



Försättsblad till skriftlig tentamen vid Linköpings Universitet

10A
Aven

Datum för tentamen	2010-10-16
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	732G25
Provkod	TENT
Kursnamn/benämning Provnamn/benämning	Statistisk analys av sammansatta Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	5
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, carita.lilja@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.
Övrigt	G=12, VG=16
Vilken typ av papper ska användas, rutigt eller linjerat	
Antal exemplar i påsen	18

Tentamen i Statistisk analys av samhällsdata, 2010-10-16

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.

Jourhavande lärare: Lotta Hallberg

Redovisa och motivera kort alla dina lösningar

1

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

Ålder	Dödsrisk
0	0.35
1	0.05
2	0.10
3	0.20
4	0.35
5	0.60
6	0.75
7	0.85
8+	1.00

- a) 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
- b) Hur stor är den återstående livslängden för ett nyfött djur? 0.5p
- c) Beräkna den sannolika livslängden (medianåldern) 1.5p

Formler :

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

2

I nedanstående tabell ges antalet insjuknade per vecka.

Antal insjuknade y	6	7	9	12	14	19	20	24	30
Veckonummer t	1	2	3	4	5	6	7	8	9

Skatta tillväxtfaktorn och tolka den i de två modellerna:

$$y = \beta_0' + \beta_1' t + \varepsilon$$

$$y = \beta_0 \cdot \beta_1^t \cdot \delta$$

Dvs skatta och tolka β_1 och β_1' . ε och δ är slumpkomponenter. 4p

3

Klassisk komponentuppdelning.

Skatta och tolka säsongskomponenterna i följande tidsserie med halvårsdata. Använd multiplikativ modell. Det är tillåtet att räkna med endast 3 decimaler.

1,9 6,2 3,4 8,1 5,9 8,1 7,0 10,0

3p

4

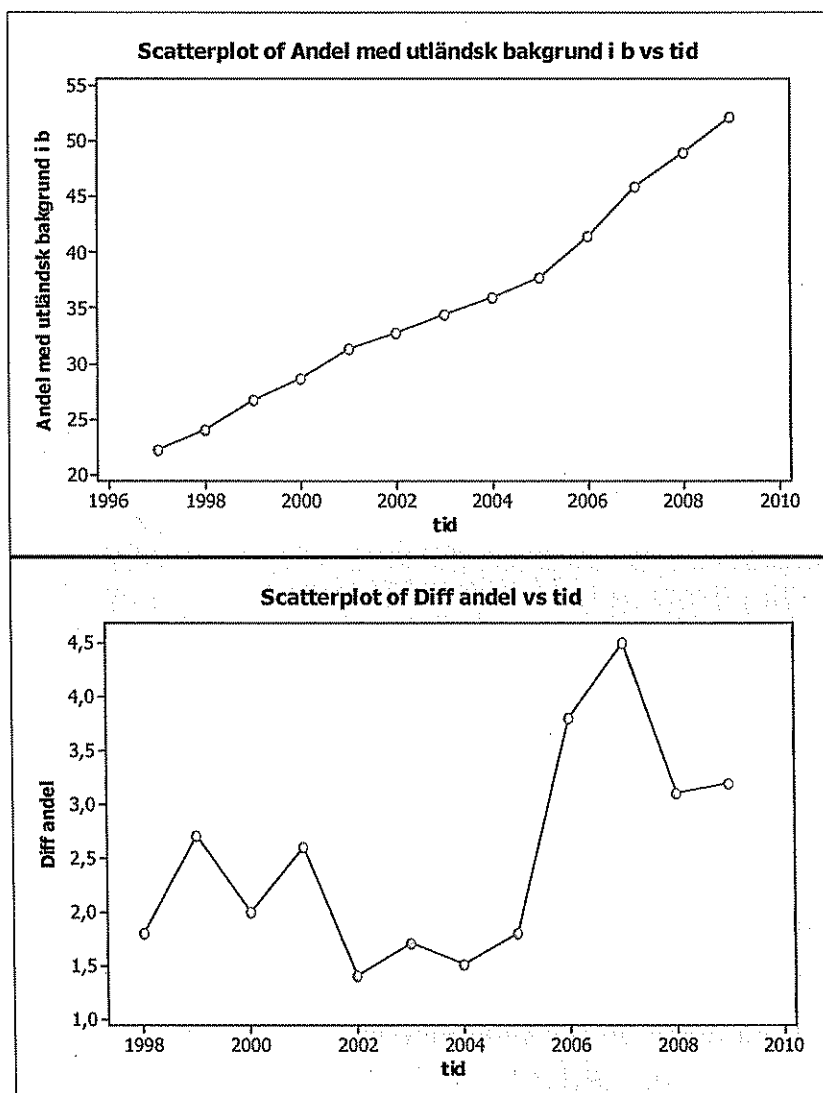
Följande data är hämtad från SCBs hemsida.

