NTNU
Kunnskap for en bedre verden

Institutt for matematiske fag

Eksamensoppgave i **TMA4245 Statistikk**

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

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

Eksamensdato: 7. juni 2019

Eksamenstid (fra–til): 09:00 – 13:00

Hjelpemiddelkode/Tillatte hjelpemidler: Hjelpemiddelkode C:

– Tabeller og formler i statistikk, Akademika,
– Et gult ark (A5 med stempel) med egne håndskrevne formler og notater,
– Bestemt, enkel kalkulator

Annen informasjon:

Alle svar må grunngis.

Du må ha med nok mellomregninger til at tenkemåten din kommer klart fram.

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

Målform/språk: bokmål

Antall sider: 6

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 Elektrisk sparkesykkel

Et sykehus som behandler pasienter med bruddskader, ønsker å studere sammenhengen mellom bruk av elektrisk sparkesykkel og bruddskader. De registrerer derfor om hver pasient med bruddskader fikk skaden ved bruk av elektrisk sparkesykkel. I 2018 behandlet sykehuset n pasienter med bruddskader. Anta videre at sannsynligheten for at en tilfeldig valgt pasient med bruddskade hadde vært utsatt for en ulykke ved bruk av elektrisk sparkesykkel er p.

La X være antall pasienter med bruddskader som var involvert i ei ulykke ved bruk av elektrisk sparkesykkel. Da er X binomisk fordelt med n forsøk og konstant suksessannsynlighet p.

a) Anta bare i dette punktet at $n = 17$ og $p = 0.2$.

Finn sannsynligheten for at nøyaktig 4 av pasientene med bruddskader var involvert i en ulykke med elektrisk sparkesykkel.

Finn sannsynligheten for at minst 4 av pasientene med bruddskader var involvert i en ulykke med elektrisk sparkesykkel.

Gitt at minst 4 av pasientene med bruddskader var involvert i en ulykke med elektrisk sparkesykkel, finn sannsynligheten for at minst 6 av pasientene med bruddskader var involvert i en ulykke med elektrisk sparkesykkel.

Anta videre at sykehuset i 2018 behandlet $n = 215$ pasienter med bruddskader, og at 54 av disse var involvert i en ulykke med elektrisk sparkesykkel.

b) Formuler sentralgrenseteoremet.

Foreslå en rimelig estimator $\widehat{p}$ for p og ta utgangspunkt i denne, og utled et uttrykk for et tilnærmet 95 % konfidensintervall for p.

Bruk verdiene gitt over til å finne tallverdier for intervallet.

Oppgave 2 Lading av elbil

Et borettslag med 17 husstander, der alle har elbil, tilbyr lading av elbil til beboerne sine. Anta at årlig strømforbruk $X_1, X_2, \ldots, X_{17}$ til hver av husstandene er uavhengige og normalfordelte med ukjent forventningsverdi μ kilowattimer og ukjent standardavvik σ kilowattimer. Borettslaget har tidligere gått ut fra at forventet strømforbruk til en tilfeldig valgt husstand er 3000 kilowattimer. De mistenker at det reelle strømforbruket er høyere, og har derfor leid inn et konsulentselskap til å undersøke dette.

Konsulentselskapet har samlet inn strømforbruket $x_1, x_2, \ldots, x_{17}$ fra 17 husstander som i gjennomsnitt brukte $\bar{x} = \frac{1}{17}\sum_{i=1}^{17} x_i = 3200$ kilowattimer i året med et utvalgsstandardavvik på $s = \sqrt{\frac{1}{16}\sum_{i=1}^{17}(x_i - \bar{x})^2} = 300$ kilowattimer til lading av elbil.

a) Formuler spørsmålet ovenfor som en hypotesetest.

Utfør hypotesetesten du har spesifisert med et signifikansnivå $\alpha = 0.05$. Oppgi spesielt sannsynlighetsfordeligen til testobservatoren du bruker.

Kan borettslaget basert på resultatet av hypotesetesten konkludere med at strømforbruket er høyere enn 3000 kilowattimer?

