Using Covariates in Loglinear Models with Sampling Zeros; A Cautionary Note

Jos Dessens, Wim Jansen, Peter G.M. van der Heijden

Dept. of Methodology & Statistics, Social Sciences Faculty, Utrecht University, P.O.Box 80140, 3508 TC Utrecht, The Netherlands. E-mail: j.dessens@fsw.ruu.nl

Abstract: When analyzing loglinear models, most computer packages will first construct the contingency table and subsequently fit the required loglinear model. Some computer packages allow the incorporation of covariates, such as age and/or income, into the loglinear model. In the SPSS modules LOGLINEAR and GENLOG this option is implemented by calculating for each pattern of the classifying variables the mean value(s) of the covariate(s), while this covariate vector is subsequently added to the design matrix describing the loglinear model.

In this article it is shown that the use of covariates in the case of contingency tables containing sampling zeros will lead to incorrect results for the SPSS modules LOGLINEAR and GENLOG. Parameter estimates, deviances and the number of degrees of freedom may be highly incorrect. This is illustrated in the case of the loglinear uniform association model, using an example from a handbook on categorical data analysis. A second, more complex example is discussed where, as a result of this incorrect handling of covariates, wrong results have been published in the literature.

It is concluded that (conditional) multinomial logit models should be used. While SPSS currently does not include such models, it is demonstrated how to obtain correct results for the case that covariates are functions of the classifying variables.

(SSNinCSDA 27, 239-245 (April 1998))

Keywords: SPSS, modules LOGLINEAR and GENLOG, loglinear model, incorrect and correct results, multinomial logit models

Received: September 1997 *Revised:* February 1998

I. Introduction

Several computer packages provide modules to fit loglinear models. SPSS, for instance, provides modules LOGLINEAR and GENLOG. An important feature of these SPSS modules is that an option is available for the incorporation of covariates of a continuous measurement level. For each pattern of the classifying variables the mean value(s) of the covariate(s) is (are) calculated, and subsequently used in estimating the model parameters.

This approach is illustrated for the data from Holmquist et al. (1967) on the agreement of diagnosis between two pathologists. These data were also used by Agresti (1990; 367-370). The data are given in Table 1.

Table 1. Pathologist Ratings of Carcinoma.

	Pathologist B				
Pathologist A	1	2	3	4	5
1	22	2	2	0	0
2	5	7	14	0	0
3	0	2	36	0	0
4	0	1	14	7	0
5	0	0	3	0	3

Source: Holmquist et al. (1967), cited in Agresti (1990).

Here 118 observations are classified on two variables: pathologist A, with levels i (i=1,...,5) and pathologist B, with levels j (j=1,...,5). Let the model

to be fitted be independence. We denote this model by

$$\log m_{ij} = u + u_{A(r)} + u_{B(c)} \qquad (1)$$

where m_{ij} is the expected frequency for cell (i,j). In order to identify the model the parameters $u_{A(r)}$ and $u_{B(c)}$, SPSS uses so-called 'deviation contrasts': $\Sigma u_{A(r)} = \Sigma u_{B(c)} = 0$.

In matrix terms, model (1) can be described by $\log \mathbf{m} = \mathbf{Xb}$. Here $\mathbf{X}$ is a 25 x 9 design matrix that has the following form for the independence model (1):

$$\log \begin{pmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ m_{15} \\ m_{21} \\ m_{22} \\ m_{23} \\ m_{24} \\ m_{25} \\ m_{31} \\ m_{32} \\ m_{33} \\ m_{34} \\ m_{35} \\ m_{41} \\ m_{42} \\ m_{43} \\ m_{44} \\ m_{45} \\ m_{51} \\ m_{52} \\ m_{53} \\ m_{54} \\ m_{55} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & -1 & -1 & -1 & -1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & -1 & -1 & -1 & -1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 1 & -1 & -1 & -1 & -1 & 0 & 1 & 0 & 0 \\ 1 & -1 & -1 & -1 & -1 & 0 & 0 & 1 & 0 \\ 1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} u \\ u_{A(1)} \\ u_{A(2)} \\ u_{A(3)} \\ u_{A(4)} \\ u_{B(1)} \\ u_{B(2)} \\ u_{B(3)} \\ u_{B(4)} \end{pmatrix}$$

Thus the first element of the parameter vector $\mathbf{b}$ is the general mean u in (1), elements two to five are $u_{A(1)}$ to $u_{A(4)}$, where $u_{A(5)}$ can be found by $u_{A(5)} = -(u_{A(1)} + ... + u_{A(4)})$, and element six to nine are $u_{B(1)}$ to $u_{B(4)}$, where $u_{B(5)}$ can be found by $u_{B(5)} = -(u_{B(1)} + ... + u_{B(4)})$.

