



Probs exam 1 - probability and statistics 1 quizzes

Bachelor of Science in Computer Science (Dedan Kimathi University of Technology)



Scan to open on Studocu

a). Describe what is meant by each of the following terms.

i). A random sample →

[1 Marks]

ii). A statistic.

[2 Marks]

b). From national statistics, it is known that 7% of all drivers in a country are young drivers. It is also known that 18% of all drivers involved in road accidents are young drivers (less than 25 years old). Define the two events:

A : a randomly chosen driver is involved in a road accident. ✓

Y : a randomly chosen driver is a young driver. ✓ $7/100 \times 100 \times 100$

i). Determine the conditional probability $P[A | Y]$ as a function of $P[A]$. [2 Marks]

ii). Comment on the result from part (i). [1 Marks]

c). The following data are available on three television factories that produce all the televisions used in a country.

Factory	% of total production	Probability of defect (Def)
A	0.35	0.020
B	0.40	0.015
C	0.25	0.010

A television is selected at random and found to have a defect (Def).

- i). Derive the expression to correctly calculate the probability that the selected television was made in factory B. [2 Marks]

- ii). Calculate, by using your answer to part (i), the probability that the selected television was produced by Manufacturer B. [3 Marks]

- d). Suppose 30% of the women in a class received an A on the test and 25% of the men received an A. The class is 60% women. Given that a person chosen at random received an A, what is the probability this person is a woman? [3 Marks]

- e). Three babies are given a weekly health check at a clinic, and then returned randomly to their mothers. What is the probability that at least one baby goes to the right mother? [4 Marks]

- f). A Banking Analyst is assessing the performance of a newly developed credit risk model against experts' knowledge. The credit scores produced on a sample of twelve customers by the experts (x) and the model (y) are the following:

x	65.8	63.7	67.6	64.4	68.2	62.9	70.5	66.4	68.0	67.1	69.5	71.8
y	68.2	66.2	68.1	66.0	69.1	66.1	68.7	65.9	69.3	67.2	67.9	70.4

- i). Fit a linear regression line of y on x . [3 Marks]
- ii). Calculate Pearson's correlation coefficient between the experts' and the model's scores. [2 Marks]

- g). Consider the following distribution table.

x	2 - 4	4 - 6	6 - 8	8 - 10	10 - 12	12 - 14	14 - 16	16 - 18	18 - 20
f	18	24	47	80	102	66	40	21	15

Using the above data, calculate the variance using the moments. [4 Marks]

QUESTION TWO (20 Marks) (Optional)

- a). Consider the following data, and the corresponding sums derived from the data:

$$x_i: 10.0 \ 6.9 \ 11.4 \ 12.6 \ 10.3 \ 12.4 \ 9.8$$

$$\sum x_i = 73.4; \sum x_i^2 = 792.22; \sum x_i^3 = 8,750.972.$$

- i). Determine the third moment about the mean for these data. [2 Marks]
- ii). a). Write down the mathematical definition of the coefficient of skewness of a set of data. [1 Marks]
- b). Determine the coefficient of skewness and interpret it. [3 Marks]

b). Consider the following sample with 20 observations x_i :

1 1 5 7 9 11 11 14 14 19
20 21 23 28 28 31 39 41 43 47

$$\sum_{i=1}^{20} x_i = 413 \text{ and } \sum_{i=1}^{20} x_i^2 = 12,311$$

Calculate the:

- i). the mean of this sample. [1 Marks]
- ii). the standard deviation of this sample. [2 Marks]
- iii). the median of this sample. [1 Marks]
- iv). the interquartile range of this sample. [2 Marks]

c). Consider an insurer that offers two types of policy: home insurance and car insurance. 70% of all customers have a home insurance policy, and 80% of all customers have a car insurance policy. Every customer has at least one of the two types of policies.

Calculate the probability that a randomly selected customer:

- i). does not have a car insurance policy. [1 Mark]
- ii). has car insurance and home insurance. [2 Marks]
- iii). has home insurance, given that he has car insurance. [2 Marks]
- iv). does not have car insurance, given that he has home insurance. [3 Marks]

QUESTION THREE (20 Marks) (Optional)

a). Two groups of students sat the same exam. The marks in the first group of 64 students had an average of 52 and a standard deviation of 9. The marks in the second group of 42 students had an average of 45 and a standard deviation of 8.

