1 Statistiske undersøkelser

Overskriftene viser til delkapitlene i boka. Først kommer en oppsummering av hele kapittelet. Under hver overskrift står gjennomregnede eksempler med løsning, og deretter oppgaver du skal løse uten hjelpemidler. Fasiten står bakerst.

Formelark

DEN STATISTISKE UNDERSØKELSESPROSESSEN

Statistisk spørsmål Besvares med data som **varierer**. Sier hvem det gjelder, og hva som måles.

Undersøkelsesprosessen

1. Formulere spørsmål
2. Samle inn data
3. Analysere data
4. Tolke resultatene

- Hvordan dataene er samlet inn, avgjør hva du kan konkludere med.

VARIABLER OG DATASETT

Datasett Én rad per **observasjon**, én kolonne per **variabel**.

Typer variabler **Kategorisk**: kategorier. **Numerisk**: tall, **diskret** (telles) eller **kontinuerlig** (måles).

Andel

$$\frac{\text{antall i kategorien}}{\text{antall observasjoner}}$$

- Gjennomsnitt bare for numeriske variabler. Postnumre og draktnumre er kategoriske.

DATAKILDER

Primær- og sekundærdata **Primær**: samlet inn selv. **Sekundær**: samlet inn av andre.

Prosentpoeng

$$\text{ny andel} - \text{gammel andel}$$

Endring i prosent

$$\frac{\text{ny} - \text{gammel}}{\text{gammel}} \cdot 100\,\%$$

Vurdere en kilde Hvem, hvorfor, hvordan, når og hva?

- Et manglende tall i en tabell betyr ikke 0.

DATA I PYTHON

Relativ frekvens

$$\frac{\text{frekvens}}{n}$$

Rydde data Skrivemåte, urimelige verdier, manglende verdier.

- Frekvensene summerer til n, de relative frekvensene til 1.
- En verdi fjernes ikke fordi den er stor, bare fordi den er feil.

UTVALG

Populasjon og utvalg p i populasjonen, $\hat{p}$ i utvalget. Resultatet varierer fra utvalg til utvalg, mindre når n er stort.

Systematisk utvalg

$$k = \frac{N}{n}$$

Stratifisert utvalg

$$\frac{n}{N} \cdot \text{antall i laget}$$

Utvalgsmetoder Tilfeldige: enkel, systematisk, stratifisert, klynge. Ikke tilfeldige: bekvemmelighet, selvseleksjon.

- Stratifisert: noen fra alle lag. Klynge: alle fra noen klynger.

FEILKILDER OG EKSPERIMENTER

Svarprosent

$$\frac{\text{svarte}}{\text{spurt}} \cdot 100\,\%$$

Feilkilder Dekning, frafall, selvseleksjon. Ledende spørsmål, sosial ønskelighet, hukommelse, målefeil.

Tilfeldig og systematisk feil Bare tilfeldig feil blir mindre med større utvalg.

Observasjon og eksperiment Observasjonsstudie: sammenheng. Randomisert eksperiment med kontrollgruppe: årsak.

- Et stort utvalg er ikke nødvendigvis representativt.
- Tilfeldig utvalg: si noe om populasjonen. Tilfeldig fordeling: si noe om årsak.

UNDERSØKELSEN

De fire stegene Spørsmål, data, analyse, tolkning. Skriv ned valgene og ryddingen, og si hvem resultatet gjelder for.

1.1.1 Fra spørsmål til svar

EKSEMPEL *Statistisk eller ikke?*

Hvilke av spørsmålene er statistiske?

a) Hvor høy er du?

b) Hvor høye er elevene på Vg2 på skolen?

c) Hva koster et månedskort på bussen for ungdom i byen din?

d) Hvor mye bruker elevene i klassen på mat i kantina i løpet av en uke?

Løsning

a) Ikke statistisk. Det har ett svar.

b) Statistisk. Høyden varierer fra elev til elev, og vi må måle mange elever.

c) Ikke statistisk. Prisen er fastsatt, og vi kan slå den opp.

d) Statistisk. Beløpet varierer fra elev til elev og fra uke til uke.

EKSEMPEL *Gjør spørsmålet statistisk*

Gjør om spørsmålet «Hvor lang tid bruker jeg på skoleveien?» til et statistisk spørsmål.

Løsning

For eksempel: «Hvor mange minutter bruker elevene på skolen vår på skoleveien en vanlig morgen?» Svaret varierer fra elev til elev, og spørsmålet sier hvem det gjelder (elevene på skolen) og hva som måles (minutter på skoleveien).

EKSEMPEL *Les av prikkplottet*

Bruk prikkplottet av søvnen i klassen fra starten av delkapittelet.

a) Hvor mange elever sov 8 timer eller mer?

b) Hvor mange prosent av klassen er det?

c) Hvor stor er forskjellen mellom den som sov lengst, og den som sov kortest?

Løsning

a) $3 + 1 + 1 + 1 = 6$ elever.

b) $\frac{6}{24} \cdot 100\,\% = 25\,\%$.

c) $9{,}5 - 4{,}5 = 5$ timer.

EKSEMPEL *Søvnundersøkelsen*

Elevrådet vil vite om elevene på skolen sover for lite. Beskriv hvordan de kan gå gjennom de fire stegene.

Løsning

1. **Spørsmål:** «Hvor mange timer sover elevene på skolen vår natt til en skoledag, og hvor stor andel sover minst 8 timer?»
2. **Data:** De lager et kort, anonymt spørreskjema: «Når la du deg, og når sto du opp i dag?» Det sendes til elever fra alle trinn og linjer, ikke bare til elevrådets egne klasser.
3. **Analyse:** De regner ut søvnen for hver elev, lager et prikkplott, finner gjennomsnittet og andelen som sov minst 8 timer.
4. **Tolkning:** De svarer på spørsmålet, for eksempel: «Omtrent én av fire sov minst 8 timer natt til tirsdag.» De skriver også hva som kan ha påvirket svaret: Dataene gjelder én natt, elevene har anslått tidene selv, og ikke alle svarte.

EKSEMPEL *Kritiske spørsmål til en påstand*

En nettavis skriver: «Ny undersøkelse: Ungdom bruker sju timer på mobilen hver dag.» Hvilke spørsmål bør du stille før du tror på påstanden?

Løsning

- **Spørsmål:** Hvem er «ungdom»? Gjelder det alle dager, eller bare helg? Regnes all skjermtid med, også skolearbeid?
- **Data:** Hvem ble spurt, og hvor mange? Er tiden målt med mobilens skjermtidsfunksjon, eller har ungdommene anslått den selv?
- **Analyse:** Er sju timer et gjennomsnitt? Da kan noen få som bruker mobilen svært mye, trekke tallet opp.
- **Tolkning:** Hvor mye varierer det? Sier undersøkelsen noe om ungdom i Norge, eller bare om dem som svarte?

1.1 Er spørsmålet statistisk? Begrunn kort.

a) Hvor mange søsken har du?

b) Hvor mange søsken har elevene på Vg2 på skolen?

c) Når åpner skolebiblioteket?

d) Hvor lenge er elevene på biblioteket når de først er der?

e) Hvor mye regn kom det i byen din i går?

1.2 Gjør om spørsmålene til statistiske spørsmål. Si hvem det gjelder, og hva som skal måles.

a) Hvor mange skritt gikk jeg i går?

b) Hvor mye koster lunsjen min i kantina?

c) Hvor gammel var mamma da hun tok lappen?

1.3 Hvilket steg i undersøkelsesprosessen hører aktiviteten til?

a) Du regner ut hvor mange prosent som svarte ja.

b) Du bestemmer at du vil undersøke hvor mye elevene bruker på klær.

c) Du spør 30 tilfeldige elever fra elevlista.

d) Du skriver at svaret bare gjelder elevene på Vg2.

1.4 Sett stegene i riktig rekkefølge.

A. Lage et prikkplott av svarene. B. Skrive at elevene som svarte, bruker i snitt 40 minutter på skoleveien, men at bare to klasser svarte. C. Spørre: «Hvor lang tid bruker elevene på skoleveien?» D. Dele ut et spørreskjema i to klasser.

1.5 20 elever skrev ned hvor mange timer de brukte på lekser forrige uke:

0, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 5, 5, 6, 7, 9

a) Tegn et prikkplott av dataene.

b) Hvor mange elever brukte mer enn 5 timer?

c) Hvor mange prosent brukte 2 timer eller mindre?

d) Hvor stor er forskjellen mellom den som brukte mest tid, og den som brukte minst?

1.6 Hvilken type variasjon er det?

a) To elever i klassen er ikke like høye.

b) Du veier deg om morgenen og om kvelden og får to forskjellige tall.

c) To elever tar tiden på det samme 60-meterløpet og får 8,2 og 8,4 sekunder.

d) To meningsmålinger samme uke gir et parti 24 % og 27 %.

