

# Tentamen

Linköpings Universitet, Institutionen för datavetenskap, Statistik

---

Kurskod och namn: 732G21 Sambandsmodeller

Datum och tid: 2015-03-07, 8-13

Jourhavande lärare: Karl Wahlin

Tillåtna hjälpmedel: Valfri räknedosa, kursbok Applied Linear Statistical Models (Kutner, Nachtsheim m fl) utan anteckningar. Det är dock tillåtet att ha

Betygsgränser: markeringar och flikar. Flikarna får ha en liten anteckning.

Tentamen omfattar totalt 20p. Godkänt från 12 poäng, väl godkänt från 16 poäng.

Siffrorna i uppgifterna är påhittade.

---

**Redovisa och motivera tydligt alla dina lösningar!**

## Uppgift 1 (7p)

Tabellen nedan redovisar hur försäljningen av en produkt (Sale, uttryckt i 1000-tals produkter) beror på investeringar i reklam (Advert, uttryckt i 10000-tals dollar)

Advert	Sale
0,06	0,61
3,55	6,58
1,02	2,30
2,66	3,51
4,83	8,29

- Rita ett diagram Sale mot Advert. Passar en linjär modell? Motivera. (1p)
- Anpassa den enkla linjära regressionsmodellen där Pay är responsvariabel och Spend är förklaringsvariabel och redovisa den anpassade regressionsekvationen. Tolka regressionskoefficienterna. (3p)

- c) Beräkna ett konfidensintervall för antalet produkter som man kan sälja om man investerar 30000 dollar i reklam. Tolka konfidensintervallet. (3p)

## Uppgift 2 (6p)

I denna uppgift ska ni undersöka sambandet mellan variablerna

- nox: Kväveoxidnivå
- lstat: Nivå av fattigdom (högre siffror betyder fattigare områden)
- Rate: Nivå av kriminalitet (1=hög, 0=låg eller medel)

som visar läget i olika områden i närheten av Boston.

1. En linjär modell med responserna Nox och förklaringsvariabler lstat, Rate och Rate\*lstat har anpassats och man har fått utskriften nedan.
  - a) Redovisa de anpassade regressionsekvationerna för områden med låg eller medel nivå av kriminalitet och för områden med hög nivå av kriminalitet. Tolka regressionskoefficienten som motsvarar Rate\*lstat. (1p)
  - b) Se plotten nedan och motivera om man kan påstå att feltermens varians är konstant. (1p)

### Regression Analysis: nox versus Rate; lstat

Analysis of Variance

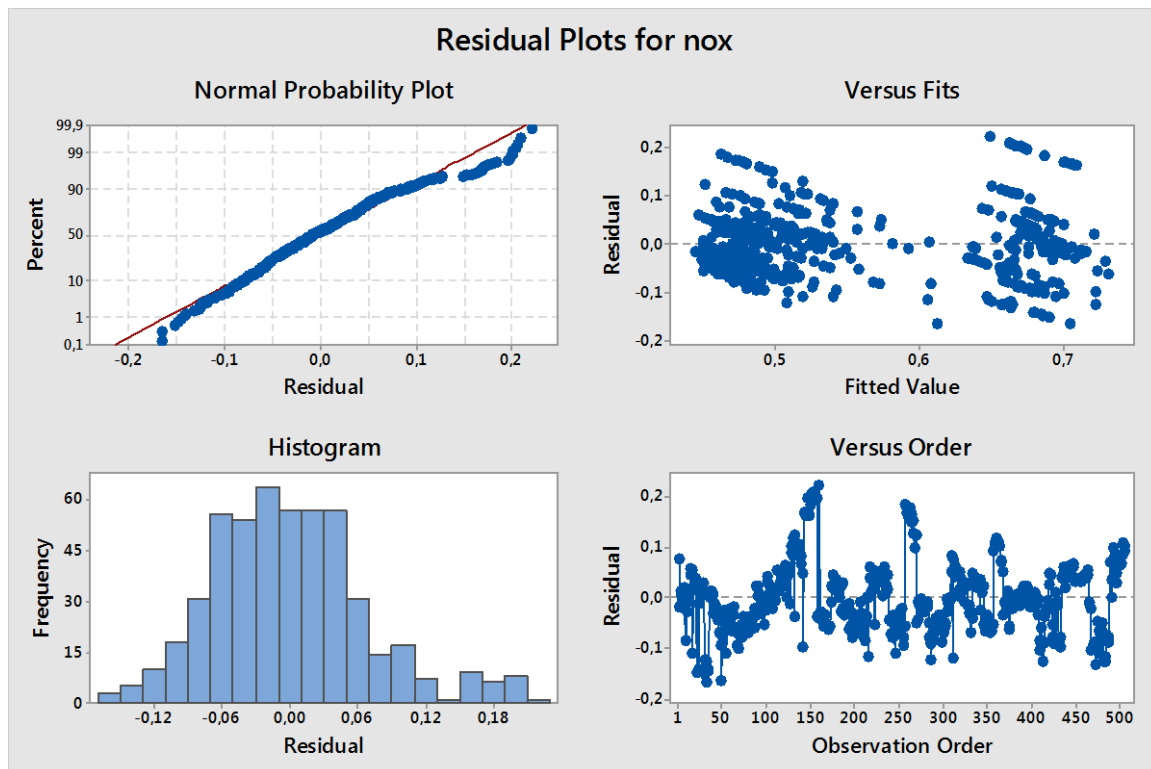
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	4,35173	1,45058	299,76	0,000
Rate	1	0,72863	0,72863	150,57	0,000
lstat	1	0,31544	0,31544	65,18	0,000
Rate*lstat	1	0,04639	0,04639	9,59	0,002
Error	502	2,42923	0,00484		
Lack-of-Fit	467	2,21693	0,00475	0,78	0,863
Pure Error	35	0,21229	0,00607		
Total	505	6,78096			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,0695636	64,18%	63,96%	63,52%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0,43239	0,00816	53,00	0,000	
Rate	0,1965	0,0160	12,27	0,000	6,05
lstat	0,005851	0,000725	8,07	0,000	2,80
Rate*lstat	-0,00315	0,00102	-3,10	0,002	9,79



