TMA4240 F18
21.10.2016

Ett utvalg : intervall for μ [9.4]

Populasjon: kvinnelige studenter ved NTNU

Spørsmål: Hva er gjennomsnittshøyde? μ

Utvalg: $n = 149$ kvinner som svarte på spørreund. i TMA4240. Gjsn. høyde var 169.5 cm.

a) Formelt:

$X_1, X_2, \ldots, X_n$ u.i.f $f(x)$

$E(X_i) = \mu$, $Var(X_i) = \sigma^2 = (6.6)^2$ kjent

b) Estimator for μ: $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$

Vi observerer $\hat{\mu} = 169.5$ cm

c) Hvilken fordeling har $\bar{X}$?

Hvis vi antar at $X_1, \dots, X_n$ u.i.f $N(\mu, \sigma^2)$ vil også $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

Sjekk normplott(data) ⇒ ser ok ut for våre data.

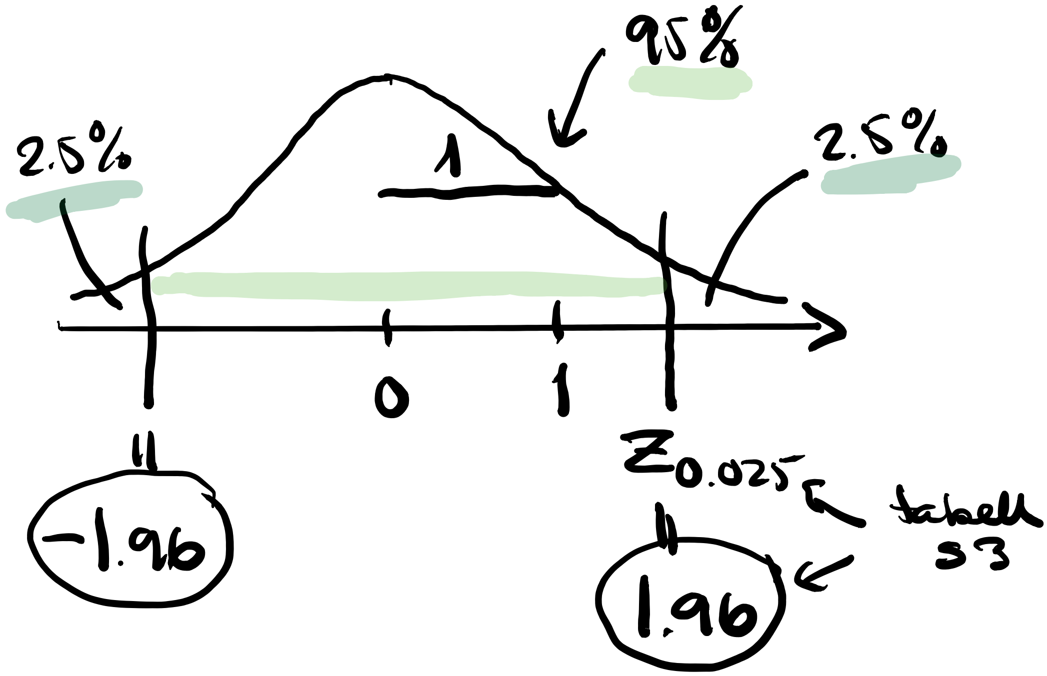
Hvis $n \geq 30$ og vi ikke antar noe om fordelingen til X_i'ene vil sentralgrenseteoremet gi oss at

$$Z = \frac{\bar{X} - E(\bar{X})}{SD(\bar{X})} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

er tilnærmet standardnormal $N(0,1)$

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

d) Intervall for $Z \sim N(0,1)$



$$P(-1.96 < Z < 1.96) = 0.95$$

e) Intervall for μ

kombiner d) med at $Z = \dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

$$P\left(-1.96 < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$$

og gjøre matematiske manipulasjoner for å få μ alene i midten.

$\cdot \frac{\sigma}{\sqrt{n}}$

$$P\left(-1.96 \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$-\bar{X}$

$$P\left(-1.96 \cdot \frac{\sigma}{\sqrt{n}} - \bar{X} < -\mu < 1.96 \cdot \frac{\sigma}{\sqrt{n}} - \bar{X}\right) = 0.95$$

$\cdot(-1)$

$$P\Big(\underbrace{\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}}_{\hat{\mu}_L \text{ lower}} < \mu < \underbrace{\bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}}_{\hat{\mu}_U \text{ upper}}\Big) = 0.95$$

Intervallet $[\hat{\mu}_L, \hat{\mu}_U]$ innsatt verdier for $\bar{x}$, σ, n kalles 95% konfidensintervall for μ. Vi sier at vi har 95% tillit (confidence) til at den sanne μ ligger i dette intervallet.

