

Reeksamen august 2004, Opgave 4

(a)

$$\bar{x} = 1639/12 = 136.583$$

$$\bar{y} = 26470/12 = 2205.83$$

$$SSD_x = 246267 - 1639^2/12 = 22406.92$$

$$SSD_y = 63627352 - 26470^2/12 = 5238943.7$$

$$SPD_{xy} = 3918147 - 1639 \times 26470/12 = 302786.17$$

$$\hat{\beta} = 302786.17/22406.92 = \mathbf{13.513}$$

$$\hat{\alpha} = 2205.83 - 136.583 \times 13.513 = \mathbf{360.174}$$

$$SSD_{\text{res}} = 5238943.7 - 13.513^2 \times 22406.92 = 1147373.90$$

$$\hat{\sigma}^2 = 1147373.90/10 = \mathbf{114737.39}$$

Til udregning af 95% konfidensgrænser for α og β får vi brug for 97.5%-fraktilen i t-fordelingen med 10 frihedsgrader, som er 2.228.

Standardafvigelsen for $\hat{\beta}$ estimeres til

$$\sqrt{\frac{114737.39}{22406.92}} = 2.2629$$

så vi får

$$\beta = 13.513 \pm 2.228 \times 2.2629 = 13.513 \pm 5.042$$

svarende til 95% sikkerhedsintervallet **[8.471,18.555]**.

Standardafvigelsen for $\hat{\alpha}$ estimeres til

$$\sqrt{\left(\frac{1}{12} + \frac{136.583^2}{22406.92}\right) \times 114737.39} = 324.17$$

så vi får

$$\alpha = 360.174 \pm 2.228 \times 324.17 = 360.174 \pm 722.298$$

svarende til 95% sikkerhedsintervallet **[-362.1,1082.5]**.

(b)

At hypotesen $\alpha = 0$ godkendes på 95% niveauet følger af at 0 ligger i 95% sikkerhedsintervallet for α . Der er således ikke grund til at udregne selve T-størrelsen, men hvis man gør det alligevel får man

$$t = \frac{360.174}{324.17} = 1.111.$$

(som altså er insignifikant). Vi kan altså godkende hypotesen om at regressionslinien går gennem (0,0), svarende til at prisen følger boligarealet proportionalt. Det forekommer ikke helt realistisk, og godkendelsen af denne hypotese kan måske tilskrives datasættets beskedne størrelse.

(c)

F-teststørrelsen for reduktion af den udvidede model til den simple regressionsmodel bliver (idet dimensionen af den udvidede model ses at være $1+1+3-1=4$)

$$f = \frac{((1147374 - 397019)/2)}{397019/(12 - 4)} = 7.5599.$$

Denne størrelse skal vurderes i en F-fordeling med (2,8) frihedsgrader. I denne fordeling 95%-fraktilen 4.46 og 99%-fraktilen er 8.65. Vi får altså en P-værdi under 0.05, formodentlig ret tæt på 0.01. Dette tyder på, at antallet af soveværelser har indflydelse på prisen. Hvilken vej denne indflydelse går har vi ikke den store mulighed for at undersøge ved hjælp af en lommeregner. Men et godt gæt er naturligvis, at prisen stiger med antallet af soveværelser.

(I 13-tals besvarelsen begrundes den sidste påstand med et x - y plot, hvor det ses at i hvert fald to af de tre huse med 4 soveværelser ligger markant over regressionslinien).