Assume that for the data in Table 1 there exists a covariate Z whose values are collected into a 118-elements vector **z**. I.e., for every observation in Table 1 there is an observed value on a continuous covariate. Then SPSS allows the inclusion of this covariate in the analysis by constructing a grouped 25-element vector $\mathbf{z_g}$ to be concatenated on $\mathbf{X}$, where every element of $\mathbf{z_g}$ is the average of the observed covariate values of the variable Z.

II. The problem: covariates in the presence of sampling zeros

Which value for $\mathbf{z_g}$ should be imputed when an element refers to a sampling zero, affects estimates for both the parameters and the expected frequencies. This is clear from the likelihood equation for a covariate, that is $\mathbf{z_g'n} = \mathbf{z_g'm}$, where n is the 25-element vector of observed frequencies.

It would seem very difficult to give a motivation that holds in every possible situation. For example, when the mean of Z differs from zero, this option for the missing entry would seem unwise. Possibly a better option would be to enter the average of Z, or even a value derived from the values of Z in adjacent cells. More research is needed before it can be decided how to deal with sampling zeros. Perhaps we should even conclude that this problem for sampling zeros puts this whole approach of dealing with covariates into question.

III. Special cases that go wrong: the uniform association model and other examples

In principle, two types of covariates can be distinguished. The first type is independent of the values of the classifying variables and has been discussed in sections 1 and 2. Variables like age or income are examples of this type of covariate. The second type is a function of the values of the classifying variables. The use of these covariates allows the fitting of the uniform association model. We will give an example of the use of such a covariate, and show why zero-imputation for the covariate values for sampling-zero-cells is the wrong strategy.

When there are no sampling zeros, an important consequence of the possibility of incorporating covariates into loglinear models is that this allows loglinear association and related models to be fitted (Goodman, 1979). In the case of two classifying variables the most simple loglinear association model is that of constant or uniform association.

The uniform association model can be defined as:

$$\log m_{ij} = u + u_{A(r)} + u_{B(c)} + i*j*\phi \qquad (2)$$

If i and j are unit spaced fixed values (for example 1, 2, 3, 4, 5), model (2) implies that for adjacent cells the log odds-ratio is constant and equal to ϕ.

For Table 1 the model implies that, in matrix terms, the design matrix X given in section 1 must be supplemented with one extra column, having values i*j.

SPSS claims that its covariate procedure can be used to fit the uniform association model. First, with a compute-statement a new variable should be constructed having as values the product i*j. Second, this variable should be treated as a covariate in the loglinear analysis. Indeed, when there are no sampling zeros, for every cell (i,j) the covariate values will be equal to i*j, and the column $\mathbf{z}_g$ will contain the correct values.

If there are sampling zeros the corresponding values for $\mathbf{z}_g$ should be i*j. However, in the presence of sampling zeros SPSS imputes a zero value into $\mathbf{z}_g$. It is precisely this zero imputation that is responsible for the fact that SPSS will produce incorrect results.

Example 1

The example provided in section 1 may be used as an illustration that in the presence of sampling zeros, the fitting of association models with SPSS will produce incorrect results. The model fitted by Agresti is a loglinear uniform association model with a diagonal parameter:

$$\log m_{ij} = u + u_{A(r)} + u_{B(c)} + i*j*\phi + \delta_{(i-j)} \qquad (3)$$

Because the data concern agreement between two pathologists, the inclusion of a diagonal parameter seems obvious.

In matrix format model (3) can be defined as:

$$\log \begin{pmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ m_{15} \\ m_{21} \\ m_{22} \\ m_{23} \\ m_{24} \\ m_{25} \\ m_{31} \\ m_{32} \\ m_{33} \\ m_{34} \\ m_{35} \\ m_{41} \\ m_{42} \\ m_{43} \\ m_{44} \\ m_{45} \\ m_{51} \\ m_{52} \\ m_{53} \\ m_{54} \\ m_{55} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 2 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 3 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 4 & 0 \\ 1 & 1 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & 5 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 2 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 4 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 6 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 8 & 0 \\ 1 & 0 & 1 & 0 & 0 & -1 & -1 & -1 & -1 & 10 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 3 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 6 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 9 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 12 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & -1 & -1 & -1 & 15 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 4 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 8 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 12 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 16 & 1 \\ 1 & 0 & 0 & 0 & 1 & -1 & -1 & -1 & -1 & 20 & 0 \\ 1 & -1 & -1 & -1 & -1 & 1 & 0 & 0 & 0 & 5 & 0 \\ 1 & -1 & -1 & -1 & -1 & 0 & 1 & 0 & 0 & 10 & 0 \\ 1 & -1 & -1 & -1 & -1 & 0 & 0 & 1 & 0 & 15 & 0 \\ 1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 1 & 20 & 0 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 25 & 1 \end{pmatrix} \begin{pmatrix} u \\ u_{A(1)} \\ u_{A(2)} \\ u_{A(3)} \\ u_{A(4)} \\ u_{B(1)} \\ u_{B(2)} \\ u_{B(3)} \\ u_{B(4)} \\ \phi \\ \delta \end{pmatrix}$$

