



## Försättsblad till skriftlig tentamen vid Linköpings Universitet

<b>Datum för tentamen</b>	2013-06-11
<b>Sal (1)</b> Om tentan går i flera salar ska du bifoga ett försättsblad till varje sal och <u>ringa in</u> vilken sal som avses	TER3
<b>Tid</b>	8-12
<b>Kurskod</b>	732G05
<b>Provkod</b>	TENB
<b>Kursnamn/benämning</b> <b>Provnamn/benämning</b>	Regressions- och tidsserieanalys Tentamen
<b>Institution</b>	IDA
<b>Antal uppgifter som ingår i tentamen</b>	4
<b>Jour/Kursansvarig</b> Ange vem som besöker salen	Tommy Schyman Patrik Waldmann
<b>Telefon under skrivtiden</b>	Tommy: 0768 – 30 31 09 Patrik: 072 742 68 70
<b>Besöker salen ca kl.</b>	09:30, 10:30
<b>Kursadministratör/kontaktperson</b> (namn + tfnr + mailaddress)	annelie.almquist@liu.se, tel 282934
<b>Tillåtna hjälpmedel</b>	Valfri räknedosa, kursboken (Bowerman et al) som får innehålla markeringar av text samt flärpar med anteckning. I bilaga 1 finns formler för index.
<b>Övrigt</b>	
<b>Vilken typ av papper ska</b>	Rutigt

# Tentamen

Linköpings Universitet, Institutionen för datavetenskap, Statistik

---

Kurskod och namn:	732G05 Regressions- och tidsserieanalys
Datum och tid:	2013-06-11, 08-12
Jourhavande lärare:	Tommy Schyman och Patrik Waldmann
Tillåtna hjälpmedel:	Valfri räknedosa, kursboken (Bowerman et al) som får innehålla markeringar av text samt flärpar med anteckning. I bilaga 1 finns formler för index.
Betygsgränser:	Tentamen omfattar totalt 20p. Godkänt från och med 12p, väl godkänt från och med 16p.

---

Redovisa och motivera tydligt alla dina lösningar!

## Uppgift 1 (7p)

Vid ersättning för mark i samband med vägbyggen får markägaren ersättning per kvadratmeter men ersättningsnivån skiljer sig mellan olika län. Nedan följer några exempel på ersättning ( $y$ ) för olika stora ytor ( $x$ ) från 8 länsstyrelser. Din uppgift är att:

- Skatta parametrarna i en enkel linjär regressionsmodell för  $y$  och  $x$ . (3p)
- Testa med t-test om regressionsparametrarna är signifikanta på 0.05-nivån. (2p)
- Beräkna prediktionen och 95 % prediktionsintervall för en yta på 550 kvm. (2p)

Ersättning (tSEK)	80.2	4.1	17.3	50.5	53.7	4.9	113.5	65.3
Yta (kvm)	892	126	218	314	593	234	1087	357

## Uppgift 2 (5p)

Ett amerikanskt företag utförde en undersökning av hur fem variabler påverkade 290 slumpmässigt utvalda individers vilja att skänka pengar (dollar) till välgörenhet.

- a) Din uppgift är att utföra ett partiellt F-test mellan den hela modellen och den reducerade modellen baserat på utskrifterna nedan. (3p)

### Regression Analysis: dollar versus utb; ink; lån; stadstrl; antbarn

The regression equation is

$$\text{dollar} = 1,70 + 0,892 \text{ utb} + 0,847 \text{ ink} + 0,857 \text{ lån} + 0,966 \text{ stadstrl} + 0,015 \text{ antbarn}$$

Predictor	Coef	SE Coef	T	P
Constant	1,7013	0,1203	14,15	0,000
utb	0,8919	0,1313	6,79	0,000
ink	0,8469	0,1258	6,73	0,000
lån	0,8568	0,1159	7,39	0,000
stadstrl	0,9665	0,1170	8,26	0,000
antbarn	0,0147	0,1190	0,12	0,901

S = 2,04323    R-Sq = 48,7%    R-Sq(adj) = 47,8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	1127,24	225,45	54,00	0,000
Residual Error	284	1185,64	4,17		
Total	289	2312,87			

