

TENTAMEN I STATISTISK ANALYS AV SAMHÄLLSDATA, 2016-08-11

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

I tabellen nedan visas viss data över befolkningen i Linköping år 2005.

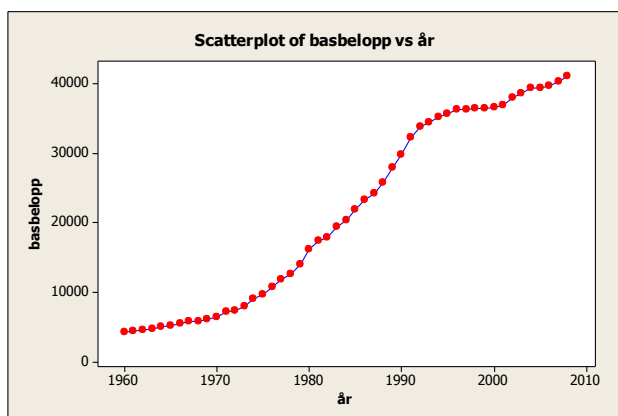
Åldersintervall	Totala antalet kvinnor i intervallet	Totala antalet levande födda barn i intervallet
15-19	4217	7
20-24	5799	75
25-29	5052	214
30-34	4552	336
35-39	4462	157
40-44	4438	22

Inga barn föddes av kvinnor äldre än 44 år. Totalt fanns det 69005 män och 68270 kvinnor.

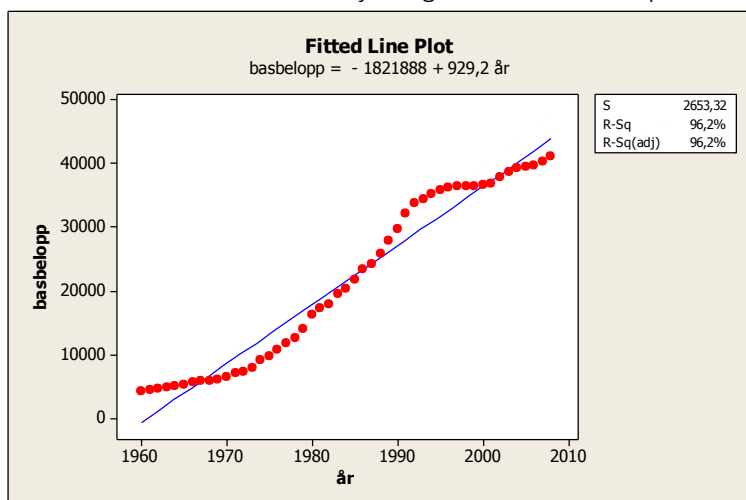
- a) Beräkna det allmänna födelsetalet. 0.5p
- b) Beräkna den generella fruktsamhetskvoten 0.5p
- c) Beräkna den summerade fruktsamhetskvoten och tolka den. 2p
- d) Beräkna bruttoreproduktionen. 1p

2

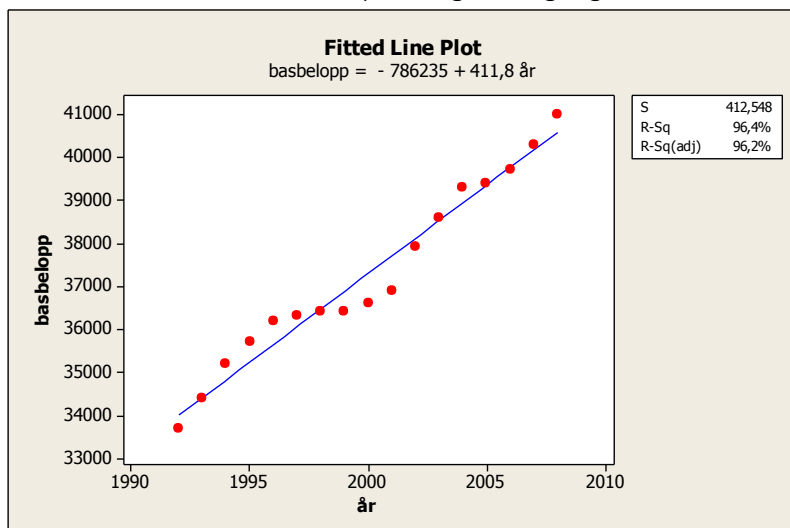
I grafen nedan ser vi basbeloppet i kr i Sverige mellan 1960 och 2008.



