Eksamen juni 2005, Opgave 4

(a)

Med betegnelser som i forelæsningsnoternes kapitel 8 får vi

$\bar{x} = 47.0/12 = 3.9167$
$\bar{y} = 12.748/12 = 1.0623$
$\text{SSD}_x = 197.44 - 47.0^2/12 = 13.3567$
$\text{SSD}_y = 16.3600 - 12.748^2/12 = 2.8174$
$\text{SPD}_{xy} = 55.6984 - 47.0 \times 12.748/12 = 5.7687$
$\text{SSD}_{\text{res}} = 2.8174 - \frac{5.7687^2}{13.3567} = 0.32589$

Vi får således parameterestimaterne

$\hat{\beta} = \frac{5.7687}{13.3567} = 0.4319$
$\hat{\alpha} = 1.0623 - 3.9167 \times 0.4319 = -0.6293$
$\hat{\sigma}^2 = 0.32589/(12-2) = 0.03259$

Til udregning af 95% sikkerhedsgrænser for β får vi brug for følgende størrelser:

Estimeret standardafvigelse for $\hat{\beta}$:

$$\sqrt{\frac{0.03259}{13.3567}} = 0.04940$$

97.5% fraktilen i T–fordelingen med 10 frihedsgrader: 2.228.

95% sikkerhedsgrænser for β er således givet ved

$$\beta = 0.4319 \pm 2.228 \times 0.04940 = 0.4319 \pm 0.1101$$

svarende til konfidensintervallet [0.3218,0.5420].

(b)

Estimatet for den forventede værdi af effekten ved vindstyrke 4 er naturligvis

$$\hat{\alpha} + \hat{\beta} \times 4 = -0.6293 + 0.4319 \times 4 = 1.0983\,.$$

Den estimerede standardafvigelse på denne størrelse er (Stat. side 79)

$$\sqrt{\left(\frac{1}{12} + \frac{(4-3.9167)^2}{13.3567}\right) \times 0.03259} = 0.052275$$

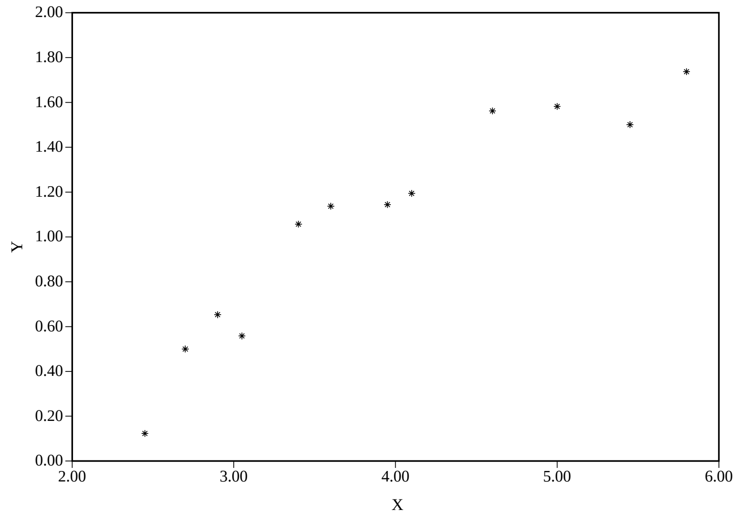
så 95% konfidensgrænserne er givet ved

$$\alpha + \beta \times 4 = 1.0983 \pm 2.228 \times 0.052275 = 1.0983 \pm 0.1165$$

svarende til konfidensintervallet [0.9818,1.2148].

(c)

Tegningen



tyder på at sammenhængen mellem x og y snarere er beskrevet ved en konkav funktion end en linie. En model som (for $\gamma > 0$) tillader en række funktioner af et lignende udseende er netop den multiple regressionsmodel

$$y_i \sim N\left(\alpha + \beta x + \gamma \log(x), \sigma^2\right)$$

Udskriften af estimaterne i denne model viser at testet for $\gamma = 0$ fører til forkastelse med en P–værdi på 0.001688. Vi må konkludere, at den lineære regressionsmodel ikke er god nok. Den krumning som tegningen antyder (som en "sur parabel", idet $\hat{\gamma} = 4.260 > 0$) kan ikke ignoreres.

Så ville det naturligvis være rart hvis vi kunne bruge en simpel regression med $\log(x)$ som forklarende variabel i stedet for. Men det kan vi åbenbart heller ikke, da testet for $\beta = 0$ fører til forkastelse med $P = 0.027243$. Dog er der her tale om en væsentlig svagere signifikans. Så hvis man skulle vælge mellem x og $\log(x)$ som forklarende variabel i en lineær regression, skulle man helt klart vælge $\log(x)$. Man kunne naturligvis også forsøge sig med $\sqrt{x}$, x^2 osv.