i Institutt for matematiske fag

Eksamensoppgave i TMA4240 Statistikk

Eksamensdato: Torsdag 9. desember

Eksamenstid (fra-til): 15.00-19.00

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 uavhengige og normalfordelte stokastiske variabler. La X ha forventningsverdi lik -2 og standardavvik lik 2, og la Y ha forventningsverdi lik 1 og standardavvik lik 1.

Oppgave: Finn følgende sannsynligheter. Oppgi svarene med tre desimaler etter komma.

- $P(X > 0) =$ ☐
- $P(2X - Y \geq -1) =$ ☐
- $P(Y \geq X) =$ ☐

Maks poeng: 5

1(b) **Innledning**: Anta at vi har en urne med 15 røde og 20 blå kuler. Fra denne urna trekker vi uten tilbakelegging ut 6 kuler, og antar at trekningene gjøres slik at alle kuler har like stor sannsynlighet for å bli trukket ut. Vi er så interessert i hvilke 6 kuler vi har trukket ut.

Oppgave: Finn følgende antall. Oppgi svarene som heltall.

- Når vi regner hver kule som unik, på hvor mange måter kan man trekke ut de 6 kulene? ☐
- Når vi regner hver kule som unik, på hvor mange måter kan man trekke ut de 6 kulene slik at det er trukket ut minst 2 kuler av hver farge? ☐

Maks poeng: 5

1(c) **Innledning**: La $X_1, X_2, \ldots, X_{50}$ være et tilfeldig utvalg fra en sannsynlighetsfordeling med forventningsverdi $\mu = \mathrm{E}[X_i] = 3.5$ og varians $\sigma^2 = \mathrm{Var}[X_i] = 1.5$. La $\bar{X}$ betegne gjennomsnittet av X_i-ene, dvs

$\bar{X} = \frac{1}{50} \sum_{i=1}^{50} X_i.$

Oppgave: Bruk sentralgrenseteoremet til å finne tilnærmede verdier for følgende sannsynligheter. Oppgi svarene med tre desimaler etter komma.

- $P(\bar{X} \leq 4.0) =$
- $P\left(\sum_{i=1}^{50} X_i \leq 185\right) =$
- $P\left(\bar{X} \leq 4.0 \middle| \sum_{i=1}^{10} X_i = 48.2\right) =$

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
0	0.4	0.1
1	0.3	0
2	0	0.2

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

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

Maks poeng: 5

2 **Innledning**: La X være en kontinuerlig fordelt stokastisk variabel som er gammafordelt med parametre $\alpha = 2$ og $\beta = \frac{1}{2}$. Sannsynlighetstettheten til X er dermed gitt som

$$f(x) = \begin{cases} 4xe^{-2x} & \text{for } x \geq 0, \\ 0 & \text{ellers.} \end{cases}$$

Oppgave: Finn følgende størrelser.

- $P(X \leq 2)$
- $P(X^2 - 4X < -3)$
- $\mathrm{E}[\sqrt{X}]$

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

Format

Words: 0

Maks poeng: 10

3(a) **Innledning**: Det har brutt ut et virus i samfunnet. Myndighetene ønsker å finne andelen smittede i befolkningen på et tidspunkt ved å teste et tilfeldig utvalg av N personer.

La p være sannsynligheten for at en tilfeldig utvalgt person fra befolkningen er smittet. Anta at testen ikke er helt pålitelig, sannsynligheten for et "falskt negativt" resultat er 0.02. Dette betyr at sannsynligheten for at en syk person tester positiv er 0.98. Du kan anta at sannsynligheten for et "falskt positivt" resultat, det vil si at en frisk person tester positivt, er null.

La X være en stokastisk variabel som er lik antall av de N personene i utvalget som tester positivt.

Oppgave:

- Forklar hvorfor X er binomisk fordelt med parametre N og $0.98p$.
- Vis at $\hat{p} = \frac{X}{0.98N}$ er en forventningsrett estimator for p og regn ut standardavviket til denne estimatoren.

