NTNU
Kunnskap for ei betre verd

Institutt for matematiske fag

Eksamensoppgåve i **TMA4245 Statistikk**

Fagleg kontakt under eksamen: Sara Martino[a], Torstein Fjeldstad[b]
Tlf: [a] 994 03 330, [b] 962 09 710

Eksamensdato: 7. juni 2019
Eksamenstid (frå–til): 09:00 – 13:00
Hjelpemiddelkode/Tillatne hjelpemiddel: Hjelpemiddelkode C:

– Tabeller og formler i statistikk, Akademika,
– Eit gult ark (A5 med stempel) med eigne handskrivne 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: 6
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 Elektrisk sparkesykkel

Eit sjukehus som behandlar pasientar med bruddskadar, ynskjer å studere samanhengen mellom bruk av elektrisk sparkesykkel og bruddskadar. Dei registrerer difor om kvar pasient med bruddskadar fekk skaden ved bruk av elektrisk sparkesykkel. I 2018 behandla sjukehuset n pasientar med bruddskadar. Anta vidare at sannsynet for at ein tilfeldig vald pasient med bruddskade var råka av ei ulukke under bruk av elektrisk sparkesykkel er p.

La X vere talet på pasientar med bruddskadar som var involvert i ei ulukke ved bruk av elektrisk sparkesykkel. Då er X binomisk fordelt med n forsøk og konstant suksessannsyn p.

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

Finn sannsynet for at nøyaktig 4 av pasientane med bruddskadar var involvert i ei ulukke med elektrisk sparkesykkel.

Finn sannsynet for at minst 4 av pasientane med bruddskadar var involvert i ei ulukke med elektrisk sparkesykkel.

Gjeve at minst 4 av pasientane med bruddskadar var involvert i ei ulukke med elektrisk sparkesykkel, finn sannsynet for at minst 6 av pasientane med bruddskadar var involvert i ei ulukke med elektrisk sparkesykkel.

Anta vidare at sjukehuset i 2018 behandla $n = 215$ pasientar med bruddskadar, og at 54 av desse var involvert i ei ulukke med elektrisk sparkesykkel.

b) Formuler sentralgrenseteoremet.

Foreslå ein rimeleg estimator $\widehat{p}$ for p og ta utgangspunkt i denne og utlei eit uttrykk for eit tilnærma 95 % konfidensintervall for p.

Nytt verdiane gjeve over til å finne talverdiar for intervallet.

Oppgåve 2 Lading av elbil

Eit burettslag med 17 husstandar, der alle har elbil, tilbyr lading av elbil til bebuarane sine. Anta at årleg straumforbruk $X_1, X_2, \ldots, X_{17}$ til kvar av husstandane er uavhengige og normalfordelte med ukjend forventningsverdi μ kilowattimar og ukjent standardavvik σ kilowattimar. Burettslaget har tidlegare gått ut ifrå at forventa straumforbruk til ein tilfeldig vald husstand er 3000 kilowattimar. Dei mistenkjer at det reelle straumforbruket er høgare, og har difor hyra inn eit konsulentselskap til å undersøkje dette.

Konsulentselskapet har samla inn straumforbruket $x_1, x_2, \ldots, x_{17}$ frå 17 husstandar som i gjennomsnitt nytta $\bar{x} = \frac{1}{17}\sum_{i=1}^{17} x_i = 3200$ kilowattimar i året med eit utvalsstandardavvik på $s = \sqrt{\frac{1}{16}\sum_{i=1}^{17}(x_i - \bar{x})^2} = 300$ kilowattimar til lading av elbil.

a) Formuler spørsmålet ovanfor som ein hypotesetest.

Utfør hypotesetesten du har spesifisert med eit signifikansnivå $\alpha = 0.05$. Oppgje særskild sannsynsfordelinga til testobservatoren du nyttar.

Kan burettslaget basert på resultatet av hypotesetesten konkludere med at straumforbruket er høgare enn 3000 kilowattimar?

Oppgåve 3 Avrenning

Årleg avrenning Y (millimeter per år) er eit mål på kor stor del av årleg nedbør (millimeter per år) innan eit avgrensa område, ofte kalla nedbørfelt, som renn ut i tilsluttande vassdrag. Differansen mellom årleg nedbør og årleg avrenning antas å ha fordampa frå nedbørfeltet.

