NTNU
Kunnskap for ei betre verd

Institutt for matematiske fag

# Eksamensoppgåve i **TMA4240 / TMA4245 Statistikk**

**Fagleg kontakt under eksamen:** Håkon Tjelmeland
**Tlf:** 48 22 18 96

**Eksamensdato:** 10. august 2017
**Eksamenstid (frå–til):** 09.00-13.00
**Hjelpemiddelkode/Tillatne hjelpemiddel:**
*Tabeller og formler i statistikk*, Akademika,
K. Rottmann: *Matematisk formelsamling*,
Kalkulator Casio fx-82ES PLUS, CITIZEN SR-270X, CITIZEN SR-270X College eller HP30S,
Gult stempla A5-ark med eigne handskrivne notat.

**Annan informasjon:**
Alle svar skal grunngjevast og besvarelsen skal innehalde naturlege mellomrekningar.

**Målform/språk:** nynorsk
**Sidetal:** 3
**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 Sannsyn**

Anta at den stokastiske variabelen $X$ har sannsynstettleik

$$f(x;\theta) = c\exp\{-(x-\theta)\}\,,\;\; x \geq \theta,$$

medan $f(x;\theta) = 0$ for $x < \theta$. $\theta$ er ein konstant parameter og $c$ er ein positiv konstant.

**a)** Bestem konstanten $c$, og finn sannsynet for at $X > \theta + 1$.

**b)** Utlei eit uttrykk for sannsynsmaksimeringsestimatoren $\hat{\Theta}$ for parameteren $\theta$ basert på $n$ uavhengige observasjonar, $x_1, \ldots, x_n$, av $X$. Hint: Skissér rimelegheitsfunksjonen (likelihood-funksjonen).

**c)** La $X_1, X_2, \ldots, X_{10}$ vere 10 uavhengige stokastiske variablar som kvar har same sannsynstettleik som $X$. Bestem sannsynlighetstettleiken for den stokastiske variabelen $W = \min\{X_1, X_2, \ldots, X_{10}\}$. Rekn ut sannsynet for at $W > \theta + 1$.

**Oppgåve 2 Utmatting**

Rotorfesta på eit helikopter er utsett for brot på grunn av utmatting. For å betre utmattingsfastheita kan ein foreta ei overflateherding av festa ved å bruke ein av dei to metodane A og B skildra under.

Metode A, og kalla ”shot-peening”, går ut på at materialet blir utsett for ei type sandblåsing med ganske grove kuler.

Ved Metode B (nitrering) blir overflaten varma opp til ca. 500 grader celsius og tilført ammoniakk. Dette gjev eit tynt, slitesterkt sjikt.

Ein er interessert i å undersøke om metode A gjev betre utmattingsfastheit enn metode B. Ved fabrikken som produserer rotorfesta, blir 13 rotorfeste tilfeldig valt ut til eit forsøk. De 7 første rotorfesta blir behandla etter metode A, og dei øvrige etter metode B. Dei 13 rotorfesta blir sett i ein utmattingsrigg og utsett for sykliske belastningar med spenningsvidde 700 MPa (MegaPascal). Erfaring har vist at den naturlege logaritmen til talet på syklar før eit brot oppstår vil vere tilnærma ein normalfordelt stokastisk variabel.

Den naturlige logaritmen til talet på syklar før brot for rotorfesta som er behandla etter metode A, er $X_1, \ldots, X_7$. For rotorfesta behandla etter metode B får ein tilsvarande $Y_1, \ldots, Y_6$. Vi antar at alle dei 13 observasjonane er uavhengige, normalfordelte stokastiske variable med ukjente forventningsverdiar og variansar:

$$E(X_i) = \mu_A\,,\ \mathrm{Var}(X_i) = \sigma_A^2\,,\ i = 1, \ldots, 7.$$
$$E(Y_j) = \mu_B\,,\ \mathrm{Var}(Y_j) = \sigma_B^2\,,\ j = 1, \ldots, 6.$$

Anta i dei to neste punkta at $\mu_A = \mu_B = \mu$ og $\sigma_A = \sigma_B = \sigma$, men at deira felles verdiar er ukjente.

