Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Faglige kontakter under eksamen:
Bo Lindqvist 975 89 418

EKSAMEN I FAG TMA4240 STATISTIKK

Torsdag 5. august 2004

Tid: 09:00–14:00

Tillatte hjelpemidler:
Tabeller og formler i statistikk (Tapir Forlag).
K. Rottmann: *Matematisk formelsamling.*
Kalkulator: HP30S.

BOKMÅL

Sensur: 26. august 2004.

Oppgave 1 Forurensning

En målestasjon for luftforurensning registrerer innholdet av såkalt inhalerbart støv (dvs. partikler med midlere diameter opptil 1/100 mm) i lufta. La X (med enhet milliondels gram pr m^3) være en måling av innholdet av inhalerbart støv på en tilfeldig valgt dag. Målinger av X på ulike dager antas stokastisk uavhengige.

a) Under vanlige forhold antas at X er normalfordelt med forventning $\mu = 35$ og varians $\sigma^2 = 25$. (Dette skal antas bare i dette punktet).

Finn sannsynligheten for at en måling X er over 40, $P(X > 40)$.

Finn også sannsynligheten for at X er mellom 30 og 40, $P(30 < X < 40)$.

Finn sannsynligheten for at summen av målinger på to ulike dager er over 80. Anta at de to målingene er uavhengige.

I en kommune finnes n slike målestasjoner. Fra disse får vi i løpet av en dag målinger $X_1, X_2, \ldots, X_n$ som antas uavhengige og identisk normalfordelte med forventning μ og varians σ^2, der både μ og σ er ukjente parametre. Her vil μ angi graden av forurensning.

Anta at $n = 5$ og at det en dag gjøres følgende målinger av $X_1, \dots, X_5$:

$252, 311, 268, 287, 302$

For senere bruk oppgis at $\sum_{i=1}^{5} X_i = 1420$ og $\sum_{i=1}^{5}(X_i - \bar{X})^2 = 2342$.

b) Hvilke egenskaper bør en god estimator ha?

Sett opp en forventningsrett estimator $\hat{\mu}$ for μ basert på $X_1, \dots, X_n$. Finn variansen til $\hat{\mu}$. (Vis hvordan du regner den ut).

Sett også opp en estimator for σ^2 og bruk denne til å finne en estimator for variansen til $\hat{\mu}$.

c) Utled et 95% konfidensintervall for μ.

Regn ut tallsvar for intervallgrensene når dataene er som gitt ovenfor.

Regjeringen har fastsatt regler som krever spesielle tiltak for å redusere forurensningen dersom $\mu \geq 300$. Kommunen skal på grunnlag av estimatoren $\hat{\mu}$ beslutte om slike tiltak skal settes i verk. Man stiller følgende krav til den beslutningsregel som skal brukes:

(i) Hvis $\mu \geq 300$ skal det være minst 90% sannsynlighet for å sette i verk tiltak.

(ii) Hvis $\mu < 300$ vil man helst unngå å sette i verk tiltak.

d) Forklar hvorfor den ønskede regelen kan baseres på en test av

$$H_0 : \mu \geq 300 \text{ mot } H_1 : \mu < 300$$

med signifikansnivå $\alpha = 0.10$. Hvilken beslutning skal kommunen velge dersom H_0 forkastes? Hva vil feil av type I og type II for testen bety i kommunens beslutningsproblem?

e) Gjennomfør testingen når observerte målinger er som gitt tidligere i oppgaven. Hvilken beslutning velger kommunen i dette tilfellet?

Anta at det egentlig er 6 målestasjoner i kommunen, men at det på den dagen de 5 observasjonene ovenfor ble gjort, skjedde en feil ved avlesningen fra den siste stasjonen.

f) Utled et intervall som med en sannsynlighet på 95% inneholder den ukjente målingen ved denne stasjonen.

Hva kaller vi et slikt intervall? Hvorfor er dette intervallet bredere enn konfidensintervallet fra punkt c)?

Oppgave 2 Sykehjemmet

Vi ser på dødsfall om natten ved sykehjemmet "Aftensol". Ved sykehjemmet er det tre sykepleiere i rene nattevaktstillinger, Anne, Bernt og Cecilie. Hver natt er en av dem på vakt gjennom hele natten, og det er da ingen andre ansatte tilstede ved hjemmet. Anne jobber i 100% nattevaktstilling, mens Bernt og Cecilie jobber i 50% nattevaktstillinger.

