

TENTAMEN I SAMBANDSMODELLER, 2017-06-12

- 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!

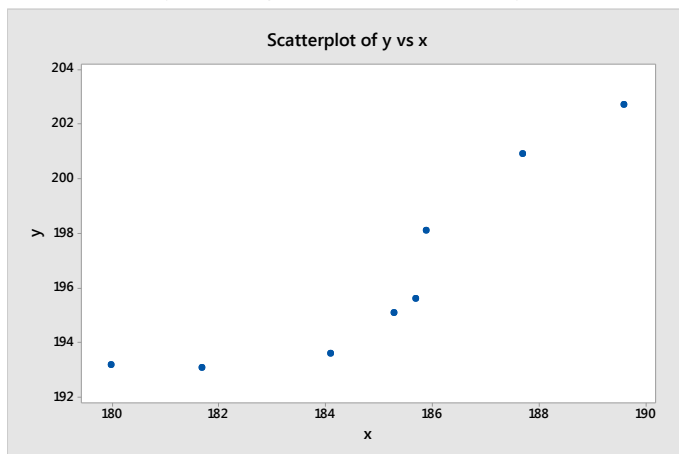
1

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 med endast x som förklarande variabel.
Signifikansnivå 5%. 2p
- I modell 3 är F-testet signifikant men ingen av x-variablerna är signifikanta. Förklara orsaken till denna motsägelse. 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

2

- Vilken typ av regression är lämplig att använda om responsvariabeln har värdena 'Ja' eller 'Nej'? Hur hanteras denna nominella variabel? 1p
- Vilken typ av regression är lämplig att använda om responsvariabeln är normalfördelad? 1p
- Ge exempel på två olika länkfunktioner inom generaliserade linjära modeller. 1p

3

Låt X vara antalet olyckstillfällen under en månad på en viss väglänk. X kan antas vara Poissonfördelad med parameter λ . Man observerade antalet olyckstillfällen under 10 månader. De observerade värdena blev 0, 0, 2, 1, 1, 0, 3, 0, 1, 2 olyckstillfällen

Det gäller att $E[X] = Var[X] = \lambda$. Härled maximumlikelihood-skattningen av variansen för X .

Sannolikhetsfunktionen för en Poissonfördelad slumpvariabel är $f(x) = \frac{\lambda^x}{x!} e^{-\lambda}$, $x = 0, 1, \dots$

3p

4

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

- 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
- 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 anläggningarna är slumpmässigt utvalda. Skatta varianskomponenten för de slumpmässiga effekterna med ett 95% konfidensintervall genom att använda Satterthwait's metod. 2p
- Anta att vi fortfarande har situationen som i c-uppgiften. Beräkna ett 95% konfidensintervall för den överlag förväntade föroreningsnivån bland alla anläggningar. 2p