0.1 Introduksjon

I ei *undersøking* hentar vi inn informasjon. Denne informasjonen kan gjerne være tal eller ord, og kallast *data*. Ei samling av innhenta data kallast eit *datasett*.

For eksempel, tenk at du spør to mennesker om de iliker kaviar. Den eine svarer "ja", den andre "nei". Da er "ja" og "nei" dataa (svara) du har samla inn, og {"ja", "nei"} er datasettet ditt.

Statistikk handler grovt sett om to ting; *å presentere* og *å tolke* innsamla data. For begge desse formåla har vi nokre verktøy som vi i komande seksjonar skal studere ved hjelp av nokre forskjellige eksempel på undersøkingar. Desse finn du på side 2.

Det er ikkje nokre fullstendige fasitsvar på korleis ein presenterer eller tolker data, men to retningslinjer bør du alltid ha med deg:

- La det alltid komme tydeleg fram kva du har undersøkt, og kva data som er innhenta.
- Tenk alltid over kva metodar du bruker for å tolke dataa.

Språkboksen

Personar som deltek i ei undersøking der ein skal svare på noko, kallast *respondentar*.

Undersøking 1

10 personar testa kor mange sekund dei kunne halde pusten. Resultata blei desse:

47 124 61 38 97 84 101 79 56 40

Undersøking 2

15 personar blei spurd kor mange epler dei et i løpet av ei veke. Svara blei desse:

7 4 5 4 1 0 6 5 4 8 1 6 8 0 14

Undersøking 3

300 personar ble spurd kva deira favorittdyr er.

- 46 personer svarte tiger
- 23 personer svarte løve
- 17 personer svarte krokodille
- 91 personer svarte hund
- 72 personer svarte katt
- 51 personer svarte andre dyr

Undersøking 4

Mobiltelefonar med smartfunksjonar (app-baserte) kom på det norske markedet i 2009. Tabellen[1] under viser det totale salget mobiltelefonar i tidsperioden 2009-2014, og andelen med og utan smartfunkskjonar.

År	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**
totalt	2 365	2 500	2 250	2 200	2 400	2 100
u. sm.f.	1 665	1 250	790	300	240	147
m. sm.f.	700	1 250	1 460	1 900	2 160	1 953

[1]Tala er henta frå medienorge.uib.no.

0.2 Presentasjonsmetoder

Skal vi presentere våre undersøkingar, bør vi vise datasetta slik at det er lett for andre å sjå kva vi har funne. Dette kan vi gjere blant anna ved hjelp av *frekvenstabellar*, *søylediagram*, *sektordiagram* eller *linjediagram*.

0.2.1 Frekvenstabell

I ein frekvenstabell sett ein opp dataa i ein tabell som viser kor mange gongar kvart unike svar dukkar opp. Dette antalet kallast *frekvensen*.

Undersøking 2

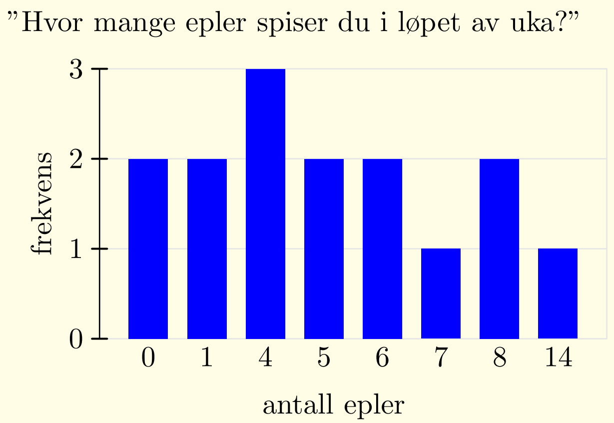
I vår undersøking har vi to 0, to 1, tre 4, to 5, to 6, én 7, to 8 og én 14. I ein frekvenstabell skriv vi da

