

FORELESNING 7

EIVIND ERIKSEN

FEB 09 2012

MET 434/

STATISTIKK

PLAN:

① Kvartiler, Prosentiler

② Boksplott

③ Eksempler : SAS JMP.

} [T] Kap. 3.4

JMP-fil finnes på
It's Learning

Kvartiler

Prosentiler

Median

Kvartiler:

Q_1, Q_2, Q_3

"
Median

Prosentiler:

$P_1, P_2, \dots, P_{99}$

$P_{25} = Q_1$ $P_{75} = Q_3$

$P_{50} = Q_2 = \text{Median}$

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) $Q_2 = \text{Median}$
- Øvre kvartil Q_3 , skiller de nedre 75% fra de øvre 25%

Kvartilene deler de sorterte dataene i fire like store deler.

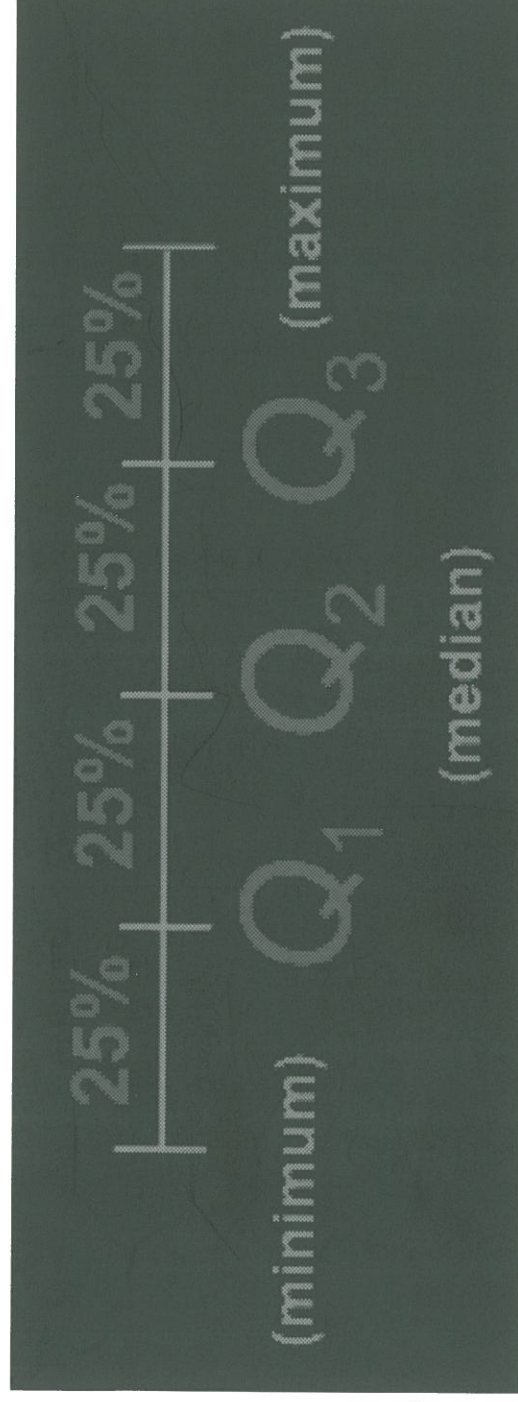
$$n = 1280$$

observasjoner

$$Q_1 \approx \frac{1280}{4} = 320$$

$$Q_2 \approx \frac{1280}{2} = 640$$

$$Q_3 \approx \frac{1280 \cdot 3}{4} = 960$$



Persentiler

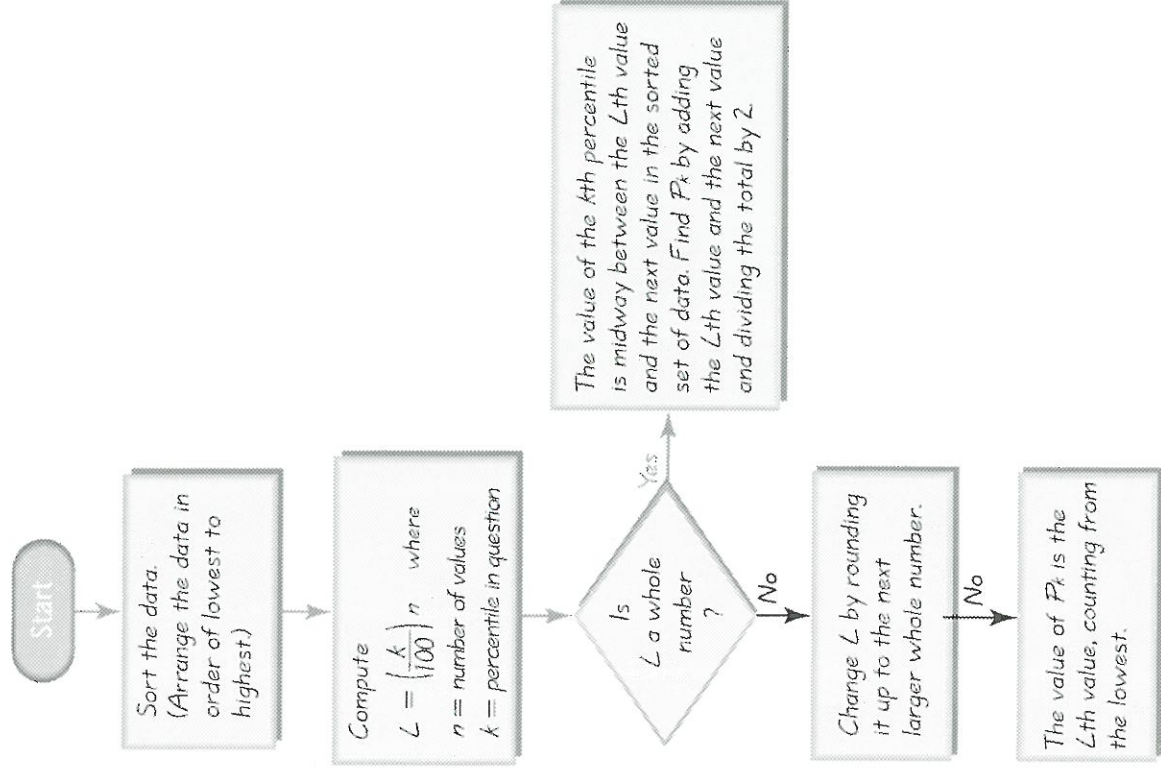


Persentil

- Kvartiler gir firdeling
- Med persentiler deler man opp i 100 like store deler
- $P_1, P_2, \dots, P_{99}$
- $Q_1 = P_{25}$ og $Q_3 = P_{75}$
- Bruk Figur 3-6

$$L = \frac{k}{100} \cdot n$$

$n = \text{antall observasjoner}$
 $k: P_k, \text{ } k^{\text{te}} \text{ persentil}$
 $n \cdot k. \leftarrow \text{persentil}$



