Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Faglig kontakt under eksamen:
Jan Terje Kvaløy 73 59 16 96
Stian Lydersen 73 59 70 53

EKSAMEN I FAG SIF 5505 STATISTIKK

Lørdag 9. januar 1999
Tid: 09:00–14:00

Hjelpemidler:
Godkjent lommekalkulator, utdelt ordliste, Statistiske tabeller og formler (Tapir forlag).

Oppgave 1 Medisinkonsentrasjon

Ved behandling av visse kreftformer får pasientene kurer der en bestemt type medisin blir injisert i blodet i løpet av 24 timer. Alle pasienter får tilført samme dose medisin. Ved avslutningen av kuren blir konsentrasjonen av medisin i blodet målt. Medisinkonsentrasjonen måles i milligram medisin per liter blod. For at behandlingen skal ha ønsket effekt bør medisinkonsentrasjonen ved avslutningen av kuren helst overstige 5 mg/l. På grunn av bivirkninger blir det ansett som uheldig om medisinkonsentrasjonen overstiger 12 mg/l. La Y betegne målt medisinkonsentrasjon ved avslutningen av en kur, og anta at Y er normalfordelt med forventning μ og varians σ^2. Målt medisinkonsentrasjon ved avslutningen av ulike kurer antas uavhengige. Anta i første omgang at $\mu = 8$ og $\sigma^2 = 2^2$.

a) Beregn sannsynlighetene $P(Y \leq 12)$, $P(Y > 5)$ og $P(5 < Y \leq 12)$.

Dersom en pasient går gjennom 8 kurer, hva er sannsynligheten for at målt medisinkonsentrasjon ved slutten av samtlige 8 kurer er i intervallet (5,12]?

Følgende hendelser er definert:

A_1: Målt medisinkonsentrasjon ved slutten av en kur overstiger 5 mg/l (dvs $Y > 5$).
A_2: Målt medisinkonsentrasjon ved slutten av en kur er mindre eller lik 12 mg/l (dvs $Y \leq 12$).

b) Er A_1 og A_2 disjunkte? (Begrunn svaret)

Er A_1 og A_2 uavhengige? (Begrunn svaret)

Følgende hendelse er definert:

A_3: Målt medisinkonsentrasjonen ved slutten av en kur er mellom 5 mg/l og 12 mg/l (dvs $5 < Y \leq 12$).

Uttrykk A_3 ved A_1 og A_2.

Anta nå at μ er ukjent, mens $\sigma^2 = 2^2$ fremdeles antas kjent. Fra åtte ulike kurer har man registrert dataene:

kur i	1	2	3	4	5	6	7	8
y_i	7.1	9.2	10.8	12.0	6.1	8.2	8.7	7.7

c) Skriv opp en rimelig estimator for μ, og regn ut estimatet.

Utled et 95% konfidensintervall for μ. Hva blir intervallet med de oppgitte dataene?

Legene har etterhvert funnet ut at i stedet for å gi alle pasienter samme dose medisin, vil det være gunstigere å justere dosene etter hvor syk pasienten er og hvor godt han/hun tåler bivirkningene. La x være dosen. Vi antar at x kan kontrolleres, dvs x er ikke stokastisk.

Man antar at en god lineær regresjonsmodell for sammenhengen mellom x og Y vil være

$$Y = \beta x + E,$$

der β er en ukjent konstant og E er en normalfordelt stokastisk variabel med forventningsverdi 0 og kjent varians $\sigma_E^2 = 2^2$.

d) Hvorfor er det i dette tilfellet rimelig å ikke ha med noe konstantledd i den lineære regresjonsmodellen?

Vis at sannsynlighetsmaksimeringsestimatoren (SME) for β basert på n uavhengige observasjoner blir

$$\hat{\beta} = \frac{\sum_{i=1}^{n} Y_i x_i}{\sum_{i=1}^{n} x_i^2}$$

der x_i og Y_i er henholdsvis dose og målt medisinkonsentrasjon for observasjon nummer i.

Regn ut forventningen og variansen til $\hat{\beta}$.

Det har i løpet av ti kurer på ulike pasienter blitt observert følgende sammenhørende verdier for x og Y:

kur i	1	2	3	4	5	6	7	8	9	10
x_i	4.5	4.0	5.5	7.0	8.0	8.5	9.0	6.5	6.0	5.0
y_i	6.2	5.2	7.3	8.7	9.0	10.5	10.3	8.2	7.4	7.0

Det oppgis at $\sum_{i=1}^{10} y_i x_i = 536.4$ og $\sum_{i=1}^{10} x_i^2 = 436$.

