M.A. Examination, 2023

Semester-II

Subject: ECONOMICS

Paper-C-9 (Computer Application)

Time: Three Hours **Full Marks: 40**

Questions are of value as indicated in the margin

Answer ***any four*** questions

1. (a) What will be your unit of analysis against each of the following broad study objectives?

(i) Estimating the population growth rate across districts in India using time series data.

(ii) The financial performance of women Self-Help Groups in Birbhum district.

(iii) Effect of various inputs on productivity of paddy in agricultural lands in Birbhum district.

(iv) Effect of age, gender and education on mobile data consumption pattern among adults.

(v) Effect of improved cook stove in reducing indoor air pollution and health hazards in India.

(b) What types of charts would you use in describing the following information?

(i) Relative importance of different countries in India's total petroleum import during 2022.

(ii) Daily highest and lowest temperature in Sriniketan during last one year.

(iii) Gender and class wise distribution of Economics (honours) students in Visva-Bharati during the academic year 2022-23.

(iv) Relationship between Block-wise rainfall and paddy productivity in West Bengal during the previous Kharif season

(v) Market share of different Electric Car producers in India during the last financial year.

5+5 = 10

2. (a) Suppose a detailed household survey is being planned to understand the livelihood scenario in a remote village. The villagers have a limited set of livelihood options comprising of *agriculture*, *daily labour*, *migrant labour*, *petty business* and *receipt of government assistance*. Very few villagers might be earning through other activities which cannot be specified prior to the survey. Also, none of the villagers have educational qualification above graduation level. Create a survey module for the sample households that can record member-wise details like age, gender, education, primary activity (if any) and one subsidiary activity (if any). Members of a household may be listed through a 'member_serial' ignoring their names. Also, specify the code lists that you might use for the module.

(b) If the data has to be entered in an Excel Sheet using the module you specified in part (a), explain how you would assign the variables to successive columns and what would be data range under each of these columns. 5+5 = 10

3. Following is part of an OLS regression result produced by Excel using a household survey dataset, where the dependent variable is household's monthly per-capita consumption expenditure (**mpce**):

Variable description	Regressor	Coefficient	t-value	p-value
Total number of household members	fsize	-365	-3.77	0.003
Age of the household head (Years)	hdage	10.5	2.48	0.014
Gender of the household head (0=Male; 1= Female)	hdsex	-0.41	-1.30	0.195
Location status (0 = Urban; 1 = rural)	rural	-1495	-5.74	0.001
Amount of agricultural land (Katha)	agri_land	155	3.64	0.005
Total value of livestock,holding (in thousand rupees)	livestock	22.2	0.66	0.508
	Constant	5388	3.55	0.006
Number of observations=1136, R-squared=0.238, F-value=50.297, Prob>F=0.000				

(a) Describe your understanding of the behaviour of ***mpce*** with respect to each of the regressors separately.

(b) What is the marginal effect of *fsize* on *mpce*? Do you have any plausible explanation for its sign?

(c) Draw an appropriate probability distribution for the estimated coefficient of the regressor '*hdage*'. Indicate the area outside the corresponding estimated t-statistic. 5+2+3= 10

4. (a) A regression output produced by Excel, typically provides the ANOVA for the regression. Following is an example of such a result:

	A	B	C	D	E	F
1	ANOVA					
2		*df*	*SS*	*MS*	*F*	*Significance F*
3	Regression	5	25831	5166	6.52	0.0015
4	Residual	17	13471	792		
5	Total	22	39302			

(i) What do you mean by 'significance F'? How significant is the regression result here?

(ii) What is meant by F-value? How is it computed?

(iii) What should be the value of R^2 in this regression result? (give formula only, using cell references)

(b) In a survey of 150 households, information on household's daily expenditure on food (Rs.) is collected along with family size. It is observed that family size ranges from 1 to 10. Then food expenditure is regressed on family size (fsize) in a quadratic form. The partial regression result is produced below:

Variable	*Coeff.*	*Std. Err.*	*t Stat*	*P-value*
Intercept	-20	27.19	-0.77	0.455
fsize	215	12.38	17.47	0.000
fsize-square	-7	1.20	-5.79	0.000

On the basis of the above result, draw a graph showing estimated daily expenditure on food against family size for those households.. Explain its shape. (2+2+1) +5 = 10

5. (a) Suppose a Stata dataset contains two variables x and y. It is known that they are related by the following equation:

$$y_i = e^{\alpha+\beta x_i+u_i}$$

where u_i is the random error term following CLRM specification and α, β are the parameters. Write down the set of commands in Stata that would help you to obtain the estimated parameters using OLS method.