antall epler	frekvens
0	2
1	2
4	3
5	2
6	2
7	1
8	2
14	1

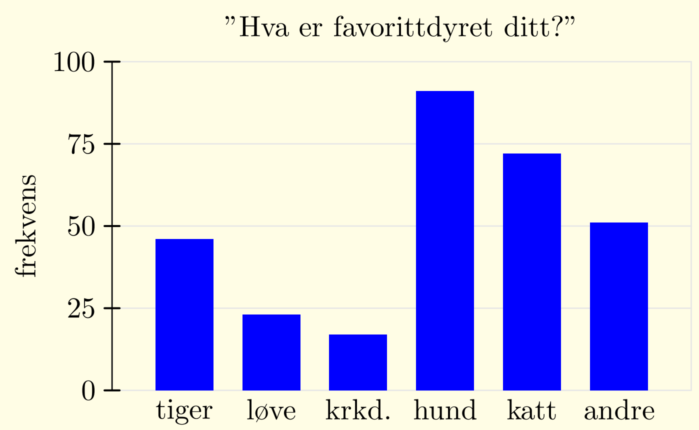
0.2.2 Søylediagram (stolpediagram)

Med eit søylediagram presenterer vi dataa med søyler som viser frekvensen.

Undersøking 2



Undersøking 3

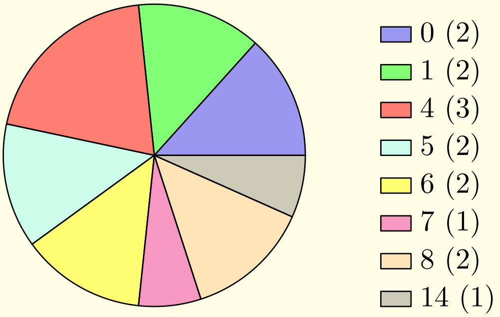


0.2.3 Sektordiagram (kakediagram)

I eit sektordiagram visast frekvensane som sektorar av ein sirkel.

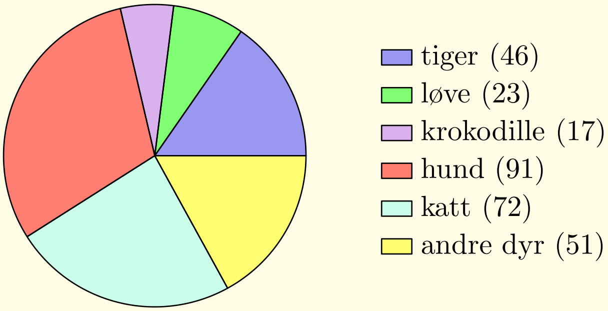
Undersøking 2

Epler spist i løpet av uka (frekvens i parantes)



Undersøking 3

Favorittdyr (frekvens i parantes)



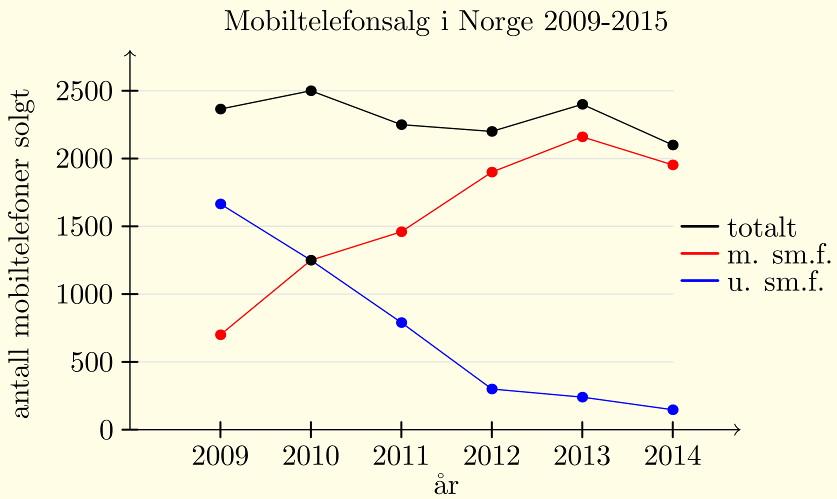
Å lage et sektordiagram for hand

Skal du sjølv teikne eit sektordiagram, treng du kunnskapar om vinklar og om brøkandelar. Sjå *Seksjon* **??**, MB, s. 76 og oppgåve **??**.

