Erstatt denne teksten med ditt innhold...

Institutt for matematiske fag

Eksamensoppgave i TMA4240/45 Statistikk

Eksamensdato: 06.08.2020

Eksamenstid (fra-til): 09:00–13:00

Hjelpemiddelkode/Tillatte hjelpemidler: A / Alle hjelpemidler tillatt

Faglig kontakt under eksamen: Geir-Arne Fuglstad og Håkon Tjelmeland
Tlf.: 45 27 08 06 og 48 22 18 96

Teknisk hjelp under eksamen: NTNU Orakel
Tlf: 73 59 16 00

ANNEN INFORMASJON:

Gjør dine egne antagelser og presiser i besvarelsen hvilke forutsetninger du har lagt til grunn i tolkning/avgrensing av oppgaven. Faglig kontaktperson skal kun kontaktes dersom det er direkte feil eller mangler i oppgavesettet.

Lagring: Besvarelsen din i Inspera Assessment lagres automatisk. Jobber du i andre programmer – husk å lagre underveis.

Juks/plagiat: Eksamen skal være et individuelt, selvstendig arbeid. Det er tillatt å bruke hjelpemidler. Alle besvarelser blir kontrollert for plagiat. *Du kan lese mer om juks og plagiering på eksamen her*.

Varslinger: Hvis det oppstår behov for å gi beskjeder til kandidatene underveis i eksamen (f.eks. ved 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. Det vil i tillegg bli sendt SMS til alle kandidater for å sikre at ingen går glipp av viktig informasjon. Ha mobiltelefonen din tilgjengelig.

Vekting av oppgavene: Det er 20 deloppgaver som teller likt ved sensur.

Fritekstoppgaver: Følg instruksene og beskriv med egne ord de stegene du har tatt for å komme frem til svaret.

OM LEVERING:

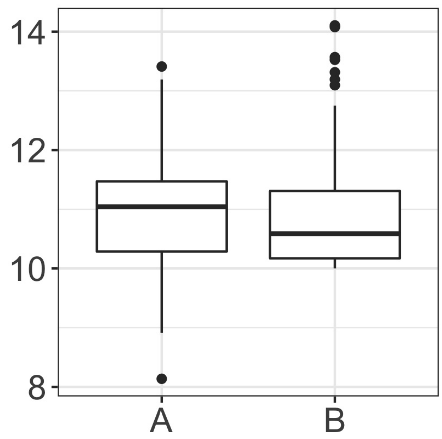
Besvarelsen din leveres automatisk når eksamenstida er ute og prøven stenger, forutsatt at minst én oppgave er besvart. Dette skjer selv om du ikke har klikket «Lever og gå tilbake til Dashboard» på siste side i oppgavesettet. Du kan gjenåpne og redigere besvarelsen din så lenge prøven er åpen. Dersom ingen oppgaver er besvart ved prøveslutt, blir ikke besvarelsen din levert.

Trekk fra eksamen: Ønsker du 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.

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

1

(a)



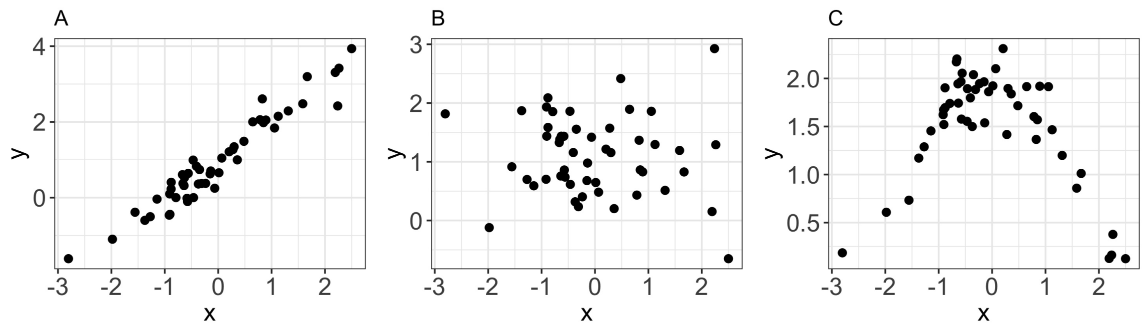
Boksplottene over beskriver to ulike datasett: Datasett A og Datasett B.
Kommenter forskjeller/likheter i lokasjon, spredning og skjevhet i de to datasettene.

Skriv ditt svar her

Ord: 0

Maks poeng: 5

(b)



Figuren over viser tre ulike kryssplott (spredningsplott). I hvert plott er observasjoner av to stokastiske variabler, X og Y, plottet mot hverandre.

Hva kan du si om sammenhengen mellom X og Y i de tre ulike situasjonene? Er de uavhengige? Er de korrelerte?

Skriv ditt svar her

Ord: 0

Maks poeng: 5

2

(a) Fotballaget "United" vinner 60% av kampene der de skårer kampens første mål, men bare 10% av kampene der motstanderlaget skårer det første målet. 5% av "United" sine kamper ender med stillingen 0-0.

