



Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA
 Centro Tecnológico para Informática Agropecuária - NTIA
 Rodovia SP 340, Km 105,4 - Caixa Postal 5010
 13031 Campinas SP - Fone (0192) 42-9753

COMUNICADO TÉCNICO

CT-NTIA-8904

Agosto 1989

Uso do SOC para ajuste de modelos log-lineares a dados categorizados

Isildinha Marques dos Reis, FSP/USP
 Julio da Motta Singer, IME/USP

LOGLIN.AJU e LOGLIN.HIP são subrotinas criadas a partir do módulo CM do pacote científico SOC, para a análise de dados categorizados através de Modelos Log-Lineares.

Os dados devem ser apresentados na forma de uma tabela de contingência bidimensional (sxr), onde as s linhas referem-se às subpopulações definidas pelas combinações dos níveis das variáveis explicativas e as r colunas correspondem às categorias de resposta obtidas pelas combinações dos níveis das variáveis respostas. O modelo probabilístico de geração dos dados corresponde a um produto de s distribuições Multinomiais independentes.

A subrotina LOGLIN.AJU possibilita a obtenção de estimativas através de Máxima Verossimilhança (MV) e de Mínimos Quadrados Generalizados (MQG) dos parâmetros de modelos log-lineares (hierárquicos ou não), expressos na formulação matricial de logitos múltiplos

$$A * \log PI = XL * BETA$$

onde $PI = (PI_1' \dots PI_s')$: (srxl) é o vetor de probabilidades associadas às sr caselas; $XL [s(r-1) \times t]$ é a matriz de especificação do modelo, $BETA (t \times 1)$, t menor ou igual a s(r-1), é o vetor de parâmetros a ser estimado e A é a matriz que define os s(r-1) logitos múltiplos; o logito para a categoria de resposta j na subpopulação i é

$$\log (PI_{ij} / PI_{ir}) \text{ para } i=1, \dots, s \text{ e } j=1, \dots, (r-1).$$

Essa formulação para o modelo log-linear é equivalente a formulação usual encontrada na maioria dos livros texto como Bishop et al. (1975) por exemplo. Para maiores detalhes sobre essa equivalência (incluindo exemplos) ver Reis (1989).

Inicialmente são obtidas as estimativas dos parâmetros pelo método de MQG, empregando-se a solução proposta por Grizzle, Starmer & Koch (1969). Em seguida as estimativas de MV são calculadas através do processo iterativo de Newton-Raphson. As estimativas de MQG podem ser utilizadas como estimativas iniciais no processo iterativo. Detalhes sobre esses métodos de estimação podem ser encontrados em Imrey et al. (1981, 1982), Koch et al. (1985) ou Reis (1989), por exemplo.

O programa fornece as estimativas de MQG e MV dos parâmetros, respectivos desvios padrões, valores dessas estimativas em unidades dos desvios padrões correspondentes, estatísticas Qui-quadrado para testar o ajuste do modelo (Wald, Pearson e Razão de Verossimilhança), logitos múltiplos observados e ajustados por MQG e MV e respectivos desvios padrões e proporções observadas e ajustadas por MQG e MV e respectivos desvios padrões. Opcionalmente são fornecidas as estimativas das matrizes de covariância para proporções observadas, ajustadas e para as estimativas dos parâmetros.

A subrotina LOGLIN.HIP possibilita testar hipóteses do tipo $H_0: C \cdot BETA = 0$, onde C ($c \times t$) é uma matriz de constantes conhecidas de posto c menor ou igual a t , através da estatística de Wald.

Pessoas interessadas em cópias do programa e da documentação devem enviar um disquete 5 1/4" D/D para:

Setor de Difusão de Tecnologia - NTIA / EMBRAPA
Caixa Postal 5010
13031 - Campinas - SP

Referências Bibliográficas:

- BARTLETT, M. S. (1935) Contingency table interactions. J.R. Statist. Soc., Suppl., 2: 248-252.
- BISHOP, Y. M. M., FIENBERG, S. E.; HOLLAND, P. W. (1975) Discrete Multivariate Analysis: Theory and Practice. Cambridge, MIT Press 557 p.
- GRIZZLE, J. E., STARMER, C. F. & KOCH, G. G. (1969) The analysis of categorical data by linear models. Biometrics, 25: 489-504.

IMREY et al. (1981) Categorical data analysis. some reflections on the log linear model and logistic regression. Part I: Historical and methodological overview. Int. Statist. Rev., 49: 265-283.

IMREY et al. (1982) Categorical data analysis: some reflections on the log linear model and logistic regression. Part II: Data analysis. Int. Statist. Rev., 50: 35-63.

KOCH, G. G. et al. (1981) Analyse of Categorical Data. Montréal. Les Presses de l'Université de Montréal. 290 p.

REIS, I. M. (1989) Modelos log-lineares para análise de dados categorizados. São Paulo. 136 p. Dissertação de mestrado - Instituto de Matemática e Estatística da USP.

