AFRICA UNIVERSITY

COLLEGE OF BUSINESS, PEACE, LEADERSHIP AND GOVERNANCE DEPARTMENT

OF COMPUTER SCIENCE AND INFORMATION SYSTEMS

MAIN EXAMINATION

COURSE CODE:	NMMS204
NARRATION:	QUANTITATIVE ANALYSIS II
TIME ALLOWED:	7 HOURS

INSTRUCTIONS TO CANDIDATES

1) ANSWER ANY ONE QUESTION

2) ALL QUESTIONS CARRY EQUAL MARKS

QUESTION 1

(a) For each of the following, identify the independent variable and dependent variable and determine whether each is categorical or quantitative;

(i) Among 3rd grade students, there will be a difference in graded reading performance scores between males and females. (5 marks)

(ii) Students in larger classes tend to score lower on standardised mathematics tests. (5 marks)

(iii) Students whose parents are educators will have a higher academic self-efficacy than students whose parents are not educators. The academic self-efficacy scale ranges from 5 = low to 15 = high. (5 marks)

(iv) For females in public schools, researchers found that one's mathematics attitude predicts well one's mathematics achievement. (Mathematics attitude is a latent variable and is formed as a composite score from responses to several indicators and has a range of 20 = low to 85 = high). (5 marks)

(b) Explain the difference between the coefficient of determination and correlation coefficient. (9 marks)

(c) An investigator claims that the average salary for a group of workers is $35 000. To test this claim, current salaries for 474 workers were recorded. A one sample test was run using SPSS and the following results were achieved.

One sample statistics

	N	**Mean ($)**	**Std deviation ($)**	**Std error mean ($)**
Current salary	474	34 419.57	17 075.66	784 311

One sample test (Test value = $35 000)

	t		**Sig (2-tailed)**	**Mean difference**	**95% confidence interval of the difference**	
					Lower	**Upper**
Current salary	-0.740		0.460	-$580.432	-$2121.60	-$960.77

(i) Briefly interpret the results reported in table entitled "One sample statistics". (6 marks)

(ii) Using the information presented in table entitled "One sample test (Test value = $35 000)", test at 5% level of significance the null hypothesis that the average current salary is $35 000 against the alternative hypothesis that the average current salary is not equal to $35 000. (9 marks)

(d) An investigator claims that the average current salary for men and women workers is equal. Current salaries for 216 female and 258 male workers were recorded and the following results were achieved using SPSS;

Group statistics

	Gender	N	Mean ($)	Std deviation ($)	Std error mean ($)
Current salary	Female	216	26 031	2 558	514.2
	Male	258	41 441	19 499	1213.9

Independent sample test (t-test for equality of mean)

		t	df	Sig (2-tailed)	Mean diff ($)	Std error diff ($)	95% confidence interval of the difference	
							Lower ($)	Upper ($)
Current salary	Equal variance assumed	-10	472	0.00	-15 409	1407.9	-18176	-12643
	Equal variance not assumed	-11	344	0.00	-15 409	1318.4	-18 000	-12816

Note: Mean diff =Mean difference; Std error diff = Std error difference

(i) Discuss the results reported in Group statistics table. (6 marks)

(ii) Given the results above, is there evidence that the average current salary is equal for both female and male workers? (10 marks)

[Total 60 marks]

QUESTION 2

(a) An investigator in the insurance industry wants to determine whether the claim amount claimed is influenced by the type of claim. Data was recorded for the cost of claims in thousand dollars and the type of claim, and the following results were generated from SPSS;

One way ANOVA

	Sum of squares	df	Mean square	F	Sig
Between groups	25 043 718	4	6 260 929	412.12	0.000
Within groups	66 995 929	4 410	15 191		
Total	92 039 647	4 414			

(i) Discuss the difference between ANOVA and contingency table analysis. (6 marks)

(ii) Discuss the conditions that should hold before interpreting ANOVA results presented above. (15 marks)

(iii) Using 5% level of significance, test the hypothesis that the claim amount claimed depends on the type of claim? (10 marks)

(b) An investigator in the insurance industry would like to find out whether there is a relationship between gender and type of the claim. Data on type of claim against gender was recorded and the following results were achieved through SPSS;

Chi-Square tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-square	4.526	4	0.340
Likelihood Ratio	4.528	4	0.339
Linear –by-Linear association	0.667	1	0.414
N of valid cases	4415		

Is there enough statistical evidence to conclude that the type of the claim is related to the gender of the claimant? (6 marks)

(c) The results below were generated from quantitative data analysis software package, EVIEWS 7. The results are based on the study to examine the impact of domestic and external debt on a particular economy.

