Plan:

(1) Beskrivelse av datasettet

(2) Konfidens intervall

Pensum:

[L] 2.1-2.5, 6.3

Repetisjon:

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

Vi antar at $X_1, X_2, \ldots, X_n$ er uavhengige, identisk distribuerte stokastiske variabler med ukjente parametre $\mu = E(X_i)$, $\sigma^2 = Var(X_i)$

Eks: n uavhengige målinger / n uavhengige trekninger

a) Estimator for μ:

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$\bar{X}$ forventningsrett estimator

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

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

standardfeil SE = standardavvik
- minst mulig
- går mot 0 når $n \to \infty$

b) Estimator for σ^2:

$$\hat{\sigma^2} = S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n \cdot \bar{X}^2\right)$$

S^2 forventningsrett

$E(S^2) = \sigma^2$

Vi antar at $X_1, X_2, \ldots, X_n$ er uavhengige, identisk distribuerte binomiske variabler med parametre p, $n=1$. Da er $X = X_1 + \ldots + X_n$ binomisk med parametre n, p.

c) Estimator for p:

$$\hat{p} = \frac{X}{n} = \frac{X_1 + \ldots + X_n}{n}$$

$\hat{p}$ forventningsrett

$E(\hat{p}) = p$

$Var(\hat{p}) = \frac{pq}{n} \Rightarrow SE(\hat{p}) = \sqrt{\frac{pq}{n}}$ $(q = 1-p)$

b) Når vi observerer verdiene $x_1, \dots, x_n$ til $X_1, \dots, X_n$ (henter inn et datasett) og setter de inn i en estimator, får vi et punktestimat (verdi).

Estimator	$\longrightarrow$	Estimat (punktestimat)
Stokastisk variabel		tall-verdi
$\bar{X} = \frac{X_1 + \dots + X_n}{n}$		$\bar{x} = \frac{x_1 + \dots + x_n}{n}$

① Beskrivelse av utvalget / datasettet (Kap. 2)

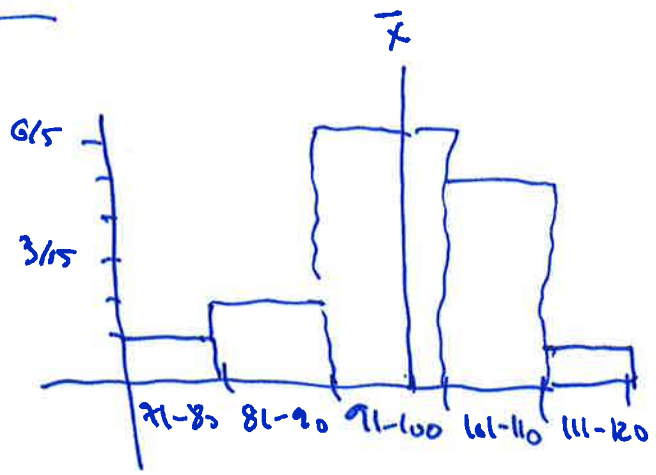
Eks:

104	109	111	109	87
91	103	99	108	96
79	87	94	92	97

$\leftarrow$ $x_1, \dots, x_{15}$ $n = 15$

a) Frekvenstabell / histogram

	antall	relativ frekvens
71-80	1	1/15
81-90	2	2/15
91-100	6	6/15
101-110	5	5/15
111-120	1	1/15
	15	



b) Sentralmål: $\bar{x} = \frac{104 + \dots + 97}{15} = \underline{97.73}$

b) Sentralmål: typisk verdi (sentral verdi)

Gjennomsnitt: $\bar{x} = \frac{x_1 + \dots + x_n}{n} \approx \underline{97.73}$

Median: median = $\underline{97}$

79 87 87 91 92 94 96 97 99 103 104 108 109 109 111

↑ observasjon nr. 8 er median.

Generelt: observasjon $\frac{n+1}{2}$ når observasjonen er skrevet i stigende rekkefølge.

På kalkulator:

Slette statistikk-minne: $\boxed{\text{C STAT}}$

Legge inn data: 104 $\boxed{\Sigma+}$ 109 $\boxed{\Sigma+}$ ....

Gjennomsnitt: $\boxed{\bar{x}\ \bar{y}} \to \underline{97.73}$

c) Kvartiler og boksplott

Nedre kvartil = 25-prosentil = 91

Øvre kvartil = 75-prosentil = 108

Kvartilbredde: $108 - 91 = \underline{17}$

Eks: 79 87 87 91 92 94 96 97 99 103 104 108 109 109 111

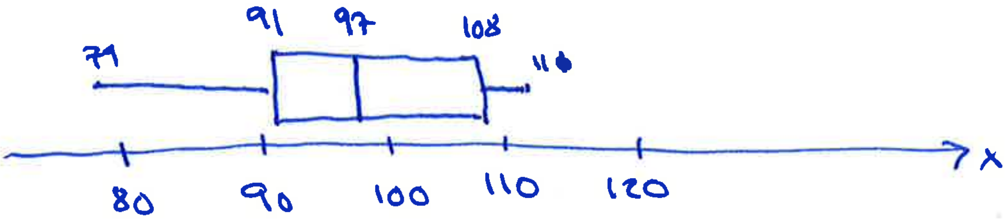
nedre kvartil: nr $\frac{n+1}{4} = 4$

median (nr 8): nr $\frac{n+1}{2}$

øvre kvartil: nr $\frac{3}{4}(n+1) = 12$

Boksplott:

median = 97
nedre kvartil = 91
øvre kvartil = 108
minste = 79
største = 111



d) Variasjonsmål:

Std. avvik: (utvalg)
$$S_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \longleftarrow$$
typisk avvik fra gjennomsnitt.

