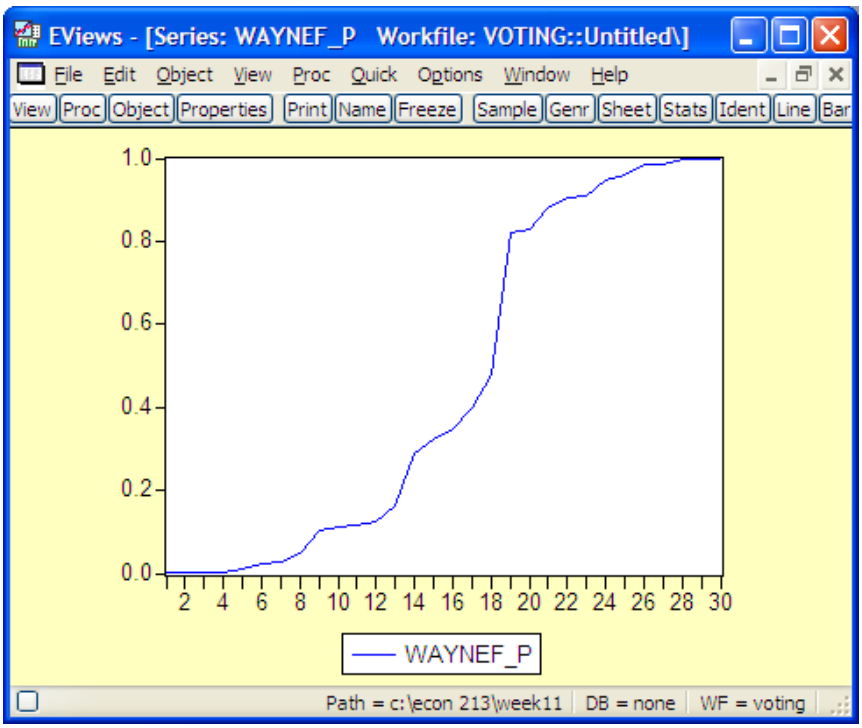
Make sure that "Probit" is selected by "Binary estimation method" (see above) and then click "OK".

Compare your regression results with those in TABLE 12-D on page 267 of the text and confirm that they are identical. Name your equation "EQ2".

4B. Use the Probit estimates to get "fitted" values for each of the observations. To do that, repeat the analysis of Exercises #3A-3B above, except name the forecasted variable "waynef_p" (for Probit). When the "Forecast" window appears, make sure that you select the "Probability" option (see below).



Sort your data in ascending order of the "waynef_p" variable, and then graph the corresponding fitted values. You should obtain the following graph:



4C. Answer Question #15.

5A. Question #11 below asked you to use the estimated Linear Probability model to predict the value of the dependent variable (WAYNE) for a 35 year-old male who has an income of thirty thousand dollars. In this exercise, you will do the same thing using the Probit estimates of EQ2.

To do that, follow the same procedure you used in getting your predicted value for Question #11. As you proceed, you will notice that the "Forecasting Equation" representation for EQ2 has an unusual specification. It reads as follows:

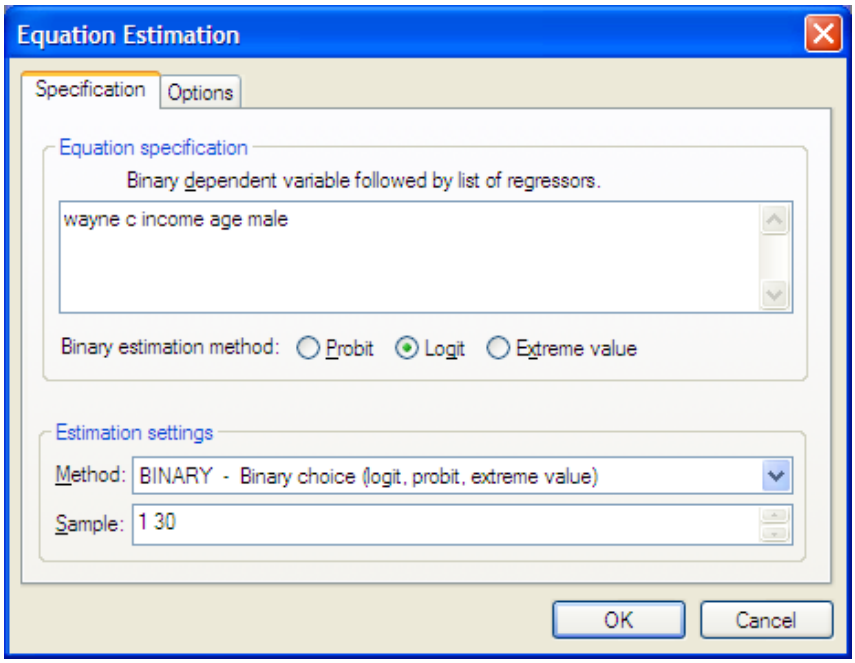
```
WAYNE = 1-@CNORM(-(C(1) + C(2)*INCOME + C(3)*AGE + C(4)*MALE))
```

