Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Fagleg kontakt under eksamen:
Jo Eidsvik 90127472
Ola Diserud 93218823

EKSAMEN I FAG TMA4240 STATISTIKK

Mandag 12.des 2011

Tid: 09:00–13:00

Tiletne hjelpemiddel:
Gult, stempla A5-ark med eige handskrivne notat.
Tabeller og formler i statistikk (Tapir Forlag).
K. Rottmann: *Matematisk formelsamling.*
Kalkulator: HP30S.

NYNORSK

Sensur: 5.jan 2012.

Oppgåve 1 Oljeleiting

a) La A og B vere to hendingar i eit utfallsrom. Det er gjeve at $P(A) = 0.3$, $P(B) = 0.3$ og $P(A \cup B) = 0.6$.

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

Oljefeltet Aldous Major / Avaldsnes blei funnet på grunn av mindre liknande funn i næleiken. Avhengnadsstruktur i modellen for oljefunn gav auka sannsyn for å finne olje på Aldous Major / Avaldsnes.

Vi tenkjer oss to oljefelt. Hendinga A = oljefunn på felt 1, medan den komplementære hendinga A^c = ingen olje på felt 1. Tilsvarande er hendinga B = oljefunn på felt 2, medan B^c = ingen olje på felt 2. Vi får gjeve 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) Teikn eit Venn diagram for hendingane.

Finn sannsynet for olje på felt 1.

Anta at ein har funne olje på felt 2. Kva er nå sannsynet for olje på felt 1?

Anta at ein har påvist at felt 2 ikkje inneheld olje. Kva er nå sannsynet for olje på felt 1?

Er hendingane A og B uavhengige?

Vi ser for oss ein kostnad $K = 100$ millionar kroner ved å leite etter olje. Dersom du ikkje finn olje får du ingen gevinst, men betaler denne kostnaden. Dersom du finn olje, får du ein stor gevinst, som overstiger kostnaden K. Anta at profitten ved oljefunn på felt 1 er $R_1 = 500 - K = 400$ millionar kroner, mens profitten ved oljefunn på felt 2 er $R_2 = 1100 - K = 1000$ millionar kroner.

c) Rekn ut forventa profitt ved å leite etter olje på felt 1.

Anta at du har funne olje på felt 2. Kva er nå forventa profitt ved å leite etter olje på felt 1?

Anta at du kan velje mellom følgjande leitestrategiar: Ikkje leite nokon stad, leite ved felt 1, eller leite ved felt 2. Dersom du vel å leite ved felt 1 eller 2, kan du avhengig av utfallet stoppe, eller leite vidare på det andre feltet. Avgjeringa tas frå forventa verdi. Kva for ein leitestrategi gir høgast forventa profitt, og kva er forventa profitt under denne strategien?

Oppgåve 2 Lønningar

La den tilfeldige (stokastiske) variabelen Y skildre ein persons årslønn for 2011 i kNOK (1 kNOK er 1000 norske kronar). Paretos lov påstår da at

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

kor k er minimumsinntekta for heile populasjonen.

a) Anta at Paretos lov stemmer. Vis at sannsynsfordelinga (tettleiken) 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 vidare at $k = 214.9$ kNOK, dvs første lønnstrinn i Statens lønnsregulativ for 2011. Frå eit tilfeldig utval på 30 årslønnar finn du at $\sum_{i=1}^{30} y_i = 13611$ og $\sum_{i=1}^{30} \ln y_i = 174.7$.

b) Vis at sannsynsmaksimeringsestimatoren (maximum likelihood estimator) for θ, frå eit tilfeldig utval med storleik n, blir

