11

OPPGAVEHEFTE · STATISTIKK

Det utforskende arbeidet

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. Mange av oppgavene har flere gode svar. Fasiten viser ett av dem.

Formelark

PROBLEMSTILLING OG PLAN

God problemstilling Statistisk, avgrenset, målbar, gjennomførbar og interessant.

Data Primærdata eller sekundærdata. Ta hensyn til personvern.

ANALYSE OG VURDERING

Metode Etter spørsmålet og variabeltypene.

Vurdering Gyldighet, usikkerhet og årsak.

PRESENTASJON

Rapporten Innledning, data og metode, resultater, drøfting, konklusjon, kilder.

Usikkerhet i ord «Vi er 95 % sikre på ...», «henger sammen med», «gir ikke grunnlag for å si ...»

> - Et godt utforskende arbeid svarer grundig på et avgrenset spørsmål og er ærlig om usikkerheten.

11.1.1 Problemstilling og plan

EKSEMPEL *Fødsler*

Gjør temaet «fødsler» om til en problemstilling med underspørsmål.

Løsning

Problemstilling: Hvilke kjennetegn har kommunene der det blir født mange barn i forhold til folketallet?

Underspørsmål:

1. Hvor mye varierer antall fødte per 1000 innbyggere mellom kommunene i 2025?
2. Henger fødselstallet sammen med inntekt, utdanning eller hvor mange som bor i hver husholdning?
3. Er forskjellene mellom små kommuner større enn mellom store, og hvorfor?

Populasjonen er de 357 kommunene. Hovedvariabelen er **fødselsraten**: antall levendefødte i 2025 per 1000 innbyggere. Den er bedre enn antall fødte, fordi store kommuner ellers alltid ville fått flest. ■

EKSEMPEL *Variabeltabell*

Lag en tabell over variablene i fødselsundersøkelsen.

Løsning

Variabel	Type	Kilde
Fødte i 2025	Numerisk, diskret	SSB 04231
Folketall 2026	Numerisk, diskret	SSB 07459
Fødselsrate	Numerisk, kontinuerlig	Regnes ut
Medianinntekt	Numerisk, kontinuerlig	SSB 06944
Høyere utdanning (%)	Numerisk, kontinuerlig	SSB 09429
Personer per husholdning	Numerisk, kontinuerlig	Regnes ut

Alle dataene er sekundærdata for hele populasjonen, så det er ikke noe utvalg. Folketallet er fra 1. januar 2026 og inntekten fra 2024, så årene er ikke helt like. Det tar vi med som en begrensning. ■

EKSEMPEL *En plan for arbeidet*

Lag en plan for fødselsundersøkelsen med metoder og tidsbruk. Arbeidet skal gjøres på fire skoletimer.

Løsning

Time	Arbeid	Metoder
1	Hente og koble data, regne ut fødselsrate og personer per husholdning	pandas
2	Beskrive fødselsraten	Histogram, gjennomsnitt, median, kvartiler, uteliggere
3	Sammenhenger med de andre variablene	Punktdiagram, r, regresjon
4	Vurdering og rapport	Feilkilder, årsak, store talls lov

Underveis noterer vi alle valg og begrunnelser, så de kan brukes i rapporten og på eksamen. ■

11.1 Skriv de fem kjennetegnene ved en god problemstilling i et utforskende arbeid.

11.2 Er problemstillingen statistisk? Begrunn.

a) Hvor høy er Galdhøpiggen?

b) Hvor mange timer i uka bruker elevene på Vg2 på lekser?

c) Er Messi bedre enn Ronaldo?

11.3 Gjør temaet «trening» om til en avgrenset problemstilling med to underspørsmål.

11.4 Forklar forskjellen på primærdata og sekundærdata, og gi ett eksempel på hver.

11.5 Nevn én fordel og én ulempe med sekundærdata fra SSB.

11.6 Hvorfor er fødselsraten (fødte per 1000 innbyggere) bedre enn antall fødte når vi sammenlikner kommuner?

11.7 Lag en variabeltabell (variabel, type, kilde) for problemstillingen «Sover elever som bruker mobilen i senga, mindre enn andre elever?»

11.8 Nevn tre regler for personvern når du samler inn data om andre elever.

11.9 I en klasse på 25 elever spør du om kjønn, alder og om eleven har en diagnose. Hva er problemet?

11.10 Hva blir utskriften?

Python

```
fodte = [7, 120]
folketall = [477, 15000]
for f, n in zip(fodte, folketall):
    print(round(f / n * 1000, 1))
```

11.11 Hvorfor er det bedre å svare grundig på et lite spørsmål enn overfladisk på et stort?

11.1.2 Analyse og vurdering

EKSEMPEL *Vurdering av fødselsundersøkelsen*

Vurder analysen av fødselsraten etter de tre punktene.

Løsning

Gyldighet: Dataene er fra SSB og dekker alle kommunene, så de er pålitelige og ikke et utvalg. Men folketallet er fra 2026, inntekten fra 2024 og fødslene fra 2025. Fødselsraten måler fødte per innbygger, ikke hvor mange barn hver kvinne får. Det er et annet mål.

