

11.10: Bla bla bla data:

(se motsvarande pdf)

1: a) se pdf (ta medelvärde)

b) Vi skall skatta β_0 och β_1 . Enligt föreläsning skall vi använda

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \quad \text{och} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

där $S_{xx} = \sum_{k=1}^n (x_k - \bar{x})^2$ och $S_{xy} = \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})$.

Här är (x_k, y_k) våra datapunkter.

Data ger att $S_{xx} \approx 188.8$ $S_{xy} \approx 99061$

så att $\hat{\beta}_1 \approx 524.7$ och $\hat{\beta}_0 \approx -4.62$

Vår skattade linje blir alltså

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = -4.62 + 524.7 \cdot x$$

c) Vad menas? Vi skall använda

regressionslinjen och sätta in $x = 3.450$

för att få fram $\hat{y}(x=3.45)$.

Vi får

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 3.45 = 1805.6 //$$

d) Vad blir förklaringsgraden?

L:

$$\text{förklaringsgraden} = R^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}} \approx 0.99995$$

där $S_{yy} = \sum_{k=1}^n (y_k - \bar{y})^2 \approx 51981000$ (störningen minimal)

e)

Bestäm ett 99.9 % K.I. för β_1 .

L: Vi använder referensvariabeln

$$\tilde{R} = \frac{\hat{\beta}_1 - \beta_1}{S_r \sqrt{S_{xx}}} \sim t(n-2) = t(22)$$

OBS: Notationen för ref.-var och förklaringsgraden krökar. Använd kontext!

$$1 - \alpha = \mathbb{P}(-t_{\alpha/2}^{(22)} \leq \frac{\hat{\beta}_1 - \beta_1}{S_r \sqrt{S_{xx}}} \leq t_{\alpha/2}^{(22)})$$

$$= \mathbb{P}\left(\hat{\beta}_1 - t_{\alpha/2}^{(22)} \frac{S_r}{\sqrt{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2}^{(22)} \frac{S_r}{\sqrt{S_{xx}}}\right)$$

där $S_r^2 = \frac{1}{n-2} \sum_{k=1}^n (y_k - \hat{y}_k)^2 \approx 396.3$ (stort?)

Enligt tabell (sid 699) så är

$$t_{0.0005}^{(22)} \approx 3.79$$

så att ett 99.9 % numeriskt K.I. för β_1 blir

$$I_{\beta_1} = \hat{\beta}_1 \pm t_{0.0005}^{(22)} \cdot \frac{S_r}{\sqrt{S_{xx}}} \approx 524.7 \pm 3.79 \cdot \frac{\sqrt{396.3}}{\sqrt{188.8}}$$

$$\approx [519, 530]. //$$

11,42: Beträkta datamängden:

$x: 80 \quad 75 \quad 75 \quad \dots$

$y: 5.0 \quad 4.7 \quad 4.5 \quad \dots$

a) Ignorera att $y = \beta_0 + \beta_1 x$ är korrekt och skatta β_0 och β_1 ändå.

1: Som innan

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \approx 0.00043 \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \approx 4.85$$

och $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x //$

b) + c) Hitta $\hat{y}_u$ och residualerna.

1: $\hat{y}_u = \hat{\beta}_0 + \hat{\beta}_1 x_u$ så att

$$e_u = y_u - \hat{y}_u = y_u - (\hat{\beta}_0 + \hat{\beta}_1 x_u) //$$

d) Plotta resultatet

e) Vilken typ av samband tror du gäller?

1: Kanske $y = \beta_0 + \beta_1 x^2$ eller $y = \beta_0 + \beta_1 (x-a)^2$
eller $y = \beta_0 + \beta_1 (x-a)^4$ etc. där $a \approx 95$.

f) skip

g) Vad blir förklaringsgraden?

1: $R^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}} \approx 0.000026$ (extremt litet) //

h) Testa $H_0: \beta_1 = 0$

$$H_1: \beta_1 \neq 0$$

på signifikansnivån 0.1

L: Vi beräknar p-värdet. Ett tvåsidigt test gör att

$$p\text{-värde} = 2 \mathbb{P}(T(E) \geq |T(E)|).$$

Här väljer vi test-statistikan

$$T(E) = \frac{\hat{\beta}_1(E)}{S_r(E)/\sqrt{S_{xx}}} \sim t(n-2) = t(16)$$

så att

$$T(E) = \frac{\hat{\beta}_1(E)}{S_r(E)/\sqrt{S_{xx}}} \approx \frac{0.00043}{\sqrt{7.4}/\sqrt{6257}} \approx 0.0125$$

där

$$S_r^2(E) = \frac{1}{n-2} \sum_{k=1}^n (y_k - \hat{y}_k)^2 \approx 7.4.$$

$$p\text{-värde} = 2 \mathbb{P}(T(E) \geq 0.0125)$$

$$= 2 (1 - \mathbb{P}(T(E) \leq 0.0125))$$

$$\approx 2 (1 - 0.5049) \approx 0.99$$

↑
Matlab

Extremt högt! //