

Plan

Kap. 9 Estimering: Mandag 7. mars, torsdag 8.mars og mandag 14.

Kap. 10 Hypotesetesting: Tirsdag 15. og **onsdag 16. 12.15-14 i R8 og onsdag 30 mars 12.15-14 i R8**

Kap. 11 Lineær regresjon (++) 4.april, 5.april, 11.april, 12.april
Oppsummering og eksamsoppgåverekning: 26. og 27. mai (10-12).

Referansegruppemøte: Onsdag 9. mai 13.15-14. Gje
tilbakemelding!

- Kva er gjennomsnittshøgda for kvinnelege NTNU-studentar. **Estimering**
- Er eg høgare enn gjennomsnittleg kvinnelege NTNU-studentar. **Hypotesetest**
- Kva er forskjellen i kvalitet på laks for lagringsmetode 1 og lagringsmetode 2. **Estimering**
- Er kvaliteten på laks ved lagringsmetode 2 betre enn ved lagringsmetode 1. **Hypotesetest**

- Antar fordeling frå kjennskap til fenomenet; $X_i \sim f(x; \theta)$.
 - Har data $x_1, x_2, \dots, x_n$
 - *Estimerer/ berekner θ frå data v.h.a. ein estimator $\hat{\theta}$.*
-
- Kva er ein god estimator (**forventningsrett og minst mogeleg varians**).
 - Korleis finne ein estimator (**SME (forts. i dag)**)
 - Korleis kvantifisere usikkerheita i estimat (i dag: **Konfidensintervall**)

Finn den verdien for parameteren θ (studentpåske $\theta = p$) som gjev høgast sannsyn for å observere dei dataene vi har observert.

Korleis

- 1 Finn likelihoodfunksjonen

$$L(\theta; x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n; \theta)$$

Sannsynet /sanns.tettheten for våre data.

- 2 Finn toppunkt av likelihoodfunksjonen: $\underset{\theta}{\operatorname{argmax}} L(\theta; x_1, \dots, x_n)$

Finn den verdien for parameteren θ (student påske $\theta = p$) som gjev høgast sannsyn for å observere dei dataene vi har observert.

OPPSKRIFT

- 1 Finn likelihoodfunksjonen

$$L(\theta; x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n; \theta) \stackrel{\text{uavh}}{=} \prod_{i=1}^n f(x_i; \theta)$$

- 2 Finn toppunkt av likelihoodfunksjonen:

- Tar ln av L ; $l(\theta; x_1, x_2, \dots, x_n) = \ln(L(\theta; x_1, x_2, \dots, x_n))$.
Reknetriks som nesten alltid blir brukt. L og l har same toppunkt.

- Deriverer og set lik 0; $\frac{\partial}{\partial \theta} l(\theta; x_1, x_2, \dots, x_n) = 0$

- Løyer ut for $\theta = h(x_1, x_2, \dots, x_n)$.

- 3 **Estimator:** $\hat{\theta} = h(X_1, X_2, \dots, X_n)$ (stok.var.)

Estimat: $\theta^* = h(x_1, x_2, \dots, x_n)$ (talverdi)

Eksempel NTNU kvinner, case 1 og 2

Høgde kvinnelege NTNU-DATA

- CASE 1

- $n = 68$
- $\bar{x} = 169.1$, kjent varians $\sigma^2 = 6.0^2$

- CASE 2

- $n=5$
- $\bar{x} = 169.1$, kjent varians $\sigma^2 = 6.0^2$

ANTAR

- $X_i \stackrel{u.i.f.}{\sim} N(\mu, \sigma^2)$ for $i = 1, 2, \dots, n$

ESTIMATOR

- $\hat{\mu} = \bar{X} \sim N(\mu, \sigma^2/n)$
- CASE 1: $Var(\hat{\mu}) = 0.7^2$
- CASE 2: $Var(\hat{\mu}) = 2.7^2$

Viktige observatorar

Dersom X_i , $i = 1, 2, \dots, n$ er uavhengig identisk fordelt med $E(X_i) = \mu$ og $\text{Var}(X_i) = \sigma^2$, så

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{n \rightarrow \infty} N(0, 1)$$

Dersom $X_i \stackrel{u.i.f.}{\sim} N(\mu, \sigma^2)$ og $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

Student-t fordelt med $\nu = n - 1$ fridomsgrader.

Dersom $X_i \stackrel{u.i.f.}{\sim} N(\mu, \sigma^2)$ og $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Kji-kvadrat fordelt med $\nu = n - 1$ fridomsgrader.

- X_{ny} ; en ny observasjon
- $X_{ny} \sim N(\mu, \sigma^2)$
- μ ukjent, men har estimat frå $n = 68$ data;
 $\mu^* = 169.1$. Varians antatt kjent $\sigma^2 = 6.0^2$.
- X_{ny} uavhengig av X_i -ane i $\bar{X}$.
- $\bar{X} \sim N(\mu, \sigma^2/n)$
- Ser på $X_{ny} - \bar{X} \sim N(0, \sigma^2 + \sigma^2/n) = N(0, \sigma_{diff}^2) = N(0, 6.04^2)$
- $Z = \frac{X_{ny} - \bar{X}}{\sigma_{diff}} \sim N(0, 1)$
- **Prediksjonsintervall:** $\alpha = 0.05$
- $[\bar{x} - z_{\alpha/2}\sigma_{diff}; \bar{x} + z_{\alpha/2}\sigma_{diff}] = [157.2180.9]$

Konfidensintervall: Dersom forsøket blir repetert, vil sann parameter vere i KI i $1 - \alpha$ del av forsøk.