國立政治大學 110 學年度第 1 學期 期中考 考試命題紙

考試科目： Regression Analsis (I)　　開課班別： 商院選修　　命題教授: 吳漢銘

考試日期：11 月 11 日 (四) 9:10-10:30	※准帶項目打「O」，否則打「×」				1. 需加發計算紙或答案紙請備註。
本試題共 3 頁，印刷份數: 36 份	Calculator	Book Notes	Dictionary	Cell phone Laptop	2. 為環保節能減碳，試題一律採雙面印刷，如有特殊印製需求，請註記：
備註：注意事項要看!! (§1∼§3)	O	×	×	×	

Note: (1) Fill in your name and student ID on the answer sheet。(2) Answer the questions in English。(3) Answer the questions in the order in which they appear。(4) Pencils are permitted for use。(5) Hand in the question, the answer sheets and the sketch papers。(6) The calculation process is required.

1. (10%) For the given sample observations $\{(X_i, Y_i), i = 1, \cdots, n\}$, we assume a normal error regression model as $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, where ϵ_i are independent normally distributed with mean 0 and variance σ^2. Find the MLEs of the parameters β_0 and β_1.

2. Let $\{(X_i, Y_i), i = 1, \cdots, n\}$ be the observed data and we would like to perform a simple linear regression analysis. Please answer the following questions.

   (a) (8%) Which plots can be used to conduct the diagnostics for predictor variable?

   (b) (12%) The residuals can be used to examine six important types of departures from the simple linear regression model with normal errors. What are those six important types of departures?

   (c) (10%) Describe the Brown-Forsythe Test with a level of significant α (including at least the assumption for the data, the null hypothesis, the test statistics and the decision rule.)

3. (25%) In the textbook, we have already learned some transformations for X and/or Y to ensure that the assumptions for a simple linear regression normal error model are adequate. The transformations are

$$\log_{10}(X), 1/X, \sqrt{X}, X^2, \exp(X), \exp(Y), \log_{10}(Y), 1/Y, \sqrt{Y}, Y^\lambda.$$

   Four real world cases given below are analyzed each by a simple linear regression normal error model.

   (a) A research would like to study the regression relationship between alpha counts per second (Y) and plutonium activity (X) to estimate the activity of plutonium in the material under study.

   (b) A chemist studied the concentration of a solution (Y) over time (X). Fifteen identical solutions were prepared. The 15 solutions were randomly divided into five sets of three, and the five sets were measured, respectively, after 1, 3, 5, 7, and 9 hours.

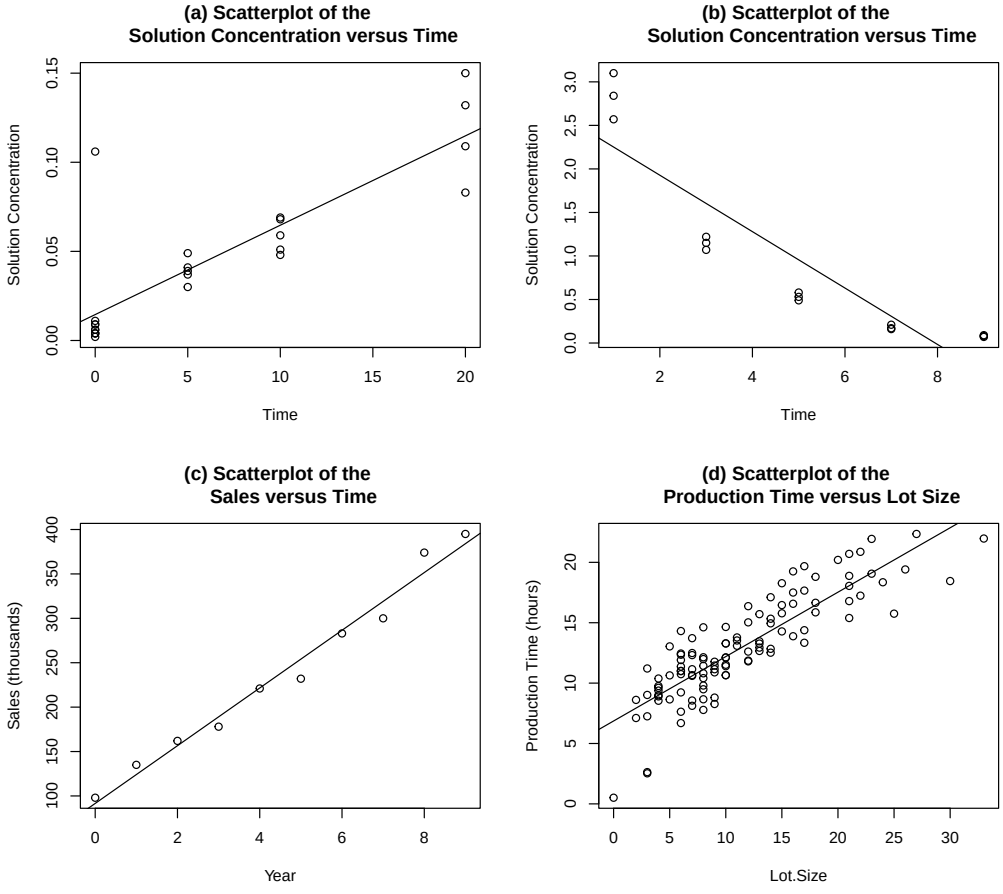
   (c) A marketing researcher studied annual sales of a product that had been introduced 10 years ago. The data is collected, where X is the year (coded) and Y is sales in thousands of units.

