

UGANDA MARTYRS UNIVERSITY
NKOZI, RUBAGA AND MASAKA CAMPUS
UNIVERSITY EXAMINATIONS
FACULTY OF SCIENCE
DEPARTMENT OF MATHEMATICS AND STATISTICS
END OF SEMESTER TWO FINAL ASSESSMENT YEAR ONE
COMPUTATIONAL STATISTICS

INSTRUCTIONS

DATE OF EXAMINATION: 22/07/2022

TIME ALLOCATED: 3 HOURS

1. Carefully read through ALL the questions before attempting
2. **SECTION A IS COMPULSORY ANSWER ANY 3 QUESTIONS IN SECTION B**
3. Ensure that your **Registration number** is indicated on all pages of your answer sheets.
4. All questions require clear workings and all numbers carry equal marks
5. SECTION A requires you to select what are the best options, SECTION B requires clear workings and all numbers carry equal marks

SECTION ONE

[25 marks]

QUESTION ONE

Which of the following points do not reflect statistics?

- ☐ They can be inferential
- ☐ It describes ways of analyzing qualitative data
- ☐ They can be descriptive
- ☐ It describes methods for analyzing quantitative data
- ☐ It is a while subject of study
- ☐ It describes methods of collecting, quantitative data

QUESTION TWO

What is the first stage in Statistics?

- ☐ Collect data
- ☐ Organize data
- ☐ Identify the group of people to be studied
- ☐ Summarize data
- ☐ Review the materials
- ☐ Analyze data

This scenario applies to Questions 3 and 4: A study was done to compare the lung capacity of coal miners to the lung capacity of farm workers. The researcher studied 200 workers of each type. Other factors that might affect lung capacity are smoking habits and exercise habits. The smoking habits of the two worker types are similar, but the coal miners generally exercise less than the farm workers.

QUESTION THREE

Which of the following is the explanatory variable in this study?

- ☐ Exercise
- ☐ Lung Capacity
- ☐ Smoking or not
- ☐ Occupation

QUESTION FOUR

Which of the following confounding variable in this study?

- ☐ Exercise
- ☐ Lung Capacity
- ☐ Smoking or not
- ☐ Occupation

QUESTION FIVE

Significance testing is based on which of the following principles?

- ☐ Verification
- ☐ Constructivism
- ☐ Evaluation
- ☐ Falsification
- ☐ Construction
- ☐ Idealism

QUESTION SIX

A scatter plot of number of teachers and number of people with college degrees for cities in California reveals a positive association. The most likely explanation for this positive association is:

- ☐ Teachers encourage people to get college degrees, so an increase in the number of teachers is causing an increase in the number of people with college degrees.

- ☐ Larger cities tend to have both more teachers and more people with college degrees, so the association is explained by a third variable, the size of the city.
- ☐ Teaching is a common profession for people with college degrees, so an increase in the number of people with college degrees causes an increase in the number of teachers.
- ☐ Cities with higher incomes tend to have more teachers and more people going to college, so income is a confounding variable, making causation between number of teachers and number of people with college degrees difficult to prove.

QUESTION SEVEN

The value of a correlation is reported by a researcher to be $r = -0.5$. Which of the following statements is correct?

- ☐ The x-variable explains 25% of the variability in the y-variable.
- ☐ The x-variable explains -25% of the variability in the y-variable.
- ☐ The x-variable explains 50% of the variability in the y-variable.
- ☐ The x-variable explains -50% of the variability in the y-variable.

QUESTION EIGHT

If there is a probability of 5 percent in how many cases would a result arise solely due to chance?

- ☐ 1/50
- ☐ 10/100
- ☐ None of these
- ☐ 5/100
- ☐ 50/50
- ☐ 0.0005

QUESTION NINE

What is the conventional level of significance typically adopted in psychology?

