



Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2012-11-30
Sal (1) Om tentan går i flera salar ska du bifoga ett försättsblad till varje sal och <u>ringa in</u> vilken sal som avses	TER2
Tid	14-18
Kurskod	732G05
Provkod	TENB
Kursnamn/benämning Provnamn/benämning	Regressions- och tidsserieanalys Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	4
Jour/Kursansvarig Ange vem som besöker salen	Tommy Schyman, Patrik Waldmann
Telefon under skrivtiden	Tommy: 076- 8303109 Patrik: 072-7426870
Besöker salen ca kl.	15:15, 16:30
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	carita.lilja@liu.se tel 1463
Tillåtna hjälpmedel	Valfri räknedosa, kursboken (Bowerman et al) som får innehålla markeringar av text samt flärpar med anteckning.
Övrigt	I bilaga 1 finns formler för index.
Vilken typ av papper ska användas, rutigt eller linjerat	Rutigt
Antal exemplar i påsen	

Tentamen

Linköpings Universitet, Institutionen för datavetenskap, Statistik

- Kurskod och namn: 732G05 Regressions- och tidsserieanalys
- Datum och tid: 2012-11-30, 14-18
- 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 (6p)

Nedan visas hur kostnaden i dollar (y) förhåller sig till antalet fel som behöver rättas i ett dokument (x) för ett företag som ägnar sig åt korrekturrättning. Din uppgift är att:

- Anpassa en linjär regressionsmodell. (2p)
- Beräkna ett 95% prediktionsintervall för 10 fel. (2p)
- Beräkna residualerna och plotta dessa (ungefärligt) mot de sorterade x -värdena. Vilka slutsatser kan dras från residualplotten? (2p)

Index	Y	X
1	128	7
2	213	12
3	191	10
4	178	10
5	250	14
6	446	25
7	540	30
8	457	25
9	324	18
10	177	10
11	75	4
12	107	6

Uppgift 2 (5p)

Ett konsultföretag har fått i uppgift att undersöka hur olika faktorer påverkar priset hos amerikanska begagnade bilar som tillverkats 2005. Följande variabler ingick i studien:

- Price: responsvariabeln pris.
- Mileage: antalet miles som bilen körts.
- Make: fem indikatorvariabler för tillverkare (Saturn, Pontiac, and Chevrolet etc.).
- Cylinder: antalet cylindrar.
- Liter: storleken på motorn i liter.
- Doors: antalet dörrar.
- Cruise: indikator variabel för automatisk farthållare (1 = cruise)
- Sound: indikator variabel för uppgraderad ljudutrustning (1 = upgraded)
- Leather: indikator variabel för läderklädsel (1 = leather)

Först utförde man en multipel regression med samtliga förklarande variabler i Minitab och erhöll utskriften nedan.

a) Förklara vad VIF är och hur man ska tolka VIF-värdena i utskriften.

(1p)

Regression Analysis: Price versus Mileage; Cylinder; ...

The regression equation is

Price = 15717 - 0,18035 Mileage - 215 Cylinder + 4760 Liter - 1710 Doors
- 504 Cruise - 44 Sound + 25 Leather + 16272 Make_Cadillac
- 2237 Make_Chevrolet - 1779 Make_Pontiac + 14639 Make_SAAB
- 2312 Make_Saturn

Predictor	Coef	SE Coef	T	P	VIF
Constant	15717	1290	12,19	0,000	
Mileage	-0,18035	0,01497	-12,05	0,000	1,007
Cylinder	-215,0	404,0	-0,53	0,595	21,033
Liter	4760,4	460,0	10,35	0,000	17,307
Doors	-1710,1	155,2	-11,01	0,000	1,166
Cruise	-503,8	346,7	-1,45	0,147	1,500
Sound	-44,1	277,3	-0,16	0,874	1,123
Leather	25,5	299,8	0,08	0,932	1,204
Make_Cadillac	16271,6	662,4	24,56	0,000	2,635
Make_Chevrolet	-2237,3	483,9	-4,62	0,000	3,760
Make_Pontiac	-1779,0	496,5	-3,58	0,000	2,507
Make_SAAB	14639,4	589,6	24,83	0,000	2,835
Make_Saturn	-2311,6	655,5	-3,53	0,000	1,988

S = 3463,85 R-Sq = 87,9% R-Sq(adj) = 87,7%

Efter detta utförde man tre olika metoder för modelljämförelse.

b) Den första utskriften visar de två bästa modellerna med olika antal variabler. Förklara vad R-sq, R-Sq(adj), Mallows C-p och s står för. Redogör också för vilken modell som är bäst.