1.7 Jonas vil vite hvor mange elever som bruker skolebiblioteket. Han legger ut en avstemning på bibliotekets Instagram-konto. 85 % av dem som svarer, bruker biblioteket hver uke.

a) Hvilket steg i undersøkelsesprosessen er det feil i?

b) Forklar hvorfor 85 % trolig er for høyt.

1.8 6 tilfeldige elever på skolen sov i gjennomsnitt 6,8 timer natt til onsdag. Emma skriver: «Elevene på skolen sover 6,8 timer.» Nevn to ting som er galt med påstanden.

1.9 En nettavis skriver: «Halvparten av elevene skulker minst én time i uka.» Skriv tre spørsmål du vil stille for å vurdere påstanden.

1.10 Hva blir utskriften?

Python

```
sovn = [6.0, 7.5, 5.5, 8.0, 6.5, 7.0]
antall = 0
for t in sovn:
    if t < 7:
        antall = antall + 1
print(antall)
print(antall / len(sovn) * 100)
```

1.11 Hva blir utskriften?

Python

```
lekser = [2, 0, 3, 1]
sum_timer = 0
for t in lekser:
    sum_timer = sum_timer + t
print(sum_timer)
print(sum_timer / len(lekser))
```

1.12 Endre programmet i oppgaven over slik at det teller hvor mange elever som ikke gjorde lekser i det hele tatt.

1.13 Tre grupper trakk hver 5 tilfeldige elever fra den samme klassen og regnet ut gjennomsnittlig søvn. De fikk 6,4, 7,3 og 7,0 timer.

a) Hvorfor fikk gruppene forskjellige svar?

b) Hvilket av svarene er riktig?

1.14 To meningsmålinger i samme uke ga et parti 22 % og 25 % oppslutning. En kommentator sier at partiet har gått fram. Drøft påstanden.

1.15 Et datasett har disse opplysningene om hver elev på skolen: trinn, alder, skolevei i minutter og om eleven har deltidsjobb. Skriv to statistiske spørsmål som datasettet kan svare på.

1.16 Elevrådet vil undersøke hvor lang tid elevene bruker på skoleveien. Beskriv hva de bør gjøre i hvert av de fire stegene i undersøkelsesprosessen.

1.1.2 Variabler og datasett

EKSEMPEL *Værdata*

En værstasjon registrerer hver dag høyeste temperatur, nedbør i millimeter og vindretning (nord, sør, øst eller vest). Datasettet dekker hele januar. Hva er en observasjon, og hvor mange observasjoner og variabler er det?

Løsning

En observasjon er én dag. Januar har 31 dager, så det er 31 observasjoner. Variablene er temperatur, nedbør og vindretning, altså 3 variabler (pluss datoen, som skiller radene fra hverandre).

■

EKSEMPEL *Hvilken type?*

Hvilken type er variabelen?

a) Antall mål i en fotballkamp

b) Høyden til en elev

c) Favorittsjanger i musikk

d) Postnummeret der eleven bor

Løsning

a) Numerisk diskret. Mål telles.

b) Numerisk kontinuerlig. Høyde måles og kan være 172,4 cm.

c) Kategorisk.

d) Kategorisk. Postnummeret er et tall, men det er et navn på et område. Gjennomsnittet av to postnumre er ikke et sted midt imellom.

■

EKSEMPEL *Transport til skolen*

I tabellen fra elevrådet har de 12 elevene disse svarene (i samme rekkefølge):

- Transport: buss, gå, sykkel, buss, buss, gå, buss, sykkel, gå, buss, buss, sykkel
- Søsken: $1, 2, 0, 1, 3, 1, 2, 1, 0, 2, 1, 4$

a) Hvor stor andel av elevene tar buss?

b) Hvor stor andel av elevene har minst to søsken?

Løsning

a) 6 av 12 elever tar buss: $\frac{6}{12} = 0{,}5 = 50\,\%$.

b) 5 elever har minst to søsken (2, 3, 2, 2 og 4): $\frac{5}{12} \approx 0{,}42 = 42\,\%$.

■

1.17 Hvilken type er variabelen? Svar kategorisk, numerisk diskret eller numerisk kontinuerlig.

a) Antall SMS-er sendt i går

b) Temperaturen ute klokka 12

c) Favorittfag

d) Hvor lang tid det tar å lade mobilen

e) Antall elever i en klasse

f) Klassen eleven går i (2STA, 2STB, ...)

1.18 Er variabelen kategorisk eller numerisk? Begrunn.

a) Telefonnummer

b) Skostørrelse

c) Romnummer på skolen

d) Antall timer fravær

1.19 Et datasett har én rad for hver av de 28 elevene i en klasse, og kolonnene navn, høyde, hårfarge og antall søsken.

a) Hvor mange observasjoner er det?

b) Hvilke kolonner er variabler vi kan regne gjennomsnitt av?

1.20 En kafé registrerer hvert salg: tidspunkt, hva som ble kjøpt (kaffe, te, kakao), pris i kroner og betalingsmåte (kort, Vipps, kontant). Hva er én observasjon, og hvilken type er hver variabel?

1.21 20 elever svarte på hvordan de kommer seg til skolen. 8 tar buss, 5 går, 4 sykler, og resten blir kjørt.

a) Hvor mange blir kjørt?

b) Hvor mange prosent tar buss?

c) Hvor mange prosent går eller sykler?

1.22

Elev	Trinn	Søsken	Skolevei (min)
1	Vg1	2	15
2	Vg2	0	30
3	Vg2	1	10
4	Vg3	1	25
5	Vg1	3	20
6	Vg2	1	20

a) Hvor mange prosent går på Vg2?

b) Hva er gjennomsnittlig skolevei?

c) Hva er gjennomsnittlig antall søsken?

d) Hvorfor gir det ikke mening å regne ut gjennomsnittet av trinn?

1.23 Lise sier: «Antall søsken er en kategorisk variabel, fordi ingen kan ha 1,5 søsken.» Hva er feil?

1.24 I en klasse ble draktnumrene til alle som spiller fotball registrert. Gjennomsnittet er 11,4. Hvorfor sier ikke dette tallet noe?

1.25 Gjør om spørsmålet slik at svaret blir en numerisk variabel: «Trener du mye?»

1.26 Gjør om spørsmålet slik at svaret blir en kategorisk variabel: «Hvor mange minutter leste du i går?»

1.27 Hva blir utskriften?

Python

```
fag = ["matte", "norsk", "matte",
       "kunst", "matte", "norsk"]
print(fag.count("matte"))
print(len(fag))
print(fag.count("norsk") / len(fag))
```

1.28 Hva blir utskriften?

Python

```
svar = ["ja", "nei", "ja", "ja"]
ja = svar.count("ja")
nei = svar.count("nei")
print(ja - nei)
print(ja / len(svar) * 100)
```

1.29 Skriv et program som regner ut andelen elever som går til skolen, når transportmidlene er lagret i lista `transport`.

1.30 Høyde kan registreres i centimeter (172) eller som «lav», «middels» eller «høy».

a) Hvilken type variabel får vi i hvert tilfelle?

b) Hvilken måte gir mest informasjon? Forklar.

1.31 Du vil undersøke om elever som har deltidsjobb, sover mindre enn andre. Foreslå tre variabler du vil registrere, og si hvilken type hver av dem er.

1.2.1 Datakilder

EKSEMPEL *Primær eller sekundær?*

Er det primærdata eller sekundærdata?

a) Du måler pulsen til klassekameratene før og etter en løpetur.

b) Du laster ned nedbøren i byen din de siste ti årene fra Meteorologisk institutt.

c) Du bruker tall fra en forskningsrapport om ungdoms søvn.

d) Du teller hvor mange biler som passerer skolen mellom 8 og 9.

Løsning

a) Primærdata. Du måler selv.

b) Sekundærdata. Meteorologisk institutt har målt.

c) Sekundærdata. Forskerne har samlet inn dataene.

d) Primærdata. Du teller selv.

■

EKSEMPEL *Velg datakilde*

Hvilken type data passer best til spørsmålet? Begrunn.

a) Hvor mange elever på skolen vår går til skolen?

b) Har andelen nordmenn som bruker Facebook, endret seg de siste årene?

Løsning

a) Primærdata. Ingen andre har trolig disse tallene, og elevene er lette å spørre.

b) Sekundærdata. Du trenger svar fra mange nordmenn over flere år, og det har SSB.

■

EKSEMPEL *TikTok og Snapchat*

Bruk tabellen over.

a) Hvor mange prosentpoeng økte andelen som brukte TikTok, fra 2023 til 2025?

b) Hvor mange prosent økte andelen?

c) Hvilken av appene i tabellen gikk mest ned fra 2023 til 2025, målt i prosentpoeng?

Løsning

a) $74 - 66 = 8$ prosentpoeng.

b) $\frac{8}{66} \cdot 100\,\% \approx 12\,\%$.

c) Endringene er -2 (Snapchat), -5 (Instagram), $+8$ (TikTok) og -10 (Facebook) prosentpoeng. Facebook gikk mest ned. YouTube kan vi ikke ta med, fordi det mangler tall for 2023.

