Modelos Log-lineares usando SAS

EXEMPLO 1 - Bicicleta

Deseja-se estudar se existe associação entre o tipo de bicicleta ("moutain " ou outra) e o uso de capacete de segurança. Para tal foi selecionada uma amostra de 100 pessoas que andam de bicicleta e as classificou segundo tipo de capacete e se usa capacete ou não

Programa:

```
/* *********   Exemplo - Bicicleta   ****************** */
/*  ************ Entrada de dados  ****************  */

data bicicleta;
   input tipo $ capacete $ freq;
   cards;
   moutain sim 34
   moutain nao 32
   outra   sim 10
   outra   nao 24
;
run;
/* ***** Modelo Saturado ***** */

proc catmod data=bicicleta;
   title ' Modelo Saturado ' ;
   weight freq;
   model tipo*capacete=_response_/noresponse noiter noparm;
   loglin tipo| capacete;
run;

/* ***** Modelo de independencia ***** */

proc catmod data=bicicleta;
      Title 'Modelo de independencia ' ;
      weight freq;
    model tipo*capacete=_response_/noresponse noiter noparm;
    loglin tipo capacete;
run;
```

Modelo Saturado

The CATMOD Procedure

Data Summary			
Response	tipo*capacete	**Response Levels**	4
Weight Variable	freq	**Populations**	1
Data Set	BICICLETA	**Total Frequency**	100
Frequency Missing	0	**Observations**	4

Population Profiles	
Sample	**Sample Size**
1	100

Response Profiles		
Response	**tipo**	**capacete**
1	moutain	nao
2	moutain	sim
3	outra	nao
4	outra	sim

Maximum Likelihood Analysis
Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
tipo	1	11.29	0.0008
capacete	1	3.28	0.0700
tipo*capacete	1	4.33	0.0374
Likelihood Ratio	0	.	.

Modelo de independencia

The CATMOD Procedure

Data Summary			
Response	tipo*capacete	**Response Levels**	4
Weight Variable	freq	**Populations**	1
Data Set	BICICLETA	**Total Frequency**	100
Frequency Missing	0	**Observations**	4

Population Profiles	
Sample	**Sample Size**
1	100

Response Profiles		
Response	**tipo**	**capacete**
1	moutain	nao
2	moutain	sim
3	outra	nao
4	outra	sim

Maximum Likelihood Analysis
Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
tipo	1	9.87	0.0017
capacete	1	1.43	0.2313
Likelihood Ratio	1	4.56	0.0328

Analysis of Maximum Likelihood Estimates					
Parameter		**Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**
tipo	**moutain**	0.3779	0.1124	11.29	0.0008
capacete	**nao**	0.2037	0.1124	3.28	0.0700
tipo*capacete	**moutain nao**	-0.2340	0.1124	4.33	0.0374

##

EXEMPLO 2 - Uso de Álcool, Cigarro e Maconha
(*Agresti – cap.7*)

Uma pesquisa que perguntou a alunos do último ano de uma escola secundária perto de Dayton, Ohio, se eles já haviam usado álcool, cigarro ou maconha. Designando as variáveis nesta tabela de contingência 2 × 2 × 2 por X para uso de álcool, Y para uso de cigarro e Z para uso de maconha.

Tabela 7.1 - Consumo de álcool (X), cigarro (Y) e maconha (Z) para alunos do último ano do ensino médio.

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

Source: Thanks to Prof. Harry Khamis, Wright State University and United Health Services in Dayton for these data (the `Substance` data file at the text website).

