Tutorial 2

Inferential Statistics, Statistical Modelling & Survey Methods

(BS2506)

Pairach Piboonrungroj
(Champ)
pairach@piboonrungroj.com

1.(a) Petrol price

Average Price	1.00	0.99	0.92	1.11	1.40	1.47
Year	2000	2001	2002	2003	2004	2005
Year (x)	-5	-3	-1	1	3	5

Code the year variable into integers that sum to zero and estimate the average price for the year 2006

Note: when the values of the independent variables sum to zero, the formulas to estimate the intercept the intercept and slope coefficients become much simpler, as given below

1 (a)

$$\hat{\beta}_0 = \bar{y} = \frac{6.89}{6} = 1.148$$

$$\hat{\beta}_1 = \frac{\sum xy}{\sum x^2} = \frac{3.77}{70} = 0.054$$

$$\hat{y} = 1.148 + 0.054x$$

Predict a 2006 price using x = 7

$$\hat{y}_{2006} = 1.148 + 0.054(7) = 1.526$$

2(a)

Y	X_0	X_1	X_2
10.5	1	1	0
-11.3	1	1	0
-15.42	1	0	1
11.63	1	0	1
-9.19	1	1	0
5.32	1	1	0
7.50	1	0	1

2(a)

Y	X_0	X_1	X_2	$X_1 + X_2$
10.5	1	1	0	1
-11.3	1	1	0	1
-15.42	1	0	1	1
11.63	1	0	1	1
-9.19	1	1	0	1
5.32	1	1	0	1
7.50	1	0	1	1

2(a)

$$x_0 = x_1 + x_2$$

So there is **EXACT MULTICOLINEARITY**.

It is not possible to include all three repressors
and estimate the parameters.

3 (a)

Variable	B	SE B
(Constant)	55.0	47.47
X1	0.925	0.217
X2	-5.205	1.229
X3	17.446	12.78

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

$$\hat{y} = 55 + 0.925 x_1 - 5.205 x_2 + 17.446 x_3$$

A house size increase by 1 square meter (holding other variables constant), price are expected to increase by £925.

3.b.

At $\alpha = 0.05$

$$t_{\alpha/2, df} = t_{0.025, 15-4=11} = 2.201$$

Testing for the significance of the house size

$$t_{\beta_i} = \frac{\beta_i}{SE_{\beta_i}}$$

$$t_{house\_size} = \frac{\beta_{house\_size}}{SE_{\beta_{house\_size}}} = \frac{0.925}{0.217} = 4.263$$

Thus, house size is significant

3(c)

Testing for the significance of the house age

$$t_{\beta_i} = \frac{\beta_i}{SE_{\beta_i}}$$

$$t_{house\_age_i} = \frac{\beta_{house\_age}}{SE_{\beta_{house\_age}}} = \frac{-5.205}{1.229} = -4.235$$

$$t_{\alpha/2, df} = t_{0.025, 15-4=11} = 2.201$$

Thus, house age is significant

3(d)

Testing for the significance of the number of room

$$t_{\beta_i} = \frac{\beta_i}{SE_{\beta_i}}$$

$$t_{number\_of\_room} = \frac{\beta_{number\_of\_room}}{SE_{\beta_{number\_of\_rooom}}} = \frac{17.446}{12.78} = 1.365$$

$$t_{\alpha/2, df} = t_{0.025, 15-4=11} = 2.201$$

Thus, number of rooms is NOT significant

3(e)

Estimate the price of a 20 years old house which has 4 bedrooms with 200 square meters.
$x_1 = 200$, $x_2 = 20$, $x_3 = 4$

$$\hat{y} = 55 + 0.925x_1 - 5.205x_2 + 17.446x_3$$

$$\hat{y} = 55 + 0.925(200) - 5.205(20) + 17.446(4)$$

$$\hat{y} = 205.684 thousand\_pounds$$

$$\hat{y} = 205,684\, pounds$$

3(f)

Testing for the significance of the number of room

$$\hat{y} = 35.818 + 41.826x_3$$

$$t_{\beta_i} = \frac{\beta_i}{SE_{\beta_i}}$$

$$t_{number\_of\_room} = \frac{\beta_{number\_of\_room}}{SE_{\beta_{number\_of\_rooom}}} = \frac{41.826}{18.3} = 2.285$$

At $\alpha = 0.05$ $t_{\alpha/2, df} = t_{0.025, 15-2=13} = 2.1604$

Thus, number of rooms is significant
(when we exclude house size that has a high correlation with).