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

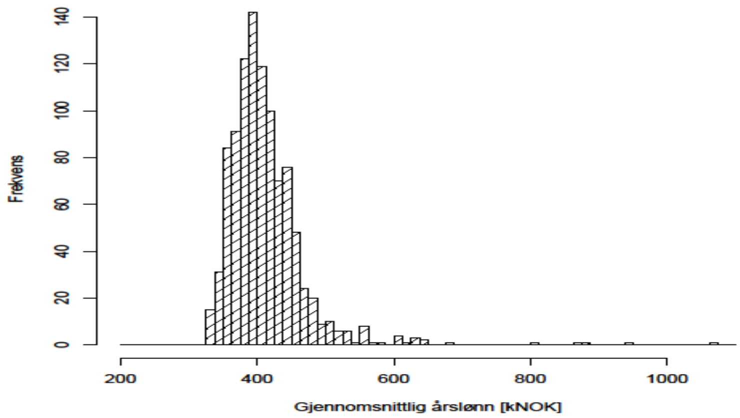
Kva blir sannsynet for at ein tilfeldig person skal ha årslønn under 472.2 kNOK, dvs grensa for toppskatt, dersom vi brukar punktestimatet for θ som parameterverdi?

c) For 2010 var gjennomsnittslønna for alle norske tilsette 435 kNOK, med standardavvik 516 kNOK. Gjennomsnittlig årslønn for 2010 hadde da stige med 3.6 % frå til 2009.

Du skal nå teste om forventa årslønn for 2011 og ligg 3.6 % over året før (2010), eller om finanskrisa har gitt ein dårligare lønnsutvikling. Du kan her anta at standardavviket i årsinntekt for 2011 er kjend, og lik det for 2010. Formuler først hypotesane for denne testen. Kva for nokre antakingar er påkravd for at ein slik test skal kunne gjennomførast?

d) Gjennomfør testen med eit signifikansnivå lik 0.05. Kva blir konklusjonen på testen?

Frå sannsynsfordelinga i oppgave a), med $k = 214.9$ og θ lik punktestimatet frå b), trekkjer vi 1000 tilfeldige utval av storleik $n = 30$. Figur 1 viser eit histogram over gjennomsnittsverdiane frå disse utvalene. Ser det ut som om $n = 30$ er ein tilstrekkeleg utvalstorleik for å bruke sentralgrenseteoremet for denne fordelinga? Grunngi svaret.



Oppgåve 3 Gruvedrift

I ressursevaluering av mineralførekomstar samlar ein inn ulike data for å undersøkje utvinningspotensialet med gruvedrift. Her ser vi på kjerneprøvar som gjev Y_i = mengde mineral per volumeining av prøve $i = 1, \ldots, n$. I tillegg veit vi bergart x_i, $i = 1, \ldots, n$, som kategoriske variablar. Vi ønskjer å tilpasse ein lineær regresjonsmodell for responsen Y_i, der bergartstype x_i brukast som forklaringsvariabel (kovariat). Statistisk skildrast 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 variablar med forventning 0 og kjend varians $\sigma^2 = 0.5^2$.

Ved hjelp av minste kvadrats metode kan vi finne estimatorar for β_0 og β_1. Dei 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 ein forventningsrett estimator for β_1.

Eit firma samlar inn $n = 7$ kjerneprøvar, der responser $y_1, \ldots, y_7$, er målte i områder med ulike klassar bergart x_i, $i = 1, \ldots, 7$. Resultatet blir som følgjer.

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) Utlei eit 90 % konfidensintervall for β_1. Finn tallverdiar for intervallet.

Medarbeidarane vurderer om konfidensintervallet kunne blitt kortare dersom dei i staden tok tre kjerneprøvar i bergart 1 og 3, og berre ein prøve i bergart 2? Eller kva om dei tok ein prøve i bergart 1, ein i 3, og fem i bergart 2? Gi reflekterte svar på disse spørsmåla.

c) Firmaet skal ta ein ny kjerneprøve ein stad der bergart er $x_0 = 3$.

Utlei eit 90 % prediksjonsintervall for den nye kjerneprøven, og finn tallverdiar frå data i tabellen over.