persentil

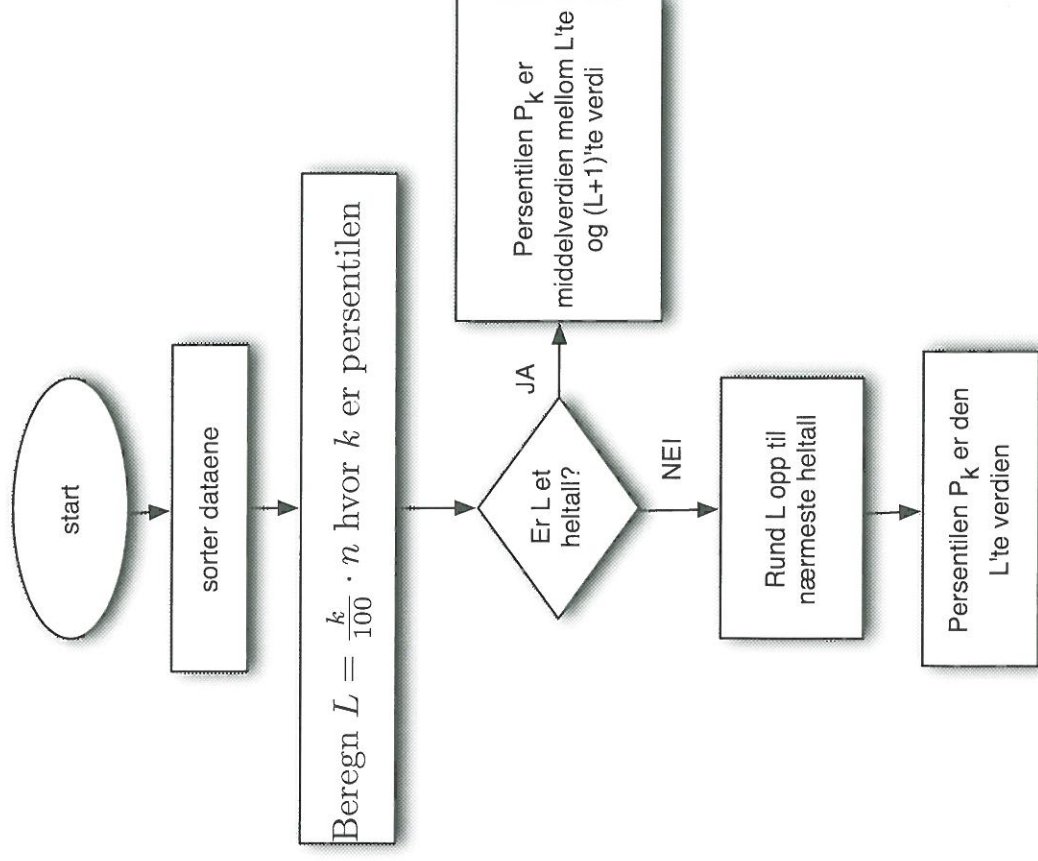
Persentiler

For å finne P_5 når $n = 1280$

$$L = \frac{5}{100} \cdot 1280 = 64$$

Persentil

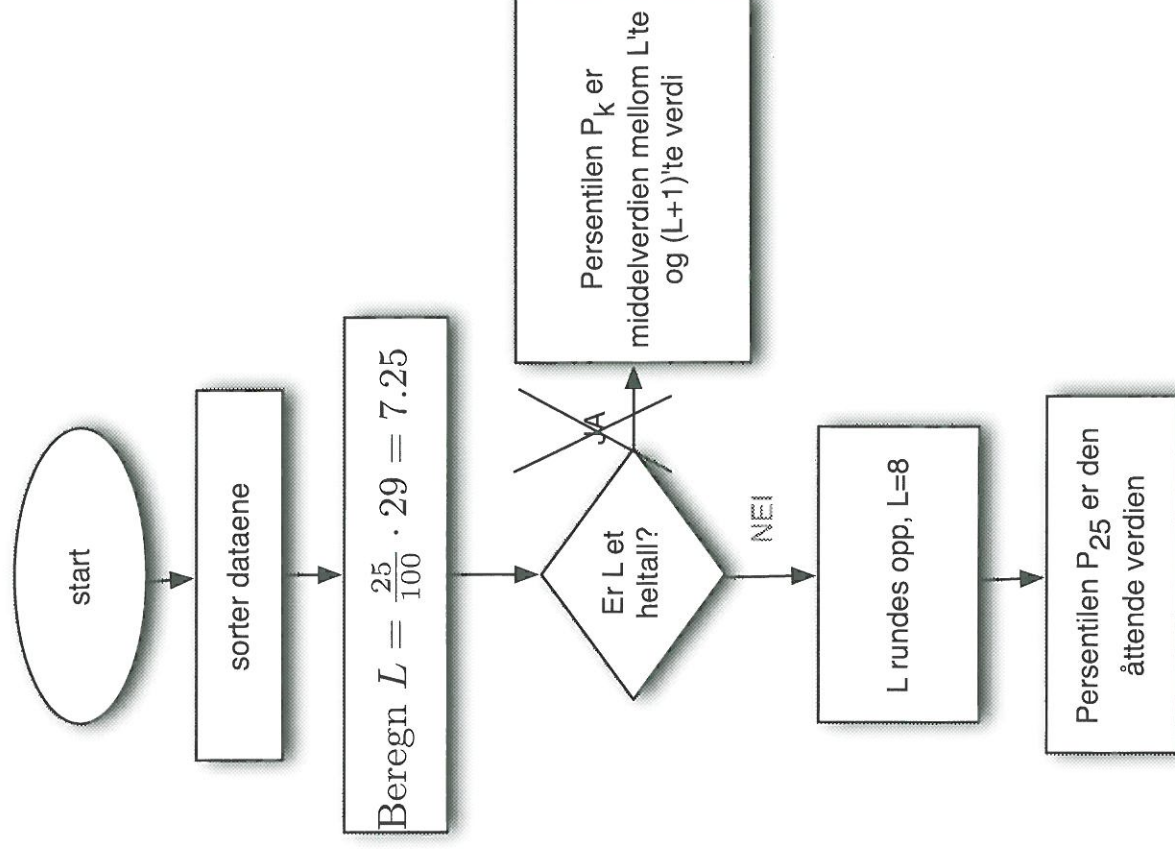
- Mange måter å beregne kvartiler/persentiler på
- SPSS, JMP, Excel er uenige
- Vi bruker Figur 3-5 i boka



