The Analysis of Weighted Poisson Data

D-96-12
A.W. Vogelesang
Leidschendam, 1996
SWOV Institute for Road Safety Research, The Netherlands

Report documentation

Number:	D-96-12
Title:	The Analysis of Weighted Poisson Data
Author(s):	A. W. Vogelesang
Research manager:	S. Oppe
Project number SWOV:	74.207
Project code client:	This research was funded by the Dutch Ministry of Transport and Public Works
Keywords:	Mathematical model, statistics, analysis (math), method.
Contents of the project:	This report offers a description of the SWOV-program WPM ('Weighted Poisson Methods'). A comparision is made with the well-known SAS-GENMOND procedure, in order to define WPM in terms of a SAS-GENMOD procedure. Technical issues are raised that have to do with methodological differences between the two procedures.
Number of pages:	24 pp. + 21 pp.
Price:	Dfl. 22,50
Published by:	SWOV, Leidschendam, 1996

SWOV Institute for Road Safety Research
P.O. Box 1090
2260 BB Leidschendam
The Netherlands
Telephone 31703209323
Telefax 31703201261

Contents

1. Introduction

The aim of this paper is to describe the SWOV- program WPM. 'Weighted Poisson Methods', developed by De Leeuw and Oppe: De Leeuw, (1975), De Leeuw & Oppe (1976), and Oppe (1981, 1992, 1993), to compare it with the well- known SAS-GENMOD (SAS/STAT Software, 1993) procedure and to define WPM in terms of a SAS GENMOD procedure. Issues are raised that have to do with methodological differences between the two procedures. Data, calculations, results and SAS-setups are given in Appendices.

WPM was inspired by Goodman's (1970) hierarchical analysis of cross-classified data, it was new with respect to the possibility of differentially weighting Poisson distributed data and is similar to Andersen (1977, 1981). It has a very simple input with user defined orthogonal contrast vectors, and provides significance tests for every contrast specified. In practice, evaluation of the model fit to the data is done using the modified chi-squared method, but there is also a maximum likelihood (ML) version available.

1.1 Minimum Modified Chi-Squared Statistic

A comparison of chi-squared statistics is given in Agresti (1990, Chapters 12-13). The minimum *modified* chi-squared statistic is discussed in §13.5.1:

$$X^2[\pi(\theta), \mathbf{p}] = n \sum \frac{[p_i - \pi_i(\theta)]^2}{\pi_i(\theta)}; \quad X^2_{\text{mod}}[\pi(\theta), \mathbf{p}] = n \sum \frac{[p_i - \pi_i(\theta)]^2}{p_i} .$$

Neyman, in 1949, introduced minimum modified chi-squared statistics and showed that they are best asymptotically normal estimators. Bhapkar, in 1966, showed that minimum modified chi-squared estimators are identical to WLS-estimators. This statistic is then identical to the WLS residual χ^2-statistic for testing the fit of the model. When the model does not hold, estimators obtained by different models can be quite different (see Agresti, 1990).

Comparison of WPM- and SAS-results will be done using the ML versionand the minimum modified chi-squared method. Both methods are different from the *default* procedures for loglinear analysis in SAS, in more respects. GENMOD is primarily designed for generalised linear modelling (Poisson *regression*). To be able to define the WPM program in terms of a SAS procedure, we have to go through some theory first. The model, its roots, and the differences with respect to the SAS-GENMOD procedure for Poisson regression are treated in Section 2, together with differences due to using different chi-squared test statistics, the Likelihood Ratio (LR), Pearson's χ^2 or the related Wald statistic. Orthogonal contrasts are given in Section 3, test statistics using orthogonal contrasts in WPM are given in Section 4. The Generalised Linear Model, with concepts as 'link'- function and 'offset', is described in section 5. Section 6 describes the differences between *types* of sums of squares (Type 1 - 4), because SAS distinguishes between these and the distinction is relevant. Moreover, to mimic WPM in SAS we need Type 3 sums of squares, whereas Type 1 analysis is the default with SAS. Section 7 gives two different ways of restricting the *number* of parameters, these are the μ-model and the ANOVA-model (Freund & Littell, 1981). ANOVA restrictions mimic those in well-known model equations for analysis of variance: parameter estimates are departures from the grand mean μ. In the MEANS- or μ-model, the last level of each variable is set zero.

1.2. Examples

For illustration of the procedures, more examples are presented. The data are given in *Appendix 1*. Examples 1- 4 serve to illustrate computation of parameter estimates using either form of parametrisation (*Appendices* 2-6). Example 4 also illustrates the conversion from one parametrisation into another one (*Appendix* 7). WPM estimates are different from those obtained from SAS-GENMOD. This is because SAS-GENMOD is a procedure for generalised linear modelling (Poisson regression), whereas Goodman's procedure is a hierarchical decomposition of the logarithm of the probability that an observation will fall in cell (*i,j*) of an cross-classification. The decomposition is into main effects and interaction effects, in the same way as ANOVA is a hierarchical decomposition into main and interaction effects. In order to correct for small sample bias, Goodman (1970) and De Leeuw & Oppe (1976) added 0.5 to each cell count. This is not possible with SAS-GENMOD, because SAS-GENMOD only accepts integers. WPM results can be obtained from SAS-GENMOD by specifying orthogonal contrast vectors for the desired effects.
Computation of the Goodman parameters is exemplified in *Appendix* 6 (Example 4, ANOVA-model, without adding the 0.5) and in *Appendix* 9 (Goodman's data, including the 0.5). Orthogonal contrasts are given with the setups for the examples. Their orthonormal equivalents constitute the WPM-designmatrix. Clarifying comments are given with the text. Results are slightly different, because the procedures are not identical and because with WPM, 0.5 is added to each observation (see above). Using Goodman's 'Knowledge of Cancer data', we mimic WPM in SAS and compare results with those from WPM and from Goodman (1970). Setups and results are presented in *Appendix* 9. For Oppe's (1993) BAG-data we compare the results of Poisson regression using orthogonal contrasts, Type 3 analysis and Wald statistics, with the results obtained using WPM (*Appendix* 10).

The discussion is illustrated in the following examples (*Appendix* 1):
- Example 1: a 2×2 cross-classification, unweighted;
- Example 2: a 2×2 cross-classification, weighted (very simple);
- Example 3: Example 1, differentially weighted;
- Example 4: a 2×4 cross-classification, unweighted;
- Example 5: Oppe's BAG-data 2x2×4 cross-classification, weighted;
- Example 6: Goodman's data: the 'Knowledge of Cancer Data'

Nearly all examples are analyzed using WPM-ML. Only one set of data, Oppe's BAG data, is analyzed using the modified chi-squared method. The data are given in *Appendix* 1, the SAS-setups in *Appendix* 8.

2. Weighted Poisson Model & Loglinear Analysis

The use of the Poisson model for contingency tables goes back to Sir Ronald Fisher. When the parameter of interest is the ratio of Poisson means or the value of a Poisson mean as a fraction of the total, it is usually appropriate to condition on the observed total. Conditioning on the total leads to multinomial or binomial response models of the log-linear type (McCullagh et al., 1989, p. 213). The connection between the two stems from the fact that the binomial and multinomial distributions can be derived from a set of independent Poisson random variables conditionally on their total being fixed.

Dyke and Patterson (1952) analyzed cross-classified survey data concerning the proportion of subjects who have a good knowledge of cancer. The recorded explanatory variables were exposures to various information sources, newspapers, radio, solid reading, lectures. A factorial model was postulated in which the logit of success, $\log\{p/(1-p)\}$, is expressed linearly as a combination of the four information sources and interactions among them. Success in this context is interpreted as 'good knowledge of cancer'. This data is the running example in Goodman (1970).

WPM was designed for the analysis of Poisson distributed data in cross-classifications (cf. Andersen, 1977), with the possibility of differentially weighting the cells of the cross-classification. It is also referred to as a 'multiplicative Poisson model', which means that main effects and interaction effects are multiplicative instead of additive (as in the ANOVA-model). Under Poisson sampling, cell counts are independent Poisson variables. The cell count is denoted m_{ij} $(i = 1, \ldots, r; j = 1, \ldots c)$, has expected value μ_{ij}, and the probability function for μ_{ij} has the Poisson form.

WPM is similar to Goodman's (1970) direct approach, it is a *weighted* version of Goodman's model. It is an ANOVA-like decomposition of the expectation of the logarithm of the cell count into main effects and interaction effects. The analysis is symmetric in all variables.

