KENYATTA UNIVERSITY
UNIVERSITY EXAMINATIONS 2010/2011
INSITUTIONAL BASED PROGRAMME (DECEMBER SESSION)

EXAMINATION FOR THE DEGREE OF MASTER OF SCIENCE

SMA /SS/570: STATISTICAL COMPUTING

DATE: **Friday 29th April 2011** **TIME:** **9.00a.m – 12.00p.m**

INSTRUCTIONS

Answer QUESTION ONE and ANY OTHER THREE questions

1 Question One (30 Marks)

(a) The R functions $which.min(x)$ and $which.max(x)$ determine the location, i.e., index of the (first) minimum or maximum of a numeric vector x. Determine the output of the following R program by determining the element(s) in the objects $yyy, exx, output.one$ and $output.two$. *(4 marks)*

```
yyy <- c(2:5,0:3,8)
yyy
eex <- yyy + 3
eex
output.one <- which.min(eex)-1
output.one
output.two <-2*which.max(eex)
output.two
```

(b) Consider the following data called *ExamData* which we assume to be saved as a text file in a folder called *StatisticalData* in the *C* directory:

weight	51	59	64	76	93	106	125	149	171	199
age	2	4	6	8	10	12	14	16	18	20

Table 1: Weight and Age Data

Write a **single** well commented program in R that does the following:

(i) Reads the above data into R. *(2 marks)*

(ii) Sorts the weight into decreasing order. *(1 marks)*

(iii) Computes the standard deviation of weight. *(1 marks)*

(iv) Plots the scatter diagram of *Weight* against *Age* where: *(4 marks)*

X-axis is labeled as "Age".

Y-axis is labelled as "Weight" .

The main title of the graph is "A graph of weight against age".

The subtitle is " Question I Graph"

(v) Superimposes a straight line of gradient 8.1848 and Y-intercept of 19.2667 onto the graph in (iv) above. *(2 marks)*

(vi) Summarizes each of the variables *Weight* and *Age* in a histogram. *(2 marks)*

(c) The following data represent the Withdrawals and Balances (in '000 Kshs) of some Bank customers:

balance	51	59	64	76	93	106	125	149	171	199
withdrawal	5	1	6	4	10	12	14	12	15	10

Table 2:

Consider the following output from SPSS:

ANOVA[b]					
Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	12635.723	1	12635.723	9.714	0.014[a]
Residual	10406.377	8	1300.797		
Total	23042.100	9			
a. Predictors: (Constant), withdrawal b. Dependent Variable: balance					

Table 3: SPSS Output

Coefficients[a]							
	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
Model	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	37.639	25.666		1.466	.181		
withdrawal	8.052	2.583	.741	3.117	.014	1.000	1.000
a. Dependent Variable: balance							

Table 4: SPSS Output

(i) Write down the fitted model. *(2 marks)*

(ii) Estimate the withdrawal of a customer who had a balance of 187000 Kshs. *(2 marks)*

(iii) Comment on the significance of each of the model parameters. *(2 marks)*

(iv) Is the overall regression model significant at 5% level of significance? Explain.*(3 marks)*

(v) Write down a well commented R program that could have given the above output.*(5 marks)*

2 Question Two (20 Marks)

For data that is normally distributed, the standard error for the sample mean is given by

$$\frac{S}{\sqrt{n}}$$

while that for the median is given by

$$\sqrt{\frac{\pi}{2}}\frac{S}{\sqrt{n}}$$

Assuming that the data in table (5) is normally distributed, write a well commented program in R that:

(i) Determines the median for the data. *(2mark)*

(ii) Draws (without replacement) 10000 samples of size 31 each and determines the following: *(18 marks)*

(a) The median for **each sample**

(b) The 98% confidence interval for **each sample**

(c) Determines the number of times the confidence intervals in *(ii)b* above contain the true median computed in *(i)* above.

salary (in Euros)												
51	43	58	52	50	49	52	49	49	54	52	42	53
51	50	47	52	53	44	46	45	54	44	53	48	54
43	55	52	51	52	49	52	46	49	50	52	55	50
49	54	49	47	53	48	55	50	49	53	44	52	44
51	52	48	47	56	52	50	58	52	48	52	54	54

Table 5: Salary for Company Employees

3 Question Three (20 Marks)

An actuary is interested in testing whether the average age of his business customers is significantly different from 35 years of age using 10% level of significance. The data are in table (6) which we assume to be saved as a text file named as *customers* in a folder called **BUSINESS** in the *C* directory

(a) Write down the null and alternative hypothesis to be tested. *(2 mks)*

(b) Write down the mathematical expression for the test statistic (also called the calculated Z, Z_{calc}). *(3 mks)*

(c) Explain how you would test the null hypothesis *(3 mks)*

(d) Write a well commented R program that:

(i) Reads the data into R. *(2 mks)*

(ii) Computes Z_{calc} and the critical Z-value, Z_{crit}. *(10 mks)*

30	29	29	28	33	31	26	36	34	32	27	37
33	26	35	40	35	35	37	32	31	26	30	38
32	23	25	27	22	27	20	32	30	36	25	29

Table 6:

4 Question Four (20 Marks)

The following data set gives the average *weight*, *waist* and *height* for American women aged 30 and 39 years of age. Note that *one* is added to the data frame to take care of the constant when fitting a linear regression equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$$

where Y represents *weight*, X_1 represents *waist* and X_2 represents *height*.

Using **matrices only**, write a well commented R programme that does the following:

(i) Reads in the data.*(2 marks)*

(ii) Computes the following:

(a) The model parameters, $\hat{\beta} = (X'X)^{-1}X'Y$. *(5 marks)*

(b) The vector of predicted responses, $\hat{Y} = X\hat{\beta}$.*(3 marks)*

(c) The regression sum of squares, $\hat{\beta}'X'Y - n\bar{Y}^2$.*(5 marks)*

(d) The error sum of squares, $Y'Y - \hat{\beta}'X'Y$.*(5 marks)*

Y	X		
weight	one	waist	height
115	1	40	158
117	1	25	159
120	1	29	160
123	1	36	161
126	1	35	162
129	1	30	163
132	1	36	164
135	1	35	165
139	1	40	166
142	1	39	167
146	1	48	168
150	1	45	169
154	1	55	170
159	1	52	171
164	1	50	172

Table 7:

5 Question Five (20 Marks)

In a study of the effectiveness of certain exercises in weight reduction, a group of 15 persons engaged in these exercises for one month and showed the results (weight in pounds) in table (8).

weight before	211	180	171	182	160	185	198	172	155	214	194	182	167	185	203
weight after	189	173	172	179	161	181	192	166	154	209	192	182	164	187	201

Table 8:

A researcher is interested in testing whether the exercises are effective in weight reduction.

(a) Show how the data in table (8) can be captured in SPSS by creating a table with specified variable(s). Clearly state the measurement level of the specified variable(s). *(5 marks)*

(b) The following are some results from SPSS concerning the data in table (8):

Test of Homogeneity of Variances			
weight			
Levene Statistic	df1	df2	Sig.
0.160	1	28	0.692

Table 9:

ANOVA					
weight					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	108.300	1	108.300	0.397	0.534
Within Groups	7638.667	28	272.810		
Total	7746.967	29			

Table 10:

(i) Interpret table (9). *(4 marks)*

(ii) Interpret table (10) and comment on whether the exercises are effective in weight reduction.*(11 marks)*