TECHNICAL UNIVERSITY OF MOMBASA

Faculty of Applied & Health Sciences

DEPARTMENT OF MATHEMATICS & PHYSICS

UNIVERSITY EXAMINATION FOR DEGREE OF:

BACHELOR OF SCIENCE MATHEMATICS & COMPUTER SCIENCE (BMCS)

AMA 4320: STATISTICAL MODELLING

END OF SEMESTER EXAMINATION
SERIES: APRIL 2015
TIME ALLOWED: 2 HOURS

Instructions to Candidates:
You should have the following for this examination
- *Mathematical tables*
- *Scientific Calculator*

This paper consist of **FIVE** questions
Answer question **ONE (COMPULSORY)** and any other **TWO** questions
Maximum marks for each part of a question are as shown
This paper consists of **FOUR** printed pages

Question One (Compulsory)

a) You are required to write R code that generates a vector with a sequence of numbers from 1 to 10 with 0.5 steps. Name your vector X. **(5 marks)**

```
> x[x >= 5]
```

b) Using Vector X above what results would you get after running this code, **(5 marks)**

c) Loop statements in R will execute a block of commands until the condition is no longer satisfied below is a loop statement:

```
>t<-5
>while (t<10) {t<-t+1; print(t);}
```

(i) Identify the conditions expression in the loop statement **(2 marks)**
(ii) List the results of the loop statement above **(3 marks)**

d) (i) Write a code that can generate 1000 random variates from normal distribution, name the variates y **(3 marks)**

(ii) Write a code to get the median and inter quartile range (IQR) of 1000 random variates you generated previously. **(2 marks)**

e) A statistician was asked to load data (from SPSS Formant) into R, here are the codes he sed:

>Library (Foreign)

>Setwd ("D;/Moses/Teaching/TUM/statistical_modelling/Data")

Explain these two lines of R syntax **(5 marks)**

f) (i) A researcher proposes that left handness is associated with risk of car accident. To prove his theory, he got data on whether a driver was left or right handness and if they had an accident or not what statistical test would he use to test the null hypothesis of no association between right or left handness and risk of committing an accident? **(2 marks)**

(ii) Suppose the dependent variable (accident) was labeled 'y' and the independent variable (handness) was labeled "x". Write a code in R to test the null hypothesis of no association between right or left handness and risk of committing an accident using the statistical test you named above. **(3 marks)**

Question Two

The rate of crime in United States has been suggested to be linearly associated with the poverty rate. The rate of crime and poverty variably are labeled as 'crime' and 'poverty' respectively. To test the null hypothesis of no linear relationship between rate of crime and poverty, a researcher obtained the following results:

Coefficients:

(intercept)	Poverty
-86.20	49.03

a) Write the code in R used to produce the above results **(3 marks)**

b) State THREE assumptions the researcher made before using the linear regression **(3 marks)**

c) Write the linear equation from these results **(5 marks)**

d) The researcher was interested in showing that there was a statistically significant linear relationship between the two variables. Below are the model summary results he got;

Residents

Min	IQ	Median	3Q	Max
-794.16	-257.78	-22.91	165.50	1713.93

Coefficients

	Estimate	Std.Error	t-value	Pr (>|t|)
(Intercept)	-86.20	176.99	-0.487	0.628403
Poverty	49.03	11.83	4.145	0.000134

Residual standard error: 383.4 on 49 degrees of freedom. Multiple R-squared; 0.2596, adjusted R-squared; 0.2445. F statistic; 17.18 on 1 and 49 DF, P-value 0.0001342

Was there evidence to show that there was a linear relationship **(5 marks)**

e) Constant a 95% confidence interval for the regression coefficient. What can you conclude from the 95% confidence interval **(4 marks)**

Question Three

You are needed to solve the linear equations below in R.

$3x + 4y + 3z = 4$
$2x + y + 2z = 8$
$5x + 4y + 5z = 7$

The first step is to create a matrix. You do this by first loading the 'matrix' package.

```
>Library (Matrix)
```

a) Write a code to create the 3 x 3 matrix that includes the x, y and z data. Call the matrix 'c' **(9 marks)**

b) Write a code to create the vector containing the results of the three equations (4, 8 and 7) call the vector X **(3 marks)**

c) Next you write the following code; >solve (C)
What is the result of running this code on matrix C? **(2 marks)**

d) The results of these equations are given below:

```
     3 x 1 matrix of class "dgematrix"
     {1}
[1] 3.363636
[2] -4.318182
[3] 1.863636
```

Write the code used to produce these results **(6 marks)**

Question Four

Grade point average (GPA) is a grading system used in some countries for admission to college. A social scientist is interested in studying the effect of GPA on college admission. GPA is a continuous variable assumed to have a normal distribution ranging from 1 to 6. College admission is a binary variable coded as 1 for those admitted and 0 for those not admitted to college. The social scientists hired a Statistician to help with the analysis. Below are the first lines codes in R written by the Statistician to export data into R;

```
>data_binary_<_read.spss ("binary.sav",to.data.frame=TRUE)
>attach (data_binary)
>Names (data_binary)
```

a) This data export code depends on particular R package, name this package **(2 marks)**

b) Name the appropriate statistical test that the Statistician used for this analysis **(2 marks)**

c) Given that, the binary admission variable was labeled 'admit' and the GPA grade variable was labeled 'gpa', write the code required to run the statistical test you listed above **(2 marks)**

d) Below are the analysis results produced by the hired Statistician data: gpa by admit
t = -3.6379, dF = 250.049, Rvalue = 0.0003339 alternative hypothesis; true difference in means is not equal to 0:

95% confidence interval

-0.22429214	-0.06673381

Sample estimates

Mean in group O	Mean in group 1
3.343700	3.489213

What was the GPA mean difference between the two groups (those admitted and those not admitted) and the 95% confidence interval of the difference? **(6 marks)**

e) Was there any evidence against the null hypothesis of no mean GPA difference between the two groups **(8 marks)**

Question Five

Analysis of variance (ANOVA) is used to test group differences on the mean of a continuous variable divided up by a categorical variable with more than two levels.

a) State THREE assumptions underlying ANOVA **(6 marks)**

A researcher believes that the level of haemoglobin among first, second and third year diploma students does not differ across the year of academic level. To test this hypothesis he sampled 912 students from first year to third year students and obtained their haemoglobin levels. He then used the below R codes to run ANOVA

```
>MyModel <- aov (hg~class)
>anova (MyModel)
```

b) From the codes above, identify the dependent and independent variables **(4 marks)**

c) Here are the output of the analysis;
Analysis of variance Table

Response: hg

	DF	Sum sq	Mean sq	Fvalue	Pr (>F)
Class	1	27.5	27.5086	Omitted	0.01798
Residuals	888	4347.4	4.8957		

The F value has been omitted from the output compute the omitted F value **(3 marks)**

d) Was the researcher right to believe that levels of haemoglobin didn't differ across the three years of study? What can you conclude from the results **(7 marks)**