

i Eksamensoppgave i TMA4245 Statistikk**Dato:** 11. mai 2026**Tid:** 09.00-13.00**Faglig kontakt:** Håkon Tjelmeland, Geir-Arne Fuglstad**Møter i eksamenslokalet:** Nei**Hjelpemiddelkode/Tillatte hjelpemidler:** C

- Tabeller og formler i statistikk (Akademika).
- Bestemt enkel kalkulator i henhold til NTNUs generelle reglement.

Tabeller og formler i statistikk (Akademika).

Bestemt enkel kalkulator i henhold til NTNUs generelle reglement.

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.

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»

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

FAGSPESIFIKK INFORMASJON

Papirhåndtegninger: I oppgavene 7, 8 og 9 er det lagt opp til å besvare på ark. Andre oppgaver skal besvares direkte i Inspira. Nederst i oppgaven 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».

Merk at det kun er tillatt med blå eller svart kulepenn.

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 oppgave 7, 8 og 9 teller hver av deloppgavene (7, 8, 9a, 9b, 9c, 9d og 9e) 10 poeng.

Endring av målformer på eksamensoppgaven

Dersom eksamensoppgaven er skrevet på flere målformer, kan dette endres ved å gå i «hamburgermenyen» i øvre høyre hjørne og velg «Språk».

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 Inspira. Merk at det kan ta én virkedag før eventuelle papirhåndtegninger vil være tilgjengelige under «tidligere prøver».

- 1 **Innledning:** La X og Y være uavhengige og normalfordelte stokastiske variabler. Anta at $E[X] = 2$, $\text{Var}[X] = 9$ og $E[Y] = -1$, $\text{Var}[Y] = 16$.

Oppgave: Finn følgende sannsynligheter. Angi svarene med tre desimaler etter komma. Eventuell mellomregning bør gjøres med minst fem desimaler etter komma.

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

$$P(X - Y < 6): \boxed{}$$

Maks poeng: 5

- 2 Innledning:** La X og Y være kontinuerlige stokastiske variabler med simultan sannsynlighetstetthet gitt ved

$$f_{X,Y}(x,y) = \begin{cases} \frac{4}{5}(1+xy) & \text{hvis } x,y \in (0,1), \\ 0 & \text{ellers.} \end{cases}$$

Det oppgis at de tilhørende marginale sannsynlighetstetthetene for X og Y er henholdsvis

$$f_X(x) = \begin{cases} \frac{2}{5}(2+x) & \text{hvis } x \in (0,1), \\ 0 & \text{ellers,} \end{cases} \quad \text{og} \quad f_Y(y) = \begin{cases} \frac{2}{5}(2+y) & \text{hvis } y \in (0,1), \\ 0 & \text{ellers.} \end{cases}$$

Oppgave: Finn følgende sannsynligheter. Angi svarene som desimaltall med tre desimaler etter komma. Eventuell mellomregning bør gjøres med minst fem desimaler etter komma.

$$P(Y \leq \frac{1}{2}): \boxed{}$$

$$P(X \leq \frac{1}{2} | Y \leq \frac{1}{2}): \boxed{}$$

Maks poeng: 5

- 3 Innledning:** Anta at vi uten tilbakelegging trekker tilfeldig fem kort fra en vanlig korstokk bestående av 52 kort. Hvert kort har altså en verdi og en sort. Det er 13 verdier og fire sorter (spar, kløver, hjerter eller ruter).

Oppgave: Finn følgende antall. Oppgi svarene som heltall.

Hvor mange uordnede utvalg kan vi trekke fra kortstokken? $\boxed{}$

På hvor mange måter kan vi få tress (tre kort av samme verdi, og de to andre kortene av forskjellige andre verdier)? $\boxed{}$

Maks poeng: 5

- 4 **Innledning:** La X og Y være uavhengige og normalfordelte variabler. Anta at X har forventningsverdi lik μ og varians lik 4 mens Y har forventningsverdi lik 2μ og varians lik 9.

Anta at verdien til μ er ukjent og la $\hat{\mu}$ være sannsynlighetsmaksimeringsestimatoren (SME) for μ basert på de to stokastiske variablene X og Y .

Oppgave: Utled en formel for $\hat{\mu}$ og marker alternativet som er korrekt.

