NTNU

Kunnskap for ei betre verd

Institutt for matematiske fag

Eksamensoppgåve i **TMA4245 Statistikk**

Fagleg kontakt under eksamen: Gunnar Taraldsen[a], Torstein Fjeldstad[b]

Tlf: [a]464 32 506, [b]962 09 710

Eksamensdato: 23. mai 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 håndskrivne formlar og notat,
– Bestemd, enkel kalkulator

Annan informasjon:

Alle svar må grunngjevast.

Du må ha med nok mellomrekningar til at tenkemåten din kjem klart fram.

Oppgåva består av 10 delpunkt som har lik vekt ved sensur.

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

Anta at den stokastiske variabelen X har sannsynstettleik

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

a) Finn den kumulative fordelinga til X og skisser denne.

Finn $P(X \geq 0.5)$ og $P(X \leq 0.7 | X \geq 0.5)$.

b) Finn sannsynstettleiken til $Y = -\ln(X)$.

Kva kjend fordeling har Y? Finn eit uttrykk for $\mu = \mathrm{E}(Y)$.

Gje ei kort tolkning av forventingsverdi.

Oppgåve 2 Ulovlig nedlasting av filmar

Bedrifta KTK slår ned på ulovleg nedlasting av filmar gjort via nettverket tilgjengeleg for sine tilsette.

a) Anta at 20 % av alle mannlege tilsette i KTK har lasta ned minst ein film ulovleg på KTK sitt nettverk og at 17 % av alle kvinnelege tilsette i KTK har gjort det same. Du kan anta at kvinneandelen i KTK er 67 %.

Definer relevante hendingar.

Rekn ut sannsynet for at ein tilfeldig vald tilsett har lasta ned minst ein film.

Gjeve at ein tilfeldig vald tilsett har lasta ned minst ein film, finn sannsynet for at det er ei kvinneleg tilsett.

Anta at dei IT-ansvarlege sender ut ein e-post kvar gong dei blir merksame på eit brot på KTK sine IT-retningsliner. La X vere talet på e-postar dei IT-ansvarlege sender ut i eit tilfeldig vald år og anta at denne er poissonfordelt med forventingsverdi μ. Du kan anta at talet utsende e-postar kvart år er uavhengig av talet på e-postar send dei andre åra. I tidsperioden 2009 - 2016 var den sanne forventingsverdien $\mu = 18$.

I 2017 blei det send ut 13 e-postar. Dei IT-ansvarlige vil teste om det er grunnlag for å hevda at forventa tal på e-postar har gått ned i 2017. Det vil seie dei IT-ansvarlige vil utføre følgjande hypotesetest:

$$H_0 : \mu = 18 \qquad \text{mot} \qquad H_1 : \mu < 18.$$

b) Finn kritisk verdi c slik at $P(X \leq c | H_0 : \mu = 18) \leq 0.10$.

Kva er forkastingsområdet til testen definert over med signifikansnivå 10 %?

Vil du behalde eller forkaste nullhypotesen basert på den observerte verdien?

c) Utlei eit uttrykk for sannsynet for type-II feil når forventa tal på e-postar send ut i 2017 er $\mu = \mu_1$, der $\mu_1 < 18$.

Finn sannsynet for type-II feil når $\mu_1 = 14$.

Oppgåve 3 Vekst av soyabønneplanter

Me ynskjer å undersøkja veksthastigheita hos soyabønneplantar som veks under visse ytre føresetnadar. Eit større tal på frø såast samstundes. Ei veke etter at den første planta er byrja å spira målast høgda til ei tilfeldig vald plante. Deretter målar me høgda til ei tilfeldig vald plante kvar veke på det same tidspunktet. Følgjande tabell gjev dei målte høgdene $h_1, h_2 \ldots, h_{11}$ i cm og tilhøyrande vekenummer $t_1, t_2, \ldots, t_{11}$:

t (veke)	1	2	3	4	5	6	7	8	9	10	11
h (cm)	2	4	11	16	19	25	27	32	34	39	42

Anta at høgdene $H_1, H_2, \ldots, H_{11}$ blir trekte uavhengig frå ei normalfordeling med forventingsverdi $\mu_i = a + b(t_i - 6)$ for $i = 1, 2, \ldots, 11$, der a og b er ukjende, og ukjend standardavvik σ.

a) Plott dei observerte høgdene mot vekenumra.

Diskuter om det er rimeleg med ein enkel lineær regresjonsmodell for dei observerte dataa.

Det er gjeve at

$$\sum_{i=1}^{11} h_i = 251, \quad \sum_{i=1}^{11} h_i^2 = 7577, \quad \sum_{i=1}^{11} (t_i - 6) = 0,$$

$$\sum_{i=1}^{11} (t_i - 6)^2 = 110 \quad \text{og} \quad \sum_{i=1}^{11} (t_i - 6) h_i = 449.$$

b) Vis at minste kvadraters metode (method of least squares) estimatorane for a og b, når $\sum_{i=1}^{11}(t_i - 6) = 0$, er gjeve ved:

$$\hat{a} = \frac{1}{11}\sum_{i=1}^{11} H_i \qquad , \qquad \hat{b} = \frac{\sum_{i=1}^{11}(t_i - 6)\,H_i}{\sum_{i=1}^{11}(t_i - 6)^2}.$$

Nytt dei gjevne talverdiane til å finna talverdiar for $\hat{a}$ og $\hat{b}$.

Skisser den tilpassa lina i figuren du laga i **a)**. Diskuter resultatet kort.

c) Skriv opp rimeligheitsfunksjonen (likelihood function) for situasjonen skildra over.

Vis at sannsynsmaksimeringsestimatorane for a og b er identiske med minste kvadraters metode estimatorane.

I resten av oppgåva kan du nytte, utan bevis, at

$$S^2 = \frac{1}{11-2}\sum_{i=1}^{11}\Big(H_i - \hat{a} - \hat{b}(t_i - 6)\Big)^2$$

er ein forventingsrett estimator for σ^2, og at

$$V = \frac{(11-2)\cdot S^2}{\sigma^2}$$

er khikvadratfordelt med $11 - 2 = 9$ fridomsgrader. Dessuten er $(11-2)\cdot S^2/\sigma^2$ uavhengig av $\hat{a}$ og $\hat{b}$. Det er gjeve at innsett dei observerte verdiane blir estimert varians $s^2 = 1.88$.

d) Utlei uttrykk for forventingsverdi og varians til $\hat{b}$.

Utlei eit 95 % konfidensintervall for veksthastigheit per veke. Rekn ut konfidensintervallet numerisk.

Gje ei kort tolking av eit konfidensintervall.

Oppgåve 4 Streik?

I ei fagforening som består av 100 medlemmar utføres ei prøveavstemning bland 20 tilfeldig valde medlem i spørsmålet om streik. Det viser seg at 3 er for streik, medan 17 er imot. La X vere talet på personar som stemde "for".

Kva er punktsannsynet til X? Grunngje svaret ditt.

Er det rimeleg å anta at fleirtalet er imot streik? Svaret skal grunngjevast ved å utføra ein hypotesetest ved utrekning av ein p-verdi.

Du kan nytte at

$$\sum_{x=0}^{3} \frac{\binom{50}{x}\binom{50}{20-x}}{\binom{100}{20}} \approx 0.0004.$$