Institutt for matematiske fag

Eksamensoppgave i TMA4240 Statistikk

Faglig kontakt under eksamen: John Tyssedal, Haakon Bakka.

Tlf.: John Tyssedal: 41645376. Tlf: Haakon Bakka: 97955667.

Eksamensdato: 17.12.2014

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 (uten forside): 4

Antall sider vedlegg: 0

Kontrollert av:

Dato Sign

Oppgave 1

I en fotballsesong i tippeligaen skal hvert lag spille 30 kamper, 15 hjemmekamper og 15 bortekamper. Ved starten av sesongen regner et av fotballagene med at sannsynlighetene for seier (S), uavgjort (U) eller tap (T) i en hjemmekamp er gitt som P(S) = 0.6, P(U) = 0.2 og P(T) = 0.2. I en bortekamp regner de med at sannsynlighetene er P(S) = 0.4, P(U) = 0.2 og P(T) = 0.4. Anta videre at resultatet i hver kamp er uavhengig av resultatene i de andre kampene.

a) Lagene får 3 poeng for seier, 1 poeng for uavgjort og 0 poeng for tap. La X være tallet på poeng laget får i en hjemmekamp og la Y være tallet på poeng det får i en bortekamp. Vis at forventning og varians til X er gitt ved 2 og 1.6 og at forventning og varians til Y er gitt ved 1.4 og 1.84.

La $X_i,\ i = 1,2,\cdots,15$ være poengene laget får i de 15 hjemmekampene og $Y_i,\ i = 1,2,\cdots,15$ være poengene de får i de 15 bortekampene. La videre $X_H = \sum_{i=1}^{15} X_i$ og $Y_B = \sum_{i=1}^{15} Y_i$. Laget regner videre med at dersom de greier å få minst 60 poeng så er det godt nok til medalje.

b) Finn forventning og varians til X_H og Y_B. Anta deretter at X_H og Y_B er tilnærmet normalfordelte og finn tilnærmet sannsynligheten for at laget greier å få minst 60 poeng.

Det viser seg at poengsankingen går trått og da det er 8 kamper igjen, 4 hjemmekamper og 4 bortekamper, regner de med at de må vinne minst 7 av de 8 siste kampene for å redde plassen i serien. De bestemmer seg for å gjøre en sannsynlighetsvurdering der de bruker sannsynlighetene for seier gitt i innledningen av oppgaven. Dersom sannsynligheten for å redde plassen i serien blir mindre enn 0.05 bestemmer de seg for å sparke treneren.

c) La N_H og N_B være tallet på hjemme- og bortekamper som laget vinner i de 8 siste kampene. Hvilken fordeling har N_H og N_B? Blir treneren sparket?

Oppgave 2

Ingolv har alltid vært fasinert av vulkaner og vulkanutbrudd. Etter videregående bestemmer han seg derfor for å dra til Island og ta en 5-årig master der. La X_t være tallet på vulkanutbrudd på Island i et tidsintervall av lengde t. Ut i fra historiske data viser det seg at det er rimelig å gå ut i fra at X_t er poissonfordelt med parameter λt, der $\lambda = 0.3$ og t er tallet på år. Det vil si at $P(X_t = x) = \frac{(0.3t)^x e^{-0.3t}}{x!},\ x = 0,1,2,\cdots$.

a) Hva er sannsynligheten for at Ingolv skal få oppleve minst et vulkanutbrudd i en 5 års periode?

Anta at det skjedde et vulkanutbrudd nøyaktig et år etter at Ingolv kom til Island. Hva er da sannsynligheten for at han får oppleve minst et til i det femårige oppholdet sitt?

b) La T være tiden i år fra Ingolv kommer til Island til det har skjedd to vulkanutbrudd. Hvilken fordeling har T?
Forklar hvorfor $P(T \le t) = 1 - P(X_t < 2)$. Bruk dette til å finne ut tilnærmet hvor lenge Ingolv må oppholde seg på Island for at sannsynligheten for å få med seg 2 vulkanutbrudd er minst 0.80.

