

Oppsummering av statistikk 2P/2PY

Innhold:

Frekvenstabeller

Sentralmål og spredningsmål

Gruppedelt materiale

Diagrammer

FREKVENSTABELLER

Frekvenstabeller: i en frekvenstabell teller vi opp hvor mange det er av hver verdi, dette antallet er det vi mener med frekvens. Tabellen kan man for eksempel bruke til å beregne relativ frekvens, kumulativ frekvens, relativ frekvens i prosent og relativ kumulativ frekvens.

Relativ frekvens er det samme som å finne prosentfaktor:

$$\text{relativ frekvens} = \frac{\text{frekvens}}{\text{sum}}$$

Ved å gange med 100% gjør vi om fra prosentfaktor til prosent (altså relativ frekvens i prosent):

$$\text{relativ frekvens i prosent} = \text{relativ frekvens} \cdot 100\%$$

Kumulativ frekvens finner vi ved å samle sammen frekvensene fra toppen og nedover. **Relativ kumulativ frekvens** finner du ved å samle opp relativ frekvens fra toppen og nedover eller du kan bruke den kumulative frekvensen når du deler på summen:

$$\text{relativ kumulativ frekvens} = \frac{\text{kumulativ frekvens}}{\text{sum}}$$

Eksempel:

Trine har spurt 20 elever i klassen hvor mange ganger de hjalp til med husarbeid hjemme i løpet av en uke. Resultatene ser du nedenfor.

10 14 14 8 5 10 10 10 0 8 8 7 7 0 8 7 7 7 5 7

Tegn av og fyll ut tabellen nedenfor.
Forklar hva tallene i nest siste rad i tabellen betyr

Antall ganger en elev hjelper til med husarbeid hjemme	Frekvens	Kumulativ frekvens	Relativ frekvens	Relativ kumulativ frekvens
0				
5				
7				
8				
10				
14				

Teller opp antall elever for hver verdi gitt ovenfor og fyller inn tabellen.

Antall ganger en elev hjelper til med husarbeid hjemme	Frekvens	Kumulativ frekvens	Relativ frekvens	Relativ kumulativ frekvens
0	2	2	$2/20 = 0,10$	0,10
5	2	$2+2=4$	$2/20 = 0,10$	$0,10+0,10=0,20$
7	6	$4+6=10$	$6/20=0,30$	$0,20+0,30=0,50$
8	4	$10+4=14$	$4/20=0,20$	$0,50+0,20=0,70$
10	4	$14+4=18$	$4/20=0,20$	$0,70+0,20=0,90$
14	2	$18+2=20$	$2/20=0,10$	$0,90+0,10=1,00$
SUM	$N=20$		1	

Tallene i nest nederste rad:

- frekvens viser at det er 4 elever som hjelper til hjemme 10 ganger i løpet av uka.
- kumulativ frekvens viser at det er 18 elever som hjelper til hjemme 10 ganger eller mindre i løpet av uka.
- relativ frekvens viser at det er 0,20 eller 20% av elevene som hjelper til hjemme 10 ganger i uka.
- relativ kumulativ frekvens viser at det er 0,90 eller 90% av elevene som hjelper til hjemme 10 ganger eller mindre i løpet av uka.

SENTRALMÅL OG SPREDNINGSMÅL

Gjennomsnitt:

$$g = \frac{\text{Summen av alle tallene}}{\text{Antall verdier}}$$

Typetall er det tallet det er flest av, altså den verdien som har størst frekvens om du har det i en frekvenstabell.

Medianen er tallet i midten når vi sorterer verdiene i stigende rekkefølge. Er det flere tall i midten tar du gjennomsnittet av disse to.