Oppgave 3 Avrenning

Årlig avrenning Y (millimeter per år) er et mål på hvor stor del av den årlige nedbøren (millimeter per år) innen et avgrenset område, ofte kalt nedbørfelt, som renner ut i tilsluttende vassdrag. Differansen mellom årlig nedbør og årlig avrenning antas å ha fordampet fra nedbørfeltet.

Anta følgende lineære sammenheng mellom årlig avrenning Y og årlig nedbør x innen et nedbørfelt

$$Y = \beta_0 + \beta_1 x + \varepsilon, \tag{1}$$

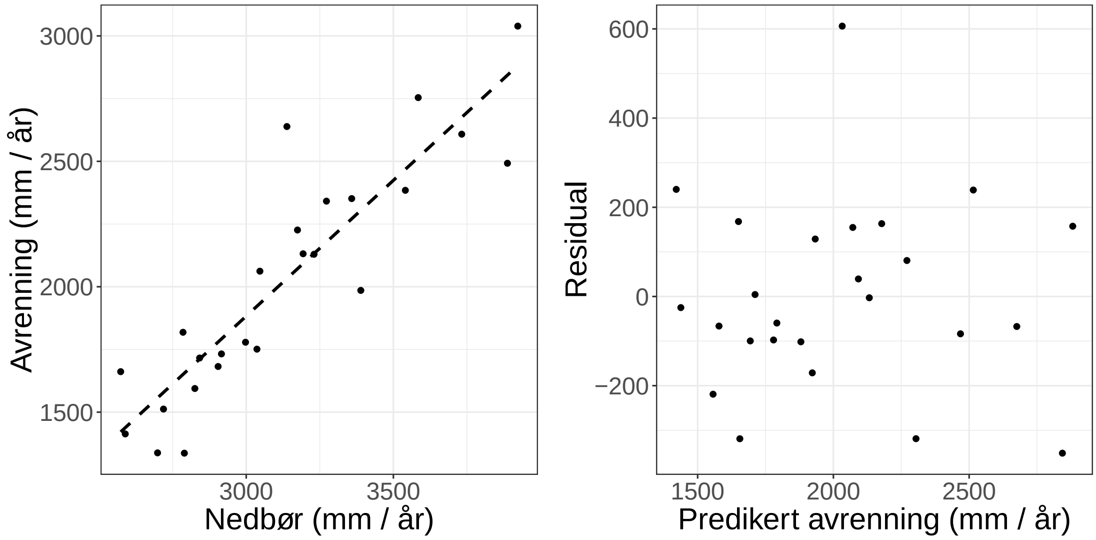
der β_0 og β_1 er ukjente konstanter og ε er normalfordelt med forventningsverdi 0 og ukjent varians σ^2.

Hydrologer har samlet inn uavhengige observasjoner fra det aktuelle nedbørfeltet over en periode på 25 år, det vil si et tilfeldig utvalg $(x_1, Y_1), (x_2, Y_2), \ldots, (x_{25}, Y_{25})$ fra modellen definert i (1). Det er gitt at følgende er forventningsrette estimatorer for henholdsvis β_1, β_0 og σ^2:

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum_{i=1}^{25}(x_i - \bar{x})Y_i}{\sum_{i=1}^{25}(x_i - \bar{x})^2} \\
\widehat{\beta}_0 &= \bar{Y} - \widehat{\beta}_1 \bar{x} \\
S^2 &= \frac{1}{23}\sum_{i=1}^{25}\left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2 .
\end{aligned} \tag{2}$$

I Figur 1a er de observerte verdiene $(x_1, y_1), (x_2, y_2), \ldots, (x_{25}, y_{25})$ vist sammen med den tilpassede regresjonslina $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta} x_i$.

I Figur 1b er tilpasset avrenning $\widehat{y}_i$ plottet mot residualene $e_i = y_i - \widehat{y}_i$.



