i Forside

Institutt for matematiske fag

Eksamensoppgave i ISTA1001, ISTA1002, ISTA1003, ISTG1001, ISTG1002, ISTG1003, ISTT1001, ISTT1002, ISTT1003, VB6200 Statistikk

Eksamensdato: 16.12.2022

Eksamenstid (fra-til): 09:00 – 12:00

Hjelpemiddelkode/Tillatte hjelpemidler: C
Godkjent kalkulator
Digitale ressurser under eksamen: Formelark og fem tabeller (binomisk kumulativfordeling, Poisson kumulativfordeling, normal kumulativfordeling, normalfordeling kritisk verdi og t-fordeling kritisk verdi) er lagt ved eksamen som pdf-filer

Faglig kontakt under eksamen:
Thea Bjørnland (41123849)
Charles Curry (48201626)

Faglig kontakt møter i eksamenslokalet: NEI

ANNEN INFORMASJON:

Henvend deg til en eksamensvakt hvis du ønsker å kontakte faglig kontaktperson under eksamen. Noter spørsmålet ditt på forhånd. Faglig kontaktperson skal kun kontaktes dersom det er direkte feil eller mangler i oppgavesettet.

Vekting av oppgavene: er oppgitt for hver oppgave. Det gis ikke minus-poeng for gale eller manglede svar.

Lagring: Besvarelsen din i Inspera Assessment lagres automatisk hvert 15. sekund.

Varslinger: Hvis det oppstår behov for å gi beskjeder til kandidatene underveis i eksamen (f.eks. ved klare mangler eller feil i oppgavesettet), vil dette bli gjort via varslinger i Inspera. Et varsel vil dukke opp som en dialogboks på skjermen i Inspera. Du kan finne igjen varselet ved å klikke på bjella øverst i høyre hjørne på skjermen.

Trekk fra eksamen: Blir du syk under eksamen, eller av andre grunner ønsker å levere blankt/trekke deg, gå til «hamburgermenyen» i øvre høyre hjørne og velg «Lever blankt». Dette kan ikke angres selv om prøven fremdeles er åpen. Tilkall så eksamensvakt og følg instruksene fra eksamensvakten.

Tilgang til besvarelse: Du finner besvarelsen din i Arkiv etter at sluttida for eksamen er passert.

1 Oppgave 1: Normalfordelingen

Oppgaven består av 4 deloppgaver.

a) La Z være en standard normalfordelt stokastisk variabel. Hva er forventning og varians til Z?
Velg ett alternativ:

- E(Z) = 0 og Var(Z) = 2
- E(Z) = 0 og Var(Z) = 0
- E(Z) = 1 og Var(Z) = 1
- E(Z) = 1 og Var(Z) = 2
- E(Z) = 1 og Var(Z) = 0
- E(Z) = 0 og Var(Z) = 1 ✔

b) La X være en normalfordelt stokastisk variabel med forventningsverdi 3 og standardavvik 2.
Hva er $P(X \leq 1)$?
Velg ett alternativ

- 0.3421
- 0.4672
- 0.1587 ✔
- 0.6587
- 0.9878
- 0.2648

c) La Z være en standard normalfordelt stokastisk variabel, og la Y = 2+4Z. Hvilken fordeling har Y?

Velg ett alternativ

- Y er normalfordelt med forventningsverdi 6 og standardavvik 4
- Y er normalfordelt med forventningsverdi 2 og standardavvik 2
- Vi har ikke nok informasjon til å vite hvilken fordeling Y har
- Y er normalfordelt med forventningsverdi 6 og standardavvik 2
- Y er normalfordelt med forventningsverdi 2 og standardavvik 4 ✔
- Y er normalfordelt med forventningsverdi 0 og standardavvik 2

d) La $X_1, X_2, \ldots, X_{50}$ være uavhengige stokastiske variabler der hver $X_i \sim \mathbf{Eksponential}(2)$ for $i = 1, 2, \ldots, 50$. La $Y = X_1 + X_2 + \cdots + X_{50}$. Hvilket utsagn stemmer?

Velg ett alternativ

- Y er tilnærmet Poissonfordelt med en rate lik 100
- Y er tilnærmet normalfordelt med forventningsverdi 0.5 og varians 0.5
- Y er tilnærmet normalfordelt med forventningsverdi 25 og varians 12.5 ✔
- Vi har ikke nok informasjon til å si noe om fordelingen til Y
- Y er tilnærmet Poissonfordelt med en rate lik 2
- Y er tilnærmet normalfordelt med forventningsverdi 100 og varians 50

2 Oppgave 2: Estimatorer for lengde

Oppgaven består av 2 deloppgaver.

