THE CATHOLIC UNIVERSITY OF EASTERN AFRICA

A. M. E. C. E. A

P.O. Box 62157

00200 Nairobi - KENYA

Telephone: 891601-6

Ext 1022/23/25

MAIN EXAMINATION

SEPTEMBER –DECEMBER 2021

FACULTY OF SCIENCE

DEPARTMENT OF MATHEMATICS

REGULAR PROGRAMME

MAT 365: TESTS OF HYPOTHESES II

Date: DECEMBER 2021	**Duration: 2 Hours**
INSTRUCTIONS: Answer Question ONE and any TWO Questions	

Q1. a) (i) Define non-parametric test **(3 marks)**

(ii) What are the advantages and disadvantages of non-parametric tests? **(4 marks)**

b) According to the genetic theory, the number of colour strain pink, white and blue in a certain flower should appear in the ratio 3:2:5. For 100 plants, the results were as follows:

Colour	Pink	White	Blue
Number of plants	24	14	62

Perform the χ^2 - test at 5% level of significance to investigate the validity of the theory **(6 marks)**

c) The following data shows the people's preference for red or white wine of 1000 randomly chosen individual form Nairobi East and Nairobi West areas

		Preferences	
		Red	White
Area	East	238	158
	West	304	300

Test whether the preference for red or white wine lives at $\alpha = 2.5\%$. **(6 marks)**

d) Briefly describe the sign test. **(5 marks)**

e) The following is an arrangement of men M and women W, lined up, at the over 60 years old entrance, to purchase tickets for a Sofa Paka versus Kong'lo match
MMMMWWWMMMMMMMWWWMMWWWW

Test the randomness of the arrangement at the 5% level of significance. **(6 marks)**

Q2. a) Given a set of paired observations $(x_i, y_i), i = 1, 2, \mathrm{K}, n$, show that the rank correlation coefficient is given by $r_s = \dfrac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$ where d_i is the difference between ranks assigned to x_i and y_i. **(10 marks)**

b) The following figures released by the Federal Trade Commission, show the milligrams of tar and nicotine found in 10 bounds of cigarettes.

Cigarettes Brand	Tar content	Nicotine content
Viceroy	14	0.9
Marlboro	16	1.1
Chesterfield	28	1.6
Kool	17	1.3
Kent	15	1.0
Raleigh	13	0.8
Old Gold	24	1.5
Philip Morris	25	1.4
Oasis	18	1.2
Players	31	2.0

i) Calculate the rank correlation coefficient to measure the degree of relationship between the tar and nicotine content in cigarettes **(6 marks)**

ii) Stating your hypotheses and using a 5% level of significance, test the significance of your value. **(4 marks)**

1. a) The students at a medical seminar on smoking comprised of 4 smokers and 6 non-smokers. Their pulse rates, in beats per minute, were measured with the following results.

Non-smokers	60	63	68	78	91	94
Smoker	64	76	92	102		

Use Kolmogorov-Smirnov test to test whether the two samples have different distribution functions. **(8 marks)**

b) The weights (in grams) of species of mice from a certain area were found to be:

106	90	95	80	99	119	83	100	88	100	70

It is thought that the median of the population form which this sample is taken is 100. Test, at the 10% level, the following hypothesis

i) H_0: median = 100

H_1: median < 100

ii) H_0: median = 100

H_1: median $\neq$ 100 **(12marks)**

2. a) The dice is rolled 60 times. The results are shown below:

Score	1	2	3	4	5	6
Frequency	10	11	9	6	10	14

Test whether the dice is fair at 5% level of significance **(10marks)**

b) The PASS leader who spends a lot of time wearing high heels counts the number of kilometers she can walk in two different pair of shows before her feet get sore. Random samples of n_1=7 days when wearing pair A and n_2= 9 days when wearing pair B are taken. The table shows the number of kilometers she walked before her feet started to hurt.

	Kilometers walked							
Pair A	1.5	2.3	2.8	1.9	2.4	2.1		
Pair B	1.1	2.2	1.8	2.3	0.9	1.4	1.7	1.6

Using the Wilcoxon Rank Sum test at 5% level of significance, test whether there is a difference in the distance she can walk in the two different pairs of shoes. **(10 marks)**

3. a) Define a contingency table? **(2 marks)**

b) Consider an $r \times s$ contingency table in which the $(i-j)^{th}$ cell frequency is n_{ij}. Show that to test the null hypothesis of homogeneity we can use the statistics

$$Q = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{\left(n_{ij} - \frac{n_{i\bullet}n_{\bullet j}}{n}\right)^2}{\left(\frac{n_{i\bullet}n_{\bullet j}}{n}\right)}$$

Where $n_{i\bullet}$ and $n_{\bullet j}$ are the marginal total of $i-th$ row and $j-th$ column respectively.

(8 marks)

c) 300 people of different ages were interviewed and asked which genre of books they mostly read (fiction/non-fiction/science fiction). The results are as follows:

		Book type			
		Fiction	**Non-fiction**	**Science fiction**	**Total**
Age	0-25 years	23	16	41	80
	26-50 years	54	38	38	130
	51+ years	29	43	18	90
Total		**100**	**97**	**97**	**300**

Test at the 5% significance level, determine whether the type of book is independent of age. **(10 marks)**

END