Plan

1. Konfidensintervall for μ når σ er kjent
2. Student t-fordeling
3. Konfidensintervall for μ når σ er ukjent

Repetisjon:

a) Estimatorer: Stokastisk variabel som brukes til å estimere en ukjent parameter.

Antar: $X_1, X_2, \ldots, X_n$ er uavhengige, identisk fordelte stokastiske variable med

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

Eks: n uavhengige målinger / trekninger

For μ: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$

$E(\bar{X}) = \mu$ forv. rett

$Var(\bar{X}) = \sigma^2/n$

$SE(\bar{X}) = \sigma/\sqrt{n}$ standard error = std. avvik

For σ^2: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$

$E(S^2) = \sigma^2$ forv. rett

Antar: $X_1, X_2, \ldots, X_n$ er uavhengige binomisk fordelte stok. variable med $n=1$, p slik at $X = X_1 + X_2 + \ldots + X_n$ er binomisk med parametre (n, p).

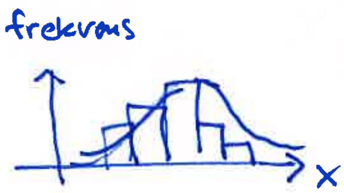
For p: $\hat{p} = \frac{X}{n}$

$E(\hat{p}) = p$ forv. rett

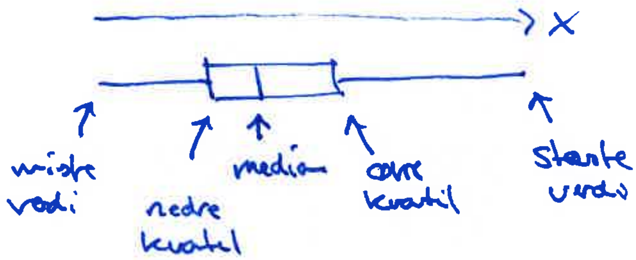
$Var(\hat{p}) = \frac{pq}{n}$

$SE(\hat{p}) = \sqrt{\frac{pq}{n}}$

b) Beskrivelse av datasett:



- frekvenstabell / histogram
  (sannsynlighetsfordeling)



- boksplott
  (sentralmål / spredningsmål)

- variasjonskoeffisient: $\frac{s_x}{\bar{x}}$

Eivind Eriksen, Kontor B4-032, eivind.eriksen@bi.no
http://www.dr-eriksen.no/teaching/MET1190/

① Konfidensintervall for μ når σ er kjent

Eks:

171 184 169 175
187 190 168 176

$\leftarrow$ $x_1\ x_2\ x_3\ x_4$
$x_5\ x_6\ x_7\ x_8$

Estimator:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{8}(X_1 + X_2 + \dots + X_8)$$

$$E(\bar{X}) = \mu$$

$$SE(\bar{X}) = \sigma/\sqrt{n} = \sigma/\sqrt{8}$$

Punktestimat:

$$\bar{x} = \frac{171 + 184 + \dots + 176}{8} = 177.5$$

Konfidensintervall: Velger konfidensnivå α
$\rightarrow$ lager $(1-\alpha)\cdot 100\%$ konfidensintervall

$\alpha = 0.05 \rightarrow 95\%$ konfidensintervall

Defn:

$$\boxed{\text{Et } (1-\alpha)\cdot 100\% \text{ konfidensintervall for } \mu \text{ er et intervall } [A, B] \text{ slik at } P(A \le \mu \le B) = 1-\alpha}$$

Merk: A, B er stokastiske variabler
μ er den sanne, ukjente verdien til en parameter

$\cancel{P(-1 \le Z \le 1)}$

$P(z_1 \le 3 \le z_2)$

Forutsetninger: Konfidensintervall for μ når σ er kjent

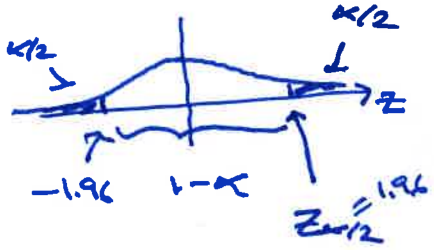
- De stokastiske variablene $X_1, X_2, \ldots, X_n$ er uavhengige og identisk fordelte med $E(X_i) = \mu$ og $Var(X_i) = \sigma^2$.
- Estimatoren $\bar{X}$ er tilnærmet normalfordelt (ok for $n \geq 30$ eller når hver X_i er normalt.)

