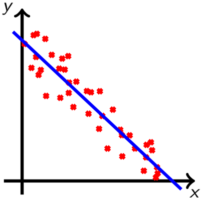
Egenskaper til estimatorene i enkel lineær regresjon

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland
Institutt for matematiske fag
Norges teknisk-naturvitenskapelige universitet

Enkel lineær regresjon

- ⋆ Situasjon: Har observasjonspar $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- ⋆ Tilpasser en rett linje til de observerte parene
- ⋆ Sannsynlighetsmodell: $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$,
  $Y_1, Y_2, \ldots, Y_n$ uavhengige
- ⋆ Husk:
  - α og β er ukjente parametre (tall)
  - betrakter Y_i-ene som stokastiske
  - betrakter x_i som tall (altså ikke stokastiske variabler)



- ⋆ SME:

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})Y_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$\widehat{\alpha} = \overline{Y} - \widehat{\beta}\overline{x}$$

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

- ⋆ Fokus i denne videoen:
  - $E[\widehat{\alpha}]$ og $\text{Var}[\widehat{\alpha}]$
  - $E[\widehat{\beta}]$ og $\text{Var}[\widehat{\beta}]$
  - sannsynlighetsfordelingen for $\widehat{\alpha}$ og for $\widehat{\beta}$
  - $E[\widehat{\sigma}^2]$

Utregning av $E[\widehat{\beta}]$ og $\text{Var}[\widehat{\beta}]$

$$E[aX] = aE[X]$$
$$E[b] = b$$
$$E[X+Y] = E[X] + E[Y]$$
$$\text{Var}[aX] = a^2\text{Var}[X]$$
$$\text{Var}[b] = 0$$
Hvis X, Y uavh.:
$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$$

- ⋆ $Y_1, Y_2, \ldots, Y_n$ er uavhengige og $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$
- ⋆ Estimator for β

$$\widehat{\beta} = \frac{\sum_{i=1}^n (x_i - \overline{x}) Y_i}{\sum_{i=1}^n (x_i - \overline{x})^2}$$

$$\begin{aligned}
E[\widehat{\beta}] &= E\left[\frac{\sum_{i=1}^n (x_i - \overline{x}) Y_i}{\sum_{i=1}^n (x_i - \overline{x})^2}\right] \\
&= \frac{1}{\sum_{i=1}^n (x_i - \overline{x})^2} E\left[\sum_{i=1}^n (x_i - \overline{x}) Y_i\right] \\
&= \frac{\sum_{i=1}^n E[(x_i - \bar{x}) Y_i]}{\sum_{i=1}^n (x_i - \overline{x})^2} \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x}) E[Y_i]}{\sum_{i=1}^n (x_i - \overline{x})^2} \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x})(\alpha + \beta x_i)}{\sum_{i=1}^n (x_i - \overline{x})^2} \\
&= \frac{\alpha \sum_{i=1}^n (x_i - \bar{x}) + \beta \sum_{i=1}^n (x_i - \bar{x}) x_i}{\sum_{i=1}^n (x_i - \overline{x})^2} \\
&= \beta \frac{\sum_{i=1}^n (x_i - \bar{x}) x_i}{\sum_{i=1}^n (x_i - \overline{x})^2} \\
&= \beta
\end{aligned}$$

$$\begin{aligned}
\text{Var}[\widehat{\beta}] &= \text{Var}\left[\frac{\sum_{i=1}^n (x_i - \overline{x}) Y_i}{\sum_{i=1}^n (x_i - \overline{x})^2}\right] \\
&= \left(\frac{1}{\sum_{i=1}^n (x_i - \overline{x})^2}\right)^2 \text{Var}\left[\sum_{i=1}^n (x_i - \overline{x}) Y_i\right] \\
&= \frac{\sum_{i=1}^n \text{Var}\left[(x_i - \overline{x}) Y_i\right]}{\left(\sum_{i=1}^n (x_i - \overline{x})^2\right)^2} \\
&= \frac{\sum_{i=1}^n (x_i - \overline{x})^2 \text{Var}\left[Y_i\right]}{\left(\sum_{i=1}^n (x_i - \overline{x})^2\right)^2} \\
&= \frac{\sum_{i=1}^n (x_i - \overline{x})^2 \sigma^2}{\left(\sum_{i=1}^n (x_i - \overline{x})^2\right)^2} \\
&= \frac{\sigma^2 \sum_{i=1}^n (x_i - \overline{x})^2}{\left(\sum_{i=1}^n (x_i - \overline{x})^2\right)^2} \\
&= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2}
\end{aligned}$$

