國立政治大學 110 學年度第 1 學期 小考 (3) 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：12 月 23 日 (四) 11:10-12:00	※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: 30 份	Calculator	Book Notes	Dictionary	Cell phone Laptop	
備註：注意事項要看!! (範圍: §6∼7)	O	×	×	×	

Note: (1) Fill in your name and student ID。(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. (7) Use $\underset{\sim}{\beta}$ or $\underset{\sim}{X}$ to represent a vector $\boldsymbol{\beta}$ or a matrix $\mathbf{X}$.

1. (10%) Consider the multiple linear regression model for a given data $\{Y_i, X_{i1}, X_{i2}, \cdots, X_{ip}\}_{i=1}^{n}$, someone would like to perform a F-test for Lack of Fit for this model. Please state (a) the general (multiple) linear regression model for this data; (b) the mean response function; (c) the test hypothesis (H_0, H_a); (d) the test statistic; and (e) the decision rule.

2. (5%) What is the extra sums of squares and what does it measure?

3. (10%) When the regression model contains three X variables, a variety of decompositions of $SSR(X_1, X_2, X_3)$ into extra sums of squares can be obtained. Please give three examples.

4. (10%) Consider the first-order regression model with three predictor variables, someone would like to use extra sums of squares in testing whether both $\beta_2 X_2$ and $\beta_3 X_3$ can be dropped from the full model. Please state (a) the test hypothesis (H_0, H_a); (b) the full model and the reduced model; (c) the general linear test statistics; and (d) the decision rule.

5. (5%) What is the definition of the coefficient of partial determination (take $R^2_{Y1|2}$ as an example and express it in terms of the extra sum of squares) and what does it measure?

6. (20%) Consider the multiple regression analysis, what is the multicollinearity problem? What are the effects of multicollinearity when conduct the multiple regression analysis? (Hint: you cannot just say that the multicollinearity has effects on the regression coefficients, for example, you need to describe what does it result in on the regression coefficients.)

考試日期：12 月 23 日 (四) 11:10-12:00	※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: 30 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!! (範圍: §6∼7)	O	×	×	×	

7. **Commercial properties**. A commercial real estate company evaluates vacancy rates, square footage, rental rates, and operating expenses for commercial properties in a large metropolitan area in order to provide clients with quantitative information upon which to make rental decisions. The data below are taken from 81 suburban commercial properties that are the newest, best located, most attractive, and expensive for five specific geographic areas. Shown here are the age (X_1), operating expenses and taxes (X_2), vacancy rates (X_3), total square footage (X_4), and rental rates (Y).

   (a) (10%) Obtain the analysis of variance table that decomposes the regression sum of squares into extra sums of squares associated with X_4; with X_1 given X_4; with X_2 , given X_1, and X_4; and with X_3, given X_1, X_2 and X_4. (Hint: $SSR(X_4), SSR(X_1|X_4), \cdots$)

   (b) (10%) Test whether X_3 can be dropped from the regression model given that X_1, X_2 and X_4 are retained. Use the F^* test statistic and level of significance 0.01. State the alternatives, decision rule, and conclusion. (Hint: $F(0.99; 1, 76) = 6.980578; F(0.99; 2, 76) = 4.89584; F(0.99; 3, 76) = 4.050282; F(0.99; 1, 75) = 6.985359; F(0.99; 2, 75) = 4.899877; F(0.99; 3, 75) = 4.054022$)

   (c) (10%) Test whether both X_2 and X_3 can be dropped from the regression model given that X_1 and X_4 are retained; use $\alpha = 0.01$. State the alternatives, and decision rule. (Hint: specify $df1$ and $df2$ in $F(0.99; df1, df2)$ as a critical value.)

   (d) (10%) Using the given R report sheet below, calculate the coefficient of partial determination $R^2_{Y2|14}$ and interpret. (Hint: Answer "There was not sufficient information provided." if the information provided was not sufficient to calculate $R^2_{Y2|14}$.)

```
> summary(m4)
lm(formula = Y ~ X4)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.378e+01  2.903e-01  47.482  < 2e-16 ***
X4          8.437e-06  1.498e-06   5.632 2.63e-07 ***
---
Residual standard error: 1.462 on 79 degrees of freedom
Multiple R-squared:  0.2865,    Adjusted R-squared:  0.2775
F-statistic: 31.72 on 1 and 79 DF,  p-value: 2.628e-07

> summary(m14)
lm(formula = Y ~ X1 + X4)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.436e+01  2.771e-01  51.831  < 2e-16 ***
X1          -1.145e-01  2.242e-02  -5.105 2.27e-06 ***
X4           1.045e-05  1.363e-06   7.663 4.23e-11 ***
---
Residual standard error: 1.274 on 78 degrees of freedom
Multiple R-squared:  0.4652,    Adjusted R-squared:  0.4515
F-statistic: 33.93 on 2 and 78 DF,  p-value: 2.506e-11

> summary(m124)
lm(formula = Y ~ X1 + X2 + X4)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.237e+01  4.928e-01  25.100  < 2e-16 ***
X1          -1.442e-01  2.092e-02  -6.891 1.33e-09 ***
X2           2.672e-01  5.729e-02   4.663 1.29e-05 ***
X4           8.178e-06  1.305e-06   6.265 1.97e-08 ***
---
Residual standard error: 1.132 on 77 degrees of freedom
Multiple R-squared:  0.583,     Adjusted R-squared:  0.5667
F-statistic: 35.88 on 3 and 77 DF,  p-value: 1.295e-14

> summary(m1234)
lm(formula = Y ~ X1 + X2 + X3 + X4)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.220e+01  5.780e-01  21.110  < 2e-16 ***
X1          -1.420e-01  2.134e-02  -6.655 3.89e-09 ***
X2           2.820e-01  6.317e-02   4.464 2.75e-05 ***
X3           6.193e-01  1.087e+00   0.570     0.57
X4           7.924e-06  1.385e-06   5.722 1.98e-07 ***
---
Residual standard error: 1.137 on 76 degrees of freedom
Multiple R-squared:  0.5847,    Adjusted R-squared:  0.5629
F-statistic: 26.76 on 4 and 76 DF,  p-value: 7.272e-14
```

```
> anova(m4)
Analysis of Variance Table
Response: Y
          Df  Sum Sq Mean Sq F value    Pr(>F)
X4         1  67.775  67.775  31.723 2.628e-07 ***
Residuals 79 168.782   2.136
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1  ' '  1

> anova(m124)
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
X1         1 14.819  14.819  11.566  0.001067 **
X2         1 72.802  72.802  56.825 7.841e-11 ***
X4         1 50.287  50.287  39.251 1.973e-08 ***
Residuals 77 98.650   1.281
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1  ' '  1

> anova(m14)
Analysis of Variance Table
Response: Y
          Df  Sum Sq Mean Sq F value    Pr(>F)
X1         1  14.819  14.819  9.1365  0.003389 **
X4         1  95.231  95.231 58.7160 4.225e-11 ***
Residuals 78 126.508   1.622
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1  ' '  1

> anova(m1234)
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
X1         1 14.819  14.819 11.4649  0.001125 **
X2         1 72.802  72.802 56.3262 9.699e-11 ***
X3         1  8.381   8.381  6.4846  0.012904 *
X4         1 42.325  42.325 32.7464 1.976e-07 ***
Residuals 76 98.231   1.293
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1  ' '  1
```