



MACHAKOS UNIVERSITY

University Examinations 2021/2022 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

SECOND/THIRD YEAR SECOND SEMESTER EXAMINATION FOR

BACHELOR OF SCIENCE IN PUBLIC HEALTH

HPH 214: BIOSTATISTICS

DATE: 16/12/2021

TIME: 8.30-10.30 AM

INSTRUCTION:

Answer Question One and Any Other Two Questions

QUESTION ONE (30 MARKS)

- a) Define the term statistics and explain three major characteristics of statistics. (4 marks)
- b) Distinguish between the following terms as used in scientific research:
- i. Descriptive and Inferential statistics; (2 marks)
 - ii. Type 1 and Type II error; (2 marks)
 - iii. Quantitative and Qualitative variable; (2 marks)
 - iv. Open-ended and closed questions. (2 marks)
- c) Distinguish between a bar chart and Histogram as used in data presentation. (2 marks)
- d) The data given below represents marks obtained by 9 public health students in a test.
45, 32, 37, 46, 39, 36, 41, 48, 36
Calculate the geometric mean of the data. (4 marks)
- e) The mean systolic blood pressure was found to be 129.4 and 133.6 mm Hg with standard deviations of 10.6 and 15.2 mm Hg, respectively, for two groups of 12 and 15 men. What is the mean and pooled standard deviation systolic pressure of all the 27 men? (6 marks)
- f) Assume that in Machakos county the mean systolic blood pressure of persons aged 20 to 40 is 130 mm Hg with a standard deviation of 10 mm Hg. A random sample of 64 persons aged

20 to 40 from village x of the county has a mean systolic blood pressure of 132 mm Hg. Does the mean systolic blood pressure of the dwellers of the village (aged 20 to 40) differ from that of the inhabitants of the county (aged 20 to 40) in general, at a 5% Level of significance? (6 marks)

QUESTION TWO (20 MARKS)

- a) Define the term variable as used in statistics and give two examples. (3 marks)
- b) Explain the four scales of measurement for statistical data, giving an example in each case. (8 marks)
- c) Define the following terms as used in significance test:
 - i. Null hypothesis (1 mark)
 - ii. Alternative hypothesis (1 mark)
 - iii. P-value (1 mark)
- d) Coronary heart disease (CHD) begins in young adulthood and is the fifth leading cause of death among adults aged 20 to 24 years. Studies of serum cholesterol levels among college students, however, are very limited. A 2020 study looked at a large sample of students from Machakos University and reported that the mean serum cholesterol level among women is 168 mg/dl with a standard deviation of 27 mg/dl. A more recent study at Machakos University investigated the lipid levels in a cohort of sedentary university students. The mean total cholesterol level among $n = 71$ females was $\bar{X} = 173.7$. Is there evidence that the mean cholesterol level among sedentary students differs from this average over all students? Test the significance at $\alpha = 0.05$. (6 marks)

QUESTION THREE (20 MARKS)

- a) Differentiate between the following terms as used in sampling:
 - i. Sampling frame and sampling interval. (2 marks)
 - ii. Convenience sampling and quota sampling. (2 marks)
- b) A Population has three strata with sizes, means and variances as given below:

Stratum	Size (N_n)	$\bar{X}_n$	S_n^2
1	15000	12.5	9.00
2	23000	7.00	4.50
3	32000	5.00	2.25

If the sample of size $n = 10000$ is to be selected from the population, determine;

- i. The proportional allocation. (2 marks)
 - ii. The variance of estimate of the population mean under proportional allocation and simple random sampling without replacement (srswor) and conclude which scheme is more efficient. (6 marks)
- c) Consider a population with 6 units from which a sample of size 3 is to be drawn (selected). Let the values for the population units be 3, 4, 5, 6, 7, and 8. Find;
- i. The population mean and variance. (2 marks)
 - ii. The sampling interval (1 mark)
 - iii. Using a systematic sample consisting of 3, 5, 7 drawn from the population calculate the variance (mean square among units). (2 marks)
 - iv. The variance of estimate of the population mean under systematic sampling. (3 marks)

QUESTION FOUR (20 MARKS)

- a) The table below represents Blood pressure levels of 60 first-year male medical students (in mm Hg);

Class limit	90-99	100-109	110-119	120-129	130-139	140-149	150-159
Number of students	2	6	17	16	12	6	1

Using the direct method, determine;

- i. Mean (2 marks)
 - ii. Median (3 marks)
 - iii. Mode (3 marks)
 - iv. Interquartile Range. (2 marks)
 - v. Standard deviation. (3 marks)
- b) A sample of the various prices for a particular drug has been conducted in 16 stores which were selected at random in a neighborhood of a city. The following prices were noted: 95, 108, 97, 112, 99, 106, 105, 100, 99, 98, 104, 110, 107, 111, 103, 110. Assuming that the prices of this product follow a normal law of variance of 25 and an unknown mean:
- i. What is the distribution of the sample mean? (2 marks)
 - ii. Determine the confidence interval at 95% for the population mean. (5 marks)

QUESTION FIVE (20 MARKS)

- a) Differentiate between the dependent variable and the independent variable as defined on a linear regression equation and give an example. (2 marks)
- b) The table given below represents the Resting Metabolic Rate (RMR) which is related to body weight for a sample of 10 individuals who are residents of Machakos county.

Body Weight (kg)	57.6	64.9	59.2	60.0	72.8	77.1	82.0	86.2	91.6	99.8
RMR (kcal/24hrs)	1325	1365	1342	1316	1382	1439	1536	1466	1519	1639

- Compute the Pearson's correlation co-efficient between body weight and Resting Metabolic Rate (RMR). (6 marks)
- Interpret the correlation coefficient obtained in (i). (2 marks)
- Fit a regression line to the data. (5 marks)
- What is the expected RMR if a person has body weight of 65 kg and the standard error of prediction? (5 marks)