Anta følgjande lineære samanheng mellom årleg avrenning Y og årleg nedbør x innan eit nedbørfelt

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

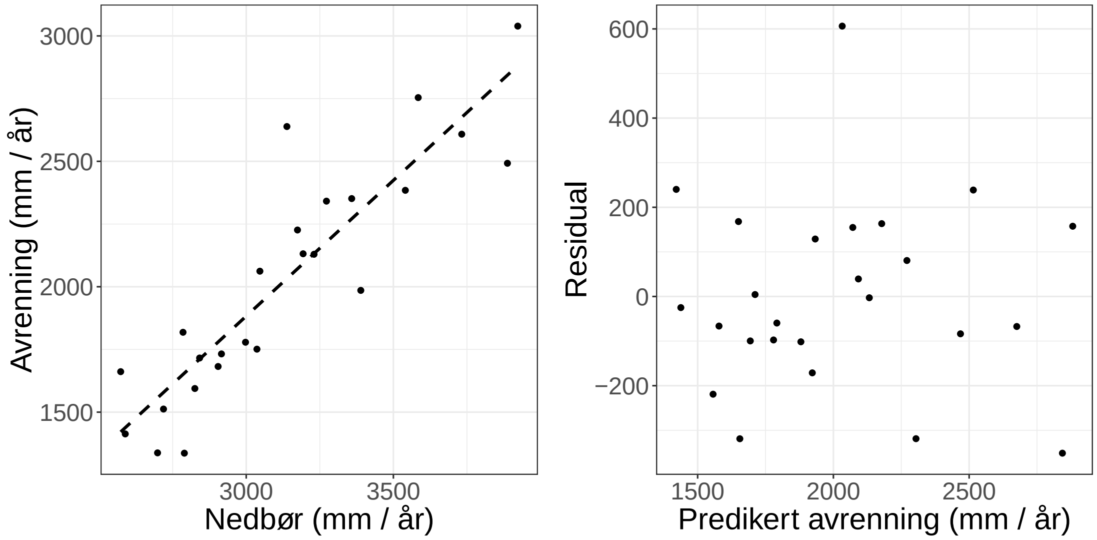
der β_0 og β_1 er ukjende konstantar og ε er normalfordelt med forventningsverdi 0 og ukjend varians σ^2.

Hydrologar har samla inn uavhengige observasjonar frå det aktuelle nedbørfeltet over ein periode på 25 år, det vil seie eit tilfeldig utval $(x_1, Y_1), (x_2, Y_2), \ldots, (x_{25}, Y_{25})$ frå modellen definert i (1). Det er gjeve at følgjande er forventningsrette estimatorar for høvevis 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 dei observerte verdiane $(x_1, y_1), (x_2, y_2), \ldots, (x_{25}, y_{25})$ viste saman med den tilpassa regresjonslina $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta} x_i$.

I Figur 1b er tilpassa avrenning $\widehat{y}_i$ plotta mot residuala $e_i = y_i - \widehat{y}_i$.



(a) Nedbør x_i mot avrenning y_i.

(b) Tilpassa avrenning $\widehat{y}_i$ mot residual e_i.

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

a) Forklar kort korleis minste kvadraters (eng.: least squares) metode kan nyttast til å finne estimatorar for β_0 og β_1, og illustrer med ein figur. Du skal ikkje utleie uttrykka for estimatorane.

Ta utgangspunkt i den tilpassa modellen vist i Figur 1.

Drøft kort om det er rimeleg å nytte ein lineær regresjonsmodell. Oppgje særskild (kort) kva antakingar som må vere oppfylt for å kunne nytte ein lineær regresjonsmodell.

Anta at me no er interessert i å predikere framtidig avrenning for eit nytt år Y_0 gjeve åreleg nedbør $x = x_0$, frå 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 gjeve at $\widehat{Y}_0 = \widehat{\beta}_0 + \widehat{\beta}_1 x_0$ er ein rimeleg punktestimator for forventa avrenning $\mu_{Y|x_0} = \beta_0 + \beta_1 x_0$ når årleg nedbør er x_0. Det er gjeve at $\widehat{\beta}_0 = -1364$ og $\widehat{\beta}_1 = 1.08$.

