EKSAMEN/EXAM STA-3300

Eksamen i/Examination in	**: STA-3300 Brukerkurs i statistikk 2**
Dato/Date	**:Tirsdag 6. desember 2005**
Tid/Time	**:Kl 09:00 til 13:00**
Sted/Site	**:Åsgårdveien 9**
Tillatte hjelpemidler/Aids allowed	**:Alle skriftlige og alle kalkulatorer/**
	All notes, books and pocket calculator

Oppgavesettet er på 8 sider inkludert førstesiden/

The exam paper consists of 8 pages including the front page

Kontaktperson under eksamen/ Contact person during the exam:

Amanuensis Elinor Ytterstad, tlf 77 64 4015

Eksamenssettet består av oppgave 1 a-h og oppgave 2 a og b /

The exam paper consists of Problem 1 a-h and Problem 2 a and b

Oppgave/Problem 1

Hardhet til tolv stålplater med forskjellig kopperinnhold (%) og herdetemperatur (F) er målt med følgende resultat:
/
Twelve specimens of sheet steel, having different copper contents (%) and annealing temperatures (Fahrenheit), are measured for hardness with the following results (Norwegian decimal comma):

Specimen ID	hardness Y	copper X_1	temp X_2	temp (in thousands degrees) X_5
1	78,9	0,02	1000	1,0
2	65,1	0,02	1100	1,1
3	55,2	0,02	1200	1,2
4	56,4	0,02	1300	1,3
5	80,9	0,10	1000	1,0
6	69,7	0,10	1100	1,1
7	57,4	0,10	1200	1,2
8	55,4	0,10	1300	1,3
9	85,3	0,18	1000	1,0
10	71,8	0,18	1100	1,1
11	60,7	0,18	1200	1,2
12	58,9	0,18	1300	1,3

De vedlagte SPSS-utskriftene viser resultater fra regresjonsanalyse for følgende modeller:
/
The accompanying SPSS-outputs are results from fitting the following regression models:

Model 1: $Y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + E$

Model 2: $Y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_2^2 + E$

Model 3: $Y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_2^2 + \beta_4 \cdot x_1 \cdot x_2 + E$

Model 4: $Y = \beta_0 + \beta_1 \cdot x_1 + \beta_5 \cdot x_5 + \beta_6 \cdot x_5^2 + E$

a) Kommentèr spredningsplottene Y mot x_1 og Y mot x_2. Hvilke variable ser ut til å forklare stålets hardhet.
/
Comment on the scatter plots of Y against x_1 and Y against x_2. Which variables seem to be reasonable to include in a model explaining hardness of steel?

b) Regn ut R^2 i model 1 og forklar hva denne verdien sier om modellen

Calculate R^2 for model 1 and explain what this value is telling us about the model

c) Skriv ned estimert regresjonsligningen for model 2. Test om noen av variablene kan fjernes fra modellen. Formuler null-hypoteser ut utfør testene. Hva er dine konklusjoner fra hver av testene.
/
Write down the equation for the fitted model 2. Test whether any of the variables may be deleted from this model. Formulate the null-hypothesis and perform the tests. What are your conclusions from each test.

d) Bruk de vedlagte utskriftene, regn ut og fyll inn kvadratsummer og frihetsgrader som mangler i tabellen under.
/
Use the accompanying outputs, calculate and fill in the missing sum-of-squares and degrees of freedom in the table below.