- ☐ $p > 0.01$
- ☐ $p < 0.05$
- ☐ $p > 0.001$
- ☐ $p > 0.005$
- ☐ $p > 0.5$
- ☐ $p < 0.5$

This scenario applies to Questions 10 to 12: A randomized experiment was done by randomly assigning each participant either to walk for half an hour three times a week or to sit quietly reading a book for half an hour three times a week. At the end of a year the change in participants' blood pressure over the year was measured, and the change was compared for the two groups.

QUESTION TEN

This is a randomized experiment rather than an observational study because:

- ☐ Blood pressure was measured at the beginning and end of the study.
- ☐ The two groups were compared at the end of the study.
- ☐ The participants were randomly assigned to either walk or read, rather than choosing their own activity.
- ☐ A random sample of participants was used.

QUESTION ELEVEN

The two treatments in this study were:

- ☐ Walking for half an hour three times a week and reading a book for half an hour three times a

week.

- ☐ Having blood pressure measured at the beginning of the study and having blood pressure measured at the end of the study.
- ☐ Walking or reading a book for half an hour three times a week and having blood pressure measured.
- ☐ Walking or reading a book for half an hour three times a week and doing nothing.

QUESTION TWELVE

If a statistically significant difference in blood pressure change at the end of a year for the two activities was found, then:

- ☐ It cannot be concluded that the difference in activity caused a difference in the change in blood pressure because in the course of a year there are lots of possible confounding variables.
- ☐ Whether or not the difference was caused by the difference in activity depends on what else the participants did during the year.
- ☐ It cannot be concluded that the difference in activity caused a difference in the change in blood pressure because it might be the opposite, that people with high blood pressure were more likely to read a book than to walk.
- ☐ It can be concluded that the difference in activity caused a difference in the change in blood pressure because of the way the study was done.

QUESTION THIRTEEN

A national random sample of 20 ACT scores from 2020 is listed below. Calculate the sample mean and standard deviation

29, 26, 13, 23, 23, 25, 17, 22, 17, 19, 12, 26, 30, 30, 18, 14, 12, 26, 17, 18

- ☐ 20.50, 5.79
- ☐ 20.50, 5.94
- ☐ 20.85, 5.79
- ☐ 20.85, 5.94

QUESTION FOURTEEN

Which of the following represents a valid probability table?

(i). Outcomes	1	2	3	4	5
Probability	1/5	1/5	1/5	1/5	1/5

(ii). Outcomes	1	2	3	4	5
Probability	1/2	1/5	1/10	1/10	1/10

Circle the best choice: A. (i) B. (ii) C. (i) and (ii) D. Not enough information

QUESTION FIFTEEN

In 3 tosses of a coin which of following equals the event "exactly two heads"?

A = THH, HTH, HHT, HHH

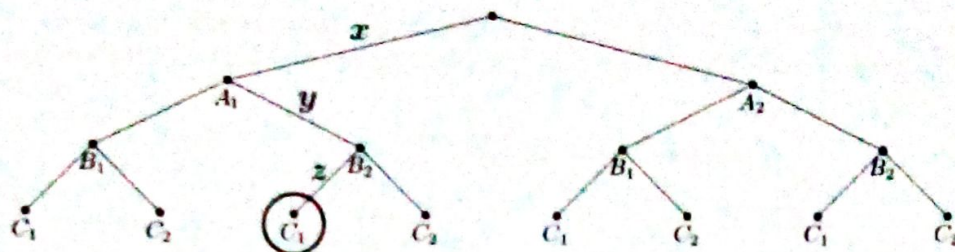
B = THH, HTH, HHT

C = HTH, THH

Circle the best answer: A B C B and C

QUESTION SIXTEEN

These questions all refer to the following figure. For each one circle the best answer.



- (i) The probability x represents
- (ii). The probability y represents
- (iii). The probability z represents
- (iv). The circled node represents the event

QUESTION SEVENTEEN

True or false: Setting the prior probability of a hypothesis to 0 means that no amount of data will make the posterior probability of that hypothesis the maximum over all hypotheses.

Circle one:

True

False

QUESTION EIGHTEEN

True or false: It is okay to have a prior that depends on more than one unknown parameter.

