



## Försättsblad till skriftlig tentamen vid Linköpings Universitet

<b>Datum för tentamen</b>	2011-06-08
<b>Sal (2)</b> Om tentan går i flera salar ska du bifoga ett försättsblad till varje sal och <u>ringa in</u> vilken sal som avses	TER4 TERC
<b>Tid</b>	14-18
<b>Kurskod</b>	732G25
<b>Provkod</b>	TENT
<b>Kursnamn/benämning</b> <b>Provnamn/benämning</b>	Statistisk analys av samhällsdata Tentamen
<b>Institution</b>	IDA
<b>Antal uppgifter som ingår i tentamen</b>	3 upg, 6 sidor exkl försättsblad
<b>Jour/Kursansvarig</b> Ange vem som besöker salen	Lotta Hallberg
<b>Telefon under skrivtiden</b>	
<b>Besöker salen ca kl.</b>	10
<b>Kursadministratör/kontaktperson</b> (namn + tfnr + mailaddress)	carita.lilja@liu.se , 1463
<b>Tillåtna hjälpmedel</b>	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. Anteckning på flärp tillåten.
<b>Övrigt</b>	G=12, VG=16
<b>Vilken typ av papper ska användas, rutigt eller linjerat</b>	
<b>Antal exemplar i påsen</b>	

## Tentamen i Statistisk analys av samhällsdata, 2011-06-08

**Skrivtid:** kl: 14-18

**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. Anteckning på flärp tillåten.

**Jourhavande lärare:** Lotta Hallberg

**Redovisa och motivera kort alla dina lösningar**

---

### 1

Data nedan är hämtade från SCB:s hemsida. Data visar Hushållens konsumtionsutgifter (ENS95), löpande priser, mnkr efter ändamål COICOP och tid. Speciellt livsmedel här.

- a) Räkna om KPI för åren 2004 till 2008 så att 2004 blir basår. 1p  
b) Använd KPI för att deflatera löpande prisserien för livsmedel nedan 1p  
c) Hur stor är den procentuella reella förändringen för konsumtionsutgifterna för livsmedel mellan åren 2004 och 2008. 1p

	2004	2005	2006	2007	2008
<b>011 livsmedel</b>	136306	138831	145511	153026	162671
KPI	279.14	280.41	284.22	290.51	300.50

COICOP - Classification of individual consumption by purpose.

### 2

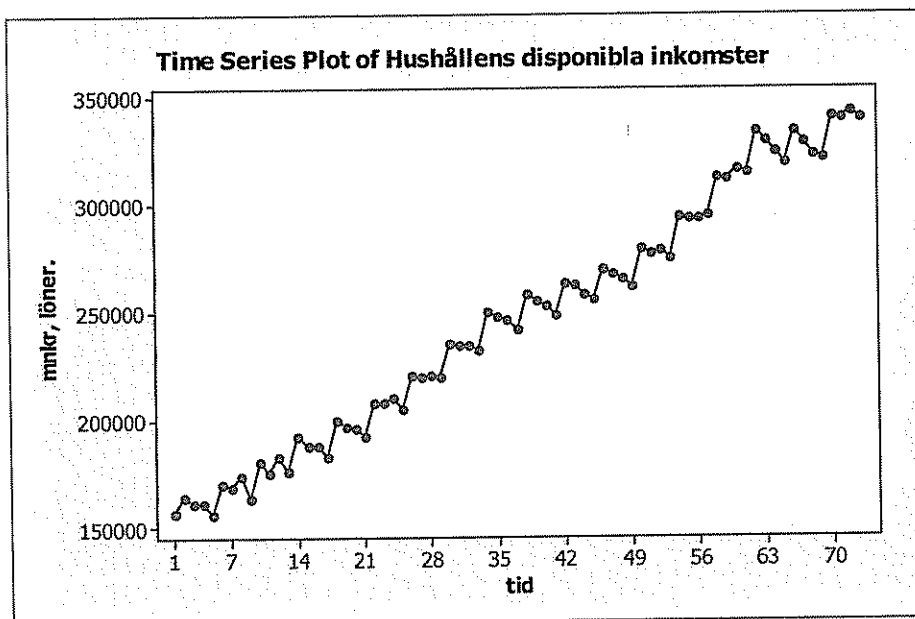
Du har tillgång till följande data om befolkningen i Bangladesh 1974.

