

Försättsblad till skriftlig tentamen vid Linköpings universitet



Datum för tentamen	2017-06-08
Sal (2)	G32(24) <u>TER2(64)</u>
Tid	14-18
Kurskod	732G05
Provkod	TENB
Kursnamn/benämning Provnamn/benämning	Regressions- och tidsserieanalys Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	4
Jour/Kursansvarig Ange vem som besöker salen	Annika Tillander
Telefon under skrivtiden	073-2129393
Besöker salen ca klockan	kl. 15, kl. 16 och kl. 17
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Anna Garabska Eklund ankn. 2362 anna.grabska.eklund@liu.se
Tillåtna hjälpmedel	Miniräknare Kursbok: Bowerman, O'Connel, Koehler: Forecasting, Time series, and Regression (alla upplagor tillåtna - får innehålla markeringar, understrykningar och flärpar, men inte anteckningar) A4-blad: med vad som helst skrivet på fram- och baksida (för hand/skrivet på dator/kopierat etc. - inga begränsningar).
Övrigt	
Antal exemplar i påsen	

Försättsblad till skriftlig tentamen vid Linköpings universitet



Datum för tentamen	2017-06-08
Sal (2)	G32(24) TER2(64)
Tid	14-18
Kurskod	732G05
Provkod	TENB
Kursnamn/benämning Provnamn/benämning	Regressions- och tidsserieanalys Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	4
Jour/Kursansvarig Ange vem som besöker salen	Annika Tillander
Telefon under skrivtiden	073-2129393
Besöker salen ca klockan	kl. 15, kl. 16 och kl. 17
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Anna Garabska Eklund ankn. 2362 anna.grabska.eklund@liu.se
Tillåtna hjälpmedel	Miniräknare Kursbok: Bowerman, O'Connel, Koehler: Forecasting, Time series, and Regression (alla upplagor tillåtna - får innehålla markeringar, understrykningar och flärpar, men inte anteckningar) A4-blad: med vad som helst skrivet på fram- och baksida (för hand/skrivet på dator/kopierat etc. - inga begränsningar).
Övrigt	
Antal exemplar i påsen	

Annika Tillander
IDA/Statistik

Tentamen Statistik, 732G05
2017-06-08

Skrivtid: 14.00-18.00

Tillåtna hjälpmedel:

Miniräknare

Kursbok Bowerman, O'Connel, Koehler: Forecasting, Time series, and Regression (alla upplagor tillåtna - får innehålla markeringar, understrykningar och flärpar, men inte anteckningar)

A4-blad med vad som helst skrivet på fram- och baksida (för hand/skrivet på dator/kopierat etc. - inga begränsningar).

Betyg: För godkänt betyg krävs 25 av 42 poäng. För väl godkänt betyg krävs 34 av 42 poäng.

Redovisa utförligt dina lösningar och tolka resultat.

Lösningsförslag till denna tentamen läggs upp på kursens LISAM-sida den 9/6.

Lycka till!

Uppgift 1 (5p)

Tabell 1. Visar priser och kvantitet för tre sportrelaterade produkter för de tre senaste åren.

Å	Träningskort		Träningskläder		Proteinpulver	
	<i>p</i>	<i>q</i>	<i>p</i>	<i>q</i>	<i>p</i>	<i>q</i>
2014	600	2	699	2.8	31.3	80
2015	650	2	749	3.0	31.7	120
2016	750	2	799	4.1	35.6	110

a) 4p
Vad blir ett kedjeindex för de tre produkterna i tabell 1 med 2014 som basår beräknat med Laspeyres?

b) 1p
Hur stor har prisutvecklingen varit från 2014 till de två efterkommande åren?

Uppgift 2 (8p)

a) 2p
Vilka är modellantagandena för linjär regression?

b) 4p
Visa hur modellantagandena kontrolleras grafiskt, dels hur det ser ut när antagandena är uppfyllda och dels när de inte är uppfyllda.

c) 2p
Resultat från en multipel regressionsanalys som undersöker vad som påverkar inköp av antikviteter där de förklarande variablerna är ålder, inkomst och utbildningslängd i år visas i Minitabutskriften nedan. Finns det något problem med analysen, i så fall hur visar det sig och hur kan det åtgärdas?

The regression equation is
 $\text{Expend. antiques} = 0,007 + 0,1046 \text{ Age} - 0,000026 \text{ Income} + 0,2037 \text{ Edu in years}$

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0,007	0,223	0,03	0,974	
Age	0,1046	0,0901	1,16	0,249	611,28
Income	-0,000026	0,000078	-0,34	0,736	655,59
Edu in years	0,2037	0,0864	2,36	0,020	33,63

S = 0,943990 R-sq = 92,16% R-sq(adj) = 91,91%

Analysis of Variance

Source	DF	SS	MS	F-Value	P-Value
Regression	3	1005,34	335,114	376,06	0,000
Error	96	85,55	0,891		
Total	99	1090,89			

Uppgift 3 (15p)

Följande modell skattar kostnad för byggande av kärnkraftverk i USA

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 D_1 + \beta_4 D_2 + \beta_5 x_1 * D_1 + \varepsilon$$

där y = kostnaden i milj. dollar i 1973 års pris, x_1 = tidpunkt, x_2 = nettoeffekten hos kraftverket i megawatt, $D_1 = 1$ indikerar användningen av kyltorn i anläggningen annars 0 och $D_2 = 1$ indikerar att kraftverket byggs i USAs nordöstra del annars 0. Nedan visas delar från Minitabutskriften baserat på 32 tidpunkters data.

