



Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2010-06-14
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	VALMAT
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	Kalle Wahlin
Telefon under skrivtiden	0709-719096
Besöker salen ca kl.	10.30
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Carita Lilja, 1463, carita.lilja@liu.se
Tillåtna hjälpmedel	Valfri räknedosa, kursbok (Kutner m fl) utan anteckningar
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 (HSTB20) Sambandsmodeller

Datum och tid: 2010-06-14, 8-13

Jourhavande lärare: Kalle Wahlin

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

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

Siffrorna i uppgifterna är påhittade.

Redovisa och motivera tydligt alla dina lösningar!

Uppgift 1 (4p)

En dagligvarukedja vill undersöka om bredden på hyllutrymme påverkar försäljningen av en viss vara, eftersom man tänker sig att en vara som står i breda rader syns bättre och därmed lättare attraherar kunder. 12 av kedjans butiker valdes slumpmässigt ut, och följande resultat noterades

Butik	Hyllutrymme (cm)	Veckovis försäljning (tkr)
1	75	16
2	75	22
3	75	14
4	150	19
5	150	24
6	150	26
7	225	23
8	225	27
9	225	28
10	300	26
11	300	29
12	300	31

- a) Åskådliggör relationen mellan hyllutrymme och veckovis försäljning i en lämplig figur. Tolka din figur med avseende på om och i så fall hur sambandet ser ut. (1p)

En regressionsanpassning av relationen mellan hyllutrymme och veckovis försäljning görs i Minitab. Följande resultat erhålles.

Regression Analysis: Veckovis försäljning (tkr) versus Hyllutrymme (cm)					
The regression equation is					
Veckovis försäljning (tkr) = 14.5 + 0.0493 Hyllutrymme (cm)					
Predictor	Coef	SE Coef	T	P	
Constant	14.500	2.178	6.66	0.000	
Hyllutrymme (cm)	0.04933	0.01061	4.65	0.001	
S = 3.08058 R-Sq = 68.4% R-Sq(adj) = 65.2%					
Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	1	205.35	205.35	21.64	0.001
Residual Error	10	94.90	9.49		
Total	11	300.25			

- Råder det något samband mellan hyllutrymme och veckovis försäljning på 5% signifikansnivå? Motivera. (1p)
- Tolka med ord hur den veckovisa försäljningen ändras om hyllutrymmet ökar med en cm. (1p)
- Bestäm en punktskattning av den förväntade försäljningen i en butik som skyltar varan 175 cm brett. (1p)

Uppgift 2 (6p)

En forskare försöker utreda vilka faktorer som påverkar vilket resultat man fått på högskoleprovet. Hon drar ett slumpmässigt urval av personer som går eller nyss har slutat gymnasiet och samlar förutom provresultatet in information om ålder, kön (1 = kvinna 0 = man), senaste genomsnittsbetyg från gymnasieskolan samt huruvida man bor i tätort eller på landsbygden (1 = tätort 0 = landsbygd). Forskaren provar flera olika regressionsmodeller, se utskrifter nedan.

Modell 1:

The regression equation is
 Resultat = - 2.90 + 0.176 Ålder + 0.0628 Kön + 0.0693 Genomsnittsbetyg
 - 0.0705 Landsbygd/tätort

