MAASAI MARA UNIVERSITY

REGULAR UNIVERSITY EXAMINATIONS
2023/2024 ACADEMIC YEAR
THIRD YEAR SECOND SEMESTER

SCHOOL OF PURE APPLIED AND HEALTH SCIENCES
BSC COMPUTER SCIENCE PROGRAMME

COURSE CODE: COM 3207

COURSE TITLE: COMPUTER APPLICATIONS 2

DATE: 6th June 2024 **TIME: 1100-1300 HRS**

INSTRUCTIONS TO CANDIDATES

Answer Question ONE and any other TWO questions

QUESTION ONE

a. Group the following variables into the appropriate level of measure **(3 marks)**

   i. Medal award in an Olympic game
   ii. Temperature in ^0C
   iii. Religion

b. Give the meaning of the following terms as used in data management **(4 marks)**

   i. Selection
   ii. Sorting
   iii. Aggregating
   iv. Restructuring

c. A data set contains Five variables (Gender, Age, Weight, Religion and Marital_status). The variables Gender, Religion and Marital status were coded as follows: Gender (1 = Female, 2=Male), Religion (1 = Christian, 2= Catholic, 3 = Muslim, 4 = Other religions), Marital status (1=Single, 2 = Married, 3 = Widowed, 4 = Separated). Write down SPSS code that will do the following;

   i. Select Females who are Muslims and weigh at least 35 Kg **(2 marks)**

   ii. Filter out Married youths (A youth is a person between 18 – 35 year) **(2 marks)**

   iii. A 32-year-old single lady is looking for a Muslim guy who is older than her. Kindly write a code to assist him with this for the given data set **(2 marks)**

d. There is a believe that weight difference is due to gender. To ascertain this believe you decided to pick both 15 male and female students to participate in your study. The output for your analysis was as follows;

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					
								95% Confidence Interval of the Difference	
		F	Sig.	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
weight of individuals in Kg	Equal variances assumed	.031	.862	28	.695	-1.600	4.045	-9.887	6.687
	Equal variances not assumed			27.998	.695	-1.600	4.045	-9.887	6.687

i. State the null and alternative hypothesis for this test **(2 marks)**

ii. Which assumption is performed on this test? Comment on it**.** **(2marks)**

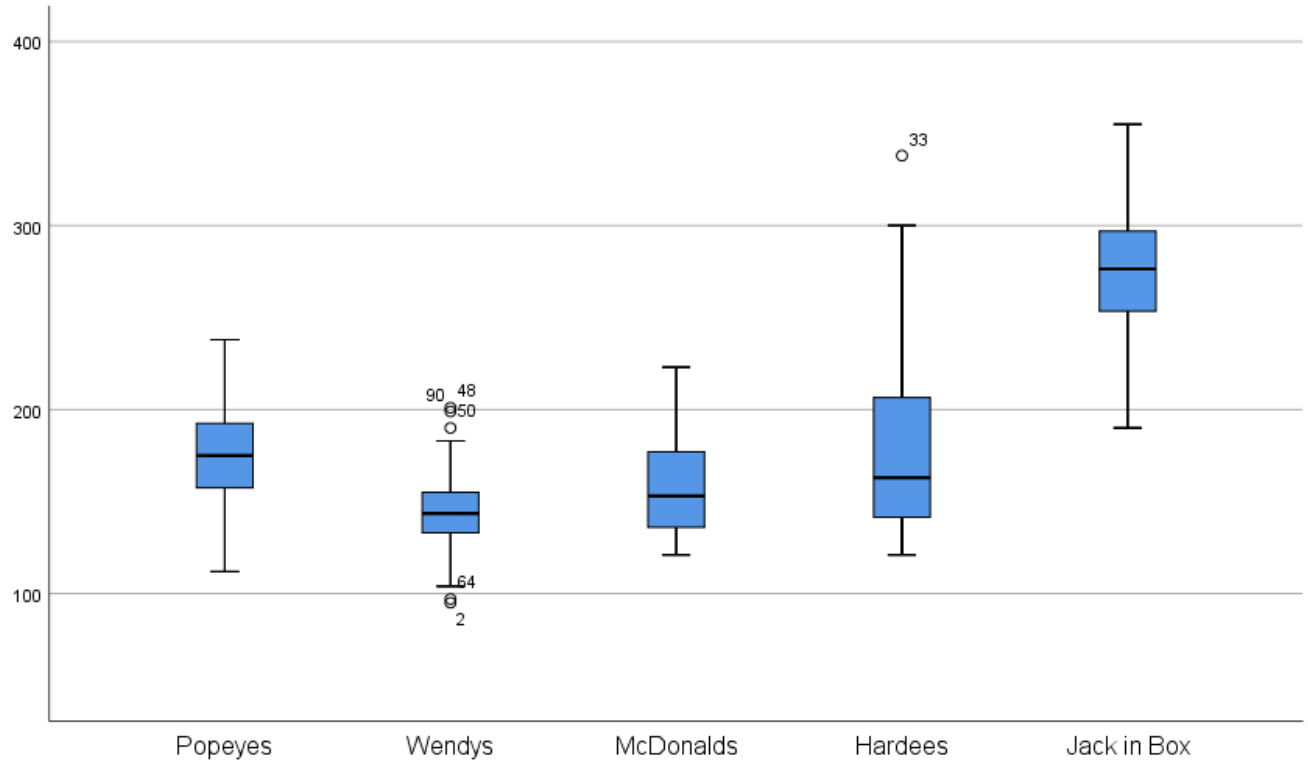
iii. Which assumption is ignored and what is the effect? **(2 marks)**