■

EKSEMPEL *Nettavisens avstemning*

En nettavis har en avstemning: «Hvilken app bruker du mest?» 12 000 lesere svarer, og 45 % svarer Facebook. Overskriften er: «Facebook er fortsatt nordmenns favoritt.» Vurder kilden med de fem spørsmålene.

Løsning

- **Hvem:** Nettavisen, en kjent avsender, men avstemningen er ikke en vitenskapelig undersøkelse.
- **Hvorfor:** For å engasjere leserne, ikke for å måle hva nordmenn mener.
- **Hvordan:** Bare lesere av akkurat denne avisen, som selv velger å svare. Mange svar hjelper ikke når de som svarer, ikke ligner på befolkningen. Kanskje kan samme person stemme flere ganger.
- **Når:** Ett tidspunkt, ukjent hvem som var inne da.
- **Hva:** «Bruker mest» er ikke det samme som «bruker», og avstemningen sier ingenting om alder.

Overskriften holder ikke. Avstemningen sier bare noe om dem som valgte å svare.

■

1.32 Er det primærdata eller sekundærdata?

a) Du måler hvor lang tid klassekameratene bruker på å løse en sudoku.

b) Du henter folketallet i kommunen din fra SSB.

c) Du bruker temperaturmålinger fra Meteorologisk institutt.

d) Du spør 30 elever hvor mange timer de trener i uka.

e) Du bruker tall fra en rapport fra Folkehelseinstituttet.

1.33 Nevn én fordel og én ulempe med

a) primærdata,

b) sekundærdata.

1.34 Hvilken type data passer best til spørsmålet? Begrunn kort.

a) Hvor mange elever i klassen har deltidsjobb?

b) Hvordan har folketallet i Norge endret seg siden 1950?

c) Hvor høy puls har elevene etter å ha gått opp trappene på skolen?

1.35 Tabellen viser andelen 16–24-åringer som brukte appen en gjennomsnittsdag (SSB, tabell 14511):

App	2023	2025
Snapchat	91	89
Instagram	83	78
TikTok	66	74
Facebook	60	50

a) Hvor mange prosentpoeng endret andelen seg for hver app?

b) Hvilken app økte?

c) Hvor mange prosent gikk andelen som brukte Instagram, ned?

1.36 Andelen som bruker en app, går fra 40 % til 50 %.

a) Hvor mange prosentpoeng har andelen økt?

b) Hvor mange prosent har andelen økt?

1.37 Andelen som røyker daglig, går fra 12 % til 9 %.

a) Hvor mange prosentpoeng har andelen gått ned?

b) Hvor mange prosent har andelen gått ned?

1.38 Renta økte fra 4 % til 5 %. Ola sier at renta økte med 1 %. Hva er feil, og hva er riktig?

1.39 I en SSB-tabell står det «.» i stedet for et tall for ett av årene. Emma regner med at verdien er 0. Hva er galt?

1.40 Hvilke fem spørsmål bør du stille til en datakilde?

1.41 En nettavis lar leserne stemme: «Bør mobilen forbys i skolen?» 15 000 svarer, og 70 % svarer ja. Overskriften er: «Klart flertall vil forby mobilen.» Vurder påstanden.

1.42 En sjokoladeprodusent skriver: «Undersøkelse viser at sjokolade gjør deg gladere.» Hvilke spørsmål vil du stille til undersøkelsen?

1.43 I klassen brukte 20 av 25 elever TikTok i går. SSB oppgir 74 % for 16–24-åringer en gjennomsnittsdag.

a) Hvor mange prosent av klassen brukte TikTok?

b) Gi to grunner til at tallene kan være forskjellige.

1.44 Hva blir utskriften?

Python

```
start = [40, 12, 4]
slutt = [50, 9, 5]
for i in range(3):
    print(slutt[i] - start[i])
```

1.45 Hva blir utskriften?

Python

```
gammel = 60
ny = 45
poeng = ny - gammel
prosent = poeng / gammel * 100
print(poeng)
print(prosent)
```

1.46 Du vil finne ut om elever på skolen spiser frokost. Beskriv hvordan du kan bruke primærdata, og hvordan du kan bruke sekundærdata. Hva er fordelen med hver løsning for akkurat dette spørsmålet?

1.47 Et lite parti går fra 3 % til 6 % oppslutning, og et stort parti går fra 25 % til 28 %.

a) Hvor mange prosentpoeng og hvor mange prosent økte hvert parti?

b) Hvorfor kan de to partiene velge ulike måter å omtale økningen på?

1.2.2 Data i Python

EKSEMPEL *Transport, ryddet*

Etter at skrivemåten er rettet (se under), har de 20 elevene disse svarene:

> buss, gå, sykkel, buss, buss, gå, buss, sykkel, gå, buss, buss, sykkel, bil, buss, gå, buss, sykkel, buss, gå, buss

Lag en frekvenstabell med frekvens og relativ frekvens.

Løsning

Transport	Frekvens	Relativ frekvens
buss	10	0,50
gå	5	0,25

Transport	Frekvens	Relativ frekvens
sykkel	4	0,20
bil	1	0,05
Sum	20	1,00

Kontroll: Frekvensene summerer til $n = 20$, og de relative frekvensene til 1.

■

EKSEMPEL *Skoleveien sortert*

19 av elevene oppga skolevei i minutter:

25, 10, 15, 40, 20, 8, 35, 12, 5, 30, 45, 19, 10, 250, 28, 14, 33, 7, 22

Sorter verdiene. Hva er minste og største verdi? Er det noe som ser rart ut?

Løsning

5, 7, 8, 10, 10, 12, 14, 15, 19, 20, 22, 25, 28, 30, 33, 35, 40, 45, 250

Minste verdi er 5 minutter, største er 250 minutter. 250 minutter er over fire timer hver vei, og langt unna alle de andre. Det er trolig en skrivefeil for 25.

■

EKSEMPEL *Rydd transporten*

Før rydding har variabelen transport disse ulike verdiene og frekvensene:

Verdi	Frekvens
buss	8
gå	5
sykkel	3
Buss	1
BUSS	1
sykel	1
bil	1

Hvor mange tar buss, og hvor mange sykler, etter rydding?

Løsning

«buss», «Buss» og «BUSS» er samme kategori: $8 + 1 + 1 = 10$ tar buss. «sykel» er en skrivefeil for sykkel: $3 + 1 = 4$ sykler. Det stemmer med frekvenstabellen over.

■

1.48 20 elever svarte på hvilken årstid de liker best:

sommer, vinter, sommer, høst, vår,
sommer, sommer, vinter, vår, sommer,
høst, sommer, vinter, sommer, vår,
sommer, høst, vinter, sommer, vår

Lag en frekvenstabell med frekvens og relativ frekvens.

1.49 I en frekvenstabell med 25 observasjoner har en kategori frekvensen 7.

a) Hva er den relative frekvensen?

b) Hvor mange prosent er det?

1.50 Den relative frekvensen til en kategori er 0,35, og det er 40 observasjoner. Hva er frekvensen?

1.51 En frekvenstabell har kategoriene A, B og C med relative frekvenser 0,42 og 0,31 for A og B. Hva er den relative frekvensen til C?

1.52 Sorter tallene fra minst til størst, og finn minste og største verdi:

42, 17, 8, 33, 25, 8, 51, 19, 30, 12

1.53 Er verdien trolig en feil? Begrunn.

a) Høyde til en elev: 1,78 cm

b) Antall timer trening forrige uke: 18

c) Karakter i matte: 9

d) Pris for en buss-billett: 45 kr

1.54 Rydd svarene, og lag en frekvenstabell:

Ja, nei, JA, ja, Nei, ja , jaa, nei, Ja, ja

1.55 Lise har laget en frekvenstabell for 30 elever. Frekvensene er 12, 9 og 7. Hva tyder det på?

1.56 8 elever oppga skolevei i minutter: 12, 20, 8, 15, 150, 10, 25, 18.

a) Hvilken verdi er trolig en feil?

b) Regn ut gjennomsnittet med og uten denne verdien.

1.57 I et datasett med 50 elever mangler 4 svar på søvn. Hvor mange observasjoner bruker du når du regner ut gjennomsnittlig søvn?

1.58 Hvorfor er det viktig å skrive ned hva du har ryddet i et datasett?

1.59 Hva blir utskriften?

Python

```
tall = [14, 3, 22, 8]
print(sorted(tall))
print(min(tall), max(tall))
```

1.60 Hva blir utskriften?

Python

```
fag = ["gym", "matte", "gym",
       "norsk", "gym", "matte"]
n = len(fag)
for k in ["gym", "matte", "norsk"]:
    print(k, fag.count(k) / n)
```

1.61 Hva blir utskriften?

Python

```
svar = ["Ja", "ja", "JA", "nei"]
ryddet = []
for s in svar:
    ryddet.append(s.lower())
print(ryddet)
print(ryddet.count("ja"))
```

1.62 Skriv et program som teller hvor mange elementer i lista `svar` som er `"buss"`, uansett om det er skrevet med store eller små bokstaver.

