

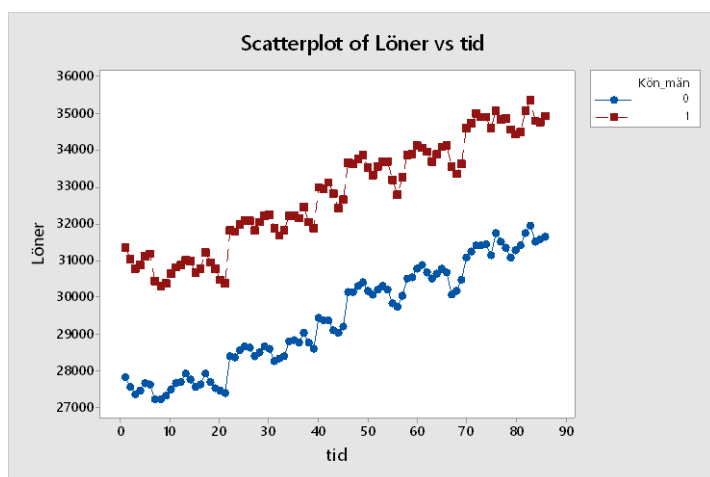
TENTAMEN I STATISTISK ANALYS AV SAMHÄLLSDATA, 2018-10-10

Skrivtid: kl: 14-18
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

Genomsnittlig månadslön för månadsavlönade, statlig sektor (KLS) polisväsendet heltidsanställning Månad 2010M01 - 2017M02 ges i grafen nedan uppdelat på män och kvinnor. De röda prickarna är lön för män.



Följande modell har anpassats.

Regression Analysis: Löner versus tid; Kön_män

Analysis of Variance

Source	DF	Adj SS	Adj MS
Regression	2	826456922	413228461
Error	169	24301569	143796
Total	171	850758492	

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
379,205	97,14%	97,11%	97,04%

Coefficients

Term	Coef	SE Coef
Constant	26962,5	65,1
tid	56,84	1,16
Kön_män	3355,2	57,8

Regression Equation

$$\text{Löner} = 26962,5 + 56,84 \text{ tid} + 3355,2 \text{ Kön_män}$$

Hur stor är den genomsnittliga skillnaden i lön mellan män och kvinnor enligt modellen ovan? Hur stor är felmarginalen för denna skillnad om 95% konfidensgrad används? 2p

2

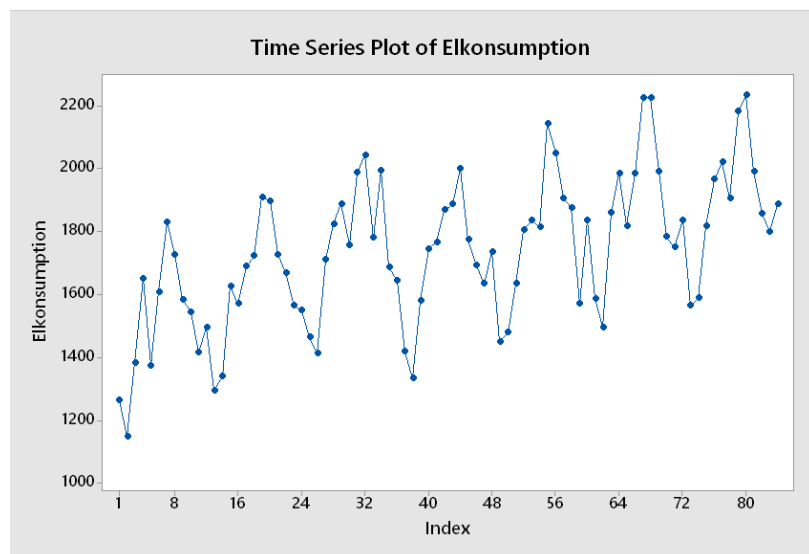
I tabellen nedan visas totalförsäljningen (i löpande priser) för insektsmedel för hästar och för hundar. Även priser för medlet 'Stopp o bit' för hästar från gruppen insektsmedel för hästar samt 'fästinghalsband' för hundar från gruppen insektsmedel för hundar är givna. Data är från åren 2014 till 2016

År	Tot. förs. insektsmedel för hästar	Tot. förs. insektsmedel för hundar	Pris, Stopp o bit	Pris, fästinghalsband
2014	1205	730	297	95
2015	1310	855	317	115
2016	1560	945	325	125

- a) Använd varorna 'Stopp o bit' och fästinghalsband som representantvaror för sina grupper och beräkna ett kedjeprisindex av Laspeyre-typ för företagets priser med basår 2014. 3p
- b) Hur har företagets priser för grupperna insektsmedel för hästar och hundar förändrats mellan 2014 och 2016? Använd det sammanvägda indexet från a) 1p

3

Ekonomen i ett företag vill prediktera elkostnaderna för nästa år. Man har data för de sista 84 månaderna. De sista 21 värdena i serien och en tidsseriegraf visas nedan. Anta att enheten på konsumtionen är kilo-wattimmar kWh.



