

Sensurdokument for eksamen i TMA4240 Statistikk, 22. desember 2023

Dette sensurdokumentet beskriver hvordan eksamensbesvarelser blir vurdert. Dette dokumentet bør lese sammen med oppgaveteksten og løsningsskissen. Eksamenssettet består av 10 deloppgaver (1, 2, 3a, 3b, 3c, 3d, 4a, 4b, 4c, 4d), og eksamensbesvarelser blir vurdert ved å gi en poengsum for besvarelsene for hver deloppgave. For hver deloppgave kan det gis inntil 10 poeng. Disse poengene blir så til slutt omgjort til en bokstavkarakter ved å følge koverteringsregelen angitt i slutten av dette dokumentet.

I det følgende angis hvilke aspekter som vurderes ved sensur av de ulike deloppgavene, samt hvor mange poeng man får ved å mestre hvert av disse aspektene.

- **Oppgave 1:** Fem aspekter vurderes ved sensur av denne oppgaven.
 - 3 poeng: Slå opp i tabell for sannsynligheter i en standard normalfordeling.
 - 2 poeng: Standardisere for å få en standard normalfordelt variabel.
 - 1 poeng: Bruke komplementærsetningen korrekt.
 - 2 poeng: Regne ut forventningsverdien $E[13X - 2Y]$ og variansen $\text{Var}[13X - 2Y]$ ved hjelp av regneregler.
 - 2 poeng: Bruke at X og Y er uavhengige til å få $P(Y \geq 4|X \geq 2) = P(Y \geq 4)$ eller at $P(Y \geq 4|X \geq 2) = P(Y \geq 4) \cdot P(X \geq 2)$, eller på annen korrekt måte bruke uavhengigheten til å bestemme den siste sannsynligheten.
- **Oppgave 2:** Fire aspekter vurderes ved sensur av denne deloppgaven.
 - 2 poeng: Finne fra python-koden at $y = \varphi\sqrt{u}$.
 - 3 poeng: Fra $y = \varphi\sqrt{u}$ finne $F_Y(y)$ og $f_Y(y)$. Trekker 1 poeng dersom man ikke regner ut $f_Y(y)$.
 - 3 poeng: Regne ut formel for $F_X^{-1}(u)$ og vise at man forstår hvordan denne brukes til å simulere en realisasjon av X .
 - 2 poeng: Skrive python-kode for å simulere realisasjoner av X .
- **Oppgave 3a:** Fire aspekter vurderes ved sensur av denne deloppgaven.
 - 3 poeng: Finne (ubetinget) sannsynlighet fra $F_X(x)$.
 - 3 poeng: Uttrykke betinget sannsynlighet ved hjelp av ubetingede sannsynligheter.
 - 2 poeng: Derivere for å finne $f_X(x)$.
 - 2 poeng: Integrere for å finne $E[X^2]$. Det gis full uttelling ved å skrive opp integralet og korrekt løse dette ved delvis integrasjon eller ved å benytte gammafunksjonen. Det gis også full uttelling dersom man gjenkjenner den oppgitte fordelingen som en weibullfordeling og korrekt benytter formler for forventningsverdi og varians for denne til å finne $E[X^2]$. Gir 1 poeng for korrekt å skrive opp integralet (inkludert grenser) for forventningsverdien, forutsatt at integralet er relatert til situasjonen i oppgaven. Man får altså null poeng for bare å skrive opp den generiske formelen $E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx$.
- **Oppgave 3b:** Tre aspekter vurderes ved sensur av denne deloppgaven.
 - 2 poeng: Angi resultat i sentralgrenseteorem.
 - 4 poeng: Beskrive praktisk konsekvens av sentralgrenseteoremet.
 - 4 poeng: Anvende sentralgrenseteoremet i den aktuelle situasjonen. Man må fra besvarelsen demonstrere, på et eller annet vis, at man har forstått at det er X_i^2 man her tar gjennomsnittet av.

• **Oppgave 3c:** Fire aspekter vurderes ved sensur av denne deloppgaven.

