

FORELESNING 4

MET 3431

Eivind Eriksen, JAN 31 2012

STATISTIKK

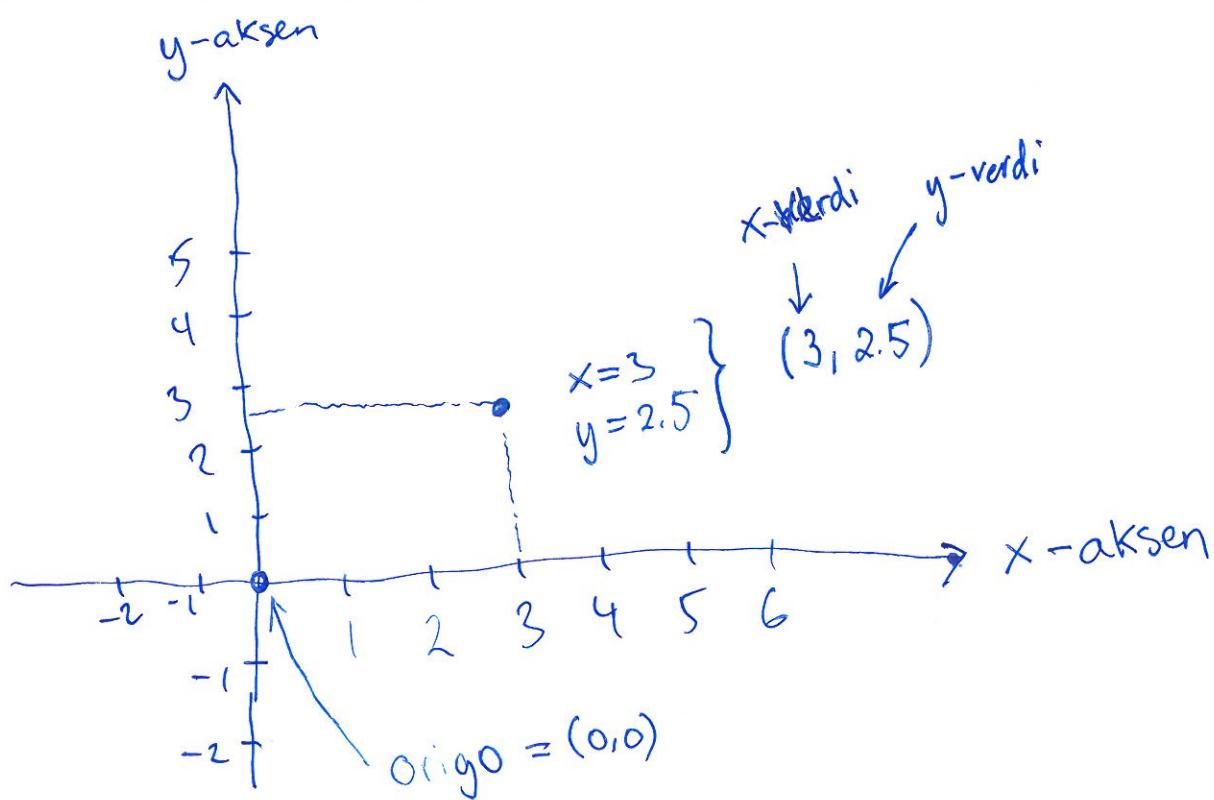
PLAN:

