



Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2011-04-07
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	TER1
Tid	8-13
Kurskod	732G21
Provkod	TEN1
Kursnamn/benämning Provnamn/benämning	Sambandsmodeller Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	3
Jour/Kursansvarig Ange vem som besöker salen	Oleg Sysoev/Lotta Hallberg
Telefon under skrivtiden	0735673482
Besöker salen ca kl.	10
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	carita.lilja@liu.se tel 1463
Tillåtna hjälpmedel	Valfri räknedosa, kursbok (Kutner m fl) utan anteckningar. Markering av texten i boken med färg är tillåtet.
Övrigt	
Vilken typ av papper ska användas, rutigt eller linjerat	rutigt
Antal exemplar i påsen	

Tentamen

Linköpings Universitet, Institutionen för datavetenskap, Statistik

Kurskod och namn: 732G21 Sambandsmodeller
Datum och tid: 2011-04-06, 8-13
Jourhavande lärare: Oleg Sysoev
Tillåtna hjälpmedel: Valfri räknedosa, kursbok (Kutner m fl) utan anteckningar. Markering av texten i boken med färg är tillåtet.
Betygsgränser: Tentamen omfattar totalt 20p. Godkänt från och med 12p, väl godkänt från och med 16p.

Redovisa och motivera tydligt alla dina lösningar!

Uppgift 1 (5p)

Tabellen nedan redovisar löner av lärare från flera skolor i USA (**Pay**, i 1000-tal dollar) och motsvarande omkostnader per elev (**Spend**, i 1000-tal dollar)

Pay	Spend
19,58	3,34
20,32	3,55
26,8	4,64
25,85	4,16
21,69	3,56

1. Rita ett diagram Pay mot Spend. Passar en linjär modell? Motivera (0.5p)
2. Anpassa den enkla linjära regressionsmodellen där Pay är responsvariabel och Spend är förklaringsvariabel och redovisa den anpassade regressionsekvationen. Redogör uträkningar! Tolka regressionskoefficienterna. (2.5p)
3. Beräkna konfidensintervall för lärarens lön om man vet att skolan spenderar 4000 dollar per elev. Använd konfidensgrad 99% och tolka konfidensintervallet. (2p)

Uppgift 2 (5p)

I denna uppgift ska ni undersöka sambandet mellan variablerna

- Price: Pris av ett hus (100-tal dollar)
 - SqFt: Husets area (kvadratfot)
 - Feats : **0**, om huset har mellan 0 och 3 utrustningsenheter (standardnivå), **1**, om huset har mellan 4 och 8 utrustningsenheter (komfortnivå)
1. En modell med responsen Price och förklaringsvariabler Feats och SqFtxFeats= $SqFt * Feats$ har anpassats och man har fått utskriften nedan.
 - a. Redovisa de anpassade regressionsekvationer för hus som har standardnivå av utrustning. Gör samma sak för hus med komfortnivå av utrustning. Tolka regressionskoefficienten som motsvarar SqFtxFeats. (1p)
 - b. Kan man påstå att feltermens varians är konstant eller att feltermen är oberoende eller att det finns multikollinearitet i modellen? Motivera. (1p)

Regression Analysis: PRICE versus Sqft; Feats; SqFtxFeats

The regression equation is

$$PRICE = 40,3 + 0,606 Sqft + 46,7 Feats - 0,0045 SqFtxFeats$$

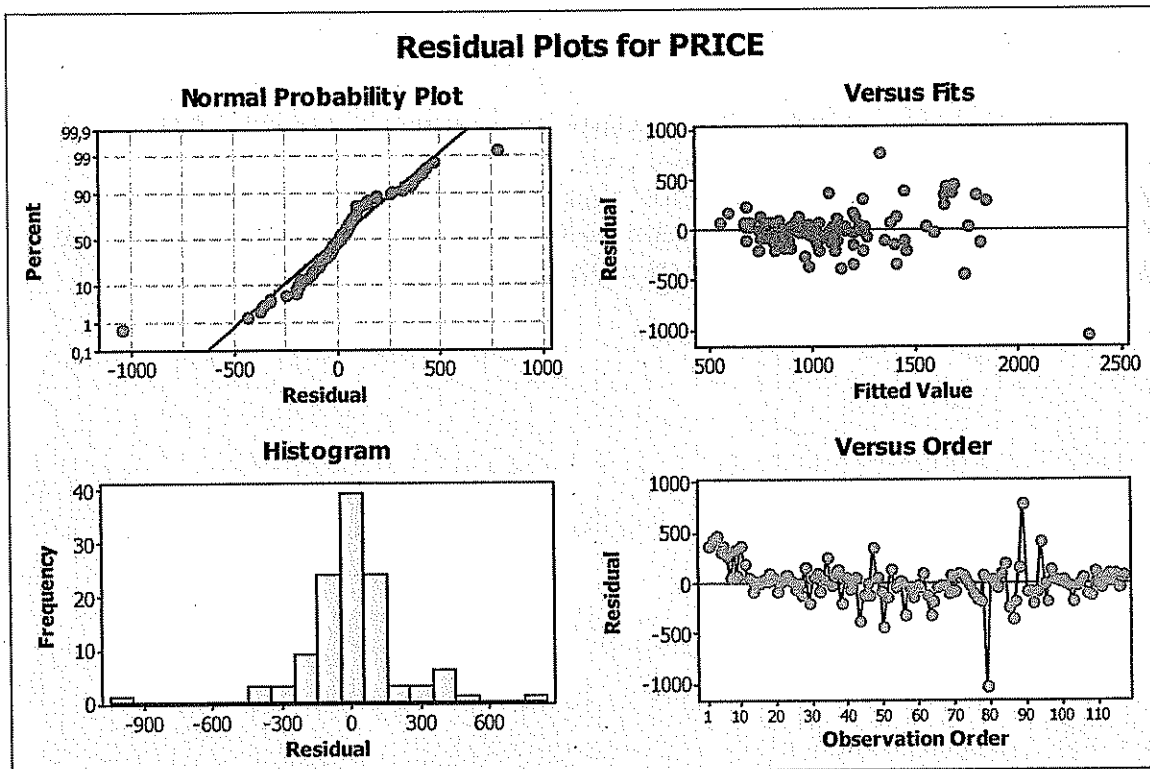
Predictor	Coef	SE Coef	T	P	VIF
Constant	40,33	65,51	0,62	0,539	
Sqft	0,60573	0,04003	15,13	0,000	1,209
Feats	46,69	48,05	0,97	0,333	1,588
SqFtxFeats	-0,00448	0,01506	-0,30	0,767	1,725

