



EXP

Försättsblad till skriftlig tentamen vid Linköpings Universitet

(fylls i av ansvarig)

Datum för tentamen	2009-03-26
Sal	TER2
Tid	08-12
Kurskod	732G05
Provkod	TENB
Kursnamn/benämning	Statistik I, grundkurs
Institution	IDA
Antal uppgifter som ingår i tentamen	4
Antal sidor på tentamen (inkl. försättsbladet)	8
Jour/Kursansvarig	Lotta Hallberg
Telefon under skrivtid	0702 642030
Besöker salen ca kl.	10,00
Kursadministratör (namn + tfnr + mailadress)	Carita Lilja 013-2811463 carli@ida.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. Formelsamling. Tabeller
Övrigt (exempel när resultat kan ses på webben, betygsgränser, visning, övriga salar tentan går i m.m.)	G=12, VG=16

Tentamen i Regressions- och tidsserieanalys, 2009-03-26

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. Formelsamling. Tabeller

Jourhavande lärare: Lotta Hallberg

Redovisa och motivera kort alla dina lösningar

I uppgifterna ges vissa översättningar till engelska inom parentes.

1

Man vill prediktera underhållskostnaden per månad för en maskin. Följande data finns tillgängliga.

Kostnad i tkr	8.8	8.5	8.3	9.5	10.0	11.25	13.1
Månad	1	2	3	4	5	6	7

a) Sätt upp en enkel linjär (simple linear) regressionsmodell och skatta (estimate) de båda beta-parametrarna med minsta kvadratmetoden (least squares estimates). Pröva på 5% signifikansnivå om lutningen är signifikant. Du får använda direkt att $SSE=3.53$.

3p

b) Prediktera kostnaden för månad nr 8.

1p

2

I tabellen nedan visas totalförsäljningen i löpande priser för ett postorderföretag som säljer rid och hästartiklar, samt priser för varan ridbyxor från varugrupp ridartiklar och varan regntäcke från varugrupp hästartiklar under åren 2003-2005

År	Tot. förs. Ridartiklar Milj kr	Tot. förs. Hästartiklar Milj kr	Pris, ridbyxor kr	Pris, regntäcke kr
2003	127.0	177.0	1050	1750
2004	127.3	178.8	1125	1970
2005	133.0	186.8	1200	2210

Använd varorna *ridbyxor* och *regntäcke* som representantvaror för sina respektive varugrupper och beräkna ett kedjeprisindex av Laspeyre-typ för företagets priser för 2003-2005 med basår 2003.

3p

3

En stor livsmedelskedja vill granska sin produktivitet och sina kostnader noga. Data som ska analyseras nedan är tagna från en affär för varje vecka under 1 år, dvs 52 observationer. Variablerna är: x_1 = antal sålda enheter under veckan, x_2 =indirekt kostnad av totala arbetskraften i procent, $x_3=1$ om veckan har en helgdag och 0 annars, $t=1$ till 52 dvs den är en tidsvariabel för veckorna. x_1x_2 =interaktion mellan x_1 och x_2 .

Responsvariabeln Y =totala antalet arbetskraftstimmar under veckan.

Nedan följer fem anpassade (fitted) modeller:

Modell 1

Regression Analysis: Y versus x1; x2; x3; t; x1x2

The regression equation is

$$Y = 4507 - 0,00032 x_1 - 60 x_2 + 623 x_3 + 0,93 t + 0,000133 x_1x_2$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	4507	1346	3,35	0,002	
x1	-0,000319	0,004392	-0,07	0,942	141,616
x2	-60,0	175,7	-0,34	0,734	57,098
x3	623,42	65,76	9,48	0,000	1,082
t	0,926	1,602	0,58	0,566	1,416
x1x2	0,0001334	0,0005745	0,23	0,817	215,548

S = 145,686 R-Sq = 69,1% R-Sq(adj) = 65,8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	2185813	437163	20,60	0,000
Residual Error	46	976323	21224		
Total	51	3162136			

