NTNU

Kunnskap for en bedre verden

Institutt for matematiske fag

# Eksamensoppgave i **TMA4240/45 Statistikk**

**Faglig kontakt under eksamen:** Øyvind Bakke

**Tlf:** 990 41 673

**Eksamensdato:**

**Eksamenstid (fra–til):**

**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:** 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** Sannsynlighet

La $X$ være en stokastisk variabel med punktsannsynlighet $p(x) = P(X = x)$ som vist i tabellen:

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

**a)** Finn $k$ slik at $p(x)$ er en gyldig sannsynlighetsfordeling.

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

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

**Oppgave 2** Levetid

Anta at levetiden $T$, målt i timer (behøver ikke være et heltall), til en tilfeldig valgt elektronisk komponent er eksponentialfordelt med sannsynlighetstetthet

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

der $\beta > 0$ er en parameter.

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

Vis at den kumulative fordelingsfunksjonen til $T$ er gitt 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 si $k$ slik at $P(T \leq k) = 0.5$.

**b)** Vis at eksponentialfordelingen er uten hukommelse, det vil si

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

for $s \geq 0$.

Vis at $\frac{2}{\beta}T$ er khikvadratfordelt med 2 frihetsgrader. (Vink: du kan bruke at $\Gamma(1) = 1$ i formelen i formelsamlingen for sannsynlighetstettheten til en khikvadratfordelt variabel.)

La $T_1$, $T_2$, $\ldots$, $T_n$ være uavhengige observasjoner av levetiden til $n$ elektroniske komponenter.

**c)** Grunngi kort hvorfor $\frac{2}{\beta}\sum_{i=1}^{n} T_i$ er khikvadratfordelt med $2n$ frihetsgrader. (Vink: du kan bruke resultatene fra oppgave **b)** uten bevis.)

Bruk dette til å utlede et $(1-\alpha)100$ %-konfidensintervall for $\beta$.

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

**Oppgave 3** Søvnløshet

Anta at tiden (i minutt), $X$, det tar for en pasient å sovne etter å tatt sovemedisin mot søvnløshet er normalfordelt med forventningsverdi $\mu$ og standardavvik $\sigma$.

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

Finn sannsynligheten for at pasienten sovner innen 20 minutt etter å ta tatt sovemedisinen.

Gitt at pasienten ikke har sovnet innen de første 20 minuttene, finn sannsynligheten for pasienten ikke har sovnet innen de første 40 minuttene.

Anta at to pasienter tar sovemedisinen uavhengig av hverandre en vilkårlig dag. Finn sannsynligheten for at totaltiden det tar før de sovner er over 90 minutt.

To farmasiselskap konkurrerer om å være markedsledende innen sovemedisin og hevder begge at deres medisin er bedre enn konkurrenten sin. Anta at begge selskapene gir sovemedisinen sin til $n$ unike pasienter og måler tiden det tar før hver pasient sovner.

Anta at tidene $X_1, X_2, \ldots, X_n$ det tar før pasientene til selskap A sovner er et tilfeldig utvalg med ukjent forventningsverdi $\mu$ og kjent standardavvik $\sigma = 10$. Videre kan du anta at de tilsvarende tidene $Y_1, Y_2, \ldots, Y_n$ for selskap B er uavhengige og normalfordelt med ukjent forventningsverdi $\theta$ og kjent standardavvik $\tau = 12$. Du kan anta at de to utvalgene er uavhengige av hverandre.

Selskap B ønsker å utføre hypotesetesten

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

med et signifikansnivå $\alpha = 0.1$ basert på de tilfeldige utvalgene $X_1, X_2, \ldots, X_n$ og $Y_1, Y_2, \ldots, Y_n$.

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

Utled et uttrykk for det minste antallet pasienter $n$ begge selskapene må gi sovemedisinen sin til dersom de krever at testens styrke skal være minst 95 % dersom den sanne differansen er $\mu - \theta = 5$ minutt.

**Oppgave 4** Bordtennisball

En produsent produserer bordtennisballer til profesjonelt bruk. Et av kravene for at en bordtennisball skal kunne brukes i konkurranse er at diameteren er 40 millimeter. Anta at diameteren, i millimeter, til en tilfeldig valgt bordtennisball er normalfordelt med ukjent forventningsverdi $\mu$ og kjent standardavvik $\sigma = 0.5$.