Velg ett alternativ:

- ☐ $\hat{\mu} = \frac{1}{4}(2X + Y)$
- ☐ $\hat{\mu} = \frac{1}{11}(3X + 4Y)$
- ☐ $\hat{\mu} = \frac{1}{22}(4X + 9Y)$
- ☐ $\hat{\mu} = \frac{1}{25}(9X + 8Y)$
- ☐ $\hat{\mu} = \frac{1}{2}(X + Y)$
- ☐ $\hat{\mu} = \frac{1}{26}(8X + 9Y)$
- ☐ $\hat{\mu} = \frac{1}{17}(9X + 4Y)$

Maks poeng: 5

- 5 Innledning:** La $X_1, X_2, \dots, X_n$ være et tilfeldig utvalg fra en normalpopulasjon med forventningsverdi lik μ og varians σ^2 . Anta at verdien til μ er ukjent, mens verdien til variansen σ^2 er kjent.

Anta at vi skal benytte observerte verdier for $X_1, X_2, \dots, X_n$ til å teste

$$H_0 : \mu = \mu_0 \text{ mot } H_1 : \mu < \mu_0,$$

der μ_0 er et angitt kjent tall. La som vanlig $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ og $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

Oppgave: Angi hvilke (en eller flere) av følgende størrelser som oppfyller kravene til å være en testobservator i denne situasjonen.

Velg ett eller flere alternativer

☐ $\frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$

☐ $\bar{X}$

☐ $\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$

☐ $\frac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}}$

☐ $\bar{X} - \mu$

☐ $\bar{X} - \mu_0$

☐ $\frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}}$

Maks poeng: 5

- 6 Innledning:** La $X_1, X_2, \dots, X_n$ være et tilfeldig utvalg fra en normalpopulasjon med forventningsverdi lik μ og varians lik σ^2 . Anta at verdien til μ er ukjent, mens verdien til variansen σ^2 er kjent.

Vi ønsker så å benytte observerte verdier for $X_1, X_2, \dots, X_n$ til å teste

$$H_0 : \mu = 2.5 \text{ mot } H_1 : \mu < 2.5.$$

Anta videre at vi benytter

$$Z = \frac{\bar{X} - 2.5}{\sqrt{\frac{\sigma^2}{n}}}$$

som testobservator, der vi som vanlig har at $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

Anta at vi velger å forkaste H_0 dersom $Z \leq -2.0$.

Oppgave: For $n = 30$ og $\sigma^2 = 0.25$, bestem følgende sannsynligheter. Angi svarene med tre desimaler etter komma. Eventuell mellomregning bør gjøres med minst fem desimaler etter komma.

Sannsynligheten for type **I**-feil:

Sannsynligheten for type **II**-feil når $\mu = 2.2$:

Maks poeng: 5

- 7 **Innledning:** Betrakt en vanlig modell for enkel lineær regresjon. Dvs. vi har uavhengige stokastiske variabler $Y_1, Y_2, \dots, Y_n$, der vi for $i = 1, 2, \dots, n$ har at

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \quad \text{der } \varepsilon_i \sim N(0, \sigma^2),$$

der α, β og $\sigma^2 > 0$ er ukjente parametre.

Som estimatorer for parametrene brukes

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \hat{\alpha} = \bar{Y} - \hat{\beta} \bar{x} \quad \text{og} \quad S^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2.$$

Oppgave: Forklar hva et residualplott er og hva det brukes til. Angi spesielt hva man plotter langs x-aksen og langs y-aksen i et residualplott. Benytt notasjonen introdusert over, eller introduser din egen notasjon, slik at du kan svare presist.

Gi minst to eksempler på hvordan et residualplott kan se ut dersom regresjonsmodellen **ikke** passer med de observerte verdiene $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Velg de to eksemplene slik at eksemplene illustrerer hvordan et residualplott kan se ut når ulike antagelser i regresjonsmodellen brytes.

Gi også et eksempel på hvordan et residualplott kan se ut dersom regresjonsmodellen passer med de observerte verdiene.

Maks poeng: 10

- 8 Innledning:** La $X_1, X_2, \dots, X_n$ være et tilfeldig utvalg fra en gammafordeling med parametre $\alpha = 4$ og β , der sannsynlighetstettheten er som gitt i "Tabeller og formler i statistikk". Verdien til parameteren β antar vi her er ukjent. Ved å argumentere med sentralgrenseteoremet kan man som kjent da utlede et tilnærmet 95%–konfidensintervall for β . For n stor nok får man at (merk at du ikke trenger å utlede dette)

$$P\left(\frac{\bar{X}}{4+3.92/\sqrt{n}} \leq \beta \leq \frac{\bar{X}}{4-3.92/\sqrt{n}}\right) \approx 0.95,$$

slik at et tilnærmet 95%–konfidensintervall for β er gitt ved

$$\left[\frac{\bar{x}}{4+3.92/\sqrt{n}}, \frac{\bar{x}}{4-3.92/\sqrt{n}} \right].$$

Oppgave: For en gitt verdi for β , forklar hvordan man ved hjelp av stokastisk simulering kan finne et forventningsrett estimat for den eksakte sannsynligheten for at den sanne verdien β blir inneholdt i konfidensintervallet.

