

Eksamen er todelt. Første del inneholder oppgaver som skal besvares basert på informasjon i dette oppgavesettet. Andre del inneholder spørsmål omkring analyser av datasettet som ble levert ut to uker før eksamen. *Begrunn kort dine svar, som regel holder det med 2-3 setninger. I oppgave 5 kan du bruke 5-10 setninger per oppgave.*

Del 1

PISA (Programme for International Student Assessment) er en internasjonal undersøkelse som måler 15-åringers kompetanse i blant annet matematikk. Den gir også informasjon om elevenes interesser og deres hjemmemiljø. Vi skal se på PISA fra året 2015, som ble gjennomført i 72 land. Til sammen deltok nesten 540 000 elever, og disse ble trukket ut slik at de skulle representere rundt 28 millioner 15-åringer som gikk på skole i deltakerlandene.

Vi skal se nærmere på følgende variabler:

- *Skolearbeid*: 'Hvor mange timer bruker du på skolearbeid utenom skoletiden?'.
• *Matematikk*: Antall poeng på matematikktesten i PISA.
• *Bøker*: 'Hvor mange bøker har dere hjemme?'. Det var seks svaralternativ: Færre enn 11 bøker, 11-25 bøker, 26-100 bøker, 101-200 bøker, 201-500 bøker og Flere enn 500 bøker.
• *Stille sted*: 'Har du et rolig sted du kan gjøre lekser på hjemme?'. Det var to svaralternativ: Ja og Nei.

Oppgave 1

- (a) Hva er populasjonen i PISA undersøkelsen?
- (b) På hvilket målenivå er *Skolearbeid*?
- (c) På hvilket målenivå er *Bøker*?
- (d) På hvilket målenivå er *Stille sted*?
- (e) Nevn to typer grafer som brukes for å fremstille fordelingen til variabler som er på samme målenivå som *Matematikk*.
- (f) Vi regner ut gjennomsnittet for *Matematikk* i det norske utvalget. Er dette tallet en observator eller en parameter?

Oppgave 2

Stolpediagrammet i Figur 1 bakerst i oppgavesettet viser fordelingen til *Bøker* blant de 5278 norske elevene.

- (a) Bestem medianverdien til *Bøker* blant de 5278 norske elevene.
- (b) I en gruppe av seks studenter, så er antallet bøker hjemme lik 8, 18, 13, 15, 180 og 200. Finn medianverdien i denne gruppen.
- (c) Beregn standardavviket til de seks observasjonene fra forrige oppgave.

Oppgave 3

La oss anta at *Matematikk* (antall poeng på mattetesten) til en tilfeldig valgt norsk 15-åring er normalfordelt med $\mu = 499$ poeng og standardavvik $\sigma = 85$ poeng.

- (a) z -verdien til Lars sin mattescore var $z = -1$. Hvor mange poeng fikk Lars?
- (b) Hva vil du si er en uvanlig høy mattescore?
- (c) Hvor stor andel av norske 15-åringer skårer under 329 poeng?
- (d) Dersom man lager et 95% konfidensintervall for μ basert på skårene til de 5278 norske elevene, hva er sannsynligheten for at det ikke inneholder 499?
- (e) Det var 2612 jenter i utvalget, og 2666 gutter. Standardavvikene for mattescore var $s_J = 80.2$ for jentene og $s_G = 89.6$ for guttene. For hvilket kjønn varierte matematikkprestasjonene mest?
- (f) En forsker påstår at norske 15 åringer i gjennomsnitt skårer mer enn 499 på PISA prøven. Skriv opp hypotesene H_0 og H_A for å utføre en test for denne påstanden.
- (g) Dersom forskeren utfører testen på signifikansnivå 10%, hva er da sannsynligheten for at han finner støtte for sin feilaktige påstand?
- (h) Kan forskeren oppleve å gjøre en Type II feil i dette scenariet? Begrunn.

Oppgave 4

Vi skal se på sammenhengen mellom *Skolearbeid* og *Matematikk* blant norske 15-åringer.

- (a) Finn et potensielt argument for at korrelasjonen mellom *Skolearbeid* og *Matematikk* vil være positiv.

- (b) Finn et potensielt argument for at korrelasjonen mellom *Skolearbeid* og *Matematikk* vil være negativ.
- (c) Regresjon av *Matematikk* på *Skolearbeid* og kjønn gir dette resultatet

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	527.5409	2.2101	238.69	0.0000
kjønnjente	-2.8600	2.1927	-1.30	0.1922
skolearbeid	-0.8922	0.0851	-10.48	0.0000

Nevn minst to antagelser for regresjon som bør holde for at du skal kunne stole på resultatene i denne tabellen.