### Regression Analysis: dollar versus utb; ink; lån; stadstrl

The regression equation is

$$\text{dollar} = 1,70 + 0,892 \text{ utb} + 0,848 \text{ ink} + 0,858 \text{ lån} + 0,966 \text{ stadstrl}$$

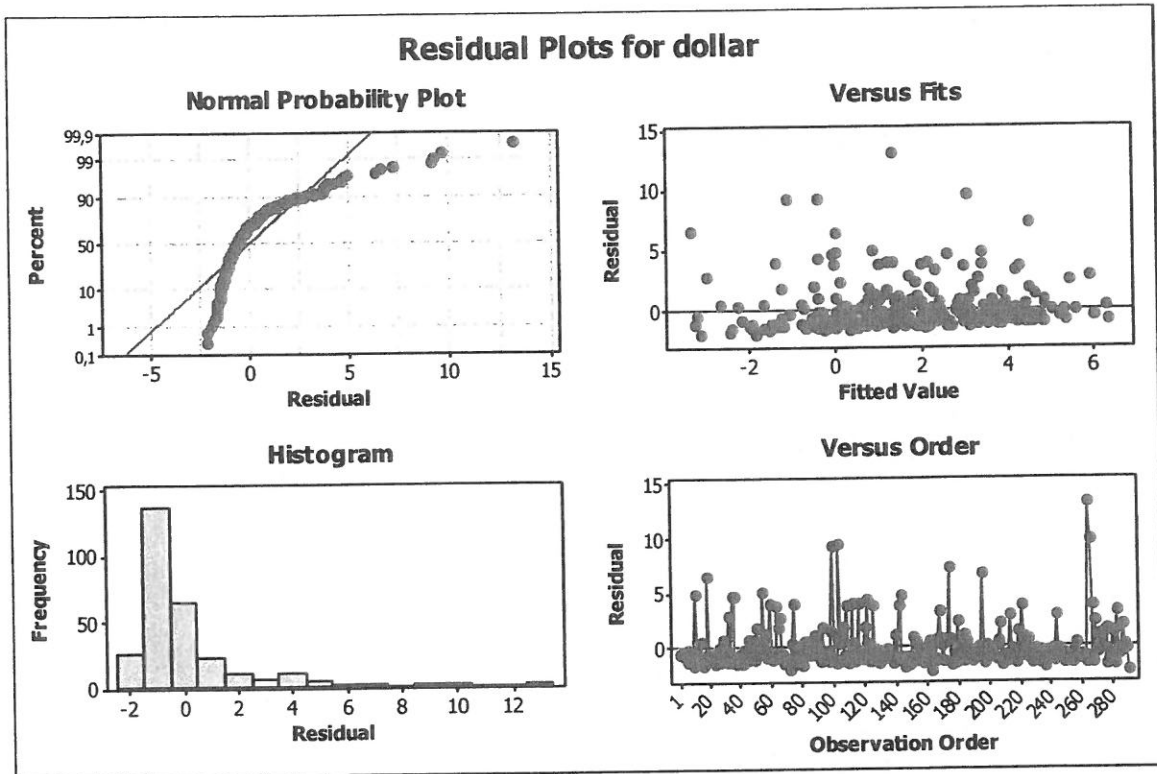
Predictor	Coef	SE Coef	T	P
Constant	1,7010	0,1200	14,17	0,000
utb	0,8919	0,1310	6,81	0,000
ink	0,8480	0,1252	6,77	0,000
lån	0,8582	0,1151	7,46	0,000
stadstrl	0,9660	0,1168	8,27	0,000

S = 2,03970    R-Sq = 48,7%    R-Sq(adj) = 48,0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	1127,17	281,79	67,73	0,000
Residual Error	285	1185,70	4,16		
Total	289	2312,87			

