

Institutt for matematiske fag

## Eksamensoppgave i **TMA4255 Anvendt statistikk**

**Faglig kontakt under eksamen:** Anna Marie Holand

**Tlf:** 951 38 038

**Eksamensdato:** 30. mai 2014

**Eksamenstid (fra–til):** 09:00-13:00

**Hjelpemiddelkode/Tillatte hjelpemidler:** Alle trykte og håndskrevne hjelpemidler. Spesiell kalkulator.

### **Annen informasjon:**

- I utskrift fra MINITAB er komma brukt som desimalseparator.
- Signifikansnivå 5% skal brukes hvis ikke annet er spesifisert.
- Alle svar må begrunnes.

**Målform/språk:** bokmål

**Antall sider:** 6

**Antall sider vedlegg:** 0

**Kontrollert av:**

---

Dato

Sign



## Oppgave 1 Darwins maisplanter

Darwin (1876) studerte veksten av par av maisplanter, der den ene planten var fremstilt ved kryssbefruktning og den andre ved selvbefruktning. Målet hans var å demonstrere at kryssbefruktede planter har større fitness (f.eks. overlevelse og vekst) enn selvbefruktede planter.

Femten par av kryssbefruktede og selvbefruktede planter ble dyrket under identiske forhold. For hvert par ble høyden (i tommer) til hver plante registrert.

For par  $i$  la  $X_{1i}$  betegne høyden av planten fremstilt ved kryssbefruktning og  $X_{2i}$  betegne høyden av planten fremstilt ved selvbefruktning,  $i = 1, \dots, 15$ . Videre 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 kryssbefruktede planter er høyere enn selvbefruktede planter? Skriv ned null hypotesen og den alternative hypotesen, velg en testobservator og gjennomfør en hypotesetest. Bruk signifikansnivå  $\alpha = 0.05$ . Spesifiser hvilke antagelser du gjør.

- b) Anta at  $X_1$  og  $X_2$  ikke er normalfordelte. Utfør en fortegnstest (sign test) for å teste om kryssbefruktede planter blir høyere enn selvbefruktede planter. Kommenter dine resultater/funn.

**Oppgave 2 Gråspurv**

Gråspurv (*Passer domesticus*) populasjoner langs kysten av Midt- og Nord-Norge har blitt studert av forskere ved NTNU gjennom flere tiår.

Gråspurvhanner har en svart brystflekk (svarte fjær i brystet). Brystflekken er variabel i størrelse og tidligere studier indikerer at størrelsen på brystflekken er et signal på sosial status og fysisk form. I en studie ble disse variablene målt.

- $y$ , synlig brystflekkstørrelse målt i mm.
- $x_1$ , tarslengde (lengden av foten) målt i mm.
- $x_2$ , nebbhøyden målt i mm.
- $x_3$ , total brystflekkstørrelse målt i mm.

Forskere tror at størrelsen på den synlige brystflekken,  $y$ , er avhengig av tarslengde,  $x_1$ , (som indikerer generell kroppstørrelse), nebbhøyde,  $x_2$ , (som sier noe om evnen til å spise mat av en bestemt type) og den totale brystflekken,  $x_3$ , (som sier noe om den potensielle synlige brystflekken).

Disse trekkene ble målt for  $n = 901$  gråspurvhanner. Vi analyserer dataene med en multiple lineær model:

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

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

En MINITAB utskrift fra statistisk analyse finnes i Figur 1 og Figur 2 og et plot av standardiserte residualer finnes i Figur 3.

- a) Utfør en 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%.

Hvor mange prosent av variasjonen forklarer 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 fra statistiske analyser 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 fra statistiske analyser av gråspurvdataene.

b) Hva er et passende estimat for  $\sigma^2$ ?

I utskriften fra en utført variansanalyse av den tilpassede multiple lineære modellen i Figur 2 er fire numeriske verdier byttet ut med spørsmålsteget. Beregn numerisk verdi for hver av disse, og forklar hva hver av disse betyr.

c) Bruk den estimerte regresjonsmodellen i Figur 1 til å beregne et punktestimat for synlig brystflekkstørrelse,  $\hat{y}_0$ , for observert tarslengde  $x_1^0 = 20$ , nebbhøyde  $x_2^0 = 8$  og total brystflekkstørrelse  $x_3^0 = 19$ . Det er gitt at det estimerte standardavviket til  $\hat{y}_0$  er 0.097.