0.2.4 Linjediagram

I eit linjediagram legg vi inn dataa som punkt i eit koordinatsystem, og trekk ei linje mellom dei. Linjediagram brukast oftast når det er snakk om ei form for utvikling.

Undersøking 4



0.3 Tolking av tendenser; sentralmål

I eit datasett er det gjerne svar som er heilt eller tilnærma like, og som gjentar seg. Dette betyr at vi kan seie noko om hva som gjelder for mange; ein *tendens*. Dei matematiske omgrepa som fortel noko om dette kallast *sentralmål*. Dei vanlegaste sentralmåla er *typetal*, *gjennnomsnitt* og *median*.

0.3.1 Typetal

0.1 Typetal

Typetalet er verdien det er flest eksemplar av i datasettet.

Undersøking 2

I datasettet er det verdien 4 som opptrer flest (tre) gongar. (Dette kan vi ssjå frå sjølve datasett på s. 2, frå frekvenstabellen på s. 3, frå søylediagrammet på s. 4 eller sektordiagrammet på s. 5.)

4 er altså typetallet.

0.3.2 Gjennomsnitt

Når eit datasett består av svar i form av tal, kan vi finne summen av svara. Når vi spør kva gjennomsnittet er, spør vi om dette:

"Vis alle svara var like, og summen den same, kva verdi måtte alle svarene da ha hatt?"

Dette er jo ingenting anna enn divisjon[1]:

0.2 Gjennomsnitt

$$\text{gjennomsnitt} = \frac{\text{summen av verdiane frå datasettet}}{\text{antall verdier}}$$

Undersøking 1

Vi summerer verdiane frå datasettet, og deler med antall verdiar:

$$\begin{aligned}\text{gjennomsnitt} &= \frac{47 + 124 + 61 + 38 + 97 + 84 + 101 + 79 + 56 + 40}{10} \\ &= \frac{727}{10} \\ &= 72{,}7\end{aligned}$$

Altså, i gjennomsnitt heldt dei 10 deltakarane pusten i 72,7 sekund.

[1]sjå MB, s. 23.

Undersøking 2

Metode 1

$$
\begin{aligned}
\text{gjennomsnitt} &= \frac{7+4+5+4+1+0+6+5+4+8+1+6+8+0+14}{15} \\
&= \frac{73}{15} \\
&\approx 4.87
\end{aligned}
$$

Metode 2

Vi utvidar frekvenstabellen frå side 3 for å finne summen av verdiene frå datasettet (vi har også tatt med summen av frekvensane):

Antall epler	Frekvens	antall · frekvens
0	2	$0 \cdot 2 = 0$
1	2	$1 \cdot 2 = 2$
4	3	$4 \cdot 3 = 12$
5	2	$5 \cdot 2 = 10$
6	2	$6 \cdot 2 = 12$
7	1	$7 \cdot 1 = 14$
8	1	$8 \cdot 2 = 16$
14	1	$14 \cdot 1 = 14$
sum	15	73

No har vi at

$$
\begin{aligned}
\text{gjennomsnitt} &= \frac{73}{15} \\
&\approx 4{,}87
\end{aligned}
$$

Altså, i gjennomsnitt et dei 15 respondentane 4,87 epler i veka.

Undersøking 4

(Utrekning utelatt. Verdiane er runda ned til næraste éinar).

