Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Bokmål

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

SIF5062/SIF5506 Statistikk

Lørdag 25. mai 2002
Kl. 9–14

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

Sensur: 29. juni 2002

Alle punktene teller likt ved vurderingen av besvarelsen. Alle svar skal begrunnes.

Oppgave 1

På Botanisk forskningsstasjon er det plantet et felt med en sjelden grasart. På et bestemt tidspunkt er lengden X målt i cm av et tilfeldig valgt grasstrå eksponentielt fordelt, dvs. at X har sannsynlighetstetthet f gitt ved at $f(x) = \frac{1}{\beta}e^{-x/\beta}$ for $x \geq 0$ og kumulativ fordelingsfunksjon F gitt ved at $F(x) = 1 - e^{-x/\beta}$ for $x \geq 0$.

a) Anta (bare i dette punktet) at $\beta = 10$. Regn ut

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

Vi antar at lengdene av stråene er uavhengige. Planen er å måle lengdene i et tilfeldig utvalg av stråene for å estimere β.

b) Utled sannsynlighetsmaksimeringsestimatoren (SME) for β basert på lengdene i et tilfeldig utvalg på k strå.

Ved en misforståelse slår vaktmesteren graset samme dag som målingene skal gjøres, og pløyer deretter opp feltet. Forskeren planlegger nå å måle lengden Y på et utvalg av stråene som ligger i grasklipperens oppsamler, og basere estimeringen på disse lengdene. Alle stråene ble klipt i samme høyde c, og bare de som var høyere enn klippehøyden ble klipt. Gitt at et strå hadde lengde $X > c$, ligger det altså i oppsamleren, og har lengde $Y = X - c$.

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

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

Hvilken kjent fordeling har Y?

Dessverre viser det seg at stråene i oppsamleren har tørket inn og krympet, slik at det ikke er mulig å måle lengdene. Men det er mulig å telle stråene, og oppsamleren var tom da vaktmesteren startet slåingen. Dessuten kjenner forskeren antall strå som ble plantet ut i feltet og som var i live denne dagen (døde strå ble fjernet fra feltet fortløpende).

d) La Z være antall strå i oppsamleren. Anta at det var n strå i live på feltet. Hvilke betingelser må være oppfylt for at Z skal være binomisk fordelt?

Anta at Z er binomisk fordelt. Uttrykk suksessannsynligheten ved β og c.

Foreslå en estimator for β basert på Z.

Oppgave 2

På et analyselaboratorium drives en kontinuerlig kvalitetskontroll. Hver gang en serie prøver analyseres, analyseres også en kontrolløsning med kjent konsentrasjon 0,10 mg/l. Utfallet av analysen av kontrolløsningen kan regnes som normalfordelt med forventningsverdi μ og varians σ^2, der σ^2 er variansen i målefeilen ved analysemetoden og μ under normale omstendigheter er lik 0,10 mg/l.

En *alarmhendelse* A er definert ved at målt verdi X i kontrolløsningen avviker mer enn 2 standardavvik fra konsentrasjonen 0,10 mg/l, altså $|X - 0{,}1| > 2\sigma$ (dvs. $X < 0{,}1 - 2\sigma$ eller $X > 0{,}1 + 2\sigma$). En *aksjonshendelse* B er definert ved at målt verdi X i kontrolløsningen avviker mer enn 3 standardavvik fra 0,10 mg/l, dvs. $|X - 0{,}1| > 3\sigma$.

a) Anta (bare i dette punktet) at $\sigma = 0{,}01$ mg/l.

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

Anta så at en urenhet har sneket seg inn i prøven, slik at $\mu = 0{,}11$ mg/l, og beregn nå $P(B)$.

Anta i resten av oppgaven at $\mu = 0{,}10$ mg/l. For å estimere σ^2 benyttes resultatene X_1, X_2, $\ldots$, X_n, som er uavhengige, fra flere analyser av kontrolløsningen. To estimatorer for σ^2 er foreslått,

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

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

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

der $\overline{X}$ er empirisk middelverdi (utvalgsmiddelverdi).

I de neste punktene kan du bruke uten bevis at $\sum_{i=1}^{n}(X_i - \mu)^2/\sigma^2$ og $\sum_{i=1}^{n}(X_i - \overline{X})^2/\sigma^2$ er χ^2-fordelte (khikvadratfordelte) med henholdsvis n og $n-1$ frihetsgrader.

b) Hvilke to egenskaper kjennetegner en god estimator?

Hvilken av de to estimatorene $\widehat{\sigma^2}$ og S^2 vil du anbefale?

Resultatene målt i mg/l av 20 analyser av kontrolløsningen 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 oppgis at $\sum_{i=1}^{20} x_i = 1{,}9240$ og at $\sum_{i=1}^{20} x_i^2 = 0{,}1866$.

c) Utled et 90 %-konfidensintervall for σ^2 ved å benytte din favorittestimator fra (b).

d) En ønsker et 90 %-konfidensintervall med lengde høyst 50 % av punktestimatet. Hvor mange observasjoner trengs? Oppgi svaret avrundet opp til nærmeste tall delelig på 10 (10, 20, 30, $\ldots$). (Vink: Prøv deg fram med tall fra tabellen i *Tabeller og formler i statistikk.*)

Oppgave 3

Et kommunalt kloakkrenseanlegg har en utslippstillatelse for fosfor på 0,20 mg/l. En miljøorganisasjon mener anlegget er nedslitt og ikke oppfyller utslippskravene. For å underbygge sine påstander, får de analysert fosforinnholdet i ti prøver av utslippene fra renseanlegget, $X_1,\ X_2,\ \ldots,\ X_{10}$. Det er rimelig å anta at prøvene er uavhengige og normalfordelte med forventningsverdi lik det virkelige fosforinnholdet og *kjent* standardavvik på 0,02 mg/l.

Resultatene målt i mg/l av prøvene ble:

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

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

Miljøorganisasjonen ønsker å påvise at utslippene fra anlegget er høyere enn 0,20 mg/l.

a) Formuler problemstillingen som en hypotesetest og beregn et forkastningsområde. Benytt signifikansnivå 0,05. Hva blir konklusjonen på testen når dataene er som over?

Miljøorganisasjonen mener konklusjonen fra forsøket kom av at forsøket var for dårlig planlagt, og at ti prøver ikke var tilstrekkelig for å avsløre et for høyt utslipp.

b) Hvor stor er sannsynligheten for å oppdage avviket ved bruk av testen over på ti nye prøver hvis utslippet fra anlegget er på 0,21 mg/l?

Hvor mange observasjoner er nødvendig for at et utslipp på 0,21 mg/l skal oppdages med 90 % sannsynlighet?