Ett regressionproblem

Talen 12.58 -18.45 17.89 -2.8 -17.6 ..... -3.89 24.61 70.9 -16.84

är de fem först och fem sista i en serie av 250 successiva differenser i Dow Jones index Sept 93- Aug 94.

Talen -25.4 -9.3 -1 22.9 -1.4 .... 2.3 -12 9.5 16.7 -0.6

är motsvarande differenser i det australiska All-ordinaries Index men en dag efteråt. Det är inte orimligt att tänka sig att den australiska förändringen i index styrs av förändringen i Dow Jones. Man ansätter därför modellen

$$y_i = m + bx_i + \varepsilon_i$$

där y_i är den i:te differensen i det australiska indexet och x_i är motsvarande differens (en dag tidigare) i Dow-Jones och där ε_i är N(0,σ) oberoende.

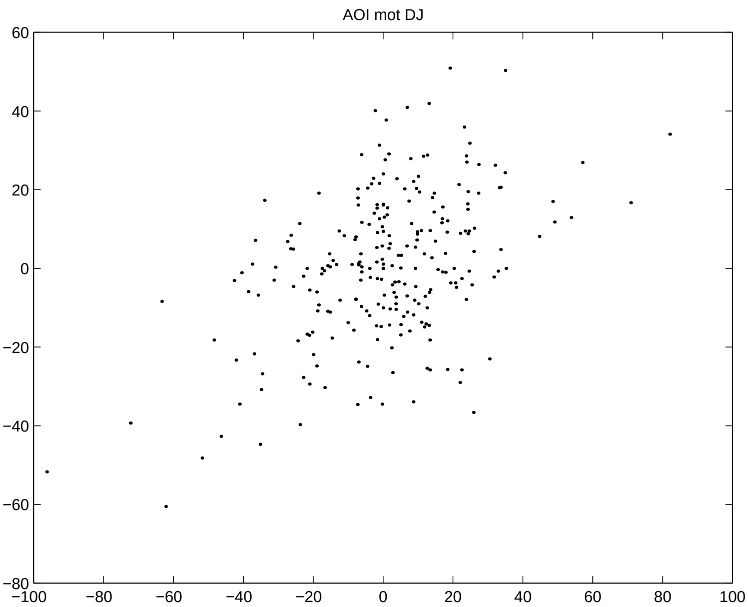
Räknehjälp: $\sum y_i = 138.9, \sum y_i^2 = 85392.55, \sum_i x_i = 208.26, \sum_i x_i^2 = 126111.0826,$ $\sum y_i x_i = 48419.053.$

a) Skatta m och b samt variansen σ^2.

b) Det vore rimligt att tro på proportionellt samband, d.v.s. att $m = 0$. Testa denna hypotes på 5 % signifikansnivå genom att bilda ett lämpligt konfidensintervall.

c) Testa om det överhuvudtaget föreligger en linjär regression, dvs testa hypotesen att $b = 0$, genom att bilda ett lämpligt 95 % konfidensintervall.

d) Antag att $m = 0$, dvs antag att sambandet är proportionellt, $y_i = bx_i + \varepsilon_i$. Skatta b med hjälp av minsta-kvadratmetoden.



LÖSNING

a) Enkel linjär regressionsmodell. Vi sätter $\beta = b$ och $\alpha = m + \beta\bar{x}$ och har på vanligt sätt modellen

$$Y_i = \alpha + \beta(x_i - \bar{x}) + \varepsilon_i$$

och vi skattar $\alpha^* = \bar{y} = 138.9/250 = 0.5556$ och $\beta^* = S_{xy}/S_{xx} = 0.3835$ eftersom $S_{xy} = 48419.053 - 250\bar{x} \cdot \bar{y} = 48419.053 - 250 \cdot \frac{208.26}{250}\frac{138.9}{250} = 48303.34$ och på liknande sätt $S_{xx} = 125937.59$. Vi har vidare att $S_{yy} = 85392.55 - \frac{138.9^2}{250} = 85315.38$ och därmed får vi variansskattningen $\sigma^{*2} = \frac{1}{248}(S_{yy} - S_{xy}^2/S_{xx}) = 269.3$. Skattningen av b är densamma som skattningen av β, medan $m^* = \alpha^* - \beta^*\bar{x} = 0.236$ eftersom $\bar{x} = 0.83304$.

b) Ett 95 % konfidensintervall av $m = \alpha - \beta\bar{x}$ ges av $m^* \pm t_{0.025}(248)\sigma^*\sqrt{1/250 + \bar{x}^2/S_{xx}} = 0.236 \pm 2.044$.
(i regressionen sätter vi $x = 0$). Hypotesen $m = 0$ kan inte förkastas.

c) Ett konfidensintervall för $b = \beta$ ges av $\beta^* \pm t_{0.025}(n-2)\sigma^*/\sqrt{S_{xx}} = 0.3835 \pm 1.96\sqrt{\frac{269.3}{125937.59}} = 0.3835 \pm 0.0906$
Intervallet innehåller inte talet 0. Hypotesen att $\beta = 0$ förkastas, det föreligger en signifikant regression.

d) Enligt minsta-kvadratmetoden skall $Q = \sum_i (y_i - bx_i)^2$ minimeras. Men

$$\frac{dQ}{db} = -2\sum_i (y_i - bx_i)x_i = 0$$

för $b^* = \sum_i x_i y_i / \sum_i x_i^2 = 48419.053/126111.0826 = 0.3839$

Figuren visar data med inlagd skattad regressionslinje.