Institutt for matematiske fag

Eksamensoppgåve i **TMA4245 Statistikk**

Fagleg kontakt under eksamen:

Tlf:

Eksamensdato:

Eksamenstid (frå–til):

Hjelpemiddelkode/Tillatne hjelpemiddel:

Annan informasjon:

Målform/språk: nynorsk

Sidetal: 8

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

a) For å vera ein gyldig sannsynsfordeling må me ha $\sum_x p(x) = 1$, me må difor ha $k = 0.1$.

Den kumulative fordelingsfunksjonen til X er gjeve som

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < -1 \\ 0.3 & -1 \leq x < 0 \\ 0.9 & 0 \leq x < 1 \\ 1 & x \geq 1. \end{cases}$$

Forventningsverdien er gjeve som

$$\mathrm{E}(X) = \sum_x x \cdot p(x) = -1 \cdot 0.3 + 0 \cdot 0.6 + 1 \cdot 0.1 = -0.2.$$

For å finne variansen nyttar me at $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$ der

$$\mathrm{E}(X^2) = \sum_x x^2 \cdot p(x) = (-1)^2 \cdot 0.3 + 0^2 \cdot 0.6 + 1^2 \cdot 0.1 = 0.4.$$

Variansen til X er derfor

$$\mathrm{Var}(X) = 0.4 - (-0.2)^2 = 0.36.$$

Oppgåve 2

a) For $t \geq 0$ har me

$$\begin{aligned} P(T \leq t) &= \int_0^t \frac{1}{\beta} e^{-x/\beta} \mathrm{d}x \\ &= \left[-e^{-x/30}\right]_0^t \\ &= 1 - e^{-t/30} \end{aligned}$$

og $P(T \leq t) = 0$ for $t < 0$.

Me har

$$P(T < 20) = P(T \leq 20) = F(20) = 1 - e^{-20/30} = 0.487$$

og

$$\begin{aligned} P(T < 20 \cup T > 40) &= P(T < 20) + P(T > 40) \\ &= P(T < 20) + 1 - P(T < 40) \\ &= 0.487 + 1 - (1 - e^{-40/30}) \\ &= 0.750. \end{aligned}$$

For at $P(T \leq k) = 0.5$ må me ha

$$1 - e^{-k/30} = 0.5 \Leftrightarrow k = -30 \cdot \ln(0.5) \approx 20.79.$$

b) Me har

$$\begin{aligned} P(T \geq t+s \mid T \geq t) &= \frac{P(T \geq t+s \cap T \geq t)}{P(T \geq t)} \\ &= \frac{P(T \geq t+s)}{P(T \geq t)} \\ &= \frac{e^{-(t+s)/\beta}}{e^{-t/\beta}} \\ &= e^{-s/\beta} \\ &= P(T \geq s). \end{aligned}$$

Me ynskjer å vise at sannsynstettleiken til Y er khikvadratfordelinga med 2 fridomsgrader, det vil seie Y har sannsynstettleik

$$f_Y(y) = \frac{1}{2} \cdot e^{-y/2}$$

for $y \geq 0$. Sidan $Y = \frac{2}{\beta}T$ er ein ein-til-ein transformasjon kan me nytte transformasjonsformelen som gir

$$\begin{aligned} f_Y(y) &= f_X\left(\frac{y\beta}{2}\right)\left|\frac{\mathrm{d}}{\mathrm{d}y}\frac{\beta}{2}y\right| \\ &= \frac{1}{\beta}e^{-\frac{y\beta/2}{\beta}} \cdot \frac{\beta}{2} \\ &= \frac{1}{2}e^{-y/2} \end{aligned}$$

for $y \geq 0$ og 0 elles som var det me skulle vise.

c) Sidan $T_1, T_2, \ldots, T_n$ er uavhengig av kvarandre er $\frac{2}{\beta}T_1, \frac{2}{\beta}T_2, \ldots \frac{2}{\beta}T_n$ uavhengige av kvarandre. Det er kjend at summen av uavhengige khikvadratfordelte variablar og er khikvadratfordelt med summen av dei individuelle fridomsgradene. Sidan $\frac{2}{\beta}T_i$ er khikvadratfordelt med 2 fridomsgrader er $\frac{2}{\beta}\sum_{i=1}^{n} T_i$ khikvadratfordelt med $2n$ fridomsgrader.

