SOME REMARKS ABOUT A SIMULATION STUDY FOR A LATENT CLASS MODEL UNDER SPARSENESS

Adalberto González Debén*, Hannelore Liero**, Jesús E. Sánchez García*, Henning Läuter**
*ICIMAF, Havana, Cuba
**University of Potsdam, Germany

RESUMEN
En el presente trabajo se hace un estudio teórico con vistas a explicar algunos resultados de la simulación que se realizó en el artículo "Sobre la estimación de los parámetros del modelo de clases latentes en condiciones de raleza" (González, Sánchez & Hernández, 2002). La herramienta fundamental es la aplicación de resultados acerca de la normalidad asintótica de los estimadores máximo verosímiles.

ABSTRACT
A theoretical study is accomplished in this paper in order to explain some simulation results exposed in the paper "Sobre la estimación de los parámetros del modelo de clases latentes en condiciones de raleza" (About the parameter estimation of the latent class model under sparseness) (González, Sánchez & Hernández, 2002). The main tool is the application of results concerning asymptotic normality of the maximum likelihood estimators

KEY WORDS: Maximum likelihood estimation, Fisher information matrix, Simulation Study

MSC 62H17

1 INTRODUCTION

In the paper "Sobre la estimación de los parámetros del modelo de clases latentes en condiciones de raleza" (About the parameter estimation of the latent class model under sparseness) by González, Sánchez & Hernández (2002) the following latent class model with four (manifest) variables A, B, C, and D and two latent classes is considered. Let X be the (unobservable) random variable characterizing the latent class and define

$$\pi = P(X=1) = 1 - P(X=2).$$

Suppose that the manifest variables are conditionally independent, then we have for $j,k,l,m \in \{0,1\}$

$$\begin{aligned} &P(A=j, B=k, C=l, D=m) = p(j,k,l,m) \\ &= \pi p_{A/1}^{j}(1-p_{A/1})^{1-j}\, p_{B/1}^{k}(1-p_{B/1})^{1-k}\, p_{C/1}^{l}(1-p_{C/1})^{1-l}\, p_{D/1}^{m}(1-p_{D/1})^{1-m} \\ &+ (1-\pi)\, p_{A/2}^{j}(1-p_{A/2})^{1-j}\, p_{B/2}^{k}(1-p_{B/2})^{1-k}\, p_{C/2}^{l}(1-p_{C/2})^{1-l}\, p_{D/2}^{m}(1-p_{D/2})^{1-m} \end{aligned} \qquad (1)$$

where

$$\begin{aligned} p_{A/1} &= P(A=1|X=1), \quad p_{A/2} = P(A=1|X=2), \\ p_{B/1} &= P(B=1|X=1), \quad p_{B/2} = P(B=1|X=2), \\ p_{C/1} &= P(C=1|X=1), \quad p_{C/2} = P(C=1|X=2), \\ p_{D/1} &= P(D=1|X=1), \quad p_{D/2} = P(D=1|X=2). \end{aligned}$$

are the conditional probabilities.

The estimation of the probability π and of the conditional probabilities by the maximum likelihood method leads to highly nonlinear system of equations which can be solved only by iterative procedures.

A question arising in this context is what happens when there is some sparseness in the cells of the resulting table. In the above mentioned paper the authors investigated the effect of sparseness on the estimation of π and of $p_{A/1}$ by a simulation study. For that purpose different sparseness constellations were generated by varying the following four factors:

- Sample size
- Strength of relation between manifest and latent variables
- Size of the latent class
- Noise

On the basis of the simulations empirical versions of the bias and the mean squared error were computed. It was observed in the results that for some constellations there was a large number of the so-called indeterminate solutions of the estimation procedure.

A program for MatLab 5.3 was written for the generation of data and LEM for MSDOS (Vermunt, 1997) was used for the fitting of the latent class model. Por each treatment all th necessary replications were generated up to 10 000 fitte models with principal latent class.

The aim of the present paper is to get by some theoretical investigations an explanation for some of the simulations results. The main tool to do this is the application of results about the asymptotic normality of maximum likelihood estimators.

