NTNU

Kunnskap for en bedre verden

Institutt for matematiske fag

Eksamensoppgave i **TMA4240 Statistikk**

Faglig kontakt under eksamen: Sara Martino[a], Torstein Fjeldstad[b]

Tlf: [a]994 03 330 , [b]962 09 710

Eksamensdato: 28. november 2018

Eksamenstid (fra–til): 09:00–13:00

Hjelpemiddelkode/Tillatte hjelpemidler: Hjelpemiddelkode C:

– Tabeller og formler i statistikk, Akademika,
– Ett gult ark (A5 med stempel) med egne håndskrevne formler og notater,
– Bestemt, enkel kalkulator

Annen informasjon:

Alle svar må begrunnes.

Du må ha med nok mellomregninger til at tenkemåten din klart fremgår.

Oppgavesettet består av 4 oppgaver med tilsammen 10 delpunkter som har lik vekt ved sensur.

Målform/språk: bokmål

Antall sider: 5

Antall sider vedlegg: 0

Kontrollert av:

Informasjon om trykking av eksamensoppgave	
Originalen er:	
1-sidig ☐	**2-sidig** ☒
sort/hvit ☒	**farger** ☐
skal ha flervalgskjema ☐	

Dato Sign

Oppgave 1 La X være en stokastisk variabel med sannsynlighetstetthet

$$f(x) = \begin{cases} \frac{1+x}{2} & x \in (-1, 1) \\ 0 & \text{ellers} \end{cases}$$

Betrakt $Y = X^2$.

a) Finn den kumulative fordelingen til Y.

Finn sannsynlighetstettheten til Y.

Finn $E(2Y - Y^2)$.

Oppgave 2 Anta at antall biler som passerer et punkt på en spesifikk veistrekning inn i en kommune følger en poisson-prosses:

$$X(t) = \{\text{antall biler som passerer i løpet av } t \text{ minutter}\},$$

med parameter λ, det vil si $X(t) \sim \text{Poisson}(\lambda t)$.

a) Hvilke egenskaper må en poisson-prosess oppfylle?

Gi en kort tolkning av parameteren λ i situasjonen beskrevet over.

b) Anta (kun) i dette punktet at $\lambda = 1.5$.

Hva er sannsynligheten for at nøyaktig 2 biler passerer punktet i løpet av en periode av lengde 1 minutt?

Hva er sannsynligheten for at minst 2 biler passerer punktet i løpet av en periode av lengde 2 minutter?

Betrakt nå 10 ikke-overlappende perioder, hver av lengde 1 minutt. Hva er sannsynligheten for at det i minst en av disse periodene passerer flere enn 5 biler?

Anta nå at parameteren λ er ukjent. For å estimere λ har en representant fra kommunen besøkt den spesifikke veistrekningen n ganger og talt antall biler som passerer. For et gitt besøk står representanten der i t_i minutter og observerer at X_i biler passerer punktet, for $i = 1, \ldots n$. Anta at de stokastiske variablene $X_1, \ldots, X_n$ er uavhengige.

Det er oppgitt at sannsynlighetsmaksimeringsestimatoren for λ er

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

Kommunen ønsker å redusere trafikken på veistrekningen og bestemmer derfor at dersom det, i gjennomsnitt, passerer flere enn 1.5 biler per minutt vil kommunen sette opp en bomstasjon for å redusere trafikken. Kommunen vil altså utføre følgende hypotesetest

$$\mathrm{H}_0 : \lambda = \lambda_0 = 1.5 \qquad \text{mot} \qquad \mathrm{H}_1 : \lambda > 1.5$$

Representanten fra kommunen har besøkt veistrekningen $n = 10$ ganger og har hver gang talt biler i 10 minutter. Kall hvert av besøkene på 10 minutter et delforsøk. Resultatene i de 10 delforsøkene er:

Tabell 1: Antall observerte bilpasseringer i hvert delforsøk.

t_i	10	10	10	10	10	10	10	10	10	10
x_i	17	14	25	22	18	17	19	27	14	19

der $\sum_{i=1}^{10} x_i = 192$ og $\sum_{i=1}^{10} t_i = 100$.

c) Ta utgangspunkt i $\widehat{\lambda}$ og spesifiser en rimelig testobservator. Oppgi hvilken fordelingen denne har under nullhypotesen. Begrunn svaret ditt. (Hint: du kan bruke, uten bevis, at poissonfordelingen med parameter λ kan tilnærmest med en normalfordeling dersom $\lambda > 14$.)