<b>Åldersgrupp</b>	<b>Män i tusental</b>	<b>Kvinnor i tusental</b>	<b>Födslar i tusental</b>
0-4	7457	7192	--
5-9	5820	5625	--
10-14	4730	4533	--
15-19	4017	3777	756.9
20-24	3362	3101	1046.0
25-29	2891	2636	819.5
30-34	2391	2161	565.1
35-39	1983	1793	353.2
40-44	1635	1484	141.6
45-49	1343	1222	16.5
50-54	1066	974	--
55-59	858	784	--
60-64	658	601	--
65-69	487	440	--
70-74	339	304	--
75+	399	334	--

- a) Bestäm födelsetalet (Crude birth rate) CBR 1p
- b) Bestäm den generella fruktsamhetskvoten GBR 1p
- c) Bestäm den åldersspecifika fruktsamheten för gruppen 20-29år 1p
- d) Hur många barn förväntas en kvinna föda i Bangladesh 1974? 2p

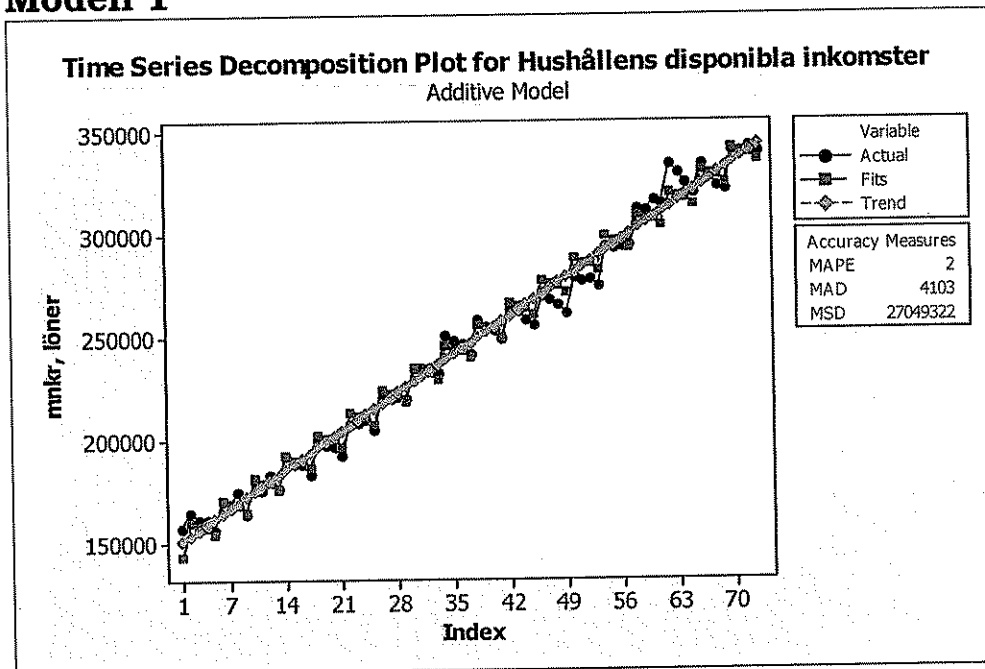
### 3

Nedan visas en tidsserie över Hushållens disponibla inkomster (ENS95), mnkr, löner. Kvartalsdata 1993K1 till 2011K1



- a) Tolka säsongskomponenten för kvartal 4 i modell 1 och 2. 1p
- b) Hur mycket kommer den disponibla inkomsten öka per år enligt trendlinjen. Välj modell 1 eller 2 vid denna tolkning. 1p
- c) Gör prognos för kvartal 2 och 3 år 2011 dels med modell 1 och dels med modell 2. 2p
- d) Tolka Durbin-Watson statistikan i modell 2. 1p
- e) Eftersom både modell 1 och 2 uppvisade autokorrelerade residualer så har 3 SARIMA-modeller anpassats (modell 3 till modell 5). Jämför dessa tre modeller noggrant. Vilken anser du vara den bästa modellen. Motivera. 2p
- f) Låt  $y_t$  vara den disponibla inkomsten. Låt vidare  $z_t$  vara  $\text{diff}^4_1$ , dvs  $z_t$  är resultatet av att  $y_t$  är differentierad först med en tidsförskjutning och sedan med 4 tidsförskjutningar (lag). Uttryck  $z_t$  i  $y_t$ . 2p
- g) Modell 4 är den enklaste av de tre SARIMA-modellerna därför ska vi titta närmare på den. Sist i tentan visas vissa data från slutet på tidsserien. Även prognoser som Minitab har beräknat. Visa hur prognoserna är beräknade. 3p

