M.A. Examination, 2017
Semester-II
Economics
Course-VI
(Computer Application)

Time: 3 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 describe the following data?
   (i) *Weekly rate of inflation over one year*
   (ii) *Share of different manufacturers in the small car market of India in 2011*
   (iii) *Daily maximum and minimum temperatures in Sriniketan for one year.*
   (iv) *State Domestic Product of different states of India in 2016-17*
   (v) *Sex distribution of students in five different classes in the Department of Economics in 2016-17 academic session.*

   (b) The following questions were part of a questionnaire used in a survey among the households in a forest village. Show how you will format your Excel sheet to enter the data so that anybody can obtain all the necessary information without any need to check the filled-up questionnaires

   Q(i). Main source of fuel for the household (Put √): *1. Fuel wood; 2. LPG 3. Kerosene 4. Other*

   Q(ii) Average expenses on fuel: (Rs.)_____per Week / Month (Put √)

   Q (iii). Sources of earning during last one year (*put √ on corresponding option/options)*

   1. Agriculture; 2.Daily wage labourer 3. Forest collection 4. Salaried employment 5. Any Other

5+5=10

2. (a) How would you generate an integer random number between 0 to 100 in a cell in Excel? How can you preserve its value so that the generated number is not changed even if the page is refreshed?

   (b) Suppose data on a variable X is provided in column A in an Excel sheet. You want to create a scatter diagram in (X,Y) plane so that the scatter shows a positive linear relation between the variables with a slope of 2. How can you generate the variable Y in Excel?

   (c) The following macroeconomic data for a hypothetical closed economy is ready in an Excel sheet (the first column and first row are labels for rows and colums):

	A	B	C	D
1	National Income=	1000	Disposable Income=	
2	MPC out of disposable income=	0.8	Government Expenditure=	
3	Tax rate=	0.2		
4	Autonomous investment=	200		

Write the formulae you will put in cells D1 and D2. After entering it, what value you would expect to appear in D2? 2+4+4=10

P.T.O.

3. Suppose you need to estimate the parameters of a production function

$$Q = A.L^{\alpha}.K^{\beta}.u \qquad \text{(where u is an error term)}$$

Raw data on production (Q_i), labour input (L_i) and capital employed (K_i) is ready in an Excel sheet. You have to estimate the parameters of the model. You can do it either in Excel or in Stata. If you use Excel, write the steps sequentially that you will execute. If you use Stata, write the corresponding command lines for Stata (assuming the data is already imported in Stata and is in the background). 10

4. (a) Explain the concept of 'p-value' as reported in regression results in any statistical software.

(b) In a OLS regression with 25 degrees of freedom, t-values of two regressors are found to be $\mathbf{t_1=2.32}$ and $\mathbf{t_2= - 0.83}$. The corresponding p-values are reported as **0.028** and **0.414**. 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) 5+5=10

5. Following is the result of a regression run in Excel:

	A	B	C	D	E	F
1	***Regression Statistics***					
2	Multiple R	**??**				
3	R Square	0.35				
4	Adjusted R Square	0.25				
5	Standard Error	18.29				
6	Observations	**??**				
7	**ANOVA**					
8		***df***	***SS***	***MS***	***F***	*Significance F*
9	Regression	4	4580	1145	**??**	0.0231
10	Residual	25	**??**	334		
11	Total	29	**??**			
12		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	
13	Intercept	216.2742	**??**	10.0220	0.0001	
14	X1	0.9140	0.4557	2.0059	0.0558	
15	X2	-1.1086	1.3353	-0.8303	0.4143	
16	X3	1.4585	0.6268	2.3269	0.0284	
17	X4	0.8514	1.1926	0.7140	0.4819	

(i) Write down the formula for filling up the missing values in the cells marked by interrogation sign (cells **B2, B6, C10, C11, C13** and **E9).** Please note that only formula is required for these cells, NOT the values.

(ii) On how many observations the regression was run?

(iii) Which of the explanatory variables are significant and at what level of significance?

(iv) What is your conclusion regarding the overall significance of the regression?

(v) With the same data, but with a different model, the regression sum of square is found to have reduced to 4200 (instead of current 4580). What will be the change in Total Sum of Squares (TSS)? Why? 3+1+2+2+2=10

6. Data on consumption and income of two groups of people are available in an Excel sheet. The entire data set is given in three columns – column A containing income (Rs.), column B containing consumption (Rs.) and column C containing the code of the group (=1, if the observation belongs to group I; =2, if the observation belongs to group II). The scatter projection of consumption against income (for the entire data) shows usual upward rising shape, but in two distinct sets. One set clearly lies above the other, but showing similar slope. With this background, answer the following questions:
   (i) Write down the equation that you would like to estimate with the given data such that the equation is best fit in the given scatter diagram.
   (ii) Do you need to modify your data in the Excel sheet for the estimation you propose? Explain the modification you will carry out.
   (iii) With the parameter estimates of your model, describe diagrammatically how your regression line/lines fit into the scatter diagram (constructing the scatter following the description in the question). Explain the implication of the parameter estimates with the diagram. 2+2+6=10

7. (a) Suppose $\mathbf{x_1}$, $\mathbf{x_2}$, $\mathbf{x_3}$ are three continuous variables that are stored in a STATA data file named **test.dta** which can be accessed through the path **"C:\Users\Personal\Desktop\test.dta"**. The variable $\mathbf{x_2}$ contains some missing values (dots) which should be considered as zero. Write down the set of commands that will execute the following steps in STATA software:
   (i) Load the data from the file **test.dta** in STATA data sheet
   (ii) Generate a new variable named **y** that is defined as $y = \mathbf{x_1}+3\mathbf{x_2}$
   (iii) Generate summary statistics of variable **y**

   (b) In a STATA dataset named '**hhdata.dta'**, which can be accessed through the same path given in the previous part of this question, there are three variables 'xx', 'yy' and 'zz'. The first is a continuous variable with a range from 1,000 to 25,000. The second is a discrete variable with a range from 1 to 14. The third is a discrete variable with integer values 1,2,3 and 4 only. Write the command lines in STATA for the following tasks:
   (i) Renaming the variable 'xx' as "hhconsumption" and 'yy' as "hhsize"
   (ii) Generate a new variable 'pcc' which is equal to *hhconsumption* divided by *hhsize*
   (iii) Define a label 'caste' which describes the codes 1,2,3 and 4 by "SC", "ST" , "OBC" and "General" respectively.
   (iv) Assign the label *caste* to the values of the numeric variable 'zz'
   (v) Scatter plot of *pcc* against *hhsize* 5+5=10

8. Following is the description of a Stata dataset:
   Location of data file on computer: **"C:\Users\Desktop\STATA\test2.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 (considered as head 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

**** 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

Carefully read the description of the data as given above and then write the Stata command lines for a Do-file that will execute the following tasks:

(a) Open the data file, generate a variable '**totinc**' which represents the total income of the household. Generate a variable '**pcc**' that calculates per-capita income of the household. Run a regression to check whether pcc is explained by **totinc** and **land**. Save the data file as **test3.dta** in the same location. Erase all data from Stata datasheet.

(b) Open the data file **test3.dta.** Transform the data into individual level data. Generate a dummy variable for individuals who have salaried employment. Run an appropriate regression to check whether education and age has any significant impact on the probability of getting salaried employment. 5+5=10

_____