

PLAN:

① STANDARDAVVIK, VARIANS

[T] 3.3

② Z-verdier, kvartiler

[T] 3.4

HUSK:

ARBEIDSKRAV 2

Frest: mandag 13/02.

Repetisjon:

Gitt et datasett hentet inn fra en
stikprøve = utvalg på n personer $\rightarrow$
har vi

$$\text{Std. avvik} = s = \sqrt{s^2}$$

$$\text{Varians} = s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots}{n-1}$$

i	X	...
1	x_1	
2	x_2	...
3	x_3	
4	x_4	
5	x_5	
...	...	
n	x_n	

Alternative formler:

$$\text{Std. avvik} = s = \sqrt{s^2}$$

$$\text{Varians} = s^2 = \frac{n \sum x_i^2 - (\sum x_i)^2}{n \cdot (n-1)} = \frac{n(x_1^2 + x_2^2 + \dots + x_n^2) - (x_1 + x_2 + \dots + x_n)^2}{n \cdot (n-1)}$$

Parametre: Dersom hele populasjonen består av N personer,
så er parametrene

$$\text{Std. avvik: } \sigma = \sqrt{\sigma^2}$$

$$\text{Varians: } \sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

Husk:

$$\mu = \frac{\sum x_i}{N}$$

gj. snitt i
populasjonen

Øvingsoppgave

I en stikprøve finner man følgende lønn for utvalgte fotballtrenere:

$$x_1 = 150.000 \quad x_2 = 300.000 \quad x_3 = 360.000$$

Fin variens og std. avvik for stikprøven.

Løsning:

Varians:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2}{3-1}$$

$$\bar{x} = \frac{150.000 + 300.000 + 360.000}{3}$$

$$= 270.000$$

$$= \frac{(-120.000)^2 + (30.000)^2 + (90.000)^2}{2}$$

$$= 11.700.000.000$$

$$150.000 - 270.000 = -120.000$$

Std. avvik:

$$s = \sqrt{s^2} = \sqrt{11.700.000.000} \approx 108.200$$

Utregning på kalk.

120.000

+/-

x²

+

30.000

x²

+

90.000

x²

=

%

2

=

√x

VARIANS

STD. AVVIK

$$1 + 2 \cdot 3 = 1 + 6 = 7 \leftarrow$$

Test for å se om AOS-modus er innstilt!

Trykke:

2ND

og

FORMAT

↑

Hvis ikke AOS så

2ND

og

SET

Øvingsoppgave

Example

Alderen på lederne i en bedrift er: 35 32 55 48 41 47.

Beregn standardavviket.

$$\bar{X} = \frac{35 + 32 + 55 + 48 + 41 + 47}{6} = 43$$

$$S^2 = \frac{(35-43)^2 + (32-43)^2 + (-8)^2 + (-11)^2 + 12^2 + 5^2 + (-2)^2 + 4^2}{6-1} = \frac{374}{5}$$

$$= \underline{74.8} \quad (\text{varians})$$

$$s = \sqrt{74.8} \approx \underline{8.6} \quad (\text{std. avvik})$$

Øvingsoppgave

Example

Alderen på lederne i en bedrift er: 35 32 55 48 41 47.

Beregn standardavviket.

$$\begin{aligned}s &= \sqrt{\frac{(35-43)^2 + (32-43)^2 + (55-43)^2 + (48-43)^2 + (41-43)^2 + (47-43)^2}{5}} \\&= \sqrt{\frac{(-8)^2 + (-11)^2 + 12^2 + 5^2 + (-2)^2 + 4^2}{5}} \\&= \sqrt{\frac{64+121+144+25+4+16}{5}} \\&= \sqrt{\frac{374}{5}} = \sqrt{74.8} = 8.6\end{aligned}$$

Shortcut formelen for s

$$s = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}}$$

Example

Stikkprøve med tre tall: **3, 7, -4**.

$$\sum x^2 = 3^2 + 7^2 + (-4)^2 = 9 + 49 + 16 = 74$$

$$\sum x = 3 + 7 + (-4) = 6$$

$$\begin{aligned} s &= \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}} \\ &= \sqrt{\frac{3 \cdot 74 - 6^2}{3(3-1)}} = \sqrt{\frac{186}{6}} = \sqrt{31} = 5.6 \end{aligned}$$

Varsians

- *Varsiansen* er s^2
- Eng: Variance

Example

Stikkprøve med tre tall: **3, 7, -4** har variansen $5.6^2 = 31.0$.