2. MODELS AND MAXIMUM LIKELIHOOD ESTIMATOR

We start our considerations with the correct model, that is: model (1). The unknown (nine dimensional) parameter is $\vartheta = (p_{A/1}, p_{B/1}, \ldots, p_{A/2}, \ldots, p_{D/2}, \pi)^t$.
Given the observations $y_i = (y_{i1}, y_{i2}, y_{i3}, y_{i4})$, $y_{iq} \in \{0,1\}$, $q = 1, \ldots, 4, i = 1, \ldots, n$, the likelihood function has the form:

$$\prod_{i=1}^{n} (\pi p_{A/1}^{y_{i1}} (1-p_{A/1})^{1-y_{i1}} p_{B/1}^{y_{i2}} (1-p_{B/1})^{1-y_{i2}} p_{C/1}^{y_{i3}} (1-p_{C/1})^{1-y_{i3}} p_{D/1}^{y_{i4}} (1-p_{D/1})^{1-y4}$$
$$+ (1-\pi) p_{A/2}^{y_{i1}} (1-p_{A/2})^{1-y_{i1}} p_{B/2}^{y_{i2}} (1-p_{B/2})^{1-y_{i2}} p_{C/2}^{y_{i3}} (1-p_{C/2})^{1-y_{i3}} p_{D/2}^{y_{i4}} (1-p_{D/2})^{1-y_{i4}},$$

and for the log likelihood function we obtain:

$$L(\vartheta, y) = \sum_{j=0}^{1} \sum_{k=0}^{1} \sum_{l=0}^{1} \sum_{m=0}^{1} H_{jklm}(y) \log p(j,k,l,m),$$

where $H_{jklm}(y)$, $y = ((y_{11}, \ldots, y_{14}), \ldots, (y_{n1}, \ldots, y_{n4}))$, are the cell frequencies, that is:

$$H_{jklm}(y) = \sum_{i=1}^{n} 1(y_{i1} = j, y_{i2} = k, y_{i3} = l, y_{i4} = m).$$

the maximum likelihood estimator (m.l.e.) $\hat{\vartheta}_n$ is defined by

$$L(\hat{\vartheta}_n, y) \geq L(\vartheta, y) \text{ for all } \vartheta.$$

Sometimes, this models is not adequate in practice. For example, if answers to questions are given randomly or in cases where the same value is added to each cell in order to avoid sparseness. We describe such situations by a **model with noise**; that is: our observations follow the model:

$$P(A=j, B=k, C=l, D=m) = p^*(j,k,l,m) = \tau p(j,k,l,m) + (1-\tau)\frac{1}{16}. \tag{2}$$

Here τ is the parameter of noise, for $\tau = 1$ we get the correct model. In the simulation study $\tau = 7/8$ and $\tau = 3/4$ were chosen.

3. BEHAVIOR OF THE ESTIMATORS IN THE CORRECT MODEL

A consequence of the high non-linearity of the likelihood equations is that one cannot compute the bias and the variance of the resulting estimators. One possibility to get a deeper insight of the properties of the estimator is to investigate its Fisher information matrix (Cox & Hinkley, 1982).

It is known that under weak assumptions on the underlying model for the maximum likelihood the following asymptotic result holds (Kendall & Stuart, 1973):

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{d} N(0, I^{-1}(\vartheta)),$$

where $I^{-1}(\vartheta)$ is the inverse of the Fisher information matrix in model (1). The elements of the Fisher information matrix are defined by:

$$I_{rs}(\vartheta) = \sum_{j=0}^{1}\sum_{k=0}^{1}\sum_{l=0}^{1}\sum_{m=0}^{1} \frac{\partial p(j,k,l,m)}{\partial \vartheta_s} \frac{\partial p(j,k,l,m)}{\partial \vartheta_r} \frac{1}{p(j,k,l,m)}.$$

Let $I_{rs}^{-1}(\vartheta)$ be the elements of the inverse of the Fisher information matrix. Then $n^{-1}I_{11}^{-1}(\vartheta)$ and $n^{-1}I_{99}^{-1}(\vartheta)$ can be considered as asymptotic expressions for the variance of the estimator $\hat{\vartheta}_{n1}$ for $p_{A/1}$ and $\hat{\vartheta}_{n9}$ for π, respectively.

