

TENTAMEN I REGRESSIONS- OCH TIDSSERIEANALYS, 2014-06-13

Skrivtid: kl 8-12

Hjälpmedel: Räknedosa. Bowerman, B.J., O'Connell, R, Koehler, A.: *Forecasting, Time Series and Regression*. 4th ed. Duxbury, 2005 som inte får innehålla anteckningar men får ha markeringar och flärpar. Flärpar får ha en liten anteckning. Med tentan vidhäftat formelblad.

Jourhavande lärare: Lotta Hallberg

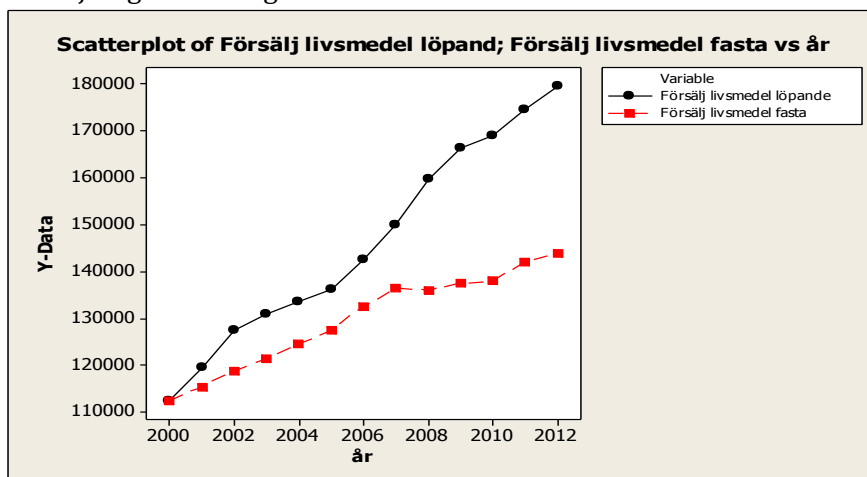
Redovisa och motivera kort alla dina lösningar

1

Följande data är hämtat från SCB hemsida.

Försäljning (inkl moms) av livsmedel Handel År 2000 – 2012. (Enhet saknas men anta att det är i miljoner kr)

Nedan visas en graf över försäljningen i löpande och fasta priser. Även en tabell med försäljningsvärdena ges.



ÅR	FÖRSÄLJ LIVSMEDEL FASTA	FÖRSÄLJ LIVSMEDEL LÖPANDE
2000	112418	112418
2001	115365	119381
2002	118802	127495
2003	121352	130829
2004	124511	133648
2005	127556	136137
2006	132570	142630
2007	136510	150017
2008	135974	159633
2009	137408	166362
2010	138019	168955
2011	142081	174535
2012	143724	179490

- a) Förklara vad som menas med att försäljningsvärdena är i **löpande** priser respektive **fasta** priser. Ange också vilket år som är basår. 2p
- b) Beräkna implicitprisindex med hjälp av de båda försäljningsserierna ovan för år 2009 till år 2012. 1p

Följande regressionsekvation har anpassats:

The regression equation is
Försälj livsmedel fasta = - 5142450 + 2628 År

Predictor	Coef	SE Coef	T	P
Constant	-5142450	281917	-18,24	0,000
År	2628,2	140,5	18,70	0,000

S = 1895,94 R-Sq = 97,0% R-Sq(adj) = 96,7%

Durbin-Watson statistic = 0,686989

- c) Hur stor är ökningen i försäljning av livsmedel per år enligt modellen ovan? 1p
- d) Tolka förklaringsgraden. 1p
- e) Kan residualerna antas vara okorrelerade? Förklara. 1p

2

Följande data är antalet dagar till distribution av en viss vara.

Beräkna ett fem-punkters centrerat glidande medelvärde för tidsserien:

38, 40, 25, 23, 35, 38, 35, 32, 41, 33, 31, 21

2p

3

Man vill undersöka hur försäljningen för en viss typ av affär påverkas av antalet hushåll i närområdet samt av hur affären är placerad. Under juli ett visst år i en viss stad valde man slumpmässigt ut 5 affärer som låg utmed en gata, 5 affärer som låg i centrala staden samt 5 affärer som låg i ett köpcentrum.

Variablerna är betecknade:

Y= försäljning i 1000-tal dollar

x= antal hushåll i 1000-tal

D1= 1 om affär placerad i köpcentrum och 0 annars

D2= 1 om affär placerad i centrala staden och 0 annars

Så vi ser att den kvalitativa variabeln som beskriver placering av affär är hanterad som två dummyvariabler D1 och D2.

Vidare är:

xD1 och xD2 interaktionstermer mellan x och D1, D2.

