

Frame

1	1	10366	1
1	2	8162	0
1	3	5969	0
1	4	4496	0
1	5	3512	0
1	6	2201	0
1	7	1421	0
1	8	1121	0
1	9	826	2
2	1	3121	0
2	2	2937	0
2	3	2288	0
2	4	2015	0
2	5	1648	1
2	6	1310	2
2	7	927	0
2	8	710	3
2	9	606	0
3	1	3577	0
3	2	3286	1
3	3	2546	1
3	4	2219	2
3	5	1826	0
3	6	1386	1
3	7	988	2
3	8	684	4
3	9	449	3
4	1	4217	0
4	2	4214	0
4	3	3185	0
4	4	2560	4
4	5	1893	0
4	6	1384	2
4	7	849	2
4	8	470	2
4	9	280	5
5	1	5683	0
5	2	6385	1
5	3	5483	1
5	4	4687	6
5	5	3646	5
5	6	2411	12
5	7	1567	9
5	8	857	7
5	9	416	7
6	1	3042	0
6	2	4050	1
6	3	4290	4
6	4	4268	9
6	5	3529	9
6	6	2424	11
6	7	1409	10
6	8	663	5
6	9	284	3
7	1	670	0
7	2	1166	0
7	3	1482	0
7	4	1580	4
7	5	1336	6
7	6	924	10
7	7	556	7
7	8	255	4
7	9	104	1

```
Call: glm(formula = y ~ a + b + offset(log(n)), family = "poisson")
Deviance Residuals:
```

Min	1Q	Median	3Q	Max
-1.832904	-0.8559731	-0.3807713	0.4241323	2.176178

```
Coefficients:
```

	Value	Std. Error	t value
(Intercept)	-7.2810177	0.1724004	-42.2331955
a1	-2.1787235	0.5027471	-4.3336372
a2	-0.9586813	0.3663622	-2.6167581
a3	-0.0796293	0.2562401	-0.3107605
a4	0.1302123	0.2492969	0.5223182
a5	0.7221383	0.1716511	4.2070126
a6	0.9374875	0.1689896	5.5476038
b1	-3.1187052	0.8933875	-3.4908761
b2	-2.1717629	0.5299118	-4.0983481
b3	-1.4171269	0.3886831	-3.6459703
b4	0.0842177	0.2294316	0.3670711
b5	0.1235435	0.2421091	0.5102803
b6	1.0901213	0.2050192	5.3171658
b7	1.3289269	0.2177310	6.1035266
b8	1.7861285	0.2292702	7.7904966

```
(Dispersion Parameter for Poisson family taken to be 1)
```

