Enkel lineær regresjon

⋆ Situasjon:

- har observert $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- ønsker å tilpasse en rett linje til disse dataene

⋆ Stokastisk modell:

- betrakter $y_1, y_2, \ldots, y_n$ som realisasjoner av stokastiske variabler Y_1, $Y_2, \ldots, Y_n$
- betrakter $x_1, x_2, \ldots, x_n$ som (kjente) tall
- antar Y_1, $Y_2, \ldots, Y_n$ uavhengige og

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

- ukjente parametre: α, β og σ^2

⋆ Metoder for å estimere parametrene

- sannsynlighetsmaksimeringsprinsippet
- minste kvadraters metode

Enkel lineær regresjon — SME

- ⋆ SME

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

- ⋆ Har argumentert for at $\widehat{\alpha}$ og $\widehat{\beta}$ er normalfordelte, og vist at

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

$$\mathrm{E}\left[\widehat{\alpha}\right] = \alpha \quad \text{og} \quad \mathrm{Var}\left[\widehat{\alpha}\right] = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2},$$

- ⋆ Estimatoren $\widehat{\sigma}^2$ var forventingsskjev,

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

Bruker i stedet den forventningsrette estimatoren

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

Inferens om α og β

⋆ Standardiserer $\widehat{\alpha}$ og $\widehat{\beta}$,

$$\frac{\widehat{\alpha}-\alpha}{\sqrt{\frac{\sigma^2\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i-\bar{x})^2}}} \sim n(z;0,1) \quad \text{og} \quad \frac{\widehat{\beta}-\beta}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}} \sim n(z;0,1)$$

⋆ Hvis σ^2 er ukjent erstattes denne med estimatoren S^2,

$$\frac{\widehat{\alpha}-\alpha}{\sqrt{\frac{S^2\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i-\bar{x})^2}}} \sim t_{n-2} \quad \text{og} \quad \frac{\widehat{\beta}-\beta}{\sqrt{\frac{S^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}} \sim t_{n-2}$$

⋆ Dette er utgangspunkt for utledning av konfidensintervall for α og β, og for å lage testobservator i en hypotesetest.

Inferens om α og β

- Standardiserer $\widehat{\alpha}$ og $\widehat{\beta}$,

$$\frac{\widehat{\alpha}-\alpha}{\sqrt{\frac{\sigma^2\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i-\bar{x})^2}}} \sim n(z;0,1) \quad \text{og} \quad \frac{\widehat{\beta}-\beta}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}} \sim n(z;0,1)$$

- Hvis σ^2 er ukjent erstattes denne med estimatoren S^2,

$$\frac{\widehat{\alpha}-\alpha}{\sqrt{\frac{S^2\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i-\bar{x})^2}}} \sim t_{n-2} \quad \text{og} \quad \frac{\widehat{\beta}-\beta}{\sqrt{\frac{S^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}} \sim t_{n-2}$$

- Dette er utgangspunkt for utledning av konfidensintervall for α og β, og for å lage testobservator i en hypotesetest.
  - for eksempel: $H_0 : \alpha = 0$ mot $H_1 : \alpha \neq 0$
  - testobservator:

$$T = \frac{\widehat{\alpha}}{\sqrt{\frac{S^2\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i-\bar{x})^2}}} \sim t_{n-2} \quad \text{når } H_0 \text{ er riktig}$$

Plan for i dag

- ⋆ Konfidensintervall for forventningsverdien til en ny observasjon,

$$\mu_{Y_0|x_0} = \mathrm{E}[Y_0|x_0] = \alpha + \beta x_0$$

  - Estimator for $\mu_{Y_0|x_0}$: $\widehat{\mu}_{Y_0|x_0} = \widehat{\alpha} + \widehat{\beta} x_0$

- ⋆ Prediksjonsintervall for en ny observasjon Y_0 med $x = x_0$

- ⋆ Hvordan vurdere modellantagelsene?
  - residualplott

- ⋆ Starte på repetisjon