



Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2010-03-27
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	TER1
Tid	8-12
Kurskod	732G05
Provkod	TENB
Kursnamn/benämning Provnamn/benämning	Regressions- och tidsserieanalys Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	3
Jour/Kursansvarig Ange vem som besöker salen	Lotta Hallberg
Telefon under skrivtiden	
Besöker salen ca kl.	10
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Carita Lilja, 1463, carli@ida.liu.se
Tillåtna hjälpmedel	Räknedosa. Bowerman, B.J., O'Connell, R, Koehler, A.: <i>Forecasting, Time Series and Regression</i> , alla upplagor tillåtna, som inte får innehålla anteckningar men får ha markeringar och flärpar. Formelsamling. Tabeller
Övrigt	G=12 VG=16
Vilken typ av papper ska användas, rutigt eller linjerat	

Antal exemplar i påsen	
-------------------------------	--

Tentamen i Regressions- och tidsserieanalys, 2010-03-27

Skrivtid: kl: 8-12

Hjälpmedel: Räknedosa. Bowerman, B.J., O'Connell, R, Koehler, A.: *Forecasting, Time Series and Regression*, alla upplagor tillåtna, som inte får innehålla anteckningar men får ha markeringar och flärpar. Formelsamling. Tabeller

Jourhavande lärare: Lotta Hallberg

Redovisa och motivera kort alla dina lösningar

1

Vid ett sjukhus vill en forskare bygga en modell för att kunna prediktera överlevnadstiden hos patienter som genomgått en viss leveroperation. Ett antal patienter slumpades ut och följande variabler mättes på varje patient då det gått en längre tid efter operation.

Y = överlevnadstiden, min värde = 181,0 max värde = 2343,0 i datamaterialet

lnY = naturliga logaritmen på överlevnadstiden

koag = blodkoaguleringsvärde

index = prognosindex

ensym = ensymfunktionsvärde

lever = leverfunktionsvärde

Ålder (ligger mellan 30 och 70år)

Kön = 0 för man och 1 för kvinna

Alk-mod = 1 om tidigare måttlig konsumtion av alkohol och 0 annars

Alk-hög = 1 om tidigare hög konsumtion av alkohol och 0 annars

De två sista variablerna är en dummykodning på en kvalitativ variabel 'alkoholkonsumtion' med nivåerna ingen, måttlig och hög konsumtion.

Nedan följer en mängd utskrifter. Uppgifterna a-f är insprängda mellan vissa utskrifter.

Först följer en korrelationsmatris

Correlations: lnY; koag; index; ensym; lever; Alder; Kön

	lnY	koag	index	ensym	lever	Ålder
koag	0,246					
index	0,470	0,090				
ensym	0,654	-0,150	-0,024			
lever	0,649	0,502	0,369	0,416		
Ålder	-0,145	-0,021	-0,048	-0,013	-0,207	
Kön	0,232	0,042	0,124	0,140	0,296	0,006

Eftersom ensym är den variabel som korrelerar starkast med lnY så anpassade forskaren en enkel linjär regressionsmodell.

Modell 1

The regression equation is

$$\ln Y = 5,26 + 0,0151 \text{ ensym}$$

Predictor	Coef
Constant	5,2643
ensym	0,015124

S = 0,3755 R-Sq =

Analysis of Variance

Source	DF	SS	MS
Regression	*	*	5,4762
Residual Error	*	*	*
Total	53	12,8077	

a) Pröva på 5% nivå om regressionskoefficienten (lutningsparametern) är 0.

2p

b) Beräkna och tolka förklaringsgraden.

1p

c) Bland de modeller som är numrerade 1 till 11 ska du välja den du tycker är bäst på ett vettigt sätt. Motivera noggrant ditt val.