I de siste konkurransene har flere bordtennisspillere hevdet at bordtennisballen har hatt feil diameter. Produsenten ønsker derfor å utføre en hypotesetest for å undersøke om påstanden til spillerene 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 et signifikansnivå $\alpha = 0.1$ basert på et tilfeldig utvalg $X_1, X_2, \ldots, X_{10}$. Det er gitt 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?

Utled et uttrykk for et 95 %-prediksjonsintervall for diameteren $X_0$ til en ny bordtennisball, uavhengig av $X_1, X_2, \ldots, X_{10}$.

**Oppgave 5** Lineær regresjon

Anta følgende enkle lineære regresjonsmodell,

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

der $\beta_0$ og $\beta_1$ er ukjente parametre, $x$ er en kjent forklaringsvariabel, og $\varepsilon$ er antatt normalfordelt med forventningsverdi 0 og kjent varians $\sigma^2$.

Anta at vi fra modellen definert i (4) har et tilfeldig utvalg $(x_1, Y_1)$, $(x_2, Y_2)$, $\ldots$, $(x_n, Y_n)$.

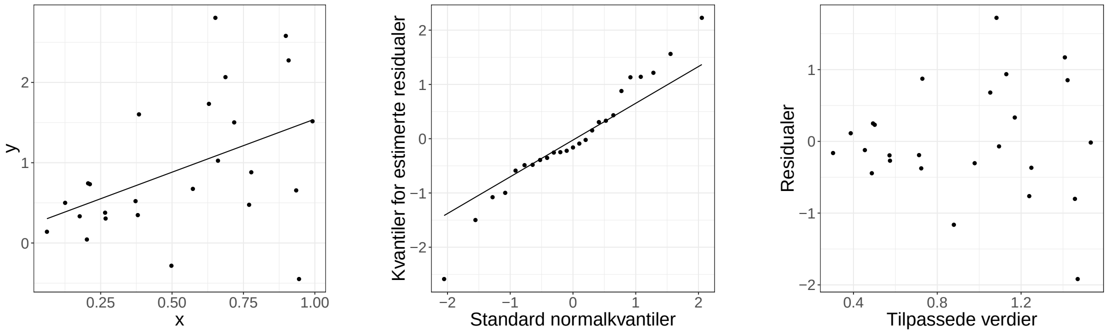
**a)** Anta (kun i dette punktet) at $\beta_0 = 0$.

Vis at sannsynlighetsmaksimeringsestimatoren for $\beta_1$ basert på det tilfeldige utvalget $(x_1, Y_1)$, $(x_2, Y_2)$,$\ldots$, $(x_n, Y_n)$ da 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) tilpasset et tilfeldig utvalg $(x_1, Y_1)$, $(x_2, Y_2)$,..., $(x_{25}, Y_{25})$ vist sammen med et normalsannsynlighetsplott (Q–Q-plott) av residualene, og de tilpassede observasjonene plottet mot residualene. Det er gitt at $\widehat{\beta}_0 = 0.22$ og $\widehat{\beta}_1 = 1.32$.




**b)** Finn estimert forventet verdi for $Y$ når det er gitt at $x = 0.25$.

Basert på figuren, diskuter kort om den tilpassede modellen er rimelig for observasjonene. Oppgi spesielt hvilke antagelser som må være oppfylt ved bruk av en enkel lineær regresjonsmodell.

## Oppgave 6

La $X_1$, $X_2$ og $X_3$ være uavhengige poissonfordelte stokastiske variabler med forventningsverdi henholdsvis $\mu_1$, $\mu_2$ og $\mu_3$.

La

$$Z_i = \begin{cases} 1 & \text{hvis } X_i = X_3 = 0 \\ 0 & \text{ellers.} \end{cases} \tag{5}$$

for $i = 1, 2$.

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

Bruk dette til å utlede et uttrykk for kovariansen mellom $Z_1$ og $Z_2$, det vil si $\mathrm{Cov}(Z_1, Z_2)$.