Me har difor at

$$\mathrm{P}\left(\chi^2_{1-\alpha/2,2n} \leq \frac{2\sum_{i=1}^{n} T_i}{\beta} \leq \chi^2_{\alpha/2,2n}\right) = 1 - \alpha.$$

Ved å løyse for β får me

$$\mathrm{P}\left(\frac{2\sum_{i=1}^{n} T_i}{\chi^2_{\alpha/2,2n}} \leq \beta \leq \frac{2\sum_{i=1}^{n} T_i}{\chi^2_{1-\alpha/2,2n}}\right) = 1-\alpha.$$

det vil seie at eit $(1-\alpha)\cdot 100$ % konfidensintervall for β er

$$\left[\frac{2\sum_{i=1}^{n} T_i}{\chi^2_{\alpha/2,2n}}, \frac{2\sum_{i=1}^{n} T_i}{\chi^2_{1-\alpha/2,2n}}\right]$$

der $P(V \geq \chi^2_{\alpha/2,\nu}) = \alpha/2$ der V khikvadratfordelt med ν fridomsgrader. Innsatt tala i oppgåva har me

$$\left[\frac{2\cdot 30}{59.342}, \frac{2\cdot 30}{24.433}\right] = [1.011, 2.456]$$

Oppgåve 3

a) Sannsynet for å sovne innan 20 minutt er

$$P(X \leq 20) = P\left(Z \leq \frac{20-35}{10}) = P(Z \leq -1.5\right) = 0.067.$$

Sannsynet for å ikkje ha sovna innan 40 minutt, gjeve at ein var våken etter 20 minutt er

$$\begin{aligned}
(PX \geq 40 \mid X \geq 20) &= \frac{P(X \geq 40 \cap X \geq 20)}{P(X \geq 20)} \\
&= \frac{P(X \geq 40)}{P(X \geq 20)} \\
&= \frac{1-P(Z \leq \frac{40-35}{10})}{1-P(X \leq 20)} \\
&= \frac{1-0.691}{0.933} \\
&= 0.331.
\end{aligned}$$

La X_1 vere tida det tar før den første pasienten sovner og X_2 tida før den andre pasienten sovner, som er uavhengige og normalfordelt med forventning 35 minutt og standardavvik 10 minutt. Sidan ein lineærkombinasjon av uavhengige og normalfordelte stokastiske variablar er og normalfordelt, det vil seie $X_1 + X_2$ er normalfordelt med forventningsverdi 70 og standardavvik $\sqrt{2 \cdot 10^2}$.

$$\begin{aligned} P(X_1 + X_2 \geq 90) &= 1 - P(X_1 + X_2 \leq 90) \\ &= 1 - P\left(Z \leq \frac{90 - 70}{\sqrt{2 \cdot 10^2}}\right) \\ &= 1 - P(Z \leq 1.41) \\ &= 1 - 9207 \\ &= 0.0793. \end{aligned}$$

b) Me nyttar $\bar{X} - \bar{Y}$ som estimator for $\mu - \theta$. Sidan $\bar{X}$ og $\bar{Y}$ er uavhengigeg og normalfordelte stokastiske variablar har me at $\bar{X} - \bar{Y}$ er normalfordelt med forventning 0 og standaravvik $\sqrt{\frac{10^2}{n} + \frac{12^2}{n}}$ under nullhypotesen. Ved eit signifikansnivå $\alpha = 0.1$ forkaster me nullhypotesen dersom

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{10^2}{n} + \frac{12^2}{n}}} \geq z_{0.1} = 1.282.$$

