

# Statistikk-quiz Samling 1.

Lederen for et kjøpesenter ønsker å finne ut hva kundene synes om senterets åpningstider. Hun utformer et spørreskjema med fire spørsmål: Alder, kjønn, og

- Hyppighet: *Hvor ofte er du på kjøpesenteret?*. Det er tre svaralternativ: *Mindre enn en gang i uka, en-to ganger i uka, oftere.*
- Tilfredshet med åpningstider: *Hvor tilfreds er du med åpningstidene?*. Det er fire svaralternativ: *åpner for sent, stenger for tidlig, tilfreds, vet ikke.*

1. Hva vil du si er populasjonen i dette studiet?
2. Lederen legger ut ark med spørreskjemaene i infoskranken. De som ønsker å svare er med i trekningen av et gavekort. Nevn et hovedproblem med denne fremgangsmåten. Er utvalget representativt?
3. Man ønsker et utvalg stratifisert etter kjønn. Hvordan kan man oppnå det?
4. La oss anta at lederen skaffer til veie et utvalg på en fornuftig måte, og at 36 respondenter svarer på undersøkelsen. Hvilken programvare kan hun bruke for å analysere dataene?
5. Hvor mange rader og søyler blir det i regnearket?
6. Etter at arkene er lest inn i programvare, så beregner lederen gjennomsnittsalderen. Er dette tallet en parameter eller en observator?
7. Lederen ønsker å bestemme hvor mange % av kundene på senteret som er menn. Er denne andelen en parameter eller en observator?
8. Påstand: *Populasjonsparameteren varierer fra utvalg til utvalg.* Sann eller usann? Begrunn.
9. Påstand: *Observatoren varierer fra utvalg til utvalg.* Sann eller usann? Begrunn.
10. Senterlederen lager et histogram over aldersammensetningen. Kalles dette deskriptiv statistikk eller inferens? Begrunn.
11. På bakgrunn av gjennomsnittsalderen i utvalget så estimeres alderen i populasjonen. Den blir 46 år, med en feilmargin<sup>1</sup> på 6 år. Hva kalles denne prosessen?

---

<sup>1</sup>Dette kalles konfidensintervall

12. Team 1 består av 8 studenter. Alderen deres er 20, 20, 19, 23, 29, 27, 25 og 21 år. Beregn medianalder.
13. R: Beregn standardavviket til alder i Team 1. Bruk funksjonen  $sd()$ .
14. Team 2 består av fem studenter som alle er 23 år. Beregn standardavviket til alder i Team 2.
15. Team 3 består av seks studenter med alderen 18, 19, 27, 35, 29 og 20 år. Team 4 består av fem studenter med alderen 29, 27, 26, 30 og 31 år. Hvilket av disse teamene har størst standardavvik? Du trenger ikke beregne noe her, bare argumentere.
16. Team 4 består av syv studenter med alderen: 18, 19, 47, 20, 22, 21, 20, 19 og 20 år. Foretrekker du gjennomsnitt eller median som et mål på senter her? Begrunn.
17. La oss tenke oss at startlønn i et stor utvalg av sykkelbud er normalfordelt med  $\bar{x} = 450000$  kr og at standardavviket er  $s = 15000$  kr.
  - (a) Vil du karakterisere en startlønn på 470000 kr som uvanlig?
  - (b) Anslå hvor mange prosent av sykkelbud som har startlønn under 420000 kr.
18. En fordeling er venstreskjev. Vil  $z$ -scoren til medianen være negativ, null eller positiv?  
\*\*\*\*
19. **REMA 1000** Renate er leder på Rema 1000 på Klepp. Hun skriver masteroppgave der hun definerer problemstillingen til å være *Hva er viktigst for deltidsansatte i detaljvarehandelen i Norge, lønn eller fleksibel arbeidstid?* Hva er populasjonen i studiet?
20. Renate utarbeider en spørreundersøkelse. Fra alle landets Rema butikker velger hun tilfeldig ut 25 butikker. Så kontakter hun butikklederne ved disse 25 butikkene, og får e-mail lister til alle deltidsansatte. Hun sender link til et online spørreskjema til disse epost adressene. Hva heter den type utvalg som Renate får?
21. Er utvalget representativt? Finn to svakheter.
22. Renate får svar fra 67 respondenter. Gjennomsnittet for timer jobbet i måneden er  $\bar{x} = 40$  timer med standardavvik  $s = 12$  timer. Er  $s$  en parameter eller en observator?
23. Fabian er deltidsansatt og jobber 60 timer i måneden. På bakgrunn av Renates tallmateriale, vil du si at han jobber uvanlig mye for å være deltidsansatt?
24. Seks deltidsansatte på Klepp jobbet i januar 23, 55, 60, 70, 65 og 60 timer. Beregn standardavviket  $s$  i R.  
\*\*\*\*
25. På øya Pingelab er det svært mange fargeblinde. I et utvalg av 240 innbyggere så var 26 personer fargeblinde. Beregn punktestimatet  $\hat{p}$  for andel fargeblinde.

