i Forside

Eksamensoppgave i TMA4240/TMA4245 Statistikk
Dato: 04.08.2025
Tid: 09:00

Faglig kontakt: Geir-Arne Fuglstad og Jarle Tufto
Møter i eksamenslokalet: Nei

Hjelpemiddelkode/Tillatte hjelpemidler: C

- Tabeller og formler i statistikk (Akademika)
- Bestemt enkel kalkulator i henhold til NTNUs generelle reglement
- Ett A5-ark med egne formler og notater

ANNEN INFORMASJON

Les oppgavene nøye og gjør dine egne antagelser. Presiser i besvarelsen hvilke forutsetninger du har lagt til grunn i tolkning/avgrensing av oppgaven.

Alle svar skal begrunnes og besvarelsen skal inneholde naturlig mellomregning slik at det er helt klart hvilke regneregler som benyttes.

Dersom det er direkte feil eller mangler i oppgavesettet og du ikke kan gjøre dine egne antagelser, vises det til informasjon om klage på formelle feil på NTNU sin nettside om «Begrunnelse og klage».

FAGSPESIFIKK INFORMASJON

Papirhåndtegninger:
Oppgave 1 til 6 skal besvares direkte i Inspera. Oppgavene 7a, 7b, 7c, 7d, 8a, 9a og 9b skal alle besvares på separate ark. Nederst i hver oppgave finner du en sjusifret kode. Fyll inn denne koden øverst til venstre på arkene du ønsker å levere.

Det anbefales å gjøre dette underveis i eksamen. Dersom du behøver tilgang til kodene etter at eksamenstiden har utløpt, må du klikke «Vis besvarelse».

Du er selv ansvarlig for å fylle inn riktige koder på eventuelle håndtegningsark. Les derfor informasjonen på omslagsarket nøye. Eksamenskontoret kan ikke garantere at feilaktig utfylte ark blir lagt til besvarelsen din.

Vekting av oppgavene: I oppgave 1 til 6 teller hver oppgave 5 poeng. I oppgavene 7, 8 og 9 teller hver av deloppgavene (7a, 7b, 7c, 7d, 8a, 9a og 9b) 10 poeng.

Trekk fra/avbrutt eksamen:
Dersom du ønsker å levere blankt/avbryte eksamen, gå til "hamburgermenyen" i øvre høyre hjørne og velg «Lever blankt». Dette kan ikke angres selv om prøven fremdeles er åpen.

Tilgang til besvarelse:
Etter eksamen finner du besvarelsen din under tidligere prøver i Inspera. Merk at det kan ta én virkedag før eventuelle håndtegninger vil være tilgjengelige under «tidligere prøver».

1 # 1a)

Innledning:
La X og Y være uavhengige normalfordelte stokastiske variabler. Det er kjent at X har forventningsverdi 1 og varians 9 og at Y har forventningsverdi 0 og varians 16.

Oppgave:
Bestem følgende sannsynligheter. Angi svarene med minst to desimaler etter komma. Eventuell mellomregning bør gjøres med fire desimaler etter komma.

$\mathrm{P}(X+Y>1) =$ ☐

$\mathrm{P}((X+Y)^2 \geq 1) =$ ☐

Maks poeng: 5

2 # 2a)

Innledning:
La X og Y være to stokastiske variabler. Det er kjent at X har forventningsverdi 1 og varians 4, at Y har forventningsverdi -1 og varians 3, og at kovariansen mellom X og Y er -2.

Oppgave:
Regn ut følgende størrelser. Oppgi svarene med minst to desimaler etter komma. Eventuell mellomregning bør gjøres med fire desimaler etter komma.

$\mathrm{E}[2X+Y] =$ ☐

$\mathrm{Var}[2X+Y] =$ ☐

Maks poeng: 5

3

3a)

Innledning:
La X og Y være to uavhengige poissonfordelte stokastiske variabler. Det er kjent at X har forventningsverdi 6 og at Y har forventningsverdi 4.

Oppgave:
Bestem følgende sannsynligheter. Angi svarene med minst to desimaler etter komma. Eventuell mellomregning bør gjøres med fire desimaler etter komma.

$\mathrm{P}(X \leq 7) =$ ☐

$\mathrm{P}(X + Y \leq 7) =$ ☐

Maks poeng: 5

4

4a)

Innledning:
La X være en stokastisk variabel med sannsynlighetstetthet $f(x) = \frac{1}{5}\exp(-x/5)$, $x > 0$.

Oppgave:
Bestem følgende sannsynligheter. Angi svarene med minst to desimaler etter komma. Eventuell mellomregning bør gjøres med fire desimaler etter komma.

