

Et regressionsanalyse-eksempel

Anscombes data

I nedenstående tabel er vist fire forskellige (konstruerede) sæt af talpar (x, y) ; hvert sæt kan underkastes en lineær regressionsanalyse.^[1]

1. Hvis man ikke tænkte nærmere over det, kunne man måske finde på at bære sig ad som om tallene y_1, y_2, \dots, y_{11} i et givet datasæt var observerede værdier af uafhængige stokastiske variable Y_1, Y_2, \dots, Y_{11} , hvor Y_i antages at være normalfordelt med middelværdi $\alpha + \beta x_i$ og varians σ^2 , $i = 1, 2, \dots, 11$.

Udregn for hvert datasæt estimererne $\hat{\alpha}$, $\hat{\beta}$ og s^2 over parametrene α , β og σ^2 .

2. Lav for hvert datasæt et *scatterplot*, dvs. en tegning med punkterne (x_i, y_i) , og indtegn den estimerede regressionslinje.
3. Hvad kan man lære heraf?

Tabel 1 Anscombes data.

datasæt 1		datasæt 2		datasæt 3		datasæt 4	
x	y	x	y	x	y	x	y
10	8.04	7	7.26	11	7.81	8	6.58
8	6.95	4	3.10	4	5.39	8	5.76
13	7.58	14	8.10	5	5.73	8	7.71
9	8.81	9	8.77	13	12.74	8	8.84
11	8.33	8	8.14	14	8.84	8	8.47
14	9.96	10	9.14	12	8.15	8	7.04
6	7.24	13	8.74	10	7.46	8	5.25
4	4.26	11	9.26	9	7.11	19	12.50
12	10.84	6	6.13	6	6.08	8	5.56
7	4.82	12	9.13	7	6.42	8	7.91
5	5.68	5	4.74	8	6.77	8	6.89

[1] F. J. Anscombe. Graphs in statistical analysis. *The American Statistician*, 27:17–21, 1973.