- c) Använd utskriften ovan och konstruera 99-procentiga simultana konfidensintervall för regressionskoefficienter för variablerna Rate och Rate\*Istat. Indikerar dessa intervall att variablerna Rate och Rate\*Istat kan uteslutas från modellen? Motivera. (2p)
- d) En logistisk modell med responsvariabel Rate och förklaringsvariabler nox och Istat har anpassats (utskriften nedan). Undersök om det finns en regressionsrelation genom att genomföra ett likelihood ratio-test med konfidensgraden 90 procent. Ange hypoteser, beslutsregel och slutsatsen. Vilket är testets p-värde? (2p)

### Binary Logistic Regression: Rate versus nox; Istat

Method

Link function Logit  
 Rows used 506

Response Information

Variable	Value	Count	
Rate	1	174	(Event)
	0	332	
	Total	506	

Deviance Table

Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	2	399,521	199,760	399,52	0,000
nox	1	249,584	249,584	249,58	0,000

lstat	1	4,195	4,195	4,19	0,041
Error	503	251,774	0,501		
Total	505	651,294			

#### Model Summary

Deviance	Deviance	
R-Sq	R-Sq(adj)	AIC
61,34%	61,04%	257,77

#### Coefficients

Term	Coef	SE Coef	Coef	VIF
Constant	-18,18		1,69	
nox	29,17		2,94	1,04
lstat	0,0519		0,0256	1,04

### Uppgift 3 (7p)

En grisuppfödare vill maximera viktökningen bland sina grisar och vill undersöka vilket foder som leder till att grisarna går upp mest i vikt. Uppfödaren väljer ut 21 identiska grisar och delar upp dessa i fyra grupper. Efter två månader vägs grisarna och deras viktökning i kilo noteras:

Foder 1	Foder 2	Foder 3	Foder 4
2.3	4.5	0.9	5.7
2.1	6.4	0.7	7.3
1.7	6.2	2.3	7.9
3.6	3.9	1.2	6.2
2.4	4.9	0.6	7.6
	5.2		

- Pröva om det finns signifikant skillnad mellan fodren med ett test. Välj 5 procents signifikansnivå. (3p)
- Pröva om det är skillnad i populationsmedelvärde mellan foder 1 och foder 2 på 5 procents signifikansnivå. (2p)
- Låt  $\mu_i$  vara populationsmedelvärde för foder  $i$ . Pröva nollhypotesen  $\frac{\mu_1 + \mu_2}{2} = \frac{\mu_3 + \mu_4}{2}$  på 5 procents signifikansnivå. (2p)