Variasjonsbredde er den største verdien minus den laveste verdien.

$$\text{Variasjonsbredde} = \text{størst} - \text{minst}$$

Kvartilbredde finner vi ved å først sortere i stigende rekkefølge, så dele tallmaterialet i to like store deler og til slutt dele disse to delene i to igjen. Da skal du ha fire like store deler. Tallet i midten av nederste del er nedre kvartil, tallet i midten av øverste del er øvre kvartil.

$$\text{Kvartilbredde} = \text{øvre kvartil} - \text{nedre kvartil}$$

Standardavviket sier noe om hvor langt verdiene er fra gjennomsnittet, altså hvor stor spredning det er i tallmaterialet. Formelen for standardavvik kan se vanskelig ut, men vanligvis bes man ikke om å regne ut standardavviket på del 1. På del 2 kan dere bruke statistikk i Geogebra. I formelen er x verdien du måler (f.eks. karakteren), g er gjennomsnittet, f er frekvensen for denne verdien og N er antall verdier.

$$\text{Standardavvik} = \sqrt{\left(\frac{(x - g)^2 \cdot f}{N}\right)}$$

Eksempel:

Trine har spurt 20 elever i klassen hvor mange ganger de hjalp til med husarbeid hjemme i løpet av en uke. Resultatene ser du nedenfor.

10 14 14 8 5 10 10 10 0 8 8 7 7 0 8 7 7 7 5 7

Bestem medianen, gjennomsnittet, typetallet og variasjonsbredden for dette datamaterialet.

Bestem også kvartilbredden og standardavviket.

Sorterer tallene i stigende rekkefølge:

0, 0, 5, 5, 7, 7, 7, 7, 7, 7, **7, 8**, 8, 8, 8, 10, 10, 10, 10, 14, 14

Med 20 verdier er medianen på plass 10 og 11, der står tallene 7 og 8. Finner gjennomsnittet av disse to:

$$\text{Medianen er } \frac{7 + 8}{2} = 7,5$$

Øvre kvartil er 10 (to tiere i midten) og nedre kvartil er 7 (to syvere i midten), da er kvartilbredden:

$$\text{Øvre kvartil} - \text{nedre kvartil} = 10 - 7 = 3$$

Gjennomsnittet:

$$\begin{aligned} g &= \frac{0 + 0 + 5 + 5 + 7 + 7 + 7 + 7 + 7 + 7 + 7 + 8 + 8 + 8 + 8 + 10 + 10 + 10 + 10 + 14 + 14}{20} \\ &= \frac{0 + 2 \cdot 5 + 6 \cdot 7 + 4 \cdot 8 + 4 \cdot 10 + 2 \cdot 14}{20} = 7,6 \end{aligned}$$

Typetall er den verdien det er flest av, altså 7.

Variasjonsbredden:

$$14 - 0 = 14$$

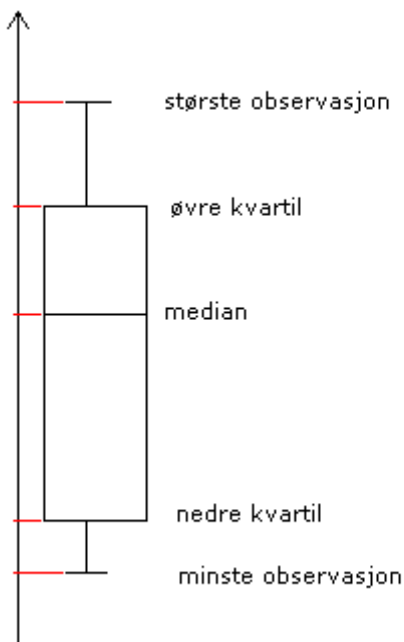
Standardavviket:

