NTNU

Kunnskap for en bedre verden

Institutt for matematiske fag

# Eksamensoppgave i **TMA4245 Statistikk**

**Faglig kontakt under eksamen:** Håkon Tjelmeland

**Tlf:** 48 22 18 96

**Eksamensdato:** ??. august 2014

**Eksamenstid (fra–til):** 09:00–13:00

**Hjelpemiddelkode/Tillatte hjelpemidler:** C:

- *Tabeller og formler i statistikk*, Tapir forlag.
- K. Rottman: *Matematisk formelsamling*.
- Kalkulator CASIO fx-82ES PLUS, CITIZEN SR-270X, CITIZEN SR-270X College eller HP30S.
- Et stemplet gult A5-ark med egne håndskrevne formler og notater.

**Annen informasjon:**
Alle svar skal begrunnes og besvarelsen skal inneholde naturlig mellomregning.

**Målform/språk:** bokmål

**Antall sider:** 6

**Antall sider vedlegg:** 0

**Kontrollert av:**

Dato Sign

## Oppgave 1 Medisinmengder

En bedrift produserer en type medisin i pulverform. Medisinen selges på flasker som hver er ment å ta 500 gram pulver. Bedriften anvender en tappemaskin som utporsjonerer pulveret i riktige doser. Når tappemaskinen innstilles på $\mu$ gram, kan de utporsjonerte pulvermengdene $X_1, X_2, \ldots$, antas uavhengige og normalfordelte med forventningsverdi $\mu$ gram og standardavvik $\sigma = 12$ gram. Vekten av flaskene (uten pulver) som pulveret fylles i, $Y_1, Y_2, \ldots$, kan antas å være uavhengige og normalfordelte med forventningsverdi 50 gram og standardavvik 2 gram. Pulvervekt og flaskevekt antas å være uavhengige av hverandre.

I punkt a) og b) i denne oppgaven er $\mu$ satt lik 510 gram.

**a)** Hva er sannsynligheten for at ei tom flaske veier mer enn 53 gram?

Begrunn av vekten av ei flaske fylt med pulver er normalfordelt med forventningsverdi lik 560 gram og standardavvik $\sqrt{148}$ gram.

Hva er sannsynligheten for at vektforskjellen mellom to flasker fylt med pulver skal være større enn 15 gram?

Ei flaske fylt med pulver som veier mindre enn 540 gram sies å være undervektig. Kunder som kommer til å kjøpe ei eller flere undervektige flasker, har rett til å returnere disse. Flaskene pakkes i kartonger med 24 flasker i hver kartong.

**b)** Hva er sannsynligheten for at ei flaske fylt med pulver skal være undervektig?

La $U$ betegne antall undervektige flasker i en tilfeldig valgt kartong. Hvilken sannsynlighetsfordeling har $U$? Grunngi svaret.

Hva er sannsynligheten for at en tilfeldig valgt kartong skal inneholde ei eller flere undervektige flasker?

Det anses som dårlig reklame for bedriften at for mange flasker er undervektige. Bedriftsledelsen forlanger derfor at høyst 1% av flaskene i det lange løp skal være undervektige.

**c)** Hvordan må da maskinen innstilles for at dette kravet skal være oppfylt?

Når dette kravet er oppfylt, hvor mange undervektige flasker må en kunde forvente å få ved kjøp av 50 kartonger?

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\mu_i$ | 470 | 475 | 480 | 490 | 495 | 500 | 505 | 510 | 515 | 520 |
| $V_i$ | 488.1 | 455.4 | 497.8 | 475.5 | 502.3 | 469.6 | 517.2 | 484.5 | 506.2 | 520.1 |

Tabell 1: Verdier benyttet for $\mu_1, \mu_2, \ldots, \mu_{10}$ og målte utporsjonerte kvanta $v_1, v_2, \ldots, v_{10}$.

