



Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2013-06-12
Sal (1) Om tentan går i flera salar ska du bifoga ett försättsblad till varje sal och <u>ringa in</u> vilken sal som avses	TER3
Tid	8-13
Kurskod	732G21
Provkod	TEN1
Kursnamn/benämning Provnamn/benämning	Sambandsmodeller Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	2
Jour/Kursansvarig Ange vem som besöker salen	Lotta Hallberg
Telefon under skrivtiden	
Besöker salen ca kl.	10
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	annelie.almquist@liu.se
Tillåtna hjälpmedel	Räknedosa. Läroboken: <i>Applied linear statistical models</i> av Kutner, Nachtsheim m fl som inte får innehålla anteckningar men får ha markeringar och flärpar. Flärpar får ha en liten anteckning.
Övrigt	G=12, VG=16
Vilken typ av papper ska	

TENTAMEN I SAMBANDSMODELLER, 2013-06-12

Skrivtid: kl: 8-13

Hjälpmedel: Räknedosa. Läroboken: *Applied linear statistical models* av Kutner, Nachtsheim m fl som inte får innehålla anteckningar men får ha markeringar och flärpar. Flärpar får ha en liten anteckning.

Jourhavande lärare: Lotta Hallberg.

Redovisa och motivera kort alla dina lösningar

Tolka (om möjligt) alla dina resultat!

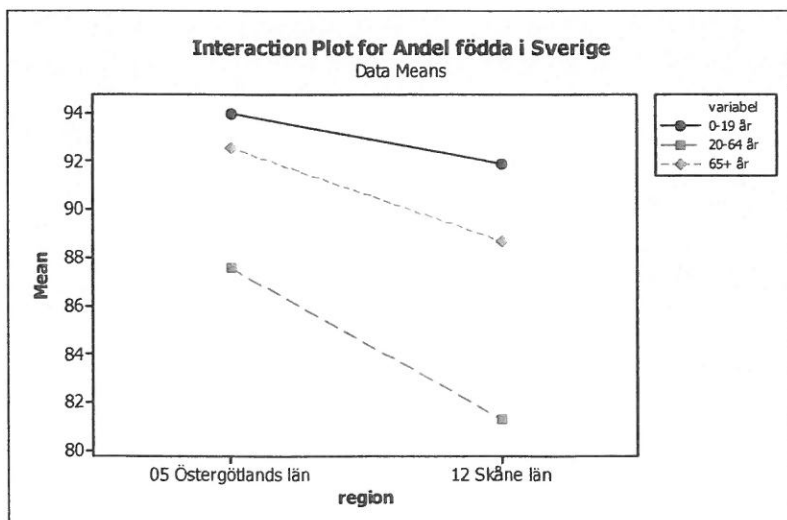
1

I denna uppgift ska andelen födda i Sverige i Sveriges befolkning studeras. Vi ska undersöka om det finns en skillnad i andelen födda i Sverige mellan Skånes län och Östergötlands län. Andra variabler som ska beaktas är ålderskategorier, och tid. Data är hämtade från SCBs hemsida.

Variabeldeklaration:

Variabelnamn	Anm	Värden
tid	år 1997 till 2011	1997; 1998; 1999; 2000; 2001; 2002; 2003; 2004; 2005; 2006; 2007; 2008; 2009; 2010; 2011
variabel	tre åldersgrupper	0-19 år; 20-64 år; 65+ år
region	länen	05 Östergötlands län; 12 Skåne län
0-19	dummy	1 om 0-19 år, 0 annars
20-24	dummy	1 om 20-24 år, 0 annars
Östergötland	dummy	1 om Östergötlands län, 0 om Skånes län
(0-19)*österg	interaktion	
(20-24)*österg	interaktion	
Andelen födda i Sverige		värden i procentenheter mellan 0 och 100

Följande graf kan vara till hjälp för att förstå data-materialet.



Först har två regressionsmodeller anpassats:

Modell 1

Regression Analysis: Andel födda i Sverige versus Östergötland

The regression equation is

$$\text{Andel födda i Sverige} = 87,3 + 4,08 \text{ Östergötland}$$

Predictor	Coef	SE Coef	T	P
Constant	87,3133	0,5987	145,84	0,000
Östergötland	*	*	*	*

$$S = 4,01620 \quad R\text{-Sq} = 20,9\% \quad R\text{-Sq(adj)} = 20,0\%$$

Andelen i summorna nedan är Andelen födda i Sverige

$$\begin{aligned} \sum \text{andelen} &= 8041,7 & \sum \text{Östergötland} &= 45 & \sum \text{Östergötland}^2 &= 45 \\ \sum \text{andelen}^2 &= 720337 & \sum \text{andelen} * \text{Östergötland} &= 4112,6 & & \end{aligned}$$

Modell 2

Regression Analysis: Andel födda i Sv versus Östergötland; 0-19; 20-24

The regression equation is

$$\text{Andel födda i Sverige} = 88,6 + 4,08 \text{ Östergötland} + 2,29 \text{ 0-19} - 6,19 \text{ 20-24}$$

Predictor	Coef	SE Coef	T	P
Constant	88,6144	0,3704	239,26	0,000
Östergötland	4,0778	0,3704	11,01	0,000
0-19	2,2867	0,4536	5,04	0,000
20-24	-6,1900	0,4536	-13,65	0,000

$$S = 1,75685 \quad R\text{-Sq} = 85,2\% \quad R\text{-Sq(adj)} = 84,7\%$$

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	1528,12	509,37	165,03	0,000
Residual Error	86	265,44	3,09		
Total	89	1793,56			

- I modell 1 har utskriften reducerats. Ta hjälp av summorna ovan och skatta regressionskoefficienten för Östergötland med minsta kvadratmetoden. 1p
- Pröva om det är skillnad i andelen födda i Sverige mellan Skånes län och Östergötlands län i modell 1. 2p
- Pröva med ett partiellt F- test om dummy-variablerna för åldersgrupper ska läggas till modellen. 1p

Nedan följer en modell utökad med två interaktionstermer samt tidsvariabeln.

