Noregs teknisk-naturvitskaplege universitet
Institutt for matematiske fag

Nynorsk

Fagleg kontakt under eksamen:
Øyvind Bakke, telefon 73 59 81 26
Odd Kolbjørnsen, telefon 73 59 35 27

SIF5062/SIF5506 Statistikk

Laurdag 25. mai 2002
Kl. 9–14

Hjelpemiddel: *Tabeller og formler i statistikk,* Tapir Akademisk Forlag
K. Rottmann: *Matematisk formelsamling*
Bestemt, enkel kalkulator

Sensur: 29. juni 2002

Alle punkta tel likt ved vurderinga av eksamenssvaret. Alle svara skal grunngjevast.

Oppgåve 1

På Botanisk forskingsstasjon er det planta eit felt med ein sjeldsynt grasart. På eit visst tidspunkt er lengda X målt i cm av eit tilfeldig valt grasstrå eksponentielt fordelt, dvs. at X har sannsynstettleik f gjeve ved at $f(x) = \frac{1}{\beta}e^{-x/\beta}$ for $x \geq 0$ og kumulativ fordelingsfunksjon F gjeve ved at $F(x) = 1 - e^{-x/\beta}$ for $x \geq 0$.

a) Lat (berre i dette punktet) $\beta = 10$. Rekn ut

$$P(X \leq 4), \quad P(X > 7) \quad \text{og} \quad P(X > 7 \mid X > 4).$$

Vi reknar med at lengdene av stråa er uavhengige. Planen er å måle lengdene i eit tilfeldig utval av stråa for å estimere β.

b) Utlei sannsynsmaksimeringsestimatoren (SME) for β basert på lengdene i eit tilfeldig utval på k strå.

Ved ei misforståing slår vaktmeisteren graset same dagen som målingane skal gjerast, og pløyer opp feltet etterpå. Forskaren planlegg no å måle lengda Y på eit utval av stråa som ligg i oppsamlaren på grasklipparen, og basere estimeringa på desse lengdene. Alle stråa vart klipte i same høgd c, og berre dei som var høgre enn klippehøgda vart klipte. Gjeve at eit strå hadde lengd $X > c$, ligg det altså i oppsamlaren, og har lengd $Y = X - c$.

c) Vis at $P(X > c) = e^{-c/\beta}$.

Finn $P(Y > y)$, der $y > 0$.

Kva kjend fordeling har Y?

Diverre viser det seg at stråa i oppsamlaren har tørka inn og krympa, slik at det ikkje er mogleg å måle lengdene. Men det er mogleg å telje stråa, og oppsamlaren var tom då vaktmeisteren starta slåtten. Dessuten kjenner forskaren talet på strå som vart planta ut i feltet og som var i live denne dagen (daude strå vart fjerna frå feltet etter kvart).

d) Lat Z vere talet på strå i oppsamlaren. Det var n strå i live på feltet. Kva for føresetnader må vere oppfylte for at Z skal vere binomisk fordelt?

Vi reknar med at Z er binomisk fordelt. Uttrykk suksessannsynet ved β og c.

Føreslå ein estimator for β basert på Z.

Oppgåve 2

På eit analyselaboratorium driv ein ein kontinuerleg kvalitetskontroll. Kvar gong ein serie prøver blir analysert, blir òg ei kontrolløysing med kjend konsentrasjon 0,10 mg/l analysert. Utfallet av analysen av kontrolløysinga kan reknast som normalfordelt med forventningsverdi μ og varians σ^2, der σ^2 er variansen i målefeilen ved analysemetoden og μ under normale omstende er lik 0,10 mg/l.

Ei *alarmhending* A er definert ved at målt verdi X i kontrolløysinga avvik meir enn 2 standardavvik frå konsentrasjonen 0,10 mg/l, altså $|X - 0{,}1| > 2\sigma$ (dvs. $X < 0{,}1 - 2\sigma$ eller $X > 0{,}1 + 2\sigma$). Ei *aksjonshending* B er definert ved at målt verdi X i kontrolløysninga avvik meir enn 3 standardavvik frå 0,10 mg/l, dvs. $|X - 0{,}1| > 3\sigma$.

a) Lat (berre i dette punktet) $\sigma = 0{,}01$ mg/l.

