

Fall 2025, MLE GVPT 729a Problem Set 1

NAME: Mya Zepp

Please read all of the questions and tables carefully. Be as brief as possible; extra calculations and incorrect information, even in the presence of correct information, will result in point deductions. If you need to make any assumptions, be sure to make them explicit. Show all formulas and all work.

1. Imagine you hypothesize that the probability of supporting a law to lower the voting age to 16 is lower for people age 65 and older than it is for people 30-64 years old. Using the table of survey results below answer questions 1a, 1b, and 1c. Please show all of your work.

Table 1. Probability of Supporting a Law to Lower the Voting Age to 16 by Age (cell counts)

Age	Oppose	Support	Total
18-29	75	125	200
30-64	250	250	500
65 and older	200	100	300
Total	525	475	1000

- 1a. What is the probability of supporting the law? (1 point)**

$$475/1000 = 0.475$$

47.5%

- 1b. What is the probability of being 18-29 or supporting the law? Write the general (i.e. using symbols) formula you used to obtain your answer and show all of your work. (2 points)**

Probability of supporting the law = 47.5%

Probability of being 18-29 = 200/1000 = 20%

$$0.475 + 0.2 = 0.675$$

67.5%

- 1c. In terms of the substantive results (ignore statistical significance), does the evidence support your hypothesis? Explain, using details, how you know (no more than 2 sentences). (5 points)**

$$300/1000 = 0.3$$

$$500/1000 = 0.5$$

$$475/1000 = 0.475$$

$$0.3 * 0.475 = 0.1425 = 14.25$$

$$0.5 * 0.475 = 0.2375 = 23.75$$

The evidence supports my hypothesis because the probability that someone is 30-64 years old and supports the law is 23.75% while the probability that someone is 65 and older and supports the law is 14.25. This means that people age 30-64 are more likely to say they support changing the law to lower the voting age.

2. Provide a political science example of an ordered categorical variable. (1 point)

A scale of political identity (left-leaning, moderate, right-leaning) from 1 to 7 is a ordered categorical variable.

3. Which of the following is NOT a feature of the linear probability model when run on binary dependent variables? Circle the letter and the description. (2 points)

- a) possibility of predicted probabilities greater than 1 or less than 0
- b) homoscedasticity
- c) non-normality of the disturbances
- d) R^2 values that tend to be low
- e) marginal effects that are constant across the range of the X

4. Write out the equation for the effect of X_2 on Y from the following OLS model:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{23} X_2 * X_3 + \beta_4 X_4 + u. \quad (3 \text{ points})$$

$$\frac{\partial Y}{\partial X_2} = \beta_2 + \beta_{23} X_3$$

5. Based on the information below, answer Q5a-Q5d.

Dependent variable: European Parliament Feeling Thermometer, ranging from 0 (coldest) to 100 (warmest)

Independent variable: Female, coded 1 for female and 0 for other; age 16-90.

The following results were obtained from a survey of 1,500 citizens age 16 to 90 in European Union member nations.

<u>OLS Regression Results Q5</u>		
<u>Independent Variables</u>	<u>Coefficient</u>	<u>p value</u>
Constant	49	0.000
Female	-7	0.000
Age	0.2	0.000

5a. Determine whether the following statement is True or False and fully explain why, in no more than 1 sentence: The interpretation of the coefficient on the constant term is useful on its own. (2 points)

False

5b. Using the information above, what is the substantive meaning of the coefficient on the Female variable? Answer in no more than 1 sentence and be sure to tailor the answer to the variables used in this example. (2 points)

People who have more positive feelings toward the European Parliament are seven percentage points less likely to identify as female compared to those who have less positive feelings toward the European Parliament on average.

5c. True or False. Using the information above, because the coefficient on Female is larger than the coefficient on Age, the effect of being female on the European Parliament Feeling Thermometer is larger than the effect of age. Just state True or False. (1 point)

False

5d. Using the information above, might the positive coefficient on Age be the result of older people being more knowledgeable about the European Parliament? In no more than 1 sentence, explain your answer. (3 points)

Yes, it may be a result of being more knowledgeable about the European Parliament in addition to having voted for longer compared to younger people.

6. Answer the questions below based on the following information and the following table.

Data: 2020, individual level U.S. survey data

Method: Probit

Dependent Variable: Voted or not, coded as 1 for voted and 0 for did not vote

Independent Variables:

Age (coded 18-92);

Education (coded in 5 categories (1 = No high school degree, 5 = Graduate Degree));

South (state is in the south and was part of the confederacy or not, 1=state is in the south, 0=state is not in the south).

Table 3. Probability of Voting Results for Q6

Independent Variables	Coefficient	Std. Error	Z value	p value
Constant	-3.79	0.04	-94.75	0.000
Age	0.04	0.00	100.00	0.000
Education	0.48	0.23	2.05	0.040
South	-0.40	0.16	1.88	0.030

N = 98,857

Log likelihood = -56,089.376

Chi Sq = 13,196

p value = 0.000

a) Interpret the coefficient on the Education variable (3 points)

An increase in education is associated with a 0.48 percentage point increase in the likelihood of having voted on average.

b) At what level can the hypothesis that the coefficient on education is less than or equal to zero be rejected? State the exact level of significance. (1 point)

p < 0.05