Usikkerhet: I en kommune med 500 innbyggere svarer én fødsel til 2 fødte per 1000. Små kommuner kan derfor få svært høye eller lave rater av rene tilfeldigheter, slik vi så for andelen gutter i 5.1.1. Standardavviket for fødselsraten er 2,87 for kommuner under 2000 innbyggere og 1,56 for kommuner over 5000. Uteliggerne er nesten bare små kommuner, og vi bør være forsiktige med å tolke dem.

Årsak: Sammenhengen med personer per husholdning går trolig delvis motsatt vei: Når det blir født barn, blir husholdningene større (omvendt årsak). Den viktigste bakenforliggende variabelen er trolig **alderen** i kommunen: Kommuner med mange unge voksne mellom 25 og 40 år får flere barn. Det forklarer også Oslo, som har mange unge voksne, men små husholdninger fordi mange bor alene. Vi har ikke aldersdata for kommunene i datasettet, så det er et naturlig neste steg. ∎

11.12 Hvilken graf og hvilke mål passer?

a) Hvor mange minutter elevene bruker på skoleveien.

b) Hvilket parti elevene ville stemt på.

c) Sammenhengen mellom høyde og lengdehopp.

d) Skoleveien for elever som tar buss og elever som sykler.

11.13 Når bør du bruke median i stedet for gjennomsnitt?

11.14 I klasse A (20 elever) har 60 % deltidsjobb, og i klasse B (30 elever) har 40 % det. Hvor stor andel av alle elevene har deltidsjobb?

11.15 Skriv de tre spørsmålene en vurdering av en undersøkelse skal svare på.

11.16 Hva er en bakenforliggende variabel? Gi et eksempel fra fødselsundersøkelsen.

11.17 Forklar hvorfor sammenhengen mellom fødselsrate og personer per husholdning delvis kan være omvendt årsak.

11.18 Hvorfor er det mest ekstreme fødselsrater i små kommuner?

11.19 Hvilken feilkilde er det mest snakk om?

a) Svarene ble samlet inn på foreldremøtet.

b) «Hvor mye trener du?» uten tidsperiode.

c) Elevene veide seg på ulike vekter.

11.20 En elev finner $r = 0{,}39$ mellom inntekt og fødselsrate i kommunene. Hvor stor andel av variasjonen forklarer inntekten?

11.21 En elev tar bort alle kommuner under 5000 innbyggere fordi sammenhengen da blir sterkere. Når er det greit, og når er det et problem?

11.1.3 Presentere og kommunisere usikkerhet

EKSEMPEL *Konklusjon for fødselsundersøkelsen*

Skriv en konklusjon på fire–fem setninger for fødselsundersøkelsen, med riktig grad av sikkerhet.

Løsning

Fødselsraten varierte fra 0 til 14,7 fødte per 1000 innbyggere mellom kommunene i 2025, men de fleste kommunene lå mellom 7 og 10. De ekstreme verdiene var nesten bare i små kommuner, der tilfeldigheter betyr mye. Kommuner med høy medianinntekt og store husholdninger hadde som regel høyere fødselsrate, men sammenhengene var moderate ($r = 0{,}39$) og forklarte bare omtrent 15 % av variasjonen. Siden dette er en observasjonsstudie, kan vi ikke si at inntekt eller husholdningsstørrelse er årsaken. Alderen i befolkningen er trolig en viktig bakenforliggende variabel, og det bør undersøkes videre. ∎

11.22 Skriv de seks delene i en rapport om et utforskende arbeid.

11.23 Hvor i rapporten hører det til?

a) «Vi trakk 80 tilfeldige elever fra elevlista.»

b) «Medianen var 35 minutter.»

c) «Utvalget er lite, så …»

11.24 Skriv om: «Vi har bevist at trening gir bedre søvn.» (Observasjonsstudie, moderat sammenheng.)

11.25 En test gir p-verdi 0,18. Skriv konklusjonen i en setning.

11.26 45 av 120 tilfeldige elever sier at de leser bøker på fritiden. Skriv resultatet med usikkerhet.

11.27 Hva skal være med på en god figur i en rapport?

11.28 Nevn tre spørsmål sensor kan stille på muntlig eksamen om et utforskende arbeid.

11.29 Hvorfor er det lurt å drøfte svakheter i eget arbeid på eksamen?

11.30 Hva blir utskriften?

Python

```
x = 45
n = 120
p = x / n
m = 1.96 * (p * (1 - p) / n)**0.5
print(round(p - m, 2), round(p + m, 2))
```

11.31 Hva er galt med å oppgi «internett» som kilde?

Fasit

11.1.1 Problemstilling og plan

11.1
Statistisk, avgrenset, målbar, gjennomførbar og interessant.

11.2
a) Nei, svaret er ett fast tall. b) Ja, svaret avhenger av data som varierer. c) Nei, slik det står. Det må avgrenses og gjøres målbart, for eksempel «Hvem hadde flest mål per kamp i ligaen i perioden ...?»

