AMREF INTERNATIONAL UNIVERSITY

SCHOOL OF PUBLIC HEALTH

DEPARTMENT OF COMMUNITY HEALTH

END OF SEMESTER EXAMINATION APRIL 2024

HSM 712: **BIOSTATISTICS**

DATE: April 2024

TIME: Three Hours **Start:** 1600 Hours **Finish**: 1900 Hours

INSTRUCTIONS

1. This exam is marked out of 100 marks
2. This Examination comprises TWO Sections
3. Section A: Compulsory Question (25 marks)
4. Section B: Long Answer Questions (75 marks)
5. All questions in Section A are compulsory and Answer any THREE questions in Section B
6. This online exam shall take 3 Hours
7. Late submission of the answers will not be accepted
8. Ensure your web-camera is on at all times during the examination period
9. No movement is allowed during the examination
10. Idling of your machine for 5 min or more will lead to lock out from the exam
11. The Learning Management System (LMS) has inbuilt integrity checks to detect cheating
12. Any aspect of cheating detected during and or after the exam administration will lead to nullification of your exam
13. In case you have any questions call the invigilator for this exam on Tel. +254722774221/ +254725984499 and the Head of Department on Tel +254720573449
14. For adverse incidences please write an email to: amiu.examinations@amref.ac.ke and jarim.omogi@Amref.ac.ke and michel.mutabazi@mcampus.amref.ac.ke

Question 1

a) Consider the data given in the table below.

Heights in cm	Number of patients
100 -104	12
105 - 109	15
110 -114	3
115 - 119	11
120 - 124	9
Total	50

Determine and interpret the coefficient of variation (6 marks)

b) Consider the following contingency table:

	Boys	Girls	Total
Drinking	60	10	70
Not Drinking	20	110	130
Total	80	120	200

Test the hypothesis that the gender and drinking are independent. (7 marks)

c) The following table shows the scores (X) in Anatomy and scores (Y) in Physiology of 10 students in Nursing program.

X	78	45	36	78	62	90	65	75	39	41
Y	84	55	50	60	82	86	58	60	47	51

Find the rank correlation coefficient (6 marks)

d) Systolic blood pressure of 100 males was taken. Mean blood pressure was found to be 128 mm of mercury (Hg) and standard deviation of 13 mm of mercury. Find the 95% confidence limits of blood pressure within which the population mean would lie. (6 marks)

SECTION B: ANSWER ANY THREE (3) QUESTIONS

Question 2

The following table gives the heights and weights of 12 patients chosen at random from a certain health facility.

Height X in cm.	168	150	174	144	158	168	177	156	150	161	156	163
Weight Y in Kg.	77	75	90	67	78	84	89	80	66	73	69	76

a) Calculate the Pearson's coefficient of correlation between the two variables and interpret the result. (8 marks)
b) Calculate the determination coefficient and interpret the finding. (4 marks)
c) Find the best regression line that can be used to predict the weight of a patient when we know his height, that is Y=a + b X (8 marks)
d) Estimate or predict the weight of a patient whose height is 150 cm. (5 marks)

Question 3

a) A certain drug is claimed to be effective in curing flu. In an experiment of 164 persons, a half of them were given the drug and a half were given sugar pills. The patient reaction to the treatment is as follows. Of the patients on test drug, 52 were cured, 10 got worse and 20 showed no change. Of the patients who were on sugar pills, 44 were cured, 12 got worse and 26 showed no change.

i. Construct a 2 x 3 table of the results for the patients on different treatments. (6 marks)
ii. Test the hypothesis that the test drug is no better than the sugar pill for curing flu. (11 marks)

b) Explain the properties of a binomial distribution. (8 marks)

Question 4

a) Discuss the use of statistical methods in clinical trials giving relevant examples. (9 marks)

b) In a study of the effect of diet on low-density lipoprotein cholesterol, Rassias et al used, as subjects 12 mildly hypercholesterolemic men and women. The plasma cholesterol levels (mmol/l) of the subjects were as follows: 6.0, 6.4, 7.0, 5.8, 6.0, 5.8, 5.9, 6.7, 6.1, 6.5, 6.3, 5.8. Let us assume that these 12 subjects behave as a simple random sample of subjects from a normally distributed population of similar subjects. We wish to estimate, from the data of this sample, the variance of the plasma cholesterol levels in the population with a 95 percent confidence interval. (16 marks)

Question 5

a) If 30% of the patients who are exposed to HIV/AIDS become infected, suppose we select 5 patients from this population. Estimate the probability that;

i. None will become infected
ii. One patient will become infected
iii. More than 4 patients will become infected
iv. At least three patients will be infected (16 marks)

c) Explain how you can apply t-tests in research. (9 marks)

Question 6

a) Nancy conducted a study to determine weight loss, body composition, body fat distribution, and resting metabolic rate in obese subjects before and after 12 weeks of treatment with a very-low-calorie diet (VLCD) and to compare hydro-densitometry with bioelectrical impedance analysis. The 17 subjects (9 women and 8 men) participating in the study were from an outpatient, hospital-based treatment program for obesity. The women's weights before and after the 12-week VLCD treatment are shown in the table below. Do these data provide sufficient evidence to allow the conclusion that the treatment is effective in causing weight reduction in obese women?