Metode for å finne konfidensintervallet:

$\boxed{E(\bar{X}) = \mu \quad SE(\bar{X}) = \sigma/\sqrt{n}}$

1. Velg α.
2. Se på $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

Standard normalfordelt



$P(A \leq \mu \leq B) = 1 - \alpha \leftarrow$ $\boxed{\text{defn. av konf. intervall}}$

$P(A - \bar{X} \leq \mu - \bar{X} \leq B - \bar{X}) = 1 - \alpha$

$P(\bar{X} - A \geq \bar{X} - \mu \geq \bar{X} - B) = 1 - \alpha$

$P\left(\dfrac{\bar{X}-A}{\sigma/\sqrt{n}} \geq \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \geq \dfrac{\bar{X}-B}{\sigma/\sqrt{n}}\right) = 1 - \alpha$

Husk: $Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

Eks: $\alpha = 0.05$
$\alpha/2 = 0.025$

$z_{\alpha/2} \approx 1.96$

(0.975 $\boxed{\text{INV}}$ $\boxed{Z \geq P}$)

$P\left(\dfrac{\bar{X}-B}{\sigma/\sqrt{n}} \leq Z \leq \dfrac{\bar{X}-A}{\sigma/\sqrt{n}}\right) = 1 - \alpha$

$\Downarrow$ $\alpha = 0.05$.

$\dfrac{\bar{X}-B}{\sigma/\sqrt{n}} = -1.96 = -z_{\alpha/2}$ $\qquad$ $\dfrac{\bar{X}-A}{\sigma/\sqrt{n}} = 1.96 = z_{\alpha/2}$

$\bar{X} - B = -z_{\alpha/2} \cdot \sigma/\sqrt{n}$ $\qquad$ $\bar{X} - A = z_{\alpha/2} \cdot \sigma/\sqrt{n}$

$$B - \bar{x} = z_{\alpha/2} \cdot \sigma/\sqrt{n} \qquad A - \bar{x} = -z_{\alpha/2} \cdot \sigma/\sqrt{n}$$

$$B = \bar{x} + z_{\alpha/2} \cdot \sigma/\sqrt{n} \qquad A = \bar{x} - z_{\alpha/2} \cdot \sigma/\sqrt{n}$$

$\Downarrow$

Konfidensintervall: $[A, B] = [\bar{x} - z_{\alpha/2} \cdot \sigma/\sqrt{n}, \bar{x} + z_{\alpha/2} \cdot \sigma/\sqrt{n}]$

Konklusjon:

Konfidensintervallet for μ når σ er kjent er

$$[A, B] = [\bar{x} - z_{\alpha/2} \cdot \sigma/\sqrt{n}, \bar{x} + z_{\alpha/2} \cdot \sigma/\sqrt{n}]$$

når konfidensnivået er α og $\bar{x}$ er normalfordelt.

Eks: Anta datasettet fra tidligere i forelesningen, ~~med~~ med $n = 8$, $\bar{x} = 177.5$ kommer fra en normalfordelt variabel.

Anta at $\sigma = 8$.

$\alpha = 0.05$: $\qquad z_{\alpha/2} = 1.96$

95% konfidensintervall

$z_{\alpha/2}$ $\qquad$ $\bar{x}$ Punkt-estimat $\qquad$ $SE(\bar{x})$

$$[A, B] = \left[177.5 - 1.96 \cdot \frac{8}{\sqrt{8}}, 177.5 + 1.96 \cdot \frac{8}{\sqrt{8}}\right]$$

$$\cong [177.5 - 5.5, 177.5 + 5.5]$$

$$= [172, 183]$$

95% sanns. at μ er i intervallet

Bredden på konfidensintervallet er bestemt av:

- Jo større α er, jo mindre blir bredden på konfidensintervallet
- Jo større standardfeilen til estimatoren er, jo større blir bredden på konfidensintervallet

