

## Tentamen i Sambandsmodeller, 2008-06-11

Skrivtid: kl: 8-12

Hjälpmedel: Räknedosa. Läroboken: *Applied linear statistical models* av Kutner, Nachtsheim m fl som får innehålla anteckningar. Alla upplagor tillåtna. Tre formelblad

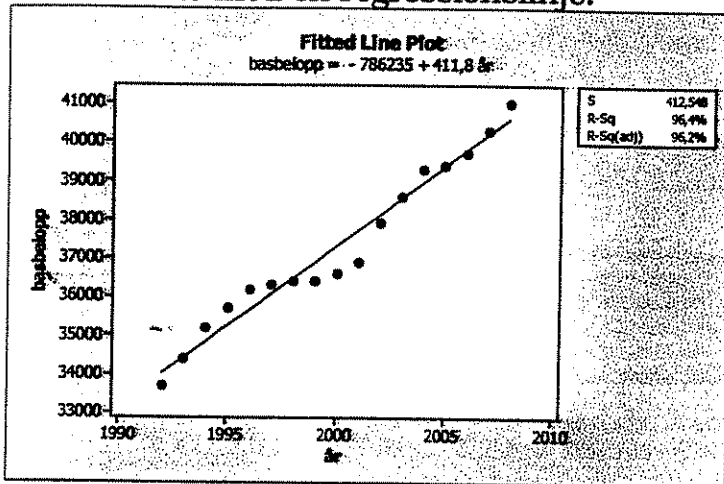
Jourhavande lärare: Lotta Hallberg, Olle Eriksson

Redovisa och motivera kort alla dina lösningar

Obs! Skriv namn och personnummer på varje papper du lämnar in

**1**

I grafen nedan ser vi basbeloppet i kr i Sverige mellan 1992 och 2008 tillsammans med en regressionslinje.



### Regression Analysis: basbelopp versus år

The regression equation is  
 $\text{basbelopp} = -786235 + 412 \text{ år}$

Predictor	Coef	SE Coef
Constant	-786235	40848
år	411,76	20,42

S = 412,548 R-Sq = \*

#### Analysis of Variance

Source	DF	SS	MS	F
Regression	*	*	69176471	*
Residual Error	*	*	*	*
Total	16	71729412		

#### Uppgifter:

- Hur mycket har basbeloppet ökat i snitt per år sedan 1992 enligt modellen ovan? 1p
- Beräkna ett 95% konfidensintervall för regressionskoefficienten i modellen ovan. Tolka intervallet. 1p
- Fyll i de rätta siffrorna som är markerade med \* i utskriften ovan. 2p

**2**

Man vill jämföra tre typer av järnvägsräls och se vilken typ som är minst benägen att få sprickbildningar. Antalet sprickor i rälsen beror antagligen på hur hårt trafikerad banan är så därför mäter man genomsnittligt antal tåg per dag för varje rälstyp. Studien varade i två år.

Nedan följer en utskrift där  
 Y= antalet sprickor i rälsen  
 X= genomsnittligt antal tåg per dag  
 Den kvalitativa variabeln 'rälstyp' är kodad med två dummyvariabler där  
 TYPB=1 om rälstyp B och 0 annars och  
 TYPC=1 om rälstyp C och 0 annars  
 Anta att Y är normalfördelad.

Model: MODEL1

		Model Crossproducts X'X X'Y Y'Y				
X'X	INTERCEP	X	TYPB	TYPC	Y	
INTERCEP	15	296.5	5	5	120	
X	296.5	6593.29	86.8	128.5	2653.1	
TYPB	5	86.8	5	0	29	
TYPC	5	128.5	0	5	51	
Y	120	2653.1	29	51	1104	

X'X Inverse, Parameter Estimates, and SSE						
	INTERCEP	X	TYPB	TYPC	Y	
INTERCEP	0.7668070767	-0.034901914	-0.160909857	0.1301721025	1.390973643	
X	-0.034901914	0.0021491326	-0.002407029	-0.020330794	0.406959751	
TYPB	-0.160909857	-0.002407029	0.4026958719	0.2227704898	-2.655794921	
TYPC	0.1301721025	-0.020330794	0.2227704898	0.5923293159	-1.649839245	
Y	1.390973643	0.406959751	-2.655794921	-1.649839245	18.538101542	

Dependent Variable: Y

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	125.46190	41.82063	24.815	0.0001
Error	11	18.53810	1.68528		
C Total	14	144.00000			

Root MSE	1.29818	R-square	0.8713
Dep Mean	8.00000	Adj R-sq	0.8362

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T	
INTERCEP	1	1.390974	1.13678764	1.224	0.2467	
X	1	0.406960	0.06018218	6.762	0.0001	
TYPB	1	-2.655795	0.82380586	-3.224	0.0081	
TYPC	1	-1.649839	0.99912057	-1.651	0.1269	

Variable	DF	Type I SS
INTERCEP	1	960.000000
X	1	107.877252
TYPB	1	12.989282
TYPC	1	4.595365