- Gjennomsnitt for totalt salg av mobilar: 2302
- Gjennomsnitt for salg av mobilar uten smartfunksjon: 732
- Gjennomsnitt for salg av mobilar med smartfunksjon: 1570

0.3.3 Median

0.3 Median

Medianen er talet som ender opp i midten av datasettet når det rangerast frå talet med lågast til høgst verdi.

Hvis datasettet har partalls antal verdiar, er medianen gjennomsnittet av de to verdiane i midten (etter rangering).

Undersøking 1

Vi rangerer datasettet frå lågast til høgst verdi:

$$38 \quad 40 \quad 47 \quad 56 \quad 61 \quad 79 \quad 84 \quad 97 \quad 101 \quad 124$$

Dei to tallene i midten er 61 og 79. Gjennomsnittet av desse er

$$\frac{61+79}{2} = 70$$

Altså er medianen 70.

Undersøking 2

Vi rangerer datasettet frå lågast til høgst verdi:

$$0 \quad 0 \quad 1 \quad 1 \quad 4 \quad 4 \quad 4 \quad 5 \quad 5 \quad 6 \quad 6 \quad 7 \quad 8 \quad 8 \quad 14$$

Tallet i midten er 5, altså er medianen 5.

Undersøking 4

(Utrekning utelatt. Verdiane er runda ned til næraste éner).

- Median for totalt salg av mobilar: 2307
- Median for salg av mobilar utan smartfunksjon: 545
- Median for salg av mobilar med smartfunksjon: 1570

0.4 Tolking av forskjellar; spreiingsmål

Ofte vil det også vere store forskjellar (stor spreiing) mellom dataa som er samla inn. Dei vanlegaste matematiska omgrepa som forteljer noko om dette er *variasjonsbredde*, *kvartilbredde*, *varians* og *standardavvik*.

0.4.1 Variasjonsbredde

0.4 Variasjonsbredde

Differansen mellom svara med høvesvis høgst og lågast verdi.

Undersøking 1

Svaret med høvesvis høgst og lågast verdi er 124 og 38. Altså er

$$\text{variasjonsbredde} = 124 - 38 = 86$$

Undersøking 2

Svaret med henholdsvis høgst og lågast verdi er 14 og 0. Altså er

$$\text{variasjonsbredde} = 14 - 0 = 14$$

Undersøking 4

- Variasjonsbredde for mobilar totalt:

$$2\,500 - 2\,100 = 400$$

- Variasjonsbredde for mobilar uten smartfunksjoner:

$$1\,665 - 147 = 518$$

- Variasjonsbredde for mobilar med smartfunksjoner:

$$2\,160 - 700 = 1460$$

0.4.2 Kvartilbredde

0.5 Kvartilbredde og øvre og nedre kvartil

Kvartilbredden til et datasett kan finnes på følgende måte:

1. Ranger datasettet frå høgst til lågast verdi.
2. Skil det rangerte datasettet på midten, slik at to nye sett oppstår. (Viss det er oddetalls antal verdiar i datasettet, utelatast medianen).
3. Finn dei respektive medianane i dei to nye setta.
4. Finn differansen mellom medianane frå punkt 3.

Om medianene frå punkt 3: Den med høgst verdi kallast *øvre kvartil* og den med lågast verdi kallast *nedre kvartil*.

Undersøking 1

1. 38 40 47 56 61 79 84 97 101 124
2. 38 40 47 56 61 79 84 97 101 124
3. Medianen i det blå settet er 47 (nedre kvartil) og medianen i det røde settet er 97 (øvre kvartil).

   38 40 47 56 61 79 84 97 101 124

4. Kvartilbredde $= 97 - 47 = 50$

Undersøking 2

1. 0 0 1 1 4 4 4 5 5 6 6 7 8 8 14
2. 0 0 1 1 4 4 4 5 5 6 6 7 8 8 14
3. Medianen i det blå settet er 1 (nedre kvartil) og medianen i det raude settet er 7 (øvre kvartil).

