KARATINA UNIVERSITY

UNIVERSITY EXAMINATIONS
2021/2022 ACADEMIC YEAR

SECOND YEAR SECOND SEMESTER REGULAR EXAMINATION

FOR THE DEGREE 0F BACHELOR OF SCIENCE IN (COMPUER SCIENCE AND INFORMATION TECHNOLOGY)

COURSE CODE: STA 205

COURSE TITLE: PROBABILITY AND STATISTICS

DATE:8th June, 2022 TIME:8:00am-12 noon

INSTRUCTION TO CANDIDATES

- SEE INSIDE

INSTRUCTIONS: ANSWER ALL QUESTIONS IN SECTION A AND ANY OTHER THREE IN SECTION B

SECTION A (31 MARKS)

QUESTION ONE (15 MARKS)

a) List and explain three methods of random sampling (6 marks)

b) The probability of a rare disease striking a given population is 0.003. A Sample of 1000 was examined. Find the:

i) Expected number suffering from the disease (1mark)

ii) Standard deviation (2marks)

iii) Probability that at most two persons were found to be stricken by the rare disease (2marks)

c) A sample of 200 people with a particular disease was selected. Out of these, 100 were given a drug and the others were not given any drug. The results were as follows;

Number of people			
	Drug	No Drug	Total
Cured	65	55	120
Not Cured	35	45	80
Total	100	100	200

Test whether the drug is effective or not at 5% level of significance (4 marks)

QUESTION TWO (16 MARKS)

a) (i) Define the term random variable (1 mark)

(ii) State the difference between discrete random variable and continuous random variable

(2 marks)

(iii) The following is a probability distribution function

$$f(x)=\begin{cases} kx(1-x^2), & 0\le x\le 1 \\ 0, & elsewhere \end{cases}$$

k being a constant. Find the:

I. Value of k (2 marks)

II. Mean of the distribution (2 marks)

b) The mean and the standard deviation grade points of a random sample of 36 college students are calculated to be 2.6 and 0.3 respectively. Find 99% confidence interval for the mean of the entire class. (3 marks)

c) Consider the data bellow

Mean temperature ($x^{\circ}C$)	7	5	5	9	10	5	18	19
Changing consumption (y)	29	30	34	27	23	25	19	18

i) Calculate the least square regression line of y on x. (5 marks)

ii) Estimate the values of y given $x=12$ (1 mark)

SECTION B (39 MARKS)

QUESTION THREE (13 MARKS)

a) It is expected that 10% of the production from industrial processing firm will be defective. Calculate the probability that in a sample of 10 units chosen at random

i) Only 2 will be defective (3 marks)

ii) At least 2 will be detective (3 marks)

b) A factory produces blades in packets of 10. The probability of a blade to be defective is 0.2%. Use the poisson distribution approximation of binomial distribution to find the number of packets having two defective blades in a consignment of 10,000 packets. (3 marks)

c) An industrial processing machine1 produces items which are normally distributed. 8% of them weigh over 64kg and 31% weigh under 45kg. Calculate the:

i) Mean weight (2 marks)

ii) Standard deviation for this distribution. (2marks)

QUESTION FOUR (13 MARKS)

a) Explain four properties of a good estimator [illegible] - unbiased - consistent - efficient - (4 marks)

b) A machine is producing ball bearings with diameter of 0.5 inches. It is known that the standard deviation of the ball bearing is 0.005 inches. A sample of 100 ball bearings is selected and their average diameter is found to be 0.498 inches. Determine the 99 per cent confidence interval (4 marks)

c) In a random sample of 400 adults and 600 teenagers who watches a certain TV program, 100 adults and 300 teenagers indicated that they liked it. Construct 95% confident limit for the difference in proportion of all adults and teenagers who watches the program and liked it. (5 marks)

QUESTION FIVE (13 MARKS)

a) Differentiate between:

i) Type I error and type II error (2 marks)

ii) One tailed test and two tailed test (2 marks)

iii) Critical region and acceptance region (2 marks)

iv) Statistics and parameter (2 marks)

$$1\frac{580-1600}{90}$$

b) The mean life time of a sample of 100 light tube produced by a company is found to be 1580 hours with standard deviation of 90 hours. test the hypothesis that the mean lifetime of the tubes produced by the company is 1,600hours ie