Till uppgifterna nedan ska du använda modellen som är anpassad ovan.

a) Prediktera antalet sprickor i rälsen för rälstyp B och X=10 samt ge ett 99% prediktionsintervall. Standardavvikelsen på prediktionen är skattad till 0,73. 2p

b) Pröva med ett test på 5% nivå om det är skillnad mellan rälstyperna. Om det är skillnad, vilken rälstyp är i så fall bäst? 2p

c) Korrelationen mellan skattningarna på parametrarna bör inte vara för stark. Skatta korrelationen mellan skattningarna på parametrarna för TYPB och TYPC. 2p

3. (5 poäng) I ett försök undersöker man hur råttor reagerar på en viss behandling. Man har råttor från 4 olika stammar och man har 3 olika behandlingar. Behandling 1 är kontroll medan behandling 2 och 3 är aktiva behandlingar. Den variabel man mäter har värdet 0 om ingen reaktion uppstår och ökande värde ju starkare reaktion som uppstår. Det tillgängliga datamaterialet visas i tabellen nedan. Det finns bara ett värde per cell. Sista raden och sista kolumnen visar marginalmedelvärden.

	beh 1	beh 2	beh 3	medelv
stam 1	12	18	21	17
stam 2	5	15	16	12
stam 3	10	16	16	14
stam 4	5	15	19	13
medelv	8	16	18	14

- Testa nollhypotesen att det inte finns någon skillnad mellan stammar på 5% risknivå.
  - Testa nollhypotesen att det inte finns någon skillnad mellan behandlingar på 5% risknivå.
  - Testa nollhypotesen att det inte finns någon skillnad mellan behandlingar på 5% risknivå med någon ickeparametrisk metod.
  - Jämför alla 3 behandlingar med varandra genom att beräkna konfidensintervall för skillanden mellan varje par av behandlingar med 95% gemensam konfidensnivå.
  - Man har valt att testa  $H_0: \mu_1 - 0.33\mu_2 - 0.67\mu_3 = 0$  för att kombinationen ser ut att vara tydligt skild från 0. Genomför testet med en metod som ger 5% risknivå att förkasta  $H_0$  om det inte finns några behandlingseffekter och man selekterar den kombination som är mest skild från 0.
4. (3 poäng) Data i uppgift 3 presenterades tyvärr på ett ofullständigt sätt, vilket du inte visste då. Egentligen finns det mer data att tillgå och siffrorna i uppgift 3 är medelvärden. Det riktiga materialet består av två observationer per cell och visas i tabellen nedan.

	beh 1	beh 2	beh 3
stam 1	15 9	21 15	21 21
stam 2	7 3	15 15	17 15
stam 3	9 11	17 15	13 19
stam 4	5 5	15 15	18 20

Här har inga medelvärdena skrivits ut eftersom de redan finns med i uppgift 3.

De behandlingarna som ingår i försöket också är de behandlingar man är intresserad av. De stammar som ingår i försöket är däremot slumpmässigt valda ur en stor population av stammar som man är intresserad av att yttra sig om. En analys av materialet i Minitab, där vissa delar ersatts med "\*\*", är:

Analysis of Variance for Y

Source	DF	SS	MS	F	P
stam	**	84.000	**	**	**
beh	**	448.000	**	**	**
stam*beh	**	40.000	**	**	**
Error	**	70.000	**		
Total	**	642.000			

- (a) Testa nollhypotesen att det inte finns någon skillnad mellan stammar på 5% risknivå.
  - (b) Testa nollhypotesen att det inte finns någon skillnad mellan behandlingar på 5% risknivå.
  - (c) Skatta  $\mu_2 - \mu_1$  och  $\mu_3 - \mu_1$  med konfidensintervall på gemensam 95% konfidensnivå.
5. (2 poäng) Beskrivningen av experimentet och faktorerna enligt uppgift 4 var korrekt, men nu får du veta att data inte presenterades på helt rätt sätt den gången heller. Några experiment genomfördes aldrig. De data som egentligen observerades visas med siffror i den följande tabellen, medan "\*" visar att man aldrig gjorde någon observation.

	beh 1	beh 2	beh 3
stam 1	15 9	21 15	21 *
stam 2	7 3	15 *	17 15
stam 3	9 11	17 15	13 19
stam 4	* 5	15 15	18 20

Medelvärden skrivs inte ut här. Om man låter varje cell ha samma vikt så får man de marginalmedelvärden som gavs i uppgift 3.

Antag att de observationer som föll bort gjorde det på ett slumpmässigt sätt. I uppgift 4 hade man ersatt saknade värden med det enda värde som kunde observeras i samma cell, men den metoden kan ifrågasättas. Du ska nu välja någon annan metod som använder de data som kunde observeras och inte kräver att man fyller ut till en balanserad försöksplan.

- (a) Testa nollhypotesen att det inte finns någon skillnad mellan stammar på 5% risknivå.
- (b) Testa nollhypotesen att det inte finns någon skillnad mellan behandlingar på 5% risknivå.