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

Format B I U x_2 x^2 I_x Ω Σ

Words: 0

Maks poeng: 10

3(b) **Innledning:** Myndighetene tester $N = 1000$ personer og for $x = 55$ av disse er testen positiv.

Oppgave: Utled et 95% konfidensinterval for p basert på disse observasjonene.

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

Format

Words: 0

Maks poeng: 10

4(a) **Innledning**: Levetiden (målt i antall måneder) til noen typer mekaniske komponenter har vist seg å følge en fordeling med sannsynlighetstetthet

$$f(x) = \begin{cases} \frac{1}{2\theta}\frac{1}{\sqrt{x}}\exp\left\{-\frac{\sqrt{x}}{\theta}\right\} & \text{for } x \geq 0, \\ 0 & \text{ellers,} \end{cases}$$

der $\theta > 0$ er en parameter som beskriver kvaliteten av denne typen komponenter. Tilhørende kumulativ fordelingsfunksjon er

$$F(x) = \begin{cases} 1 - \exp\left\{-\frac{\sqrt{x}}{\theta}\right\} & \text{for } x \geq 0, \\ 0 & \text{ellers.} \end{cases}$$

I hele denne oppgaven skal vi dessuten anta at levetiden til ulike komponenter er uavhengige av hverandre.

[Hint: Du kan i noen av punktene (a)-(e) i denne oppgaven få bruk for å benytte at

$$\int_0^\infty x^\alpha \exp\left\{-\frac{\sqrt{x}}{\theta}\right\}dx = 2\theta^{2\alpha+2}\Gamma(2\alpha+2) \text{ for } \alpha > -1, \theta > 0,$$

der $\Gamma(\alpha)$ er gammafunksjonen.]

Oppgave: For $\theta = 2.5,$ besvar følgende spørsmål.

- Hva er sannsynligheten for at en slik komponent fremdeles vil funksjonere etter 15 måneder?
- Hvis man vet at en slik komponent fremdeles funksjonerer etter 15 måneder, hva er sannsynligheten for at denne komponenten fremdeles vil funksjonere etter 20 måneder?
- Hva er forventet levetid for en slik komponent?

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

Format

Words: 0

Maks poeng: 10

4(b) **Innledning**: For å undersøke kvaliteten på en ny type av slike komponenter har man prøvet ut $n = 25$ av disse. La $X_1, X_2, \ldots, X_n$ være levetidene, slik at dette altså er et tilfeldig utvalg fra fordelingen spesifisert i (a). De observerte levetidene, $x_1, x_2, \ldots, x_n$, er:

1.01, 7.10, 9.41, 0.79, 40.06, 1.90, 2.77, 26.91, 4.04, 2.79, 6.58, 41.46, 5.68, 2.37, 5.08, 1.59, 1.00, 0.07, 1.44, 0.27, 118.22, 0.07, 5.21, 2.85, 101.87.

Det oppgis at dette gir $\sum_{i=1}^{n} \sqrt{x_i} = 70.493.$

Oppgave: Vis at sannsynlighetsmaksimeringsestimatoren for θ er

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \sqrt{X_i}.$$

Er $\hat{\theta}$ en forventningsrett estimator for θ? Begrunn svaret.

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

Format

Words: 0

Maks poeng: 10

4(c) **Innledning**: For $i = 1, 2, \ldots, n,$ definer

$$Z_i = \frac{2}{\theta}\sqrt{X_i}.$$

Oppgave: Vis ved hjelp av transformasjonsformelen at Z_i er χ^2-fordelt med 2 frihetsgrader.

Bruk så dette til å begrunne at

$$\frac{2n\hat{\theta}}{\theta}$$

er χ^2-fordelt med $2n$ frihetsgrader.