$H_0: \mu = 1600$ against $H_1: \mu \neq 1600$ (5 marks)

QUESTION SIX (13 MARKS)

a) State the difference between regression analysis and correlation analysis (2 marks)

b) Calculate the regression equation of y on x

x	1	2	3	4	5
y	2	5	3	8	7

(4 marks)

c) The following is a computer output on decision analysis done on a data where a property appraiser wants to model the relationship between the sale price of a residential property and 3 independent variables . The fit under this model was

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Where y =Sale price in dollars, x_1 =appraised level value (dollars), x_2 =appraised improvements (dollars), x_3 =Area(square feet)

ANOVA

Source	D.f	SS	MS	F	P
Model	3	8779677	----------	----------	< .10000
Error	16	1003491	----------		
Total	19	9783168			

Parameter estimates

Variable	D.f	Parameter estimates	Standard error	t
Intercept	1	1470.276	5746.32	0.26
Land Value	1	9.81449	9.512	1.56
Improved	1	9.82044	9.211	
Area	1	13.5865	6.586	2.05

i) write down the least squares prediction equation (1 mark)

ii) Fill the missing values in the first table (3marks)

iii) Compute the multiple coefficient of determination and comment on its value (3marks)

QUESTION SEVEN (13 MARKS)

a) A company carries two lines of production and has 3 salesmen. The table below shows a month's record of the number of units of the two lines sold by each salesman.

	Salesman		
	1	2	3
Line 1	20	8	15
Line 2	17	16	5

Test the claim that each salesman's ability depends on the line he is selling.

(Use $\alpha = 1\%$ level of significance).

(5 marks)

b) The mean height from a random sample of size 100 is 64 inches with standard deviation 4 inches. Test the statement that the mean height is 68 inches at 0.05 level of significance.

(3marks)

c) Kenya Bureau of Standards quality control officer visit a supermarket and weighs a sample of 10 two Kg cans of cooking fat. The following weights (in Kg) were found 1.95,1.95, 1.96, 2.01, 2.00, 1.89, 2.10, 1.97, 1.99 and 1.99. Construct a 95% confidence interval for the mean weight.

(5marks)

KARATINA UNIVERSITY

UNIVERSITY EXAMINATIONS
2020/2021 ACADEMIC YEAR

SECOND YEAR SECOND SEMESTER EXAMINATION

FOR THE DEGREE 0F BACHELOR OF SCIENCE IN

COMPUER SCIENCE AND BACHELOR OF SCIENCE

IN INFORMATION TECHNOLOGY

COURSE CODE: STA 205

COURSE TITLE: PROBABILITY AND STATISTICS

DATE: 23/06/2021 TIME: 2:00-5:00pm

INSTRUCTION TO CANDIDATES

- SEE INSIDE

minutes and a standard deviation of 8.4 minutes. Construct a 95% confidence interval for the true mean. [4 marks]

1.5 6

QUESTION FOUR [13 MARKS]

a) Explain four properties of a good estimator [4 marks]

b) It is expected that 10% of the production from continuous processing firm will be defective. Calculate the probability that in a sample of 10 units chosen at random;

i) Only 2 will be defective [2 marks]

ii) At least 2 will be detective [3 marks]

iii) At most 2 will be defective [2 marks]

c) Consider the random variable x with the cumulative distribution function (c.d.f)

$$F(x) = \begin{cases} 0 & ,x < 0 \\ 1 - e^{x}, & 0 \leq x \end{cases}$$

k-1

i. What is the pdf of x? [1 mark]

ii. Use the pdf obtained to find the $\Pr(x > 10)$. [1 mark]

Ratio function ANOVA Treatment

QUESTION FIVE [13 MARKS]

a) Give a small description of the concept of Analysis of Variance. [3 marks]

b) The following is a computer output on decision analysis done on a data where a property appraiser wants to model the relationship between the sale price of a residential property and 3 independent variables. The fit under this model was

SST SSE SST

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Where y =Sale price in dollars, x_1 =appraised level value (dollars), x_2 =appraised improvements (dollars), x_3 =Area (square feet)

