



# MERU UNIVERSITY OF SCIENCE AND TECHNOLOGY

P.O. Box 972-60200 – Meru-Kenya.

Tel: +254 (0)799529958, +254 (0)799529959, +254 (0)712524293

Website: [www.must.ac.ke](http://www.must.ac.ke) Email: [info@must.ac.ke](mailto:info@must.ac.ke)

---

## University Examinations 2023/2024

SECOND YEAR, FIRST SEMESTER EXAMINATION THE DEGREE OF BACHELOR OF SCIENCE IN COMMUNITY HEALTH, BACHELOR OF SCIENCE IN PUBLIC HEALTH

### HPP 3213/HPR 3213: MEDICAL BIOSTATISTICS

DATE: DECEMBER 2023

TIME: 2 HOURS

---

**INSTRUCTIONS:** Answer question one and any other two questions

---

#### QUESTION ONE (30 MARKS)

- Define the term *Biostatistics* and state 1 useful applications for Biostatistics in your area of specialization (2 marks)
- Distinguish between **univariate** and **multivariate** data as used in medical biostatistics giving suitable examples (2 marks)
- Discuss any three methods of collecting primary data in your area of practice (3 marks)
- A random sample of 100 hospitals in a county was selected and their average daily oxygen gas consumption (in cubic meters) in las month were recorded as follows: -

55	82	83	109	78	87	95	94	86	67
80	109	83	89	91	104	90	103	67	52
107	78	86	29	72	66	92	99	60	75
88	112	97	88	49	62	70	66	88	62
72	85	81	78	77	41	105	92	94	74
78	75	87	83	71	99	56	69	78	60
197	39	104	86	67	79	98	102	82	91
46	120	73	125	132	86	48	55	112	28
42	24	130	100	46	57	31	129	137	59
102	51	135	53	105	110	107	46	108	117

- i) Construct a stem and leaf diagram to represent these data and state the range of the data (3 marks)
  - ii) State the modal value and comment on the distribution of oxygen consumption (2 marks)
  - iii) Create a grouped frequency distribution table of these data (2 marks)
  - iv) Draw up a histogram of these data and superimpose a frequency polygon (4 marks)
  - v) Estimate the mean and variance of gas consumption in the county (4 marks)
  - vi) Compute the coefficient of variation for these data (1 mark)
- e) A group of 210 workers in certain company were categorized according to their exposure to smoking over the past 10 years. The workers were classified as belonging to one of 3 categories namely; A: Non-smoker, B: Passive smoker, C: Smoker. Based on a thorough medical examination of their respiratory system, the health of the workers was categorized as “poor” or “good”. The data on health workers was tabulated as follows:

	Poor Health	Good Health
Non-smoker	11	79
Passive smoker	19	51
Smoker	30	10

Use the chi –square test (with alpha=5%) to check if the data provides enough evidence to support the claim that exposure to cigarette smoke significantly affects the respiratory health workers (8 marks)

**QUESTION TWO (20 MARKS)**

- a) During a study, data on the following variables was collected for each patient in a hospital ward (see table). For each of the following variables, state whether it is qualitative; hence further classify into either nominal, ordinal, discrete or continuous (10 marks)

Variable	Type	Subtype
Exact age of patient (in years)		
Weight (in grams)		
Height (in meters)		
Systolic blood pressure		
Blood type		

- b) Alice suspects that most smokers in her home area are male. She interviews 12 randomly selected individuals at the shopping center and records these data on their gender (M = male, F = female) and smoking status

Respondent	1	2	3	4	5	6	7	8	9	10	11	12
Gender	M	M	F	M	F	F	F	M	F	M	M	F
Smoker	Yes	Yes	No	No	Yes	Yes	No	No	No	Yes	No	No

- i. What statistical hypothesis test should Alice conduct to validate her suspicion? (1 mark)
- ii. Organize these data into a duly filled cross-table (3 marks)
- iii. State the null and alternative hypotheses for this test (2 marks)
- iv. Perform the test proposed in part (i) at the 5% level of significance (4 marks)

### QUESTION THREE (20 MARKS)

- a) Describe what you understand by the term “sampling” as used in biostatistics. Explain why it is necessary (2 marks)
- b) Discuss the following methods of sampling as used in statistics. Where possible give a scenario when the said method would be appropriate for sample selection
  - i. Stratified random sampling (3 marks)
  - ii. Judgement sampling (3 marks)
  - iii. Multi-stage sampling (3 marks)
- c) The table below gives the frequency distribution of the Apgar scores for 100 low-birth-weight infants in a certain pediatric ward

Apgar Score	0	1	2	3	4	5	6	7	8	9
Frequency	6	1	3	4	5	10	11	23	24	13

- i) State the modal Apgar score (1 mark)
- ii) Find the mean Apgar score (2 marks)
- iii) What is the interquartile range of Apgar scores? (3 marks)
- iv) Find the standard deviation of the Apgar score (3 marks)

**QUESTION FOUR (20 MARKS)**

- a) Distinguish between the terms census and sample survey (2 marks)
- b) Give reasons why a researcher may prefer to conduct a sample survey over a census (4 marks)
- c) Victor suspects that male students perform better, on average, in Mathematics courses than female students. He collected the following marks of 11 students in his class after they received their results for a Mathematics Exam

Female	26	25	43	34	18	52
Male	23	30	18	25	28	

- i. State the null and alternative hypotheses that Victor should test (2 marks)
- ii. Propose a suitable statistical test that Victor could carry out. Explain your reasoning (2 marks)
- iii. Perform the test in part (ii) and comment on your results (5 marks)
- d) Outline the 5 step-procedure followed when conducting a statistical hypothesis test (5 marks)

**QUESTION FIVE (20 MARKS)**

- a) Distinguish between primary data and secondary data (2 marks)
- b) With the help of suitable examples, distinguish between the following terms as used in biostatistics:
  - i. Nominal and Ordinary data (2 marks)
  - ii. Scale and Ration variables (2 marks)
  - iii. Discrete and continuous variables (2 marks)
  - iv. Quantitative data and Qualitative data (2 marks)
- c) The number of patients seen by a clinician per day in 15 consecutive days was recorded as follows: 22 27 39 53 29 58 44 21 45 13 18 12 27 29 34
  - i. Arrange the data in ascending order and compute the rate (2 marks)
  - ii. Compute the lower quartile, middle quartile and upper quartile of these data (3 marks)
  - iii. Determine the Inter-Quartile Range (IQR) (1 mark)
  - iv. Represent the data in a box-and-whisker diagram and use it to comment on the shape of the distribution (4 marks)