2.1. The SAS-GENMOD Approach to WPM

To simulate a weighted Poisson analysis such as WPM in SAS-GENMOD, we specify a Type 3 analysis (see Section 6) for the orthogonal contrasts defined on the serial numbers ('No') of the observations,. This is because all effects are defined as orthogonal contrasts. We use 'No' as a 'hypervariable', subsuming all effects. We ask for 'Wald' statistics to obtain Pearson's χ^2-statistic for each effect (see *Appendix* 8), because WPM only presents Pearson's statistic. With SAS, default options are Type 1 sums of squares and the Likelihood-Ratio (LR) test statistic. The advantage of the LR-statistic is that it can be additively decomposed into contributions of constituting effects. Differences between LR-ratios are also chi-squared distributed (Goodman, 1970; McCullagh et al., 1989).

Parameter estimates are obtained using \E ('estimates'). [Note that specifying a *log* link function results in a natural log ('ln') transformation of the data.] A Goodman analysis may be characterised as a 'χ^2-decomposition', which means that the total sum of squares is decomposed into all possible main effects and interaction effects[1]. Each effect can be further decomposed into independent standard normal distributed *z*-values (or chi-squared distributed variables with one degree of freedom). Using SAS-GENMOD, data can be analysed with different link functions. For a weighted Poisson analysis corresponding to WPM we specify:

- the link function as log (see section 5);
- offset var: the variable containing ln(weight) for each observation;
- Type 3 analysis (partialised effects);
- WALD statistics (yielding Pearson's χ^2-statistic);
- \E for parameter estimates;
- orthogonal contrasts as in WPM, with additional options: \E WALD.

The Pearson χ^2-values for the orthogonal contrasts in SAS-GENMOD each have one degree of freedom, hence the *square roots* of these values are $N(0,1)$ effects. The Goodman/WPM standardised effects are obtained from the Pearson χ^2-values under Type 3 SS by taking the square root.

SAS-GENMOD *only* accepts counts. In Goodman's approach, the problem of sparse data is handled by adding 0.5 to each cell count. In transforming each observed value by f(value) = 10 × (value + 0.5), we have counts, but a factor 10 too large, which can be down-weighted again using the *offset*-option. The compensation of f(value) is to divide each observation by ln(10). This is done by adding a weighting variable 'var1' to the data set. The new variable, 'var1' has a constant value, ln(10), for each observation. In the GENMOD-setup we specify 'offset = var1'. In doing so, each (log-transformed) expectation is divided by ln(10), see *Appendix* 9. Goodness-of-fit chi-squared values must be divided by 10.

---

[1] For ease of comparison with related techniques, and because of the advantage of decomposability of effects, the Likelihood-ratio test statistic might be implemented in WPM.

2.2. Theory and Formulae for Loglinear Analysis

Theory and formulae are taken from Goodman (1970), Fienberg (1980), Agresti (1984, 1990), McCullagh and Nelder (1989). The expected value for an observation (i,j) in a two-way cross-classification under the hypothesis of independence of row and column effects is $E(y_{ij}) = m_{ij} = (x_{i+} x_{+j})/N$, where $i = 1,..., r$, $j = 1, ..., c$; x_{i+} and x_{+j} are marginals and N is total observed.

For the logarithmic model: $\log m_{ij} = \log x_{i+} + \log x_{+j} - \log N.$

In shorthand notation: $\log m_{ij} = \mu + \alpha_i + \beta_j$,

and μ is the grand mean of the logarithms of the expected frequencies under the model of independence:

$\mu = 1/IJ \, \Sigma^{I}_{i=1} \Sigma^{J}_{j=1} \log m_{ij}$, and

$\mu + \alpha_i = 1/J \, \Sigma^{J}_{j=1} \log m_{ij}$

is the mean of the logarithms of the expected frequencies in the J cells at the i^{th} level of the first variable. Mutatis mutandis,

$\mu + \beta_j = 1/I \, \Sigma^{I}_{i=1} \log m_{ij}$,

and α_i and β_j are *deviations* from the grand mean, μ: $\Sigma^{I}_{i=1} \log \alpha_i = \Sigma^{J}_{j=1} \log \beta_j = 0$. From this, parameter estimates for unsaturated models follow immediately. Below, row and column effects for Example I, a 2x2-tabel (see *Appendix* 1, Table 1.1, and *Appendix* 2, Tables 2a - 2d) are given:

$\mu = 1/4 \, \Sigma^{2}_{i=1} \Sigma^{2}_{j=1} \log m_{ij} = 1/4 \times 19.015 = 4.7539$
$\mu + \alpha_1 = 1/2 \, \Sigma^{2}_{j=1} \log m_{1j} = 1/2 \times 8.799 = 4.400$
$\mu + \alpha_2 = 1/2 \, \Sigma^{2}_{j=1} \log m_{2j} = 1/2 \times 10.216 = 5.108$
$\mu + \beta_1 = 1/2 \, \Sigma^{2}_{i=1} \log m_{i1} = 1/2 \times 9.830 = 4.915$
$\mu + \beta_2 = 1/2 \, \Sigma^{2}_{i=1} \log m_{i2} = 1/2 \times 9.185 = 4.592$

Therefore,
$\mu = 4.7539$
$\alpha_1 = 4.4000 - 4.7539 = - 0.3541$
$\alpha_2 = 5.1080 - 4.7539 = + 0.3541$
$\beta_1 = 4.9153 - 4.7539 = + 0.1614$
$\beta_2 = 4.5925 - 4.7539 = - 0.1614$

2.3. SAS: Last Level Estimates Absorbed in Intercept

In SAS, the intercept is estimated from the *last* level within each factor. The parameters for every last level are set zero in view of the number of parameters that can be uniquely estimated from the data. This is called the Means Model or μ-Model. Another strategy to restrict the number of parameters in the model is the ANOVA-model, in which the sum of parameter values within an effect must be zero. In the ANOVA-model, the grand mean determines the intercept, in the μ-Model, the grand mean *plus* the parameters of the last level of each variable determine the intercept. This is accomplished by subtracting the *last level value* from each separate variable level. Thus, the intercept depends on the model. However, the resulting model equations will be the same. Just fill in the parameter estimates provided by the program. In *Appendix* 7, it will be shown how to translate parameter estimates from one model to the other one.

2.3.1. *Intercepts in μ-model*

Example I, continued
Using the above equations, we find intercept estimates for the μ-Model:
1) μ = 4.7539 (mean only);
2) $\mu + \alpha_2$ = 4.7539 + .3541 = 5.1080 (mean + rows);
3) $\mu + \beta_2$ = 4.7539 - .1614 = 4.5925 (mean + columns);
4) $\mu + \alpha_2 + \beta_2$ = 4.7539 + .3541 - .1614 = 4.9466 (all effects);

These estimates are obtained using SAS by performing different analyses, one for each model. The intercept depends on the model and includes the last levels of all specified effects . Further estimates in the μ-model are:
- row effect: $\alpha_1 - \alpha_2$ = - 0.7082
- column effect: $\beta_1 - \beta_2$ = 0.3228

Estimates are obtained by performing analyses for each model. For each row, column, or interaction effect, the last-level value is subtracted.

2.3.2. *Parameter estimates in μ-model*

Example I, continued
(1) row effect: from each row, the last level value (.3541) is subtracted:
Row 1: $\alpha_1 - \alpha_2$ = -.3541 - (+.3541) = - .7082;
Row 2: $\alpha_2 - \alpha_2$ = 0;

(2) column effect: from each column, the value (-.1614) is subtracted:
Column 1: $\beta_1 - \beta_2$ = .1614 - (-.1614) = .3228;
Column 2: $\beta_2 - \beta_2$ = 0;

These values are obtained with SAS. ANOVA-model estimates are:
row 1 ($\mu + \alpha_1$): 5.1080 -.7082 = 4.3998, row 2 ($\mu + \alpha_2$): 5.1080;
column 1: 4.5925 + .3228 = 4.9153 ($\mu + \beta_1$), column 2 ($\mu + \beta_2$): 4.5925.

3. Orthogonal Contrast Vectors: Hierarchical Analysis

There is an intimate connection between analysis of variance and the technique of planned orthogonal comparisons ('contrasts'). Each degree of freedom associated with treatments in a fixed-effects analysis of variance corresponds to a possible comparison of means. The number of degrees of freedom for the mean square between treatments is the number of *independent* comparisons to be made on the means. Any analysis of variance is equivalent to a breakdown of the data into hierarchically ordered *sets* of orthogonal comparisons (Hays, 1988, Ch. 11.9, Ch. 16). There are as much contrasts to be tested as there are physical features to be examined. A contrast is a linear function such that the elements of the coefficient vector sum to zero for each effect. The elements constituting a contrast constitute a set of weights, $c(i)$, such that $\Sigma_i\, c(i) = 0$. For example, for the 2×4 table of Example 4, we can form one contrast between rows and three contrasts between columns. In general, we can define (r-1) independent contrasts for rows, (c-1) independent contrasts for columns, and (r-1)(c-1) interaction contrasts. All contrasts must be orthogonal, that is, the sum of the products of corresponding elements of the contrast vectors must be zero.

