Plan

Repetisjon:

i) Sentralgrenseteoremet:

$X_1, X_2, X_3, \ldots, X_n$: uavhengige, identisk fordelte stokastiske variable, med $E(X_i) = \mu$ og $Var(X_i) = \sigma^2$

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n) \to N(\mu, \sigma^2/n)$$

når $n \to \infty$

* Gode god tilnærming : $n \geq 30$
* Hvis hver X_i er normalfordelt, så er $\bar{X}$ normalfordelt for alle n.

ii) Normaltilnærming til binomisk fordeling:

$X \sim Binom(n, p) \rightsquigarrow X \simeq Y \sim N(\mu = np, \sigma^2 = npq)$

ganske god tilnærming hvis n er stor

Best tilnærming med heltallskorreksjon:

$$\boxed{P(a \leq X \leq b) \simeq P(a - 0{,}5 \leq Y \leq b + 0{,}5)}$$

Oppgavesett 7, Oppg 7:

X_1 = første kast
X_2 = andre "
⋮
X_{79} = kast nr 79

X_1 :

x	1	2	3	4	5	6
$p(x)$	1/6	1/6	1/6	1/6	1/6	1/6

$$\mu_1 = E(X_1) = 1\cdot 1/6 + \dots + 6 \cdot 1/6 = \frac{21}{6} = \underline{\underline{\frac{7}{2}}}$$

$$\sigma_1^2 = Var(X_1) = E(X_1^2) - (7/2)^2$$
$$= 1^2 \cdot \frac{1}{6} + \dots + 6^2 \cdot \frac{1}{6} - (7/2)^2$$
$$= \frac{91}{6} - \frac{49}{4} = \underline{\underline{\frac{35}{12}}}$$

$X = X_1 + X_2 + \dots + X_{79}$

uavhengige,
identisk fordelte var.

⇓

$\boxed{X \text{ tilnærmet normalfordelt ved sentralgrenseteoremet}}$

X : $\underbrace{E(X)}_{\mu} = 79 \cdot \mu_1 = 79 \cdot 7/2$

$\mu = \underline{\underline{276.5}}$

$\sigma^2 = Var(X) = 79 \cdot \sigma_1^2 \simeq 230.4$

$\Rightarrow \sigma \simeq \underline{\underline{15.2}}$

Skal finne:

$$p(X \le 300)$$
$$\simeq p(Y \le 300.5)$$
$$= p\left(Z \le \frac{300.5 - 276.5}{15.2}\right)$$
$$= G(1.58) \simeq \underline{\underline{0.943}}$$

$X \le 300$: trenger minst 80 kast for å komme over 300

$Y \sim N(\mu, \sigma^2)$, der $\mu = 276.5$, $\sigma = 15.2$

normaltilnærming m/ heltallskorreksjon

1 Populasjon og utvalg

Eks: Månedlige kostnader for BI-studenter

Populasjon: alle BI-studenter

Utvalg: de BI-studentene man henter inn data fra.

Merk: – utvalgsstørrelse n=20
– tilfeldig utvalg

Datasett:

1 Hva vi kan si om dataene i datasettet
Deskriptiv statistikk

2 Hva vi kan si om hele populasjonen basert på datasettet og utvalget
Inferens

(2) Enkel data-analyse

En variabel:

i	x_i
1	x_1
2	x_2
3	x_3
⋮	⋮
n	x_n

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$ utvalgs-gjennomsnitt

$$s_x^2 = \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \dots + (x_n-\bar{x})^2}{n-1}$$ utvalgs-varians

$$s_x = \sqrt{s_x^2}$$ utvalgs-standardavvik

Eks:

i	x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
1	23	−21	441
2	34	−10	100
3	55	11	121
4	60	16	256
5	48	4	16
	220		934

$$\bar{x} = \frac{220}{5} = 44$$

$$s_x^2 = \frac{934}{4} = 233.5$$

$$s_x = \sqrt{233.5} \simeq 15.28$$

På kalk:

$\boxed{C}$ $\boxed{\text{STAT}}$

23 $\boxed{\Sigma+}$
34 $\boxed{\Sigma+}$
55 $\boxed{\Sigma+}$
60 $\boxed{\Sigma+}$
48 $\boxed{\Sigma+}$

$\boxed{\bar{x}}$ 44
$\boxed{s_x}$ 15.28

Datasett: Én variabel

i	x_i	
1	15,192	$= x_1$
2	14,020	$= x_2$
3	13,850	$= x_3$
4	12,745	
5	13,520	
6	13,280	
7	10,196	
8	7,173	
9	8,826	
10	14,392	
11	11,218	
12	10,496	
13	12,646	
14	16,087	
15	7,563	
16	16,088	
17	12,713	
18	13,561	
19	19,169	
20	11,571	

$n=20$

$$\left.\begin{aligned} \bar{x} &= 12.715,3 \\ s_x &= 2948,9 \end{aligned}\right\} \text{fra kalkulator}$$

Sentralmål: i) $\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$ utvalgs-gjennoms.

