

Försättsblad till skriftlig tentamen vid Linköpings universitet



Datum för tentamen	2017-08-15
Sal (1)	TER2(20)
Tid	8-12
Kurskod	732G25
Provkod	TEN1
Kursnamn/benämning Provnamn/benämning	Statistisk analys av samhällsdata Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	5
Jour/Kursansvarig Ange vem som besöker salen	Lotta Hallberg
Telefon under skrivtiden	013-281657
Besöker salen ca klockan	10
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Annelie Almquist
Tillåtna hjälpmedel	Räknedosa. Ett handskrivet A4 blad med egna anteckningar (båda sidor)
Övrigt	
Antal exemplar i påsen	

TENTAMEN I STATISTISK ANALYS AV SAMHÄLLSDATA, 2017-08-15

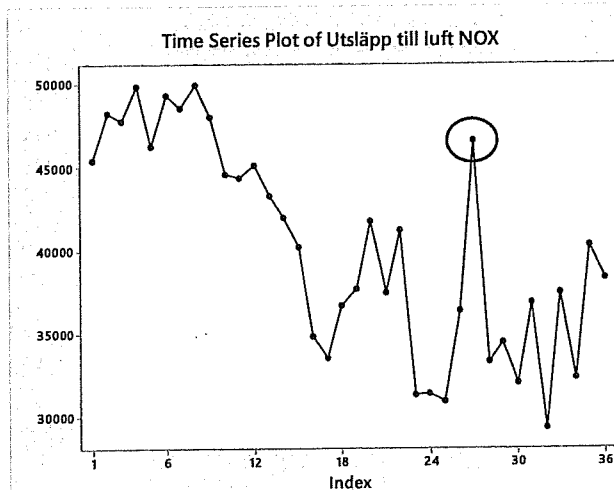
Skrivtid: kl: 8-12
Hjälpmedel: Räknedosa. Ett handskrivet A4 blad med egna anteckningar (båda sidor).
Jourhavande lärare: Lotta Hallberg
Betyg: För godkänt krävs minst 12 av 20 poäng och för väl godkänt krävs minst 16 av 20 poäng.

Redovisa och motivera kort alla dina lösningar
Tolka (om möjligt) alla dina resultat

Följande data har hämtats från SCBs hemsida.

Utsläpp till luft redovisat efter näringsgren SNI 2007 och ämne. Kvartal 2008K1 - 2016K4
Näringsgren: Transport och magasinering
Ämne: NOX kväveoxider, ton

Variabeln som ska analyseras kallas Utsläpp till luft NOX
Graf över data:



1

I denna uppgift ska variabeln Utsläpp till luft NOX analyseras.

Nedan ses utskrifter från fyra olika regressionsmodeller:

Förklarande variabler:

Tid: en numrering av tidpunkterna 1,2,...,36

Kvartal: Numrering av kvartal 1,2,3,4,1,2,3,...,3,4

Kvartal_i är indikator för kvartal i. *Kvartal_i* är 1 om kvartal i och 0 annars.

Indikator är en indikator för obs nr 27 så indikator är 1 om obs nr 27 och 0 annars. I grafen ovan ses ett extremt värde för obs nr 27 inringad med rött

Modell 1

Regression Analysis: Utsläpp till luft NOX versus tid; kvartal

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	827345680	413672840	23,01	0,000
Error	33	593253204	17977370		
Total	35	1420598884			

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	47531	2048	23,21	0,000	
tid	-464,2	68,4	-6,78	0,000	1,01
kvartal	489	636	0,77	0,447	1,01

Modell 2

Regression Analysis: Utsläpp till luft NOX versus tid; kvartal_1; kvartal_2; kvartal_3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	853047136	213261784	11,65	0,000
Error	31	567551748	18308121		
Total	35	1420598884			

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	48571	1985	24,47	0,000	
tid	-464,2	69,0	-6,72	0,000	1,01
kvartal_1	-1068	2028	-0,53	0,602	1,52
kvartal_2	55	2022	0,03	0,978	1,51
kvartal_3	1744	2018	0,86	0,394	1,50

Modell 3

Regression Analysis: Utsläpp till luf versus tid; kvartal_1; kvartal_2; kvartal_3; indikator

