

考試日期：11 月 10 日 (四) 09:10-10:50

* 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4頁· 印刷份數：60 份

Calculator	Book Notes	Dictionary	Cell phone Laptop
------------	------------	------------	-------------------

備註：注意事項要看!! (§1~§3)

O	×	×	×
---	---	---	---

Note: (1) Fill in the student ID number and your name ° (2) Answer all questions in English ° (3) Answer each question in the order it appears ° (4) It is recommended to use a dark ballpoint pen ° (5) The calculation process is required. (算至小數點以下 4 位) (6) Please return both the answer sheet and the question sheet. (7) The total score is 110.

★ For all the problems below, we assume the normal error regression model.

- (5%) What is the "simple linear regression model with distribution of error terms unspecified"? Which assumptions are made for error terms?
- (5%) In a normal error regression model, why do we need to assume that the error terms are normally distributed?
- (10%) What is the purpose of the diagnostic for the predictor variable? How do you do the diagnostic?
- (10%) As described in the textbook, what are the six types of departures from the SLR model?
- (10%) What is meant by residual plot? How can residual plots be used?
- (10%) Several types of departures may occur together in SLR, which ones tend to be serious and why? Which ones tend to be less serious and why?
- (15%) About the Brown-Forsythe Test (according to the textbook),
 - (5%) What are the null hypothesis and alternative hypothesis?
 - (5%) What is the test statistic for this test? (If you use symbols, explain them.)
 - (5%) **Crime rate.** A criminologist studying the relationship between level of education-and crime rate in medium-sized U.S. counties collected the following data for a random sample of 84 counties; X is the percentage of individuals in the county having at least a high-school diploma, and Y is the crime rate (crimes reported per 100,000 residents) last year. Assume that first-order regression model (1.1) is appropriate.

i	1	2	3	...	82	83	84
X_i	74	82	81	...	88	83	76
Y_i	8487	8179	8362	...	8040	6981	7582

We divide the data into the two groups, $X \leq 79$, $X > 79$, and conduct the Brown-Forsythe test with $\alpha = 0.05$. State the decision rule and conclusion. (The R Output (I))

考試日期：11 月 10 日 (四) 09:10-10:50

※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4頁· 印刷份數: 60 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

×

×

×

8. (25%) **Plastic hardness.** Sixteen batches of the plastic were made, and from each batch one test item was molded. Each test item was randomly assigned to one of the four predetermined time levels, and the hardness was measured after the assigned elapsed time. The results are shown below; X is the elapsed time in hours. and Y is hardness in Brinell units.

Data		$j = 1$	$j = 2$	$j = 3$	$j = 4$
Replicate		$X_1 = 16$	$X_2 = 24$	$X_3 = 32$	$X_4 = 40$
Y_{ij}	$i = 1$	199	218	237	250
	$i = 2$	205	220	234	248
	$i = 3$	196	215	235	253
	$i = 4$	200	223	230	246
\bar{Y}_j		200	219	234	249.25

$$\sum_i \sum_j Y_{ij} = 3609, \sum_i \sum_j Y_{ij}^2 = 819499.$$

The SLR analysis report and ANOVA Table conducted by R are given in Output (II) and (III). Perform the F test to determine whether or not there is lack of fit of a linear regression function; use $\alpha = 0.01$.

- (5%) What are the assumptions of the lack of fit F test?
 - (5%) State the full model, and reduced model.
 - (10%) Complete the ANOVA Table (lack of fit). (Output (IV))
 - (5%) State the decision rule, and conclusion.
9. (20%) **Production time.** In a manufacturing study, the production times for III recent production runs were obtained. The table below lists for each run the production time in hours (Y) and the production lot size (X).
- (5%) A scatter plot of the data and the residual plot are shown on Output (V). Does a linear relation appear adequate here?
 - (5%) Would a transformation on X or Y be more appropriate here? Why? What transformation would you use?
 - (10%) Find the expected value of the ordered residuals under normality for $e_1 = -0.583$. (See Output (VI) and (VII), Rank $k = 43$) (Let Z_a be the $a \times 100$ percentile of the standard normal distribution. You need to calculate MSE and a first.)

