



Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2011-01-22
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	8-12
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	Lotta Hallberg
Telefon under skrivtiden	
Besöker salen ca kl.	10
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Carita Lilja, 1463, carita.lilja@liu.se
Tillåtna hjälpmedel	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. Flärpar med liten anteckning är tillåten. Formelsamling. Tabeller
Övrigt	G=12, VG=16
Vilken typ av papper ska användas, rutigt eller linjerat	Rutigt

Antal

95

Antal exemplar i påsen	
-------------------------------	--

Tentamen i Regressions- och tidsserieanalys, 2011-01-22

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. Flärpar med liten anteckning är tillåten. Formelsamling. Tabeller

Jourhavande lärare: Lotta Hallberg

Redovisa och motivera kort alla dina lösningar

1

I tabellen nedan visas totalförsäljningen (i löpande priser) för insektsmedel för hästar och för hundar. Även priser för medlet "Stopp o bit" för hästar från gruppen insektsmedel för hästar samt "fästinghalsband" för hundar från gruppen insektsmedel för hundar är givna. Data är från åren 2008 till 2010

År	Tot. förs. insektsmedel för hästar	Tot. förs. insektsmedel för hundar	Pris, Stopp o bit	Pris, fästinghalsband
2008	1150	630	257	210
2009	1210	755	267	215
2010	1530	835	285	275

Använd varorna 'Stopp o bit' och fästinghalsband som representantvaror för sina grupper och beräkna ett kedjeprisindex av Laspeyre-typ för företagets priser med basår 2008.

3p

2

Datamängden som ska analyseras består av följande variabler:

Fat = mängden kroppsfett

Triceps = hudveckstjocklek (vet inte vad som menas med hudveck här)

Thigh = omkrets på lår

Midarm = omkrets på överarm

Dessa variabler har mätts på 20 friska kvinnor i åldern 20-34 år.

Variablen *Fat* är så komplicerad och dyr att mäta så man vill hitta en modell för att kunna prediktera den istället.

Så här ser data ut: Därefter kommer ett par utskrifter från Minitab

Fat	Triceps	Thigh	Midarm
11,9	19,5	43,1	29,1
22,8	24,7	49,8	28,2
18,7	30,7	51,9	37,0
20,1	29,8	54,3	31,1
12,9	19,1	42,2	30,9
21,7	25,6	53,9	23,7
27,1	31,4	58,5	27,6
25,4	27,9	52,1	30,6
21,3	22,1	49,9	23,2
19,3	25,5	53,5	24,8
25,4	31,1	56,6	30,0
27,2	30,4	56,7	28,3
11,7	18,7	46,5	23,0
17,8	19,7	44,2	28,6

12,8	14,6	42,7	21,3
23,9	29,5	54,4	30,1
22,6	27,7	55,3	25,7
25,4	30,2	58,6	24,6
14,8	22,7	48,2	27,1
21,1	25,2	51,0	27,5

Correlations:

	Fat	Triceps	Thigh
Triceps	0,843		
Thigh	0,878	0,924	
Midarm	0,142	0,458	0,085

Cell Contents: Pearson correlation

Regression Analysis: Fat versus Triceps; Thigh; Midarm

The regression equation is

$$\text{Fat} = 117 + 4,33 \text{ Triceps} - 2,86 \text{ Thigh} - 2,19 \text{ Midarm}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	117,08	99,78	1,17	0,258	
Triceps	4,334	3,016	1,44	0,170	708,843
Thigh	-2,857	2,582	-1,11	0,285	564,343
Midarm	-2,186	1,595	-1,37	0,190	104,606

S = 2,47998 R-Sq = 80,1% R-Sq(adj) = 76,4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	396,98	132,33	21,52	0,000
Residual Error	16	98,40	6,15		
Total	19	495,39			

- Beskriv hur nollhypotesen ser ut som hör till F-testet ovan. Förklara därefter varför vi kan ha signifikans i detta test men inte signifikans i något av t-testen. 1p
- Beräkna den multipla korrelationen R^2 för *Thigh* och de övriga två förklarande variablerna. 1p
- Prediktera *Fat* för en kvinna med *Triceps*=19, *Thigh*=44 och *Midarm*= 25. 1p