3.1 Row contrasts: *R*

Contrast R:	+1	-1	first row, second row sum is zero

There are two rows, the first row is multiplied by +1, the second row by -1. With four columns, there are several possibilities. We may test the first two columns against the last two (Contrast A1, see below). Alternatively, we may test the first column against the last three columns, the second column against the last two columns and the third column against the fourth (4-1 = 3 independent contrasts). Contrast A1 may be followed by A2 and A3, two contrasts *nested within* the two levels of A1. An interaction contrast is a test (contrast) for the interaction, e.g., between R and A; an interaction contrast is the product of two main effect contrasts over all cells, e.g., R×A1, R×A2, and R×A3, and R×B1, R×B2, and R×B3:

3.2 Column Contrasts

Alternative Column Contrasts: A, B

Contrast A1:	+1	+1	-1	-1	sum is zero
Contrast A2:	+1	-1	0	0	id.
Contrast A3:	0	0	+1	-1	id.

or:

Contrast B1:	+3	-1	-1	-1	sum is zero
Contrast B2:	0	+2	-1	-1	id.
Contrast B3:	0	0	+1	-1	id.

3.3. Interaction Contrasts

Contrast R×A1:	+1	+1	-1	-1	first row of contrast matrix
	-1	-1	+1	+1	second row contrast matrix
Contrast R×A2:	+1	-1	0	0	id.
	-1	+1	0	0	
Contrast R×A3:	0	0	+1	-1	id.
	0	0	-1	+1	

or:

Contrast R×B1:	+3	-1	-1	-1	id.
	-3	+1	+1	+1	
Contrast R×B2:	0	+2	-1	-1	id.
	0	-2	+1	+1	
Contrast R×B3:	0	0	+1	-1	id.
	0	0	-1	+1	

The test for contrast A1 and the tests of the differences between the levels nested within A1 (contrasts A2 and A3) are independent because the contrasts are independent: $\Sigma_i A1(i) \times A2(i) = 0$, $\Sigma_i A1(i) \times A3(i) = 0$, and $\Sigma_i A2(i) \times A3(i) = 0$. The same is true for the B-contrasts (B1, B2, and B3 are independent). Note that the A- and B-contrasts are *not* independent. A and B represent different hypotheses concerning the differences between the levels of the column factor and, hence, cannot both be used in one analysis. Note that the presented weights are correct up to a normalising constant. The correct weights are

$$w_g = \sum_i \frac{c_i^2}{n_i}$$

for the g^{th} contrast (Hays, 1988).

4. Test Statistics in WPM

WPM is similar to Goodman's (1970) direct estimation method and Andersen's (1977) method for Poisson analysis in cross-classifications (see above). Apart from differences in estimation procedures, hierarchical parameter estimates generally will differ from their loglinear analogues, because of the correction for small sample bias, $m_{ij} + 0.5$ (see section 1). Resulting test statistics have smaller bias and smaller mean square error (Goodman, 1970; Agresti, 1984, 1990).
In *Appendices* 9 and 10, WPM is defined in terms of SAS-GENMOD. Goodman's 'Knowledge-of-Cancer' data illustrate the decomposition when adding 0.5 to each m_{ij}. GENMOD only accepts counts. Adding '5', downweighting by ln(10), and dividing the χ^2–statistic by 10 (variances of counts are squared), gives the desired result. The setup for Oppe's BAG-data illustrates the procedure with and without adding 0.5 to each m_{ij}. Results are compared with those obtained by GENMOD.
Main effects and interactions are defined in terms of odds ratios, test statistics are based on the log-odds ratios. In principle, there are no dependent ('response') variables in Goodman's model. The analysis is a decomposition of the cell counts into main effects and interactions, as is ANOVA for normally distributed data. All effects have one degree of freedom and the resulting test statistic can be referred to percentage points of the standard normal distribution. If there are more categories for one variable, log-odds ratios become 'continuation odds-ratios' (Goodman, 1970; Agresti, 1990) with a fixed reference category (the first one). The corresponding contrast vectors are contrasts with respect to the first one. WPM also contains 'nested' contrasts, levels nested in a hierarchically higher ordered effect.

4.1. Odds Ratios and Cross-Product Ratios

Let π_{ij} denote population probabilities in a 2x2 table. Within row 1 the odds that the response is in column 2 instead of column 1 is defined to be

$$\Omega_1 = \frac{\pi_{12}}{\pi_{11}}$$

where Ω_1 is called the *odds*, the ratio of the chances for π_{12} against the chances for π_{11}. Within row 2, the corresponding odds equals

$$\Omega_2 = \frac{\pi_{22}}{\pi_{21}}$$

Each Ωi is nonnegative, with value greater than 1.0 if response 2 is more likely than response 1. The ratio of these odds,

$$\theta = \frac{\Omega_2}{\Omega_1} = \frac{(\pi_{22}/\pi_{21})}{(\pi_{12}/\pi_{11})} = \frac{(\pi_{11}\pi_{22})}{(\pi_{12}\pi_{21})}$$

is referred to as 'the odds ratio'. An alternative name is the *cross -product ratio* since θ equals the ratio of the products $\pi_{11}\pi_{22}$ and $\pi_{12}\pi_{21}$ of proportions of cells that are diagonally opposite. The variables are independent if and only if the two odds are identical ($\Omega_1 = \Omega_2$). In this case the odds ratio $\theta = 1$. In practice, the population proportions $\{\pi_{ij}\}$ are unknown parameters, and hence so is θ. For sample cell frequencies $\{m_{ij}\}$ a sample analog of θ, $\hat{\theta}$, is given below, together with $\tilde{\theta}$, which has smaller bias and smaller mean square error (see Agresti, 1984):

$$\text{sample value } \hat{\theta} = \frac{m_{11}m_{22}}{m_{12}m_{21}},$$

$$\text{preferred estimator } \tilde{\theta} = \frac{(m_{11} + 0.5)(m_{22} + 0.5)}{(m_{12} + 0.5)(m_{21} + 0.5)}$$

4.2 Log-Odds Ratios, Goodness of Fit, Adding 0.5

The odds ratio is a multiplicative function of the cell proportions. Its logarithm is an additive function, namely, $\log \theta = \log \pi_{11} - \log \pi_{12} - \log \pi_{21} + \log \pi_{22}$. Log $\hat{\theta}$ converges faster than does $\hat{\theta}$ to its asymptotic distribution. The asymptotic standard deviation of log $\hat{\theta}$, denoted by $\sigma(\log \hat{\theta})$, can be estimated by

$$\hat{\sigma}(\log \hat{\theta}) = \left(\frac{1}{m_{11}} + \frac{1}{m_{12}} + \frac{1}{m_{21}} + \frac{1}{m_{22}}\right)^{1/2}$$

An approximate 100(1-p) percent confidence interval for $\log \theta$ is given by

$$\log \hat{\theta} \pm z_{p/2}\, \hat{\sigma}(\log \hat{\theta})$$

where $z_{p/2}$ is the percentage point from the standard normal distribution corresponding to a two-tail probability equal to p. The corresponding confidence interval for θ can be obtained by exponentiating endpoints of the confidence interval for log θ. One should not form confidence intervals for θ directly using $\hat{\theta}$ and its standard error because of its slower convergence to normality and because this one is not equivalent to the one obtained using $1/\hat{\theta}$ and its standard error (Agresti, 1984, p. 17). Again, the estimates of θ and of $\sigma(\log \theta)$ have smaller asymptotic bias and mean square error if the $\{m_{ij}\}$ are replaced by $\{m_{ij} + 0.5\}$.

5. Generalised Linear Models

Loglinear models often are written in terms of the Generalised Linear Model (GLM, cf. McCullagh & Nelder, 1989). The classical linear model is of the form

$$E(\mathbf{Y}) = \boldsymbol{\mu} \qquad \text{where} \qquad \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}. \tag{1}$$

The components of **Y** are independent normal variables with constant variance σ^2. The model has three components:

1. The *random component*: the components of **Y** have independent Normal distributions with $E(\mathbf{Y}) = \boldsymbol{\mu}$ and constant variance σ^2 ;
2. The *systematic component*: covariates $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_p$ produce a linear predictor $\boldsymbol{\eta}$ given by
$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta};$$
3. The *link* between the random and systematic components is the *identity* link:
$$\boldsymbol{\mu} = \boldsymbol{\eta}.$$

The generalisation introduces a link function between the linear predictor $\boldsymbol{\eta}$ and the expected value $\boldsymbol{\mu}$ of the random component. In the classical linear model $\boldsymbol{\eta}$ is identical to $\boldsymbol{\mu}$, but in the generalised model $\boldsymbol{\eta}$ is a function of $\boldsymbol{\mu}$:

$$\eta_i = g(\mu_i)$$

and $g(.)$ is called the *link function*.

