ALUPE UNIVERSITY COLLEGE
...Bastion of Knowledge...
P. O.Box 845-50400 Busia(K)
principal@auc.ac.ke
Tel: +254 741 217 185
+254 736 044 469
off Busia-Malaba road

OFFICE OF THE DEPUTY PRINCIPAL
ACADEMICS, STUDENT AFFAIRS AND RESEARCH

UNIVERSITY EXAMINATIONS

2020 /2021 ACADEMIC YEAR

THIRD YEAR FIRST SEMESTER REGULAR EXAMINATION

FOR THE DEGREE OF BACHELOR OF SCIENCE (APPLIED STATISTICS WITH COMPUTING)

COURSE CODE: STA 318

COURSE TITLE: COMPUTING METHODS AND DATA ANALYSIS

DATE: 9/03/2021 TIME: 1400 – 1700 HRS

INSTRUCTION TO CANDIDATES

- SEE INSIDE

THIS PAPER CONSISTS OF 5 PRINTED PAGES PLEASE TURN OVER

STA 318

REGULAR – MAIN EXAM

STA 318: COMPUTING METHODS AND DATA ANALYSIS

STREAM: ASC **DURATION: 3 Hours**

INSTRUCTION TO CANDIDATES

Answer **ALL** questions from section A and any **THREE** from section B.

==========

SECTION A [31 Marks]. Answer ALL questions.

QUESTION ONE [15 Marks]

a) Define clearly the following terms. [4 Marks]
   i) Descriptive statistics
   ii) Case study.
   iii) Missing values.
   iv) Outliers.

b) Distinguish between nominal and ordinal types of data. [2 Marks]

c) Why is sample survey preferred as compared to carrying census? [2 Marks]

d) Give three objectives of data management process in a research project. [3 Marks]

e) Identify and describe briefly two major sources of data. [4 Marks]

QUESTION TWO [16 Marks]

a) State two advantages of using other statistical packages other than MS Excel. [2 Marks]

b) Write and describe clearly procedure and R codes used to read data from file in MS Excel into R and perform descriptive statistics of variables. Assume file name is "Family_Income.xlsx" and located on folder I:\Research_Project. [4 Marks]

c) Elaborate the term data cleaning. [3 Marks]

d) What is the difference between a database and a spread sheet? [2 Marks]

e) In an investigation to determine the risk of developing cancer amongst males and females in a certain locality, the following data was recorded.

	Developed Cancer	
Gender	Yes	No
Male	7	240
Female	4	734

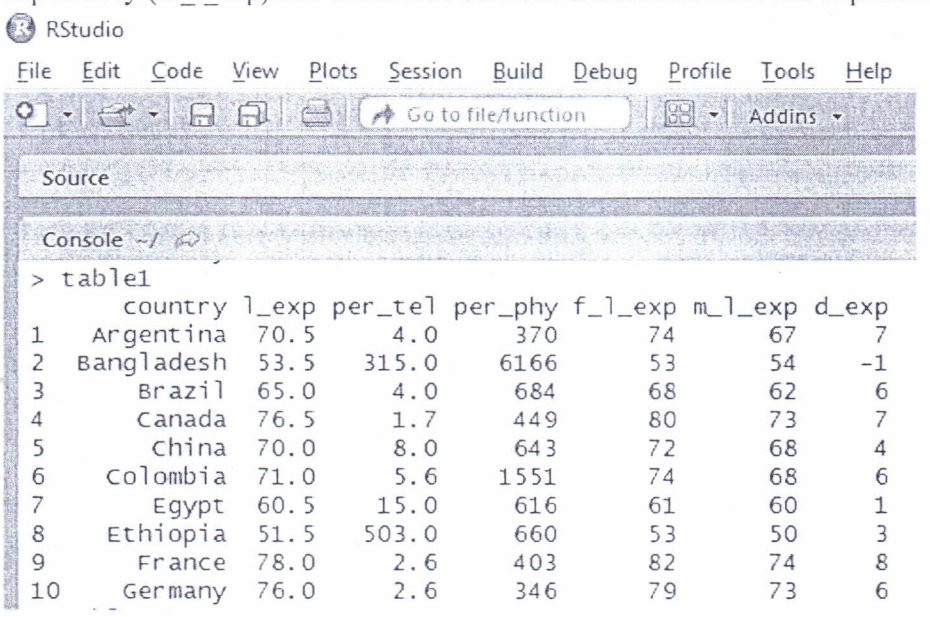
Test whether there exists any association between gender and risk of developing cancer ($\alpha = 0.05$). [5 Marks]

SECTION B [39 Marks] Answer any THREE questions]

QUESTION THREE [13 Marks]

a) Discuss the basic principles required in designing a good questionnaire. [4 Marks]

b) Consider the R output below on life expectancy at birth for ten countries with the following variables, average life expectancy at birth (l_exp), number of people per television set (per_tel), number of people physician (per_phy), female life expectancy (f_l_exp), male life expectancy (m_l_exp) and difference between female and male life expectancy(d_exp)




```
> table1
      country l_exp per_tel per_phy f_l_exp m_l_exp d_exp
1   Argentina  70.5     4.0     370      74      67     7
2  Bangladesh  53.5   315.0    6166      53      54    -1
3      Brazil  65.0     4.0     684      68      62     6
4      Canada  76.5     1.7     449      80      73     7
5       China  70.0     8.0     643      72      68     4
6    Colombia  71.0     5.6    1551      74      68     6
7       Egypt  60.5    15.0     616      61      60     1
8    Ethiopia  51.5   503.0     660      53      50     3
9      France  78.0     2.6     403      82      74     8
10    Germany  76.0     2.6     346      79      73     6
```

