

i **Forside**

Eksamensoppgåve i TMA4240/TMA4245 Statistikk

Dato: 04.08.2025

Tid: 09:00

Fagleg kontakt: Geir-Arne Fuglstad og Jarle Tufto

Kjem til eksamenslokalet: Nei

Hjelpemiddelkode/Tillatne hjelpemiddel: C

- Tabeller og formler i statistikk (Akademika)
- Bestemd enkel kalkulator i samsvar med NTNUs generelle reglement
- Eitt A5-ark med egne formlar og notat

ANNA INFORMASJON

Les oppgåvene nøye og gjer dine egne meiningar. Presiser i svara kva for føresetnadar du har lagt til grunn i tolking/avgrensing av oppgåva.

Alle svar skal grunngivast og svaret skal innehalda naturleg mellomrekning slik at det er heilt klart kva reknereglar som blir nytta.

Dersom det er direkte feil eller manglar i oppgåvesettet og du ikkje kan gjere di eiga tolking, blir det vist til informasjon om klage på formelle feil på NTNUs nettside om «Grunngiving og klage».

FAGSPESIFIKK INFORMASJON

Papirhandteikningar:

Oppgåve 1 til 6 skal svarast direkte i Inspera. Oppgåvene 7a, 7b, 7c, 7d, 8a, 9a og 9b skal alle svarast på egne ark. Nedst i oppgåva finn du ein sjuisifra kode. Fyll inn denne koden øvst til venstre på dei arka du ynskjer å levere.

Det er tilrådd å gjere dette undervegs i eksamen. Dersom du treng tilgang til kodane etter at eksamenstida har gått ut, må du klikke «Sjå levering».

Du er sjølv ansvarleg for å fyller inn riktige kodar på eventuelle handteikningsark/ark for handteikningar. Les difor informasjonen om omslagsarket nøye. Eksamenskontoret kan ikkje garantere at feilaktig utfylte ark blir lagt til svara dine.

Vekting av oppgåvene:

I oppgåve 1 til 6 tel kvar oppgåve 5 poeng. I oppgåvene 7, 8 og 9 tel kvar deloppgåve (7a, 7b, 7c, 7d, 8a, 9a og 9b) 10 poeng.

Trekk frå/avbroten eksamen:

Dersom du ynskjer å levere blankt/avbryte eksamen, gå til «hamburgarmenyen» i øvre høgre hjørne og vel «Lever blankt». Dette kan ikkje angrast sjølv om prøven framleis er open.

Tilgang til svara dine:

Etter eksamen finn du svara dine under tidlegare prøver i Inspira. Merk at det kan ta ein vrykedag før eventuelle handteikningar vert tilgjengelege under «tidlegare prøvar».

1 1a)**Innleiing:**

La X og Y vere uavhengige normalfordelte stokastiske variablar. Det er kjent at X har forventa verdi 1 og varians 9 og at Y har forventa verdi 0 og varians 16.

Oppgåve:

Bestem dei følgjande sannsyna. Oppgje svara med minst to desimalar etter komma. Eventuell mellomrekning bør gjerast med fire desimalar etter komma.

$$P(X + Y > 1) = \boxed{}$$

$$P((X + Y)^2 \geq 1) = \boxed{}$$

Maks poeng: 5

2 2a)**Innleiing:**

La X og Y være to stokastiske variablar. Det er kjent at X har forventa verdi 1 og varians 4, at Y har forventa verdi -1 og varians 3, og at kovariansen mellom X og Y er -2.

Oppgåve:

Rekn ut dei følgjande storleikane. Oppgje svara med minst to desimalar etter komma. Eventuell mellomrekning bør gjerast med fire desimalar etter komma.

$$E[2X + Y] = \boxed{}$$

$$\text{Var}[2X + Y] = \boxed{}$$

Maks poeng: 5

3 3a)

Innleiing:

La X og Y være to uavhengige poissonfordelte stokastiske variablar. Det er kjent at X har forventa verdi 6 og at Y har forventa verdi 4.

