Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Faglige kontakter under eksamen:
Mette Langaas 98847649
Eirik Mo 73593541/41106633

EKSAMEN I FAG TMA4245 STATISTIKK

Lørdag 5. juni 2004

Tid: 09:00–14:00

Tillatte hjelpemidler:
Gult A5 ark med egne håndskrevne notater.
Tabeller og formler i statistikk (Tapir Forlag).
K. Rottmann: *Matematisk formelsamling.*
Kalkulator: HP30S.

BOKMÅL

Sensur: 28. juni 2004.

Oppgave 1 Atle, du lyver!

Utspørring av deltakere i humor- og realityprogram på TV har den siste tiden blitt svært populært. Spørsmålene som stilles kan vi dele inn i tre typer, og vi definerer følgende disjunkte hendelser:

A_1="det stilles et spørsmål av ikke-sensitiv natur, f.eks. hva heter du?",
A_2="det stilles et spørsmål av delvis sensitiv natur, f.eks. hvor gammel er du?",
A_3="det stilles et spørsmål av sensitiv natur, f.eks. har du vært utro?".

I tillegg definerer vi hendelsen:
L="deltakeren lyver"

Følgende sannsynligheter er oppgitt:
$P(A_1) = 0.1$, $P(A_2) = 0.4$, $P(A_3) = 0.5$, $P(L|A_1) = 0.05$, $P(L|A_2) = 0.2$, $P(L|A_3) = 0.6$.

a) Vis de fire hendelsene i et venndiagram.

Gitt at en deltaker blir spurt et spørsmål av type A_2, hva er sannsynligheten for at deltakeren ikke lyver, $P(L'|A_2)$?

Hva er sannsynligheten for at en tilfeldig valgt deltaker lyver, $P(L)$?

Et av spørsmålene som regnes å være av delvis sensitiv natur er "hvor gammel er du?". En gruppe på n personer ble stilt dette spørsmålet, deretter ble svarene registrert og sammenlignet med informasjon i offentlige registre. La X være en stokastisk variabel som angir antall personer som lyver blant n personer, og la p være sannsynligheten for at en person lyver.

b) Under hvilke antagelser vil X være binomisk fordelt?

Vi antar at $p = 0.2$ og at vi spør $n = 20$ personer. Hva er $P(X = 4)$?

Hva er $P[(X \leq 2) \cup (X > 5)]$?

Vi antar nå at p er ukjent. For å estimere p er det foreslått to estimatorer,

$$\hat{p} = \frac{X}{n} \quad \text{og} \quad p^* = \frac{X}{n-1}.$$

c) Finn forventningsverdi og varians til hver av estimatorene $\hat{p}$ og p^*.

Hvilke to egenskaper kjennetegner en god estimator?

Hvilken av estimatorene $\hat{p}$ og p^* vil du foretrekke? Begrunn svaret.

Vi forutsetter at n er stor og velger å benytte estimatoren $\hat{p}$ videre. Da er

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

tilnærmet standard normalfordelt.

Basert på tidligere undersøkelser antar vi at $p = 0.2$. Vi ønsker å undersøke om de innsamlede data gir oss grunn til å tro at p er større enn 0.2.

d) Formulér dette som en hypotesetest ved å definere nullhypotese og alternativ hypotese.

Bruk at Z er tilnærmet standard normalfordelt til å bestemme et forkastningsområde for nullhypotesen når vi velger signifikansnivå $\alpha = 0.01$.

Hva blir konklusjonen på testen når $n = 200$ personer ble spurt og $x = 55$ personer løy?

Vil p-verdien til testen være mindre eller større enn 0.01? Begrunn svaret. (Det kreves ikke at du regner ut p-verdien.)

Oppgave 2 Aksjekurser

Selskapet Agderfrukt er notert på børsen. Vi antar at endringen X i verdien på en Agderfrukt-aksje i løpet av en dag er normalfordelt med forventningsverdi $\mu_X = 0.15$ kroner og standardavvik $\sigma_X = 0.60$ kroner. Har du en aksje i Agderfrukt vil $X > 0$ bety fortjeneste, mens $X < 0$ er tap.

a) Hva er sannsynligheten for å tape penger i løpet av en dag, dvs. $P(X < 0)$?

Hva er $P(0 \leq X \leq 0.15)$?

Hvis du kjøper 10 aksjer i Agderfrukt idag og selger imorgen, hva er forventet fortjeneste? Finn også variansen til fortjenesten.

Selskapet Trønderfrukt er også notert på børsen. Vi kaller endringen på en Trønderfrukt-aksje i løpet av en dag for Y, der Y er normalfordelt med forventningsverdi $\mu_Y = 0.15$ kroner og standardavvik $\sigma_Y = 0.80$ kroner.

b) Vi ser på aksjekursendringen på den samme dagen for Agderfrukt og Trønderfrukt og antar i dette punktet at aksjekursendringene X og Y er uavhengige.

