Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Bokmål

Faglig kontakt under eksamen:
Jan Terje Kvaløy 73 59 35 20
Stian Lydersen 73 59 35 20

EKSAMEN I FAG SIF5062/SIF5506 STATISTIKK

Tirsdag 30. mai 2000
Tid: 09:00–14:00

Hjelpemidler:
Godkjent lommekalkulator, Statistiske tabeller og formler (Tapir forlag), Rottman: Matematisk Formelsamling

Sensuren faller: 27. juni 2000.

Oppgave 1 pH-målinger

En bestemt målemetode for bestemmelse av pH-verdien i en løsning gir måleresultater som antas å være uavhengige og normalfordelte, med forventningsverdi μ lik virkelig pH og varians $\sigma^2 = 0.060^2$. La $X_1, \ldots, X_n$ være uavhengige målinger av pH i en bestemt løsning.

a) Anta (bare i dette punktet) at den virkelige pH-verdien i en løsning er 6.8.

Hva er sannsynligheten for at en bestemt måling gir et resultat som er under 6.74?

Hva er sannsynligheten for at en bestemt måling gir et resultat som er mellom 6.74 og 6.86?

Hva er sannsynligheten for at en bestemt måling, X, gir et resultat som avviker mer enn 0.06 fra μ, dvs bestem $P(|X - \mu| > 0.06)$?

Du skal estimere pH i en løsning, og bruker gjennomsnittet av 5 uavhengige målinger som estimat. La Y være gjennomsnittet av 5 uavhengige målinger.

b) Hva er sannsynligheten for at Y avviker mer enn 0.06 fra μ?

Utled et 95% konfidensintervall for μ. Hva blir konfidensintervallet når gjennomsnittet av fem uavhengige målinger ble 6.76?

Oppgave 2 Kjølevannspumper

På et bestemt sted i et prosessanlegg står to elektrisk drevne kjølevannspumper. For å ha tilstrekkelig kapasitet må begge pumpene fungere. La

$$E_i = \text{Pumpe nr } i \text{ er i feiltilstand}, \quad i = 1, 2.$$

Følgende sannsynligheter er gitt:

$$P(E_1) = P(E_2) = 0.01, \quad P(E_2|E_1) = 0.1.$$

Er hendelsene E_1 og E_2 uavhengige? Er de disjunkte? Begrunn svarene.

Finn sannsynligheten for at man ikke har nok kapasitet, dvs finn $P(E_1 \cup E_2)$.

Ledelsen er ikke fornøyd med dette, og vil installere en turbindrevet pumpe i tillegg. Det er da tilstrekkelig kapasitet hvis minst 2 av de 3 pumpene fungerer. La

$$T = \text{Den turbindrevne pumpa er i feiltilstand}$$

Det er oppgitt at den turbindrevne pumpa fungerer uavhengig av de elektrisk drevne pumpene, og at $P(T) = 0.04$.

Finn sannsynligheten for at det ikke er tilstrekkelig pumpekapasitet, dvs finn sannsynligheten for at minst 2 av de 3 pumpene ikke fungerer.

Oppgave 3 Bakterier i matvarer

I mange matvarer finnes det ofte en viss mengde uønskede bakterier, og for høye konsentrasjoner av slike bakterier kan være uheldig for konsumenten. Vi skal se på forekomsten av en bestemt type uønskede bakterier i en bestemt matvare. Studier har vist at antall bakterier, Y, i en prøve på t milligram (mg) av matvaren kan antas å være Poissonfordelt med forventning μt.

En kritisk grense for bakterieantallet i matvaren antas å være 12 bakterier per mg. Dersom antall bakterier per mg er større enn 12, kan dette være uheldig for konsumenten av matvaren.

Anta i første omgang at det er kjent at forventet antall bakterier per mg matvare er 5, dvs $\mu = 5$.

a) Finn sannsynligheten for å finne mer enn 12 bakterier i en matvareprøve på 1 mg.

Finn sannsynligheten for å finne mer enn 6 bakterier i en matvareprøve på 0.5 mg.

Dersom man tar 10 uavhengige prøver på 1 mg, hva er sannsynligheten for at minst en av de 10 prøvene innholder mer enn 12 bakterier?

I praksis vil μ, forventet antall bakterier per mg matvare, være ukjent. Basert på uavhengige målinger $Y_1, \ldots, Y_n$ av antall bakterier i n prøver på henholdsvis $t_1, \ldots, t_n$ mg ønsker vi å estimere μ.

b) Vis at sannsynlighetsmaksimeringsestimatoren (SME) for μ er

$$\hat{\mu} = \frac{Y_1 + \cdots + Y_n}{t_1 + \cdots + t_n}$$

Regn ut estimatorens forventning og varians.

