

LAIKIPIA



STAT 423

UNIVERSITY

UNIVERSITY EXAMINATIONS

SECOND SEMESTER 2023/2024 ACADEMIC YEAR

FOURTH YEAR EXAMINATION FOR THE DEGREE OF
BACHELOR OF SCIENCE (ECONOMIC & STATISTICS)

STAT 423: ADVANCED REGRESSSION ANALYSIS

STREAM: R

TIME: 2 HRS

DAY: TUESDAY[11.30A.M – 1.30P.M]

DATE: 09/04/2024

THIS QUESTION PAPER CONSISTS OF SIX (6) PAGES

PLEASE DO NOT OPEN UNTIL THE INVIGILATOR SAYS SO



INSTRUCTIONS

- (i) Answer question one and any other two questions
- (ii) Show all the workings clearly
- (iii) Do not write on the question paper

QUESTION ONE (30 MARKS)

- a) Suppose you fit the cubic model $E(Y) = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$ to 24 data points and found That $F = 37.37$
- i) Do the data provide sufficient evidence to indicate that the model contributes information for the prediction of Y? **(3 Marks)**
 - ii) Use the value of F to calculate R^2 . Interpret its value. **(3 Marks)**
- b) Consider the following outputs from Poisson Regression and logistic regression

Model 1- logistic regression

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.301468	0.431332	-0.699	0.484600
X1	-0.001928	0.009799	-0.197	0.843981
X2	0.223370	0.061134	3.654	0.000258 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 693.15 on 499 degrees of freedom

Residual deviance: 679.14 on 497 degrees of freedom

AIC: 685.14

Model 2- Poisson regression

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	9.802e-01	1.845e-01	5.312	1.08e-07 ***
X1	-2.972e-01	6.892e-02	-4.312	1.62e-05 ***
X2	-4.734e-05	3.678e-03	-0.013	0.99

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Poisson family taken to be 1)

Null deviance: 762.37 on 499 degrees of freedom

Residual deviance: 743.57 on 497 degrees of freedom

AIC: 1729.8

- i) Comment on the significance of the explanatory variables in Model 1. (2 Marks)
 - ii) Comment on the goodness of fit in Model 1 (2 Marks)
 - iii) Comment on the significance of the explanatory variables in Model 2. (2 Marks)
 - iv) Comment on the goodness of fit in model 2 without using the null deviance (2 Marks)
 - v) Compare the goodness of fit between model 1 and model 2 (2 Marks)
- c) i. Which are the 2 most likely causes of lack of fit in Poisson Regression? (2 Marks)
 ii. How do you overcome this problem of lack of fit? (2 Marks)
- d) Why does OLS regression fail to work when dealing with count data? (2 Marks)
- e) Consider the following data

income(y)	8400	10500	81930	171000	6790	57900	8100	200000	71810	12300
Age(x)	21	42	32	41	30	38	37	36	38	55

- i) Calculate the tri-cube weights that could be used to fit weighted least squares also referred to as LOWESS in local polynomial regression. (4 Marks)
- ii) Write a well commented R program that
 - I. Reads the data to R (1 Mark)
 - II. Plots the scatter plot for the above data (1 Mark)
 - III. Fits and superimposes the fitted local polynomial regression line on to the plot in (II) above. (2 Marks)

QUESTION TWO (20 MARKS)

- a) Discuss the lowess procedure in Local polynomial regression (6 Marks)
- b) The compressive strength of an alloy fastener used in aircraft construction is being studied. Ten observations selected over the range 2500-4300 psi and a number of fasteners were tested at those loads. The numbers of fasteners failing at each load were recorded. A logistic regression analysis using R language produced the following results:

Coefficients:

	Estimate	Stderror	z value	Pr(> z)
(intercept)	-5.33971	0.53570	-9.785	<2e-16
Load	0.00155	0.00016	9.829	<2e-16

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 956.17 on 689 degrees of freedom



