

TENTAMEN I SAMBANDSMODELLER, 2018-04-06

- 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

En mäklarfirma har observationer på huspriser från den senaste tiden. Priset är i USD. För att ta reda på vad som är viktigt för huspriset så har man samlat in observationer på kvadratmeteryta m_2 , tomtyta $tomt$ samt antal rum rum .

a) Du ska med hjälp av framåtvalsmetoden välja den bästa modellen för dessa förklarande variabler (Modell 1-7). Välj även bästa modell med bakåtelimineringsmetoden. Förklara hur du går tillväga.

3p

b) Mäklaren har valt modellen med endast kvadratmeteryta, m_2 . Han har också observationer på antal badrum i husen. Denna variabel är kodad som två dummy-variabler. Variabeln $mindre$ är 1 om antalet badrum är 1 eller 1.5 och noll annars. Variabeln $medel$ är 1 om antalet badrum är 2 eller 2.5 och noll annars. Båda är noll om antalet badrum ligger mellan 3 och 4. Se modell 8. Pröva med ett test om antalet badrum (kodad på detta sätt) bidrar med förklaring av priset. Signifikansnivå 5%.

2p

c) Tolka regressionkoefficienten framför variabeln $mindre$ i modell 8.

1p

```
REG-proceduren
Modell: MODEL1
Beroendevariabel: pris

Antal lästa observationer      150
Antal använda observationer   150

Rot MSE          29946      R-kvadrat      0.4856
Beroende medel   153775     Just. R-kvadr. 0.4821
Koeff.var.       19.47364
```

```
Parameterskattningar

Variabel  DF  Parameter-  Standard-  t-värde  Pr > |t|
Skärning  1   skattning  fel        7.97     <.0001
m2        1   49.37473   4.17746    11.82    <.0001
```

Modell: MODEL2
Beroendevariabel: pris

Antal lästa observationer 150
Antal använda observationer 150

Rot MSE 36507 R-kvadrat 0.2354
Beroende medel 153775 Just. R-kvadr. 0.2303
Koeff.var. 23.74079

Parameterskattningar

Variabel	DF	Parameter- skattning	Standard- fel	t-värde	Pr > t
Skärning	1	144025	3312.31777	43.48	<.0001
tomt	1	9054.72274	1341.32739	6.75	<.0001

Modell: MODEL3
Beroendevariabel: pris

Antal lästa observationer 150
Antal använda observationer 150

Rot MSE 34115 R-kvadrat 0.3324
Beroende medel 153775 Just. R-kvadr. 0.3278
Koeff.var. 22.18477

Parameterskattningar

Variabel	DF	Parameter- skattning	Standard- fel	t-värde	Pr > t
Skärning	1	37969	13776	2.76	0.0066
rum	1	15966	1860.07275	8.58	<.0001

Modell: MODEL4
Beroendevariabel: pris

Antal lästa observationer 150
Antal använda observationer 150

Rot MSE 24552 R-kvadrat 0.6565
Beroende medel 153775 Just. R-kvadr. 0.6519
Koeff.var. 15.96628

Parameterskattningar

Variabel	DF	Parameter- skattning	Standard- fel	t-värde	Pr > t
Skärning	1	61100	6566.39960	9.30	<.0001
m2	1	46.24260	3.44458	13.42	<.0001
tomt	1	7760.02369	907.21706	8.55	<.0001

Modell: MODEL5
Beroendevariabel: pris

Antal lästa observationer 150
Antal använda observationer 150

Rot MSE 30047 R-kvadrat 0.4856
Beroende medel 153775 Just. R-kvadr. 0.4786
Koeff.var. 19.53961

Parameterskattningar

Variabel	DF	Parameter- skattning	Standard- fel	t-värde	Pr > t
Skärning	1	64221	12766	5.03	<.0001
m2	1	49.67304	7.50702	6.62	<.0001
rum	1	-140.54127	2934.12098	-0.05	0.9619

Modell: MODEL6
Beroendevariabel: pris

Antal lästa observationer 150
Antal använda observationer 150

Rot MSE 28383 R-kvadrat 0.5410
Beroende medel 153775 Just. R-kvadr. 0.5347
Koeff.var. 18.45774

Parameterskattningar

Variabel	DF	Parameter- skattning	Standard- fel	t-värde	Pr > t
Skärning	1	33408	11475	2.91	0.0042
tomt	1	8534.31750	1044.16760	8.17	<.0001
rum	1	15328	1549.55022	9.89	<.0001

Modell: MODEL7
Beroendevariabel: pris

Antal lästa observationer 150
Antal använda observationer 150

Rot MSE 24613 R-kvadrat 0.6572
Beroende medel 153775 Just. R-kvadr. 0.6501
Koeff.var. 16.00583

Parameterskattningar

Variabel	DF	Parameter- skattning	Standard- fel	t-värde	Pr > t
Skärning	1	56820	10493	5.42	<.0001
m2	1	43.55130	6.19092	7.03	<.0001
tomt	1	7792.53527	911.58054	8.55	<.0001
rum	1	1261.77232	2409.07169	0.52	0.6012