Frøya vil måle lengden (i cm) på en skjøteledning så nøyaktig som mulig. La μ betegne den ukjente lengden. Frøya gjør 5 uavhengige målinger med en meterstokk. Vi kan anta at hver måling X_i $(i = 1, ..., 5)$ er normalfordelt med forventningsverdi μ cm og standardavvik 1 cm. Senere får Frøya låne en lasermåler som hun bruker til å gjøre én måling Y. Vi kan anta at Y er normalfordelt med forventningsverdi μ cm og standardavvik 0.2 cm, og at Y er uavhengig av tidligere målinger.

For å anslå lengden μ vurderer Frøya fem ulike estimatorer:

$$\hat{\mu}_1 = Y \quad \hat{\mu}_2 = \frac{\sum_{i=1}^5 X_i}{5} \quad \hat{\mu}_3 = \frac{1}{2}\left(Y + \frac{\sum_{i=1}^5 X_i}{5}\right) \quad \hat{\mu}_4 = \frac{Y+\sum_{i=1}^5 X_i}{6} \quad \hat{\mu}_5 = 5Y - \sum_{i=1}^5 X_i$$

a) Hvilken av de fem estimatorene er *ikke* forventningsrett?
Velg ett alternativ

- $\hat{\mu}_1$
- $\hat{\mu}_2$
- $\hat{\mu}_3$
- $\hat{\mu}_4$
- $\hat{\mu}_5$ ✔
- Alle estimatorene er forventningsrette

b) Hvilken estimator har lavest varians?
Velg ett alternativ

- $\hat{\mu}_1$ ✔
- $\hat{\mu}_2$
- $\hat{\mu}_3$
- $\hat{\mu}_4$
- $\hat{\mu}_5$
- Alle estimatorene har samme varians

3 Oppgave 3: Produksjon av børsteløse motorer

Oppgaven består av 5 deloppgaver.

En bedrift produserer børsteløse motorer til bruk i radiostyrte kjøretøy. Motorene gjennomgår kvalitetssikring før salg. Av erfaring vet bedriften at sannsynligheten for at en tilfeldig valgt motor er feilfri er 0.85. Sannsynligheten for at en motor har en feil som kan repareres er 0.11, og sannsynligheten for at en motor må kasseres (har en feil som ikke kan repareres) er 0.04.

a) Gitt at en tilfeldig valgt motor ikke er feilfri, hva er sannsynligheten for at den kan repareres?
Velg ett alternativ:

- 0.359
- 0.733 ✔
- 0.815
- 0.524
- 0.408
- 0.653

b) Hva er sannsynligheten for at minst 23 av 25 tilfeldig valgte motorer er feilfrie?
Velg ett alternativ

- 0.473
- 0.254 ✔
- 0.043
- 0.163
- 0.398
- 0.535

Under kvalitetskontroll er det en viss usikkerhet i om en motor skal repareres. Dersom en motor er feilfri er det en 2% sjanse for at den unødig blir sendt til reparasjon, for en motor med en feil som kan repareres er det en 90% sjanse for at motoren blir sendt til reparasjon, og for en motor som må kasseres er det en 40% sjanse for at motoren likevel blir sendt til reparasjon.

c) Hva er sannsynligheten for at en motor blir sendt til reparasjon?

Velg ett alternativ

- 0.412
- 0.243
- 0.384
- 0.564
- 0.619
- 0.132 ✔

Etter at en motor har passert kvalitetskontroll hos bedriften (er feilfri eller reparert), har motoren en Weibull-fordelt levetid (målt i timer) med parametere $\alpha = 1.5$ og $\lambda = 1/400$.

d) Hva er sannsynligheten for at en motor virker i mer enn 300 timer?

Velg ett alternativ

- 0.169
- 0.412
- 0.522 ✔
- 0.042
- 0.286
- 0.349

e) I den radiostyrte bilen DualXRacer brukes to motorer samtidig. Bilen fungerer bare dersom begge motorene fungerer. Vi kan anta at motorenes levetider er uavhengige. Hva er sannsynligheten for at bilen virker i mer enn 300 timer?

Velg ett alternativ

- 0.476
- 0.006
- 0.029
- 0.273 ✔
- 0.135
- 0.357

4 Oppgave 4: Lakselus ved oppdrettsanlegg

Oppgaven består av 4 deloppgaver.

Ved et oppdrettsanlegg for laks er det bestemt at forventet antall lakselus (av typen kjønnsmodne hunnlus) ikke skal overskride 0.5 lus per laks.