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	942048820	188409764	11,81	0,000
Error	30	478550064	15951669		
Total	35	1420598884			

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	48992	1862	26,32	0,000	
tid	-485,2	65,1	-7,46	0,000	1,03
kvartal_1	-1131	1893	-0,60	0,555	1,52
kvartal_2	13	1887	0,01	0,994	1,51
kvartal_3	601	1945	0,31	0,760	1,60
indikator	10101	4277	2,36	0,025	1,11

Modell 4

Regression Analysis: Utsläpp till luft NOX versus tid; indikator

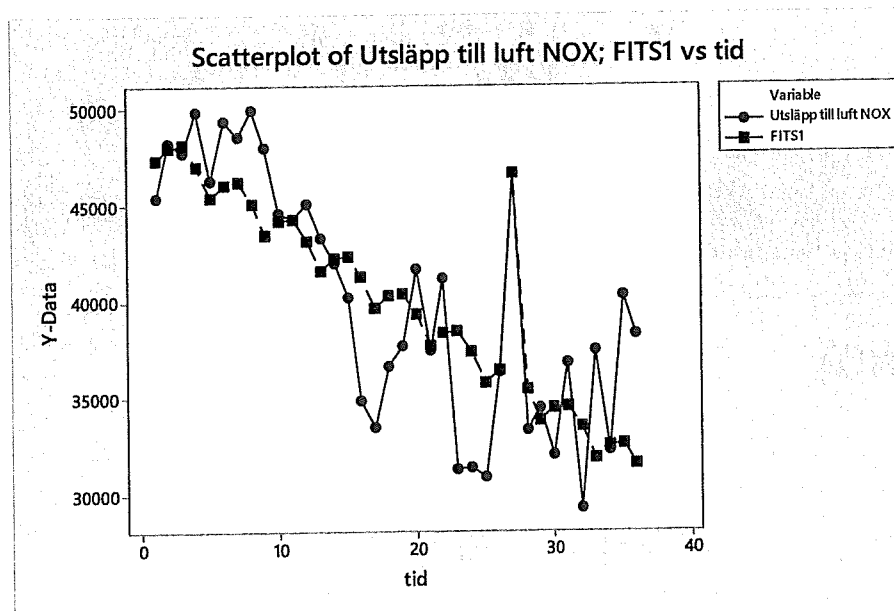
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	928462466	464231233	31,13	0,000
Error	33	492136418	14913225		
Total	35	1420598884			

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	48787	1316	37,09	0,000	
tid	-482,2	62,6	-7,71	0,000	1,02
indikator	10826	3955	2,74	0,010	1,02

- Modell 1 har ett allvarligt fel. Förklara vad som är fel och varför det är fel. 1p
- Vilken av modellerna 2, 3 eller 4 anser du vara bäst. Motivera noga. 1p
- Tolka regressionskoefficienten för kvartal_1 i modell 3. 1p
- Hur många ton NOX minskar utsläpp till luft varje år enligt modell 4? 1p
- Hur många fler ton utsläpp av NOX var det vid tidpunkt 27 jämfört med trenden enligt modell 3? Grafen nedan är kanske till hjälp. 1p
- Beräkna en prognos för tidpunkt 37 med hjälp av modell 2. 1p
- Vilket värde skulle VIF ha om alla förklarande variabler var okorrelerade? 1p