(2p)

Stepwise Regression: Price versus Mileage; Cylinder; ...

Backward elimination. Alpha-to-Remove: 0,05

Response is Price on 12 predictors, with N = 804

Step	1	2	3	4	5
Constant	15717	15733	15684	15311	15041
Mileage	-0,180	-0,180	-0,180	-0,180	-0,180
T-Value	-12,05	-12,06	-12,06	-12,07	-12,08
P-Value	0,000	0,000	0,000	0,000	0,000
Cylinder	-215	-219	-214		
T-Value	-0,53	-0,55	-0,54		
P-Value	0,595	0,584	0,592		
Liter	4760	4766	4762	4532	4451
T-Value	10,35	10,49	10,50	30,66	32,37
P-Value	0,000	0,000	0,000	0,000	0,000
Doors	-1710	-1710	-1709	-1726	-1710
T-Value	-11,01	-11,02	-11,03	-11,37	-11,28
P-Value	0,000	0,000	0,000	0,000	0,000
Cruise	-504	-507	-506	-511	
T-Value	-1,45	-1,47	-1,47	-1,49	
P-Value	0,147	0,141	0,142	0,138	
Sound	-44	-41			
T-Value	-0,16	-0,15			
P-Value	0,874	0,882			
Leather	25				
T-Value	0,08				
P-Value	0,932				
Make_Cadillac	16272	16288	16289	16130	16192
T-Value	24,56	25,79	25,81	28,93	29,10
P-Value	0,000	0,000	0,000	0,000	0,000
Make_Chevrolet	-2237	-2229	-2233	-2214	-2063
T-Value	-4,62	-4,70	-4,72	-4,69	-4,47
P-Value	0,000	0,000	0,000	0,000	0,000
Make_Pontiac	-1779	-1774	-1771	-1776	-1674
T-Value	-3,58	-3,60	-3,60	-3,61	-3,44
P-Value	0,000	0,000	0,000	0,000	0,001
Make_SAAB	14639	14647	14655	14727	14612
T-Value	24,83	25,18	25,31	26,15	26,17
P-Value	0,000	0,000	0,000	0,000	0,000
Make_Saturn	-2312	-2311	-2298	-2260	-2077
T-Value	-3,53	-3,53	-3,54	-3,51	-3,28
P-Value	0,000	0,000	0,000	0,000	0,001
S	3464	3462	3460	3458	3461
R-Sq	87,90	87,90	87,90	87,90	87,87
R-Sq(adj)	87,72	87,74	87,75	87,76	87,74

Stepwise Regression: Price versus Mileage; Cylinder; ...

Forward selection. Alpha-to-Enter: 0,05

Response is Price on 12 predictors, with N = 804

Step	1	2	3	4	5	6
Constant	19178	17250	2723	6163	12124	12865
Make_Cadillac	21758	23686	17152	16979	17597	17322
T-Value	24,83	32,57	33,29	35,33	38,66	36,99
P-Value	0,000	0,000	0,000	0,000	0,000	0,000
Make_SAAB		12245	16455	16688	16476	16065
T-Value		19,62	38,50	41,82	43,90	38,98
P-Value		0,000	0,000	0,000	0,000	0,000
Liter			4800	4819	4635	4582
T-Value			32,80	35,32	35,84	35,00
P-Value			0,000	0,000	0,000	0,000
Mileage				-0,177	-0,179	-0,179
T-Value				-11,02	-11,85	-11,91
P-Value				0,000	0,000	0,000
Doors					-1531	-1590
T-Value					-10,40	-10,69
P-Value					0,000	0,000
Make_Chevrolet						-690
T-Value						-2,38
P-Value						0,018
S	7437	6115	3996	3725	3498	3488
R-Sq	43,47	61,82	83,72	85,87	87,55	87,64
R-Sq(adj)	43,40	61,73	83,66	85,80	87,48	87,55
Mallows Cp	2896,9	1698,7	268,6	130,2	21,8	18,1