- 3 poeng: Angi H_0 og H_1 . Null uttelling dersom hypotesene ikke angis ved hjelp av θ . Maksimalt ett poeng dersom man formulerer H_1 ved hjelp en likhet. Maksimalt to poeng dersom man har ulikheten i H_1 feil vei, men beskriver at H_1 skal være at levetiden er blitt større. Maksimalt ett poeng dersom man har H_1 feil vei uten å gi noen argumenter for dette. Hvis man har hypotesene korrekt formulert, ikke trekk poeng selv om man ikke begrunner hvilken vei ulikheten i H_1 skal gå.
- 2 poeng: Angi testobservator og tilnærmet fordeling for denne når H_0 er korrekt. Maksimalt 1 poeng dersom θ inngår i formelen for testobservatoren. Maksimalt 1 poeng dersom man angir en annen fordeling enn en normalfordeling. Merk at testobservatoren man angir ikke trenger å være standardisert.
- 3 poeng: Regne ut p-verdi. Minst 1 poeng bare for å uttrykke p-verdien korrekt som en sannsynlighet.
- 2 poeng: Diskutere p-verdien. Legg vekt på om det fremgår at vedkommende har forstått hva en p-verdi er. Kun 1 poeng dersom man kun angir korrekt konklusjon uten noen diskusjon om hvorfor.

• **Oppgave 3d:** Tre aspekter vurderes ved sensur av denne oppgaven.

- 5 poeng: Vise (i tekst, ligninger eller python-kode) at man vet hva den eksakte p-verdien er.
- 3 poeng: Formulere i ord eller python-kode hvordan den eksakte p-verdien kan estimeres.
- 2 poeng: Skrive korrekt python-kode.

• **Oppgave 4a:** Fire aspekter vurderes ved sensur av denne oppgaven.

- 2 poeng: Benytte regneregler for forventningsverdi.
- 3 poeng: Benytte regneregler for varians av en sum av uavhengige stokastiske variabler. Ikke trekk poeng selv om man ikke påpeker at man benytter at Y_i 'ene er uavhengige. Dersom løsningen inneholder så lite mellomregning at det ikke er mulig å vurdere om vedkommende har forstått regnereglene, skal man legge til grunn av regnereglene ikke er forstått.
- 2 poeng: Forenkle summeuttrykk korrekt. Svarene skal forenkles så mye det går an og dersom dette ikke gjøres skal man legge til grunn at regneregler for summeuttrykk ikke er forstått.
- 3 poeng: Vurdere og konkludere med hvilken estimator som er best. Kun et poeng dersom man kun sier at man skal velge den forventningsrette estimatoren med minst varians, uten å relatere dette på noe vis til den aktuelle oppgaven. Minst to poeng dersom man klart formulerer eller viser at man ønsker å bestemme hvilken av esimatorene $\tilde{\beta}$ eller $\tilde{\tilde{\beta}}$ som har minst varians.

• **Oppgave 4b:** Tre aspekter vurderes ved sensur av denne oppgaven.

- 3 poeng: Sette opp rimelighetsfunksjonen og log-rimelighetsfunksjonen og forenkle uttrykk så mye som det er naturlig å gjøre før derivasjon. Det kreves ikke at man begrunner produkt ved uavhengighet. Maksimalt et poeng dersom man kun setter opp rimelighetsfunksjonen.
- 2 poeng: Finne den partiellderiverte ved hensyn på β . Ved forenkling av uttrykket skal løsningen inneholde tilstrekkelig mellomregning til at det er klart hvilke regneregler for summer som anvendes, og hvis dette ikke er tilstede trekkes det et poeng. Denne forenklingen kan selvfølgelig gjøres i forbindelse med løsning av ligningssystemet, så sjekk eventuelt hva som gjøres der.
- 2 poeng: Finne den partiellderiverte ved hensyn på σ^2 . Det trekkes ikke selv om man deriverer med hensyn på σ , dersom denne derivasjonen gjøres korrekt.
- 3 poeng: Sette de partiellderiverte lik null og løse ligningssystemet. Maksimalt 2 poeng dersom man åpenbart kun løser hver ligning hver for seg og ender opp med et uttrykk for $\hat{\sigma}^2$ som inneholder den sanne parameterverdien β .

• **Oppgave 4c:** Fire aspekter vurderes ved sensur av denne oppgaven.