2p

Modell 2

The regression equation is

$$\ln Y = 5,61 + 0,298 \text{ lever}$$

Predictor	Coef	SE Coef	T	P
Constant	5,6122	0,1425	39,38	0,000
lever	0,29819	0,04844	6,16	0,000

S = 0,3775 R-Sq = 42,2% R-Sq(adj) = 41,0%

Modell 3

The regression equation is

$$\ln Y = 4,35 + 0,0141 \text{ index} + 0,0154 \text{ ensym}$$

Predictor	Coef	SE Coef	T	P
Constant	4,3506	0,2144	20,30	0,000
index	0,014124	0,002364	5,98	0,000
ensym	0,015389	0,001880	8,19	0,000

S = 0,2908 R-Sq = 66,3% R-Sq(adj) = 65,0%

Modell 4

The regression equation is

$$\ln Y = 5,03 + 0,0107 \text{ ensym} + 0,209 \text{ lever}$$

Predictor	Coef	SE Coef	T	P
Constant	5,0282	0,1714	29,33	0,000
ensym	0,010731	0,002254	4,76	0,000
lever	0,20945	0,04477	4,68	0,000

S = 0,3171 R-Sq = 59,9% R-Sq(adj) = 58,4%

Modell 5

The regression equation is

$$\ln Y = 3,77 + 0,0955 \text{ koag} + 0,0133 \text{ index} + 0,0165 \text{ ensym}$$

Predictor	Coef	SE Coef	T	P
Constant	3,7662	0,2268	16,61	0,000
koag	0,09546	0,02169	4,40	0,000
index	0,013340	0,002035	6,56	0,000
ensym	0,016452	0,001630	10,09	0,000

S = 0,2493 R-Sq = 75,7% R-Sq(adj) = 74,3%

Modell 6

The regression equation is

$$\ln Y = 4,41 + 0,0110 \text{ index} + 0,0126 \text{ ensym} + 0,130 \text{ lever}$$

Predictor	Coef	SE Coef	T	P
Constant	4,4058	0,1990	22,14	0,000
index	0,011009	0,002404	4,58	0,000
ensym	0,012609	0,001955	6,45	0,000
lever	0,12977	0,04175	3,11	0,003

S = 0,2689 R-Sq = 71,8% R-Sq(adj) = 70,1%

Modell 7

The regression equation is

$$\ln Y = 4,03 + 0,0948 \text{ koag} + 0,0132 \text{ index} + 0,0164 \text{ ensym} - 0,00476 \text{ \AA}lder$$

Predictor	Coef	SE Coef	T	P
Constant	4,0281	0,2791	14,43	0,000
koag	0,09483	0,02139	4,43	0,000
index	0,013195	0,002008	6,57	0,000
ensym	0,016410	0,001607	10,21	0,000
\AA}lder	-0,004764	0,003040	-1,57	0,124

S = 0,2458 R-Sq = 76,9% R-Sq(adj) = 75,0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	9,8476	2,4619	40,75	0,000
Residual Error	49	2,9602	0,0604		
Total	53	12,8077			

Modell 8

The regression equation is

$$\ln Y = 3,78 + 0,0944 \text{ koag} + 0,0131 \text{ index} + 0,0162 \text{ ensym} + 0,0635 \text{ K\o}n$$

Predictor	Coef	SE Coef	T	P
Constant	3,7752	0,2273	16,61	0,000
koag	0,09438	0,02176	4,34	0,000
index	0,013108	0,002054	6,38	0,000
ensym	0,016225	0,001651	9,83	0,000
K\o}n	0,06348	0,06952	0,91	0,366

S = 0,2498 R-Sq = 76,1% R-Sq(adj) = 74,2%

Modell 9

The regression equation is

$$\ln Y = 4,04 + 0,0937 \text{ koag} + 0,0130 \text{ index} + 0,0162 \text{ ensym} - 0,00481 \text{ \AA}lder + 0,0651 \text{ K\o}n$$

Predictor	Coef	SE Coef	T	P
Constant	4,0397	0,2796	14,45	0,000
koag	0,09372	0,02144	4,37	0,000
index	0,012955	0,002026	6,40	0,000
ensym	0,016177	0,001627	9,94	0,000
\AA}lder	-0,004807	0,003043	-1,58	0,121
K\o}n	0,06511	0,06849	0,95	0,347

