

Försättsblad till skriftlig tentamen vid Linköpings universitet



Datum för tentamen	2019-03-22
Sal (1)	<u>G32(2)</u>
Tid	14-19
Utb. kod	732G46
Modul	TEN1
Utb. kodnamn/benämning Modulnamn/benämning	Regressions- och variansanalys Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	4
Jour/Kursansvarig Ange vem som besöker salen	Isak Hietala (endast telefonjour)
Telefon under skrivtiden	013-281970 (endast 14-19)
Besöker salen ca klockan	
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	annelie.almquist@liu.se +46 13 28 29 34
Tillåtna hjälpmedel	Valfri räknedosa, kursbok Applied Linear Statistical Models (Kutner, Nachtsheim m fl) utan anteckningar. Det är dock tillåtet att ha markeringar och flikar. Flikarna får ha en liten anteckning.
Övrigt	
Antal exemplar i påsen	

Tentamen

Linköpings universitet
Institutionen för datavetenskap, IDA
Avdelningen för Statistik och maskininlärning, STIMA

Kurskod och namn:	732G46, Regression och variansanalys
Datum och tid:	2019-03-22, 14-19
Jourhavande lärare:	Isak Hietala
Tillåtna hjälpmedel:	Räknedosa. Läroboken <i>Applied linear statistical models</i> av Kutner, Nachtsheim m fl som inte får innehålla anteckningar men får ha markeringar och flärpar. Flärpar får ha en liten anteckning.
Betygsgränser:	Tentamen omfattar totalt 20p. G från 12p och VG från 16p.
Instruktioner:	Siffrorna i uppgifterna är påhittade. 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!

Uppgift 1 (10p)

Tabell 1: Datamaterial för uppgift 1

	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

Sprickbildningar 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.

Nedan har tre regressionmodeller 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	218,7796	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	-0,59988	1,33265	-0,450	0,658	
train	0,46187	0,06148	7,513	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	-2,56176	1,04172	-2,459	0,025	
Antal tåg	0,49056	0,04387	11,182	0,000	1,02
Material	2,74980	0,62833	4,376	0,000	1,02

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

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

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

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%	89,29%	83,00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-1,22131	1,31749	-0,927	0,368	
Antal tåg	0,42830	0,05797	7,388	0,000	1,94
Material	0,05950	1,82388	0,033	0,974	9,35
Interaktion	0,13188	0,08437	1,563	0,138	9,42

Regression Equation

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

- 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 intervallet (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 *Material* = 0. (3p)

Uppgift 2 (2p)

Vid kast av en symmetriskt sexsidig tärning så är sannolikheten p att få en sexa lika med en sjättedel. Ta fram maximum-likelihood-skattningen för att få en sexa då du kastar en tärning 15 gånger.

Ledning: Antingen får du en sexa eller så får du inte en sexa

Anta att du har kastat en tärning 15 gånger och räknat antalet prickar till 4, 6, 5, 6, 1, 1, 4, 1, 4, 1, 6, 3, 3, 5, 3. Använd observationerna för att skatta sannolikheten för att få en sexa. Tror du tärningen är symmetrisk? (2p)

Uppgift 3 (5p)

Tabell 2: Datamaterial för uppgift 3

1	85.0	84.0	84.4
2	85.1	85.2	84.9
3	84.7	84.5	85.2
4	84.2	84.1	84.6

Vid ett laboratorium har man många kemister anställda. Man vill undersöka om det de mäter märkbart olika. Därför väljer man slumpmässigt ut fyra kemister och de har fått i uppgift att bestämma procentuella halten metyl-alkohol i en viss kemisk sammansättning. Varje kemist gör tre mätningar. $SSE = 0,9533$

- Sätt upp en lämplig modell och testa på lämpligt sätt med ett test om alla anställda kemister på laboratoriet mäter olika. Signifikansnivå 10%. Visa hypoteserna. (3p)
- Beräkna ett 95% konfidensintervall för det förväntade värdet på procentuella halten metyl-alkohol bland alla mätningar som görs. (2p)

Uppgift 4 (3p)

- Vilken typ av regression är lämplig att använda om responsvariabeln har värdena 'Ja' eller 'Nej'? Hur hanteras denna nominella variabel? (1p)
- Vilken typ av regression är lämplig att använda om responsvariabeln är normalfördelad? (1p)
- Ge exempel på två olika länkfunktioner inom generaliserade linjära modeller. (1p)