NANYANG TECHNOLOGICAL UNIVERSITY

SEMESTER I EXAMINATION 2024-2025

MH3510– Regression Analysis

December 2024 TIME ALLOWED: 2 HOURS

INSTRUCTIONS TO CANDIDATES

1. This examination paper contains **FOUR (4)** questions and comprises **FIVE (5)** printed pages.
2. Answer **all** questions. The marks for each question are indicated at the beginning of each question.
3. Answer each question beginning on a **FRESH** page of the answer book.
4. This is **a RESTRICTED OPEN BOOK exam. You are only allowed to bring into the examination hall ONE DOUBLE-SIDED A4-SIZE REFERENCE SHEET WITH TEXTS HANDWRITTEN OR TYPED ON THE A4 PAPER WITHOUT ANY ATTACHMENTS (e.g. sticky notes, post-it notes, gluing or stapling of additional papers).**
5. Candidates may use calculators. However, they should write down systematically the steps in the workings.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

QUESTION 1. (25 marks)

Suppose that a company wishes to study the effectiveness of a help center in improving the customer experience. To this end, the center keeps a complete log of all users and the duration of each visit. The researchers select 100 customers registering at the center against the center log to see the minutes spent (if any) and to ask them to rate the overall values of their experience on a scale from 0 to 100. The 100 customers on average spent 3 minutes at the center with a standard deviation of 5 minutes. On average they gave a rating of 55 with a standard deviation of 9.3. The correlation coefficient between mintues spent and rating is 0.45.

(i) Identify the response variable and predictor variable.

(ii) Find a regression model for the variables of interest and estimate the parameters.

(ii) Is the regression coefficient significant at 5% significance level ($t_{\infty}^{0.05} = 1.96, t_{98}^{0.05} = 1.98, t_{1}^{0.01} = 63.6$)?

(iv) What percent of the observed variation in the experience can be explained by the time spent ? Does it suggest a good fit ?

(v) Estimate the increase in the expected number of rating when there are 3 more minutes spent at the center by customer A than customer B.

QUESTION 2. (25 marks)

(a) A model for sales of a type of cell phone may be formulated as

$$y_i = \frac{\beta_0 x_{i1}}{2 + \beta_1 x_{i1} + \beta_2 x_{i2}}, \quad i = 1, \cdots, n,$$

where $\{y_i, x_i\}, i = 1, \cdots, n$ are known.

(i) Suggest a transformation for y_i such that the parameters β_0, β_1 and β_2 can be estimated via the least square approach.

(ii) Write down the estimates for β_0, β_1 and β_2.