The computation of the Fisher information matrices for the parameters used for the simulations yields the following results:

3.1 Strong cases

We set

$$p_{A/1} = 0.9, p_{A/2} = 0.1, p_{B/1} = 0.9, p_{B/2} = 0.1, p_{C/1} = 0.9, p_{C/2} = 0.1, p_{D/1} = 0.9, p_{D/2} = 0.1$$

and $\pi = 0.5$ for the symmetric case and $\pi = 0.9$ for the asymmetric case.

The eigenvalues of $I(strong, symm)$ are

2.301, 5.776, 5.556, 5.057, 4.336, 4.626, 4.428, 4.336, 5.556

and those of $I(strong, asymm)$ are

8.409, 9.728, 9.929, 10.626, 0.578, 0.821, 9.728, 0.821

From here we conclude that these matrices are non-singular and we get for the strong symmetric case:

$$I_{11}^{-1}(strong, symm) = -.2215 \text{ and } I_{99}^{-1}(strong, symm) = 0.2846$$

For the strong asymmetric case we have

$$I_{11}^{-1}(strong, asymm) = 0.1076 \text{ and } I_{99}^{-1}(strong, asymm)\ 0.1066.$$

Thus, we give in the following tables the asymptotic approximations for the variance of the estimators $\hat{\vartheta}_{n1} = \hat{p}_{A/1}$ and $\hat{\vartheta}_{n9} = \hat{\pi}$ in our simulated cases

We see:

1. Since the maximum likelihood estimators are asymptotically unbiased, we can interpret the asymptotic expressions for the variance resulting form the approach used as (theoretic) approximations of

the MSE. Our results show that the difference between this "theoretic" MSE an the one from the simulation is almost zero.
The variance in the asymmetric case is always a little bit smaller than in the symmetric one. For estimating π this corresponds to the well-known fact that the variance of the estimator for the success probability in a binomial distribution is maximal if this parameter is 0.5. For estimating $p_{A/1}$ the explanation is that in this case we have more information about class 1 than in the case where both classes have the same probability.
2. We see from the inverses of Fisher matrices that in the symmetric case the variances of the estimators for the other components of the parameter vector ϑ do not differ very much (about $n^{-1}0.22$). It seems to be clear from a heuristical point of view because we have the same information from each class and estimating 0.9 or 0.1 leads to the same variance in the binomial distribution. In the asymmetric case we see that the parameters ϑ_1, ϑ_2, ϑ_3 and ϑ_4 are estimated with a variance of about $n^{-1}0.108$, but the parameters ϑ_5, ϑ_6, ϑ_7 and ϑ_8 have variances of about $n^{-1}1.34$. This is due to the fact that π = 0.9 gives only little information about the second class.
3. Considering variances (resp. MSE's) it seems that the asymmetric is better, the fact of having indeterminate results being a contradiction to this statement. A possible explanation for this phenomenon is that we have more sparse cells in this case. This is reflected by the fact that some of the eigenvalues of the Fisher matrix are very small (in comparison to the eigenvalues in the symmetric case.)

$\hat{p}_{A/1}$				
	Asymptotic expression		Simulation MSE	
N	Symmetric	Asymmetric	Symmetric	Asymmetric
32	0.0069	0.0034	0.007	0.004
64	0.0035	0.0017	0.004	0.002
128	0.0017	0.0008	0.002	0.001
256	0.0009	0.0004	0.001	0.000

Table 1: Asymptotic approximations for the variance of the estimator $\hat{p}_{A/1}$

$\hat{\pi}$				
	Asymptotic expression		Simulation MSE	
N	Symmetric	Asymmetric	Symmetric	Asymmetric
32	0.0089	0.0033	0.008	0.005
64	0.0044	0.0017	0.004	0.002
128	0.0022	0.0008	0.002	0.001
256	0.0011	0.0004	0.001	0.000

Table 2: Asymptotic approximations for the variance of the estimator $\hat{\pi}$

3.2 Weak cases

We set $p_{A/1}$ = 0.9, $p_{A/2}$ = 0.1, $p_{B/1}$ = 0.9, $p_{B/2}$ = 0.1, $p_{C/1}$ = 0.5, $p_{C/2}$ = 0.5, $p_{D/1}$ = 0.5, $p_{D/2}$ = 0.5 and π = 0.5 for the symmetric case, and π = 0.9 for the asymmetric case.
It turns out that the Fisher information matrices are singular in the weak cases. The computed eigenvalues of $I(weak, symm)$ are