In this formulation, classical linear models have a Normal distribution in component 1 and the identity function for the link in component 3.

In a univariate 'generalised lineair model', Y is a non-lineair function of X and η is a non-lineair transformation, which is needed e.g. if Y is a sum of discrete events with $0 \leq \mu = E(Y) \leq \infty$. In using the logarithmic transformation $\log(Y)$ for Poisson distributed variables and the logistic transformation $\log p/(1-p)$ for binomial distributed variables, the range of the function will be $(-\infty, +\infty)$. Some well-known link functions are:

1. *log*	$\eta = \log(\mu)$;
2. *logit*	$\eta = \log\{\mu/(1-\mu)\}$;
3. *probit*	$\eta = \Phi^{-1}(\mu)$.
4. *identity*	$\eta = \mu$

With each link function, a different error structure (random component) is associated. The link function maps the argument on the real line.

6. Orthogonality: Partial and Sequential Sums of Squares

In order to find out what contributions particular explanatory variables have to a model (e.g., what their maximum contribution is or their unique contribution), four types of sums of squares (Type 1 - Type 4) are distinguished. As Freund & Littell (1981, p. 103) note, these approaches relate to: (1) the *orthogonality* of effects and (2) the involvement of the cell sample sizes in the linear function of the parameters tested:

SS and Associated Hypotheses for the With-Interaction Model

Effect	Type 1	Type 2	Type 3 = Type 4
A	$R(\alpha \mid \mu)$	$R(\alpha \mid \mu, \beta)$	$R(\alpha \mid \mu, \beta, \alpha\beta)$
B	$R(\beta \mid \mu, \alpha)$	$R(\beta \mid \mu, \alpha)$	$R(\beta \mid \mu, \alpha, \alpha\beta)$
A×B	$R(\alpha\beta \mid \mu, \alpha, \beta)$	$R(\alpha\beta \mid \mu, \alpha, \beta)$	$R(\alpha\beta \mid \mu, \alpha, \beta)$

Type 1 functions correspond to adding each factor *sequentially* to the model in the order listed. Type 1 SS are the ANOVA-sequence of sums of squares. It reflects differences between unadjusted means of a factor as if the data consists of a one-way structure.

Type 3 analysis is associated with '*partial*' sums of squares, like in regression analysis, where each regression coefficient is a 'partial' regression coefficient reflecting the influence of one variable *corrected for* the influence of all others. Its principal use is in situations which require a comparison of main effects even in the presence of interaction.

Type 2 functions are neither just sequential, neither completely partial. There is partialising of other effects unless they are contained in the first effect. Thus, with A, B, and A×B as effects, testing A means partialising B, but *not* partialising A×B, because A×B is contained in A (part of A).

Type 4 functions are designed primarily for situations where there are empty cells; it is based on 'estimable' functions (linear functions of the parameters). Type 4 SS and estimable functions are identical to those provided by Type 3 when there are no empty cells.

With SAS-GENMOD, Type 1 and Type 3 sums of squares can be obtained.

The difference between models and their associated sums of squares ('SS') is more easily explained using 'reduction notation' (see Freund and Littell, 1981; Searle, 1987). Also, the situations in which they should be used is treated.

Denote by Model SS_1 the sum of squares ('SS') for a regression model with $m = 5$ x-variables:

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon\,,$$

and by Model SS_2 the SS for a reduced model not containing x_4 and x_5:

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon.$$

Reduction notation is used to represent the difference between regression SS for the two models. The difference $R(\beta_4, \beta_5 \mid \beta_0, \beta_1, \beta_2, \beta_3)$ indicates the increase in sum of squares due to the addition of β_4, and β_5 to the reduced model:

$$R(\beta_4, \beta_5 \mid \beta_0, \beta_1, \beta_2, \beta_3) = \text{Model } SS_1 - \text{Model } SS_2\,.$$

The expression $R(\beta_4, \beta_5 \mid \beta_0, \beta_1, \beta_2, \beta_3)$ is also referred to as:
(1) the sums of squares due to β_4, and β_5 (or x_4 and x_5) *adjusted for* (corrected for) β_0, β_1, β_2, β_3 (or the intercept and x_1, x_2, and x_3)
(2) SS due to fitting x_4 and x_5 after fitting the intercept and x_1, x_2, x_3
(3) the effects of x_4 and x_5 *above and beyond* or *partial of* the intercept and x_1, x_2, x_3.

7. Parametrisation: μ-Model or ANOVA-Model

Any model for **ANalysis-Of-VAriance** (ANOVA) or regression analysis can be formulated in terms of the product of a *design* matrix (in the case of ANOVA and loglinear analysis) or *data* matrix (in the case of regression analysis) **X** and a vector β of parameters:

$$\mathbf{Y} = \mathbf{X}\beta ,$$

where **Y** is the vector of observations for the dependent variable. In regression analysis, **X** may be the matrix containing the independent variables. In the analysis of variance, **X** is a designmatrix, each column of which corresponds to one parameter. The number of parameters to be estimated has to be restricted in accordance with the number of independent observations. To do this, there are at least two approaches:

- **μ-model**: the parameter for the last level of each variable is set zero,
- **ANOVA-model**: the sum of the deviations from the mean is zero.

SAS uses the μ-model parametrisation. The conversion from one parametrisation to the other is exemplified in *Appendix* 7.

7.1. ANOVA-Parametrisation: Deviations from μ.

The notation for the ANOVA-model is:

$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ and $\alpha_i = \mu_i - \mu$, and

y_{ij} = j^{th} observation for i^{th} group
ε_{ij} = random error with mean = 0 en variance = σ^2
i = 1, ... , c; j = 1, ... , n_i; c = number of groups
n_i = number of observations in the i^{th} group.

In the ANOVA-model, μ serves as the 'baseline' value and the means of the respective levels are deviations from μ :

$\mu_i = \mu + \alpha_i$ under the restriction $\Sigma\alpha_i = 0$.

The **deviations from μ** are represented by the α's. Only if the α-parameters satisfy certain identification constraints, unique estimates for parameters will be available. The identification constraints for the ANOVA model are $\Sigma\alpha_i$ = 0. The mean μ is the mean over all levels:

$$\mu = (\mu_1 + \mu_2 + ... + \mu_c)/c = \{(\mu + \alpha_1) + (\mu + \alpha_2) + ... + (\mu + \alpha_c)\}/c .$$

7.2. μ-Model: Deviations from Last Levels

Mostly, the following notation is used for the μ-model or *Means*-model:

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

This model is called the 'μ-model' or 'Means model' because the group-means $\mu_1,...,\mu_c$ are the parameters that determine the model. The Grand Mean μ is the mean over all levels:

$$\mu = (\mu_1 + \mu_2 + ... + \mu_c)/c .$$

The mean of the last level (μ_c) is set to zero for each variable. **Parameter values are deviations from the last level** (which is zero). The μ-model parametrisation is defined as

$$\mu_c = 0; \ \hat{\mu}_i = \mu_i - \mu_c = \mu_i. \text{ Also,}$$

$\hat{\mu}_i = \bar{y}_i = (\Sigma_j y_{ij})/n_i$ is the mean of the n_i observations in group i.

7.3. From ANOVA-Model to μ-Model and Vice Versa

To show the connection between the ANOVA-model and the μ-model, we manipulate both sets of restrictions:

$\Sigma_i \alpha_i = 0$	in the ANOVA-model, and
$\mu_c = 0$	in the μ-model.

From the restrictions the reparametrisation follows:

(ANOVA-model) $\alpha_i - \alpha_c = \mu_i - \mu_c = \mu_i$ (μ-model).

Summation over i yields:

(ANOVA-model)	$0 - c \times \alpha_c$	$= \Sigma_i \mu_i$	(μ-model), thus
(ANOVA-model)	α_c	$= - \Sigma_i \mu_i / c$	(μ-model), thus
(ANOVA-model)	last level	= minus the mean	(μ-model).

The reparametrisation from ANOVA-model to μ-model and vice versa is treated and exemplified in *Appendix* 7. Example 4 is used to compute μ-model and ANOVA-model estimates. Also, the conversion from one model to the other is shown.

8. Examples

Five examples served to illustrate the conversion from μ-model to ANOVA-model and the difference between SAS.GENMOD and Goodman's procedure. A *weighted* version of Goodman's procedure for loglinear analysis was programmed by Oppe at SWOV, for the analysis of data that had to be weighted. The BAG ('Blood Alcohol Level') data are an example of this. In traffic research a lot of weighting is called for: correction for length of time in control, compensation for road segment length, etc.. Multiplicative Poisson models with unequal cell weights are necessary tools for the road safety researcher.

