

Institutt for matematiske fag

Eksamensoppgåve i **TMA4255 Anvendt statistikk**

Fagleg kontakt under eksamen: Anna Marie Holand

Tlf: 951 38 038

Eksamensdato: 30. mai 2014

Eksamenstid (frå–til): 09:00-13:00

Hjelpemiddelkode/Tillatne hjelpemiddel: Alle trykte og handskrivne hjelpemiddel. Spesiell kalkulator.

Annan informasjon:

- I utskrifta frå MINITAB er komma brukt som desimalskilleteikn.
- Signifikansnivå 5% skal brukast om ikke anna er spesifisert.
- Alle svar må grunngjevast.

Målform/språk: nynorsk

Sidetal: 6

Sidetal vedlegg: 0

Kontrollert av:

Dato

Sign

Oppgave 1 Darwins maisplantar

Darwin (1876) studerte veksten av par av maisplantar, der den eine planten er framstilt ved kryssbefruktning og den andre ved sjølvbefruktning. Målet hans var å demonstrere at kryssbefrukta plantar har større fitness (t.d. overlevelse og vekst) enn sjølvbefrukta plantar.

Femten par av kryssbefrukta og sjølvbefrukta plantar vart dyrka under identiske tilhøve. For kvart par vart høgda (i tommar) til kvar plante registrert.

For par i la X_{1i} vere høgda av planta framstilt ved kryssbefruktning og X_{2i} vere høgda av planta framstilt ved sjølvbefruktning, $i = 1, \dots, 15$. Vidare la $D_i = X_{1i} - X_{2i}$. Dataene fra eksperimentet er presentert under.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_{1i}	188	96	168	176	153	172	177	163	146	173	186	168	177	184	96
x_{2i}	130	163	160	160	147	149	149	122	132	144	130	144	102	124	144

Deskriptive mål er $\bar{x}_1 = \frac{1}{15} \sum_{i=1}^{15} x_{1i} = 161.53$,

$s_{x1} = \sqrt{\frac{1}{14} \sum_{i=1}^{15} (x_{1i} - \bar{x}_1)^2} = 28.94$, $\bar{x}_2 = \frac{1}{15} \sum_{i=1}^{15} x_{2i} = 140$,

$s_{x2} = \sqrt{\frac{1}{14} \sum_{i=1}^{15} (x_{2i} - \bar{x}_2)^2} = 16.64$, $\bar{d} = \bar{x}_1 - \bar{x}_2 = 21.53$,

$s_d = \sqrt{\frac{1}{14} \sum_{i=1}^{15} (d_i - \bar{d})^2} = 38.29$.

(Kilde: Darwin, C. (1876). The Effect of Cross- and Self-fertilization in the Vegetable Kingdom, 2nd Ed. London: John Murray.)

- a) Anta at X_{1i} og X_{2i} er normalfordelte, $X_{1i} \sim N(\mu_1, \sigma^2)$ og $X_{2i} \sim N(\mu_2, \sigma^2)$, $i = 1, \dots, 15$.

Basert på dette forsøket, kan Darwin konkludere med at kryssbefrukta plantar er høgare enn sjølvbefrukta plantar? Skriv ned null hypotesen og den alternative hypotesen, vel ein testobservator og gjennomfør ein hypotesetest. Bruk signifikansnivå $\alpha = 0.05$.

Spesifiser kva antakingar du gjer.

- b) Anta at X_1 og X_2 ikkje er normalfordelte. Utfør ein forteiknstest (sign test) for å teste om kryssbefrukta plantar vert høgare enn sjølvbefrukta plantar. Kommenter resultata/funna dine.

Oppgave 2 Gråspurv

Gråspurv (*Passer domesticus*) populasjonar langs kysten av Midt- og Nord-Noreg har vorte studert av forskarar ved NTNU gjennom fleire tiår.