Oppgave 3

To El-bil produsenter konkurrerer om det samme kundemarkedet med hver sine modelltyper. La oss kalle de to modelltypene for modell A og modell B. En viktig faktor for valg av modelltype er rekkevidden som bilen har før en må lade batteriet på nytt. Produsenten av modell B beskylder produsenten av modell A for å oppgi for lang rekkevidde. De hevder at rekkevidden som er oppgitt ikke er realistisk å oppnå under normale kjøreforhold. For å undersøke dette nærmere bestemmer produsenten av modell A seg for å gjøre et lite forsøk. 10 sjåfører får i oppdrag å kjøre hver sin bil i noe som ble definert som normalterreng og med en snittfart på 60 km/t. La X_i være rekkevidden for sjåfør i, $i = 1,2,\cdots,10$. De observerte rekkeviddene i km er gitt nedenfor:

x_i: 123.0, 119.3, 119.4, 118.2, 119.5, 118.5, 117.7, 118.1, 119.6, 119.4

La $E(X_i) = \mu_A$ og $Var(X_i) = \sigma^2$, $i = 1,2,\cdots,10$

a) Sett opp forventningsrette estimatorer for μ_A og σ^2 basert på hele utvalget. Hva blir estimatene når $\sum_{i=1}^{10} x_i = 1192.70$ og $\sum_{i=1}^{10} (x_i - \bar{x})^2 = 19.68$?

Produsenten av modell A oppgir at forventet rekkevidde under slike forhold er 125 km. Anta nå at $X_i \sim N(\mu_A, \sigma^2)$ og uavhengige, $i = 1,2,\cdots,10$.

b) Kan en ut i fra dataene konkludere med at produsenten har satt forventet rekkevidde for høyt? Formuler dette som ett hypotesetestingsproblem og utfør testen. Bruk signifikansnivå $\alpha = 0.01$.

I en reklame for modell A står det at bilen har lenger rekkevidde enn modell B. Produsenten av modell B bestemmer seg derfor for å gjennomføre samme forsøk som produsenten av modell A gjennomførte i samme terreng og med samme snittfart. La Y_i være rekkevidden for sjåfør i, $i = 1,2,\cdots,10$.

De observerte rekkeviddene i km er gitt nedenfor.

y_i: 109.9, 110.4, 109.7, 111.5, 112.3, 111.8, 112.1, 108.2, 109.9, 116.1

$$\sum_{i=1}^{10} y_i = 1111.9 \quad \text{og} \quad \sum_{i=1}^{10} (y_i - \bar{y})^2 = 41.75$$

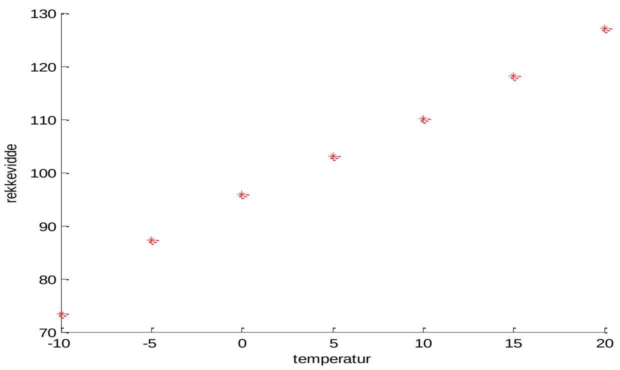
Anta at $Y_i \sim N(\mu_B, \sigma^2)$. Variansen er altså lik i de to utvalgene, men forventningene kan være forskjellige. Anta også at $X_1, \cdots, X_{10}, Y_1, \cdots, Y_{10}$ alle er uavhengige.

c) Sett opp en forventningsrett estimator for σ^2 basert på de 2 utvalgene. Lag et 95 % konfidensintervall for $\mu_A - \mu_B$ basert på de observerte dataene. Tyder intervallet på at det som står i reklamen for modell A kan være riktig? Grunngi svaret ved å trekke inn tolkningen av konfidensintervallet.

Fra forbrukerhold kommer det klage på at El-bil produsenter underkommuniserer hvor avhengig rekkevidden er av temperaturen. Produsenten av modell A utfører derfor et forsøk der en tester rekkevidden for 7 forskjellige temperaturer mellom -10^0 C og 20^0 C. Temperaturene og de oppnådde rekkeviddene er gitt i tabellen nedenfor.