**a)** Vis at
$$\hat{\mu} = \frac{1}{2}\left(\bar{X} + \bar{Y}\right)$$
er ein forventningsrett estimator for $\mu$ og finn hans varians uttrykt ved $\sigma$. Her er $\bar{X} = \sum_{i=1}^{7} X_i/7$ og $\bar{Y} = \sum_{j=1}^{6} Y_j/6$.

**b)** Foreslå så ein forventningsrett estimator $\mu^*$ for $\mu$ som har mindre varians enn $\hat{\mu}$. Vis dette ved utrekning.

I resten av oppgåva vil vi ikkje lenger anta at $\mu_A = \mu_B$ og $\sigma_A = \sigma_B$. Følgjande formel er henta direkte frå læreboka:

$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}$$

Den blir brukt til å rekne ut tilnærma tal på fridomsgrader $\nu$ for estimatorar som er vanleg å bruke ved estimering og hypotesetesting som i resten av denne oppgaven.

**c)** La $\delta = \mu_A - \mu_B$. Utlei eit $100(1 - \alpha)\%$ konfidensintervall for $\delta$. Kva blir intervallet når $\alpha = 0.1$ med dei empiriske resultata som er gjeve til slutt i oppgåva?

**d)** Tyder resultata frå forsøket på at metode A gjev betre utmattingsfastheit enn metode B? Formulér dette som eit hypotesetestingsproblem og begrunn valet av nullhypotese og alternativ hypotese. Velg signifikansnivå 0.05 og gjennomfør testen med dei resultata som er gjeve til slutt i oppgåva.

$\bar{x} = 15.22$

$\sum_{i=1}^{7}(x_i - \bar{x})^2 = 0.32$

$\bar{y} = 14.56$

$\sum_{j=1}^{6}(y_j - \bar{y})^2 = 0.47$

$\nu \approx 9$

## Oppgåve 3 Kvalitetskontroll

Ved produksjonen av ein type medisin i pulverform, som blir solgt på forsegla boksar pakka i pappesker med $k$ boksar i kvar ekse, er det viktig å sørge for at pulveret ikkje er forureina av soppsporer. Dette blir kontrollert ved ein laboratorietest. Sannsynet for at ein tilfeldig valt boks gjev eit positivt testresultat (dvs. at boksen er infisert), blir antatt å vere lik $p$, og testresultata for dei ulike boksane blir antatt å vere uavhengige.

Ein går fram på følgjande måte: Testane blir gjort av dei $k$ ($> 1$) boksane i ei eske om gangen. Vi kallar desse boksane ei $k$-gruppe. Halvparten av innhaldet i kvar boks i ei $k$-gruppe blir blanda og blandinga blir analysert. Dersom denne testen gjev positivt testresultat, noko som betyr at minst ein av boksane er infisert, blir det gjort nye testar av resten av innhaldet i dei $k$ boksane, og prøvane blir analysert enkeltvis. Dersom derimot blandinga gjev negativt testresultat, blir det antatt at ingen av boksane i $k$-gruppa er infisert. Dersom det er soppsporer i ein boks, antar vi at dei er tilnærma jamnt fordelt i boksen.

**a)** Begrunn kvifor talet på infiserte boksar i ei $k$-gruppe er binomisk fordelt, og vis at sannsynet for at ei blanding av innhaldet i $k$ boksar skal gje positivt testresultat, er

$$1 - (1-p)^k.$$

**b)** Korleis blir betinga sannsyn for ei hending $A$ gjeve ei anna hending $B$ definert? Dersom ein boks er i ei $k$-gruppe som har vist positivt testresultat, kva er då sannsynet for at boksen er infisert?

**c)** Anta at ein kvalitetskontroll blir gjennomført ved at $m$ esker med $k$ boksar i kvar eske blir tatt ut for å testast. La $X$ vere talet på prøvar som må analyserast før heile partiet på $mk$ boksar er ferdig testa. Vis at

$$E(X) = m + mk\Big(1 - (1-p)^k\Big).$$

Dersom $k = 4$, for kva verdiar av $p$ er den framgangsmåten som blir brukt å foretrekke i staden for å teste boksane enkeltvis med ein gong?