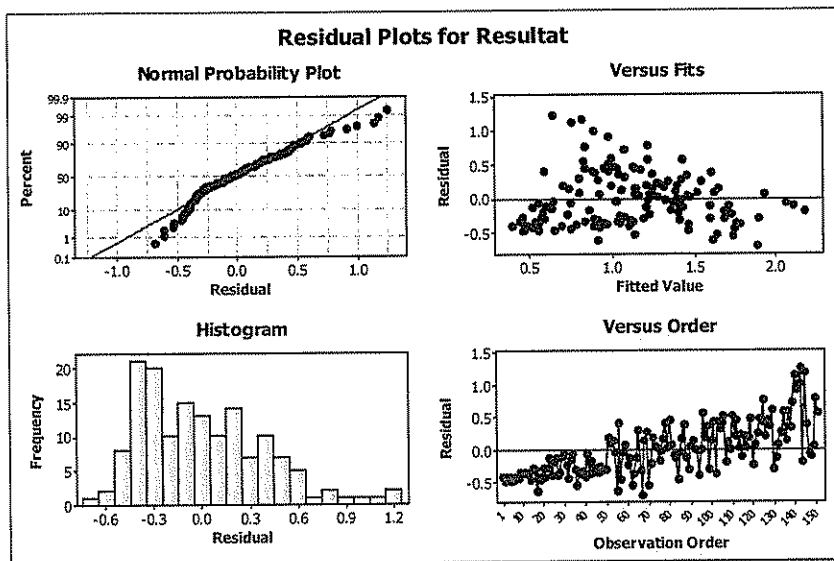
Predictor	Coef	SE Coef	T	P	VIF
Constant	-2.8991	0.4069	-7.12	0.000	
Ålder	0.17595	0.02329	7.55	0.000	1.088
Kön	0.06280	0.06627	0.95	0.345	1.026
Genomsnittsbetyg	0.069336	0.009351	7.41	0.000	1.093
Landsbygd/tätort	-0.07046	0.07053	-1.00	0.319	1.005

S = 0.395384 R-Sq = 51.2% R-Sq(adj) = 49.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	23.9859	5.9965	38.36	0.000
Residual Error	146	22.8240	0.1563		
Total	150	46.8099			

Source	DF	Seq SS
Ålder	1	15.0922
Kön	1	0.0076
Genomsnittsbetyg	1	8.7301
Landsbygd/tätort	1	0.1560



Modell 2:

The regression equation is

$$\text{Resultat} = -2.72 + 0.166 \text{ Ålder} - 0.389 \text{ Kön} + 0.0255 \text{ Kön} \cdot \text{Ålder} + 0.0692 \text{ Genomsnittsbetyg} - 0.0740 \text{ Landsbygd/tätort}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	-2.7238	0.5188	-5.25	0.000	
Ålder	0.16641	0.02915	5.71	0.000	1.695
Kön	-0.3886	0.8285	-0.47	0.640	159.665
Kön*Ålder	0.02550	0.04665	0.55	0.585	157.830
Genomsnittsbetyg	0.069189	0.009377	7.38	0.000	1.094
Landsbygd/tätort	-0.07395	0.07099	-1.04	0.299	1.013

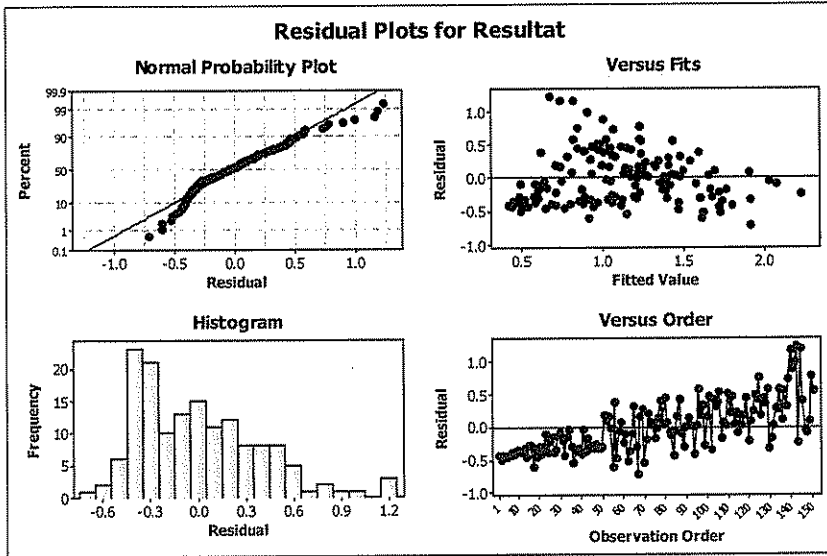
S = 0.396337 R-Sq = 61.3% R-Sq(adj) = 59.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	24.0329	4.8066	30.60	0.000
Residual Error	145	22.7771	0.1571		
Total	150	46.8099			

