# Econometrics II, ECO351, Semester I, 2025-26
# Homework I (100 points)

Instructor: M.A. Rahman

Deadline: 4:00 pm, August 29, 2025 (Room 409, ESB-2).

**Please read the instructions carefully and follow them while writing answers.**

- *Solutions to homework should be written in A4 size loose sheets. If you are not comfortable writing on white sheets, please ask for biology paper in Tarun Book Store.*
- *Questions should be answered in order as they appear in the homework. Every new question should begin in a new page. Please number all the pages of your homework solution. Please leave a margin of one inch from top and one inch from left. Staple the sheets on the top-left.*
- *Please submit computational assignments (if any) and written answers together and in the correct order. Your answer script should directly address the questions, with all code included in the appendix. All questions in this assignment must be completed using MATLAB. Work done in any other software will NOT be accepted.*

---

**1. (2+4+4+10 = 20 points)** The data in the file '`igg.xlsx`' refers to the serum concentration (grams per litre) of immunoglobulin-G (IgG) in 298 children aged from 6 months to 6 years.

$$\begin{aligned} \text{Model 1} \quad y_i &= \beta_1 + \beta_2 x + u_i, \\ \text{Model 2} \quad y_i &= \beta_1 + \beta_2 x + \beta_3 x_i^2 + \varepsilon_i, \end{aligned}$$

where $y_i = IgG$ and $x_i = age$ for the $i$-th individual. Using the data, answer the following.

(a) Give a descriptive summary (mean, median, standard deviation, maximum and minimum) of $IgG$ and $age$. Plot a histogram of both the variables. Comment on the presence/absence of skewness in histogram.

(b) Estimate Model 1 using ordinary least squares. Report the estimates, standard errors, $t$-stats, and $R^2$. Comment on the regression relationship.

(c) Estimate Model 2 using ordinary least squares. Report the estimates, standard errors, $t$-stats, and $R^2$. Comment on the regression relationship.

(d) Now assume the errors $\varepsilon$ follow a standard normal distribution. Write down the log-likelihood function for Model 2. Estimate the model using the method of maximum likelihood. Report the ML estimates and the standard errors. (You need to code the Newton-Raphson method to maximize (minimize) the log-likelihood (negative of log-likelihood).

**2. (2×10 = 20 points)** In a recent survey conducted between March 13th to March 17th, 2013 by the Pew Research Center for the People & the Press by Abt SRBI (Schulman, Ronca & Buculvas, Inc.), individuals were interviewed over the phone (landline and cell phone) about their opinion on 'Marijuana Legalization' in the United States. Overall, a sample of 1,501 adult individuals were interviewed and their responses recorded. However, after cleaning the data, we have a sample of 1240 observation that is attached in the file '`marijuana.xlsx`'. Based on the data, do the following.

Let's prepare the data for fitting a probit and a logit model.

(a) Recode the response variable $q85$ as $y_i = 1$ if 'Yes, legal' and $y_i = 0$ if 'No, illegal'.

Provide the frequency counts and percentages for each category of $y$.

(b) Label the variable 'age' as $x_2$ and 'hh1' (number of members in the household) as $x_3$.

Provide the mean and standard deviation for 'age' and 'hh1'.

(c) Recode the variable 'past use' as 1 if 'Yes' and 0 if 'No'. Label the recoded variable as $x_4$. The variable 'past use' tells if the individual has used marijuana in the past.

Provide the frequency counts and percentages for each category of 'past use'.

(d) Recode the variable 'sex' as $sex_i = 1$ if male and $sex_i = 0$ if female. Label the recoded sex variable as $x_5$. Similarly recode the variable 'parent' as $parent_i = 1$ if 'Yes' and $parent_i = 0$ if 'No'. Label the recoded parent variable as $x_6$.

Provide the frequency counts and percentages for each category of 'sex' and 'parent'.

(e) Lets work on marital status. Generate three categories as follows. Category 1 ('single') consists of 'Never been married', Category 2 ('post-married') consists of 'Divorced', ''Separated' or 'Widowed' and Category 3 ('couple') consists of 'Married' or 'Living with a partner'. Create a variable $x_7$ which takes the value 1 if an individual belongs to the category 'single' and 0 otherwise. Similarly, create another variable $x_8$ which takes the value 1 if an individual belongs to the category 'post-married' and 0 otherwise. We will use the category 'couple' as the base category and not include in the regression.

Provide the frequency counts and percentages of 'single', 'post-married', and 'couple'.

(f) Lets work on the column 'income'. Create the category 'poor' by merging 'Less than 10000', '10 to under 20000', '20 to under 30000', '30 to under 40000' and '40 to under 50000'. Similarly, create the category 'middle' by merging '50 to under 75000' and '75 to under 100000'. Finally, create the category 'rich' by merging '100 to under 150000' and '150000 or more'. Now, create a

variable $x_9$ which takes the value 1 if the individual belong the category 'poor' and 0 otherwise. Similarly, create another variable $x_{10}$ which takes the value 1 if an individual belongs to the category 'middle' and 0 otherwise. We will use the category 'rich' as the base category and not include in the regression.

Provide the frequency counts and percentages of 'poor', 'middle', and 'rich'.

