Linköpings Universitet
IDA/STIMA-Statistik
LH

732G21 Sambandsmodeller, 7.5 av 15 hp

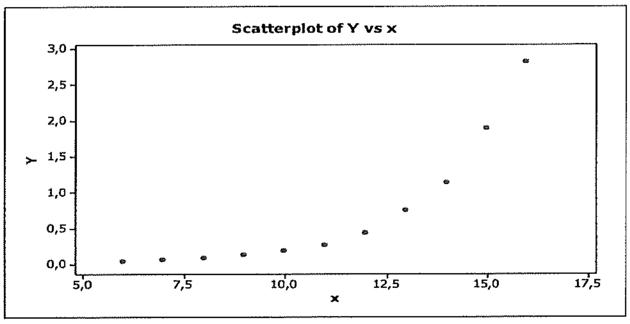
TENTAMEN I SAMBANDSMODELLER, 2019-06-12

Skrivtid: kl: 8-13
Hjälpmedel: Räknedosa. Läroboken: *Applied linear statistical models* av Kutner, Nachtsheim m fl som inte får innehålla anteckningar men får ha markeringar och flärpar. Flärpar får ha en liten anteckning.
Jourhavande lärare: Lotta Hallberg
Betygsgränser: För godkänt krävs minst 12 av 20 poäng och för väl godkänt krävs minst 16 av 20 poäng.

Redovisa och motivera kort alla dina lösningar
Tolka (om möjligt) alla dina resultat!

1

För att hitta sambandet mellan torrvikten (Y) och åldern (x) i dagar för 11 kyckling-embryon, så har tre olika modeller anpassats.
Nedan finner du ett spridningsdiagram över data och de tre anpassade modellerna.
I modell 2, x^2 = x^2.



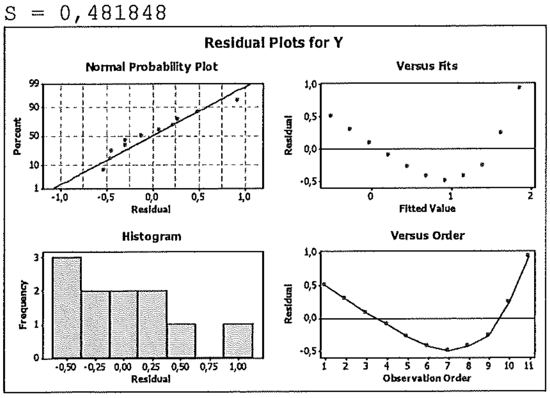
Modell 1

Regression Analysis: Y versus x

```
The regression equation is
Y = - 1,88 + 0,235 x

Predictor      Coef   SE Coef
Constant    -1,8845    0,5258
x           0,23507   0,04594

S = 0,481848
```



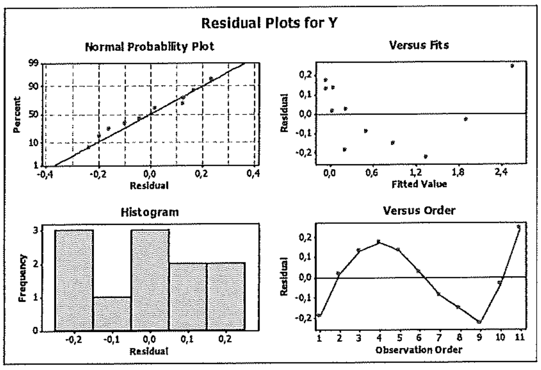
Modell 2

Regression Analysis: Y versus x; x^2

```
The regression equation is
Y = 3,26 - 0,785 x + 0,0463 x^2
```

Predictor	Coef	SE Coef
Constant	3,2603	0,6920
x	-0,7846	0,1329
x^2	0,046350	0,005991

```
S = 0,175488
```



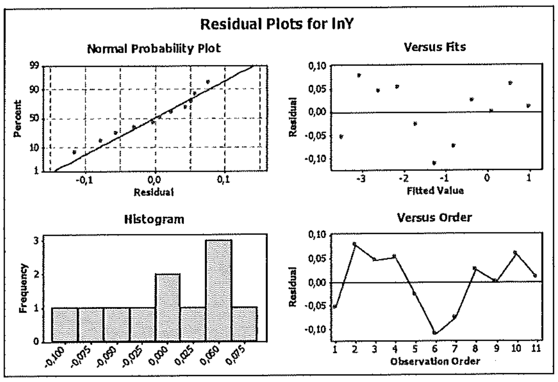
Modell 3

Regression Analysis: lnY versus x

```
The regression equation is
lnY = - 6,19 + 0,451 x
```

Predictor	Coef	SE Coef
Constant	-6,19211	0,07035
x	0,451033	0,006146

```
S = 0,0644635
```



a) Visa hur de tre modellerna ser ut, uttryckta i Y. Ange också de antaganden som måsta göras på modellerna. 3p

b) Prediktera torrvikten för en 17 dagars gammalt kyckling- embryo. En prediktion för var och en av de tre modellerna. 2p

c) Vilken av de tre modellerna ovan anpassas bäst till data? Använd all tillgänglig information för att fatta ditt beslut. 1p

