

Försättsblad till skriftlig tentamen vid Linköpings universitet



Datum för tentamen	2016-01-22
Sal (1)	<u>TER1</u>
Tid	14-19
Kurskod	732G21
Provkod	TEN1
Kursnamn/benämning Provnamn/benämning	Sambandsmodeller Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	4
Jour/Kursansvarig Ange vem som besöker salen	Lotta Hallberg
Telefon under skrivtiden	013-281657
Besöker salen ca klockan	16
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Annelie Almquist
Tillåtna hjälpmedel	Räknedosa. Läroboken: Applied linear statistical models av Kutner, Nachtsheim m fl som inte får innehålla anteckningar men får ha markeringar och flärpar. Flärpar får ha en liten anteckning. Dessutom ett handskrivet, dubbelsidigt A4 ark med egna anteckningar.
Övrigt	
Antal exemplar i påsen	

TENTAMEN I SAMBANDSMODELLER, 2016-01-22

- Skrivtid:** kl: 14-19
- Hjälpmedel:** Räknedosa. Läroboken: *Applied linear statistical models* av Kutner, Nachtsheim m fl som inte får innehålla anteckningar men får ha markeringar och flärpar. Flärpar får ha en liten anteckning. Dessutom ett handskrivet, dubbelsidigt A4 ark med egna anteckningar.
- Jourhavande lärare:** Lotta Hallberg
- Betygsgränser:** För godkänt krävs minst 12 av 20 poäng och för väl godkänt krävs minst 16 av 20 poäng.

Redovisa och motivera kort alla dina lösningar

Tolka (om möjligt) alla dina resultat!

1

Mätningarna som ska analyseras gjordes på män som var inblandade i en fysisk fitness kurs på N.C.State Univ .

Variablerna är Ålder (år) , Vikt (kg) , Syreupp = syreupptagningsförmåga (ml per kg kroppsvikt per minut) , Löptid = tid att springa 1,5 km (minuter) , Vilopuls , Löppuls = puls under löpning (här mättes syreupptagningsförmågan) samt Maxpuls under löpning.

Nedan följer en mängd utskrifter. Efter alla utskrifter kommer uppfifterna.

Correlation:

	Ålder	Vikt	Syreupp	Löptid	Vilopuls	Löppuls
Vikt	-0,234					
Syreupp	-0,305	-0,163				
Löptid	0,189	0,144	-0,862			
Vilopuls	-0,164	0,044	-0,399	0,450		
Löppuls	-0,338	0,182	-0,398	0,314	0,352	
Maxpuls	-0,433	0,249	-0,237	0,226	0,305	0,930

Följande modell har anpassats:

Regression Analysis: Syreupp versus Ålder; Vikt; Löptid; Vilopuls; Löppuls; Maxpuls

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	722,544	120,424	22,43	0,000
Ålder	1	27,746	27,746	5,17	0,032
Vikt	1	9,911	9,911	1,85	0,187
Löptid	1	250,822	250,822	46,72	0,000
Vilopuls	1	0,571	0,571	0,11	0,747
Löppuls	1	51,058	51,058	9,51	0,005
Maxpuls	1	26,491	26,491	4,93	0,036
Error	24	128,838	5,368		
Total	30	851,382			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2,31695	84,87%	81,08%	77,36%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	102,9	12,4	8,30	0,000	
Ålder	-0,2270	0,0998	-2,27	0,032	1,51
Vikt	-0,0742	0,0546	-1,36	0,187	1,16
Löptid	-2,629	0,385	-6,84	0,000	1,59
Vilopuls	-0,0215	0,0661	-0,33	0,747	1,42
Löppuls	-0,370	0,120	-3,08	0,005	8,44
Maxpuls	0,303	0,136	2,22	0,036	8,74

Regression Equation

Syreupp = 102,9 - 0,2270 Ålder - 0,0742 Vikt - 2,629 Löptid - 0,0215 Vilopuls - 0,370 Löppuls + 0,303 Maxpuls

Fits and Diagnostics for Unusual Observations

Obs	Syreupp	Fit	Resid	Std Resid	
15	51,85	46,47	5,38	2,49	R
17	40,84	46,24	-5,40	-2,43	R

R Large residual

Regression Analysis: Maxpuls versus Ålder; Vikt; Löptid; Vilopuls; Löppuls

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	2231,28	446,26	38,72	0,000
Ålder	1	21,39	21,39	1,86	0,185
Vikt	1	10,11	10,11	0,88	0,358
Löptid	1	2,25	2,25	0,20	0,662
Vilopuls	1	0,42	0,42	0,04	0,851
Löppuls	1	1462,36	1462,36	126,88	0,000
Error	25	288,14	11,53		
Total	30	2519,42			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3,39492	88,56%	86,28%	82,03%

Best Subsets Regression: Syreupp versus Ålder; Vikt; ...