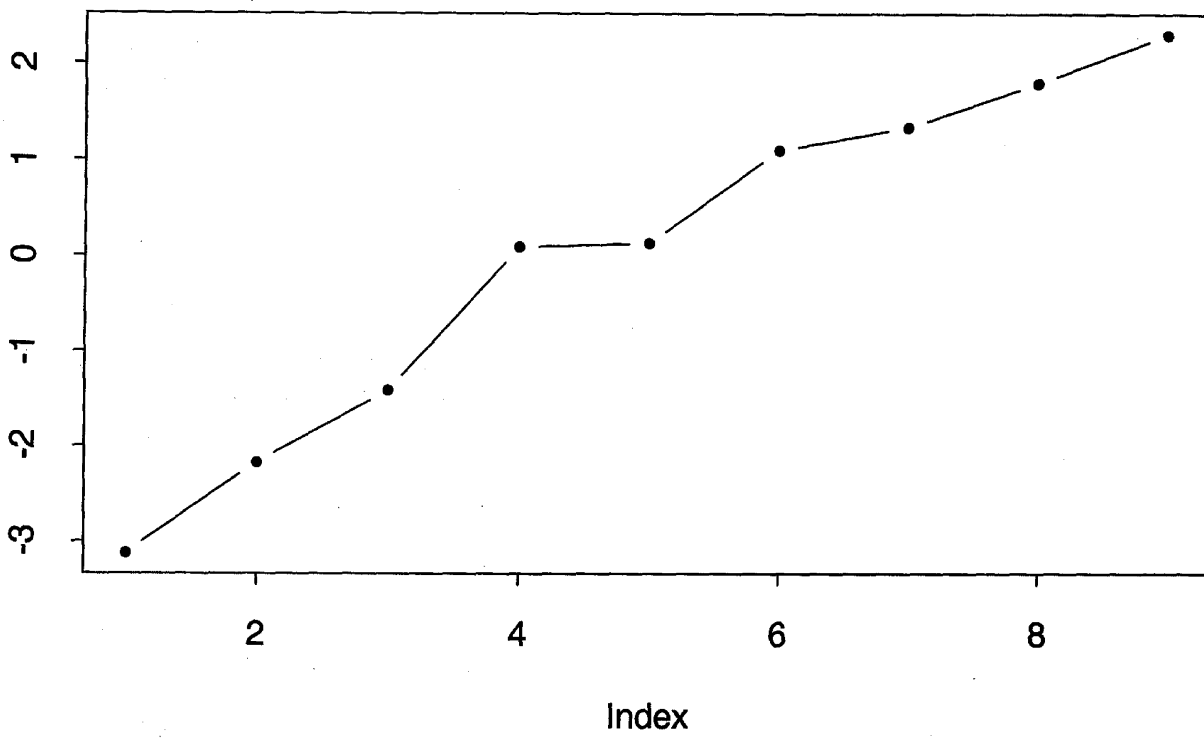
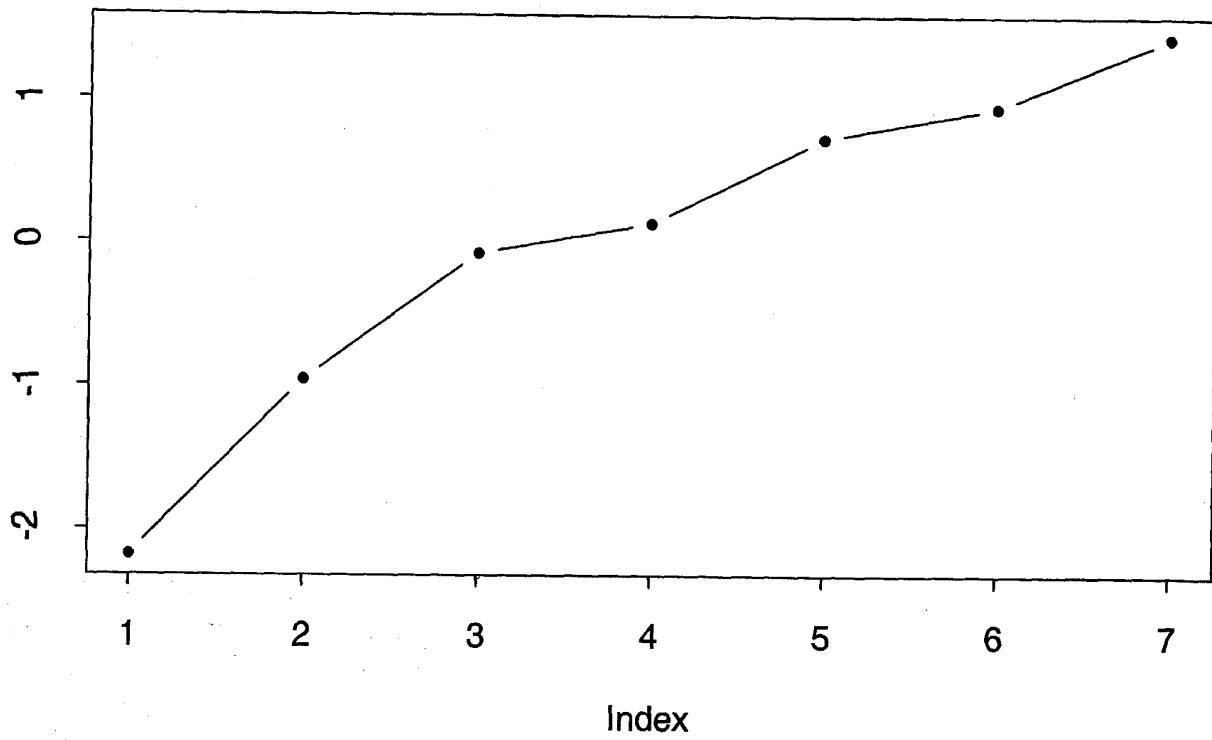
```
Null Deviance: 445.099 on 62 degrees of freedom
```

```
Residual Deviance: 51.47087 on 48 degrees of freedom
```