26. Forklar hvorfor det egentlig ikke holder å oppgi kun punktestimatet  $\hat{p}$ , hvilken informasjon er det vi trenger i tillegg?
27. Konstruer vha av R et 95 % konfidensintervall for andel fargeblinde på Pingelab<sup>2</sup>. Gi en tolkning av intervallet.
28. I et utvalg av 56 studenter var det 28 menn. Konfidensintervallet som konstrueres for andelen menn i populasjonen er (0.393, 0.607). Hva var konfidensnivået?
29. Du har lager to konfidensintervall, ett med konfidensnivå 90 %, og ett med konfidensnivå 99 %. Hvilket er bredest/lengst?
30. Vi antar at alderen på Pingelab er normalfordelt. Du tar et tilfeldig utvalg av fem øyboere: 30 44 34 41 59 35. bruk R til å lage et 95% konfidensintervall for gjennomsnittsalderen  $\mu$  på øya.<sup>3</sup>
31. Alder er egentlig ikke normalfordelt i Pingelab populasjonen. Vi trenger  $n > 30$  for å lage konfidensintervall. I fila "agePingelab.csv" som ligger ligger aldrene til 60 tilfeldig valgte øyboere. Last inn datafila i R ved å bruke `read.csv("agePingelab.csv")`. Hva er utvalgsgjennomsnittet?
32. Lag et 90% konfidensintervall for alder på Pingelab.
33. Det påstås at snittalder på Pingelab er 45 år. Test denne hypotesen ( $H_0 : \mu = 45$ ) mot en tosidig  $H_A$ , med et 5% signifikansnivå. Bruk funksjonen `t.test()`. Har vi tilstrekkelig støtte til å forkaste  $H_0$  ?  
\*\*\*\*
34. Herr Flink utfører ofte hypotesetester i sin jobb som markedsanalytiker. Til helgen skal Herr Flink foreta en hypotesetest og han vil følge prosedyren i læreboka uten å gjøre feil. Kan det da allikevel skje en type I feil? Forklar
35. Han har utført femhundre tester opp igjennom årene, og han bruker alltid et signifikansnivå på  $\alpha = 0.05$ . La oss anta at i alle testene så var  $H_0$  korrekt. Anslå hvor mange ganger Herr Flinks tester endt med en type I feil?
36. La oss anta at i alle testene så var  $H_0$  korrekt. Anslå hvor mange ganger Herr Flinks tester har endt med en type II feil?
37. Herr Flink foretar en hypotesetest vha JMP og ser at  $p$ -verdien er 0.032. Han benytter et signifikansnivå på  $\alpha = 0.05$ . Skal han beholde eller forkaste  $H_0$ ?
38. Fru Alfa foretar en hypotesetest med  $\alpha = 0.1$  som signifikansnivå, og konklusjonen blir at  $H_0$  beholdes. Hva vet du om  $p$ -verdien i denne testen?

---

<sup>2</sup>bruk funksjonen `prop.test()`

<sup>3</sup>bruk funksjonen `t.test()`

39. Herr Flink ønsker å teste om andelen forbrukere som kunne tenke seg å kjøpe rugbrød de neste to ukene er mer enn 10 %. Må han da bruke et signifikansnivå på  $\alpha = 0.1$ ?

40. En hypotesetest fører til forkastning av  $H_0$  når signifikansnivået er  $\alpha = 0.05$ . Dersom signifikansnivået hadde vært  $\alpha = 0.1$ , ville det fortsatt resultert i forkastning?

\*\*\* Regresjon

41. Vi tar et tilfeldig utvalg av ti studenter. Vi har følgende info  $x$ : antall ganger møtt på undervisning, og  $y$ : antall poeng på eksamen. Tror du det er negativ, ingen eller positiv korrelasjon mellom  $x$  og  $y$ ?

42. Her er dataene for de ti studentene:

oppmøte ( $x$ )	2	15	8	14	8	12	13	2	5	4
poeng ( $y$ )	13	26	20	27	16	24	20	11	16	20

Bruk R og funksjonen `plot(x,y)` til å lage et spredningsdiagram

43. Viser spredningsdiagrammet en lineær trend? Hvilke andre antagelser må være oppfylt for at inferensen i regresjon skal være gyldig? Ser det ut til at også disse antagelsene er oppfylt?

44. Bruk funksjonen `lm()` til å beregne regresjonslinja. Hva blir stigninstallet? Tolk dette.

45. Mikal er en student som ikke var med i det opprinnelige utvalget. Han møtte til undervisning åtte ganger. Bruk regresjonslinja til å anslå hvor mange poeng han vil få på eksamen.

46. Mikal fikk faktisk 21 poeng på eksamen. Hva er residualet til Mikal?

47. Professoren påstår at jo mer du følger med på undervisningen, jo bedre går det på eksamen. Du ønsker å foreta en hypotesetest. Skriv opp  $H_0$  og  $H_A$ .

48. Hva er p-verdien til testen?

49. Hva blir konklusjonen på testen dersom du bruker signifikansnivå  $\alpha = 0.01$ ?