Línuleg tölfræðilíkön (09.15.46)

Vikublað 2

1: Lítið á línulega aðhvarfsgreiningarlíkanið $Y_i \sim n(\alpha + \beta x_i, \sigma^2)$ með óháðar normaldreifðar hendingar Y_i. Finnið sennileikametilinn fyrir σ^2.

2: Endurtakið fyrir almennt aðhvarfsgreiningarlíkan, $Y_i \sim n(f(\mathbf{x_i}, \boldsymbol{\beta}), \sigma^2)$ þar sem f er eitthvert fall sem tengir væntigildi Y_i við óþekkta stuðla í vigrinum $\boldsymbol{\beta}$ og óháðu breyturnar í vigrinum $\mathbf{x_i}$.

3: Fjölvíða aðhvarfsgreiningu má gera með því að lágmarka kvaðratsummu frávika ekki einungis yfir tvo stika, heldur nota fleiri óháðar breytur. Þannig má t.d. meta α, β og γ með því að lágmarka $\sum_i (y_i - (\alpha + \beta x_i + \gamma z_i))^2$. Setjið tilsvarandi líkan fram á fylkjaformi og skrifið niður jöfnur til að leysa bestunarverkefnið. Athugið að pakkar eins og R leysa verkefni af þessari gerð einfaldlega með upptalningu á óháðu breytunum innan "lm" skipunarinnar ("lm(y∼x+z)").

4: Gerið einstaklinsbundnu aðhvarfsgreininguna á vikublaði 1 í R en lítið á tvær óháðar breytur, $x1 = x$ og $x2 = x^2$. Notið "lm" með (a) "summary", (b) "anova", (c) "add1" og (d) "drop1" skiparnar. Útskýrið kvaðratsummurnar í hverri einustu töflu (ekki F-prófin eða frígráðurnar).

5: Gerið einfalda aðhvarfsgreiningu með gögnunum hér fyrir neðan. Finnið, hvað er að líkaninu, lagið líkanið eins og með þarf og endurtakið síðan (3) með réttu líkani.

Valin dæmi úr bók, kafla 2: 2.6, 2.28, 2.29, 2.30, 2.31, 2.32.

English version

1: Consider the linear regression model $Y_i \sim n(\alpha + \beta x_i, \sigma^2)$ with independent random variables Y_i. Find the maximum likelihood estimator for σ^2.

2: Repeat for a general regression model, $Y_i \sim n(f(\mathbf{x_i}, \boldsymbol{\beta}), \sigma^2)$ where f is a function linking the expected value of Y_i with unknown coefficients in the vector $\boldsymbol{\beta}$ and the "independent" variables in the vectors $\mathbf{x_i}$.

3: Multiple regression is undertaken by minimizing the sum of squared errors not only over two parameters, but using more independent variables. For example α, β and γ can be estimated by minimizing $\sum_i (y_i - (\alpha + \beta x_i + \gamma z_i))^2$. Write the corresponding model in matrix form and derive equations for estimating the optimization problem. Note that a package like R will solve these problems when the independent variables are listed within the "lm" command ("lm(y∼x+z)").

4: Repeat the individualized regression in homework 1 in R, but consider two independent variables, $x1 = x$ and $x2 = x^2$. Use the "lm" function with (a) "summary", (b) "anova", (c) "add1" and (d) "drop1" commmands. Explain the sums of squares in the output from each command (not the F-tests or the degrees of freedom).

5: Consider the data set below. Analyse the data first using simple linear regression and find the problem with the model. Extend the model as needed and repeat (3) with the correct model.

Selected exercises from the book, chapter 2: 2.6, 2.28, 2.29, 2.30, 2.31, 2.32.

x	y
1	4.18
2	2.98
3	4.74
4	5.66
5	4.57
6	3.82
7	4.03
8	1.34
9	1.61
10	2.39
11	0.01
12	-2.16
13	-0.23
14	-2.41
15	-4.14
16	-3.94
17	-5.29
18	-7.41
19	-10.80
20	-10.43
21	-12.41
22	-14.64
23	-18.36
24	-19.79
25	-19.69