Kunnskap for ei betre verd

Institutt for matematiske fag

Eksamensoppgåve i **TMA4240 Statistikk**

Fagleg kontakt under eksamen: Håkon Tjelmeland[a], Sara Martino[b]
Tlf: [a] 48 22 18 96 , [b] 99 40 33 30

Eksamensdato: 30. november 2017

Eksamenstid (frå–til): 09.00-13.00

Hjelpemiddelkode/Tillatne hjelpemiddel: C
Tabeller og formler i statistikk, Akademika,
Bestemt, enkel kalkulator,
Gult stempla A5-ark med eigne handskrivne notat.

Annan informasjon:
Du skal gje grunn for alle svara og oppgåvesvara skal innehalde naturlege mellomrekningar.

Målform/språk: nynorsk

Sidetal: 4

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

Gå ut frå at me kastar to vanlege terningar. La A vere hendinga at talet på auge på dei to terningane til saman er minst seks, og la X vere talet på auge.

Finn $P(A)$. Gje tolkinga av dette talet.

Finn $\mathrm{E}[X]$. Gje tolkinga av dette talet.

Oppgåve 2 Oppfylgingsavdelinga

To operasjonsavdelingar ved eit sjukehus i ein større by har ei felles oppfylgingsavdeling. Me skal her kalle dei to operasjonsavdelingane for høvesvis avdeling A og avdeling B. Pasientar som er utskrivne frå sjukehuset etter ein operasjon ved ei av dei to avdelingane kan, viss dei ynskjer det, oppsøkje oppfylgingsavdelinga for vidare rettleiing og kontrollar. Me skal gå ut frå at talet på gongar ein pasient som er operert ved avdeling A oppsøkjer oppfylgingsavdelinga er poissonfordelt med forventingsverdi $\mu_A = 1.4$, og tilsvarande at talet på gongar ein pasient som er operert ved avdeling B oppsøkjer oppfylgingsavdelinga er poissonfordelt med forventingsverdi $\mu_B = 0.81$.

Me skal også betrakte det som kjent at av alle pasientar som blir opererte ved dei to avdelingane blir 66% opererte ved avdeling A.

a) Rekn ut sannsynet for at ein pasient som er operert ved avdeling A har inkje besøk hos oppfylgingsavdelinga.

Gjeve at ein pasient som er operert ved avdeling A har minst eit besøk hos oppfylgingsavdelinga, rekn ut sannsynet for at denne pasienten har meir enn to besøk hos oppfylgingsavdelinga.

Rekn også ut sannsynet for at ein tilfeldig valt pasient har inkje besøk til oppfylgingsavdelinga.

Gå ut frå at avdeling A eit år opererer $n_A = 16\ 302$ pasientar, mens avdeling B same året opererer $n_B = 8\ 398$ pasientar. La X vere talet på kor mange av desse $n_A + n_B$ pasientane som har inkje besøk hos oppfylgingsavdelinga og la Y vere det totale talet på besøk desse pasientane har hos oppfylgingsavdelinga. Gå dessutan ut frå at talet på besøk ulike pasientar har hos oppfylgingsavdelinga er uavhengige av kvarandre.

b) Rekn ut forventingsverdi og standardavvik for X.

Kva type sannsynfordeling har Y? Husk å gje grunn for svaret.

Rekn ut forventingsverdi og standardavvik for Y.

Komponent i	1	2	3	4	5	6	7	8	9	10
z_i	1.0	1.5	4.0	1.0	2.0	2.5	3.0	2.0	2.0	4.5
y_i	707	217	326	292	477	285	234	243	204	260

Tabell 1: Stressfaktorar og tilhøyrande observerte levetider. Det gis at $\sum_{i=1}^{10}(y_i z_i)^2 = 6\,736\,616$.

Oppgåve 3 Levetid til ein ny type mekaniske komponentar

Levetida Y (målt i talet på døgn) til ein ny type mekaniske komponentar skal bli undersøkt. Det er kjent at levetidene avheng blant anna av temperatur og trykk der desse komponentane blir nytta. Effekten av desse forholda kan bli beskrive ved ein såkalla *stressfaktor*, z. Av erfaring frå liknande komponentar går ein ut frå at sannsynstettleiken til Y for ein gjeve stressfaktor z er gjeve ved

$$f(y) = \begin{cases} \frac{2yz^2}{\theta^2} \cdot e^{-\frac{y^2 z^2}{\theta^2}} & \text{for } y > 0, \\ 0 & \text{elles,} \end{cases}$$

der $\theta > 0$ er ein parameter som beskriv kvaliteten på komponentane. I heile denne oppgåva skal me gå ut frå at levetida til ulike mekaniske komponentar er uavhengige av kvarandre.

a) Gå i dette punktet ut frå at $\theta = 1000$ og $z = 1$.