Antall lus på en tilfeldig valgt laks kan antas å være Poissonfordelt med ukjent forventningsverdi λ lus per laks. Dersom alt er som det skal i anlegget er raten λ høyst lik 0.5. Avlusing er en kostbar prosedyre, så bedriften vil kun starte med avlusing dersom de har *svært gode* indikasjoner på at raten λ overstiger 0.5.

Bedriften tar en stikkprøve bestående av 10 tilfeldige valgte laks. En stokastisk variabel X som representerer totalt antall lus på 10 tilfeldig valgte laks vil være Poissonfordelt med forventning 10λ. Bedriften observerer til sammen x=8 lus.

a) Hva blir den estimerte raten basert på disse observasjonene?
Velg ett alternativ:

- $\hat{\lambda} = 0.6$
- $\hat{\lambda} = 0.4$
- $\hat{\lambda} = 0.5$
- $\hat{\lambda} = 0.7$
- $\hat{\lambda} = 0.8$ ✔
- $\hat{\lambda} = 0.9$

For å avgjøre om avlusing skal iverksettes vil bedriften teste H_0: $\lambda = 0.5$ mot H_1: $\lambda > 0.5$. Testen skal utføres ved å beregne en p-verdi (sannsynligheten for det som er observert eller noe mer ekstremt i retning alternativhypotesen når vi antar at H_0 er sann) og forkaste H_0 dersom p-verdien er lavere enn det valgte signifikansnivået $\alpha = 0.10$.

b) Beregn p-verdien ved å regne ut $P(X \geq 8)$ under antagelsen om at null-hypotesen er sann, og konkluder om H_0 kan forkastes eller ikke.

Velg ett alternativ

- Siden p = 0.1334 forkastes ikke H_0 til fordel for H_1 ✔
- Siden p = 0.1334 forkastes H_0 til fordel for H_1
- Siden p = 0.2061 forkastes H_0 til fordel for H_1
- Siden p = 0.0487 forkastes ikke H_0 til fordel for H_1
- Siden p = 0.0487 forkastes H_0 til fordel for H_1
- Siden p = 0.2061 forkastes ikke H_0 til fordel for H_1

c) For å holde signifikansnivået α = 0.10 finner bedriften ut at H_0 kan forkastes for observasjoner på totalt 9 lus eller mer, når 10 fisk velges ut tilfeldig. Hva blir da teststyrken dersom H_1 er sann og λ = 0.85 lus per laks?

Velg ett alternativ

- 0.5734
- 0.4769 ✔
- 0.2368
- 0.3651
- 0.6791
- 0.7573

Bedriften er usikker på om de kan stole på resultatene av stikkprøven fordi utvalget var lite. De tar derfor en ny stikkprøve bestående av 20 fisk. En stokastisk variabel Y som representerer totalt antall lus på 20 tilfeldig valgte laks vil være Poissonfordelt med forventning 20λ. Bedriften observerer til sammen y=16 lus i den nye stikkprøven.

d) Hva blir nå konklusjonen av testen av H_0: λ = 0.5 mot H_1: λ > 0.5? Bruk signifikansnivå α = 0.10.

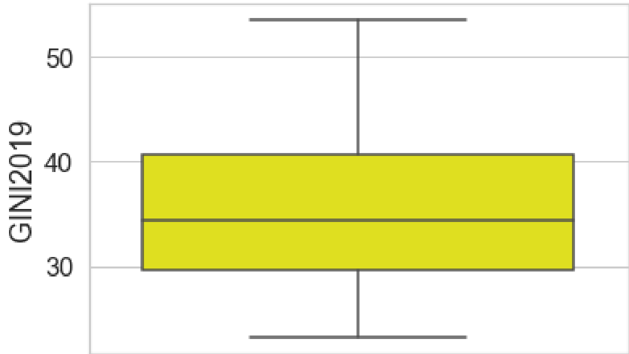
Velg ett alternativ

- Siden p = 0.0487 forkastes ikke H_0 til fordel for H_1
- Siden p = 0.1334 forkastes H_0 til fordel for H_1
- Siden p = 0.2061 forkastes H_0 til fordel for H_1
- Siden p = 0.2061 forkastes ikke H_0 til fordel for H_1
- Siden p = 0.1334 forkastes ikke H_0 til fordel for H_1
- Siden p = 0.0487 forkastes H_0 til fordel for H_1 ✔

5 Oppgave 5: Økonomisk ulikhet og pressefrihet

Oppgaven består av 5 deloppgaver.

Ginikoeffisienten er et mål på inntektsforskjeller i et land. Ginikoeffisienten tar verdier mellom 0 (fullstendig likhet ved at alle personer har samme inntekt) og 100 (størst mulig ulikhet ved at én person har all inntekt). I Norge var for eksempel Ginikoeffisienten i 2019 lik 27.7.

