AFRICA UNIVERSITY

COLLEGE OF BUSINESS, PEACE, LEADERSHIP AND GOVERNANCE DEPARTMENT

NMEC209: COMPUTER AND STATISTICAL ANALYSIS

END OF SEMESTER EXAMINATION

MAY 2021

LECTURER: DR SAUNGWEME

TIME :7 HOURS

INSTRUCTIONS

Answer **ONE** question

The marks allocated to each question are shown at the end of the question

QUESTION 1

(a) In carrying out linear regression analysis, certain assumptions have to be satisfied before interpreting the results. In the diagrams below, identify: (i) the assumption being tested; (ii) the test being used; and (iii) interpret the results of the two groups – Groups 1 and 2.

(6 marks)

a) Group 1 — Restricted diet
b) Group 2 — Ad libitum diet
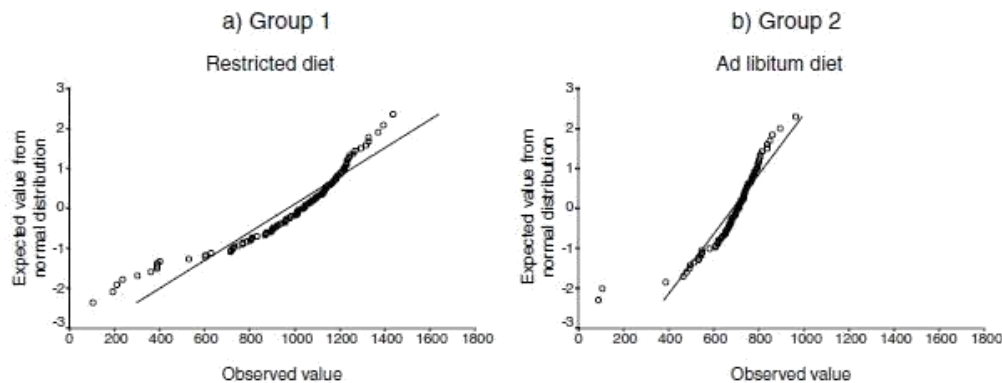
(b) A clinical trial is run to compare weight loss programs and participants are randomly assigned to one of the comparison programs and are counselled on the details of the assigned program. Participants follow the assigned program for 8 weeks. The outcome of interest is weight loss, defined as the difference in weight measured at the start of the study (baseline) and weight measured at the end of the study (8 weeks), and measured in pounds. By interpreting each of the tables below, test if there is a statistically significant difference in the mean weight loss among the four diets? [18 marks]

Weight loss program	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
					Descriptive statistics			
low calories	5	6.60	2.302	1.030	3.74	9.46	3	9
low fat	5	3.00	1.581	.707	1.04	4.96	1	5
low carbohydrates	5	3.40	1.140	.510	1.98	4.82	2	5
control	5	1.20	1.643	.735	-.84	3.24	-1	3
Total	20	3.55	2.544	.569	2.36	4.74	-1	9

Test of Homogeneity of Variances

Weight loss

Levene Statistic	df1	df2	Sig.
.695	3	16	.568

ANOVA

Weight loss	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	75.750	3	25.250	8.559	.001
Within Groups	47.200	16	2.950		
Total	**122.950**	**19**			

Homogeneous Subsets

Weight loss

	Weight loss program	N	Subset for alpha = 0.05	
			1	2
Tukey HSD[a]	control	5	1.20	
	low fat	5	3.00	
	low carbohydrates	5	3.40	
	low calories	5		6.60
	Sig.		.220	1.000
Means for groups in homogeneous subsets are displayed.				
a. Uses Harmonic Mean Sample Size = 5.000.				

