

EXP



## Försättsblad till skriftlig tentamen vid Linköpings Universitet

<b>Datum för tentamen</b>	2009-12-17
<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	TER2
<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	Lotta Hallberg
<b>Telefon under skrivtiden</b>	
<b>Besöker salen ca kl.</b>	Kl 10,00
<b>Kursadministratör/kontaktperson</b> (namn + tfnr + mailaddress)	Carita Lilja, 28 1463, carli@ida.liu.se
<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. Formelsamling. Tabeller
<b>Övrigt</b>	G=12, VG=16
<b>Vilken typ av papper ska användas, rutigt eller linjerat</b>	
<b>Antal exemplar i påsen</b>	38



## Tentamen i Regressions- och tidsserieanalys, 2009-12-17

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

### 1

Följande data erhöles i en studie där man mätte det diastoliska blodtrycket hos 8 pojkar. Även deras ålder antecknades.

Resultat:

Obs nr	1	2	3	4	5	6	7	8
Blodtryck	63	67	74	64	75	69	90	60
Ålder	5	8	11	7	13	12	12	6

Den enkla linjära regressionsmodellen  $\text{blodtryck} = \beta_0 + \beta_1 \text{ålder} + \varepsilon$  ska anpassas.

- a) Använd minsta kvadratmetoden (least squares estimate) för att skatta  $\beta_0$  och  $\beta_1$ . Tolka  $b_1$ , dvs skattningen på  $\beta_1$ . 2p

Nedan visas ANOVA-tabellen för dessa data.

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	367,50	367,50	8,23	0,028
Residual Error	6	268,00	44,67		
Total	7	635,50			

- b) Använd data i tabellen ovan och beräkna förklaringsgraden i modellen samt tolka den. 1p
- c) Den 7:e pojken har ett väldigt högt diastoliskt blodtryck. Nedan har modellen anpassats igen med denna observation borttagen. Beräkna ett 95% prediktionsintervall för blodtrycket för en 12 årig pojke. Ligger den borttagna observationen inom intervallet? 3p

### Regression Analysis: blodtryck versus ålder

The regression equation is  
 $\text{blodtryck} = 53,1 + 1,62 \text{ålder}$

Predictor	Coef	SE Coef	T	P
Constant	53,068	3,214	16,51	0,000
ålder	1,6214	0,3448	4,70	0,005

S = 2,64538 R-Sq = 81,6% R-Sq(adj) = 77,9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	154,72	154,72	22,11	0,005
Residual Error	5	34,99	7,00		
Total	6	189,71			

## 2

Nedan visas en tabell med konsumentprisindex för första halvåret 2009.

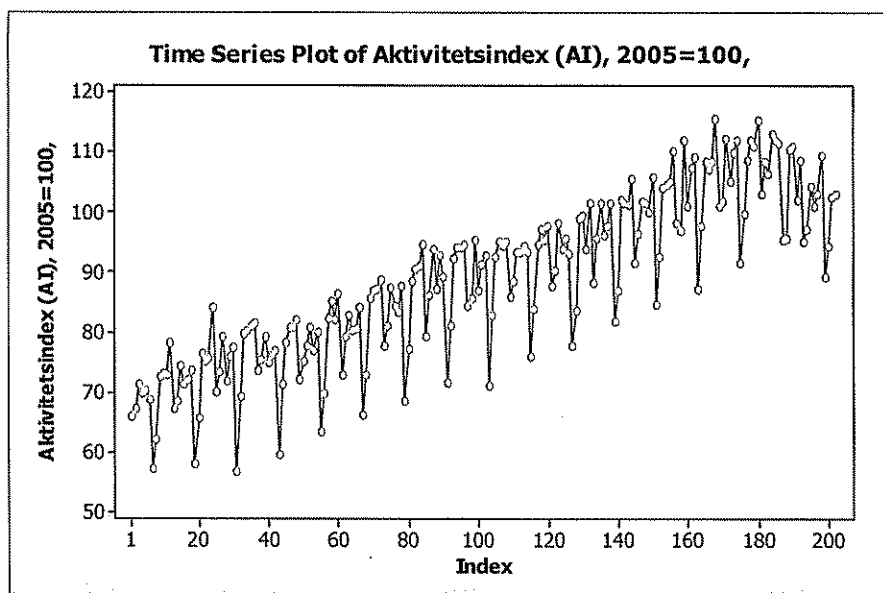
Månad 2009	Jan	Feb	Mar	Apr	Maj	Jun	Jul
KPI basår jan 1980	297,9	297,9	298,8	299,3	299,5	300,2	298,8

- Förklara i korta drag vad KPI är och vad det mäter. 1p
- Byt basår till januari 2009 i KPI-serien ovan. 1p
- Hur stor procentuell förändring har skett i KPI-serien ovan från januari 2009 till juli 2009? 1p

## 3

Nedan visas en tidsserie över aktivitetsindex Månad 1993M01-2009M10. Följande beskrivning är citerat direkt från SCBs hemsida.