11.3
For eksempel: «Trener elevene på Vg1 like mye som anbefalingen på 60 minutter fysisk aktivitet om dagen?» Underspørsmål: Hvor mange minutter trener elevene en vanlig uke? Er det forskjell mellom elever som driver med organisert idrett og de andre?

11.4
Primærdata samler du inn selv, for eksempel en spørreundersøkelse i klassen. Sekundærdata er samlet inn av andre, for eksempel folketall fra SSB.

11.5
Fordel: store, pålitelige data, ofte for hele befolkningen. Ulempe: de svarer bare på de spørsmålene SSB har stilt, og er ofte oppsummert (for eksempel per kommune, ikke per person).

11.6
Store kommuner får alltid flest fødte fordi de har flest innbyggere. Raten gjør kommuner av ulik størrelse sammenliknbare.

11.7
Søvn (numerisk kontinuerlig, spørreundersøkelse), mobil i senga (ja/nei, kategorisk, spørreundersøkelse), gjerne trinn (kategorisk, spørreundersøkelse).

11.8
For eksempel: Spør bare om det du trenger, undersøkelsen skal være anonym, ikke spør om sensitive opplysninger uten avklaring, deltakelsen skal være frivillig, og rådataene skal slettes etterpå.

11.9
Diagnoser er sensitive helseopplysninger. Med kjønn og alder i en liten klasse kan elevene i tillegg kjennes igjen.

11.10
14.7 og 8.0

11.11
Et lite, presist spørsmål kan besvares med data og analyse innenfor tiden du har, og konklusjonen kan begrunnes og vurderes. Et stort spørsmål gir vage svar som ikke kan støttes av dataene.

11.1.2 Analyse og vurdering

11.12
a) Histogram eller boksplott, median og kvartiler (trolig skjev). b) Søylediagram, andeler. c) Punktdiagram, r og regresjonslinje. d) Boksplott side om side, median per gruppe.

11.13
Når fordelingen er skjev eller har uteliggere, fordi medianen ikke påvirkes mye av ekstreme verdier.

11.14
$\frac{12+12}{50} = 48\,\%$, ikke gjennomsnittet 50 %.

11.15
Gyldighet (måler vi det vi vil, er dataene pålitelige og representative?), usikkerhet (hvor mye kan tilfeldigheter forklare?) og årsak (kan sammenhengen skyldes noe annet?).

11.16
En variabel som påvirker begge variablene vi ser på, slik at de henger sammen uten at den ene påvirker den andre. Alderen i kommunen kan påvirke både fødselsraten og inntekten.

11.17
Når det blir født barn, blir husholdningene større. Da kan mange fødsler føre til store husholdninger, ikke bare omvendt.

11.18
Med få innbyggere er det få fødsler, og noen få fødsler mer eller mindre gir stor forskjell i raten. Tilfeldig variasjon betyr mer når tallene er små (store talls lov).

11.19
a) Skjevt utvalg. b) Uklart spørsmål. c) Målefeil.

11.20
$R^2 = 0{,}39^2 \approx 0{,}15$, altså omtrent 15 %.

11.21
Det er greit når valget er begrunnet på forhånd (små kommuner har stor tilfeldig variasjon) og nevnes i rapporten. Det er et problem hvis eleven prøver mange varianter og velger den som gir det peneste resultatet.

11.1.3 Presentere og kommunisere usikkerhet

11.22
Innledning, data og metode, resultater, drøfting, konklusjon, kilder.

11.23
a) Data og metode b) Resultater c) Drøfting

11.24
For eksempel: «Elever som trener mye, sover som regel litt lengre, men sammenhengen er moderat. Det er en observasjonsstudie, så forskjellen kan skyldes andre forhold.»

11.25
«Dataene gir ikke grunnlag for å si at det er en forskjell.» (Ikke «Det er ingen forskjell.»)

11.26
$\hat{p} = 0{,}375$ og feilmarginen er $1{,}96 \cdot \sqrt{\frac{0{,}375 \cdot 0{,}625}{120}} \approx 0{,}087$. «Vi er 95 % sikre på at mellom 29 % og 46 % av elevene leser bøker på fritiden.»

11.27
Aksetitler med enheter, kilde, og en tydelig hovedsak uten unødvendig pynt. Ingen kuttede akser eller 3D som villeder.

11.28
For eksempel: «Hvorfor valgte du median?», «Kan resultatet skyldes tilfeldigheter?», «Betyr sammenhengen at det ene fører til det andre?», «Hva ville du undersøkt videre?»

11.29
Det viser at du forstår begrensningene i data og metoder, som er en sentral del av statistikkfaget. Alle undersøkelser har svakheter.

11.30
0.29 0.46

11.31
Leseren kan ikke finne igjen dataene eller vurdere hvor pålitelige de er. Oppgi hvem som har laget dataene, tabellnummer eller nettadresse, og datoen du hentet dem.