2

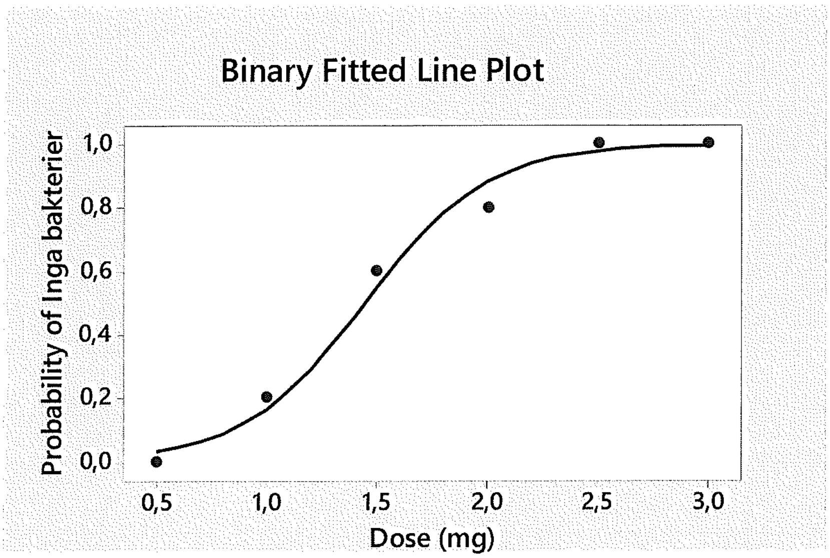
En forskare inom medicin vill undersöka hur dosering av en viss medicin påverkar floran av en viss bakterie hos vuxna. Forskaren utför därför ett experiment på 30 vuxna med 6 olika dosnivåer under fjorton dagar. De 30 vuxna tilldelas slumpmässigt de 6 dosnivåerna om 5 vuxna för varje dos.

Efter fjorton dagar undersöks om det finns några bakterier eller inte.

Responsvariabeln *No/Inga bacteria* är 1 om inga bakterier upptäcktes och 0 annars.

Följande modell har anpassats till dessa data.

Dose (mg)	No Bacteria	Trials
0,5	0	5
1,0	1	5
1,5	3	5
2,0	4	5
2,5	5	5
3,0	5	5



Binary Logistic Regression: No Bacteria versus Dose (mg)

Method

Link function	Logit
Rows used	6

Response Information

Variable	Value	Count	Event Name
No Bacteria	Event	18	Inga bakterier
	Non-event	12	
Trials	Total	30	

Deviance Table

Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	1	22,7052	22,7052	22,71	0,000
Dose (mg)	1	22,7052	22,7052	22,71	0,000
Error	4	0,9373	0,2343		
Total	5	23,6425			

Model Summary

Deviance R-Sq	Deviance R-Sq(adj)	AIC
96,04%	91,81%	21,68

Coefficients

Term	Coef	SE Coef	VIF
Constant	-5,25	1,99	
Dose (mg)	3,63	1,30	1,00

Odds Ratios for Continuous Predictors

	Odds Ratio	95% CI
Dose (mg)	37,5511	(2,9645; 475,6528)

Regression Equation

P(Inga bakterier) = exp(Y')/(1 + exp(Y'))

Y' = -5,25 + 3,63 Dose (mg)

Goodness-of-Fit Tests

Test	DF	Chi-Square	P-Value
Deviance	4	0,94	0,919
Pearson	4	0,70	0,951
Hosmer-Lemeshow	4	0,70	0,951

a) Tolka oddskvoten. 1p

b) För vilken dos är sannolikheten för att det inte ska finnas några bakterier större än 0,6? 2p

c) Prediktera sannolikheten för inga bakterier då dosnivån är 1,75mg. 2p

3

Ett företag inom kemi vill undersöka om föroreningen hos spillvattnet vid olika anläggningar skiljer sig åt. Fem spillvattenprover togs slumpmässigt från var och en av fyra anläggningar. Föroreningshalten mättes (lb/gal).
Resultat:

Anläggning	Prov 1	Prov 2	Prov 3	Prov 4	Prov 5	medelvärde	standardavvikelse
1	1,65	1,72	1,50	1,37	1,60	1,568	0,1366
2	1,70	1,85	1,46	2,05	1,80	1,772	0,2160
3	1,40	1,75	1,38	1,65	1,55	1,546	0,1592
4	2,10	1,95	1,65	1,88	2,00	1,916	0,1689

SSTO=0,9417

a) Sätt upp en en-vägs variansanalysmodell där anläggning är en faktor med fyra nivåer. Sätt upp lämpliga hypoteser och testa med ett F-test om föroreningen i spillvattnet skiljer sig åt mellan anläggningarna. 2p

b) Använd Tukey's metod för att undersöka vilka anläggningar som skiljer sig åt (om det nu finns några skillnader). Familjekonfidensgrad 95% 2p

c) Anta nu istället att det finns väldigt många anläggningar och de fyra anläggningarna är slumpmässigt utvalda. Skatta varianskomponenten för de slumpmässiga effekterna med ett 95% konfidensintervall genom att använda Satterthwait's metod. 2p

4

Låt X vara antalet inkommande samtal till en telefonväxel under en minut. X kan antas vara Poissonfördelad med parameter λ. Man observerade antalet samtal under 8 slumpmässigt valda minuter. De observerade värdena blev 2, 4, 4, 0, 0, 3, 5, 2 samtal.
Det gäller att $E[X] = \lambda$. Härled maximumlikelihood-skattningen av det förväntade antalet samtal under en minut.
Sannolikhetsfunktionen för en Poissonfördelad slumpvariabel är $f(x) = \frac{\lambda^x}{x!}e^{-\lambda}, \quad x = 0,1,\ldots$

3p