Me krev no at testen sin styrke er minst 95 % når $\mathrm{H}_1 : \mu - \theta = 5$ er sann, det vil seie

$$\begin{aligned} 0.95 &\geq P(\text{forkast } \mathrm{H}_0 \text{ når } \mathrm{H}_1 : \mu - \theta = 5) \\ &= P\left(\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{10^2}{n} + \frac{12^2}{n}}} \geq 1.282 \text{ når } \mathrm{H}_1 : \mu - \theta = 5\right) \\ &= P\left(\bar{X} - \bar{Y} \geq 1.282\sqrt{\frac{10^2}{n} + \frac{12^2}{n}} \text{ når } \mathrm{H}_1 : \mu - \theta = 5\right) \\ &= P\left(Z \geq \frac{1.282\sqrt{\frac{10^2}{n} + \frac{12^2}{n}} - 5}{\sqrt{\frac{10^2}{n} + \frac{12^2}{n}}} \text{ når } \mathrm{H}_1 : \mu - \theta = 5\right) \\ &P\left(Z \geq 1.282 + -\frac{5}{\sqrt{\frac{10^2}{n} + \frac{12^2}{n}}} \text{ når } \mathrm{H}_1 : \mu - \theta = 5\right), \end{aligned}$$

altså $P\left(Z \leq 1.282 - \frac{5}{\sqrt{\frac{10^2}{n} + \frac{12^2}{n}}} \text{ når } \mathrm{H}_1 : \mu - \theta = 5\right) \leq 0.05$. Det betyr at me må ha

$$z_{-0.05} = -1.645 \leq 1.282 - \frac{5}{\sqrt{\frac{10^2}{n} + \frac{12^2}{n}}}$$

$$2.927\sqrt{\frac{10^2}{n} + \frac{12^2}{n}} \geq 5$$

$$\frac{10^2}{n} + \frac{12^2}{n} \leq \left(\frac{5}{2.926}\right)^2$$
$$n \geq \frac{2.926^2 \cdot (10^2 + 12^2)}{5^2} \approx 83.56.$$

Kvart selskap må gi sovemedisinen til minst 84 personer.

Oppgåve 4

a) Under nullhypotesen har me at $Z = \frac{\bar{X}-40}{\sqrt{0.05/10}}$ er standard normalfordelt. Me har observert

$$Z_{\text{obs}} = \frac{40.2 - 40}{\sqrt{0.5^2/10}} = 1.27$$

Frå definisjon av p-verdi har me

$$\begin{aligned} p - \text{verdi} &= P(|Z_{\text{obs}}| \geq 1.27) \\ &= 2 \cdot P(Z \geq 1.27) \\ &= 2 \cdot (1 - P(Z \leq 1.27)) \\ &= 2 \cdot (1 - 0.898) \\ &= 0.204. \end{aligned}$$

Altså vil me ikkje forkaste nullhypotesen ved eit signifikansnivå $\alpha = 0.1$.

Ta utgangspunkt i $\bar{X} - X_0$ som er ein lineærkombinasjon av uavhengige og normalfordelte stokastiske variablar og difor og normalfordelt med forventningsverdi 0 og varians $\sigma^2(\frac{1}{n} + 1)$. Me har difor

$$P(-z_{0.025} \leq \frac{\bar{X} - X_0}{\sqrt{\sigma^2(\frac{1}{n} + 1)}} \leq z_{0.025}) = 0.95$$

Ved å isolere X_0 får me

$$P\left(\bar{X} - z_{0.025}\sqrt{\sigma^2(\frac{1}{n} + 1)} \leq X_0 \leq \bar{X} + z_{0.025}\sqrt{\sigma^2(\frac{1}{n} + 1)}\right) = 0.95,$$

det vil seie at eit 95 % prediksjonsintervall for X_0 er

$$\left[\bar{X} - z_{0.025}\sqrt{\sigma^2(\frac{1}{n} + 1)}, \bar{X} + z_{0.025}\sqrt{\sigma^2(\frac{1}{n} + 1)}\right].$$

Oppgåve 5

a) Rimelegheitsfunksjonen er gjeve ved

$$\begin{aligned} L(\beta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y_i-\beta x_i)^2} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta x_i)^2}. \end{aligned}$$

Log-rimelegheitsfunksjonen er

$$l(\beta) = -n\ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta x_i)^2.$$

Deriverer med hensyn på β og set lik 0.

$$\frac{\partial l(\beta)}{\partial \beta} = \frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(y_i - \beta x_i)x_i = 0.$$

Me får då $\beta = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$, det vil seie at sannsynsmaksimeringsestimatoren til β er