Response is Syreupp

Vars	R-Sq	R-Sq (adj)	PRESS	R-Sq (pred)	Mallows Cp	S	r	t	d	s	s	s
1	74,3	73,5	250,9	70,5	13,7	2,7448						
1	15,9	13,1	834,3	2,0	106,3	4,9675						
2	76,4	74,7	250,8	70,5	12,4	2,6774	X	X				
2	76,1	74,4	242,8	71,5	12,8	2,6934			X	X		
3	81,1	79,0	205,1	75,9	7,0	2,4406	X	X	X			
3	81,0	78,9	212,9	75,0	7,1	2,4478			X	X	X	
4	83,7	81,2	188,6	77,8	4,9	2,3116	X	X	X	X		
4	81,6	78,8	212,3	75,1	8,1	2,4513	X	X	X	X		
5	84,8	81,8	181,6	78,7	5,1	2,2752	X	X	X	X	X	
5	83,7	80,4	202,4	76,2	6,8	2,3558	X	X	X	X	X	
6	84,9	81,1	192,8	77,4	7,0	2,3169	X	X	X	X	X	X

- a) Om modell ska väljas enligt framåtvalsmetoden. Vilken variabel ska då först testas för att läggas till? 1p
- b) Om bakåtvalsmetoden ska användas för val av modell, vilken variabel ska då först väljas bort? 1p
- c) Välj ett lämpligt mått i utskriften för 'Best subsets' för att välja den bästa modellen. Vilken modell väljer du? 1p
- d) Beräkna VIF för Maxpuls. 1p
- e) Det är något konstigt med de skattade värdena för Maxpuls i den anpassade modellen. Förklara vad som är konstigt och vad det kan bero på. 1p
- f) Beräkna ett prediktionsintervall för Syreupptagningsförmågan hos en man som är 40 år och väger 80 kg. Har en löptid på 10 min och vilopuls på 65. Löp- och maxpuls är 175. Standard error of fit är 1.923. Använd den fulla modellen som är anpassad även om du inte tycker den är bra. 2p

2

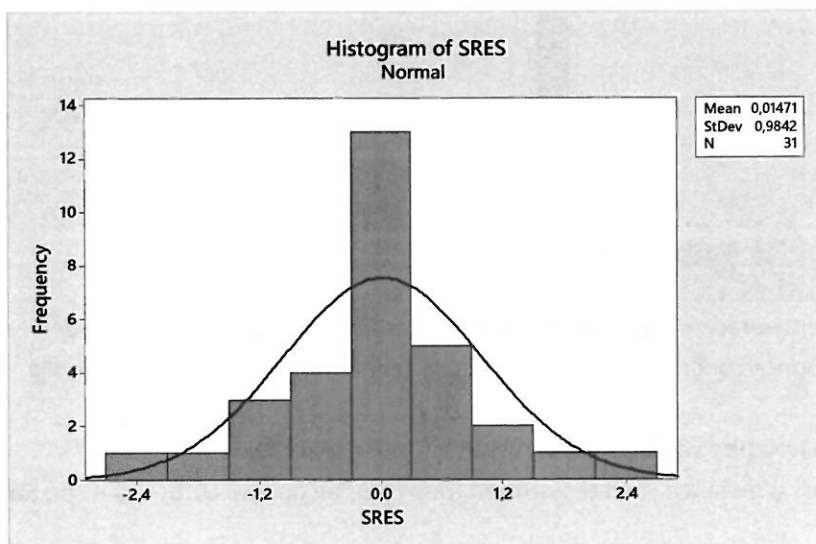
Fortsättning på uppgift 1. Nu ska modellen utvärderas. Även om du valt en annan modell i uppgift 1 så är det modellen med alla sex förklarande variabler som ska utvärderas här. Uppgifterna kommer efter utskrifter och grafer.