Copy this into the white workspace area under the main Eview taskbar. Calculate the scalar predicted value in the usual fashion -- by substituting the appropriate values for INCOME, AGE, and MALE.

5B. Answer Questions #16-24.

The next set of exercises demonstrates how to use EViews to estimate a Logit model (cf. pages 268-270 in the text).

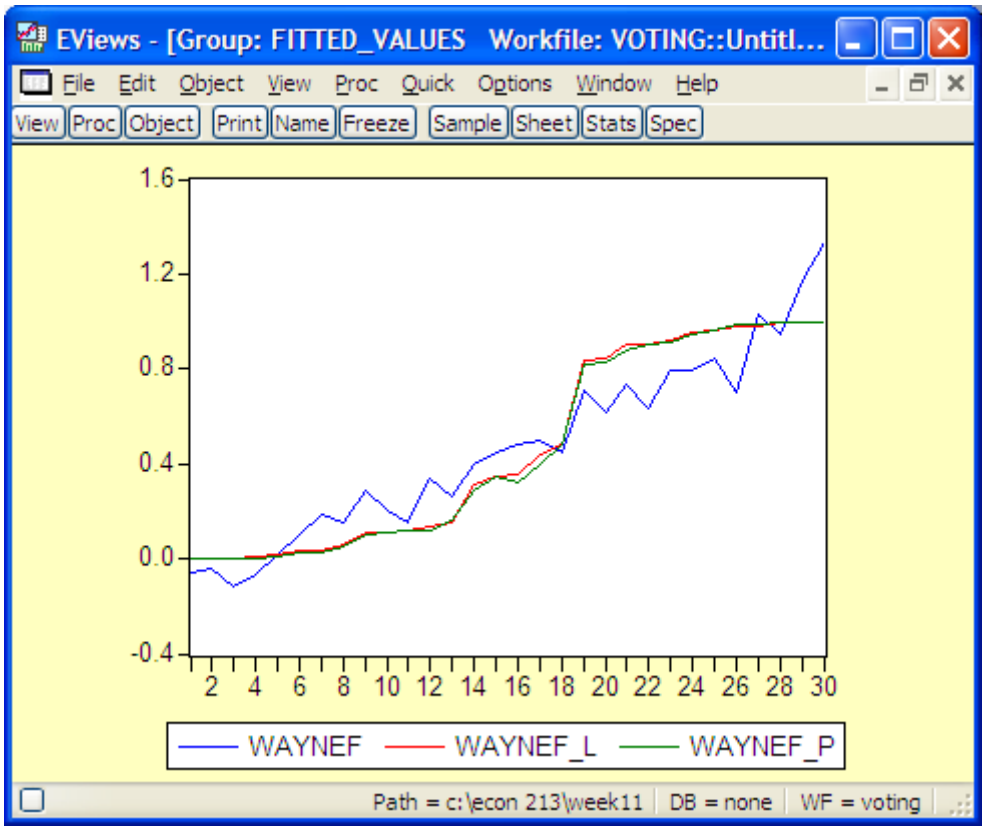
6A. Repeat the procedure you used to estimate a Probit model (cf. Exercise #4A above). When you get to the "Equation Specfication" window, choose the "Logit" option (see below):



Click "OK". Name your equation "EQ3". Compare your results with those in TABLE 12-E on page 270 in the text.

6B. Follow the same procedure you used in Exercise #4B above to produce a graph of "fitted" values for the Logit model. Call the forecasted series "waynef_L" (for Logit).

6C. Sort the data into ascending order of "waynef_L" values. Select the three "fitted" value series and produce the following graph of the three series:



6D. Answer Questions #25-27.

