NTNU

Kunnskap for ei betre verd

Institutt for matematiske fag

Eksamensoppgåve i **TMA4245 Statistikk**

Fagleg kontakt under eksamen: Håkon Tjelmeland

Tlf: 48 22 18 96

Eksamensdato: ??. august 2014

Eksamenstid (frå–til): 09:00–13:00

Hjelpemiddelkode/Tillatne hjelpemiddel: 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.
- Eit stempla gult A5-ark med eigne handskrivne formlar og notat.

Annan informasjon:
Du skal gi grunn for all svar og inkludere naturleg mellomrekning.

Målform/språk: nynorsk

Sidetal: 6

Sidetal vedlegg: 0

Kontrollert av:

Dato Sign

Oppgåve 1 Medisinmengder

Ei bedrift produserer ein type medisin i pulverform. Medisinen seljast på flasker som kvar er meint å ta 500 gram pulver. Bedrifta nytter ein tappemaskin som porsjonerer ut pulveret i riktige doser. Når tappemaskinen stillast inn på μ gram, kan ein gå ut frå at dei utporsjonerte pulvermengdene $X_1, X_2, \ldots$, er uavhengige og normalfordelte med forventingsverdi μ gram og standardavvik $\sigma = 12$ gram. Ein kan vidare gå ut frå at vekta av flaskene (uten pulver) som pulveret fylles i, $Y_1, Y_2, \ldots$, er uavhengige og normalfordelte med forventingsverdi 50 gram og standardavvik 2 gram. Ein kan óg gå ut frå at pulvervekt og flaskevekt er uavhengige av kvarandre.

I punkt a) og b) i denne oppgåva er μ satt lik 510 gram.

a) Kva er sannsynet for at ei tom flaske veier meir enn 53 gram?

Forklar kvifor vekta av ei flaske fylt med pulver er normalfordelt med forventingsverdi lik 560 gram og standardavvik $\sqrt{148}$ gram.

Kva er sannsynet for at vektforskjellen mellom to flasker fylt med pulver skal vere større enn 15 gram?

Ei flaske fylt med pulver som veier mindre enn 540 gram seies å vere undervektig. Kundar som kjem til å kjøpe ei eller fleire undervektige flasker, har rett til å returnere desse. Flaskene pakkes i kartongar med 24 flasker i kvar kartong.

b) Kva er sannsynet for at ei flaske fylt med pulver skal vere undervektig?

La U vere talet på undervektige flasker i ein tilfeldig valt kartong. Kva sannsynsfordeling har U? Gi grunn for svaret.

Kva er sannsynet for at ein tilfeldig valt kartong skal ha ei eller fleire undervektige flasker?

Det er dårleg reklame for bedrifta at for mange flasker er undervektige. Bedriftsleiinga forlanger difor at høgst 1% av flaskene i det lange løp skal vere undervektige.

c) Korleis må då maskina stillast inn for at dette kravet skal vere oppfylt?

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

i	1	2	3	4	5	6	7	8	9	10
μ_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: Verdiar nytta for $\mu_1, \mu_2, \ldots, \mu_{10}$ og målte utporsjonerte kvanta $v_1, v_2, \ldots, v_{10}$.

Ein kunde har kjøpt n flasker med pulver og ønsker på det grunnlag å anslå μ. Dei andre parametrane har verdiar som gitt tidlegare i oppgåva. Kunden veier pulveret i kvar flaske, $X_1, X_2, \ldots, X_n$, og nytter estimatoren

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

d) Utlei eit $100(1-\alpha)\%$ konfidensintervall for μ basert på pulvervektene i kvar flaske, $X_1, X_2, \ldots, X_n$.

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

Kor stor måtte n ha vore for at lengda på dette 95%-konfidensintervallet skulle blitt kortare enn 8 gram?