Utskriften här är kopierad från uppgift 1 ovan

Fits and Diagnostics for Unusual Observations

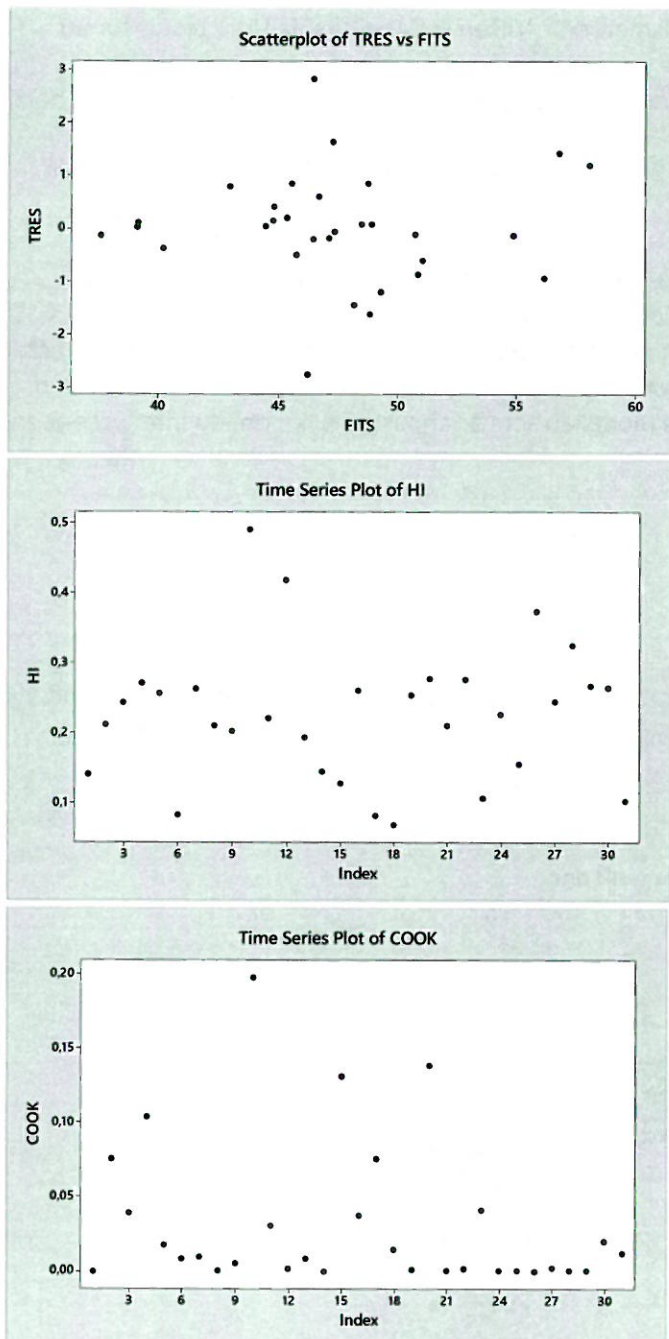
Obs	Syreupp	Fit	Resid	Std Resid	
15	51,85	46,47	5,38	2,49	R
17	40,84	46,24	-5,40	-2,43	R

R Large residual



SRES = standardiserade residualer

TRES = studentiserade deleted residualer



- Testa med hjälp av studentiserade deleted residualer om det finns några avvikande värden i Y led genom att använda Bonferroni metodik med 10% familjekonfidensgrad. Använd lämplig graf ovan. 1p
- Undersök om det finns avvikande värden i x led. Använd lämplig graf ovan. 1p
- I den sista grafen ges Cook's distance. Vad testar man med hjälp av dessa? Utför testet på 50% nivå. 1p

3

Vid ett löpande band produceras en viss typ av mobiltelefon. Vid en kvalitetskontroll förfar man på följande vis. Vid en viss tidpunkt så undersöker man telefoner och räknar antalet telefoner tills man får en med någon defekt. Den defekta inkluderas i antalet. Detta görs 10 gånger. Resultat:

2 1 35 43 46 3 36 38 51 17

Låt X vara detta antal som är beskrivet ovan. X är då en geometriskt fördelad slumpvariabel. Sannolikhetfunktionen är given av

$$f(x) = (1 - \pi)^{x-1}\pi, \quad x = 1, 2, \dots$$

Maximum Likelihood skatta π . Sätt även in observationerna i skattningen.

3p

4

I en studie i brottslighet så drog man slumpmässigt 40 personer från olika områden enligt tabellen nedan. Man frågade varje person om hur stort deras förtroende i rättsväsendet är. Stora värden är stort förtroende. Som man kan se i tabellen så har man valt personer som har dömts 0, 1, eller fler än 2 gånger och även valt personer från tre olika socialklasser.

Antal ggr dömd	Socialklass		
	Låg	Medel	Hög
0	4 14 15 19 17 17 16	7 10 12 15 16	8 10 19 17
1	2 7 18	6 19 12 12	7 6 5 3 16
2	7 8 2 11 12	1 2 4	4 2 8 9

ANOVA-tabell

Källa	df	SS
Dömd	2	400.000
Socialklass	2	22.739
Dömd*Socialklass	4	109.930
Error	31	704.080

- Rita en så kallat treatment means plot. Beskriv vad du ser. 1p
- Har vi ett balanserat eller obalanserat experiment? Förklara hur du ser det. 1p
- Vilken faktor är viktigast? Är någon faktor inte intressant? Utför två test för att checka detta. 1p
- Skatta konfidensintervall för marginalväntevärdena för de båda faktorerna. Använd Scheffés metod. Är Scheffés metod den mest optimala att använda här? Motivera. 2p
- Pröva om interaktionen är signifikant. Ta därefter bort interaktionstermen och pröva igen om Dömd är signifikant. 2p