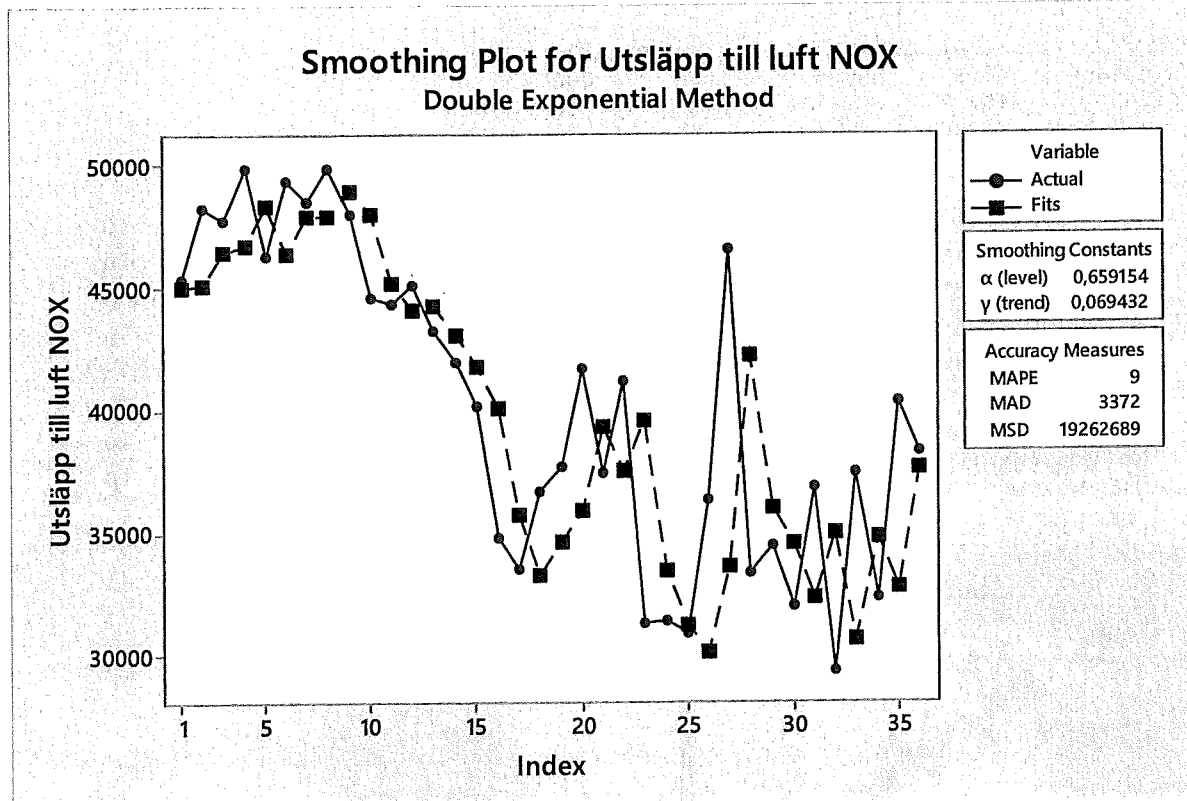


Denna graf visar anpassade värden (rött) för modell 3.

2

Även i denna uppgift ska variabeln Utsläpp till luft NOX analyseras.
Följande modell har anpassats

Double Exponential Smoothing for Utsläpp till luft NOX



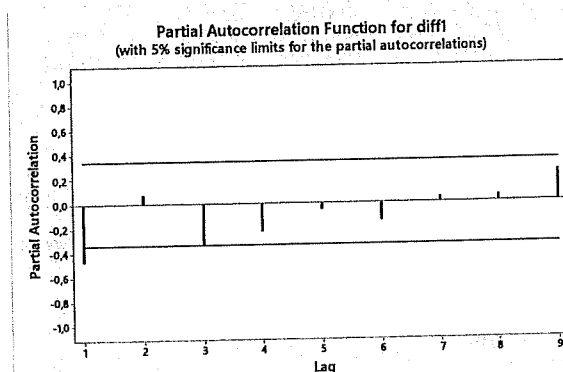
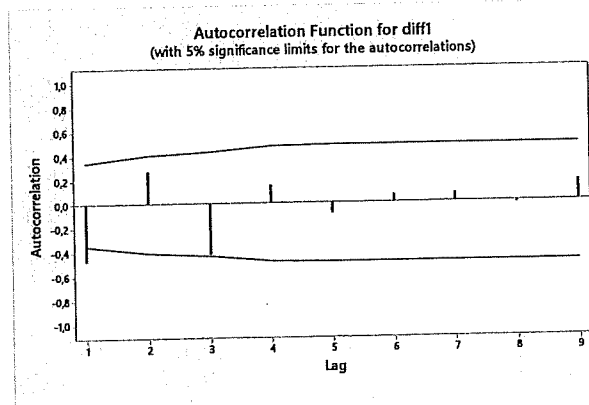
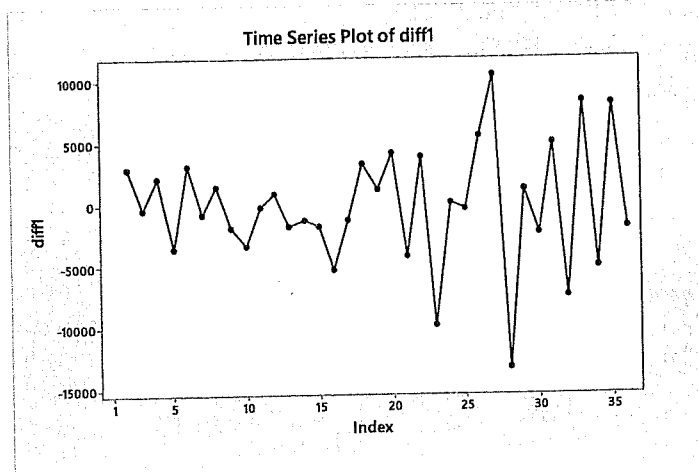
Utsläpp till luft NOX	l_T	b_T	T
36826,0	35269,9	-358,456	31
29267,8	31191,4	-616,742	32
37440,5	35100,3	-302,518	33
32257,8	33123,6	-418,763	34
40283,0	?	?	35
38291,1	?	?	36

