

TENTAMEN I STATISTISK ANALYS AV SAMHÄLLSDATA, 2015-08-11

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!

1

Anta att kvaliteten på en produkt ska mätas med ett visst mått, Score2 men att detta mått är mycket dyrt att ta fram. Det finns då ett annat mått som är betydligt billigare Score1. Score2 kan uppskattas med Score1 via en enkel linjär regressionsanalys.

Regression Analysis: Score2 versus Score1

The regression equation is
Score2 = 1.12 + 0.218 Score1

Predictor	Coef	SE Coef
Constant	1.1177	0.1093
Score1	0.21767	0.01740

S = 0.127419 R-Sq = 95.7% R-Sq(adj) = 95.1%

Analysis of Variance

Source	DF	SS	MS
Regression	1	2.5419	2.5419
Residual Error	7	0.1136	0.0162
Total	8	2.6556	

- Pröva på 5% signifikansnivå om Score1 är en signifikant förklarande variabel för Score2. 1p
- Hur mycket av variationen i Score2 förklaras av Score1. 1p
- Prediktera storleken på Score2 då Score1 är 5. 1p

2

Utjämna tidsserien 116 108 95 125 133 110 150 108 med ett 3 punkters centrerat glidande medelvärde. 2p

3

I en påhittad hundpopulation är dödsriskerna enligt tabellen:
Som vanligt är den övre gränsen i åldern inte inkluderad i intervallet.

Ålder (år)	Dödsrisk (%) ${}_n q_x$
0-1 (1 ej inkl)	15
1-3	16
3-6	7
6-9	11
9-12	44
12+	100

Beräkna den återstående medellivslängden för de 6 åldersintervallen. Svara också på frågan: Hur gammal kan husse förvänta sig att hunden blir om den lyckats uppnå en ålder på 9 år?

Sätt förslagsvis $l_0 = 1000$ och ${}_n a_x = 0,5$.

3p

Använd följande formler:

$${}_n d_x = {}_n q_x \cdot l_x \qquad {}_n L_x = \frac{n(l_x + l_{x+n})}{2} \qquad L_{12+} \approx 3 \cdot l_{12}$$

$$T_x = T_{x+n} + {}_n L_x \qquad T_{12+} = L_{12+} \qquad e_x = \frac{T_x}{l_x}$$

4

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

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

- Skatta de fyra säsongskomponenterna i den klassiska komponentuppdelningen i en multiplikativ modell. Tolka en av komponenterna 3p
- Den skattade trendfunktionen för den säsongrensade serien $Dese$ är:

$$Dese_t = 31,34 + 0,662 \cdot tid$$

Beräkna prognos för Y_t i tidpunkt 13 med hjälp av resultatet i a- och b-uppgiften. 1p

- Även en tidsserieregression har anpassats till tidsserien.

$$Y_t = 47,0 - 15,5 kv_1 - 23,4 kv_2 - 22,7 kv_3 + 0,625 tid$$

där kv_1, kv_2 och kv_3 är dummies för kvartal.

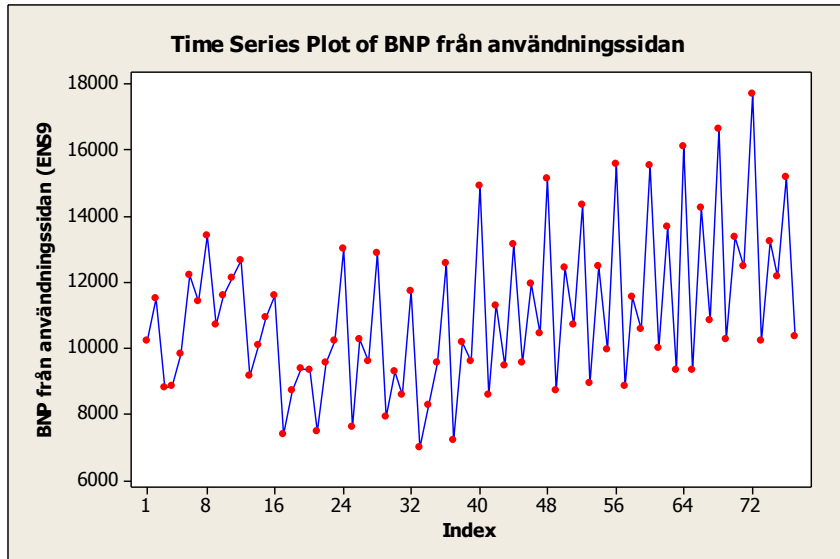
Beräkna prognos för Y_t i tidpunkt 13 med hjälp av denna anpassade modell. 1p