Man har till slut bestämt att följande enkla modell är bäst:

Regression Analysis: Fat versus Thigh

The regression equation is

$$\text{Fat} = -23,6 + 0,857 \text{ Thigh}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	-23,634	5,657	-4,18	0,001	
Thigh	0,8565	0,1100	7,79	0,000	1,000

S = 2,51024

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	381,97	381,97	60,62	0,000
Residual Error	18	113,42	6,30		
Total	19	495,39			

Följande summor har också beräknats:

$$\sum Fat = 403,9 \quad \sum Thigh = 1023,4 \quad \sum Fat * Thigh = 21113,5 \quad \sum Thigh^2 = 52888$$

- d) Visa att skattningarna med minsta kvadrat (least squares) metoden på regressionskoefficienterna β_0 och β_1 i modellen ovan blir som i utskriften. Ta hjälp av summorna ovan. 2p
- e) Prediktera *Fat* på en kvinna med låromfång = *Thigh* på 55. Beräkna även ett prediktionsintervall med konfidensgrad 95%. 2p
- f) Beräkna R^2 i modellen och tolka den. Vad kallar man R^2 vanligen på svenska? 1p
- g) Man kom fram till modellen ovan genom bakåtelimineringmetoden, Förklara hur man gick tillväga. 1p

3

Utgjämna tidsserien 5 10 8 9 15 17 16 med ett 3 punkters (period) centrerat glidanden medelvärde (centered moving average). Rita även en graf med tidsserie och centrerat glidande medelvärde inlagt. 2p

4

Följande tidsserier är hämtade från SCB:s hemsida. De beskriver antalet döda i bronkit, emfysem och astma, åren 1969 till 1996, hela riket. Data är uppdelat på kön.

Frågeställningar: Hur har antalet dödsfall i befolkningen förändrats mellan 1969 och 1996?

Har skillnaden i antalet dödsfall mellan könen förändrats mellan 1969 och 1996?

Antal döda är antalet döda i bronkit, emfysem och astma. Variabeln är uppdelat på kön.

Variabeln kön är omgjord till en dummy variabel och heter *Dummy kön* som har värdet 1 för kvinnor och värdet 0 för män.

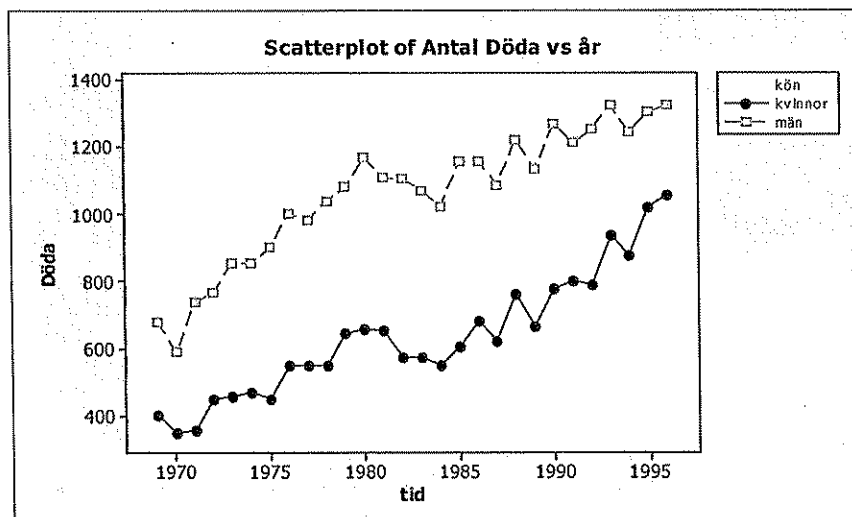
Tot antal döda är summan av *Antal döda* över kön

Diff m-k är differens mellan antal döda män och antal döda kvinnor

Tid och *Tid1* räknar åren: 1, 2, ..., 28

Nedan följer 6 olika modeller:

Uppgifterna kommer efter utskriften.