iv. Using the confidence interval, comment on this believe. **(2 marks)**

v. Will the decision to reject the null hypothesis lead to any type of error? If yes which type of error **(2 mark)**

vi. Which test is appropriate if the required assumptions for this test does not hold **(1 marks)**

e. The figure below shows the results of a statistical analysis of number of customers who visited a restaurant in a certain month use it to answer the questions that follows;



i. Which restaurant had the highest number of extreme observations? **(1 mark)**

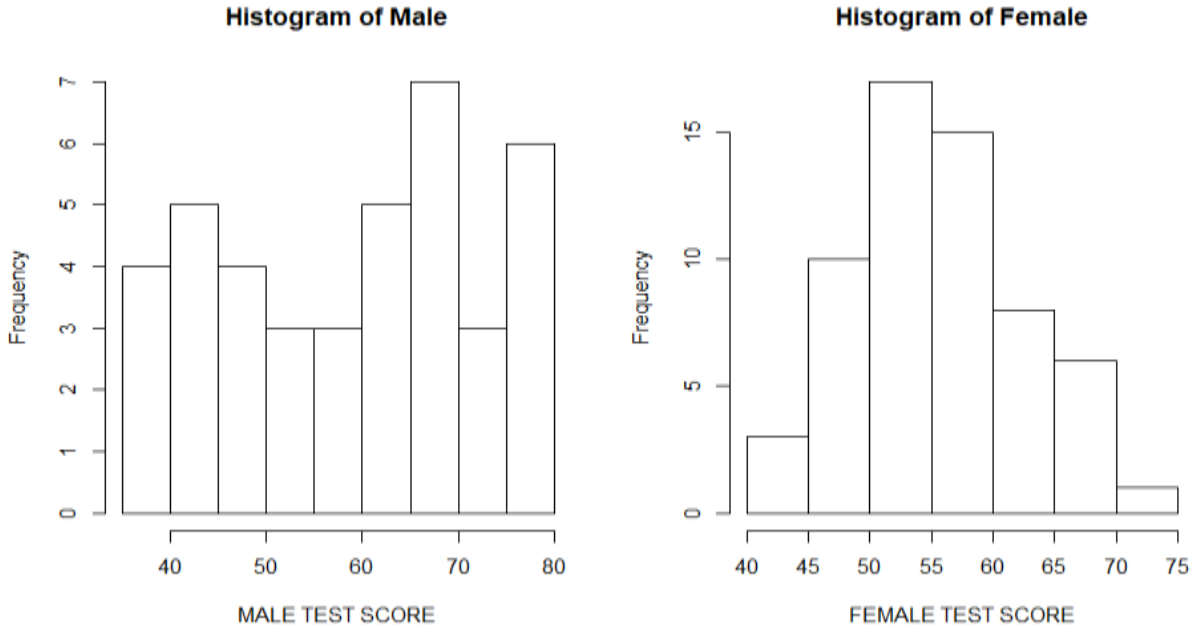
ii. In which restaurants were the number of customers who visited normally distributed? **(2 marks)**

Give two reasons why standard deviation is not a good measure of dispersion in measuring the average variation for Hardees restaurant **(2 marks)**

iii. Giving appropriate reason suggest the best measure of central tendency for Wendy's restaurant **(1 mark)**