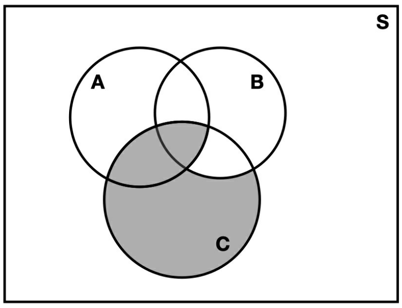
Dersom "United" skårer det første målet i 30% av sine kamper, hvor stor andel av kampene sine vinner de?

Velg ett alternativ:

- 0.245
- 0.250
- 0.700
- 0.180

Maks poeng: 5

(b)



Hvilken hendelse viser det skraverte området i venndiagrammet?

Velg ett alternativ:

- $A \cap B \cap C'$
- $(A \cap C) \cup (C \cap (A \cup B)')$
- $B' \cup (A \cup C)$
- $(A \cup C) \cup (B \cup C)$

Maks poeng: 5

(c) Vi ser på to hendelser, A og B, slik at P(A) > 0 og P(B) > 0. Påstanden "Dersom A og B er uavhengige hendelser, så er P(A|B)=P(B|A)" er

Velg ett alternativ:

- Alltid sann
- Sann kun dersom A og B er disjunkte hendelser
- Sann kun dersom A og B er like sannsynlige
- Alltid usann

Maks poeng: 5

3 La X og Y være to diskrete stokastiske variabler med simultanfordeling som gitt i tabellen under

x / y	0	1	2	3
0	0.1	0.05	0.20	0.10
1	0.05	0.15	0.30	0.05

Tabell: Simultan punktsannsynlighet $P(X = x, Y = y)$

(a)

Regn ut marginal punktsannsynlighet $P(X = 0)$. Oppgi svaret med to desimaler.

$P(X = 0) =$ ☐

Maks poeng: 5

(b) Hva er $P(Y \leq 1|X = 0)$?

Velg ett alternativ:

- 0.222
- 0.333
- 0.150
- 0.450

Merk: Svaralternativene er avrundet til 3 desimaler

Maks poeng: 5

(c) Er X og Y uavhengige stokastiske variabler? Forklar hvordan du kom fram til svaret ditt.

Skriv ditt svar her

Maks poeng: 5

(d) Vi ser på funksjonen $g(X, Y) = X + 2XY$. Hva blir forventingsverdien $\mathrm{E}(g(X, Y))$? Forklar og vis utregningene dine.

Skriv ditt svar her

Maks poeng: 5

4 Levetiden til en elektrisk komponent, uttrykt ved den stokastiske variabelen X, er eksponensialfordelt med parameter $\lambda = 0.8$.

Sannsynlighetstettheten er gitt ved

$$f(x) = 0.8 \cdot e^{-0.8x}$$

for $x \geq 0$

(a) Hva er forventet levetid til en slik elektrisk komponent? Skriv svaret avrundet til to desimaler.

$E(X) =$ ☐ .

Maks poeng: 5

(b) Hva er sannsynligheten for at komponenten virker i over ett år?

Velg ett alternativ:

- ○ 0.449
- ○ 0.359
- ○ 0.641
- ○ 0.551

Merk: Svaralternativene er avrundet til 3 desimaler

Maks poeng: 5

(c) I en maskin er 5 slike elektriske komponenter koblet i serie. La X_i være levetiden til komponent i, $i = 1, \ldots, 5$, og anta at komponentenes levetider er uavhengige. Hva er sannsynligheten for at maskinen fungerer i minst ett år?

Velg ett alternativ:

- ○ 0.015
- ○ 0.005
- ○ 0.018
- ○ 0.007

Merk: Svaralternativene er avrundet til 3 desimaler

Maks poeng: 5

(d) En annen maskin inneholder kun en komponent av denne typen, men når denne komponenten slutter å virke blir den umiddelbart byttet ut med en ny komponent av samme type. Anta at vi har tilgjengelig fire slike reservekomponenter, i tillegg til komponenten som står i når vi starter maskinen. Levetiden til maskinen blir dermed

$$Y = X_1 + X_2 + X_3 + X_4 + X_5$$

der X_1, X_2, X_3, X_4 og X_5 altså er uavhengige og eksponensialfordelte med parameter $\lambda = 0.8$.

Beskriv en fremgangsmåte du kan bruke for å bevise at Y er gammafordelt. Forklar også hvordan du samtidig kan bestemme parameterverdiene α og β i denne gammafordelingen.

Skriv ditt svar her

Maks poeng: 5

5 Dr. Hansen ønsker å undersøke om et intensivt treningsopplegg har en blodtrykksdempende effekt på pasienter med høyt systolisk blodtrykk. La den stokastiske variabelen X representere en pasients systoliske blodtrykk før treningsopplegget, og la Y representere pasientens blodtrykk etter fullført treningsopplegg. Dr. Hansen antar at X er normalfordelt med forventningsverdi μ_x og varians σ_x^2, og at Y er normalfordelt med forventningsverdi μ_y og varians σ_y^2.

