ℹ Institutt for matematiske fag

Eksamensoppgave i TMA4240 Statistikk

Eksamensdato: Torsdag 9. desember

Eksamenstid (fra-til): 09.00-13.00

Hjelpemiddelkode/Tillatte hjelpemidler: Hjelpemiddelkode C.

- Tabeller og formler i statistikk (Fagbokforlaget).
- Et gult ark (A5 med stempel) med egne håndskrevne formler og notater.
- Bestemt, enkel kalkulator.

Faglig kontakt under eksamen: Håkon Tjelmeland / Sara Martino
Tlf.: 4822 1896 / 9940 3330

ANNEN INFORMASJON:

Skaff deg overblikk over oppgavesettet før du begynner på besvarelsen din.

Generell informasjon:

- I oppgave 1 skal du kun oppgi korrekt svar (altså uten begrunnelse) direkte i Inspera. Følg instruksene i oppgaveteksten angående antall desimaler i svaret, hvis ikke kan svaret bli vurdert som feil. I alle mellomregninger må du bruke minst to desimaler mer enn hva du skal ha i svaret.
- I oppgavene 2, 3 og 4 skal alle svar begrunnes og besvarelsen skal inneholde mellomregning slik at det er helt klart hvordan man har tenkt. Besvarelsen på disse oppgavene skal skrives for hånd på utleverte ark, se informasjon under InsperaScan lenger ned på denne siden.

Les oppgavene nøye, 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. Henvend deg til en eksamensvakt hvis du ønsker å kontakte faglærer. Noter gjerne spørsmålet ditt på forhånd.

InsperaScan: Besvarelsen av oppgave 2, 3 og 4 skal skrives for hånd på utleverte 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».

Vekting av oppgavene: Vekt ved sensur for hver deloppgave er angitt i oppgavesettet.

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. Du kan finne igjen varselet ved å klikke på bjella øverst til høyre.

Trekk fra/avbrutt eksamen: Blir du syk under eksamen, eller av andre grunner 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 i arkivet i Inspera. Merk at det kan ta én virkedag før eventuelle håndtegninger vil være tilgjengelige i arkivet.

1(a) **Innledning**: La X og Y være to uavhengige og normalfordelte stokastiske variabler, der X har forventningsverdi lik 2 og standardavvik lik 1, mens Y har forventningsverdi lik -1 og standardavvik lik 3.

Oppgave: Regn ut følgende størrelser. Angi svarene med tre desimaler etter komma.

- $P(X \geq 3) =$ [ ]
- $P(2X - Y \leq 0) =$ [ ]
- $P\left(X^2 \leq \frac{1}{4}\right) =$ [ ]

Maks poeng: 5

1(b) **Innledning**: Anta at det er 55 studenter som tar et emne ved NTNU. Av disse 55 studentene er det 25 gutter og 30 jenter. I dette emnet skal det nå oppnevnes en referansegruppe bestående av 5 studenter.

Oppgave: Finn følgende antall. Oppgi svarene som heltall.

- Hvor mange ulike referansegrupper kan man danne i dette emnet? [ ]
- Hvor mange ulike referansegrupper kan man danne i dette emnet dersom man krever at gruppa skal ha minst to medlemmer av hvert kjønn? [ ]

Maks poeng: 5

1(c) **Innledning**: La X være binomisk fordelt med $n = 25$ forsøk og sannsynlighet for suksess $p = 0.33.$ Som kjent kan en binomisk fordeling med god approksimasjon tilnærmes med en normalfordeling når $np \geq 5$ og $n(1-p) \geq 5,$ et vilkår som er oppfylt for fordeling til $X.$

Oppgave: Benytt normalapproksimasjon til å finne (tilnærmede) verdier for følgende sannsynligheter. Husk å benytte kontinuitetskorreksjon når du regner ut de tilnærmede verdiene. Angi svarene med tre desimaler etter komma.

- $P(X < 5) =$ ☐
- $P(X > 10) =$ ☐
- $P(X > 10 | X \geq 4) =$ ☐

Maks poeng: 5

1(d)

Innledning: La X og Y være to diskrete stokastiske variabler med simultanfordeling $P(X = x, Y = y)$ som gitt i tabellen under:

$x \setminus y$	0	1	2
0	0.22	0	0.31
1	0.32	0.15	0

Oppgave: Regn ut følgende størrelser. Angi svarene med tre desimaler etter komma.