Modell 3

Regression Analysis: Andel födda i Sverige versus 0-19; 20-24; ...

The regression equation is

$$\begin{aligned} \text{Andel födda i Sverige} &= 627 + 3,16 \text{ 0-19} - 7,40 \text{ 20-24} + 3,85 \text{ Östergötland} \\ &\quad - 1,75 \text{ int1} + 2,42 \text{ int2} - 0,269 \text{ tid} \end{aligned}$$

Predictor	Coef	SE Coef	T	P
Constant	627,42	47,51	13,21	0,000
0-19	3,1600	0,3548	8,91	0,000
20-24	-7,4000	0,3548	-20,86	0,000
Östergötland	3,8533	0,3548	10,86	0,000
(0-19)*österg	-1,7467	0,5018	-3,48	0,001
(20-24)*österg	2,4200	0,5018	4,82	0,000
tid	-0,26881	0,02371	-11,34	0,000

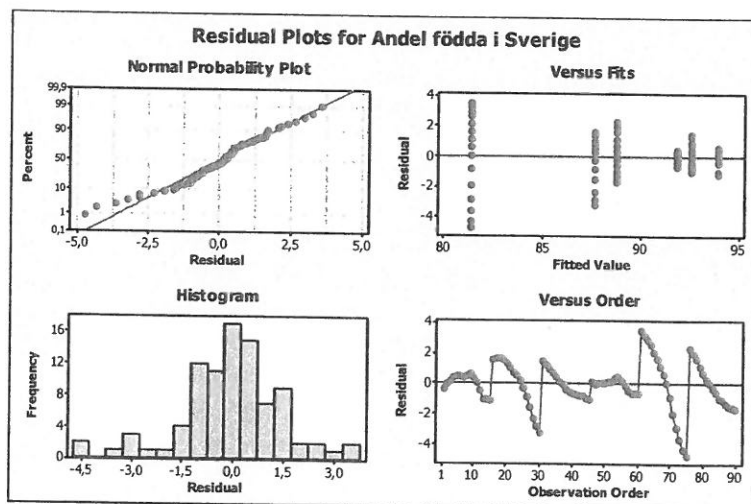
S = 0,971739 R-Sq = 95,6% R-Sq(adj) = 95,3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	6	1715,19	285,86	302,73	0,000
Residual Error	83	78,37	0,94		
Total	89	1793,56			

Frågorna nedan gäller modell 3.

- d) Hur många procentenheter sjunker andelen födda i Sverige varje år enligt modell 3? 1p
- e) Tolka regressionskoefficienten 3,85 vid variabeln Östergötland. 1p
- f) Prediktera andelen födda i Sverige i Skåne län i åldersgruppen 65+ år 2012. 1p
- g) Prediktera med ett 95% prediktionsintervall 'andelen födda i Sverige' i Östergötland i åldersgruppen 20-24 år 2012. Du får använda direkt att medelvärdet för prediktionen är 0,315. 2p
- h) De fyra graferna nedan gäller för residualerna från modell 3. Utför analys av residualerna. Är modellen godkänd? 1p
- i) Tolka förklaringsgraden. 1p



2

En tillverkare av frukostflingor ville undersöka effekten av en prisreduktion. Man planerade därför ett experiment: Responsvariabel i experimentet är Y = försäljningsvolym i hundratal. Man valde slumpmässigt 2 stycken städer bland väldigt många städer. Denna faktor kallas 'city' nedan. Den andra faktorn 'price' har två nivåer. Nivå 0 är ordinarie pris och nivå 1 är 10% prisreduktion. För var och en av de 4 kombinationerna (cellerna) drogs 3 affärer slumpmässigt.

Resultat:

	Price 0 Ordinarie pris	Price 1 10% prisreduktion
City 1	14 14 15	19 17 14
City 2	9 7 12	12 13 14

The GLM Procedure
Class Level Information

Class	Levels	Values
city	2	1 2
price	2	0 1

Number of Observations Read	12
Number of Observations Used	12

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	84.6666667	28.2222222	8.06	0.0084
Error	8	28.0000000	3.5000000		
Corrected Total	11	112.6666667			

R-Square	Coeff Var	Root MSE	Y Mean
0.751479	14.03122	1.870829	13.33333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
city	1	56.3333333	56.3333333	16.10	0.0039
price	1	27.0000000	27.0000000	7.71	0.0240
city*price	1	1.3333333	1.3333333	0.38	0.5543

Source	DF	Type III SS	Mean Square	F Value	Pr > F
city	1	56.3333333	56.3333333	16.10	0.0039
price	1	27.0000000	27.0000000	7.71	0.0240
city*price	1	1.3333333	1.3333333	0.38	0.5543

- Rita ett så kallat *Treatment means plot*. 1p
- Pröva om försäljningsvolymen har förändrats vid prisreduktionen. Visa hypoteser. Signifikansnivå 5%. 1p
- Pröva om det är skillnad i försäljningsvolym mellan alla städer. Visa hypoteser. Signifikansnivå 5%. 1p
- Skatta varianskomponenten för städer med ett 95% konfidensintervall. Använd Satterthwaites metod. 2p
- Skatta differensen i priseffekt med ett 95% konfidensintervall. 2p
- Testa om varianskomponenten, $\sigma_{\alpha\beta}^2$, för interaktionen är noll. Anta därefter att den är noll och testa återigen om försäljningsvolymen har förändrats vid prisreduktionen. Signifikansnivå 5%. 2p