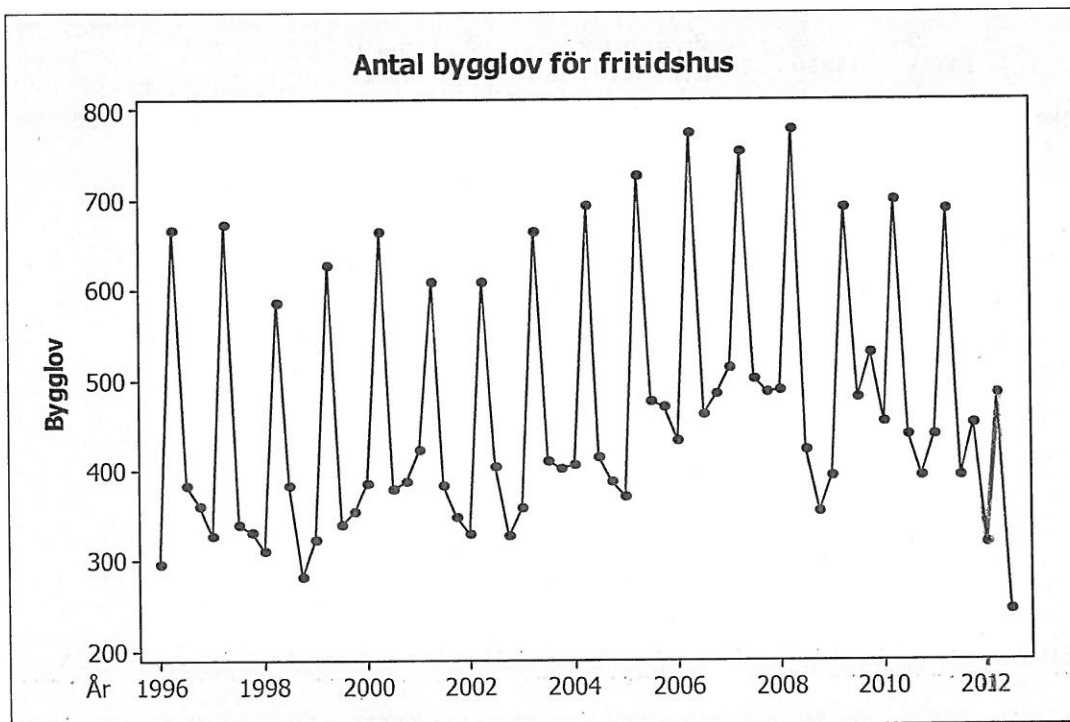
Uppgift 3 (3p)

År	Kilopris på köttfärs	KPI (livsmedel, basår = 1980)
2005	54,10	238,10
2006	57,50	239,90
2007	61,60	244,77
2008	66,50	261,69

- Beräkna ett index som beskriver prisutvecklingen på köttfärs mellan 2005 och 2008. Tolka indexet. (1p)
- Beräkna ett index som beskriver prisutvecklingen på köttfärs relativt prisutvecklingen för den större varugruppen. Tolka indexet. (2p)

Uppgift 4 (6p)

Nedan visas en tidsserie över antal utfärdade bygglov för fritidshus i Sverige från kvartal 1 1996 till kvartal 3 2012.



På nästkommande sidor kommer två utskrifter från Minitab som ni ska använda för att lösa nedanstående uppgifter.

- Tolka trenden och säsongparametrarna för kvartal 1 och 2 i **Modell 1**. (1p)
- Testa om trenden är signifikant i **Modell 1**. (1p)
- Undersök med hjälp av ett test om säsongsmönstret är signifikant. (2p)

Modell 1

The regression equation is

$$\text{Bygglov} = 352 + 1,34 \text{ Tid} - 8,8 \text{ Kvartal}_1 + 272 \text{ Kvartal}_2 + 5,0 \text{ Kvartal}_3$$

Predictor	Coef	SE Coef
Constant	351,67	20,54
Tid	1,3370	0,3939
Kvartal_1	-8,85	21,71
Kvartal_2	272,35	21,70
Kvartal_3	5,01	21,71

S = 62,3075 R-Sq = 80,6% R-Sq(adj) = 79,3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	997121	249280	64,21	0,000
Residual Error	62	240698	3882		
Total	66	1237819			

Modell 2

The regression equation is
Bygglov = 419 + 1,35 Tid

Predictor	Coef	SE Coef
Constant	419,48	33,47
Tid	1,3464	0,8556

S = 135,442 R-Sq = 3,7% R-Sq(adj) = 2,2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	45422	45422	2,48	0,120
Residual Error	65	1192397	18345		
Total	66	1237819			

- d) Tidsserieregession är inte den mest lämpliga metoden för denna tidsserie. Varför inte, och vilken metod skulle antagligen passa bra på denna tidsserie? Motivera. (2p)

Bilaga 1

Formelsamling för index

Sammansatta 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 vikterna.

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:

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