S = 0,2460 R-Sq = 77,3% R-Sq(adj) = 75,0%

Modell 10

The regression equation is

$$\ln Y = 4,05 + 0,0908 \text{ koag} + 0,0130 \text{ index} + 0,0161 \text{ ensym} + 0,0110 \text{ lever} - 0,00458 \text{ \AA}lder$$

Predictor	Coef	SE Coef	T	P
Constant	4,0474	0,2967	13,64	0,000
koag	0,09081	0,02896	3,14	0,003
index	0,012969	0,002300	5,64	0,000
ensym	0,016130	0,002107	7,66	0,000
lever	0,01104	0,05301	0,21	0,836
\AA}lder	-0,004579	0,003196	-1,43	0,158

S = 0,2482 R-Sq = 76,9% R-Sq(adj) = 74,5%

Modell 11

The regression equation is

$$\ln Y = 4,03 + 0,0948 \text{ koag} + 0,0130 \text{ index} + 0,0162 \text{ ensym} - 0,0030 \text{ lever} \\ - 0,00486 \text{ \u00c5lder} + 0,0662 \text{ K\u00f6n}$$

Predictor	Coef	SE Coef	T	P
Constant	4,0346	0,2975	13,56	0,000
koag	0,09480	0,02933	3,23	0,002
index	0,013013	0,002304	5,65	0,000
ensym	0,016249	0,002114	7,69	0,000
lever	-0,00302	0,05526	-0,05	0,957
\u00c5lder	-0,004859	0,003215	-1,51	0,137
K\u00f6n	0,06620	0,07202	0,92	0,363

S = 0,2486 R-Sq = 77,3% R-Sq(adj) = 74,4%

d) Forskaren tycker att modell 7 ovan \u00e4r bra s\u00e5 han vill d\u00e5 g\u00e5 vidare och pr\u00f6va om den kvalitativa variabeln alkoholkonsumtion skulle vara intressant att l\u00e4gga till modellen. Se modell 12 nedan. Denna \u00e4r ju kodad som tv\u00e5 dummies. Pr\u00f6va med ett l\u00e4mpligt test om denna variabel tillf\u00f6r signifikant information till modellen. 2p

e) Prediktera \u00f6verlevnadstiden med hj\u00e4lp av modell 12 nedan f\u00f6r en 50 \u00e5rig manlig patient med tidigare h\u00f6g alkoholkonsumtion. V\u00e4rderna p\u00e5 testerna \u00e4r ensym=80, index=60 och koag=7. 2p

Modell 12

The regression equation is

$$\ln Y = 4,02 + 0,0735 \text{ koag} + 0,0139 \text{ index} + 0,0155 \text{ ensym} - 0,00369 \text{ \u00c5lder} \\ + 0,0592 \text{ Alk-mod} + 0,379 \text{ Alk-h\u00f6g}$$

Predictor	Coef	SE Coef	T	P
Constant	4,0208	0,2391	16,82	0,000
koag	0,07352	0,01888	3,89	0,000
index	0,013896	0,001734	8,01	0,000
ensym	0,015488	0,001389	11,15	0,000
\u00c5lder	-0,003691	0,002629	-1,40	0,167
Alk-mod	0,05916	0,06751	0,88	0,385
Alk-h\u00f6g	0,37869	0,08836	4,29	0,000

S = 0,2098 R-Sq = 83,8% R-Sq(adj) = 81,8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	6	10,7386	1,7898	40,65	0,000
Residual Error	47	2,0691	0,0440		
Total	53	12,8077			

f) Forskaren har h\u00f6rt talas om VIF-v\u00e4rde och g\u00f6r d\u00e4rf\u00f6r f\u00f6ljande anpassning. Ber\u00e4kna med hj\u00e4lp av utskriften nedan VIF-v\u00e4rdet f\u00f6r variabeln lever och tolka resultatet. 2p