Før legene gir en pasient en viss dose x_0 ønsker de å vite noe om hvilken målt medisinkonsentrasjon Y_0 man kan regne med at dette vil gi. Du skal hjelpe legene ved å lage et 95% prediksjonsintervall.

e) Hva er tolkningen av et 95% prediksjonsintervall?

Utled et 95% prediksjonsintervall for Y_0 når $x_0 = 8$ ved å bruke de oppgitte dataene.

Oppgave 2 Frimerkesamleren

Det norske 50 øres frimerket med posthornmotiv fra 1910 ble trykket i ark a 100 frimerker. Pga en feil på trykkeplaten, ble ett av de 100 trykket med påskrift "50 ØRF" i steden for "50 ØRE". Hvert ark a 100 frimerker inneholdt altså ett frimerke med varianten ØRF.

Stian samler på frimerker, og finner en sigareske med 100 brukte 50 øres frimerker av denne utgaven på loftet. La X være antall "ØRF" blant de 100.

a) Under hvilke forutsetninger vil X være binomisk fordelt med parametre $n = 100$ og $p = 0.01$?

Anta nå at disse forutsetningene er oppfylt.

Hva blir $P(X = 0)$ og $P(X > 1 | X > 0)$?

En frimerkehandler selger dette frimerket i pakker a 100 brukte frimerker. Stian ønsker å undersøke om ØRF variantene er helt eller delvis plukket ut fra disse, slik at færre enn 1% av de frimerkene som frimerkehandleren selger, er ØRF. For å undersøke dette kjøper han $n = 400$ frimerker, og registrerer antall "ØRF", X, blant disse.

b) Forklar hvorfor vi *ikke* kan anta at X er tilnærmet normalfordelt.

Formuler Stians problemstilling som et hypotesetestingsproblem. Skriv opp nullhypotese og alternativ hypotese. Hva blir resultatet av testen hvis signifikansnivået er 5%, $n = 400$ og $X = 0$?

I steden for å kjøpe et forhåndsbestemt antall n frimerker, vil Jan Terje gjøre følgende: Han kjøper en og en pakke a 100 frimerker, og undersøker dem etterhvert. Hvis han får et ØRF, avslutter han innkjøpene. Dog har han ikke råd til å kjøpe mer enn til sammen k pakker, så dersom han ikke har fått noen ØRF etter k pakker (totalt $100k$ frimerker), avslutter han innkjøpene. La Z være antall pakker a 100 frimerker han må kjøpe etter denne prosedyren.

c) Finn punktsannsynligheten til Z som funksjon av p og k. Hva blir $\mathrm{E}(Z)$, hvis $p = 0.01$ og $k = 4$?

Oppgave 3 Forsikringsselskapet

Et forsikringsselskap regner med at utbetalingen X etter en industribrann er eksponensialfordelt, slik at sannsynlighetstettheten til X blir:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{hvis } x > 0, \\ 0 & \text{ellers,} \end{cases}$$

Selskapet er spesielt interessert i de høyeste utbetalingene, fordi de evt vil reassurere i andre selskap. La X_1 og X_2 være to uavhengige utbetalinger. Finn sannsynlighetstettheten til $V = \max(X_1, X_2)$. Finn $\mathrm{E}(V)$. Sammenlikn med $\mathrm{E}(X)$ og $2\mathrm{E}(X)$ og kommenter.

Oppgave 4 Lønnsstatistikk

Eva lurer på om Northwest Oil i Offshore driver med kvinnediskriminerende avlønning av sine sivilingeniører. For å undersøke dette, får hun NIF-avdelingen i selskapet til å innhente årslønn for n tilfeldig valgte sivilingeniører av hvert kjønn i selskapet. (For å få sammenlignbare tall, innhentes årslønn for kvinner og menn med samme eksamensår.) Eva bruker følgende modell, der X og Y betegner lønn for hhv kvinnelige og mannlige sivilingeniører.

$X_1, \ldots, X_n$ antas uavhengige $\mathrm{N}(\mu_1, \sigma^2)$
$Y_1, \ldots, Y_n$ antas uavhengige $\mathrm{N}(\mu_2, \sigma^2)$

Formuler Evas problemstilling som et hypotesetestingsproblem. Skriv opp uttrykket for testobservatoren. Regn ut verdien av testobservatoren når de oppgitte årslønningene (i 1000 kr) er som nedenfor. Har Eva grunnlag for å påstå at det drives kvinnediskriminering mht lønn, ved et signifikansnivå på 5%?

8 kvinner	480	305	290	295	385	360	240	290
8 menn	325	335	325	240	520	420	305	390

Det oppgis at $\bar{x} = 330.625$, $\bar{y} = 357.500$, $\sum_{i=1}^{8}(x_i - \bar{x})^2 = 39571.875$ og $\sum_{i=1}^{8}(y_i - \bar{y})^2 = 50550$