"typisk verdi"

median: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$

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

Spredningsmål: i) S_x : utvalgsstandardavvik

"variasjon" eller "spredning"

ii) variasjonsbredde : $x_{(n)} - x_{(1)}$

Datasett : flere variabler:

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

$\bar{x}, \bar{y}$: utvalgs-~~standard~~ gjennomsnitt

S_x, S_y : utvalgs-standardavvik

$$S_{xy} = \frac{(x_1-\bar{x})\cdot(y_1-\bar{y}) + (x_2-\bar{x})(y_2-\bar{y}) + \dots}{n-1}$$

utvalgskovarians

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$$

korrelasjons-koeffisient i utvalget

Eks:

i	x_i	y_i	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$
1	23	42	−21	16	−336
2	34	27	−10	1	−10
3	55	20	11	−6	−66
4	60	19	16	−7	−112
5	48	22	4	−4	−16
	220	130			−540

$\bar{x} = 44$ $\quad \bar{y} = 26$ $\quad S_{xy} = -135$

$$S_{xy} = \frac{(x_1-\bar{x})(y_1-\bar{y}) + \dots + (x_n-\bar{x})(y_n-\bar{y})}{n-1}$$

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y} = \frac{-135}{15.3 \cdot 9.46} \approx -0.9$$

$\boxed{\bar{x}}$ 44 ← $\bar{x}$

$\boxed{\bar{x},\bar{y}}$ $\boxed{\text{SWAP}}$ 26 ← $\bar{y}$

$\boxed{S_x, S_y}$ 15.28 ← S_x

$\boxed{S_x, S_y}$ $\boxed{\text{SWAP}}$ 9.46 ← S_y

$\boxed{\hat{x}, r}$ $\boxed{\text{SWAP}}$ −0.934 ← r_{xy}

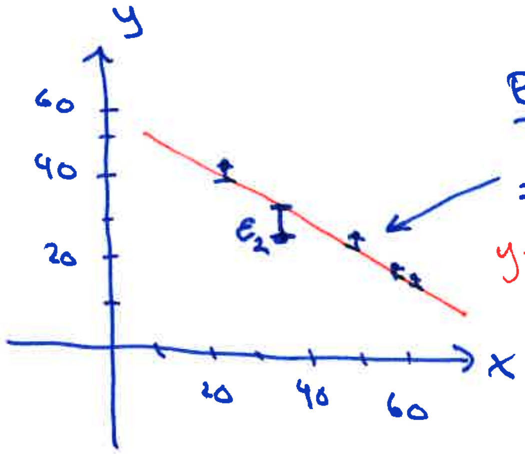
På kalk:

$\boxed{\text{⬑ STAT}}$

23	INPUT	42	Σ+
34	"	27	"
55	"	20	"
60	"	19	"
48	"	22	"

Spredningsdiagram:

i	x_i	y_i
1	23	42
2	34	27
3	55	20
4	60	19
5	48	22



Regr. linjen:

$= 51.4 - 0.578x$

$y = \alpha + \beta x$

$r_{xy} = -0.93$

$$\boxed{\begin{aligned} \beta &= r_{xy} \cdot \frac{s_y}{s_x} \\ \alpha &= \bar{y} - \beta \bar{x} \end{aligned}}$$

Minste kvadraters metode:

$$\varepsilon_1 = y_1 - (\alpha + \beta x_1)$$
$$\varepsilon_2 = y_2 - (\alpha + \beta x_2)$$
$$\vdots$$
$$\varepsilon_n = y_n - (\alpha + \beta x_n)$$

Minste kvadraters metode:

Velge α, β slik at $\varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_n^2$ er minst mulig

$\varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_n^2$:

Jo mindre denne summen av kvadrater er, jo bedre passer den rette linjen med dataene.

Regresjonslinjen:

den rette linjen som passer best mulig med datapkt., dvs løsningen ved minste kvadraters metode

På kalk:

[$\hat{y}, m$] [SWAP]

$\beta \rightarrow -0.578$

[$\bar{x}w, b$] [SWAP]

$\alpha \rightarrow 51.4$

$$\boxed{\begin{aligned} \beta &= r_{xy} \cdot \frac{s_y}{s_x} \\ \alpha &= \bar{y} - \beta \cdot \bar{x} \end{aligned}}$$

3 Sum-notasjon

$\sum$: summetegn (stor sigma, sum)

det som skal summeres:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \dots + x_n$$

$\sum$ = Summetegn

$$\sum_{i=1}^{n} \Big\} \quad i = 1, 2, 3, \dots, n$$

Eks:

$$\sum_{i=1}^{5} i^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2$$
$$= 1 + 4 + 9 + 16 + 25 = \underline{55}$$

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

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

Regneregler:

i) $\sum_{i=1}^{n}(a_i \pm b_i) = \sum_{i=1}^{n} a_i \pm \sum_{i=1}^{n} b_i$

ii) $\sum_{i=1}^{n}(c \cdot a_i) = c \cdot \sum_{i=1}^{n} a_i$ (c er konstant)

iii) $\sum_{i=1}^{n} c = n \cdot c$

Merk: $\sum_{i=1}^{n} a_i^2 \neq \left(\sum_{i=1}^{n} a_i\right)^2$

$a_1^2 + a_2^2 + \dots + a_n^2 \neq (a_1 + a_2 + \dots + a_n)^2$

Eks: $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}\left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right)\right)$

$= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} 2x_i\bar{x} + \sum_{i=1}^{n}\bar{x}^2\right]$

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

$\sum_{i=1}^{n}(a_i + b_i) = (a_1 + b_1) + (a_2 + b_2) + \dots + (a_n + b_n)$

$= a_1 + a_2 + \dots + a_n + b_1 + b_2 + \dots + b_n$

$= \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i$

$\sum_{i=1}^{n}(ca_i) = ca_1 + ca_2 + \dots + ca_n = c(a_1 + a_2 + \dots + a_n)$

$= c \cdot \sum_{i=1}^{n} a_i$