



Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2011-06-18
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	4
Jour/Kursansvarig Ange vem som besöker salen	Oleg Sysoev
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: 2011-06-18, 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 (6p)

I ett experiment undersöktes hur vikten hos kvinnor minskar pga motion. Fem kvinnor fick äta samma mat och springa ett givet avstånd varje dag i en vecka. Tabellen nedan redovisar hur viktändring beror på avståndet.

Avstånd(km)	Viktändring (kg)
1	-0,4
3	-3,6
10	-5
15	-5,4
20	-5,6

1. Rita ett diagram Viktändring mot Avstånd. Passar en linjär modell? Motivera (0.5p)
2. För att öka anpassningsgraden i modellen var det bestämt att försöka använda antingen $1/\text{Avstånd}$ eller $\sqrt{\text{Avstånd}}$ som förklaringsvariabel i den enkla linjära modellen. Vilken modell (av dem två) tycker du passar bäst? Motivera. (1p)
3. Anpassa den bästa modellen som du har valt i steg 2 och redovisa den anpassade regressionsekvationen. Redovisa uträkningar! Tolka regressionskoefficienterna. (2.5p)
4. Beräkna ett prediktionsintervall för viktändring för en kvinna som kommer att springa 5 km per dag. Använd konfidensgrad 95% och tolka prediktionsintervallet. (2p)

Uppgift 2 (4p)

I denna uppgift ska ett datamaterial analyseras som beskriver hur brottsligheten i USA beror på ett antal faktorer. Varje observation motsvarar en US delstat, variablerna är:

- R – antalet offer per miljon invånare
- Ex0 – finansiering av polis (\$), per capita
- NW – Antalet icke-vita personer per 1000 invånare
- W – medianinkomst (1000-tal \$)

1. En modell med responsten R och förklaringsvariabler Ex0, NW och W anpassades och man fick följande utskrift:

Regression Analysis: R versus Ex0; NW; W

The regression equation is

$$R = 18,9 + 1,04 \text{ Ex0} + 0,0531 \text{ NW} - 0,0426 \text{ W}$$

Predictor	Coef	SE Coef	T	P
Constant	18,88	39,27	0,48	0,633
Ex0	1,0431	0,2611	3,99	0,000
NW	0,05308	0,05765	0,92	0,362
W	-0,04263	0,09732	-0,44	0,664

$$S = 28,0370 \quad R\text{-Sq} = 50,9\% \quad R\text{-Sq}(\text{adj}) = 47,4\%$$

Analysis of Variance

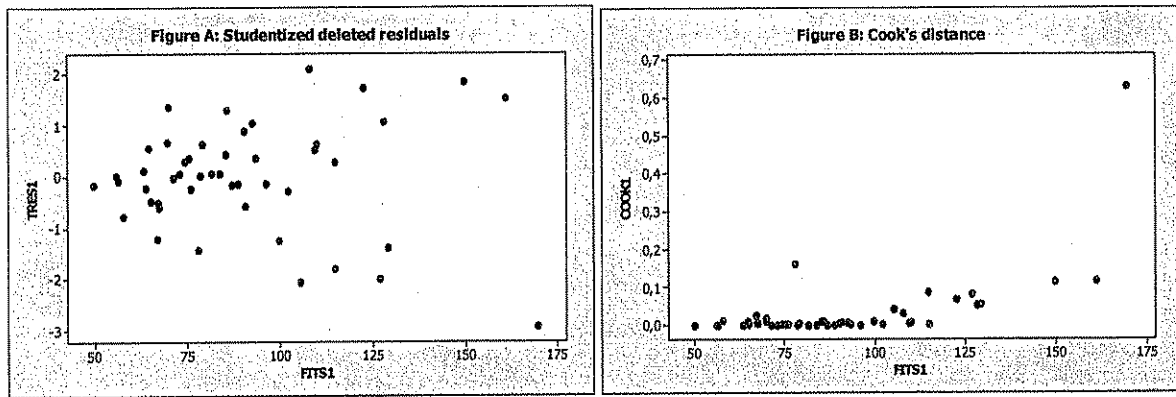
Source	DF	SS	MS	F	P
Regression	3	35008	11669	14,85	0,000
Residual Error	43	33801	786		
Total	46	68809			