Output from R

Output (I):

Brown-Forsythe Test (alpha = 0.05)

data : residuals and group

statistic : 0.03721278
num df : 1
denom df : 81.40787
p.value : 0.8475129

Output (II):

Call:

lm(formula = Y ~ X)

Residuals:

Min	1Q	Median	3Q	Max
-5.1500	-2.2188	0.1625	2.6875	5.5750

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	168.60000	2.65702	-----	< 2e-16 ***
X	2.03438	0.09039	22.51	2.16e-12 ***

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

Residual standard error: 3.234 on 14 degrees of freedom

Multiple R-squared: ----, Adjusted R-squared: ----

F-statistic: ---- on 1 and -- DF, p-value: 2.159e-12

Output (III):

Analysis of Variance Table

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X	1	5297.5	-----	-----	2.159e-12 ***
Residuals	--	-----	-----		

Output (IV):

> anova(Reduced.model, Full.model)

Analysis of Variance Table

Model 1: Y ~ X

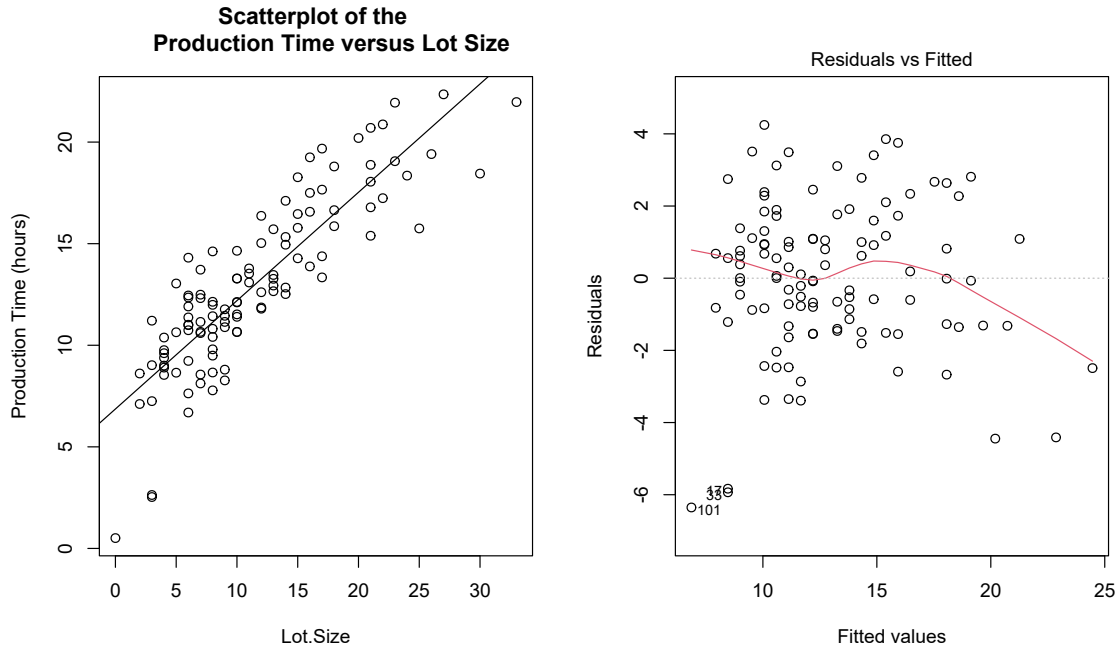
Model 2: Y ~ factor(X)

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	(1)	(2)				
2	12	128.75	(3)	(4)	(5)	0.4622

Note: SSPE = 128.75, df of SSPE = 12, (1) df of SSE, (2) SSE, (3) df of Lack of Fit, (4) SSLF, (5) F test statistic.

Output from R

Output (V):



Output (VI):

Analysis of Variance Table

Response: Production.Time

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lot.Size	1	1391.16	1391.16	310.2	< 2.2e-16 ***
Residuals	109	488.84	4.48		

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

Output (VII):

	Production.Time	Lot.Size	Residual	Rank
1	14.28	15	-0.583	43
2	8.80	9	-2.863	9
3	12.49	7	1.894	91
...				
109	16.37	12	3.107	104
110	11.45	9	-0.213	49
111	15.78	15	0.917	73

注意： 1、考試求公平及公正，請同學務必自律，維護學校與學生之榮譽。

2、考試時不得有交談、窺視、夾帶、抄襲、傳遞、代考或其它作弊等舞弊行為，考畢務必交卷，不得攜卷出場，違者依考場規則議處。