Utregning av $E[\widehat{\alpha}]$ og $\text{Var}[\widehat{\alpha}]$

$$E[\widehat{\beta}] = \beta$$
$$\text{Var}[\widehat{\beta}] = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2}$$

$$E[aX] = aE[X]$$
$$E[b] = b$$
$$E[X+Y] = E[X] + E[Y]$$
$$\text{Var}[aX] = a^2\text{Var}[X]$$
$$\text{Var}[b] = 0$$
Hvis X, Y uavh.:
$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$$

- ⋆ $Y_1, Y_2, \ldots, Y_n$ er uavhengige og $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$
- ⋆ Estimator for α

$$\widehat{\alpha} = \overline{Y} - \widehat{\beta}\overline{x}$$

$$\begin{aligned}
E[\widehat{\alpha}] &= E\left[\overline{Y} - \widehat{\beta}\overline{x}\right] \\
&= E[\overline{Y}] + E[-\widehat{\beta}\overline{x}] \\
&= E\left[\frac{1}{n}\sum_{i=1}^n Y_i\right] - E[\widehat{\beta}]\overline{x} \\
&= \frac{1}{n}E\left[\sum_{i=1}^n Y_i\right] - \beta\overline{x} \\
&= \frac{1}{n}\sum_{i=1}^n E[Y_i] - \beta\overline{x} \\
&= \frac{1}{n}\sum_{i=1}^n (\alpha + \beta x_i) - \beta\overline{x} \\
&= \frac{1}{n}\cdot n\alpha + \beta\cdot\frac{1}{n}\sum_{i=1}^n x_i - \beta\overline{x} \\
&= \alpha
\end{aligned}$$

$$\begin{aligned}
\text{Var}[\widehat{\alpha}] &= \text{Var}\left[\overline{Y} - \widehat{\beta}\overline{x}\right] \\
&= \text{Var}\left[\overline{Y} + (-\widehat{\beta}\overline{x})\right] \\
&= \text{Var}[\overline{Y}] + \text{Var}[-\widehat{\beta}\overline{x}] + 2\text{Cox}[\overline{Y}, -\widehat{\beta}\overline{x}] \\
&= \text{Var}\left[\frac{1}{n}\sum_{i=1}^n Y_i\right] + (-\overline{x})^2\text{Var}[\widehat{\beta}] - 2\overline{x}\text{Cov}[\overline{Y}, \widehat{\beta}] \\
&= \left(\frac{1}{n}\right)^2 \text{Var}\left[\sum_{i=1}^n Y_i\right] + \frac{\overline{x}^2\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2} \\
&= \frac{1}{n^2}\sum_{i=1}^n \text{Var}[Y_i] + \frac{\overline{x}^2\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2} \\
&= \sigma^2\left(\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^n (x_i - \overline{x})^2}\right) \\
&= \sigma^2 \cdot \frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i - \overline{x})^2}
\end{aligned}$$

Sannsynlighetsfordeling for $\widehat{\beta}$

- Modell: $Y_1, Y_2, \ldots, Y_n$ uavhengige og $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$
- Vi har funnet

$$\widehat{\beta} = \frac{\sum_{i=1}^n (x_i - \overline{x}) Y_i}{\sum_{i=1}^n (x_i - \overline{x})^2}$$

$$E[\widehat{\beta}] = \beta \quad \text{og} \quad \text{Var}[\widehat{\beta}] = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2}$$

- Husk teorem: La $X_1, X_2, \ldots, X_n$ være uavhengige og $X_i \sim N(\mu_i, \sigma_i^2), i = 1, 2, \ldots, n$. La $a_1, a_2, \ldots, a_n$ og b være konstanter, og

$$Y = \sum_{i=1}^n a_i X_i + b = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n + b.$$

Da er

$$Y \sim N\left(\sum_{i=1}^n a_i \mu_i + b, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$

- Merk: Vi kan skrive $\widehat{\beta}$ på formen

$$\widehat{\beta} = \sum_{i=1}^n \frac{x_i - \overline{x}}{\sum_{j=1}^n (x_j - \overline{x})^2} Y_i = \sum_{i=1}^n a_i Y_i \quad \text{der} \quad a_i = \frac{x_i - \overline{x}}{\sum_{j=1}^n (x_j - \overline{x})^2}$$