Dependent Variable: GDP
Method: Least Squares
Sample (adjusted): 1981Q2 1988Q3
Included observations: 30 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.001557	0.001983	-0.785207	0.4397
DODEBT	-0.017489	0.038342	-0.456129	0.6522
EXDEBT	-0.659846	0.184305	-3.580187	0.0014
INVEST	0.319789	0.116443	2.746319	0.0110
LABOUR	0.041902	0.012852	3.260445	0.0032

R-squared	0.814040	Mean dependent var	0.006803
Adjusted R-squared	0.552287	S.D. dependent var	0.008709
S.E. of regression	0.005828	Akaike info criterion	-7.301414
Sum squared resid	0.000849	Schwarz criterion	-7.067881
Log likelihood	114.5212	Hannan-Quinn criter.	-7.226705
F-statistic	9.943409	Durbin-Watson stat	2.352489
Prob(F-statistic)	0.000059		

Where:

GDP = Log of gross domestic product

DODEBT = Log of domestic debt

EXDEBT = Log of external debt

INVEST = Log of investment

LABOUR = log of labour

REQUIRED

Interpret the results considering that GDP is the dependent variable. (23 marks)

[Total 60 marks]

QUESTION 3

(a) A logistic regression analysis was employed to analyse the socio-economic factors that influence people's decision to become fishermen. The standard model of logit estimation was specified as follows;

$$P_i = Prob(Y_i = 1) = \frac{1}{1 + e^{-(\beta_\alpha + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})}}$$

Where: Y_i = 1 when a person decided to become fishermen and 0, otherwise

P_i = is the probability that a person becomes a fisherman

$\beta_\alpha - \beta_k$ = coefficients of the independent variables

$X_{1i} - X_{ki}$= factor variables from respondent I (independent variables)

The logistic regression model was estimated and the results are presented as follows:

Dependent variable: Being a fishermen				
Variable	**Coefficient**	**Standard Error**	**T-value**	**Prob**
Intercept	3.262	2.651	1.230	0.219
Age of respondent	-0.020	0.067	-0.298	0.765
Years of education of respondent	-0.397	0.190	-2.090	0.037
Marital status	0.286	1.204	0.238	0.812
Household size	0.546	0.300	1.817	0.069
Monthly income	0.002	0.001	1.277	0.202
Other income	-3.475	1.351	-2.571	0.010
Access to credit	3.184	1.595	1.965	0.050
LR statistic	27.77			
Prob (LR statistic)	0.00003			
McFadden R-squared	0.73177			

REQUIRED

Interpret the above logistic regression output. (25 marks)

(b) The monthly sales data for a local retail shop was recorded and presented in table below:

Month	**Year 1**	**Year 2**	**Year 3**	**Year 4**
January	280	382	480	536
February	271	344	452	456
March	290	365	433	446
April	411	512	552	573
May	281	393	352	365
June	308	394	318	351
July	284	361	263	292
August	384	466	218	358
September	449	540	333	378
October	518	571	342	394
November	590	599	294	405
December	801	685	402	539

(i) Use the data in table to create quarterly values for each year. (4 marks)

(ii) Use the information in b(i) to present a de-seasonalised series for the given sales data. (15 marks)

(c) (i) Explain the effect of the presence of outliers in regression analysis. (6 marks)
(ii) Clearly explain the Box plot procedure for testing the existence of outliers in a given data set. (6 marks)

(d) Explain any two statistical ways that can be used to test for normality in data sets. (4 marks)

[Total 60 marks]

END OF EXAMINATION