Noregs teknisk–
naturvitskaplege universitet
Institutt for matematiske fag

Nynorsk

Fagleg kontakt under eksamen:
Håkon Tjelmeland 73 59 35 38

EKSAMEN I EMNE ST1201 STATISTISKE METODAR

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

Hjelpemiddel: Alle trykte og handskrivne hjelpemiddel tillatne.
Alle kalkulatorar tillatne.

Sensur er ferdig: 24. desember 2004.

Oppgåve 1 Hamburgarvekt

Lise som studerer på NTNU er ein flittig gjest på eit gatekjøkken der dei serverer hamburgarar som dei hevdar veg 100 gram. Lise har ein mistanke om at gatekjøkkenet lurer kundane sine ved at hamburgarane dei sel i røynda veg mindre enn kva dei lovar. Før ho eventuelt skuldar gatekjøkkenet for dette vil ho vere sikker i si sak. Ho kjøper difor ti hamburgarar (på forskjellige tidspunkt) og kontrollveg desse. Målingane hennar blei

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

Lise går ut ifrå at vekta til ein tilfeldig hamburgar er normalfordelt med (ukjent) forventning μ og (ukjent) varians σ^2 og formulerer situasjonen som eit hypotesetestingsproblem,

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

a) Skriv ned testobservatoren og beslutningsregelen Lise bør bruke.

Kva blir konklusjonen på testen når signifikansnivået blir valt til $\alpha = 0.05$?

Ved nærare ettertanke blir Lise i tvil om føresetnaden om normalfordeling kan forsvarast. Ho vel difor i staden kun å gå utifrå at vekta av ein tilfeldig hamburgar har ei sannsynsfordeling som er symmetrisk rundt ein (ukjent) forventningsverdi μ. Ho vil så bruke forteiknstesten for å avgjere om ho kan forkaste H_0.

b) For signifikansnivå $\alpha = 0.05$, vis kva beslutningsregelen då blir.

Kva blir no konklusjonen på testen med observasjonane til Lise?

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

c) Finn teststyrken også for testen i punkt **b**) når vekta av ein tilfeldig hamburgar er $\mathrm{N}(95, 7^2)$-fordelt.

Samanlikn teststyrken for dei to testane og kommenter.

Kva for ein av dei to testane vil du rå Lise til å bruke? Grunngi svaret!

Oppgåve 2 Kabelproduksjon

Ein fabrikk produserer kabel og iblant oppstår det feil på den produserte kabelen. Lat Z vere lengda (i kilometer) på kabelen mellom to etterfølgjande feil. Vi skal gå ut ifrå at feila oppstår uavhengig av kvarandre, dvs. at påfølgjande observasjonar av Z langs kabelen, $Z_1, Z_2, Z_3, \ldots$, er uavhengige stokastiske variablar.

Vi skal gå ut ifrå at lengda mellom to etterfølgjande feil er eksponensialfordelt med forventning β, dvs.

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

Vi skal i denne oppgåva gå ut ifrå at ein har gjort observasjonar $Z_1, Z_2, \ldots, Z_n$ og estimerer den ukjente parametaren β ved

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

a) Vis at $\widehat{\beta}$ er forventningsrett og finn variansen til $\widehat{\beta}$.

Bruk Cramer-Raos ulikskap til å vise at $\widehat{\beta}$ er ein *beste* estimator, dvs. vis at det ikkje fins andre forventningsrette estimatorar for β som har mindre varians enn $\widehat{\beta}$.

Lat no $V_i = 2Z_i/\beta$ for $i = 1, 2, \ldots, n$.

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

c) Lei ut eit $(1-\alpha)\cdot 100\%$-konfidensintervall for β.

Rekn ut konfidensintervallet numerisk når $\alpha = 0.10$, $n = 10$ og dei observerte verdiane for Z_i-ane er $118.64, 163.58, 1.89, 25.58, 14.37, 159.32$, $142.01, 200.53, 20.24$ og 52.62.

Oppgåve 3 Arvelover for blodtyper hos menneske

Kvart menneske har ein av fire ulike blodtyper, som blir kalla A, B, AB og 0. Blodtypen til eit menneske blir arva frå foreldra ifølgje visse sannsynslover. Inntil 1924 var det to teoriar for detaljane i desse sannsynslovene. Desse blei kalla høvesvis ein-loci- og to-loci-teorien. Lat θ_A, θ_B, θ_{AB} og θ_0 vere sannsynet for at eit tilfeldig valgt individ i ein stor og genetisk stabil populasjon skal ha blodgruppe høvesvis A, B, AB og 0. Ifølgje ein-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) parametrar $p \in (0,1)$ og $q \in (0,1)$. Ifølgje 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) parametrar $a \in (0,1)$ og $b \in (0,1)$.

I 1924 publiserte F. Bernstein ein artikkel der han samanlikna dei to teorianе med blodtype-observasjonar hos 502 japanarar som budde i Korea. For dei $n = 502$ japanarane han undersøkte, lat X_A, X_B, X_{AB} og X_0 vere talet på personar for kvar av de fire blodtypene. F. Bernstein observerte $x_A = 212$, $x_B = 103$, $x_{AB} = 39$ og $x_0 = 148$.

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

a) Finn sannsynsmaksimeringsestimatorane for a og b.

Vis at sannsynsmaksimeringsestimata for a og b med F. Bernsteins data blir høvesvis 0.500 og 0.283.

b) Formuler ein goodness-of-fit-test for situasjonen, dvs. angi H_0 og H_1, angi ein test-observator og kva for ei fordeling han har når H_0 er riktig, angi kva beslutningsregelen blir og bestem kva konklusjonen blir når observasjonane er som gjevne over og ein bruker signifikansnivå $\alpha = 0.10$.

Tilsvarande som du no har gjort for to-loci-teorien kan ein sjølvsagt også lage ein hypotesetest for å teste ein-loci-teorien. Konklusjonen blir då den motsette av kva du fann (eller i alle fall, burde ha funne) over, men dette skal du sleppe å gjennomføre no.