- ① Spredningsplott [T] 2.4
- ② Oppsaver Kap. 2
- ③ Sentralmål [T] 3.2

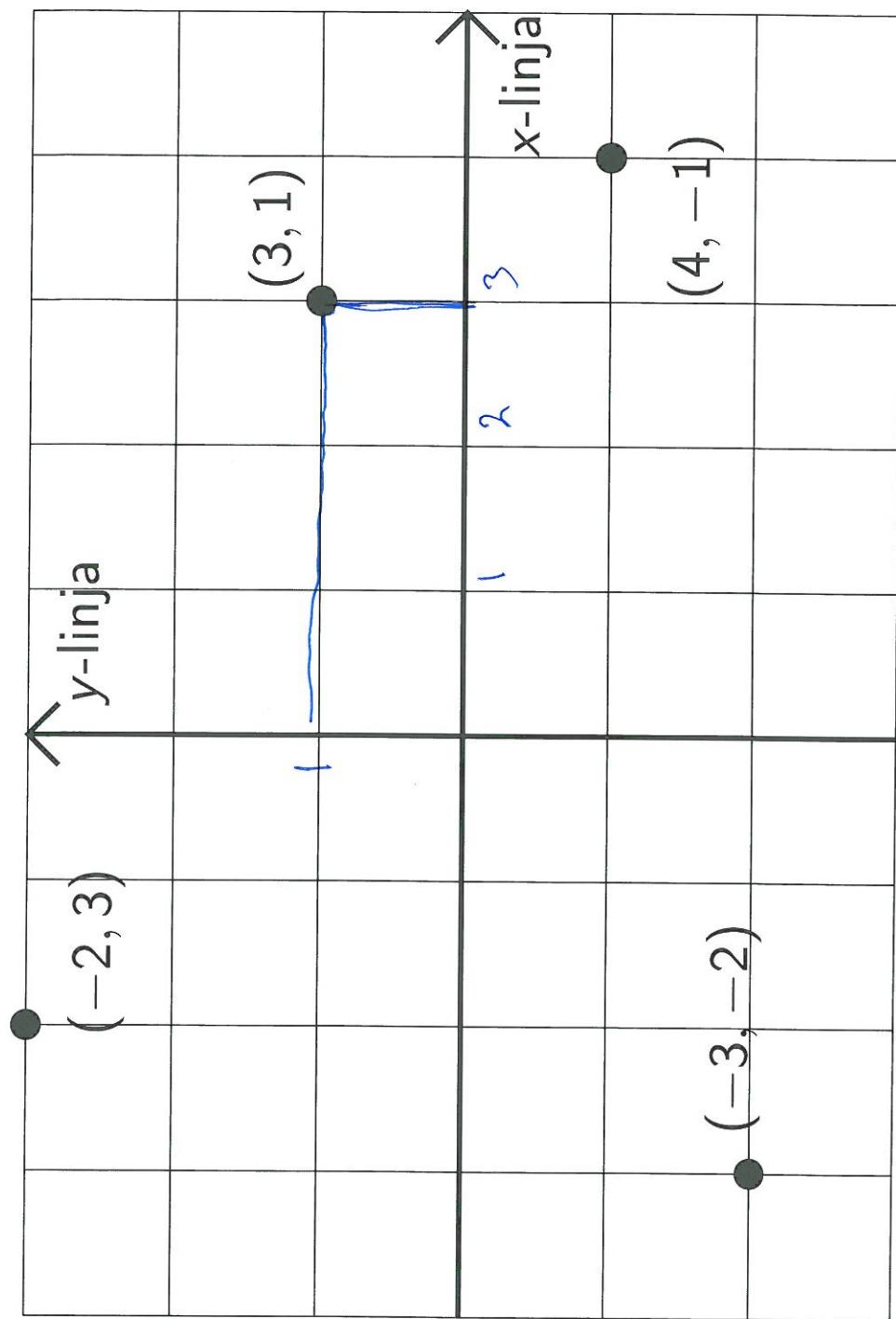
① Spredningsplott

Egner seg når vi vil se på sammenhengen mellom to variable.