1.63 Et spørreskjema spør: «Hvor mye trener du?» Svarene er blant annet «mye», «3 ganger i uka», «ca 5 timer» og «litt». Forklar hvorfor svarene er vanskelige å analysere, og foreslå et bedre spørsmål.

1.3.1 Populasjon og utvalg

EKSEMPEL *Finn populasjonen og utvalget*

Hva er populasjonen, og hva er utvalget?

a) En meningsmåling spør 1000 tilfeldig valgte stemmeberettigede om hva de vil stemme ved stortingsvalget.

b) Elevrådet spør 60 av de 600 elevene på skolen om de spiser frokost.

c) En fabrikk tester 50 av de 10 000 batteriene som ble laget i dag.

Løsning

a) Populasjon: alle stemmeberettigede i Norge. Utvalg: de 1000 som ble spurt.

b) Populasjon: de 600 elevene på skolen. Utvalg: de 60 som ble spurt.

c) Populasjon: de 10 000 batteriene fra i dag. Utvalg: de 50 som ble testet. Her er en totalundersøkelse umulig hvis testen ødelegger batteriet.

■

EKSEMPEL *Andelen i utvalget*

23 av de 60 elevene i utvalget sier at de spiser frokost hver dag. Hva er $\hat{p}$? Hva kan vi si om p?

Løsning

$\hat{p} = \frac{23}{60} \approx 0{,}38$, altså omtrent 38 %. Andelen p blant alle de 600 elevene kjenner vi ikke, men 38 % er vårt beste anslag. Et annet utvalg ville gitt et litt annet svar, så p er neppe akkurat 38 %.

■

1.64 Hva er populasjonen, og hva er utvalget?

a) En forsker måler blodtrykket til 200 tilfeldige nordmenn over 60 år.

b) En lærer spør 5 elever i en klasse på 25 om de har gjort leksene.

c) En butikk sjekker 30 av 2000 melkekartonger for å se om de er utgått på dato.

1.65 Er det en totalundersøkelse eller en utvalgsundersøkelse?

a) Folkeregisteret registrerer alle fødsler i Norge.

b) En meningsmåling spør 1000 velgere.

c) Alle elevene i en klasse leverer karakteren sin, og læreren regner ut snittet i klassen.

1.66 Nevn to grunner til at man ofte velger en utvalgsundersøkelse i stedet for en totalundersøkelse.

1.67 I en måling blant 1200 velgere svarer 276 at de vil stemme på parti B.

a) Hva er $\hat{p}$?

b) Hvorfor er ikke $\hat{p}$ nødvendigvis lik den andelen som faktisk stemmer på parti B ved valget?

1.68 Er tallet en parameter eller et tall fra et utvalg?

a) 52 % av de 500 spurte er for sommertid hele året.

b) Helsesykepleieren har målt alle de 600 elevene på skolen, og gjennomsnittshøyden er 172 cm.

c) 3 av 20 testede lyspærer var defekte.

1.69 Hva betyr det at et utvalg er enkelt tilfeldig?

1.70 Forklar med egne ord hvorfor en kokk bare trenger å smake på én skje suppe, og hva det har med meningsmålinger å gjøre.

1.71 To meningsmålinger spør tilfeldige velgere om samme spørsmål samme uke. Den ene har $n = 100$, den andre $n = 2000$. Hvilken vil du stole mest på? Begrunn.

1.72 Per sier: «Et utvalg på 1000 personer er mye bedre i Norge enn i USA, fordi det er en større del av befolkningen.» Stemmer det?

1.73 En skole har 500 elever, og 30 % av dem sykler. Ti grupper trekker hvert sitt tilfeldige utvalg på 10 elever.

a) Hvor mange syklister forventer du omtrent i hvert utvalg?

b) Vil alle gruppene få akkurat det tallet? Forklar.

1.74 Hva blir utskriften?

Python

```
svar = [0, 1, 1, 0, 1, 0, 0, 0]
print(sum(svar))
print(sum(svar) / len(svar))
```

1.75 Hva blir utskriften?

Python

```
utvalg1 = [1, 0, 1, 1, 0]
utvalg2 = [0, 0, 1, 0, 0]
for u in [utvalg1, utvalg2]:
    print(sum(u) / len(u))
```

1.76 Hva gjør `random.sample(elever, 10)` i Python, og hvorfor passer den til å trekke et enkelt tilfeldig utvalg?

1.77 En klasse vil finne ut hvor mange elever på skolen som har deltidsjobb. De spør alle i egen klasse. Kan de bruke resultatet for hele skolen? Begrunn.

1.78 Et utvalg på 50 elever gir $\hat{p} = 0{,}46$ for andelen som går til skolen. Lag en setning som forteller resultatet på en ærlig måte.

1.3.2 Utvalgsmetoder

EKSEMPEL *Stratifisert utvalg på skolen*

Elevrådet vil ha et stratifisert utvalg på 60 elever, med like stor andel fra hvert trinn. Hvor mange skal trekkes fra hvert trinn?

Løsning

Utvalget er $\frac{60}{600} = 0{,}1$, altså 10 % av elevene. Fra hvert trinn trekkes 10 %:

- Vg1: $0{,}1 \cdot 250 = 25$
- Vg2: $0{,}1 \cdot 200 = 20$
- Vg3: $0{,}1 \cdot 150 = 15$

Kontroll: $25 + 20 + 15 = 60$. Innenfor hvert trinn trekkes elevene tilfeldig. ■

EKSEMPEL *Hvilken metode?*

Hvilken utvalgsmetode er brukt?

a) Elevrådet trekker tilfeldig tre klasser og spør alle elevene i dem.

b) Læreren spør de ti første elevene som kommer inn i klasserommet.

c) En avis ber leserne svare på en avstemning på nettsidene sine.

d) Et byrå deler velgerne inn etter fylke og trekker tilfeldig fra hvert fylke, i forhold til folketallet.

Løsning

a) Klyngeutvalg.

b) Bekvemmelighetsutvalg. De som kommer tidlig, er kanskje mer opptatt av skolen enn andre.

c) Selvseleksjon.

d) Stratifisert utvalg, med fylkene som lag. ■

1.79 Hvilken utvalgsmetode er brukt?

a) Alle som handler i en butikk mellom 12 og 13, blir spurt.

b) Hvert 20. navn i telefonkatalogen, med tilfeldig start.

c) En skole trekker tilfeldig fire klasser og spør alle elevene der.

d) En avis lar leserne stemme på nett.

e) Et byrå trekker tilfeldig fra hver aldersgruppe, i forhold til hvor mange som er i gruppen.

1.80 Fra en liste på 900 navn skal det trekkes et systematisk utvalg på 45.

a) Hva er k?

b) Startpunktet er 7. Hvilke tre nummer kommer med først?

1.81 En skole har 300 elever på Vg1, 240 på Vg2 og 180 på Vg3. Det skal trekkes et stratifisert utvalg på 60 elever med like stor andel fra hvert trinn. Hvor mange skal trekkes fra hvert trinn?

1.82 En kommune har 12 000 innbyggere i sentrum og 4 000 i distriktet. Et stratifisert utvalg skal ha 400 personer. Hvor mange skal trekkes fra hvert område?

1.83 Forklar forskjellen på et stratifisert utvalg og et klyngeutvalg.

1.84 Hvorfor gir et bekvemmelighetsutvalg ofte et skjevt resultat? Gi et eksempel.

1.85 Et TV-program ber seerne sende SMS for å si om de er for eller mot en ny lov. 85 % er mot. Hva kalles denne utvalgsmetoden, og hvorfor bør vi ikke stole på tallet?

1.86 Hvilken metode passer best? Begrunn kort.

a) Undersøke elevers syn på skolemat ved 30 skoler i et fylke, med lite tid til reising.

b) Undersøke hvor mye folk i ulike aldersgrupper bruker strømmetjenester.

1.87 Hvorfor må startpunktet i et systematisk utvalg velges tilfeldig?

1.88 I en klasseliste står elevene sortert etter karakter, best først. Du trekker hver 5. elev fra starten. Hva er problemet?

1.89 Et stratifisert utvalg har 20 elever fra hvert av de tre trinnene. Andelen som er for et forslag, er 0,5 på Vg1 (300 elever), 0,6 på Vg2 (200 elever) og 0,8 på Vg3 (100 elever). Hvor stor andel av alle elevene er trolig for?

1.90 Hva blir utskriften?

Python

```
N = 40
k = 8
for nr in range(2, N + 1, k):
    print(nr)
```

1.91 Hva blir utskriften?

Python

```
lag = [250, 200, 150]
n = 60
N = sum(lag)
for antall in lag:
    print(antall * n / N)
```

1.92 Beskriv hvordan du vil trekke et utvalg på 40 elever for å undersøke hvor mye elevene på skolen trener. Hvilken metode velger du, og hvorfor?