Bedrifta vil kontrollere at porsjoneringsmaskina er riktig justert. Når maskina stilles inn på μ gram kan ein gå ut frå at dei utporsjonerte kvanta med pulver er uavhengige og normalfordelte med forventingsverdi lik $\beta\mu$ og standardavvik lik 12, der $\beta = 1$ dersom maskina er riktig justert. For å undersøke om maskina er riktig justert utføres følgande forsøk: Maskina stilles inn på i alt k verdiar $\mu_1, \mu_2, \ldots, \mu_k$. Til kvar innstilling μ_i svarer då ein stokastisk variabel V_i, som representerer det utporsjonerte kvantum. Når forsøket er gjennomført, har ein 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 sannsynsmaksimeringsestimatoren (SME) for β.

Gjennomfør estimeringa når $k = 10$ og resultatet av forsøket blei som gitt i tabell 1. Det gis at $\sum_{i=1}^{k} v_i\mu_i = 2440526$ og $\sum_{i=1}^{k} \mu_i^2 = 2462800$.

Oppgåve 2

La X og Y vere diskret fordelte stokastiske variablar der $X, Y \in \{0, 1, 2\}$. La $f(x, y) = \mathrm{P}(X = x, Y = y)$ vere simultan punktsannsyn for X og Y og gå ut frå at $f(x, y)$ er som angitt i følgjande 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) punktsannsyn for X og for Y.

Er X og Y uavhengige? Gi grunn for svaret!

Oppgåve 3 Bremselengder

Vi skal i denne oppgåva gå ut frå at bremselengda, Y, målt i meter for ein bil som køyrer x km/time er normalfordelt med forventingsverdi βx^2 og standardavvik σx. Ein bil som for eksempel køyrer i 50 km/time vil dermed ha ei bremselengd som er normalfordelt med forventingsverdi 2500β og standardavvik 50σ. Modellen har to parametrar, β og σ^2, og desse vil avhenge av forsøksopplegget, som for eksempel dekkenes eigenskapar, veidekke og vêr- og føreforhold.

Gå ut frå at verdien til β er ukjent og skal estimerast. For å estimere β utførast n bremseprøver med ulik fart, men elles under identisk forsøkopplegg. La x_i vere farten nytta ved bremseprøve nummer i, og la Y_i vere tilhøyrande målt bremselengd. Vi skal gå ut frå at bremseprøvene utførast på ein slik måte at det er rimeleg å se på $Y_1, Y_2, \ldots, Y_n$ som uavhengige stokastiske variablar.

For å estimere β definerer vi estimatorane

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

Det oppgis at

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

Gå ut frå at ein gjer $n = 20$ bremseprøver og at desse resulterte i verdiane gitt 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: Fart, x_i, nytta under bremseprøvene og tilhøyrande målte bremselengder, y_i.

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

Kva for ein av dei to estimatorane $\widehat{\beta}$ og $\widetilde{\beta}$ vil du føretrekkje når bremseprøvene er som skildra over? Gi grunn for svaret.

Modellen for Y_i spesifisert over kan alternativt formulerast som at

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

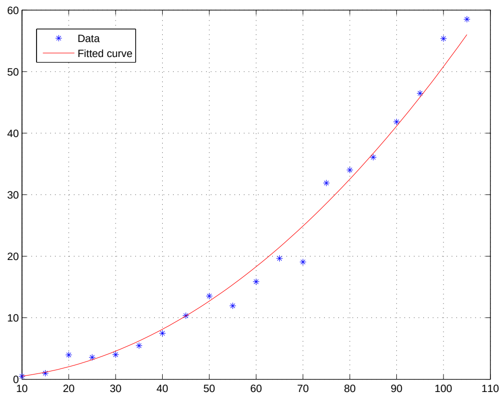
der $\varepsilon_1, \ldots, \varepsilon_n$ er eit tilfeldig utval frå ei normalfordeling med forventingsverdi lik 0 og standardavvik lik σ. Dersom vi nytter $\widehat{\beta}$ som estimator for β kan vi nytte