Programa:

```
/* *********   Exemplo - Uso de Drogas   ****************** */
/*  ************ Entrada de dados      ******************  */

/* X (A) - uso de álcool   */
/* Y (C) - uso de cigarros */
/* Z (M) - uso de maconha  */

data dados;
  input x $ y $ z $ freq;
  datalines;
s s s 911
s s n 538
s n s  44
s n n 456
n s s   3
n s n  43
n n s   2
n n n 279
;
run;

proc freq data=dados order=data;
  weight freq;
  table z*x*y;
run;
```

```
/* ***** Modelo Saturado ***** */

proc catmod data=dados order=data;
   title ' Modelo Saturado ' ;
   weight freq;
   m0: model x*y*z=_response_/ noresponse noiter noparm ;
   loglin x|y|z;
run;

/* ***** Modelo Associação Homogênea   ***** */
proc catmod order=data;
   title ' Modelo Associacao Homogenea ' ;
   weight freq;
   model x*y*z=_response_/ noprofile noresponse noiter noparm pred=freq;
   loglin x|y x|z y|z;
run;

/* ***** Modelos Independência Condicional   ***** */

/* de Y e Z dado X  */
proc catmod data=dados order=data;
   weight freq;
   model x*y*z=_response_/noresponse noiter noparm pred=freq;
   loglin x|y x|z ;
run;


/* de X e Z dado Y  */
proc catmod order=data;
   weight freq;
   model x*y*z=_response_/noresponse noiter noparm;
   loglin x|y y|z ;
run;

/* de X e Y dado Z  */
proc catmod order=data;
   weight freq;
   model x*y*z=_response_/noresponse noiter noparm;
   loglin x|z y|z ;
run;


/* ***** Modelos Independência Conjunta   ***** */

/* X conjuntamente indep. de Y e Z */
proc catmod order=data;
   weight freq;
   model x*y*z=_response_/noresponse noiter noparm;
   loglin x y|z ;
run;

/* Y conjuntamente indep. de X e Z */
proc catmod order=data;
   weight freq;
   model x*y*z=_response_/noresponse noiter noparm;
   loglin y x|z ;
run;
```

```
/* Z conjuntamente indep. de X e Y */
proc catmod order=data;
   weight freq;
   model x*y*z=_response_/noresponse noiter noparm;
   loglin z x|y ;
run;


/* ***** Modelo Independência mútua   ***** */

proc catmod order=data;
 weight freq;
   model x*y*z=_response_/noresponse noiter noparm;
   loglin x y z  ;
run;



/* ***** Determinando as Estimativas dos Parametros   ***** */
/* *****        Modelo Associação Homogênea          * **** */

proc catmod order=data;
   title ' Modelo Associacao Homogenea ' ;
   weight freq;
   model x*y*z=_response_/ noprofile noresponse noiter ;
   loglin x|y x|z y|z;
run;

/* ***** Desterminando Frequências Esperadas   ***** */
proc catmod order=data;
   title ' Modelo Associacao Homogenea ' ;
   weight freq;
   model x*y*z=_response_/ noprofile noresponse noiter noparm pred=freq;
   loglin x|y x|z y|z;
run;

/* *****       Imprimindo os valores Preditos   ***** */
/* *****        Modelo Associação Homogênea          * **** */

proc catmod order=data;
   title ' Modelo Associacao Homogenea ' ;
   weight freq;
   model x*y*z=_response_/ noprofile noresponse noiter p;
   loglin x|y x|z y|z;
run;
```

Modelo de Interação Tripla - Saturado

Maximum Likelihood Analysis
Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
x	1	157.46	<.0001
y	1	3.37	0.0662
x*y	1	24.46	<.0001
z	1	99.76	<.0001
x*z	1	37.74	<.0001
y*z	1	29.76	<.0001
x*y*z	1	0.39	0.5316
Likelihood Ratio	0	.	.

Modelo de Associação Homogênea

Maximum Likelihood Analysis
Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
x	1	174.68	<.0001
y	1	26.42	<.0001
x*y	1	139.32	<.0001
z	1	101.87	<.0001
x*z	1	41.29	<.0001
y*z	1	302.14	<.0001
Likelihood Ratio	1	0.37	0.5408

Modelos de Independência Condicional

Modelo de Independência Condicional de X e Y dado Z

Maximum Likelihood Analysis
Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
x	1	198.52	<.0001
z	1	155.20	<.0001
x*z	1	83.00	<.0001
y	1	292.68	<.0001
y*z	1	401.17	<.0001
Likelihood Ratio	2	187.75	<.0001

Modelo de Independência Condicional de X e Z dado Y

Maximum Likelihood Analysis
Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
x	1	579.20	<.0001
y	1	10.93	0.0009
x*y	1	295.07	<.0001
z	1	207.36	<.0001
y*z	1	401.17	<.0001
Likelihood Ratio	2	92.02	<.0001

