TMA4245 Statistikk
Eksamen 21. mai 2013

Korrigert 10. juni 2013

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgave 1

La X og Y vere to uavhengige normalfordelte stokastiske variablar. Gå ut frå at X har forventningsverdi lik 0 og standardavvik lik 1, medan Y har forventningsverdi lik 1 og standardavvik lik 2.

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

Oppgave 2

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

Er hendingane A og B disjunkte? Er hendingane A og B uavhengige?

Oppgave 3

Levetida T (målt i døgn) til ein ny type ventilar som eventuelt skal nyttast på oljeplattformer i Nordsjøen skal undersøkjast. Det er velkjent at levetida blir påverka av blant anna temperatur og trykk der ventilen nyttast, og av den kjemiske samansetninga av olja som går gjennom ventilane. Gå ut frå at effekten av desse faktorane målast som ein *stress*-faktor z, og at ein veit av røynsle at kumulativ fordelingsfunksjon for levetida T er gjeve ved

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

der θ er ein ukjend parameter. Parameteren θ er altså karakteristisk for ein bestemt type ventilar, medan z skildrar miljøet der ventilen nyttast.

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

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

La T_1, T_2 og T_3 vere levetidene til tre ventilar som alle opererer under tilhøve med $z = 2.0$, og gå ut frå at T_1, T_2 og T_3 er uavhengige stokastiske variablar. Når $\theta = 2 \cdot 10^6$, finn sannsynet for at minst to av dei tre levetidene er større enn 1000 døgn.

b) Vis at sannsynstettleiken til T er gjeve ved

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

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

La ein stokastisk variabel V vere definert ved

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

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

Nytt 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 undersøkje kvaliteten på denne type ventilar har ein prøvd ut $n = 10$ ventilar. La $z_1, z_2, \ldots, z_n$ vere stress-faktorane som desse ventilane opererer under, og la $T_1, T_2, \ldots, T_n$ vere tilhøyrande levetider, og gå ut frå at $T_1, T_2, \ldots, T_n$ er uavhengige stokastiske variablar. Dei observerte levetidene er gjeve i følgjande 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 er oppgjeve at $\sum_{i=1}^{n} z_i t_i^2 = 23\ 287\ 125$.

d) Utlei sannsynsmaksimeringsestimatoren (*maximum likelihood*-estimatoren) for θ og vis at han kan skrivast på forma

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

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

e) Grunngje at

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

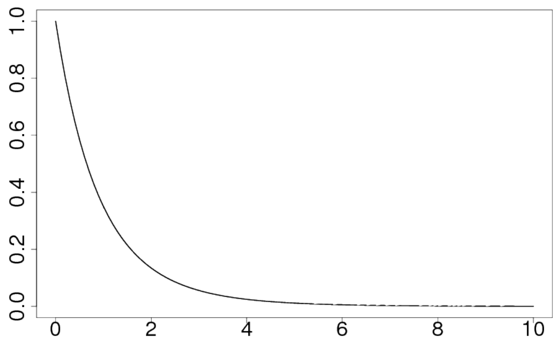
Nytt så dette til å utleie et $(1 - \alpha) \cdot 100\%$-konfidensintervall for θ. Kva blir konfidensintervallet når observasjonane er som gjeve over og $\alpha = 0.05$?

Ut frå observasjonane ynskjer ein óg å lage eit prediksjonsintervall for levetida, T_0, til ein ny ventil som skal operere under tilhøve med stress-faktor lik z_0. For å gjere dette kan ein 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 teljar og nemnar i brøken altså er uavhengige stokastiske variablar. Teljar og nemnar er begge χ^2-fordelte og har høvesvis 2 og $2n$ fridomsgradar. Når $n = 10$ kan det visast at sannsynstettleiken til Y blir som vist i figur 1.

Merk at nokre kvantilar i denne fordelinga er oppgjeve til høgre i same 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: Sannsynstettleiken for Y og nokre kvantilar i denne fordelinga.

f) Utlei eit 90% prediksjonsintervall for T_0. Kva blir prediksjonsintervallet når observasjonane er som gjeve over og $z_0 = 3.0$?

Oppgave 4

Som ledd i ei undersøking av den fysiske forma av mannlege soldatar vart $n = 42$ tilfeldig valde mannlege soldatar plukka ut til å gjennomføre ei rekkje fysiske testar. I denne oppgåva skal vi sjå på resultatet av to av testane som vart utførte, talet på armhevingar soldatane greidde på 2 minuttar og tida (målt i sekund) dei brukte på å springe 3 kilometer. Vi går ut frå ein enkel lineær regresjonsmodell for desse observasjonane, der talet på armhevingar er forklaringsvariabel (uavhengig variabel) og springetida er responsvariabel (avhengig variabel). Vi har dermed modellen

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

der Y_i og x_i er høvesvis springetida og talet på armhevingar, og der vi går ut frå at $\varepsilon_1, \ldots, \varepsilon_n$ er uavhengige og normalfordelte med forventningsverdi 0 og varians σ^2. Verda til alle dei tre parametrane β_0, β_1 og σ^2 er ukjende. For å estimere desse skal vi nytte dei vanlege estimatorane

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

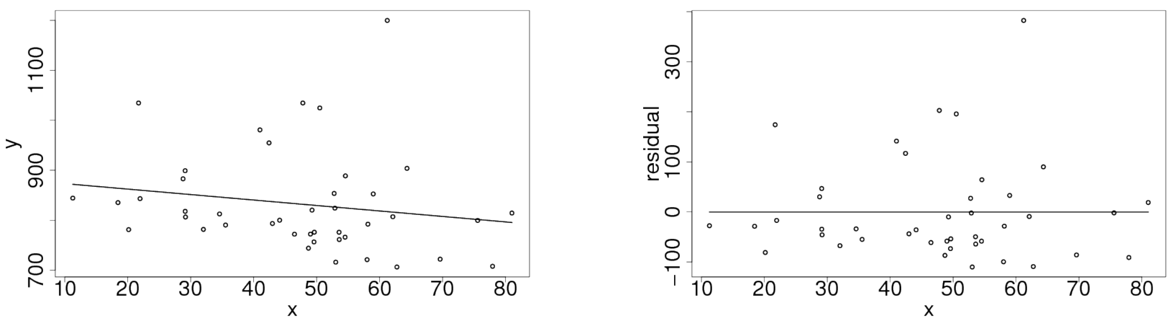
og

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

Figur 2(a) viser dei $n = 42$ observasjonane, samt estimert regresjonslinje. Observasjonane gjev $\sum_{i=1}^n (x_i - \bar{x}) y_i = -12163.6$, $\sum_{i=1}^n (x_i - \bar{x})^2 = 11113.9$ og $\sum_{i=1}^n \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i \right)^2 = 414563.1$.

Vidare i oppgåva kan du utan bevis nytte at

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



(a) (b)

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

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

Ein ynskjer å undersøkje om observasjonane gjev grunnlag for å påstå at forventa springetid minkar med aukande tal på armhevingar.

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

Kva blir konklusjonen på hypotesetesten når observasjonane er som gjeve over.

b) Modellen vi nyttar i denne oppgåva går ut frå at residuala, ε_i, er normalfordelte med forventning lik null og med same standardavvik σ. Vurder ut frå plottet i figur 2(a), og ut frå plottet av estimerte residual $\widehat{\varepsilon}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i$ i figur 2(b), om dette synast tilfredsstilt i denne undersøkinga.

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**) Ikkje forkast H_0