Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Faglige kontakter under eksamen:
Jo Eidsvik 90127472
Ola Diserud 93218823

EKSAMEN I FAG TMA4240 STATISTIKK

Mandag 12.des 2011

Tid: 09:00–13:00

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

BOKMÅL

Sensur: 5.jan 2012.

Oppgave 1 Oljeleting

a) La A og B være to hendelser i et utfallsrom. Det oppgis at $P(A) = 0.3$, $P(B) = 0.3$ og $P(A \cup B) = 0.6$.

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

Oljefeltet Aldous Major / Avaldsnes ble funnet på grunn av mindre lignende funn i nærområdet. Avhengighetstruktur i modellen for oljefunn ga økt sannsynlig for å finne olje på Aldous Major / Avaldsnes.

Vi tenker oss to oljefelt. Hendelsen A = oljefunn på felt 1, mens den komplementære hendelsen A^c = ingen olje på felt 1. Tilsvarende er hendelsen B = oljefunn på felt 2, mens B^c = ingen olje på felt 2. Vi får oppgitt at $P(A \cap B) = 0.05$, $P(A^c \cap B) = 0.1$, $P(A \cap B^c) = 0.15$ og $P(A^c \cap B^c) = 0.7$.

b) Tegn et Venn diagram for hendelsene.

Finn sannsynligheten for olje på felt 1.

Anta at man har funnet olje på felt 2. Hva er nå sannsynligheten for olje på felt 1?

Anta at man har påvist at felt 2 ikke inneholder olje. Hva er nå sannsynligheten for olje på felt 1?

Er hendelsene A og B uavhengige?

Vi ser for oss en kostnad $K = 100$ millioner kroner ved å lete etter olje. Dersom man ikke finner olje får man ingen gevinst, men betaler denne kostnaden. Dersom man finner olje, får man en stor gevinst, som overstiger kostnaden K. Anta at fortjenesten ved oljefunn på felt 1 er $R_1 = 500 - K = 400$ millioner kroner, mens fortjenesten ved oljefunn på felt 2 er $R_2 = 1100 - K = 1000$ millioner kroner.

c) Regn ut forventet fortjeneste ved å lete etter olje på felt 1.

Anta at man har funnet olje på felt 2. Hva er nå forventet fortjeneste ved å lete etter olje på felt 1?

Anta at du kan velge mellom følgende letestrategier: Ikke lete noe sted, lete ved felt 1, eller lete ved felt 2. Hvis man velger å lete ved felt 1 eller 2, kan man avhengig av utfallet stoppe, eller lete videre på det andre feltet. Beslutning tas utfra forventet verdi. Hvilken letestrategi gir høyest forventet fortjeneste, og hva er forventet fortjeneste under denne strategien?

Oppgave 2 Lønninger

La den tilfeldige (stokastiske) variabelen Y beskrive en persons årslønn for 2011 i kNOk (1 kNOK er 1000 norske kroner). Paretos lov påstår da at

$$P(Y \geq y) = \left(\frac{k}{y}\right)^{\theta},$$

hvor k er minimumsinntekten for hele populasjonen.

a) Anta at Paretos lov stemmer. Vis at sannsynlighetsfordelingen (tettheten) til Y blir

$$f(y) = \theta k^{\theta} \left(\frac{1}{y}\right)^{\theta+1}, \quad y \geq k, \;\; \theta > 2.$$

Finn forventningsverdi og varians til Y.

Anta videre at $k = 214.9$ kNOK, dvs første lønnstrinn i Statens lønnsregulativ for 2011. Fra et tilfeldig utvalg på 30 årslønner finner dere at $\sum_{i=1}^{30} y_i = 13611$ og $\sum_{i=1}^{30} \ln y_i = 174.7$.

b) Vis at sannsynlighetsmaksimeringsestimatoren (maximum likelihood estimator) for θ, fra et tilfeldig utvalg på størrelse n, blir