En kunde har kjøpt $n$ flasker med pulver og ønsker på det grunnlaget å anslå $\mu$. De andre parametrene har verdier som angitt tidligere i oppgaven. Kunden veier pulveret i hver flaske, $X_1, X_2, \ldots, X_n$, og benytter estimatoren

$$\widehat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

**d)** Utled et $100(1-\alpha)\%$ konfidensintervall for $\mu$ basert på pulvervektene i hver flaske, $X_1, X_2, \ldots, X_n$.

Bestem konfidensintervallet numerisk når $n = 24$, $\alpha = 0.05$, og $\bar{x} = \frac{1}{24}\sum_{j=1}^{24} x_i = 513.4$, hvor $x_i$ betegner målt vekt for flaske nr. $i$.

Hvor stor måtte $n$ ha vært for at lengden på dette 95%-konfidensintervallet skulle blitt kortere enn 8 gram?

Bedriften vil kontrollere at porsjoneringsmaskinen er riktig justert. Når maskinen innstilles på $\mu$ gram kan de utporsjonerte kvanta med pulver antas uavhengige og normalfordelte med forventningsverdi lik $\beta\mu$ og standardavvik lik 12, der $\beta = 1$ dersom maskinen er riktig justert. For å undersøke om maskinen er riktig justert utføres følgende forsøk: Maskinen innstilles på i alt $k$ verdier $\mu_1, \mu_2, \ldots, \mu_k$. Til hver innstilling $\mu_i$ svarer da en stokastisk variabel $V_i$, som representerer det utporsjonerte kvantum. Når forsøket er gjennomført, har en målt utporsjonerte kvanta $v_1, v_2, \ldots, v_k$, som er å betrakte som utfall av $V_1, V_2, \ldots, V_k$.

**e)** Finn sannsynlighetsmaksimeringsestimatoren (SME) for $\beta$.

Gjennomfør estimeringen når $k = 10$ og resultatet av forsøket ble som angitt i tabell 1. Det oppgis at $\sum_{i=1}^{k} v_i\mu_i = 2440526$ og $\sum_{i=1}^{k} \mu_i^2 = 2462800$.

**Oppgave 2**

La $X$ og $Y$ være diskret fordelte stokastiske variabler der $X, Y \in \{0, 1, 2\}$. La $f(x, y) = \mathrm{P}(X = x, Y = y)$ være simultan punktsannsynlighet for $X$ og $Y$ og anta at $f(x, y)$ er som angitt i følgende tabell.

| $x \backslash y$ | 0 | 1 | 2 |
|---|---|---|---|
| 0 | 0.10 | 0.25 | 0.15 |
| 1 | 0.06 | 0.15 | 0.09 |
| 2 | 0.04 | 0.10 | 0.06 |

**a)** Finn $\mathrm{P}(X > Y)$.

Finn (marginal) punktsannsynlighet for $X$ og for $Y$.

Er $X$ og $Y$ uavhengige? Begrunn svaret!

**Oppgave 3 Bremselengder**

Vi skal i denne oppgaven anta at bremselengden, $Y$, målt i meter for en bil som kjører $x$ km/time antas å være normalfordelt med forventningsverdi $\beta x^2$ og standardavvik $\sigma x$. En bil som for eksempel kjører i 50 km/time vil dermed ha en bremselengde som er normalfordelt med forventningsverdi $2500\beta$ og standardavvik $50\sigma$. Modellen har to parametre, $\beta$ og $\sigma^2$, og disse vil avhenge av forsøksbetingelsene, som for eksempel dekkenes egenskaper, veidekke og vær- og føreforhold.

Anta nå at verdien til $\beta$ er ukjent og skal estimeres. For å estimere $\beta$ gjøres $n$ bremseprøver med ulike hastigheter, men forøvrig under identiske forsøksbetingelser. La $x_i$ betegne hastigheten benyttet ved bremseprøve nummer $i$, og la $Y_i$ være tilhørende målt bremselengde. Vi skal anta at bremseprøvene utføres på en slik måte at det er rimelig å betrakte $Y_1, Y_2, \ldots, Y_n$ som uavhengige stokastiske variabler.