$\mathrm{P}(X > 5) =$ ☐

$\mathrm{P}(X > 10 | X > 5) =$ ☐

Maks poeng: 5

5

5a)

Innledning:
La A og B være hendelser i et utfallsrom *S*. Sannsynlighetene til hendelsene er $\mathrm{P}(A) = 0.2$ og $\mathrm{P}(B) = 0.9$.

Oppgave:
Velg i hver del under det kulepunktet som angir det mest korrekte utsagnet om hendelsene A og B.

Velg ett alternativ:

- Vi har ikke tilstrekkelig informasjon til å bestemme om hendelsene er disjunkte eller ikke disjunkte
- Hendelsene er ikke disjunkte
- Hendelsene er disjunkte

Velg ett alternativ

- Hendelsene er uavhengige
- Vi har ikke tilstrekkelig informasjon til å bestemme om hendelsene er uavhengige eller avhengige
- Hendelsene er avhengige

Maks poeng: 5

6 6a)

Innledning:

Et fotballag har 20 spillere. De 20 spillerene kan deles inn i 3 målvakter, 7 forsvarsspillere og 10 angrepsspillere. Blant disse 20 spillerene skal det velges ut et startlag med 11 spillere.

Oppgave:

Bestem følgende størrelser.

Antall måter å velge 11 spillere der

a) 1 spiller er målvakt og 10 spillere ikke er målvakter:

b) 1 spiller er målvakt, 4 spillere er forsvarsspillere og 6 spillere er angrepsspillere:

Maks poeng: 5

Innledning:

I denne oppgaven mener vi med en gammafordeling med parametere $a > 0$ og $b > 0$, en sannsynlighetstetthet

$$f(x) = \frac{1}{b^a \Gamma(a)} x^{a-1} \exp(-x/b), \quad x > 0,$$

der $\Gamma(a)$ betegner gammafunksjonen evaluert i a.

Anta at vi har et tilfeldig utvalg $X_1, \ldots, X_n$ fra en gammafordeling med parametere $a = 5$ og b ukjent. Vi skal i denne oppgaven se på tilnærmet og eksakt konfidensintervall for b.

Vi gjør et utvalg av størrelse $n = 10$ og observerer $\bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = 13.049$.

7
7a)

- Vis at sannsynlighetsmaksimeringsestimatoren til b er gitt ved $\hat{B} = \frac{1}{5n} \sum_{i=1}^{n} X_i$.
- Regn ut estimatet $\hat{b}$ for verdiene i oppgaven.
- Vis at $\hat{B}$ er forventningsrett for b, og at variansen er gitt ved $\mathrm{Var}[\hat{B}] = \frac{b^2}{5n}$.

Maks poeng: 10

8
7b)

- Forklar hvorfor vi kan betrakte $\hat{B}$ som omtrent normalfordelt.
- Gjør om nødvendig ytterligere approksimasjoner, og forklar i så fall disse.
- Regn ut et tilnærmet 95% konfidensintervall for b.

Maks poeng: 10

9
7c)

I stedet for å gjøre en tilnærming basert på normalfordelingen er det mulig å jobbe med eksakt fordeling for $\hat{B}$.

- Vis ved bruk av momentgenererende funksjoner at $\sum_{i=1}^{n} X_i$ er gammafordelt.
- Vis at $\frac{10n\hat{B}}{b}$ er kjikvadratfordelt med $10n$ frihetsgrader.

Maks poeng: 10

10 7d)

Basert på resultatet i 7c skal vi konstruere et eksakt 95% konfidensintervall for b.

- Utled et eksakt 95% konfidensintervall for b med utgangspunkt i fordelingen til $\frac{10n\hat{B}}{b}$ fra 7c.
- Regn ut numerisk verdi av konfidensintervallet.
- Sammenlign de numeriske verdiene av det eksakte konfidensintervallet og det tilnærmede konfidensintervallet i oppgave 7b. Diskuter hvor god tilnærmingen var.

Maks poeng: 10

11 8a)

Innledning:
La den stokastiske variabelen X ha en sannsynlighetstetthet gitt ved

$$f(x) = \frac{3}{x^4}, \quad x > 1.$$

Anta at U er en kontinuerlig uniformt fordelt stokastisk variabel på intervallet $(0, 1)$.