Calculate the average and standard deviation of the combined data set.

[4 Marks]

b). The following data refers to examination marks verses hours of study per week of a sample of 8 candidates that sat for statistics examinations in 2019:

Exam mark (Y)	61	61	84	70	72	92	72	72
Hours of study(X)	23	16	34	23	27	34	18	23

Calculate the spearman's rank correlation coefficient coefficient of correlation. comment on the relationship between the two variables. [5 Marks]

c). Three persons A, B and C are being considered for the appointment as the chairman for a company whose chance of being selected for the post are in the proportion 4:2:3 respectively. The probability that A, if selected will introduce democratization in the company structure is 0.3 the corresponding probabilities for B and C doing the same are respectively 0.5 and 0.8. What is the probability that democratization would be introduced in the company? [5 Marks]

- d). The table below gives the probability distribution function of a random variable X .

$X = x$	0	1	2	3
$\Pr(X = x)$	p	$2q$	$p + q$	q

Given that the mean of X is 1.375, find the values of p and q .

[6 Marks]

QUESTION FOUR (20 Marks) (Optional)

- a). An examination consists of multiple-choice questions, each having five possible answers. Suppose you are a student taking the exam, and that you reckon you have probability 0.75 of knowing the answer to any question that may be asked and that, if you do not know, you intend to guess an answer with probability $1/5$ of being correct. What is the probability you will give the correct answer to a question? [3 Marks]
- b). The probability distribution of a discrete random variable, X , is given by the table below.

x	0	1	2	3	4
$\Pr(X = x)$	0.2	$0.6p^2$	0.1	$1 - p$	0.1

- i). Determine the value of p . [2 Marks]
- ii). Let $p = \frac{2}{3}$. Calculate $E(X)$. [2 Marks]
- iii). Find $\Pr(X \geq E(X))$. [2 Marks]
- c). Calculate the coefficient of rank correlation of the scores obtained by 7 students in an essay writing competition by two judges, X and Y .

X	15	12	20	16	18	20	26
Y	10	15	11	11	25	18	30

[6 Marks]

- d). In a sample of 100 households in a specific city, the following distribution of number of people per household was observed:

Number of people x	1	2	3	4	5	6	7
Number of households f_x	7	7	20	7	18	10	5

The mean number of people per household was found to be 4.0. However, the frequencies for two and four members per household (f_2 and f_4 respectively) are missing.

- i). Calculate the missing frequencies f_2 and f_4 . [3 Marks]
- ii). Find the median of these data, and hence comment on the symmetry of the data. [2 Marks]

QUESTION FIVE (20 Marks) (Optional)

- a). A mathematics teacher recorded the length of time, y minutes, taken to travel to school when leaving home x minutes after 7 am on seven selected mornings. The results are as follows.

x	0	10	20	30	40	50	60
y	16	27	28	39	39	48	51

- i). Plot the data on a scatter diagram. [2 Marks]
- ii). Calculate the equation of the least squares regression line of y on x , writing your answer in the form $y = a + bx$. [3 Marks]
- iii). The mathematics teacher needs to arrive at school no later than 8.40am. The number of minutes by which the mathematics teacher arrives early at school, when leaving home x minutes after 7am, is denoted by z .
- a). Deduce that

$$z = (100 - a) - (1 + b)x$$

[2 Marks]

- b). Hence estimate, to the nearest minute, the latest time that the mathematics teacher can leave home without then arriving late at school. [2 Marks]

- b). The number of deliveries made to a business on any given work day can be represented by the random variable X . The probability distribution of X is shown in the table below.

x	0	1	2	3
$\Pr(X = x)$	0.3	p	0.2	q

The mean number of deliveries is 1.1.

- i). Find the variance of X . [3 Marks]
- ii). Find the probability that on a work day when there is a delivery, the number of deliveries is two or more. [4 Marks]
- c). Data on a sample of 29 claim amounts give a sample mean of Ksh.461.5 and a sample standard deviation of Ksh.618.8.

One claim amount of Ksh.3,657.50 is identified as an outlier and after investigation is found to be in error. Calculate the revised sample mean and standard deviation if this erroneous amount is removed. [4 Marks]

$$s^2 = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f} \right)^2$$

$$\hat{b} = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$\hat{a} = \frac{1}{n} [\sum y - \hat{b} \sum x]$$

$y = bx$