På kalk: $\boxed{S_x \; S_y} \rightarrow \underline{\underline{9.57}}$ (i Ex.)

Kvartilbredde: øvre kvartil − nedre kvartil $= \underline{\underline{17}}$

Variasjonsbredde: største − minste verdi

② Konfidensintervall

Eks: Vi gjør 8 uavhengige målinger og observerer følgende:

171	184	169	175
187	190	168	176

Vi ønsker å estimere μ:

– punktestimat:

Bruker estimator $\hat{\mu} = \bar{X} = \frac{1}{8}(X_1 + X_2 + \dots + X_8)$

Dette gir estimat: $\bar{x} = \frac{1}{8}(171 + \dots + 176) = \underline{\underline{177{,}5}}$

– konfidensintervall: Vi ønsker å finne et intervall

$[A, B]$

slik $P(A \le \mu \le B) = 0{,}95$

$\alpha = 0.05$

$\Updownarrow$

$(1-\alpha) \cdot 100\%$ konfidens-intervall

〃

95%

Merk: μ er en gitt, men ukjent verdi

A, B er stokastiske variable.

$P(1 \le X \le 2)$
er slike sannsynligheter
vi typisk har regnet
ut tidligere, her er det
motsatt

Konstruksjon av konfidensintervall for μ:

Forutsetninger:

- $X_1, X_2, \ldots, X_n$ er uavhengige, identisk fordelte stokastiske variabler med $E(X_i) = \mu$ og $Var(X_i) = \sigma^2$
- $\bar{X}$ er tilnærmet normalfordelt ($n \geq 30$ eller $X_1, \ldots, X_n$ er normalfordelt)

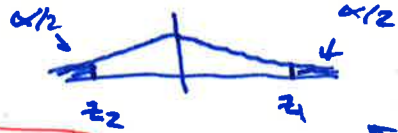
Metode for å finne A og B:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \text{ er std. normalfordelt}$$

$\bar{X}$ normalfordelt

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

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



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

$\|$

$\cancel{P(A \leq \mu}$

$Z_2 = -z_{\alpha/2}$
$= -1{,}96$

$Z_1: \; G(z_1) = 1 - \alpha/2$

$z_1 = z_{\alpha/2}$

$\alpha = 0{,}05$: $z_1 = z_{\alpha/2}$
$= 1{,}96$

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

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

$$P\Bigg(\underbrace{\frac{\bar{x} - A}{\sigma/\sqrt{n}}}_{= z_1 = 1{,}96} \geq \underbrace{\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}_{= Z} \geq \underbrace{\frac{\bar{x} - B}{\sigma/\sqrt{n}}}_{= z_2 = -1{,}96}\Bigg) = 1 - \alpha$$

$$\frac{\bar X - A}{\sigma/\sqrt{n}} = 1{,}96 \qquad \bar X - A = 1.96\cdot \sigma/\sqrt{n}$$
$$A = \underline{\bar X - 1.96\cdot\sigma/\sqrt{n}}$$

$$\frac{\bar X - B}{\sigma/\sqrt{n}} = -1.96 \qquad \bar x - B = -1.96\cdot\sigma/\sqrt{n}$$
$$B = \underline{\bar X + 1.96\cdot\sigma/\sqrt{n}}$$

<u>For $\alpha = 0{,}05$:</u> $[A,B] = [\bar X - 1.96\cdot\frac{\sigma}{\sqrt{n}}, \bar X + 1.96\cdot\frac{\sigma}{\sqrt{n}}]$

<u>For generell α:</u> $[A,B] = [\bar X - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}, \bar X + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}]$

<u>Formel for konfidensintervall for μ når σ er kjent:</u>

Under forutsetningene ovenfor (dvs X_i uavhengige, identisk fordelte stokastiske variabler med $\mu = E(X_i)$ og $\sigma^2 = Var(X_i)$, og $\bar X$ er tilnærmet normalfordelt):

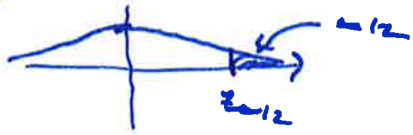
$$\left[\bar X - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}, \bar X + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right]$$

<u>Eks:</u> $n=8$ $\bar X = 177{,}5$

$z_{\alpha/2} = 1{,}96$ når $\alpha = 0{,}05$

Anta at $\underline{\sigma = 8}$.

Std. normalf.



z slik at $G(z) = 1 - \alpha/2$

95% konfidensintervall for μ: $\left[177.5 - 1.96\cdot\frac{8}{\sqrt{8}}, 177{,}5 + 1.96\cdot\frac{8}{\sqrt{8}}\right]$

$$= \underline{[171{,}96, 183.04]}$$

Oppsummering:

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

Forutsetninger:

i) $X_1, \ldots, X_n$ uavhengige stokastiske variabler med identisk fordeling, med $E(X_i) = \mu$ og $Var(X_i) = \sigma^2$

ii) $n \geq 30$ eller hver X_i normalfordelt slik at $\bar{X}$ er tilnærmet normalfordelt

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

$\Downarrow$

Konfidensintervall: $\left[\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} , \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right]$