Weights (kg) of obese women before and after 12-week VLCD treatment									
Before	117.3	111.4	98.6	104.3	105.4	100.4	81.7	89.5	78.2
After	83.3	85.9	75.8	82.9	82.3	77.7	62.7	69.0	63.9

(15 marks)

b) Explain any two formula you can apply to determine the sample size in research. (10 marks)

Statistical formulas and tables may be used.

1. $z = \frac{(x-\mu)}{\sigma}$

2. $\chi^2 = \sum \frac{(o-e)^2}{e}$

3. $\phi^2 = \frac{\chi^2}{N}$

4. $b = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}$ and $a = \bar{Y} - b\bar{X}$ for the line $y = a + bx$

5. $r_s = 1 - \frac{6\sum_{1}^{n} d_i^2}{n(n^2-1)}$

6. $r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \bar{X}\bar{Y}}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2\right)\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \bar{Y}^2\right)}}$

[redacted]

Formulas

- $Z = \dfrac{\overline{X} - \mu_o}{\delta / \sqrt{n}}$ $\qquad Z = \dfrac{X - \mu}{\delta}$
- $s = \sqrt{\dfrac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}}$ $\qquad \overline{X} = \dfrac{\sum X}{n}$
- $Z = \dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\delta_1^2}{n_1} + \dfrac{\delta_2^2}{n_2}}};$
- $t = \dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 \left(\dfrac{1}{n_1} + \dfrac{1}{n_2} \right)}},$ $\qquad s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$
- $t = \dfrac{\overline{X} - \mu_o}{s / \sqrt{n}},$ $\qquad t = \dfrac{\overline{X} - \mu}{s_{\overline{x}}}, \quad s_{\overline{x}} = \dfrac{s}{\sqrt{n-1}}$
-
- $t = \dfrac{(\overline{X}_1 - \overline{X}_2)}{S(\overline{X}_1 - \overline{X}_2)}$; where, $S(\overline{X}_1 - \overline{X}_2) = \sqrt{\dfrac{S_p^2}{n_1} + \dfrac{S_p^2}{n_2}}$; with, $S_p = \sqrt{\dfrac{(n_1 - 1)s^2{}_1 + (n_2 - 1)s^2{}_2}{n_1 + n_2 - 2}}$
- $P(X = r) = \dfrac{n!}{(n-r)!r!} p^r q^{n-r}$ OR $P(X = r) = \dbinom{n}{r} p^r q^{n-r}$
- $P(x) = \dfrac{e^{-\lambda} \lambda^x}{x!}$
- $S_B^2 = SSB = \sum \dfrac{n(\overline{X} - \overline{X}_{GM})^2}{k-1}$ $\qquad S_W^2 = SSW = \dfrac{\sum (n_i - 1) S_i^2}{\sum (n_i - 1)}$
- $\overline{X}_{GM} = \dfrac{\sum X}{N}$ $\qquad F = \dfrac{SSB}{SSW} = \dfrac{S_B^2}{S_W^2}$
- $T = \sum \dfrac{T_i^2}{n_i} - \overline{X}_{GM}$ $\qquad T_{SS} = \sum y^2{}_{ij} - \overline{X}_{GM}$
- $C.I = \overline{X} \pm Z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$ (large sample) or $C.I = \overline{X} \pm t_{df, n-1} \dfrac{s}{\sqrt{n}}$ (small sample)
- $C.I = p_s \pm Z_{\alpha/2} \sqrt{\dfrac{p_s(1 - p_s)}{n}}$ or $(p_{s_1} - p_{s_2}) \pm Z \sqrt{\overline{p}(1 - \overline{p}) \left(\dfrac{1}{n_1} + \dfrac{1}{n_2} \right)}$
- $Expected \ Value = \dfrac{Row \ Sum \times Column \ Sum}{Grand \ Total};$ $\qquad \chi^2 = \sum_{i=1}^{n} \dfrac{(O_i - E_i)^2}{E_i}$