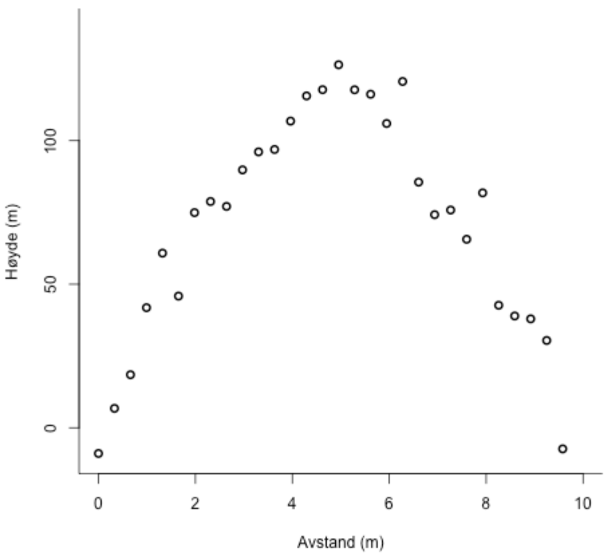
Oppgave:

- Bestem kumulativ fordelingsfunksjon til X.
- Vis at hvis $Y = (1-U)^{-1/3}$ så har Y samme fordeling som X.
- Bruk dette til å skrive en python-funksjon simX som generer simuleringer av X basert på simuleringer av U. Funksjonen skal ta som input antall simuleringer n, og gi som output n simuleringer av X.

Maks poeng: 10

Innledning:

Vi er interessert i bevegelsen av en kule som skytes ut fra en kanon. Denne bevegelsen beskrives ved høyden av kulen i ulike avstander fra kanonen. For å studere denne bevegelsen måler vi høyden av kulen ved ulike avstander. Vi samler inn et datasett bestående av $n = 30$ målinger av høyde (i meter), $y_1, \ldots, y_n$, med tilhørende avstander (i meter) henholdsvis, $x_1, \ldots, x_n$. Målingene er vist i figur 1 under.



Figur 1: Kryssplott av $y_1, \ldots, y_n$ mot $x_1, \ldots, x_n$.

Som figur 1 viser er det stor usikkerhet i målingene og vi skal betrakte to ulike lineære regresjonsmodeller for å beskrive bevegelsen.

Modell 1:

$$Y_i = \alpha_1 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n,$$

der α_1 og β_1 er ukjente koeffisienter, og $\epsilon_1, \ldots, \epsilon_n$ er uavhengige normalfordelte stokastiske variabler med forventningsverdi 0 og ukjent varians $\sigma_1^2 > 0$.

Modell 2:

$$Y_i = \alpha_2 + \beta_2 z_i^2 + \epsilon_i, \quad i = 1, \ldots, n,$$

der $z_i = x_i - 4.9$, α_2 og β_2 er ukjente koeffisienter, og $\epsilon_1, \ldots, \epsilon_n$ er uavhengige normalfordelte stokastiske variabler med forventningsverdi 0 og ukjent varians $\sigma_2^2 > 0$.

Følgende størrelser er oppgitt for datasettet:

- $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 4.7895$,
- $\overline{z^2} = \frac{1}{n} \sum_{i=1}^n z_i^2 = 8.1859$,
- $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = 71.0258$,
- $\sum_{i=1}^n (x_i - \bar{x}) y_i = 242.104$,
- $\sum_{i=1}^n \left(z_i^2 - \overline{z^2} \right) y_i = -8178.659$,
- $\sum_{i=1}^n (x_i - \bar{x})^2 = 245.210$, og

- $\sum_{i=1}^{n}\left(z_i^2-\overline{z^2}\right)^2=1610.039$.

12 ## 9a)

I denne oppgaven trenger dere ikke utlede estimatorene. Bruk formelsamling.

- Regn ut minste kvadratsumestimatene til α_1, β_1, α_2 og β_2, og angi regresjonslinjene.
- Skisser målingene og begge regresjonslinjene i én figur.
- Diskuter hvilke antagelser som er brutt i modell 1, og diskuter om antagelsene i modell 2 virker å være oppfyllt.

Maks poeng: 10

13 ## 9b)

I denne oppgaven bruker vi modell 2.

Produsenten av måleutstyret som er brukt for å måle verdiene $y_1,...,y_n$ påstår at feilen i målingene skal ha et standardavvik på 5 meter. Du mener at usikkerheten må være mye høyere enn dette, og ønsker å overbevise produsenten om at det er en feil med utstyret. For å gjøre dette ønsker du å gjennomføre en hypotesetest med signifikansnivå 0.05.

I tillegg til opplysningene i starten av oppgaven er det oppgitt at $\sum_{i=1}^{n}(y_i-\hat{\alpha}_2-\hat{\beta}_2 z_i^2)^2=2615.707$.

- Formuler passende nullhypotese og alternativ hypotese.
- Bestem en forkastningsregel.
- Gjennomfør hypotesetesten og gi konklusjonen.

Maks poeng: 10