Durbin-Watson statistic = 2,27001

Modell 2

Regression Analysis: Y versus x2; x3; t

The regression equation is

$$Y = 4409 - 23,3 x_2 + 634 x_3 + 2,00 t$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	4408,7	179,0	24,63	0,000	
x2	-23,33	25,91	-0,90	0,373	1,214
x3	634,35	64,52	9,83	0,000	1,018
t	2,001	1,491	1,34	0,186	1,199

S = 147,346 R-Sq = 67,0% R-Sq(adj) = 65,0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	2120016	706672	32,55	0,000
Residual Error	48	1042119	21711		
Total	51	3162136			

Durbin-Watson statistic = 2,06634

Modell 3

Regression Analysis: Y versus x1; x2; x3

The regression equation is

$$Y = 4150 + 0,000787 x_1 - 13,2 x_2 + 624 x_3$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	4149,9	195,6	21,22	0,000	
x1	0,0007871	0,0003646	2,16	0,036	1,009
x2	-13,17	23,09	-0,57	0,571	1,020
x3	623,55	62,64	9,95	0,000	1,014

S = 143,289 R-Sq = 68,8% R-Sq(adj) = 66,9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	2176606	725535	35,34	0,000
Residual Error	48	985530	20532		
Total	51	3162136			

Durbin-Watson statistic = 2,29835

Modell 4

Regression Analysis: Y versus x1; x3; t

The regression equation is
 $Y = 4060 + 0,000721 x1 + 620 x3 + 0,48 t$

Predictor	Coef	SE Coef	T	P	VIF
Constant	4060,3	112,1	36,23	0,000	
x1	0,0007213	0,0003919	1,84	0,072	1,160
x3	620,37	62,44	9,94	0,000	1,003
t	0,484	1,428	0,34	0,736	1,158

S = 143,602 R-Sq = 68,7% R-Sq(adj) = 66,7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	2172303	724101	35,11	0,000
Residual Error	48	989833	20622		
Total	51	3162136			

Durbin-Watson statistic = 2,31733

Modell 5

Regression Analysis: Y versus x1; x3

The regression equation is
 $Y = 4058 + 0,000770 x1 + 620 x3$

Predictor	Coef	SE Coef	T	P	VIF
Constant	4058,4	110,9	36,59	0,000	
x1	0,0007704	0,0003609	2,13	0,038	1,002
x3	619,63	61,83	10,02	0,000	1,002

S = 142,299 R-Sq = 68,6% R-Sq(adj) = 67,3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	2169932	1084966	53,58	0,000
Residual Error	49	992204	20249		
Total	51	3162136			

Durbin-Watson statistic = 2,32198

- För att testa om det finns ett tidsberoende mellan veckorna så kan man gå tillväga på två olika sätt. Visa hur detta går till genom att använda de båda metoderna på två lämpligt valda modeller ovan. 2p
- Hur mycket ökar antalet arbetskraftstimmar i snitt en vecka som innehåller en helgdag. Använd valfri modell ovan. 1p
- Använd ett partiellt F-test för att pröva om man kan ta bort x1 och x1x2 från modell 1. 2p
- Endast modell 1 har höga VIF-värden. Hur ska detta tolkas och vilken är den troliga orsaken. 1p
- Använd modell 2 och prediktera totala antalet arbetskraftstimmar nästkommande vecka. Anta att denna vecka har en helgdag samt att den indirekta kostnaden är 7,5 %. 1p

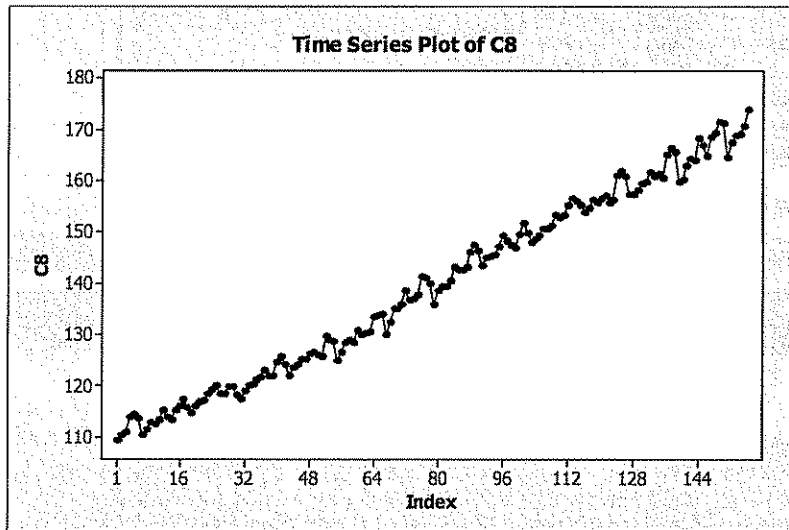
4

Följande data är hämtade från SCB hemsida.