Kvartil eksempel

Example

- Vi ville finne nedre kvartil
 $Q_1 = P_{25}$
- Stikkprøven har størrelse
 $n = 29$
- Da er $k = 25$ i Figur 3-6
- Konklusjon: Nedre kvartil er
8. verdi fra bunn



Kvartil eksempel

Example

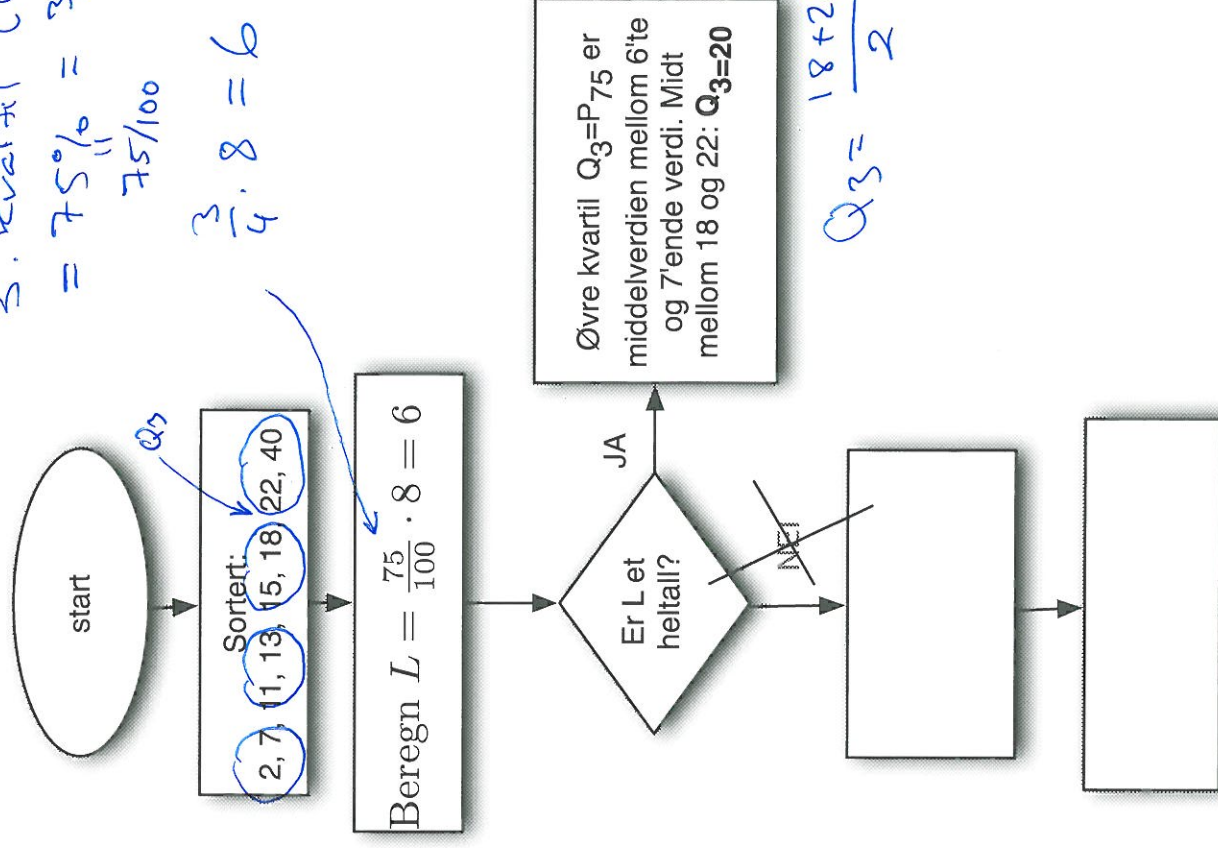
Stikkprøven:

13 22 15 18 11 7 2 40

- Finn øvre kvartil $Q_3 = P_{75}$
- Stikkprøven har størrelse $n = 8$
- Da er $k = 75$ i Figur 3-6
- Øvre kvartil ligger midt mellom 6. og 7. verdi
- Konklusjon: $Q_3 = 20$

$$\begin{aligned} 3. \text{ kvartil } (Q_3) \\ = 75\% = \frac{3}{4} \\ 75/100 \end{aligned}$$

$$\frac{3}{4} \cdot 8 = 6$$



Om uteliggere

En uteligger (ekstrem observasjon) kan ha drastisk effekt på

- Gjennomsnittet
- Standardavviket
- På skalaen til histogrammet, slik at det ikke gir et godt bilde av fordelingen

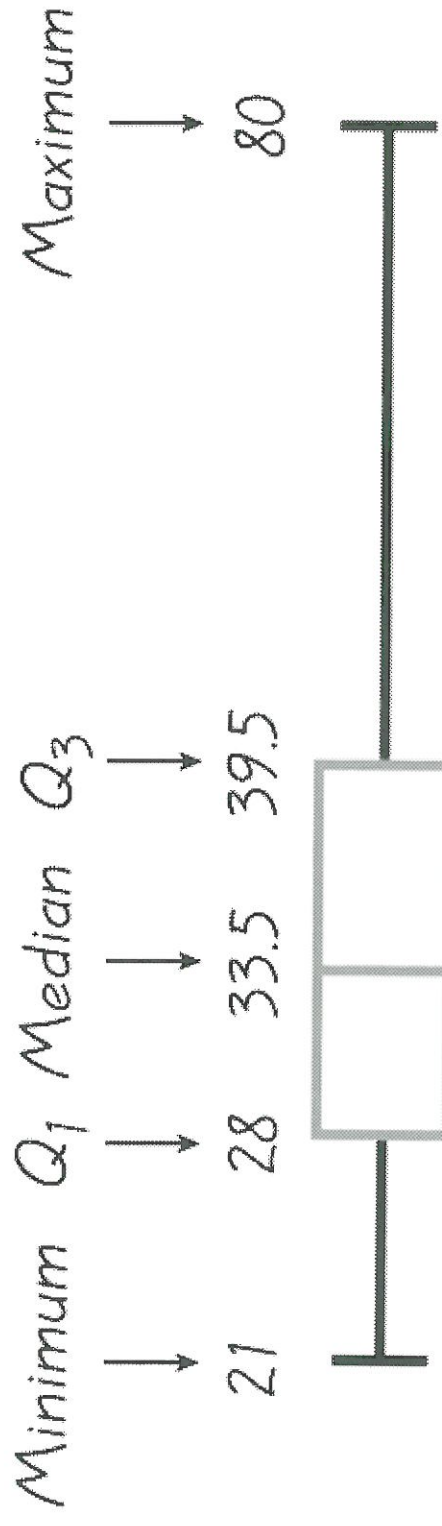
5-talls oppsummering og boksplott

5-talls oppsummering

Minimumsverdi, Nedre kvartil, Median, Øvre kvartil og maksimumsverdien Q_1 Q_2 Q_3