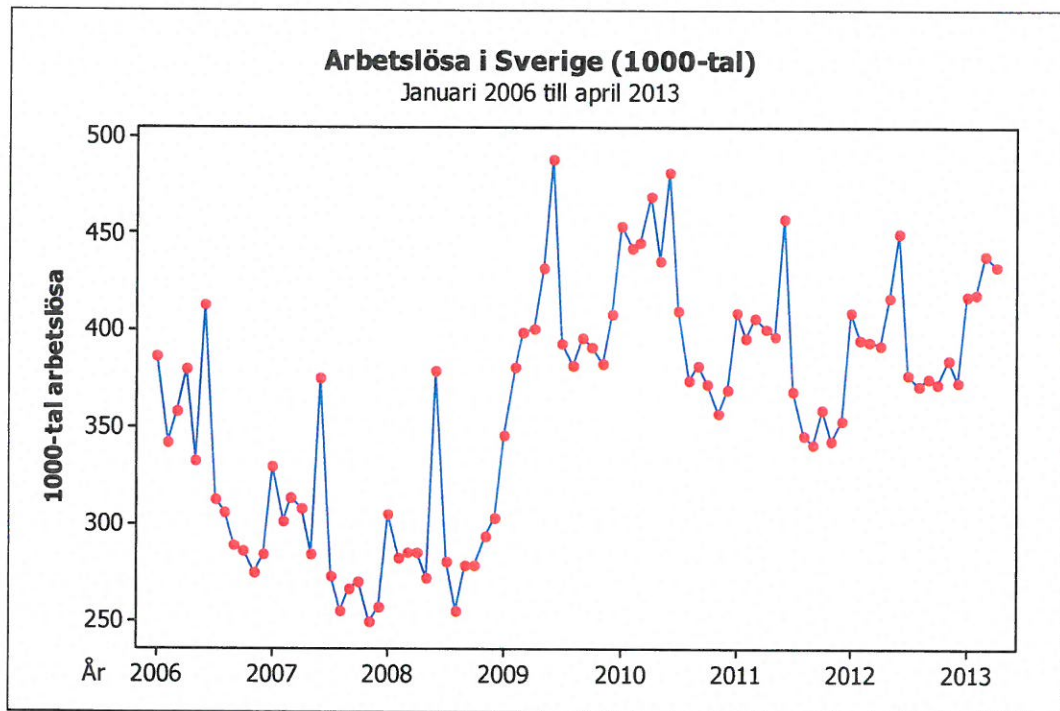
b) Nedan följer fyra residualplottar från analysen av den hela modellen. Tolka plottarna. (2p)





### Uppgift 3 (5p)

Nedan visas en tidsserie över tusentals arbetslösa i åldrarna 15-74 år i Sverige från januari 2006 till april 2013 (månadsdata), samt utskrift från multiplikativ och additiv komponentuppdelning på den aktuella tidsserien. Uppgifter finns på nästa sida.



Data 1000-tal arbetslösa  
Length 88  
NMissing 0

Fitted Trend Equation

$$Y_t = 297,09 + 1,43 * t$$

Seasonal Indices

Period	Index
1	24,2894
2	9,5561
3	18,0498
4	14,8352
5	14,4207
6	79,9915
7	-12,5939
8	-32,5148
9	-24,8481
10	-29,1023
11	-36,7398
12	-25,3439

Accuracy Measures

MAPE	8,90
MAD	30,95
MSD	1406,32

Data 1000-tal arbetslösa  
Length 88  
NMissing 0

Fitted Trend Equation

$$Y_t = 296,65 + 1,44 * t$$

Seasonal Indices

Period	Index
1	1,07235
2	1,02513
3	1,05157
4	1,04404
5	1,03609
6	1,21587
7	0,95941
8	0,91103
9	0,93212
10	0,92957
11	0,89645
12	0,92638

Accuracy Measures

MAPE	9,01
MAD	31,39
MSD	1439,00

- a) Tolka säsongskomponenten för april för båda komponentuppdelningarna. (1p)
- b) Motivera vilken av de två komponentuppdelningarna som bör användas för den aktuella tidsserien. (1p)
- c) Använd den valda komponentuppdelningen och beräkna prognoser för de två kommande månaderna (maj 2013 och juni 2013). (1p)

Nedan visas utskrifter där tidsserieregression samt Holt-Winters metod har tillämpats på tidsserien. Uppgift kommer längst ner på sidan.

