Institutt for matematiske fag

Eksamensoppgåve i **TMA4240 / TMA4245 Statistikk**

Fagleg kontakt under eksamen: Jo Eidsvik

Tlf: 901 27 472

Eksamensdato: 15. august 2016

Eksamenstid (frå–til): 09.00-13.00

Hjelpemiddelkode/Tillatne hjelpemiddel: C: *Tabeller og formler i statistikk* (Tapir forlag, Fagbokforlaget), *Matematisk formelsamling* (K. Rottmann), eit stempla gult A5-ark med egne handskrivne notater, bestemt enkel kalkulator: HP30S, Citizen SR-270X, Citizen SR-270X College eller Casio fx-82ES PLUS.

Annan informasjon:
Alle svar skal grunngjevast, og svara skal innehalde naturlege mellomrekninger.

Målform/språk: nynorsk

Sidetal: 4

Sidetal vedlegg: 0

Kontrollert av:

Dato Sign

Oppgåve 1 Terningspel

Sjå på kast av ein vanlig terning. La X_i vere ein stokastisk variabel som syner talet på auge i kast i, der $x_i \in \{1, \ldots, 6\}$, og $P(X_i = l) = \frac{1}{6}$, $l = 1, \ldots, 6$. Det er oppgitt at $E(X_i) = 3.5$ og $\text{Var}(X_i) = 1.71^2$.

a) Vi ser på summen av to uavhengige terningkast: $Y_2 = X_1 + X_2$.

Rekn ut sannsynet for at summen er 12.

Rekn ut sannsynet for at summen er større eller lik 10.

Rekn ut den forventa summen.

I eit spel kaster ein terningen gjenteke gonger, og ein forsøkjer å oppnå høg score. La score etter kast k vere Z_k. Før kast $k+1$ vel ein sjølv å kaste ein gong til, eller om spelet skal avsluttast med sin noverande score Z_k. Ein starter spelet med score $Z_0 = 0$ ved $k = 0$. Dersom kast $k+1$ vert $x_{k+1} \in \{1, \ldots, 5\}$, så leggjast talet til scoren ($Z_{k+1} = Z_k + X_{k+1}$). Dersom kastet vert $x_{k+1} = 6$, så set ein scoren $Z_{k+1} = 0$, og spelet avsluttast.

b) Rekn ut forventa score etter første kast.

Dersom ein spelar har score 30, og vel å fortsetje og kaste, kva er forventa score etter neste kast?

Ein speler bruker optimal forventningsverdi til å velje mellom *i*) fortsetje og kaste eller *ii*) stoppe med noverande score. For kva score er det optimalt å fortsetje og kaste terning?

Oppgåve 2 Hotellpris

Harald vurderer hotellferie i Barcelona på seinsumaren. Ein kan antake at prisen på eit hotellrom kan beskrivast ved ei normalfordeling med forventning μ Euro per natt og standardavvik 20 Euro. Ein kan og antake ein kronekurs på 9.2 kroner per Euro.

a) I dette punktet er $\mu = 100$.

Kva er fordelinga til prisen, i kroner, på eit hotellrom per natt?

Kva er sannsynet for at eit hotellrom kostar meir enn 1000 kroner?

Kva er sannsynet for at eit hotellrom kostar meir enn 1100 kroner, gjeve at det kostar meir enn 1000 kroner?

Harald meiner forventa pris er 100 Euro, medan ein ven seier at det har vorte dyrare. Dei er einige om at standardavviket er kjend og lik 20 Euro. Harald samler inn data for å undersøkje om forventa pris har vorte høgare enn 100 Euro. Han ringjer rundt på staden og samlar inn prisdata, $x_1, \ldots, x_{20}$, for 20 hotellrom. Han får at $\bar{x} = 120$ Euro.

b) Formuler undersøkinga som ein hypotesetest.

Nytt antakinga om normalfordelte data og tala ovanfor til å gjennomføre hypotesetesten på signifikansnivå 0.05.

c) Vi antek så at sann forventning er lik 110 Euro.

Rekn ut teststyrken.

Kor mange data måtte Harald ha samla inn for å få ein teststyrke på 0.95?