1.93 Hvorfor blir ikke en nettavstemning med 50 000 svar mer pålitelig enn en tilfeldig meningsmåling med 1000 svar?

1.3.3 Representativitet og feilkilder

EKSEMPEL *Hvor mye kan frafallet bety?*

Elevrådet trekker 200 elever tilfeldig og spør om de ønsker lengre lunsjpause. 120 svarer, og 84 av dem svarer ja.

a) Hva er svarprosenten?

b) Hvor stor andel av dem som svarte, sa ja?

c) Hva ville andelen ja vært blant alle de 200 hvis alle som ikke svarte, egentlig mente nei? Og hvis alle mente ja?

Løsning

a) $\frac{120}{200} \cdot 100\,\% = 60\,\%$

b) $\frac{84}{120} = 0{,}70 = 70\,\%$

c) Alle nei: $\frac{84}{200} = 42\,\%$. Alle ja: $\frac{84 + 80}{200} = 82\,\%$.

Den sanne andelen blant de 200 ligger altså et sted mellom 42 % og 82 %. Bare hvis de som ikke svarte, mener omtrent det samme som de som svarte, er 70 % et godt anslag. ■

EKSEMPEL *Skriv om spørsmålet*

Spørsmålet under er ledende. Skriv det om slik at det blir nøytralt.

> «Mange eksperter mener at mobilbruk skader konsentrasjonen. Synes du mobilen bør forbys i timene?»

Løsning

Innledningen om eksperter dytter mot ja. Et nøytralt spørsmål:

> «Hva mener du om mobilbruk i timene? Bør den være tillatt, tillatt med begrensninger, eller forbudt?»

Svaralternativene dekker hele spennet, og ingen av dem er framhevet. ■

1.94 Hva betyr det at et utvalg er representativt?

1.95 Hvilken feilkilde er det? Velg mellom dekning, frafall, ledende spørsmål, sosial ønskelighet og målefeil.

a) En undersøkelse om pendling ringer bare til fasttelefoner.

b) Spørsmålet er: «Synes du ikke også at bussbillettene er for dyre?»

c) Folk blir spurt ansikt til ansikt om de har kjørt for fort den siste måneden.

d) Bare 15 % av de utvalgte svarer på en undersøkelse om fritidsaktiviteter.

e) Et termometer i solveggen brukes til å måle lufttemperaturen.

1.96 En undersøkelse sendes til 1200 personer, og 420 svarer. Hva er svarprosenten?

1.97 500 elever trekkes ut. 350 svarer, og 210 av dem er fornøyd med skolemiljøet.

a) Hva er svarprosenten?

b) Hvor mange prosent av dem som svarte, er fornøyd?

c) Hva er laveste og høyeste mulige andel fornøyde blant alle de 500?

1.98 Forklar forskjellen på tilfeldig og systematisk feil. Hvilken av dem blir mindre når utvalget blir større?

1.99 Er feilen tilfeldig eller systematisk?

a) En vekt viser alltid 200 g for lite.

b) To tilfeldige utvalg av samme størrelse gir litt ulike svar.

c) En spørreundersøkelse om trening har bare med folk fra et treningssenter.

1.100 Skriv om spørsmålet slik at det blir nøytralt: «Hvor mye irriterer det deg at kantina er så dyr?»

1.101 Hvorfor svarer færre at de har jukset på en prøve når læreren spør muntlig, enn når de svarer anonymt? Hva kalles denne feilkilden?

1.102 En nettavis har en avstemning med 40 000 svar. En tilfeldig meningsmåling har 1000 svar. Hvilken vil du stole mest på, og hvorfor?

1.103 I 1936 fikk Literary Digest 2,4 millioner svar, men spådde feil vinner av presidentvalget. Nevn to grunner.

1.104 En undersøkelse om søvn sendes ut klokka 07 : 30 en mandag. Hvem tror du svarer minst, og hvordan kan det påvirke resultatet?

1.105 Hva blir utskriften?

Python

```
spurt = 250
svarte = 150
for ja in [90, 120]:
    print(ja / svarte * 100)
```

1.106 Hva blir utskriften?

Python

```
spurt = 200
svarte = 160
ja = 100
lav = ja / spurt
hoy = (ja + spurt - svarte) / spurt
print(lav, hoy)
```

1.107 En skole vil vite hvor mange elever som opplever mobbing, og legger ut et skjema som elevene kan svare på hvis de vil. Nevn to feilkilder, og foreslå en bedre metode.

1.108 En undersøkelse har høy svarprosent, 95 %. Betyr det at resultatet er riktig? Begrunn.

1.3.4 Observasjonsstudier og eksperimenter

EKSEMPEL *Observasjon eller eksperiment?*

Er det en observasjonsstudie eller et eksperiment?

a) Forskere sammenlikner helsen til personer som selv har valgt å bli vegetarianere, med andre.

b) En lærer lar halvparten av klassen, trukket tilfeldig, øve med en app i tre uker, og sammenlikner prøveresultatene.

c) SSB undersøker sammenhengen mellom utdanning og inntekt.

Løsning

a) Observasjonsstudie: deltakerne har valgt selv.

b) Eksperiment: læreren bestemmer hvem som får appen.

c) Observasjonsstudie: SSB registrerer det som allerede er.

■

EKSEMPEL *Finn den bakenforliggende variabelen*

Foreslå en bakenforliggende variabel som kan forklare sammenhengen.

a) Elever som har egen PC, får bedre karakterer.

b) Byer med mange brannstasjoner har mange branner.

c) Personer som bruker solkrem, får oftere hudkreft.

Løsning

a) Foreldrenes inntekt og utdanning kan påvirke både om eleven har egen PC, og hvor mye hjelp og ro eleven får til skolearbeid.

b) Størrelsen på byen: store byer har både flere brannstasjoner og flere branner.

c) Hvor mye sol personen får: de som er mye i sola, bruker mer solkrem og har større risiko for hudkreft.

■

1.109 Hva er forskjellen på en observasjonsstudie og et eksperiment?

1.110 Er det en observasjonsstudie eller et eksperiment?

a) Forskere sammenlikner vekten til personer som går til jobben, med personer som kjører.

b) Halvparten av klassen, trukket tilfeldig, får lekser på nett, resten på papir.

c) En lege sammenlikner pasienter som selv har valgt to ulike behandlinger.

d) Forsøkspersoner får tilfeldig koffein eller placebo før en reaksjonstest.

1.111 Hva er en bakenforliggende variabel? Gi et eksempel.

1.112 Foreslå en bakenforliggende variabel.

a) Barn med store føtter leser bedre.

b) Kommuner med mange leger har mange dødsfall.

c) Elever som har mange bøker hjemme, får bedre karakterer.

1.113 Nevn fire kjennetegn på et godt eksperiment.

1.114 Hvorfor får kontrollgruppen ofte placebo i medisinske forsøk?

1.115 Hvorfor bør deltakerne fordeles tilfeldig på gruppene i et eksperiment?

1.116 En observasjonsstudie viser at elever som har deltidsjobb, får dårligere karakterer. Kan vi konkludere med at deltidsjobb gir dårligere karakterer? Begrunn.

1.117 En trener lar alle spillerne på laget gjøre en ny oppvarming i en måned, og skadene går ned. Hva mangler for at vi kan si at det er oppvarmingen som virker?

1.118 30 elever skal fordeles tilfeldig på tre like store grupper. Hvor mange blir det i hver gruppe, og hvordan kan du gjøre fordelingen tilfeldig uten PC?

1.119 Hva blir utskriften?

Python

```
elever = ["A", "B", "C", "D", "E", "F"]
print(elever[:2])
print(elever[2:])
```

1.120 Hva blir utskriften?

Python

```
deltakere = list(range(1, 21))
forsok = deltakere[:10]
kontroll = deltakere[10:]
print(forsok[0], kontroll[0])
print(len(forsok) + len(kontroll))
```

1.121 Hva er forskjellen på et tilfeldig utvalg og en tilfeldig fordeling på grupper? Hva gjør hver av dem mulig å si?

1.122 Planlegg et eksperiment som undersøker om elever husker flere ord når de leser dem høyt, enn når de leser dem stille. Beskriv grupper, fordeling og hvordan du måler resultatet.

1.123 Hvorfor kan vi ikke gjøre et eksperiment for å finne ut om røyking gir lungekreft? Hvordan ble det likevel vist?

1.4.1 Undersøkelse: Klassens spørreundersøkelse

EKSEMPEL *Leggetid på Vg2*

En gruppe i 2STB lurte på hvor sent elevene legger seg.