Til å estimere $\beta$ definerer vi estimatorene

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i^2} \quad \text{og} \quad \widetilde{\beta} = \frac{\sum_{i=1}^{n} Y_i x_i^2}{\sum_{i=1}^{n} x_i^4}.$$

Det oppgis at

$$\mathrm{E}\left[\widehat{\beta}\right] = \beta \quad \text{og} \quad \mathrm{Var}\left[\widehat{\beta}\right] = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}.$$

Anta at det gjøres $n = 20$ bremseprøver og at disse resulterte i verdiene angitt i tabell 2. Det oppgis at $\bar{y} = \frac{1}{20}\sum_{i=1}^{20} y_i = 21.022$, $\sum_{i=1}^{20} x_i^2 = 82750$, $\sum_{i=1}^{20} x_i^4 = 573216250$, $\sum_{i=1}^{20} x_i^6 = 4.7222 \cdot 10^{12}$ og $\sum_{i=1}^{20}(y_i/x_i)^2 = 2.1656$.

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 |
| $y_i$ | 0.47 | 0.99 | 3.96 | 3.56 | 4.00 | 5.46 | 7.47 | 10.33 | 13.53 | 11.93 |

| $i$ | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 | 100 | 105 |
| $y_i$ | 15.85 | 19.64 | 19.06 | 31.90 | 34.01 | 36.09 | 41.83 | 46.48 | 55.36 | 58.52 |

Tabell 2: Hastigheter, $x_i$, benyttet under bremseprøvene og tilhørende målte bremselengder, $y_i$.

**a)** Finn forventningsverdi og varians for $\widetilde{\beta}$.

Hvilken av de to estimatorene $\widehat{\beta}$ og $\widetilde{\beta}$ vil du foretrekke når bremseprøvene er som beskrevet over? Begrunn svaret.

Modellen for $Y_i$ spesifisert over kan alternativt formuleres som at

$$Y_i = \beta x_i^2 + x_i \varepsilon_i \quad \text{for } i = 1, \ldots, n$$

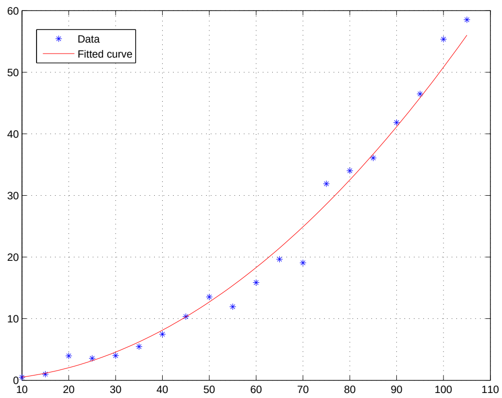
der $\varepsilon_1, \ldots, \varepsilon_n$ er et tilfeldig utvalg fra en normalfordeling med forventningsverdi lik 0 og standardavvik lik $\sigma$. Dersom vi benytter $\widehat{\beta}$ som estimator for $\beta$ kan vi benytte

$$\widehat{\varepsilon}_i = \frac{y_i - \widehat{\beta} x_i^2}{x_i}.$$

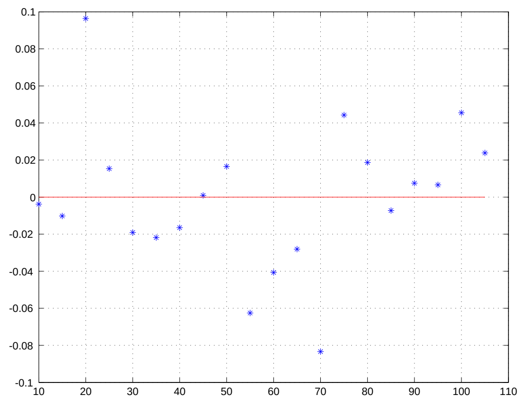
som et anslag for $\varepsilon_i$. I figur 1(a) er de observerte data plottet sammen med estimert regresjonskurve $\widehat{y} = \widehat{\beta} x^2$, i figur 1(b) er punktene $(x_i, \widehat{\varepsilon}_i), i = 1, \ldots, n$ plottet og i figur 1(c) vises et normalsannsynlighetsplott (normalkvantil-kvantilplott, QQ-plott) for $\widehat{\varepsilon}_i, i = 1, 2, \ldots, n$.