The regression equation is

$$\text{lever} = -1,68 + 0,358 \text{ koag} + 0,0192 \text{ index} + 0,0241 \text{ ensym} - 0,0170 \text{ \u00c5lder} \\ + 0,361 \text{ K\u00f6n}$$

Predictor	Coef	SE Coef	T	P
Constant	-1,6810	0,7382	-2,28	0,027
koag	0,35774	0,05659	6,32	0,000
index	0,019158	0,005347	3,58	0,001
ensym	0,024060	0,004295	5,60	0,000
\u00c5lder	-0,017004	0,008033	-2,12	0,039
K\u00f6n	0,3608	0,1808	2,00	0,052

S = 0,6494 R-Sq = 66,7% R-Sq(adj) = 63,2%

2

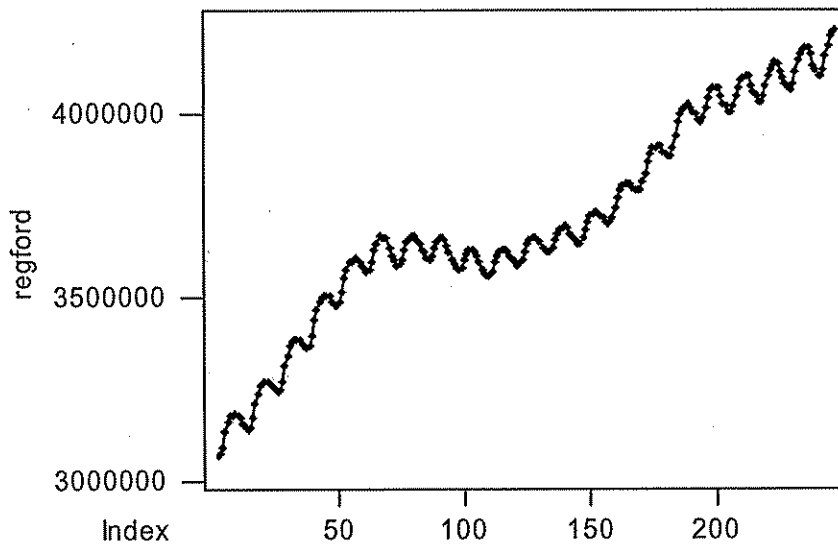
Betrakta följande tidsserie med kvartalsdata från tre år. Serien har alltså 12 termer.

30, 26, 24, 50, 36, 28, 32, 52, 38, 28, 30, 54

- a) Plotta tidsserien. 1p
b) Skatta de fyra säsongskomponenterna i en multiplikativ modell. Tolka en av komponenterna 4p

3

Tidsserieplotten nedan visar antalet registrerade personbilar per månad i Sverige från januari 1985 till och med augusti 2005:



a) Kommentera tidsseriens struktur utifrån bilden: Verkar det finnas någon trend? Hur är det med konjunktur-/cykliska svängningar? Säsongsmönster?

1p

b) Vi vill göra en komponentuppdelning, men vilken typ av modell verkar vara lämplig? Motivera ditt val av additiv eller multiplikativ modell. 1p

Här följer utskrift från en komponentuppdelning i Minitab enligt multiplikativ modell:

Time Series Decomposition

Data regford
Length 248,000
NMissing 0

Trend Line Equation

$$Y_t = 3234066 + 3761,80 * t$$

Seasonal Indices

Period	Index
1	0,989931
2	0,989461
3	0,992466

4	1,00000
5	1,00454
6	1,00768
7	1,00845
8	1,00872
9	1,00598
10	1,00213
11	0,997208
12	0,993432

Accuracy of Model

MAPE: 2
MAD: 63557
MSD: 5,68E+09

- c) Tolka i ord säsongseffekterna för period 4 (april) och period 9 (september).
1p
- d) Ge en rimlig prognos av antalet registrerade personbilar för nästa månad,
dvs september 2005.
1p