

2024

ECONOMICS — HONOURS — PRACTICAL

Paper : DSE-A-1P

(Applied Econometrics)

Full Marks : 30

*The figures in the margin indicate full marks.*Answer **any three** questions (using STATA or R).

Create a log file (in smcl or log format in STATA and R respectively) and put your registration number and roll number, without any blank space, as file name. If your Registration Number is XXX-XXXX-XXXX-XX and Roll Number is XXXXXX-XX-XXXX your file name will be REGXXX-XXXX-XXXX-XXROLLXXXXXX-XX-XXXX.

1. Consider dataset in question no. 1 and answer the following questions :

- Calculate the detailed summary statistics of wage and family income (*faminc*).
- Generate a variable *dce* by taking the difference between wage and reported wage (*repwage*) at the time of interview. Find the variance of *dce*.
- Summarize wage if it is at most 5.6.
- Generate a variable *famincsq*, twice square root of family income.
- Give the information of the dataset.
- Draw a bar diagram for wage and *faminc*.
- Make a list of wages greater than 25,000 (Rs.).
- Draw the histogram for wage. 2+2+1+1+1+1+1+1

2. In question no. 2 dataset sex denotes the gender of the person, for male sex = 1 and for female sex = 2. Use this dataset to answer the following questions :

- Create a variable which will take the value 1, if the person is female and 0 otherwise.
- Get the spreadsheet of all the variables in the dataset. What type of variable is 'hhid'?
- Get the frequency distribution of various categories of household type (*hhd-type*).
- Find the mean and median age.
- Draw a scatter diagram between wage and age. Give a suitable title of the diagram.
- Make a list of male person whose general education (*gen-edu*) is 10. 2+(1+1)+1+2+2+1

Please Turn Over

(0989)

3. Consider the dataset given in question no. 3 and answer the following questions :

- Create two variables, one indicating natural log of wage as *lwage* and another indicating square of experience (*exper*) as *expersq*.
- Run a linear regression to estimate the model where *lwage* is regressed on *exper* (experience), *expersq*, *educ*(education), *age*, *kidslt6* (number of kids is less than 6 years) and *kdsge6* (number of kids whose age lies between 6 to 18 years).
- Interpret the estimated coefficients of *expersq* and *kidslt6*.
- Find 90% confidence intervals of *educ*.
- Check the overall significance of the model.
- Find the predicted values of *lwage* as *lwagehat* and residuals as *res*, of the model.
- Make a formal test to check whether the residuals are homoscedastic.
- Present the correlation matrix of the variables. 1+1+2+1+1+2+1+1

4. Dataset provide in question no. 4 represents annual data of per capita gross domestic product (*pcgdp*) of a country for the period (year) 2011-12 to 2023-24. Use this dataset and answer the following questions :

- Set the data as yearly time series data.
- Create two period lagged values of *pcgdp*.
- Make the best fitted line for *pcgdp*.
- Make a line plot of *pcgdp* and fit a straight line on the line plot.
- Estimate the average rate of growth of *pcgdp* for the entire period.
- Get the difference between *pcgdp* and estimated value of average rate of growth of *pcgdp* and name it as *res*. Is this difference significant?
- Make a list of year and *pcgdp* for the period of 2015-16 to 2023-24. 1+1+1+2+2+2+1

5. Dataset of question no. 5 pertains to the data of 9 regions (country) for 28 years (year). Use this dataset to answer the following questions :

- Generate a series to convert string variable "country" to numeric.
- Set the dataset as panel data by setting the new series for "country" as panel variable and year as time variable.
- Draw line plots of "output per worker" for the panel.
- Get the descriptive statistics of the panel data.
- Suppose "output per worker" is explained by "wage workers" and "vulnerable employment". Estimate the pooled regression model and interpret the results.
- Estimate above model by assuming significant differences among regions (country) but no significant temporal effects. Interpret the results.
- Perform a test to compare the models estimated in (d) and (e). Which model is better? 1+1+1+1+2+2+2