Oppgåve:

Bestem dei følgjande sannsyna. Oppgje svara med minst to desimalar etter komma. Eventuell mellomrekning bør gjerast med fire desimalar etter komma.

$$P(X \leq 7) = \boxed{}$$

$$P(X + Y \leq 7) = \boxed{}$$

Maks poeng: 5

4 4a)

Innleiing:

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

Oppgåve:

Bestem dei følgjande sannsyna. Oppgje svara med minst to desimalar etter komma. Eventuell mellomrekning bør gjerast med fire desimalar etter komma.

$$P(X > 5) = \boxed{}$$

$$P(X > 10 | X > 5) = \boxed{}$$

Maks poeng: 5

5 5a)**Innleiing:**

La A og B vere hendingar i eit utfallsrom S . Sannsynet til hendingane er $P(A) = 0.2$ og $P(B) = 0.9$.

Oppgåve:

Vel i kvar del under det kulepunktet som angir det mest korrekte utsegnen om hendingane A og B .

Vel eitt alternativ

- Hendingane er disjunkte
- Vi har ikkje tilstrekkeleg informasjon til å avgjere om hendingane er disjunkte eller ikkje disjunkte
- Hendingane er ikkje disjunkte

Vel eitt alternativ

- Hendingane er avhengige
- Hendingane er uavhengige
- Vi har ikkje tilstrekkeleg informasjon til å avgjere om hendingane er uavhengige eller avhengige

Maks poeng: 5

6 6a)**Innleiing:**

Eit fotballag har 20 spelarar. Dei 20 spelarane kan delast inn i 3 målvakter, 7 forsvarsspelarar og 10 angrepsspelarar. Blant desse 20 spelarane skal det veljast ut eit startlag med 11 spelarar.

Oppgåve:

Bestem dei følgjande storleikane.

Talet på måtar å velje 11 spelarar der

a) 1 spelar er målvakt og 10 spelarar ikkje er målvakter:

b) 1 spelar er målvakt, 4 spelarar er forsvarsspelarar og 6 spelarar er angrepsspelarar:

Maks poeng: 5

Innleiing:

I denne oppgåva meiner vi med ei gammafordeling med parameterane $a > 0$ og $b > 0$, ein sannsynstettleik

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

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

Anta at vi har eit tilfeldig utval X_1, \dots, X_n frå ein gammafordeling med parametrar $a = 5$ og b kjent. Me skal i denne oppgåva sjå på tilnærma og eksakt konfidensintervall for b .

Me gjer eit utval av storleik $n = 10$ og observerer $\bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = 13.049$.

7 7a)

- Vis at sannsynsmaksimeringsestimatorens til b er gitt ved $\hat{B} = \frac{1}{5n} \sum_{i=1}^n X_i$.
- Rekn ut estimatet \hat{b} for verdiane i oppgåva.
- Vis at \hat{B} er forventningsrett for b , og at variansen er gitt ved $\text{Var}[\hat{B}] = \frac{b^2}{5n}$.

Maks poeng: 10

8 7b)

- Forklar kvifor vi kan sjå på \hat{B} som om lag normalfordelt.
- Gjer om nødvendig ytterligare tilnærmingar, og forklar desse i så fall.
- Rekn ut eit tilnærma 95% konfidensintervall for b .

Maks poeng: 10

9 7c)

I staden for å gjere ei tilnærming basert på normalfordelinga, er det mogleg å arbeide med den eksakte fordelinga for \hat{B} .

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

Maks poeng: 10

10 7d)

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

- Utlei eit eksakt 95% konfidensintervall for b med utgangspunkt i fordelinga til $\frac{10n\hat{B}}{b}$ frå 7c.
- Rekn ut den numerisk verdien av konfidensintervallet.
- Samanlikn dei numeriske verdiane av det eksakte konfidensintervallet og det tilnærma konfidensintervallet i oppgave 7b. Diskuter kor god tilnærminga var.

