NTNU
Kunnskap for ei betre verd

Forelesing 24
Lineær regresjon

Torstein Fjeldstad
Institutt for matematiske fag, NTNU
10.04.2018

I dag

— Praktisk informasjon
— Fullføre kap. 10 (hypotesetest)
— Lineær regresjon

Informasjon I

- Det vil vere vikar tysdag 17.04
- Siste førelesing er tysdag 24.04
- Du er sjølve ansvarleg for at du har godkjend øvingsopplegg
- Eksamen onsdag 23.05
- Det vil bli spørjetimer/statistikklab før eksamen. Tidspunkt vil bli lagt ut på heimesida snart

Informasjon II

Tentativ plan

Dato	Plan
10.04	Avslutning av hypotesetest og introduksjon til lineær regresjon
11.04	Eigenskapar til estimatorar (lineær regresjon)
17.04	Prediksjonsintervall og residualplott
18.04	Avslutning av lineær regresjon og oppsummering av pensum
24.04	Oppgåverekning

Informasjon III

Statistikklab:

Veke	Tysdag (14:15 - 18:00)	Onsdag (14:15 - 18:00)	Torsdag (12:15 - 16:00)
15	R2 og S5	R2 og S4	Kun R2
16	R2 og S5	R2 og S7	R2 (14:15 - 18:00) S2 (14:15 - 16:00)
17	R2 og S5	R2 og S7	R2 og S2

Informasjon IV

Eksamensforelesing? (Dato vil bli bestemd på Kahoot i morgon)

1. Fredag 11.05
2. Laurdag 12.05
3. Tysdag 15.05
4. Laurdag 19.05

Val av tal på observasjonar n I

Vil undersøkje om ein ny medisin B er betre enn eksisterande medisin A

- Eksisterande medisin A verker på 70% av pasientane
- Ny medisin B påstås å verke på meir enn 70% av pasientane

Korleis kan me konkludere om B er betre enn A?
Kor mange pasienter må me gi medisin B for å kunne vere "sikre" på konklusjonen vår?

Val av tal på observasjonar n II

Hypotesetest:

$$H_0 : p = p_0 = 0.7 \qquad \text{mot} \qquad H_1 : p > 0.7$$

Hugs:

$$\frac{X - np}{\sqrt{np(1-p)}} \approx n(z; 0, 1) \quad \text{alltid}$$

$$Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} \approx n(z; 0, 1) \quad \text{når } H_0 \text{ er riktig}$$

Val av tal på observasjonar n III

Forkast H_0 dersom

$$Z = \frac{X - np_0}{\sqrt{np_0(1 - p_0)}} > z_\alpha$$

Velg α slik at

$$P(\text{Type I-feil}) = P(\text{forkast } H_0 \text{ når } H_0 \text{ riktig}) \leq \alpha$$

er liten. Itillegg ynskjer me å velge n slik at me kontrollerer

$$P(\text{Type II-feil}) = P(\text{ikkje forkast } H_0 \text{ når } H_1 \text{ riktig}) \leq \beta.$$

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 (i dag): 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:

$$n\alpha + \beta\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{1}$$

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

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

SME enkel lineær regresjon II

SME

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

I morgon

- Lineær regresjon
- Kahoot