EKSAMENSOPPGAVE I STA-2000

Eksamen i	:	**STA-2000, Brukerkurs i statistikk II**
Eksamensdato	:	**Fredag 27. februar 2004**
Tid	:	**Kl 0900-1300**
Sted	:	**Samvirkegården 4. etg.**
Tillatte hjelpemidler	:	**Alle trykte og skrevne, samt kalkulator**

Oppgavesettet er på 10 sider inkl. forside

Kontaktperson under eksamen:

Amanuensis Elinor Ytterstad, tlf 44 015

(Oppgavesettet består av punktene a til j)

I en australsk studie ble det samlet inn data fra 200 pasienter som ventet på rettslig behandling av erstatningskrav. Pasientene var hodeskadet etter ulykker.
Data inneholder alder, kjønn, mental kapasitet før og etter skaden og forventet erstatningsbeløp. Vi skal undersøke i hvilken grad størrelsen av forventet erstatningsbeløp kan forklares av mental kapasitet, alder og kjønn. Det er 100 pasienter av hvert kjønn hvorav 20 i hver av fem aldersgruppe.

Forklaring til variablene:

- ID: pasient nr
- Y=AMOUNT: forventet erstatningsbeløp
- Z_S=FEMALE: 0 hvis mann, 1 hvis kvinne
- AGEGROUP: aldersgruppe
  - 1: 0-11 år
  - 2: 12-20 år
  - 3: 21-35 år
  - 4: 36-60 år
  - 5: 61- år
- X_1= MENTAL1: mental kapasitet før hodeskade
- X_2= MENTAL2: mental kapasitet etter hodeskade

Et utdrag av datamaterialet:

ID	AMOUNT	SEX	FEMALE	AGEGROUP	MENTAL1	MENTAL2
1	45186	M	0	1	49	46
2	29886	M	0	1	50	48
.....						
21	42428	M	0	2	80	69
22	33480	M	0	2	64	65
.....						
41	22647	M	0	3	106	84
42	24795	M	0	3	71	56
.....						
61	27809	M	0	4	83	64
62	4759	M	0	4	87	68
.....						
81	33876	M	0	5	75	48
82	15345	M	0	5	70	46
.....						
101	10000	F	1	1	45	47
102	39160	F	1	1	56	58
.....						
121	20025	F	1	2	60	64
122	15500	F	1	2	77	76
.....						
141	23400	F	1	3	70	67
142	28650	F	1	3	73	72
.....						
161	8200	F	1	4	71	63
162	5450	F	1	4	63	59
.....						
181	12675	F	1	5	76	59
182	10080	F	1	5	55	46
.....						
200	16485	F	1	5	77	61

Forklaring til andre variable som brukes i modeller i denne oppgaven:

Z_{A1}= 1 hvis AGEGROUP=1, og 0 ellers
Z_{A2}= 1 hvis AGEGROUP =2 , og 0 ellers
Z_{A3}= 1 hvis AGEGROUP =3, og 0 ellers
Z_{A4}= 1 hvis AGEGROUP =4, og 0 ellers
Z_O= 1 hvis AGEGROUP ≥4 og 0 ellers

På de siste sider presenteres spredningsplott og resultater av regresjonsanalyser fra fem modeller:

1. $Y=\beta_0+\beta_1 Z_S+\beta_{11} Z_{A1}+\beta_{12} Z_{A2}+\beta_{13} Z_{A3}+\beta_{14} Z_{A4}+\beta_3 X_1+\beta_4 X_2+E$
2. $Y=\beta_0+\beta_1 Z_S+\beta_2 Z_O+\beta_3 X_1+\beta_4 X_2+\beta_5 Z_S Z_O+\beta_6 X_1 Z_O+\beta_7 X_2 Z_O+E$
3. $Y=\beta_0+\beta_1 Z_S+\beta_2 Z_O+\beta_3 X_1+\beta_5 Z_S Z_O+\beta_6 X_1 Z_O+E$
4. $Y=\beta_0+\beta_1 Z_S+\beta_2 Z_O+\beta_5 Z_S Z_O+\beta_6 X_1 Z_O+E$
5. $Y=\beta_0+\beta_1 Z_S+\beta_2 Z_O+\beta_6 X_1 Z_O+E$