Analysis of Variance

Source	DF	SS	MS	F-Value
Regression	A	704116	D	F
Error	B	193056	E	
Total	31	C		

S = **G** R-sq = 78,48% R-sq(adj) = **H**

Coefficients

Term	Coef	SE Coef	T-Value
Constant	-139.30678	76.27088	-1.826
x1	13.06970	2.41920	5.402
x2	0.39215	0.08413	4.661
D1	143.21901	69.08832	2.073
D2	105.85012	38.37323	2.758
x1*D1	-3.71434	3.68273	-1.009

a)

4p

Fyll i de saknade värdena som är ersatta med bokstäver i fetstil från **A-H**. Redovisa dina beräkningar.

b)

5p

Testa, på 5% signifikansnivå, om modellen som helhet är signifikant.

c)

6p

Gör prediktion med 95% prediktionsintervall för kostnaden att bygga ett kärnkraftverk vid tidpunkt 33 med en nettoeffekt på 1200 megawatt i nordöstra USA dels med kyltorn i anläggningen och dels utan kyltorn. Distance value med kyltorn är 0.735 och distance value utan kyltorn är 0.372.

Uppgift 4 (14p)

Nedanstående data visar antal personskador kvartalsvis relaterade till hoverboards i Region Östergötland

År	Kvartal	Antal personskador
2016	1	7
	2	10
	3	15
	4	21
2017	1	26

a)

5p

Skatta linjära regressionsmodellen för antal personskador beroende av tid, rita upp data med skattade regressionslinjen.

b)

5p

Tolka lutningskoefficienten (b_1) och testa den på 5% signifikansnivå.

c)

4p

Närliggande region har motsvarande data men från och med fjärde kvartalet 2014 till och med första kvartalet 2017. De är inte säkra på att en linjär modell är det som bäst beskriver sambandet utan testat tre olika modeller:

1. $y = \beta_0 + \beta_1 x + \varepsilon$

2. $y = \beta_0 \beta_1^x \varepsilon$

3. $y = \beta_0 x^{\beta_1} \varepsilon$

Nedan resultat för de olika modellerna men i annan ordning och med bokstavsbezeichnung.

A. The regression equation is

$$\text{Antal personskador} = -2.3144 + 2.4512 \text{ Tid}$$

$$S = 3,25 \quad R\text{-sq} = 85,44\% \quad R\text{-sq(adj)} = 83,62\%$$

B. The regression equation is

$$\log(\text{Antal personskador}) = 0.3574 + 0.8992 \log(\text{Tid})$$

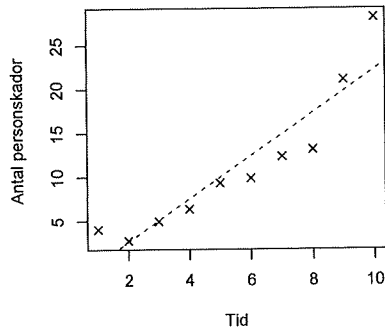
$$S = 0.1468 \quad R\text{-sq} = 81,05\% \quad R\text{-sq(adj)} = 78,68\%$$

C. The regression equation is

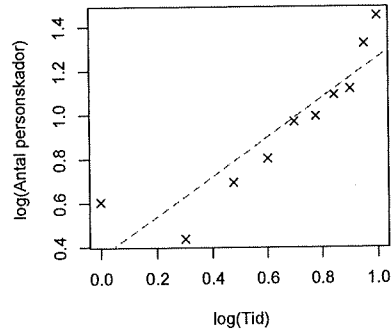
$$\log(\text{Antal personskador}) = 0.386368 + 0.101989 \text{ Tid}$$

$$S = 0.08047 \quad R\text{-sq} = 94,31\% \quad R\text{-sq}(\text{adj}) = 93,60\%$$

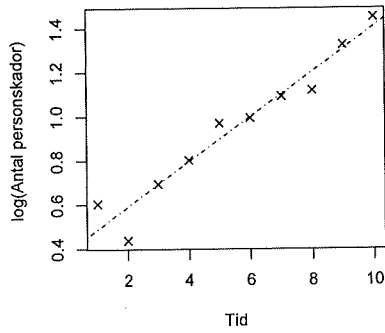
A: $y = a_0 + a_1 x$



B: $\log(y) = a_0 + a_1 \log(x)$



C: $\log(y) = a_0 + a_1 x$



För samtliga modeller tolka regressionskoefficienten. Vilket modellnummer hör ihop med vilken modellbokstav? Vilken modell är bäst och motivera ditt svar?