



Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2013-01-11
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	TER4
Tid	8-12
Kurskod	732G05
Provkod	TENB
Kursnamn/benämning Provnamn/benämning	Regressions- och tidsserieanalys Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	3
Jour/Kursansvarig Ange vem som besöker salen	Tommy Schyman Patrik Waldmann
Telefon under skrivtiden	Tommy: 0768 – 30 31 09 Patrik: 072 742 68 70
Besöker salen ca kl.	09:30, 10:30
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. I bilaga 1 finns formler för index.
Övrigt	

**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:	2013-01-11, 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 (8p)

Nedan följer några statistiska analyser av demografiska data från 1990. Man var främst intresserad av att analysera sambandet mellan 91 länders bruttonationalprodukt (BNP) per capita och följande förklaringsvariabler:

- Dödskvot per 1000 invånare (DK)
- Födelsekvot per 1000 invånare (FK)
- Spädbarnsdödlighet under 1 år per 1000 invånare (SPD)
- Medellivslängd för män (ML M)
- Medellivslängd för kvinnor (ML K)
- Indikatorvariabel för region Östeuropa (ÖE)
- Indikatorvariabel för region Sydamerika (SA)
- Indikatorvariabel för region Mellanöstern (MÖ)
- Indikatorvariabel för region Asien (AS)
- Indikatorvariabel för region Afrika (AF)

Indikatorvariablerna är jämförda mot region Västeuropa som därför inte är med i någon modell. Först beräknade man korrelationskoefficienterna (se nedan) och baserat på detta beslöt man sig för att anpassa en enkel linjär regression mellan BNP och ML K. Vissa värden saknas i utskriften från Minitab. Din uppgift är att

- a) Pröva på 5% nivå om regressionskoefficienten (lutningsparametern) är 0. (2p)
- b) Beräkna samt tolka förklaringsgraden och Mallows Cp (använd information från Modell 5 nedan). (2p)

Correlations: BNP; DK; FK; SBD; ML M; ML K; ÖE; SA; MÖ; AS; AF

	BNP	DK	FK	SBD	ML M	ML K	ÖE	SA	MÖ	AS
DK	-0,303									
FK	-0,629	0,506								
SBD	-0,602	0,678	0,857							
ML M	0,643	-0,754	-0,866	-0,935						
ML K	0,65	-0,715	-0,894	-0,954	0,982					
ÖE	-0,157	-0,015	-0,348	-0,267	0,205	0,267				
SA	-0,197	-0,11	-0,008	-0,033	0,053	0,088	-0,129			
MÖ	0,072	-0,315	0,12	-0,059	0,131	0,061	-0,116	-0,137		
AS	-0,181	-0,136	-0,013	0,105	-0,025	-0,108	-0,141	-0,166	-0,15	
AF	-0,394	0,542	0,718	0,628	-0,721	-0,698	-0,215	-0,253	-0,228	-0,277

Regression Analysis: BNP versus ML K

Predictor	Coef	SE Coef	T	P
Constant	-25469		X	X
ML K	472,67		X	X

S = 6184,84 R-Sq = X C-p = X

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	X	X	2491238936	X	X
Residual Error	X	X	X		
Total	90	5895688821			

Efter detta anpassade man 5 olika multipla regressionsmodeller (Modell 1-5).

c) Din uppgift är att välja den bästa modellen samt motivera ditt val.

(2p)

Modell 1

Regression Analysis: BNP versus DK; SBD; ML K; ML M; ÖE; SA; MÖ; AS; AF

The regression equation is

$$\text{BNP} = -36774 + 420 \text{ DK} + 43,8 \text{ SBD} + 794 \text{ ML K} - 152 \text{ ML M} - 15747 \text{ ÖE} - 12614 \text{ SA} \\ - 4718 \text{ MÖ} - 8631 \text{ AS} - 8156 \text{ AF}$$

Predictor	Coef	SE Coef	T	P
Constant	-36774	14430	-2,55	0,013
DK	420,0	174,0	2,41	0,018
SBD	43,82	29,82	1,47	0,146
ML K	793,8	307,7	2,58	0,012
ML M	-152,2	288,5	-0,53	0,599
ÖE	-15747	1596	-9,87	0,000
SA	-12614	1729	-7,30	0,000
MÖ	-4718	2216	-2,13	0,036
AS	-8631	2192	-3,94	0,000
AF	-8156	2231	-3,66	0,000

S = 3787,97 R-Sq = 80,3% R-Sq(adj) = 78,1% C-p = 9,2

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	9	4733443799	525938200	36,65	0,000
Residual Error	81	1162245022	14348704		
Total	90	5895688821			