Du kan vidare i oppgåva nytte (utan bevis) at $\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}$, det vil seie khikvadratfordelt med $n-2$ fridomsgrader. Du kan og nytte at $\bar{Y}$ og $\widehat{\beta}_1$, og $\widehat{Y}_0$ og $\frac{(n-2)S^2}{\sigma^2}$ er uavhengige stokastiske variablar.

b) Kva er estimert forventa avrenning for eit 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 eit uttrykk for eit 95 % prediksjonsintervall for ein ny observasjon Y_0 gjeve $x = x_0$.

Oppgåve 4 Vekt gullbarre

Thomas har arva ein gullbarre etter besteforeldra sine som han ynskjer å selja. Gullbarren er av reint gull og veg μ gram. Før Thomas sel gullbarren ynskjer han å sikre seg at han sel den for ein korrekt pris og bestemmer seg difor for å måle vekta til gullbaren på to måtar.

Først nyttar Thomas si eiga kjøkkenvekt til å vege gullbaren. Anta at vekta målt på kjøkkenvekta X er ein normalfordelt stokastisk variabel med forventningsverdi μ og standardavvik 1 gram. Deretter går han til ein forhandlar for å måle vekta profesjonelt. La Y vere vekta målt hos forhandlaren, 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 vekta μ gram til gullbarren vil Thomas samanlikne to ulike estimatorar

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

a) Beskriv kort kva som kjenneteikner ein god estimator.

Kva for ein av dei to estimatorane $\widehat{\mu}$ og $\tilde{\mu}$ bør Thomas velge? Grunngje svaret ditt.

Oppgåve 5 Momentgenererande funksjon

Anta at $X_1, X_2, \ldots, X_n$ er eit tilfeldig utval frå gammafordelinga med parameter α og β, det vil seie med sannsynstettleik

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

der $\alpha > 0$ er ein kjend parameter og $\beta > 0$ er ein ukjend parameter. Du kan nytte (utan bevis) at den momentgenererande funksjonen til ein 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 momentgenererande funksjon at $Y = X_1 + X_2 + \cdots + X_n$ er gammafordelt med parametrar $n\alpha$ og β, det vil seie at Y har sannsynstettleik

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

Anta at me har ein observasjon y frå (3). Utlei eit uttrykk for sannsynsmaksimeringsestimatoren for β basert på dette. Hugs at α (og n) er kjend.

Oppgåve 6 Sannsyn

La X og Y vere to uavhengige normalfordelte stokastiske variable med kjende forventningsverdiar, høvevis, $\mu_X = 1$ 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 kvifor $X - Y$ og er normalfordelt og bruk dette til å finne $P(-1 \leq X - Y \leq 1)$.

Anta at me har to uavhengige tilfeldige utval $X_1, X_2, \ldots, X_{10}$ og $Y_1, Y_2, \ldots, Y_{15}$ frå dei to fordelingane spesifisert over.

b) Utlei sannsynet for at høgst 5 av $X_1, X_2, \ldots, X_{10}$ er mindre enn eller lik x, uttrykt ved den kumulative fordelingsfunksjonen til X.

Utlei eit uttrykk for $P(\max\{X_1, X_2, \ldots, X_{10}, Y_1, Y_2, \ldots, Y_{15}\} \leq z)$ uttrykt ved dei kumulative fordelingsfunksjonane til X og Y.

Oppgåve 7 Hypotesetest uniformfordelinga

La X_1 og X_2 vere to uavhengige uniformt fordelte stokastiske variable på intervallet $[\theta, \theta + 1]$, det vil seie dei har sannsynstettleik

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

der θ er ein ukjend konstant.

Følgjande hypotese skal undersøkjast

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

To forkastingsreglar 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 ein ukjend kritisk verdi som skal avgjerast.

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

For forkastningsregel 1, finn eit uttrykk for testen sin styrke som ein funksjon av $\theta = \tau$, det vil seie $P(\text{forkast H}_0 \text{ når } \theta = \tau)$ for $\tau > 0$, og skisser grafen til denne funksjonen.

Me krev no at forkastingsregel 1 og 2 skal ha identisk sannsyn for type I-feil. Finn k.

Vink: det kan vere nyttig å skissere sannsynstettleiken til X_1 og simultantettleiken til (X_1, X_2).