Plan for dagen

- To utval med kjent varians
- To utval med felles ukjent varians
- To utval med ulik ukjent varians
- Parutval
- To andelar
  - Eksamen juni 2007, 1c)

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

Lagring av fisk

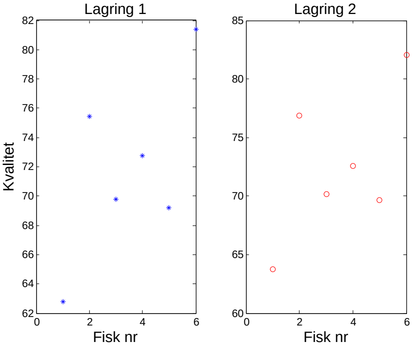
To lagringsmetodar; metode 1 og metode 2.
Antar:

- $X_1 \sim N(\mu_1, \sigma^2)$ for $i = 1, 2, \ldots, n_1$
- $X_2 \sim N(\mu_2, \sigma^2)$ for $i = 1, 2, \ldots, n_2$

Ønsker å finne eit K.I. for $\mu_1 - \mu_2$ med data:

- $n_1 = 6$, $\bar{x}_1 = 71.89$ og $s_1^2 = 6.28^2$
- $n_2 = 6$, $\bar{x}_2 = 72.50$ og $s_2^2 = 6.54^2$

Lagring av fisk



Konfidensintervall

Har eit nivå av trygghet for at sann parameter ligg i intervallet.

1. Finn observator der parameter av interesse og estimator inngår:
   - $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$,
     når SGT eller $X_i \overset{u.i.f.}{\sim} N(\mu, \sigma^2)$
   - $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$,
     når $X_i \overset{u.i.f.}{\sim} N(\mu, \sigma^2)$ og $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
   - $X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$,
     når $X_i \overset{u.i.f.}{\sim} N(\mu, \sigma^2)$ og $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
2. Har at (f.eks) $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$.
3. Løyser ut for parameter, (f.eks. μ); $P(\hat{\mu_L} < \mu < \hat{\mu_L}) = 1 - \alpha$.
4. Konfidensintervall; sett inn for data: $[\mu_L^*, \mu_U^*]$

Finn 95% K.I for $\mu_1 - \mu_2$, lik og kjent varians

Har $\alpha = 0.05$ og at $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$

- Sett inn for Z
- Løys ut for $(\mu_1 - \mu_2)$:

$$P\left[(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2}\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2}\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right] = 1 - \alpha$$

Konfidensintervall:

$$\left[(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}; (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$$

KI, to utval med ulik og ukjent varians

Har $Var(\bar{X}_1 - \bar{X}_2) = \sigma_1^2/n_1 + \sigma_2^2/n_2$.
Kan vise at

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \overset{approx}{\sim} T_\nu$$

Fridomsgrader to utval med ulik varians

$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

ν estimat, rund ned til nærmaste heiltal.

Bildekk

Du skal samanlikne slitasje av to typar bildekk, dekk A og dekk B. Du har ti frivillige test-sjåførar som skal bruke dekkene på sine bilar i ein månad.

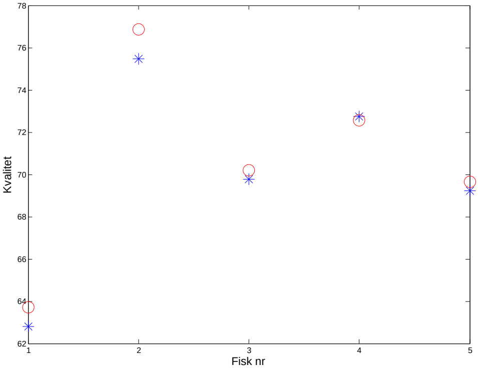
Alt 1: Fem bilar med dekk A og fem bilar med dekk B.

Alt 2: Kvar bil har to av dekk A og to av dekk B.

Kva gjer du? Kvifor?

- Slitasje vil avhenge mykje av kva dekka har vore utsatt for (antall mil, offansiv kjøring, vegkvalitet).
- Med alternativ 2 kan vi samanlikne dekk som har vore utsatt for det same.

Lagring av fisk, differansar



REP: Geometrisk fordeling

Fordelinga til antall terningkast inntil første sekesar.

- Uavhengige forsøk med suksess / ikkje-suksess.
- Likt suksess-sannsyn i alle forsøka.
- X: antall forsøk t.o.m. første suksess.
- X er då geometrisk fordelt

REP: Geometrisk fordeling, pengespelet

- Uavhengige forsøk med suksess / ikkje-suksess.
  Svarar feil / svarar rett
- Likt suksess-sannsyn i alle forsøka.
  Likt sannsyn p for å svare feil.
- X: antall forsøk t.o.m. første suksess.
  t.o.m. første feile svar.
- X er då geometrisk fordelt

REP, Sannsynsmaksimeringsestimator, SME

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

OPPSKRIFT

- Finn likelihoodfunksjonen
  $L(\theta; x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n; \theta) \overset{uavh}{=} \prod_{i=1}^{n} f(x_i; \theta)$
- 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; $\frac{\partial}{\partial \theta} l(\theta; x_1, x_2, \ldots, x_n) = 0$
  - Løyser ut for θ.

REP, Bernoulli-prosess, binomisk fordeling

Bernoulli-prosess

- n uavhengige forsøk.
- Kvart forsøk resulterer i suksess eller ikkje-suksess.
- Sannsynet p for suksess er likt i alle forsøk.

La X vere antall suksess. Då er $X \sim bin(n, p)$

Bernoulli-prosess

- n uavhengige forsøk.
  n_1 uavhengige deltakarar
- Kvart forsøk resulterer i suksess eller ikkje-suksess.
  Kvar deltakar klarer enten ferre enn 5 oppgåver (C) eller 5 eller fleire oppgåver (C').
- Sannsynet p for suksess er likt i alle forsøk.
  $P(C) = q_1$ i alle forsøka

La X vere antall suksess. Då er $X \sim bin(n, p)$
Lar Z_1 vere antall deltakarar med ferre enn 5 rette. Då er $Z_1 \sim bin(n_1, q_1)$

Sentralgrenseteoremet, teorem 8.2

La X_i, $i = 1, 2, \ldots, n$ vere uavhengige identisk fordelte (u.i.f.) stokastiske vaiable med $E(X_i) = \mu$ og $Var(X_i) = \sigma^2 < \infty$ (endeleg varians). La $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ og $Z = \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}$. Då

$$Z \rightarrow N(0, 1)$$

når $n \rightarrow \infty$.
Dette tilsvarer

$$\bar{X} \rightarrow N(\mu, \frac{\sigma^2}{n})$$

når $n \rightarrow \infty$.

PS: Gjeld uansett fordeling for X_i

Lært i dag

- To utval med kjent varians $X_i \sim N(\mu, \sigma^2)$ bruker Z.
- To utval med felles ukjent varians $X_i \sim N(\mu, \sigma^2)$ bruker T med S_p^2.
- To utval med ulik ukjent varians $X_i \sim N(\mu, \sigma^2)$ approx T.
- Parutval (bildekk)
- To andelar SGT $\Rightarrow$ normaltilnærming og Z