1. **Spørsmål:** «Hvor stor andel av elevene på Vg2 legger seg etter midnatt på skoledager?»
2. **Data:** De laget et skjema i Forms med spørsmålet «Når la du deg i går kveld?» og fire svaralternativer. Lenken ble lagt ut i klassens Teams-kanal og delt på Snapchat. De fikk 48 svar.
3. **Analyse:**

Leggetid	Frekvens
Før kl. 22	3
22–23	12
23–24	19
Etter 24	14
Sum	48

4. **Tolkning:** «29 % av elevene på Vg2 legger seg etter midnatt.»

a) Kontroller at 29 % stemmer.

b) Hvilke feilkilder har undersøkelsen?

c) Hvordan burde konklusjonen vært formulert?

Løsning

a) $\frac{14}{48} \approx 0{,}29 = 29\,\%$. Det stemmer.

b) Feilkilder:

- **Utvalg:** Lenken gikk til egen klasse og til venner på Snapchat (bekvemmelighet og selvseleksjon). Andre Vg2-klasser er kanskje ikke med, og vennene kan gå på andre trinn.
- **Frafall:** Vi vet ikke hvor mange som fikk lenken og ikke svarte.
- **Spørsmålet:** «I går kveld» er én kveld. Var det en fredag, sier det lite om skoledager.
- **Svarene:** Noen husker kanskje feil eller runder av.

c) For eksempel: «Blant de 48 som svarte, la 14 (29 %) seg etter midnatt kvelden før. Utvalget er ikke tilfeldig og dekker trolig mest vår egen klasse og vennene våre, så vi vet ikke om det gjelder alle på Vg2.»

■

1.124 Skriv de fire stegene i den statistiske undersøkelsesprosessen, og si kort hva dere gjør i hvert steg.

1.125 Hva er galt med spørsmålene? Skriv dem om.

a) «Hvor mye sover du?»

b) «Er du enig i at skolen starter for tidlig?»

c) «Navn og klasse:»

1.126 En gruppe vil undersøke hvor mange elever på skolen som har deltidsjobb. Foreslå et statistisk spørsmål, to variabler (med type) og en utvalgsmetode.

1.127 En gruppe fikk disse svarene på hvordan elevene kommer seg til skolen:

Transport	Frekvens
buss	18
gå	9
sykkel	6
bil	3

a) Hvor mange svarte?

b) Lag en kolonne med relative frekvenser.

c) Hvor mange prosent går eller sykler?

1.128 En gruppe la ut et skjema i klassens kanal og fikk 22 svar av 28 elever. Hva er svarprosenten, og hvem gjelder resultatet for?

1.129 I en undersøkelse om skolevei har én elev svart 600 minutter. Hva gjør dere, og hvordan skriver dere det i rapporten?

1.130 Skriv en ærlig konklusjon: Av 45 elever som svarte på et skjema delt på Snapchat, sa 27 at de leser bok minst én gang i uka.

1.131 Nevn tre feilkilder som ofte finnes i spørreundersøkelser laget av elever, og si hvilken vei hver av dem kan trekke resultatet.

1.132 Hva blir utskriften?

Python

```
svar = [12, 25, 8, 300, 15, 20]
ryddet = []
for t in svar:
    if t <= 180:
        ryddet.append(t)
print(len(ryddet))
print(sum(ryddet) / len(ryddet))
```

1.133 Beskriv hva rapporten fra en spørreundersøkelse bør inneholde.

Fasit

1.1.1 Fra spørsmål til svar

1.1

a) Nei, ett svar om én person.
b) Ja, antallet varierer fra elev til elev.
c) Nei, åpningstiden er fastsatt.
d) Ja, tiden varierer fra elev til elev og fra besøk til besøk.
e) Nei, én måling på ett sted på én dag.

1.2

Eksempler:

a) Hvor mange skritt går elevene i klassen på en vanlig skoledag?
b) Hvor mye bruker elevene på skolen på lunsj i kantina per uke?
c) Hvor gamle var foreldrene til elevene i klassen da de tok førerkort?

1.3

a) Analysere data
b) Formulere spørsmål
c) Samle inn data
d) Tolke resultatene

1.4

C, D, A, B

1.5

a) Én prikk per elev over tallinja, stablet: én ved 0, tre ved 1, fem ved 2, fire ved 3, to ved 4, to ved 5 og én ved hver av 6, 7 og 9.
b) 3 elever
c) $\frac{9}{20} \cdot 100\,\% = 45\,\%$
d) $9 - 0 = 9$ timer

1.6

a) Mellom individer
b) Fra gang til gang
c) Ved måling
d) Fra utvalg til utvalg

1.7

a) Samle inn data.
b) De som følger bibliotekets konto, er trolig elever som bruker biblioteket. De ligner ikke på elevene flest, så andelen blir for høy.

1.8

Gjennomsnittet sier ikke at alle sover 6,8 timer; søvnen varierer fra elev til elev. Dessuten er det bare 6 elever og én natt, så et nytt utvalg kunne gitt et annet gjennomsnitt.

1.9

Eksempler: Hvem ble spurt, og hvor mange? Hvordan er «skulke» definert? Svarte elevene selv, eller er tallene hentet fra fraværsføringen? Gjelder det én skole eller hele landet?

1.10

```
3
50.0
```

1.11

```
6
1.5
```

1.12

```
lekser = [2, 0, 3, 1]
antall = 0
for t in lekser:
    if t == 0:
        antall = antall + 1
print(antall)
```

Utskriften blir `1`.

1.13

a) De trakk forskjellige elever, og søvnen varierer fra elev til elev (variasjon fra utvalg til utvalg).
b) Ingen av dem er «det riktige». Alle er anslag for gjennomsnittet i klassen. Det riktige gjennomsnittet finner vi bare ved å ta med alle elevene.

1.14

Målingene er gjort samme uke med ulike utvalg. Forskjellen kan skyldes variasjon fra utvalg til utvalg, ikke at flere støtter partiet. Vi trenger flere målinger over tid før vi kan si at partiet har gått fram.

1.15

Eksempler: Hvor mange minutter bruker elevene på skoleveien? Har elever på Vg3 deltidsjobb oftere enn elever på Vg1?

1.16

1. Spørsmål: «Hvor mange minutter bruker elevene på skolen på skoleveien en vanlig morgen?»
2. Data: Spør tilfeldig valgte elever fra alle trinn, gjerne flere dager.
3. Analyse: Lag et prikkplott, finn gjennomsnittet og hvor mye tiden varierer.
4. Tolkning: Svar på spørsmålet, og si hva som gjør svaret usikkert (hvem som svarte, anslåtte tider, uvanlige dager).

1.1.2 Variabler og datasett

1.17

a) Numerisk diskret
b) Numerisk kontinuerlig
c) Kategorisk
d) Numerisk kontinuerlig
e) Numerisk diskret
f) Kategorisk

1.18

a) Kategorisk. Det er et navn, vi regner ikke med det.
b) Numerisk. Størrelsen sier noe om lengden på foten, og vi kan regne gjennomsnitt.
c) Kategorisk. Rom 214 er ikke «dobbelt så mye» som rom 107.
d) Numerisk.

1.19

a) 28
b) Høyde og antall søsken.

1.20

Én observasjon er ett salg. Tidspunkt: numerisk kontinuerlig. Vare: kategorisk. Pris: numerisk. Betalingsmåte: kategorisk.

1.21

a) $20 - 8 - 5 - 4 = 3$
b) $\frac{8}{20} \cdot 100\,\% = 40\,\%$
c) $\frac{5+4}{20} \cdot 100\,\% = 45\,\%$

1.22

a) $\frac{3}{6} \cdot 100\,\% = 50\,\%$
b) $\frac{120}{6} = 20$ minutter
c) $\frac{8}{6} \approx 1{,}33$
d) Trinn er en kategorisk variabel. Vi kan telle hvor mange som går på hvert trinn.

1.23

Antall søsken er numerisk diskret: Verdiene telles, og vi kan regne med dem. Gjennomsnittet kan godt være 1,5 selv om ingen enkeltperson har 1,5 søsken.

1.24

Draktnummeret er et navn på en spiller, ikke et antall eller en måling. Gjennomsnittet av navn betyr ingenting.

1.25

For eksempel: «Hvor mange timer trente du forrige uke?»

1.26
For eksempel: «Hvor mye leste du i går: ingenting, under en halvtime eller mer enn en halvtime?»

1.27
```
3
6
0.3333333333333333
```

1.28
```
2
75.0
```

1.29
```
antall = transport.count("gå")
print(antall / len(transport))
```

1.30
a) Centimeter: numerisk kontinuerlig. Lav/middels/høy: kategorisk med rekkefølge.
b) Centimeter. Fra høyden i centimeter kan vi dele inn i lav, middels og høy, men ikke omvendt.

1.31
For eksempel: deltidsjobb (kategorisk: ja/nei), timer jobb per uke (numerisk kontinuerlig) og søvn natt til i dag i timer (numerisk kontinuerlig).

1.2.1 Datakilder

1.32
a) Primær b) Sekundær c) Sekundær d) Primær e) Sekundær

1.33
a) Fordel: passer nøyaktig til spørsmålet ditt. Ulempe: tar tid, og du får ofte få svar.
b) Fordel: ofte store, nøye innsamlede datasett som er gratis. Ulempe: samlet inn til et annet formål, så definisjonene passer ikke alltid.