To-dimensjonalt koordinatsystem



Koordinatsystem



Øvingsoppgaver

Tegn inn følgende punkt i et koordinatsystem

① $(0, -0)$

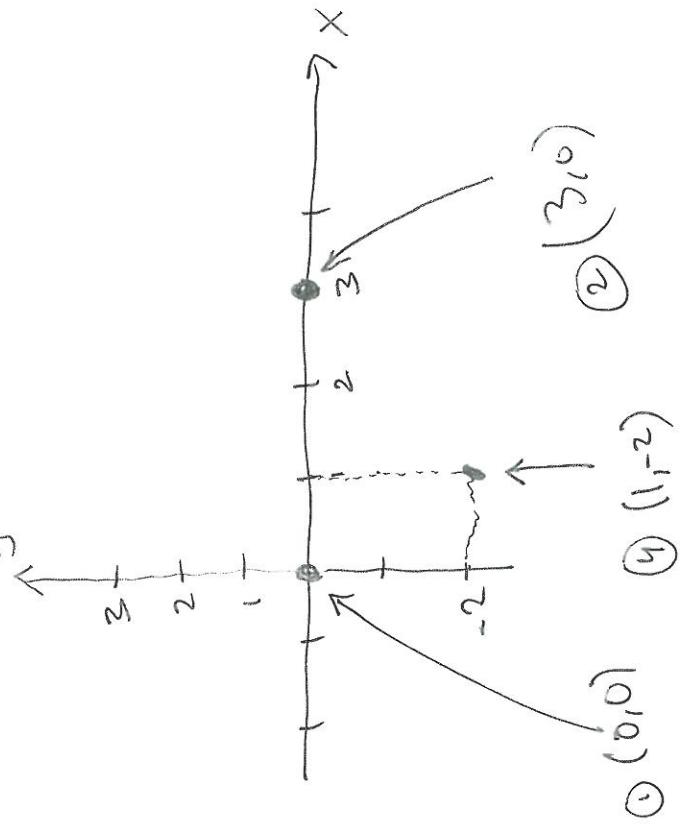
② $(3, 0)$