↖

$$\begin{aligned} s &= \text{std. avvik} = 5.6 \\ s^2 &= \text{varians} = 5.6^2 \end{aligned}$$

Standardavviket i populasjonen

I populasjonen

- s er en *observer*
- Standardavviket i populasjonen har det greske symbolet σ
- σ er en *parameter* som har formelen

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

std.-avvik

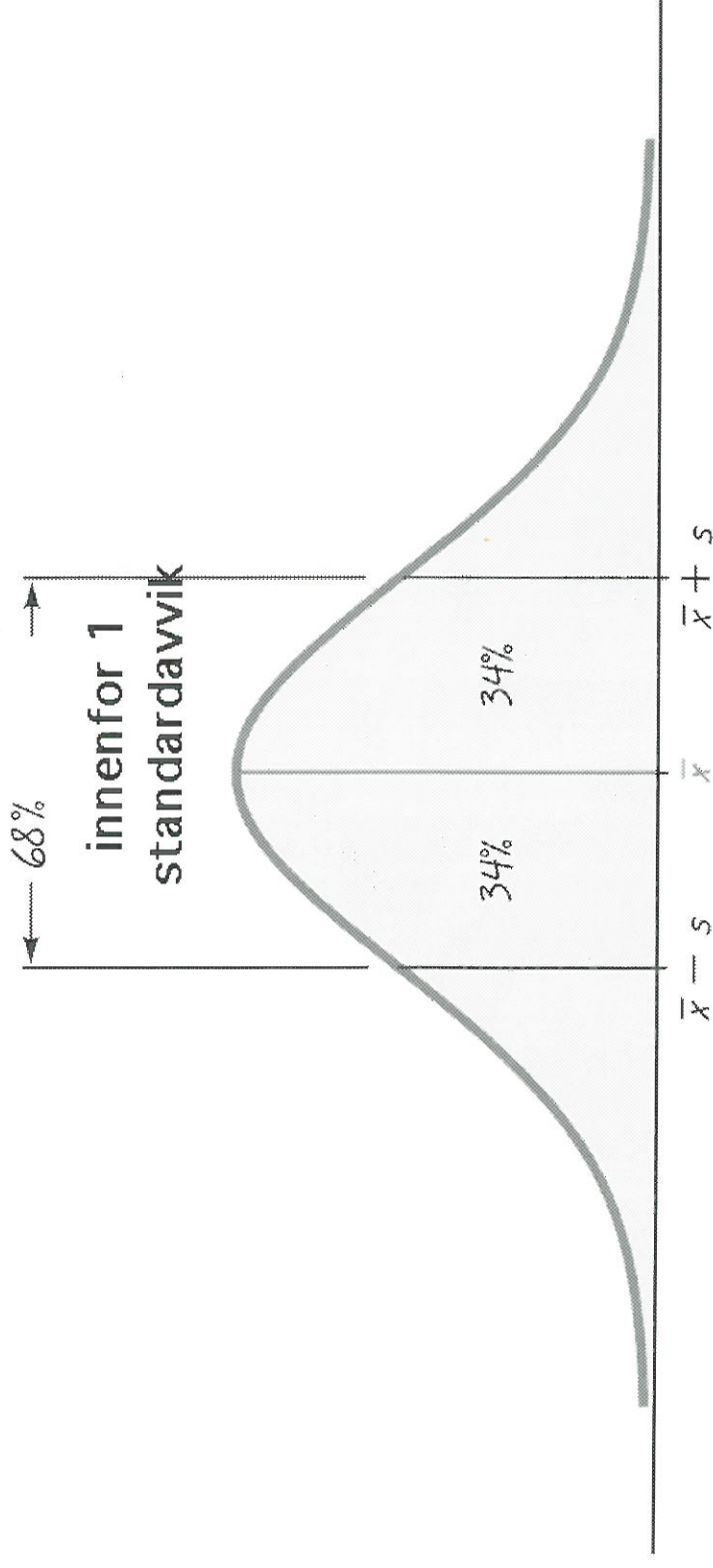
$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

varians

Empirisk regel for s

68%

Er dataene tilnærmet normalfordelt så vil ca 68% av dataene ligge 1 standardavvik ifra gjennomsnittet.



