Institutt for matematiske fag

Eksamensoppgåve i **TMA4240 Statistikk**

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

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

Eksamensdato: 28. november 2018

Eksamenstid (frå–til): 09:00–13:00

Hjelpemiddelkode/Tillatne hjelpemiddel: Hjelpemiddelkode C:

– Tabeller og formler i statistikk, Akademika,
– Eitt gult ark (A5 med stempel) med eigne handskrivne formlar og notat,
– Bestemd, enkel kalkulator

Annan informasjon:

Alle svar må grungjevast.

Du må ha med nok mellomrekningar til at tenkjemåten din klart framgår.

Oppgåvesettet innheld 4 oppgåver med tilsaman 10 delpunkt som har lik vekt ved sensur.

Målform/språk: nynorsk

Sidetal: 5

Sidetal vedlegg: 0

Kontrollert av:

Informasjon om trykking av eksamensoppgåve	
Originalen er:	
1-sidig ☐	**2-sidig** ☒
svart/kvit ☒	**fargar** ☐
skal ha fleirvalskjema ☐	

Dato Sign

Oppgåve 1 La X vere ein stokastisk variabel med sannsynstettleik

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

Sjå på $Y = X^2$.

a) Finn den kumulative fordelinga til Y.

Finn sannsynstettleiken til Y.

Finn $E\left(2Y - Y^2\right)$.

Oppgåve 2 Anta at talet på bilar som passerer eit punkt på ein spesifikk vegstrekning inn i ein kommune følgjer ein poisson-prosses:

$$X(t) = \{\text{talet på bilar som passerer i løpet av } t \text{ minutt}\},$$

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

a) Kva for eigenskapar må ein poisson-prosess oppfylle?

Gje ei kort tolking av parameteren λ i situasjonen skildra over.

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

Kva er sannsynet for at nøyaktig 2 bilar passerer punktet i løpet av ein periode av lengde 1 minutt?

Kva er sannsynet for at minst 2 bilar passerer punktet i løpet av ein periode av lengde 2 minutt?

Sjå no på 10 ikkje-overlappande periodar, kvar av lengde 1 minutt. Kva er sannsynet for at det i minst ein av desse periodane passerer fleire enn 5 bilar?

Anta no at parameteren λ er ukjend. For å estimera λ har ein representant frå kommunen vitja den spesifikke vegstrekninga n gonger og tald talet på bilar som passerer. For eit gjeve besøk står representanten der i t_i minutt og observerer at X_i biler passerer punktet, for $i = 1, \ldots n$. Anta at dei stokastiske variablane $X_1, \ldots, X_n$ er uavhengige.

Det er gjeve at sannsynsmaksimeringsestimatoren for λ er

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

Kommunen ynskjer å redusera trafikken på vegstrekninga og har difor avgjort at dersom det, i gjennomsnitt, passerer fleire enn 1.5 bilar per minutt vil kommunen setje opp ein bomstasjon for å redusera trafikken. Kommunen vil altså utføre følgjande hypotesetest

$$H_0 : \lambda = \lambda_0 = 1.5 \qquad \text{mot} \qquad H_1 : \lambda > 1.5$$

Representanten frå kommunen har vitja vegstrekninga $n = 10$ gonger og har kvar gong tald bilar i 10 minutt. Kall kvart av besøka på 10 minutt eit delforsøk. Resultata i dei 10 delforsøka er:

Tabell 1: Tal på observerte bilpasseringar i kvart 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 utgongspunkt i $\widehat{\lambda}$ og spesifiser ein rimeleg testobservator. Oppgi kva fordeling denne har under nullhypotesen. Grunngje svaret ditt. (Hint: du kan nytte, utan bevis, at poissonfordelinga med parameter λ kan tilnærmast med ei normalfordeling dersom $\lambda > 14$.)

Nytt denne testobservatoren til å utføra hypotesetesten ved signifikansnivå 1 %. Vil kommunen byggje ein bomstasjon basert på resultatet av testen?

Ein alternativ måte å gjennomføre testen er ved å definere

$$Z = \{\text{talet på delforsøk der representanten tel fleire enn } \lambda_0 t \text{ bilar}\},$$

der λ_0 er verdien til det forventa tal på bilpasseringar per minutt ein ynskjer å teste og $t = 10$ minutt er lengda av kvart delforsøk. Kommunen bestem seg for å forkasta nullhypotesen dersom $Z \geq k$, for ein bestemt konstant k.

d) Ta utgongspunkt 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 observasjonane i Tabell 1. Vil kommunen byggje ein bomstasjon eller ikkje?

Oppgåve 3 Oscar er ein aktiv idrettsutøvar som har som mål å drive med idretten sin på høgast mogleg nivå. For å kunne prestere på eit høgast mogleg nivå har han bestemd seg for å ta eit ulovleg dopingpreparat for å betre prestasjonsevna si. La x vere dosen av det ulovlege preparatet Oscar tar. Anta at dosen x kan kontrollerast, og difor ikkje er stokastisk.