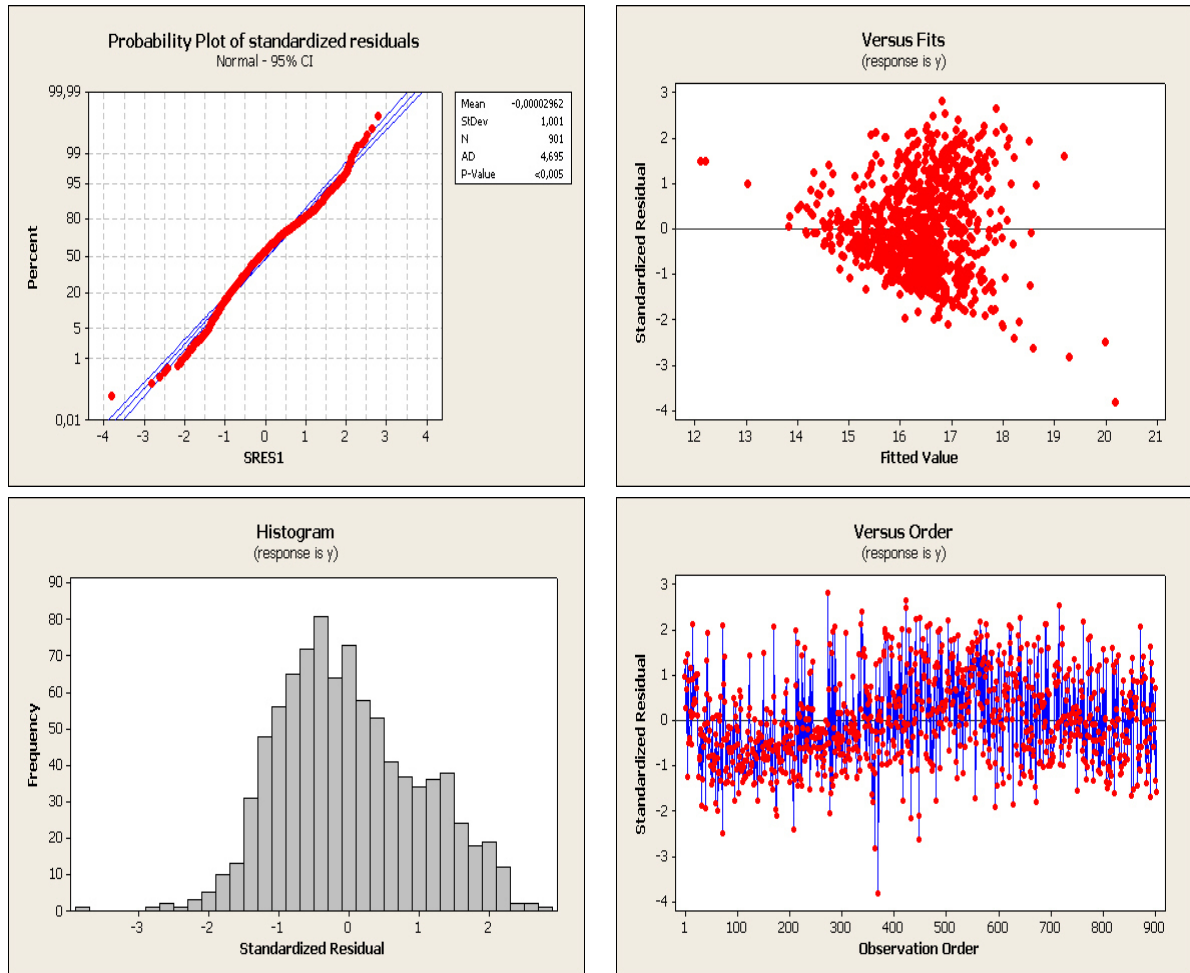
Beregn et 95% konfidensintervall til forventet verdi av  $\hat{y}_0$  og et 95% prediksjonintervall for  $y_0$ .

Hva er forskjellen i tolkningen av disse to intervallene?

d) Hvilke modellantagelser er gjort i den lineære regressjonen?

