



Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2012-10-20
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	G36
Tid	8-12
Kurskod	732G25
Provkod	TEN1
Kursnamn/benämning Provnamn/benämning	Statistisk analys av samhällsdata Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	4
Jour/Kursansvarig Ange vem som besöker salen	Lotta Hallberg
Telefon under skrivtiden	
Besöker salen ca kl.	10
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	carita.lilja@liu.se tel:1463
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. Anteckning på flärp tillåten.
Övrigt	G=12p, VG=16p
Vilken typ av papper ska användas, rutigt eller linjerat	Rutigt
Antal exemplar i påsen	

TENTAMEN I STATISTISK ANALYS AV SAMHÄLLSDATA, 2012-10-20

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. Anteckning på flärp tillåten.

Jourhavande lärare: Lotta Hallberg.

Redovisa och motivera kort alla dina lösningar

Tolka (om möjligt) alla dina resultat!

1

Nedan ses data hämtade från SCBs hemsida. Data är för Sverige, 2001, hela riket, antalet levande födda barn uppdelat på moderns ålder. Även antalet kvinnor i de givna åldersintervallen är också givna.

moderns ålder	tid	Levande födda
14 år	2001	3
15 år	2001	16
16 år	2001	60
17 år	2001	146
18 år	2001	312
19 år	2001	655
20 år	2001	1070
21 år	2001	1554
22 år	2001	2072
23 år	2001	2514
24 år	2001	3168
25 år	2001	3920
26 år	2001	5014
27 år	2001	5980
28 år	2001	6433
29 år	2001	7053
30 år	2001	7294
31 år	2001	7090
32 år	2001	6406
33 år	2001	6052
34 år	2001	5501
35 år	2001	4875
36 år	2001	4012
37 år	2001	3296
38 år	2001	2201
39 år	2001	1678
40 år	2001	1167
41 år	2001	820
42 år	2001	504
43 år	2001	317
44 år	2001	143
45 år	2001	79
46 år	2001	32
47 år	2001	15
48 år	2001	6
49 år	2001	8

moderns ålder_	totala antal kvinnor
14 år	54530
15 år	53064
16 år	51455
17 år	49387
18 år	48409

19 år	49007
20 år	49649
21 år	51266
22 år	50769
23 år	49454
24 år	50866
25 år	52449
26 år	55418
27 år	58554
28 år	58340
29 år	59554
30 år	60135
31 år	58859
32 år	58045
33 år	61301
34 år	64450
35 år	65301
36 år	65572
37 år	66064
38 år	61110
39 år	58702
40 år	57123
41 år	56698
42 år	57017
43 år	57014
44 år	58320
45 år	58053
46 år	57888
47 år	56698
48 år	58076
49 år	57929

Total medelfolkmängd år 2001 = 8897013

Totala antalet kvinnor år 2001 = 4495900

Totala antalet kvinnor mellan 14 och 49 år = 2036526

Totala antalet levande födda = 91466

- Bestäm födelsetalet (Crude birth rate) CBR samt den generella fruktsamhetskvoten GBR och redogör för ev för- och nackdelar med de båda måtten. 2p
- Bestäm den åldersspecifika fruktsamheten för åldersgruppen 25-29år 1p
- Bestäm den summerade fruktsamheten (Total fertility rate) TFR och tolka resultatet. 2p

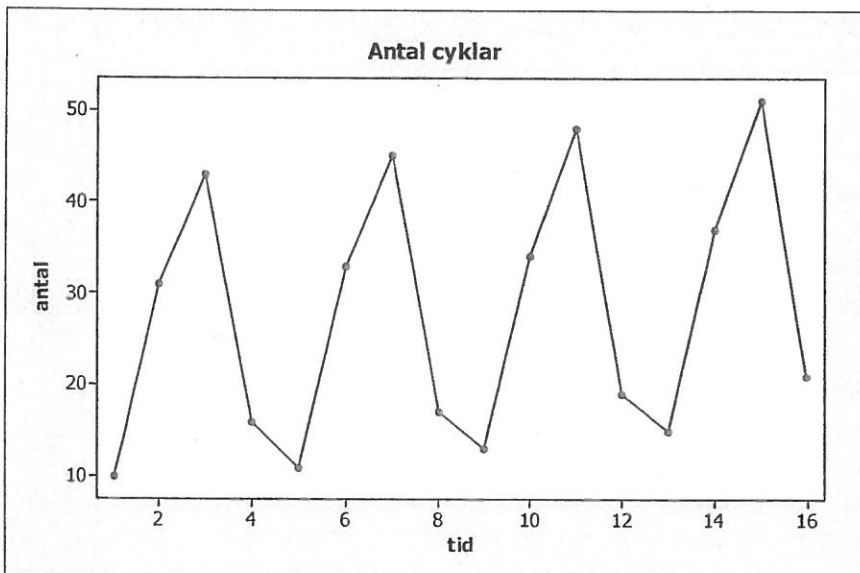
Du får i uppgift c) använda avrundade 'asfr' i promille som är givna nedan om du kan beskriva hur de är beräknade. Data ska läsas radvis.

