ERHVERVSØKONOMI/MATEMATIK-STUDIET

Besvarelse - Statistik 2. år - 14. juni 2000

Opgave 4

1) Lad $x_i \quad (i = 1, \ldots, 11)$ betegne "liter dieselolie påfyldt" og lad $y_i \quad (i = 1, \ldots, 11)$ betegne "km kørt". Den foreslåede model går så ud på at $y_1, \ldots, y_{11}$ er observationer af uafhængige normalfordelte variable $Y_1, \ldots, Y_{11}$ med samme varians σ^2 og middelværdi

$$EY_i = \alpha + \beta x_i$$

(idet x'erne opfattes som kendte tal).

Tegningen af y'er mod x'er (som bør udføres, selvom den ikke er gengivet her) indeholder intet som kan forkaste denne model (ingen "krumning", konstant "bredde").

2) Vi får

$$\begin{aligned}
SSD_x &= SS_x - \frac{S_x^2}{11} = 24103 - \frac{509^2}{11} = 550.18 \\
SSD_y &= \ldots = 3423903 - \frac{6053^2}{11} = 93102.18 \\
SPD_{xy} &= 286800 - \frac{509 \cdot 6053}{11} = 6711.18
\end{aligned}$$

Heraf fås

$$\begin{aligned}
\hat{\beta} &= \frac{6711.18}{550.18} = \underline{12.198} \text{ (usikkerhedsangivelse følger)} \\
\hat{\alpha} &= \frac{1}{11}(6053 - 509 \cdot 12.198) = \underline{-14.16}
\end{aligned}$$

$$\begin{aligned} SSD_{res} &= 93102.18 - 12.198^2 \cdot 550.18 = 11240.23 \\ \hat{\sigma}^2 &= \frac{1}{11-2} \cdot 11240.23 = \underline{1248.9} \end{aligned}$$

Til angivelse af usikkerhed på $\hat{\beta}$ kan benyttes at

$$\begin{aligned} \sqrt{var(\hat{\beta})} &= \sqrt{\frac{\sigma^2}{SSD_x}} \approx \\ \sqrt{\frac{\hat{\sigma}^2}{SSD_x}} &= \sqrt{\frac{1248.9}{550.18}} = 1.507 \end{aligned}$$

så β-estimatet skal forsynes med usikkerhedsangivelse således:

$$\begin{aligned} \beta &= 12.198 \pm 2.262 \cdot 1.507 \\ &= \underline{12.198 \pm 3.409} \end{aligned}$$

(her er 2.262 97.5% fraktilen i en t-fordeling med 9 frihedsgrader).

3) Hypotesen $\alpha = 0$ kan testes ved et t-test

$$t = \frac{-14.16}{\sqrt{\left(\frac{1}{11} + \frac{(509/11)^2}{550.18}\right) \cdot 1248.9}} = -0.201,$$

som er klart insignifikant i en hvilken som helst t-fordeling. Vi godkender altså hypotesen $\alpha = 0$.

4) I modellen med $\alpha = 0$ fås, efter passende udregninger som ikke gengives (de er gennemført i forbindelse med en opgave, og kan let udledes af den generelle teori)

$$\hat{\beta} = \frac{SP_{xy}}{SS_x} = \frac{286800}{24103} = \underline{11.899}$$

Variansen på $\hat{\beta}$ bliver

$$
\begin{aligned}
var(\hat{\beta}) &= \frac{1}{SS_x}\sigma^2 \approx \frac{1}{SS_x}\hat{\sigma}^2 = \frac{SSD_{res}/10}{SS_x} = \\
\frac{(SS_y - \hat{\beta}^2 SS_x)/10}{SS_x} &= \frac{(3423903 - 11.899^2 \cdot 24103)/10}{24103} = 0.04668, \\
\sqrt{var(\hat{\beta})} &\approx 0.216
\end{aligned}
$$

Ved anvendelse af sædvanlige t-fordelingsbaserede konfidensgrænser fås således 95%-sikkerhedsintervallet ($f = 10$)

$$
\begin{aligned}
\beta &= 11.899 \pm 2.228 \cdot 0.216, \quad \text{eller} \\
\beta &\in [11.418, 12.380]
\end{aligned}
$$

Betydningen af dette interval er at det med sandsynlighed 0.95 indeholder den sande værdi af β. Vi kan altså med rimelig sikkerhed sige at dieselolieforbruget svarer til mellem 11.4 og 12.4 km pr. liter.

13-bemærkningen:

Man bemærker at vi under den naturlige hypotese $\alpha = 0$ får en betydeligt mindre usikkerhed på β end i den oprindelige model.