

Faglig kontakt under eksamen:

Bo Lindqvist
Tlf. 975 89 418

BOKMÅL

EKSAMEN I FAG TMA4255
ANVENDT STATISTIKK

Onsdag 26. mai 2010

Tid: 09:00–13:00

Tillatte hjelpemidler:

Alle trykte og håndskrevne hjelpemidler. Spesiell kalkulator.

Sensur: 16. juni 2010

Eksamen består av 8 punkt som alle teller likt ved sensuren.

Oppgave 1

En biolog skal undersøke hvordan veksten av blåskjell påvirkes av belysningen når visse andre faktorer holdes konstante. Veksten (y) for 10 blåskjell blir målt under ulik grad av belysning (x). Observasjonene (y_i, x_i) for $i = 1, 2, \dots, 10$ er gitt i tabellen nedenfor:

i	1	2	3	4	5	6	7	8	9	10
x_i	1	1	2	2	3	3	4	4	5	5
y_i	16	18	17	20	25	21	23	20	17	19

MINITAB-utskriften og plottene på neste side viser resultatet av å tilpasse en multippel lineær regresjonsmodell der den forventede vekst y er et annengradspolynom i belysningen x . Mer presist antas modellen:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i \quad (1)$$

for $i = 1, \dots, 10$, der $\epsilon_1, \dots, \epsilon_{10}$ er uavhengige og $N(0, \sigma^2)$.

I MINITAB-utskriften er kovariatene x^2 betegnet med $x * x$.

Regression Analysis: Y versus x; x*x

The regression equation is

$$Y = 10,1 + 7,36 x - 1,14 x^2$$

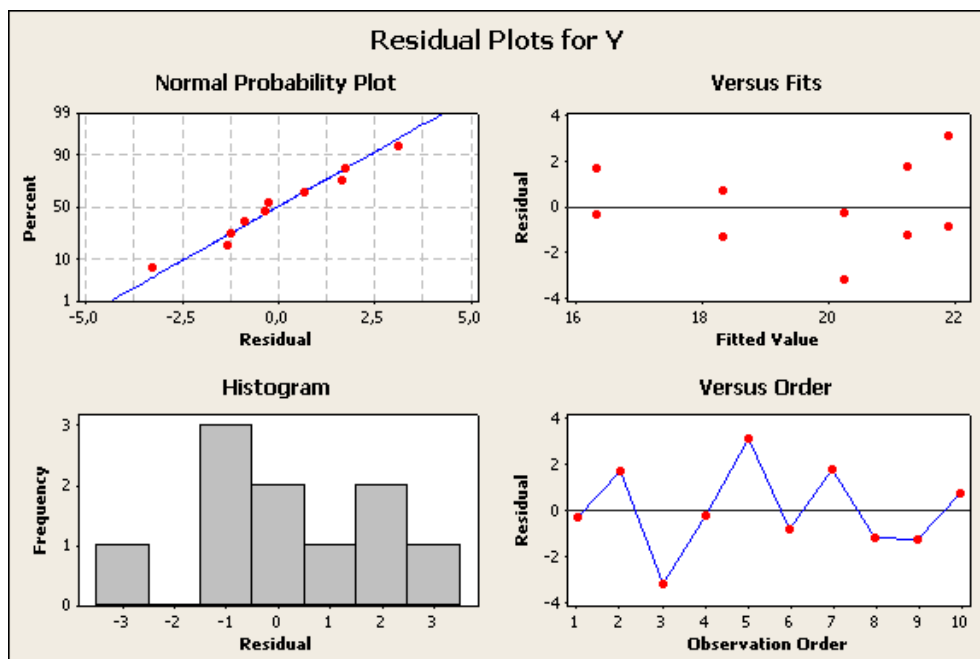
Predictor	Coef	SE Coef	T	P
Constant	10,100	3,183	3,17	0,016
x	7,357	2,425	3,03	0,019
x*x	-1,1429	0,3966	-2,88	0,024

S = 2,09859 R-Sq = 57,4% R-Sq(adj) = 45,3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	41,571	20,786	4,72	0,050
Residual Error	7	30,829	4,404		
Total	9	72,400			

Source	DF	Seq SS
x	1	5,000
x*x	1	36,571



- a) Bruk utskriften og plottene fra MINITAB til å diskutere kort om modellen (1) og de tilhørende antagelsene gir en god beskrivelse av dataene.

Hva betyr F -verdien 4.72 i tabellen under “Analysis of Variance” i utskriften? Sett opp den nullhypotesen som testes med denne testobservatoren. Hvordan kan du konkludere fra den oppgitte p -verdien?

- b) Man er interessert i å teste $H_0 : \beta_2 = 0$ mot $H_1 : \beta_2 < 0$. Sett opp en testobservator for dette og finn kritisk verdi når signifikansnivået er 5%. Hva blir konklusjonen på testen? Hva er p -verdien for denne testen?