Source	DF	Seq SS
Ex0	1	32533
NW	1	2324
W	1	151

Pröva om man kan ta bort NW och W från modellen genom att använda ett test med signifikansnivå 1%. Ange en hypotes och en mothypotes, beskriv beslutsregler och dra slutsatser. (2p)

2. Analysera följande utskrifter:

- a. Använd figur A för att identifiera om det finns uteliggare. Redovisa uträkningar (1p)
- b. Använd figur B och konstatera om det finns observationer som påverkar alla anpassade värdena. Motivera. (1p)



Uppgift 3 (4p)

Vid en fabrik tillverkas elspolar och man vill kontrollera tillverkningen genom att följa upp en viss karakteristika hos de tillverkade spolarna. Från fyra maskiner tar man från vardera på måfå ut 10 tillverkade spolar och erhåller resultat som sammanfattas i följande balanserade variansanalys:

ANOVA: karakt versus maskin

Factor	Type	Levels	Values
maskin	fixed	4	1 2 3 4

Analysis of Variance for karakt

Source	DF	SS	MS	F	P
maskin	3	602,50	200,83	28,09	0,000
Error	36	257,40	7,15		
Total	39	859,90			

Source	Variance component	Error term	Expected Mean Square for Each Term (using restricted model)
1 maskin		2	(2) + 10Q[1]
2 Error	7,150		(2)

Means

maskin	N	karakt
1	10	205,90
2	10	203,60
3	10	200,00
4	10	210,70

- Den modell som analyserats är en modell med fixa effekter, dvs att analyserna avser just dessa fyra maskiner. Sätt upp modellen vid envägs variansanalys som en effektmodell och definiera och tolka de olika parametrarna. (1p)
- Vad testar F-testet? Ange noll- och mothypotes samt dra slutsats på signifikansnivån 1 %. (1p)
- Jämför de fyra maskinerna med hjälp av Tukeys metod, så att alla jämförelserna tillsammans görs på signifikansnivån 1 % (familjekonfidensgraden är 99 %). (2p)

Uppgift 4 (6p)

Det visar sig att de fyra maskinerna som studerats ovan är ett urval ur ett stort antal maskiner som finns i fabriken. Vi har därför en situation med slumpmässiga effekter och gör då om analysen:

ANOVA: karakt versus maskin

Factor	Type	Levels	Values
maskin	random	4	1 2 3 4

Analysis of Variance for karakt

Source	DF	SS	MS	F	P
maskin	3	602,50	200,83	28,09	0,000
Error	36	257,40	7,15		
Total	39	859,90			

Source	Variance component	Error term	Expected Mean Square for Each Term (using restricted model)
1 maskin	19,368	2	(2) + 10(1)
2 Error	7,150		(2)

Means

maskin	N	karakt
1	10	205,90
2	10	203,60
3	10	200,00
4	10	210,70

- Sätt upp modellen med slumpmässiga effekter (modell II) och definiera och tolka de olika parametrarna. Använd utskriften för att beräkna punktskattningarna av de tre parametrarna (med beteckningar från boken: μ , σ_{μ}^2 , σ^2). (2p)
- Ange hypotesen vid F-testet och dra slutsats (på nivån 1 %). Blir det i praktiken någon annan slutsats än den i uppgift 1? (1p)
- Beräkna ett 99 % konfidensintervall för det genomsnittliga värdet hos karakteristikan hos hela maskinparken. Tolka intervallet. (1p)
- Skatta med ett 90% konfidensintervall kvoten $\sigma_{\mu}^2 / (\sigma^2 + \sigma_{\mu}^2)$. Tolka intervallet. (2p)