Fra en gruppe pasienter med høyt blodtrykk trekker Dr. Hansen tilfeldig ut 10 deltakere til en studie. Hun måler pasientenes blodtrykk før og etter treningsopplegget er gjennomført.

La X_i og Y_i representere blodtrykket til person i før og etter fullført treningsopplegg, for $i = 1, \ldots, 10$. Blodtrykksmålingene (x_i og y_i) av de ti pasientene er gitt i tabellen:

	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$	$i=8$	$i=9$	$i=10$
x_i	142.1	168.3	145.3	176.3	184.0	150.5	168.6	158.7	160.3	147.1
y_i	136.0	137.5	132.7	179.7	188.1	148.9	165.1	153.5	161.1	129.8

Tabell: Systolisk blodtrykk målt før (x_i) og etter (y_i) gjennomføring av treningsopplegget.

(a) Er det rimelig å anta at X og Y er uavhengige stokastiske variabler? Diskuter svaret ditt.

Skriv ditt svar her

Maks poeng: 5

(b) Differensen $D = Y - X$ er også en normalfordelt stokastisk variabel. Argumenter for at forventningsverdien er gitt ved $\mu_D = \mu_y - \mu_x$. Bruk observasjonene i tabellen til å regne ut estimater for forventningsverdien og standardavviket til D. Du trenger ikke vise utregninger.

Skriv ditt svar her

Maks poeng: 5

(c) Ved å ta utgangspunkt i det tilfeldige utvalget $D_1, \ldots, D_{10}$, der $D_i = Y_i - X_i$, for $i = 1, \ldots, 10$, ønsker Dr. Hansen å teste

$H_0 : \mu_y - \mu_x = 0$ mot $H_1 : \mu_y - \mu_x < 0$

med signifikansnivå $\alpha = 0.1$.

Skriv ned en passende testobservator for denne testen, og angi hvilken fordeling denne testobservatoren har når H_0 er sann. Hva blir forkastningsområdet?

Basert på observasjonene i tabellen, regn ut og skriv ned observert verdi av testobservatoren og konklusjonen av testen. Har Dr. Hansen grunn til å tro at treningsopplegget virker?

Skriv ditt svar her

Maks poeng: 5

(d) Med bakgrunn i denne studien, diskuter og eksemplifiser forskjellen på type-1 feil og type-2 feil.

Skriv ditt svar her

Maks poeng: 5

(e) Når vi gjør en ensidig hypotesetest slik som i denne oppgaven, så representerer p-verdien:

Velg ett alternativ:

- Sannsynligheten for det vi har observert, eller noe mer ekstremt i retning av den alternative hypotesen, når vi antar at nullhypotesen er sann.
- Sannsynligheten for at de observerte resultatene er statistisk signifikante, gitt at nullhypotesen er sann
- Sannsynligheten for at nullhypotesen er sann, gitt observasjonene
- Sannsynligheten for at alternativhypotesen er sann, gitt observasjonene

Maks poeng: 5

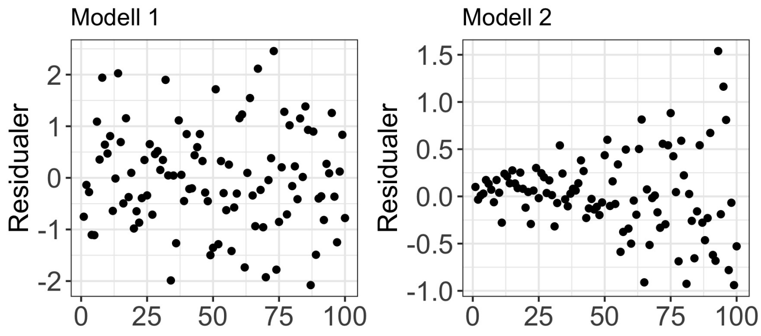
6

(a)

Vi har tilpasset en enkel lineær regresjonsmodell $y = \alpha + \beta x + \epsilon$ til to ulike datasett. Hvert datasett består av 100 observasjoner. Figurene under viser residualplott for de to modellene. Residualet til observasjon i er definert ved

$\hat{\epsilon}_i = y_i - \hat{\alpha} - \hat{\beta} x_i$

der $\hat{\alpha}$ og $\hat{\beta}$ er parameterestimater.



Hvilke(n) av antagelsene i lineær regresjon kan du sjekke ved å plotte residualene? For hvert av de to datasettene, vurder om datasettet virker å tilfredstille disse modellantagelsene.

Skriv ditt svar her

Maks poeng: 5

(b) Vi har tilpasset en enkel lineær regresjonsmodell $y = \alpha + \beta x + \epsilon$ til et datasett bestående av 15 observasjoner. Vi har regnet ut $\hat{\beta} = 1.109$, $s^2 = 0.762$, og $\sum_{i=1}^{15}(x_i - \bar{x})^2 = 18.028$. Regn ut et 95% konfidensintervall for β.

Velg ett alternativ:

- [0.706, 1.512]
- [0.622, 1.596]
- [0.745, 1.473]
- [0.665, 1.553]

Maks poeng: 5