



# MACHAKOS UNIVERSITY

University Examinations 2016/2017

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

FIRST YEAR SECOND SEMESTER EXAMINATION FOR DEGREE IN  
BACHELOR OF SCIENCE POPULATION HEALTH

PPH 102: BIostatISTICS

DATE: 6/6/2017

TIME: 8:30 – 10:30 AM

---

**INSTRUCTION:**

*Answer Question ONE which is compulsory and any other TWO Questions*

**QUESTION ONE (30 MARKS)**

- a) Discuss three advantages of sampling over census (6 marks)
- b) Differentiate the following terms as they apply in scientific research (6 marks)
- i) Response and independent variables
  - ii) Descriptive and inferential statistics
  - iii) Null hypothesis and Alternative hypothesis
- c) Two different types of drugs A and B were tried on certain patients for increasing weight, 5 persons were given drug A and 7 persons were given drug B. The increase in weight (in Kgs) is given below:

Drug A	8	12	13	9	3	-	-
Drug B	10	8	12	15	6	8	11

Determine whether the two drugs differ significantly with regard to their effect in increasing weight at  $\alpha = 5\%$  (7 marks)

- d) Highlight the meaning of the following terms as used in epidemiology
- i) Predisposing factors
  - ii) Non-Enabling/disabling factors
  - iii) Precipitation factors
  - iv) Reinforcing factors
- (4 marks)
- e) Based on the recent observations 5% of the Machakos University female students have low hemoglobin (Hgb). The University clinic laboratory can detect 80% of the low Hgb when it is present. It also gives 12% false positive results. A female student in Machakos University volunteers for Hgb test in the University clinic laboratory, determine the following probabilities;
- (i) That the test result will be positive (Low Hgb)
  - (ii) That, given a positive result, she has a low Hgb;
  - (iii) That, given a negative result, she has a normal Hgb;
  - (iv) That she was misclassified
- (7 marks)

**QUESTION TWO (20 MARKS)**

- a) The data below is a summary of copy typing speed per minute before and after a training.

Participant	A	B	C	D	E	F	G	H
Rate Before	85	124	172	123	111	139	99	77
Rate after	105	160	175	143	156	127	95	100

Test the hypothesis that on average the training did not result to any improvement on the typing speed. (8 marks)

- b) Below are the Friesian daily milk production during a particular week in the year 2014.

<i>Sales</i> <i>(KSH.000)</i>	11-20	21-30	31-40	41-50	51-60
Frequency	3	6	11	3	2

Determine;

- i) Average production per day
  - ii) Median
  - iii) Mode and
  - iv) Standard deviation.
- (12 marks)

**QUESTION THREE (20 MARKS)**

- a) Highlight two properties of a good measure of data variation (2 marks)
- b) A mixed choir of 5 boys and 7 girls is to be chosen from 9 boys and 11 girls. Determine the number of way in which the choir can be chosen. (4 marks)
- c) The lecturer presumed that the average class attendance was 65 students. He did call register for 10 days and observed following: 66, 65, 69, 70, 69, 71, 70, 63, 64 and 68. The t-test output was as given

The sample t-test output, for the test value=65						
	t	df	Sig. (2- tailed)	Mean difference	95% CI of the difference	
					Lower	Upper
Attendance	2.825	9	0.020	2.500	0.4979	4.5021

- i) State the hypothesis for the above scenario
  - ii) Based on the t-value and the sig.(2-tailed) value make statistical conclusion
  - iii) Interpret the 95% CI lower and upper difference values (6 marks)
- d) Discuss any four types of test-statistics and highlight when appropriate to use each (8 marks)

**QUESTION FOUR (20 MARKS)**

A cohort study of smoking and bladder cancer was conducted in a small island population. There were a total of 1000 people on the island where 400 were smokers while 600 were not. Fifty of the smokers and ten of the non-smokers developed bladder cancer. Using this information;

- a) Construct a two by two table and label the cells and margins.
- b) Determine
  - i) Absolute Risk (Actual)
  - ii) Relative Risk (risk ratio)
  - iii) Attributable Risk
  - iv) Odds Ratio (8 marks)
- c) The function  $\hat{y} = 3.5 + 1.23x_1 + 0.78x_2$  is model of predicting the expected yield (kg '000') per hectare of a certain crop as influenced by fertilizer  $k63(x_1)$  and the humidity level ( $x_2$ ). Interpret the values 3.5, 1.23 and 0.78 (6 marks)

- d) Differentiate the following terms as apply in scientific social research.
- i. Type I and Type II error
  - ii. Sample and population
  - iii. Two tailed and one tailed test
- (6 marks)

**QUESTION FIVE (20 MARKS)**

- a) Highlight three principles of experimental research (3 marks)
- b) By citing examples distinguish between discrete and numeric variables (6 marks)
- c) A manufacturer wished to compare the performance of different machine brands, their performance difference is summarized in ANOVA table below.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean sum of Squares	F – value
Between machines	$4 - 1 = 3$	540.69	180.23	$F = \frac{180.23}{7.15} = 25.207$
Error	$16 - 4 = 12$	85.75	7.15	
Total	$16 - 1 = 15$	626.44		

- i) Based on the table above state the number of machine brands compared (1mark)
- ii) Explain the two sources of variation in column one of the table (2 marks)
- iii) When is it appropriate to use ANOVA and F-test for data analysis (4 marks)
- iv) Was there a significant difference in their performance at  $\alpha=5\%$  (4 marks)