Predictor	Coef	SE Coef	T	P
Constant	267,72	17,37	15,42	0,000
Tid	1,4252	0,1705	8,36	0,000
Januari	53,77	20,99	2,56	0,012
Februari	40,17	20,98	1,91	0,059
Mars	48,86	20,98	2,33	0,023
April	50,98	20,97	2,43	0,017
Maj	41,76	21,69	1,93	0,058
Juni	108,11	21,68	4,99	0,000
Juli	16,74	21,67	0,77	0,442
Augusti	-3,01	21,67	-0,14	0,890
September	1,29	21,66	0,06	0,953
Oktober	0,32	21,66	0,01	0,988
November	-7,60	21,66	-0,35	0,727

S = 40,5187    R-Sq = 61,7%    R-Sq(adj) = 55,6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	12	198581	16548	10,08	0,000
Residual Error	75	123132	1642		
Total	87	321713			

Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
Maj 2013	436,32	17,37	(401,73; 470,92)	(348,50; 524,14)
Jun 2013	502,67	17,29	(468,23; 537,10)	(414,91; 590,42)

Data 1000-tal arbetslösa  
Length 88

Smoothing Constants

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

Accuracy Measures

MAPE 4,715  
MAD 17,464  
MSD 529,598

Forecasts

Period	Forecast	Lower	Upper
89	434,553	391,767	477,339
90	512,213	468,756	555,669

- d) Vilken av dessa tre modeller (den valda komponentuppdelningen, tidsserieregressionen och Holt-Winters) är den bästa för denna tidsserie? (2p)

Ledning: Kolla på prognoser, jämförelsemått och tidsseriens utseende.

### Uppgift 4 (3p)

Nedan visas medianvärdet för den disponibla inkomsten i tusentals kronor per person i Sverige mellan åren 2007 och 2011. I tabellen finns även KPI för de aktuella åren.

År	Medianinkomst	KPI (basår = 1980)
2007	203,0	290,51
2008	203,4	300,61
2009	207,6	299,66
2010	210,4	303,46
2011	215,2	311,43

- a) Beräkna ett enkelt index för medianinkomstens utveckling under de aktuella åren. (1p)
- b) Räkna om medianinkomsten till fasta priser (2007 års priser) med hjälp av KPI. (2p)

## Bilaga 1

### Formelsamling för index

#### Sammanstatta fastbasindex:

$$I_t = i_{1,t} * w_1 + i_{2,t} * w_2 + \dots + i_{n,t} * w_n$$

där  $n$  är antalet ingående varor/tjänster,  $i_{1,t}, \dots, i_{n,t}$  är enkla prisindex för ingående varor, alla med basår  $t_0$  och  $w_1, \dots, w_n$  väljs enligt ett viktsystem:

$$\text{Laspeyre: } \frac{p_{i,t_0} * q_{i,t_0}}{\sum p_{j,t_0} * q_{j,t_0}}$$

$$\text{Paasche: } \frac{p_{i,t_0} * q_{i,t}}{\sum p_{j,t_0} * q_{j,t}}$$

#### Kedjeprisindex:

$$I_t = L_{0,1} * L_{1,2} * \dots * L_{t-1,t} * 100$$

där

$$L_{t-1,t} = \sum_{i=1}^n \frac{p_{i,t}}{p_{i,t-1}} * w_{i,t-1,t}$$

är årslänken från år  $t-1$  till  $t$  för  $n$  ingående varor/tjänster.  $w_{i,t-1,t}$  väljs enligt ett viktsystem:

$$\text{Laspeyre: } w_{i,t-1,t}^L = \frac{\text{Försäljningsvärdet för vara } i \text{ år } t-1}{\text{Totala försäljningsvärdet år } t-1}$$

$$\text{Paasche: } w_{i,t-1,t}^P = \frac{\text{Försäljningsvärdet för vara } i \text{ år } t \text{ i priser för år } t-1}{\text{Totala försäljningsvärdet år } t \text{ i priser för år } t-1}$$

Med representantvaror byts "Försäljningsvärdet för vara  $i$ " mot "Försäljningsvärdet för varugrupp  $i$ " i viktorna.

#### Implicitprisindex:

$$I_t = \frac{\text{Försäljningsvärdet av varan/tjänsten/gruppen år } t \text{ i löpande priser}}{\text{Försäljningsvärdet av varan/tjänsten/gruppen år } t \text{ i fasta priser}} * 100$$

#### Relativprisindex:

$$I_t^R = \frac{\text{Prisindex för aktuell vara/tjänst/grupp}}{\text{Prisindex för den större jämförelsegruppen, t. ex. KPI}} * 100$$