1.34
a) Primærdata: lett å spørre klassen, og ingen andre har tallene.
b) Sekundærdata: SSB har folketallet for mange år.
c) Primærdata: må måles.

1.35
a) Snapchat -2, Instagram -5, TikTok $+8$, Facebook -10 prosentpoeng.
b) TikTok
c) $\frac{5}{83} \cdot 100\,\% \approx 6{,}0\,\%$

1.36
a) 10 prosentpoeng
b) $\frac{10}{40} \cdot 100\,\% = 25\,\%$

1.37
a) 3 prosentpoeng
b) $\frac{3}{12} \cdot 100\,\% = 25\,\%$

1.38
Renta økte med 1 prosentpoeng. I prosent økte den med $\frac{1}{4} \cdot 100\,\% = 25\,\%$.

1.39
Punktumet betyr at tallet mangler (for eksempel fordi kategorien ikke var med det året), ikke at verdien er 0. Året må holdes utenfor sammenlikningen.

1.40
Hvem har samlet inn dataene? Hvorfor? Hvordan? Når? Hva måler dataene egentlig?

1.41
De som svarer, er lesere av avisen som selv velger å stemme, ikke et utvalg som ligner på befolkningen. Mange svar hjelper ikke når utvalget er skjevt. Kanskje kan man stemme flere ganger, og vi vet ikke hvem som svarte. Påstanden gjelder bare dem som stemte.

1.42
For eksempel: Hvem har gjort undersøkelsen, og har de interesse av resultatet? Hvem ble spurt, og hvor mange? Hvordan ble «gladere» målt? Er det sjokoladen som gjør folk gladere, eller noe annet? Når ble den gjort?

1.43
a) $\frac{20}{25} \cdot 100\,\% = 80\,\%$
b) For eksempel: Klassen er få elever på én skole og har bare de yngste i aldersgruppa; «i går» er ikke det samme som en gjennomsnittsdag; tallene er fra ulike år; i en liten gruppe varierer andelen mye fra klasse til klasse.

1.44
```
10
-3
1
```

1.45
```
-15
-25.0
```

1.46
Primærdata: et anonymt spørreskjema til tilfeldig valgte elever på skolen, gjerne flere dager. Det svarer nøyaktig på spørsmålet om vår skole.
Sekundærdata: for eksempel Ungdata eller Folkehelseinstituttet, som har spurt mange ungdommer. De er store og godt innsamlet, men sier noe om ungdom generelt, ikke om akkurat vår skole.

1.47
a) Det lille: 3 prosentpoeng og 100 %. Det store: 3 prosentpoeng og 12 %.
b) Hvert parti velger målet som høres best ut: Det lille sier «doblet» eller «100 %», det store kan si «3 prosentpoeng». Les derfor alltid de faktiske andelene.

1.2.2 Data i Python

1.48

Årstid	Frekvens	Relativ frekvens
sommer	9	0,45
vinter	4	0,20
vår	4	0,20
høst	3	0,15
Sum	20	1,00

1.49
a) $\frac{7}{25} = 0{,}28$
b) 28 %

1.50
$0{,}35 \cdot 40 = 14$

1.51
$1 - 0{,}42 - 0{,}31 = 0{,}27$

1.52
8, 8, 12, 17, 19, 25, 30, 33, 42, 51.
Minste: 8, største: 51.

1.53
a) Ja, trolig 1,78 m (178 cm).
b) Nei, mye, men mulig.
c) Ja, karakterskalaen går fra 1 til 6.
d) Nei, rimelig.

1.54
Etter rydding: ja 7 (Ja, JA, ja, ja , jaa, Ja, ja), nei 3.

Svar	Frekvens	Relativ frekvens
ja	7	0,7
nei	3	0,3

1.55
Summen er 28, ikke 30. To elever mangler: Kanskje er en kategori glemt, eller to svar mangler. Lise må sjekke dataene.

1.56
a) 150 (trolig 15).
b) Med: $\frac{258}{8} = 32{,}25$ minutter. Uten: $\frac{108}{7} \approx 15{,}4$ minutter.

1.57
$50 - 4 = 46$

1.58
Så andre (og du selv) kan se hva som er endret og hvorfor, og vurdere om valgene var rimelige. Ryddingen kan endre resultatet mye, så den må kunne etterprøves.

1.59
```
[3, 8, 14, 22]
3 22
```

1.60
```
gym 0.5
matte 0.3333333333333333
norsk 0.16666666666666666
```

1.61
```
['ja', 'ja', 'ja', 'nei']
3
```

1.62
```
antall = 0
for s in svar:
    if s.lower() == "buss":
        antall = antall + 1
print(antall)
```

1.63
Svarene har ulike enheter (ganger, timer) og ord som betyr ulike ting for ulike personer («mye», «litt»), så de kan ikke telles eller regnes med på en god måte. Bedre: «Hvor mange timer trente du forrige uke? Skriv bare et tall.»

1.3.1 Populasjon og utvalg

1.64
a) Populasjon: alle nordmenn over 60 år. Utvalg: de 200 som ble målt.
b) Populasjon: de 25 elevene i klassen. Utvalg: de 5 som ble spurt.
c) Populasjon: de 2000 melkekartongene. Utvalg: de 30 som ble sjekket.

1.65
a) Totalundersøkelse
b) Utvalgsundersøkelse
c) Totalundersøkelse (for klassen)

1.66
For eksempel: Det er billigere, det går raskere, og noen ganger er en totalundersøkelse umulig (for eksempel når testen ødelegger produktet).

1.67
a) $\hat{p} = \frac{276}{1200} = 0{,}23$
b) $\hat{p}$ gjelder bare utvalget. Et annet utvalg ville gitt et litt annet svar (utvalgsvariasjon), og folk kan ombestemme seg før valget.

1.68
a) Utvalg b) Parameter (alle elevene er målt) c) Utvalg

1.69
Alle i populasjonen har like stor sjanse til å komme med, og alle mulige grupper på n er like sannsynlige.

1.70
Er suppa godt rørt, ligner én skje på hele gryta. Et tilfeldig utvalg er som en skje fra en godt rørt gryte: Det ligner på populasjonen. Uten røring (tilfeldighet) kan skjeen gi et feil inntrykk, uansett hvor stor den er.

1.71
Den med $n = 2000$. Begge er tilfeldige, men utvalgsvariasjonen er mye mindre i store utvalg, så den treffer som regel nærmere den sanne andelen.

1.72
Nei. Så lenge utvalget er tilfeldig, er det antallet i utvalget som avgjør presisjonen, ikke hvor stor del av befolkningen det er. 1000 tilfeldige personer gir omtrent like god presisjon i begge land.

1.73
a) $0{,}3 \cdot 10 = 3$
b) Nei. På grunn av utvalgsvariasjon vil noen få 1 eller 2, andre 4 eller 5. I gjennomsnitt over mange utvalg blir det omtrent 3.

1.74
```
3
0.375
```

1.75
```
0.6
0.2
```

1.76
Den trekker 10 ulike elementer tilfeldig fra lista `elever`, uten tilbakelegging. Alle elementene har like stor sjanse, og ingen kan komme med to ganger, slik et enkelt tilfeldig utvalg krever.

1.77
Ikke uten videre. Klassen er et utvalg, men ikke tilfeldig. Elevene går på samme trinn og linje og kan skille seg fra resten av skolen (for eksempel har eldre elever oftere jobb). De bør trekke et tilfeldig utvalg fra hele skolen.

1.78
For eksempel: «I et tilfeldig utvalg på 50 elever gikk 46 % til skolen. Andelen blant alle elevene er trolig i nærheten av dette, men tallet er usikkert fordi utvalget er lite.»

1.3.2 Utvalgsmetoder

1.79
a) Bekvemmelighetsutvalg b) Systematisk c) Klyngeutvalg d) Selvseleksjon e) Stratifisert

1.80
a) $k = \frac{900}{45} = 20$
b) 7, 27 og 47

1.81
Andelen er $\frac{60}{720} = \frac{1}{12}$. Vg1: 25, Vg2: 20, Vg3: 15.

1.82
Andelen er $\frac{400}{16\,000} = 0{,}025$. Sentrum: 300, distriktet: 100.

1.83
I et stratifisert utvalg trekkes noen tilfeldig fra hvert lag, så alle lag er med. I et klyngeutvalg trekkes noen klynger tilfeldig, og alle i de valgte klyngene er med, mens de andre klyngene ikke er med.

1.84
De som er lette å få tak i, ligner ofte ikke på populasjonen. Eksempel: Spør man dem som sitter på biblioteket om hvor mye de leser, får man trolig for høye tall.

1.85
Selvseleksjon. De som er sterkt engasjerte (ofte de som er mot), svarer oftere enn andre, og bare de som ser programmet kan svare. Utvalget ligner ikke på befolkningen.