Boksplott

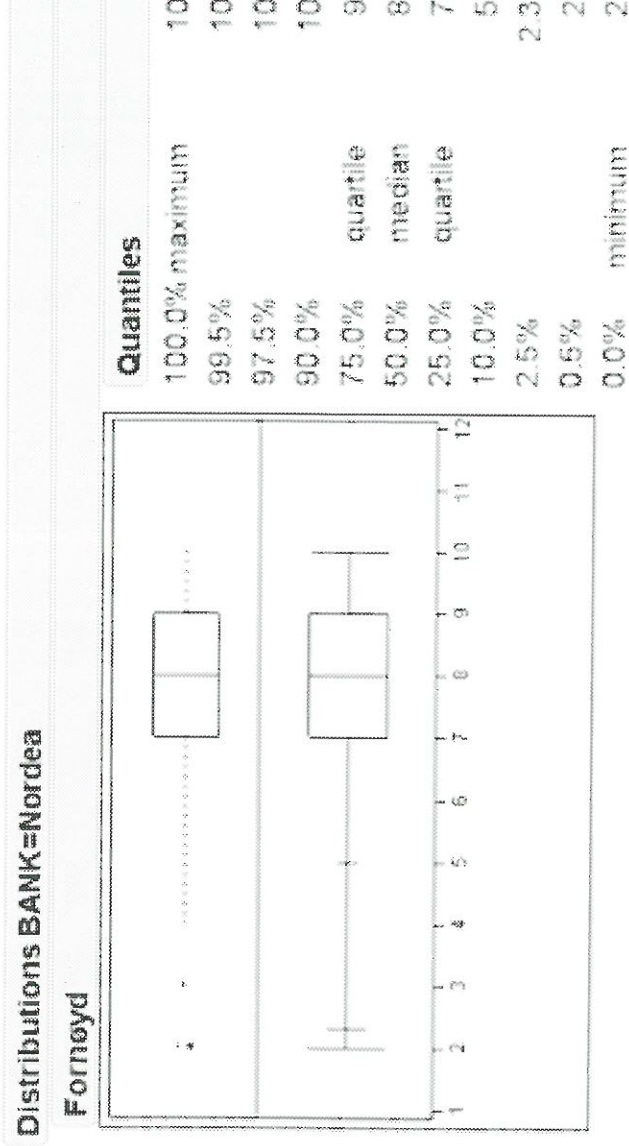
Er en graf som viser 5-talls oppsummeringen.



