Institutt for matematiske fag

Eksamensoppgave i **TMA4240 / TMA4245 Statistikk**

Faglig kontakt under eksamen: Jo Eidsvik

Tlf: 901 27 472

Eksamensdato: 15. august 2016

Eksamenstid (fra–til): 09.00-13.00

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

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

Målform/språk: bokmål

Antall sider: 4

Antall sider vedlegg: 0

Kontrollert av:

Dato Sign

Oppgave 1 Terningspill

Anta at en vanlig terning kastes. La X_i være stokastisk variabel som viser antall øyne i kast i, der utfallet $x_i \in \{1, \ldots, 6\}$, og $P(X_i = l) = \frac{1}{6}$, $l = 1, \ldots, 6$. Dette betyr at $E(X_i) = 3.5$ og $\mathrm{Var}(X_i) = 1.71^2$.

a) Vi ser på summen av 2 uavhengige terningkast: $Y_2 = X_1 + X_2$.

Regn ut sannsynligheten for at summen er 12.

Regn ut sannsynligheten for at summen er større eller lik 10.

Regn ut den forventede summen.

I et spill kastes terningen gjentatte ganger, og man forsøker å oppnå høy score. La score etter kast k være Z_k. Før kast $k+1$ velger man selv om det skal kastes et kast til, eller om spillet skal avsluttes og da beholder man sin nåværende score Z_k. Man starter spillet med score $Z_0 = 0$ ved $k = 0$. Dersom utfallet av kast $k+1$ er $x_{k+1} \in \{1, \ldots, 5\}$, så adderes utfallet til scoren ($Z_{k+1} = Z_k + X_{k+1}$). Dersom utfallet blir $x_{k+1} = 6$, så settes scoren $Z_{k+1} = 0$, og spillet avsluttes.

b) Regn ut forventet score etter første kast.

Dersom en spiller har score 30, og velger å fortsette kasting, hva er forventet score etter neste kast?

En spiller bruker optimal forventningsverdi til å ta valget om å *i*) fortsette kasting eller *ii*) stoppe med nåværende score. Ved hvilke score er det optimalt å fortsette kast av terning?

Oppgave 2 Hotellpris

Harald vurderer hotellferie i Barcelona på sensommeren. Anta at prisen på et hotellrom kan beskrives ved en normalfordeling med forventning μ Euro per natt og standardavvik 20 Euro. Anta en kronekurs på 9.2 kroner per Euro.

a) Anta i dette punktet at $\mu = 100$.

Hva er fordelingen til prisen, i kroner, på et hotellrom per natt?

Hva er sannsynligheten for at et hotellrom koster mer enn 1000 kroner?

Hva er sannsynlighteten for at et hotellrom koster mer enn 1100 kroner, gitt at det koster mer enn 1000 kroner?

Harald mener forventet pris er 100 Euro, mens en venn sier at det har blitt dyrere. De er enige om at standardavviket er kjent og lik 20 Euro. Harald samler inn data for å undersøke om forventet pris har blitt høyere enn 100 Euro. Han ringer rundt på stedet og samler inn prisdata, $x_1, \ldots, x_{20}$, for 20 hotellrom. Det oppgis at $\bar{x} = 120$ Euro.

b) Formuler undersøkelsen som en hypotesetest.

Bruk antakelsen om normalfordelte data og de oppgitte tallverdiene til å gjennomføre hypotesetesten på signifikansnivå 0.05.

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

Regn ut teststyrken.

Hvor mange data måtte Harald ha samlet inn for å få en teststyrke på 0.95?

Oppgave 3 Medisinkonsentrasjon

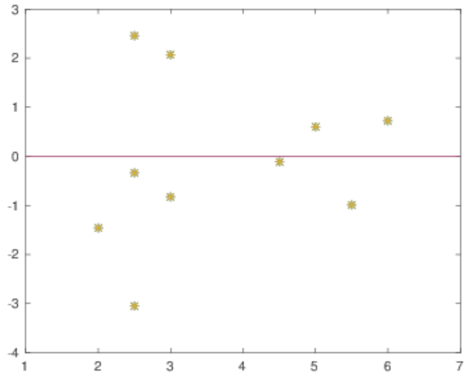
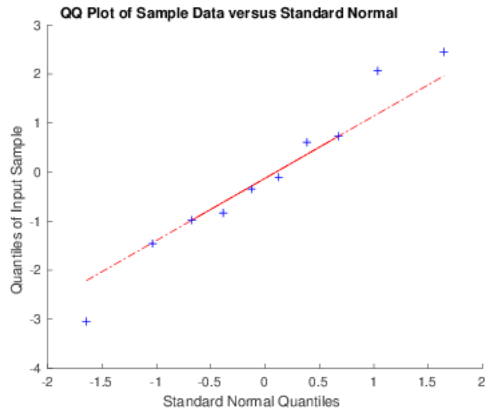
Vi skal i denne oppgaven betrakte en sykdom hvor behandlingen består i at pasienten får injisert en medisin i blodet. La x være dosen en pasient får. Vi skal anta at denne dosen kan kontrolleres slik at vi ikke betrakter x som stokastisk. Et døgn etter at man har injisert medisinen måler man konsentrasjonen, Y, av medisinen i blodet. Vi skal anta følgende lineære regresjonsmodell for sammenhengen mellom x og Y,

$$Y = bx + \varepsilon,$$

der b er en ukjent konstant og ε er en normalfordelt stokastisk variabel med forventningsverdi 0 og varians σ^2. Vi skal i hele denne oppgaven anta at $\sigma^2 = 2.0^2$ er kjent.

Vi skal anta at vi har observerte verdier for $n = 10$ pasienter og la x_i og Y_i for $i = 1, 2, \ldots, n$ være tilhørende verdier av injisert dose og målt konsentrasjon i blodet. Det antas at $Y_1, Y_2, \ldots, Y_n$ er uavhengige stokastiske variabler og at regresjonsmodellen beskrevet over gjelder for hver av dem. Observerte verdier for de n pasientene 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 residualet ε_i er plottet langs y-aksen. Til høyre: Normalsannsynlighetsplott (normalkvantil-kvantilplott, QQ-plott) for estimerte residualer.

Det oppgis 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 dataene gitt over viser figur 1 residualplott og normalsannsynlighetsplott (normalkvantil-kvantilplott, QQ-plott) for de estimerte residualene.

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

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

For å estimere parameteren b er 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 oppgis 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) Hvilken av de tre estimatorene vil du foretrekke når $n = 10$ og observasjonene er som gitt over? Begrunn svaret.

c) Skriv opp rimelighetsfunksjonen (likelihood function) for b for situasjonen beskrevet over. Benytt så rimelighetsfunksjonen til å utlede sannsynlighetsmaksimeringsestimatoren (SME) for b.

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

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

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

Beregn intervallet numerisk når observasjonene er som angitt ovenfor og $\alpha = 0.10$.

Det er viktig at konsentrasjonen av medisin i blodet ikke blir for høy da dette kan gjøre at pasienten får alvorlige bivirkninger av medisinen. Etter å ha observert resultatene for de første $n = 10$ pasientene (angitt over), får legene inn en ny pasient og etter å ha undersøkt denne pasienten vurderer legene at det er viktig at målt medisinkonsentrasjon i blodet for denne pasienten ikke overstiger 10.0.

e) Bestem den høyeste dosen x_0 den nye pasienten kan få injisert hvis det kreves en sannsynlighet på minst 0.95 for at målt medisinkonsentrasjon i blodet etter et døgn ikke overstiger 10.0.