- $P(X = 1) =$ ☐
- $P(Y = 0 | X = 0)$ = ☐
- $E(2XY) =$ ☐

Maks poeng: 5

2 **Innledning**: La X være en kontinuerlig fordelt stokastisk variabel. Anta videre at X har fordeling

$$f(x) = \begin{cases} \frac{2x}{10} e^{-x^2/10} & \text{for } x \geq 0, \\ 0 & \text{ellers.} \end{cases}$$

Oppgave: Finn følgende størrelser:

- $P(X \geq 2)$
- $P(X^2 - 4X \leq -1)$
- Medianen til X.

Det anbefales at du løser denne oppgaven med papir og blyant.

Format

Words: 0

Maks poeng: 10

3(a) **Innledning**: Levetiden til en elektronisk komponent T antas å være eksponensialfordelt med parameter λ. Sannsynlighetstettheten til T er dermed gitt som

$$f(t;\lambda) = \begin{cases} \frac{1}{\lambda}\exp(-\frac{t}{\lambda}) & \textbf{for } t \geq 0, \\ 0 & \textbf{ellers.} \end{cases}$$

La $T_1, T_2, \ldots, T_n$ være et tilfeldig utvalg fra denne fordelingen.

Oppgave:

- Vis først at $\frac{2T_i}{\lambda}$ er gammafordelt.
- Bruk så dette til å vise at $V = \frac{2\sum_{i=1}^{n} T_i}{\lambda}$ er χ^2-fordelt med $2n$ frihetsgrader.

Det anbefales at du løser denne oppgaven med papir og blyant.

Format

Words: 0

Maks poeng: 10

3(b) **Innledning**: Anta at 20 slike elektroniske komponenter er blitt testet og gjennomsnittet av de observerte levetidene er $\bar{t} = 10.5$.

Oppgave: Utled et 90% konfidensintervall for λ og bruk observert verdi for $\bar{t}$ til å regne ut intervallet numerisk.

Det anbefales at du løser denne oppgaven med papir og blyant.

Format

Words: 0

Maks poeng: 10

4(a) **Innledning**: Anta at du er blitt bedt om å undersøke antall ulykker på byggeplasser i Norge og har samlet inn informasjon om ulykker fra $n = 34$ byggeplasser. Anta at ulykker på ulike byggeplasser skjer uavhengig av hverandre og at ulykker på byggeplass nummer i skjer ifølge en poissonprosess med intensitet λm_i, der m_i angir antall ansatte på denne byggeplassen og λ er en parameter som antas felles for alle byggeplasser. La X_i være antall ulykker på byggeplass nummer i i løpet av $t = 1$ år. Dermed er X_i poissonfordelt med forventningsverdi lik λm_i og har punktsannsynlighet

$$f(x_i) = \frac{(\lambda m_i)^{x_i}}{x_i!} e^{-\lambda m_i}, x_i = 0, 1, 2, \ldots$$

Oppgave: For $\lambda = 0.020$ og $m_i = 50$, besvar følgende spørsmål.

- Hva er sannsynligheten for at det blir minst en ulykke på byggeplass nummer i?
- Gitt at det er minst en ulykke på byggeplass nummer i, hva er sannsynligheten for at det er mer enn en ulykke på denne byggeplassen?
- Gitt at det er minst en ulykke på byggeplass nummer i, hva er sannsynligheten for at det er maksimalt tre ulykker på denne byggeplassen?

Det anbefales at du løser denne oppgaven med papir og blyant.

Format

Words: 0

Maks poeng: 10

4(b) **Innledning:** I resten av oppgaven skal vi anta at verdien til parameteren λ er ukjent og at vi ønsker å estimere denne fra observerte verdier $x_1, x_2, \ldots, x_n$ av antall ulykker på hver byggeplass. Vi skal anta at antall ansatte på hver byggeplass, $m_1, m_2, \ldots, m_n$, er kjent.

