NTNU
Kunnskap for ei betre verd

Forelesing 25
Lineær regresjon

Torstein Fjeldstad
Institutt for matematiske fag, NTNU
11.04.2018

I dag

— Eigenskapar til estimatorane (lineær regresjon)

Enkel lineær regresjon

Situasjon: har observert $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

— $x_1, x_2, \ldots, x_n$ er kjende tal

— $y_1, y_2, \ldots, y_n$ er realisasjonar frå uavhengige stokastiske variablar $Y_1, Y_2, \ldots, Y_n$ med

$$Y_i \sim n(y_i; \alpha + \beta x_i, \sigma)$$

Mål: estimere α, β og σ^2

— Minste kvadraters metode

— Sannsynsmaksimeringsprinsippet

SME enkel lineær regresjon I

Maksimer log-rimelighetsfunksjon

$$l(\alpha, \beta, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2$$

Får føgljande likningssystem:

$$n\alpha + \beta\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \qquad (1)$$

$$\alpha\sum_{i=1}^{n} x_i + \beta\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \qquad (2)$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2 = n \qquad (3)$$

SME enkel lineær regresjon II

Definer:

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

SME

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) Y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \qquad \widehat{\alpha} = \bar{Y} - \widehat{\beta}\bar{x} \qquad \widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \widehat{\alpha} - \widehat{\beta} x_i)^2$$

Eigenskapar til estimatorane

$$\widehat{\beta} \sim n\left(z; \beta, \sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}\right)$$

Eigenskapar til estimatorane

$$\widehat{\beta} \sim n\left(z; \beta, \sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}\right)$$

$$\widehat{\alpha} \sim n\left(z; \alpha, \sqrt{\frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}}\right)$$

Eigenskapar til estimatorane

$$\widehat{\beta} \sim n\left(z; \beta, \sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}\right)$$

$$\widehat{\alpha} \sim n\left(z; \alpha, \sqrt{\frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}}\right)$$

Merk

$$E(\widehat{\sigma}^2) = \frac{n-2}{n}\sigma^2,$$

me nyttar derfor

$$S^2 = \frac{n}{n-2}\widehat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \widehat{\alpha} - \widehat{\beta}x_i)^2$$

Neste veke

- Prediksjonsintervall
- Residualplott
- Oppsummering av pensum