1.86
a) Klyngeutvalg: trekk noen skoler og spør elevene der.
b) Stratifisert med aldersgrupper som lag, siden bruken trolig varierer med alder.

1.87
Ellers bestemmer den som trekker, hvem som kommer med. Med tilfeldig start har alle i lista like stor sjanse.

1.88
Med start på nummer 1 kommer den beste eleven med, og rekkefølgen henger sammen med karakter. Utvalget kan bli skjevt. Startpunktet må være tilfeldig, og helst bør lista ikke være sortert etter noe som har med undersøkelsen å gjøre.

1.89
$\frac{0{,}5 \cdot 300 + 0{,}6 \cdot 200 + 0{,}8 \cdot 100}{600} = \frac{350}{600} \approx 0{,}58$

1.90

```
2
10
18
26
34
```

1.91

```
25.0
20.0
15.0
```

1.92
For eksempel et stratifisert utvalg med trinn som lag, trukket tilfeldig fra elevlista, siden treningsvaner kan variere med alder. Antallet fra hvert trinn regnes ut i forhold til størrelsen på trinnet.

1.93
Nettavstemningen bygger på selvseleksjon, så de som svarer, ligner ikke på befolkningen. Mange svar fra et skjevt utvalg gir et presist, men galt svar. Den tilfeldige målingen har mer utvalgsvariasjon, men treffer i gjennomsnitt riktig.

1.3.3 Representativitet og feilkilder

1.94
At det ligner på populasjonen i det som betyr noe for spørsmålet som undersøkes.

1.95
a) Dekning b) Ledende spørsmål c) Sosial ønskelighet d) Frafall e) Målefeil

1.96
$\frac{420}{1200} \cdot 100\,\% = 35\,\%$

1.97
a) 70 %
b) $\frac{210}{350} = 60\,\%$
c) Laveste: $\frac{210}{500} = 42\,\%$. Høyeste: $\frac{210 + 150}{500} = 72\,\%$.

1.98
Tilfeldig feil varierer fra utvalg til utvalg og blir mindre med større utvalg. Systematisk feil går i samme retning hver gang og blir ikke mindre med større utvalg.

1.99
a) Systematisk b) Tilfeldig c) Systematisk

1.100
For eksempel: «Hva synes du om prisene i kantina? For høye, passe eller lave?»

1.101
Folk vil ikke innrømme noe som ser dårlig ut når de kan bli gjenkjent. Det kalles sosial ønskelighet.

1.102
Den tilfeldige meningsmålingen. Nettavstemningen har selvseleksjon, og den systematiske feilen forsvinner ikke uansett hvor mange som svarer.

1.103
Utvalget ble trukket fra telefonkataloger og bilregistre, som under depresjonen mest omfattet velstående folk (dekning). Bare omtrent en fjerdedel svarte, og de som svarte, var trolig annerledes enn de andre (frafall).

1.104
De som sover lengst om morgenen, eller som har det travelt, svarer kanskje sjeldnere. Da blir de som står tidlig opp overrepresentert, og resultatet kan vise mindre søvn enn det egentlig er.

1.105

```
60.0
80.0
```

1.106

```
0.5 0.7
```

1.107
Selvseleksjon (de som svarer, er ikke tilfeldig valgt) og sosial ønskelighet (noen vil ikke innrømme at de blir mobbet). Bedre: Trekk et tilfeldig utvalg, eller la alle svare anonymt i en skoletime, og spør nøytralt.

1.108
Ikke nødvendigvis. Frafallet er lite, men utvalget kan likevel være skjevt (dekning), og svarene kan være skjeve (ledende spørsmål, sosial ønskelighet, målefeil).

1.3.4 Observasjonsstudier og eksperimenter

1.109
I en observasjonsstudie registrerer vi det som skjer, uten å påvirke det. I et eksperiment bestemmer vi hvem som får en behandling.

1.110
a) Observasjonsstudie b) Eksperiment c) Observasjonsstudie d) Eksperiment

1.111
En variabel som påvirker begge variablene vi ser på, og kan gi en sammenheng uten at den ene er årsak til den andre. Eksempel: Temperatur påvirker både is-salg og antall drukningsulykker.

1.112
a) Alder (eldre barn har større føtter og leser bedre).
b) Folketallet (store kommuner har mange leger og mange dødsfall).
c) Foreldrenes utdanning og interesse for skole.

1.113
Kontrollgruppe, randomisering (tilfeldig fordeling), blinding (gjerne med placebo) og mange nok deltakere.

1.114
Fordi det å tro at man får behandling, kan påvirke resultatet. Med placebo vet ingen hvem som får den virkelige medisinen, og forskjellen mellom gruppene skyldes virkestoffet.

1.115
Da blir gruppene i gjennomsnitt like i alt annet, også det vi ikke har tenkt på. Forskjellen i resultatet må da skyldes behandlingen eller tilfeldigheter.

1.116
Nei. Det er en observasjonsstudie, så den viser en sammenheng. Det kan være bakenforliggende variabler, for eksempel motivasjon for skolen eller familiens økonomi, eller årsaken kan gå motsatt vei.

1.117
En kontrollgruppe som ikke gjør den nye oppvarmingen, fordelt tilfeldig. Uten den vet vi ikke om skadene hadde gått ned uansett (for eksempel på grunn av vær, kampprogram eller tilfeldigheter).

1.118
10 i hver gruppe. Skriv navnene på lapper, bland dem i en bolle, og trekk: de første 10 til gruppe 1, de neste 10 til gruppe 2 og resten til gruppe 3.

1.119

```
['A', 'B']
['C', 'D', 'E', 'F']
```

1.120

```
1 11
20
```

1.121
Et tilfeldig utvalg trekkes fra populasjonen og gjør det mulig å si noe om populasjonen. En tilfeldig fordeling på grupper i et eksperiment gjør det mulig å si noe om årsak.

1.122
For eksempel: 30 elever fordeles tilfeldig på to grupper. Begge får de samme 20 ordene og like lang tid. Den ene gruppen leser høyt, den andre stille. Etter ti minutter skriver alle ned så mange ord de husker. Antallet riktige sammenliknes mellom gruppene.

1.123
Det er uetisk å la folk røyke for å se om de blir syke. Det ble vist med mange observasjonsstudier som pekte i samme retning, og med kunnskap om hvordan stoffene i røyken skader cellene.

1.4.1 Undersøkelse: Klassens spørreundersøkelse

1.124
1. Formulere spørsmål: statistisk spørsmål, populasjon og variabler.
2. Samle inn data: spørreskjema og utvalg.
3. Analysere data: rydde, frekvenstabell, andeler, figurer.
4. Tolke resultatene: svare på spørsmålet i konteksten og vurdere feilkilder.

1.125
a) Mangler tidsrom og enhet: «Omtrent hvor mange timer sov du i natt? Skriv bare et tall.»
b) Ledende: «Hva synes du om når skolen starter? For tidlig, passe eller for sent?»
c) Ikke anonymt. Spør heller om trinn hvis det trengs.

1.126
For eksempel: «Hvor stor andel av elevene på skolen har deltidsjobb, og hvor mange timer jobber de i uka?» Variabler: deltidsjobb (kategorisk: ja/nei) og timer jobb forrige uke (numerisk). Utvalg: stratifisert på trinn, trukket tilfeldig fra elevlista.

1.127
a) 36
b) buss 0,50, gå 0,25, sykkel 0,17, bil 0,08
c) $\frac{15}{36} \cdot 100\,\% \approx 41{,}7\,\%$

1.128
$\frac{22}{28} \cdot 100\,\% \approx 78{,}6\,\%$. Resultatet gjelder de 22 som svarte i klassen, og bare for klassen hvis de som ikke svarte, ligner på de andre.

1.129
600 minutter er ti timer, og det er urimelig. Verdien tas ut av analysen (eller rettes hvis det er tydelig hva eleven mente), og i ryddeloggen står det at én verdi på 600 minutter ble tatt ut fordi den er urimelig, slik at n ble én mindre.

1.130
For eksempel: «Blant de 45 som svarte, leste 27 (60 %) bok minst én gang i uka. Skjemaet ble delt på Snapchat, så de som svarte, er ikke et tilfeldig utvalg, og vi vet ikke om resultatet gjelder alle elevene.»

1.131
For eksempel: Bekvemmelighetsutvalg (venner og egen klasse) kan gi svar som ligner gruppens egne. Selvseleksjon gjør at de mest engasjerte svarer, og trekker mot ytterpunktene. Sosial ønskelighet gjør at folk svarer det som ser bra ut (for eksempel mer trening og mindre skjermtid enn i virkeligheten).

1.132
```
5
16.0
```

1.133
Spørsmål og populasjon, metode (skjema, utvalg, hvor mange spurt og svarte, rydding), resultater (frekvenstabell, andeler, figurer), konklusjon i konteksten med hvem den gjelder for, feilkilder, og hva man ville gjort annerledes.