Modell: MODEL8
Beroendevariabel: pris

Antal lästa observationer 150
Antal använda observationer 150

Variansanalys

Källa	DF	Summa av kvadrater	Medel- kvadrat	F-värde	Sh. > F
Modell	3	1.306386E11	43546204009	49.92	<.0001
Fel	146	1.2735E11	872260043		
Korrigerad total	149	2.579886E11			

Rot MSE 29534 R-kvadrat 0.5064
Beroende medel 153775 Just. R-kvadr. 0.4962
Koeff.var. 19.20601

Parameterskattningar

Variabel	DF	Parameter- skattning	Standard- fel	t-värde	Pr > t	Typ I SS
Skärning	1	77029	16751	4.60	<.0001	3.547014E12
m2	1	43.01660	5.56636	7.73	<.0001	1.252713E11
mindre	1	-9688.92100	11113	-0.87	0.3847	5106693725
medel	1	5090.02578	9311.85140	0.55	0.5855	260623678

2

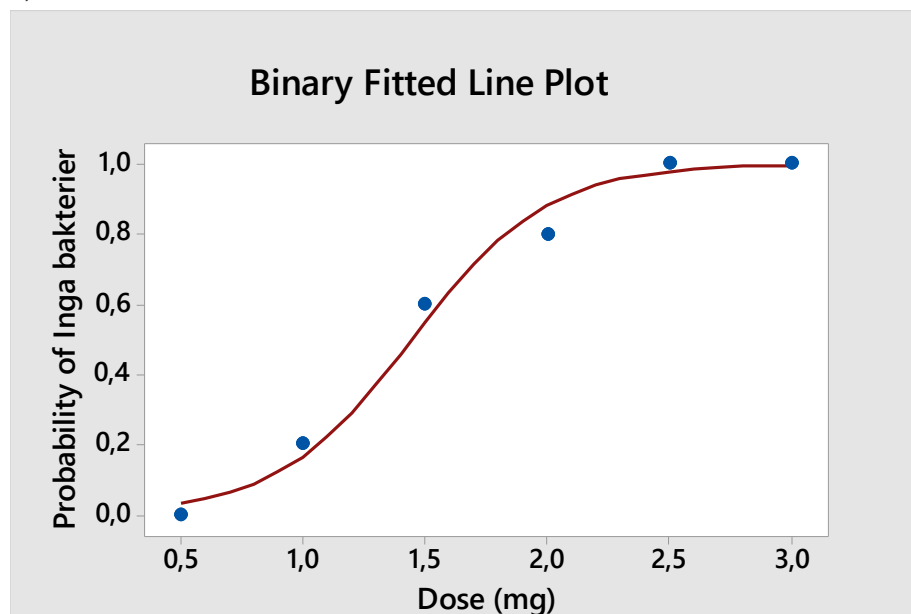
En forskare inom medicin vill undersöka hur dosering av en viss medicin påverkar florans av en viss bakterie hos vuxna. Forskaren utför därför ett experiment på 30 vuxna med 6 olika dosnivåer under fjorton dagar. De 30 vuxna tilldelas slumpmässigt de 6 dosnivåerna om 5 vuxna för varje dos.

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	Deviance	
R-Sq	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

$P(\text{Inga bakterier}) = \frac{\exp(Y^*)}{1 + \exp(Y^*)}$

$Y^* = -5,25 + 3,63 \text{ Dose (mg)}$

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

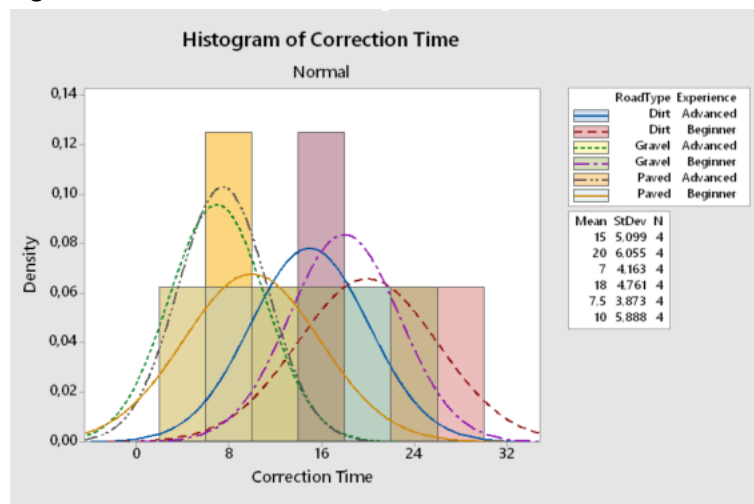
- Tolka oddskvoten. 1p
- För vilken dos är sannolikheten för att det inte ska finnas några bakterier större än 0,5? 2p
- Prediktera sannolikheten för inga bakterier då dosnivån är 1,25mg. 2p

3

En säkerhets-analytiker vill jämföra hur erfarna (advanced) och mindre erfarna (beginners) bilförare kör på tre olika typer av vägar. De tre typerna är Asfalterade (paved), oljegrus (gravel) samt grusväg (dirt). Som responsvariabel används *Tid i sekunder till undanmanöver = Y* (correction time).

