| تمهيدى ماجستير احصاء | الفصل الدراسى الثانى 1/6/2015 | جامعة القاهرة |
|---|---|---|
| الزمن 3 ساعات | مادة: التقديرات المعدلة فى تحليل الانحدار STATA 619 | معهد الدراسات و البحوث الاحصائية |

**Q1) (15 marks):** for the correlation matrix:

$$\begin{array}{c} \\ y \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} \begin{bmatrix} 1 & 0.346 & 0.593 & 0.666 & 0.726 \\ & 1 & 0.09 & -0.15 & 0.502 \\ & & 1 & -0.024 & 0.369 \\ & & & 1 & 0.416 \\ & & & & 1 \end{bmatrix}$$

a) Write the algorithm for partial correlation
b) In a partial correlation procedure which independent variable enters first in the equation and why?
c) Based on your answer in (b) which independent variable should enter second and why?

**Q2) (20 marks):** Given the following simple regression equation and the ANOVA table : $\hat{Y} = 2.58\text{-}0.324\ x$ ,

| x | 9 | 9 | 9 | 7 | 7 | 7 | 5 | 5 | 5 | 3 | 3 | 3 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| y | .07 | .09 | .08 | .16 | .17 | .21 | .49 | .58 | .53 | 1.22 | 1.15 | 1.07 | 2.84 | 2.57 | 3.1 |

ANOVA Table

| Source | DF | SS | MS | F |
|---|---|---|---|---|
| Regression | 1 | 12.597 | 12.597 | 55.99 |
| Error | 13 | 2.925 | 0.225 | |
| Total | 14 | 15.522 | | |

i. Graphically analyze the residuals
ii. Test whether this regression equation is a good fit for the data

**Q3): (15 marks):** a) Define multicollinearity problems and its consequences. What are the remedial measurements for multicollinearity
b) Define Ridge trace and show how it can be used to select the ridge constant?
c) Define $R_k^2$ and show how it can be used to compute $VIF_k$

**Q4) (10 marks):** From this output:

1- Regression Analysis: y versus x1

The regression equation is: y = 20.2 + 3.43 x1

S = 4.11052 R-Sq = 97.8% R-Sq(adj) = 97.2%

ANOVA Table

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Regression | 1 | 3004.4 | 3004.4 | 177.81 | 0.000 |
| Error | 4 | 67.6 | 16.9 | | |
| Total | 5 | 3072.0 | | | |

2- Regression Analysis: y versus x2

The regression equation is: y = 91.2 - 1.34 x2

S = 6.29758 R-Sq = 94.8% R-Sq(adj) = 93.5%

ANOVA Table

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Regression | 1 | 2913.4 | 2913.4 | 73.46 | 0.001 |
| Error | 4 | 158.6 | 39.7 | | |
| Total | 5 | 3072.0 | | | |

3- Regression Analysis: y versus x1, x2

The regression equation is: y = 33.9 + 2.78 x1 - 0.264 x2

S = 4.54876 R-Sq = 98.0% R-Sq(adj) = 96.6%

ANOVA Table

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Regression | 2 | 3009.9 | 1505.0 | 72.73 | 0.003 |
| Error | 3 | 62.1 | 20.7 | | |
| Total | 5 | 3072.0 | | | |

4- Correlations: x1, x2

Pearson correlation of x1 and x2 = -0.975

P-Value = 0.001

i. Identify presence of multicollinearity among independent variables for this data using several informal methods

ii. Which model do you prefer and why?

**Q5) (10 marks):** a) what are the criteria used to choose the best model using all possible regression models? Explain them.

b) Write a strategic plan for building a regression model and discuss three ways for model validation

مع اطيب التمنيات بالنجاح