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

Format

Words: 0

Maks poeng: 10

4(d) **Innledning**: Man vurderer å ta den nye typen av komponenter i bruk i stedet for en gammel type som har vært lenge i bruk. For den gamle typen komponenter som har vært lenge i bruk vet man at verdien til θ er 2.6. Man ønsker derfor å undersøke om de observerte levetidene gitt i (b), som er for komponenter av den nye typen, gir grunnlag for å påstå at forventet levetid for den nye typen komponenter er større enn forventet levetid for den gamle typen komponenter.

For å besvare oppgaven kan du benytte tabellen på nest siste siden av oppgavesettet. Tabellen inneholder sannsynligheter i en χ^2-fordeling med 50 frihetsgrader.

Oppgave: Formuler situasjonen beskrevet over som et hypotesetestingsproblem. Formuler hypotesene H_0 og H_1, angi en testobservator og en tilhørende beslutningsregel, og bestem p-verdien så godt du kan ut fra tabellen på nest siste side i oppgavesettet.

Basert på denne p-verdien, vil du si at det er grunnlag for å påstå at forventet levetid for den nye typen komponenter er større enn forventet levetid for den gamle typen komponenter? Begrunn svaret.

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

Format

Σ

Words: 0

Maks poeng: 10

4(e) **Innledning**: Anta at du får anledning til å prøve ut flere komponenter av den nye typen, og at du ønsker å benytte dette til å redusere sannsynligheten for type II-feil i testen formulert i (d).

Du kan forutsette at man i hypotesetesten bruker signifikansnivå $\alpha = 0.05.$

For å besvare oppgaven kan du benytte tabellen på den siste siden av oppgavesettet. Tabellen inneholder kvantiler i en χ^2-fordeling for ulike antall frihetsgrader.

Oppgave: Dersom den sanne verdien for θ for den nye typen komponenter er 3.4, hvor mange levetider for komponenter av den nye typen må man observere for at sannsynligheten for å oppdage at H_0 er feil skal være minst 0.50?

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

Format

Words: 0

Maks poeng: 10

Sannsynligheter i χ^2-fordeling med 50 frihetsgrader

$P(X^2 \leq z)$ når $X^2 \sim \chi^2_{50}$

z	$P(X^2 \leq z)$
45.0	0.3262
45.5	0.3457
46.0	0.3654
46.5	0.3854
47.0	0.4055
47.5	0.4257
48.0	0.4460
48.5	0.4663
49.0	0.4865
49.5	0.5066
50.0	0.5266
50.5	0.5464
51.0	0.5659
51.5	0.5852
52.0	0.6041
52.5	0.6226
53.0	0.6408
53.5	0.6585
54.0	0.6758
54.5	0.6927
55.0	0.7090
55.5	0.7248
56.0	0.7401
56.5	0.7548
57.0	0.7691
57.5	0.7827
58.0	0.7958
58.5	0.8084
59.0	0.8204
59.5	0.8318
60.0	0.8428

i

Kvantiler i χ^2-fordelingen

$$P(X^2 > \chi^2_{\alpha,\nu}) = \alpha$$

$\nu\backslash\alpha$	**0.95**	**0.50**	**0.05**
60	43.188	59.335	79.082
61	44.038	60.335	80.232
62	44.889	61.335	81.381
63	45.741	62.335	82.529
64	46.595	63.335	83.675
65	47.450	64.335	84.821
66	48.305	65.335	85.965
67	49.162	66.335	87.108
68	50.020	67.335	88.250
69	50.879	68.334	89.391
70	51.739	69.334	90.531
71	52.600	70.334	91.670
72	53.462	71.334	92.808
73	54.325	72.334	93.945
74	55.189	73.334	95.081
75	56.054	74.334	96.217
76	56.920	75.334	97.351
77	57.786	76.334	98.484
78	58.654	77.334	99.617
79	59.522	78.334	100.749
80	60.391	79.334	101.879
81	61.261	80.334	103.010
82	62.132	81.334	104.139
83	63.004	82.334	105.267
84	63.876	83.334	106.395
85	64.749	84.334	107.522
86	65.623	85.334	108.648
87	66.498	86.334	109.773
88	67.373	87.334	110.898
89	68.249	88.334	112.022