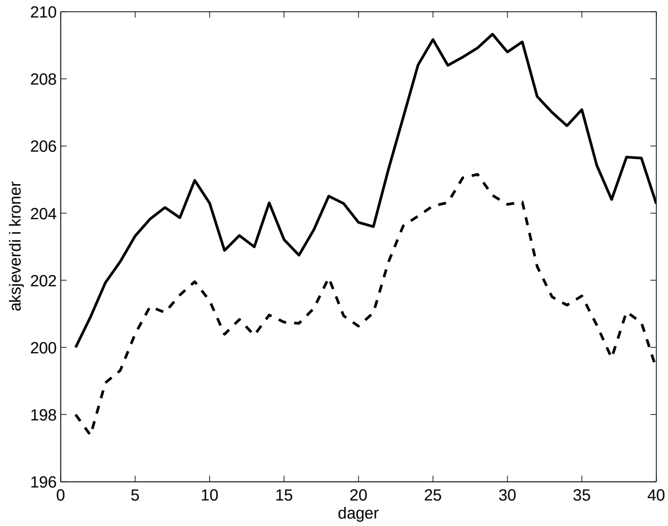
Idag er verdien til en Agderfrukt-aksje den samme som verdien til en Trønderfrukt-aksje. Vi ønsker å undersøke tre mulige strategier for aksjekjøp, der vi kjøper aksjer idag og selger imorgen.

i) Kjøp to aksjer i Agderfrukt.

ii) Kjøp en aksje i Agderfrukt og en aksje i Trønderfrukt.

iii) Kjøp to aksjer i Trønderfrukt.

Dersom du vil ha minst mulig risiko for investeringen din, hvilken av de tre investeringsstrategiene over vil du velge? Begrunn svaret.



Figuren viser utviklingen av aksjekursen til Agderfrukt (stiplet) sammen med aksjekursen til Trønderfrukt (heltrukket).

Kursendringen dag i for Agderfrukt kaller vi X_i, og vi antar at X_i er normalfordelt med forventning $\mu_X = 0.15$ kroner og standardavvik $\sigma_X = 0.60$ kroner.

Kursendringen dag i for Trønderfrukt kaller vi Y_i, og vi antar at Y_i er normalfordelt med forventning $\mu_Y = 0.15$ kroner og standardavvik $\sigma_Y = 0.80$ kroner.

Kursendringer for ulike dager antas å være uavhengige.

Vi sammenlikner de to selskapene ved å måle differansen mellom de daglige kursendringene, $D_i = X_i - Y_i$, og ta gjennomsnitt. Vi ser på 10 dager og får $\overline{D} = \frac{1}{10}\sum_{i=1}^{10} D_i = \frac{1}{10}\sum_{i=1}^{10}(X_i - Y_i)$.

c) Gir figuren grunn til å tro at endringene i de to aksjekursene samme dag, X_i og Y_i, er uavhengige?

Korrelasjonen mellom X_i og Y_i for disse to selskapene, $\rho(X_i, Y_i)$, er enten -0.5, 0.0 eller 0.5. Hvilken av disse verdiene virker mest rimelig fra figuren? Begrunn kort.

Hva blir forventningsverdi og varians for $\overline{D}$? Benytt verdien for korrelasjonen, $\rho(X_i, Y_i)$, som du valgte over.

Oppgave 3 Bølgehøyde

For å kunne dimensjonere en oljeplattform er det viktig å vite hvor store bølgene kan bli i området der plattformen skal plasseres. Det settes derfor ut en bølgehøydemåler. La X være største bølgehøyde en tilfeldig valgt dag. Vi antar at sannsynlighetstettheten til X er gitt ved

$$f(x;\theta) = \frac{2x}{\theta} e^{-\frac{x^2}{\theta}}, \qquad x \geq 0, \qquad \theta > 0.$$

Det oppgis at $\mathrm{E}[X^2] = \theta$ og $E[X^4] = 2\theta^2$.

a) Vis at den kumulative fordelingsfunksjonen, $F(x) = P(X \leq x)$, er $F(x) = 1 - e^{-\frac{x^2}{\theta}}$. (Hint: Bruk substitusjon med $u = x^2$).

Gitt at største bølgehøyde er større enn 10 meter, finn sannsynligheten for at den er større enn 15 meter hvis $\theta = 25$, dvs. $P(X > 15 | X > 10)$?

I resten av oppgaven regnes θ som ukjent.

Vi har observert største bølgehøyde i n dager. La X_i være største bølgehøyde på dag i. Vi antar at $X_1, ..., X_n$ er uavhengige og identisk fordelte med sannsynlighetstetthet $f(x;\theta)$.

b) Finn sannsynlighetsmaksimeringsestimatoren (SME) $\hat{\theta}$ for θ.

Er estimatoren $\hat{\theta}$ forventningsrett?

Finn også variansen til $\hat{\theta}$.

c) Bruk sentralgrenseteoremet til å argumentere for at

$$Z = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \theta}{\sqrt{\frac{\theta^2}{n}}}$$

er tilnærmet standard normalfordelt.

Bruk Z til å finne et tilnærmet 95% konfidensintervall for θ.

Sannsynligheten for at største bølgehøyde en tilfeldig valgt dag overskrider 10 meter er $P(X > 10) = e^{-\frac{100}{\theta}}$. Bruk det tilnærmede konfidensintervallet for θ til å finne et tilnærmet 95% konfidensintervall for $e^{-\frac{100}{\theta}}$.