En balanserad två-vägs ANOVA med interaktion ska anpassas som modell. Se modell för Formel 19,23 i boken. Där alfa-effekterna är typ av erfarenhet och beta-effekterna är vägtyp. Nedan ses en graf över *Y* för de sex faktor-kombinationerna (cellerna).

Figur 1:



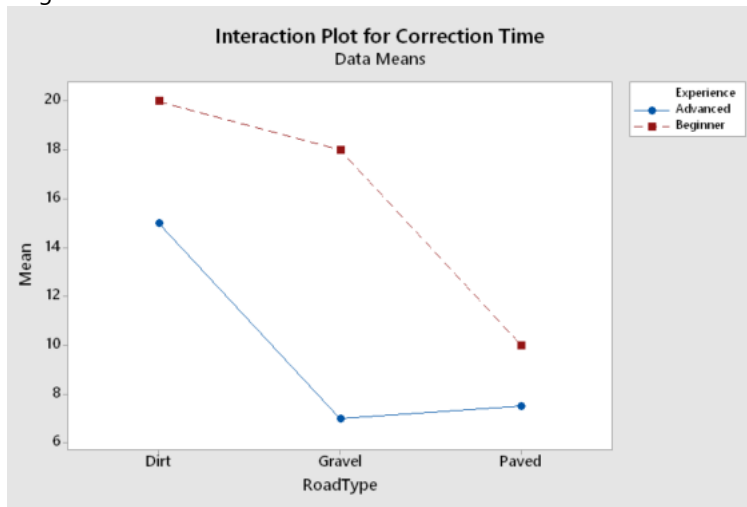
Results for Experience = Advanced Statistics

Variable	RoadType	Mean	StDev
Correction Time	Dirt	15,00	5,10
	Gravel	7,00	4,16
	Paved	7,50	3,87

Results for Experience = Beginner Statistics

Variable	RoadType	Mean	StDev
Correction Time	Dirt	20,00	6,06
	Gravel	18,00	4,76
	Paved	10,00	5,89

Figur 2:



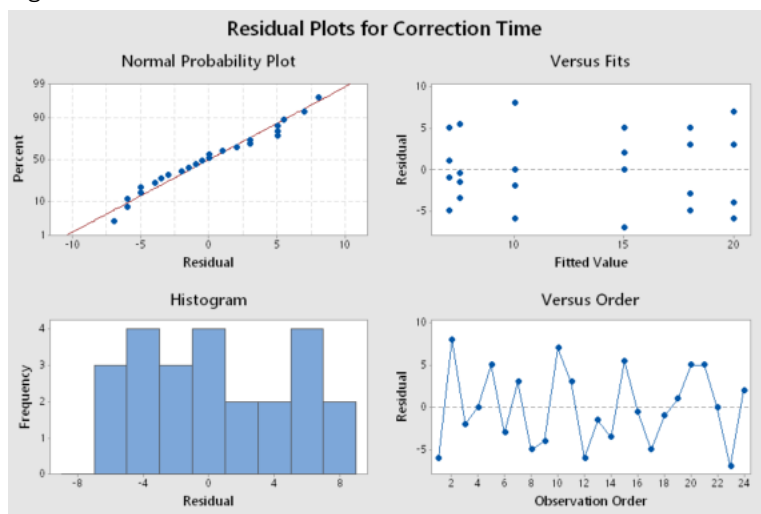
Factor Information

Factor	Type	Levels	Values
Experience	Fixed	2	Advanced; Beginner
RoadType	Fixed	3	Dirt; Gravel; Paved

Analysis of Variance for Correction Time

Source	DF	SS	MS
Experience	1	228,17	228,17
RoadType	2	308,33	154,17
Experience*RoadType	2	76,33	38,17
Error	18	457,00	25,39
Total	23	1069,83	

Figur 3:



- Punktkatta huvudeffekterna; $\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3$ samt μ . 2p
- Interaktionsplotten i Figur 2 tyder på att det finns interaktion. Testa på 5 % signifikansnivå om det finns interaktionseffekt. 2p
- Använd figur 1 och figur 3 för att avgöra om modellens antaganden angående variansen är rimlig att anta. 1p
- Skriv upp designmatrisen för modell 19,23 i boken. 1p

4

Vid ett löpande band produceras en viss typ av kretskort. Vid en kvalitetskontroll förfar man på följande vis: Vid en viss tidpunkt så undersöker man kretskort och räknar antalet kort tills man får en med någon defekt. Den defekta inkluderad i antalet. Detta görs 15 gånger. Resultat:

2 1 35 43 51 46 3 36 38 5 51 7 17 65 25

Låt X vara detta antal som är beskrivet ovan. X är då en geometriskt fördelad slumpvariabel med parameter $\pi > 0$. Sannolikhetfunktionen är given av

$$f(x) = (1 - \pi)^{x-1}\pi, \quad x = 1, 2, \dots$$

Maximum-Likelihood-skatta π . Sätt även in observationerna i skattningen.

3p