Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Bokmål

Faglig kontakt under eksamen:
Håkon Tjelmeland 73 59 35 38

EKSAMEN I EMNE ST1201 STATISTISKE METODER

Fredag 3. desember 2004
Tid: 09:00–13:00

Hjelpemidler: Alle trykte og håndskrevne hjelpemidler tillatt.
Alle kalkulatorer tillatt.

Sensur er ferdig: 24. desember 2004.

Oppgave 1 Hamburgervekt

Lise som studerer på NTNU er en flittig gjest på et gatekjøkken hvor de serverer hamburgere som påstås å veie 100 gram. Lise har en mistanke om at gatekjøkkenet lurer kundene sine ved at hamburgerne de selger i virkeligheten veier mindre enn hva som loves. Før hun eventuelt beskylder gatekjøkkenet for dette vil hun være sikker i sin sak. Hun kjøper derfor ti hamburgere (på forskjellige tidspunkt) og kontrollveier disse. Hennes målinger ble

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
97.8	89.2	106.1	98.8	85.4	90.2	91.6	89.6	104.8	100.7

Det opplyses at dette gir $\sum_{i=1}^{10} x_i = 954.2$ og $\sum_{i=1}^{10} x_i^2 = 91511.58$.

Lise antar at vekten til en tilfeldig hamburger er normalfordelt med (ukjent) forventning μ og (ukjent) varians σ^2 og formulerer situasjonen som et hypotesetestingsproblem,

$$H_0 : \mu = 100 \quad \text{mot} \quad H_1 : \mu < 100.$$

a) Skriv ned testobservatoren og beslutningsregelen Lise bør benytte.

Hva blir konklusjonen på testen når signifikansnivået velges til $\alpha = 0.05$?

Ved nærmere ettertanke blir Lise i tvil om antagelsen om normalfordeling kan forsvares. Hun velger derfor i stedet kun å anta at vekten av en tilfeldig hamburger har en sannsynlighetsfordeling som er symmetrisk fordelt rundt en (ukjent) forventningsverdi μ. Hun vil så benytte fortegnstesten for å avgjøre om hun kan forkaste H_0.

b) For signifikansnivå $\alpha = 0.05$, vis hva beslutningsregelen da blir.

Hva blir nå konklusjonen på testen med Lises observasjoner?

Dersom vekten av en tilfeldig hamburger er $\mathrm{N}(95, 7^2)$-fordelt kan det vises at testen i punkt **a)** har teststyrke lik 0.67.

c) Bestem teststyrken også for testen i punkt **b)** når vekten av en tilfeldig hamburger er $\mathrm{N}(95, 7^2)$-fordelt.

Sammenlign teststyrken for de to testene og kommenter.

Hvilken av de to testene vil du anbefale Lise å benytte? Begrunn svaret!

Oppgave 2 Kabelproduksjon

En fabrikk produserer kabel og iblant oppstår det feil på den produserte kabelen. La Z betegne lengden (i kilometer) på kabelen mellom to etterfølgende feil. Vi skal anta at feilene oppstår uavhengig av hverandre, dvs. at påfølgende observasjoner av Z langs kabelen, $Z_1, Z_2, Z_3, \ldots$, er uavhengige stokastiske variabler.

Vi skal anta at lengden mellom to etterfølgende feil er eksponensialfordelt med forventning β, dvs.

$$f_Z(z;\beta) = \begin{cases} \frac{1}{\beta} e^{-\frac{z}{\beta}} & \text{for } z > 0, \\ 0 & \text{ellers.} \end{cases}$$

Vi skal i denne oppgaven anta at man har gjort observasjoner $Z_1, Z_2, \ldots, Z_n$ og estimerer den ukjente parameteren β ved

$$\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n} Z_i.$$

a) Vis at $\widehat{\beta}$ er forventningsrett og bestem dens varians.

Benytt Cramer-Raos ulikhet til å vise at $\widehat{\beta}$ er en *beste* estimator, dvs. vis at det ikke finnes andre forventningsrette estimatorer for β som har mindre varians enn $\widehat{\beta}$.

La så $V_i = 2Z_i/\beta$ for $i = 1, 2, \ldots, n$.

b) Vis at V_i er χ^2-fordelt med to frihetsgrader. Benytt så dette til å forklare at $2n\widehat{\beta}/\beta$ er χ^2-fordelt med $2n$ frihetsgrader.

c) Utled et $(1-\alpha)\cdot 100\%$-konfidensintervall for β.

Beregn konfidensintervallet numerisk når $\alpha = 0.10$, $n = 10$ og de observerte verdier for Z_i-ene er $118.64, 163.58, 1.89, 25.58, 14.37, 159.32$, $142.01, 200.53, 20.24$ og 52.62.

Oppgave 3 Arvelover for blodtyper hos mennesker

Hvert menneske har en av fire ulike blodtyper, som kalles A, B, AB og 0. Blodtypen til et menneske arves fra ens foreldre ifølge visse sannsynlighetslover. Inntil 1924 fantes det to teorier for detaljene i disse sannsynlighetslovene. Disse ble kalt henholdsvis en-loci- og to-loci-teorien. La θ_A, θ_B, θ_{AB} og θ_0 være sannsynligheten for at et tilfeldig valgt individ i en stor og genetisk stabil populasjon skal ha blodgruppe henholdsvis A, B, AB og 0. Ifølge en-loci-teorien er

$$\theta_A = p(2-p-2q), \quad \theta_B = q(2-2p-q), \quad \theta_{AB} = 2pq \text{ og } \theta_0 = (1-p-q)^2,$$

for visse (ukjente) parametre $p \in (0,1)$ og $q \in (0,1)$. Ifølge to-loci-teorien er derimot

$$\theta_A = a(1-b), \quad \theta_B = (1-a)b, \quad \theta_{AB} = ab \text{ og } \theta_0 = (1-a)(1-b),$$

for visse andre (ukjente) parametre $a \in (0,1)$ og $b \in (0,1)$.

I 1924 publiserte F. Bernstein en artikkel der han sammenlignet de to teoriene med blodtype-observasjoner hos 502 japanere som bodde i Korea. For de $n = 502$ japanerne han undersøkte, la X_A, X_B, X_{AB} og X_0 være antallet som hadde hver av de fire blodtypene. F. Bernstein observerte $x_A = 212$, $x_B = 103$, $x_{AB} = 39$ og $x_0 = 148$.

Vi skal vurdere om de observerte data gir grunnlag for å konkludere med at to-loci-teorien er feil. Vi antar at (X_A, X_B, X_{AB}, X_0) er multinomisk fordelt med n forsøk.

a) Finn sannsynlighetsmaksimeringsestimatorene for a og b.

Vis at sannsynlighetsmaksimeringsestimatene for a og b med F. Bernsteins data blir henholdsvis 0.500 og 0.283.

b) Formuler en goodness-of-fit-test for situasjonen, dvs. angi H_0 og H_1, angi en testobservator og hvilken fordeling denne har når H_0 er riktig, gi hva beslutningsregelen blir og bestem hva konklusjonen blir når observasjonene er som gitt over og man benytter signifikansnivå $\alpha = 0.10$.

Tilsvarende som du nå har gjort for to-loci-teorien kan man selvfølgelig også lage en hypotesetest for å teste en-loci-teorien. Konklusjonen blir da den motsatte av hva du fant (eller i hvert fall, burde funnet) over, men dette skal du slippe å gjennomføre nå.