For å kontrollere konsentrasjonen av det ulovlege preparatet måler Oscar konsentrasjonen av preparatet ei veke etter inntak, Y. Anta følgjande lineære samanheng mellom dosemengda x og observert konsentrasjon Y

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

der ϵ er normalfordelt med forventning 0 og kjend varians $\sigma^2 = 4^2$. Anta at β er ein ukjend konstant som skildrar samanhengen mellom ein dose x og konsentrasjonen Y etter ei veke.

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

Ein gjeve dag tar Oscar ein dose av storleik $x = 30$, altså er Y normalfordelt med forventning $0.5 \cdot 30$ og varians 4^2 i dette punktet.

Finn sannsynet for at observert konsentrasjon av preparatet er høgare enn 20 ei veke seinare, det vil seie finn $P(Y > 20)$.

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

Gjeve at konsentrasjonen av preparatet er høgare enn 10 etter ei veke, finn sannsynet for at konsentrasjon av preparatet er høgare enn 20 etter ei veke, det vil seie finn $P(Y > 20 | Y > 10)$.

Tuva nyttar og det same ulovlege preparatet. Ho vil saman med Oscar undersøkje samanhengen mellom innteke dose og konsentrasjonen etter ei veke, det vil seie dei vil estimere β.

Grunna genetiske varisjoner har Tuva ein naturlig referansekonsentrasjon c_0 av dopingpreparatet i kroppen, noko Oscar ikkje har. Anta at Tuva og Oscar tar ein identisk dose x. Konsentrasjonen av preparatet Tuva har ei veke etter inntak, Z, av ein dose x er

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

der τ er normalfordelt med forventning 0 og kjend varians $\sigma^2 = 4^2$. Anta at c_0 er ein kjend konstant som beskriv referansekonsentrasjonen av preparatet hos Tuva og at β er den same ukjende konstanten som i (1).

I ei tilfeldig vald veke tar Oscar og Tuva ein eksakt lik dose x av dopingpreparatet, og dei målar konsentrasjonen av preparatet i kroppen etter ei veke, høvesvis Y og Z.

Anta at dei begge nyttar preparatet n gonger. For enkelheit skuld, anta at dei tar preparatet berre når den førre dosen har forsvunne heilt ut av kroppen. Oscar har altså eit tilfeldig utval $(x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n)$ frå modellen definert i (1), og tilsvarande har Tuva eit tilfeldig utval $(x_1, Z_1), (x_2, Z_2), \ldots, (x_n, Z_n)$ frå modellen definert i (2). Anta vidare at dei to utvala er uavhengig av kvarandre.

b) Vis at sannsynsmaksimeringsestimatoren basert på det tilfeldige utvalet $(x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n), (x_1, Z_1), (x_2, Z_2), \ldots, (x_n, Z_n)$ 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

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

c) Utlei eit $(1-\alpha) \cdot 100$ % konfidensintervall for β.

Det er gjeve 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 talverdiane for eit 90 % konfidensintervall for β.

Me er no interessert i å utføre følgjande hypotesetest

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

Vil du forkaste nullhypotesen basert på observasjonane ved eit signifikansnivå $\alpha = 0.1$? Grunngje svaret ditt.

Oppgåve 4 Eva jobbar ved eit oppdrettsanlegg for laks og har som arbeidsoppgåve å kontrollere fisken som blir send ut på marknaden. Bedrifta hennar antar at vekta til ein tilfeldig vald laks X (i kilogram) er normalfordelt med ukjend forventning μ kilogram og ukjend standardavvik σ kilogram.

a) Anta i dette punktet at Eva har fanga 10 laksar som ho har målt vekta til, $X_1, X_2, \ldots, X_{10}$. Anta at vekta til dei 10 laksane er uavhengig av kvarandre.

Nytt det tilfeldige utvalet $X_1, X_2, \ldots, X_{10}$ til å oppgje eit uttrykk for eit 95 % prediksjonsintervall for vekta i kilogram til ein ny laks X_0, der X_0 er uavhengig av $X_1, X_2, \ldots, X_{10}$.

Nytt 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 talverdiane for prediksjonsintervallet.

Oppdrettsanlegget har i utgangspunktet gått ut i frå at forventa vekt er 5 kilogram, men på bakgrunn av tilbakemeldingar frå butikkar som sel laks produsert ved Eva sitt oppdrettsanlegg har Eva no ein mistanke om at forventa vekt er høgare enn 5 kilogram. Eva ynskjer difor å testa hypotesen

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

med eit signifikansnivå $\alpha = 0.05$ basert på eit 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 ein tilfeldig vald laks er 5.5 kilogram. Utlei eit uttrykk for det minste talet på laks n Eva må vege dersom ho krev at testen sin styrke skal vere minst 95 % når den sanne forventningsverdien er 5.5 kilogram.

Finn den minste talverdien til n som oppfyller krava spesifisert.