③ $(0, -2)$

④ $(1, -2)$

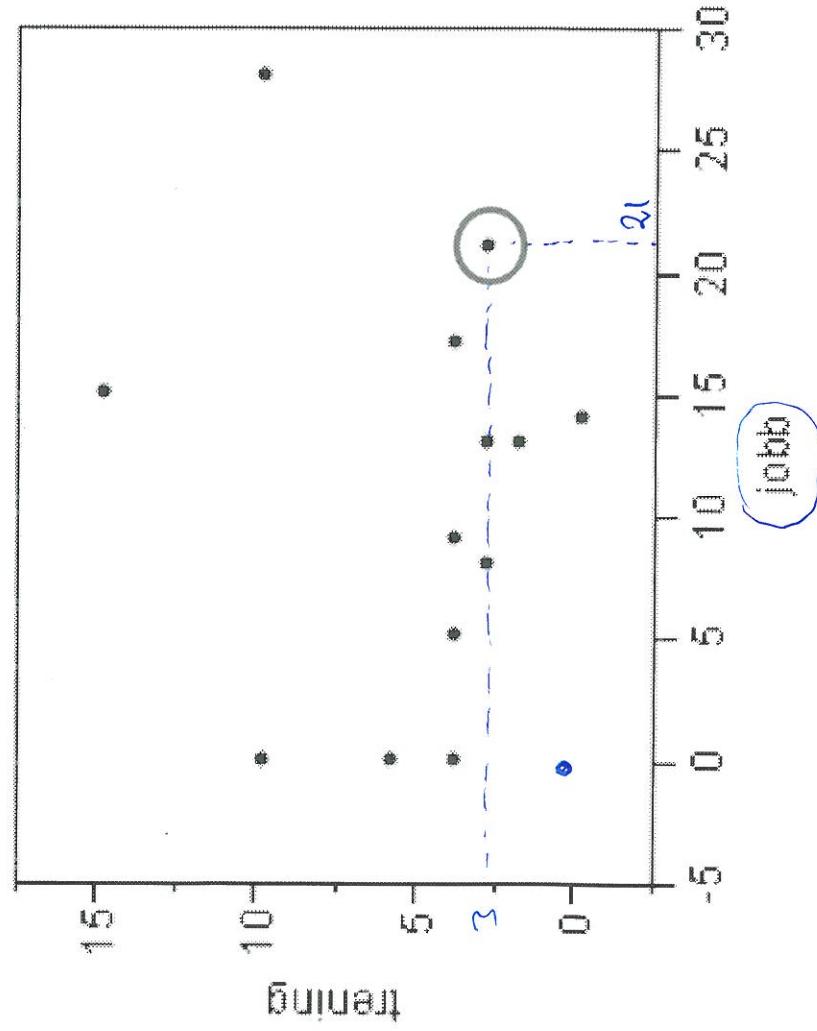
⑤ $(-3, 1)$

⑥ $(-2, -2)$



Spredningsplott

t/uke

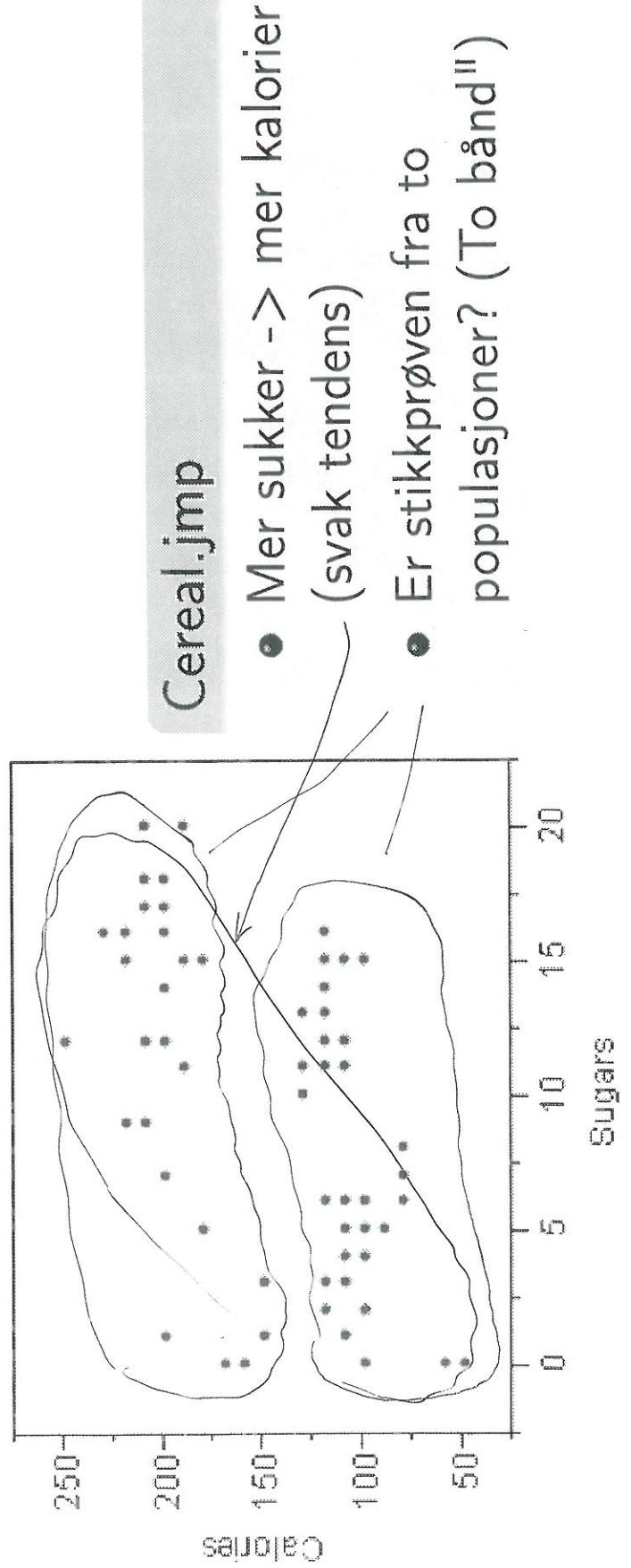


En person...