QUESTION TWO [24 MARKS]

a) State two assumptions made about the probability distribution of the random errors ε in a linear regression model. [2 marks]

b) Give any three reasons for sampling over census. [3 marks]

c) Distinguish between random sampling and non-random sampling [2 marks]

d) Consider the data bellow

Mean temperature (x°C)	7	5	5	8	10	5	18	19
Changing consumption (y)	29	30	34	27	23	25	19	18

i) Calculate the least square regression line of y on x. [4 marks]

ii) Estimate the values of y given $x=12$ [1 mark]

e) State the difference between type I error and type II error in test of hypothesis. [2 marks]

SECTION B [39 MARKS]

QUESTION THREE [13 MARKS]

a) A random sample of 400 adults and 600 teenagers who watches a certain TV program,100 adults and 300 teenagers indicated that they liked it. Construct 95% confident limit for the difference in proportion of all adults and teenagers who watches the program and liked it. [4 marks]

b) An experiment was conducted to investigate the viability of equipment designed to measure the foreign of an audio source. Three independent measurements were recorded by this equipment, they were 4.1, 5.2 and 10.2. Estimate the variance with 90% confidence interval. [5 marks]

c) A paint manufacturer wants to determine the average drying time of a new interior wall paint. If for 12 tests areas of equal size, he obtained a mean drying time of 66.3

3

ANOVA

Source	D.f	SS	MS	F	P
Model	3 SST	8779677	______	______	< .10000
Error	16 SSE	1003491	______		
Total	19. SST	9783168			

Parameter estimates

Variable	D.f	Parameter estimates	Standard error	t
Intercept	1	1470.276	5746.32	0.26
Land Value	1	9.81449	9.512	1.56
Improved	1	9.82044	9.211	
Area	1	13.5865	6.586	2.05

i) Write down the least squares prediction equation [2 marks]

ii) Fill the missing values in the first table [3 marks]

c) Here are the number of hours that 10 students spent working at their jobs the week before a final exam, and their scores on that exam. Calculate the correlation coefficient r and coefficient of determination for these data. [5 marks]

Hours	20	15	0	30	20	15	0	30	0	0
Score	87	71	89	60	86	89	91	65	94	92

5

QUESTION SIX [13 MARKS]

a) The random variable X has a probability density function given by:

$$f(x)=\begin{cases} kx(1-x^2), & 0\le x\le 1 \\ 0, & elsewhere \end{cases}$$

k being a constant. Find the:

i. Value of k [2 marks]

ii. Mean of the distribution [2 marks]

b) Suppose you plan to survey employees to find their medical expenses, you want to be 95% confident that the sample mean is within ± sh.500. A pilot study shows that the standard deviation is about sh.8000. What sample size will you use? [3 marks]

c) In a frequently traveled stretch of the Uganda-Nairobi highway, where the posted speed is 70kph, it is thought that vehicles travel on the average of at least 75kph. To check this claim the following radar measurements of the speeds are obtained for 10 vehicles travelling on this stretch of the highway. 80, 74, 79, 66, 69, 77, 81, 79, 65, 78. Do the data provide sufficient evidence to indicate that the mean speed at which vehicle travel on this stretch is almost 75kph? Test the approximate hypothesis using $\alpha=0.01$.

[6 marks]

QUESTION SEVEN [13 MARKS]

A random sample of insurance policies on the contents of private houses was examined for each of three insurance companies and the sum insured under each policy noted, the results (in "00,000") are shown below.

										10	
Company 1	36	28	32	43	30	21	33	37	26	34	320
Company 2	26	21	31	29	27	35	23	33		8	225
Company 3	39	28	45	37	21	49	34	38	44	9	335

280

Use ANOVA to test if there is mean difference among the companies. Use $\alpha=0.05$.

