Institutt for matematiske fag

Eksamensoppgåve i **TMA4240/45 Statistikk**

Fagleg kontakt under eksamen: Øyvind Bakke

Tlf: 990 41 673

Eksamensdato:

Eksamenstid (frå–til):

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: 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

Merk! Studentane finn sensur i Studentweb. Har du spørsmål om sensuren må du kontakte instituttet ditt. Eksamenskontoret vil ikkje kunne svare på slike spørsmål.

Oppgåve 1 Sannsyn

La X vere ein stokastisk variabel med punktsannsyn $p(x) = P(X = x)$ som vist i tabellen:

x	-1	0	1
$p(x)$	0.3	0.6	k

a) Finn k slik at $p(x)$ er ei gyldig sannsynsfordeling.

Finn den kumulative fordelingsfunksjonen til X, $F(x)$.

Finn $\mathrm{E}(X)$ og $\mathrm{Var}(X)$.

Oppgåve 2 Levetid

Anta at levetida T, målt i timar (treng ikkje vere heiltalig), til ein tilfeldig vald elektronisk komponent er eksponentialfordelt med sannsynstettleik

$$f(t;\beta) = \begin{cases} \frac{1}{\beta}e^{-t/\beta} & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{1}$$

der $\beta > 0$ er ein parameter.

a) Anta (berre i dette punktet) at $\beta = 30$.

Vis at den kumulative fordelingsfunksjonen til T er gjeve som

$$P(T \leq t) = \begin{cases} 1 - e^{-t/30} & t \geq 0 \\ 0 & t < 0. \end{cases}$$

Finn $P(T < 20)$ og $P(T < 20 \cup T > 40)$.

Finn medianen til T, det vil seie k slik at $P(T \leq k) = 0.5$.

b) Vis at eksponentialfordelinga er utan hukommelse, det vil seie

$$P(T \geq t + s \mid T \geq t) = P(T \geq s)$$

for $s \geq 0$.

Vis at $\frac{2}{\beta}T$ er khikvadratfordelt med 2 fridomsgrader. (Vink: du kan nytte at $\Gamma(1) = 1$ i formelen i formelsamlinga for sannsynstettleiken til ein khikvadratfordelt variabel.)

La $T_1, T_2, \ldots, T_n$ vere uavhengige observasjonar av levetida til n elektroniske komponentar.

c) Grunngje kort kvifor $\frac{2}{\beta}\sum_{i=1}^{n} T_i$ er khikvadratfordelt med $2n$ fridomsgrader. (Vink: du kan nytte resultata frå oppgåve **b)** utan bevis.)

Nytt dette til å utleie eit $(1-\alpha)100$ %-konfidensintervall for β.

Kva blir 95 %-konfidensintervallet dersom $n = 20$ og $\sum_{i=1}^{20} t_i = 30$?

Oppgåve 3 Søvnløyse

Anta at tida (i minutt), X, det tar for ein pasient å sovne etter å tatt sovemedisin mot søvnløyse er normalfordelt med forventningsverdi μ og standardavvik σ.

a) Anta berre i dette punktet at $\mu = 35$ og $\sigma = 10$.

Finn sannsynet for at pasienten sovnar innan 20 minutt etter å ta tatt sovemedisinen.

Gjeve at pasienten ikkje har sovna innan dei første 20 minutta, finn sannsynet for pasienten ikkje har sovna innan dei første 40 minutta.

Anta at to pasientar tar sovemedisinen uavhengig av kvarandre ein vilkårleg dag. Finn sannsynet for at totaltida det tar før dei sovner er over 90 minutt.

To farmasiselskap konkurrerer om å vera marknadsleiande innan sovemedisin og hevdar begge at medisinen deira er betre enn konkurrrenten sin. Anta at begge selskapa gir sovemedisinen sine til n unike pasientar og måler tida det tar før kvar pasient sovnar.

Anta at tidene $X_1, X_2, \ldots, X_n$ det tar før pasientane til selskap A sovnar er eit tilfeldig utval med ukjend forventningsverdi μ og kjent standardavvik $\sigma = 10$. Vidare kan du anta at dei tilsvarande tidene $Y_1, Y_2, \ldots, Y_n$ for selskap B er uavhengige og normalfordelt med ukjend forventningsverdi θ og kjent standardavvik $\tau = 12$. Du kan anta at dei to utvala er uavhengige av kvarandre.

Selskap B ynskjer å utføre hypotesetesten

$$\mathrm{H}_0\colon \mu = \theta \qquad \text{mot} \qquad \mathrm{H}_1\colon \mu > \theta. \tag{2}$$

med eit signifikansnivå $\alpha = 0.1$ basert på dei tilfeldige utvala $X_1, X_2, \ldots, X_n$ og $Y_1, Y_2, \ldots, Y_n$.

b) Anta at den sanne differansen mellom effekten til dei to selskapa er $\mu - \theta = 5$ minutt.

Utlei eit uttrykk for det minste talet på pasientar n begge selskapa må gje sovemedisinen sin til dersom dei krev at testen sin styrke skal vere minst 95 % dersom den sanne differansen er $\mu - \theta = 5$ minutt.