Metoden *klassisk komponentuppdelning* har anpassats till tidsserien. Trendkomponent och säsongskomponenter ges av utskriften nedan.

Fitted Trend Equation

$$Y_t = 1520,2 + 5,119 \times t$$

Elkonsump	Tid	Månad
1985	64	4
1819	65	5
1984	66	6
2225	67	7
2226	68	8
1991	69	9
1783	70	10
1750	71	11
1836	72	12
1564	73	1
1589	74	2
1817	75	3
1966	76	4
2020	77	5
1905	78	6
2181	79	7
2233	80	8
1991	81	9
1856	82	10
1800	83	11
1887	84	12

Seasonal Indices

Period	Index
1	-284,627
2	-298,481
3	-40,273
4	53,998
5	67,748
6	62,665
7	297,811
8	267,457
9	62,332
10	3,477
11	-127,856
12	-64,252

Följande uppgifter gäller modellen ovan.

- Vilken månad är konsumtionen som högst? 0,5p
- Hur mycket ökar konsumtionen i snitt på ett år? 1p
- Beräkna de fyra sista anpassade värdena för Elkonsumtionen samt residualerna för modellen. 2p
- Beräkna prognoser för de sista fyra månaderna. 2p
- Ekonomen vill göra prognos för ett helt år. Verkar det rimligt att göra det för denna tidsserie? 0,5p

4

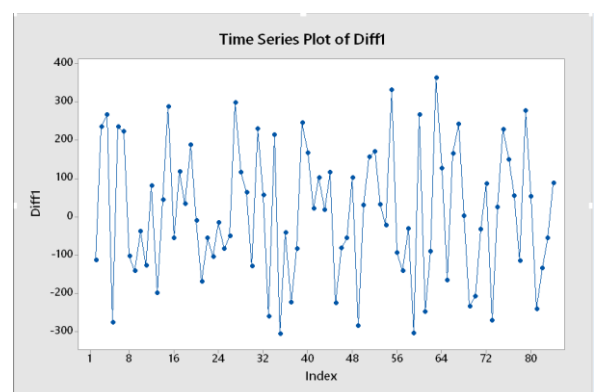
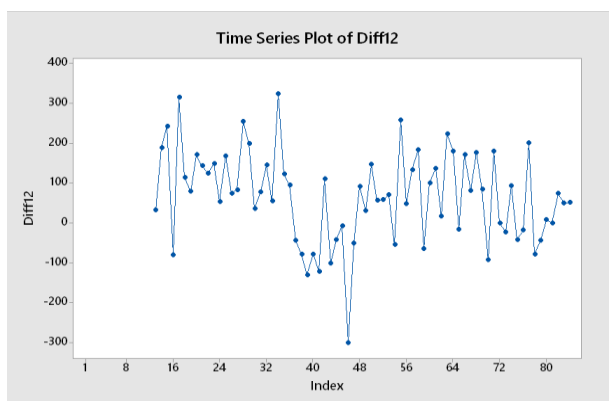
Vi fortsätter att analysera tidsserien från uppgift 3.

Nu ska en lämplig SARIMA-modell anpassas.

Eftersom vi ser en kraftig säsongvariation i tidsserien Elkonsumtion så differentieras den för säsong.

Den differentierade tidsserien kallas nu Diff12. Men man kan även se en tydlig trend så tidsserien

differentieras även för endast trend. Denna differentierade tidsserie kalla Diff1. Resultaten ses i graferna nedan.



- Rent visuellt, ser Diff12 och/eller Diff1 stationär ut? Vilka tre krav ställer vi på en stationär tidsserie? 2p

Nedan har fyra olika SARIMA-modeller anpassats.