Jobber 21 timer i uka og
trener 3 timer i uka.

t/uken

Sammenheng mellom sukker og kalorier



Fila cereal.jmp

(2) Oppgaver fra Kap. 2

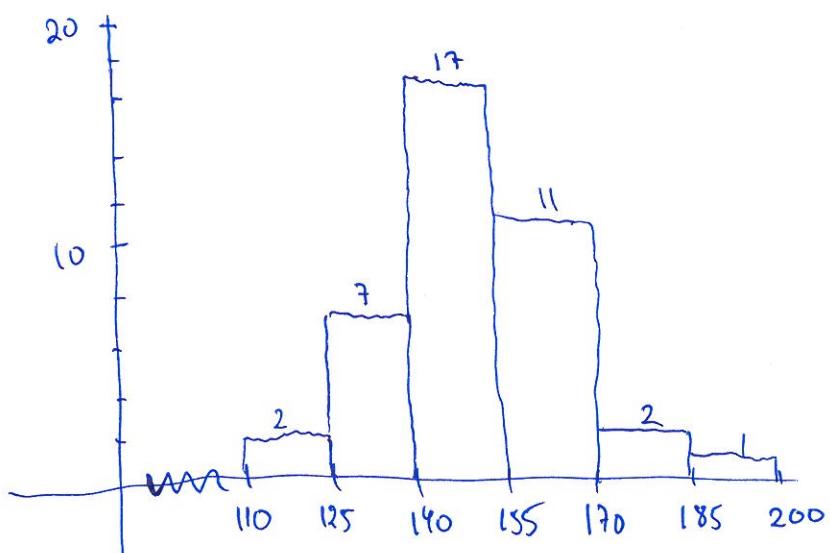
Datasett fra
2.2 Oppg. 18;

Frekvens -
tabell:

{ Intervallbredde
= 15 }

	frekvens	rel. frekvens
110-124	2	$2/40 = 5\%$
125-139	7	$7/40 = 17.5\%$
140-154	17	$17/40 = 42.5\%$
155-169	11	$11/40 = 27.5\%$
170-184	2	$2/40 = 5\%$
185-199	1	$1/40 = 2.5\%$
Sum	40	$40/40 = 100\%$

Histogram:



③

Sentral mål

Kap. 3.2

Hvor er senteret eller "midten" av dataene?

Nøkeltall: Sentral mål

- * gennomsnitt (aritmetisk gennomsnitt)
= (aritmetic) mean
- * median
- * modus = (mode)

- 1 Sec 3-2: Hvordan beskrive senteret i dataene
- 2 Sec 3-3: Hvordan beskrive spredningen i dataene

Todeling av statistikk

Deskriptiv statistikk

Oppsummering og beskrivelse av den stikkprøven du har.

Statistisk inferens

Handler om hvordan vi bruker stikkprøven til å si noe om hele populasjonen.

Mål for sentr i dataene

- Gjennomsnitt
- Median
- Modus

Notasjon

Σ står for å addere opp
 x er variabelen som representerer datavertiene. Ofte
brukes en indeks i for å angi dataverti nummer i : x_i

\bar{x} er gjennomsnittet.

n Stikkprøvens størrelse

N Populasjonens størrelse

Example

Notasjon gjør det enklere å oppgi formler. For eksempel så er
gjennomsnittet definert ved formelen

Formelen for
gjennomsnitt av
et utvalg:

$$\bar{x} = \frac{\sum x_i}{n}$$

x = variablene vi
ser på
 x_i = ~~de enkelte~~ verdiene til x
 $(x_1, x_2, x_3, x_4, \dots)$

$$\sum x_i = x_1 + x_2 + x_3 + \dots$$

Notasjon for gjennomsnitt

Gjennomsnitt i stikkprøven

Vi skriver \bar{x} for gjennomsnittet i stikkprøven:

$$\bar{x} = \frac{\sum x}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Gjennomsnitt i populasjonen

Vi skriver μ for gjennomsnittet i populasjonen:

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

*Gjennomsnitt
(my) ↗* *Summen av alle x-verdier
i hele populasjonen*

Merk at μ er som oftest ukjent.

