NTNU
Kunnskap for ei betre verd

Institutt for matematiske fag

Eksamensoppgåve i TMA4240 Statistikk

Fagleg kontakt under eksamen: John Tyssedal, Haakon Bakka.

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

Eksamensdato: 17.12. 2014

Eksamenstid (frå-til): 09:00-13:00

Hjelpemiddelkode/Tilletne 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:

Alle svar skal grunngjevast og besvarelsen skal innehalde naturleg mellomrekning

Målform/språk: nynorsk

Sidetal (utan framside): 4

Sidetal vedlegg: 0

Kontrollert av:

Dato Sign

Oppgåve 1

I ein fotballsesong i tippeligaen skal kvart lag spele 30 kampar, 15 heimekampar og 15 bortekampar. Ved starten av sesongen reknar eit av fotballaga med at sannsyna for siger (S), uavgjort (U) eller tap (T) i ein heimekamp er gitt som P(S) = 0.6, P(U) = 0.2 og P(T) = 0.2. I ein bortekamp reknar dei at sannsyna er P(S) = 0.4, P(U) = 0.2 og P(T) = 0.4. Gå vidare ut i frå at resultatet i kvar kamp er uavhengig av resultata i dei andre kampane.

a) Laga får 3 poeng for siger, 1 poeng for uavgjort og 0 poeng for tap. La X vere talet på poeng laget får i ein heimekamp og la Y vere talet på poeng det får i ein bortekamp. Vis at forventing og varians til X er gitt ved 2 og 1.6 og at forventing og varians til Y er gitt ved 1.4 og 1.84.

La X_i, $i = 1, 2, \cdots, 15$ vere poenga laget får i dei 15 heimekampane og Y_i, $i = 1, 2, \cdots, 15$ poenga dei får i dei 15 bortekampane. La vidare $X_H = \sum_{i=1}^{15} X_i$ og $Y_B = \sum_{i=1}^{15} Y_i$. Laget reknar vidare med at dersom det greier å få minst 60 poeng så er det godt nok til medalje.

b) Finn forventing og varians til X_H og Y_B. Gå deretter ut i frå at X_H og Y_B er tilnærma normalfordelte og finn tilnærma sannsynet for at laget greier å få minst 60 poeng.

Det viser seg at poengsankinga går tregt og når det er 8 kampar att, 4 heimekampar og 4 bortekampar, reknar dei med at dei må vinne 7 av dei 8 siste kampane for å redde plassen i serien. Dei bestemmer seg for å gjere ei sannsynsvurdering der dei brukar sannsyna for siger gitt i innleinga av oppgåva. Dersom sannsynet for å redde plassen i serien blir mindre enn 0.05 bestemmer dei seg for å sparke trenaren.

c) La N_H og N_B vere talet på heime- og bortekampar som laget vinn i dei 8 siste kampane. Kva fordeling har N_H og N_B? Blir trenaren sparka?

Oppgåve 2

Ingolv har alltid vore fasinert av vulkanar og vulkanutbrot. Etter vidaregåande bestemmer han seg difor for å dra til Island og ta ein 5-årig master der. La X_t vere talet på vulkanutbrot på Island i eit tidsintervall av lengde t. Ut i frå historiske data viser det seg at det er rimeleg å gå ut i frå at X_t er poissonfordelt med parameter λt, der $\lambda = 0.3$ og t er talet på år. Det vil sei at

$$P(X_t = x) = \frac{(0.3t)^x e^{-0.3t}}{x!}, \quad x = 0, 1, 2, \cdots$$

a) Kva er sannsynet for at Ingolv skal få oppleve minst eit vulkanutbrot i ein 5 års periode?

Gå ut i frå at det skjedde eit vukanutbrot nøyaktig eit år etter at Ingolv kom til Island. Kva er då sannsynet for at han skal få oppleve minst eit til i det femårige opphaldet sitt?

b) La T vere tida frå Ingolv kjem til Island til det skjer to vulkanutbrot. Kva fordeling har T?
Forklar kvifor $P(T \le t) = 1 - P(X_t < 2)$. Bruk dette til å finne ut tilnærma kor lenge Ingolv må opphalde seg på Island for at sannsynet for å få med seg 2 vulkanutbrot er minst 0.80.

Oppgåve 3

To El-bil produsentar konkurrerer om det same kundemarkedet med kvar sine modelltypar. La oss kalle dei to modelltypane for modell A og modell B. Ein viktig faktor for val av modelltype er rekkevidda som bilen har før ein må lade på nytt. Produsenten av modell B skuldar produsenten av modell A for å oppgje for lang rekkevidde. Dei hevdar at rekkevidda som er oppgitt ikkje er realistisk å oppnå under normale køyreforhold. For å undersøke dette nærare bestemmer produsenten av modell A seg for å gjere eit lite forsøk. 10 sjåførar får i oppdrag å køyre kvar sin bil i noko som blei definert som normalterreng og med ein snittfart på 60 km/t. La X_i vere rekkevidda for sjåfør i, $i = 1,2,\cdots,10$. Dei observerte rekkeviddene i km er gitt nedanfor:

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 forventingsrette estimatorar for μ_A og σ^2. Kva blir estimata 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 oppgjev at forventa rekkevidde under slike forhold er 125 km. Gå no ut i frå at $X_i \sim N(\mu_A, \sigma^2)$, $i = 1,2,\cdots,10$.

b) Kan ein ut i frå dataane konkludere med at produsenten har sett rekkevidda for høgt? Formuler dette som eit hypotesetestingsproblem og utfør testen. Bruk signifikansnivå $\alpha = 0.01$.