Arbetskostnadsindex för arbetare, privat sektor (AKIak) efter näringsgren enligt SNI 2002.

Månad 1996M01-2008M12. Ger tidpunkt 1 till 156

Handel, reparation av fordon, hotell och restauranger



Fyra modeller är anpassade i Minitab och tidsserien ligger i C8.
Uppgifterna kommer efter alla utskrifter

Modell 1

Time Series Decomposition for C8

Multiplicative Model

Data C8
Length 156
NMissing 0

Fitted Trend Equation

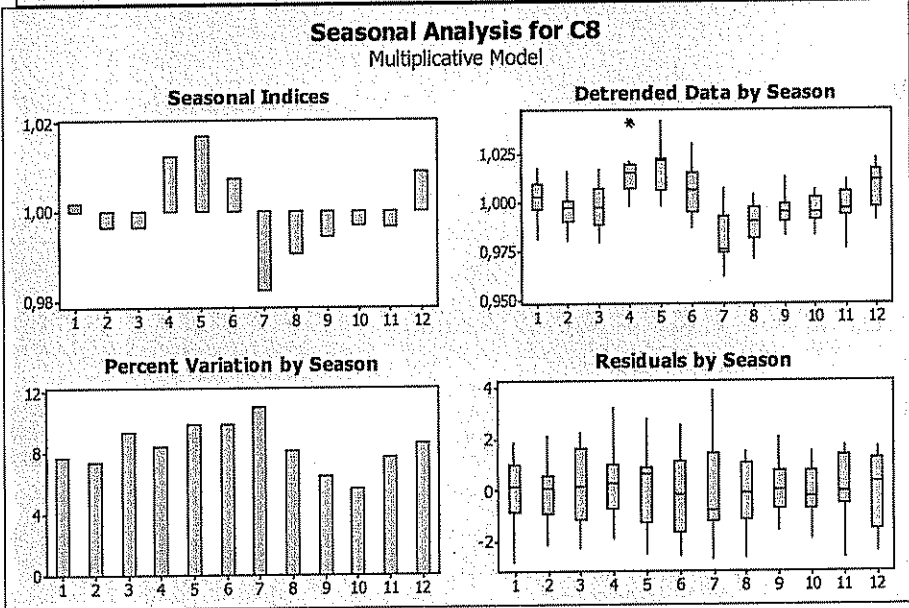
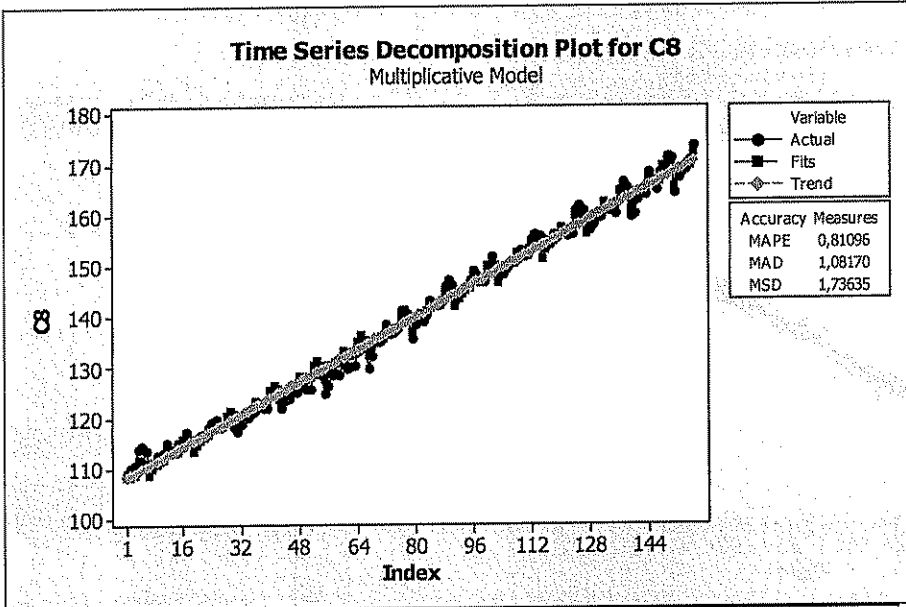
$$Y_t = 107,764 + 0,402 * t$$