Source	DF	Seq SS
Ålder	1	15.0922
Kön	1	0.0076

Kön*Ålder 1 0.0634
 Genomsnittsbetyg 1 8.6992
 Landsbygd/tätort 1 0.1705



Modell 3:

The regression equation is

$$\text{Resultat} = - 2.83 + 0.174 \text{ Ålder} + 0.0684 \text{ Genomsnittsbetyg} - 0.0736 \text{ Landsbygd/tätort}$$

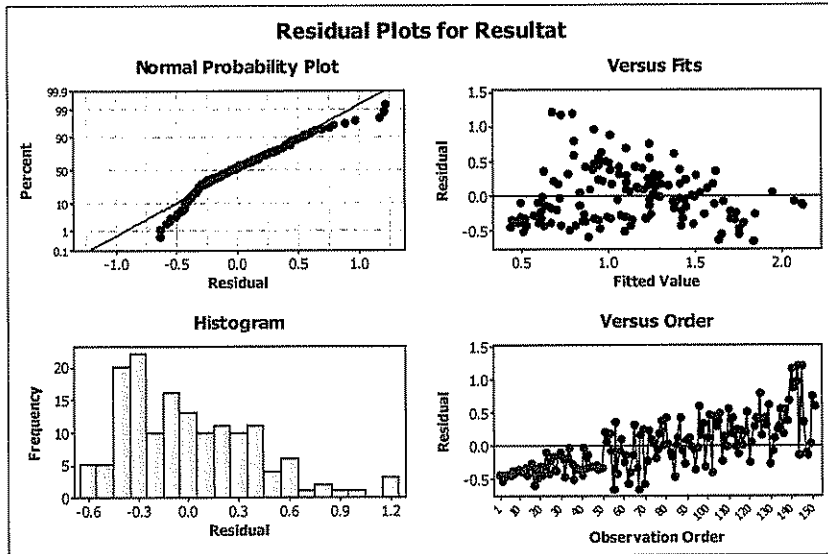
Predictor	Coef	SE Coef	T	P	VIF
Constant	-2.8273	0.3996	-7.07	0.000	
Ålder	0.17417	0.02321	7.51	0.000	1.080
Genomsnittsbetyg	0.068397	0.009295	7.36	0.000	1.081
Landsbygd/tätort	-0.07356	0.07043	-1.04	0.298	1.003

S = 0.395247 R-Sq = 50.9% R-Sq(adj) = 49.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	23.8455	7.9485	50.88	0.000
Residual Error	147	22.9644	0.1562		
Total	150	46.8099			

Source	DF	Seq SS
Ålder	1	15.0922
Genomsnittsbetyg	1	8.5829
Landsbygd/tätort	1	0.1704



Modell 4:

The regression equation is

$$\text{Resultat} = -2.87 + 0.173 \text{ Ålder} + 0.0688 \text{ Genomsnittsbetyg}$$