Skriv en pythonfunksjon som beregner et slikt estimat for gitte verdier av β og n . **Merk:** Du skal altså først beskrive med ord hvordan du kan gå frem, og deretter implementere fremgangsmåten i python-kode.

Maks poeng: 10

- 9(a) Innledning:** La X betegne levetiden til en ny type maskin. Med levetid mener vi her tiden fra maskinen tas i bruk til den slutter å fungere som den skal. Det er velkjent at levetiden vil påvirkes av forholdene der maskinen benyttes, slik som for eksempel temperatur og eventuelt mengde støv i lufta. Vi antar her at disse faktorene kan måles som en *stress-faktor* $z > 0$, og av erfaring vet man at sannsynlighetstettheten til X er gitt ved

$$f_X(x) = \begin{cases} \frac{2z}{\theta} x \exp\left\{-\frac{zx^2}{\theta}\right\}, & x > 0, \\ 0, & \text{ellers,} \end{cases}$$

der $\theta > 0$ er en ukjent parameter, mens stressfaktoren z har en kjent verdi.

Oppgave: Finn kumulativ fordelingsfunksjon $F_X(x)$. Skisser funksjonene $f_X(x)$ og $F_X(x)$ når $\theta = 12$ og $z = 1.5$.

Finn sannsynligheten $P(X \leq 4 | X \geq 2)$ når $\theta = 12$ og $z = 1.5$.

Maks poeng: 10

9(b) Innledning: La den stokastiske variabelen Y være definert ved

$$Y = \frac{2zX^2}{\theta}.$$

Oppgave: Bevis at Y er kjikvadratfordelt med to frihetsgrader. Bruk så dette til å vise at

$$E[X^2] = \frac{\theta}{z} \quad \text{og} \quad \text{Var}[X^2] = \frac{\theta^2}{z^2}.$$

Maks poeng: 10

9(c) Innledning: For å undersøke kvaliteten på denne typen nye maskiner benytter man n maskiner. La $z_1, z_2, \dots, z_n$ betegne verdiene til stress-faktorene som disse maskinene benyttes under, og la $X_1, X_2, \dots, X_n$ være de tilhørende levetidene. Vi antar at $X_1, X_2, \dots, X_n$ er uavhengige stokastiske variabler.

Det er foreslått to estimatorer for å estimere verdien til θ basert på $X_1, X_2, \dots, X_n$,

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n z_i X_i^2 \quad \text{og} \quad \theta^* = \frac{\sum_{i=1}^n z_i^2 X_i^2}{\sum_{i=1}^n z_i}.$$

Oppgave: Hvilken av de to estimatorene vil du foretrekke? Husk å begrunne svaret.

Maks poeng: 10

9(d) Innledning: Uansett hvilken av estimatorene $\hat{\theta}$ og θ^* du i forrige punkt fant til å være den beste, skal du videre i oppgaven ta utgangspunkt i estimatoren $\hat{\theta}$.

Oppgave: Basert på estimatoren $\hat{\theta}$, bestem en pivotal som kan benyttes til å utlede et konfidensintervall for θ . Hvilken sannsynlighetsfordeling har denne pivotalen? **Hint:** Du kan benytte resultatet vi etablerte i b), samt kjente egenskaper for kjikvadratfordelingen.

Utlø et $(1 - \alpha) \cdot 100\%$ konfidensintervall for θ .

Maks poeng: 10

- 9(e) Innledning:** Vi ønsker nå å utlede også et prediksjonsintervall for en ny levetid X_0 for en maskin som skal operere under forhold gitt ved en stress-faktor $z = z_0$. For å utlede prediksjonsintervallet skal vi ta utgangspunkt i størrelsen

$$V = \frac{\sum_{i=1}^n Y_i}{Y_0},$$

der $Y_0 = 2z_0X_0^2/\theta$ og $Y_i = 2z_iX_i^2/\theta$ for $i = 1, 2, \dots, n$. For $q \in (0, 1)$ lar vi v_q betegne kvantilen i fordelingen til V slik at

$$P(V > v_q) = q.$$

Oppgave: Utled et $(1 - \alpha) \cdot 100\%$ prediksjonsintervall for X_0 ved å ta utgangspunkt i V . Merk at uttrykket du kommer frem til skal inneholde kvantiler i fordelingen til V .

Forklar hvordan du ved hjelp av stokastisk simulering kan estimere de kvantilene i fordelingen til V som inngår i uttrykket for prediksjonsintervallet.

Maks poeng: 10