$$-0.4 \times 10^{-9},\ 0.7 \times 10^{-11},\ 1.30,\ 1.60,\ 1.88,\ 2.05,\ 4.34,\ 4.36,\ 5.56$$

that is, without computational errors

$$0,\ 0,\ 1.30,\ 1.60,\ 1.88,\ 2.05,\ 4.34,\ 4.36,\ 5.56$$

and those of *I(weak, asymm)*

$$-0.8 \times 10^{-9},\ 0.16 \times 10^{-9},\ 10.29,\ 9.11,\ 5.70,\ 2.83,\ 3.66,\ 0.29,\ 0.30$$

and

$$0,\ 0,\ 10.29,\ 9.11,\ 5.70,\ 2.83,\ 3.66,\ 0.29,\ 0.30,$$

respectively.

In the weak cases the parameters ϑ_s are not identifiable. For example, vector ϑ = (5/42, 2/25, 1/2, 1/2, 37/42, 23/25, 1/2, 1/2, 1/2) leads to the same cell probabilities as in the weak symmetric case. This explains why we have so many indeterminate simulation results in the weak cases.
What does "weak" mean? In the weak cases we have

$$P(C=l) = P(C=l / X=1)\pi + P(C=l / X=2)(1-\pi) = 0.5$$

and the same occurs with $P(D=m) = 0.5$.

So, the random variables C and D do not depend on X. The model is overparametrized, and the conclusion for the practical application is: if one is not convinced that the variables of interest depend on a latent class, then it is better to choose another model.

4. BEHAVIOR OF THE ESTIMATORS IN THE NOISE MODEL

We introduce the following notation to handle the noise: p^* is the (15 × 1)-vector of the cell probabilities $p^*(1,1,1,1), p^*(1,1,1,0), \ldots\ldots, p^*(0,0,0,1)$ [1]. The corresponding vector of the cell probabilities in the correct model is denoted by *p*. The relationships between *p*, *p** and the parameter ϑ are described by the functions $\varphi : [0,1]^9 \longrightarrow [0,1]^{15}$, $\eta : [0,1]^9 \longrightarrow [0,1]^{15}$:

$$p = \varphi(\vartheta),\ p^* = \eta(\vartheta) = \tau\varphi(\vartheta) + \frac{1-\tau}{16}\mathbf{1},$$

where **1** is the vector consisting of 1's. Let h_n be the vector of the relative cell frequencies. It is well known that (under mild regularity conditions)

$$\sqrt{n}(h_n - p^*) \xrightarrow{d} \mathbf{N}(0,\Sigma) \tag{3}$$

with

$$\Sigma_{tu} = \begin{cases} p_t^*(1-p_t^* & if\ t=u \\ -p_t^* p_u^* & if\ t \neq u \end{cases}$$

$(t,u=1,\ldots,15)$. Since h_n is the m.l.e. for p^* the m.l.e. for ϑ, say $\hat{\vartheta}_n$, satisfies the equation

$$h_n = \eta(\vartheta_n^*) .$$

Further, under mild conditions, ϑ_n^* is consistent and we have

$$\sqrt{n}(h_n - p^*) = \sqrt{n}(\eta(\vartheta_n^*) - \eta(\vartheta)) = R(\vartheta)\sqrt{n}(\vartheta_n^* - \vartheta) + o_P(1) . \tag{4}$$

Here $R(\vartheta)$ is the (15 × 9)-matrix of partial derivatives

$$R_{ts}(\cdot) = \frac{\partial \eta_t(\vartheta)}{\partial \vartheta_s}$$

at ϑ. On the other hand we have the asymptotic normality of the m.l.e.:

$$\sqrt{n}(\vartheta_n^* - \vartheta) \xrightarrow{d} \mathbf{N}(0, I_*^{-1}(\vartheta)), \tag{5}$$

---

[1] Note that $p_{16}^* = p^*(0,0,0,0) = 1 - \sum_{r=1}^{15} p_r^*$; $(p_1^*, \ldots, p_{16}^*)$ is not considered to avoid singularities.

where $I_*^{-1}(\vartheta)$ is the inverse of the Fisher information in model (2), i.e.:

$$I_{*rs}(\vartheta)=\sum_{j=0}^{1}\sum_{k=0}^{1}\sum_{l=0}^{1}\sum_{m=0}^{1}\frac{\partial p^*(j,k,l,m)}{\partial\vartheta_s}\frac{\partial p^*(j,k,l,m)}{\partial\vartheta_r}\frac{1}{p^*(j,k,l,m)}.$$