EKS: 95% konfidensintervall (KI) for forventet høyde kvinner NTNU basert på utvalget vårt:

$\bar{x} = 169.5$ cm

$\sigma = 6.6$ cm

$n = 149$

$z_{0.025} = 1.96$

$$\hat{\mu}_L = \bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$= 169.5 - \underbrace{1.96 \cdot \frac{6.6}{\sqrt{149}}}_{1.06}$$

$$= 168.4 \text{ cm}$$

$$\hat{\mu}_U = \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} = 169.5 + 1.06$$

$$= 170.6 \text{ cm}$$

$\Rightarrow$ 95% konfidensintervall

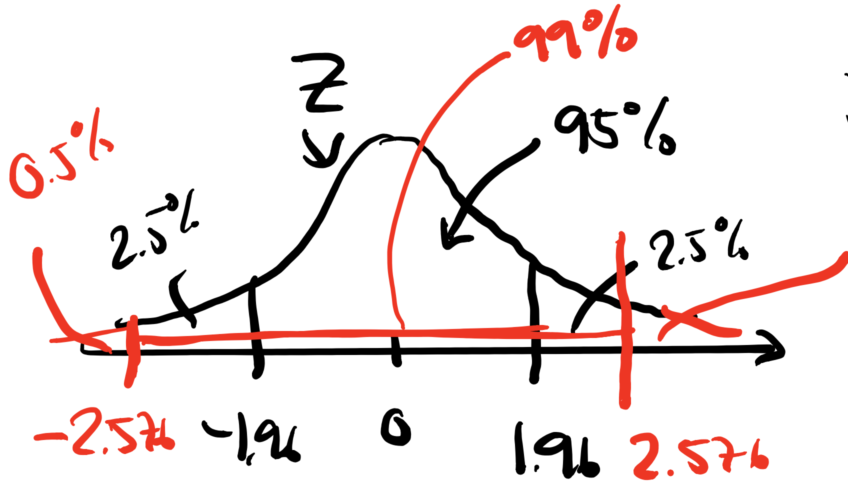
$$[168.4, 170.6] \text{ cm}$$

f) Hvordan lage et 99% KI for μ?

Er et 99% KI bredere enn et 95%?

$\Downarrow$

ja, 99% bredere enn 95%



99% → 0.5% oppe og nede

$$z_{0.005} = 2.576$$

99% KI:

$$\hat{\mu}_L = \bar{x} - \overset{2.576}{\overset{\|}{z_{0.005}}} \cdot \frac{\sigma}{\sqrt{n}}$$

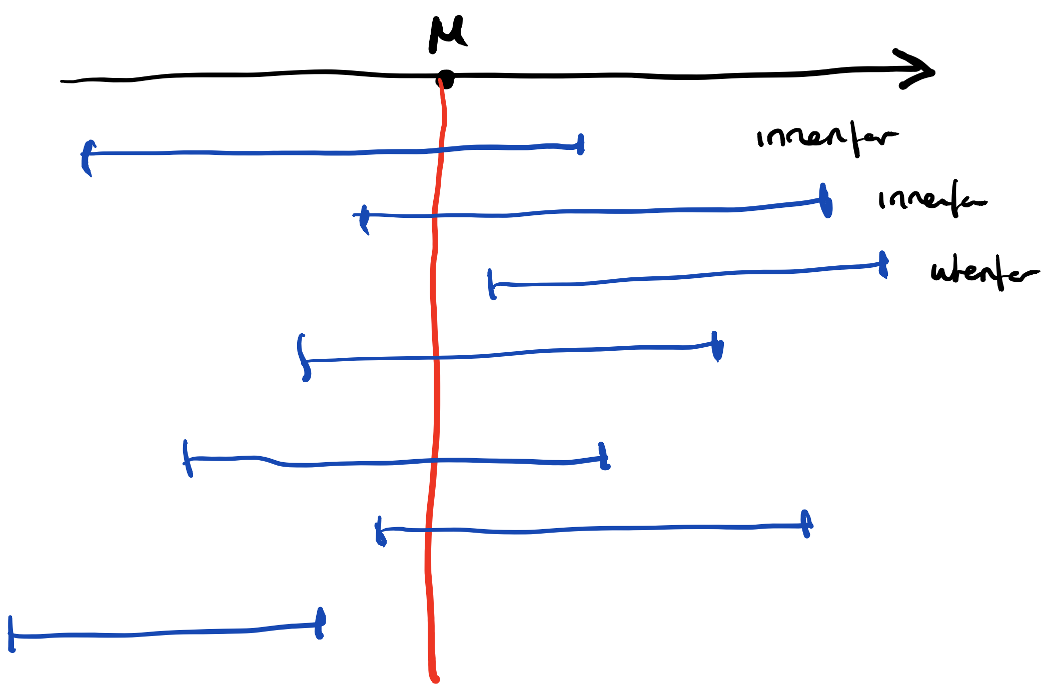
tall: $169.5 - \overbrace{2.576 \cdot \frac{6.6}{\sqrt{49}}}^{1.37} = \underline{\underline{168.1}}$

$$\hat{\mu}_U = \bar{x} + z_{0.005} \cdot \frac{\sigma}{\sqrt{n}}$$

tall: $169.5 + 2.576 \cdot \frac{6.6}{\sqrt{49}} = \underline{\underline{170.9}}$

g) Hva betyr 95% tillit?