Sist kommer ett spridningsdiagram över data

Följande 3 modeller har anpassats:

Modell 1

Regression Analysis: Y versus x; D1; D2; xD1; xD2

The regression equation is

$$Y = 7,9 + 0,921 x + 42,7 D1 + 10,3 D2 - 0,092 xD1 - 0,034 xD2$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	7,90	17,04	0,46	0,654	
x	0,9207	0,1234	7,46	0,000	11,723
D1	42,73	21,50	1,99	0,078	33,340
D2	10,26	21,28	0,48	0,641	32,658
xD1	-0,0917	0,1416	-0,65	0,533	53,014
xD2	-0,0336	0,1382	-0,24	0,813	59,686

S = 6,79953 R-Sq = 98,8% R-Sq(adj) = 98,1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	33296,1	6659,2	144,03	0,000
Residual Error	9	416,1	46,2		
Total	14	33712,2			

Modell 2

Regression Analysis: Y versus x; D1; D2

The regression equation is

$$Y = 15,0 + 0,869 x + 28,4 D1 + 6,86 D2$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	14,978	6,188	2,42	0,034	
x	0,86859	0,04049	21,45	0,000	1,447
D1	28,374	4,461	6,36	0,000	1,646
D2	6,864	4,770	1,44	0,178	1,882

S = 6,34941 R-Sq = 98,7% R-Sq(adj) = 98,3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	33269	11090	275,07	0,000
Residual Error	11	443	40		
Total	14	33712			

Modell 3

Regression Analysis: Y versus x

The regression equation is

$$Y = 14,9 + 0,937 x$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	14,87	13,13	1,13	0,278	
x	0,93712	0,07305	12,83	0,000	1,000

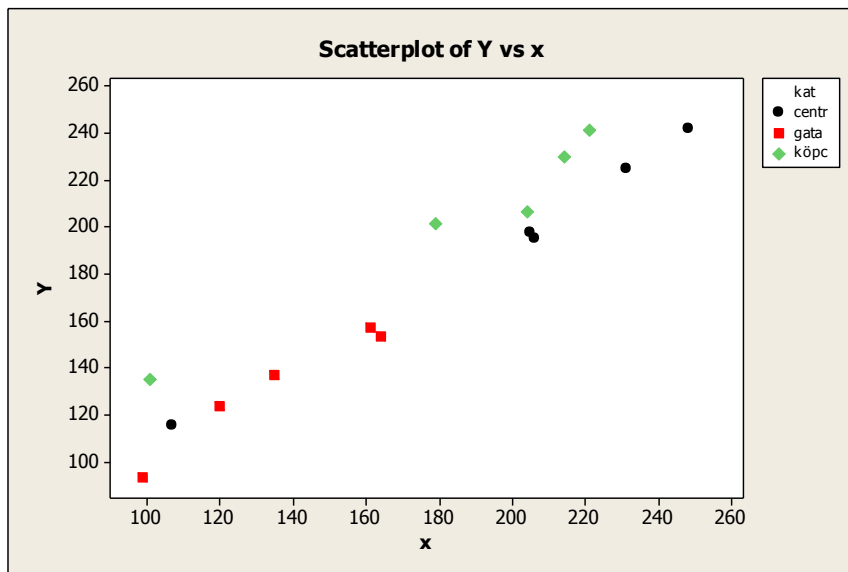
S = 13,7779 R-Sq = 92,7% R-Sq(adj) = 92,1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	31244	31244	164,59	0,000
Residual Error	13	2468	190		
Total	14	33712			

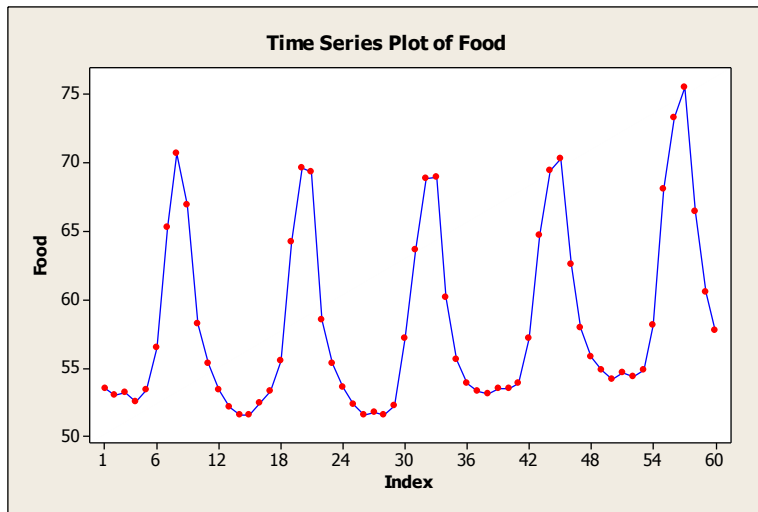
Uppgifter:

- a) Pröva med ett test om de båda interaktionstermerna i modell 1 kan tas bort. 5% signifikansnivå. 2p
- b) Skriv upp modellen för försäljningen i en affär placerad i ett köpcentrum med ett godtyckligt antal hushåll med hjälp av modell 1. 1,5p
- c) Studera VIF i de tre modellerna och förklara vilka värden som är bra och vilka som är dåliga. Vad mäter VIF och hur beräknas de? 2p
- d) Vilken av modellerna är bäst om man använder justerad R^2 som ett jämförande mått? 0,5p
- e) Tolka regressionskoefficienten för D2 i modell 2. 1p
- f) Prediktera försäljningen för en affär som har 150000 hushåll i sin närhet och ligger i ett köpcentrum enligt modell 2. 1p



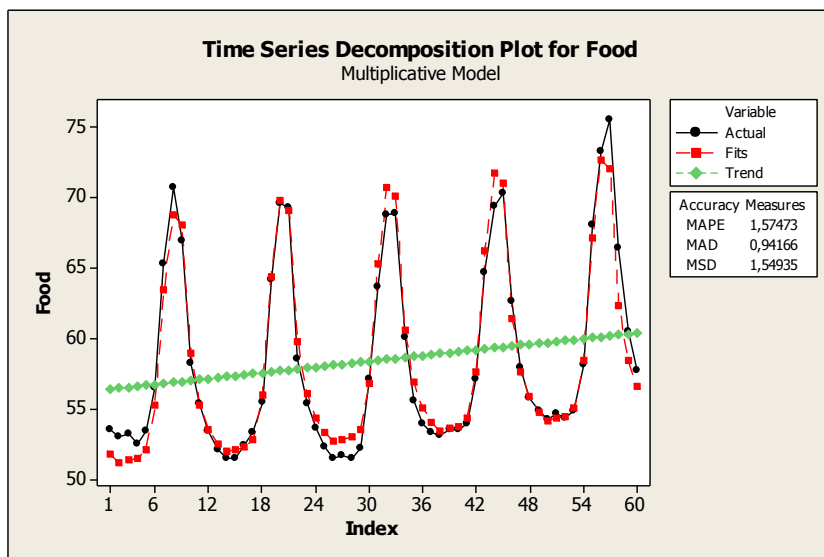
4

Följande tidsserie ska analyseras:



Tidsserien Food är sysselsättningen i en industri som tillverkar helfabricerad mat, månadsdata för fem år.

Modell 1:



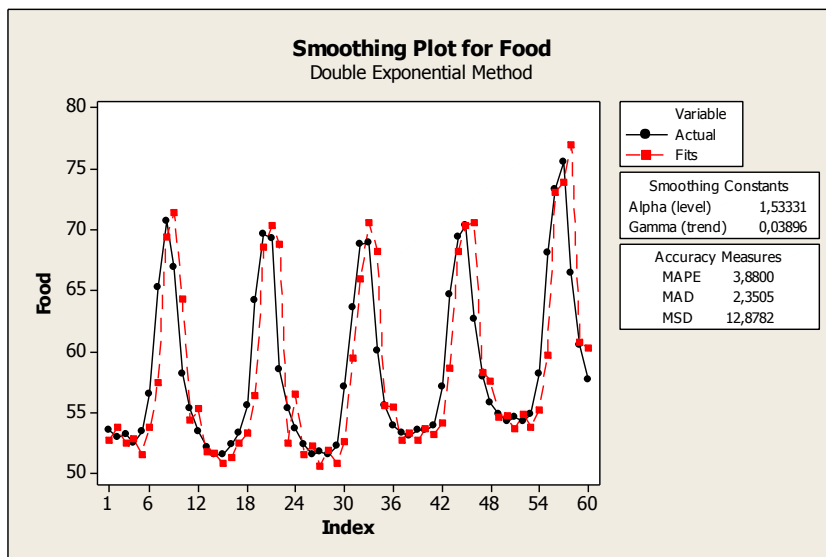
Fitted Trend Equation

$$Y_t = 56,305 + 0,0679 * t$$

Seasonal Indices

Period	Index
1	0,91852
2	0,90692
3	0,90865
4	0,91040
5	0,91855
6	0,97342
7	1,11791
8	1,20940
9	1,19646
10	1,03381
11	0,96865
12	0,93732

Modell 2



- a) Modell 1. Tolka säsongskomponenten för juni och juli. 1p
- b) Modell 1. Beräkna prognoser för Food för månad 61 och 62. 1p
- c) Modell 2. Förklara varför detta är en dålig modell. Ge förslag på förbättring. 2p