På siste side finnes resultater av en variansanalyse.

a) Forklar forskjellen mellom modell 1 og 2

b) Studèr de utvalgte spredningsplottene på side 5. Hva kan en si om variablene alder, kjønn og mental kapasitet før skade i forhold til forventet erstatningsbeløp. Skriv opp de variablene du kan se av disse plottene bør inkluderes i en modell.

c) Undersøk i modell 1 om det er signifikant forskjell mellom aldersgruppe 4 og 5 når det gjelder størrelsen av forventet erstatningsbeløp. Still opp nullhypotesen og forklar hvordan signifikanstesten utføres.

d) Regn ut og fyll inn det som mangler i ANOVA-tabellen nedenfor.

Modell	df	SS	MS
Regresjon			
Z_S		2661354629	
$Z_O \mid Z_S$		4268878029	
$Z_S \cdot Z_O \mid Z_S, Z_O$		80729431	
$X_1 \cdot Z_O \mid Z_S \cdot Z_O, Z_S, Z_O$			
$X_1 \mid X_1 \cdot Z_O, Z_S \cdot Z_O, Z_S, Z_O$			
Residual	194		81544363.47
Total	199		

e) Test hypotesen $H_0: \beta_3 = \beta_6 = 0$ i modell 3. Bruk 5% signifikansnivå.
Utfør en added-last-test av interaksjon mellom kjønn- og alder-variabelene i modell 4. Er endringen i forventet erstatningsbeløp fra ung til gammel den samme hos kvinner og menn?

f) Hvilken modell vil du velge som den beste. Forklar hvilke kriterier du bruker for valg av modell. Hvilke variable har betydning for størrelsen på forventet erstatningsbeløp?

g) Studer Jackknife-residualene side 9 og forklar årsaken til at det dannes to loddrette linjer i spredningsplottene fra modell 5.

h) Undersøk vha variansanalysen side 10 om det er signifikante forskjeller mellom aldergrupper og kjønn når det gjelder størrelsen av forventet erstatningsbeløp. Formulér hypoteser og utfør testene på 5% nivå.

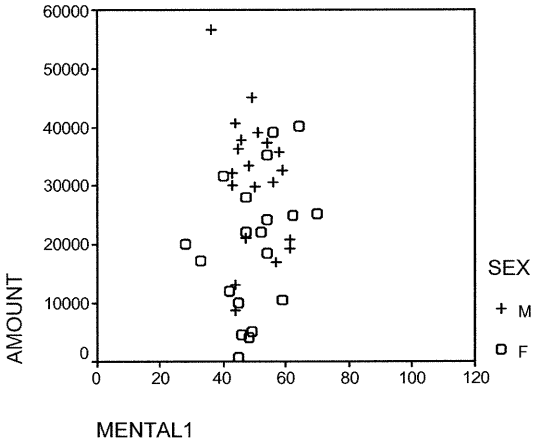
i) Hvilke likhetstrekk er det mellom ANOVA-tabellen fra variansanalysen og tabellen under punkt d)?

j) Skriv opp regresjonsmodellen som svarer til variansanalysemodellen uten interaksjonsleddet FEMALE*AGEGROUP.

Utvalgte spredningsplott:

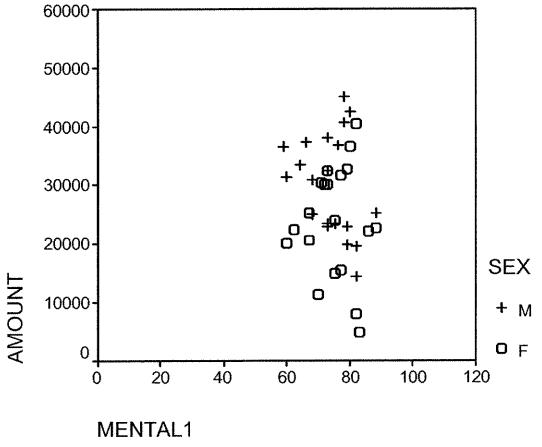
AGEGROUP 1

(0-11 years old)



