



SOUTH EASTERN KENYA UNIVERSITY

UNIVERSITY EXAMINATIONS 2016/2017

FIRST SEMESTER EXAMINATION FOR THE DEGREES OF BACHELOR OF SCIENCE IN STATISTICS & ACTUARIAL SCIENCE

STA 302: LINEAR MODELING 1

DATE: 15TH DECEMBER, 2016

TIME: 10.30-12.30PM

INSTRUCTIONS

ANSWER QUESTION ONE AND ANY OTHER TWO QUESTIONS

QUESTION ONE (30 marks)

- a) A simple linear regression model is of the form $Y = \beta_0 + \beta_1 X + \varepsilon$. Define, Y, X and ε . (3marks)
- b) Using part (a) above write the general linear regression model for n observations. State three assumptions for this model to be valid. (4marks)
- c) Define the term **residual** and briefly explain its significance in regression analysis. (4marks)
- d) State and explain the three properties of residuals. (5marks)
- e) A statistics professor wants to use the number of hours a student studies for a statistics final exam(X) to predict the final exam (Y). A regression model was fit based on data collected for a class during the previous semester, with the following results:
 $\hat{Y}_i = 35.0 + 3X_i$.
What is the interpretation of the intercept, b_0 and the slope b_1 ? (2marks)
- f) The table below shows the mean yields of soya bean plants (grams per plant) obtained in response to the indicated levels of Ozone exposure over the growing season. Perform linear regression analysis to estimate the parameters of the regression model. State the regression model. (5marks)

Ozone(x)	0.020	0.080	0.112	0.115
Yield(Y)	243	238	228	202

- g) Using the regression model obtained in part (f) above, compute the estimated values and residuals for the linear regression of Soya bean yield on Ozone dosage. (3marks)

- h) In the un weighted regression problem with $n=17$, $p'=2$, the results included $\hat{\sigma} = 0.379$, $\hat{e}_i = 1.37$ and $h_{ii} = 0.0635$, where $i = 12$ represents the 12th observation. Compute the standardized residual, r_{12} and the student's test statistic, t_i . (4marks)

QUESTION TWO (20marks)

- a) The following data indicate the relationship between X, the specific gravity of a wood sample, and Y, its maximum crushing strength in compression parallel to the grain.

X _i	0.41	0.46	0.44	0.47	0.42	0.39	0.41	0.44	0.43	0.44
Y _i	1850	2620	2340	2690	2160	1760	2500	2750	2730	3120

- (i) Estimate the regression coefficients. (6marks)
(ii) Predict the maximum crushing strength of a wood sample whose specific gravity is 0.43. (2marks)
- b) In an un weighted regression problem with $n=54$, $p'=5$, the results included $\hat{\sigma} = 4.0$ and the following statistics for four of the cases.

\hat{e}_i	1.0	1.732	9.000	10.295
h_{ii}	0.9	0.75	0.25	0.185

For each of these four cases, Compute the standardized residual, r_{12} and the student's test Statistic, t_i . Test for each of the four cases to be an outlier at $\alpha = 5\%$. (12marks)

QUESTION THREE (20marks)

- a) Three methods of packaging frozen foods were compared. The response variable was ascorbic acid (mg/100g). The data are shown in table below. Test the three means of the three methods. That is, test $H_0: \mu_1 = \mu_2 = \mu_3$ versus $H_1: \mu_1 \neq \mu_2 \neq \mu_3$ at $\alpha = 1\%$ significance level. (12marks)

A	B	C
14.29	20.06	20.03
19.10	20.63	26.23
19.10	18.01	22.74
16.25	19.55	24.04
15.10	19.47	23.37
16.50	19.07	25.01
19.50	18.38	23.27

- b) A certain airplane part must withstand extremes of temperature. The part can be made from a number of metal alloys; the one to be chosen must have the greatest strength Y for a given density X. An experiment is designed involving 5 alloys and 5 parts per alloy. In hopes of obtaining a lighter part, the density of each alloy is deliberately varied within a

safe range. The data are analyzed by covariance procedures to yield the following information.

Source	df	SS _(x)	SP	SS _(y)
Alloys	4	200	300	2500
Error	20	300	1200	7500

Complete the analysis of covariance and test H_0 versus H_1 at $\alpha = 5\%$ (8marks)

QUESTION FOUR (20 marks)

An experiment was conducted involving three different advertising media, each used for five fast food restaurants of a certain franchise. The 15 restaurants were located in different but comparable cities, and they were randomly assigned to the 3 advertising media; radio, newspaper and television. All advertising took place during the same time periods. Profits Y in thousands of dollars were recorded for the same time period. Although all restaurants were of the same size, they employed different number of workers. Since additional employees may affect profits, the number of employees was used as a concomitant variable X.

Medium

I		II		III	
X	Y	X	Y	X	Y
10	32	21	24	35	18
14	20	26	20	38	12
19	13	31	7	43	3
26	7	36	4	47	-6
26	2	41	-5	52	-10

- What assumptions must be made in order to perform the analysis of covariance? (3marks)
- Compute the uncorrected sums of squares and products. (7marks)
- Compute the corrected sum of squares and products. (5marks)
- Compute the adjusted sums of squares. (5marks)

QUESTION FIVE (20 marks)

The table below shows data that was collected and recorded during a research experiment.

X	1	2	3	4	5	6	7	8	9	10
Y	20.6	30.6	55.2	71.6	97.5	131.8	156.3	197.2	238.8	291.8

- Sketch a plot for the data. (3marks)
- Use the data to fit a polynomial regression model defined by;

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + e$$
 (14marks)
- (I) State the estimated quadratic equation. (1mark)
 (II) Use the estimated equation in (I) to predict the values of Y at X=12. (2marks)