Finn kumulativ fordelingsfunksjon for Y, $F(y)$.

Bestem sannsynet for at levetida Y er større enn 500 døgn.

Bestem medianen til Y.

For å undersøkje kvaliteten på den nye typen komponentar har ein målt levetidene til $n = 10$ slike komponentar. La $z_1, z_2, \ldots, z_n$ vere stressfaktorane som desse ble nytta under og la $Y_1, Y_2, \ldots, Y_n$ vere tilhøyrande levetider. Dei observerte verdiane er gjeve i Tabell 1.

b) Vis at sannsynsmaksimeringsestimatoren (SME) for θ er

$$\widehat{\theta} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i z_i)^2}.$$

Rekn ut estimatet.

For $i = 1, 2, \ldots, n$ la $U_i = \frac{2Y_i^2 z_i^2}{\theta^2}$.

c) Nytt transformasjonsformelen til å finne sannsynstettleiken til U_i og bruk dette til å vise at U_i er χ^2-fordelt med 2 fridomsgradar.

Bruk så dette til å grunngjeve at

$$\frac{2n\widehat{\theta}^2}{\theta^2} \sim \chi^2_{2n}.$$

d) Utlei frå resultatet i **c**) eit $(1-\alpha)\cdot 100\%$-konfidensintervall for θ.

Rekn ut intervallet numerisk for dataa i tabell 1 når $\alpha = 0.05$.

Oppgåve 4 Infeksjon etter operasjon

For eit par år sidan blei ein operasjonsavdeling omorganisert for å redusere risikoen for at pasientar får infeksjonar i forbindelse operasjonane.

La p_F vere sannsynet for at ein tilfeldig valt pasient fikk infeksjon før omorganiseringa og la p_E vere tilsvarande sannsyn etter omorganiseringa. For ein periode på eitt år før omorganiseringa, la n_F og X_F vere høvesvis talet på pasientar som blei opererte ved operasjonsavdelinga og kor mange av desse som fikk infeksjon. Tilsvarande, la n_E og X_E vere høvesvis talet på pasientar som blei opererte ved avdelinga i ein periode på eitt år etter omorganiseringa og kor mange av desse som fikk infeksjon. Me går ut frå at ulike pasientar får infeksjon uavhengig av kvarandre.

I denne oppgåva skal me nytte $\widehat{p}_F = \frac{X_F}{n_F}$ som estimator for p_F, og tilsvarande $\widehat{p}_E = \frac{X_E}{n_E}$ som estimator for p_E.

a) Formuler sentralgrenseteoremet.

Gå ut frå at både n_F og n_E er store. Grunngje då at $\widehat{p}_E - \widehat{p}_F$ er tilnærma normalfordelt. (*Hint: grunngje først at $\widehat{p}_E$ og $\widehat{p}_F$ er tilnærma normalfordelte*)

Vis at

$$\mathrm{E}\left[\widehat{p}_E - \widehat{p}_F\right] = p_E - p_F \quad \text{and} \quad \mathrm{Var}\left[\widehat{p}_E - \widehat{p}_F\right] = \frac{p_E(1-p_E)}{n_E} + \frac{p_F(1-p_F)}{n_F}.$$

	Talet på pasientar	Pasientar som fekk infeksjon
Før	2 021	186
Etter	1 919	135

Tabell 2: Talet på pasientar på operasjonsavdelinga i ein periode på eitt år før og i en periode på eitt år etter omorganiseringa, og kor mange av desse pasientane som fekk ein infeksjon i forbindelse med sin operasjon.

Me ynskjer nå å formulere ein hypotesetest for å teste om omorganiseringa av operasjonsavdelinga har vore vellykka.

b) Formuler nullhypotese og alternativ hypotese for denne situasjonen. Angje kva for ein testobservator du vil nytte og forklar kva for ein sannsynfordeling denne testobservatoren (tilnærma) har når nullhypotesen er riktig.

Rekn ut p-verdien til denne hypotesetesten når dataa er som gjeve i tabell 2. Diskuter om du vil konkludere med at omorganiseringa har vore vellykka eller ikkje.

Gå ut frå at operasjonsavdelinga bestemmer seg for å fortsetje med den nye organiseringa og at dei ynskjer å finne eit prediksjonsintervall for talet på dei av pasientane som blir opererte neste år som vil få ein infeksjon i forbindelse med sin operasjon.

c) Føresett at talet på pasientar som blir opererte neste år er $m = 2\ 000$, utlei formel for eit $(1-\alpha)\cdot 100\%$-prediksjonsintervall for kor mange av desse m pasientane som får ein infeksjon i forbindelse med sin operasjon.

Nytt verdiane i tabell 2 og $\alpha = 0.10$ til å rekne ut prediksjonsintervallet numerisk.