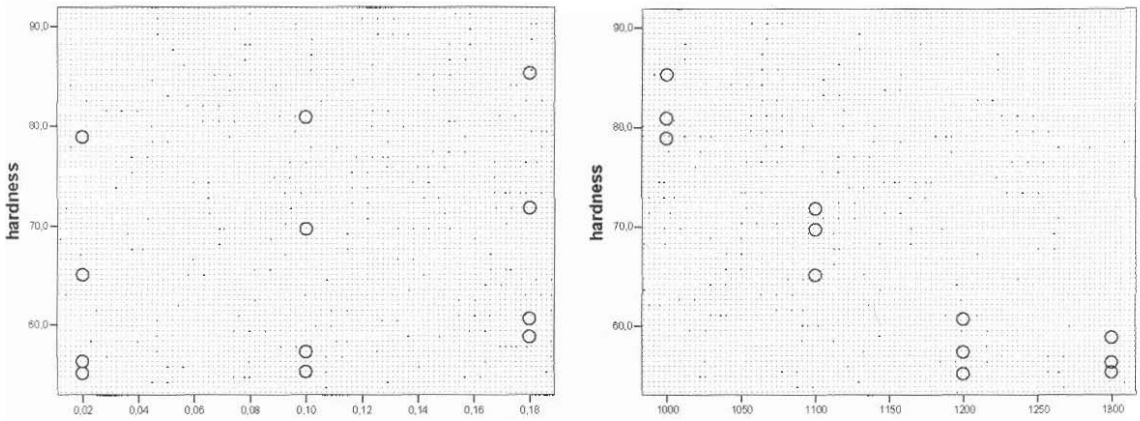
	SS	df
$SS(x_1, x_2)$		
$SS(x_2^2 \mid x_1, x_2)$		
$SS(x_1 \cdot x_2 \mid x_1, x_2, x_2^2)$	4,16	
$SSE(x_1, x_2, x_2^2, x_1 \cdot x_2)$	17,78	
SSY	1281,53	11

e) Test hypotesen: $H_0 : \beta_4 = 0$ i modell 3. Bruk 5% signifikansnivå.
Hva blir konklusjonen av denne testen?
Forklar sammenhengen mellom en t-test og en partiell F-test for denne hypotesen
Test også $H_0 : \beta_3 = \beta_4 = 0$ i modell 3 med 5% signifikansnivå.
Hva blir konklusjonen av denne testen?
/
Test the hypothesis: $H_0 : \beta_4 = 0$ in model 3 with a 5% significance level.
What is the conclusion of this test?
Explain the connection between a t-test and a partial F-test in testing this hypothesis.
Also test $H_0 : \beta_3 = \beta_4 = 0$ in model 3 with a 5% significance level.
What is the conclusion of this test?

f) Forklar hvilke kriterier som kan benyttes i valg av beste modell.
Hvilken av modellene 1, 2 eller 3 er best?
Er det noen uteliggerobservasjoner i datamaterialet? Forklar.
/
Explain which criteria you can use for choosing a best model.
What is the best model of 1, 2 and 3?
Are there any outlier observations? Explain.

g) I en tidligere og større undersøkelse av hardhet til stål med forskjellig kopperinnhold, kom man frem til at $\beta_1 = 30$ i modell 2. Test om β_1 er endret her, i.e. test $H_0 : \beta_1 = 30$ mot $\beta_1 \neq 30$. Bruk 5% signifikansnivå.
/
In a previous and larger study of hardness associated with copper content, they found $\beta_1 = 30$ in model 2. Test whether there are reasons to believe β_1 is different in this study, i.e. test $H_0 : \beta_1 = 30$ against $\beta_1 \neq 30$. Use 5% significance level.

h) Beskriv forskjellen mellom modell 2 og 4.
Regn ut predikert hardhet til stål som er herdet på 1000 grader Fahrenheit og inneholder 0,1 % kopper, for både modell 2 og 4.
Kommenter svaret, er disse to predikerte verdier forskjellige eller identiske og hvorfor?
/
Describe the difference between model 2 and 4?
Calculate from both model 2 and 4 the predicted hardness of steel annealed at 1000 degrees Fahrenheit, containing 0,1 % copper.
Comment on these two predicted values, are they different or similar and why?



