UNIVERSITY EXAMINATIONS: 2013/2014

EXAMINATION FOR THE MASTERS OF SCIENCE (MSC) IN COMMERCE

MSF 502 ECONOMETRICS I

DATE: APRIL, 2014 **TIME: 3 HOURS**

INSTRUCTIONS: Answer Question One and Any Other Three Questions

---

QUESTION ONE

a) Briefly differentiate the following estimation methods as used in econometrics: OLS, MLE and GMM [6 Marks]

b) Irene, an MSC student from KCA University, is writing her MSC thesis titled: "Impact of job satisfaction on employee retention at KCA University." She has the following three objectives for her thesis:

(i) To determine the impact of reward systems on employee retention

(ii) To assess the impact of capacity building on employee retention

(iii) To examine the impact of working environment on employee retention

She collected quantitative data from the target population to fit the following population regression equation: $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$

Where:

y = Employee retention

X_1 = Benefits (reward system)

X_2 = Training (Capacity building)

X_3 = Type of office(Working environment) and;

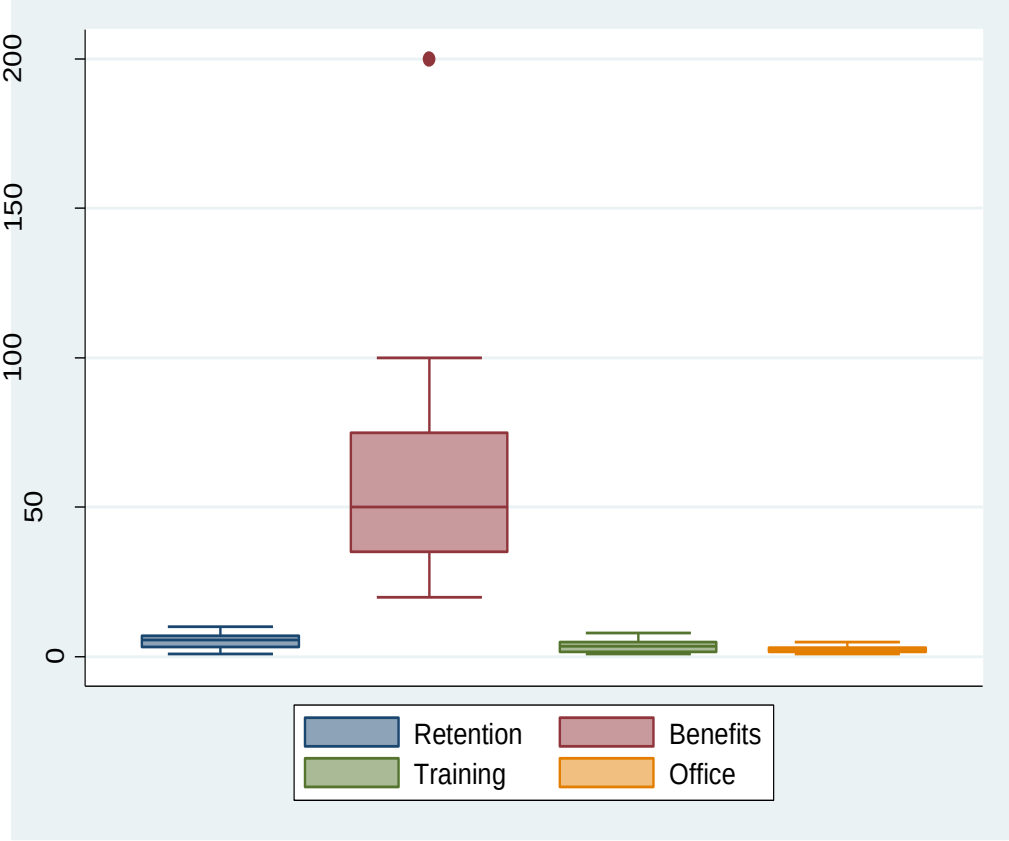
ε = Error term

Use the information provided to answer the following questions:

1) Using relevant equations, differentiate between a population regression equation and an estimated regression equation [2 Marks]

2) Diagrammatically show the relationship between the variables mentioned in the title and the objectives of study by showing clearly the outcome and the explanatory variables [4 Marks]