(b) A file named *File1.dta* is loaded in Stata. Following is a Stata command and the corresponding result as is shown in the Results window:

```
. merge 1:1 hhid using "C:\Desktop\Projects\File2.dta"

    Result                      # of obs.
    -----------------------------------------
    not matched                        14
    from master                        4         (_merge==1)
    from using                         10        (_merge==2)

    matched                            71        (_merge==3)
    -----------------------------------------
```

(P.T.O)

(i) What is attempted with *File1* and *File2* with the command? Is the attempt successful?
(ii) What is the 'master' file and what is the 'using' file?
(iii) What is the total number of observations in the whole dataset ?
(iv) What is the unique identifier across the two files?
(v) What is '_merge' ? what are its value labels and their meanings? 5+5 =10

6. (a) Consider the following part of a household questionnaire used for primary survey with coded options:

Q1. Structure of the house (put √): *1=Kuchha* *2=Semi Pucca* *3=Pucca* *4=Other*
Q2. Main source of light (put √): *1=Electricity* *2=Solar unit* *3=Kerosene* *4=Other*
Q3. Whether the household has a sanitary latrine (put √) *:1=Yes* *2=No*

The collected information is recorded in a Stata dataset with variable names q1, q2 and q3 respectively. Write the set of Stata commands that will assign appropriate variable labels as well as value labels under each variable.

(b) Write the Stata commands lines that will accomplish the following tasks:
(i) Generate a serial number (1,2,3...) for the existing dataset with the name 'slno'.
(ii) Substitute any missing value in the variables **aa**, **bb** and **cc** with the value zero.
(iii) Change the character of the variables **xx** and **yy** from alphabetic to numeric.
(iv) Create a new variable **zz** such that it contains the arithmetic mean of **aa**, **bb** and **cc**.
(v) Arrange the variables in the dataset such the serial number appears first, followed by the variables **xx** and **yy.** 5+5 =10

7. Following is the description of variables in a Stata dataset:
Location of data file on computer: "C:\Desktop\Sem2\Survey.dta"
Number of observation: 100 [household level survey data]

Variable	Description
sl	Unique identification number of the survey household
age'i'	Age of the 'i'-th member of the household [10 repetition of 'i' = 1,2,.....10]
sex'i'	Sex of the 'i'-th member of the household [10 repetition of 'i' ; each coded: 1= Male; 2= Female]
edu'i'	Number of years of education of the i-'th member of the household [10 repetition of 'i'] [0 = illiterate]
main_earning'i'	Main earning activity of 'i'-th member of the household [10 repetition of 'i'] (code 1 to 5, as described at the end of this table); blank if no earning activity
other_earning'i'	supplementary earning activity of 'i'-th member of the household (code 1 to 5, as described at the end of this table); blank if no supplementary earning activity.
land_bigha	Amount of agricultural land in Bigha (1 Bigha = 20 Katha)
land_katha	Amount of agricultural land in Katha
cow	Value of cows possessed (Rs.) [blank if no cow]
goat	Value of goats possessed (Rs.) [blank if no goat]
birds	Value of chicken/duck possessed (Rs.) [blank if no such birds]

Note: *all variables except* **sl, age1, sex1** *and* **edu1** *may have missing values*

Codes for earning:
1=Agriculture ; 2=Daily labour; 3= Migrant Labour 4= Own business 5= Salaried employment ; 6= Any other

Following the above details, sequentially write the command lines for a Stata DO-file that will execute the following tasks:

(a) Open the data file, generate new variables (with names of your choice) representing (i) family size (ii) total agricultural landholding and (iii) total value of livestock holding for each household. Store these newly generated variables along with the household identifier in

(b) Open the original data file and keep only the variables containing individual information. Reorganize the data at individual level. Assign value labels to individuals' main and supplementary occupations. Store the individual level information in a separate .dta file in the same folder using the filename 'Indiv_profile' such that it can be merged with the household level information using appropriate identifiers. Clear the data from memory.

(c) Merge the two files 'HH_profile.dta' and 'Indiv_profile.dta' with an appropriate Stata command. 4+4+2 =10

8. Refer to the dataset described in the previous question (Q#7). A model suggests that individual's decision to work as migrant labour may be affected by factors like age, gender, years of schooling, family size as well as asset possession such as land and livestock holding. Write down the set of Stata commands sequentially to check the effect of these factors on one's decision towards opting for migrant labour as his/her main earning activity. You need to check the marginal effect of each of these regressors after the model estimation. 10