Circle one:

True

False

Questions 19 and 20: A newspaper article reported that "Children who routinely compete in vigorous after-school sports on smoggy days are three times more likely to get asthma than their non-athletic peers."

QUESTION NINETEEN

Of the following, which is the most important additional information that would be useful before making a decision about participation in school sports?

- ☐ Where was the study conducted?
- ☐ How many students in the study participated in after-school sports?
- ☐ What is the baseline risk for getting asthma?
- ☐ Who funded the study?

QUESTION TWENTY

The newspaper also reported that "The number of children in the study who contracted asthma was relatively small, 265 of 3,535 ". Which of the following is represented by $265/3535 = .075$?

- ☐ The overall risk of getting asthma for the children in this study.
- ☐ The baseline risk of getting asthma for the "non-athletic peers" in the study.
- ☐ The risk of getting asthma for children in the study who participated in sports.
- ☐ The relative risk of getting asthma for children who routinely participate in vigorous after-school sports on smoggy days and their non-athletic peers.

QUESTION TWENTY ONE

Questions 21 to 24: A survey asked people how often they exceed speed limits. The data are then categorized into the following contingency table of counts showing the relationship between age group and response.

Age	Exceed Limit if Possible?		Total
	Always	Not Always	
Under 30	100	100	200
Over 30	40	160	200
Total	140	260	400

QUESTION TWENTY TWO

Among people with age over 30, what is the "risk" of always exceeding the speed limit?

- ☐ 0.20
- ☐ 0.40
- ☐ 0.33
- ☐ 0.50

QUESTION TWENTY THREE

Among people with age under 30 what are the odds that they always exceed the speed limit?

- ☐ 1 to 2
- ☐ 2 to 1
- ☐ 1 to 1
- ☐ 50%

QUESTION TWENTY FOUR

What is the relative risk of always exceeding the speed limit for people under 30 compared to people over 30%.

- ☐ 2.5
- ☐ 0.4
- ☐ 0.5
- ☐ 30%

QUESTION TWENTY FIVE

What is the effect of an outlier on the value of a correlation coefficient?

- ☐ An outlier will always decrease a correlation coefficient.
- ☐ An outlier will always increase a correlation coefficient.
- ☐ An outlier might either decrease or increase a correlation coefficient, depending on where it is in relation to the other points.
- ☐ An outlier will have no effect on a correlation coefficient.

SECTION TWO

QUESTION ONE

Suppose you are appointed a sales manager of a particular brand which has been sold at a substantial reduction in price. You have analyzed the sales figures for the past two years and found that the percentage of sales has increased steadily. You have also gone further by tabulating the sales data by areas of the city and found that the increase in sales has been greatest in areas with higher incomes. Questions have often been raised regarding the consumer acceptance of the brand, for instance whether the consumers can tell the difference between this brand and other related brands. Management has therefore decided that information be collected about consumer acceptance of this brand in the hope that questions such as those given above can be answered.

Required

- (i). State with reasons the most appropriate method that can be used to collect the above data.
- (ii). State three other alternatives that could be used in collecting the data clearly given their merits and demerits.
- (iii). Give three methods that could be used to summarize the data collected above. (b). New Vision employs 1000 employees in three departments; production, administration and sales. 600 are in production and 300 in administration. There are 110 male graduates, 110 female graduates and 290 undergraduate females. The remaining employees are undergraduate males. In the production department there are 350 undergraduate males, 150 undergraduate females and 50 male graduates. In administration department, there are 100 undergraduate males, 110 undergraduate females and 50 graduate males. Summarize the information in a table form showing all the necessary details. Provide suitable secondary statistics to describe the distribution of the employees in the departments.