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	[a]			
2	,991[b]	,983	,976	1,6560
3	,993[c]	,986	,978	1,5937

a. Predictors: (Constant), temp, X2, copper content %, X1

b. Predictors: (Constant), temp, X2, copper content %, X1, X2 ^2

c. Predictors: (Constant), temp, X2, copper content %, X1, X2 ^2, X1*X2

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	161,336	11,433		14,112	,000
	copper content %, X1	32,969	16,754	,208	1,968	,081
	temp, X2	-,086	,010	-,925	-8,735	,000
2	(Constant)	553,245	62,824		8,806	,000
	copper content %, X1	32,969	7,319	,208	4,505	,002
	temp, X2	-,774	,110	-8,369	-7,030	,000
	X2 ^2	,000299	,000048	7,450	6,258	,000
3	(Constant)	543,973	60,892		8,933	,000
	copper content %, X1	125,688	72,788	,794	1,727	,128
	temp, X2	-,766	,106	-8,282	-7,216	,000
	X2 ^2	,000299	,000046	7,450	6,503	,000
	X1*X2	-,081	,063	-,595	-1,280	,241

a. Dependent Variable: hardness

ANOVA[d]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1152,189	2	576,094	40,087	,000[a]
	Residual	129,340	9	14,371		
	Total	1281,529	11			
2	Regression	1259,590	3	419,863	153,098	,000[b]
	Residual	21,940	8	2,742		
	Total	1281,529	11			
3	Regression	1263,750	4	315,937	124,389	,000[c]
	Residual	17,779	7	2,540		
	Total	1281,529	11			

a. Predictors: (Constant), temp, X2, copper content %, X1

b. Predictors: (Constant), temp, X2, copper content %, X1, X2 ^2

c. Predictors: (Constant), temp, X2, copper content %, X1, X2 ^2, X1*X2

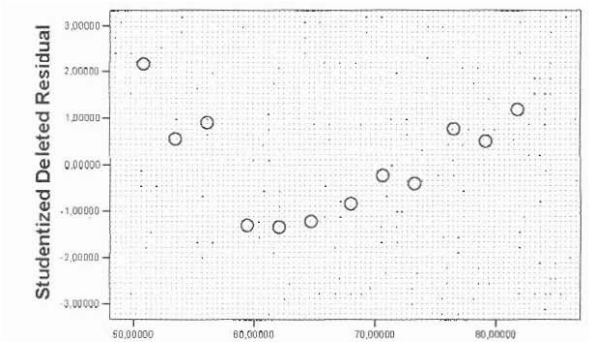
d. Dependent Variable: hardness

Modell 2:

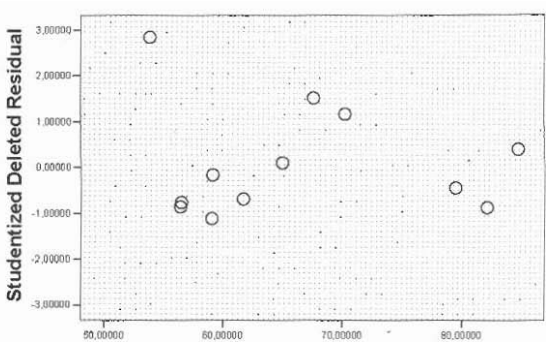
ID	Cook's	*leverage hi
1	0,04	0,44
2	0,001	0,31
3	0,09	0,31
4	0,85	0,44
5	0,09	0,32
6	0,11	0,18
7	0,07	0,18
8	0,07	0,32
9	0,04	0,44
10	0,14	0,31
11	0,06	0,31
12	0,006	0,44

*Leverage as describe in the text book

Model 1



Model 2



Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
4	(Constant)	553,245	62,824		8,806	,000
	copper content %, X1	32,969	7,319	,208	4,505	,002
	x5	-773,583	110,036	-8,369	-7,030	,000
	x5*x5	299,167	47,806	7,450	6,258	,000

a. Dependent Variable: hardness

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
4	Regression	1259,590	3	419,863	153,098	,000[a]
	Residual	21,940	8	2,742		
	Total	1281,529	11			