Husk: $\bar{X} \pm z_{\alpha/2} \cdot SE(\bar{X})$,

$SE(\bar{X}) = \sigma/\sqrt{n}$

Bredden til konfidensintervallet: $2 \cdot z_{\alpha/2} \cdot SE(\bar{X})$

Oppsummering: $\boxed{\text{Konfidensintervall for } \mu \text{ når } \sigma \text{ er kjent}}$

Forutsetninger: ① $X_1, \ldots, X_n$ uavh. stokastiske variable som er identisk fordelt med $E(X_i) = \sigma$, $Var(X_i) = \sigma^2$
② $n \geq 30$ eller hver X_i normalfordelt, slik at $\bar{X}$ er tilnærmet normalfordelt

Input: Konfidensnivå: α → $(1-\alpha) \cdot 100\%$ konfidensintervall

Konfidensintervall: $\Big[\underbrace{\bar{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n}}_{A},\ \underbrace{\bar{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n}}_{B}\Big]$

② Konfidensintervall for μ når σ er ukjent

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$E(\bar{X}) = \mu$

$SE(\bar{X}) = \sigma/\sqrt{n} \leftarrow$ ukjent

↓

Må bruke

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

til å estimere σ^2.

Konfidensintervall (konfidensnivå α): $\left[\bar{x} - t^{n-1}_{\alpha/2}\cdot s/\sqrt{n},\ \bar{x} + t^{n-1}_{\alpha/2}\cdot s/\sqrt{n}\right]$

$t^{n-1}_{\alpha/2}$: Kvantil i t-fordeling (med n-1 frihetsgr.)

$s = s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

utvalgs-std.avvik til x i datasettet

Eks:

171	184	169	175
187	190	168	176

$\leftarrow$ datasett for x ($n=8$)

Anta σ ukjent

$\bar{x} = 177.5$

$s_x \cong 8.47$

$t^{n-1}_{\alpha/2}$: $n-1 = 8-1 = 7$

velger $\alpha = 0.05$

$t^{7}_{0.025} \simeq 2.36$

$\boxed{df, t \rightleftarrows p}$ t-fordeling

7 $\boxed{INV}$ $\boxed{df, t \rightleftarrows p}$ 0.975 $\boxed{=}$ $\rightarrow$ 2.36

$$\left[\bar{x} - t^{n-1}_{\alpha/2}\cdot s/\sqrt{n},\ \bar{x} + t^{n-1}_{\alpha/2}\cdot s/\sqrt{n}\right] = \left[177.5 \pm 2.36\cdot\frac{8.47}{\sqrt{8}}\right]$$
$$= [177.5 - 7.1,\ 177.5 + 7.1]$$
$$= [170.4,\ 184.6]$$

③ T-fordeling

χ^2_k : chi-kvadrat fordeling med k frihetsgrader

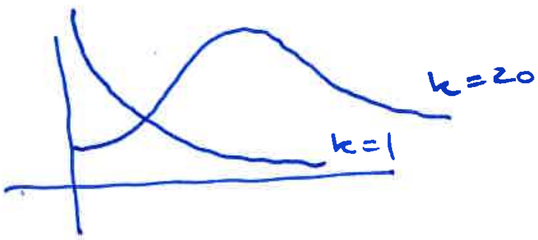
$Z_1, Z_2, \ldots, Z_k$: uavhengige, std. normalfordelte var.

$\Downarrow$

$$\boxed{\chi^2_k = Z_1^2 + Z_2^2 + \ldots + Z_k^2}$$

defn. på χ^2_k-fordeling

$\chi^2_k \geq 0$



$E(\chi^2_k) = k$

$Var(\chi^2_k) = 2k$

Resultat: $X_1, X_2, \ldots, X_n$ uavhengige normalfordelte med $E(X_i) = \mu$, $Var(X_i) = \sigma^2$

$\Downarrow$

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

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma^2} = \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \ldots + (x_n-\bar{x})^2}{\sigma^2}$$

$$= \left(\frac{x_1-\bar{x}}{\sigma}\right)^2 + \left(\frac{x_2-\bar{x}}{\sigma}\right)^2 + \ldots + \left(\frac{x_n-\bar{x}}{\sigma}\right)^2$$

Fortsettelse om t-fordeling kommer i neste forelesning.