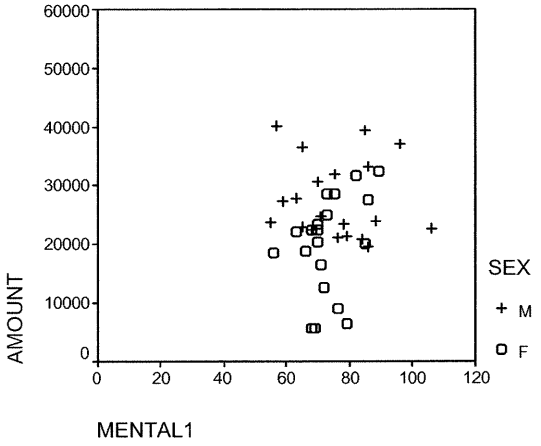
AGEGROUP 2

(12-20 years old)



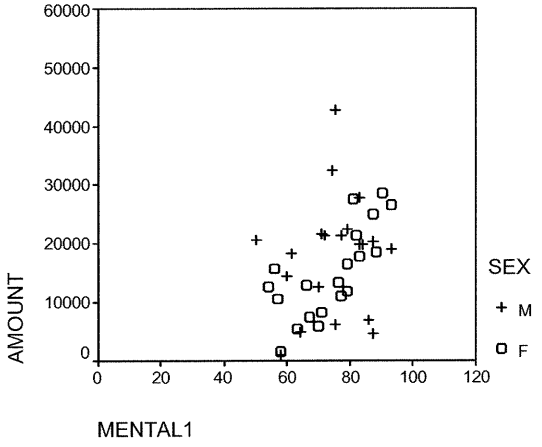
AGEGROUP 3

(21-35 years old)



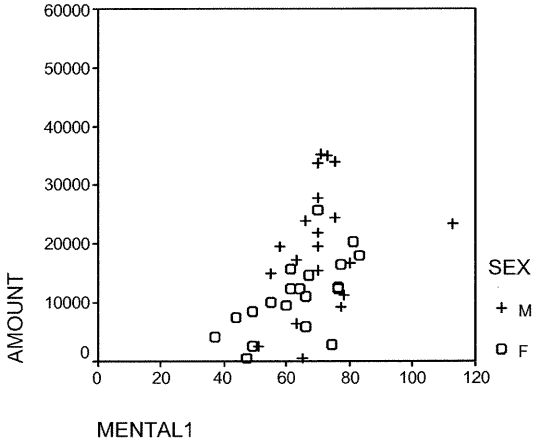
AGEGROUP 4

(36-60 years old)



AGEGROUP 5

(61+ years old)



Model 1, ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7,90E+09	7	1128046848	13,532	,000 [a]
	Residual	1,60E+10	192	83360433,18		
	Total	2,39E+10	199			

R^2=0.330

Model 1, Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5827,044	4745,614		1,228	,221
	Z_S	-5579,819	1727,872	-,255	-3,229	,001
	Z_{A1}	15029,508	2923,769	,550	5,140	,000
	Z_{A2}	13447,851	3306,387	,492	4,067	,000
	Z_{A3}	8940,036	2597,097	,327	3,442	,001
	Z_{A1}	1277,989	2415,499	,047	,529	,597
	X_1	296,883	122,381	,383	2,426	,016
	X_2	-161,505	131,865	-,213	-1,225	,222

Model 2, ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
2	Regression	8,12E+09	7	1159618691	14,106	,000 [a]
	Residual	1,58E+10	192	82209376,40		
	Total	2,39E+10	199			

R^2=0.340

Model 2, Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
2	(Constant)	29030,103	3927,681		7,391	,000
	Z_S	-8294,657	1951,684	-,379	-4,250	,000
	Z_O	-32302,8	7343,610	-1,448	-4,399	,000
	X_1	11,976	152,599	,015	,078	,938
	X_2	-5,432	158,965	-,007	-,034	,973
	Z_SZ_O	4876,537	3330,212	,178	1,464	,145
	X_1Z_O	356,787	210,320	1,158	1,696	,091
	X_2Z_O	-89,449	215,010	-,227	-,416	,678

.