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} \ln y_i - n \ln k}.$$

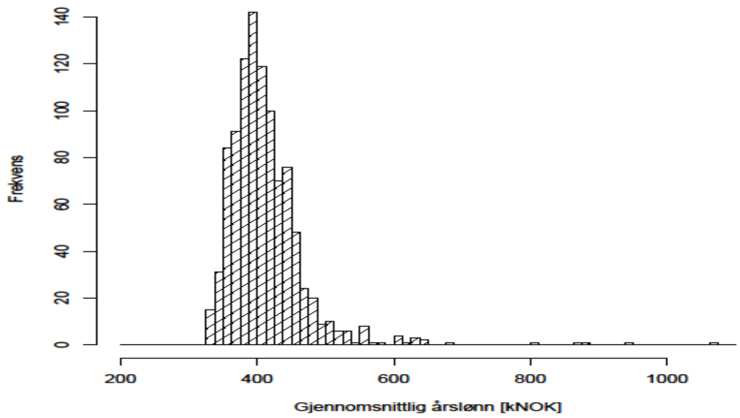
Hva blir sannsynligheten for at en tilfeldig person skal ha årslønn under 472.2 kNOK, dvs innslagspunktet for toppskatt, hvis vi bruker punktestimatet for θ som parameterverdi?

c) For 2010 var gjennomsnittslønna for alle norske ansatte 435 kNOK, med standardavvik 516 kNOK. Gjennomsnittlig årslønn for 2010 hadde da steget med 3.6 % i forhold til 2009.

Du skal nå teste om forventet årslønn for 2011 også ligger 3.6 % over året før (2010), eller om finanskrisa har gitt en dårligere lønnsutvikling. Du kan her anta at standardavviket i årsinntekt for 2011 er kjent, og lik det for 2010. Formuler først hypotesene for denne testen. Hvilke antagelser må gjøres for at en slik test skal kunne utføres?

d) Utfør testen med et signifikansnivå lik 0.05. Hva blir konklusjonen på testen?

Fra sannsynlighetsfordelingen i oppgave a), med $k = 214.9$ og θ lik punktestimatet fra b), trekker vi 1000 tilfeldige utvalg av størrelse $n = 30$. Figur 1 viser et histogram over gjennomsnittsverdiene fra disse utvalgene. Ser det ut som om $n = 30$ er en tilstrekkelig utvalgsstørrelse for å bruke sentralgrenseteoremet for denne fordelingen? Begrunn svaret.



Oppgave 3 Gruvedrift

I ressursevaluering av mineralforekomster samler man inn ulike data for å undersøke utvinningspotensialet i forbindelse med gruvedrift. Her ser vi på kjerneprøver som gir Y_i = mengde mineral per volumenhet av prøve $i = 1, \ldots, n$. I tillegg vet vi bergart x_i, $i = 1, \ldots, n$ som kategoriske variable. Vi ønsker å tilpasse en lineær regresjonsmodell for responsen Y_i, der bergart x_i brukes som forklaringsvariabel (kovariat). Statistisk beskrives modellen ved:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \qquad i = 1, \ldots, n,$$

der ϵ_i, $i = 1, \ldots, n$, er uavhengige og normalfordelte stokastiske variabler med forventning 0 og antatt kjent varians $\sigma^2 = 0.5^2$.

Ved hjelp av minste kvadraters metode kan vi finne estimatorer for β_0 og β_1. De er:

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

der $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ og $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$.

a) Vis at $\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$.

Vis at $\hat{\beta}_1$ er en forventningsrett estimator for β_1.

Et firma samler inn $n = 7$ kjerneprøver, der responser $y_1, \ldots, y_7$, er målt i områder med ulike klasser bergart x_i, $i = 1, \ldots, 7$. Resultatet oppsummeres som følger.

Type bergart (x)	1	1	2	2	2	3	3
Resultat kjerneprøve (y)	2.7	3.1	3.9	3.2	4.2	4.1	5.5

b) Utled et 90 % konfidensintervall for β_1. Finn tallverdier for intervallet.

Medarbeiderne vurderer om konfidensintervallet kunne blitt kortere dersom de istedet tok tre kjerneprøver i bergart 1 og 3, og bare en prøve i bergart 2? Eller hva om de tok en prøve i bergart 1 og en i 3, men fem i bergart 2? Gi reflekterte svar på disse spørsmålene.

c) Firmaet skal ta en ny kjerneprøve et sted der bergart er $x_0 = 3$.

Utled et 90 % prediksjonsintervall for den nye kjerneprøven, og finn tallverdier basert på data i tabellen over.