Bruk denne testobservatoren til å utføre hypotesetesten ved signifikansnivå 1 %. Vil kommunen bygge en bomstasjon basert på resultatet av testen?

En alternativ måte å gjennomføre testen er ved å definere

$$Z = \{\text{antall delforsøk der representanten teller flere enn } \lambda_0 t \text{ biler}\},$$

der λ_0 er verdien til det forventede antall bilpasseringer per minutt en ønsker å teste og $t = 10$ minutt er lengden av hvert delforsøk. Kommunen bestemmer seg for å forkaste nullhypotesen dersom $Z \geq k$, for en bestemt konstant k.

d) Ta utgangspunkt i Z og finn den minste verdien av k slik at signifikansnivået til testen er $\alpha \leq 0.01$.

Utfør testen basert på Z og observasjonene i Tabell 1. Vil kommunen bygge en bomstasjon eller ikke?

Oppgave 3 Oscar er en aktiv idrettsutøver som har som mål å drive med idretten sin på høyeste mulige nivå. For å kunne prestere på et høyest mulig nivå har han bestemt seg for å ta et ulovlig dopingpreparat for å forbedre prestasjonsevnen sin. La x være dosen av det ulovlige preparatet Oscar tar. Anta at dosen x kan kontrolleres, og derfor ikke er stokastisk.

For å kontrollere konsentrasjonen av det ulovlige preparatet måler Oscar konsentrasjonen av preparatet en uke etter inntak, Y. Anta følgende lineære sammenheng mellom dosemengden x og observert konsentrasjon Y

$$Y = \beta x + \epsilon \tag{1}$$

der ϵ er normalfordelt med forventning 0 og kjent varians $\sigma^2 = 4^2$. Anta at β er en ukjent konstant som gir sammenhengen mellom en dose x og konsentrasjonen Y etter en uke.

a) Anta (kun) i dette punktet at $\beta = 0.5$.

En gitt dag tar Oscar en dose av størrelse $x = 30$, altså er Y normalfordelt med forventning $0.5 \cdot 30$ og varians 4^2 i dette punktet.

Finn sannsynligheten for at observert konsentrasjon av preparatet er høyere enn 20 en uke senere, det vil si finn $P(Y > 20)$.

Finn sannsynligheten $P(Y < 10 \cup Y > 20)$.

Gitt at konsentrasjonen av preparatet er høyere enn 10 etter en uke, finn sannsynligheten for at konsentrasjon av preparatet er høyere enn 20 etter en uke, det vil si finn $P(Y > 20 | Y > 10)$.

Tuva benytter også det samme ulovlige preparatet. Hun vil sammen med Oscar undersøke sammenhengen mellom inntatt dose og konsentrasjonen etter en uke, det vil si de vil estimere β.

Grunnet genetiske varisjoner har Tuva en naturlig referansekonsentrasjon c_0 av dopingpreparatet i kroppen, noe Oscar ikke har. Anta at Tuva og Oscar tar en identisk dose x. Konsentrasjonen av preparatet Tuva har en uke etter inntak, Z, av en dose x er

$$Z = c_0 + \beta x + \tau \tag{2}$$

der τ er normalfordelt med forventning 0 og kjent varians $\sigma^2 = 4^2$. Anta at c_0 er en kjent konstant som beskriver referansekonsentrasjonen av preparatet hos Tuva og at β er den samme ukjente konstanten som i (1).

I en tilfeldig valgt uke tar Oscar og Tuva en eksakt lik dose x av dopingpreparatet, og de måler konsentrasjonen av preparatet i kroppen etter en uke, henholdsvis Y og Z.