Till dessa data har en enkel linjär regressionsmodell anpassats vilket visas i figuren nedan.



Eftersom detta inte alls ser ut att vara en bra modell så har endast observationerna från 1992 och framåt använts och då blev anpassningen enligt figuren nedan.



Regression Analysis: basbelopp versus år

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	69176471	69176471	406,45	0,000
Residual Error	15	2552941	170196		
Total	16	71729412			

Term	Coef	SE Coef
Constant	-786235	40848
år	411,76	20,42

S = 412,548 R-Sq = 96,4% R-Sq(adj) = 96,2%

Durbin-Watson statistic = 0,552711

The regression equation is
basbelopp = - 786235 + 412 år

Uppgifter:

- a) Hur mycket har basbeloppet ökat i snitt per år sedan 1992 enligt modellen ovan? 1p
b) Beräkna ett 95% konfidensintervall för regressionskoefficienten i modellen ovan. Tolka intervallet. 2p
c) Pröva om residualerna kan antas vara okorrelerade i modellen ovan. 1p

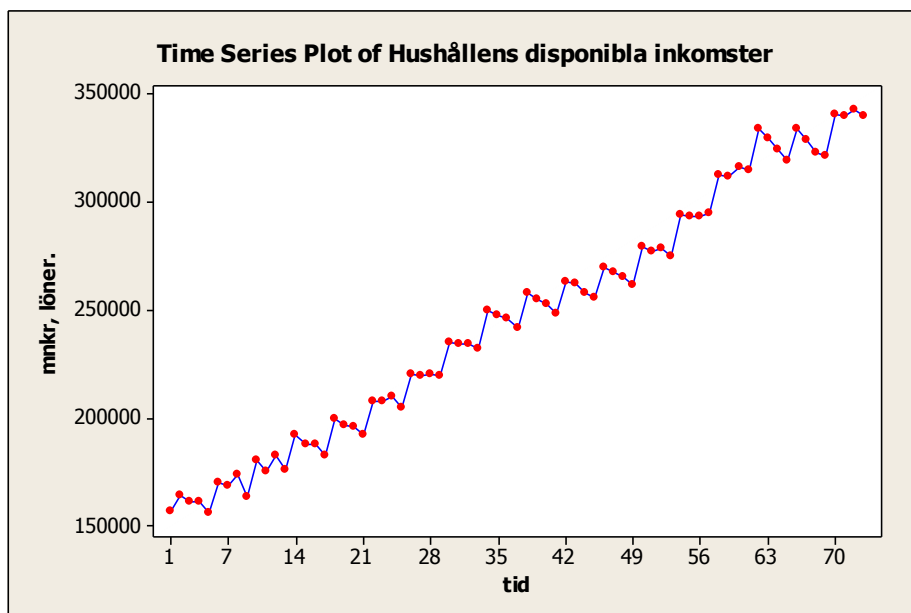
De 10 sista åren är basbeloppet:

36400, 36600, 36900, 37900, 38600, 39300, 39400, 39700, 40300, 41000

- d) Använd dubbel exponentiell utjämning för linjär trend och visa hur de 3 sista åren utjämnas. Du kan använda direkt $l_{2005} = 39109,3$ och $b_{2005} = 286,153$. Gör även prognos för 2009, 2010 samt 2011 då du står vid tidpunkt 2008. Låt $\alpha = \gamma = 0,3$ 3p
e) Beräkna 3 punkters centrerat glidande medelvärden beräknas för åren 2005, 2006 samt 2007. 2p

3

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



Nedan har tre SARIMA-modeller anpassats.

- a) Jämför de tre modellerna noggrant. Vilken anser du vara den bästa modellen? Motivera. 2p
b) Låt y_t vara den disponibla inkomsten. Låt vidare z_t vara diff4_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
c) Modell 2 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

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 2

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 3

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 2

Forecasts from period 73

95% Limits

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

diff1_4	RES11	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