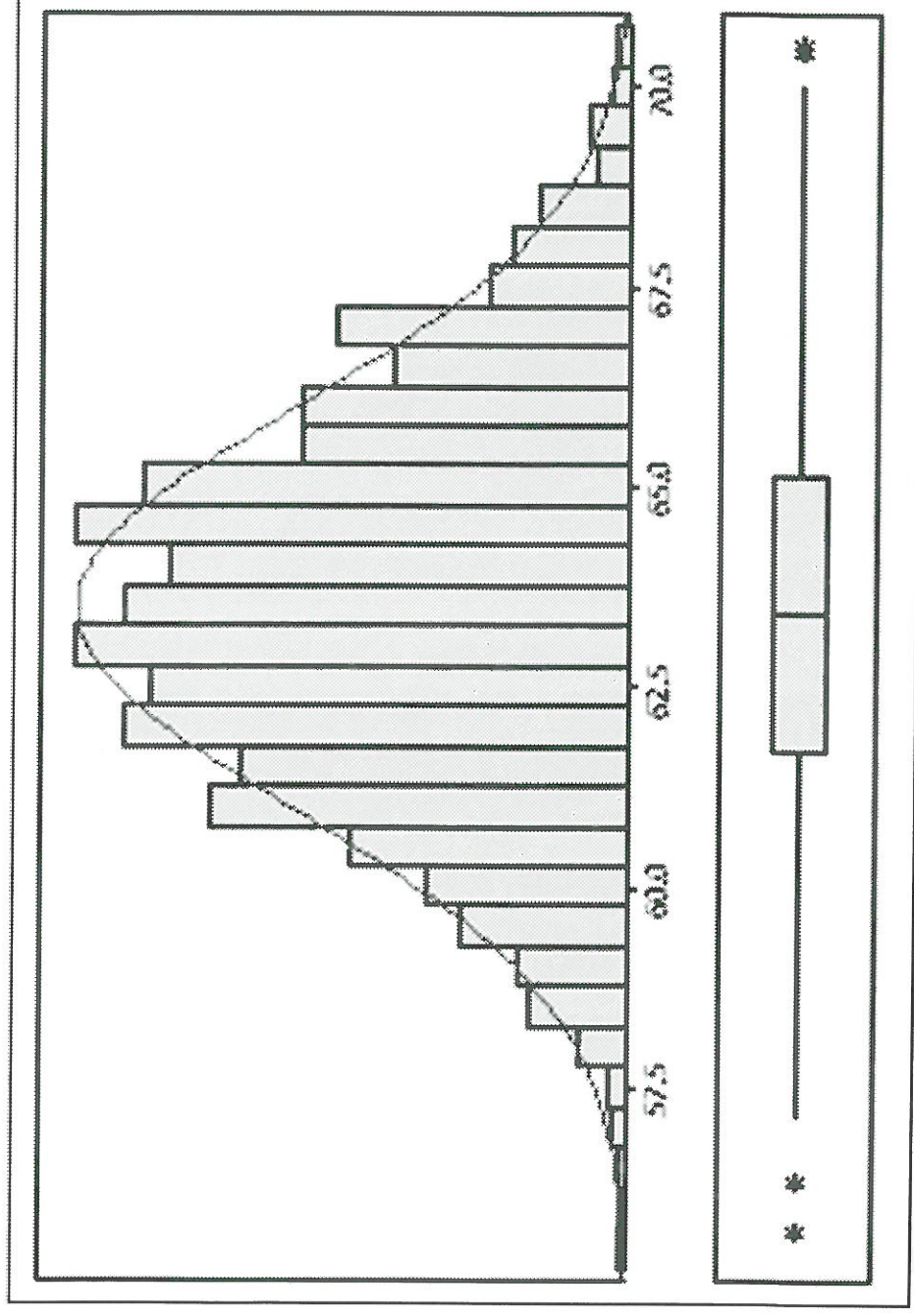
Boksplott i JMP

JMP har to boksplott

- JMPs outlier box plot er det som boka (section 3-5) kaller *Modified box plot*. Det brukes til å oppdage uteliggere.
- JMPs quantile box plot er det som boka kaller *box plot* og som ble vist på forrige slide.