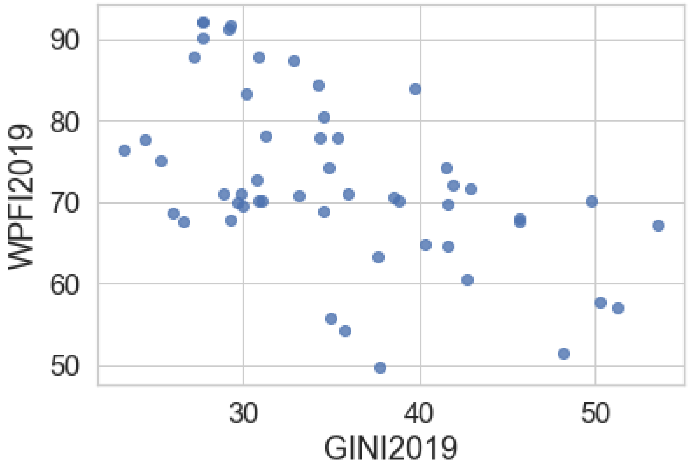


Boksplottet illustrerer Ginikoeffisienten for 2019 i n = 52 land.

a) Bruk boksplottet til å anslå median Ginikoeffisient for de 52 landene.
Velg ett alternativ:

- 30
- 50
- 35 ✔
- 45
- 25
- 40

World Press Freedom Index, forkortet WPFI, er et mål på graden av pressefrihet i et land og varierer mellom 0 (ingen pressefrihet) og 100 (fullstendig pressefrihet). Norge var det landet med høyest score på pressefrihet i 2019 (WPFI = 92.2).



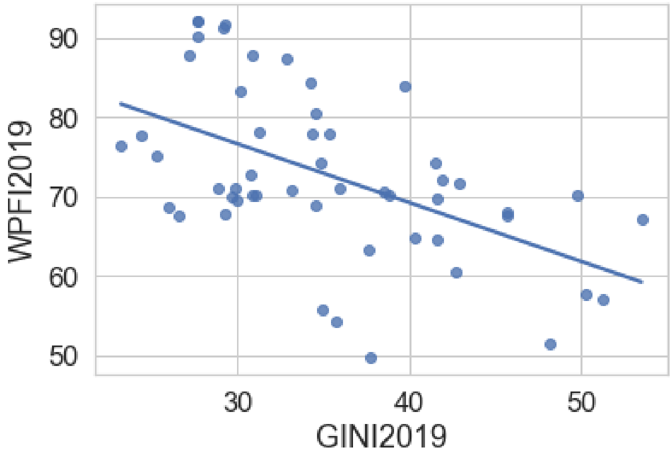
Figuren viser et kryssplott av Ginikoeffisienten og WPFI i n = 52 land i 2019.

b) Hva er sant om den empiriske korrelasjonen mellom variablene GINI2019 og WPFI2019?
Velg ett alternativ

- Korrelasjonen er omtrent 0
- Korrelasjonen er omtrent 0.5
- Korrelasjonen er omtrent 1
- Korrelasjonen er omtrent -0.5 ✔
- Korrelasjonen er omtrent -1
- Vi har ikke nok informasjon til å si noe om korrelasjonen

Utskriften viser resultatet av å tilpasse en lineær regresjonsmodell til dataene i Python. Responsen Y er WPFI og forklaringsvariabelen x er Ginikoeffisienten. Figuren under viser et kryssplott av dataene (n = 52) med den estimerte regresjonslinja.

	coef	**std err**	**t**
Intercept	98.8508	6.053	16.332
GINI2019	-0.7399	0.168	-4.412



c) Bruk utskriften til å finne ligningen for den estimerte regresjonslinja.
Velg ett alternativ

- $\hat{y} = 98.8508 + 6.053x$
- $\hat{y} = -0.7399 + 0.168x$
- $\hat{y} = 0.168 - 4.412x$
- $\hat{y} = 98.8508 - 0.7399x$ ✔
- $\hat{y} = 6.053 + 0.168x$
- $\hat{y} = 16.332 - 4.412x$

d) Hva blir predikert WPFI for et land med en Ginikoeffisient lik 42.6?
Velg ett alternativ

- 36.12
- 85.63
- 58.23
- 67.33 ✔
- 42.95
- 72.14

e) Regn ut et 90% konfidensintervall for stigningstallet i regresjonslinja.

Velg ett alternativ

- [-0.642, -0.258]
- [-0.734, -0.266]
- [-1.021, -0.458]
- [-1.130, -0.080]
- [-1.232, -0.068]
- [-0.821, -0.476]