4 (a)

	Sum of Squares	DF	Mean squares (Sum squares /DF)	F (MS Regression / MS Residual)
Regression	253,388	3		
Residual	59,234	11		
Total				

4 (a)

	Sum of Squares	DF	Mean squares (Sum squares /DF)	F (MS Regression / MS Residual)
Regression	253,388	3	84,462.67	**15.69**
Residual	59,234	11	5,384.91	
Total	312,622			

At $\alpha = 0.01$

$$F_{3,11,0.01} = 6.217 < 15.69$$

The model is significant

4(b)

Multiple coefficient of determination (R2)

$$= \frac{\text{Sum Squares Total}}{\text{Sum Squares Total}} - \frac{\text{Sum Square Error}}{\text{Sum Squares Total}}$$

= 1 – (59,234 / 312,622)
= 1 – 0.189
= 0.811

4(c)

Adjusted multiple coefficient of determination (Adjusted R2) – adjusted by DF

= 1 – (**Mean** Square Error / **Mean** Squares Total)
= 1 – (5384.91 /22,330.14)
= 1 – 0.241
= 0.759

5(a)

no	Leverage (y)	Country
1	0.67	M
2	0.62	M
3	0.78	M
4	0.54	M
5	0.65	M
6	0.29	I
7	0.32	I
8	0.29	I
9	0.35	I
10	0.37	I

5(a)

no	Leverage (y)	Country	Country (x)
1	0.67	M	1
2	0.62	M	1
3	0.78	M	1
4	0.54	M	1
5	0.65	M	1
6	0.29	I	0
7	0.32	I	0
8	0.29	I	0
9	0.35	I	0
10	0.37	I	0

5(a)

no	Leverage (y)	Country	Country (x)	xy	x^2	y^2
1	0.67	M	1	0.67	1	0.4489
2	0.62	M	1	0.62	1	0.3844
3	0.78	M	1	0.78	1	0.6084
4	0.54	M	1	0.54	1	0.2916
5	0.65	M	1	0.65	1	0.4225
6	0.29	I	0	0	0	0.0841
7	0.32	I	0	0	0	0.1024
8	0.29	I	0	0	0	0.0841
9	0.35	I	0	0	0	0.1225
10	0.37	I	0	0	0	0.1369

5(a)

no	Leverage (y)	Country	Country (x)	xy	x^2	y^2
1	0.67	M	1	0.67	1	0.4489
2	0.62	M	1	0.62	1	0.3844
3	0.78	M	1	0.78	1	0.6084
4	0.54	M	1	0.54	1	0.2916
5	0.65	M	1	0.65	1	0.4225
6	0.29	I	0	0	0	0.0841
7	0.32	I	0	0	0	0.1024
8	0.29	I	0	0	0	0.0841
9	0.35	I	0	0	0	0.1225
10	0.37	I	0	0	0	0.1369
Sum	**4.88**		**5**	**3.26**	**5**	**2.686**
Average	**0.488**		**0.500**	**0.326**	**0.500**	**0.269**

5(a)

$$\bar{x} = 0.5, \bar{y} = 0.488$$

$$\hat{\beta}_1 = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$$

$$\hat{\beta}_1 = \frac{3.26 - 10(0.5)(0.488)}{5 - 10(0.5)^2} = 0.328$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

$$\hat{\beta}_0 = 0.488 - 0.328(0.5) = 0.324$$

$$\hat{y} = 0.324 + 0.328x$$

5 (b)

$$\hat{y} = 0.324 + 0.328x$$

Assume x = 0 for India then

$$\hat{y} = 0.324$$

Thus x = 1 for Malaysia then

$$\hat{y} = 0.652$$

5 (b)

$$\hat{y} = 0.652 + (0.652 - 0.324)x$$

$$\hat{y} = 0.652 - (0.328)x$$

$$\hat{y} = \beta_0 - \beta_1 x$$

Assume x = 1 for India then

$$\hat{y} = 0.324$$

Thus x = 0 for Malaysia then

$$\hat{y} = 0.652$$

5(c)

Testing if slope coefficient is significant

H_0: $\beta_1 = 0$

H_1: $\beta_1 \neq 0$

Statistics: t-statistics $t_{\beta_i} = \dfrac{\beta_i}{SE_{\beta_i}}$

SE is still unknown

5(c)

To calculate SE

$$SE(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1}$$

$$\sqrt{Var(\hat{\beta}_1)} = \frac{S^2}{\sum (x_i - \bar{x})^2}$$

$$S^2 = \frac{\sum y^2 - \hat{\beta}_0 \sum y - \hat{\beta}_1 \sum xy}{n-2}$$

$$S^2 = \frac{2.685 - 0.324 \times 4.88 - 0.328(3.26)}{8}$$

$$S^2 = 0.004$$

And

$$\sum (x - \bar{x})^2 = \sum x^2 - n\bar{x}^2$$

$$= 5 - 10(0.5)^2 = 5 - 2.5 = 2.5$$

Then

$$\sqrt{Var(\hat{\beta}_1)} = \frac{S^2}{\sum (x_i - \bar{x})^2}$$

$$\sqrt{Var(\hat{\beta}_1)} = \frac{0.004}{2.5} = 0.02$$

Then

$$SE(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)}$$

$$SE(\hat{\beta}_1) = \sqrt{0.002} = 0.040$$

So now we can calculate t-statistics

5(c)

So now we can calculate t-statistics

$$t_{\beta_i} = \frac{\beta_i}{SE_{\beta_i}}$$

$$t_{\beta_i} = \frac{0.328}{0.040} = 8.2$$

At $\alpha = 0.05$

$$t_{\alpha/2, df} = t_{0.025, 10-2=8} = 2.3060$$

Thus slope coefficient is significant at 5% significant

Thank you

To download this slides visit;

www.pairach.com/teaching