(a) Nedbør x_i mot avrenning y_i. (b) Tilpasset avrenning $\widehat{y}_i$ mot residualer e_i.

Figur 1: Observasjoner (x_i, y_i) og tilpasset regresjonslinje $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta} x_i$, og tilpasset avrenning $\widehat{y}_i$ mot residualer e_i.

a) Forklar kort hvordan minste kvadraters (eng.: least squares) metode kan brukes til å finne estimatorer for β_0 og β_1, og illustrer med en figur. Du skal ikke utlede uttrykkene for estimatorene.

Ta utgangspunkt i den tilpassede modellen vist i Figur 1.

Drøft kort om det er rimelig å bruke en lineær regresjonsmodell. Oppgi spesielt (kort) hva slags antakelser som må være oppfylt for å kunne bruke en lineær regresjonsmodell.

Anta at vi nå er interessert i å predikere fremtidig avrenning for et nytt år Y_0 gitt årlig nedbør $x = x_0$, fra modellen definert i (1), der (x_0, Y_0) er uavhengig av $(x_1, Y_1), (x_2, Y_2), \ldots, (x_{25}, Y_{25})$. Det er gitt at $\widehat{Y}_0 = \widehat{\beta}_0 + \widehat{\beta}_1 x_0$ er en rimelig punktestimator for forventet avrenning $\mu_{Y|x_0} = \beta_0 + \beta_1 x_0$ når årlig nedbør er x_0. Det er gitt at $\widehat{\beta}_0 = -1364$ og $\widehat{\beta}_1 = 1.08$.

Du kan videre i oppgaven bruke (uten bevis) at $\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}$, det vil si khikvadratfordelt med $n-2$ frihetsgrader. Du kan også bruke at $\bar{Y}$ og $\widehat{\beta}_1$, og $\widehat{Y}_0$ og $\frac{(n-2)S^2}{\sigma^2}$ er uavhengige stokastiske variabler.

b) Hva er estimert forventet avrenning for et år der det er observert $x = 2000$ millimeter nedbør?

Vis at

$$\mathrm{E}\left(\widehat{Y}_0 - Y_0\right) = 0$$

og

$$\mathrm{Var}\left(\widehat{Y}_0 - Y_0\right) = \sigma^2 \left(1 + \frac{1}{25} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{25}(x_i - \bar{x})^2}\right).$$

Bruk dette til å finne et uttrykk for et 95 % prediksjonsintervall for en ny observasjon Y_0 gitt $x = x_0$.

Oppgave 4 Vekt gullbarre

Thomas har arvet en gullbarre etter besteforeldrene sine som han ønsker å selge. Gullbarren er av rent gull og veier μ gram. Før Thomas selger gullbarren ønsker han å sikre seg at han selger den for en korrekt pris og bestemmer seg derfor for å måle vekten til gullbaren på to måter.

Først bruker Thomas sin egen kjøkkenvekt til å veie gullbaren. Anta at vekten målt på kjøkkenvekten X er en normalfordelt stokastisk variabel med forventningsverdi μ og standardavvik 1 gram. Deretter går han til en forhandler for å måle vekten profesjonelt. La Y være vekten målt hos forhandleren, og anta at Y er normalfordelt med forventningsverdi μ og standardavvik 0.5 gram. Anta at $\mathrm{Cov}(X, Y) = -0.2$.

For å estimere den sanne vekten μ gram til gullbarren vil Thomas sammenligne to ulike estimatorer

$$\widehat{\mu} = Y \qquad \text{og} \qquad \tilde{\mu} = \frac{1}{2}(X + Y).$$

a) Beskriv kort hva som kjennetegner en god estimator.

Hvilken av de to estimatorene $\widehat{\mu}$ og $\tilde{\mu}$ bør Thomas velge? Grunngi svaret ditt.