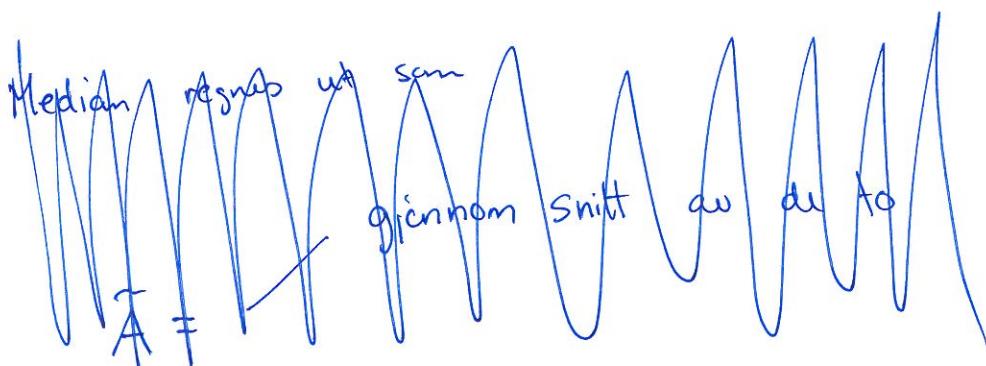
Hvor mange desimaler?

Når man oppgir gjennomsnittet, så bruk 1 mer desimal enn det er i dataene.

Eks:

	A = Alder
1	11
2	15
3	23
4	26
5	19
6	22
7	15
8	25
9	24

$$\text{Regn ut } \bar{A} = \frac{\sum A_i}{n} = \frac{A_1 + A_2 + \dots + A_9}{9}$$
$$= \frac{11 + 15 + 23 + \dots + 24}{9} = \frac{180}{9} = 20,0$$



Median

Medianen

Medianen er den midterste verdien i dataene når vi har sortert dem.

- Symbolet for medianen er \tilde{x} (x tilde)
- Medianen blir ikke så påvirket av uteliggere som gjennomsnittet

Example

Alder på ledere i en bedrift: 35 32 55 48 41 47

32 35 41 47 48 55

$$\tilde{x} = \text{median} = \frac{41 + 47}{2} = 44.0$$

Median

Medianen

Medianen er den midterste verdien i dataene når vi har sortert dem.

- Symbolet for medianen er \tilde{x}
- Medianen blir ikke så påvirket av uteliggere som gjennomsnittet

 - midterste verdie ved odder antall observations
- gjennomsnitt av de to midterste verdier ved partall antall observations

Example

Alder på ledere i en bedrift: 35 32 55 48 41 47

- Sortert: 32, 35, 41, 47, 48, 55
- De to midterste verdiene er 41 og 47
- Median $\tilde{x} = (41 + 47)/2 = 44$ år

Medianen er robust mot ekstreme observasjoner

Uteligger:

En ekstrem verdi, dvs. et tall som ligger langt ifra "hopen" av tall i stikkprøven. Påvirker gjennomsnittet mer enn medianen.

Example

Årsinntekt (i 1000 kr) i en bygd med 10 voksne

567	320	432	410	340	276	389	220	693	360
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Gjennomsnitt $\bar{x} = 400.7$ om median $\tilde{x} = (360 + 389)/2 = 374.5$.

- En ekstremt rik person flytter hjem til bygda.
- Han har årsinntekt på 3500 (3.5 millioner).
- Gjennomsnittet 400.7 \rightarrow 682.5. Stor forandring
- Median 374.5 \rightarrow 389. Liten forandring
- Her gir medianen en bedre pekepinn enn gjennomsnittet på hva en typisk inntekt er

Modus

Modus

Er den verdien som opptrer oftest. Brukes helst på kvalitative data.

Examples

- Det partiet som fikk mest stemmer...

- To stikkprøver og deres moduser.

A.	5.40	1.10	0.42	0.73	0.48	1.10	←	Modus = 1.1
B.	27	27	55	55	55	88	88	99