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

$$\begin{aligned} \mathrm{E}(\widetilde{\beta}) &= \mathrm{E}\left(\frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}\right) \\ &= \frac{\sum_{i=1}^{n} x_i \mathrm{E}(Y_i)}{\sum_{i=1}^{n} x_i^2} \\ &= \frac{\sum_{i=1}^{n} x_i \beta x_i}{\sum_{i=1}^{n} x_i^2} \\ &= \beta. \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(\widetilde{\beta}) &= \mathrm{Var}\left(\frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}\right) \\ &= \frac{\sum_{i=1}^{n} x_i^2 \mathrm{E}(Y_i)}{(\sum_{i=1}^{n} x_i^2)^2} \\ &= \frac{\sum_{i=1}^{n} x_i^2 \sigma^2}{(\sum_{i=1}^{n} x_i^2)^2} \\ &= \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}. \end{aligned}$$

b) Forventa estimert verdi er $0.2 + 1.32 \cdot 0.25 = 0.55$.

Følgjande antakingar må vere oppfylt for å kunne nytte ein enkel lineær regresjonsmodell

- lineær samanheng mellom x og y: det verker å vere ein svak stigande lineær trend mellom x og y.
- konstant varians σ^2: residualplottet viser at variansen auker med auka verdiar $\widehat{y}_i$. Ein kan og sjå den same trenden frå plottet av x mot y. Altså er ikkje variansen konstant, men ein funksjon av x.
- residuala er normalfordelt: normalsannsynsplottet indikerer avvik frå dei teoretiske kvantilane til normalfordelinga, derfor er det ein indikasjon på at residuala ikkje er normalfordelt.
- Y_i-ane er uavhengige: dette er vanskeleg å avgjere, men det er ikkje noko som tyder på at dei ikkje er det.

Oppgåve 6

a) Kva Z_i kan ta ein av to verdiar (0 eller 1) og me har kun eit forsøk med suksess (1) eller ikkje suksess (0). Difor må Z_i vere bernoullifordelt med suksessannsyn

$$\begin{aligned} p_i &= P(Z_i = 1) \\ &= P(X_i = 0 \cap X_3 = 0) \\ &= P(X_i = 0)P(X_3 = 0) \\ &= e^{-\mu_i}e^{-\mu_3} \\ &= e^{-(\mu_i+\mu_3)}. \end{aligned}$$

For ein bernoullifordelt variabel er det kjend $\mathrm{E}(Z_i) = p_i$ og $\mathrm{Var}(Z_i) = p_i(1 - p_i)$.

Frå definisjonen av kovarians har me at $\mathrm{Cov}(Z_1, Z_2) = E(Z_1Z_2) - E(Z_1)E(Z_2)$, der

$$
\begin{aligned}
\mathrm{E}(Z_1 Z_2) &= \sum_{z_1} \sum_{z_2} z_1 z_2 P(Z_1 = z_1, Z_2 = z_2) \\
&= 0 \cdot 0 \cdot P(Z_1 = 0, Z_2 = 0) + 1 \cdot 0 \cdot P(Z_1 = 1, Z_2 = 0) \\
&\quad + 0 \cdot 1 \cdot P(Z_1 = 0, Z_2 = 1) + 1 \cdot 1 \cdot P(Z_1 = 1, Z_2 = 1) \\
&= P(Z_1 = 1, Z_2 = 1) \\
&= P(X_1 = 0, X_2 = 0, X_3 = 0) \\
&= P(X_1 = 0) P(X_2 = 0) P(X_3 = 0) \\
&= e^{-\mu_1} e^{-\mu_2} e^{-\mu_3} \\
&= e^{-(\mu_1 + \mu_2 + \mu_4)}.
\end{aligned}
$$

Til saman har me derfor at

$$
\begin{aligned}
\mathrm{Cov}(Z_1, Z_2) &= \mathrm{E}(Z_1 Z_2) - E(Z_1) E(Z_2) \\
&= e^{-(\mu_1 + \mu_2 + \mu_3)} - e^{-(\mu_1 + \mu_3)} e^{-(\mu_2 + \mu_3)} \\
&= e^{-(\mu_1 + \mu_2 + \mu_3)} (1 - e^{-\mu_3}).
\end{aligned}
$$