The presentation of the first two examples has two objectives. First of all, it serves to illustrate the notation and computation. In the second place it serves to illustrate the weighting procedure and the *offset* option in SAS-GENMOD, a procdure that mimics the option with the same name in GLIM (Aitkin et al., 1989). Both in SAS.GENMOD and in GLIM, a linear predictor and a vector of expected values are prepared. Apart from the exponential transformation, the two vectors are equivalent if no weights are involved. If the analysis includes weights for the data, the offset option or the weight function can be used.

Offset

The offset option comes into effect *before* the analysis. Constant weights, such as ln(10) in our case, are applied to the linear predictor, they are '*offset*' (set apart) from the calculations needed to fit a generalised linear model. These calculations involve the technique of iteratively reweighted least squares. The use of an offset variable is illustrated in Examples 2 and 3, in the BAG-data (*Appendix* 3, 4, 8, respectively), and, especially, in the Goodman data (*Appendix* 9). If the weights for rows are proportional (as in *Appendix* 4, Table 4b), predicted values ('Linear Predictor' in GLIM) are proportional. For example, in the Goodman data (*Appendix* 9), we added 5 to each cell count, which had to be downweighted to 0.5 afterwards. To accomplish this, we prepare a vector of ln(10) in the data step, we declare it an 'offset' that has to be subtracted from the linear predictor, before expected values are computed.

Parameter Estimation
Example 4 (*Appendix* 5) is used to illustrate parameter estimation, both in the μ-model and in the ANOVA-model. The difference between both parametrisations is illustrated in Tables *5d* - *5e*. For each cell, it is shown which parameters contribute to the expected cell value. The same is done for the marginals. From these tables, it is clear, that the SAS-intercept is estimated from the lastmost (South-East) cell, while in the ANOVA-model we have to take the sum over all cells. It is also indicated, from which cells other parameters are estimated: this follows from the restrictions in either model.

The parametrisation for both models is given in *Appendix* 5, estimation of parameter values in *Appendix* 6, again for Example 4. ANOVA-model effects are departures from the grand mean, μ-model restrictions are deviances from the last level. It is spelled out for all effects for Example 4. The conversion from one parametrisation to the other one is given in *Appendix* 7, in formulas and in the parameters estimates for Example 4. It is shown that the differences between the successive levels w.r.t. the last level are the same for both parametrisations.

Orthogonal Contrast Vectors
Parameter estimation using orthogonal contrast vectors is exemplified in *Appendices* 9 *and* 10. Orthogonal contrasts can be defined for any variable in the analysis, but not *over* variables, i.e., for interaction effects. Therefore, we defined a 'hypervariable', a variable that subsumes all (combinations of) effects. We named the variable 'No', the serial number of the levels of all variables (cf. *Appendix* 8, Exhibits 8.1*d*, 8.1*e*, 8.3*a* - 8.3*c*). More complex contrasts are combinations of contrasts (cf. *Appendix* 8, Exhibits 8.3*c*, 8.3*d*). The analysis of the BAG-data , with three variables and their interactions, is completely described using orthogonal contrast vectors (*Appendix* 10). The same was done for the Goodman (1970) data, the variables Knowledge (good/poor; dependent variable) of certain subjects from Solid/Non-Solid reading, from Newspapers (Y/N), from Lectures (Y/N), or from Radio (Y/N) (see *Appendix* 9).

Output Definition
For the BAG-data, we compared the SAS-GENMOD sums of squares (Type 3 SS) and Wald statistics (that yield Pearson χ^2-squared values) with the WPM-values, they are pretty much the same (see *Appendix* 10). The default with SAS is LR-statistic (not the Pearson χ^2-statistic) and Type 1 SS, instead of the partialised effects (Type 3), needed for orthogonal contrasts.

9. Conclusions

In principle, SAS-GENMOD and WPM do the same job - as far as the analysis of Poisson distributed data in cross-classifications is concerned, but it is difficult to compare the two programs, because

1. difference in methods (estimation procedure, output statistics)
2. difference in parametrisation (μ-model, ANOVA-model)

The presentation of output is also very different. With SAS, the output is extensive, quite clear for the expert, but not always so for the novice. For example, it is not immediately obvious that sums of squares in SAS-GENMOD are sequential sums of squares. With WPM, the minimum-chi-squared method is very quick, easy to use, the output is clear after some oral explanation, and *is* frequently used. However, manuals are not available, and it is not immediately obvious that, in using contrast vectors, results will be so different from those obtained using SAS. WPM-ML is more sophisticated, not so easy to use and lacks a manual.

WPM is a SWOV-program. It has benefits and shortcomings. The differences with respect to SAS-GENMOD concern

- parametrisation (ANOVA-model, μ-model)
- difference in statistics used, *e.g.*, Pearson's χ^2 vs LR statistic
- adequate description of procedures and algorithms
- adding a constant (0.5) in view of the estimation procedure
- differences in estimation procedure
- options available in one program but not in the other one.

In this case, we may conclude that weighted Poisson analysis in cross-classifications can be satisfactorily performed using SAS-GENMOD, as well as using WPM. SAS-GENMOD has more possibilities but is not easy to use. As we have seen, WPM is a special form of Poisson analysis - as is Poisson regression. WPM is not expected to yield the same results as SAS-GENMOD. SAS-GENMOD is a procedure for Poisson regression, for which either sequential SS or partial SS can be used. WPM is a procedure for weighted Poisson analysis in cross-classifications using using partialised SS only. Sequential and partial procedures need not yield the same results (see *Appendix* 10).

10. References

Aitkin, M.A., Anderson, D.A., Francis, B.J., and Hinde, J.P.(1989). *Statistical Modelling in GLIM*. Oxford: Oxford University Press.

Andersen, E.B. (1977). *Multiplicative Poisson models with unequal cell rates*. Scand. J. Statist. **4**.

Andersen, E.B. (1981). *Contingency Tables*. In: Fleischer, G.A. (Ed.). (1981). *Contingency Table Analysis for Road Safety Studies*. Alphen aan den Rijn (The Netherlands): Sijthof & Noordhoff, pp. 3-34.

Agresti, A. (1984). *Analysis of Ordinal Categorical Data*. New York: Wiley.

Agresti, A. (1990, 2nd ed.). *Ordinal Analysis of Categorical Data*. New York: Wiley.

Bishop, Y.M.M, Fienberg, S.E, & Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Mass.: The MIT Press.

Dyke, G.V. and Patterson, H.D. (1952). *Analysis of factorial arrangements when the data are proportions*. Biometrics, **8**, 1-12.

Fienberg, S.E. (1987). *The Analysis of Cross-Classified Data*. Cambridge (Mass.): MIT Press.

Freund, R.J. and Littell, R.J. (1981). *SAS for Linear Models: A Guide to the ANOVA and GLM Procedures*. SAS Series in Statistical Applications, The SAS Institute Inc.: Cary, NC, USA.

Goodman, L.A. (1970). *The multivariate analysis of qualitative data: Interactions among multiple classifications*, JASA, **65**, 226-256.

Hays, W.L. (1988). *Statistics* (4th ed.). London: Holt, Rinehart & Winston.

Leeuw, J. de. (1975). *Maximum Likelihood Estimation for Weighted Poisson Models*, RN005-75, Department of Data Theory, University of Leiden, The Netherlands.

Leeuw, J. de and Oppe, S. (1976). *De Verkeersveiligheid in de provincie Noord-Brabant II, Appendix II.II: Analyse van kruistabellen: log-lineaire Poisson modellen voor gewogen aantallen*. Voorburg: SWOV.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models* (2nd edition). London: Chapman and Hall.

Oppe, S. (1981). *Methods for the Analysis of Contingency Tables in Road Safety Research*. In: Fleischer, G.A. (Ed.). (1981). *Contingency Table Analysis for Road Safety Studies*. Alphen aan den Rijn (The Netherlands): Sijthof & Noordhoff, pp. 3-34.

Oppe, S. (1992). *A comparison of some statistical techniques for road accident analysis*. Accid. Anal. & Prev., **24**, 397-423.

Oppe, S. (1993). *Analysetechnieken voor multivariate analyse van verkeersveiligheidgegevens.* SWOV-Report D-93-11. Leidschendam. SWOV.

SAS/STAT Software. (1993). *The GENMOD Procedure.* SAS Technical Report P243. Release 6.09. SAS Institute Inc.. Cary, NC, USA.
Searle, S.R. (1987). *Linear Models for Unbalanced Data*, New York: Wiley.