0,1	0,3	1,2	3	6,4	13,4	21,6
30,3	40,8	50,8	62,3	74,7	90,5	102,1
110,3	118,4	121,3	120,5	110,4	98,7	85,4
74,7	61,2	49,9	36	28,6	20,4	14,5
8,8	5,6	2,5	1,4	0,6	0,3	0,1
0,1						

2

I figuren nedan visas försäljningen av antal cyklar i en viss affär. Det är kvartalsdata över fyra år. Två modeller är anpassade. Y = antal cyklar.

- Prediktera antal sålda cyklar för första kvartalet år fem med båda modellerna. 2p
- Tolka för båda modellerna den skattade säsongeffekten för kvartal 3. 1p
- Avgör om residualerna kan antas vara okorrelerade över tiden i modell 1. 1p



Modell 1

Regression Analysis: y versus tid; kv_1; kv_2; kv_3

The regression equation is

$$y = 13,3 + 0,500 \text{ tid} - 4,50 \text{ kv}_1 + 16,5 \text{ kv}_2 + 29,0 \text{ kv}_3$$

Predictor	Coef	SE Coef	T	P
Constant	13,2500	0,5056	26,20	0,000
tid	0,50000	0,03769	13,27	0,000
kv_1	-4,5000	0,4900	-9,18	0,000
kv_2	16,5000	0,4827	34,19	0,000
kv_3	29,0000	0,4782	60,64	0,000

S = 0,674200 R-Sq = 99,8% R-Sq(adj) = 99,8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	2990,00	747,50	1644,50	0,000
Residual Error	11	5,00	0,45		
Total	15	2995,00			

Durbin-Watson statistic = 2,2

Modell 2 Time Series Decomposition for y

Additive Model

Data y
Length 16
NMissing 0

Fitted Trend Equation

$$Y_t = 23,531 + 0,496 * t$$

Seasonal Indices

Period	Index
1	-14,8438
2	6,4063
3	18,4063
4	-9,9688

Accuracy Measures

MAPE 2,04255
MAD 0,46967
MSD 0,36983

3

Försäljningen av gasolgrillar antas ha haft en så kraftig uppgång under perioden 2001-2004 att man kan tala om en exponentiellt ökande trend. Det finns dock en kraftig säsongvariation över året. Under tre år har säsongkomponenterna skattats till 37% över trend-nivån under andra och tredje kvartalet, och c:a 37% under trend-nivån under första och fjärde kvartalet. Originaldata förutsätts vara fria från konjunkturvariation och presenteras nedan. Säsongrensa först data, skatta och tolka sedan tillväxtfaktorn i försäljningsutvecklingen på lämpligt sätt. (tillväxtfaktorn = β_1 i growth curve model)

Tidsperiod.	Försäljning.(tkr)
apr 2001–sep 2001	9759.7
okt 2001–mar 2002	7861.5
apr 2002–sep 2002	12452.4
okt 2002–mar 2003	7056.9
apr 2003–sep 2003	17874.2
okt 2003–mar 2004	12421.6