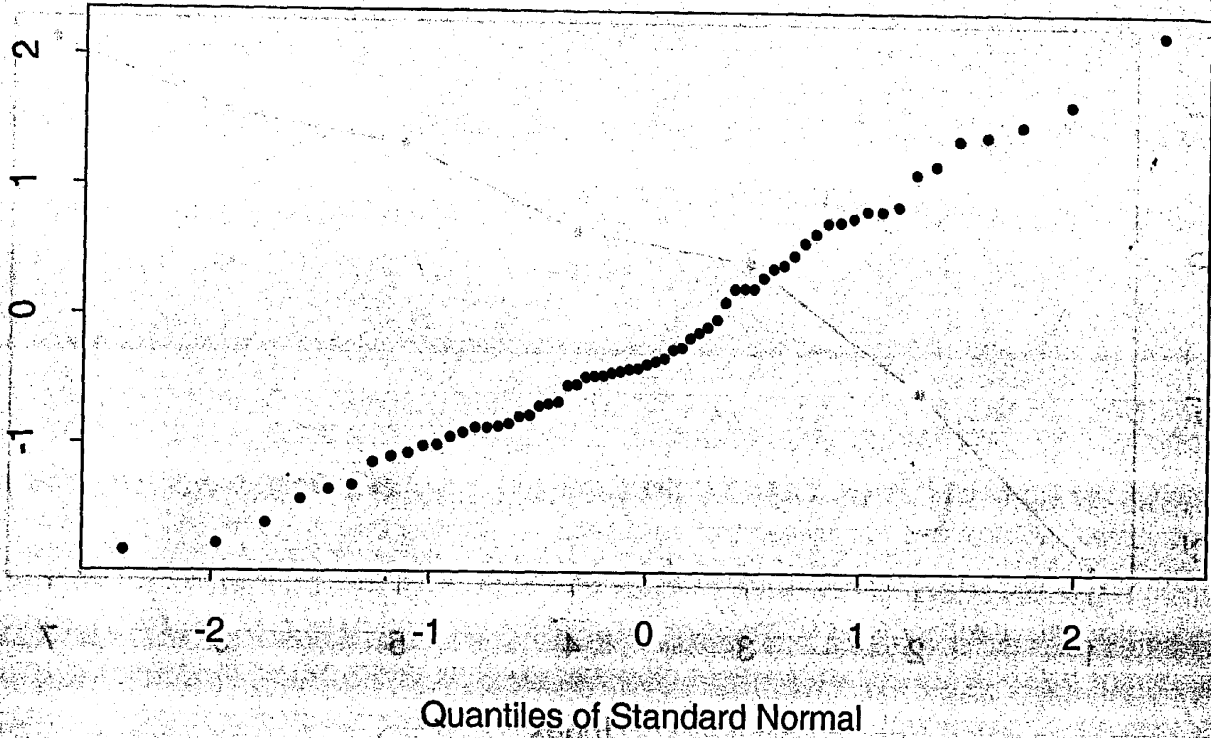
```
Number of Fisher Scoring Iterations: 5
```

```
> > > > > > > [1] 65.59833
```