Anta at de begge bruker preparatet n ganger. For enkelhets skyld, anta at de tar preparatet kun når den forrige dosen har forsvunnet helt ut av kroppen. Oscar har altså et tilfeldig utvalg $(x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n)$ fra modellen definert i (1), og tilsvarende har Tuva et tilfeldig utvalg $(x_1, Z_1), (x_2, Z_2), \ldots, (x_n, Z_n)$ fra modellen definert i (2). Anta videre at de to utvalgene er uavhengig av hverandre.

b) Vis at sannsynlighetsmaksimeringsestimatoren basert på det tilfeldige utvalget $(x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n), (x_1, Z_1), (x_2, Z_2), \ldots, (x_n, Z_n)$ for β er

$$\widehat{\beta} = \frac{\sum_{i=1}^n x_i(Y_i + Z_i) - c_0 \sum_{i=1}^n x_i}{2\sum_{i=1}^n x_i^2}$$

Vis at $\widehat{\beta}$ er en forventningsrett estimator for β med varians

$$\mathrm{Var}(\widehat{\beta}) = \frac{\sigma^2}{2\sum_{i=1}^n x_i^2} = \frac{4^2}{2\sum_{i=1}^n x_i^2}$$

c) Utled et $(1-\alpha) \cdot 100$ % konfidensintervall for β.

Det er oppgitt at $n = 10$, $c_0 = 5$, $\sum_{i=1}^{10} x_i = 982$, $\sum_{i=1}^{10} x_i^2 = 97324$, $\sum_{i=1}^{10} x_i y_i = 68586$ og $\sum_{i=1}^{10} x_i z_i = 72398$.

Finn tallverdiene for et 90 % konfidensintervall for β.

Vi er nå interessert i å utføre følgende hypotesetest

$$\mathrm{H}_0 : \beta = 0.5 \qquad \text{mot} \qquad \mathrm{H}_1 : \beta \neq 0.5$$

Vil du forkaste nullhypotesen basert på observasjonene ved et signifikansnivå $\alpha = 0.1$? Begrunn svaret ditt.

Oppgave 4 Eva jobber ved et oppdrettsanlegg for laks og har som arbeidsoppgave å kontrollere fisken som blir sendt ut på markedet. Bedriften hennes antar at vekten til en tilfeldig valgt laks X (i kilogram) er normalfordelt med ukjent forventning μ kilogram og ukjent standardavvik σ kilogram.

a) Anta i dette punktet at Eva har fanget 10 lakser som hun har målt vekten til, $X_1, X_2, \ldots, X_{10}$. Anta at vekten til de 10 laksene er uavhengig av hverandre.

Bruk det tilfeldige utvalget $X_1, X_2, \ldots, X_{10}$ til å oppgi et uttrykk for et 95 % prediksjonsintervall for vekten i kilogram til en ny laks X_0, der X_0 er uavhengig av $X_1, X_2, \ldots, X_{10}$.

Bruk at $\sum_{i=1}^{10} x_i = 53.37$ kilogram og $\sqrt{\frac{1}{9}\sum_{i=1}^{10}(x_i - \bar{x})^2} = 0.73$ kilogram til å finne tallverdiene for prediksjonsintervallet.

Oppdrettsanlegget har i utgangspunktet gått ut i fra at forventet vekt er 5 kilogram, men på bakgrunn av tilbakemeldinger fra butikker som selger laks produsert ved Eva sitt oppdrettsanlegg har Eva nå en mistanke om at forventet vekt er høyere enn 5 kilogram. Eva ønsker derfor å teste hypotesen

$$\mathrm{H}_0 : \mu = 5 \text{ kilogram} \qquad \text{mot} \qquad \mathrm{H}_1 : \mu > 5 \text{ kilogram}$$

med et signifikansnivå $\alpha = 0.05$ basert på et tilfeldig utval $X_1, X_2, \ldots, X_n$.

b) Anta i dette punktet at standardavviket til X er kjent og lik 1 kilogram.

Anta at den sanne forventningsverdien til en tilfeldig valgt laks er 5.5 kilogram. Utled et uttrykk for det minste antall laks n Eva må veie dersom hun krever at testen sin styrke skal være minst 95 % når den sanne forventningsverdien er 5.5 kilogram.

Finn den minste tallverdien til n som oppfyller kravene spesifisert.