Model 3, ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
3	Regression	8,08E+09	5	1616384919	19,822	,000 [a]
	Residual	1,58E+10	194	81544363,47		
	Total	2,39E+10	199			

R^2=0.338

Model 3, Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
3	(Constant)	29038,632	3903,857		7,438	,000
	Z_S	-8329,975	1648,873	-,381	-5,052	,000
	Z_O	-31284,5	7141,100	-1,402	-4,381	,000
	X_1	7,130	56,139	,009	,127	,899
	Z_SZ_0	3764,510	2626,989	,138	1,433	,153
	X_1Z_O	282,743	97,761	,918	2,892	,004

Model 4, ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
4	Regression	8,08E+09	4	2020152271	24,899	,000[a]
	Residual	1,58E+10	195	81132933,46		
	Total	2,39E+10	199			

R^2=0.338

Model 4, Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
4	(Constant)	29511,850	1162,848		25,379	,000
	Z_S	-8333,183	1644,515	-,381	-5,067	,000
	Z_O	-31757,7	6076,761	-1,423	-5,226	,000
	Z_SZ_O	3767,719	2620,232	,138	1,438	,152
	X_1Z_O	289,873	79,834	,941	3,631	,000

Model 5, ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
5	Regression	7,91E+09	3	2637618146	32,334	,000[a]
	Residual	1,60E+10	196	81574880,97		
	Total	2,39E+10	199			

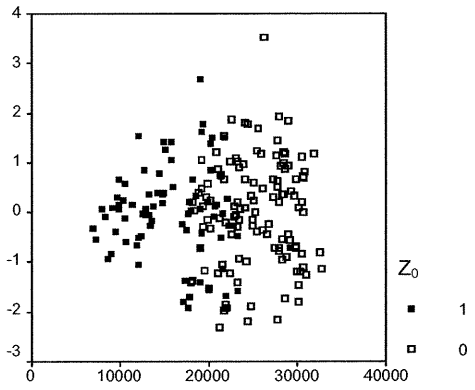
R^2=0.331

Model 5, Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
5	(Constant)	28769,780	1044,895		27,534	,000
	Z_S	-6849,044	1283,768	-,313	-5,335	,000
	Z_O	-28874,8	5752,111	-1,294	-5,020	,000
	X_1Z_O	275,708	79,439	,895	3,471	,001

Jackknife residuals

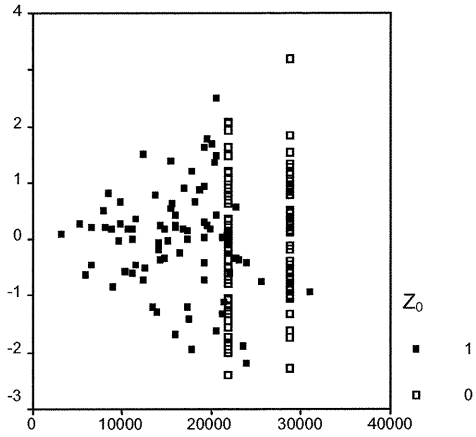
model 1



Unstandardized Predicted Value $\hat{Y}$

Jackknife residuals

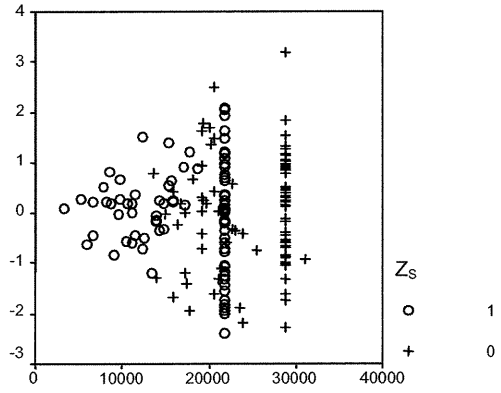
Model 5



Unstandardized Predicted Value $\hat{Y}$

Jackknife residuals

Model 5



Unstandardized Predicted Value $\hat{Y}$

Resultat etter variansanalyse med AMOUNT som avhengig variabel:

Tests of Between-Subjects Effects

Dependent Variable: AMOUNT

Source	Type III Sum of Squares [a]	df	Mean Square	F	Sig.
FEMALE	2661354629	1	2661354629		
AGEGROUP	4501415183	4	1125353796		
FEMALE * AGEGROUP	356214046	4	89053511,50	1,033	,392
Error	1,638E+10	190	86223932,90		
Corrected Total	2,390E+10	199			

a. R Squared = ,315 (Adjusted R Squared = ,282)

