

Försättsblad till skriftlig tentamen vid Linköpings universitet



Datum för tentamen	2019-06-07
Sal (2)	G35(1) <u>TER4(37)</u>
Tid	14-18
Utb. kod	732G05
Modul	TENB
Utb. kodnamn/benämning Modulnamn/benämning	Regressions- och tidsserieanalys Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	4
Jour/Kursansvarig Ange vem som besöker salen	Annika Tillander
Telefon under skrivtiden	013-28 12 14
Besöker salen ca klockan	OBS! Endast telefonjour
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Anna Grabska Eklund 013-28 23 62 anna.grabska.eklund@liu.se
Tillåtna hjälpmedel	Miniräknare Kursbok: Bowerman, O'Connel, Koehler: Forecasting, Time series, and Regression (alla upplagor tillåtna - får innehålla markeringar, understrykningar och flärpar, men inte anteckningar) Ett A4-blad: med vad som helst skrivet på fram- och baksida (för hand/skrivet på dator/kopierat etc. - inga begränsningar).
Övrigt	
Antal exemplar i påsen	

Tentamen

Linköpings universitet, Institutionen för datavetenskap, STIMA



Kurskod och namn:	732G05/G41, Regression och tidsserieanalys
Datum och tid:	2019-06-07 kl. 14-18
Jourhavande lärare:	Annika Tillander
Tillåtna hjälpmedel:	Miniräknare av valfri modell Kursbok Bowerman, O'Connel, Koehler: Forecasting, Time series, and Regression (alla upplagor tillåtna - får innehålla markeringar, understrykningar och flärpar, men inte anteckningar) A4-blad ett dubbelsidigt med vad som helst skrivet på fram- och baksida (för hand/skrivet på dator/kopierat etc. - inga begränsningar)
Betygsgränser:	Tentamen omfattar totalt 40 poäng, G från 24p, VG från 32p
Övrigt:	Siffrorna i uppgifterna är fiktiva alternativt modifierade. Saknas någon siffra för att kunna lösa uppgiften? Skriv då tydligt ut att du saknar denna information, anta ett godtyckligt värde och lös uppgiften med detta antagande.

Redovisa, tolka och motivera tydligt alla dina lösningar

Lycka till!

Uppgift 1 (8p)

Tabell 1. Priser (p) och kvantitet (q) för ett års varukorg bestående av tre varor (smågodis, chips och läsk) och konsumentprisindex (KPI) för de tre senaste åren.

	Godis p	Godis q	Chips p	Chips q	Läsk p	Läsk q	KPI
2016	49	14.3	17	2.7	16	54.1	316.43
2017	47	12.1	18	2.9	12	57.2	322.11
2018	49	16.7	22	2.4	22	61.3	328.40

a) 4p

Beräkna ett prisindex för varukorgen i tabell 1 med 2016 som basår och Laspeyre-vikter och tolka resultatet.

b) 4p

Beräkna ett relativprisindex för prisindexet i a) med 2017 som basår och tolka resultatet.

Uppgift 2 (17p)

En studie med 180 försökspersoner syftar till att undersöka blodsockernivån (y) utifrån 3 förklaringsvariabler, **Dryck** i deciliter (x_1), **Tid sedan måltid** antal timmar sedan senaste matintag (x_2) och **Aktivitetsnivå** i tre kategorier där *Nivå 0* = stillasittande, *Nivå 1* = lätt aktivitet som t.ex. promenad och *Nivå 2* = högre aktivitet som t.ex. jogga. Variabeln **Aktivitetsnivå** är dummy-kodad där $D_1 = 1$ om *Nivå 1* annars 0 och $D_2 = 1$ om *Nivå 2* annars 0. Två olika modeller föreslås (Minitab-utskrift sidan 3):

$$\text{Modell 1 } y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 D_1 + \beta_4 D_2 + \varepsilon$$

$$\text{Modell 2 } y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 D_1 + \beta_4 D_2 + \beta_5 x_2 \times D_1 + \beta_6 x_2 \times D_2 + \varepsilon$$

a) 4p

Fyll i de saknade värdena som är ersatta med bokstäver i fetstil från **A-H** i Minitab-utskriften för Modell 1. Redovisa dina beräkningar.

b) 5p

Testa på 5% signifikansnivå om en modell med interaktionstermer (modell 2) är signifikant bättre än en modell utan interaktionstermer (modell 1).

