國立政治大學 111 學年度第 1 學期 Quiz(1) 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：10 月 20 日 (四) 10:10-11:50	※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3頁, 印刷份數: 60 份	Calculator	Book Notes	Dictionary	Cell phone Laptop	
備註：注意事項要看!! (§1~§2)	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 120.

1. (15%) Explain the following (the symbols in the formula need to be explained if you write it down):
   (a) (5%) What is the "simple linear regression model with distribution of error terms unspecified"? Which assumptions are made for error terms?
   (b) (5%) What is the meaning of the regression parameters in a simple linear regression model?
   (c) (5%) In a normal error regression model, why do we need to assume that the error terms are normally distributed?

2. (15%) Assume the simple linear regression model is the normal error regression model, find the estimation of parameters using method of maximum likelihood.

3. (20%) Let b_1 be the point estimator of β_1 in the normal error regression model.
   (a) (5%) Show that b_1 is a linear combination of the observation Y_i.
   (b) (5%) Show that the sampling distribution of b_1 is normal.
   (c) (5%) Derive the mean and the variance of b_1.
   (d) (5%) Derive the $(1-\alpha)\%$ confidence interval for β_1.

4. (20%) One would like to use the general linear test approach to test whether or not $\beta_1 = 0$ for the simple linear regression.
   (a) (5%) What is the full model? What is the error sum of squares for this model?
   (b) (5%) What is the reduced model? What is the error sum of squares for this model?
   (c) (5%) What is test statistic?
   (d) (5%) Show that the above test statistic is identical to the analysis of variance test statistic.

考試日期：10 月 20 日 (四) 10:10-11:50	※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3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~§2)	O	×	×	×	

5. (15%) There are three common misunderstandings about the coefficient of determination (R^2) described in the textbook. Please correct these misunderstandings by giving the explanations or counterexamples (反例).

   (a) (5%) Misunderstanding 1 : A high R^2 indicates that useful predictions can be made.

   (b) (5%) Misunderstanding 2: A high R^2 indicates that the estimated regression line is a good fit.

   (c) (5%) Misunderstanding 3: A R^2 near zero indicates that X and Y are not related.

6. (20%) **Airfreight breakage**. A substance used in biological and medical research is shipped by airfreight to users in cartons of 1,000 ampules. The data below, involving 10 shipments, were collected on the number of times the carton was transferred from one aircraft to another over the shipment route (X) and the number of ampules found to be broken upon arrival (Y). Assume that first-order regression model (1.1) in textbook is appropriate.

i	1	2	3	4	5	6	7	8	9	10	
X_i	1	0	2	0	3	1	0	1	2	0	$\sum X_i = 10, \sum X_i^2 = 20, \sum Y_i = 142$
Y_i	16	9	17	12	22	13	8	15	19	11	$\sum Y_i^2 = 2194, \sum X_i Y_i = 182$

   (a) (5%) Obtain the estimated regression function.

   (b) (5%) Plot the estimated regression function and the data (請用手畫).

   (c) (5%) Obtain an estimate of the expected number of broken ampules when $X = 1$ transfer is made.

   (d) (5%) How to interpret regression coefficient of X for this example?

7. (15%) Refer to **Airfreight breakage Problem**

   (a) (5%) Set up the ANOVA table.

   (b) (5%) Conduct an F test to decide whether or not there is a linear association between the number of times a carton is transferred and the number of broken ampules; control the a risk at 0.05. State the alternatives, decision rule, and conclusion.

   (c) (5%) Calculate R^2 and r. How to interpret R^2 for this example?

Output from R

```
> Airfreight.lm <- lm(y ~ x)
> summary(Airfreight.lm)

Call:
lm(formula = y ~ x)

Residuals:
   Min     1Q Median     3Q    Max
  -2.2   -1.2    0.3    0.8    1.8

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.2000     0.6633  15.377 3.18e-07 ***
x               XXXX       XXXX   XXXX  2.75e-05 ***
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1  ' '  1

Residual standard error: 1.483 on 8 degrees of freedom
Multiple R-squared:  XXX,    Adjusted R-squared:  0.8885
F-statistic: XXX on 1 and X DF,  p-value: 2.749e-05

> anova(Airfreight.lm)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
x          1  160.0   160.0     XXX  2.749e-05 ***
Residuals  X   XXX      XXX
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1  ' '  1
```

Notes

- $s^2(b_1) = \dfrac{MSE}{\sum(X_i - \bar{X})^2}$.
- $\sum(X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2$.
- $\sum(X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - n\bar{X}\bar{Y}$.
- `qf(0.95, 1, 8)= 5.3177, qf(0.05, 8, 1)= 0.1881.`
- `qt(0.95, 8) = 1.8595, qt(0.05, 8) = -1.8595, qt(0.95, 9) = 1.8331, qt(0.05, 9) = -1.8331`