- (d) Dersom vi kontrollerer for skolearbeid, ser det ut til å være kjønnsforskjell i *Matematikk*?
- (e) R^2 adjusted for denne modellen er 0.0231. Hvor mye av variasjon i matematikkscore kan ikke forklares av kjønn og *Skolearbeid*?
- (f) Bruk modellen til å beregne forventet mattescore for en jente som bruker 20 timer i uka på skolearbeid utenom skoletiden.
- (g) Du ønsker å forklare variasjon i *Stille sted* ved hjelp av prediktoren *Bøker*. Hvilken type regresjonsmodell kan du da bruke?

Oppgave 5

- (a) Forklar hva en meta-analyse er for noe, og nevnt noen typiske trinn i en slik analyse.
- (b) Beskriv den såkalte replikasjonskrisen i psykologi, og nevnt årsaker til at den kunne finne sted.
- (c) Forklar hva et måleinstrument sin reliabilitet er for noe. Hva er det vanligste målet for reliabilitet?
- (d) Forklar hva et måleinstrument sin validitet er for noe. På hvilke måter kan vi bedømme validiteten?
- (e) Forklar hva en latent variabel er for noe, og hvordan den typisk måles. Gi også tre eksempler på en latent variabel i psykologi.

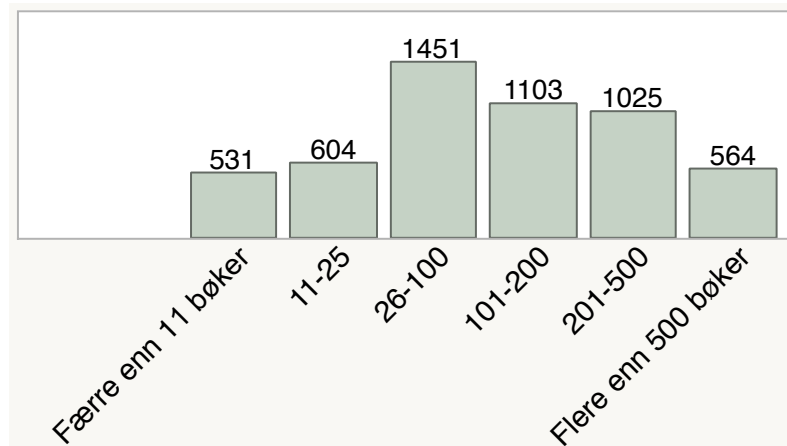
Del 2

Oppgave 6

- (a) Hvor mange respondenter hadde missing på en eller flere variabler?
- (b) Hvor mange prosent av respondentene var kvinner?
- (c) Hva var standardavviket for alder?
- (d) Ser alder ut til å være tilnærmet normalfordelt?
- (e) Hva er gjennomsnittet til NONREACT?
- (f) Hva er korrelasjonene mellom NONREACT og AWARENESS?
- (g) Hva er gjennomsnittet til DESCRIBE hos menn og hos kvinner?
- (h) Er det signifikant kjønnsforskjell i variabelen DESCRIBE?
- (i) Anta at pooled standardavvik til DESCRIBE er $s_p = 6.4$. Beregn effektstørrelsen Cohens d for DESCRIBE ift til kjønn. Vil du karakterisere kjønnsforskjell i DESCRIBE som liten, medium eller stor?
- (j) Hvor stor andel av variasjon i DESCRIBE kan forklares av alder og kjønn?
- (k) Er en eldre person forventet å kunne uttrykke sine følelser bedre eller dårligere enn en yngre person, når vi tar høyde for kjønn?
- (l) Hva er reliabiliteten til AWARENESS måleinstrumentet?
- (m) Hvilket AWARENESS item ville du ha fjernet dersom man kun ser på de standardiserte faktorladningene?
- (n) Hva kan skje med reliabiliteten når man fjerner items fra et konstrukt? Hva kan skje med validiteten?
- (o) Hva blir korrelasjonen mellom NONREACT og OBSERVE når vi behandler dem som latente variable? Kommenter forskjellen til korrelasjonen vi fikk når vi brukte summeskårer.
- (p) $fa()$ funksjonen med 5 faktorer resulterer i en tabell med faktorladninger. Hvilke av de 39 items lader kun på en faktor? Hvordan tolker du at et item lader kun på 1 faktor?
- (q) $fa()$ funksjonen med 5 faktorer resulterer i en tabell med faktorladninger. Hvilke av de 39 itemene lader på mer enn tre faktorer? Hvordan tolker du at et item lader på mange faktorer?

- (r) Reflekter omkring strukturen i faktorladningene fra `fa()` funksjonen med 5 faktorer. Kan vi si at vi tilnærmet har en 5-faktor struktur i vårt datasett?
- (s) `fa.parallel()` anslår hvilket antall faktorer ?
- (t) Nevn en annen metode enn parallel-analyse for å anslå antall faktorer i et instrument.

SLUTT.



Figur 1: Antall bøker hjemme hos norske elever. Antallet i hver kategori er skrevet på toppen av stolpene.