Beräkna l_{35} , b_{35} , l_{36} och b_{36} samt beräkna en prognos NOX utsläpp för tidpunkt 37. 3p
Det räcker att du använder två decimaler på utjämningskonstanterna.

3

Även i denna uppgift ska variabeln Utsläpp till luft NOx analyseras.

En ARIMA-modell ska anpassas. Eftersom tidsserien inte ser stationär ut så har den differentierats. Den differentierade serien kallas här *diff1*



ARIMA Model: Utsläpp till luft NOx

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0,3030	0,1955	1,55	0,131
MA 1	0,9409	0,1065	8,83	0,000
Constant	-299,03	70,57	-4,24	0,000

Differencing: 1 regular difference

Number of observations: Original series 36, after differencing 35

Residuals: SS = 566219672 (backforecasts excluded)
MS = 17694365 DF = 32

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	15,5	20,9	*	*
DF	9	21	*	*
P-Value	0,078	0,467	*	*

Resid	tid
5675,94	33
-2019,44	34
7994,37	35
3397,35	36

I uppgift 2 finner du även observationerna för tidpunkterna 33-36.

- Beräkna prognos för tidpunkt nr 37 och 38 med hjälp utav modellen ovan. 3p
- Sätt upp rätt hypoteser för testet Modified Box-Pierce vid lag 12 1p

4

För en viss djurart och ett visst år finns följande dödsrisker.

Ålder år	Dödsrisk
0	0.30
1	0.06
2	0.11
3	0.22
4	0.35
5	0.61
6	0.76
7	0.90
8+	1.00

Konstruera en livslängdstabell med $l_0 = 10000$. Dvs beräkna e_x (Ange en decimal). Sätt $L_{8+} = 10$.
Använd formlerna nedan. Redovisa alla steg i uträkningen. 3p

Formler :

$$e_x = \frac{T_x}{l_x} \quad T_x = T_{x+n} + {}_nL_x \quad {}_nL_x = \frac{n(l_x + l_{x+n})}{2}$$

5

I tabellen nedan visas totalförsäljningen i löpande priser för ett postorderföretag som säljer rid- och hästartiklar, samt priser för varan ridbyxor från varugrupp ridartiklar och varan regntäcke från varugrupp hästartiklar under åren 2014-2016

År	Total försäljning Ridartiklar milj kr	Total försäljning Hästartiklar milj kr	Pris, ridbyxor kr	Pris, regntäcke kr
2014	127.0	161.0	950	1100
2015	127.3	162.5	980	1170
2016	133.0	169.8	1010	1225

Beräkna en sammansatt-prisindexserie (fastbasindex) med basår 2014 för postorderföretaget för tidpunkterna 2014-2016. Använd ridbyxor och regntäcke som representantvaror för respektive varugrupp. Använd Laspeyre's viktsystem. 3p