(*Knowledge for Development*)

KIBABII UNIVERSITY

UNIVERSITY EXAMINATIONS
2020/2021 ACADEMIC YEAR

THIRD YEAR SECOND SEMESTER
MAIN EXAMINATIONS

FOR THE DEGREE OF BACHELOR OF EDUCATION SCIENCE

COURSE CODE: **SBL321**

COURSE TITLE: **BIOSTATISTICS**

DATE: MONDAY 4th, October 2021. **TIME:** 2:00 -4:00 p.m.

INSTRUCTIONS TO CANDIDATES

Answer Question one (1) and any other two (2) Questions. Question one is compulsory and carries 30 marks, the other Questions carry 20 marks each.

TIME: 2 Hours

This paper consists of 4 printed pages. Please Turn Over

QUESTION ONE

a) The following data shows the annual income (in Ksh 1,000s)taken from nine randomly chosen researchers from the National Research Institute.

37 102 34 12 111 56 72 17 33

i. Calculate the sample mean income. (3 Marks)
ii. Calculate the sample standard deviation of these incomes. (4 Marks)
iii. What population could this sample represent? (1 Marks)

b) What istype-1 error and what is its role in hypothesis testing. (3 Marks)

c) Discuss the role of the α-level in hypothesis testing? (4 Marks)

d) What is the role of the p-value in hypothesis testing? (4 Marks)

e) The regression line relating estimated mean hourlywages in Ksh in 2010 in Kenya to years of education is provided as y=-0.75+0.75x.
   a. What is the estimated mean hourly wage (in 2010) for persons with 12 years of education? (2 Marks)
   b. What is the estimated difference in hourly wages (in 2010) forpersons with 16 years of education versus 12 years of education? (3 Marks)

f) Differentiate with examples, between Ratio scale data and Interval scale data. (6 Marks)

QUESTION TWO

a) In a high school in Kenya, dietary counseling is being tested to measure the program's long-term impact on student's fat intake. Of the 300 students at the school, 150 are randomized to receive five one-hour sessions of dietary counseling; the other 150 students receive no counseling. Six months after the last counseling sessions, all students are asked to keep a food diary for one week. Each student's average fat intake in grams is calculated at the end of this week. The results of this exercise are as follows:

Intervention group $\bar{X}_1 = 54.8$ grams, $s1 = 28.1$ grams, $n1 = 146$
Control group $\bar{X}_2 = 62.8$ grams, $s2 = 34.7$ grams, $n2 = 142$

Construct a 95% CI for the population mean difference in daily fat intake for the intervention group as compared to the control group. (10 Marks)

b) Compute a p-value for testing the null of no association between counseling and average fat intake. Is this consistent with the confidence interval estimated in the section a) problems? (10 Marks)

QUESTION THREE

A random sample of 960 high school students in Coast region was collected to determine the association between post-traumatic stress induced by recurrent terrorist threats and drug abuse.Two of the findings from this study are that 35% of the population knew at least one person who had been killed in a terrorist attack and that 10% of the sample had used marijuana in the 30 days prior to the study.

Estimate a 95% confidence interval for the proportion of all the students who:

a) Knew at least one person killed in a terrorist attack (10 Marks)
b) Had used marijuana in the prior 30 day period (10 Marks)

QUESTION FOUR

a) Eight countries were selected from Africa each of these countries was matched with a country from Europe, based on the Infant Mortality Rate (IMR) in 2020.

Pair	**IMR: State A**	**IMR: State B**	**Difference (A - B)**
1	80	76	4
2	130	112	18
3	88	97	-9
4	98	67	31
5	103	107	-4
6	121	116	5
7	83	94	-11
8	93	78	15
Mean	99.5	93.4	6.1
SD	17.9	18.1	14.5

i. Estimate a 95% confidence for the true difference in mean IMR between the two continental groupings. (6 Marks)
ii. State your null and alternative hypotheses for the corresponding hypothesis test. (3 Marks)
iii. Report a p-value for the hypothesis test. (3 Marks)

b) A healthcare information company is interested in estimating the average charge for a standard patient visit to Kakamega Referral Hospital, after applying the discount negotiated with theUniversal Health Care Plan. Data is collected from 16 randomly selected different medical conditions costs in the hospital. The following are some summarystatistics:- Mean charge: Ksh 25.50; - SD of charges: Ksh 2.10

i. Assuming the charge data is normally distributed for all the medical conditions treated in the hospital, estimate a range of amounts that most (95%) of the conditions are charged for a standard patient visit (3 Marks)

ii. Assuming the charge data is normally distributed for all conditions, estimate a 95% confidence interval for the mean amount charged by the hospital (5 Marks)

QUESTION FIVE

a. Define selection bias in a study. Can you suggest an example (hypothetical or otherwise) of possible selection bias? (3 Marks)

b. Give a concise explanation of the utility of having a randomized control group in a scientific study. (2 Marks)

c. Referring to each of the "almost random" assignment methods (alphabetical, sequential, telephone/social security number), can you suggest how each method could yield a biased (non-random) assignment? (6 Marks)

d. Compare block randomization to simple randomization (6 Marks)

e. Write a brief description on Latin Square Design (3 Marks)