

Eksamen juni 2006, Opgave 4

(a)

Vi udregner følgende størrelser:

$$\bar{x} = 1003/195 = 5.1436$$

$$\bar{y} = 18767/195 = 96.241$$

$$\text{SSD}_x = 6301 - 1003^2/195 = 1141.98$$

$$\text{SSD}_y = 2018875 - 18767^2/195 = 212719.67$$

$$\text{SPD}_{xy} = 82516 - 1003 \times 18767/195 = -14013.75$$

$$\text{SSD}_{\text{res}} = 212719.67 - 14013.75^2/1141.98 = 40750.64$$

$$\hat{\beta} = -14013.75/1141.98 = -12.271$$

$$\hat{\alpha} = 96.241 + 12.271 \times 5.1436 = 159.36$$

$$\hat{\sigma}^2 = 40750.60/193 = 211.14$$

95% sikkerhedsgrænser for hældningen: Den estimerede standardafvigelse for $\hat{\beta}$ er

$$\sqrt{\text{vâr}(\hat{\beta})} = \sqrt{211.14/1141.98} = 0.42999$$

og da 97.5% fraktilen i T-fordelingen med 193 frihedsgrader er (tæt på) 1.972, får vi

$$\beta = -12.271 \pm 1.972 \times 0.42999 = -12.271 \pm 0.848$$

svarende til sikkerhedsintervallet $[-13.119, -11.423]$.

(b)

Estimatet for prisen på en 4 år gammel bil er naturligvis (i 1000 kr.)

$$159.36 - 4 \times 12.271 = 110.28.$$

Den estimerede standardafvigelse på dette estimat er

$$\sqrt{\left(\frac{1}{195} + \frac{(4 - 5.1436)^2}{1141.98}\right) 211.14} = 1.151,$$

så vi får 95% sikkerhedsgrænserne

$$110.28 \pm 1.972 \times 1.151 = 110.28 \pm 2.27$$

svarende til sikkerhedsintervallet $[108.01, 112.55]$.

(c)

Vi bemærker, at det nye variansestimater (141.34) er væsentligt mindre end det gamle (211.14), så den nye model forklarer priserne bedre end den gamle.

Oplysningen om den nye varians gør det muligt at rekonstruere residu-alkvadratsummen og dermed F-testet for den tilsvarende modelreduktion. Men da dette er ækvivalent med T-testet for $\gamma = 0$, som næsten kan aflæses direkte af tabellen, udfører vi dette i stedet:

$$T = \frac{-0.194}{0.0197} = -9.848.$$

Denne teststørrelse skal vurderes i T-fordelingen med 192 frihedsgrader, hvor den er ekstremt signifikant (0.1%-fraktilen ligger omkring -3.79 ifølge tabellen).

Antallet af kilometer har således væsentlig indflydelse på prisen, idet den falder med 194 kr. for hver 1000 km der er kørt. Til gengæld har bilens alder mindre indflydelse på prisen i den nye model end i den gamle (med et prisfald på 8340 kr. pr. år, mod 12270 kr. i den gamle model). Der må antages at være en positiv korrelation mellem alderen og km-tallet (multikollinearitet), som bevirker at disse to forklarende variable til dels kan erstatte hinanden. Men ingen af dem kan fuldt ud erstatte den anden. Som det ovenfor udførte T-test viser, kan km-tallet ikke undværes. Og alderen kan endnu mindre undværes, idet testet for $\beta = 0$ fører til en endnu mere ekstrem forkastelse ($T = -8.34/0.533 = -15.65$).