Antall ganger husarbeid	Frekvens	$(x - g)^2 \cdot f$
0	2	$(0 - 7,6)^2 \cdot 2 = 115,52$
5	2	$(5 - 7,6)^2 \cdot 2 = 13,52$
7	6	$(7 - 7,6)^2 \cdot 6 = 0,72$
8	4	$(8 - 7,6)^2 \cdot 4 = 0,32$
10	4	$(10 - 7,6)^2 \cdot 4 = 11,52$
14	2	$(14 - 7,6)^2 \cdot 2 = 81,92$
SUM	20	223,52

$$SD = \sqrt{\frac{223,52}{20}} = 3,34$$

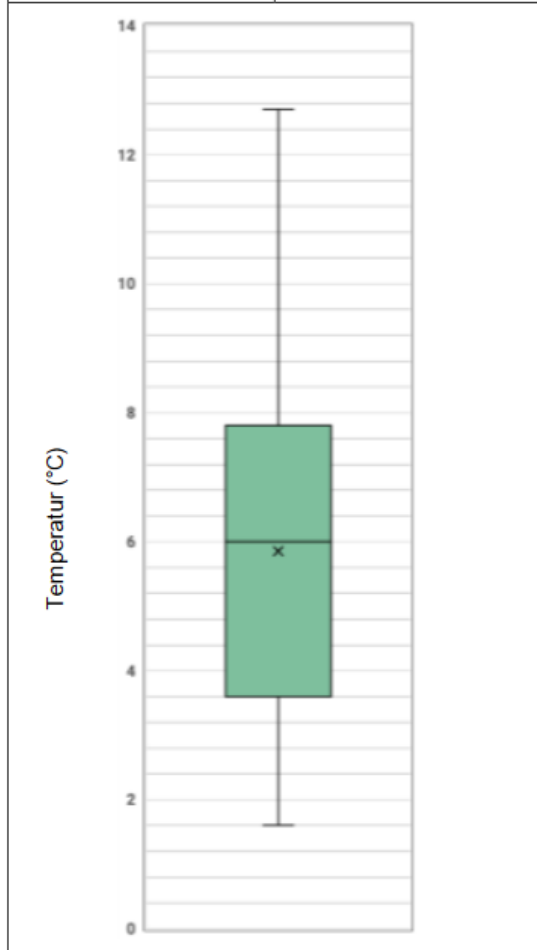
Sentralmål og spredningsmål i boksplott

Et boksplott deler tallmaterialet i fire like store deler akkurat som vi gjør for å finne kvartilbredden. Boksen i midten inneholder 50% av verdiene og streken i denne markerer medianen. Øvre ende av boksen viser øvre kvartil, nedre ende viser nedre kvartil. Linjene i hver retning fra boksen viser til største og minste verdi.

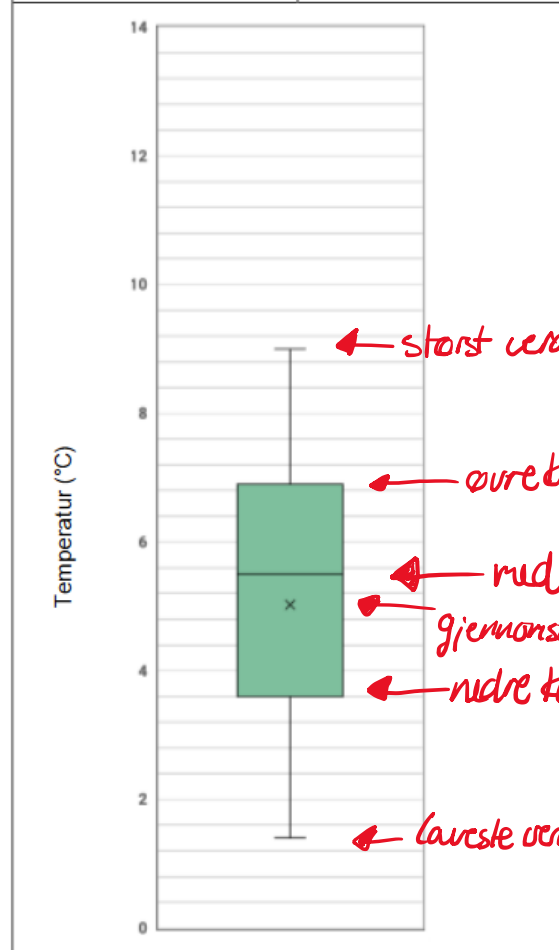


Inge har sett på maksimumstemperaturen i Trondheim og Oslo hvert døgn i januar 2020 og laget tabellene og diagrammene nedenfor.