**b)** Beskriv kort hva et normalsannsynlighetsplott (normalkvantil-kvantilplott, QQ-plott) generelt kan brukes til og hvordan man skal tolke plottet.

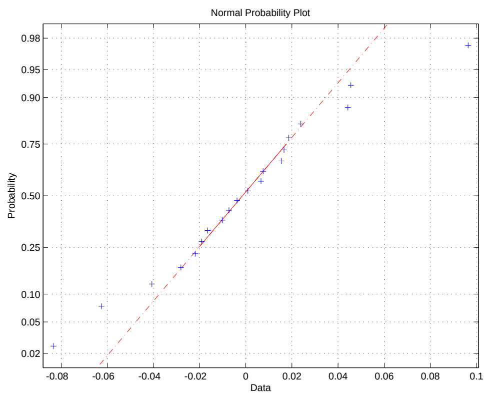
Diskuter med utgangspunkt i plottene i figur 1 hvorvidt de observerte verdiene ser ut til å passe med den spesifiserte modellen.




(a)




(b)




(c)

Figur 1: Plott av (a) data og estimert regresjonskurve, (b) residualer $\widehat{\varepsilon}_i, i = 1, 2, \ldots, 20$, og (b) normalsannsynlighetsplott (normalkvantil-kvantilplott, QQ-plott) for residualene.

Uavhengig av hva du svarte i punktene over skal du videre i denne oppgaven benytte $\widehat{\beta}$ som estimator for $\beta$ og forutsette at modellen angitt i innledningen til oppgaven er korrekt.

**c)** Anta i dette punktet at verdien til $\sigma^2$ er kjent.

Utled et $(1-\alpha)\cdot 100\%$-prediksjonsintervall for en ny bremselengde $Y_0$ ved hastighet $x_0$ når man har tilgjengelig observasjoner $(x_i, Y_i), i = 1, 2, \ldots, n$ som beskrevet over.

Hvis du fikk velge hvilke hastigheter $x_1, x_2, \ldots, x_n$ bremseprøvene skulle utføres for, hvordan ville du velge disse for at prediksjonsintervallet skulle bli kortest mulig? Kommenter.

I neste punkt i oppgaven skal vi forutsette at begge parametrene $\beta$ og $\sigma^2$ er ukjente. Du skal fremdeles benytte $\widehat{\beta}$ som estimator for $\beta$, og for å estimere $\sigma^2$ skal du benytte

$$\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{Y_i - \widehat{\beta}x_i^2}{x_i}\right)^2.$$

Det oppgis at $(n-1)\widehat{\sigma}^2/\sigma^2$ er $\chi^2$-fordelt med $n-1$ frihetsgrader, samt at $\widehat{\beta}$ og $\widehat{\sigma}^2$ er uavhengige.

Anta at det er kjent at en gammel dekktype har $\beta = 0.0053$ under spesifiserte forsøksbetingelser. Nå er det utviklet en ny dekktype og produsenten påstår at den nye dekktypen har kortere bremselengder enn den gamle dekktypen under de samme forsøksbetingelsene. For å vurdere denne påstanden gjøres $n$ bremseprøver med den nye dekktypen der vi lar $x_1, x_2, \ldots, x_n$ angi hastighetene som benyttes og $Y_1, Y_2, \ldots, Y_n$ tilhørende målte bremselengder.

**d)** Formuler spørsmålet angitt over som et hypotesetestingsproblem. Spesifiser $H_0$ og $H_1$, angi testobservator og vis hvilken sannsynlighetsfordeling denne har under $H_0$. Bestem forkastningskriteriet når signifikansnivået er $\alpha$.

Hva blir konklusjonen på hypotesetesten når bremseprøvene for den nye dekktypen er som angitt i tabell 2 og $\alpha = 0.05$. (Husk at verdier til en del summer for dette datasettet er oppgitt i innledningen til oppgaven.)