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

Questions are of value as indicated in the margin
Answer *any four* questions

1. (a) What types of charts would you prefer to visualize the following information?
   (i) *Daily opening and closing value of a stock during last one month in a stock market*
   (ii) *Composition of male and female employees in three offices*
   (iii) *Relative market share of different mobile service providers in India*
   (iv) *State Domestic Product of different states of India during 2020-21*
   (v) *Relationship between monthly household expenditure and income for 500 households*

 (b) Suppose 'x' and 'y' are two values in the cells A1 and B1 respectively in an Excel sheet. In cell C1, what formula would you enter that computes :
   (i) *Natural log of x*
   (ii) *Antilog of x*
   (iii) *Fourth root of 'x'*
   (iv) $e^x + \sqrt{y}$
   (v) *Value '10' raised to the power of 5*   5+5=10

Q2. An OLS regression result is produced in Excel and is described below with some cell values omitted (with interrogation mark '?')

	A	B	C	D	E	F
1	Multiple R	0.242421				
2	R Square	0.058768				
3	Adjusted R Square	?				
4	Standard Error	6.923368				
5	Observations	306				
6	ANOVA					
7		df	SS	MS	F	Significance F
8	Regression	4	900.8371	225.20927	4.6984	0.001083
9	Residual	301	?	47.93303		
10	Total	?	15328.68			
11						
12		Coefficients	Standard Error	t Stat	P-value	
13	Intercept	?	8.636340	-1.275470	0.203127	
14	Regressor 1	-0.000054	0.000020	?	0.006309	
15	Regressor 2	0.045694	0.021137	2.161791	0.031422	
16	Regressor 3	0.016212	?	1.831304	0.068044	
17	Regressor 4	-0.064860	0.034223	-1.895300	0.059011	

(a) What formulae would you put in cells B3, B10, B13, C9, C16 and D14? *[assume that it is part of an Excel spreadsheet and hence use cell references in your formulae, not the values]*.
(b) What is your conclusion against the null hypothesis that 'Intercept' is zero?
(c) Which of the regressors are significant at 10% level but not at 5% or 1% level of significance?

[P.T.O]

(d) What do you understand by the value of 'Standard Error' given in cell B4?
(e) What percentage of total variation in the dependent variable is explained by this regression model?   6+1+1+1+1=10

3. (a) Explain the concept of 'p-value' as reported in regression results in a statistical software with the help of a diagram, showing the corresponding distribution of the test statistic.

(b) In an OLS regression with 45 degrees of freedom, t-values of two regressors X1 and X2 are found to be t_1= -1.34 and t_2= 2.23. The corresponding p-values are 0.18 and 0.028 respectively. Draw a t-distribution and show the position of t-values and the corresponding p-values by shaded area (draw separate diagrams for the two t-values).

(c) Discuss on the significance of the two regressors X1 and X2 specifying the corresponding null hypotheses.   3+4+3 = 10

4 (a) Following is an OLS regression result with data from 226 rural households. Households' monthly expenditure (rs.) on cooking fuel is regressed upon some explanatory variables which are described below. Based on the given result, describe your conclusion regarding the effect of the regressors on households' fuel expenditure.

Variable description	Variable Name	Coefficient	t-value	P-value
Total number of family members	FSIZE	95	3.17	0.002
Total agricultural land (in Bigha)	LAND	-15	0.38	0.345
Whether there is LPG connection in the household (Dummy: 1= Yes; 0= No)	LPG	275	-2.83	0.006
Age of the household head (completed years)	HDAGE	12	-0.15	0.740
Education of the household head (years of schooling)	HDEDU	- 15	-1.21	0.172
Intercept term in the regression	CONSTANT	260	3.96	0.001

(b) Following is the summary of a regression result where the dependent variable is 'Household expenditure on food'. The variables used as regressors are described in the first column of the following table:

Variable description	Variable	Coefficient	t-value	P-Value
Household size	FSIZE	66	-5.92	0.000
Square of household size	FSIZESQR	4	6.05	0.000
Value of livestock (Rs.)	LIVESTOCK	-51	-1.34	0.182
Per-capita landholding (Bigha)	PCLAND	40	1.93	0.127
Intercept Term	CONSTANT	500	8.68	0.000

Discuss the regression result and show the estimated relationship between 'Household expenditure on food' and household size graphically.   5+5=10

5. (a) Consider a data set containing three variables such that –
$$Y = f(X, D)$$
where D (0,1) is recorded as a dummy variable. It is known that the functional relationship is linear. You want to check whether the dummy variable affects the slope, intercept or both slope and intercept in the relationship between Y and X. State the model you would like to estimate. Explain your conclusion relating to the different possible outcomes (significance/insignificance of the parameter estimates).   [P.T.O]

(b) '*Monthly household expenditure*' is regressed on the number of household members and the information whether any household member has salary earning from a regular service. The result is described below.