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

Dei observerte rekkeviddene i km er gitt nedanfor.

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

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

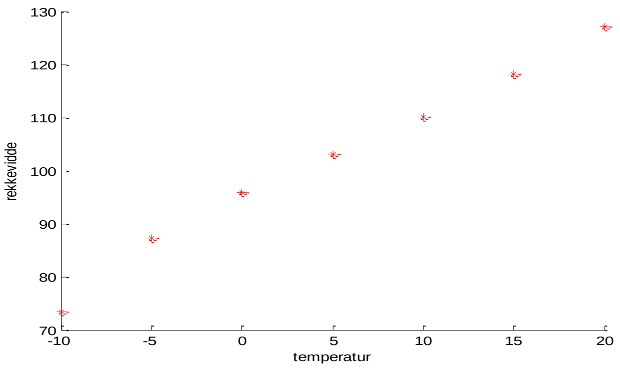
Gå ut i frå at $Y_i \sim N(\mu_B, \sigma^2)$. Variansen er altså lik i dei to utvala, men forventingane kan vere forskjellige. Gå og ut i frå at $X_1, \cdots, X_{10}, Y_1, \cdots, Y_{10}$ alle er uavhengige.

c) Sett opp ein forventingsrett estimator for σ^2 basert på dei 2 utvala. Lag eit 95 % konfidensintervall for $\mu_A - \mu_B$ basert på dei observerte dataane. Tyder intervallet på at det som står i reklamen for modell A kan vere rett. Grunngje svaret ved å trekke inn tolkningen av konfidensintervallet.

Frå forbrukarhald kjem det klage på at El-bil produsentar underkommuniserer kor avhengig rekkevidda er av temperatuten. Produsenten av modell A utfører difor eit forsøk der ein testar rekkevidda for 7 forskjellige temperaturar mellom -10^0 C og 20^0 C. Temperaturane og rekkeviddene er gitt i tabellen nedanfor.

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

Eit plott av rekkevidde mot temperatur er gitt nedanfor.

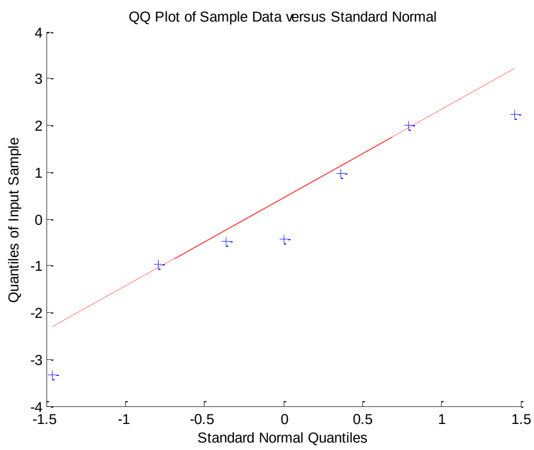


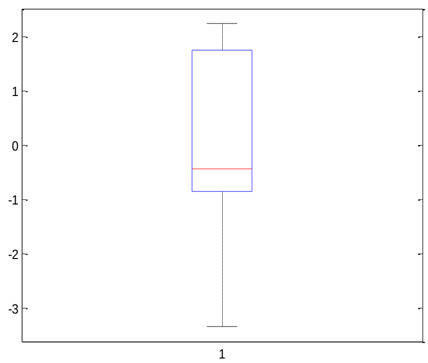
La t_i, $i = 1, 2, \cdots, 7$ vere dei 7 temperaturverdiane og R_i, $i = 1, 2, \cdots, 7$ vere dei tilhøyrande rekkeviddene. Ut i frå plottet synest samanhengen å vere tilnærma linear og dei bestemmer seg for å tilpasse ein modell av typen $R_i = \beta_0 + \beta_1 t_i + \varepsilon_i$, $i = 1, 2, \cdots, 7$ der ein går ut i frå at feil-ledda er uavhengige og oppfyller $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$, $i = 1, 2, \cdots, 7$.

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 $\dfrac{\sigma_\varepsilon^2}{\sum_{i=1}^{7}(t_i - \bar{t})^2}$. Estimata for β_0, β_1 og σ_ε^2 er gitt ved 93.7, 1.7 og 4.5 i gitt rekkjefølgje. Bruk dette til å konstruere eit 95% konfidensintervall for β_1.

Avvika for kvar t_i mellom observert rekkevidde og det tilhøyrande punktet på den estimerte linja gitt ved $93.7 + 1.7t_i$, $i = 1, 2, \cdots, 7$ blir kalla residuala. Desse kan fortelje om føresetnaden at $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ er rimeleg. Eit QQ-plott og eit box-plott av desse er gitt nedanfor.





e) Kommenter plotta.

Du kan i resten av punktet gå ut i frå at $\sigma^2 = \sigma_\varepsilon^2 = 4$ og at alle R_i, $i = 1, 2, \cdots, 7$ er uavhengige av $X_1, \cdots, X_{10}, Y_1, \cdots, Y_{10}$. Det viser seg at dei 10 verdiane for modell A oppgitt i innleinga var samla inn i 15^0C medan dei 10 verdiane for modell B var samla inn i 10^0C. Det er difor rimeleg at det er $\mu_A - 5\beta_1$ som bør samanliknast med μ_B. Bruk $\bar{X} - 5\hat{\beta}_1$ som estimator for $\mu_A - 5\beta_1$ og vurder om det er noko grunnlag for å sei at rekkevidda med modell A er forskjellig frå rekkevidda med modell B ved 10^0C. Svar på spørsmålet ved å utføre ein hypotetest. Velg signifikansnivå sjølv.