Appendices 1- 10

Appendix 1. Datafiles for Examples

Table 1.1: SAS-data for Example 1

Example 1: unweighted data			
1	125	1	1
1	40	1	2
1	165	2	1
1	170	2	2

First column are weights
Second column are counts
Last two columns are design vectors.
Third column gives index for rows
Fourth column gives index for columns

Table 1.2: SAS-data for Example 2

Example 2: simple table to show weighting			
30.0	300	1	1
3.0	30	1	2
0.6	6	2	1
500.0	5000	2	2

First column are weights
Second column are counts
(count/weight is 10 for each cell)
Last two columns are design vectors

Table 1.3*a*: SAS-data for Example 3

Example 3: Example 1 weighted			
1	125	1	1
1	40	1	2
2	165	2	1
2	170	2	2

First column are weights
Last two columns are design vectors

Table 1.3*b*: SAS-data for Example 3

Example 3: Orthogonal contrasts setup		
1	125	1
1	40	2
2	165	3
2	170	4

First column are weights
Third column gives index for cells 'No'
No indices for rows and columns

Table 1.4*a*: SAS-data for Example 4

Example 4: 2x4 Table, Weight = 1			
1	233	1	1
1	67	1	2
1	225	1	3
1	225	1	4
1	125	2	1
1	40	2	2
1	165	2	3
1	170	2	4

Weights are constant (first column)
Last two columns are design vectors

Table 1.4*b*: SAS-data for Example 4

Example 4. Weight = 500

500	233	1	1
500	67	1	2
500	225	1	3
500	225	1	4
500	125	2	1
500	40	2	2
500	165	2	3
500	170	2	4

Weights are constant (first column)
Last two columns are design vectors

This set yields exactly the same results as Table 1.4*a*, apart from the mean, μ (and the intercept)

Table 1.4*c*: SAS-data for Example 4

Example 4: Unweighted

233	1	1
67	1	2
225	1	3
225	1	4
125	2	1
40	2	2
165	2	3
170	2	4

This set yields exactly the same results as Table 1.4*a*

Table 1.5: Oppe's BAG-data weighted: 'BAGw'

No	Weights	Counts	Row	Col	Categ
1	.275	2275	1	1	1
2	.268	339	1	1	2
3	.317	263	1	1	3
4	.372	163	1	1	4
5	.265	448	1	2	1
6	.199	33	1	2	2
7	.229	11	1	2	3
8	.556	10	1	2	4
9	.236	1838	2	1	1
10	.25	350	2	1	2
11	.286	247	2	1	3
12	.291	145	2	1	4
13	.233	452	2	2	1
14	.280	38	2	2	2
15	.273	20	2	2	3
16	.425	9	2	2	4

First column is index for cells: 'No'
Second column are weights
Third column are counts
Last three columns are design vectors

This data (as well as the next) serve to illustrate the use of orthogonal contrast vectors. Orthogonal contrast vectors are defined within a variable. Therefore, we constructed 'No'. 'No' is a design vector indicating the ordering of the 16 cells. Within 'No', we can test (contrast) all kinds of effects. These effects are different combinations of levels within 'No'.

Appendix 1. Datafiles for Examples (continued)

Table 1.6. Goodman's 'Knowledgde of Cancer' Data

No	Freq	Newsp	Lect	Radio	Solid	Knowl
1	23	1	1	1	1	1
2	8	1	1	1	1	2
3	8	1	1	1	2	1
4	4	1	1	1	2	2
5	27	1	1	2	1	1
6	18	1	1	2	1	2
7	7	1	1	2	2	1
8	6	1	1	2	2	2
9	102	1	2	1	1	1
10	67	1	2	1	1	2
11	35	1	2	1	2	1
12	59	1	2	1	2	2
13	201	1	2	2	1	1
14	177	1	2	2	1	2
15	75	1	2	2	2	1
16	156	1	2	2	2	2
17	1	2	1	1	1	1
18	3	2	1	1	1	2
19	4	2	1	1	2	1
20	3	2	1	1	2	2
21	3	2	1	2	1	1
22	8	2	1	2	1	2
23	2	2	1	2	2	1
24	10	2	1	2	2	2
25	16	2	2	1	1	1
26	16	2	2	1	1	2
27	13	2	2	1	2	1
28	50	2	2	1	2	2
29	67	2	2	2	1	1
30	83	2	2	2	1	2
31	84	2	2	2	2	1
32	393	2	2	2	2	2

First columns is index for cells: 'No'
Second column are counts
Remaining columns are design vectors
Each variable has two levels

'No' is a design vector indicating the ordering of the 32 cells.
Within 'No', we can test (contrast) all kinds of effects.
These effects are different combinations of levels within 'No'.

Appendix 2. Example 1: 2x2-Table, Unweighted SAS-Analysis

Table 2*a*. Observed

	B1	B2	Total
A1	125	40	165
A2	165	170	335
Total	290	210	500

Cell values are y_{ij}

Table 2*b*. Logarithms of Observed

	B1	B2	Total
A1	4.828	3.689	8.517
A2	5.106	5.136	10.242
Total	9.934	8.825	18.759

Logarithm is to the base *e*: $\ln(m_{ij})$

Table 2*c*. Expected under 'Independence'

	B1	B2	Total
A1	4.561	4.238	8.799
A2	5.269	4.947	10.216
Total	9.830	9.185	19.015

Independence of row and column effects. Grand mean $\mu = 1/4\ \Sigma\ln(\text{cells}) = 19.016 / 4 = 4.7539$. The intercept is determined from cell (2,2): 4.947; $e^{4.947} = 140.7$ is expected value under independence (cf. Table 2*d*), (see section 2.3)

Table 2*d*: $e^{log(m_{ij})}$: Exponentials of Expected under Independence

	B1	B2	Total
A1	95.7	69.3	165
A2	194.3	140.7	335
Total	290	210	500

Values are exponentials of expected values: $e^{log(mij)}$ or 'Xbeta' in SAS (Xbeta=$\mathbf{X}\beta$).

Appendix 2. Example 1: 2x2-Table, Unweighted (continued)
SAS-Analysis

Table 2*e*. Expected under Row Effects

	B1	B2	Total
A1	4.413	4.413	8.826
A2	5.121	5.121	10.242
Total	9.534	9.534	19.068

Cell values are $log\ m_{ij} = \mu + \alpha_i$
Intercept = 5.121 ~ cell (2,2), (see §2.3)

Difference between row estimates is
4.413 - 5.121 = -.7082: row effect

Table 2*f*. Expected under Column Effects

	B1	B2	Total
A1	4.977	4.654	9.631
A2	4.977	4.654	9.631
Total	9.954	9.308	19.262

Cell values are $log\ m_{ij} = \mu + \beta_j$
Intercept = 4.654 ~ cell(2,2)

Difference between column estimates is
4.977 - 4.654 = .3228: column effect

Table 2*g*. Expected under Row & Column Effects

	B1	B2	Total
A1	4.561	4.238	8.789
A2	5.269	4.947	10.216
Total	9.830	9.185	19.115

Cell values are $log\ m_{ij} = \mu + \alpha_i + \beta_j$
Compare with model of Independence
Intercept = 4.947 ~ cell (2,2)

Table 2*h*. Expected under 'Intercept Only'

	B1	B2	Total
A1	4.828	4.828	9.656
A2	4.828	4.828	9.656
Total	9.656	9.656	19.312

Intercept Only: only grand mean effect
Cell values are $(\Sigma log\ m_{ij})/IJ$ = (log 500)/4
= 4.828 = μ

Appendix 3. Example 2: 2×2 Table, Weighted SAS-Analysis

Table 3*a*: *Observed*

	B1	B2	Tot
A1	300	30	330
A2	6	5000	5006
Tot	306	5030	5336

Data only

Table 3*b*: *Weights*

	B1	B2
A1	30	3
A2	0.6	500

Weights for data

Table 3*c*: *Weighted Data (Obs / Weight)*

	B1	B2	Tot
A1	10	10	20
A2	10	10	20
Tot	20	20	40

Cells values: obs / weight
Data are *downweighted* by weights

Table 3*d*: *Independence: Expected*

	B1	B2	Tot
A1	2.303	2.303	4.606
A2	2.303	2.303	4.606
Tot	4.606	4.606	9.212

Intercept: 2.303 [= ln(10)]

Table 3*e*: *Saturated Model: Expected*

	B1	B2	Tot
A1	2.303	2.303	4.606
A2	2.303	2.303	4.606
Tot	4.606	4.606	9.212

After weighting, equal expected values
~ Model of Independence
Row Effect: 0
Column Effect: 0
Interaction: 0

Appendix 4. Example 3 (= Example 1, Weighted)
SAS-Analysis

Table 4*a*. = **Table 1***a*: Original Data

	B1 w=1	B2 w=1	Total
A1(w=1)	125	40	165
A2(w=2)	165	170	335
Total	290	210	500

Cell values are x_{ij} (counts)

Weights are 1 for columns (w_j= 1, j = 1, 2)
Weights are 1, 2 for rows (w_{1j} = 1, w_{2j} = 2)

Table 4*b*. Weights for Intercept Only Model

	B1 w=1	B2 w=1	Total
A1(w=1)	1/6	1/6	2/6
A2(w=2)	2/6	2/6	4/6
Total	3/6	3/6	1

Cell values are w_{ij} / Σw_{ij}, e.g.,
Cell (2,1): w_{21} =2; Σw_{ij}= 6.
Weights are equal for columns
Weights are proportional for rows