Modell 2

Regression Analysis: BNP versus DK; SBD; ML K; ÖE; SA; MÖ; AS; AF

The regression equation is

$$\text{BNP} = -37105 + 425 \text{ DK} + 43,0 \text{ SBD} + 658 \text{ ML K} - 15552 \text{ ÖE} - 12547 \text{ SA} - 5066 \text{ MÖ} \\ - 8981 \text{ AS} - 8189 \text{ AF}$$

Predictor	Coef	SE Coef	T	P
Constant	-37105	14353	-2,59	0,012
DK	425,2	173,0	2,46	0,016
SBD	43,00	29,65	1,45	0,151
ML K	658,3	168,7	3,90	0,000
ÖE	-15552	1546	-10,06	0,000
SA	-12547	1716	-7,31	0,000
MÖ	-5066	2107	-2,40	0,018
AS	-8981	2080	-4,32	0,000
AF	-8189	2220	-3,69	0,000

S = 3771,26 R-Sq = 80,2% R-Sq(adj) = 78,3% C-p = 7,5

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	8	4729451010	591181376	41,57	0,000
Residual Error	82	1166237811	14222412		
Total	90	5895688821			

Modell 3

Regression Analysis: BNP versus ML K; ÖE; SA; MÖ; AS; AF

The regression equation is

$$\text{BNP} = -2428 + 270 \text{ ML K} - 15900 \text{ ÖE} - 14426 \text{ SA} - 8549 \text{ MÖ} - 12333 \text{ AS} - 11355 \text{ AF}$$

Predictor	Coef	SE Coef	T	P
Constant	-2428	5161	-0,47	0,639
ML K	270,32	65,01	4,16	0,000
ÖE	-15900	1591	-10,00	0,000
SA	-14426	1568	-9,20	0,000
MÖ	-8549	1661	-5,15	0,000
AS	-12333	1682	-7,33	0,000
AF	-11355	1950	-5,82	0,000

S = 3896,10 R-Sq = 78,4% R-Sq(adj) = 76,8% C-p = 11,0

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	6	4620602736	770100456	50,73	0,000
Residual Error	84	1275086085	15179596		
Total	90	5895688821			

Modell 4

Regression Analysis: BNP versus DK; FK; SBD; ML K; ÖE; SA; MÖ; AS; AF

The regression equation is

$$\text{BNP} = -33933 + 412 \text{ DK} - 44,2 \text{ FK} + 44,3 \text{ SBD} + 626 \text{ ML K} - 15553 \text{ ÖE} - 12192 \text{ SA} - 4542 \text{ MÖ} - 8823 \text{ AS} - 7612 \text{ AF}$$

Predictor	Coef	SE Coef	T	P
Constant	-33933	15958	-2,13	0,037
DK	412,3	176,0	2,34	0,022
FK	-44,21	95,20	-0,46	0,644
SBD	44,35	29,93	1,48	0,142
ML K	626,4	182,9	3,42	0,001
ÖE	-15553	1553	-10,01	0,000
SA	-12192	1887	-6,46	0,000
MÖ	-4542	2398	-1,89	0,062
AS	-8823	2117	-4,17	0,000
AF	-7612	2553	-2,98	0,004

S = 3789,43 R-Sq = 80,3% R-Sq(adj) = 78,1% C-p = 9,3

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	9	4732547855	525838651	36,62	0,000
Residual Error	81	1163140966	14359765		
Total	90	5895688821			

Modell 5

Regression Analysis: BNP versus DK; FK; ...

The regression equation is

$$\text{BNP} = -33781 + 408 \text{ DK} - 41,9 \text{ FK} + 45,1 \text{ SBD} - 146 \text{ ML M} + 758 \text{ ML K} - 15740 \text{ ÖE} \\ - 12274 \text{ SA} - 4236 \text{ MÖ} - 8495 \text{ AS} - 7610 \text{ AF}$$

Predictor	Coef	SE Coef	T	P
Constant	-33781	16035	-2,11	0,038
DK	408,0	177,0	2,30	0,024
FK	-41,89	95,75	-0,44	0,663
SBD	45,06	30,10	1,50	0,138
ML M	-146,1	290,3	-0,50	0,616
ML K	758,1	319,8	2,37	0,020
ÖE	-15740	1604	-9,81	0,000
SA	-12274	1903	-6,45	0,000
MÖ	-4236	2485	-1,70	0,092
AS	-8495	2225	-3,82	0,000
AF	-7610	2565	-2,97	0,004

S = 3807,02 R-Sq = 80,3% R-Sq(adj) = 77,9% C-p = 11,0

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	10	4736218254	473621825	32,68	0,000
Residual Error	80	1159470567	14493382		
Total	90	5895688821			

Efter att ha tittat ytterligare en gång på korrelationskoefficienterna beslöt man sig för att titta närmare på kollineariteten.

d) Beräkna VIF-värdet för ML K baserat på utskriften nedan.