5

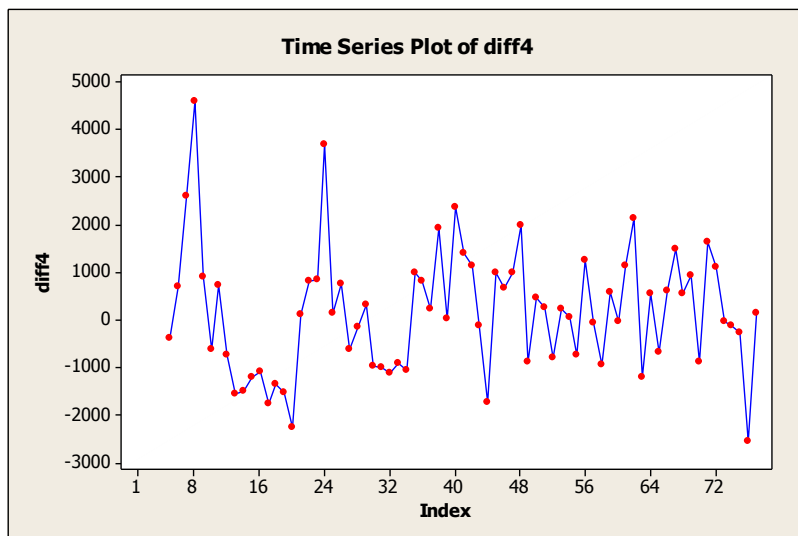
Följande tidsserie är hämtad från SCBs hemsida. Data är:

*BNP från användningssidan, försörjningsbalans, fasta priser referensår 2011, mnkr
Statliga myndigheters fasta bruttoinvesteringar.*

Data är mellan 1993K1 och 2012K1. Kvartalsdata. Vi vill nu bland annat beräkna prognoser för kvartal 2 och 3 år 2012.



Nedan ses den differentierade serien för kvartal



Modell 1

ARIMA Model: BNP från användningssidan (ENS9)

Relative change in each estimate less than 0,0010

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	1	0,5450	0,2830	1,93	0,058
MA	1	0,2242	0,3283	0,68	0,497
Constant		67,9	111,9	0,61	0,546

Differencing: 0 regular, 1 seasonal of order 4

Number of observations: Original series 77, after differencing 73

Residuals: SS = 105986406 (backforecasts excluded)
MS = 1514092 DF = 70

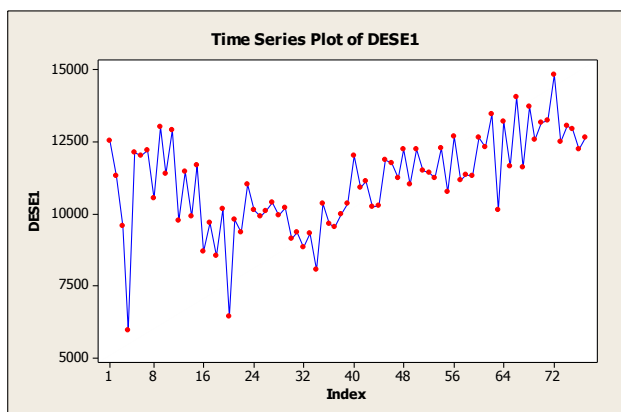
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	14,6	36,1	52,8	62,0
DF	9	21	33	45
P-Value	0,104	0,021	0,016	0,047

BNP	diff4	Resid	tid
13216	-129	-302,33	74
12176	-272	-337,34	75
15147	-2563	-2558,25	76
10324	126	881,46	77

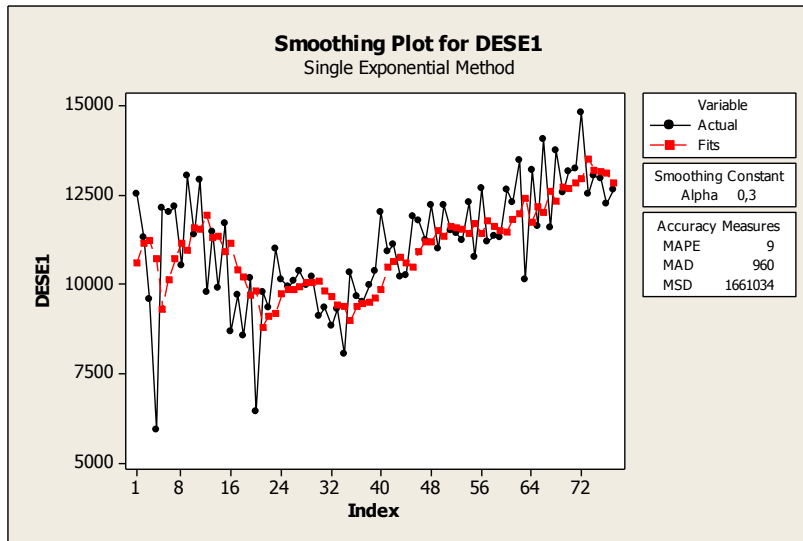
- a) Seriens värden är i fasta priser med referensår 2011. Vad betyder det? 1p
b) Gör prognos för kvartal 2 och 3 år 2012 genom att använda modell 1. 2p

I nästa modell (som vi kallar modell 2) så har först serien säsongrensats med klassisk additiv komponentuppdelning. Säsongrensad serie kallas här DESE1. Sedan har enkel exponentiell utjämning utförts på DESE1.



Säsongskomponenter

Period	Index
1	-2309,02
2	184,55
3	-772,64
4	2897,11



DESE1	utjämnad serie l_t	t
13031,5	13152,7	74
12948,6	13091,5	75
12249,9	12839,0	76
12633,0	12777,2	77

- c) Gör prognos för BNP för kvartal 2 och 3 år 2012 genom att använda modell 2. 2p
d) Visa hur l_{76} och l_{77} har beräknats. (Utgå från att l_{75} är beräknad) 2p