



Exp

Försättsblad till skriftlig tentamen vid Linköpings Universitet

(fylls i av ansvarig)

Datum för tentamen	<i>2009-04-07</i>
Sal	<i>TER1</i>
Tid	<i>8-13</i>
Kurskod	<i>732G21</i>
Provkod	<i>TEN1</i>
Kursnamn/benämning	<i>Sambandsmodeller</i>
Institution	<i>IDA</i>
Antal uppgifter som ingår i tentamen	<i>4</i>
Antal sidor på tentamen (inkl. försättsbladet)	<i>6</i>
Jour/Kursansvarig	<i>Kalle Wahlin</i>
Telefon under skrivtid	<i>0709-719096</i>
Besöker salen ca kl.	<i>10</i>
Kursadministratör (namn + tfnr + mailadress)	<i>Carita Lilja, 1463, carli@ida.liu.se</i>
Tillåtna hjälpmedel	<i>Kutner, Nachtsheim, Neter, Li: Applied Linear Statistical Models (utan anteckningar) samt valfri räknedosa.</i>
Övrigt (exempel när resultat kan ses på webben, betygsgränser, visning, övriga salar tentan går i m.m.)	

Tentamen

Linköpings Universitet, Institutionen för datavetenskap, Statistik

Kurskod och namn: 732G21 (HSTB20) Sambandsmodeller

Datum och tid: 2009-04-07, 8-13

Jourhavande lärare: Kalle Wahlin, 0709-719096

Tillåtna hjälpmedel: Valfri räknedosa, kursbok (Kutner m fl) utan anteckningar.

Betygsgränser: Tentamen omfattar totalt 20p. Godkänt från 12p.

Redovisa och motivera tydligt alla dina lösningar!

Uppgift 1 (6.5p)

En läkare drar ett slumpmässigt urval om 7 personer bland ett lands befolkning, och vill undersöka om det finns något samband mellan ålder och syreupptagningsförmåga (uttryckt i procent). Följande data erhålles.

Person nr	Ålder (x)	Syreupptagningsförmåga (y)
1	19	94
2	23	96
3	34	91
4	38	88
5	45	74
6	61	71
7	88	52

- Plotta sambandet mellan ålder och syreupptagningsförmåga. Vilken modell verkar lämplig för att undersöka om det finns något samband? Uttryck den valda modellen med de beteckningar som använts under kursen. (0.5p)
- Beräkna punktskattningar av parametrarna i den valda modellen. (1p)
- Tolka modellens parametrar i ord så att läkaren förstår det eventuella sambandet mellan ålder och syreupptagningsförmåga. (0.5p)
- Vilken syreupptagningsförmåga kan läkaren förvänta sig för en person som är 42 år gammal? (0.5p)

- e) Plotta residualerna mot x. Verkar modellen välanpassad? (1p)
- f) Läkaren kommer nu att tänka på att syreupptagningsförmågan kan vara olika beroende på om individen är rökare eller ej. Beskriv med de beteckningar som använts under kursen hur modellen kan utvecklas så att olika intercept och olika lutning för rökare respektive ickेरökare erhålles i en enda modellanpassning. (1p)
- g) Läkaren funderar också över om det finns någon samspelseffekt mellan ålder och rökning. Utveckla modellen du föreslagit i deluppgift f) så att även denna samspelseffekt kommer med. (1p)
- h) Nu har en tid förflutit sedan studien gjordes, och läkaren går tillbaka för att studera personerna. Hon noterar att två av dem avlidit och börjar fundera på om låg syreupptagningsförmåga ökar dödsrisken. Beskriv med de beteckningar som använts under kursen en lämplig typ av regressionsmodell för att undersöka om syreupptagningsförmågan kan förklara huruvida patienten avlidit eller ej. (1p)

Uppgift 2 (3.5p)

Ett bageri levererar bröd till många stormarknader i ett land. Bageriet vill veta om den höjd på hylla på vilken man skyltar bröden påverkar brödförsäljningen. Man studerar tre olika hyllhöjder: låg (L), mellan (M) och hög (H) hylla. För varje hyllhöjd har sex stormarknader med likvärdig försäljning slumpmässigt valts ut, och varje stormarknad har skyltat bröden på sin anvisade hyllhöjd under en månads tid varpå man efter månadens slut samlar in informationen om brödförsäljning bland de 18 stormarknaderna. Följande data (summerad försäljning uttryckt i 1000-tals kronor) erhålles.