SSTr		k-1	$\frac{SSTR}{}$
SSE		n-k	$\frac{SSE}{n-k}$
Total	SST 6	n-1	

MR =

STA 205 CAT 1

Question One

a) Define a random variable ⇒ (1 mark)

b) The probability of a rare disease striking a given population is 0.003. A Sample of 1000 was examined. Find the:

i) Expected number suffering from the disease (1mark)

ii) Standard deviation (2marks)

iii) Probability that at most two persons were found to be stricken by the rare disease (2marks)

c) The random variable X has a probability density function given by:

$$f(x)=\begin{cases} kx(1-x^2), & 0 \le x \le 1 \\ 0, & elsewhere \end{cases}$$

k being a constant. Find the:

(i) Value of k 4 (2 marks)

(ii) Mean of the distribution (2 marks)

(iii) Variance of the distribution (3 marks)

d) An industrial process mass produces items which are normally distributed. 11.5% of them weigh over 20kg and 5.89%weigh under 10kg. Calculate the:

i) Mean weight (3 marks)

0.08

ii) Standard deviation for this distribution. (3marks)

0.1207

0.645

Question Two

a) State the difference between

i) a parameter and a statistic (2 marks)

ii) random sampling and non-random sampling (2 marks)

b) state and explain two examples of random sampling (2 marks)

c) Kenya Bureau of Standards quality control officer visit a supermarket and weigh a sample of 10 two Kg cans of cooking fat. The following weights (in Kg) were found 1.95, 1.96, 2.01, 2.00, 1.89, 2.10, 1.97, 1.99 and 1.99. Construct a 95% confidence interval for the mean weight. (5 marks)

$= N = \bar{x} + z + \frac{\sigma}{\sqrt{n}}$

$N = \bar{x} + z_c + \frac{\sigma}{\sqrt{n}}$

0.9988
0.5
0.4938

KARATINA UNIVERSITY

UNIVERSITY EXAMINATIONS
2020/2021 ACADEMIC YEAR

SECOND YEAR SECOND SEMESTER EXAMINATION

FOR THE DEGREE 0F BACHELOR OF SCIENCE IN

COMPUER SCIENCE AND BACHELOR OF SCIENCE

IN INFORMATION TECHNOLOGY

COURSE CODE: STA 205

COURSE TITLE: PROBABILITY AND STATISTICS

DATE: 23/06/2021 TIME: 2:00-5:00pm

INSTRUCTION TO CANDIDATES

- SEE INSIDE

INSTRUCTIONS:

1. The paper comprises 7 questions.
2. Attempt question **one** and **two** (**compulsory**) and **3 other** questions (13 marks each).
3. Electronic, Scientific calculators may be used.
4. Statistical tables are provided.
5. Observe further instructions on the answer booklet.

SECTION A [31MARKS]

QUESTION ONE [18 MARKS]

a) Differentiate between a discrete random variable and a continuous random variable. [2 marks]

b) The probability of a rare disease striking a given population is 0.003. A Sample of 10000 was examined. Find the:

i) Expected number suffering from the disease. [2 mark]

ii) Standard deviation. [2 marks]

c) State three differences between regression analysis and correlation analysis. [3marks]

d) An insurance company takes a keen interest in the age at which a person is insured. Consequently, a survey conducted on prospective clients indicated that for clients having the same age the probability that they will be alive in 30 years' time is $\frac{2}{3}$. This probability was established using the actuarial tables. If a sample of 5 people was insured now, find the probability of having the following possible outcomes in 30 years:

i. All are alive. [1 mark]

ii. At least 3 are alive. [4 marks]

iii. At most one is alive. [2 marks]

Define a hypothesis giving any one type of hypothesis. [2 marks]

QUESTION TWO [14 MARKS]

a) State two assumptions made about the probability distribution of the random errors, ε in a linear regression model. [2 marks]

b) Give any three reasons for sampling over census. [3 marks]

c) Distinguish between random sampling and non-random sampling [2 marks]

d) Consider the data bellow

Mean temperature ($x°C$)	7	5	5	9	10	5	18	19	
	49	25	25	81	100	25	324	361	990
Changing consumption (y)	29	30	34	27	23	25	19	18	
	36	35	39	36	33	30	37	37	
x	203	150	136	243	230	125	342	343	1771

i) Calculate the least square regression line of y on x. [4 marks]

ii) Estimate the values of y given $x = 12$ [1 mark]

e) State the difference between type I error and type II error in test of hypothesis. [2 marks]

SECTION B [39 MARKS]