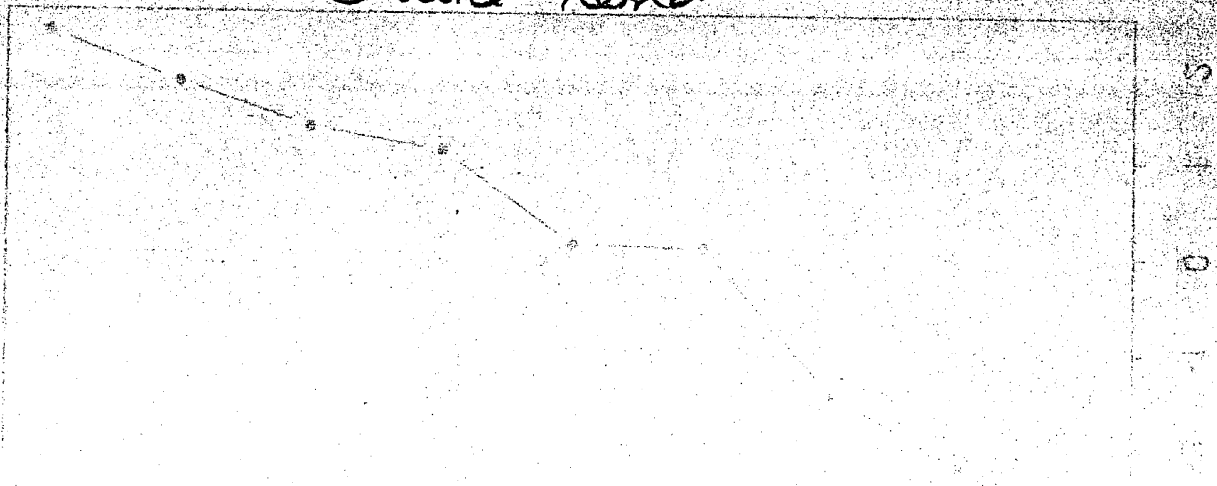
```
> [1] 0.953562
```



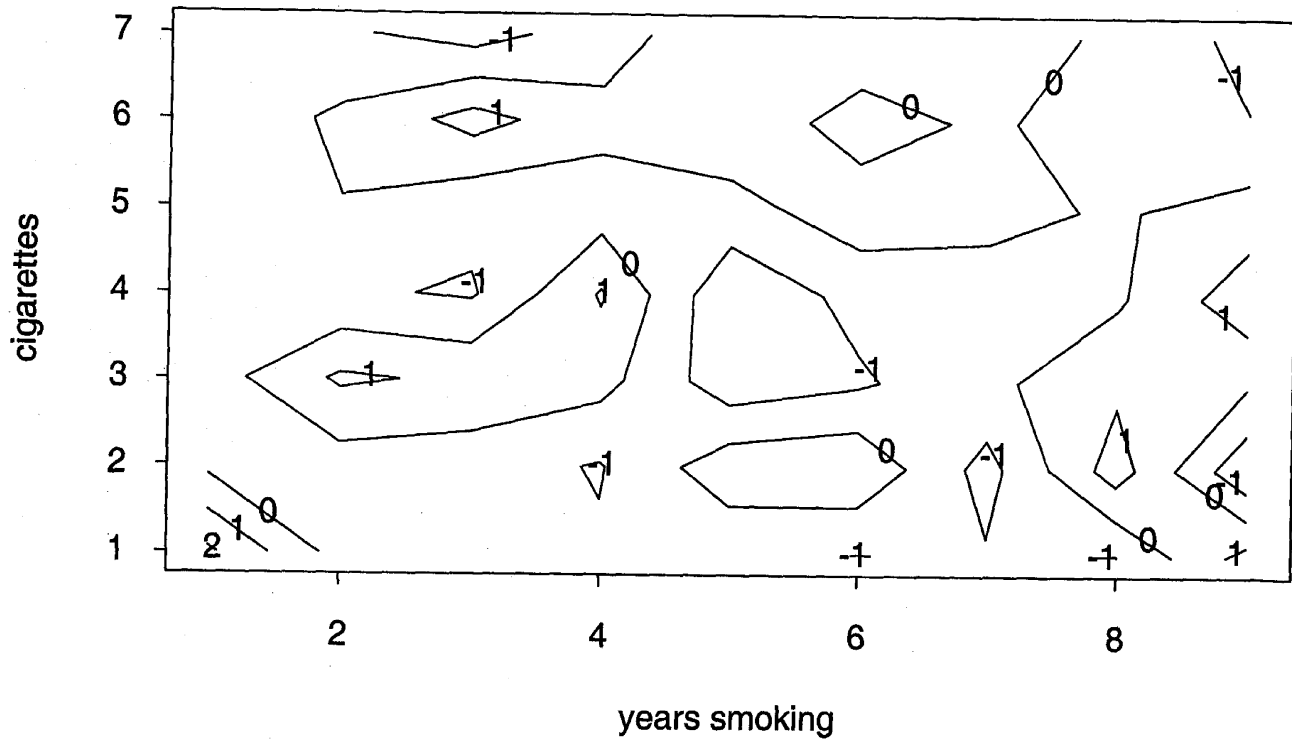
From data



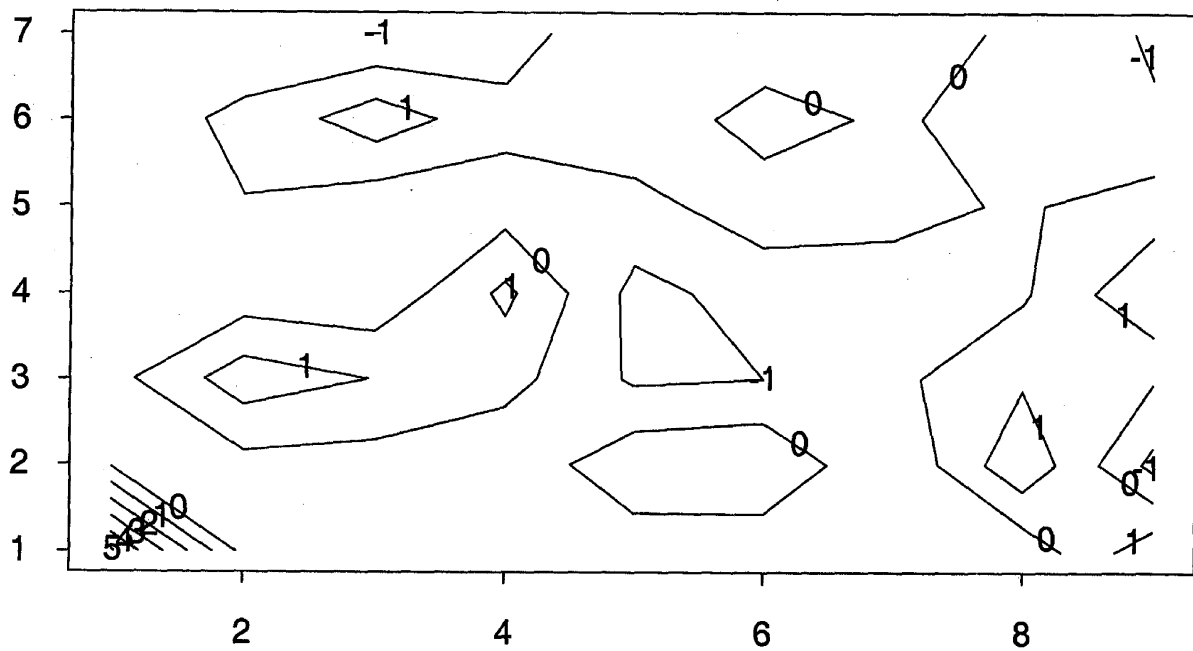
Deviance resid



Frome data - deviance residuals



Pearson residuals



(4)

24 Oct 01

Use of Poisson

Consider $Y_1, \dots, Y_k \sim \text{IP}(\mu_k)$

What is $P(Y_1, \dots, Y_k \mid Y_1 + \dots + Y_k = m)$? \otimes

$$P(Y_1 + \dots + Y_k = m) = \frac{1}{m!} \mu_+^m e^{-\mu_+}$$

$$\otimes = \frac{\prod_k \frac{1}{y_k!} \mu_k^{y_k} e^{-\mu_k}}{\frac{1}{m!} \mu_+^m e^{-\mu_+}} \quad y_1 + \dots + y_k = m$$

$$= m! \left(\prod_k \frac{1}{y_k!} \right) \left(\frac{\mu_k}{\mu_+} \right)^{y_k}$$

Write $\pi_k = \mu_k / \mu_+$

Likelihood for Poisson

$$P(y_1, \dots, y_k) = P(y_1, \dots, y_k \mid y_+ = m) P(m)$$

Suppose $P(m)$ does not depend on θ

(5)

24 Oct 01

led to the multinomial $\underline{Y} : M(m, \underline{\pi})$

The binomial, $B(m, \pi)$, is a particular case

Sometimes work with the scaled multinomial

$$\underline{Y}/m$$

$$E\{\underline{Y}/m\} = \underline{\pi}$$

$$\text{Var}\{\underline{Y}/m\} = \frac{1}{m} [\text{diag } \underline{\pi} - \underline{\pi} \underline{\pi}']$$

$$\text{var } Y_k = m \pi_k (1 - \pi_k) \quad \text{cp. Binomial}$$

$$\text{cov}\{Y_j, Y_k\} = -m \pi_j \pi_k \quad j \neq k$$

In our case $\underline{\pi}$ will depend on \underline{x} .