4p

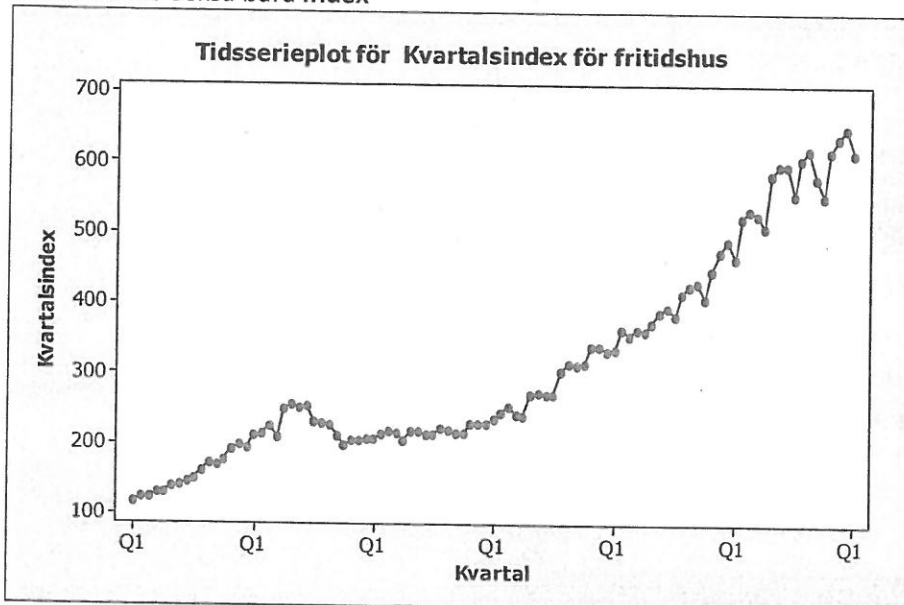
4

I grafen nedan visas

Fastighetsprisindex för fritidshus. Kvartal 1986K1-2010K1

(Q1 som syns i vissa grafer nedan betyder kvartal 1)

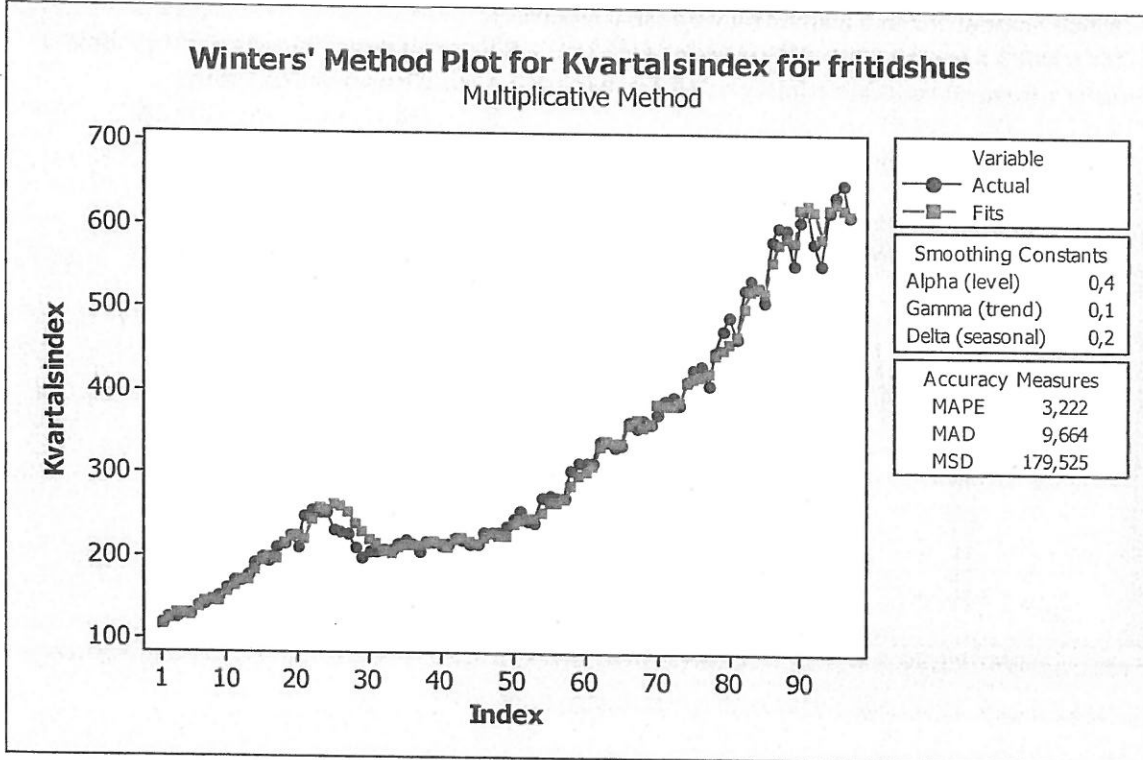
Serien kallas också bara **index**



Nedan har två typer av modeller anpassats och från modell 1 och 3 har prognoser beräknats.

