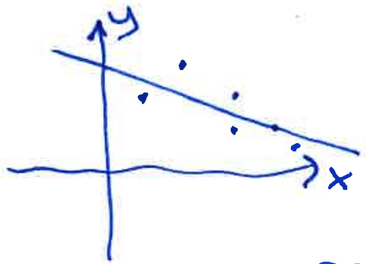
Plan

1. Estimatorer.
2. Beskrivelse av datasett.

Repetisjon:

i	x_i	y_i
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3
$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n

Spredningsdiagram:



Regresjonslinjen: passer best mulig med dataene

$y = \alpha + \beta x$ der

$$\begin{cases} \beta = r_{xy} \cdot s_y / s_x \\ \alpha = \bar{y} - \beta \cdot \bar{x} \end{cases}$$

Sentralmål: "typisk verdi"

* utvalgsgjennomsnitt

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

* median $x_{(1)} \le x_{(2)} \le \dots \le x_{(n)}$

median:
$$\begin{cases} x_{(\frac{n+1}{2})} & n \text{ odde} \\ \dfrac{x_{(n/2)} + x_{(n/2+1)}}{2} & n \text{ partall} \end{cases}$$

Spredningsmål: "variasjon"

* utvalgs varians / std. avvik

$$s_x^2 = \frac{1}{n-1} \cdot \left[(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \right]$$
$$= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$s_x = \sqrt{s_x^2}$$

Sammenhengen mellom x og y:

* utvalgskovarians / utvalgs korrelasjon

$$s_{xy} = \frac{1}{n-1} \left[(x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y}) \right]$$
$$= \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$$

$-1 \le r_{xy} \le 1$

$r_{xy} > 0$: pos. samvh.

$r_{xy} < 0$: neg. samvh.

Oppgavesett 8:

8.) $\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$

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

Alt 1: $\sum_{i=1}^{n} \cdot (x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} 2x_i\bar{x} + \sum_{i=1}^{n} \bar{x}^2$

$= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n \cdot \bar{x}^2$

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

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

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

$= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$ ✓

Alt 2: $\sum_{i=1}^{n} (x_i - \bar{x})^2 = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2$

$= (x_1^2 - 2x_1\bar{x} + \bar{x}^2) + (x_2^2 - 2x_2\bar{x} + \bar{x}^2) + \dots + (x_n^2 - 2x_n\bar{x} + \bar{x}^2)$

$= (x_1^2 + x_2^2 + \dots + x_n^2) - 2\bar{x}(x_1 + x_2 + \dots + x_n) + n \cdot \bar{x}^2$

$= (x_1^2 + x_2^2 + \dots + x_n^2) - 2\bar{x}(\, n \cdot \bar{x} \,) + n\bar{x}^2$

$= x_1^2 + x_2^2 + \dots + x_n^2 - n\bar{x}^2$

9) $\frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 \geq 0 \Rightarrow \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) \geq 0 \quad | \cdot (n-1)$

$\Rightarrow \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \geq 0 \Rightarrow \sum_{i=1}^{n} x_i^2 \geq n\bar{x}^2 = n \cdot \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$

$= \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$

1 Estimatorer

Populasjon: Alle BI-studenter

Variabel: X { Månedlig kostnader til en tilfeldig valgt BI-student

X : normalfordelt med $E(X) = \mu$, $Var(X) = \sigma^2$

Tilfeldig utvalg: n BI studenter

$X_1, X_2, \ldots, X_n$: trukket uavhengig fra fordelingen $N(\mu, \sigma^2)$

Datasett:

$x_1, x_2, \ldots, x_n$

Estimator: stokastisk variabel som brukes til å estimere en ukjent parameter

Eks: $\hat{\mu} = \dfrac{X_1 + X_2 + \ldots + X_n}{n} = \bar{X}$

Krav til estimatoren:

1) Forventningsrett: $E(\bar{X}) = \mu$

2) $Var(\bar{X})$ minst mulig

3) $Var(\bar{X}) \to 0$ når $n \to \infty$

$\hat{\theta}$: Estimator for en parameter θ

Eks: $\theta = \mu$
$\hat{\theta} = \bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$

i) $E(\hat{\theta}) = \theta$
ii) $Var(\hat{\theta})$ minst mulig
iii) $Var(\hat{\theta}) \to 0$ når $n \to \infty$

Eks: $X_1, X_2, \dots, X_n$ trekkes fra $N(\mu, \sigma^2)$ uavhengige

Velger $\hat{\mu} = \bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$

Sjekk:

i) $E(\bar{X}) = E\left(\frac{1}{n}(X_1 + \dots + X_n)\right) = \frac{1}{n}E(X_1 + X_2 + \dots + X_n)$
$= \frac{1}{n}\left(E(X_1) + E(X_2) + \dots + E(X_n)\right)$
$= \frac{1}{n}(\mu + \mu + \dots + \mu) = \frac{1}{n} \cdot n\mu = \underline{\mu}$ ok

ii) $Var(\bar{X}) = Var\left(\frac{1}{n}(X_1 + \dots + X_n)\right)$
$= \frac{1}{n^2} \cdot Var(X_1 + \dots + X_n) = \frac{1}{n^2}\left(Var(X_1) + \dots + Var(X_n)\right)$
$= \frac{1}{n^2} \cdot n\sigma^2 = \underline{\sigma^2/n}$ ok

iii) $Var(\bar{X}) = \sigma^2/n \to 0$ når $n \to \infty$ ok

$X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$
uavhengige

Estimator for σ^2:

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

Stokastisk variabel
estimator for σ^2

Forventningsrett:

$$E(S^2) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right] = \frac{1}{n-1} E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]$$

$$= \frac{1}{n-1}\cdot E\left[(X_1-\bar{X})^2 + (X_2-\bar{X})^2 + \ldots + (X_n-\bar{X})^2\right]$$

$$= \frac{1}{n-1}\left[E\left[(X_1-\bar{X})^2\right] + E\left[(X_2-\bar{X})^2\right] + \ldots + E\left[(X_n-\bar{X})^2\right]\right]$$

$$= \frac{1}{n-1}\cdot\left[E(X_1^2 - 2\bar{X}X_1 + \bar{X}^2) + \ldots + E(X_n^2 - 2X_n\bar{X} + \bar{X}^2)\right]$$

$$= \frac{1}{n-1}\cdot\left[E(X_1^2) + E(X_2^2) + \ldots + E(X_n^2) - E(n\bar{X}^2)\right]$$

$$= \frac{1}{n-1}\cdot\left[n(\sigma^2+\mu^2) - n\,E(\bar{X}^2)\right] = \frac{1}{n-1}\left[n(\sigma^2+\cancel{\mu^2}) - n\left(\frac{\sigma^2}{n}+\cancel{\mu^2}\right)\right]$$

$$= \frac{1}{n-1}(n\sigma^2 - \sigma^2) = \cancel{\frac{1}{n-1}}\,\cancel{(n-1)}\,\sigma^2 = \sigma^2 \quad \checkmark$$

③ $E(\bar{X}^2) - E(\bar{X})^2 = Var(\bar{X}) = \sigma^2/n \Rightarrow E(\bar{X}^2) = \sigma^2/n + \mu^2$

① $-2X_1\bar{X} - 2X_2\bar{X} \ldots - 2X_n\bar{X} + n\cdot\bar{X}^2$

$= -2\bar{X}(X_1+X_2+\ldots+X_n) + n\cdot\bar{X}^2 = -2\bar{X}\cdot n\bar{X} + n\bar{X}^2$

$= -2n\bar{X}^2 + n\bar{X}^2 = -n\bar{X}^2$

② $E(X_i^2) - E(X_i)^2 = Var(X_i) = \sigma^2 \Rightarrow E(X_i^2) = \sigma^2 + \mu^2$

Estimering av p ved binomisk fordeling:

$X \sim \text{Binom}(n, p)$: $\hat{p} = \frac{X}{n}$

i) $E(\hat{p}) = E(\frac{1}{n} \cdot X) = \frac{1}{\cancel{n}} \cdot \cancel{n}p = p$ ✓ ok.

ii) $Var(\hat{p}) = Var(\frac{1}{n} X) = \frac{1}{n^{\cancel{2}}} \cdot \cancel{n}pq$

$= \frac{pq}{n} = \frac{p(1-p)}{n} \to 0$ når $n \to \infty$ ✓

2 Beskrivelse av datasett.

a) Frekvenstabell / histogram.

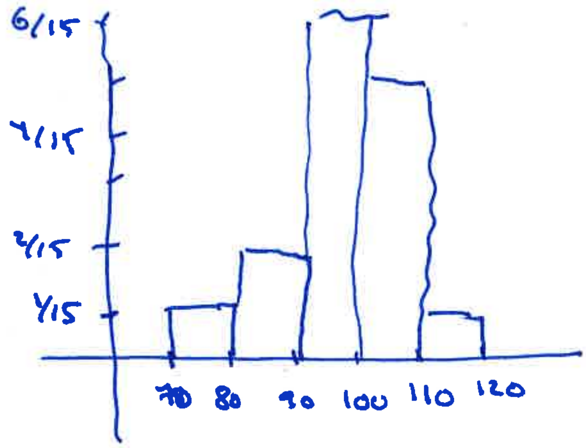
Eks: Datasett

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

$n = 15$

Frekvenstabell: Minste verdi 79
Største " 111

	antall	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	1 = 100%

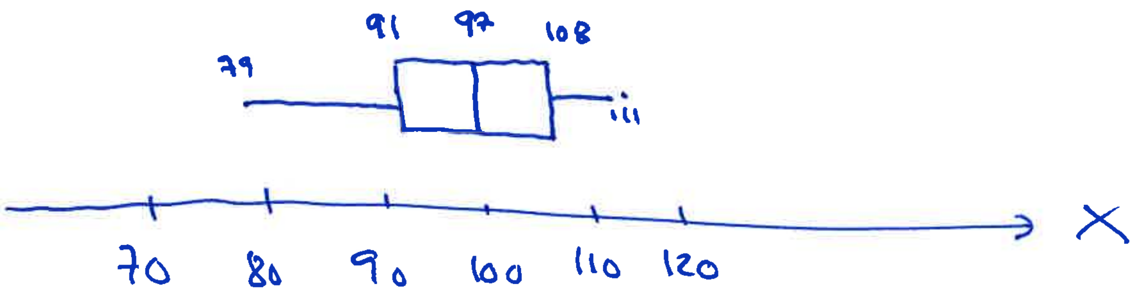


Boksplott: 79 87 87 91 92 94 96 97 99
103 104 108 109 109 111

Median: $x_{(\frac{n+1}{2})} = x_{(8)} = 97$

Nedre kvartil: $x_{(\frac{n+1}{4})} = x_{(4)} = 91$

Øvre kvartil: $x_{(\frac{3\cdot(n+1)}{4})} = x_{(12)} = 108$



Sentralmål: $\bar{x}$ gj.snitt
median
modus = den verdien med størst frekvens

Spredningsmål: s_x std.avvik / s_x^2 varians
Kvartilbredde = øvre kvartil − nedre kvartil
Variasjonsbredde = største − minste verdi