

TENTAMEN I SAMBANDSMODELLER, 2015-01-16

Skrivtid: kl: 14-19

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

Redovisa och motivera kort alla dina lösningar

Tolka (om möjligt) alla dina resultat!

1

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 olika anläggningar och 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

- Sätt upp en en-vägs variansanalysmodell där anläggning är en faktor med fyra nivåer. Testa med ett F-test om föroreningen i spillvattnet skiljer sig åt mellan anläggningarna. 2p
- Använd Tukey's metod för att undersöka vilka anläggningar som skiljer sig åt (om det nu finns några skillnader). Familjekonfidensgrad 95% 2p
- Anta nu istället att det finns väldigt många anläggningar och de fyra är slumpmässigt utvalda. Skatta varianskomponenten för de slumpmässiga effekterna med ett 95% konfidensintervall genom att använda Satterthwait's metod. 2p
- Om vi återigen antar att vi endast är intresserade av de fyra anläggningarna men vi tror inte att data är normalfördelat. Hur skulle vi då kunna gå tillväga för att utföra ett parameterfritt test? 1p

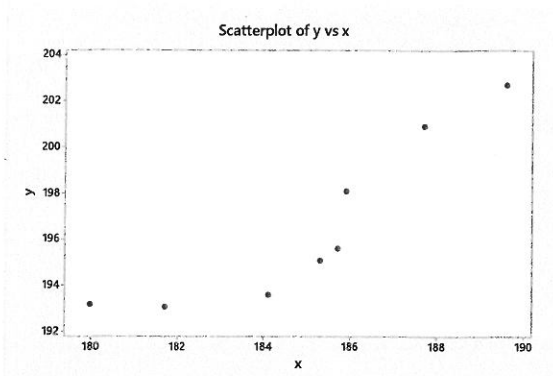
2

Ett företag vill skatta sambandet mellan:

x = kostnad för all typ av konstruktionsmaterial

Y = totala konstruktionskostnaden

Därför har följande regressionsmodeller anpassats. I utskriften är $x^2=x^2$ och $x^3=x^3$.



Modell 1

Regression Analysis: y versus x

Analysis of Variance

Source	DF	SS	MS	F-Value	P-Value
Regression	1	76,02	76,015	25,39	0,002
Error	6	17,96	2,994		
Total	7	93,98			

Model Summary

S	R-sq	R-sq(adj)
1,73029	80,89%	77,70%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	-1,2	39,2	-0,03	0,977
x	1,069	0,212	5,04	0,002

Regression Equation

$$y = -1,2 + 1,069 x$$

Modell 2

Regression Analysis: y versus x; x²

Analysis of Variance

Source	DF	SS	MS	F-Value	P-Value
Regression	2	87,197	43,598	32,14	0,001
Error	5	6,782	1,356		
Total	7	93,979			

Model Summary

S	R-sq	R-sq(adj)
1,16464	92,78%	89,90%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	4423	1541	2,87	0,035
x	-46,9	16,7	-2,81	0,038
x ²	0,1298	0,0452	2,87	0,035

Regression Equation

$$y = 4423 - 46,9 x + 0,1298 x^2$$

Modell 3

Regression Analysis: y versus x; x²; x³

Analysis of Variance

Source	DF	SS	MS	F-Value	P-Value
Regression	3	89,463	29,821	26,41	0,004
Error	4	4,516	1,129		
Total	7	93,979			

Model Summary

S	R-sq	R-sq(adj)
1,06253	95,19%	91,59%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	173135	119093	1,45	0,220
x	-2785	1933	-1,44	0,223
x ²	14,9	10,5	1,43	0,226
x ³	-0,0267	0,0188	-1,42	0,230

Regression Equation

$y = 173135 - 2785 x + 14,9 x^2 - 0,0267 x^3$

- Testa med ett partiellt F-test om en modell med ett tredjegradspolynom ger en bättre anpassning än den linjära modellen. Signifikansnivå 5%. 2p
- Om avseendet med modellen anpassningen är att prediktera den totala kostnaden, vilken modell anser du då vara den bästa? Motivera väl. 1p
- I modell 3 är F-testet signifikant men ingen av x-variablerna är signifikanta. Förklara vilken orsaken till denna motsägelse är. 1p
- Prediktera den totala kostnaden då material kostnaden är 190 genom att använda modell 2. Beräkna även ett 95% prediktionsintervall. Du får använda direkt att $s^2(\hat{Y}_h) = 1,49$ 3p

3

På ett sjukhus vill ledningen undersöka hur patienternas tillfredsställelse på sjukhuset hänger ihop med deras ålder samt deras ångestnivå.

Därför valde man slumpmässigt ut 46 patienter och mätte dessa variabler på dem.

$Y = 1$ om de är *tillfreds* och 0 om ej tillfreds

Ålder är mätt i år.

Ångest är mätt i ett index (höga värden är stor ångest).

Följande modell anpassades:

Binary Logistic Regression: tillfreds versus ångest; ålder

Method

Link function Logit

Rows used 46

Response Information

Variable	Value	Count	
tillfreds	1	22	(Event)
	0	24	
	Total	46	

Model Summary

Deviance	Deviance	
R-Sq	R-Sq(adj)	AIC
38,55%	35,41%	45,13

Coefficients

Term	Coef	SE Coef	VIF
Constant	12,30	4,22	
ångest	-2,30	1,74	1,04
ålder	-0,1898	0,0649	1,04

Odds Ratios for Continuous Predictors

	Odds Ratio	95% CI
ångest	0,1002	(0,0033; 3,0109)
ålder	0,8271	(0,7283; 0,9394)

Regression Equation

$$P(1) = \exp(Y') / (1 + \exp(Y'))$$

$$Y' = 12,30 - 2,30 \text{ ångest} - 0,1898 \text{ ålder}$$

- Tolka oddskvoten för ålder. 1p
- Prediktera sannolikheten att vara tillfreds för en 30 årig patient med ångestnivå 2. 1p
- Använd utskriften ovan för att avgöra om de två variablerna är signifikanta för modellen. 1p

4

Låt X vara normalfördelad med väntevärde μ och varians σ^2 . För att skatta variansen så tas ett stickprov om 10 observationer. Väntevärdet är okänt.

Observationerna är: 18,54 20,19 18,80 17,56 22,29 18,19 16,33 9,85 21,54 14,69

$$\sum x = 177,98 \quad \sum x^2 = 3284,34$$

Härled och visa att Maximum Likelihood-skattningen för σ^2 är $\frac{1}{10} \sum_{i=1}^{10} (x_i - \bar{x})^2$. Sätt slutligen in observationerna i skattningen. 3p

Täthetsfunktion till X är $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$, $-\infty < x < \infty$

Du behöver inte visa att det är ett maximum du får.