Fra lignende studier har man funnet at koeffisienten foran andregradsleddet i modellen (1) er -1.0 . Er det grunnlag for å tro noe annet i dette tilfellet? Formuler dette spørsmålet som et hypotesetestingsproblem og utfør testen med et signifikansnivå på 5%.

- c) Sett opp estimatet for σ som kan leses ut av MINITAB-utskriften.

Man ønsker også et 95% konfidensintervall for σ . Du skal finne dette.
(Hint: Bruk at SSE/σ^2 er kjikvadrat-fordelt.)

Hvordan kan konfidensintervallet brukes til å teste nullhypotesen $H_0 : \sigma = 1$? Hva blir i så fall den alternative hypotesen, og hva blir signifikansnivået?

- d) Vis at den estimerte regresjonsligningen har maksimum for $x_0 = 3.23$.

Man ønsker et prediksjonsintervall for responsen y_0 ved denne x -verdien. Beregn først punktprediksjonen \hat{y}_0 . Det oppgis at estimert standardavvik for \hat{y}_0 er 1.024. Bruk dette til å beregne et 95% prediksjonsintervall for y_0 .

Oppgave 2

En gruppe amerikanske biologer studerer hvordan zooplankton lever i to innsjøer ved navn Rose og Dennison. De setter opp tolv tanker i sitt laboratorium med vann fra de to innsjøene, seks tanker for hver innsjø. De tilsetter ett av tre næringstilskudd og etter 30 dager teller de antall zooplankton i en enhet vann fra hver tank. Dataene er gitt som følger, der næringstilskuddene (se variabelen Supplement) er benevnt 1,2,3, og variabelen som betegner innsjø er kalt 'Lake'.

Row	Zooplankton	Supplement	Lake
1	34	1	Rose
2	43	1	Rose
3	57	1	Dennison
4	40	1	Dennison
5	85	2	Rose
6	68	2	Rose
7	67	2	Dennison

8	53	2	Dennison
9	41	3	Rose
10	24	3	Rose
11	42	3	Dennison
12	52	3	Dennison

Nedenfor er utskrift av en to-veis ANOVA ved bruk av MINITAB

Two-way ANOVA: Zooplankton versus Supplement; Lake

Source	DF	SS	MS	F	P
Supplement	2	1918,50	959,250	9,25	0,015
Lake	1	21,33	21,333	0,21	0,666
Interaction	2	561,17	280,583	2,71	0,145
Error	6	622,00	103,667		
Total	11	3123,00			

S = 10,18 R-Sq = 80,08% R-Sq(adj) = 63,49%

a) Skriv ned modellen og modellantagelsene som er brukt i MINITAB-analysen.

Hvordan fortolker du resultatene om hovedeffekter og interaksjon for de to faktorene Supplement og Lake?

Basert på resultatene ovenfor beslutter biologene å se bort fra faktoren Lake og isteden betrakte problemet som et en-veis ANOVA problem med den ene faktoren Supplement.

b) MINITABs ANOVA-tabell for en-veis analysen med Supplement som den eneste faktoren er av formen:

Source	DF	SS	MS	F	P
Supplement	?	1918,50	?	?	0,014
Error	?	?	?		
Total	?	3123,00			

Forklar hvorfor kvadratsummen (SS) svarende til faktoren Supplement, såvel som Total kvadratsum, er uendret fra to-veis tilfellet.

Fyll så inn de korrekte tallene i hvert av feltene i tabellen ovenfor som er merket med et spørsmålsteget (?)

Hva testes med F -observatoren og hva blir konklusjonen i det foreliggende tilfellet?

Hva er det nye estimatet for σ ?

Oppgave 3

For å undersøke sammenhengen mellom høyt blodtrykk og røyking ble følgende opplysninger innsamlet fra 180 tilfeldig valgte personer:

	Ikke-røyker	Moderat røyker	Storrøyker
Høyt blodtrykk	20	36	32
Normalt blodtrykk	48	26	18

- a) Test nullhypotesen at det er uavhengighet mellom forekomsten av høyt/normalt blodtrykk og røykevane. Hva blir konklusjonen? Bruk signifikansnivå 1%.

Oppgave 4

En bedrift som produserer store stålplater ønsker å ta i bruk kvalitetskontroll. Målet er å etablere et kontrollkort (control chart) for antall overflatedefekter pr plate. Dataene er som følger:

Plate nummer	1	2	3	4	5	6	7	8	9	10	11	12
Antall defekter	4	2	1	3	0	4	6	3	2	2	1	2

- a) Hvilken type kontrollkort ville du bruke her? Begrunn svaret. Beregn kontrollgrensene ved å bruke de oppgitte dataene.

Ser prosessen ut til å være i kontroll?