If the Holmquist et al. data are analyzed using the SPSS LOGLINEAR and GENLOG modules, the different versions of the SPSS package give different results, all of which also differ from that reported by Agresti. The results are given in Table 2. The results from Table 2 are rather confusing. SPSS Deviances, number of Degrees of Freedom, and parameter estimates differ from those given in Agresti. The authors of the SPSS software may be aware of this possibility, because SPSS LOGLINEAR produces the following warning:

> Warning # 12758. Command name: LOGLINEAR
>
> For the model specified, LOGLINEAR has encountered either many sampling zeroes or a mixture of large and small observed counts. Parameter estimates or degrees of freedom could be incorrect.

Table 2: Results as reported by Agresti (1990), SPSS LOGLINEAR [1], SPSS/PC+ 5.01 LOGLINEAR [2], SPSS 6.1 LOGLINEAR [3], and SPSS 6.1 GENLOG [4] for the Holmquist et al. Data

	Deviance	DF	Uniform Association parameter estimate (standard error)	Diagonal parameter estimate (standard error)
Agresti (1990)	8.41	14	1.15 (.34)	1.07 (.40)
SPSS LOGLINEAR	1.03	3	1.14 (.44)	0.68 (.46)
SPSS/PC+ 5.01 LOGLINEAR	1.03	3	1.14 (.44)	0.68 (.46)
SPSS 6.1 LOGLINEAR	1.03	14	1.14 (.44)	0.68 (.46)
SPSS 6.1 GENLOG	1.03	14	1.14 (.44)	0.68 (.46)

The number of degrees of freedom varies for the different versions of SPSS. In the SPSS 6.1 LOGLINEAR and GENLOG versions, the number of degrees of freedom equals that in Agresti. The SPSS Warning # 12758 does not contribute to the researchers understanding of what is going on. Are the parameter estimates and number of degrees of freedom to be trusted or not?

In the Agresti example, 12 from the 25 patterns cells are zero. The value of the uniform association covariate, which is the product of the values of the ratings of both pathologists, is incorrectly set to zero for patterns with a sampling zero, and this explains the difference between the deviance found by Agresti and the SPSS deviances, and the differences between the parameter estimates. The model fitted by SPSS is *not* the uniform association model, and consequently the deviances and parameter estimates are wrong. Incidentally, it also seems to affect the number of degrees of freedom in a rather unpredictable way. In the SPSS 6.1 modules this error has been fixed. For the diagonal parameter (δ) none of the diagonal cells has a sampling zero. Consequently, SPSS creates the correct vector for the diagonal parameter in the matrix to be concatenated to **X**. But, of course, the parameter estimate is as incorrect as all other parameter estimates in the model, since the incorrect column in the design matrix for the uniform association parameter affects all parameter estimates in the model.

Example 2

We will now show a published example where the results are incorrect due to the incorrect handling of covariates for sampling zeros in SPSS. Junger et al. (1995) found a relatively strong relation between the commitment of property crimes and the involvement in traffic accidents. This relation was found for both boys and girls, and for four age categories (12-14, 15-17, 18-20, 21-24). From survey data, information on these four variables was available for 2918 respondents. The relation between crime and traffic accidents was modelled as an uniform association parameter in a loglinear model. The 'basic' uniform association model appears as:

$$\begin{aligned}\log m_{ijkl} = {} & u + u_{P(i)} + u_{T(j)} + u_{S(k)} + u_{A(l)} \\ & + u_{PS(ik)} + u_{PA(il)} + u_{TS(jk)} + u_{TA(jl)} \\ & + u_{SA(kl)} + u_{PSA(ikl)} + u_{TSA(jkl)} + i*j*\phi \quad (4)\end{aligned}$$

P = property crimes, i=0,8
T = traffic accidents, j=0,2
S = sex, k=1,2
A = age, l=1,4

The vector corresponding to the parameter ϕ contains the figures 0 through 16 and has to be added as a covariate to the loglinear model with main and interaction effects.