Vi ser på en tilfeldig valgt natt og definerer følgende hendelser:

$A =$ Anne er på vakt,
$B =$ Bernt er på vakt,
$C =$ Cecilie er på vakt,
$D =$ det skjer et dødsfall.

Anta at alle dødsfall er naturlige. Det er da rimelig å gå ut fra at sannsynligheten for dødsfall er den samme uansett hvilken sykepleier som er på vakt, dvs. at $P(D|A) = P(D|B) = P(D|C)$. Anta at den felles verdi for disse er 0.06.

a) Tegn de 4 hendelsene definert ovenfor i et venndiagram.

Hva er sannsynlighetene $P(A), P(B)$ og $P(C)$?

Finn $P(D)$. Er hendelsene D og C uavhengige? Begrunn svaret.

I den siste tiden har det vært 10 dødsfall om natten ved sykehjemmet, og hele 7 av disse har skjedd når Cecilie har vært på vakt. Det er derfor satt igang etterforskning for eventuelt å avdekke om Cecilie har noe med dødsfallene å gjøre.

Anta i det følgende at alle dødsfallene er naturlige, og at de har skjedd på forskjellige netter.

La X være en stokastisk variabel som beskriver antall av $n = 10$ naturlige dødsfall som skjer på Cecilies vakter.

b) Forklar hvorfor det kan antas at X er binomisk fordelt med $n = 10$ og $p = 0.25$. (Det er ikke tilstrekkelig å skrive opp de generelle forutsetningene for en binomisk fordeling, betingelsene må relateres direkte til situasjonen som er beskrevet.)

Hva er sannsynligheten for at 7 eller flere av 10 dødsfall om natten skjer på Cecilies vakter?

La oss tenke oss at det rundt om på sykehjem i Norge jobber 300 andre sykepleiere i tilsvarende stilling som Cecilie. Hva er sannsynligheten for at minst en av de 300 sykepleierne opplever at 7 eller flere av 10 naturlige dødsfall skjer på sine vakter?

Gir svarene i dette punktet grunn til å styrke mistanken mot Cecilie? Begrunn svaret.

Oppgave 3 Alpinulykker

Sikkerhet er en av de høyest prioriterte oppgavene i norske alpinanlegg. Vi antar at antall alpinulykker som krever legebehandling i alpinanlegget "Alpinfjellet" i en periode på t skidager, X, er Poisson-fordelt med forventningsverdi $\mu = \lambda t$. Her er λ skadefrekvens pr skidag og t er eksponeringstid i antall skidager. En skidag er definert som "en person i alpinanlegget en hel dag".

Det følger altså at punktsannsynligheten for X er

$$f(x) = \frac{(\lambda t)^x}{x!}\exp(-\lambda t); \quad \text{for} \quad x = 0, 1, 2, \ldots$$

a) Alpinanleggenes Landsforening har, basert på data fra noen av de største skianleggene i Norge, anslått at $\lambda = 1/1000$. Vi antar i dette punktet at det er kjent at $\lambda = 1/1000$ for "Alpinfjellet".

Hvis vi ser på $t = 2000$ skidager, hva er da sannsynligheten for at det skjer akkurat én ulykke, $P(X = 1)$?

Hvis du tilbringer 10 skidager i "Alpinfjellet", hva er sannsynligheten for at du utsettes for en eller flere ulykker?

Hvor mange skidager må du tilbringe i "Alpinfjellet" for at din sannsynlighet for minst en ulykke skal bli større enn 0.1?

Ved "Alpinfjellet" har man aktivt registert samhørende verdier av antall ulykker, X_i, og eksponering i antall skidager, t_i $(i = 1, ..., n)$, for n tilfeldig valgte dager anlegget var åpent. Vi antar at $X_1, \ldots, X_n$ er uavhengige.

b) Finn sannsynlighetsmaksimeringsestimatoren (SME) for λ basert på de n observasjonsparene $(X_1, t_1), (X_2, t_2), \ldots, (X_n, t_n)$.

Er estimatoren forventningsrett?