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

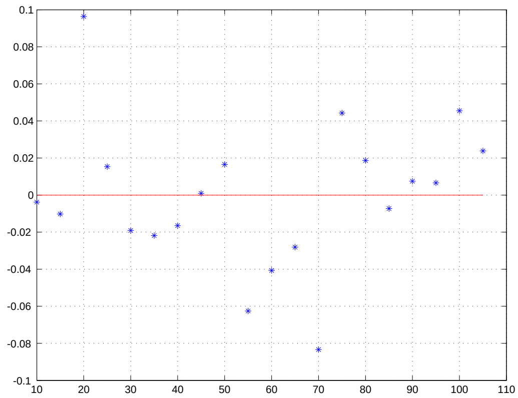
som eit anslag for ε_i. I figur 1(a) er dei observerte data plotta saman med estimert regresjonskurve $\widehat{y} = \widehat{\beta} x^2$, i figur 1(b) er punkta $(x_i, \widehat{\varepsilon}_i), i = 1, \ldots, n$ plotta og i figur 1(c) visast eit normalsannsynsplott (normalkvantil-kvantilplott, QQ-plott) for $\widehat{\varepsilon}_i, i = 1, 2, \ldots, n$.

b) Beskriv kort kva eit normalsannsynsplott (normalkvantil-kvantilplott, QQ-plott) generelt kan brukast til og korleis ein skal tolke plottet.

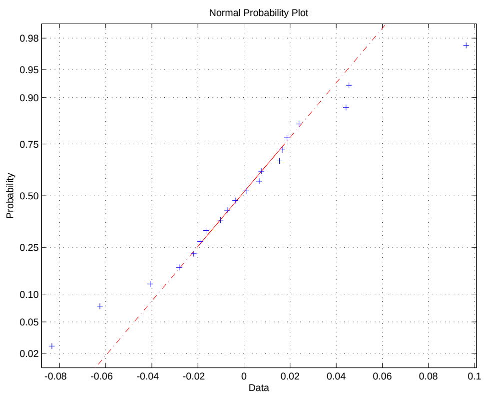
Diskuter med utgangspunkt i plotta i figur 1 om de observerte verdiane ser ut til å passe med den spesifiserte modellen.



(a)



(b)



(c)

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

Uavhengig av kva du svarte i punkta over skal du vidare i denne oppgåva nytte $\widehat{\beta}$ som estimator for β og gå ut frå at modellen gitt i innleiinga til oppgåva er korrekt.

c) Gå i dette punktet ut frå at verdien til σ^2 er kjent.

Utlei eit $(1-\alpha)\cdot 100\%$-prediksjonsintervall for en ny bremselengd Y_0 ved fart x_0 når man har tilgjengeleg observasjonar $(x_i, Y_i), i = 1, 2, \ldots, n$ som skildra over.

Hvis du fekk velje kva for fart $x_1, x_2, \ldots, x_n$ bremseprøvene skulle utførast for, korleis ville du velje desse for at prediksjonsintervallet skulle bli så kort som mogleg? Kommenter.

I neste punkt i oppgåva skal vi gå ut frå at begge parametrane β og σ^2 er ukjente. Du skal fremleis nytte $\widehat{\beta}$ som estimator for β, og for å estimere σ^2 skal du nytte

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

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

Gå ut frå at det er kjent at ein gamal dekktype har $\beta = 0.0053$ under eit spesifiserte forsøksopplegg. Nå er det utvikla ein ny dekktype og produsenten påstår at den nye dekktypen har kortere bremselengder enn den gamle dekktypen under same forsøkopplegg. For å vurdere denne påstanden gjer ein n bremseprøver med den nye dekktypen der vi lar $x_1, x_2, \ldots, x_n$ vere fart som nyttes og $Y_1, Y_2, \ldots, Y_n$ tilhøyrande målte bremselengder.

d) Formuler spørsmålet gitt over som eit hypotesetestingsproblem. Spesifiser H_0 og H_1, angi testobservator og vis kva for ein sannsynsfordeling denne har under H_0. Bestem forkastningskriteriet når signifikansnivået er α.

Kva blir konklusjonen på hypotesetesten når bremseprøvene for den nye dekktypen er som gitt i tabell 2 og $\alpha = 0.05$. (Hugs at verdiar til ein del summar for dette datasettet er gitt i innleiinga til oppgåva.)