Write commands with brief explanations that would;

i) Assign the country's names to the 'names' attribute of this vector. Also create a vector 'l_exp', 'per_tel', 'per_phy', 'f_l_exp','m_l_exp'. [2 Marks]

ii) Bind all the variables in i) above [2 Marks]

iii) Create d_exp, the difference between female and male life expectancy. [2 Marks]

iv) Generate output above, sort and descriptive statistics [3 Marks]

QUESTION FOUR [13 Marks]

a) What is a scatter diagram? What does it show? [3 Marks]

b) A study was conducted to find out whether there is any relationship between the weight and blood pressure of an individual. The following set of data was arrived at from a clinical study.

Weight	78	86	72	82	80	86	84	89
Blood pressure	140	160	134	144	180	176	174	178

i) Sketch a scatter diagram and make the necessary comments. [4 Marks]

ii) Describe procedure you would use to plot the scatter diagram above in SPSS. [6 Marks]

QUESTION FIVE [13 Marks]

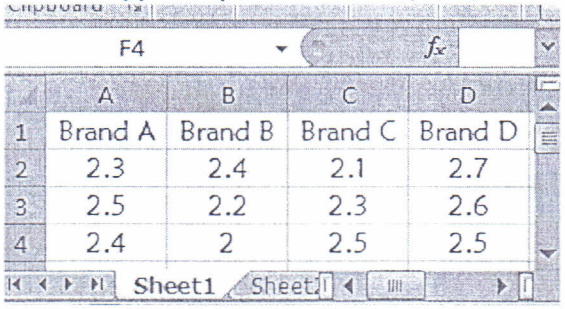
a) Why is Analysis of Variance (ANOVA) usually preferred than t – test? [2 Marks]

b) A survey was conducted to find out whether students prefer a particular type of music than the other: Classical, Rock, Pop and Jazz. Sixty students were randomly selected and asked to rate the one particular type on music, the following results were obtained. $MSS_{TR} = 9530$ and $MSS_E = 252$

i) Write H_o and H_1 [2 Marks]

ii) Sketch an ANOVA table with all the parameters. [3 Marks]

c) An experiment was conducted on muzzle velocities and the following data from four different types of powder brands (A, B, C and D) were obtained. ($\alpha = 0.05$).



F4

	A	B	C	D
1	Brand A	Brand B	Brand C	Brand D
2	2.3	2.4	2.1	2.7
3	2.5	2.2	2.3	2.6
4	2.4	2	2.5	2.5

Sheet1 Sheet2

Describe the procedure including codes how you would use to perform ANOVA in R. [6 Marks]

QUESTION SIX [13 Marks]

a) A researcher wanted to explore there are gender differences in engagement in out-of-school science activities. The variable SCIEACT is a score derived from responses to nine items on how often the student engages in particular science activity, such as watching TV programme about science and attending science club.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
SCIEACT	20	3.50	1.573	1	7
Gender	20	1.70	.470	1	2

Mann-Whitney Test

Ranks

	Gender	N	Mean Rank	Sum of Ranks
SCIEACT	Male	6	8.33	50.00
	Female	14	11.43	160.00
	Total	20		

Test Statistics[a]

	SCIEACT
Mann-Whitney U	29.000
Wilcoxon W	50.000
Z	-1.100
Asymp. Sig. (2-tailed)	.271
Exact Sig. [2*(1-tailed Sig.)]	.312[b]

a. Grouping Variable: Gender

b. Not corrected for ties.

i) Write procedure used to perform the test that produces the output above [5 marks]
ii) State the hypothesis for the test [2 marks]

b) Write a brief report about the results and make appropriate conclusions [6 marks]

QUESTION SEVEN [13 Marks]

a) i) State and discuss the steps followed when carrying out simplex algorithm. [5 Marks]
ii) A furniture makes desks, tables, and chairs. Each product needs the limited resources of lumber, carpentry and finishing; as described in the table. At most 5 tables can be sold per week

Resource	Desk	Table	Chair	Max available
Lumber (board feet)	8	6	1	48
Finishing hours	4	2	1.5	20
Carpentry hours	2	1.5	0.5	8
Max demand	Unlimited	5	Unlimited	
Price ($)	60	30	20	

Formulate a linear programming (LP) model. [3 Marks]

b) The table below shows data on weight gained in grams by patients from two different health facilities.

Patient P	1003.8	905.8	1011.4	690.2	1086.2	1001.07	1302.3	595.2
Patient Q	303.5	1103.4	1515.3	1522.2	581.7	506.8		

i) Write an R code to test whether the distribution of weights gained by patients in the two health facilities is the same or not using Mann-Whitney U test($\alpha = 0.05$) [3 Marks]
ii) Assume that from analysis output you get a p-value of 0.0047, what do you conclude? [2 Marks]