Maks poeng: 10

11 8a)

Innleiing:

La den stokastiske variabelen X ha ein sannsynstettleik gitt ved

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

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

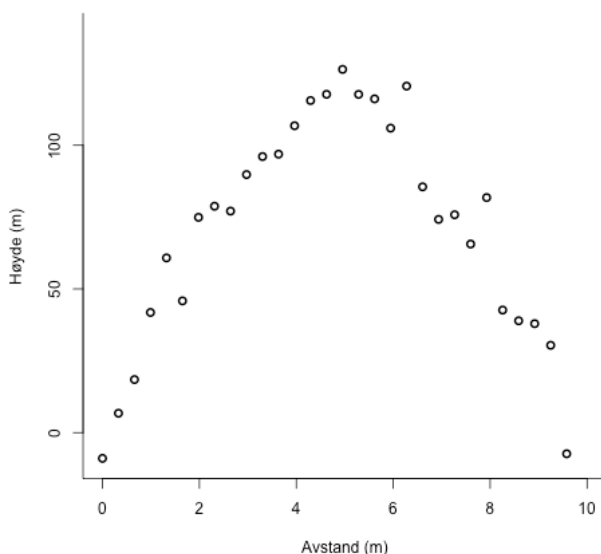
Oppgave:

- Bestem kumulativ fordelingsfunksjon til X .
- Vis at dersom $Y = (1 - U)^{-1/3}$ så har Y same fordeling som X .
- Bruk dette til å skrive ein Python-funksjon `simX` som genererer simuleringar av X basert på simuleringar av U . Funksjonen skal ta som input talet på simuleringar n , og gi som output n simuleringar av X .

Maks poeng: 10

Innleiing:

Vi er interesserte i rørsla til ei kule som blir skoten ut frå ein kanon. Denne rørsla blir skildra ved høgda til kula ved ulike avstandar frå kanonen. For å studere denne rørsla måler vi høgda til kula ved ulike avstandar. Vi samlar inn eit datasett som består av $n = 30$ målingar av høgde (i meter), y_1, \dots, y_n , med tilhøyrande avstander (i meter) høvesvis, x_1, \dots, x_n . Målingane er viste i figur 1 under.



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

Som figur 1 viser, er det stor usikkerheit i målingane, og vi skal vurdere to ulike lineære regresjonsmodellar for å skildre rørsla.

Modell 1:

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

der α_1 og β_1 er ukjende koeffisientar, og $\epsilon_1, \dots, \epsilon_n$ er uavhengige normalfordelte stokastiske variablar med forventa verdi 0 og ukjent varians $\sigma_1^2 > 0$.

Modell 2:

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

der $z_i = x_i - 4.9$, α_2 og β_2 er ukjende koeffisientar, og $\epsilon_1, \dots, \epsilon_n$ er uavhengige normalfordelte stokastiske variablar med forventa verdi 0 og ukjent varians $\sigma_2^2 > 0$.

Følgjande storleikar er oppgjevne for datasettet:

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

$$\cdot \sum_{i=1}^n (z_i^2 - \bar{z}^2)^2 = 1610.039.$$

12 9a)

I denne oppgåva treng de ikkje å utleie estimatorane. Bruk formelsamlinga.

- Rekn ut minste kvadratsumestimatane til α_1 , β_1 , α_2 og β_2 , og oppgje regresjonslinjene.
- Skisser målingane og begge regresjonslinjene i éin figur.
- Diskuter kva for føresetnader som er brotne i modell 1, og drøft om føresetnadene i modell 2 verkar å vere oppfylte.

Maks poeng: 10

13 9b)

I denne oppgåva brukar vi modell 2.

Produsenten av måleutstyret som er brukt for å måle verdiane y_1, \dots, y_n påstår at feilen i målingane skal ha eit standardavvik på 5 meter. Du meiner at usikkerheita må vere mykje større enn dette, og ønskjer å overtyde produsenten om at det er ein feil med utstyret. For å gjere dette ønskjer du å gjennomføre ein hypotesetest med signifikansnivå **0.05**.

I tillegg til opplysningane i starten av oppgåva er det oppgjeve at

$$\sum_{i=1}^n (y_i - \hat{\alpha}_2 - \hat{\beta}_2 z_i^2)^2 = 2615.707.$$

- Formuler ein passende nullhypotese og alternativ hypotese.
- Bestem ein forkastingsregel.
- Gjennomfør hypotesetesten og gje konklusjonen.

Maks poeng: 10