

Försättsblad till skriftlig tentamen vid Linköpings universitet



Datum för tentamen	2016-06-14
Sal (1)	<u>TER3</u>
Tid	8-13
Kurskod	732G21
Provkod	TEN1
Kursnamn/benämning	Sambandsmodeller
Provnamn/benämning	Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	3
Jour/Kursansvarig	Lotta Hallberg
Ange vem som besöker salen	
Telefon under skrivtiden	013-281657
Besöker salen ca klockan	10
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, 2016-06-14

- Skriptid:** kl: 8-13
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

Sprickbildning i järnvägsräls är ett besvärligt problem för dem som har ansvar för underhåll av järnvägsbanorna. Man vill jämföra sprickbildningen i olika sorters järnmaterial genom att lägga ut två kilometer räls på ett antal platser och notera antalet sprickor som uppstår under två år. Man måste då också ta hänsyn till hur stor tågtrafiken är på de olika platserna.

Datamateriel:

Row	Sprickor	Antal tåg	Material
1	8	16,9	1
2	11	23,6	1
3	7	14,4	1
4	10	17,2	1
5	4	9,1	1
6	17	33,3	1
7	9	21,4	1
8	16	27,3	1
9	10	18,6	1
10	5	12,1	1
11	5	17,8	0
12	9	24,4	0
13	5	13,5	0
14	6	20,1	0
15	4	11,0	0
16	9	19,6	0
17	8	25,4	0
18	16	35,5	0
19	7	16,8	0
20	11	31,2	0

Nedan har tre regressionsmodeller anpassats

Modell 1

Regression Analysis: Sprickor versus Antal tåg

Analysis of Variance					
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	218,78	218,779	56,44	0,000
Error	18	69,77	3,876		
Total	19	288,55			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1,96880	75,82%	74,48%	70,41%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-0,60	1,33	-0,45	0,658	
Antal tåg	0,4619	0,0615	7,51	0,000	1,00

Regression Equation

Sprickor = -0,60 + 0,4619 Antal tåg

Modell 2

Regression Analysis: Sprickor versus Antal tåg; Material

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	255,74	127,871	66,26	0,000
Error	17	32,81	1,930		
Total	19	288,55			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1,38921	88,63%	87,29%	84,33%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-2,56	1,04	-2,46	0,025	
Antal tåg	0,4906	0,0439	11,18	0,000	1,02
Material	2,750	0,628	4,38	0,000	1,02

Regression Equation

Sprickor = -2,56 + 0,4906 Antal tåg + 2,750 Material

$$X'X = \begin{pmatrix} 20,0 & 409,2 & 10,0 \\ 409,2 & 9397,8 & 193,9 \\ 10,0 & 193,9 & 10,0 \end{pmatrix}$$

$$(X'X)^{-1} = \begin{pmatrix} 0,562307 & -0,0214727 & -0,145952 \\ -0,021473 & 0,0009973 & 0,002134 \\ -0,145952 & 0,0021343 & 0,204567 \end{pmatrix}$$

Modell 3

Regression Analysis: Sprickor versus Antal tåg; Material; Interaktion

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	260,088	86,6959	48,74	0,000
Error	16	28,462	1,7789		
Total	19	288,550			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1,33375	90,14%	88,29%	83,00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-1,22	1,32	-0,93	0,368	
Antal tåg	0,4283	0,0580	7,39	0,000	1,94
Material	0,06	1,82	0,03	0,974	9,35
Interaktion	0,1319	0,0844	1,56	0,138	9,42

Regression Equation

Sprickor = -1,22 + 0,4283 Antal tåg + 0,06 Material + 0,1319 Interaktion

- Ställ upp den skattade regressionsmodellen från ovanstående analys med de två förklaringsvariablerna 'Antal tåg' och 'Material'. Skriv ut de anpassade modellerna för Material 1 resp. för Material 0 och tolka sambanden. 2p
- Skatta skillnaden mellan de två materialen med ett 99 % konfidensintervall. Tolka konfidensintervallet. 2p
- Tolka regressionskoefficienterna i modell 2. 1p
- Pröva med ett test om 'Material' och 'Interaktion' kan tas bort samtidigt från modell 3. 2p
- Beräkna ett 80% prediktionsintervall för Sprickor i modell 2 då 'Antal tåg'=10 och materialtyp är 0. 3p

2

Anta att du har ett stickprov av storlek 15 från en slumpvariabel X som är $N(0, \sigma^2)$. Härled maximumlikelihood-skattningen av σ^2 .

Täthetsfunktion till X är $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-0)^2}$, $-\infty < x < \infty$ 3p

3

Ett företag inom kemi vill undersöka om föroreningen hos spillvattnet vid olika anläggningar skiljer sig åt. Fem spillvattenprover togs slumpmässigt från var och en av fyra olika anläggningar och föroreningshalten mättes (lb/gal).

Resultat:

Anläggning	Prov 1	Prov 2	Prov 3	Prov 4	Prov 5	medelvärde	standardavvikelse
1	1,65	1,72	1,50	1,37	1,60	1,568	0,1366
2	1,70	1,85	1,46	2,05	1,80	1,772	0,2160
3	1,40	1,75	1,38	1,65	1,55	1,546	0,1592
4	2,10	1,95	1,65	1,88	2,00	1,916	0,1689

SSTO=0,9417

- Sätt upp en en-vägs variansanalysmodell där anläggning är en faktor med fyra nivåer. Testa med ett F-test om föroreningen i spillvattnet skiljer sig åt mellan anläggningarna. 2p
- Använd Tukey's metod för att undersöka vilka anläggningar som skiljer sig åt (om det nu finns några skillnader). Familjekonfidensgrad 95% 2p
- Anta nu istället att det finns väldigt många anläggningar och de fyra är slumpmässigt utvalda. Skatta varianskomponenten för de slumpmässiga effekterna med ett 95% konfidensintervall genom att använda Satterthwait's metod. 2p
- Om vi återigen antar att vi endast är intresserade av de fyra anläggningarna men vi tror inte att data är normalfördelat. Hur skulle vi då kunna gå tillväga för att utföra ett parameterfritt test? 1p

