TMA4245 Statistikk
Eksamen 21. mai 2013

Korrigert 10. juni 2013

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgave 1

La X og Y være to uavhengige normalfordelte stokastiske variabler. Anta at X har forventningsverdi lik 0 og standardavvik lik 1, mens Y har forventningsverdi lik 1 og standardavvik lik 2.

Skisser sannsynlighetstetthetene for X og Y i et felles plott. Finn sannsynlighetene $\mathrm{P}(X \leq 1.2)$, $\mathrm{P}(Y > 2)$ og $\mathrm{P}(X + Y \leq 2)$.

Oppgave 2

La A og B være to hendelser i et utfallsrom S, der $\mathrm{P}(A) = 0.2$, $\mathrm{P}(B) = 0.5$ og $\mathrm{P}(A \cup B) = 0.6$.

Er hendelsene A og B disjunkte? Er hendelsene A og B uavhengige?

Oppgave 3

Levetiden T (målt i antall døgn) til en ny type ventiler som eventuelt skal benyttes på oljeplattformer i Nordsjøen, skal undersøkes. Det er velkjent at levetiden påvirkes av blant annet temperatur og trykk der ventilen benyttes, og av den kjemiske sammensetningen av olja som går gjennom ventilene. Anta at effekten av disse faktorene måles som en *stress*-faktor z, og at en vet av erfaring at kumulativ fordelingsfunksjon for levetiden T er gitt ved

$$F(t) = \mathrm{P}(T \leq t) = \begin{cases} 1 - e^{-\frac{zt^2}{\theta}} & \text{for } t > 0, \\ 0 & \text{ellers,} \end{cases}$$

der θ er en ukjent parameter. Parameteren θ er altså karakteristisk for en bestemt type ventiler, mens z beskriver miljøet der ventilen benyttes.

a) Når $z = 2.0$ og $\theta = 2 \cdot 10^6$, finn sannsynlighetene

$$\mathrm{P}(T > 1000) \quad \text{og} \quad \mathrm{P}(T > 2000 | T > 1000).$$

La T_1, T_2 og T_3 være levetidene til tre ventiler som alle opererer under forhold med $z = 2.0$, og anta at T_1, T_2 og T_3 er uavhengige stokastiske variabler. Når $\theta = 2 \cdot 10^6$, finn sannsynligheten for at minst to av de tre levetidene er større enn 1000 døgn.

b) Vis at sannsynlighetstettheten til T er gitt ved

$$f(t) = \begin{cases} \frac{2zt}{\theta} e^{-\frac{zt^2}{\theta}} & \text{for } t > 0, \\ 0 & \text{ellers.} \end{cases}$$

Skisser $f(t)$ for $t \in [0, 3000]$ når $z = 2.0$ og $\theta = 2 \cdot 10^6$, og skraver i denne skissen arealet som er lik sannsynligheten $\mathrm{P}(T > 1000)$.

La en stokastisk variabel V være definert ved

$$V = \frac{2zT^2}{\theta}.$$

c) Bruk transformasjonsformelen til å vise at V er χ^2-fordelt med 2 frihetsgrader.

Benytt så dette til å vise at $\mathrm{E}\left(T^2\right) = \frac{\theta}{z}$ og $\mathrm{Var}\left(T^2\right) = \left(\frac{\theta}{z}\right)^2$.

For å undersøke kvaliteten på denne type ventiler har man prøvd ut $n = 10$ ventiler. La $z_1, z_2, \ldots, z_n$ betegne stress-faktorene som disse ventilene opererer under, og la $T_1, T_2, \ldots, T_n$ betegne tilhørende levetider, og anta at $T_1, T_2, \ldots, T_n$ er uavhengige stokastiske variabler. De observerte levetidene er gitt i følgende tabell:

Ventil i	1	2	3	4	5	6	7	8	9	10
z_i	1.0	3.4	1.9	2.4	1.2	4.0	3.2	2.2	1.4	3.2
t_i	1297.2	834.2	1265.8	331.7	1937.8	727.6	869.6	746.7	1965.3	280.9

Det oppgis at $\sum_{i=1}^{n} z_i t_i^2 = 23\ 287\ 125$.

d) Utled sannsynlighetsmaksimeringsestimatoren (*maximum likelihood*-estimatoren) for θ og vis at den kan skrives på formen

$$\widehat{\theta} = \frac{1}{n} \sum_{i=1}^{n} z_i T_i^2.$$

Er $\widehat{\theta}$ forventningsrett? Finn $\mathrm{Var}(\widehat{\theta})$.

e) Begrunn at

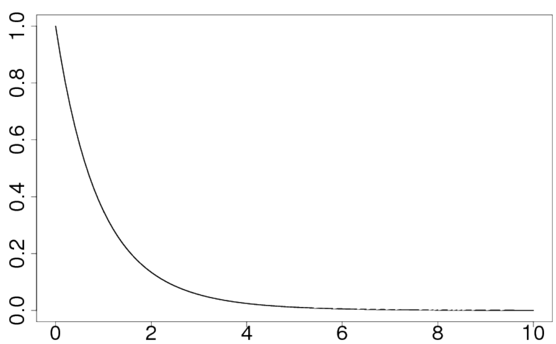
$$U = \sum_{i=1}^{n} \frac{2z_i T_i^2}{\theta} \sim \chi^2_{2n}.$$

Benytt så dette til å utlede et $(1-\alpha)\cdot 100\%$-konfidensintervall for θ. Hva blir konfidensintervallet når dataene er som gitt over og $\alpha = 0.05$?

