Institutt for matematiske fag

Eksamensoppgave i **TMA4240 Statistikk**

Faglig 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 (fra–til): 09.00-13.00

Hjelpemiddelkode/Tillatte hjelpemidler: C

Tabeller og formler i statistikk, Akademika,
Bestemt, enkel kalkulator,
Gult stemplet A5-ark med egne håndskrevne notater.

Annen informasjon:
Alle svar skal begrunnes og besvarelsen skal inneholde naturlig mellomregning.

Målform/språk: bokmål

Antall sider: 4

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

Anta at vi kaster to vanlige terninger. La A være hendelsen at antall øyne på de to terningene til sammen er minst seks, og la X betegne antall øyne.

Finn $P(A)$. Gi tolkningen av dette tallet.

Finn $\mathrm{E}[X]$. Gi tolkningen av dette tallet.

Oppgave 2 Oppfølgingsavdelingen

To operasjonsavdelinger ved et sykehus i en større by har en felles oppfølgingsavdeling. Vi skal her kalle de to operasjonsavdelingene for henholdsvis avdeling A og avdeling B. Pasienter som er utskrevet fra sykehuset etter en operasjon ved en av de to avdelingene kan, hvis de ønsker det, oppsøke oppfølgingsavdelingen for videre veiledning og kontroller. Vi skal anta at antall ganger en pasient som er operert ved avdeling A oppsøker oppfølgingsavdelingen er poissonfordelt med forventningsverdi $\mu_A = 1.4$, og tilsvarende at antall ganger en pasient som er operert ved avdeling B oppsøker oppfølgingsavdelingen er poissonfordelt med forventningsverdi $\mu_B = 0.81$.

Vi skal også betrakte det som kjent at av alle pasienter som opereres ved de to avdelingene blir 66% operert ved avdeling A.

a) Regn ut sannsynligheten for at en pasient som er operert ved avdeling A har ingen besøk hos oppfølgingsavdelingen.

Gitt at en pasient som er operert ved avdeling A har minst et besøk hos oppfølgingsavdelingen, regn ut sannsynligheten for at denne pasienten har mer enn to besøk hos oppfølgingsavdelingen.

Regn også ut sannsynligheten for at en tilfeldig valgt pasient har ingen besøk til oppfølgingsavdelingen.

Anta at avdeling A et år opererer $n_A = 16\ 302$ pasienter, mens avdeling B samme året opererer $n_B = 8\ 398$ pasienter. La X være antall av disse $n_A + n_B$ pasientene som har ingen besøk hos oppfølgingsavdelingen og la Y være det totale antall besøk disse pasientene har hos oppfølgingsavdelingen. Anta dessuten at antall besøk ulike pasienter har hos oppfølgingsavdelingen er uavhengige av hverandre.

b) Regn ut forventningsverdi og standardavvik for X.

Hva slags type sannsynlighetsfordeling har Y? Husk å grunngi svaret.

Regn ut forventningsverdi 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: Stressfaktorer og tilhørende observerte levetider. Det oppgis at $\sum_{i=1}^{10}(y_i z_i)^2 = 6\,736\,616$.

Oppgave 3 Levetid til en ny type mekaniske komponenter

Levetiden Y (målt i antall døgn) til en ny type mekaniske komponenter skal undersøkes. Det er kjent at levetidene er avhengig av blant annet temperatur og trykk der disse komponentene brukes. Effekten av disse forholdene beskrives ved en såkalt *stressfaktor*, z. Av erfaring fra lignende komponenter antar man at sannsynlighetstettheten til Y for en gitt stressfaktor z er gitt ved

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

der $\theta > 0$ er en parameter som beskriver kvaliteten på komponentene. I hele denne oppgaven skal vi anta at levetiden til ulike mekaniske komponenter er uavhengige av hverandre.

a) Anta i dette punktet at $\theta = 1000$ og $z = 1$.

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

Bestem sannsynligheten for at levetiden Y er større enn 500 døgn.

