**AMREF INTERNATIONAL UNIVERSITY**

**SCHOOL OF PUBLIC HEALTH**

**DEPARTMENT OF COMMUNITY HEALTH**

**END OF SEMESTER EXAMINATIONS - AUGUST 2024**

**MPH 712**: **BIOSTATISTICS**

**DATE:** August 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/ and the Head of Department on +254725984499 email: michel.mutabazi@mcampus.amref.ac.ke
14. For adverse incidences please write an email to: amiu.examinations@amref.ac.ke

1. The height of 100 male patients in a hospital grouped in the table below:

| Height (in) | Frequencies |
|---|---|
| 60-62 | 5 |
| 63-65 | 18 |
| 66-68 | 42 |
| 69-71 | 27 |
| 72-74 | 8 |

Determine and interpret the following measures:

i) Median (3 marks)

ii) Standard deviation (3 marks)

iii) Skewness (2 marks)

2. Outline the steps followed to determine the confidence interval for the difference of two independent populations means. (6 marks)

3. For a random sample of 31 individuals selected from the population of 12- 40 years with foetal alcohol syndrome, the mean height is 147.4 cm. The population standard deviation of 6 cm. Test, the claim that the mean height of this group is 160.0 cm using 5% level of significance and show all the steps. (4 marks)

4. Consider the following contingency table:

| | Boys | Girls | Total |
|---|---|---|---|
| Smoking | 60 | 10 | 70 |
| Not smoking | 20 | 110 | 130 |
| Total | 80 | 120 | 200 |

i. Calculate $\chi^2$ using the above data. (4 marks)

ii. Given that the tables give $\chi^2$ =3.84 at 0.05 level of significance, test the hypothesis that the gender and smoking are independent. (3 marks)

**SECTION B: ANSWER ANY THREE (3) QUESTIONS**

5. a) We have a random sample of 25 fifth grade pupils who can do 15 pushups on the average, with a standard deviation of 9, after completing a special physical education program. Does this value of 15 differ significantly from the population value of 12? Use 5% level of significance.

(5 marks)

b) The following table shows the ages (X) and blood pressure (Y) of 8 persons.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| X | 52 | 63 | 45 | 36 | 72 | 65 | 47 | 25 |
| Y | 62 | 53 | 51 | 25 | 79 | 43 | 60 | 33 |

a) Find the Spearman's rank correlation coefficient and interpret the result (13 marks)

b) Find the regression equation $Y = a + bX$ (5 marks)

c) Find the expected blood pressure of a person who is 49 years old. (2 marks)

6. a) A study on retinol levels in well-nourished and undernourished children was conducted to examine if a difference exists. Realizing the importance of conducting such a study on a large number of subjects so that the impact of uncertainties is minimized, 100 well-nourished and 70 undernourished children were included in this study. For well-nourished, mean = 30.1 and Standard deviation = 2.9 and for undernourished, mean = 29.7 and Standard deviation = 3.1

Assuming the equality of variances from which these samples were taken, test the null hypothesis that the mean samples are equal (there is no difference between the means of the two samples) at 95% level of significance. (10 marks)

b) The table below shows the commission paid to salesmen by a certain private health facility in May 2024.

| Commission in Euros | Number of Salesmen |
|---|---|
| 10-14 | 12 |
| 15-19 | 14 |
| 20-24 | 20 |
| 25-29 | 53 |
| 30-34 | 77 |
| 35-39 | 96 |
| 40-44 | 54 |
| 45-49 | 37 |
| 50-54 | 19 |
| 55-59 | 18 |

Compute the following

| | | |
|---|---|---|
| i. | Mean | (2 marks) |
| ii. | Median | (2 marks) |
| iii. | Mode | (2 marks) |
| iv. | Interquartile range | (2 marks) |
| v. | Variance | (2 marks) |
| vi. | Standard deviation | (2 marks) |
| vii. | Skewness | (3 marks) |

7. 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. (10 marks)
b) Calculate the determination coefficient and interpret the finding. (5 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 (5 marks)
d) Estimate or predict the weight of a patient whose height is 180 cm. (5 marks)

8. 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 |

(16 Marks)

b) Explain three types of chi-square tests and how you can apply each one in research. (9 marks)

9. a) The probability that a patient will have adequate hospitalization insurance is 0.43. If 2000 patients visited the hospital, determine the number of patients who did not have adequate hospitalization insurance. (5 marks)

b) In a sample of 100 patients who are exposed to hepatitis B, 25 of them become infected. Suppose that we select 6 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
v. At most one patient will be infected (15 marks)

c) Explain the properties of the Normal Distribution. (5 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. $r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \bar{X}\bar{Y}}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2\right)\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \bar{Y}^2\right)}}$

$$\text{Median} = L_1 + \left( \frac{\frac{N}{2} - c}{f} \right) \times i$$

$$\text{Mode} = L_1 + \left( \frac{f_2}{f_0 + f_2} \right) \times i$$

$$Q_1 = L_{Q_1} + \left( \frac{\frac{N}{4} - (\Sigma f)_{Q_1}}{f_{Q_1}} \right) \times i$$

$$Q_2 = L_{Q_2} + \left( \frac{\frac{N}{2} - (\Sigma f)_{Q_2}}{f_{Q_2}} \right) \times i$$

$$Q_3 = L_{Q_3} + \left( \frac{\frac{3N}{4} - (\Sigma f)_{Q_3}}{f_{Q_3}} \right) \times i$$

$$\bar{x} = \frac{\sum f_i x_i}{\sum f} = \frac{\sum f_i x_i}{N}$$

$s^2 = [\sum f (Xi - \bar{x})^2] / (n - 1)$ variance for the sample

$\sigma^2 = [\sum f (Xi - \bar{x})^2] / n$ variance for the population

Skewness = 3(Mean - Median) / Standard Deviation

Coefficient of Variation = (standard Deviation / Mean) x 100%

[redacted]

**Formulas**

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