						95% Confidence Interval	
	(I) Weigh loss program	(J) Weigh loss program	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
Tukey HSD	low calories	low fat	3.600[*]	1.086	.021	.49	6.71
		low carbohydrates	3.200[*]	1.086	.042	.09	6.31
		control	5.400[*]	1.086	.001	2.29	8.51
	low fat	low calories	-3.600[*]	1.086	.021	-6.71	-.49
		low carbohydrates	-.400	1.086	.982	-3.51	2.71
		control	1.800	1.086	.377	-1.31	4.91
	low carbohydrates	low calories	-3.200[*]	1.086	.042	-6.31	-.09
		low fat	.400	1.086	.982	-2.71	3.51
		control	2.200	1.086	.220	-.91	5.31
	control	low calories	-5.400[*]	1.086	.001	-8.51	-2.29
		low fat	-1.800	1.086	.377	-4.91	1.31
		low carbohydrates	-2.200	1.086	.220	-5.31	.91

Multiple Comparisons

Dependent Variable: weight loss

*. The mean difference is significant at the 0.05 level.

(c) The total revenue flows of Halsteds Mutare were recorded for the period from 2016 to 2020 and a summary of the revenues flows, in millions (US$) is presented in table below:

Quarter	Year				
	2016	2017	2018	2019	2020
1	1026.10	1172.50	1286.90	1462.40	1776.25
2	1056.20	1249.60	1317.10	1452.50	1808.25
3	1182.30	1346.70	1449.20	1631.60	1941.75
4	2861.40	3402.80	3893.30	4200.70	4128.75

Use the information in the table to present the four-point de-seasonalised time series.

[36 marks]

TOTAL 60 MARKS

QUESTION 2

(a) Three road repair compounds were tested at four different locations of particular highway. At each location four different sections of the road were repaired with each section using one of three components. Data was then collected on the number of days on traffic usage before additional repair was required.

		LOCATION			
		1	2	3	4
Compound	A	99	73	85	103
	B	82	72	85	97
	C	81	79	82	86

Using SPSS or otherwise, test:

(i)   if there is a significant difference in the location.   [20 marks]

(ii)   if there is a significant difference in the compound.   [10 marks]

(b) An experiment of car cleaning was performed to examine the impact of extroversion, sex and age on the amount of time spent in car cleaning. The following results were obtained:

Correlations

		gender	age in years	extroversion	minutes per week
gender	Pearson Correlation	1	-.053	.403**	.661**
	Sig. (2-tailed)	.	.744	.010	.000
	N	40	40	40	40
age in years	Pearson Correlation	-.053	1	.397*	.234
	Sig. (2-tailed)	.744	.	.011	.147
	N	40	40	40	40
extroversion	Pearson Correlation	.403**	.397*	1	.670**
	Sig. (2-tailed)	.010	.011	.	.000
	N	40	40	40	40
minutes per week	Pearson Correlation	.661**	.234	.670**	1
	Sig. (2-tailed)	.000	.147	.000	.
	N	40	40	40	40

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.799a	.638	.608	13.021

a. Predictors: (Constant), age in years, gender, extroversion

ANOVAb

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10746.290	3	3582.097	21.126	.000a
	Residual	6104.085	36	169.558		
	Total	16850.375	39			

a. Predictors: (Constant), age in years, gender, extroversion
b. Dependent Variable: minutes per week

Coefficientsa

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	Upper Bound	Collinearity Statistics Tolerance	VIF
1	(Constant)	17.905	5.319		3.366	.002	7.137	28.674		
	extroversion	.708	.127	.670	5.566	.000	.451	.966	1.000	1.000
2	(Constant)	15.680	4.436		3.534	.001	6.691	24.669		
	extroversion	.509	.115	.482	4.423	.000	.276	.743	.838	1.194
	gender	19.180	4.473	.467	4.288	.000	10.118	28.243	.838	1.194

a. Dependent Variable: minutes per week

Required:

(i) To describe the assumptions of linear multiple regression analysis and how these can

be tested using SPSS. [12 marks]

(ii) To interpret the regression results presented in all the tables above. [18 marks]

TOTAL 60 MARKS

QUESTION 3

(a) For each of the following scenarios, determine and justify a statistical test that would be

suitable:

(i) Assess the association between recreational reading interest and academic reading interest. Both types of interest are measured with Garfield reading interest subscale which ranges from 10 = low to 40 = high. [4 marks]

(ii) Does high school student's decision to apply for college admission (yes vs no) differs depending upon whether that student's mother attended college (yes vs no)?