Thus, from (3), (4) and (5) it follows that

$$\Sigma = R(\vartheta)I_*^{-1}(\vartheta)R^t(\vartheta) \qquad (6)$$

The difference $h_n - p$ can be decomposed into

$$\sqrt{n}(h_n - p) = \sqrt{n}(h_n - p^*) + \sqrt{n}(p^* - p) = Z_n + \sqrt{n}B.$$

By (3) the first summand converges in distribution to $\mathbf{N}(0,\Sigma)$. The second summand is equal to

$$B=(1-\tau)(\frac{1}{16}\mathbf{1}-p).$$

Furthermore, by definition of the estimator $\hat{\vartheta}_n$ we have

$$\sqrt{n}(h_n - p) = \sqrt{n}(\varphi(\hat{\vartheta}_n)-\varphi(\vartheta)) = M(\tilde{\vartheta})\sqrt{n}(\hat{\vartheta}_n - \vartheta) + o_P(1). \qquad (7)$$

Here $M(\tilde{\vartheta})$ is the (15 × 9)-matrix of partial derivatives

$$M_{ts}(\cdot)=\frac{\partial\varphi_t(\vartheta)}{\partial\vartheta_s}$$

at a point $\tilde{\vartheta}$, which lies between $\hat{\vartheta}_n$ and ϑ. Observe that $R(\vartheta)=\tau M(\vartheta)$. So we have

$$M(\tilde{\vartheta})\sqrt{n}(\hat{\vartheta}_n - \vartheta) - \sqrt{n}B \xrightarrow{d} \mathbf{N}(0,\Sigma). \qquad (8)$$

Let us assume that we can replace $M(\tilde{\vartheta})$ by $M(\vartheta)$ (This is a very rough approximation!), then we have

$$M(\vartheta)\sqrt{n}(\hat{\vartheta}_n - \vartheta) \approx Z_n + \sqrt{n}B$$
$$M(\vartheta)^t M(\vartheta)\sqrt{n}(\hat{\vartheta}_n - \vartheta) \approx M(\vartheta)^t Z_n + \sqrt{n}M(\vartheta)^t B$$

From this we get as an approximation for the variance of the estimator $\hat{\vartheta}_n$

$$\begin{aligned}\mathbf{Var}\ \hat{\vartheta}_n &\approx n^{-1}(M(\vartheta)^t M(\vartheta))^{-1}M(\vartheta)^t\Sigma M(\vartheta)(M(\vartheta)^t M(\vartheta))^{-1}\\ &= n^{-1}(M(\vartheta)^t M(\vartheta))^{-1}M(\vartheta)^t R(\vartheta)I_*^{-1}(\vartheta)R(\vartheta)^t M(\vartheta)(M(\vartheta)^t M(\vartheta))^{-1}\\ &= n^{-1}\tau^2(M(\vartheta)^t M(\vartheta))^{-1}M(\vartheta)^t M(\vartheta)I_*^{-1}(\vartheta)M(\vartheta)^t M(\vartheta)(M(\vartheta)^t M(\vartheta))^{-1}\\ &= n^{-1}\tau^2 I_*^{-1}(\vartheta). \end{aligned} \qquad (9)$$

And as approximation for the bias we get
:

$$\textbf{Bias}\,(\hat{\vartheta}_n) \approx (M(\vartheta)^t M(\vartheta))^{-1} M(\vartheta)^t B$$

$$= (1-\tau)(M(\vartheta)^t M(\vartheta))^{-1} M(\vartheta)^t (\frac{1}{16}\mathbf{1} - p). \qquad (10)$$

Computing these approximations for the bias in the strong case we get the following results:

$\hat{p}_{A/1},\ \tau = 7/8$				
Asymptotic expression for Bias			Simulation Bias	
Symmetric	Asymmetric	n	Symmetric	Asymmetric
		32	0.026	0.000
		64	0.026	0.009
-0.0207	-0.01147	128	0.027	0.014
		256	0.029	0.015