- 3 poeng: Begrunne at oppgitt uttrykk er χ_n^2 -fordelt. Det kreves at man i besvarelsen gjør de to trinnene adskilt slik at det klart fremgår hva som skjer, eventuelt at man tar utgangspunkt i at $\varepsilon_i = Y_i - \beta x_i$ er uavhengige og identisk normalfordelt og argumenterer med formel på side 27 i formelsamlinga. Dersom man tar utgangspunkt i formel på side 27 i formelsamlinga om at $\sum_{i=1}^n (X_i - \mu)^2 / \sigma^2 \sim \chi_n^2$ og benytter denne for $Y_i \sim N(\beta x_i, \sigma^2)$ gis det kun 1 poeng, siden forutsetningene for denne formelen da ikke er oppfylt i oppgaven.
- 2 poeng: Skrive at man mister en frihetsgrad når β erstattes med $\hat{\beta}$. Ingen uttelling for å argumentere med formelen $\sum_{i=1}^n (X_i - \bar{X})^2 / \sigma^2$ på side 27 i formelsamlinga, siden forutsetningene for denne formelen ikke er oppfylt i oppgaven.
- 3 poeng: Vise at $\hat{\sigma}^2$ er forventningsskjev. Her må det klart fremgå at man benytter at

$$E \left[\sum_{i=1}^n \left(\frac{Y_i - \hat{\beta} x_i}{\sigma} \right)^2 \right] = n - 1.$$

Man trekker ikke poeng selv om man ikke eksplisitt begrunner dette ved at forventningsverdien i en χ^2 -fordeling er lik antall frihetsgrader. Annen korrekt utregning av forventningsverdien godtas selvfølgelig også.

- 2 poeng: Foreslå forventningsrett estimator for σ^2 og vise at den er forventningsrett. Maksimalt 1 poeng dersom man ikke viser at den foreslåtte estimatoren faktisk er forventningsrett.

• **Oppgave 4d:** Tre aspekter vurderes ved sensur av denne deloppgaven.

- 4 poeng: Forklare hvordan residualplott konstrueres for den aktuelle regresjonsmodellen. Ut fra besvarelsen må det klart fremgå hva som er plottet langs hver av de to aksene. Maksimalt 3 poeng dersom residualet (kun) er formulert ved hjelp av β , og ikke $\hat{\beta}$. Maksimalt 1 poeng dersom man kun forklarer for en standard regresjonsmodell.
- 4 poeng: Diskutere residualplottet i figur 3. Ved sensur skal man ignorere det som eventuelt sies om vurdering av residualplott generelt, uansett om dette er korrekt eller feil. Sensur vurderes kun basert på hva som sies om det aktuelle residualplottet. Ingen uttelling hvis man bare sier at alt ser bra ut, uten noen begrunnelse for dette.
- 1 poeng: Foreslå ny modell. Her aksepteres alle modeller som er noenlunde rimelig ut fra hva men observerer i residualplottet dersom forslaget er fornuftig begrunnet.

En av deloppgavene (oppgave 3d) viste seg ved sensur å falle betydelig vanskeligere for svært mange studenter enn hva som var hensikten da oppgavesettet ble laget. For å ta hensyn til dette ved sensur, på en måte som var rettferdig også for studenter som fikk til oppgave 3d, ble det ved karakterfastsettingen sett bort fra (for hver enkelt student) den deloppgaven som var gitt færrest poeng.

Matematisk kan konvergeringsregelen fra poenger på hver deloppgave til bokstavkarakter beskrives på følgende måte. La $x_1, x_2, \dots, x_{10}$ betegne antall poeng gitt på hver av de 10 deloppgavene. En total poengsum for besvarelsen beregnes da ved formelen

$$z = \frac{10}{9} \left[\sum_{i=1}^n x_i - \min\{x_1, x_2, \dots, x_{10}\} \right]. \quad (1)$$

Denne totale poengsummen, z , konverteres så til bokstavkarakter ved å benytte følgende regel:

Poengintervall	Bokstavkarakter
(86, 100]	A
(73, 86]	B
(60, 73]	C
(48, 60]	D
(36, 48]	E
[0, 36]	F