For et konkret datasett vil den sanne μ enten være innenfor eller utenfor KI. Samler vi inn et nytt tilfeldig utvalg kan vi lage et nytt KI, osv.



Av disse KI vil 95% dekke den sanne μ. Derfor sier vi at vi har 95% tillit til at μ er inneholdt i et 95% KI.

Hva hvis variansen, σ^2, er ukjent? [9.4 + 8.6]

Hvis σ^2 er ukjent, kan vi finne en estimator for σ^2:

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

Vi hadde at

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

Vi erstatter σ med S

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

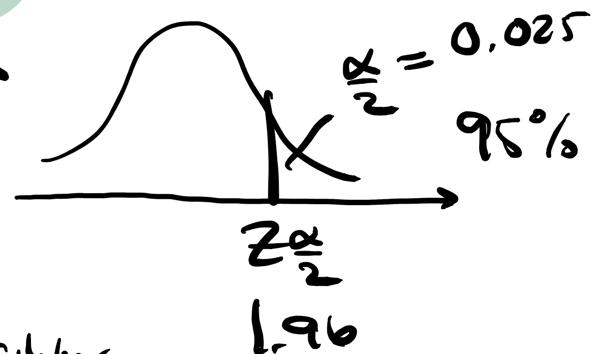
T er Student-t fordelt med parameter $\nu = (n-1)$ frihetsgrader

gresk "ny"

$E(T) = 0$, $Var(T) = \frac{\nu}{\nu - 2}$ for $\nu \geq 3$

Konfidensintervall: $(1-\alpha) \cdot 100\%$

N: $\bar{X} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \Leftarrow \sigma$ kjent



men nå, litt mer usikker
fordi vi estimerer σ; så $z_{\frac{\alpha}{2}}$ byttes med $t_{\frac{\alpha}{2}, n-1}$

T: $\bar{X} \pm t_{\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{n}}$

$n = 10$ obs $\quad t_{0.025,9} = 2.262$

$20 \quad t_{0.025,19} \quad 2.093$

$\vdots \qquad\qquad \vdots$

$149 \quad t_{0.025,148} = 1.976$

$\vdots$

$\infty \quad t_{0.025,\infty} = z_{0.025} = 1{,}96$

tabell

<u>EKS</u>: kvinnelige studenter:

95%, σ ukjent

$n = 149$

$\bar{x} = 169.5$

$s = 6.6$ (estimat)

$t_{0.025,148} = 1{,}976$

$$\bar{x} \pm 1976 \cdot \frac{6.6}{\sqrt{149}}$$

$$= \underline{\underline{[168.4, 170.6]}}$$

s regnet ut fra $\frac{1}{148} \sum_{i=1}^{149} (x_i - \bar{x})^2$

$\bar{x} = 169.5$

ved å bruke det

tilfeldige utvalget vårt.

Sammenligne N og t-intervall for vårt eksempel:

$\boxed{\text{sagt på forelesn. men skrevet på notes etter forelesningen}}$

1) Det var juks at vi visste hva σ var. Jeg hadde egentlig satt inn den s jeg regnet ut fra dataene. Grunnen til det var at jeg ville ha fokus på de andre tallene som endret seg.

2) t-fordelingen har tyngre haler enn $N(0,1)$ og derfor må vi lengre fra 0 for å finne tallet $t_{\frac{\alpha}{2},n-1}$ som har areal $\frac{\alpha}{2}$ til høyre enn for $z_{\frac{\alpha}{2}}$. Det betyr at $t_{\frac{\alpha}{2},n-1} > z_{\frac{\alpha}{2}}$ som vil gi et bredere intervall for t- enn N-intervallet

3) For oss var $t_{0.025,148} = 1.976$ og $z_{0.025} = 1.96$ $\Rightarrow$ ganske like 95% KI.

N 95% $[168.44, 170.56]$

t 95% $[168.43, 170.57]$

95-99, $\frac{\alpha}{2}$ og $(1-\alpha)\cdot 100\%$

$\underbrace{(1-\alpha)\cdot 100}\%$ KI

$95 \Leftrightarrow \quad \alpha = 0.05 \Leftrightarrow \frac{\alpha}{2} = 0.025$

$99 \Leftrightarrow \quad \alpha = 0.01 \Leftrightarrow \frac{\alpha}{2} = 0.005$