Ejemplo:Cigarrillo–Alcohol–Marihuana

Los datos corresponden a un estudio realizado en la Wright State University School of Medicine and the United Health Services en Dayton Ohio, realizado a fin de relacionar el uso de alcohol (A), marihuana (M) y cigarrillos (C) en los estudiantes. Se realizó una encuesta entre los estudiantes del último año y se les preguntó si alguna vez habían consumido alguna de las tres cosas.

Table 7.3. Alcohol (A), Cigarette (C), and Marijuana (M) Use for High School Seniors

Alcohol Use	Cigarette Use	Marijuana Use	
		Yes	No
Yes	Yes	911	538
	No	44	456
No	Yes	3	43
	No	2	279

Table 7.4. Fitted Values for Loglinear Models Applied to Table 7.3

Alcohol Use	Cigarette Use	Marijuana Use	Loglinear Model				
			(A, C, M)	(AC, M)	(AM, CM)	(AC, AM, CM)	(ACM)
Yes	Yes	Yes	540.0	611.2	909.24	910.4	911
		No	740.2	837.8	438.84	538.6	538
	No	Yes	282.1	210.9	45.76	44.6	44
		No	386.7	289.1	555.16	455.4	456
No	Yes	Yes	90.6	19.4	4.76	3.6	3
		No	124.2	26.6	142.16	42.4	43
	No	Yes	47.3	118.5	0.24	1.4	2
		No	64.9	162.5	179.84	279.6	279

(AM, CM)

AC

$$1.0 = \frac{909.24 \times 0.24}{45.76 \times 4.76} = \frac{438.84 \times 179.84}{555.16 \times 142.16}$$

marginal AC

$$2.7 = \frac{(909.24 + 438.84)(0.24 + 179.84)}{(45.76 + 555.16)(4.76 + 142.16)}$$

Table 7.5. Estimated Odds Ratios for Loglinear Models in Table 7.4

	Conditional Association			Marginal Association		
Model	*AC*	*AM*	*CM*	*AC*	*AM*	*CM*
(A, C, M)	1.0	1.0	1.0	1.0	1.0	1.0
(AC, M)	17.7	1.0	1.0	17.7	1.0	1.0
(AM, CM)	1.0	61.9	25.1	2.7	61.9	25.1
(AC, AM, CM)	7.8	19.8	17.3	17.7	61.9	25.1
(ACM) level 1	13.8	24.3	17.5	17.7	61.9	25.1
(ACM) level 2	7.7	13.5	9.7			

Table 7.7. Goodness-of-Fit Tests for Loglinear Models Relating Alcohol (A), Cigarette (C), and Marijuana (M) Use

Model	G^2	X^2	df	P-value*
(A, C, M)	1286.0	1411.4	4	<0.001
(A, CM)	534.2	505.6	3	<0.001
(C, AM)	939.6	824.2	3	<0.001
(M, AC)	843.8	704.9	3	<0.001
(AC, AM)	497.4	443.8	2	<0.001
(AC, CM)	92.0	80.8	2	<0.001
(AM, CM)	187.8	177.6	2	<0.001
(AC, AM, CM)	0.4	0.4	1	0.54
(ACM)	0.0	0.0	0	—

* P-value for G^2 statistic.

Planteamos primero el modelo (A,C,M)

```
* A = Alcohol
* C = Cigarrillo
* M = Marihuana
  man 3
  dim 2 2 2
  lab A C M
  mod {A,C,M}
  dat [ 911 538
          44 456
           3 43
           2 279]
```

```
Number of iterations = 2
Converge criterion   = 0.0000000000
X-squared            = 1411.3860 (0.0000)
L-squared            = 1286.0200 (0.0000)
Cressie-Read         = 1316.2009 (0.0000)
Dissimilarity index  = 0.2875
Degrees of freedom   = 4
Log-likelihood       = -3950.08844
Number of parameters = 3 (+1)
Sample size          = 2276.0
BIC(L-squared)       = 1255.0993
AIC(L-squared)       = 1278.0200
BIC(log-likelihood)  = 7923.3674
AIC(log-likelihood)  = 7906.1769
Eigenvalues information matrix
  2220.4205  2051.8191  1119.8361
```