Basert på plottene presentert i Figur 3, tror du at disse antagelsene er oppfylt? Begrunn svaret.

Hvordan tror du dette vil påvirke resultatene gitt i utskriftene?



Figur 3: Residualplott (normal plott basert på standardiserte residualer i øvre venstre panel, standardiserte residualer mot tilpassede verdier i øvre høyre panel, histogram basert på standardiserte residualer i nedre venstre panel og standardiserte residualer mot rekkefølgen på observasjonene i nedre høyre panel) for regressjonsmodellen i ligning (1) for gråspurvdtaene.

### Oppgave 3 Støping av stempelringer

En fabrikk produserer stempelringer til en bilmotor ved en støpeprosess. Fabrikken er interessert i å kontrollere diameteren på insiden av stempelringene.

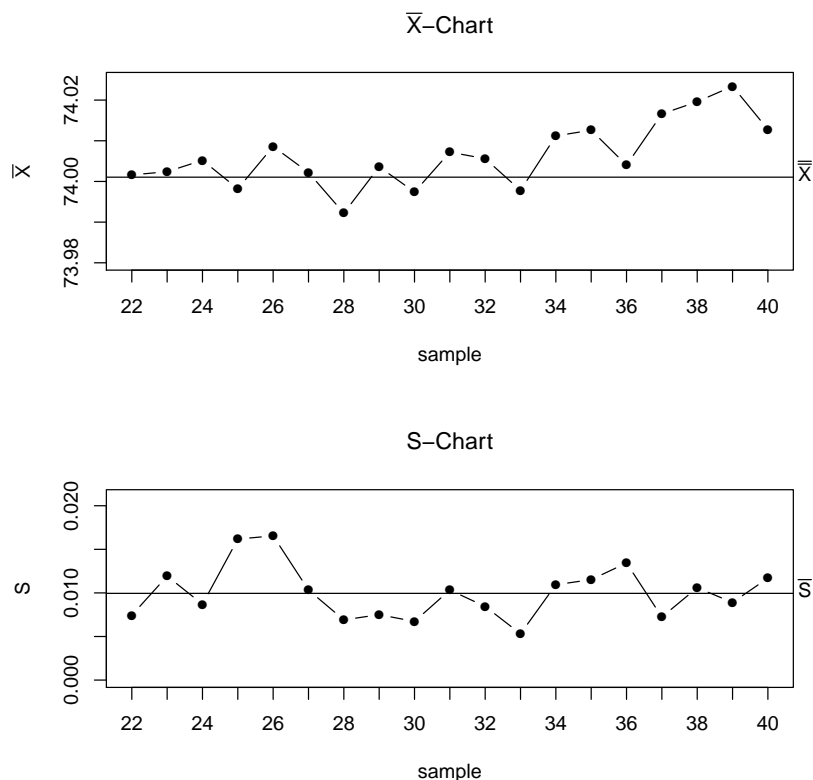
21 utvalg ble tatt, hvert av størrelse 5. Prosessen antas å være i kontroll når utvalget ble tatt.

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

Basert på disse 21 utvalgene, finner vi at  $\bar{\bar{x}} = 74.001$  og  $\bar{s} = 0.00995$ .

- a) Konstruer et  $S$ -chart og et  $\bar{X}$ - $S$ -chart (med  $3\sigma$  grenser).

Nye utvalg ble tatt fra prosessen, vist i figuren under. Ser prosessen ut til å være i kontroll? Begrunn svaret.



**Oppgave 4 Godt eller dårlig temperament hos ektemenn og hustruer**

I en studie undersøkte man om temperamentet til ektemenn og hustruer var uavhengige. 111 ektepar ble tilfeldig valgt og en slektning av ekteparet kryssklassifiserte ektemannen og hustruen til enten å ha et godt eller dårlig temperament.

	God hustru	Dårlig hustru
God ektemann	24	27
Dårlig ektemann	34	26

- a) Er det grunn til å tro at temperamentet (godt/dårlig) til ektemannen er avhengig av temperamentet (godt/dårlig) til hustruen? Skriv ned null hypotesen og den alternative hypotesen og utfør en hypotesetest basert på tabellen ovenfor. Bruk signifikansnivå 5%.