

Plan for dagen

Ser på normalfordelte gjennomsnitt (pga SGT eller fordi data frå normalfordeling)

- KI vs hypotesetest (pengespelet, to andelar, SGT)
- Eit utval, kjent varians
- Eit utval, ukjent varians
- To utval, kjent varians
- To utval, lik ukjent varians
- To utval ukjent varians
- Para utval

På torsdag:

- Ein andel
- To andelar
- Varians

- 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**

Metode p-verdi

- 1 Antar H_0 er sann.
- 2 Finn p -verdi: $P(\text{vårt estimat eller meir ekstremt} \mid H_0 \text{ er sann})$
- 3 Forkastar H_0 dersom liten p -verdi ($< \alpha$).

Metode forkastningsområde

- 1 Antar H_0 er sann.
- 2 Finn testobservator og område for 'testobservasjon' (evt. estimat) som fører til forkastning.
- 3 Forkastar H_0 dersom 'testobservasjon'/estimat i forkastningsområdet.

Dersom $X_i \overset{u.i.f.}{\sim} N(\mu, \sigma^2)$, kjent σ^2 eller pga SGT

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

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 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

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

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

Eksamen juni 07, pengespelet

- Spm.lagar A: Tilfeldig deltakar klarer < 5 med sanns. q_1 .
- Spm.lagar B: Tilfeldig deltakar klarer < 5 med sanns. q_2 .
- Z_1 av $n_1 = 64$ klarer ferre enn 5.
- Z_2 av $n_2 = 64$ klarer ferre enn 5.

Utled tilnærma 95% KI for $d = q_1 - q_2$ basert på normaltilnærming til binomisk.

- $Z_1 \sim \text{bin}(n_1, q_1)$, stor n_A , $Z_1 \overset{\text{approx}}{\sim} N(n_1 q_1, n_1 q_1(1 - q_1))$
- $Z_2 \sim \text{bin}(n_2, q_2)$, stor n_B , $Z_2 \overset{\text{approx}}{\sim} N(n_2 q_2, n_2 q_2(1 - q_2))$
- $\hat{q}_1 = Z_1/n_1 \overset{\text{approx}}{\sim} N(q_1, q_1(1 - q_1)/n_1)$
- $\hat{q}_2 = Z_2/n_2 \overset{\text{approx}}{\sim} N(q_2, q_2(1 - q_2)/n_2)$
- $\hat{d} = \hat{q}_1 - \hat{q}_2 \overset{\text{approx}}{\sim} N(q_1 - q_2, q_1(1 - q_1)/n_1 + q_2(1 - q_2)/n_2)$
- $$Z = \frac{\hat{d} - d}{\sqrt{q_1(1 - q_1)/n_1 + q_2(1 - q_2)/n_2}}$$

Ser på $d = \mu_1 - \mu_2$

Konfidensinterval for d

$(1 - \alpha) = 0.95$ konfidensinterval:

$[0.08, 0.41]$

Hypotesetest

$H_0 : d = 0$

$H_1 : d \neq 0$

Med signifikansnivå $\alpha = 0.05$

Forkastar H_0 .

Påstand som skal testast: Ingelin er høgare enn gj.snittlege NTNU kvinne.

μ : gj.snitt for NTNU kvinner.

Hypoteser

- $H_0: \mu = \mu_0 = 172$
- $H_1: \mu < \mu_0$

Moglege beslutningar

- Forkastar H_0 , og aksepterer H_1 .
Påstand 'bevist' ved data.
- Forkaster ikkje H_0 .
Data underbygger ikkje påstand.

- μ_1 : Gj.snitt kvalitet på fisk, tradisjonell lagring.
- μ_2 : Gj.snitt kvalitet på fisk, ny lagring.

Ny metode bedre kvalitet?

Hypoteser

- $H_0: \mu_1 = \mu_2$
- $H_1: \mu_2 > \mu_1$

Moglege beslutningar

- Forkastar H_0 , og aksepterer H_1 .
Det er ulik kvalitet. Innfører ny lagringsmetode.
- Forkaster ikkje H_0 .
Kan ikkje bevisе at $\mu_2 > \mu_1$. Fortsett med eksisterande lagring.

Høgdehypotese, kjent varians (eit utval kjent varians)

Høgde kvinnlege NTNU stud: $X_i \sim N(\mu, 6.0^2)$, $n = 97$, $\alpha = 0.05$

Hypoteser

- $H_0: \mu = \mu_0 = 172$
- $H_1: \mu < \mu_0$

Forkastningsområde

- Testobservator: $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$
- Under H_0 : $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$
 $P(Z < -z_\alpha) = \alpha$
- Forkaster H_0 dersom $z_{obs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z_\alpha = -1.645$
- Forkaster H_0 dersom $\bar{x} < \bar{x}_{grense} = \mu_0 - z_\alpha \sigma / \sqrt{n} = 168.5$

Frå data: $\bar{x} = 169.1$, dvs $z_{obs} = -5.77 \Rightarrow$ forkastar $H_0 \Rightarrow H_1$

Fiskehypotese, kjent varians (to utval, kjent varians)

Metode 1: $X_1 \sim N(\mu_1, \sigma_1^2)$, n_1 data

Metode 2: $X_2 \sim N(\mu_2, \sigma_2^2)$, n_2 data

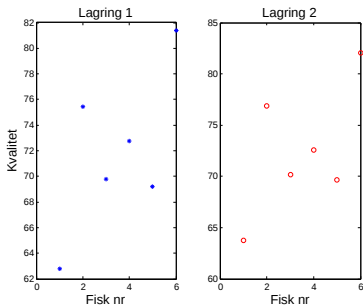
Hypoteser

- $H_0: \mu_1 - \mu_2 = 0$
- $H_1: \mu_1 - \mu_2 < 0$

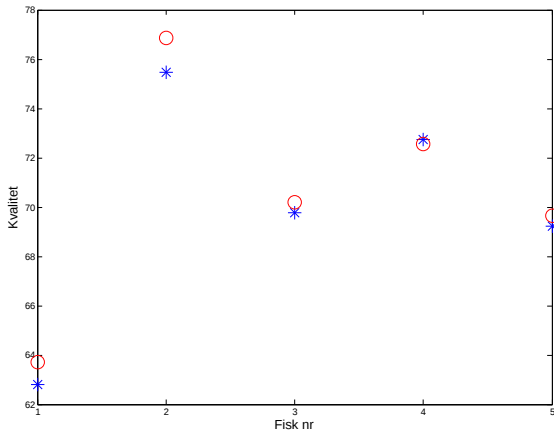
Forkastningsområde

- Testobservator: $Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1)$
- Under H_0 : $Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1)$
 $P(Z < -z_\alpha) = \alpha$
- Forkaster H_0 dersom $z_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} < -z_\alpha$

Lagring av fisk



Lagring av fisk, differansar



Konfidensintervall vs hypotesetesting

Eit $(1 - \alpha)$ KI for μ $[a, b]$

Hypotesetest $H_0 : \mu = \mu_0$ få α nivå.

Forkastar H_0 dersom μ_0 ikkje er i KI.

Ser på normalfordelte gjennomsnitt (pga SGT eller fordi data frå normalfordeling)

- Eit utval, kjent varians. Observator: Z
- Eit utval, ukjent varians, norm.ford. data. Observator: T
- To utval, kjent varians, norm.ford. data. Observator: Z
- To utval, lik ukjent varians Observator, norm.ford. data.: T med S_{pooled}
- To utval ukjent varians, norm.ford. data. Observator: T
- Para utval, norm.ford. data. Observator: T