Trondheim	
Median:	6,0 °C
Gjennomsnitt:	5,9 °C
Variasjonsbredde:	11,1 °C
Kvartilbredde:	4,2 °C



Oslo	
Median:	5,5
Gjennomsnitt:	5
Variasjonsbredde:	$9 - 1,4 = 7,6$
Kvartilbredde:	$6,9 - 3,6 = 3,3$



a) Bruk diagrammet til høyre til å bestemme medianen, gjennomsnittet, variasjonsbredden og kvartilbredden for maksimumstemperaturene i Oslo. Forklar hvordan du finner disse verdiene.

b) Forklar hvordan vi kan se ut fra diagrammene at det i begge byene var omtrent 8 dager med en maksimumstemperatur på 3,6 °C eller lavere.

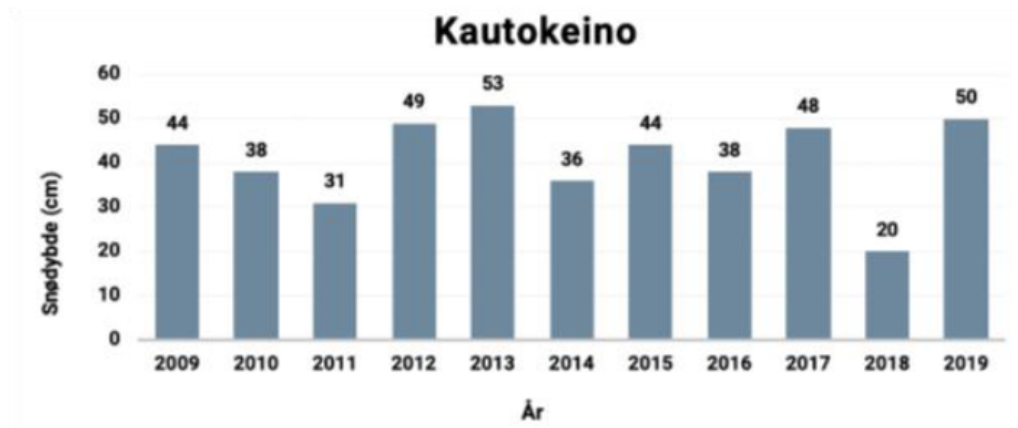
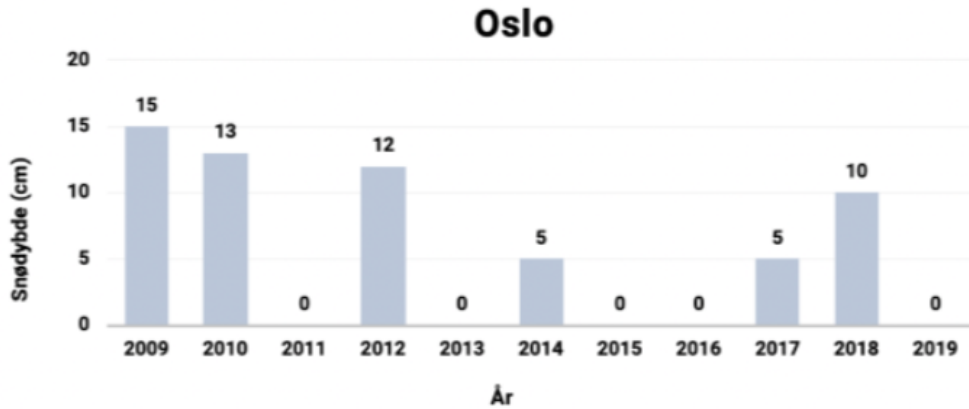
b) Diagrammet viser data for hele januar, altså 31 dager. Deler vi dette i fire deler skal én del være under nedre kvartil, hver del tilsvarer 8 dager. Nedre kvartil er 3,6 for begge byer og det er da åtte dager under denne temperaturen i begge byer.