⑥

24 Oct 01

Product multinomial

$$\pi_{ij}$$

p.m.f.

$$\left[\prod_i \frac{y_{i+!}}{\prod_j \pi_{ij}} \prod_j \pi_{ij}^{y_{ij}} \right]$$

vs. multinomial $K = IJ$

Lizards Multinomial

Each classified by perch height and perch diameter

$$K \leq 4$$

$$\pi_{ab} \quad a, b = 1, 2$$

H: $\pi_{ab} = \pi_{.a} \pi_{.b}$?

In this case row totals were fixed by design.

Table 2-2
Counts for Structural Habitat Categories for *sagrei* Adult Male An
Bimini (Schoener [1968])

(a) Observed values

Perch Height (feet)	Perch Diameter (inches)		Totals
	≤ 4.0	> 4.0	
> 4.75	32	11	43
≤ 4.75	86	35	121
Totals	118	46	164

⑦

24 Oct 01

The fitted values are important

Data	32	11	Totals
	86	35	121
			164

Fitted values	30.9	12.1	43
	82.1	33.9	121
	118	46	164

$$\chi^2 = .178$$

$$p\text{-value } .673$$

Analysis of Deviance Table

Poisson model

Response: y

Terms added sequentially (first to last)

	Df	Deviance	Resid.	Df	Resid.	Dev	Pr (Chi)
NULL				3		71.53126	
A	1	38.64061		2		32.89065	0.0000000
B	1	32.71267		1		0.17798	0.0000000
A:B	1	0.17798		0		0.00000	0.6731154

(8)

24 Oct 2001

Loglinear model

$I \times J \times K \times L$ table

Counts Y_{ijkl}

$$m_{ijkl} = E Y_{ijkl}$$

Write, if additive in effects,

$$\log m_{ijkl} = \mu + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)}$$

Need constraints

$$\sum_i u_{1(i)} = 0, \sum_j u_{2(j)} = 0, \sum_k u_{3(k)} = 0, \sum_l u_{4(l)} = 0$$

Might require interactions, e.g.

$$u_{12(ij)} = \dots$$

①

24 Oct. 2001

log linear model

Contingency table

- Counts classified by 2 or more factors

 $t \times l$ \tilde{Y} : counts $t \times l$ $\tilde{m} = E\tilde{Y}$: expected cell frequencies $t \times p$ \tilde{X} : design matrix

each row represents a cell made of 0's and 1's

 $t \times l$

$$\tilde{\lambda} = \log \tilde{m} = \tilde{X} \tilde{\beta}$$

 $\tilde{m} = \tilde{m}(\tilde{\beta})$, $\tilde{\beta}$ subject to constraintsFit the model as if the Y 's are independent
Poisson

$$l = \sum_{i=1}^t y_i \log m_i - m_i + C$$

$$\text{Deviance} \quad 2 \sum_i \left[y_i \log y_i / \hat{m}_i - (y_i - \hat{m}_i) \right]$$

$$\sim \chi_{t-r}^2$$

 r : number of constraints

(2)

Data broken down by

24 Oct, 2001

Example, 3 categorical variables

$$\log m_{ijk} = \mu + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + \dots + u_{123(ijk)}$$

$$\sum_i u_{1(i)} = 0, \dots, \sum_i u_{12(ij)} = 0, \dots$$

$$\sum_i u_{123(ijk)} = 0$$

"Old statisticians never die, they just get broken down by age and sex."

29 Oct 01

An interesting log-linear example

Chromosome mapping.

Barley powdery mildew fungus

Loci A B C D E F

$$2^6 = 64 \text{ cells}$$

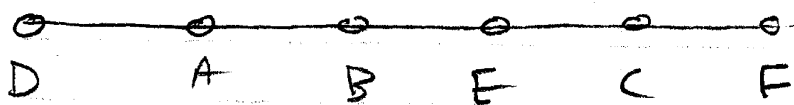
$n = 70$ many cells empty

Context suggested interactions of order ≥ 3 were 0

$$\binom{6}{2} = 15 \text{ second order interactions}$$

Use "backward elimination"

Selected model: CF, CE, BE, AD, AB



Edwards (1995) Introduction to Graphical Modelling.