<i>Låg (L)</i>	<i>Mellan (M)</i>	<i>Hög (H)</i>
58.2	73.0	52.4
53.7	78.1	49.7
55.8	75.4	50.9
55.7	76.2	54.0
52.5	78.4	52.1
58.9	82.1	49.9

För att undersöka om det finns någon skillnad i genomsnittlig försäljning (μ_L , μ_M och μ_H) mellan de tre hyllhöjderna anpassas modellen

$$y = \beta_0 + \beta_M D_M + \beta_H D_H + \varepsilon$$

där

$$D_M = \begin{cases} 1 & \text{om medelhög hyllhöjd} \\ 0 & \text{annars} \end{cases}$$

$$D_H = \begin{cases} 1 & \text{om hög hyllhöjd} \\ 0 & \text{annars} \end{cases}$$

Följande anpassning erhålles.

The regression equation is
Försäljning = 55.8 + 21.4 M - 4.30 H

Predictor	Coef	SE Coef	T	P	VIF
Constant	55.800	1.013	55.07	0.000	
M	21.400	1.433	14.93	0.000	1.333
H	-4.300	1.433	-3.00	0.009	1.333

S = 2.48193 R-Sq = 96.1% R-Sq(adj) = 95.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	2273.9	1136.9	184.57	0.000
Residual Error	15	92.4	6.2		
Total	17	2366.3			

Source	DF	Seq SS
M	1	2218.4
H	1	55.5

Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	77.200	1.013	(75.040; 79.360)	(71.486; 82.914)

Values of Predictors for New Observations

New Obs	M	H
1	1.00	0.000000

- Tolka parametrarna i regressionsmodellen i ord. (0.5p)
- Testa hypotesen $H_0: \beta_M = \beta_H = 0$. (0.5p)
- Vad är den genomsnittliga försäljningen när man använder hyllhöjd = hög? Svara med en punktskattning. (0.5p)
- Rapportera ett 95% intervall för den genomsnittliga försäljningen vid hyllhöjd = mellan. (0.5p)
- Finns det några indikationer på multikollinearitet i den valda modellen? (0.5p)

- f) Antag att vi misstänker att några enskilda observationer har stor påverkan på regressionsanpassningen. Föreslå i ord en metod för att undersöka detta. (1p)

3

Man är intresserad av attityden i en viss fråga från fyra olika politiska partier. Man valde därför 5 politiker slumpmässigt från vardera partiet. De fick svara på en skala från 1-100, där 100 är positivt inställd. Följande resultat erhöles:

Parti nr:	1	2	3	4
\bar{Y}_i	85	80	95	50
S_i	6	7	4	10

- a) Sätt upp en envägs-variansanalysmodell och pröva om det är skillnad mellan partierna. Visa att $MSE=50.25$. 2p
- b) Använd Tukey's metod för att pröva vilka partier som skiljer sig åt. Nivå är 5%. 2p

4

Man vill jämföra om husläkares skicklighet skiljer sig åt beroende på var läkaren tidigare har praktiserat. Därför har man valt tre kliniker och tre sjukhus

Faktor sjukhus

		Sjukhus 1	Sjukhus 2	Sjukhus 3
Faktor	Barnmedicin	91,7 74,9	86,3 88,1	82,3 78,7
		88,2 79,5	92,0 69,5	89,8 84,5
klinik	Kvinnoklinik	80,1 76,2	71,3 73,4	90,1 65,6
		70,3 89,5	76,9 87,2	74,6 79,1
	Diabetes och	71,5 49,8	80,2 76,1	48,7 54,4
	högt blodtryck	55,1 75,4	44,2 50,5	60,1 70,8

The ANOVA Procedure

Dependent Variable: skickl

Source	DF	Sum of		F Value	Pr > F
		Squares	Mean Square		
Model	7	2862.630000	408.947143	3.44	0.0088
Error	28	3328.500000	118.875000		
Corrected Total	35	6191.130000			

R-Square	Coeff Var	Root MSE	skickl Mean
0.462376	14.66440	10.90298	74.35000

Source	DF	Anova SS	Mean Square
sjukhus	2	104.630208	52.315104
klirik	2	2468.214583	1234.107292
sjukhus*klirik	3	289.785208	96.595069

- Betrakta de båda faktorerna som fixa och pröva på lämplig nivå :
om vi har interaktion mellan faktorerna, om sjukhusen skiljer sig åt samt
om det är skillnad mellan kliniker. 2p
- Anta att vi i a)-uppgiften inte har kunnat påvisa några interaktioner.
Vilken typ av kontraster är vi då intresserade av? 1p
- Upprepa analysen i a)-uppgiften men där båda faktorerna ses som
slumpmässiga. 2p
- Anta att man i en viss stad vill uttala sig om husläkarnas skicklighet.
Staden har endast tre sjukhus men en mängd kliniker. Förklara vilka
faktorer som ska vara fixa respektive slumpmässiga om det är så att man
vill kunna uttala sig om alla husläkare i staden. 1p