   (d) In a manufacturing study, the production times for 111 recent production runs were obtained. The data consists of records for each run the production time in hours (Y) and the production lot size (X).

考試日期：11 月 11 日 (四) 9:10-10:30	※准帶項目打「O」，否則打「×」				1. 需加發計算紙或答案紙請備註。 2. 為環保節能減碳，試題一律採雙面印刷，如有特殊印製需求，請註記：
本試題共 3 頁，印刷份數: 36 份	Calculator	Book Notes	Dictionary	Cell phone Laptop	
備註：注意事項要看!! (§1∼§3)	O	×	×	×	

Based on the scatterplots of Y versus X with a regression line, please indicates whether the transformations are needed for Y and/or X and conclude which transformations are possible for each case. That is, fill in the blank spaces with the transformation methods in the following table in the answer sheet. Mark the blank by "×" if the transformation is not necessary. You don't have to specify the λ value when you think the Box-Cox transformation is appropriate.

Case	Transformation of X	Transformation of Y
(a)		
(b)		
(c)		
(d)		



考試日期：11 月 11 日 (四) 9:10-10:30	※准帶項目打「O」，否則打「×」				1. 需加發計算紙或答案紙請備註。 2. 為環保節能減碳，試題一律採雙面印刷，如有特殊印製需求，請註記：
本試題共 3 頁，印刷份數: 36 份	Calculator	Book Notes	Dictionary	Cell phone Laptop	
備註：注意事項要看!! (§1∼§3)	O	×	×	×	

4. Suppose that we obtain a data set that can be expressed in the form:

$$\{(X_j, Y_{ij}) : i = 1, \cdots, n_j; j = 1, \cdots, c\}, \text{ where } c > 2.$$

Someone would like to use F test for lack of fit to determine whether a simple linear regression model adequately fits the data, where X is the predictor variables and Y is the response.

(a) (5%) What are the assumptions of the lack of fit test?

(b) (5%) What is the full model used for the lack of fit test?

(c) (5%) What is the reduced model used for the lack of fit test?

(d) (5%) What is the null hypothesis for the lack of fit test?

(e) (10%) The Growth rate data are available on the effect of dietary supplement on the growth rates of rats. Here X = dose of dietary supplement and Y = growth rate. The following table presents the data in a form suitable for the analysis ($c = 6, n = 12$). Construct a general ANOVA Table (including Source of Variation, Sum of Square (SS), Degree of Freedom (df), Mean Square (MS) and F statistics) for testing lack of fit of a simple linear regression function.

Data		$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$
Replicate		$X_1 = 10$	$X_2 = 15$	$X_3 = 20$	$X_4 = 25$	$X_5 = 30$	$X_6 = 35$
Y_{ij}	$i=1$	73	85	90	87	75	65
	$i=2$	78	88	91	86		63
	$i=3$			91			

(f) (5%) State the test statistics, decision rule and conclusion. (for all j at 5% level of significance)

(Some numbers: error sum of squares for the reduced model ($SSE(R)$) = 891.73, regression sum of squares (SSR) = 204.27, total sum of squares ($SSTO$) = 1096.00, $F(0.95; 5, 5) = 5.050, F(0.95; 6, 4) = 6.163, F(0.95; 4, 6) = 4.534, F(0.95; 1, 10), F(0.95; 10, 1) = 241.881, F(0.95; 2, 10) = 4.103, F(0.95; 2, 9) = 4.256, F(0.95; 2, 8) = 4.459$; $\hat{Y}_{ij} = 92.003 - 0.498X_j$)