Anja Lyng Bækken

Gjennomsnitt fra frekvenstabell:

Gjennomsnitt og standardavvik på del 2 – Geogebra

Diagrammene nedenfor viser snødybden i Oslo og i Kautokeino julaften de 11 siste årene.



- a) Bestem gjennomsnittet og standardavviket for snødybdene i Oslo og for snødybdene i Kautokeino.

Etter å ha regnet ut gjennomsnittet for Oslo og Kautokeino kom Isak med følgende påstand:

«Siden gjennomsnittet for Kautokeino ble høyere enn gjennomsnittet for Oslo, må standardavviket for Kautokeino også bli høyere enn standardavviket for Oslo. Det er alltid slik at det datamaterialet som har høyest gjennomsnitt, også har høyest standardavvik.»

- b) Er påstanden riktig? Begrunn svaret ditt.

Gjennomsnitt og median i klassedelt materiale

Tabellen nedenfor viser en oversikt over høydene til elevene ved en skole.

Høyde i cm	Frekvens
[150,160)	10
[160,170)	30
[170,180)	50
[180,200)	10

- Bestem gjennomsnittshøyden.
- I hvilken klasse ligger medianen? Anslå en omtrentlig verdi, hvilke antakelser gjør du?
- Lag et histogram som viser fordelingen av høydene.

a) Finner først klassemidtpunkt i hver av gruppene:

Høyde i cm	Frekvens	Klassemidtpunkt
[150,160)	10	155
[160,170)	30	165
[170,180)	50	175
[180,200)	10	190
SUM	100	

Tar så klassemidtpunkt og ganger med frekvensene før vi legger sammen og deler på antall verdier:

$$g = \frac{10 \cdot 155 + 30 \cdot 165 + 50 \cdot 175 + 10 \cdot 190}{100} = 171,5$$

b)

1. Medianen skal ligge på plass 50 og 51 ettersom det er 100 deltakere totalt. Vi teller oss frem til at dette i gruppen 170 til 180 cm. Ettersom det er 40 personer før denne gruppen så skal vi henholdsvis 10 og 11 plasser inn i denne gruppen.

2. Klassebredden er 10 cm og det er 50 personer i denne gruppen. Antar vi at høyden til personene er jevnt fordelt i dette intervallet kan vi så finne hvor mange cm høyden øker per person ved å ta klassebredden delt på frekvensen:

$$\text{Økning per person i gruppen} = \frac{\text{Klassebredde}}{\text{Frekvens}} = \frac{10 \text{ cm}}{50} = 0,20 \text{ cm}$$

3. Da finner vi høyden til person 10 og 11 ved å ta nedre klassegrense og legge til høyden 0,20 cm 10 og 11 ganger:

$$\text{Person nr 50} = 170 + \frac{10}{50} \cdot 10 = 172 \quad \text{Person nr 51} = 170 + \frac{10}{50} \cdot 11 = 172,20$$

Medianen er da snittet av disse 172,1 cm.

c) På del 2 gjøres histogram i Geogebra. For histogram på del 1 må du først beregne søylehøyden:

Høyde i cm	Frekvens	Klassebredde	Søylehøyde
[150,160)	10	10	$10/10=1$
[160,170)	30	10	$30/10=3$
[170,180)	50	10	$50/10=5$
[180,200)	10	20	$20/10=2$

Diagrammer

Søylediagram

Kurve/linjediagram

Kakediagram/sektordiagram

Et idrettslag har 240 medlemmer. Idrettslaget har fire forskjellige aktivitetsgrupper.

Medlemmene fordeler seg slik:

Aktivitetsgruppe	Antall medlemmer
Langrenn	60
Hopp	40
Freestyle	80
Alpint	60

Gjør beregninger og lag et sektordiagram som viser fordelingen av medlemmene på de ulike gruppene. Det skal gå klart fram hvor mange grader hver av sektorene i diagrammet er på.