Noregs teknisk–naturvitskaplege universitet
Institutt for matematiske fag

Nynorsk

Fagleg kontakt under eksamen:
Ingelin Steinsland 73 55 02 39 / 92 66 30 96
Jo Eidsvik 73 59 01 53 / 90 12 74 72

EKSAMEN I EMNE TMA4240 STATISTIKK

1. desember 2008
Tid: 09:00–13:00

Hjelpemiddel: *Tabeller og formler i statistikk*, Tapir Forlag
K. Rottmann: *Matematisk formelsamling*
Kalkulator HP30S
Gult, stempla A5-ark med eigne handskrevne notat.

Sensuren fell: 22. desember 2008

Oppgåve 1 Sykkelruter

Solan og Fabian bur i same kollektiv. Dei syklar begge til universitetet, men dei har ulike ruter. Vi antar at tida (i minutt) kvar av dei bruker er normalfordelt. For Solan $X \sim N(\mu_1, \sigma^2)$ og for Fabian $Y \sim N(\mu_2, \sigma^2)$. Merk at vi antar felles varians.

a) Anta i dette puntet at $X \sim N(6, 1^2)$ og $Y \sim N(7, 1^2)$.

- Kva er sannsynet for at Fabian bruker meir enn seks minutt på turen?
- Kva er sannsynet for at Solan bruker mindre enn sju minutt gitt at han bruker mindre enn 8 minutt?
- Ein dag startar dei samtidig, kva er sannsynet for at minst ein av dei er på universitetet før det er gått seks minutt?

b) Solan har på eit vor-spiel vedda på at hans rute er raskare enn Fabian si. Ettermiddagen etterpå bestemmer dei at begge skal samla inn data for alle sykkelturane i ei veke. Deretter skal dei gjere ein hypotesetest med signifikansnivå $\alpha = 0.05$. Dataene er gjeve i tabell 1. Anta at observasjonane er uavhengige og at $\sigma^2 = 1^2$.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
7.0	6.8	5.4	7.3	5.9	8.4	6.2	7.1	4.1	6.4	6.9	6.6	6.7	6.0	6.7

Tabell 1: *Observerte tider for Solan* $(x_1, x_2, \ldots, x_7)$ *og Fabian* $(y_1, y_2, \ldots, y_8)$*. Vi får* $\bar{x} = \frac{1}{7}\sum_{i=1}^{7} x_i = 6.31$, $\bar{y} = \frac{1}{8}\sum_{i=1}^{8} y_i = 6.81$, $\sum_{i=1}^{7}(x_i - \bar{x})^2 = 5.44$ *og* $\sum_{i=1}^{8}(y_i - \bar{y})^2 = 6.91$

Formuler hypotesetesten, og test hypotesa med gjeve data.

Kva er styrken på denne testen dersom sannsynsfordelingane til X og Y er som i punkt a).

c) Vi vil i dette punktet sjå på antakinga $\sigma^2 = 1$. På grunnlag av data i punkt b), estimer σ^2, og finn eit 95% konfidensintervall. Kan du på grunnlag av konfidensintervallet seie at variansen er signifikant forskjellig frå 1?

d) På grunn av trafikken trur Fabian at sykkelturen tar lengre tid dess seinare han kjem seg ut om morgonen. Han bruker ei veke på å samle inn data om starttidspunkt (t_i minutt etter kl $7:00$) og sykkelturtid (y_i), sjå tabell 2.

t_i	0	15	30	45	60
y_i	5.6	5.5	6.1	7.5	7.4

Tabell 2: *Starttidspunkt* t_i *i minutt etter kl* $7:00$ *og lengde på sykkeltur* y_i *i minutt. Vi får;* $\bar{t} = 30$, $\bar{y} = 6.42$, $\sum_{i=1}^{5}(t_i - \bar{t})^2 = 2250$ *og* $\sum_{i=1}^{5}(t_i - \bar{t})y_i = 84$

Sett opp ein enkel lineær regresjonsmodell for sykkelturtida, og spesifiser antakingane dine. Sett og opp minste-kvadratersestimatorar for parametra i modellen (du treng ikkje vise korleis desse kjem fram). Du kan seinare i oppgåva utan bevis bruke at desse estimatorane er forventningsrette.

Anta i resten av dette punktet, og i punkt e), at støyledda ϵ i regresjonsmodellen er normalfordelte med forventning 0 og kjent varians 0.5^2.

Formuler Fabian sin teori som ein hypotesetest, og utfør testen med signifikansnivå på 1%.

e) Ein dag startar Fabian kl $8:30$. Prediker basert på modellen i punkt d) kor lang tid han vil bruke, og finn eit 95% prediksjonsintervall for sykkelturtida denne dagen. Kommenter.