Question 2 continues on page 3

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

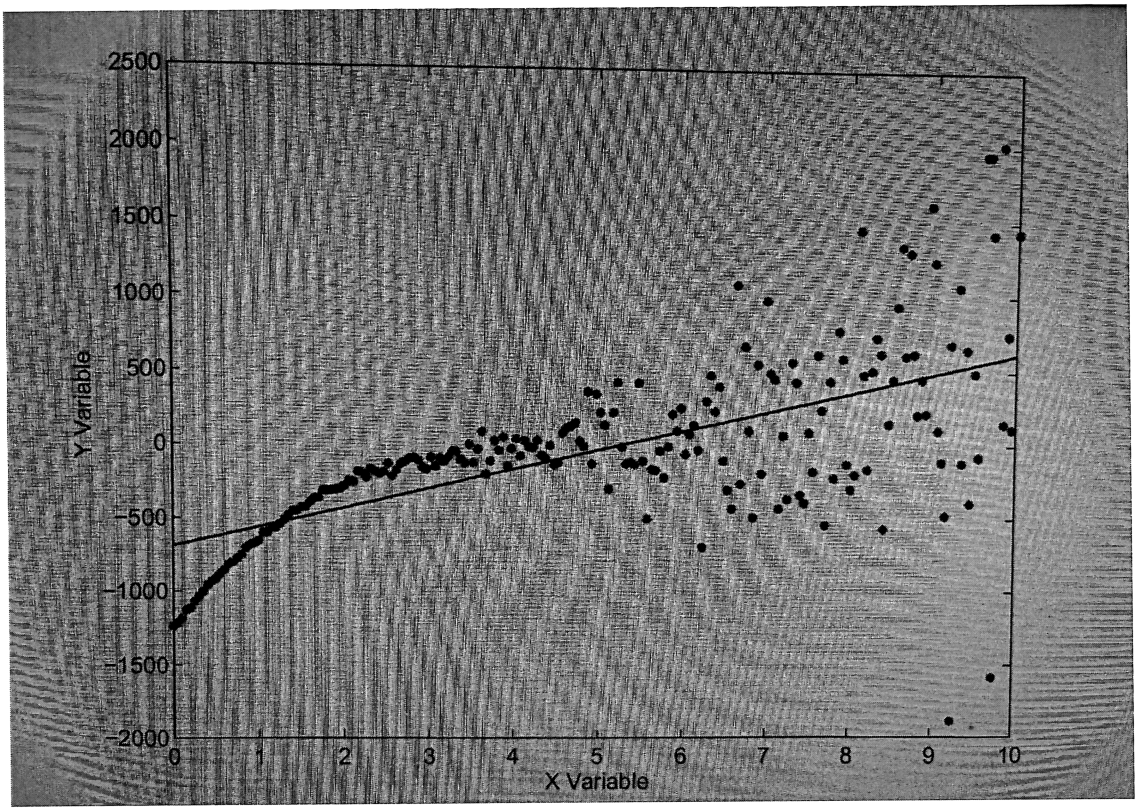


Figure 1: residuals against predictor

(b) A scatter plot is given in Figure 1. A regression line is fitted to the data. List the linear regression model assumptions which are violated.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

QUESTION 3. **(25 marks)**

A medical center was interested in the association between prostatespecific antigen (PSA) and a number of prognostic clinical measurements in men with advanced prostate cancer. Data were collected on 90 men who were about to undergo radical prostectomies. Each line of the data set provides information on four variables for each person. They are:

y	PSA level
x_1	Cancer Volume
x_2	Age
x_3	Cleason score

After fitting a linear model in R, we obtain the following model summary:

	Estimate	Std. Error	t value	Pr(> | t |)
(Intercept)	-46.58	38.91	-1.197	0.235
Volume	2.7804	0.4105	6.773	0.000
Age	-0.3450	0.4582	-0.75	0.455
Cleason	9.963	6.142	1.622	0.108

Moreover the R output indicates that $S_{yy} = 169571$ and $SSE = 95703$.

(i) Interpret the coefficient β_2 (corresponding to Age).

(ii) Test whether there is a regression relation between the response variable and the three predictor variables at the a = 0.05 level. You may use the critical value $F(0.95; 3, 86) = 2.71065$.

(iii) Determine R^2. Explain what R^2 means.

(iv) Test whether both x_2 and x_3 can be dropped from the regression model, given that x_1 is retained. Use a = 0.05. You may use the facts that $SSE(x_1) = 96867, SSE(x_1, x_2) = 96451$ and $SSE(x_1, x_3) = 95023$. Here the error sum of squares are denoted by $SSE(x_1)$ and $SSE(x_1, x_2)$ respectively when x_1 only is in the model and when x_1 and x_2 only are in the regression model. You may use the critical values F(0.95; 2, 86) = 3.102552 or F(0.975; 1, 87) = 5.2.

(v) What happens to the value of y when the cancer volume of one patient increases by two units ?

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

QUESTION 4. **(25 marks)**

A factory was having trouble in its product with production rates. After extensive discussions the factory decided to investigate the effects of six different combinations of three machines and two production lines. One of the issues that the factory had was an inability to reproduce production rates under what seemed to be the same conditions. Do each experimental runs twice to obtain an estimate of this inherent variability. The data are given as follow.

	Line	
Machine	1	2
A	5,6	10,8
B	4,6	14,16
C	14,11	12,13

Consider a two-way ANOVA model for the above study.

(i) Write down a model for the above question and specify the model assumption.

(ii) Carry out a significance test at the 5% significance level to test for any interaction between the two factors ($F_{3,6}^{0.05} = 4.76, F_{2,6}^{0.05} = 5.14, F_{2,12}^{0.05} = 3.89$).

(iii) Give the least square estimates of the main effects of the two factors, respectively.

Consider a one-way ANOVA model for the study where the factor of concern is the combination of two factors.

(iv) Write down a model for the above question and specify the model assumption.

(v) Calculate and comment on a 95% confidence interval for the difference in the production rate between machine A and machine C for Line 1 ($t_6^{0.025} = 2.447, t_{10}^{0.025} = 2.228$).

END OF PAPER

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

MH3510 REGRESSION ANALYSIS

CONFIDENTIAL

Please read the following instructions carefully:

1. **Please do not turn over the question paper until you are told to do so. Disciplinary action may be taken against you if you do so.**

2. You are not allowed to leave the examination hall unless accompanied by an invigilator. You may raise your hand if you need to communicate with the invigilator.

3. Please write your Matriculation Number on the front of the answer book.

4. Please indicate clearly in the answer book (at the appropriate place) if you are continuing the answer to a question elsewhere in the book.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library