Rekn ut $P(B)$ og $P(B \mid A)$ når $\mu = 0{,}10$ mg/l.

Tenk deg no at ei urening har snike seg inn i prøva, slik at $\mu = 0{,}11$ mg/l, og rekn no ut $P(B)$.

Lat $\mu = 0{,}10$ mg/l i resten av oppgåva. For å estimere σ^2 blir uavhengige resultat $X_1, X_2, \ldots, X_n$ frå fleire analysar av kontrolløysinga nytta. To estimatorar for σ^2 er foreslåtte,

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2,$$

der $\mu = 0{,}10$ mg/l er den kjende konsentrasjonen i kontrolløysninga, og

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

der $\overline{X}$ er den empiriske middelverdien (utvalsmiddelverdien).

I dei neste punkta kan du bruke utan bevis at $\sum_{i=1}^{n}(X_i - \mu)^2/\sigma^2$ er χ^2-fordelt (khikvadratfordelt) med n fridomsgrader og at $\sum_{i=1}^{n}(X_i - \overline{X})^2/\sigma^2$ er χ^2-fordelt med $n-1$ fridomsgrader.

b) Kva for to eigenskapar kjenneteiknar ein god estimator?

Kva for ein av dei to estimatorane $\widehat{\sigma^2}$ og S^2 vil du tilrå?

Resultata målt i mg/l av 20 analysar av kontrolløysinga er:

0,0939	0,1069	0,1002	0,1106	0,0866	0,1048	0,0837	0,0856	0,1029	0,0986
0,0887	0,0971	0,0942	0,0910	0,1025	0,0851	0,1031	0,0797	0,1053	0,1034

Det blir gjeve opp at $\sum_{i=1}^{20} x_i = 1{,}9240$ og at $\sum_{i=1}^{20} x_i^2 = 0{,}1866$.

c) Utlei eit 90 %-konfidensintervall for σ^2 ved å nytte favorittestimatoren din frå (b).

d) Ein ynskjer eit 90 %-konfidensintervall med lengd høgst 50 % av punktestimatet. Kor mange observasjonar trengst? Rund av svaret opp til næraste talet deleleg med 10 (10, 20, 30, . . .). (Vink: Prøv deg fram med tal frå tabellen i *Tabeller og formler i statistikk.*)

Oppgåve 3

Eit kommunalt kloakkreinseverk har eit utsleppsløyve for fosfor på 0,20 mg/l. Ein miljøorganisasjon meiner reinseverket er utslite og ikkje held utsleppskrava. For å underbyggje påstandane sine, får dei analysert fosforinnhaldet i ti prøver av utsleppa frå reinseverket, X_1, X_2, $\ldots$, X_{10}. Det er rimeleg å rekne med at prøvene er uavhengige og normalfordelte med forventningsverdi lik det verkelege fosforinnhaldet og *kjent* standardavvik på 0,02 mg/l.

Resultata målt i mg/l av prøvene vart:

0,2188	0,1782	0,1960	0,1884	0,2300	0,2446	0,2242	0,1950	0,2146	0,2055

Det blir gjeve opp at $\sum_{i=1}^{10} x_i = 2{,}0954$.

Miljøorganisasjonen ynskjer å godtgjere at utsleppa frå reinseverket er høgre enn 0,20 mg/l.

a) Formuler problemstillinga som ein hypotesetest og rekn ut eit forkastingsområde. Nytt signifikansnivå 0,05. Kva blir konklusjonen på testen når dataa er som over?

Miljøorganisasjonen meiner konklusjonen frå forsøket kom av at forsøket var for dårleg planlagt, og at ti prøver ikkje var nok for å avdekkje eit for høgt utslepp.

b) Kor stort er sannsynet for å oppdage avviket ved bruk av testen over på ti nye prøver dersom utsleppet frå reinseverket er på 0,21 mg/l?

Kor mange observasjonar er naudsynte for at eit utslepp på 0,21 mg/l skal oppdagast med 90 % sannsyn?