Table 4*c*. Predicted Intercept Only Model

	B1 w=1	B2 w=1	Total
A1(w=1)	83.333	83.333	166.667
A2(w=2)	166.667	166.667	333.333
Total	250	250	N = 500

Predicted values are w_{ij} / $\Sigma w_{ij} \times N$

Predicted is proportional for rows
Predicted is constant for columns

SAS: Predicted × w_{ij}^{-1} = $\mathbf{X}\beta$ ***(expected)***

Prediction cell (2,1): 1/3 × 500 = 166.67
Expectation cell (2,1): 1/2×166.67= 83.33
Pred. cell (1,2): 1/6×500 = 83.33; Exp.:83.33

Table 4*d*. Expected (XBETA) for Intercept Only Model in counts (upper entry) and logs (lower entry)

	B1 w=1	B2 w=1	Total
A1(w=1)	83.333	83.333	166.667
	4.423	4.423	8.846
A2(w=2)	83.333	83.333	166.667
	4.423	4.423	8.846
Total	166.667	166.667	333.333
	8.846	8.846	17.692

Expected = Predicted corrected for weight
$\mathbf{X}\beta$ = Predicted Values × w_{ij}^{-1}

Expected is equal for rows
Expected is equal for columns
Dividing Predicted by w_{ij}^{-1} gives constant expected values for Interc. Only Model

Without weighting, expected = 125 (for each cell), in logs: 4.828

Appendix 4. Example 3 (= Example 1, Weighted) (continued)

Column effects model: column effects only. Rows are proportional to weights (1/3; 2/3). Row totals are 166.67, resp. 333.33. Predicted value for cell (2,1) is 2× 333.33× 290 / 500, i.e., weight × row total × column total /*N*. Exponentials of cell entries are given; cf.Table 4*a*
Row effects model: analogous.

Table 4*e*. Predicted values Column Model

	B1 w=1	B2 w=1	Total
A1(w=1)	96.67	70.00	166.67
A2(w=2)	193.33	140.00	333.33
Total	290	210	N =500

Column effect ~ margin
Row effect ~ $w_{i.} \Rightarrow N \times w_{i.} / \Sigma w_{i.}$
Row 1: 500 × $w_{1.}$ × 1/3
Row 2: 500 × $w_{2.}$ × 1/3
Columns equally weighted: w_j =1

Predicted = $w_{i.}$ × rowtot × coltot / N

Table 4*f*. Expected values Column Model

	B1 w=1	B2 w=1	Total
A1(w=1)	96.667	70.000	166.667
A2(w=2)	96.667	70.000	166.667
Total	193.333	140.000	333.333

Expected = Predicted /weight
$\mathbf{X}\beta = w_{ij}^{-1} \times$ Predicted Values

Weights are 1 for columns
Weights are 1, 2 for rows

Table 4*g*. Predicted values Row Model

	B1 w=1	B2 w=1	Total
A1(w=1)	82.5	82.5	165
A2(w=2)	167.5	167.5	335
Total	250	250	500

Row effect ~ margin
Column effect ~ $w_{.j} \Rightarrow N \times w_{.j} / \Sigma w_{.j}$
Column 1: 500 × $w_{.1}$ × 1/2 = 250
Column 2: 500 × $w_{.2}$ × 1/2 = 250
Rows not equally weighted: $w_{i.}$ = 1,2

Predicted = w_{ij} × rowtot × coltot / N

Table 4*h*. Expected values Row Model

	B1 w=1	B2 w=1	Total
A1(w=1)	82.517	82.517	165.033
A2(w=2)	83.764	83.764	167.527
Total	166.281	166.281	332.560

Corrected for weights

Expected = Predicted / weight
$\mathbf{X}\beta = w_{ij}^{-1} \times$ Predicted Values,

Appendix 5. Example 4: 2×4 Table Unweighted

Table 5a: *Observed*

	B1	B2	B3	B4	Total
A1	233	67	225	225	750
A2	125	40	165	170	500
Tot	358	107	390	395	1250

Table 5b: *Expected under Independence (anti-logs)*

	B1	B2	B3	B4	Total
A1	214.8	64.2	234	237	750
A2	143.2	42.8	156	158	500
Tot	358	107	390	395	1250

Table 5c: *Expected under Independence (logs)*

	B1	B2	B3	B4	Total
A1	5.370	4.162	5.455	5.468	20.450
A2	4.964	3.757	5.050	5.063	18.830
Tot	10.334	7.919	10.505	10.531	39.290

Table 5d: *Expected under Independence (ANOVA-model parameterss)*

	B1	B2	B3	B4	Total
A1	$\mu+\alpha_1+\beta_1$	$\mu+\alpha_1+\beta_2$	$\mu+\alpha_1+\beta_3$	$\mu+\alpha_1+\beta_4$	$4\mu+4\alpha_1+(\beta_1+\beta_2+\beta_3+\beta_4)=4\mu+4\alpha_1$
A2	$\mu+\alpha_2+\beta_1$	$\mu+\alpha_2+\beta_2$	$\mu+\alpha_2+\beta_3$	$\mu+\alpha_2+\beta_4$	$4\mu+4\alpha_2+(\beta_1+\beta_2+\beta_3+\beta_4)=4\mu+4\alpha_2$
Tot	$2\mu+2\beta_1$	$2\mu+2\beta_2$	$2\mu+2\beta_3$	$2\mu+2\beta_4$	$8\mu+4(\alpha_1+\alpha_2)+2(\beta_1+\beta_2+\beta_3+\beta_4)=8\mu$

Table 5e: *Expected under Independence (μ-model parameters)*

	B1	B2	B3	B4	Total
A1	$\mu+\alpha_{1'}+\beta_{1'}$	$\mu+\alpha_{1'}+\beta_{2'}$	$\mu+\alpha_{1'}+\beta_{3'}$	$\mu+\alpha_{1'}$	$4\mu+4\alpha_{1'}+\beta_{1'}+\beta_{2'}+\beta_{3'}$
A2	$\mu+\beta_{1'}$	$\mu+\beta_{2'}$	$\mu+\beta_{3'}$	μ	$4\mu+\beta_{1'}+\beta_{2'}+\beta_{3'}$
Tot	$2\mu+\alpha_{1'}+2\beta_{1'}$	$2\mu+\alpha_{1'}+2\beta_{2'}$	$2\mu+\alpha_{1'}+2\beta_{3'}$	$2\mu+\alpha_{1'}$	$8\mu+4\alpha_{1'}+2(\beta_{1'}+\beta_{2'}+\beta_{3'})$

For the ANOVA-model, $\Sigma\alpha_i=\Sigma\beta_j=0$ and intercept is marginal total/IJ = $8\mu/8 = \mu$. For the μ-model, the intercept is estimated from cell (2,4). For this cell, parameters are μ, $\alpha_{2'}$ and $\beta_{4'}$, and since $\alpha_{2'} = \beta_{4'} = 0$, cell (2,4) gives the intercept. Next, $\alpha_{1'}$ is estimated from cell (1,4), $\beta_{1'}$ from cell (2,1), etc.

Appendix 6. Example 4: Estimates in μ- and ANOVA-Model

ANOVA-model effects: **deviances with respect to the grand mean, μ**

Grand Mean
$\mu = 1/8\ \Sigma^2{}_{i=1}\ \Sigma^4{}_{j=1} \log m_{ij} = 1/8 \times 39.29 = 4.911$

Rows
$\mu + \alpha_1 = 1/4\ \Sigma^4{}_{j=1} \log m_{1j} = 1/4 \times 20.46 = 5.115$ $\alpha_1 = 5.1150 - 4.911 = +\ 0.2028$
$\mu + \alpha_2 = 1/4\ \Sigma^4{}_{j=1} \log m_{2j} = 1/4 \times 18.83 = 4.708$ $\alpha_2 = 4.7075 - 4.911 = -\ 0.2028$

Columns
$\mu + \beta_1 = 1/2\ \Sigma^2{}_{i=1} \log m_{i1} = 1/2 \times 10.330 = 5.165$ $\beta_1 = 5.1650 - 4.911 = +\ 0.256$
$\mu + \beta_2 = 1/2\ \Sigma^2{}_{i=1} \log m_{i2} = 1/2 \times 7.920 = 3.960$ $\beta_2 = 3.9600 - 4.911 = -\ 0.952$
$\mu + \beta_3 = 1/2\ \Sigma^2{}_{i=1} \log m_{i3} = 1/2 \times 10.505 = 5.253$ $\beta_3 = 5.2525 - 4.911 = +\ 0.342$
$\mu + \beta_4 = 1/2\ \Sigma^2{}_{i=1} \log m_{i4} = 1/2 \times 10.531 = 5.265$ $\beta_4 = 5.2655 - 4.911 = +\ 0.354$

