Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Faglig kontakt under eksamen:
Magnar Lillegård, tlf. 73 59 35 38
Odd Kolbjørnsen, tlf. 73 59 35 27

Eksamen i SIF5062 og SIF5506 Statistikk

Mandag 28. mai 2001
Tid: Kl. 09.00–14.00
Tillatte hjelpemidler:
Tabeller og formler i statistikk / Statistiske tabeller og formler
Enkel lommeregner.
Sensurdato: Mandag 2. juli 2001

Oppgave 1

Vi ser på konsentrasjonen av et giftstoff i havbunnen like utenfor en fabrikk. Miljøforskriftene sier at konsentrasjonen ikke skal overstige 12 [g/cm^3]. For å kontrollere dette tas prøver av havbunnen. Anta at en prøveverdi Y er normalfordelt med forventning μ og standardavvik σ. Sett $\mu = 13$ og $\sigma = 1{,}5$ i punkt a), og la de være ukjent i resten av oppgaven.

a) Beregn $P(Y < 12)$ og $P(11 < Y < 14)$.

b) De observerte måleverdiene er

$$11{,}7 \quad 12{,}4 \quad 12{,}8 \quad 12{,}9 \quad 13{,}3.$$

Kan vi på grunnlag av dette konkludere med at giftkonsentrasjonen på havbunnen like ved fabrikken er over 12? Formuler problemstillingen som en hypotesetest og utfør testen på signifikansnivå 0,05.

c) Det blir tatt 10 nye målinger, men denne gang i ulike avstander x fra fabrikken. Målingene er

x	10	20	30	40	50	60	70	80	90	100
y	9,9	11,1	9,3	10,6	9,2	9,3	10,0	9,2	10,3	8,4

I tillegg kommer de fem målingene i b). Her er $x = 0$. Det oppgis at $\sum x_i = 550$, $\sum(x_i - \bar{x})^2 = 18\,333{,}33$, $\sum y_i = 160{,}4$ og $\sum x_i y_i = 5245$.

Vi velger å utføre en lineær regresjonsanalyse med Y som avhengig variabel og x som uavhengig variabel. Modellen er

$$E(Y \mid x) = \alpha + \beta x.$$

Beregn estimatene for α og β. Forklar hva estimatet for α beskriver i dette eksemplet.

Regresjonsanalysen gir oss ikke grunnlag for å konkludere med at $\alpha > 12$. Hvorfor ikke? Sammenlign resultatet fra denne analysen med resultatet i b) og kommenter. Hvorfor kan det skje at to slike analyser gir forskjellig konklusjon? Bruk gjerne figur i forklaringen.

Oppgave 2

Det skal tas blodprøve av samtlige soldater i en militæravdeling for å finne ut om noen er smittet av en bestemt sykdom. Sannsynligheten for at blodprøven fra en tilfeldig valgt soldat er positiv (dvs. at han er smittet) antas å være lik p, og testresultatene for ulike individer antas å være uavhengige.

Man går frem på følgende måte: Blodprøver tas av $k(> 1)$ soldater om gangen, blodprøvene blandes og blandingen analyseres. Hvis den gir positiv reaksjon, noe som betyr at minst en av soldatene er smittet, tas nye blodprøver av de k, og prøvene analyseres enkeltvis. Hvis blandingen gir negativ reaksjon, blir det ikke tatt flere prøver.

a) Begrunn hvorfor antall smittede i en k-gruppe er binomisk fordelt og vis at sannsynligheten for at en blanding av k blodprøver skal gi positiv reaksjon er

$$1 - (1-p)^k.$$

b) Hvordan defineres betinget sannsynlighet for en hendelse A gitt en annen hendelse B? Soldat 19 Haugen får beskjed om at hans k-gruppe har vist positiv reaksjon. Hva er da sannsynligheten for at han er positiv?

c) Anta at en avdeling består av m grupper, hver med k soldater, slik at det totale antallet soldater er mk. La X være antall blodprøver som må analyseres før hele avdelingen er ferdig kontrollert. Vis at

$$E(X) = mk\left(1 + \frac{1}{k} - (1-p)^k\right).$$

Hvis $k = 4$, for hvilke verdier av p er den benyttede fremgangsmåten å foretrekke i stedet for å analysere blodprøvene enkeltvis med en gang?

Oppgave 3

La X være gammafordelt med sannsynlighetstetthet f gitt ved

$$f(x) = \frac{\theta^4}{6} x^3 e^{-\theta x}, \quad x > 0.$$

a) Skriv opp uttrykket for den momentgenererende funksjonen (MGF) til en tilfeldig variabel, og vis at MGF til X er gitt ved

$$M_X(t) = \left(1 - \frac{t}{\theta}\right)^{-4}, \quad t < \theta.$$

Finn forventningen til X ved bruk av M_X.

b) La $X_1, X_2, \ldots, X_n$ være et tilfeldig utvalg fra f. Vis at en estimator $\hat{\theta}$, basert på sannsynlighetsmaksimering ("maksimum likelihood"), for den ukjente parameteren θ kan skrives som

$$\hat{\theta} = \frac{4n}{\sum_{i=1}^{n} X_i}.$$

c) Vis at $2\theta \sum_{i=1}^{n} X_i$ er χ^2-fordelt med $8n$ frihetsgrader. Hint: Bruk momentgenererende funksjon.

Bruk dette resultatet til å konstruere et 95% konfidensintervall for θ. Beregn intervallet når observasjonene er $x_1 = 0{,}8$, $x_2 = 1{,}3$ og $x_3 = 2{,}4$.

Oppgave 4

En student greier å løse fire av ni delspørsmål i de tre foregående oppgavene riktig. Anta at alle kombinasjoner av fire riktige svar er like sannsynlige. La X være antall riktige svar på oppgave 1 (tre delspørsmål). Finn fordelingen til X.