Typiske spørsmål

- Kva er gjennomsnittshøgda for NTNU-studiner.Estimering
- Er eg høgare enn gjennomsnittleg kvinneleg NTNU-student. 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

Estimering

- Har data $x_1, x_2, \ldots, x_n$
- Antar fordeling frå kjennskap til fenomenet; $x_i \sim f(x; \theta)$.
- *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 estimalor (SME, sannsynlighetsmaksimeringsestimator)
  - Kvantifisere usikkerheita i estimat (konfidensintervall)

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

Korleis

1. Finn likelihoodfunksjonen $L(\theta; x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n; \theta)$ Sannsynet /sanns.tettheten for våre data.
2. Finn toppunkt av likelihoodfunksjonen: $\underset{\theta}{argmax}\, L(\theta; x_1, \ldots, x_n)$

Sannsynsmaksimeringsestimator, SME

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

OPPSKRIFT

1. Finn likelihoodfunksjonen
   $L(\theta; x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n; \theta) \overset{uavh}{=} \prod_{i=1}^{n} f(x_i; \theta)$
2. Finn toppunkt av likelihoodfunksjonen:
   - Tar ln av L; $l(\theta; x_1, x_2, \ldots, x_n) = \ln(L(\theta; x_1, x_2, \ldots, x_n))$. *Reknetriks som nesten alltid blir brukt. L og l har same toppunkt.*
   - Deriverer og set lik 0; $\dfrac{\partial}{\partial\theta} l(\theta; x_1, x_2, \ldots, x_n) = 0$
   - Løyser ut for $\theta = h(x_1, x_2, \ldots, x_n)$.
3. Estimator: $\hat{\theta} = h(X_1, X_2, \ldots, X_n)$ (stok.var.)
   Estimat: $\theta^* = h(x_1, x_2, \ldots, x_n)$ (talverdi)

DATA

- CASE 1
  - $n = 18$
  - $\bar{x} = 170.4$, kjent varians $\sigma^2 = 6.0^2$
- CASE 2
  - n=5
  - $\bar{x} = 170.4$, kjent varians $\sigma^2 = 6.0^2$

ANTAR

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

ESTIMATOR

- $\hat{\mu} = \bar{X} \sim N(\mu, \sigma^2/n)$
- CASE 1: $Var(\hat{\mu}) = 1.4^2$
- CASE 2: $Var(\hat{\mu}) = 2.7^2$

Viktige observatorar

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

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

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

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

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

Prediksjonsintervall kvinnlege NTNU stud.

- X_{ny}; en ny observasjon
- $X_{ny} \sim N(\mu, \sigma^2)$
- μ ukjend, men har estimat frå $n = 18$ data; $\mu^* = 170.4$. Varians antatt kjent $\sigma^2 = 6.0^2$.
- X_{ny} uavhengig av X_i-ane i $\bar{X}$.
- Ser på $X_{ny} - \bar{X} \sim N(0, \sigma^2 + \sigma^2/n) = N(0, \sigma^2_{diff}) = N(0, 6.2^2)$
- $Z = \dfrac{X_{ny} - \bar{X}}{\sigma_{diff}} \sim N(0, 1)$
- Prediksjonsintervall: $\alpha = 0.05$
- $[\bar{x} - z_{\alpha/2}\sigma_{diff}; \bar{x} - z_{\alpha/2}\sigma_{diff}] = [158.3; 182.5]$

Oppsummering

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

Observatorar: Z, T og X^2