[AF]	2.9974	2	0.3452
[BC]	1.8815	4	1.0000
[BE]	4.7105	4	0.5840
[CD]	1.4225	4	1.0000
[CE]	5.5934	4	0.3920
[DE]	1.0446	4	1.0000
[EF]	4.1595	4	0.6762

Removed edge [BC]

Model: CDEF,BDEF,ABDF

Deviance: 6.6571 DF: 32 P: 1.0000

Edge	Statistic	DF	P
[AF]	2.9974	2	0.3452
[BE]	4.5564	4	0.6289
[CD]	1.4032	3	1.0000
[CE]	5.4393	4	0.5288

Removed edge [CD]

Model: CEF,BDEF,ABDF

Deviance: 8.0603 DF: 36 P: 1.0000

Edge	Statistic	DF	P
[AF]	2.9974	2	0.3452
[BE]	4.5564	4	0.6289
[CE]	4.1912	2	0.2072
[DE]	1.4387	4	0.9554

Removed edge [DE]

Model: CEF,BEF,ABDF

Deviance: 9.4990 DF: 40 P: 1.0000

Edge	Statistic	DF	P
[AF]	2.9974	2	0.3452
[BD]	2.5611	2	0.2230
[BE]	4.4793	2	0.1323
[CE]	4.1912	2	0.2072
[DF]	3.0386	3	0.6152

Removed edge [DF]

Model: CEF,BEF,ABF,ABD

Deviance: 12.5375 DF: 44 P: 1.0000

Edge	Statistic	DF	P
[AF]	0.0216	1	1.0000
[BD]	0.0581	1	1.0000
[BE]	4.4793	2	0.1323
[CE]	4.1912	2	0.2072

Removed edge [AF]

Model: CEF,BEF,ABD

Deviance: 12.5692 DF: 46 P: 1.0000

Edge	Statistic	DF	P
[BD]	0.0581	1	1.0000
[BE]	4.4793	2	0.1323
[CE]	4.1912	2	0.2072


```

MIM->fact a2b2c2d2e2f2
MIM->sread abcdef
DATA->0 0 0 0 3 0 1 0 0 1 0 0 0 1 0 0
DATA->1 0 1 0 7 1 4 0 0 0 0 2 1 3 0 11
DATA->16 1 4 0 1 0 0 0 1 4 1 4 0 0 0 1
DATA->0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
End of data. 64 cells have been read.

```

```

MIM->satmod
MIM->time
Timer facility is now: on.
MIM->step e

```

```

Coherent Backward Selection
Decomposable models, chi-squared tests.
Critical value: 0.0500
Initial model: ABCDEF
Model: ABCDEF

```

Deviance: 0.0000 DF: 0 P: 1.0000

Edge	Statistic	DF	P
[AB]	29.3360	5	0.0000 +
[AC]	0.4027	1	1.0000
[AD]	20.0479	3	0.0002 +
[AE]	4.5529	3	1.0000
[AF]	0.4027	1	1.0000
[BC]	1.1790	2	1.0000
[BD]	3.3078	3	0.6364
[BE]	6.8444	3	0.1881
[BF]	1.1790	2	1.0000
[CD]	0.3684	2	1.0000
[CE]	5.7735	4	0.3624
[CF]	48.7976	6	0.0000 +
[DE]	2.7335	4	1.0000
[DF]	0.7711	3	1.0000
[EF]	4.1392	4	0.6702

Removed edge [AC]

Model: BCDEF, ABDEF

Deviance: 0.4027 DF: 16 P: 1.0000

Edge	Statistic	DF	P
[AE]	4.3729	3	1.0000
[AF]	6.0981	3	0.4087
[BC]	1.8815	4	1.0000
[CD]	1.4225	4	1.0000
[CE]	5.5934	4	0.3920

Removed edge [AE]

Model: BCDEF, ABDF

Deviance: 4.7756 DF: 24 P: 1.0000

Edge	Statistic	DF	P
------	-----------	----	---