Seasonal Indices

Period	Index
1	1,00187
2	0,99643
3	0,99648
4	1,01247
5	1,01710
6	1,00748
7	0,98201
8	0,99012
9	0,99436
10	0,99661
11	0,99634
12	1,00873

Accuracy Measures

MAPE 0,81096
MAD 1,08170
MSD 1,73635



Model 2

Trend Analysis for C8

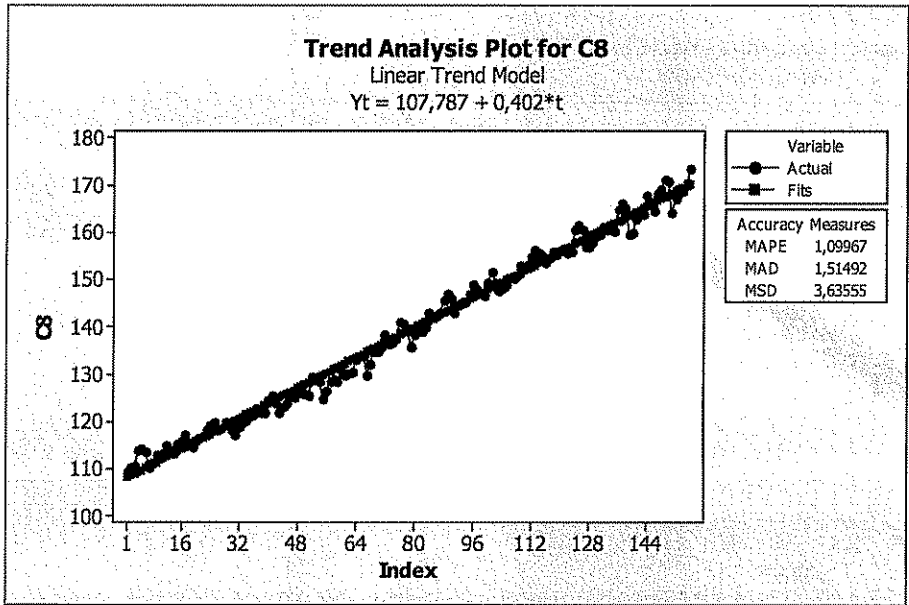
Data C8
 Length 156
 NMissing 0

Fitted Trend Equation

$$Y_t = 107,787 + 0,402 * t$$

Accuracy Measures

MAPE 1,09967
 MAD 1,51492
 MSD 3,63555



Modell 3
Winters' Method for C8

Multiplicative Method

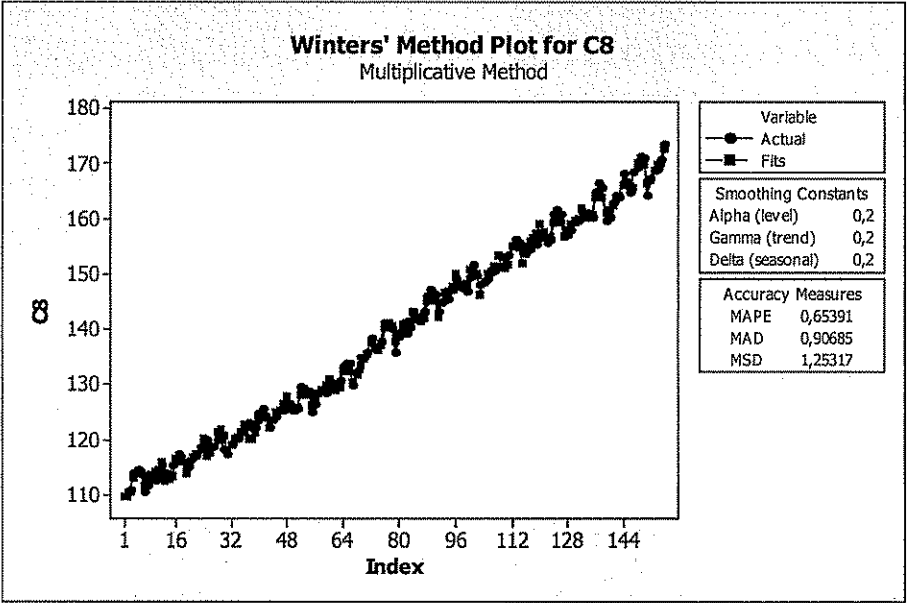
Data C8
Length 156

Smoothing Constants

Alpha (level) 0,2
Gamma (trend) 0,2
Delta (seasonal) 0,2

Accuracy Measures

MAPE 0,65391
MAD 0,90685
MSD 1,25317



Modell 4 Double Exponential Smoothing for C8

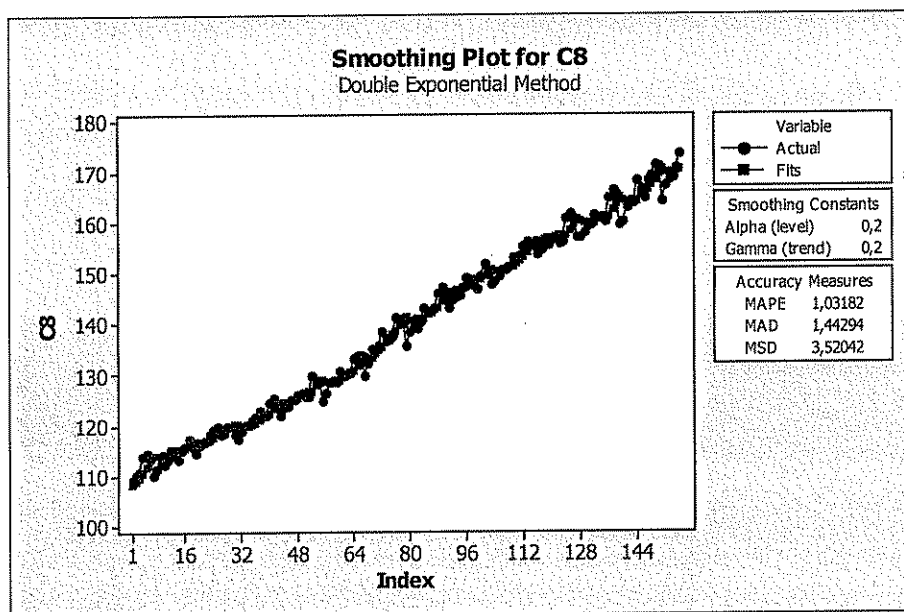
Data C8
Length 156

Smoothing Constants

Alpha (level) 0,2
Gamma (trend) 0,2

Accuracy Measures

MAPE 1,03182
MAD 1,44294
MSD 3,52042



Här är data för det sista året:

2008M01 166,7
2008M02 164,7
2008M03 168,4
2008M04 169,2
2008M05 171,4
2008M06 171,1
2008M07 164,3
2008M08 167,4
2008M09 168,7
2008M10 168,9
2008M11 170,7
2008M12 173,7

- a) Vilka är för- resp nackdelar med modell av typ 1 och 2 ovan gentemot modell 3 och 4? 2p
- b) Vilken av de fyra modellerna väljer du? Motivera väl varför du vill/inte vill ha med säsongskomponent/effekt. 1p
- c) Det finns ytterligare en modell som inkluderar säsongseffekt som skulle kunna anpassas här. Beskriv den modellen. 1p
- d) Beräkna två prognoser för december 2008 med hjälp av modell 1 och modell 2. Jämför med det sanna värdet. Vilken prognos anser du vara den mest trovärdiga? 2p