~~MAV~~

$$\bar{x} - s = 34.4$$

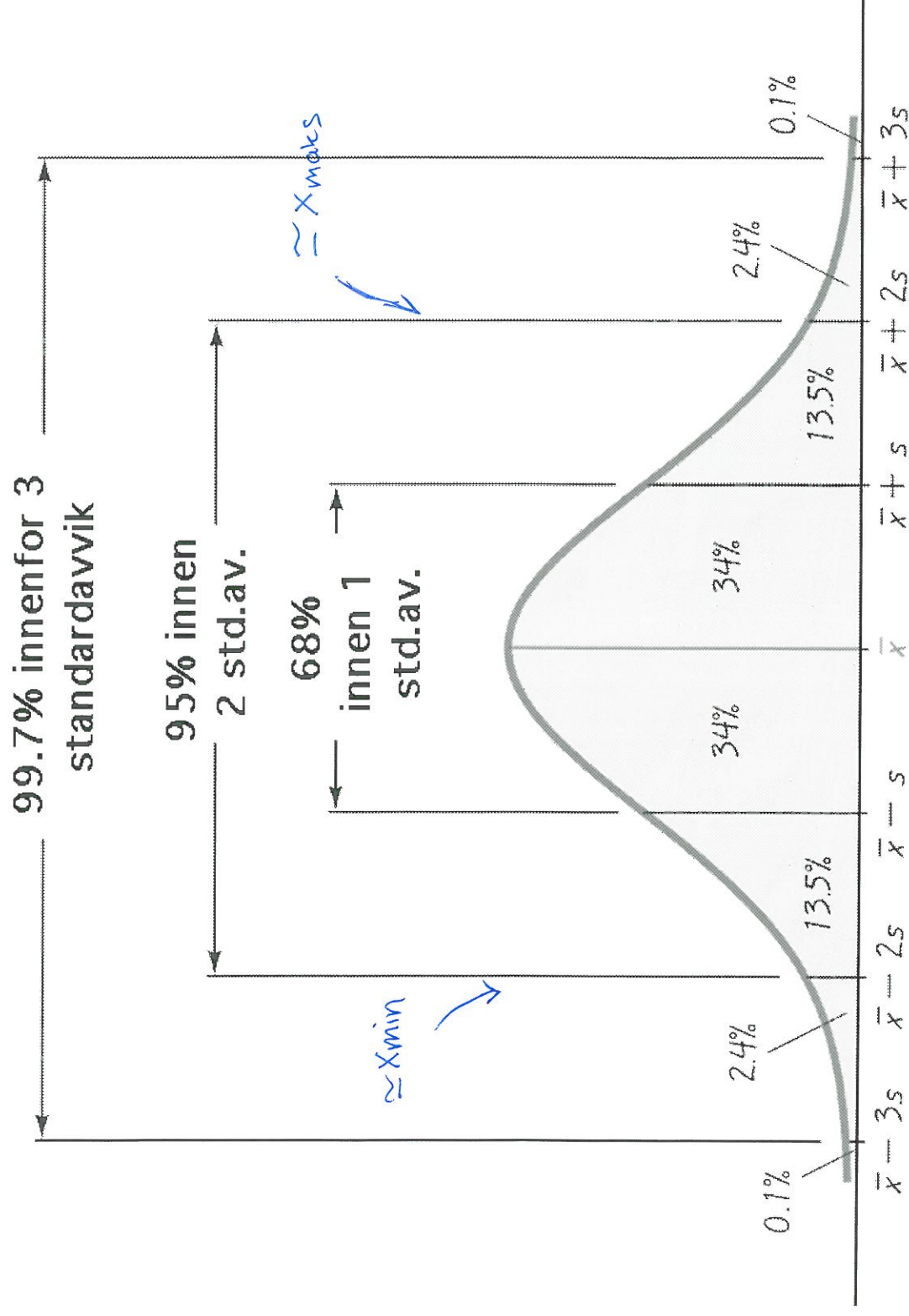
$$\bar{x} = 43$$

$$s = 8.6$$

$$\bar{x} + s = 51.6$$

68% - 95 % - 99.7 %

For normalfordelte data vil 95% av verdiene ligge innenfor 2 standardavvik.



Tommelfingerregel for s

En omtrentlig regel

Hvis du deler variasjonsbredden på 4, så har du et anslag på s .

Examples

- 1 Stikkprøve med tre tall: **3, 7, -4** har variasjonsbredde 11. Så da kan vi anslå standardavviket til å være $11/4 = 2.75$. Men det bommer endel på det faktiske standardavviket $s = 5.6$
- 2 I en stikkprøve av 56 studenter varierte antall ukentlige studietimer fra 40 til 0 timer. Med tommelfingerregelen anslår vi s til å være $40/4 = 10$ timer. Det faktiske standardavviket er $s = 8.1$ timer.

Tolkning av s

Hvordan tolke s ?