3) From the data collected, from 20 respondents, the exploratory output below was derived. Interpret the output. [4 Marks]



4) The MSC student fitted the multiple regression, and the output is given as:

Source	SS	df	MS		
Model	61.4618138	3	20.4872713	Number of obs =	20
Residual	68.7381862	16	4.29613664	F(3, 16) =	4.77
				Prob > F =	0.0147
Total	130.2	19	6.85263158	R-squared =	0.4721
				Adj R-squared =	0.3731
				Root MSE =	2.0727

retention	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
benefits	.019274	.0125077	1.54	0.143	-.0072411	.045789
training	.7903634	.2304289	3.43	0.003	.301876	1.278851
office	.0191658	.419212	0.05	0.964	-.869524	.9078555
_cons	1.308159	1.668116	0.78	0.444	-2.228089	4.844406

Required:

Use the above output to answer the following questions:

(i) Write the STATA command used to give the regression output shown above. [1 Mark]

(ii) How well does the model fit the data? [2 Marks]

(iii) Discuss the significance of the explanatory variables in the regression equation [3 Marks]

(iv) Write the fitted regression from the output given [2 Marks]

c) Suppose we have a multivariate model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_1 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + u_i$$

and we want to test whether independent variables $x_3, x_4, and\ x_5$ jointly have effect on y.

(i) Write the null hypothesis [1 Mark]

(ii) Write the restricted model. [1 mark]

(iii) How many restrictions are there in this test? [1 mark]

(iv) Now consider an estimated regression model given as:

$$\hat{Y} = 3500 - 0.148X_{1t} - 11.13X_{2t} + 2.2X_{3t}, R^2 = 0.123, n = 700$$

When the model was estimated without the second and third explanatory variables the following was obtained:

$$\hat{Y} = 3495 - 0.151X_{1t}, R^2 = 0.113, n = 700$$

Are the second and third explanatory variables jointly significant in the original function at 5% level of significance? [4 Marks]

QUESTION TWO

Given the following over 21 observations

$$(X'X)^{-1} = \begin{bmatrix} 2 & 3 & -1 \\ 3 & 1 & 6 \\ -1 & 6 & 4 \end{bmatrix}, \quad X'Y = \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix} \quad \text{and} \quad \sum e_i^2 = 10$$

a) Obtain the OLS estimates and specify the estimated function [5 Marks]

b) Determine the standard errors of the coefficient estimates [8 Marks]

c) Test the significance of the independent variables in the estimated function at 5% level of significance [10 Marks]

QUESTION THREE

Billy, a General Manager at Mumias Sugar Company, estimated a multiplicative demand function of the form: $Q_d = aP^{a_p} I^{a_1} P_0^{\ a_0}$ using a cross-section data collected in the company sales on 30th June, 2013. The estimation results are as follows:

	Constant	*Price (P)*	*Income (I)*	*Price of Other Good (P0)*
Estimated coefficient	0.022	-0.223	1.354	0.133
Standard Error	0.012	0.056	0.502	0.814
t-statistic	(1.19)	(-3.98)	-2.69	-0.13
Number of Observations, *n = 224; R-squared = 0.8515*				
Critical Students' t = 1.96 at 5% Level of Significance				

a) How would the coefficients and R^2 value be interpreted? [5 Marks]

b) What will the quantity demanded be if the values of the independent variables are:
Price = 10; Income per capita = Ksh 9,000; and the Price of the Other Good = Ksh 15?
[5 Marks]

c) How much would quantity demanded change if the price were decreased to Ksh 8 and values of other variables held constant? [5 Marks]

d) What effect would a price increase have on total revenue? Are the two goods substitutes or complements? Explain [8 Marks]

QUESTION FOUR

Consider a zoo that has only chimps and elephants where $poop_j$ is the weight (in pounds) of $poop$

animal j produces on a given day, S_j = 1 if the animal is a elephant, and zero otherwise. The amount of $poop$ animal j produces on a given day is a draw from a *Poisson distribution* (some animals on some days are constipated) with parameter λ_j where $\lambda_j = \alpha + \beta S_j$. The director of the zoo likes only chimps, but knows that chimps like to ride on the backs of elephants, so has 90 chimps and 10 elephants (lots of chimps can ride one elephant).