Variable description	Variable	Coefficient	t-value	P-Value
Household size	FSIZE	440	4.40	0.005
Dummy Variable: =1 for service holder; =0 otherwise	SRVC	421	0.42	0.678
= FSIZE x SRVC	FSIZE_SRVC	250	4.78	0.001
Intercept Term	CONSTANT	2400	2.56	0.015

Show the estimated relationship between 'Monthly household expenditure' and FSIZE for households with service holders and contrast it with that for other households. What is the estimated change in monthly expenditure of a household due to one unit increase in FSIZE if there is no service holder in that household?   5+5=10

6. A dataset is loaded in Stata with some household level variables as described below:

Variable	Variable description
fsize	Total number of household members (integer values between 1 to 14)
caste	Integer values between 1 to 4
land_katha	Households' agricultural land in Katha unit (with missing values)
land_bigha	Households' agricultural land in Bigha unit (with missing values)
livestock	Values 0 or 1 (1 = possessing some livestock; 0= No livestock possession)
child_5	Number of children below 5 years of age completed. (values between 0 to 3)
mexp	Monthly expenditure in Rupees.

(a) Describe the tasks accomplished by the following commands:

```
(i)   . label variable livestock "Whether the household has  livestock"
(ii)  . label define YN 0 "No" 1 "Yes"
(iii)   . label values livestock YN
(iv) . gen with_child = 0
(v)   . replace with_child = 1 if child_5 !=0
```

Assuming the commands described above are executed successfully, write the set of Stata commands that would do the following sequentially:

(i)   Create a new variable 'totland' that would calculate the total agricultural land of a household in katha unit (1 Bigha = 20 Katha).

(ii)   Assign value labels to the 'caste' variable such that 1,2,3 & 4 values represent 'General', 'SC', 'ST' & 'OBC' respectively

(iii)   Regress 'mexp' on household size, total agricultural land, a dummy variable representing whether the household possess livestock and a dummy variable representing whether the household has children below the age of 5 years.

                    5+5=10

7. (a) Suppose you need to estimate the parameters of a production function
$$Q = A.L^{\alpha}.K^{\beta}.u \quad \text{(where u is an error term)}$$
Raw data on production (Q), labour input (L) and capital employed (K) is already loaded in Stata. You need to estimate the parameters of the model A, α and β. Write down the corresponding command lines and describe how you would obtain the estimated parameter values from the results generated by your commands.

(b) Discuss the result shown in the following table that is generated from the Stata commands:

```
. logit y x1 x2 x3 x4
. mfx
```

                    [P.T.O]

3

variable	dy/dx	Std. Err.	z	P > z
x1	0.166	0.421	0.39	0.694
x2	-0.033	0.029	-1.16	0.244
x3	-0.495	0.583	-0.85	0.396.
x4	0.004	0.003	1.34	0.181

8. Following is the description of a Stata dataset:

Location of data file on computer: "C:\Users\Desktop\STATA\Test.dta"

Number of observation: 100 [household level survey data]

Variable	Description
hhid	Unique identification number of the survey household
fsize	Number of members in the household
earn1	Name of the first earning member of the household (considered as head of the household)
age1	Age of the first earning member of the household
sex1	Sex of the first earning member of the household [coded: 1= Male; 2= Female]
edu1	Number of years of education of the first earning member [0=illiterate]
occu1	Occupation code (1 to 5, as described at the end of this table) of the first earning member
income1	Average annual income (Rs.) of the first earning member from the occupation described in occu1
earn2*	Name of the second earning member of the household
age2*	Age of the second earning member of the household
sex2*	Sex of the second earning member of the household [coded: 1= Male; 2= Female]
edu2*	Number of years of education of the second earning member [0=illiterate]
occu2*	Occupation code (1 to 5, as described at the end of this table) of the second earning member
income2*	Average annual income (Rs.) of the second earning member from the occupation described in occu2
conexp	Monthly average consumption expenditure of the household
land*	Amount of agricultural land in the household (Bigha)

Implies some of the values of these variable might be missing (i.e, appears as 'dots')

[Codes used for occupation: 1=Own agriculture ;   2=Daily labour;   3=Own business 4= Salaried employment ;   5= Any other]

Given the description of the data above, sequentially write the command lines for a STATA Do-file that will execute the following tasks:

(a) Open the data file, calculate the total income of the households from two possible earning members.

(b) Open the data file, run a regression that explains per-capita monthly consumption expenditure as function of monthly income and family size at household level.

(c) Open the data file and reorganize the data at individual level. Run a logit regression to examine whether age, sex and educational attainment of an individual are significant determinants for choice of the occupation 'Own business'. Calculate the marginal effects (change in probabilities).   3+3+4=10

--X--