Gråspurvhannar har ein svart brystfleck (svarte fjører i brystet). Brystflekken er variabel i storleik og tidligare studiar indikerer at storleiken på brystflekken er eit signal på sosial status og fysisk form. I ein studie vart desse variablane målt.

- y , synleg brystfleckstorleik målt i mm.
- x_1 , tarslengd (lengda av foten) målt i mm.
- x_2 , nebbhøgda målt i mm.
- x_3 , total brystfleckstorleik målt i mm.

Forskarar trur at storleiken på den synlege brystflekken, y , er avhengig av tarslengda, x_1 , (som indikerer generell kroppstorleik), nebbhøgda, x_2 , (som seier noko om evna til å ete mat av ein bestemd type) og den totale brystflekken, x_3 , (som seier noko om den potensielle synlege brystflekken).

Desse trekka vart målt for $n = 901$ gråspurvhannar. Vi analyserer dataene med en multiple lineær model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i, \quad (1)$$

kor ϵ_i er u.i.f. $N(0, \sigma^2)$ for $i = 1, \dots, n$.

Ei MINITAB utskrift frå statistisk analyse finst i Figur 1 og Figur 2 og et plot av standardiserte residualer finst i Figur 3.

- a) Utfør ein t-test for å teste null hypotesen $H_0: \beta_1 = 0$ mot den alternative hypotesen $H_1: \beta_1 \neq 0$. Bruk signifikansnivå 5%.

Kor mange prosent av variasjonen forklarar regresjonen?

Predictor	Coef	SE Coef	T	P
Constant	4,966	2,020	2,46	0,014
x1	0,11584	0,06609	?	?
x2	-0,6629	0,1905	-3,48	0,001
x3	0,71509	0,04441	16,10	0,000

S = 1,54791 R-Sq = 24,1% R-Sq(adj) = 23,8%

Figur 1: Utskrift frå statistiske analysar av gråspurvdataene.

Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	3	682,54	227,51	?	0,000
Residual Error	897	?	?		
Total	900	?			

Source	DF	Seq SS
x1	1	14,11
x2	1	47,08
x3	1	621,36

Figur 2: Utskrift frå statistiske analysar av gråspurvdataene.

b) Kva er eit passende estimat for σ^2 ?

I utskriftera frå ei utført variansanalyse av den tilpassa multiple lineære modellen i Figur 2 er fire numeriske verdiar bytta ut med spørjeteikn. Rekn ut dei numeriske verdiane for kvar av desse, og forklar kva kvar av desse tyder.

c) Bruk den estimerte regresjonsmodellen i Figur 1 til å rekne ut eit punkttestimat for synleg brystflekkestorleik, $\hat{y}_0$, for observert tarslengd $x_1^0 = 20$, nebbhøgde $x_2^0 = 8$ og total brystflekkestorleik $x_3^0 = 19$. Det er gjeve at det estimerte standardavviket til $\hat{y}_0$ er 0.097.