Table 3: Approximation for the bias in the strong case

$\hat{p}_{A/1},\ \tau = 3/4$				
Asymptotic expression for Bias			Simulation Bias	
Symmetric	Asymmetric	n	Symmetric	Asymmetric
		32	0.051	0.008
		64	0.054	0.019
-0.0413	-0.02294	128	0.056	0.023
		256	0.057	0.024

Table 4: Approximation for the bias in the strong case $\hat{p}_{A/1},\ \tau = 3/4$

$\hat{\pi},\ \tau = 7/8$				
Asymptotic expression for Bias			Simulation Bias	
Symmetric	Asymmetric	n	Symmetric	Asymmetric
		32	-0.000	0.083
		64	-0.001	0.066
-0.0048	-0.0548	128	-0.000	0.058
		256	0.001	0.054

Table 5: Approximation for the bias in the strong case $\hat{\pi},\ \tau = 7/8$

$\hat{\pi},\ \tau = 3/4$				
Asymptotic expression for Bias			Simulation Bias	
Symmetric	Asymmetric	n	Symmetric	Asymmetric
		32	-0.001	0.147
		64	-0.000	0.136
-0.0095	-0.1095	128	0.000	0.133
		256	0.001	0.133

Table 6: Approximation for the bias in the strong case $\hat{\pi},\ \tau = 3/4$

Although the agreement between the simulated bias and the approximating term is not so good as in the case of the variance in the correct model these results can be interpreted as follows:

1. It is clear that the bias does not depend on the sample size. Furthermore, if the amount of noise is larger, then the bias becomes larger, as it was expected.
2. Let us consider the estimation of π. In the symmetric cases the bias is very small, almost zero. This corresponds heuristically to the fact of the simple binomial model with noise, i.e.: $\mathbf{P}(Z = 1) = \tau p + (1 - \tau)/2$. In this model the bias of the relative frequency as estimator of p is zero, if $p = 0.5$ (for all τ.) In our more complicated model one can use the following approach: By definition of the m.l.e. the estimators satisfy the following condition:

3.

$$\hat{\pi} = \sum_{j,k,l,m} \hat{p}_{1/jklm} H_{jklm} / n$$

where $p_{1/jklm}$ = $\mathbf{P}(X=1/A=j, B=k, C=l, D=m)$ (see: Goodman, 1974), equation (12). Now, simplify this equation to get a heuristic explanation and replace $\hat{p}_{1/jklm}$ by $p_{1/jklm}$. In this idealized case the bias is

$$\mathbf{E}\hat{\pi} - \pi = \sum_{j,k,l,m} p_{1/jklm} p^*(j,k,l,m) - \pi.$$

Computing this for the parameter constellations

$$\vartheta = (a, a, a, a, (1-a), (1-a)(1-a), (1-a), 0.5)$$

one gets that the bias is zero.

4. The bias of the estimator for $p_{A/1}$ is in the symmetric cases larger than the bias of the estimator for π, in the asymmetric case it is smaller. Further, comparing the simulated bias of the estimator for $p_{A/1}$ in the symmetric and the asymmetric cases, we see that it is smaller in the asymmetric model. But this is also only on the first view, namely in the asymmetric case with noise a large number o simulation results are eliminated because of indeterminate results. It is not clear why it is theoretical smaller. It could be because M is too rough an approximation.

Acknowledgments
This paper has been undertaken during stays of the authors at the University of Potsdam and ICIMAF, respectively. These stays were supported by the DAAD.

RECEIVED FEBRUARY 2009
REVISED JANUARY 2010

REFERENCES

[1] COX, D.R. and HINKLEY, D.V. (1982): **Theoretical Statistics**, Chapman & Hall, London

[2] GOODMAN, L.A. (1974): Exploratory latent structure analysis using both identifiable and unidentifiable models, **Biometrika,** 61, 215-231

[3] GONZÁLEZ, A., SÁNCHEZ, J.E. and HERNÁNDEZ, J. (2002): Sobre la estimación de los parámetros del modelo de clases latentes en condiciones de raleza. **Reporte de Investigación ICIMAF 2002-192**

[4] KENDALL, M.G. and STUART, A. (1973): **The Advanced Theory of Statistics.** Volume 2, Inference and Relationship, 3rd edition. Griffin, London

[5] VERMUNT, J. K. (1997): A general program for the analysis of categorical data at URL: http://cwis.kub.nl/~fsw_1/mto