Oppgåve 2 Ras ved sprengningsarbeid

Det overraskende raset i Løsberga ved Steinkjer førte til stenging av vei og jernbane, med kostnad ca 1 million per dag. Etter dette (og andre ras) har nasjonal rassikringsgruppe levert krav til samferdselsminister Navarsete om en milliard kroner. I denne oppgaven skal vi studere en tenkt situasjon med kjent sannsynlighet for ras i forbindelse med sprengning.

Et firma har i oppdrag å utbedre en bilvei. Arbeidet medfører sprengningsarbeid, med risiko for ras. Fra basis kunnskap om geologien i området antar de sannsynlighet for ras $p = P(X = 1) = 0.15$. Her er stokastisk variabel $X = 1$ dersom det raser, mens $X = 0$ ellers.

a) Anta, kun i dette punktet, at det i sprengningsarbeidet langs veistrekningen finnes 4 slike mulige rassteder, og at eventuelle ras her vil skje uavhengige av hverandre.

S er antall ras av de fire mulige rasene. Argumenter for at antall ras er binomisk fordelt med parametre $n = 4$ og $p = 0.15$.

Hva er sannynligheten for at det går ingen ras?

Gitt at det blir minst ett ras, hva er sannsynligheten for at det blir flere enn ett ras?

Videre i oppgaven ser vi kun på et av de mulige rasstedene. Dersom det raser her, blir det uforutsett stenging av vei, omdirigering av trafikk, dekking av skader, etc. Dette har total kostnad 40 millioner kroner. Dersom det ikke raser, er ingen skade skjedd, og kostnad 0 kroner. Det er også mulig å legge om veien i anleggsperioden. Dette har en fast kostnad på 7 millioner kroner. Et ras vil da ikke gi ytterligere kostnad.
Strategi A er å sprenge uten omlegging av veien.
Strategi B er midlertidig omlegging av veien.

b) Beregn forventet kostnad Z under *Strategi A*. Finn også standardavviket til kostnad. Bør veien legges midlertidig om? Begrunn svaret.

For 5 millioner kroner kan det gjøres grundige geologiske undersøkelser som gir et sikkert svar om det vil rase eller ikke. Basert på resultatet av en slik undersøkelse vil man vite om *strategi A* eller *strategi B* bør velges.

Hva er forventet kostnad dersom man gjennomfører denne grundige undersøkelsen? Bør den grundige undersøkelsen gjennomføres? Begrunn svaret.

For 1 million kroner kan geologer undersøke området med enkle metoder og gi en kvalifisert uttalelse om ras ($Y = 1$) eller ikke ras ($Y = 0$). Denne enkle metoden er ikke sikker, og vi antar at de treffer med sannsynlighet $P(Y = 1|X = 1) = P(Y = 0|X = 0) = \gamma > 0.5$.

Vi skal først estimere sannsynligheten γ fra data der vi kjenner utfallet av X, om det gikk ras eller ikke. Firmaet har brukt geologene i lignende, uavhengige, situasjoner 15 ganger tidligere.

I sju av tilfellene gikk det ras $X_i = 1$, $i = 1, \ldots, 7$. Uttalelsene til geologene var da som følger: $Y_1 = 0$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 1$, $Y_7 = 1$. I åtte av tilfellene gikk det ikke ras $X_i = 0$, $i = 8, \ldots, 15$. Uttalelsene til geologene var da som følger: $Y_8 = 1$, $Y_9 = 0$, $Y_{10} = 0$, $Y_{11} = 0$, $Y_{12} = 1$, $Y_{13} = 0$, $Y_{14} = 1$, $Y_{15} = 0$.

En estimator for γ er:

$$\hat{\gamma} = \frac{\sum_{i=1}^{15} I_i}{15}$$

der $I_i = 1$ dersom $Y_i = X_i$, og $I_i = 0$ dersom $Y_i \neq X_i$.
Vi antar at I_i, $i = 1, \ldots, 15$ er uavhengige.

c) Er $\hat{\gamma}$ sannsynlighetsmaksimeringsestimatoren (SME) til γ? Også kalt maximum likelihood estimator. Svar ved å finne SME til γ.

Regn ut estimatet for γ basert på estimatoren $\hat{\gamma}$.

d) Beregn forventning og varians til $\hat{\gamma}$.

Firmaet vurderer nå å samle inn slike relativt billige data fra geologene.

e) Bruk i dette punktet den estimerte verdien av γ fra punkt c).

Bruk Bayes formel til å regne ut sannsynligheten for ras når geologene uttaler ikke ras. Regn videre sannsynligheten for ras når geologene uttaler ras.

Firmaet ønsker å regne forventet kostnad før geologene eventuelt kalles inn. Argumenter for at forventet kostnad er:

$$C = 1 + \sum_{y=0}^{1} \min[7, E(Z|Y = y)]P(Y = y),$$

der Z er kostnad ved *strategi* A. Videre er $\min[a, b] = a$ dersom $a < b$, og $\min[a, b] = b$, dersom $a > b$.

Hva blir forventet kostnad C? Bør undersøkelsen gjennomføres? Begrunn svaret.