QUESTION TWO

[25 marks]

- (a). Define Probability
- (b). Using Relevant examples, how is probability applied in real life computational statistics
- (c). A group of five people simultaneously enter an elevator at the ground floor. There are 10 upper floors. The persons choose their exit floors independently of each other.
- (i). Specify an appropriate sample space and determine;
- (ii). The probability that they are all going to different floors when each person randomly chooses one of the 10 floors as the exit floor.
- (iii). How does the answer change when each person chooses with probability 1 the 10th floor as the 2 exit floor and the other floors remain equally likely as the exit floor with a probability of 1 each 18.
- (d). A parking lot has 10 parking spaces arranged in a row. There are 7 cars parked. Assume that each car owner has picked at a random a parking place among the spaces available. Specify an appropriate sample space and determine the probability that the three empty places are adjacent to each other.
- (e). A company manufactures and sells a single product. Estimated sales, costs and selling prices for the coming year are as follows:-

Sales Units	Probability	Selling Price	Probability	Variable Cost/ Unit	Probability
30,000	0.3	400	0.2	200	0.1
45,000	0.2	500	0.4	250	0.3
60,000	0.1	600	0.2	300	0.3
55,000	0.2	700	0.1	350	0.2
65,000	a	800	b	400	c

Given that the fixed costs per year are 7,000,000/ =, 8,000,000/ = and 10,000,000/ = with associated probabilities of 0.4, 0.3 and 0.3 respectively. Determine the:-

- (i). Values of a, b and c
- (ii). Total revenue of the firm
- (iii). Total costs of the firm
- (iv). Expected annual profit of the firm and advice
- (v). Worst possible scenario in the upcoming year

QUESTION THREE

(a). Define the following Terminologies;

[25 marks]

- (i). An Experiment
- (ii). Chance Experiment
- (iii). Sample Space
- (iv). Event
- (v). Outcome

(b). Assume that a hat contains 4 bills: a \$1 bill, a \$5 bill, a \$10 bill and a \$20 bill. Two bills are to be selected at random with replacement. Construct a sample space, and find the probability that:

- (i) Both bills are \$1 bills if given the first selected is a \$1 bill
- (ii) Both bills have a value greater than a \$5 bill given the second bill is a \$10 bill.
- (iii) The second bill is a five given the first bill is a \$5.
- (iv) The first bill is a five given the second bill is a \$5.

(c). A quality control inspector, is checking a sample of light bulbs for defects. The following table summarizes her findings.

Wattage	Good	Defective	Total
20	80	15	95
50	100	5	105
100	120	10	130
Total	300	30	330

If one of these light bulbs is selected at random, find the probability that the light bulb is;

- (i). Good given it is 100 watts
- (ii). Defective given it is 50 watts
- (iii). 100 watts given it is good
- (iv). 50 watts given it is defective

(d). In a factory there are three machines that make light bulbs. The machines manufacture 20%, 30% and 50% of the total production. From their production, 5% , 4 % and 2% respectively are faulty. I choose a collection of light bulbs at random from the factory's output.

- (i). If the collection contains two faulty light bulbs, what is the probability that they come from the same machine?

- (ii). If the collection contains three faulty light bulbs, what is the probability that they come from three different machines?

QUESTION FOUR

- (a). Difference between Correlation and Regression

[25 marks]

- (b). State the Key Differences between Correlation and Regression

- (c). State the Key Similarities among Correlation and Regression

(d). Carina obtains cash from an ATM (cash machine). She suspects that the rate at which she spends cash is affected by the amount of cash she withdrew at her previous visit to an ATM. To investigate this she deliberately varies the amounts she withdraws. She records, for each visit to an ATM, the amount, £ x , withdrawn and the number of hours, y , until her next visit to an ATM.

Withdrawal	1	2	3	4	5	6	7	8	9	10
x	40	10	100	110	120	150	20	90	80	130
y	56	62	195	330	94	270	48	196	214	286

- (i). Draw a scatter diagram of the data

- (ii). Calculate the equation of the regression line of y on x and draw it on your scatter diagram

(e) (i). Carina made one withdrawal immediately before giving on weekend visit to Edinburgh. Identify the most likely withdrawal, giving a reason

(ii). Following another withdrawal, Carina was confined to bed for several days with a heavy cold. Identify the most likely withdrawal, giving a reason.

- (f) (i). Interpret, in context, the gradient of the regression line.

(ii). Comment on the evidence, if any, that Carina's rate of spending cash is affected by the amount she withdraws