

Institutt for matematiske fag

Eksamensoppgave i **TMA4245 Statistikk**

Faglig kontakt under eksamen: Gunnar Taraldsen^a, Torstein Fjeldstad^b

Tlf: ^a464 32 506, ^b962 09 710

Eksamensdato: 23. mai 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.

Oppgaven består av 10 delpunkter som har lik vekt ved sensur.

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 den stokastiske variabelen X har sannsynlighetstetthet

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

- a) Finn den kumulative fordelingen til X og skisser denne.

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

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

Hvilken kjent fordeling har Y ? Finn et uttrykk for $\mu = E(Y)$.

Gi en kort tolkning av forventningsverdi.

Oppgave 2 Ulovlig nedlasting av filmer

Bedriften KTK slår ned på ulovlig nedlasting av filmer gjort via nettverket tilgjengelig for sine ansatte.

- a) Anta at 20 % av alle mannlige ansatte i KTK har lastet ned minst en film ulovlig på KTK sitt nettverk og at 17 % av alle kvinnelige ansatte i KTK har gjort det samme. Du kan anta at kvinneandelen i KTK er 67 %.

Definer relevante hendelser.

Regn ut sannsynligheten for at en tilfeldig valgt ansatt har lastet ned minst en film.

Gitt at en tilfeldig valgt ansatt har lastet ned minst en film, finn sannsynligheten for at det er en kvinnelig ansatt.

Anta at de IT-ansvarlige sender ut en e-post hver gang de blir oppmerksomme på et brudd på KTK sine IT-retningslinjer. La X være antallet e-poster de IT-ansvarlige sender ut i et tilfeldig valgt år og anta at denne er poissonfordelt med forventningsverdi μ . Du kan anta at antallet utsendte e-poster hvert år er uavhengig av antallet e-poster sendt de andre årene. I tidsperioden 2009 - 2016 var den sanne forventningsverdien $\mu = 18$.

I 2017 ble det sendt ut 13 e-poster. De IT-ansvarlige vil teste om det er grunnlag for å hevde at forventet antall e-poster har gått ned i 2017. Det vil si de IT-ansvarlige vil utføre følgende hypotesetest:

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

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

Hva er forkastningsområdet til testen definert over med signifikansnivå 10 %?

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

- c) Utled et uttrykk for sannsynligheten for type-II feil når forventet antall e-poster sendt ut i 2017 er $\mu = \mu_1$, hvor $\mu_1 < 18$.

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

Oppgave 3 Vekst av soyabønneplanter

Vi ønsker å undersøke veksthastigheten hos soyabønneplanter som vokser under visse ytre betingelser. Et større antall frø sås samtidig. En uke etter at den første planten er begynt å spire måles høyden til en tilfeldig valgt plante. Deretter måler vi høyden til en tilfeldig valgt plante hver uke på det samme tidspunktet. Følgende tabell gir de målte høydene $h_1, h_2, \dots, h_{11}$ i cm og tilhørende ukenummer $t_1, t_2, \dots, t_{11}$:

t (uke)	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 høydene $H_1, H_2, \dots, H_{11}$ blir trukket uavhengig fra en normalfordeling med forventningsverdi $\mu_i = a + b(t_i - 6)$ for $i = 1, 2, \dots, 11$, der a og b er ukjente, og ukjent standardavvik σ .

- a) Plott de observerte høydene mot ukenumrene.

Diskuter om det er rimeleg med en enkel lineær regresjonsmodell for de observerte dataene.

Det er oppgitt 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) estimatorene for a og b , når $\sum_{i=1}^{11}(t_i - 6) = 0$, er gitt ved:

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

Benytt de oppgitte tallverdiene til å finne tallverdier for $\hat{a}$ og $\hat{b}$.

Skisser den tilpassede linjen i figuren du laget i **a)**. Diskuter resultatet kort.

- c)** Skriv opp rimelighetsfunksjonen (likelihood function) for situasjonen beskrevet over.

Vis at sannsynlighetsmaksimeringsestimatorene for a og b er identiske med minste kvadraters metode estimatorene.

I resten av oppgaven kan du bruke, uten bevis, at

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

er en forventningsrett estimator for σ^2 , og at

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

er khikvadratfordelt med $11 - 2 = 9$ frihetsgrader. Dessuten er $(11 - 2) \cdot S^2 / \sigma^2$ uavhengig av $\hat{a}$ og $\hat{b}$. Det oppgis at innsatt de observerte verdiene blir estimert varians $s^2 = 1.88$.

- d)** Utled uttrykk for forventningsverdi og varians til $\hat{b}$.

Utled et 95 % konfidensintervall for veksthastighet per uke. Regn ut konfidensintervallet numerisk.

Gi en kort tolkning av et konfidensintervall.

Oppgave 4 Streik?

I en fagforening som består av 100 medlemmer foretas en prøveavstemning blant 20 tilfeldig valgte medlemmer i spørsmålet om streik. Det viser seg at 3 er for streik, mens 17 er imot. La X være antallet personer som stemte "for".

Hva er punktsannsynligheten til X ? Grunngi svaret ditt.

Er det rimelig å anta at flertallet er imot streik? Svaret skal begrunnes ved å utføre en hypotesetest ved beregning av en p -verdi.

Du kan bruke at

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