Required:

(i) Write down the density function for $poop$ for this population. [5 Marks]

(ii) Explain why this is the density function. [3 Marks]

(iii) Derive the expected amount of $poop$ produced per day. [10 Marks]

(iv) Comment briefly on the likelihood of the $poop$ being Poisson distributed rather than Normally distributed. [5 Marks]

QUESTION FIVE

a) State in algebraic notation and explain the assumption about the CLRMs disturbances, referred to by the term "homoskedasticity". [5 Marks]

b) What would the consequence be for a simple linear regression model if the errors were not homoskedastic? Explain [6 Marks]

c) Briefly explain the meaning of the term heteroskedasticity as used in econometrics [2 Marks]

d) Explain the procedures that are used to test for "heteroskedasticity" [10 Marks]

QUESTION SIX

Consider a STATA dataset with the following variables:

age	age in years
educ	highest grade completed
black	1 if race is black, 0 otherwise
other	1 if race is neither black nor white, 0 otherwise
married	1 if married, 0 otherwise
hwk	average hours worked per week
female	1 if female, 0 otherwise
drmo	alcoholic drinks per month

We are interested in estimating the effect of work hours on alcohol consumption. The theoretical effect is ambiguous: on one hand, working more leaves less time for leisure activities, so drinking may drop, but on the other hand, working more may increase stress levels, so drinking may rise. To answer this, through research, we fit a simple linear regression of the form:

$drmo = \beta_0 + \beta_1 hwk + u$; and a multiple linear regression of the form: $drmo = \beta_0 + \beta_1 hwk + \beta_2 age + \beta_3 educ + \beta_4 black + \beta_5 other + \beta_6 married + \beta_7 female + u$

to the dataset.

Suppose that if we estimate the simple linear regression we obtain the following STATA output:

```
      Source |       SS       df       MS              Number of obs =    7489
-------------+------------------------------           F(  1,  7487) =   26.51
       Model |   20239.281     1   20239.281           Prob > F      =  0.0000
    Residual |  5716797.56  7487  763.563185           R-squared     =  0.0035
-------------+------------------------------           Adj R-squared =  0.0034
       Total |  5737036.84  7488  766.164109           Root MSE      =  27.633

------------------------------------------------------------------------------
        drmo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hwk |   .0809038   .0157143     5.15   0.000     .0500995    .1117082
       _cons |    8.92635    .619701    14.40   0.000     7.711562    10.14114
------------------------------------------------------------------------------
```

Also that if we estimate the multiple linear regression model, we obtain the following STATA output:

Source	SS	df	MS	Number of obs =	7488
Model	266836.037	7	38119.4338	F(7, 7480) =	52.17
Residual	5465529.92	7480	730.685819	Prob > F =	0.0000
				R-squared =	0.0465
Total	5732365.96	7487	765.642575	Adj R-squared =	0.0457
				Root MSE =	27.031

drmo	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
hwk	.0302414	.0161589	1.87	0.061	-.0014345	.0619174
age	-.3381229	.1397713	-2.42	0.016	-.612114	-.0641318
educ	-.2849688	.1265659	-2.25	0.024	-.5330735	-.0368641
black	-3.643506	.7159052	-5.09	0.000	-5.046882	-2.240131
other	-3.100133	1.358236	-2.28	0.022	-5.762659	-.4376079
married	-3.096517	.6613095	-4.68	0.000	-4.392869	-1.800164
female	-10.73453	.6449336	-16.64	0.000	-11.99878	-9.470277
_cons	38.07319	6.498279	5.86	0.000	25.33474	50.81164

Required:

Use the STATA outputs to answer the following questions:

(i) In the simple linear regression is hwk statistically significant at 5% level? What about in the multiple regression? (Explain your answer) [4 Marks]

(ii) Explain, in words, exactly what the coefficient estimate for hwk in the multiple regression is telling us. How would you interpret $female$ in the multiple regression? [5 Marks]

(iii) Does it appear that our estimator for the effect of work hours on drinking is biased if we do not include any control variables? (Explain your answer) [4 marks]

(iv) Using an F-test, which is a better model in estimating the effect of work hours on alcohol consumption? (show all your working) [10 Marks]