Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Faglig kontakt under eksamen:
Magnar Lillegård, tlf. 73 59 01 54 / 991 61 665

Eksamen i ST0201 Brukerkurs i statistikk

Tirsdag 23. mai 2006
Tid: 09.00–13.00
Tillatte hjelpemidler: Alle trykte og skrevne. Lommeregner
Sensurdato: 13. juni 2006

Alle delspørsmål i oppgavesettet teller like mye.

Oppgave 1

Et datamateriale om for tidlig fødte barn viser svangerskapslengde (X) målt i antall uker og fødselsvekt (Y) målt i kg for 40 barn ved tre ulike sykehus A, B og C. Vi antar at alle observasjonene er uavhengige og tilnærmet normalfordelte og at fødselsvektene har samme varians ved alle tre sykehusene. En oppsummering av dataene finner du i tabellen under.

Sykehus	n	$\bar{x}$	$\bar{y}$	$\sum(x_i - \bar{x})^2$	$\sum(y_i - \bar{y})^2$	$\sum(x_i - \bar{x})(y_i - \bar{y})$
A	15	30,07	1,07	96,93	0,57	4,73
B	14	31,36	1,44	113,21	1,49	12,29
C	11	32,09	1,32	76,91	0,56	5,88

a) Estimer korrelasjonen mellom svangerskapslengde og fødselsvekt for barn født ved sykehus A. Utfør deretter en hypotesetest for å finne ut om den teoretiske korrelasjonen ρ er større enn null. Bruk signifikansnivå 0,05.

b) Skriv opp den lineære regresjonsmodellen der Y er responsvariabelen og X er forklaringsvariabelen. Estimer regresjonslinjen og variansen i modellen når du bruker datamaterialet fra sykehus A.

c) Parametrene μ_A, μ_B og μ_C er forventede fødselvekter ved henholdsvis sykehusene A, B og C. Beregn et 95% konfidensintervall for $\mu_B - \mu_C$.

d) Test hypotesen $H_0\colon \mu_A = \mu_B = \mu_C$ mot $H_1\colon$ ikke alle forventningene er like. Hva blir konklusjonen når signifikansnivået er 0,05? Hint: Du trenger å regne ut det totale gjennomsnittet i datasettet.

Oppgave 2

Vi trekker fem personer tilfeldig fra befolkningen og lar X være antall av disse som kan defineres som overvektige.

a) Hvis vi antar at X er binomisk fordelt med parametre 5 og p, bruk enten momentmetoden eller sannsynlighetsmaksimering til å begrunne at $X/5$ er en fornuftig estimator for p. Finn uttrykk for forventningen og variansen til denne estimatoren.

b) Vi ønsker å teste hypotesen $H_0\colon p \leq 0{,}2$ mot $H_1\colon p > 0{,}2$. Hvis sannsynligheten for type I-feil maksimalt skal være 0,10, hvilket område må X da ligge i for at vi skal forkaste nullhypotesen? Begrunn svaret.

c) Sett opp et uttrykk for styrkefunksjonen til testen. Beregn deretter teststyrken for $p = 0{,}5$ og $p = 0{,}8$.

Oppgave 3

Vekten X på voksne individer av en dyreart antas å ha fordelingen

$$f(x) = \frac{\lambda^3}{2} x^2 e^{-\lambda x}, \; x > 0.$$

a) Bruk kjente egenskaper ved gammafordelingen til å vise at $E(X) = 3/\lambda$ og $\mathrm{Var}(X) = 3/\lambda^2$.

b) Den stokastiske variable $\bar{X}$ er middelverdien av vektene til $n = 30$ individer av arten. Anta at $\bar{X}$ er tilnærmet normalfordelt og bruk dette til å vise hvordan du tester

$$H_0\colon \lambda = 1 \text{ mot } H_1\colon \lambda < 1.$$

Gjennomfør testen hvis $\bar{x} = 4{,}7$ og signifikansnivået er 0,01.

c) Begrunn at $(\lambda\bar{X} - 3)\sqrt{n/3}$ er tilnærmet standard normalfordelt. Bruk dette til å utlede et 95% konfidensintervall for λ. Beregn intervallet når observasjonene er som i punkt b.