- a) Studera modell 1. Visa hur modellen ser ut. Förklara hur parametrarna skattas samt hur man går tillväga för att hitta bäst anpassning. Är du nöjd med modellen. Motivera. 2p

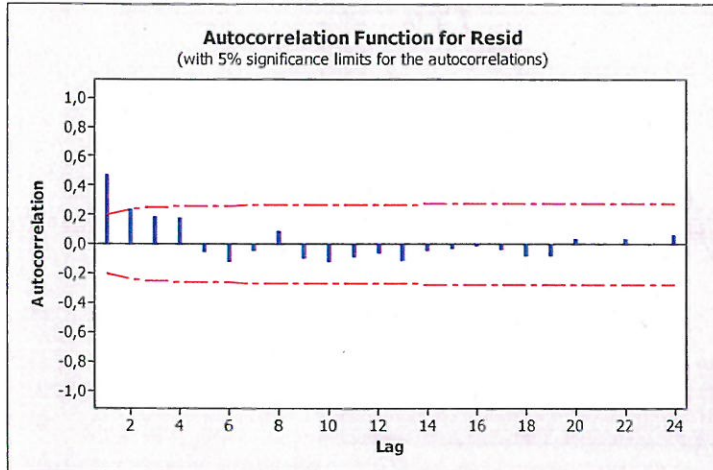
Modell 1



Forecasts

Period	Forecast	Lower	Upper
98	671,366	647,691	695,041
99	682,891	657,686	708,095
100	673,571	646,553	700,588
101	650,058	620,998	679,119

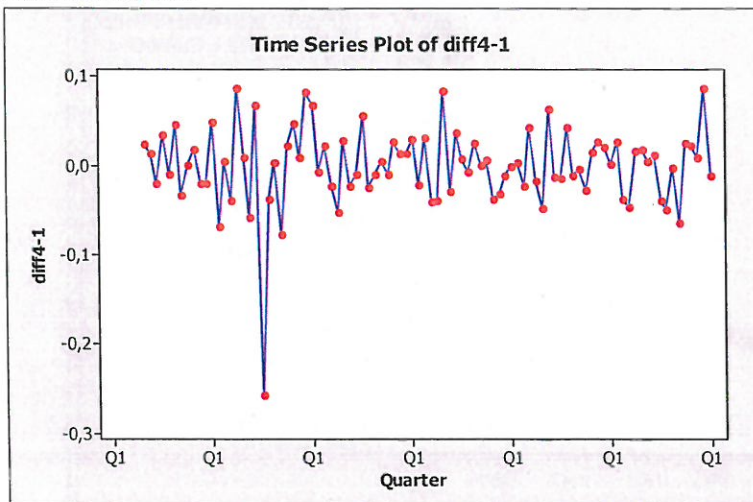
SAC för residualer från modell 1.

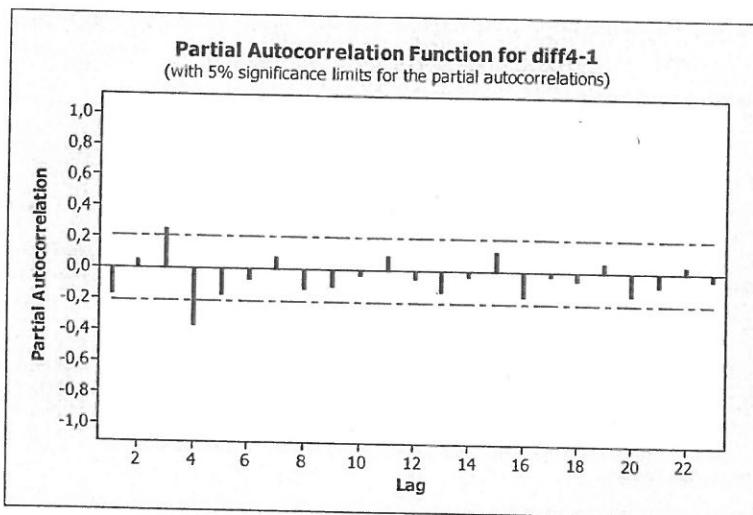
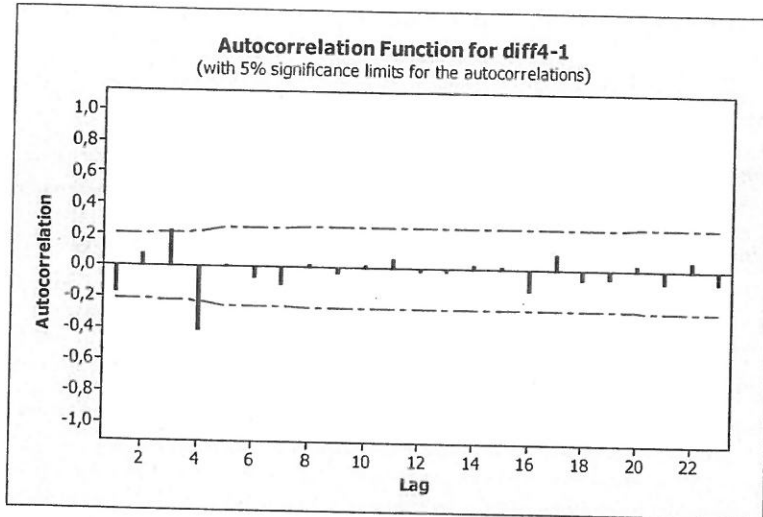