Modell1

Regression Analysis: Antal Döda versus tid; Dummy kön; kön*tid

The regression equation is

$$\text{Antal Döda} = 728 + 22,7 \text{ tid} - 400 \text{ Dummy kön} - 1,47 \text{ kön*tid}$$

Predictor	Coef	SE Coef	T	P
Constant	727,94	27,75	26,23	0,000
tid	22,718	1,672	13,59	0,000
Dummy kön	-399,80	39,24	-10,19	0,000
kön*tid	-1,472	2,364	-0,62	0,536

S = 71,4575 R-Sq = 94,1% R-Sq(adj) = 93,8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	4250714	1416905	277,49	0,000
Residual Error	52	265521	5106		
Total	55	4516235			

Modell2

Regression Analysis: Antal Döda versus tid; Dummy kön

The regression equation is

$$\text{Antal Döda} = 739 + 22,0 \text{ tid} - 421 \text{ Dummy kön}$$

Predictor	Coef	SE Coef	T	P
Constant	738,62	21,69	34,05	0,000
tid	21,982	1,175	18,70	0,000
Dummy kön	-421,14	18,99	-22,18	0,000

S = 71,0434 R-Sq = 94,1% R-Sq(adj) = 93,9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	4248735	2124368	420,90	0,000
Residual Error	53	267500	5047		
Total	55	4516235			

Modell3

Regression Analysis: tot antal döda versus tid1

The regression equation is

$$\text{tot antal döda} = 1056 + 44,0 \text{ tid1}$$

Predictor	Coef	SE Coef	T	P
Constant	1056,09	45,09	23,42	0,000
tid1	43,964	2,717	16,18	0,000

S = 116,122 R-Sq = 91,0% R-Sq(adj) = 90,6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	3531354	3531354	261,89	0,000
Residual Error	26	350593	13484		
Total	27	3881947			

Modell4

Regression Analysis: Diff m-k versus tid1

Diff m-k = 400 + 1,47 tid1

Predictor	Coef	SE Coef	T	P
Constant	399,80	32,35	12,36	0,000
tid1	1,472	1,949	0,76	0,457

S = 83,3089 R-Sq = 2,1% R-Sq(adj) = 0,0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	3958	3958	0,57	0,457
Residual Error	26	180450	6940		
Total	27	184407			

Modell5

Regression Analysis: Diff m-k versus tid1; tid1*tid1

The regression equation is

Diff m-k = 216 + 38,2 tid1 - 1,27 tid1*tid1

Predictor	Coef	SE Coef	T	P
Constant	216,08	20,39	10,60	0,000
tid1	38,215	3,241	11,79	0,000
tid1*tid1	-1,2670	0,1085	-11,68	0,000

S = 33,4281 R-Sq = 84,9% R-Sq(adj) = 83,6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	156471	78236	70,01	0,000
Residual Error	25	27936	1117		
Total	27	184407			

Modell6

Double Exponential Smoothing for tot antal döda

Data tot antal döda
Length 28

Smoothing Constants

Alpha (level) 0,3
Gamma (trend) 0,4

Accuracy Measures

MAPE 6,2
MAD 101,2
MSD 16890,8

Här ges de utjämnade värdena för de tre sista tidpunkterna:

LEVEL	TREND	tid=T
2179,58	64,6980	26
2269,09	74,6252	27
2354,30	78,8592	28

- a) Pröva om antalet döda ökar i samma takt för män och för kvinnor eller om antalet döda ökar snabbare för något av könen. Modell1 1p
- b) Bestäm två regressionsuttryck för antalet döda, ett för vardera könet. Modell1 eller Modell2. 1p
- c) Beräkna prognos för totala antalet döda för 1997 och 1998. Använd någon lämplig modell bland modell1 till modell5. 1p
- d) Beräkna prognos för totala antalet döda för 1997 och 1998 med modell 6. 1p
- e) Svara på de två frågeställningarna ovan. Utred. 2p