Oppgave: Finn rimelighetsfunksjonen $L(\lambda)$ og benytt denne til utlede sannsynlighetsmaksimeringsestimatoren, $\widehat{\lambda}$, for λ. Vis spesielt at denne estimatoren kan skrives som

$$\widehat{\lambda} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} m_i}.$$

Vis videre at $\widehat{\lambda}$ er en forventningsrett estimator for λ og at

$$\mathrm{Var}[\widehat{\lambda}] = \frac{\lambda}{\sum_{i=1}^{n} m_i}.$$

Det anbefales at du løser denne oppgaven med papir og blyant.

Format B I U Σ

Words: 0

Maks poeng: 10

4(c) **Innledning**: For 15 år siden ble det gjennomført en grundig undersøkelse om antall ulykker på norske byggeplasser, og man fant da at verdien på parameteren λ var $0.020.$ Du ønsker nå å benytte observerte verdier for X_i-ene til å vurdere om det er grunnlag for å påstå at antall ulykker på norske byggeplasser er lavere i dag enn det var for 15 år siden.

Du kan i resten av denne oppgaven forutsette (uten bevis) at

$$\frac{\widehat{\lambda}-\mathrm{E}[\widehat{\lambda}]}{\sqrt{\mathrm{Var}[\widehat{\lambda}]}}$$

er tilnærmet standard normalfordelt.

Oppgave: Formuler hypotesene H_0 og H_1 for denne situasjonen, velg en testobservator, angi en beslutningsregel og bestem testens p-verdi (tilnærmet) når de observerte verdiene er gitt i tabellen under.

Basert på denne p-verdien, vil du si at det er grunnlag for å påstå at antall ulykker i dag er lavere enn den var for 15 år siden?

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
x_i	1	0	2	3	0	0	1	2	0	0	2	0	1	0	0	1	1
m_i	28.8	55.2	49.4	55.9	34.8	34.6	28.4	77.0	81.1	29.3	62.8	22.6	106.3	44.9	27.8	49.8	73.1
i	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
x_i	0	0	1	2	2	1	1	0	2	3	0	2	1	0	1	1	2
m_i	43.8	21.7	64.5	77.6	79.6	18.2	69.5	23.1	63.8	56.2	65.0	28.9	52.8	53.4	49.0	40.6	115.0

Det oppgis at disse verdiene gir $\sum_{i=1}^{n} x_i = 33$ og $\sum_{i=1}^{n} m_i = 1784.4.$

Det anbefales at du løser denne oppgaven med papir og blyant.

Format

Words: 0

Maks poeng: 10

4(d) **Innledning**: Anta nå at man har mulighet for å innhente informasjon om antall ulykker på de $n = 34$ byggeplassene over en lengre periode enn $t = 1$ år, og at du ønsker å benytte dette til å redusere sannsynligheten for type-II feil i hypotesetesten du formulerte i (d).

Du kan forutsette at man i hypotesetesten bruker signifikansnivå $\alpha = 0.05$, og at verdiene til $m_1, m_2, \ldots, m_n$ er konstant lik verdiene angitt i tabellen i oppgave (c).

Oppgave: Dersom den sanne verdien for λ i dag er 0.018, for hvor lang tid t må du innhente informasjon om antall ulykker for at sannsynligheten for å oppdage at H_0 er feil skal være minst 0.2. (Hint: Finn først uttrykk for sannsynlighetsmaksimeringsestimatoren for λ når X_i er antall ulykker ved byggeplass nummer i i løpet av t år.)

Det anbefales at du løser denne oppgaven med papir og blyant.

Format

Words: 0

Maks poeng: 10

4(e) **Innledning**: Som et alternativ til å benytte sannsynlighetsmaksimeringsprinsippet for å konstruere en estimator for λ er det foreslått å benytte en estimator på formen

$$\tilde{\lambda} = \sum_{i=1}^{n} b_i X_i,$$

der $b_1, b_2, \ldots, b_n$ er konstanter (som kan være funksjoner av $m_1, m_2, \ldots, m_n$).

Oppgave: Vis at estimatoren $\tilde{\lambda}$ blir best ved å la alle b_i-ene være like og gitt som

$$b_i = \frac{1}{\sum_{j=1}^{n} m_j}.$$

Det anbefales at du løser denne oppgaven med papir og blyant.

Format

Words: 0

Maks poeng: 10