Modell 1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	A	B	C	D	
Error	175	55357	316,3		
Total	179	150494			

Model Summary

S	R-sq	R-sq(adj)
E	63,22%	F

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	136,74	4,35	31,46	0,000
X1	G	0,230	2,04	0,043
X2	-1,509	H	-2,85	0,005
D1	-23,51	3,80	-6,19	0,000
D2	-52,87	3,51	-15,07	0,000

Modell 2

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	107685	17947,4	72,53	0,000
Error	173	42809	247,5		
Total	179	150494			

Model Summary

S	R-sq	R-sq(adj)
15,7306	71,55%	70,57%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	195,56	9,12	21,44	0,000
X1	0,626	0,207	3,02	0,003
X2	-16,45	2,16	-7,62	0,000
D1	-84,2	11,6	-7,27	0,000
D2	-117,91	9,65	-12,22	0,000
X2 * D1	14,97	2,60	5,76	0,000
X2 * D2	15,77	2,22	7,12	0,000

c)

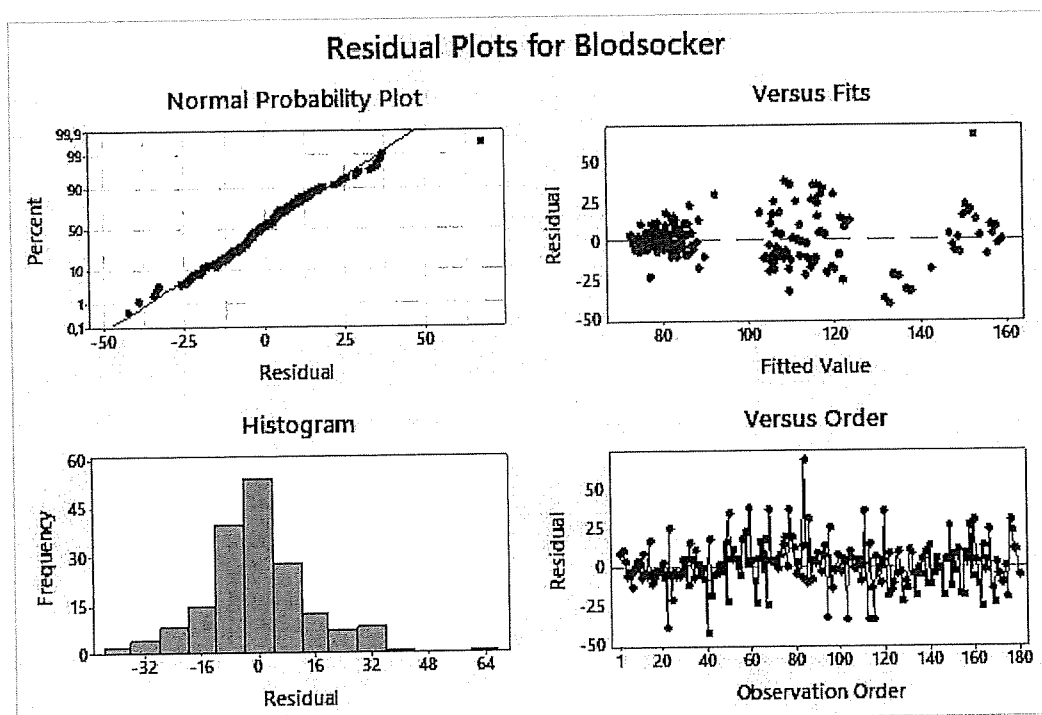
6p

Gör en prognos/prediktion för blocksockernivå med 95% konfidensintervall för en grupp som intagit 5 deciliter dryck och 4 timmar sedan matintag för var och en av de tre aktivitetsnivåerna; Nivå 0 (distance value = 0.0317), Nivå 1 (distance value = 0.0312) och Nivå 2 (distance value = 0.018).

d)

2p

Givet resultatet i figur 1, kan antagande för linjär regression ses som uppfyllda?



Figur 1. Residualplottar från Minitab för Modell 2.

Uppgift 3 (3p)

a)

1p

Vad innebär autokorrelation?

b)

1p

Hur kan autokorrelation upptäckas?

c)

1p

Vilka modeller som ingår i kursen är lämpliga respektive olämpliga om det finns autokorrelation i data?

Uppgift 4 (12p)

Antal olyckor med elsparkcyklar antas ha haft en så kraftig uppgång under perioden 2016 till 2019 att det kan ses som en exponentiellt ökande trend. Det finns dock säsongsvariation över året. Under de tre åren har säsongskomponenterna skattats till 25% över trend-nivån för sommarhalvåret och ca. 25% under trend-nivån för vinterhalvåret. Originaldata förutsätts vara fria från konjunkturvariation och presenteras i tabell 2.

Tabell 2. Antal olyckor per halvår från sommar 2016 till vinter 2018/2019.

Tid	Antal olyckor
Sommar 16	97
Vinter 16/17	78
Sommar 17	124
Vinter 17/18	76
Sommar 18	231
Vinter 18/19	189

a) 5p

Säsongsgrensa data och skatta den exponentiella tillväxtfaktorn (growth rate) per år.

b) 5p

Testa koefficienten (b_1) i a) på 5% signifikansnivå.

c) 2p

Gör en prognos/prediktion för antal olyckor med elsparkcyklar Sommar 2019 och tolka resultatet.