S = 205,323 R-Sq = 71,6% R-Sq(adj) = 70,9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	12025159	4008386	95,08	0,000
Residual Error	113	4763788	42157		
Total	116	16788947			

Source	DF	Seq SS
Sqft	1	11981915
Feats	1	39516
SqFtxFeats	1	3728



- Använd utskriften ovan och konstruera 99% simultana konfidensintervall för regressionskoefficienter som motsvarar variablerna Feats och SqftxFeats. Indikerar dessa intervall på att dessa variabler kan droppas från modellen? Motivera. (1.5p)
- En logistisk modell med responsvariabel Feats och förklaringsvariabler Price och SqFt har anpassats (utskriften nedan). Pröva om det finns regressionsrelation genom att genomföra Likelihood ratio test med konfidensgraden 90%. Ange hypotesen och mothypotesen, beslutsregeln och slutsatsen. Vilket är P-värde av testet? (1.5p)

Binary Logistic Regression: Feats versus PRICE; Sqft

Link Function: Logit

Response Information

Variable	Value	Count
Feats	1	64 (Event)
	0	53
Total		117

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-2,14572	0,738516	-2,91	0,004			
PRICE	0,0010897	0,0011610	0,94	0,348	1,00	1,00	1,00
Sqft	0,0007457	0,0008195	0,91	0,363	1,00	1,00	1,00

Log-Likelihood = -74,001

Test that all slopes are zero: G = 13,159, DF = 2, P-Value = 0,001

Uppgift 3 (10p)

En viktig egenskap hos biljardbollar är deras elasticitet, som mäts på en skala mellan 0 och 100, där höga värden är vad man strävar efter. Vid ett experiment utgår man från en traditionell plastmassa vid tillverkningen och vill se om två tillsatser A och B påverkar elasticiteten. Man gör 30 satser av plastmassor, som delas upp i tre lika delar. Den ena delen får ingen tillsats (kod 0), den andra tillsatsen A (kod 1) och den tredje tillsatsen B (kod 2).

Följande data har erhållits

Row Elasticitet Tillsats

Row	Elasticitet	Tillsats
1	51	0
2	45	0
3	49	0
4	66	0
5	53	0
6	41	0
7	58	0
8	56	0
9	60	0
10	63	0
11	75	1
12	89	1
13	73	1
14	84	1
15	66	1
16	85	1
17	73	1
18	71	1
19	78	1
20	65	1
21	39	2
22	43	2
23	51	2
24	34	2
25	54	2
26	43	2
27	42	2
28	41	2
29	37	2
30	44	2

En envägs variansanalys har gjorts i Minitab, där följande resultat har erhållits (där mycket är struket i utskriften och ersatt med --):

One-way ANOVA: Elasticitet versus Tillsats

Analysis of Variance for Elastici

Source	DF	SS	MS	F	P
Tillsats	--	5654,9	--	--	--
Error	--	1472,1	--		
Total	--	7127,0			

Individual 95% CIs For Mean

Based on Pooled StDev

Level	N	Mean	StDev	-----+-----+-----+-----	
0	10	54,200	7,927	(---*---)	
1	10	75,900	8,048		(---*---)
2	10	42,800	5,996	(---*---)	
-----+-----+-----+-----					

- a) Ange den modell som ligger bakom variansanalysen i form av en modell med behandlingseffekter (med μ och τ). Beräkna skattningar av modellens parametrar, dvs av väntevärdet, de tre effekterna samt av residualernas standardavvikelse.

3p

- b) Genomför ett test på signifikansnivån 1 % om det finns någon skillnad mellan de tre behandlingarna/tillsatserna. Var noggrann med att sätta upp noll- och mothypotes, beräkna testvariabel samt dra slutsats.

2p

- c) Vi vill jämföra de tre behandlingarna parvis med Tukeys metod. Beräkna lämpliga konfidensintervall med simultan konfidensgrad 99 %. Tolka de resultat som erhålls,

och påpeka särskilt om någon tillsats är bättre än att använda bara den traditionella massan.

3p

- d) Tillsatsen A verkar vara den bästa av de tre testade. Beräkna därför ett 99 % konfidensintervall för den förväntade elasticiteten med denna tillsats. 1p
- e) Effektmодellen i a) går att uttrycka som en multipel regressionsmodell. Sätt upp denna modell och ange förklaringsvariablerna. Skriv explicit upp X -matrisen och β -vektorn.

1p