*Aktivitetsindex är en indikator som ger ett mått på aktiviteten i den svenska ekonomin varje månad; dvs. produktion av varor och av tjänster inom både privat och offentlig sektor. Aktivitetsindex är en modellbaserad beräkning med bruttonationalprodukten (BNP) i fasta priser som målvariabel. SCB publicerar sedan januari 1998 Aktivitetsindex månadsvis.*



De sista 13 värdena i serien är:

2008M10	111,30
2008M11	102,19
2008M12	108,81
2009M01	95,35
2009M02	97,36
2009M03	104,48
2009M04	101,26
2009M05	103,11
2009M06	109,73
2009M07	89,31
2009M08	94,62
2009M09	102,77
2009M10	103,17

Nedan har två olika modeller anpassats till data.

## Modell 1

### Time Series Decomposition for Aktivitetsindex (AI), 2005=100,

Additive Model

Data Aktivitetsindex (AI), 2005=100,  
Length 202  
NMissing 0

Fitted Trend Equation

$$Y_t = 68,054 + 0,205 \cdot t$$

Seasonal Indices

Period	Index
1	-4,5723
2	-1,8167
3	3,9233
4	0,2981
5	1,5575
6	3,6517
7	-15,3402
8	-7,4485
9	3,6656
10	5,0910
11	3,6017
12	7,3887

Accuracy Measures

MAPE 2,25090  
MAD 2,05615  
MSD 8,23806

## Modell 2

### Regression Analysis: Aktivitetsindex versus t; mån\_2; ...

The regression equation is

$$\begin{aligned} \text{Aktivitetsindex (AI), 2005=100,} &= 63,4 + 0,205 t + 2,89 \text{ mån}_2 + 8,86 \text{ mån}_3 \\ &+ 5,19 \text{ mån}_4 + 6,57 \text{ mån}_5 + 7,97 \text{ mån}_6 \\ &- 10,6 \text{ mån}_7 - 3,05 \text{ mån}_8 + 8,05 \text{ mån}_9 \\ &+ 8,98 \text{ mån}_{10} + 8,28 \text{ mån}_{11} + 12,1 \text{ mån}_{12} \end{aligned}$$

Predictor	Coef	SE Coef	T	P
Constant	63,4335	0,7948	79,81	0,000
t	0,204736	0,003564	57,45	0,000
mån_2	2,893	1,012	2,86	0,005
mån_3	8,862	1,012	8,76	0,000
mån_4	5,190	1,012	5,13	0,000
mån_5	6,565	1,012	6,49	0,000
mån_6	7,967	1,012	7,87	0,000
mån_7	-10,573	1,012	-10,44	0,000
mån_8	-3,046	1,012	-3,01	0,003
mån_9	8,047	1,012	7,95	0,000
mån_10	8,984	1,013	8,87	0,000
mån_11	8,278	1,028	8,05	0,000
mån_12	12,102	1,028	11,77	0,000

S = 2,95070 R-Sq = 95,7% R-Sq(adj) = 95,4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	12	36378,3	3031,5	348,19	0,000
Residual Error	189	1645,5	8,7		
Total	201	38023,9			

a) Tolka säsongseffekten för juli månad för de båda modellerna.

1p

b) Gör prognos för AI för november och december 2009 med båda modellerna.

1p

Nedan har ytterligare en modell anpassats. Två prognoser har också beräknats.

- c) Ta hjälp av grafen ovan, prognoserna samt de jämförande måtten och avgör vilken av de tre modellerna som är bäst? 2p

### Modell 3

#### Winters' Method for Aktivitetsindex (AI), 2005=100,

Additive Method

Data Aktivitetsindex (AI), 2005=100,  
Length 202

Smoothing Constants

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

Accuracy Measures

MAPE 1,95589  
MAD 1,71860  
MSD 5,31575

Forecasts

Period	Forecast	Lower	Upper
203	101,664	97,453	105,874
204	106,332	102,056	110,609

## 4

Nedan har ett stort datamaterial från USA analyserats. Responsvariabel (dependent variable) Y är antal grova brott per 100 000 invånare = Krim/befolk.

Data är obeserverat i 440 'kommuner'. De förklarande variablerna (independent variables) är:

AndelUnga = Procent av pop mellan 18 o 34 år.

AndelGamla = Procent av pop över 65 år

Antalläkare = Antal praktiserande läkare

AndelGymn = Procent med gymnasial utbildning

AndelKandidat = Procent med kandidatexamen

AndelFattiga = Procent med inkomst under existensminimum

AndelArblösa = Procent arbetslösa

Täthet = landarea per 100 000 invånare

Reg1-Reg3 = tre dummies som delar in USA i fyra stora delområden.

Tre modeller visas nedan. Uppgifterna kommer därefter.