<u>NOTE</u>: The textbook discusses the problem with standard goodness-of-fit tests when estimating binary choice models. The following example will illustrate this: Suppose you have 100 observations of voting behaviour: 10 vote for WAYNE, 90 do not. Suppose your model predicts that nobody votes for WAYNE. Then you would get a "pseudo-R^2" of 90%. This would suggest a good fit, when the truth is quite the opposite! Accordingly, your textbook suggests (cf. page 264f.) that you separately calculate the percentage correct predictions for each of the binary outcomes. In this case, you would discover that while the model predicted 100% of those who did not vote for WAYNE, it predicted 0% correctly for those who did vote for him. While it is not difficult to get Eviews to produce these statistics for you, this information is not automatically calculated. Because it requires additional commands, I have chosen to omit this topic.

Be prepared to answer the following questions in class:

1. (Multiple choice: Select the single best answer.)
There are a variety of situations where you may want to build a regression model around a dependent variable that is not a number; the dependent variable may describe a…

A) Characteristic.
B) Choice.
C) Category.
D) All of the above.

2. (Multiple choice: Select the single best answer.)
The following are examples of regression models that would require a dummy dependent variable:

A) Whether somebody takes the bus or drives a car to work.
B) Whether a Commerce student decides to be an economics major or chooses some other major.
C) Whether a voter votes for the Labour candidate for prime minister or some other candidate.
D) All of the above.

3. (Multiple choice: Select the single best answer.)
Regression models where the dependent variable takes the values 0 or 1 are called "binary choice models."

A) True
B) False

4. (Multiple choice: Select the single best answer.)
The "Linear Probability" model looks like a standard OLS regression model.

A) True
B) False

5. (Multiple choice: Select the single best answer.)
Consider the following estimated regression equation (cf. Equation 12-3 on page 258 of the textbook):
GRADSCHOOL = -0.7 + 0.4 GPA + 0.002 INCOME.
The estimated coefficient for GPA, 0.4, can be interpreted as meaning that students who go to graduate school have GPA's that are 40% higher than those who don't.

A) True
B) False

6. (Multiple choice: Select the single best answer.)
Consider the following estimated regression equation (cf. Equation 12-3 on page 258 of the textbook):
GRADSCHOOL = -0.7 + 0.4 GPA + 0.002 INCOME.
The estimated coefficient for GPA, 0.4, can be interpreted as follows: Suppose we have two groups of students that are identical in every respect except the first group of students have GPA's that are one point higher than the second group. The probability of students in the first group going to graduate school is 40% higher than the probability of students in the second group.

A) True
B) False

7. (Multiple choice: Select the single best answer.)
Consider the following estimated regression equation (cf. Equation 12-3 on page 258 of the textbook):
GRADSCHOOL = -0.7 + 0.4 GPA + 0.002 INCOME.
The predicted value of GRADSCHOOL for a student with a GPA of 3.0 and an annual family income of 50 thousand dollars is:

A) 0.85
B) 0.70
C) 0.60
D) 0.55

8. (Multiple choice: Select the single best answer.)
Consider the following estimated regression equation (cf. Equation 12-3 on page 258 of the textbook):
GRADSCHOOL = -0.7 + 0.4 GPA + 0.002 INCOME.
You should be able to confirm that the predicted value of GRADSCHOOL for a student with a GPA of 3.0 and an annual family income of 50 thousand dollars is 0.60. This value can be interpreted as follows: It tells us that probability that this student will go to graduate school is estimated to be 60%.

A) True
B) False

9. (Multiple choice: Select the single best answer.)
Which of the following are problems with the Linear Probability model?

A) The Linear Probability model can produce estimated probabilities that are less than 0 and larger than 1.
B) The error terms in the Linear Probability model suffer from heteroscedasticity.
C) R^2 is no longer a useful "goodness-of-fit" measure.
D) All of the above.

10. (Provide a short answer in the space below.)
The estimated coefficient for AGE in the Linear Probability regression model of Exercise #2A above is 0.016. In plain English, briefly explain what this number means.

11. (Multiple choice: Select the single best answer.)
Use the regression results from the estimated Linear Probability model of Exercise #2A to predict the value of the dependent variable (WAYNE) for a 35 year-old male who has an income of thirty thousand dollars. Select the best answer below.
(HINT: You may want to review the Week #4 Assignment for how to use EViews to obtain predicted values.)