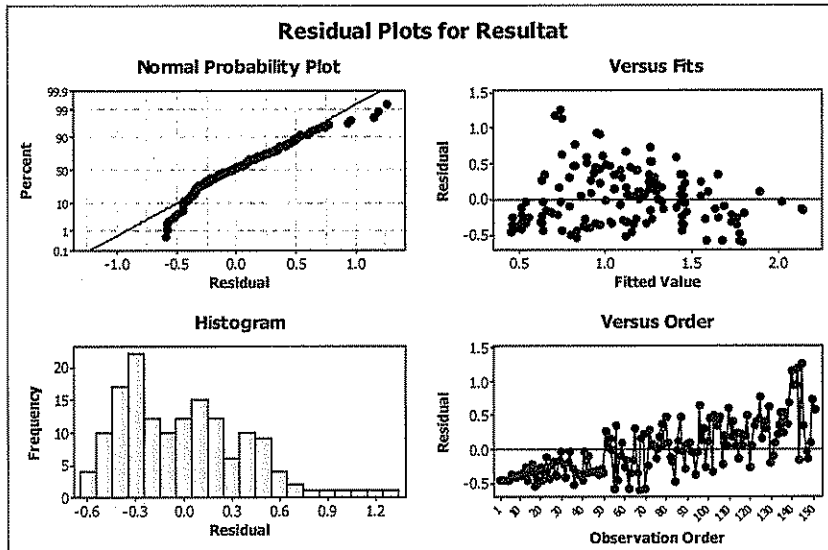
Predictor	Coef	SE Coef	T	P	VIF
Constant	-2.8657	0.3981	-7.20	0.000	
Ålder	0.17311	0.02319	7.46	0.000	1.078
Genomsnittsbetyg	0.068829	0.009289	7.41	0.000	1.078

S = 0.395369 R-Sq = 50.6% R-Sq(adj) = 49.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	23.675	11.838	75.73	0.000
Residual Error	148	23.135	0.156		
Total	150	46.810			

Source	DF	Seq SS
Ålder	1	15.092
Genomsnittsbetyg	1	8.583



- Vilken modell anser du är den bästa? Motivera! (1p)
- Hur många personer ingick i forskarens studie? (0.5p)
- Uttryck den valda modellen med de beteckningar som använts under kursen. (0.5p)
- Är modellen välanpassad? Red ut denna fråga genom att svara på om det finns några indikationer på multikollinearitet samt genom residualanalys. (1p)
- Tolka parametrarna i din valda modell i ord så att forskaren förstår sambandet mellan provresultat och var och en av variablerna. (1p)
- Baserat på din valda modell, bestäm en punktskattning av vilket resultat på högskoleprovet som en 20-årig kvinna från Stockholm med genomsnittsbetyg 15.7 kan förvänta sig. (1p)
- Forskaren misstänker att vissa personer i materialet med väldigt höga gymnasiebetyg inte är representativa för populationen, och önskar utreda om dessa utgör outliers (alltså outliers med avseende på förklarande variabel). Föreslå med ord en lämplig metod för att utreda detta. (1p)

3

På ett företag planerar man att internutbilda personalen på den nya datorutrustningen. Man vill göra en bedömning av om det är skillnad i effekten mellan fyra olika utbildningsprogram. Därför slumpades 32 personer till de fyra programmen, åtta till vardera program. Dessa personer fick efter utbildningens slut utföra ett test. Resultat av testet:

Medelvärde Standardavvikelse

Program 1	80 72 69 74 90 77 81 70	76,625	6,96804
Program 2	68 64 77 71 68 69 75 75	70,875	4,45413
Program 3	84 78 79 92 71 83 90 70	80,875	8,00781
Program 4	64 72 67 81 70 69 71 79	71,625	5,75543

- a) Pröva om det finns signifikant skillnad mellan programmen med ett test. Ange också vilken modell du har och vilka antaganden som måste vara uppfyllda för att få utföra testet. Ange hypoteserna. Visa alla uträkningar. Signifikansnivå 5%. 4p
- b) Pröva om det är skillnad i populationsmedelvärde mellan program 2 och 3. Signifikansnivå 5%. 1p
- c) Låt μ_i vara populationsmedelvärde för program nr i. Pröva nollhypotesen $\frac{\mu_1 + \mu_2}{2} = \frac{\mu_3 + \mu_4}{2}$ med signifikansnivå 5%. 2,5p
- d) Om antalet program som fanns tillgängliga var mycket stort och man hade erhållit de fyra programmen med ett slumpmässigt urval, hur ser då modellen ut? Beräkna också ett 95% konfidensintervall för σ_μ^2 . 2,5p