   0 0 1 1 4 4 4 5 6 6 7 8 8 14

4. Kvartilbredde $= 7 - 1 = 6$

Undersøking 4

(Utrekning utelatt)

- For mobilar totalt er kvartilbredden: 200
- For mobilar uten smartfunksjoner er kvartilbredden: 1010
- For mobilar med smartfunksjoner er kvartilbredden: 703

Språkboksen

Nedre kvartil, medianen og øvre kvartil blir også kalla høvesvis *1. kvartil*, *2. kvartil* og *3. kvartil*.

0.4.3 Avvik, varians og standardavvik

0.6 Varians

Differansen mellom en verdi og gjennomsnittet i et datasett kalles *avviket* til verdien.

Variansen til et datasett kan finnes på følgende måte:

1. Kvadrer avviket til hver verdi i datasettet, og summer disse.

2. Divider med antall verdier i datasettet.

Standardavviket er kvadratroten av variansen.

Eksempel

Gitt datasettet

$$2 \quad 5 \quad 9 \quad 7 \quad 7$$

Da har vi at

$$\text{gjennomsnitt} = \frac{2+5+9+7+7}{5} = 6$$

Og videre er

$$\text{variansen} = \frac{(2-6)^2 + (5-6)^2 + (9-6)^2 + (7-6)^2 + (7-6)^2}{5}$$

$$= 5$$

Da er standardavviket $= \sqrt{5} \approx 2{,}23$.

Undersøking 1

(Utrekning utelatt)

Variansen er 754,01. Standardavviket er $\sqrt{754{,}01} \approx 27{,}46$

Undersøking 2

Gjennomsnittet fant vi på side 9. Vi utvidar frekvenstabellen vår frå side 3:

antall epler	frekvens	frekvens · kvadrert avvik
0	2	$2 \cdot \left(0 - \frac{73}{15}\right)^2$
1	2	$2 \cdot \left(1 - \frac{73}{15}\right)^2$
4	3	$3 \cdot \left(4 - \frac{73}{15}\right)^2$
5	2	$2 \cdot \left(5 - \frac{73}{15}\right)^2$
6	2	$2 \cdot \left(6 - \frac{73}{15}\right)^2$
7	1	$1 \cdot \left(7 - \frac{73}{15}\right)^2$
8	2	$2 \cdot \left(8 - \frac{73}{15}\right)^2$
14	1	$1 \cdot \left(9 - \frac{73}{15}\right)^2$
sum	15	$189{,}7\overline{3}$

Altså er variansen

$$\frac{189{,}7\overline{3}}{15} \approx 12{,}65$$

Da er standardavviket $\sqrt{12{,}65} \approx 3.57$

Undersøking 4

(Utrekning utelatt)

- For mobilar totalt er variansen $17\,781{,}25$ og standardavviket ca. 133,4.
- For mobilar uten smartfunksjoner er variansen $318\,848.\overline{3}$ og standardavviket ca. 17,87
- For mobilar med smartfunksjoner er variansen $245\,847.91\overline{6}$ og standardavviket ca. 495,83.

Hvorfor innebærer variansen kvadrering?

La oss se hva som skjer hvis vi gjentar utrekningen frå *Eksempel 1* på side 15, men uten å kvadrere:

$$\frac{(2-6)+(5-6)+(9-6)+(7-6)+(7-6)}{5}$$
$$= \frac{2+5+9+7+7}{5} - 6$$

Men brøken $\frac{2+5+9+7+7}{5}$ er jo per definisjon gjennomsnittet til datasettet, og dermed blir uttrykket over lik 0. Dette vil gjelde for alle datasett, så i denne sammenhengen gir ikkje tallet 0 noen ytterligere informasjon. Om vi derimot kvadrerer avvikene, unngår vi et uttrykk som alltid blir lik 0.