a. Predictors: (Constant), x5*x5, copper content %, X1, x5

b. Dependent Variable: hardness

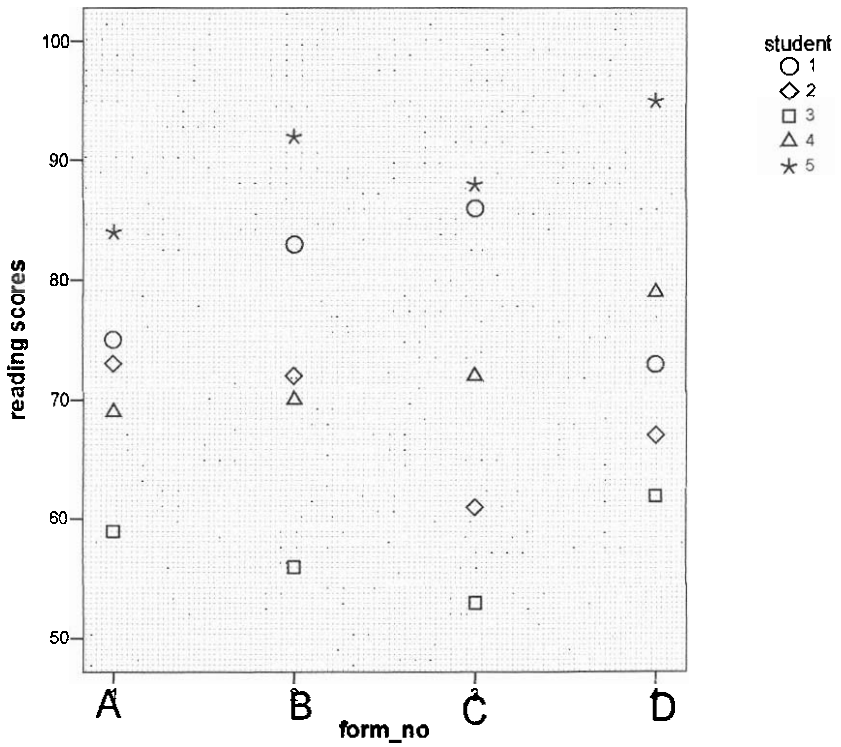
Oppgave/Problem 2

Fire ulike skjemaer for ekvivalente tester av leseferdighet er gitt til hver av fem studenter. Resultatet er gitt i tabellen under:
/
Four different, though supposedly equivalent, forms of a standardized reading achievement test were given to each of five students, and the following are the scores which they obtained:

	Student 1	Student 2	Student 3	Student 4	Student 5
Form A	75	73	59	69	84
Form B	83	72	56	70	92
Form C	86	61	53	72	88
Form D	73	67	62	79	95

a) Kommentèr datamaterialet fra hva du ser i spredningsplottet under.
Er det forskjeller mellom studenter? Forskjeller mellom skjemaer?
Fyll inn de verdiene som mangler i en-veis ANOVA-tabellen nedenfor
/
What can be said about the data from only looking at the scatter plot below?
Differences between students? Between forms?
Fill in the missing values in the one-way ANOVA-table below.

b) Bruk SPSS-utskriftene nedenfor og test om det er rimelig å anta at de fire skjemaene er ekvivalente i fastsettelse av leseferdighet.
I.e. H_0 : 'De fire skjemaene A, B, C and D gir samme gjennomsnittelige poengsum i leseferdighet.'
Test hypotesen både med en en-veis- og en to-veis variansanalyse. Hva blir konklusjonene. Hvilken modell er best?
Bruk 5% signifikansnivå.
/
Use the results from the SPSS-outputs below and test whether it is reasonable to treat the four forms as equivalent.
I.e. H_0 :"The four forms A, B, C and D give equal mean scores in reading achievment."
What are the conclusions from a one-way and a two-way analysis of variance? Which is the better model?
Use 5% significance level.



Results from one-way analysis of variance:

	SS	df	Mean SS	F	Sig.
Between groups (form)	42,950				0,685
Within	2666,000				
total					

Results from two-way analysis of variance:

	SS	df	Mean SS	F	Sig.
Form	42,950	3	14,317	0,506	0,685
Student	2326,700	4	581,675	20,572	0,000
Within	339,300	12	28,275		
total	2708,950	19			