#### Descriptive Statistics: AndelUnga; AndelGamla; Antalläkare; AndelGymn; ...

Variable	Mean	StDev	Minimum	Maximum
AndelUnga	28,568	4,191	16,400	49,700
AndelGamla	12,170	3,993	3,000	33,800
Antalläkare	988,0	1789,7	39,0	23677,0
AndelGymn	77,561	7,015	46,600	92,900
AndelKandidat	21,081	7,655	8,100	52,300
AndelFattiga	8,721	4,657	1,400	36,300
AndelArblösa	6,597	2,338	2,200	21,300
Täthet	476,0	697,2	3,1	7542,2
Krim/befolk	5729	2733	460	29599

## Modell 1

### Regression Analysis: Krim/befolk versus AndelUnga; AndelGamla; ...

The regression equation is

$$\begin{aligned} \text{Krim/befolk} = & 267 + 69,6 \text{ AndelUnga} + 67,9 \text{ AndelGamla} + 0,311 \text{ AntalLäkare} \\ & + 1,8 \text{ AndelGymn} + 40,1 \text{ AndelKandidat} + 260 \text{ AndelFattiga} \\ & + 10,8 \text{ AndelArbLösa} - 0,746 \text{ Täthet} - 1905 \text{ Reg1} - 997 \text{ Reg2} \\ & + 185 \text{ Reg3} \end{aligned}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	267	2726	0,10	0,922	
AndelUnga	69,64	33,27	2,09	0,037	1,985
AndelGamla	67,91	32,98	2,06	0,040	1,770
AntalLäkare	0,31118	0,06286	4,95	0,000	1,292
AndelGymn	1,79	30,05	0,06	0,953	4,536
AndelKandidat	40,12	22,78	1,76	0,079	3,103
AndelFattiga	259,56	34,32	7,56	0,000	2,608
AndelArbLösa	10,85	58,45	0,19	0,853	1,906
Täthet	-0,7456	0,1709	-4,36	0,000	1,449
Reg1	-1904,9	370,0	-5,15	0,000	2,511
Reg2	-996,8	344,8	-2,89	0,004	2,253
Reg3	185,4	355,3	0,52	0,602	2,920

S = 2073,79 R-Sq = 43,9% R-Sq(adj) = 42,4%

#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	11	1437818922	130710811	30,39	0,000
Residual Error	428	1840650302	4300585		
Total	439	3278469225			

## Modell 2

### Regression Analysis: Krim/befolk versus AntalLäkare; AndelKandidat; ...

The regression equation is

$$\begin{aligned} \text{Krim/befolk} = & 1370 + 0,289 \text{ AntalLäkare} + 67,9 \text{ AndelKandidat} + 338 \text{ AndelFattiga} \\ & - 0,642 \text{ Täthet} \end{aligned}$$

Predictor	Coef	SE Coef	T	P
Constant	1369,9	475,0	2,88	0,004
AntalLäkare	0,28883	0,06290	4,59	0,000
AndelKandidat	67,89	15,79	4,30	0,000
AndelFattiga	338,02	25,65	13,18	0,000
Täthet	-0,6418	0,1608	-3,99	0,000

S = 2203,05 R-Sq = 35,6% R-Sq(adj) = 35,0%

#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	1167235097	291808774	60,12	0,000
Residual Error	435	2111234128	4853412		
Total	439	3278469225			

## Modell 3

### Regression Analysis: Krim/befolk versus AntalLäkare; AndelKandidat; ...

The regression equation is

$$\begin{aligned} \text{Krim/befolk} = & 2969 + 0,324 \text{ AntalLäkare} + 46,4 \text{ AndelKandidat} + 268 \text{ AndelFattiga} \\ & - 0,731 \text{ Täthet} - 1726 \text{ Reg1} - 916 \text{ Reg2} + 297 \text{ Reg3} \end{aligned}$$

Predictor	Coef	SE Coef	T	P
Constant	2969,3	565,2	5,25	0,000
AntalLäkare	0,32445	0,06091	5,33	0,000
AndelKandidat	46,42	15,27	3,04	0,003
AndelFattiga	267,65	26,25	10,19	0,000
Täthet	-0,7307	0,1702	-4,29	0,000
Reg1	-1726,0	342,4	-5,04	0,000
Reg2	-916,2	341,0	-2,69	0,008
Reg3	297,4	326,5	0,91	0,363

S = 2077,66 R-Sq = 43,1% R-Sq(adj) = 42,2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	1413668547	201952650	46,78	0,000
Residual Error	432	1864800678	4316668		
Total	439	3278469225			

- a) Studera modell 1. Förklara kort vad multikollinjäritetsproblem innebär och vilka konsekvenser det kan få och hur det mäts. Har vi denna typ av problem i modell 1? 2p
- b) Modell 2 är resultatet av Bakåtelimineringsmetoden då Reg1-Reg3 inte var med. Förklara hur denna metod går till. 2p
- c) Pröva med ett test om Reg1-Reg3 kan läggas till modell 2. 2p
- d) Prediktera Krim/befolk med hjälp av modell 3 då andelen fattiga är 5%, antal läkare är 2000, andel med kandidat-examen är 30%, tätheten är 90 och Reg3, dvs delområde 3. 1p