ARIMA Model 1: Elkonsumtion

Relative change in each estimate less than 0,001

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
SAR 12	-0,262	0,126	-2,07	0,042
SMA 12	0,848	0,112	7,57	0,000
Constant	76,33	1,78	42,84	0,000

Differencing: 0 regular, 1 seasonal of order 12

Number of observations: Original series 84, after differencing 72

Residual Sums of Squares

DF	SS	MS
69	418631	6067,11

Back forecasts excluded

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

Lag	12	24	36	48
Chi-Square	19,10	37,35	50,52	64,43
DF	9	21	33	45
P-Value	0,024	0,015	0,026	0,030

Residualer	Tid	Månad
67,486	68	8
50,269	69	9
-116,564	70	10
-39,562	71	11
16,624	72	12
-53,937	73	1
3,310	74	2
-17,664	75	3
1,508	76	4
59,923	77	5
-89,768	78	6
-18,692	79	7
34,264	80	8
-11,706	81	9
-126,499	82	10
-12,675	83	11
-11,240	84	12

ARIMA Model 2: Elkonsumption

Estimates at Each Iteration

Unable to reduce sum of squares any further

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0,6003	0,0953	6,30	0,000
MA 1	0,9766	0,0490	19,91	0,000
Constant	2,18	1,01	2,15	0,034

Differencing: 1 regular difference

Number of observations: Original series 84, after differencing 83

Residual Sums of Squares

DF	SS	MS
80	1994165	24927,1

Back forecasts excluded

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

Lag	12	24	36	48
Chi-Square	76,04	150,03	202,82	247,75
DF	9	21	33	45
P-Value	0,000	0,000	0,000	0,000

Residualer	Tid	Månad
134,393	68	8
-106,526	69	9
-173,138	70	10
-79,402	71	11
26,086	72	12
-300,329	73	1
-107,200	74	2
106,120	75	3
113,588	76	4
73,308	77	5
-78,001	78	6
266,682	79	7
144,582	80	8
-134,191	81	9
-122,955	82	10
-97,215	83	11
23,497	84	12

ARIMA Model 3: Elkonsumption

Estimates at Each Iteration

Relative change in each estimate less than 0,001

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0,855	0,165	5,19	0,000
SAR 12	-0,256	0,139	-1,84	0,070
MA 1	0,683	0,225	3,04	0,003
SMA 12	0,828	0,122	6,80	0,000
Constant	11,091	0,609	18,20	0,000

Differencing: 0 regular, 1 seasonal of order 12

Number of observations: Original series 84, after differencing 72

Residual Sums of Squares

DF	SS	MS
67	395976	5910,09

Back forecasts excluded

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

Lag	12	24	36	48
Chi-Square	13,73	29,65	46,91	61,73
DF	7	19	31	43
P-Value	0,056	0,056	0,033	0,032

ARIMA Model 4: Elkonsumption

Estimates at Each Iteration

Relative change in each estimate less than 0,001

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	-0,083	0,155	-0,53	0,595
SAR 12	-0,239	0,148	-1,62	0,110
MA 1	0,7677	0,0993	7,73	0,000
SMA 12	0,773	0,142	5,46	0,000
Constant	-1,028	0,625	-1,64	0,105

Differencing: 1 regular, 1 seasonal of order 12

Number of observations: Original series 84, after differencing 71

Residual Sums of Squares

DF	SS	MS
66	443075	6713,26

Back forecasts excluded

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

Lag	12	24	36	48
Chi-Square	14,43	32,16	49,74	60,69
DF	7	19	31	43
P-Value	0,044	0,030	0,018	0,039

Data finns efter uppgifterna

- b) Skriv var och en av de fyra modellerna ovan på formen $ARIMA(p,d,d)(P,D,Q)x$.
Dvs ge värdena på p d q P D Q och x . 2p
- c) Prediktera Elkonsumtionen för de närmast två månaderna med hjälp av modell 1. 2p
- d) Prediktera Elkonsumtionen för de närmast två månaderna med hjälp av modell 2 2p

Tid	Månad	Elkonsump	Diff1	Diff12
68	8	2226	1	177
69	9	1991	-235	84
70	10	1783	-208	-93
71	11	1750	-33	180
72	12	1836	86	0
73	1	1564	-272	-23
74	2	1589	25	93
75	3	1817	228	-42
76	4	1966	149	-19
77	5	2020	54	201
78	6	1905	-115	-79
79	7	2181	276	-44
80	8	2233	52	7
81	9	1991	-242	0
82	10	1856	-135	73
83	11	1800	-56	50
84	12	1887	87	51