(2p)

Regression Analysis: ML K versus DK; FK; SBD; ML M; ÖE; SA; MÖ; AS; AF

The regression equation is

$$\text{ML K} = 27,8 - 0,157 \text{ DK} - 0,0763 \text{ FK} - 0,0409 \text{ SBD} + 0,743 \text{ ML M} + 0,612 \text{ ÖE} \\ - 0,094 \text{ SA} - 2,60 \text{ MÖ} - 3,40 \text{ AS} - 1,55 \text{ AF}$$

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	9	11008,8	1223,2	699,22	0,000
Residual Error	81	141,7	1,7		
Total	90	11150,5			

Uppgift 2 (4p)

I en matbutik har priserna och försäljningen av mjölk, fil och yoghurt sett ut på följande sätt mellan 2010 och 2012:

År	Mjök		Fil		Yoghurt	
	Literpris	Sålda liter	Literpris	Sålda liter	Literpris	Sålda liter
2010	8,2	225	9,9	91	11,9	74
2011	8,5	231	10,1	93	11,9	71
2012	8,9	227	10,5	92	12,9	73

Försäljningen är i tusentals liter.

Beräkna ett kedjeprisindex med Laspeyre-vikter som beskriver prisutvecklingen från 2010 till 2012.

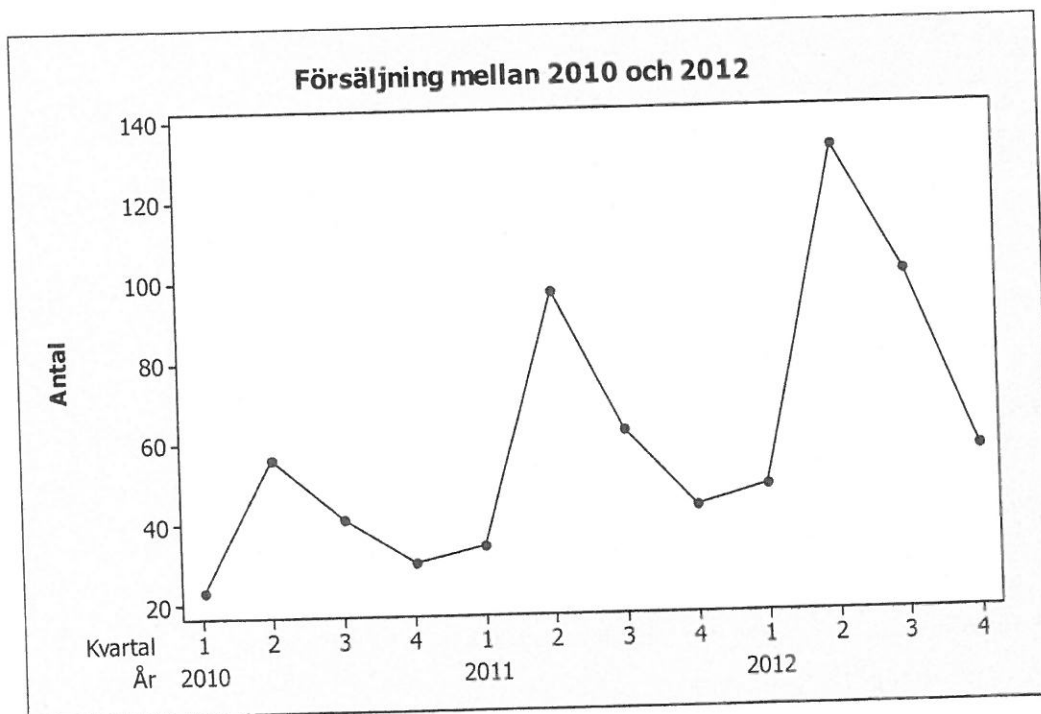
(4p)

Uppgift 3 (8p)

Nedan visas kvartalvis försäljning (i antal) för en viss vara.

Försäljning (antal)	År	Kvartal
23	2010	1
56	2010	2
41	2010	3
30	2010	4
34	2011	1
97	2011	2
62	2011	3
43	2011	4
48	2012	1
132	2012	2
101	2012	3
57	2012	4

Som synes i nedanstående graf så är det en positiv trend samt en ökande säsongvariation.



- Använd lämplig komponentuppdelningsmodell och skatta säsongskomponenterna för de fyra kvartalen. Tolka säsongskomponenten för kvartal 2. (4p)
- Säsongsrensa serien med hjälp av de skattade säsongskomponenterna från a). (1p)
- Beräkna trendkomponenten med hjälp av enkel linjär regression. (2p)
- Beräkna en punktskattning på försäljningen i kvartal 1 2013. (1p)

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:

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

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

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:

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