Modelo de Independência Condicional de C e M dado A

Maximum Likelihood Analysis
Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
x	1	213.50	<.0001
y	1	19.87	<.0001
x*y	1	295.07	<.0001
z	1	86.20	<.0001
x*z	1	82.95	<.0001
Likelihood Ratio	2	497.37	<.0001

Modelos de Independência Conjunta

Modelo de Independência Conjunta X de Y e Z

Maximum Likelihood Analysis
Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
x	1	892.31	<.0001
y	1	292.71	<.0001
z	1	207.38	<.0001
y*z	1	401.20	<.0001
Likelihood Ratio	3	534.21	<.0001

Modelo de Independência Conjunta Y de X e Z

Maximum Likelihood Analysis
Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
y	1	216.28	<.0001
x	1	198.44	<.0001
z	1	86.22	<.0001
x*z	1	82.97	<.0001
Likelihood Ratio	3	939.56	<.0001

Modelo de Independência Conjunta Z de X e Y

Maximum Likelihood Analysis
Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
z	1	55.22	<.0001
x	1	579.20	<.0001
y	1	19.87	<.0001
x*y	1	295.07	<.0001
Likelihood Ratio	3	843.83	<.0001

Modelo de Independência Mútua

Maximum Likelihood Analysis
Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
x	1	892.65	<.0001
y	1	216.28	<.0001
z	1	55.22	<.0001
Likelihood Ratio	4	1286.02	<.0001

Resumindo e Analisando os Resultados:

Complete a tabela a seguir consultando as saídas apresentadas anteriormente.

	Modelo	Equação do Modelo	G^2	Gl	p-value
1	[XYZ]		-	0	-
2	[XY] [XZ] [YZ]				
3	[XZ] [YZ]				
4	[XY] [YZ]				
5	[YX] [XZ]				
6	[X] [YZ]				
7	[Y] [XZ]				
8	[Z] [XY]				
9	[X] [Y] [Z]				

Modelo Associação Homogênea

- **Estimativas dos Parâmetros**

The CATMOD Procedure

Data Summary			
Response	x*y*z	**Response Levels**	8
Weight Variable	freq	**Populations**	1
Data Set	PRED	**Total Frequency**	2276
Frequency Missing	0	**Observations**	8

Maximum Likelihood Analysis
Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
x	1	174.68	<.0001
y	1	26.42	<.0001
x*y	1	139.32	<.0001
z	1	101.87	<.0001
x*z	1	41.29	<.0001
y*z	1	302.14	<.0001
Likelihood Ratio	1	0.37	0.5408

Analysis of Maximum Likelihood Estimates					
Parameter		**Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**
x	**s**	1.5040	0.1138	174.68	<.0001
y	**s**	0.2823	0.0549	26.42	<.0001
x*y	**s s**	0.5136	0.0435	139.32	<.0001
z	**s**	-1.1960	0.1185	101.87	<.0001
x*z	**s s**	0.7465	0.1162	41.29	<.0001
y*z	**s s**	0.7120	0.0410	302.14	<.0001

Note que:

$$\hat{\theta}_{AC} = \exp(4x0{,}5136) = \exp(2{,}0544) = 7{,}8021$$

$$\hat{\theta}_{AM} = \exp(4x0{,}7465) = \exp(2{,}9686) = 19{,}8062$$

$$\hat{\theta}_{CM} = \exp(4x0{,}7120) = \exp(2{,}848) = 17{,}2532$$

- **Frequências Esperadas**

Maximum Likelihood Predicted Values for Frequencies							
			Observed		Predicted		
x	y	z	Frequency	Standard Error	Frequency	Standard Error	Residual
s	s	s	911	23.37434	910.3832	23.3514	0.61683
s	s	n	538	20.26889	538.6168	20.2535	-0.61683
s	n	s	44	6.568819	44.61683	6.541681	-0.61683
s	n	n	456	19.09554	455.3832	19.06099	0.61683
n	s	s	3	1.730909	3.61683	1.631699	-0.61683

Podemos montar a seguinte tabela com as frequências esperadas segundo o modelo obtidas com o SAS:

		maconha			
	Cigarro	sim		não	
Alcool		sim	não	sim	não
sim					
não					

Observe que:

$$\hat{\theta}_{AC} = \exp(4x0{,}5136) = \exp(2{,}0544) = 7{,}8021$$

ou usando as frequências esperadas temos:

ou