QUESTION THREE [13 MARKS]

a) A random sample of 400 adults and 600 teenagers who watches a certain TV program,100 adults and 300 teenagers indicated that they liked it. Construct 95% 1.96. confident limit for the difference in proportion of all adults and teenagers who watches the program and liked it. [4 marks]

b) An experiment was conducted to investigate the viability of equipment designed to measure the foreign of an audio source. Three independent measurements were recorded by this equipment, they were 4.1, 5.2 and 10.2. Estimate the variance with 90% confidence interval. 1.65 [5 marks]

c) A paint manufacturer wants to determine the average drying time of a new interior wall paint. If for 12 tests areas of equal size, he obtained a mean drying time of 66.3

minutes and a standard deviation of 8.4 minutes. Construct a 95% confidence interval for the true mean. [4 marks]

QUESTION FOUR [13 MARKS]

a) Explain four properties of a good estimator [4 marks]

b) It is expected that 10% of the production from continuous processing firm will be defective. Calculate the probability that in a sample of 10 units chosen at random;

i) Only 2 will be defective $p = 0.1$ [2 marks]

$n = 10$

ii) At least 2 will be detective x [3 marks]

iii) At most 2 will be defective [2 marks]

c) Consider the random variable x with the cumulative distribution function (c.d.f)

$$F(x) = \begin{cases} 0 & ,x < 0 \\ 1 - e^{x}, & 0 \le x \end{cases}$$

i. What is the pdf of x? [1 mark]

ii. Use the pdf obtained to find the $\Pr(x > 10)$. [1 mark]

QUESTION FIVE [13 MARKS]

a) Give a small description of the concept of Analysis of Variance. [3 marks]

b) The following is a computer output on decision analysis done on a data where a property appraiser wants to model the relationship between the sale price of a residential property and 3 independent variables. The fit under this model was

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Where y =Sale price in dollars, x_1 =appraised level value (dollars), x_2 =appraised improvements (dollars), x_3 =Area (square feet)

ANOVA

Source	D.f	SS	MS	F	P
Model	3	8779677	2926,559	46.6620	< .10000
Error	16	1003491	62,718.475		
Total	19	9783168	18		

Parameter estimates

Variable	D.f	Parameter estimates	Standard error	t
Intercept	1	1470.276	5746.32	0.26
Land Value	1	9.81449	9.512	1.56
Improved	1	9.82044	9.211	
Area	1	13.5865	6.586	2.05

i) Write down the least squares prediction equation [2 marks]

ii) Fill the missing values in the first table [3 marks]

c) Here are the number of hours that 10 students spent working at their jobs the week before a final exam, and their scores on that exam. Calculate the correlation coefficient r and coefficient of determination for these data. [5 marks]

	400	225	0	900	400	225	0	900	0	0	
x Hours	20	15	0	30	20	15	0	30	0	0	= 130
y Score	87	71	89	60	86	89	91	65	94	92	= 824
xy											

QUESTION SIX [13 MARKS]

a) The random variable X has a probability density function given by:

$$f(x)=\begin{cases} kx(1-x^2), & 0\le x\le 1 \\ 0, & elsewhere \end{cases}$$

k being a constant. Find the:

i. Value of k [2 marks]

ii. Mean of the distribution [2 marks]

b) Suppose you plan to survey employees to find their medical expenses, you want to be 95% confident that the sample mean is within $\pm$ sh.500. A pilot study shows that the standard deviation is about sh.8000. What sample size will you use? [3 marks]

c) In a frequently traveled stretch of the Uganda-Nairobi highway, where the posted speed is 70kph, it is thought that vehicles travel on the average of at least 75kph. To check this claim the following radar measurements of the speeds are obtained for 10 vehicles travelling on this stretch of the highway. 80, 74, 79, 66, 69, 77, 81, 79, 65, 78. Do the data provide sufficient evidence to indicate that the mean speed at which vehicle travel on this stretch is almost 75kph? Test the approximate hypothesis using $\alpha=0.01$.

[6 marks]

QUESTION SEVEN [13 MARKS]

A random sample of insurance policies on the contents of private houses was examined for each of three insurance companies and the sum insured under each policy noted, the results (in "00,000") are shown below.

y1	Company 1	36	28	32	43	30	21	33	37	26	34
y2	Company 2	26	21	31	29	27	35	23	33		
y3	Company 3	39	28	45	37	21	49	34	38	44	

27.

Use ANOVA to test if there is mean difference among the companies. Use $\alpha=0.05$.