Bestem medianen til Y.

For å undersøke kvaliteten på den nye typen komponenter har man målt levetidene til $n = 10$ slike komponenter. La $z_1, z_2, \ldots, z_n$ betegne stressfaktorene som disse ble benyttet under og la $Y_1, Y_2, \ldots, Y_n$ være tilhørende levetider. De observerte verdiene er gitt i tabell 1.

b) Vis at sannsynlighetsmaksimeringsestimatoren (SME) for θ er

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

Regn ut estimatet.

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

c) Benytt transformasjonsformelen til å finne sannsynlighetstettheten til U_i og bruk dette til å vise at U_i er χ^2-fordelt med 2 frihetsgrader.

Bruk så dette til å begrunne at

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

d) Utled ved å ta utgangspunkt i resultatet i **c**) et $(1-\alpha)\cdot 100\%$-konfidensintervall for θ.

Regn ut intervallet numerisk for dataene i tabell 1 når $\alpha = 0.05$.

Oppgave 4 Infeksjon etter operasjon

For et par år siden ble en operasjonsavdeling omorganisert for å redusere risikoen for at pasienter får infeksjoner i forbindelse med operasjonene.

La p_F være sannsynligheten for at en tilfeldig valgt pasient fikk infeksjon før omorganiseringen og la p_E være tilsvarende sannsynlighet etter omorganiseringen. For en periode på ett år før omorganiseringen, la n_F og X_F være henholdsvis antall pasienter som ble operert ved operasjonsavdelingen og antall av disse som fikk infeksjon. Tilsvarende, la n_E og X_E være henholdsvis antall pasienter som ble operert ved avdelingen i en periode på ett år etter omorganiseringen og antall av disse som fikk infeksjon. Vi antar at ulike pasienter får infeksjon uavhengig av hverandre.

I denne oppgaven skal vi bruke $\widehat{p}_F = \frac{X_F}{n_F}$ som estimator for p_F, og tilsvarende $\widehat{p}_E = \frac{X_E}{n_E}$ som estimator for p_E.

a) Formuler sentralgrenseteoremet.

Anta at både n_F og n_E er store. Begrunn da at $\widehat{p}_E - \widehat{p}_F$ er tilnærmet normalfordelt. (*Hint: begrunn først at $\widehat{p}_E$ og $\widehat{p}_F$ er tilnærmet normalfordelt*)

Vis at

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

	Antall pasienter	Pasienter som fikk infeksjon
Før	2 021	186
Etter	1 919	135

Tabell 2: Antall pasienter på operasjonsavdelingen i en periode på ett år før og i en periode på ett år etter omorganiseringen, og antall av disse pasientene som fikk en infeksjon i forbindelse med sin operasjon.

Vi ønsker nå å formulere en hypotesetest for å teste om omorganiseringen av operasjonsavdelingen har vært vellykket.

b) Formuler nullhypotese og alternativ hypotese for denne situasjonen. Angi hvilken testobservator du vil benytte og forklar hvilken sannsynlighetsfordeling denne testobservatoren (tilnærmet) har når nullhypotesen er riktig.

Regn ut p-verdien til denne hypotesetesten når dataene er som angitt i tabell 2. Diskuter om du vil konkludere med at omorganiseringen har vært vellykket eller ikke.

Anta at operasjonsavdelingen bestemmer seg for å fortsette med den nye organiseringen og at de ønsker å finne et prediksjonsintervall for antall av pasientene som blir operert neste år som vil få en infeksjon i forbindelse med sin operasjon.

c) Forutsatt at antall pasienter som blir operert neste år er $m = 2\ 000$, utled formel for et $(1-\alpha)\cdot 100\%$-prediksjonsintervall for antall av disse m pasientene som får en infeksjon i forbindelse med sin operasjon.

Bruk verdiene i tabell 2 og $\alpha = 0.10$ til å beregne prediksjonsintervallet numerisk.