Data består av andel med utländsk bakgrund i befolkningen i stadsdelen Skäggetorp/Linköping. Andel ges i procent för åren 1997 t o m 2009.

Nedan är två modeller anpassade till data. Dessutom visas tidsserien och differentierad tidsserie. Sist följer tidsserien i siffror plus residualer från modell 2.

Din uppgift är nu att göra prognos för andelen med utländsk bakgrund i Skäggetorp 2010 och 2011 dels med modell 1 och dels med modell 2. Vilka prognoser tror du mest på? Motivera.

4p



Modell 1

ARIMA Model: Andel med utländsk bakgrund i b

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0,5458	0,2738	1,99	0,074
Constant	1,1386	0,2541	4,48	0,001

Differencing: 1 regular difference

Number of observations: Original series 13, after differencing 12

Residuals: SS = 7,67736 (backforecasts excluded)

MS = 0,76774 DF = 10

Modell 2

ARIMA Model: Andel med utländsk bakgrund i b

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
MA 1	-0,4774	0,2889	-1,65	0,129
Constant	2,4985	0,3804	6,57	0,000

Differencing: 1 regular difference

Number of observations: Original series 13, after differencing 12

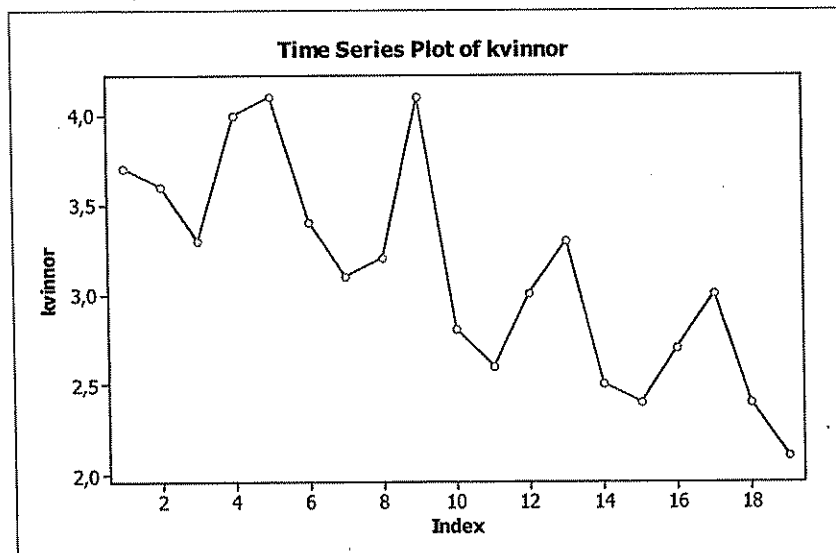
Residuals: SS = 8,03794 (backforecasts excluded)
MS = 0,80379 DF = 10

tid	Andel	Residualer från modell 2
1997	22,2	*
1998	24,0	-0,47903
1999	26,7	0,43014
2000	28,7	-0,70389
2001	31,3	0,43748
2002	32,7	-1,30739
2003	34,4	-0,17441
2004	35,9	-0,91528
2005	37,7	-0,26160
2006	41,5	1,42634
2007	46,0	1,32054
2008	49,1	-0,02895
2009	52,3	0,71528

5

I grafen nedan visas data över andelen sjukfrånvaro för statligt anställda kvinnor år 2004 kvartal 1 till år 2008 kvartal 3. Variabeln heter "kvinnor". Därefter är två modeller anpassade.

- Beskriv för- och nackdelar med de båda modellerna. Vilket/vilka mått ska användas för att avgöra vilken modell som bäst. Vilken modell väljer du? 2p
- Tolka koefficienterna framför t och kvart1 i modell 1. 1p
- Hur stor är förklaringsgraden i modell 1. Tolka den. 1p



Modell 1

Regression Analysis: kvinnor versus t; kvart1; kvart2; kvart3

The regression equation is

$$\text{kvinnor} = 3,99 - 0,0761 t + 0,339 \text{ kvart1} - 0,285 \text{ kvart2} - 0,449 \text{ kvart3}$$

Predictor	Coef	SE Coef	T	P
Constant	3,9857	0,1536	25,95	0,000
t	-0,076071	0,009916	-7,67	0,000
kvart1	0,3389	0,1577	2,15	0,050
kvart2	-0,2850	0,1574	-1,81	0,092
kvart3	-0,4489	0,1577	-2,85	0,013

S = 0,234651 R-Sq = 88,1% R-Sq(adj) = 84,6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	5,6807	1,4202	25,79	0,000
Residual Error	14	0,7709	0,0551		
Total	18	6,4516			

Modell 2