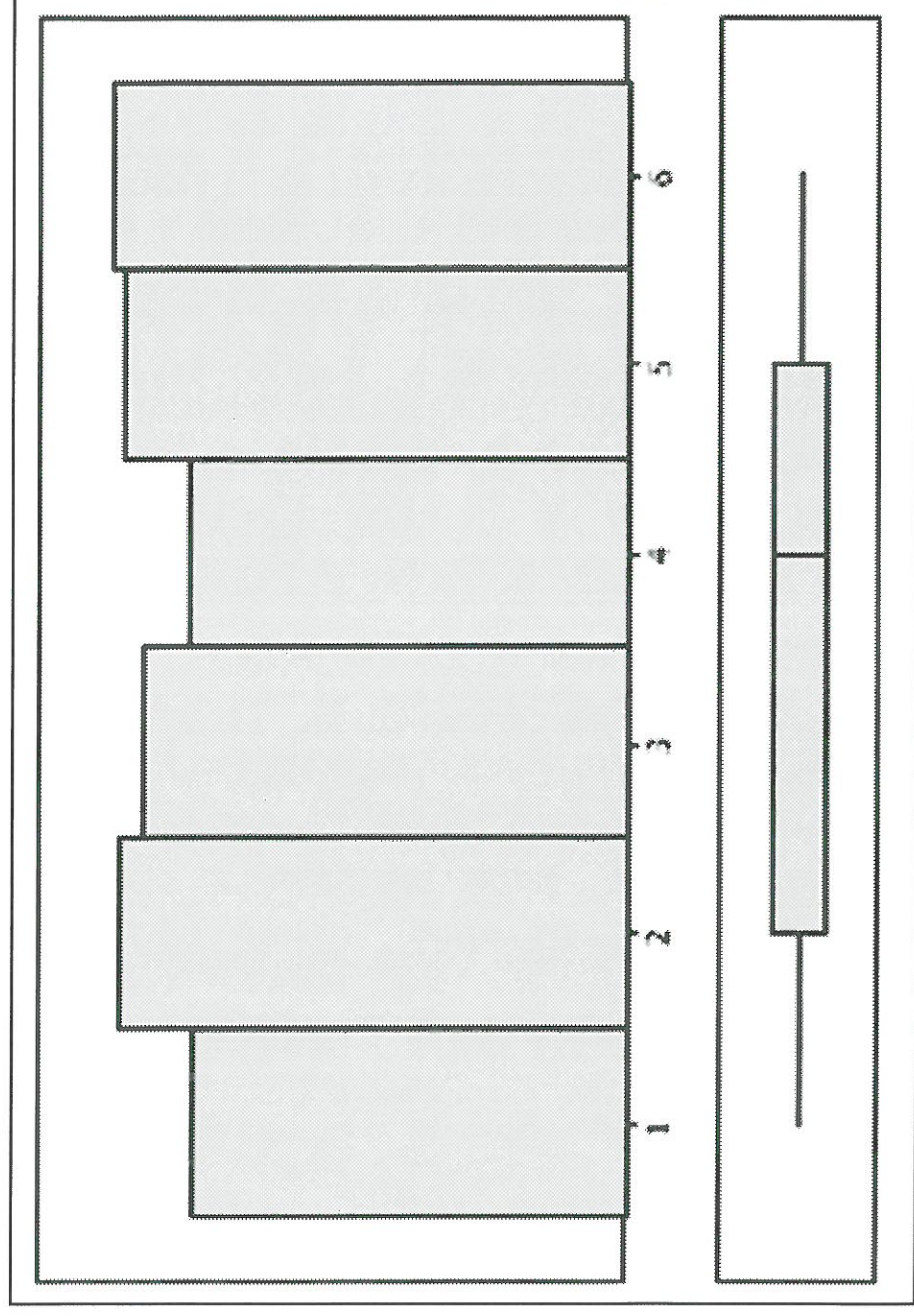


Boksplott for normalfordeling



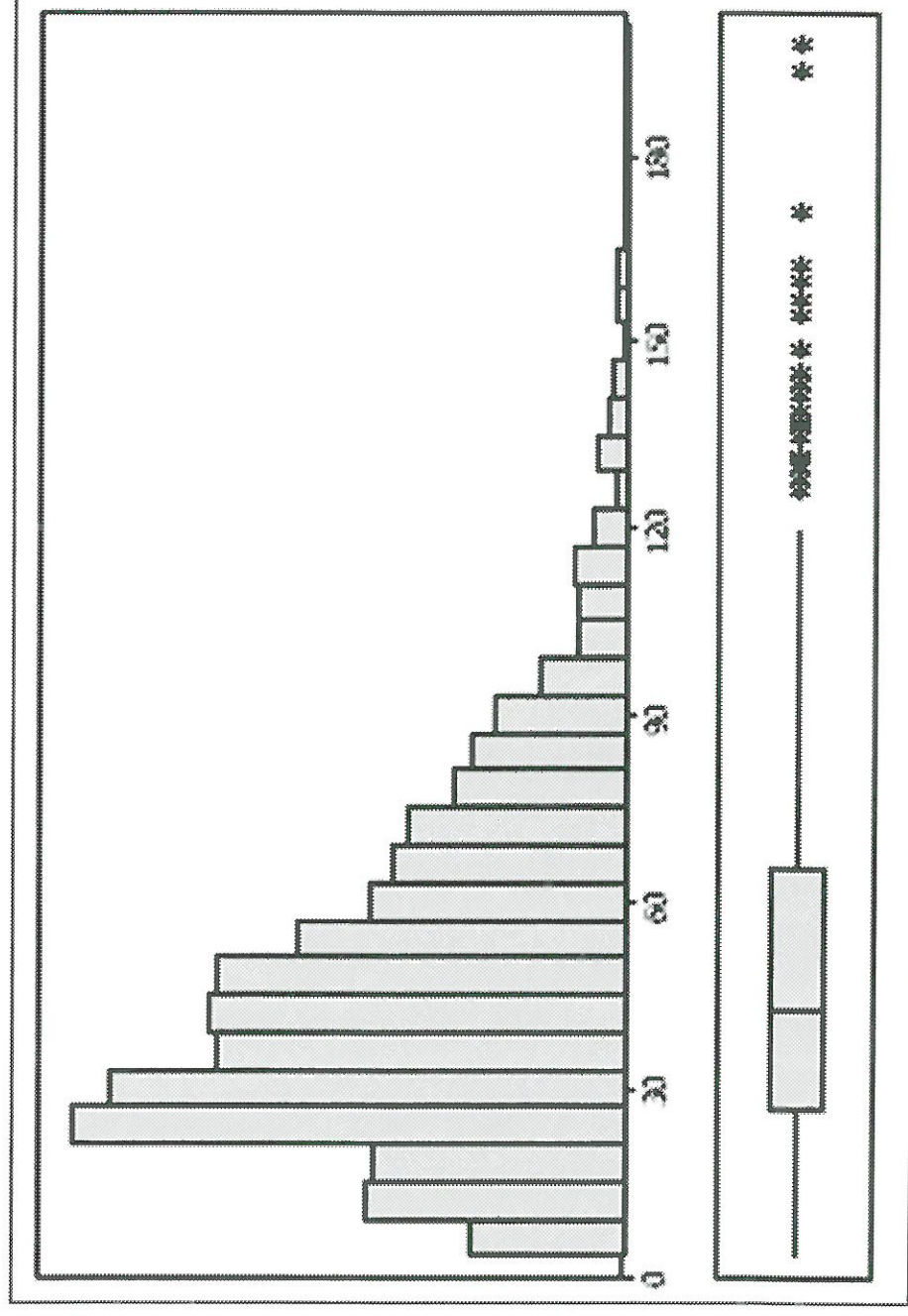
(a) Høyde (inch) for 10000 kvinner. Normalfordelt

Boksplott for uniform fordeling



(b) 1000 terningkast. Uniform fordeling

Boksplott for høreskjev fordeling



(c) Månedslønn til 1000 markedsførere. Høreskjev.