QUESTION TWO

a. The figure below shows the distribution of scores for students in a class, use it to answer the questions that follows;



i. Describe the distribution of the male and female test scores. **(2 marks)**

ii. Can we use independent sample t-test to compare the scores for males and female? Kindly justify your reasoning. **(2 marks)**

iii. From the above plot which group of students do you think performed well? Kindly justify your reasoning. **(2 marks)**

b. Below is the regression output to investigate whether height of an individual affects his/her weight. Use the outputs below to answer the questions.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	620.575	1	620.575	6.129	.020[b]
	Residual	2835.292	28	101.260		
	Total	3455.867	29			

a. Dependent Variable: weight of individuals in Kg

b. Predictors: (Constant), height of individuals inches

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	114.641	19.037			.000
	height of individuals inches	-.787	.318	-.424		.020

a. Dependent Variable: weight of individuals in Kg

i. Calculate the t values for table 2. **(2 marks)**
ii. Write down the regression equation **(2 mark)**
iii. What is the weight for an individual who is 62.5 inches tall? **(1 mark)**
iv. State three assumptions for this test **(3 marks)**
v. Can we conclude that height is a significant factor for studying weight? Justify. **(2 marks)**
vi. Calculate the R^2 value and interpret it **(4 marks)**

QUESTION THREE

a. Differentiate using examples between descriptive and inferential statistics **(4 marks)**

b. A researcher wanted to investigate if the average number of crimes reported per day is different between Narok, Nakuru and Kericho Counties. The researcher recorded the number of crimes reported in the three town in a single week. A one-way analysis of variance test was carried out on the data set and the results were as illustrated below;

Descriptives

Number of crimes reported

County	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					L. B	**U. B.**
Narok	7	11.00	2.449	.926	8.73	13.27
Nakuru	7	21.29	1.799	.680	19.62	22.95
Kericho	7	16.14	1.574	.595	14.69	17.60
Total	21	16.14	4.693	1.024	14.01	18.28

ANOVA				
Source of variatio	Sum of Squares	df	Mean Squares	F
Between groups	370.286			
Within groups				
Total	440.571			

i) State the null and alternative hypothesis for the study **(2 marks)**

ii) Complete the ANOVA table above **(7 marks)**

iii) Based on the test results above, at 95% level of confidence is there sufficient evidence to show that the number of crime reported in the three towns is different. **(3 marks)**

iv) State two assumptions that must be satisfied in order to carry out the above test **(2 marks)**

v) Give the corrective measure that would be taken to test the hypothesis made in the study if each of the assumptions in (iv) are violated **(2 marks)**

QUESTION FOUR

a. From question (2b), the investigator was not satisfied and decided to add another variable Age to the model. The output now looks as follows;

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	751.354	2	375.677	3.751	.037[b]
	Residual	2704.513	27	100.167		
	Total	3455.867	29			

a. Dependent Variable: weight of individuals in Kg

b. Predictors: (Constant), Age, height of individuals inches

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	136.265	26.770		5.090	.000
	height of individuals inches	-.873	.325	-.470	-2.686	.012
	Age	-.428	.374	-.200	-1.143	.263

a. Dependent Variable: weight of individuals in Kg

i. Which type of model is fitted above? **(1 marks)**

ii. Give any three assumptions for this type of regression **(3 marks)**

iii. By stating the appropriate null and alternative hypothesis, comment on the adequacy of this model **(4 marks)**

iv. Comment on the significance of the new variable added to the model **(2 marks)**

b. Use the following logistic regression output to answer the questions below. The dependent variable is cancer.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Gender(1)	-2.077	.925	5.035	1	.025	.125
	Age	.130	.098	1.744	1	.187	1.139
	Height	-.033	.091	.134	1	.714	.967
	Weight	-.002	.047	.002	1	.964	.998
	Constant	-1.307	9.569	.019	1	.891	.271

a. Variable(s) entered on step 1: Gender, Age, Height, Weight.

i. What are the assumptions of this test **(2 marks)**

ii. Based on the analysis, explain how the variable gender influence cancer **(2 marks)**

iii. Using the variable age, explain what Exp(B) indicates **(2 marks)**

iv. What is your comment on the significance of this variables **(1 mark)**

c. Under what circumstances do we use Z test instead of t test **(2 marks)**

d. What is type II error and how can we control it **(1 mark)**

>>>>>>>END>>>>>>>WISH YOU WELL>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>