A) 0.46
B) 0.36
C) 0.26
D) 0.16

12. (Provide a short answer in the space below.)
In plain English, briefly interpret the predicted value for WAYNE in the question above. What does that number mean?

13. (Fill in the required information in the "Hypothesis Test Table" below.)
Consider the regression results from Exercise #2A above. A researcher is studying the following statement: Age does not affect voting behaviour. Complete the corresponding "Hypothesis Test Table."
(HINT: You may want to review the Week #3 Assignment for how to conduct hypothesis tests.)

1) H_0: H_A:
2) LEVEL OF STATISTICAL SIGNIFICANCE =
3) STATISTICS:
4) DECISION =

14. (Provide a short answer in the space below.)
Do you notice anything unusual about the graph in Exercise #3B above? What is it?

15. (Provide a short answer in the space below.)
Compare the graph of "fitted" Probit values from Exercise #4B above with the graph of "fitted" Linear Probability values from Exercise #3B. Which set appears to be more reasonable? Why?

16. (Supply the correct numerical answer on the line below.)
Use the regression results from the estimated Probit model of Exercise #4A to predict the value of the dependent variable (WAYNE) for a 35 year-old male who has an income of thirty thousand dollars. Select the best answer below.

A) 0.42
B) 0.32
C) 0.22
D) 0.12

17. (Provide a short answer in the space below.)
In plain English, briefly interpret the predicted value for WAYNE in the question above. What does that number mean?

18. (Multiple choice: Select the single best answer.)
Compare your predicted values using the Linear Probability (Question #11) and Probit (Question #16) models. Are the predicted values approximately the same?

A) Yes.
B) No.

19. (Multiple choice: Select the single best answer.)
Use the regression results from the estimated Probit model (Exercise #4A) to predict the probability that an "average person" would vote for WAYNE. An "average person" is defined as a person with the average values for all variables (INCOME, AGE, and MALE). Select the best answer below.
(HINT: You may find it helpful to review the Week #5 Assignment for how to use EViews to obtain average values for calculating predictions.)

A) 0.49
B) 0.39
C) 0.29
D) 0.19

20. (Multiple choice: Select the single best answer.)
Repeat the analysis of the previous problem, only do it for a person who is one year older. Select the best answer below.

A) 0.52
B) 0.42
C) 0.32
D) 0.22

21. (Multiple choice: Select the single best answer.)
Use the regression results from the estimated Probit model (Exercise #4A) to predict the probability that a 45 year-old male with an income of a sixty thousand dollars would for WAYNE. Select the best answer below.

A) 0.95
B) 0.85
C) 0.75
D) 0.65

22. (Multiple choice: Select the single best answer.)
Repeat the analysis of the previous problem, only do it for a person who is one year older. Select the best answer below.

A) 0.96
B) 0.86
C) 0.76
D) 0.66

23. (Multiple choice: Select the single best answer.)
Compare your answers to Questions #20 and 22. Is the effect of a one-year increase in AGE the same?

A) Yes.
B) No.

24. (Multiple choice: Select the single best answer.)
The estimated coefficient for AGE in the Probit regression model of Exercise #4A above is 0.073. Do you see an obvious relationship between this number, and the estimated effects of a one-year increase in AGE that you calculated in Questions #20 and 22?

A) Yes.
B) No.

25. (Multiple choice: Select the single best answer.)
Compare the estimated coefficient values for the Probit (Exercise #4A) and Logit (Exercise #6A) models. Are the estimated values virtually identical?

A) Yes.
B) No.

26. (Multiple choice: Select the single best answer.)
Compare the graphs of the fitted values for the Probit and Logit models (cf Exericse #6C above). Are the fitted values from these two models virtually identical?

A) Yes.
B) No.

27. (Fill in the required information in the "Hypothesis Test Table" below.)
Consider the Logit regression results from Exercise #6A above. A researcher is studying the following statement: Women are more likely to vote for WAYNE than men are. Complete the corresponding "Hypothesis Test Table."

1) H_0: H_A:
2) LEVEL OF STATISTICAL SIGNIFICANCE =
3) STATISTICS:
4) DECISION =

<u>NOTE</u>: Once again, I have assembled some additional questions to help you review material from previous chapters. We will do these questions on Tuesday.