Model (4) was fitted by Junger et al. using the SPSS LOGLINEAR module (appendix, job 3)[5]. We reran the model (4) in SPSS LOGLINEAR, SPSS/PC+ 5.01 LOGLINEAR, and SPSS 6.1 GENLOG. In this four-dimensional example 28 from the 216 cells are sampling zeros. As was by now expected, the various SPSS modules produced incorrect results (Table 3).

The number of degrees of freedom varies for the different versions of SPSS. Only in the GENLOG version is this number correct. Deviances and parameter estimates are wrong. Fitting the uniform association model in a proper way (described in section 5, and cross-checked using other software

Table 3. Results from SPSS LOGLINEAR, SPSS/PC+ 5.01 LOGLINEAR, and SPSS 6.1 GENLOG on the data from Junger et al.

	Deviance	DF	Uniform Association parameter estimate (standard error)
SPSS LOGLINEAR	134.72	123	.1495 (.0161)
SPSS/PC+ 5.01 LOGLINEAR	134.72	123	.1495 (.0161)
SPSS 6.1 LOGLINEAR [6)]	134.72	129	.1495 (.0161)
SPSS 6.1 GENLOG	134.72	127	.1495 (.0161)

packages such as GLIM (Francis et al., 1993)) will give the correct results: deviance = 142.37, df = 127 and the uniform association parameter = .1414 (.0158).and the uniform association parameter = .1414 (.0158).

In conclusion, one can say that not only is the uniform association model incorrectly estimated in the presence of sampling zeros, but all other association models such as the Row and/or Column Effect Association model (Goodman, 1979) on data with sampling zeros will also lead to wrong results in SPSS. The same is true for models in which one or more of the factors is/are treated as a linear variable.

V. A remedy for the special cases within SPSS: using data in tabular format

For the special cases discussed in section 4 there is a remedy within SPSS which is quite laborious, but which works. It turns out that SPSS produces correct results, if the input data matrix is *not* a matrix of individual observations by variables, but a matrix of response patterns with their frequency of appearance (i.e. data in tabular format). So the way to handle this problem in SPSS is (a) to read the matrix of individual observations by variables (i.e., for example 1 a matrix of 128 rows and 2 variables); (b) to write to a new file the frequencies plus the response patterns (i.e. a data matrix of 25 rows and three columns, namely the values for variables A and B and their frequency of appearance); (c) to read in this new file as input data matrix, call the frequency FREQ, and state WEIGHT BY FREQ; and (d) to carry out the desired loglinear analysis with covariates.

It then turns out that the correct values will appear in the covariate vector $\mathbf{z}_g$. The exact SPSS setup for the Holmquist data is given in the appendix, job 2 (second part), and for the Junger et al. data in job 4 of the appendix.

This remedy is appropriate only for special cases, and not in the more general situation where the covariate (for example age in years) does not depend on the values of the classifying variables (see section 1 and 2). In this situation the data cannot be represented by a tabular format. For these situations models are in order that do not require the data to be aggregated on the classifying variables. Conditional multinomial logit models, as described by Logan (1983) and DiPrete (1990) are the appropriate models here. SPSS does not provide modules for the fitting of such models. In the software packages LIMDEP (Greene, 1991) and STATA (1996), options for fitting conditional multinomial logit models are given.

VI. Conclusion

We have shown that for fitting loglinear models with covariates that depend on the values of the classifying variables, SPSS LOGLINEAR and GENLOG produce wrong results in the case of patterns with sampling zeros. Deviances, number of degrees of freedom, and parameter estimates are incorrect. In section 5 we showed a 'digressive' way to obtain the correct results with SPSS.

In all situations in which the value(s) of covariate(s) *do not* depend on the patterns of the classifying variables, one should use more adequate models.

Acknowledgment

The authors wish to thank Dr. Ming Long Lam (SPSS Inc.) and two anonymous referees for their comments on a previous version.

Notes

1) SPSS LOGLINEAR is the mainframe version of the loglinear SPSS module. See SPSS (1990).

2) SPSS/PC+ 5.01 in the MS-DOS version was used for the SPSS loglinear module. See Noru©is (1992).

3) The SPSS/PC+ LOGLINEAR syntax was run under SPSS 6.0 for Windows.
4) SPSS 6.0 for Windows was used for the SPSS GENLOG module. See Norušis (1994).
5) Jobs can be obtained in electronic format on request from the first author.
6) The SPSS/PC+ LOGLINEAR syntax was run under SPSS 6.0. The difference in the number of degrees of freedom results from this.

References

Agresti, A., *Categorical Data Analysis* (John Wiley & Sons, New York, 1990).