For å sikre at grenseverdien 12 bakterier per mg ikke overskrides for ofte anbefaler Næringsmiddeltilsynet at forventet antall bakterier per mg matvare ikke bør overskride 6, dvs $\mu \leq 6$.

c) Vi ønsker for et bestemt matvareparti å undersøke om det er grunn til å påstå at Næringsmiddeltilsynets anbefaling overskrides.

Formuler dette som en hypotesetest.

Du kan i resten av dette punktet bruke uten bevis at en Poissonfordelt variabel med forventningsverdi større enn 15 er tilnærmet normalfordelt. Forklar hvorfor observatoren

$$\frac{\hat{\mu} - 6}{\sqrt{6/\sum_{i=1}^{n} t_i}}$$

er tilnærmet normalfordelt under nullhypotesen dersom $\sum_{i=1}^{n} t_i > 2.5$.

Utfør hypotesetesten på (tilnærmet) 5% nivå når vi fra $n = 10$ uavhengige målinger fra matvarepartiet har fått dataene gitt i tabellen under.

måling i	1	2	3	4	5	6	7	8	9	10
y_i	6	4	9	7	5	8	9	3	7	7
t_i	0.8	0.7	1.3	1.1	1.2	0.9	1.0	0.8	0.8	1.4

Det oppgis at $\sum_{i=1}^{10} y_i = 65$ og $\sum_{i=1}^{10} t_i = 10$.

For å effektivisere rutinene er man interessert i å ta i bruk en ny og raskere metode for å måle antall bakterier i en matvareprøve. Ulempen med denne nye målemetoden er at kun en viss andel av alle bakteriene i en prøve oppdages. Anta at hver enkelt bakterie oppdages med samme sannsynlighet p, og at alle bakterier oppdages eller oversees uavhengig av hverandre.

La som før Y være det totale antall bakterier i en prøve på t mg og la X være antallet av disse som oppdages med den nye målemetoden.

d) Forklar hvorfor X gitt $Y = y$ vil være binomisk fordelt med parametre y og p.

Finn simultanfordelingen til X og Y, og bruk denne til å vise at (marginal)fordelingen til X vil være en Poissonfordeling med forventingsverdi $\mu t p$.

Den nye målemetoden har blitt brukt til å utføre n uavhengige målinger $X_1, \ldots, X_n$ i prøver på henholdsvis $t_1, \ldots, t_n$ mg. Basert på disse målingene ønsker vi å estimere μ. Anta at p er kjent, og foreslå hvordan estimatoren for μ fra punkt b) kan modifiseres for å ta hensyn til at vi nå kjenner $X_1, \ldots, X_n$ i stedet for $Y_1, \ldots, Y_n$.

Oppgave 4 Mosjonisten

En 45-åring startet med løpetrening for 9 år siden, og har hvert år siden deltatt i samme mosjonsløp. Anvendt tid, i minutter, er gitt i tabellen nedenfor.

år i	1	2	3	4	5	6	7	8	9
alder x_i	37	38	39	40	41	42	43	44	45
tid y_i	45.54	41.38	42.50	38.80	41.26	37.20	38.19	38.05	37.45

Det oppgis at $\sum_{i=1}^{9} x_i = 369$, $\sum_{i=1}^{9} y_i = 360.37$, $\sum_{i=1}^{9}(x_i - \bar{x})^2 = 60$, $\sum_{i=1}^{9}(y_i - \bar{y})^2 = 63.28$ og $\sum_{i=1}^{9}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{9}(x_i - \bar{x})y_i = -52.57$.

Vi skal anta at observasjonene kan ses på som realisasjoner av uavhengige normalfordelte variable $Y_1, \ldots, Y_9$, hvor $\mathrm{E}(Y_i) = \alpha + \beta x_i$ og $\mathrm{Var}(Y_i) = \sigma^2$.

a) Skriv opp de vanlige forventningsrette estimatorene $\hat{\alpha}$, $\hat{\beta}$ og $\hat{\sigma}^2$ for α, β og σ^2. Regn ut estimatene for α og β for de gitte dataene. Plott datasettet og den estimerte regresjonslinjen.

Det oppgis at estimatet for σ^2 er 1.568^2.

b) Regn ut et uttrykk for variansen til estimatoren $\hat{\beta}$.

Gjennomfør en test av $H_0: \ \beta = 0$ mot $H_1: \ \beta \neq 0$, på signifikansnivå 1%.

Hva blir den praktiske fortolkningen av testen over?

Løperen ønsker å predikere anvendt tid på mosjonsløpet neste gang (alder $x_0 = 46$ år).

c) Regn ut predikert tid.

Det oppgis at $\mathrm{Var}(\hat{\alpha} + \hat{\beta}x_0) = \sigma^2(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2})$. Utled et 95% prediksjonsintervall for Y ved $x_0 = 46$ år. Hva blir intervallet med de oppgitte data?

Hvis løperen ber deg predikere anvendt tid om 15 år (alder 60 år), hva vil du svare da?