*** FREQUENCIES ***

A	C	M	observed	estimated	std. res.
1	1	1	911.000	539.983	15.966
1	1	2	538.000	740.226	-7.433
1	2	1	44.000	282.091	-14.176
1	2	2	456.000	386.700	3.524
2	1	1	3.000	90.597	-9.203
2	1	2	43.000	124.194	-7.286
2	2	1	2.000	47.329	-6.589
2	2	2	279.000	64.880	26.583

```
*** LOG-LINEAR PARAMETERS ***
* TABLE ACM [or P(ACM)] *
  effect           beta  std err  z-value   exp(beta)     Wald  df  prob
  main           5.2320                     187.1738
  A
   1             0.8926   0.0299   29.868      2.4414
   2            -0.8926                       0.4096   892.12   1 0.000
  C
   1             0.3247   0.0221   14.706      1.3836
   2            -0.3247                       0.7228   216.26   1 0.000
  M
   1            -0.1577   0.0212   -7.431      0.8541
   2             0.1577                       1.1708    55.23   1 0.000
```

```
*** (CONDITIONAL) PROBABILITIES ***
* P(ACM) *
  1 1 1          0.2373  (0.0071)
  1 1 2          0.3252  (0.0081)
  1 2 1          0.1239  (0.0048)
  1 2 2          0.1699  (0.0060)
  2 1 1          0.0398  (0.0023)
  2 1 2          0.0546  (0.0031)
  2 2 1          0.0208  (0.0013)
  2 2 2          0.0285  (0.0018)
```

Probamos con (AC,AM,CM)

```
man 3
dim 2 2 2
lab A C M
mod {AC,AM,CM}
dat [ 911 538
       44 456
        3 43
        2 279]
```

```
Number of iterations = 6
  Converge criterion   = 0.0000000246
  X-squared            = 0.4011 (0.5265)
  L-squared            = 0.3740 (0.5408)
  Cressie-Read         = 0.3913 (0.5316)
  Dissimilarity index  = 0.0011
  Degrees of freedom   = 1
  Log-likelihood       = -3307.26545
  Number of parameters = 6 (+1)
  Sample size          = 2276.0
  BIC(L-squared)       = -7.3562
  AIC(L-squared)       = -1.6260
  BIC(log-likelihood)  = 6660.9120
  AIC(log-likelihood)  = 6626.5309
  Eigenvalues information matrix
5310.35  3346.091  1750.952   500.779   176.449
26.072
```

```
*** FREQUENCIES ***
  A C M     observed  estimated  std. res.
  1 1 1     911.000    910.383      0.020
  1 1 2     538.000    538.617     -0.027
  1 2 1      44.000     44.617     -0.092
  1 2 2     456.000    455.383      0.029
  2 1 1       3.000      3.617     -0.324
  2 1 2      43.000     42.383      0.095
  2 2 1       2.000      1.383      0.524
  2 2 2     279.000    279.617     -0.037
```

```
*** LOG-LINEAR PARAMETERS ***
* TABLE ACM [or P(ACM)] *
  effect           beta  std err  z-value   exp(beta)     Wald  df  prob
  main           4.2515                     70.2130
  A
   1             1.5040   0.1138   13.215      4.4996
   2            -1.5040                      0.2222   174.64   1 0.000
  C
   1             0.2823   0.0549    5.140      1.3261
   2            -0.2823                      0.7541    26.42   1 0.000
  M
   1            -1.1960   0.1185  -10.093      0.3024
   2             1.1960                      3.3070   101.87   1 0.000
  AC
   1 1           0.5136   0.0435   11.802      1.6714
   1 2          -0.5136                      0.5983
   2 1          -0.5136                      0.5983
   2 2           0.5136                      1.6714   139.30   1 0.000
  AM
   1 1           0.7465   0.1162    6.426      2.1096
   1 2          -0.7465                      0.4740
   2 1          -0.7465                      0.4740
   2 2           0.7465                      2.1096    41.29   1 0.000
  CM
   1 1           0.7120   0.0410   17.381      2.0380
   1 2          -0.7120                      0.4907
   2 1          -0.7120                      0.4907
   2 2           0.7120                      2.0380   302.10   1 0.000
```

```
*** (CONDITIONAL) PROBABILITIES ***
* P(ACM) *
  1 1 1          0.4000  (0.0103)
  1 1 2          0.2367  (0.0089)
  1 2 1          0.0196  (0.0029)
  1 2 2          0.2001  (0.0084)
  2 1 1          0.0016  (0.0007)
  2 1 2          0.0186  (0.0028)
  2 2 1          0.0006  (0.0003)
  2 2 2          0.1229  (0.0069)
```