(g) Lets work on the column 'educ'. Create the category 'HSandBelow' by merging 'Less than HS', 'HS Incomplete' and 'HS'. Similarly, create the category 'lessThanBachelors' by merging 'Some college' and 'Associate Degree'. Finally create the category 'BachelorsandAbove' by merging 'Bachelors', 'Postgraduate Degree' and 'Some Postgraduate'. Now, create a variable $x_{11}$ which takes the value 1 if the individual belong the category 'HSandBelow' and 0 otherwise. Similarly, create another variable $x_{12}$ which takes the value 1 if an individual belongs to the category 'lessThanBachelors' and 0 otherwise. We will use the category 'BachelorsandAbove' as the base category and not include in the regression.

Provide the frequency counts and percentages of 'HSandBelow', 'lessThanBachelors', and 'BachelorsandAbove'.

(h) Lets work on the column 'race'. Create the category 'white' for white individuals, category 'black' for black individuals and category 'allOther' by merging the remaining races. Now, create a variable $x_{13}$ which takes the value 1 if the individual belong the category 'white' and 0 otherwise. Similarly, create another variable $x_{14}$ which takes the value 1 if an individual belongs to the category 'black' and 0 otherwise. We will use the category 'allOther' as the base category and not include in the regression.

Provide the frequency counts and percentages of 'white', 'black', and 'allOther'.

(i) Lets work on the column 'party'. Create the category 'democrat' for individuals whose party affiliation is 'Democrat', category 'republican' for individuals whose party affiliation is 'Republican' and category 'independentOthers' by merging the remaining party affiliations. Now, create a variable $x_{15}$ which takes the value 1 if the individual belongs the category 'democrat' and 0 otherwise. Similarly, create another variable $x_{16}$ which takes the value 1 if an individual belongs to the category 'republican' and 0 otherwise. We will use the category 'independentOthers' as the base category and not include in the regression.

Provide the frequency counts and percentages of 'democrat', 'republican' and 'independentOthers'.

**Enough of text mining, lets get to real work!**

**3. (4+8+4+4 = 20 points)** Consider the data from Question 2 and employ a binary **probit model** to perform the following tasks..

(a) Derive the probability of $y_i = 1$ and $y_i = 0$. Construct the likelihood function.

(b) Estimate the binary probit model, with $y$ as the response variable and $(1, x_2, \cdots x_{16})$ as the covariates. Report the coefficient estimates, standard errors, and $t$-values to 3 digits after the decimal.

(c) Calculate the covariate effect of increasing the age by 5 years on $\Pr(y_i = 1|\beta)$. Report the change in probability to 4 digits after the decimal.

(d) Calculate the covariate effect of being a parent on $\Pr(y_i = 1|\beta)$. Report the change in probability to 4 digits after the decimal.

**4. (4+4+4+4+4 = 20 points)** Consider the data from Question 2 and employ a binary **logit model** to perform the following tasks.

(a) Derive the probability of $y_i = 1$ and $y_i = 0$. Construct the likelihood function.

(b) Estimate the binary logit model, with $y$ as the response variable and $(1, x_2, \cdots x_{16})$ as the covariates. Report the coefficient estimates, standard errors, and $t$-values to 3 digits after the decimal.

(c) Calculate the covariate effect of increasing the age by 5 years on $\Pr(y_i = 1|\beta)$. Report the change in probability to 4 digits after the decimal.

(d) Calculate the covariate effect of being a parent on $\Pr(y_i = 1|\beta)$. Report the change in probability to 4 digits after the decimal.

(e) Is there any difference in the covariate effects for parts (c) and (d) from the probit and logit models? Explain.

**5. (4+4+4+8 = 20 points)** Consider the data provided in the file "`EducationalAttainment.xlsx`" and the ordinal probit model as discussed in the class. The description of the variables is given below.

| | |
|---|---|
| `educLevel` | Levels of educational attainment. 1 = less than high school (HS) education, 2 = HS degree, 3 = some college or associate's degree, and 4 = college or graduate degree. |
| `famIncome` | Family income in thousands of 1980 US Dollars. Take the square root of `famIncome` when estimating the ordinal probit model. |
| `educMother` | Education of mother in years. |
| `educFather` | Education of father in years. |
| `motherWorked` | Working status of mother. 1 = working, 0 = otherwise. |
| `female` | Indicator for female gender. 1 = female, 0 = male. |
| `black` | Indicator for black race. 1 = black, 0 = otherwise. |
| `urban` | Indicator that youth lived in an urban area at the age of 14. 1 = urban, 0 = otherwise. |
| `south` | Indicator that youth lived in South at the age of 14. 1 = south, 0 = otherwise. |
| `age15` | Indicator variable for age cohort 15 in 1979. 1 = age15, 0 = otherwise. |

| | |
|---|---|
| `age16` | Indicator variable for age cohort 16 in 1979. 1 = age16, 0 = otherwise. |
| `age17` | Indicator variable for age cohort 17 in 1979. 1 = age17, 0 = otherwise. |

*Please maintain this ordering when reporting the data summary and estimation results from ordinal probit regression.*

(a) For an ordinal probit model, what is $\Pr(y_i = j)$? Write the likelihood function for an ordinal probit model.

(b) Consider the data in the file "`EducationalAttainment.xlsx`" and report the descriptive summary of the variables. Specifically, report the mean (count) and standard deviation (percentage) for the continuous (categorical) variables, all in single table.

(c) Fit an ordinal probit model with `educLevel` as the dependent variable and the remaining variables as independent variables. Do not forget to include an intercept and take the square-root of `famIncome`. Report the coefficient estimates, standard errors, and $t$-values in a table to 3 digits after the decimal. Find the hit-rate and the McFadden's R-squared.

(d) Find the average covariate effect of a \$10,000 increase in `fameIncome`. Report the change in probabilities to 4 digits after the decimal.