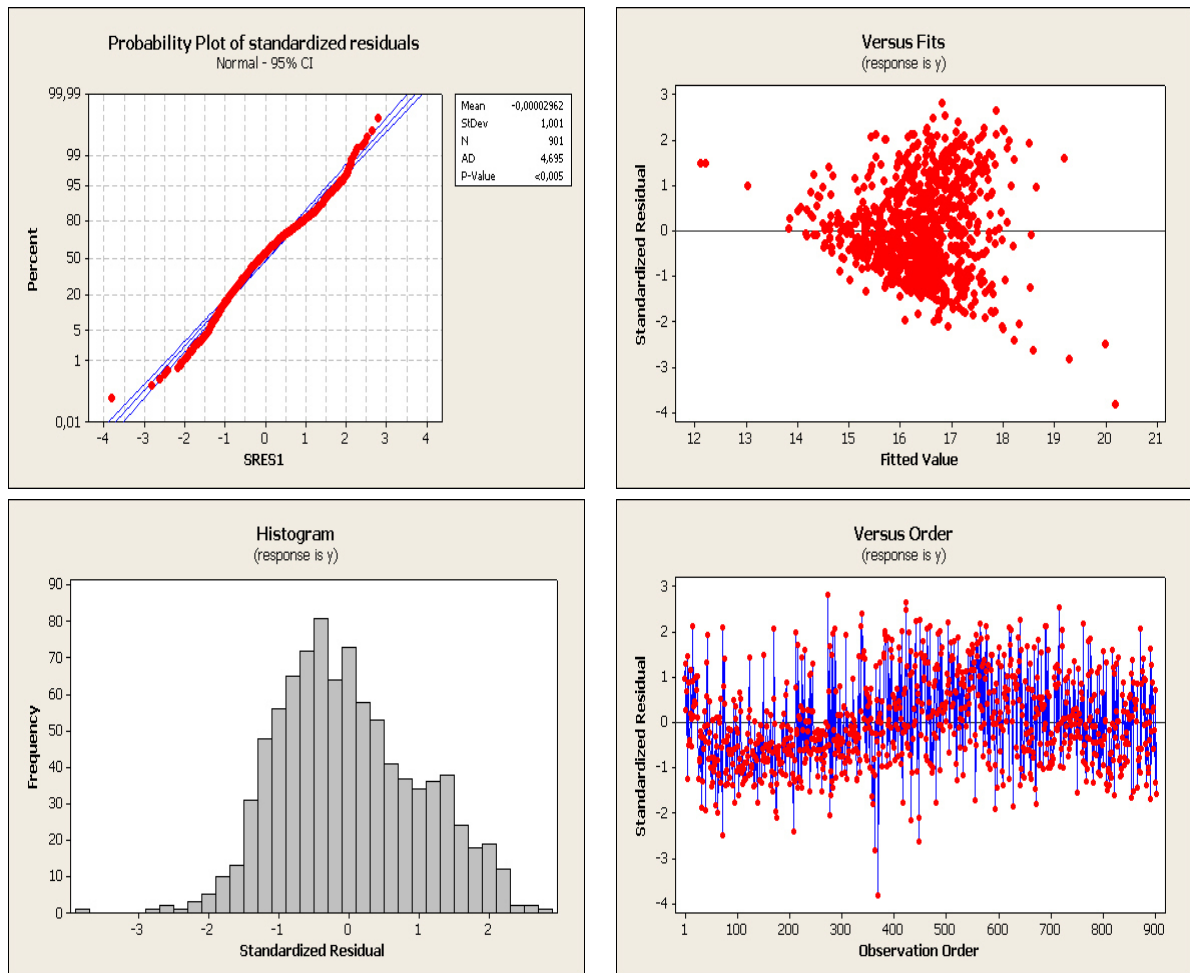
Rekn ut eit 95% konfidensintervall til forventa verdi av $\hat{y}_0$ og eit 95% prediksjonsintervall for y_0 .

Kva er skilnaden i tolkninga av desse to intervalla?

d) Kva for modellantakingar er gjord i den lineære regressjonen?

Basert på plotta presentert i Figur 3, trur du at desse antakingane er oppfylte? Grunngi svaret.

Korleis trur du dette vil påvirke resultatane gjeve i utskriftane?



Figur 3: Residualplott (normal plott basert på standardiserte residualer i øvre venstre panel, standardiserte residualer mot tilpassa verdi i øvre høgre panel, histogram basert på standardiserte residualer i nedre venstre panel og standardiserte residualer mot rekkjefølgda på observasjonane i nedre høgre panel) for regressjonsmodellen i ligning (1) for gråspurvdataene.

Oppgave 3 Støyping av stempelringar

Ein fabrikk produserer stempelringar til ein bilmotor ved ein støypeprosess. Fabrikken er interessert i å kontrollere diameteren på insiden av stempelringa.