Probamos de quitar AM

* A = Alcohol
* C = Cigarrillo
* M = Marihuana

```
man 3
dim 2 2 2
lab A C M
mod {AC,CM}
dat [ 911 538
       44 456
        3 43
        2 279]
```

```
 Number of iterations = 2
 Converge criterion   = 0.0000000000
 X-squared            = 80.8148 (0.0000)
 L-squared            = 92.0184 (0.0000)
 Cressie-Read         = 82.0368 (0.0000)
 Dissimilarity index  = 0.0349
 Degrees of freedom   = 2
 Log-likelihood       = -3353.08764
 Number of parameters = 5 (+1)
 Sample size          = 2276.0
 BIC(L-squared)       = 76.5580
 AIC(L-squared)       = 88.0184
 BIC(log-likelihood)  = 6744.8262
 AIC(log-likelihood)  = 6716.1753
 Eigenvalues information matrix
4190.794  2743.776  1399.651   351.217   183.293
```

Ahora quitamos AC

* A = Alcohol
* C = Cigarrillo
* M = Marihuana

```
man 3
dim 2 2 2
lab A C M
mod {AM,CM}
dat [ 911 538
       44 456
        3 43
        2 279]
```

```
Number of iterations = 2
  Converge criterion   = 0.0000000000
  X-squared            = 177.6149 (0.0000)
  L-squared            = 187.7543 (0.0000)
  Cressie-Read         = 178.1090 (0.0000)
  Dissimilarity index  = 0.0887
  Degrees of freedom   = 2
  Log-likelihood       = -3400.95561
  Number of parameters = 5 (+1)
  Sample size          = 2276.0
  BIC(L-squared)       = 172.2940
  AIC(L-squared)       = 183.7543
  BIC(log-likelihood)  = 6840.5621
  AIC(log-likelihood)  = 6811.9112
  Eigenvalues information matrix
    4814.082  2582.42  1755.695   272.398    26.286
```

Ahora quitamos CM

* A = Alcohol
* C = Cigarrillo
* M = Marihuana

```
man 3
dim 2 2 2
lab A C M
mod {AC,AM}
dat [ 911 538
       44 456
        3 43
        2 279]
```

```
Number of iterations = 2
  Converge criterion   = 0.0000000000
  X-squared            = 443.7611 (0.0000)
  L-squared            = 497.3693 (0.0000)
  Cressie-Read         = 453.7193 (0.0000)
  Dissimilarity index  = 0.1786
  Degrees of freedom   = 2
  Log-likelihood       = -3555.76310
  Number of parameters = 5 (+1)
  Sample size          = 2276.0
  BIC(L-squared)       = 481.9089
  AIC(L-squared)       = 493.3693
  BIC(log-likelihood)  = 7150.1771
  AIC(log-likelihood)  = 7121.5262
  Eigenvalues information matrix
 3897.094  3302.761  1803.663   263.962   26.369
```

Ninguno de los 3 modelos (AC,AM) (AC,CM) (AM,CM) ajusta bien
Nos quedamos entonces con (AC,AM,CM)