Ut fra de observerte dataene ønsker man også å lage et prediksjonsintervall for levetiden, T_0, til en ny ventil som skal operere under forhold med stress-faktor lik z_0. For å gjøre dette kan man ta utgangspunkt i den stokastiske variabelen

$$Y = n \cdot \frac{\frac{2T_0^2 z_0}{\theta}}{\sum_{i=1}^{n} \frac{2z_i T_i^2}{\theta}},$$

der teller og nevner i brøken altså er uavhengige stokastiske variabler. Teller og nevner er begge χ^2-fordelt og har henholdsvis 2 og $2n$ frihetsgrader. Når $n = 10$ kan det vises at sannsynlighetstettheten til Y blir som vist i figur 1. Merk at noen kvantiler i denne fordelinga er oppgitt til høyre i samme figur.



$P(Y > y_\alpha) = \alpha$

α	y_α
0.975	0.025
0.950	0.051
0.900	0.106
0.100	2.59
0.050	3.49
0.025	4.46

Figur 1: Sannsynlighetstettheten for Y og noen kvantiler i denne fordelinga.

f) Utled et 90% prediksjonsintervall for T_0. Hva blir prediksjonsintervallet når dataene er som gitt over og $z_0 = 3.0$?

Oppgave 4

Som ledd i en undersøkelse av den fysiske formen av mannlige soldater ble $n = 42$ tilfeldig valgte mannlige soldater plukket ut til å gjennomføre en rekke fysiske tester. I denne oppgaven skal vi betrakte resultatet av to av testene som ble utført, antall armhevinger soldatene greidde i løpet av 2 minutter og tiden (målt i sekunder) de brukte på å løpe 3 kilometer. Vi antar en enkel lineær regresjonsmodell for disse dataene, der antall armhevinger er forklaringsvariabel (uavhengig variabel) og løpetiden er responsvariabel (avhengig variabel). Vi har dermed modellen

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

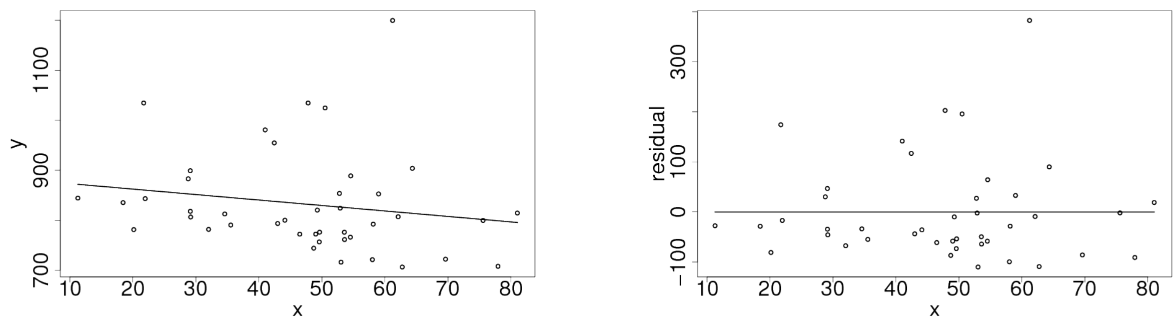
der Y_i og x_i betegner henholdsvis løpetiden og antall armhevinger, og $\varepsilon_1, \ldots, \varepsilon_n$ antas uavhengige og normalfordelte med forventningsverdi 0 og varians σ^2. Verdiene til alle de tre parametrene β_0, β_1 og σ^2 er ukjente. For å estimere disse skal vi benytte de vanlige estimatorene

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{x}$$

og

$$S^2 = \frac{1}{n-2} \sum_{i=1}^n \left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i \right)^2 .$$

Figur 2(a) viser de $n = 42$ observasjonene, samt estimert regresjonslinje. Observasjonene gir $\sum_{i=1}^n (x_i - \bar{x}) y_i = -12\,163.6$, $\sum_{i=1}^n (x_i - \bar{x})^2 = 11\,113.9$ og $\sum_{i=1}^n \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i \right)^2 = 414\,563.1$.



(a) (b)

Figur 2: (a) Observerte data og estimert regresjonslinje, med antall armhevinger på x-aksen og løpetiden på y-aksen. (b) Tilhørende residualplott med antall armhevinger på x-aksen og estimerte residual $\widehat{\varepsilon} = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i$ på y-aksen.

Videre i oppgaven kan du uten bevis benytte at

$$\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\frac{S^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}}$$

er t-fordelt med $n-2$ frihetsgrader.

Man ønsker å undersøke om de observerte data gir grunnlag for å påstå at forventet løpetid avtar med økende antall armhevinger.

a) Formuler dette som et hypotesetestingsproblem og lag en test for dette formålet med signifikansnivå 5%.

Hva blir konklusjonen på hypotesetesten når dataene er som gitt over.

b) Modellen vi benytter i denne oppgaven antar at residualene, ε_i, er normalfordelte med forventning lik null og med samme standardavvik σ. Vurder ut fra plottet i figur 2(a), samt plottet av estimerte residualer $\widehat{\varepsilon}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i$ i figur 2(b), om dette synes tilfredsstilt i denne undersøkelsen.

Fasit

1. 0.8849,0.3085,0.6736

3. **a**) 0.368,0.050,0.306 **d**) $\mathrm{E}[\widehat{\theta}] = \theta$, $\mathrm{Var}[\widehat{\theta}] = \theta^2/n$ **e**) $[1\,363\,016, 4\,856\,037]$ **f**) $[198.97, 1\,645.92]$

4. **a**) Ikke forkast H_0