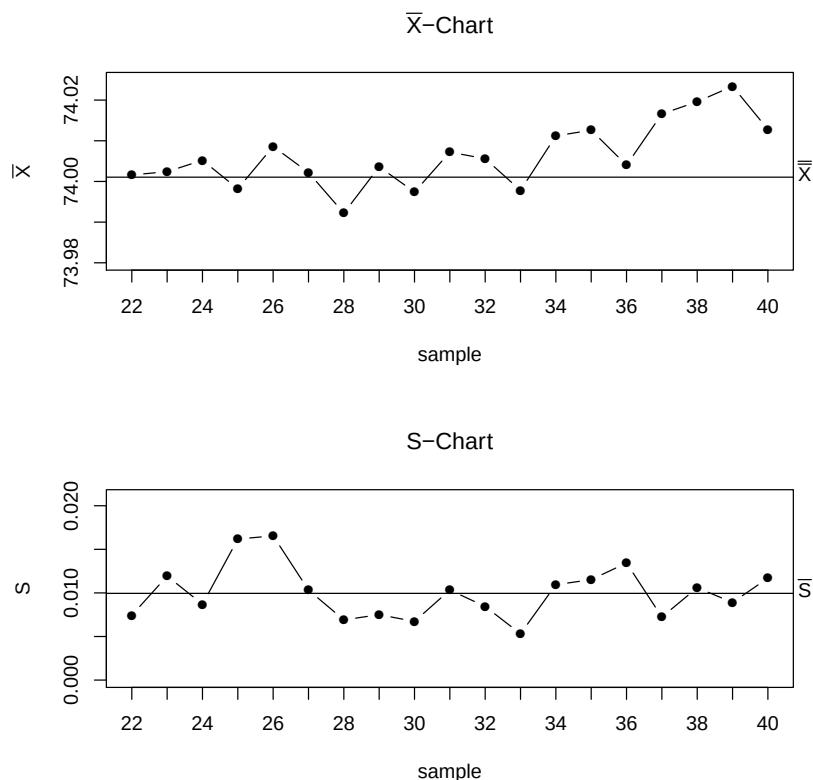
21 utval vart tekne, kvar av storleik 5. Vi antar prosessen er i kontroll då utvalet vart teke.

La X_{ij} vere målet på diameteren for stempelring j , fra utval i , kor $j = 1, 2, 3, 4, 5$ og $i = 1, 2, \dots, 21$. Vidare, $\bar{X}_i = \frac{1}{5} \sum_{j=1}^5 X_{ij}$, $S_i = \sqrt{\frac{1}{4} \sum_{j=1}^5 (X_{ij} - \bar{X}_i)^2}$, $\bar{\bar{X}} = \frac{1}{21} \sum_{i=1}^{21} \bar{X}_i$, og $\bar{S} = \frac{1}{21} \sum_{i=1}^{21} S_i$.

Basert på desse 21 utvala, finn vi at $\bar{\bar{x}} = 74.001$ og $\bar{s} = 0.00995$.

a) Konstruer eit S -chart og eit $\bar{X}$ - S -chart (med 3σ grensar).

Nye utval vart teken frå prosessen, vist i figuren under. Ser prosessen ut til å vere i kontroll? Grunngi svaret.



Oppgave 4 Godt eller dårleg temperament hos ektemenn og hustruer

I ein studie undersøkte ein om temperamentet til ektemenn og hustruer var uavhengige. 111 ektepar vart tilfeldig valt og ein slektning av ekteparet kryssklassifiserte ektemannen og hustrua til anten å ha eit godt eller dårleg temperament.

	God hustru	Dårleg hustru
God ektemann	24	27
Dårleg ektemann	34	26

- a) Er det grunn til å tru at temperamentet (godt/dårleg) til ektemannen er avhengig av temperamentet (godt/dårleg) til hustrua? Skriv ned null hypotesen og den alternative hypotesen og utfør ein hypotesetest basert på tabellen ovanfor. Bruk signifikansnivå 5%.