Oppgave 5 Momentgenererende funksjon

Anta at $X_1, X_2, \dots, X_n$ er et tilfeldig utvalg fra en gammafordeling med parametre α og β, det vil si med sannsynlighetstetthet

$$f_X(x;\alpha,\beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x > 0 \\ 0 & \text{ellers} \end{cases}$$

der $\alpha > 0$ er en kjent parameter og $\beta > 0$ er en ukjent parameter. Du kan bruke (uten bevis) at den momentgenererende funksjonen til en gammafordelt stokastisk variabel X er

$$M_X(t) = \frac{1}{(1-\beta t)^\alpha} \qquad \text{for } t < \frac{1}{\beta}.$$

a) Vis ved bruk av momentgenererende funksjon at $Y = X_1 + X_2 + \cdots + X_n$ er gammafordelt med parametre $n\alpha$ og β, det vil si at Y har sannsynlighetstetthet

$$f_Y(y;n\alpha,\beta) = \begin{cases} \frac{1}{\beta^{n\alpha} \Gamma(n\alpha)} y^{n\alpha-1} e^{-y/\beta} & y > 0 \\ 0 & \text{ellers.} \end{cases} \tag{3}$$

Anta at vi har en observasjon y fra (3). Utled et uttrykk for sannsynlighetsmaksimeringsestimatoren for β basert på dette. Husk at α (og n) er kjent.

Oppgave 6 Sannsynlighet

La X og Y være to uavhengige normalfordelte stokastiske variabler med kjente forventningsverdier, henholdsvis, $\mu_X = 1$ og $\mu_Y = 0$, og kjent standardavvik $\sigma_X = \sigma_Y = 1$.

a) Finn $P(2X > 3)$.

Finn $P(2X > 3 \mid Y > 0)$.

Forklar kort hvorfor $X - Y$ også er normalfordelt og bruk dette til å finne $P(-1 \leq X - Y \leq 1)$.

Anta at vi har to uavhengige tilfeldige utvalg $X_1, X_2, \dots, X_{10}$ og $Y_1, Y_2, \dots, Y_{15}$ fra de to fordelingene spesifisert over.

b) Utled sannsynligheten for at høyst 5 av $X_1, X_2, \dots, X_{10}$ er mindre enn eller lik x, uttrykt ved den kumulative fordelingsfunksjonen til X.

Utled et uttrykk for $P(\max\{X_1, X_2, \dots, X_{10}, Y_1, Y_2, \dots, Y_{15}\} \leq z)$ uttrykt ved de kumulative fordelingsfunksjonene til X og Y.

Oppgave 7 Hypotesetest uniformfordelingen

La X_1 og X_2 være to uavhengige uniformt fordelte stokastiske variabler på intervallet $[\theta, \theta+1]$, det vil si de har sannsynlighetstetthet

$$f(x;\theta) = \begin{cases} 1 & x \in [\theta, \theta+1] \\ 0 & \text{ellers} \end{cases}$$

der θ er en ukjent konstant.

Følgende hypotese skal undersøkes

$$\mathrm{H}_0 : \theta = 0 \qquad \text{mot} \qquad \mathrm{H}_1 : \theta > 0.$$

To forkastingsregler er foreslått

$$\begin{aligned} &\text{Forkastningsregel 1}: && \text{Forkast } \mathrm{H}_0 \text{ dersom } X_1 > 0.95 \\ &\text{Forkastningsregel 2}: && \text{Forkast } \mathrm{H}_0 \text{ dersom } X_1 + X_2 > k \end{aligned}$$

der k er en ukjent kritisk verdi som skal bestemmes.

a) Finn sannsynligheten for type I-feil for forkastningsregel 1.

For forkastningsregel 1, finn et uttrykk for testens styrke som en funksjon av $\theta = \tau$, det vil si $P(\text{forkast } \mathrm{H}_0 \text{ når } \theta = \tau)$ for $\tau > 0$, og skisser grafen til denne funksjonen.

Vi krever nå at forkastingsregel 1 og 2 skal ha identisk sannsynlighet for type I-feil. Finn k.

Vink: Det kan være nyttig å skissere sannsynlighetstettheten til X_1 og simultantettheten til (X_1, X_2).