# Modell 1



Time Series Decomposition for Hushållens disponibla inkomster

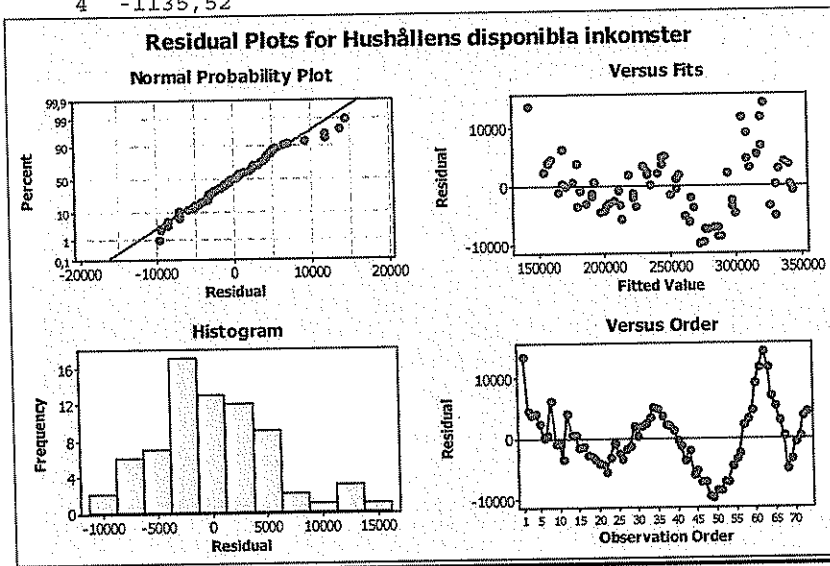
Additive Model

Data Hushållens disponibla inkomster  
 Length 73  
 NMissing 0

Fitted Trend Equation  
 $Y_t = 147508 + 2679 * t$

Seasonal Indices

Period	Index
1	-7283,02
2	6593,86
3	1824,67
4	-1135,52



## Modell 2

### Regression Analysis: Hushällens dispo versus kv2; kv3; kv4; tid

The regression equation is

$$\text{Hushällens disponibla inkomster} = 140745 + 13090 \text{ kv2} + 8259 \text{ kv3} + 5702 \text{ kv4} + 2678 \text{ tid}$$

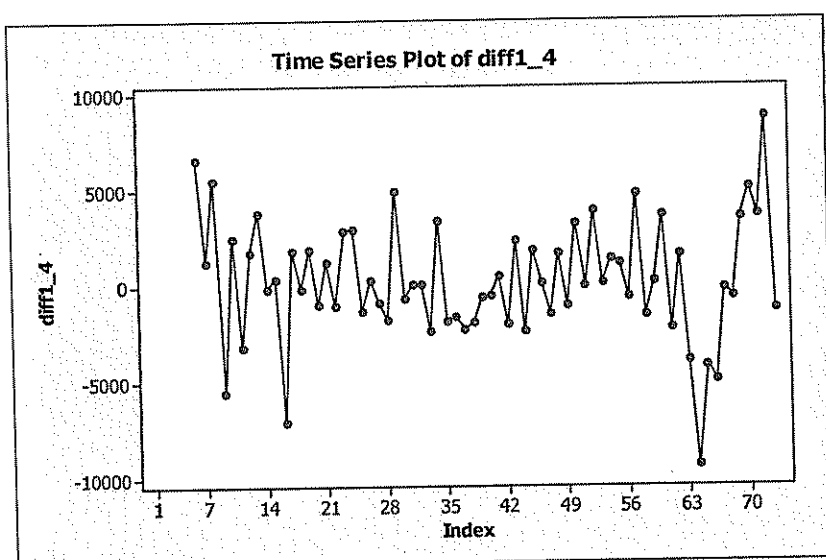
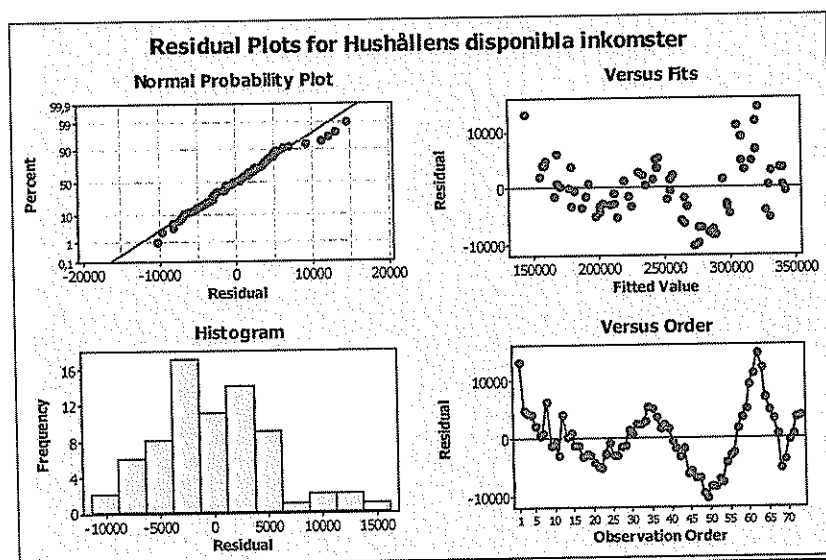
Predictor	Coef	SE Coef	T	P
Constant	140745	1657	84,96	0,000
kv2	13090	1769	7,40	0,000
kv3	8259	1769	4,67	0,000
kv4	5702	1769	3,22	0,002
tid	2678,35	29,88	89,62	0,000