- Dette gir at

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

Sannsynlighetsfordeling for $\widehat{\alpha}$

- Modell: $Y_1, Y_2, \ldots, Y_n$ uavhengige og $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$
- Vi har funnet

$$\widehat{\alpha} = \overline{Y} - \widehat{\beta}\overline{x} \quad \text{der} \quad \widehat{\beta} = \frac{\sum_{i=1}^n (x_i - \overline{x}) Y_i}{\sum_{i=1}^n (x_i - \overline{x})^2}$$

$$E[\widehat{\alpha}] = \alpha \quad \text{og} \quad \text{Var}[\widehat{\alpha}] = \sigma^2 \cdot \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \overline{x})^2}$$

- Husk teorem: La $X_1, X_2, \ldots, X_n$ være uavhengige og $X_i \sim N(\mu_i, \sigma_i^2), i = 1, 2, \ldots, n$. La $a_1, a_2, \ldots, a_n$ og b være konstanter, og

$$Y = \sum_{i=1}^n a_i X_i + b = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n + b.$$

Da er

$$Y \sim N\left(\sum_{i=1}^n a_i \mu_i + b, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$

- Merk: Vi kan skrive $\widehat{\alpha}$ på formen

$$\widehat{\alpha} = \frac{1}{n}\sum_{i=1}^n Y_i - \left(\sum_{i=1}^n \frac{x_i - \overline{x}}{\sum_{j=1}^n (x_j - \overline{x})^2} Y_i\right)\overline{x} = \sum_{i=1}^n a_i Y_i \quad \text{der} \quad a_i = \frac{1}{n} - \frac{(x_i - \overline{x})\overline{x}}{\sum_{j=1}^n (x_j - \overline{x})^2}$$

- Dette gir at

$$\widehat{\alpha} \sim N\left(\alpha, \sigma^2 \cdot \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \overline{x})^2}\right)$$

Sannsynlighetsfordeling for $\widehat{\sigma}^2$

- Modell: $Y_1, Y_2, \ldots, Y_n$ uavhengige og $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$
- Vi har funnet

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

- Merk: Vi har at

$$\frac{Y_i - (\alpha + \beta x_i)}{\sigma} \sim N(0,1)$$

$$\left(\frac{Y_i - (\alpha + \beta x_i)}{\sigma}\right)^2 \sim \chi_1^2$$

$$\sum_{i=1}^{n}\left(\frac{Y_i - (\alpha + \beta x_i)}{\sigma}\right)^2 \sim \chi_n^2$$

- Hva skjer når α og β erstattes av henholdsvis $\widehat{\alpha}$ og $\widehat{\beta}$?

Teorem

La $Y_1, Y_2, \ldots, Y_n$ være uavhengige og $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$, og la $\widehat{\alpha}$ og $\widehat{\beta}$ være gitt ved SME. Da har vi at

$$\sum_{i=1}^{n}\left(\frac{Y_i - (\widehat{\alpha} + \widehat{\beta} x_i)}{\sigma}\right)^2 \sim \chi_{n-2}^2$$

Dessuten er $\widehat{\sigma}^2$ uavhengig av estimatorene $\widehat{\alpha}$ og $\widehat{\beta}$.

- Dette gir at

$$\frac{n\widehat{\sigma}^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{Y_i - (\widehat{\alpha} + \widehat{\beta} x_i)}{\sigma}\right)^2 \sim \chi_{n-2}^2$$

Estimator for σ^2

- Husk:

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - (\widehat{\alpha} + \widehat{\beta}x_i))^2 \quad \text{og} \quad \frac{n\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}$$

- Dermed har vi at

$$E\left[\frac{n\widehat{\sigma}^2}{\sigma^2}\right] = n-2 \quad \Rightarrow \quad E[\widehat{\sigma}^2] = \frac{n-2}{n}\sigma^2$$

- En forventningsrett estimator for σ^2 er

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

- Merk at vi da har

$$\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}$$

Oppsummering

- Vi har funnet at for en enkel lineær regresjonsmodell har vi

$$\widehat{\alpha} \sim N\left(\alpha, \sigma^2 \cdot \frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i - \overline{x})^2}\right)$$

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

$$S^2 = \frac{1}{n-2}\sum_{i=1}^n (Y_i - (\widehat{\alpha} + \widehat{\beta}x_i))^2, \quad E[S^2] = \sigma^2$$

$$\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}$$

  - S^2 er uavhengig av $\widehat{\alpha}$ og av $\widehat{\beta}$