



MACHAKOS UNIVERSITY

University Examinations for 2022/2023 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

THIRD YEAR SEMESTER EXAMINATIONS FOR

BACHELOR OF SCIENCE (MATHEMATICS)

SMA 364: APPLIED STATISTICAL METHODS

DATE:

TIME:

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

QUESTION ONE (COMPULSORY) (30 MARKS)

a) Briefly differentiate between the following terms as they apply in statistics

- i. Simple Linear and Multiple Linear Regression models (2 marks)
- ii. Logistic regression and Multinomial regression (2 marks)
- iii. One way ANOVA and two way ANOVA (2 marks)

b) The data below shows the number of personnel assigned to harvest flowers in a garden and average number plots attended per day.

Number of farm attendants	1	2	3	4	5
Number plots	2	5	3	8	7

Using the data determine

- i. The linear regression equation of number plots on number of farm attendants (4 marks)
 - ii. The value of R^2 (4 marks)
- c) Explain briefly the following terms as used in applied statistics
- i) Fixed model effects (3 marks)
 - ii) Random model effects (3 marks)
 - iii) Mixed model effects (3 marks)

- d) Highlight three ways of dealing with a missing data (3 marks)
- e) Distinguish the following terms as used in applied statistics
- i. A variable and a case
 - ii. Multicollinearity and outliers (4 marks)

QUESTION TWO (20 MARKS)

- a) Discuss three methods of variable selection for model buildings (6 marks)
- b) The following are the results after conducting research on a micro finance company performance as affected by , wages, cost of capital and technology.

$$\hat{y} = -1.275 + 0.128x_1 + 0.353x_2 + 0.045x_3 + \mu$$

$$Se \ (9.348) \ (0.210) \ (0.132) \ (0.128)$$

$$R^2 = 0.928 \ n = 25$$

Where Y is the microfinance performance as measured by return on assets

x_1 is wages

x_2 is cost of capital

x_3 is the technology

μ is the error term

- i. Write the R- script for the output (2 marks)
- ii. Interpret the coefficient of determination. (1 mark)
- iii. Interpret the results using regression coefficients (5 marks)
- iv. Test the significance of the coefficients at $\alpha = 0.05$ level (6 marks)

QUESTION THREE (20 MARKS)

- a) Highlight four ways of carrying out regression diagnostics by utilizing the plots of residuals (4 marks)
- b) A researcher presumed that the average microorganism abundance per quadrat was 65. A random sample on 10 quadrats yielded the following: 66, 65, 69, 70, 69, 71, 70, 63, 64 and 68. With the test statistics output summarized in table 1.

Table 1: 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. Write the R- script for the above output
 - 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)
- c) A random sample of 400 persons was selected from each of three age groups and each person was asked to specify which of the three presidential candidates she/ he preferred. The results are shown in table 2:

Table 2

Age group	Presidential Candidate		
	A	B	C
Under 30	120	30	50
31 – 50	10	75	15
51 and above	10	30	60

Test the hypothesis that the populations were homogeneous with respect to the presidential candidate they preferred despite their age difference. (10 marks)

QUESTION FOUR (20 MARKS)

- a) Discuss four methods of variable selection for model buildings (8 marks)
- b) A contracting farming agency collected the data and predicted the yield of biashara green grams using ANOVA and regression, based on the amount of water, spacing and fertilizer applied: the following is part of the results:

SUMMARY OUTPUT

Regression Statistics					
Multiple R	0.625373				
R Square	0.391092				
Adjusted R Square	-0.21782				
Standard Error	7.63801				
Observations	7				
ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance</i>
Regression	3	112.411	<i>q</i>	<i>r</i>	0.637562
Residual	3	<i>p</i>	58.33919		
Total	6	287.4286			

i.

Determine the values of p , q and r .

(3 marks)

ii.

Write the R-script for the output

(3 marks)

iii.

Interpret the finding in this analysis

(3 marks)

iv.

Explain the difference between R Square and R Adjusted

(3 marks)

QUESTION FIVE (20 MARKS)

- Discuss two ways of dealing with the problem of an outlier in a data set (4 marks)
- Distinguish the following terms as used in applied statistics
 - Confounding variable and Noise variable
 - Univariate and Bivariate data dimensions (4 marks)
- The table below is part of an output of data analyzed whose response variable was whether or not (1=yes,0=no) a student passed an examination (y), the predictor variables included the students age (x1), gender (x2), fee balance (x3) in Ksh'000', family size (x4) and the religion coded such that (1=catholic,2=protestant,3=muslim,4=hindu) (x5)

Variables in the equation								
Step1 ^a	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Family-size	1.338	.494	4.419	1	.036	.354	.135	.932
Age	1.197	.120	2.710	1	.043	.821	.649	1.038
Fee bal	.201	.173	1.342	1	.247	1.222	.870	1.716
Religion			1.085	3	.781			
Religion(1)	-.073	1.477	.002	1	.961	.930	.051	16.801
Religion(2)	1.468	1.743	1.710	1	.047	4.342	.143	132.234
Religion(3)	-19.860	11.714	.000	1	.999	.000	.000	.
Gender(1)	.645	1.343	.231	1	.631	1.906	.137	26.476
Constant	8.329	4.262	3.819	1	.051	412.576		

- Write down the regression equation for the output. (2 marks)
- With justification highlight the variables that contributed significantly to the prediction of the examination status of a student. (4 marks)
- Interpret the regression coefficients , $\text{Exp}(\beta)$ and the 95%CI for $\text{Exp}(\beta)$ of the variables highlighted in (ii) (6 marks)