Hvis du kjenner standardavviket s så kan du anslå typiske" maksimums- og minimumsverdier:

- En typisk minimumsverdi er 2 standardavvik under gjennomsnittet $\bar{x} - 2s$
- En typisk maksimumsverdi er 2 standardavvik over gjennomsnittet $\bar{x} + 2s$
- Denne regelen er brukbar så lenge dataene er tilnærmet normalfordelt.

1 Section 3-4: standardiserte verdier
z-verdier

2 Sec 3-4: Kvartiler og persentiler

3 Sec 3-5: Boksplott og uteliggere

Å finne standardisert verdi z

z-verdier

- Smart og viktig måte å standardisere på
- Vi omregner til en z-verdi
- Da kan vi sammenligne med andre verdier, fra samme stikkprøve, men også fra andre
- Viktig å forstå z-verdier

Formelen for z-verdien til en dataverdi x er

$$z = \frac{x - \bar{x}}{s}$$

formel for z-verdi

z-verdien teller antall standardavvik ifra gjennomsnittet.

Eks: x = alderen til ledere i antall år

$\bar{x} = 43$	$x = 35$	$z = \frac{35 - 43}{8.6} \approx -0.93$
$s = 8.6$	$x = 55$	$z = \frac{55 - 43}{8.6} \approx 1.40$

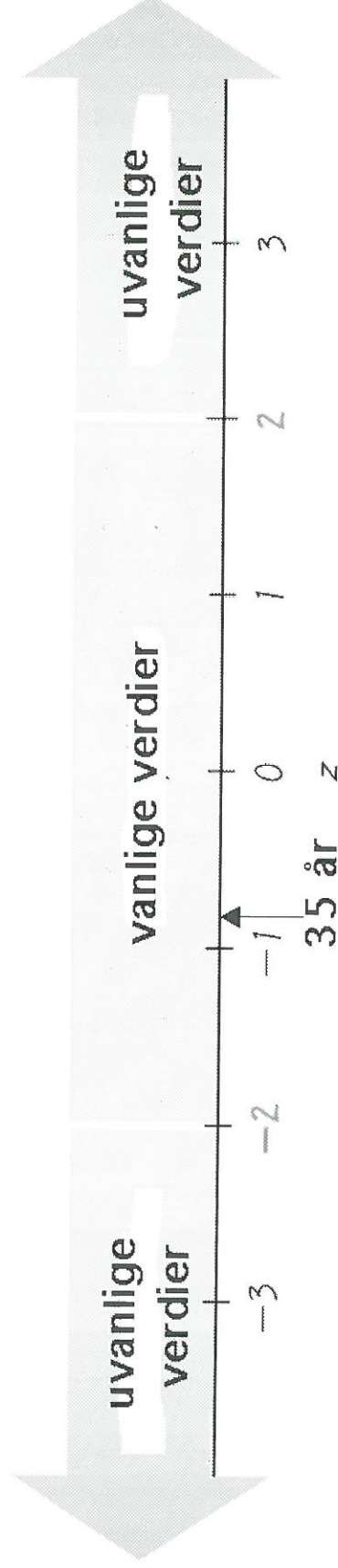
Example

Alderen på lederne i en bedrift er: 35 32 55 48 41 47.

- Lederen på 35 år, hvor ligger han plassert aldersmessig?
- Han er 8 år yngre enn gjennomsnittet, men hvor mye er det?
- De 8 årene må sammenlignes med standardavviket:

$$z = \frac{35 - 43}{8.6} = -0.9$$

- Lederen er et snaut standardavvik yngre enn gjennomsnittet.
- Han er ikke "uvanlig" ung, siden z-verdien ikke overstiger ± 2 .



z-verdier brukes også til å sammenligne verdier ifra to ulike stikkprøver.

Example

- I kommune A og B var gjennomsnittslønna 274000 kr og 317000 kr.
- Standardavvikene var kr 25000(A) og kr 35000 (B).
- Alf i kommune A tjener 360000 kroner, mens Berit i kommune B tjener 400000. Hvem tjener mest?
- Hvem tjener best i forhold til nivået i hjemkommunen?
- Svar: Alf sin z-verdi er 3.4, mens Berit har z-verdi 2.4
- Konklusjon: Relativt sett er det Alf som tjener best, når vi ser på hvor han står i forhold til innbyggerne i sin kommune

Kvartiler

Kvartiler

Sorter dataene.

- Nedre kvartil Q_1 , skiller de nedre 25% fra de øvre 75%
- Midterste kvartil Q_2 , skiller de nedre 50% fra de øvre 50% (Medianen)
- Øvre kvartil Q_3 , skiller de nedre 75% fra de øvre 25%

Kvartilene deler de sorterte dataene i fire like store deler.