DiPrete, T.A., Adding Covariates to Loglinear Models for the Study of Social Mobility, *American Sociological Review*, 55 (1990) 757-773.

Francis, B. et al. *The GLIM System. Release 4 Manual.* (Oxford University Press, Oxford, 1993).

Goodman, L.A., Simple Models for the Analysis of Association in Cross-classifications having Ordered Categories, *Journal of the American Statistical Association*, 74 (1979) 537-552.

Greene, W.H., *LIMDEP Version 6.0. Users Manual and Refrence Guide.* (Econometric Software, Inc., Bellport, NY, 1991).

Holmquist, N.S., C.A. McMahon & O.D. Williams, Variability in Classification of Carcinoma in Situ of the Uterine Cervix. *Arch. Patholog.*, 84 (1967) 334-345.

Junger, M., G.-J. Terlouw & P.G.M. van der Heijden, Crime, Accidents and Social Control, *Criminal Behaviour and Mental Health*, 5 (1995) 386-411.

Logan, J.A., A Multivariate Model for Mobility Tables, *American Journal of Sociology*, 89 (1983) 324-349.

Norušis, M.J., *SPSS/PC+ Advanced Statistics. Version 5.0* (SPSS Inc., Chicago, 1992).

Norušis, M.J., *SPSS Advanced Statistics 6.1* (SPSS Inc., Chicago, 1994).

SPSS, *SPSS Reference Guide.* (SPSS Inc., Chicago, 1990).

StataCorp., *Stata Statistical Software. Release 5.0* (Stata Corporation, College Station, TX, 1997).

Appendix

1. SPSS job used in fitting the uniform association and diagonal parameter model for the Holmquist et al. data in a contingency table format.

```
data list free/path_a path_b freq.
begin data.
1 1 22
1 2  2
1 3  2
1 4  0
1 5  0
. .  .
. .  .
5 1  0
5 2  0
5 3  3
5 4  0
5 5  3
end data.
compute u=path_a*path_b.
compute diag=0.
if (path_a=path_b) diag=1.
weight by freq.
loglinear path_a(1,5) path_b(1,5) with u diag
 /print estim
 /design path_a path_b u diag.
```

2. SPSS job used in fitting the uniform association and diagonal parameter model for the Holmquist et al. data in datamatrix format.

```
data list free/path_a path_b.
begin data.
1       1
.       .
1       1
1       2
1       2
1       3
1       3
2       1
.       .
2       1
```

```
2       2
.         .
2       2
2       3
.         .
2       3
3       2
3       2
3       3
.         .
3       3
4       2
4       3
.         .
4       3
4       4
.         .
4       4
5       3
.         .
5       3
5       5
.         .
5       5
end data.
compute u=path_a*path_b.
compute diag=0.
if (path_a=path_b) diag=1.
loglinear path_a(1,5) path_b(1,5) with u diag
 /print estim
 /design path_a path_b u diag.
crosstabs var path_a(1,5) path_b(1,5)
 /tables=path_a by path_b
 /write=all.
data list file='spss.prc' free/d1 d2 freq path_a path_b.
compute u=path_a*path_b.
compute diag=0.
if (path_a=path_b) diag=1.
weight by freq.
loglinear path_a(1,5) path_b(1,5) with u diag
 /print estim
 /design path_a path_b u diag.
```

3. SPSS job used in fitting the uniform association model for the Junger et al. data.

```
(get file 'a:crime.sys'.)
compute ua=property*traffic.
loglinear  sex(1,2) property(0,8) traffic(0,2) age(1,4) with ua
  /print estim/noprint default/criteria iterate(50)
  /design sex property traffic age
   sex by property sex by traffic sex by age
   property by age traffic by age
   sex by property by age sex by traffic by age ua.
```

4. SPSS job, circumventing SPSS setting covariate values of the uniform association variable to zero for cells with zero frequencies.

```
(get file 'a:crime.sys'.)
procedure output outfile='a:crime_ta.dat'.
crosstabs var=property(0,8) traffic(0,2) sex(1,2) age(1,4)
  /tables=property by traffic by sex by age
  /write=all.
data list free file 'a:crime_ta.dat'/del1 del2 freq property traffic sex age.
compute ua=property*traffic.
weight by freq.
loglinear  sex(1,2) property(0,8) traffic(0,2) age(1,4) with ua
  /print estim/noprint default/criteria iterate(50)
  /design sex property traffic age
   sex by property sex by traffic sex by age
   property by age traffic by age
   sex by property by age sex by traffic by age ua.
```

(The examples can be downloaded at http://www.gsf.de/MED-STAT/SSN/Jansen.html)