Nedan har två säsong ARIMA-modeller anpassats. Först har data logaritmerats med den naturliga logaritmen. Denna serie heter **ln index**. Sedan har **ln index** differentierats för säsong och för trend. Denna serie heter **diff4-1**

- b) Förklara varför serien har logaritmerats. 1p
- c) Vilken av modell 2 och 3 anser du vara bäst. Motivera. 1p
- d) För modell 3 har prognoser för **ln index** beräknats. Räkna om dessa till prognoser för index. Jämför prognosintervallen från modell 1 och modell 3. Vilka föredrar du? Motiera. 1p
- e) Låt $\text{diff4-1} = z_t$. Visa hur modell 3 ser ut för z_t . 2p

Modell 2 och 3





Modell 2 ARIMA Model: In index

Estimates at each iteration

Iteration	SSE		Parameters			
0	1,10237	0,100	0,100	0,100	0,100	0,081
1	0,28629	0,088	-0,043	0,112	0,243	-0,027
2	0,27429	-0,062	-0,034	-0,036	0,261	-0,029
3	0,26170	-0,212	-0,024	-0,183	0,282	-0,031
4	0,19901	-0,362	0,049	-0,316	0,413	-0,020
5	0,17517	-0,226	0,085	-0,166	0,481	-0,013
17	0,12538	-0,309	0,240	-0,171	0,883	-0,000
18	0,12537	-0,343	0,239	-0,207	0,884	-0,000
19	0,12536	-0,339	0,243	-0,202	0,886	-0,000
20	0,12536	-0,339	0,243	-0,202	0,886	-0,000

Relative change in each estimate less than 0,0010

Final Estimates of Parameters

Type	Coef	SE Coef
AR 1	-0,3391	0,7272
SAR 4	0,2429	0,1517
MA 1	-0,2025	0,7617

SMA 4 0,8857 0,0914
Constant -0,0001336 0,0006898

Differencing: 1 regular, 1 seasonal of order 4
Number of observations: Original series 97, after differencing 92
Residuals: SS = 0,122825 (backforecasts excluded)
 MS = 0,001412 DF = 87

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

Lag	12	24	36	48
Chi-Square	6,0	10,2	21,9	33,0
DF	7	19	31	43
P-Value	0,541	0,947	0,885	0,865

Modell 3 ARIMA Model: In index

Estimates at each iteration

Iteration	SSE	Parameters			
0	1,10237	0,100	0,100	0,090	
1	0,26502	-0,046	0,246	-0,027	
2	0,18848	0,028	0,396	-0,014	
3	0,15119	0,101	0,546	-0,006	
4	0,13435	0,169	0,696	-0,002	
5	0,12828	0,224	0,846	-0,000	
6	0,12799	0,216	0,868	-0,000	
7	0,12796	0,220	0,875	-0,000	
8	0,12795	0,224	0,879	-0,000	
9	0,12795	0,228	0,882	-0,000	
10	0,12795	0,230	0,884	-0,000	
11	0,12795	0,232	0,886	-0,000	

Unable to reduce sum of squares any further

Final Estimates of Parameters

Type	Coef	SE Coef
SAR 4	0,2317	0,1403
SMA 4	0,8855	0,0867
Constant	-0,0001128	0,0005739

Differencing: 1 regular, 1 seasonal of order 4
Number of observations: Original series 97, after differencing 92
Residuals: SS = 0,125022 (backforecasts excluded)
 MS = 0,001405 DF = 89

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

Lag	12	24	36	48
Chi-Square	7,2	11,3	23,8	36,0
DF	9	21	33	45
P-Value	0,620	0,956	0,880	0,830

Forecasts from period 97

Period	Forecast	95% Limits		Actual
		Lower	Upper	
98	6,51030	6,43683	6,58378	
99	6,53415	6,43024	6,63806	
100	6,53354	6,40628	6,66081	
101	6,49809	6,35114	6,64504	