Temperatur i ^{0}C	-10	-5	0	5	10	15	20
Rekkevidde i km	73.4	87.2	95.9	103.1	110.1	118.1	127.1

Et plott av rekkevidde mot temperatur er gitt nedenfor.

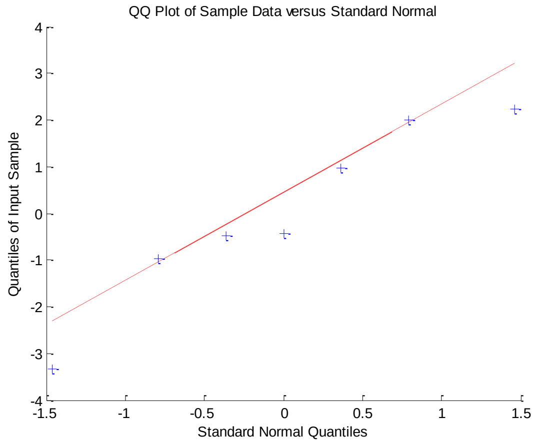


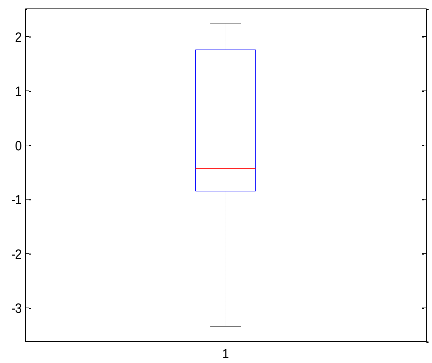
La $t_i,\ i = 1, 2, \cdots, 7$ være de 7 temperaturverdiene og $R_i,\ i = 1, 2, \ldots, 7$ være de tilhørende rekkeviddene. Ut i fra plottet synes sammenhengen å være tilnærmet lineær og de bestemmer seg for å tilpasse en modell av typen

$R_i = \beta_0 + \beta_1 t_i + \varepsilon_i,\ i = 1,2,\cdots,7$ der en antar at feil-leddene er uavhengige og oppfyller $\varepsilon_i \sim N(0,\sigma_\varepsilon^2)$. Minstekvadratsums estimatoren for β_1 er gitt ved $\hat\beta_1 = \dfrac{\sum_{i=1}^{7}(t_i - \bar t)R_i}{\sum_{i=1}^{7}(t_i - \bar t)^2}$.

d) Vis at variansen til $\hat\beta_1$ er gitt ved: $\dfrac{\sigma_\varepsilon^2}{\sum_{i=1}^{7}(t_i - \bar t)^2}$. Estimatene for β_0, β_1 og σ_ε^2 er gitt ved 93.7, 1.7 og 4.5 i gitt rekkefølge. Bruk dette til å konstruere et 95 % konfidensintervall for β_1.

Avvikene for hver t_i mellom observert rekkevidde og det tilhørende punktet på den estimerte linjen gitt ved $93.7 + 1.7t_i$ blir kalt residualene. Disse kan fortelle om forutsetningene at $\varepsilon_i \sim N(0,\sigma_\varepsilon^2)$ er rimelige. Et QQ-plott og et box-plott av disse er gitt nedenfor .





e) Kommenter plottene.

Du kan i resten av punktet anta at $\sigma^2 = \sigma_\varepsilon^2 = 4$ og at alle $R_i,\ i = 1,2,\ldots,7$ er uavhengige av $X_1,\cdots,X_{10},Y_1,\cdots,Y_{10}$. Det viser seg at de 10 verdiene for modell A oppgitt i innledningen var samlet inn i 15^0C og at de 10 verdiene for modell B var samlet inn i 10^0C. Det er derfor rimelig at det er $\mu_A - 5\beta_1$ som bør sammenlignes med μ_B. Bruk $\bar X - 5\hat\beta_1$ som estimator for $\mu_A - 5\beta_1$ og vurder om det er noe grunnlag for å si at forventet rekkevidde med modell A er forskjellig fra forventet rekkevidde med modell B ved 10^0C. Svar på spørsmålet ved å utføre en hypotesetest. Velg signifikansnivå selv.