S = 5377,16    R-Sq = 99,2%    R-Sq(adj) = 99,1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	2,33447E+11	58361829379	2018,47	0,000
Residual Error	68	1966141149	28913840		
Total	72	2,35413E+11			

Durbin-Watson statistic = 0,258596



### Modell 3

#### ARIMA Model: Hushållens disponibla inkomster

Unable to reduce sum of squares any further

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
SAR 4	0,2287	0,1662	1,38	0,173
SMA 4	0,8707	0,1084	8,03	0,000
Constant	81,09	48,85	1,66	0,102

Differencing: 1 regular, 1 seasonal of order 4

Number of observations: Original series 73, after differencing 68

Residuals: SS = 449299813 (backforecasts excluded)  
MS = 6912305 DF = 65

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

Lag	12	24	36	48
Chi-Square	14,4	22,3	38,8	44,2
DF	9	21	33	45
P-Value	0,107	0,382	0,224	0,504

### Modell 4

#### ARIMA Model: Hushållens disponibla inkomster

Unable to reduce sum of squares any further

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
SMA 4	0,8050	0,0736	10,94	0,000
Constant	91,30	74,71	1,22	0,226

Differencing: 1 regular, 1 seasonal of order 4

Number of observations: Original series 73, after differencing 68

Residuals: SS = 458964898 (backforecasts excluded)  
MS = 6954014 DF = 66

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

Lag	12	24	36	48
Chi-Square	17,0	27,6	44,8	50,6
DF	10	22	34	46
P-Value	0,075	0,188	0,102	0,296

### Modell 5

#### ARIMA Model: Hushållens disponibla inkomster

Unable to reduce sum of squares any further

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
MA 1	-0,0119	0,1241	-0,10	0,924
SMA 4	0,8022	0,0743	10,80	0,000
Constant	92,01	76,32	1,21	0,232

Differencing: 1 regular, 1 seasonal of order 4

Number of observations: Original series 73, after differencing 68

Residuals: SS = 458838843 (backforecasts excluded)  
MS = 7059059 DF = 65

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

Lag	12	24	36	48
Chi-Square	16,8	27,3	44,7	50,4
DF	9	21	33	45
P-Value	0,051	0,161	0,085	0,267

---

## Data och prognoser från modell 4

Forecasts from period 73

Period	Forecast	95% Limits		Actual
		Lower	Upper	
74	357528	352358	362698	
75	355662	348351	362973	

diff1_4	RESI1	inkomst	tid
-9378	-6061,13	324719	64
-4181	-3101,63	319592	65
-4935	-3355,14	333811	66
-171	-3116,58	329046	67
-630	-5600,40	323315	68
3485	896,94	321673	69
5036	2243,86	340928	70
3664	1063,90	339827	71
8770	4170,47	342866	72
-1236	-605,28	339988	73