Outlier box plot = Modified box plot

Boksplott for å oppdage uteliggere

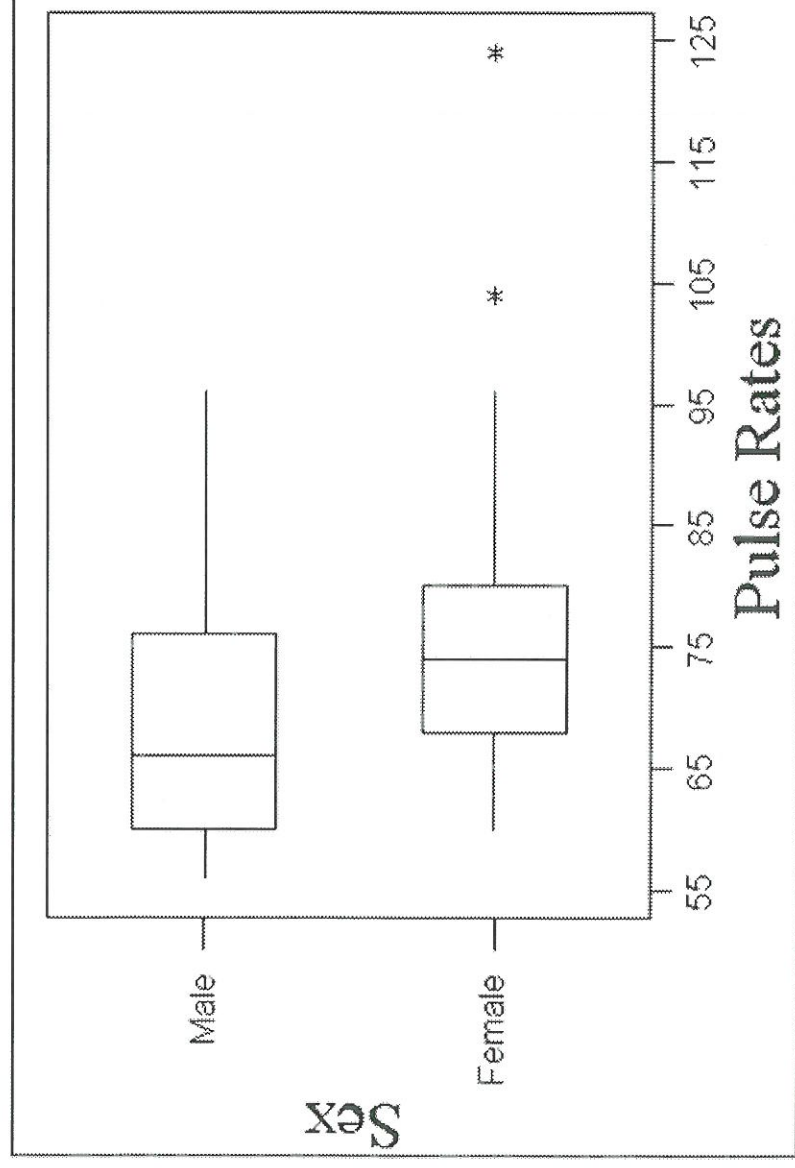
Disse boksplottene markerer en verdi som uteligger.

- Kvartilbredden er definert som $IQ = Q_3 - Q_1$
- Markert som uteligger dersom større enn $Q_3 + 1.5 \cdot IQ$
- Eller mindre enn $Q_1 - 1.5 \cdot IQ$

Uteligger:

* Større enn
 $Q_3 + 1.5 \cdot IQ$

* mindre enn
 $Q_1 - 1.5 \cdot IQ$



Eks:

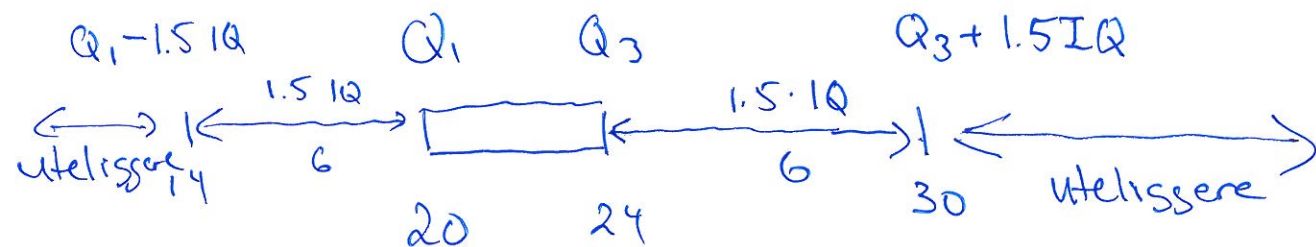
Alder fra JMP-fil

$$Q_1 = 20$$

$$Q_3 = 24$$

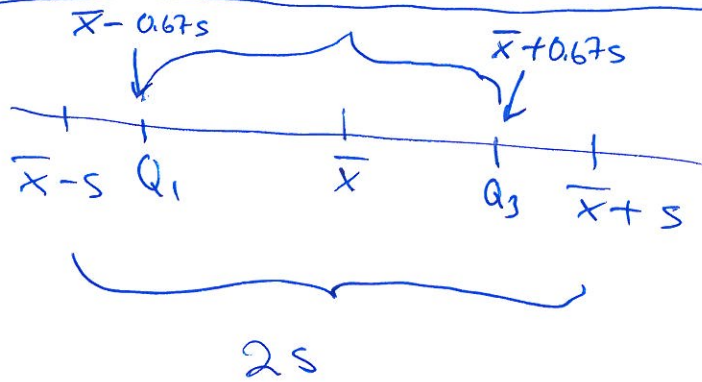
$$IQ = Q_3 - Q_1 = 24 - 20 = 4$$

(50% av studentene innenter dette intervallut på 4 År)



$$1.5 IQ = 1.5 \times 4 = 6$$

Hva sier når dataene er normalfordelt



Svarer til 68%
av observasjonene

$$IQ = 2 \times 0.67s = 1.33s$$

Grenser for uteliggere:

$$\begin{aligned} Q_3 + 1.5 IQ &= (\overline{x} + 0.67s) + 1.5 \cdot (2 \cdot 0.67s) \\ &\quad \uparrow \qquad \qquad \qquad \underbrace{\hspace{1cm}}_{IQ} \\ &\quad Q_3 \qquad \qquad \qquad \\ &= \overline{x} + 2.67s \end{aligned}$$

1 Eks med alder (JMP)

$$\overline{x} + 2.67s = 23.2 + 2.67 \cdot 5.4 = \underline{\underline{37.6}}$$