Residual deviance: 843.71 on 688 degrees of freedom AIC: 847.71

- i) Give the fitted logistic regression model and least squares fit model (2 Marks)
- ii) Does the model deviance indicate that the logistic regression model in part (i) is adequate (3 Marks)
- iii) Use the odds ratio to determine the increase in odds of failure that results by increasing the load by 140 psi. Hence find probability of failure. (3 Marks)
- iv) Find approximate 95% confidence intervals for the model parameters of the logistic regression model fit to the data and their respective probability scale. Interpret the confidence intervals. (6 Marks)

QUESTION THREE (20 MARKS)

- a) Differentiate between the following terms as used in advanced regression;
 - i. Span and Bandwidth. (2 Marks)
 - ii. Local polynomial regression and kernel regression estimation. (2 Marks)
 - iii. Poisson Regression and logistic regression. (2 Marks)

- b) Write a well commented R program that does the following
 - I. Generates variables of size 1000 each as follows
 - i) $e \sim N(0,1)$ (1 Mark)
 - ii) $x_1 \sim \chi^2(3)$ (1 Mark)
 - iii) $x_2 \sim P(10)$ (1 Mark)
 - II. Generates the response variable Y as follows

$$Y = 4.0\pi + \sqrt{2.5X_1} + 1.87X_2 + e$$
 (3 Marks)
 - III. Regress Y on x1 and x2 through local polynomial lowess mean function and plots the results in a mesh/regression surface. (6 Marks)
 - IV. Repeats part (II) above 100 times (2 Marks)



QUESTION FOUR (20 MARKS)

a) Differentiate between the following terms

- i) Parametric models
- ii) Non parametric models

(4 Marks)

b) Consider the following data

Gender	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
age	23	35	26	25	28	31	23	25	24	31	26	26	29	22	25
Number of accidents	2	1	1	0	2	1	1	1	3	1	0	0	0	2	5

Write an R program that

- i) Input the above table into R. **(1 Mark)**
- ii) Regress the data using logistic Regression by choosing the appropriate dependent variable **(4 Marks)**
- iii) Regress the data using Poisson Regression by choosing the appropriate dependent variable **(3 Marks)**

c) Suppose you fit a Poisson regression model with $\mu = \exp(\alpha + \beta x)$ to 54 data points using R and obtained the following results:

Coefficients:

	Estimate	Stderror	z value	Pr(> z)
(intercept)	3.69196	0.04541	81.302	<2e-16
X	-0.20599	0.05157	-3.994	6.49 e-05

(Dispersion parameter for Poisson family taken to be 1)
 Null deviance: 297.37 on 53 degrees of freedom
 Residual deviance: 210.42 on 52 degrees of freedom
 AIC: 493.06

- i) Give the fitted Poisson regression model and least squares fit model. **(3 Marks)**
- ii) Find approximate 90% confidence intervals for the model parameters of the Poisson regression model fit to the data and interpret the confidence intervals.

(5 Marks)



QUESTION FIVE (20 MARKS)

a) Briefly illustrate the following terms as used in advanced regression.

i) Residuals

(3 Marks)

ii) Polynomial regression

(3 Marks)

b) A regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \epsilon$ has been fit to a sample of $n=12$ observations from a study investigating the carbonation level (Y) of a soft drink beverage as it relates to the temperature (x_1) of the product and the filler operating temperature (x_2). The following quantities were obtained:

$$(X'X)^{-1} = \begin{bmatrix} 528019.0007 & -16937.7882 & -23305.8707 & 747.3367 \\ -16937.7882 & 543.3792 & 747.4860 & -23.9714 \\ -23305.8707 & 747.4860 & 1029.8937 & -33.0199 \\ 747.3367 & -23.9714 & -33.0199 & 1.0588 \end{bmatrix},$$

$$X'Y = \begin{bmatrix} 95.34 \\ 2970.37 \\ 2204.92 \\ 68721.71 \end{bmatrix} \text{ and}$$

$$SST = 342.1899$$

i) Find the least squares estimates for the regression coefficients $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_{12}$ **(4 Marks)**

ii) What is the predicted value of the carbonation level Y when $x_1 = 30$ and $x_2 = 20$. **(2 Marks)**

c) Discuss the lowess procedure for fitting a local- polynomial regression curve **(8 Marks)**