Oppgåve 3 Medisinkonsentrasjon

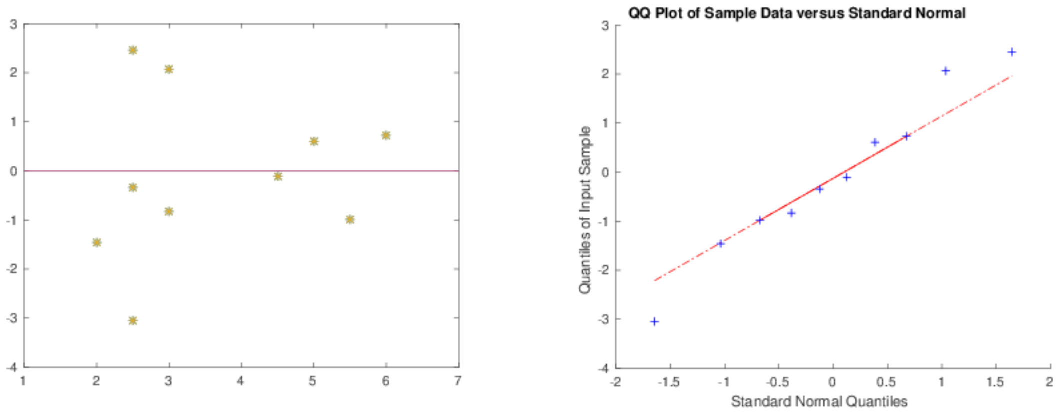
Vi skal i denne oppgåva betrakte ein sykdom der behandlinga består i at pasienten får injisert ein medisin i blodet. La x vere dosen ein pasient får. Vi antek at denne dosen kan kontrollerast slik at vi ikkje betraktar x som stokastisk. Eit døgn etter at ein har injisert medisinen måler ein konsentrasjonen, Y, av medisinen i blodet. Vi kan antake følgjande lineære regresjonsmodell for sammenheng mellom x og Y,

$$Y = bx + \varepsilon,$$

der b er ein ukjend konstant og ε er ein normalfordelt stokastisk variabel med forventningsverdi 0 og varians σ^2. I heile denne oppgåva antek vi at $\sigma^2 = 2.0^2$ er kjend.

Vi har observerte verdier for $n = 10$ pasientar. La x_i og Y_i for $i = 1, 2, \ldots, n$ vere tilhøyrande verdier av injisert dose og målt konsentrasjon i blodet. Ein kan antake at $Y_1, Y_2, \ldots, Y_n$ er uavhengige stokastiske variablar og at regresjonsmodellen gjeve ovanfor gjeld for kvar av dei. Observerte verdier for dei n pasientane er:

pasient i	1	2	3	4	5	6	7	8	9	10
x_i	2.0	5.0	2.5	2.5	4.5	5.5	3.0	6.0	3.0	2.5
y_i	2.9	11.5	5.1	7.9	9.7	11.0	8.6	13.8	5.7	2.4



Figur 1: Til venstre: Residualplott. Dose x_i er plottet langs x-aksen og estimert verdi for residual ε_i er plotta langs y-aksen. Til høgre: Normalsannsynplott (normalkvantil-kvantilplott, QQ-plott) for estimerte residualer.

Det vert oppgjeve at $\sum_{i=1}^n x_i = 36.5$, $\sum_{i=1}^n x_i^2 = 152.25$ og $\sum_{i=1}^n x_i y_i = 331.65$.

For modellen og data over viser figur 1 residualplott og normalsannsynplott (normalkvantil-kvantilplott, QQ-plott) for dei estimerte residuala.

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

Med utgangspunkt i plotta i figur 1, diskuter om dei observerte verdiane ser ut til å passe med den spesifiserte modellen.

For å estimere parameteren b vert det foreslått tre estimatorer

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

Det vert oppgjeve at $\widehat{\widehat{b}}$ er forventningsrett og at $\mathrm{Var}\left[\widehat{\widehat{b}}\right] = \sigma^2 / \sum_{i=1}^n x_i^2$.

b) Kva for ein av dei tre estimatorane vil du føretrekkje når $n = 10$ og observasjonene er som over? Grunngje svaret.

c) Skriv opp rimelighetsfunksjonen (likelihood function) for b for situasjonen over. Bruk så rimelighetsfunksjonen til å utleie sannsynmaksimeringsestimatoren (SME) for b.

I resten av denne oppgåva skal du, uavhengig av dine resultat i punkt **b**) og **c**), ta utgangspunkt i estimatoren $\hat{\hat{b}}$ gjeve over.

d) Grunngje at $\hat{\hat{b}}$ er normalfordelt.

Utlei eit $(1-\alpha)\cdot 100\%$ konfidensinterval for b uttrykt ved n, $x_1, x_2, \ldots, x_n$, $Y_1, Y_2, \ldots, Y_n$, σ^2 og α.

Berekn intervalet numerisk når observasjonene er som angjeve tidligare og $\alpha = 0.10$.

Det er viktig at konsentrasjonen av medisin i blodet ikkje blir for høg då dette kan gjere at pasienten får alvorlege bivirknadar av medisinen. Etter å ha observert resultata for dei første $n = 10$ pasientene (angjeve over), får legane inn ein ny pasient og etter å ha undersøkt denne pasienten vurderer legane at det er viktig at målt medisinkonsentrasjon i blodet for denne pasienten ikkje overstig 10.0.

e) Bestem den høgaste dosen x_0 den nye pasienten kan få injisert om ein krev eit sannsyn på minst 0.95 for at målt medisinkonsentrasjon i blodet etter eit døgn ikkje overstig 10.0.