Oppgåve 4 Bordtennisball

Ein produsent produserer bordtennisballar til profesjonelt bruk. Eit av krava for at ein bordtennisball skal kunne brukast i konkurranse er at diameteren er 40 millimeter. Anta at diameteren, i millimeter, til ein bordtennisball er normalfordelt med ukjend forventningsverdi μ og kjent standardavvik $\sigma = 0.5$.

I dei siste konkurransane har fleire bordtennisspelarar hevda at ballen har hatt feil diameter. Produsenten ynskjer derfor å utføre ein hypotesetest for å undersøkje om påstanden til spelarane stemmer. Produsenten utfører hypotesetesten

$$\mathrm{H}_0\colon \mu = 40 \text{ millimeter} \qquad \text{mot} \qquad \mathrm{H}_1\colon \mu \neq 40 \text{ millimeter} \tag{3}$$

ved eit signifikansnivå $\alpha = 0.1$ basert på eit tilfeldig utval X_1 , X_2, …, X_{10}. Det er gjeve at $\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 40.2$.

a) Utfør hypotesetesten over ved bruk av p-verdi. Vil produsenten forkaste nullhypotesen?

Utlei eit uttrykk for eit 95 %-prediksjonsintervall for diameteren X_0 til ein ny bordtennisball, uavhengig av X_1, X_2, …, X_{10}.

Oppgåve 5 Lineær regresjon

Anta følgjande enkle lineære regresjonsmodell,

$$Y = \beta_0 + \beta_1 x + \varepsilon, \tag{4}$$

der β_0 og β_1 er ukjende parametrar, x er ein kjend forklaringsvariabel og ε er anteke normalfordelt med forventningsverdi 0 og kjend varians σ^2.

Anta at me frå modellen definert i (4) har eit tilfeldig utval (x_1, Y_1), (x_2, Y_2), …, (x_n, Y_n).

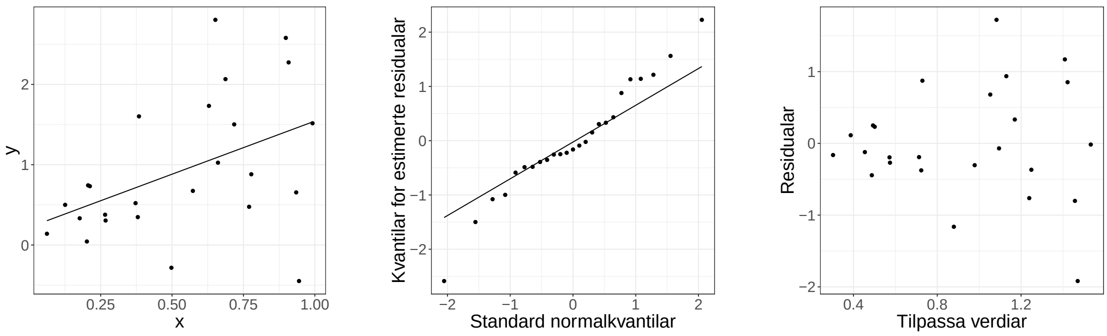
a) Anta (berre i dette punktet) at $\beta_0 = 0$.

Vis at sannsynsmaksimeringsestimatoren for β_1 basert på det tilfeldige utvalet (x_1, Y_1), (x_2, Y_2), …, (x_n, Y_n) då er

$$\widetilde{\beta}_1 = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}.$$

Finn forventningsverdi og varians til $\widetilde{\beta}_1$.

I figuren er regresjonsmodellen definert i (4) tilpassa eit tilfeldig utval (x_1, Y_1), (x_2, Y_2), $\ldots$, (x_{25}, Y_{25}) vist saman med eit normalsannsynsplott (Q–Q-plott) av residuala, og dei tilpassa observasjonane plotta mot residuala. Det er gjeve at $\widehat{\beta}_0 = 0.22$ og $\widehat{\beta}_1 = 1.32$.



b) Finn estimert forventa verdi for Y når det er gjeve at $x = 0.25$.

Basert på figuren, diskuter kort om den tilpassa modellen er rimeleg for observasjonane. Oppgi særskild kva antakingar som må vere oppfylt ved bruk av ein enkel lineær regresjonsmodell.

Oppgåve 6

La X_1, X_2 og X_3 vere uavhengige poissonfordelte stokastiske variablar med forventningsverdi høvesvis μ_1, μ_2 og μ_3.

La

$$Z_i = \begin{cases} 1 & \text{viss } X_i = X_3 = 0, \\ 0 & \text{elles,} \end{cases} \tag{5}$$

for $i = 1, 2$.

a) Vis at Z_i er bernoullifordelt, det vil seie binomisk fordelt med $n = 1$ delforsøk, med suksessannsyn $p_i = e^{-(\mu_i + \mu_3)}$, og bruk dette til å finne forventningsverdien og variansen til Z_i for $i = 1, 2$.

Nytt dette til å utleie eit uttrykk for kovariansen mellom Z_1 og Z_2, det vil seie $\text{Cov}(Z_1, Z_2)$.