[4 marks]

(iii) Does the racial/ethnic distribution of our sample match the expected population distribution, that is, are the sample and population frequencies or percentages consistent? [4 marks]

(iv) Laser toner cartridges for the Hewlett-Packard Laserjet 1200 series are estimated to last approximately 2000 pages per cartridge assuming 5% coverage per page. To test this advertising claim, 15 toner cartridges were purchased and installed in 15 HP Laserjet 1200 printers throughout a school district. Number of pages was recorded for each cartridge until visible lines of non-print became apparent. Below are the numbers of pages recorded. Is there any evidence that the advertiser' claim of 2000 pages is supported?

1985   1675   1548   1756   1865   1235   1498   1195   1611   1754 2056
1454   1332   1444   1669

[5 marks]

(v) Who goes to the local library? Local patrons are asked to complete a brief questionnaire that elicits various demographic variables. One variable is patron sex. Library researchers anticipated a 50/50 split between male and female patrons, but actual numbers show the distribution is 63 females and 37 males. How could this difference be tested against the expected 50/50 distribution? [4 marks]

(vi) Is there a difference in level of recreational reading interest between students at Vurombo Elementary School and Runyararo Elementary School? Recreational reading interest is measured using the Garfield reading interest scale which ranges from 10 = low to 40 = high. [4 marks]

(b) An experiment was performed to test the existence or nonexistence of a relationship between two design groups and the following results were obtained:

Ranks

	de...	N	Mean Rank	Sum of Ranks
sales	1	5	6.20	31.00
	2	5	4.80	24.00
	Total	10		

Test Statistics[b]

	sales
Mann-Whitney U	9.000
Wilcoxon W	24.000
Z	-.736
Asymp. Sig. (2-tailed)	.462
Exact Sig. [2*(1-tailed Sig.)]	.548

a. Not corrected for ties.

b. Grouping Variable: design

Required:

(i) Discuss the justification for using the Mann-Whitney U test in this experiment.

[4 marks]

(ii) Test at 5% level of significance if there is a significant difference in sales between the design group 1 and group 2. [8 marks]

(c) Freddy too twelve Africa University students and placed them on a liquid diet in a weight-reducing plan. Freddy then recorded the weights before and after the diet. The data are shown in the table below.

Prison	1	2	3	4	5	6	7	8	9	10	11	12
Before	186	171	177	168	191	172	177	191	170	171	188	187
After	188	177	176	169	196	172	165	190	165	180	181	172

Applying the Wilcoxon Signed-Rank test, Freddy found the following results:

```
NPar Tests

Wilcoxon Signed Ranks Test

                              Ranks
                                          N    Mean Rank   Sum of Ranks
Weight Before Diet  Negative Ranks        5a       5.50          27.50
- Weight After Diet Positive Ranks        6b       6.42          38.50
                    Ties                  1c
                    Total                 12
  a. Weight Before Diet < Weight After Diet
  b. Weight Before Diet > Weight After Diet
  c. Weight After Diet = Weight Before Diet


          Test Statisticsb
                         Weight Before
                         Diet - Weight
                           After Diet
  Z                          -.490a
  Asymp. Sig. (2-tailed)      .624
    a. Based on negative ranks.
    b. Wilcoxon Signed Ranks Test
```

Required:

(i) Briefly explain the justification of using a nonparametric test in this experiment.   [3 marks]

(ii) Using 5% significance level, test the hypothesis that the plan was successful or not.

[6 marks]

(ii) Explain any 2 possible shortfalls in this experiment and how they can be improved.

[4 marks]

(d) The element data for Africa University over six years is given as:

Year	1	2	3	4	5	6
Number of students	20500	20200	19500	19000	19100	18800

Forecast the enrolment for the seventh year.   [10 marks]

TOTAL 60 MARKS

END OF EXAMINATION