μ-model effects: **deviances with respect to the last level:**
(μ-model estimates are given with a prime, e.g., $\alpha_{1'}$)

Rows
$\alpha_2 = -\ 0.2028$
$\alpha_{1'} = \alpha_1 - \alpha_2 = 0.2028 - (-\ 0.2028) = +\ 0.4056$ (Row 1)
$\alpha_{2'} = \alpha_2 - \alpha_2 = -\ 0.2028 - (-0.2028) = 0$

Columns
$\beta_4 = +\ 0.354$
$\beta_{1'} = \beta_1 - \beta_4 = +\ 0.256 - 0.354 = -\ 0.098$ (Column 1)
$\beta_{2'} = \beta_2 - \beta_4 = -\ 0.952 - 0.354 = -\ 1.306$
$\beta_{3'} = \beta_3 - \beta_4 = +\ 0.342 - 0.354 = -\ 0.012$
$\beta_{4'} = \beta_4 - \beta_4 = 0.$

Using SAS, we find the same values (apart from rounding errors):

$\alpha_{1'} = +\ 0.4055$ (Row 1)
$\alpha_{2'} = 0;$
$\beta_{1'} = -\ 0.0984$ (Column 1)
$\beta_{2'} = -\ 1.3061$
$\beta_{3'} = -\ 0.0127$
$\beta_{4'} = 0.$

Intercepts for Poisson Regression Model:

Intercept Only (Mean):	4.911
Mean + Row Effects	4.911 - 0.2028 = 4.7082
Mean + Column Effects:	4.911 + 0.354 = 5.265
Mean + Row and Column Effects	4.911 - 0.2028 + 0.354 = 5.0622

For all models, the intercept is determined from the lastmost cell: cell (2,4).

Appendix 7. Conversion from μ-Model to ANOVA-Model

To calculate ANOVA-model parameters for one specific effect when μ-model parameters are given, proceed as follows (the proof is given below):

-1- Determine the sum of parameter estimates in the μ-model;
-2- Divide this sum by the number of levels;
-3- Change sign;
-4- Add this number to all μ-model estimates.

This procedure will be applied to Example 4, the 2×4 Table, first to rows, then to columns.

Rows
Step 1: SUM μ-model estimates: .4056 + 0 = .4056.
Step 2: DIVIDE by 2: .4056/2 = .2028.
Step 3: CHANGE sign: - .2028.
Step 4: ADD (-.2028) to all levels:
Row 1: .4056 - .2028 = .2028.
Row 2: 0 - .2028 = -.2028.

Columns
Step 1: SUM = - .098 - 1.306 - .012 = - 1.416.
Step 2: DIVIDE by 4: - 1.416/4 = - .354.
Step 3: CHANGE sign: + .354.
Step 4: ADD (.354) to all levels:
Column 1: -.098 + .354 = .256;
Column 2: -1.306 + .354 = -.952;
Column 3: -.012 + .354 = .342;
Column 4: 0 + .354 = .354, the ANOVA-model estimates we started from.

To calculate μ-model estimates when ANOVA-model estimates are given, proceed as follows:

-1- Determine the parameter estimate for the last level;
-2-Subtract this number from all parameter estimates.

This procedure will be applied to the above data, first to the rows, then to columns.

Rows
Step 1: DETERMINE last level estimate: -.2028;
Step 2: SUBTRACT (-.2028) from all parameter estimates:
Row 1: .2028 - (-.2028)= + .4056; Row 2: -.2028 -(-.2028) = 0.

Columns
Step 1:DETERMINE last level estimate: .354;
Step 2: SUBTRACT (.354) from all parameter estimates:
Column 1: .256 - (.354) = - 0.098;
Column 2: -.952 - (.354) = - 1.306;
Column 3: .342 - (.354) = - 0 .012
Column 4: .354 - (.354) = 0.

(Note that the difference between the successive levels w.r.t. the last level are the same for the μ-model and the ANOVA-model).

Appendix 7. Conversion from μ-Model to ANOVA-Model

(continued)

From ANOVA-model to μ-model and vice versa:

Let the parameters for a specific effect in the μ-model be given by

μ-MODEL:	$\beta_1, \beta_2, ..., \beta_{J-1}, \beta_J$	for which $\beta_J = 0$,

and let the parameters in the ANOVA-model be given by

ANOVA-MODEL:	$\gamma_1, \gamma_2, ..., \gamma_J$,	for which $\Sigma_j\, p_j\, \gamma_j = 0$.

Then it holds that 1) $\gamma_j - \gamma_J = \beta_j - \beta_J = \beta_j$ and that
2) $\Sigma_j\, p_j\, \gamma_j = 0$, from which

$$\Sigma_j\, p_j\, (\gamma_j - \gamma_J) = \Sigma_j\, \beta_j\, p_j \quad , \quad \text{or}$$

$$0 - J \times \gamma_J = \Sigma_j\, \beta_j\, p_j \quad , \quad \text{so that}$$

$$\gamma_J = -1/J\, \Sigma_j \beta_j\, p_j \quad .$$

WPM-Estimates for Examples 1 and 4:

Using WPM-ML, we find the following estimates for Example 4, the 2×4 Table:
Rows: 0.2161, -0.2161
Columns: 0.2338, -0.9591, 0.3552, 0.3701.
Using SAS we found, after transformation to ANOVA parameterization:
Rows: 0.2028, -0.2028
Columns: 0.256, - 0.952, 0.342, + 0.354.

For Example 1, the 2×2 Table, we find using WPM-ML:
Rows: -0.4311, 0.4311;
Columns: 0.2774, -0.2774.
The μ-model estimates, given by SAS, are:
Rows: -0.7082, 0;
Columns: 0.3228, 0.
We find the ANOVA-parameters using the transformation rule:
Rows: -0.3541, 0.3541;
Columns: 0.1614, -0.1614,

The WPM-estimates are slightly different because 0.5 is added to each observation.

Appendix 8. SAS-GENMOD Setups for Examples

Exhibit 8.1*a*: Creating SAS-File for Example 1	Comment
libname xx '[own dir]';	generates SAS-dataset xx.Ex1
filename invoer '[own.dir]Ex1';	in directory van owner ('own')
data xx.Ex1;	libname xx
infile invoer;	
input n c A B;	*n* is weighting variable
ln = log(n);	ln is offset variable
proc contents;	*c* (count) is dependent variable

Exhibit 8.1*b*: Poisson Regression for Example 1	Comment
options pagesize=59 linesize=80 nocenter;	Alternatives for Model statement ★
libname xx '[own.dir]';	(1) model c= /dist=poisson etc.
	⇒generates Intercept Only Model;
proc genmod data=xx.Ex1 order=data;	(2) model c= A /dist=etc.
class A B;	⇒generates Row Effects Model;
model c= B /dist=poisson ★	(3) model c= A B /dist=etc.
link = log ;	⇒generates Row+Column Model.
run;	

Exhibit 8.1*c*: Type 1 and Type 3 Analysis	Comment
options pagesize=59 linesize=80 nocenter;	
libname xx '[own.dir]';	
proc genmod data=xx.Ex1; order=data;	(1) Same SAS-data set as above
class A B;	
make 'obstats' out = outdata;	(2) Extensive output preparation
model c = /dist = poisson ★	(3) Intercept model as above (★)
link = log	
type1	Sequential sum of squares
type3	Partialising effects
obstats	Produces extensive output
run;	

Exhibit 8.1*d*: Type 3 Analysis and Wald Statistics: Example 3	**Comment**
options pagesize=59 linesize=80 nocenter;	
libname xx '[own.dir]';	
proc genmod data=xx.ex3 order=data,	Example 3, weighted data
class A B ;	
model c= A B A*B /dist=poisson ★	model includes interactions
link = log offset = ln	downweighting by 'ln'
type1 type3 ;	Type 1 Analysis: sequential,
contrast 'Bb' B 1 -1 / E wald;	Wald statistics ⇒ Pearson χ^2
contrast 'Aa' A 1 -1 / E wald ;	instead of Likelihood Ratio G^2
◆contrast 'inter A*B' .5 -.5 -.5 .5 /E WALD;	◆Imposible to define interaction with class-structure present
run;	

Exhibit 8.1*e*: Orthogonal Contrasts for Example 3	**Comment**
options pagesize=59 linesize=80 nocenter;	
libname xx '[own.dir]';	
proc genmod data= xx.ex3b order=data;	◆Contrast vectors defined on
◆ class No;	serial numbers of categories,
model c = A B A*B /dist=poisson	to obtain interaction contrasts
link = log offset = ln	
type1 type3 ;	Type 3 Analysis: LR ratio stat.
contrast 'A' A 1 -1;	Contrast: main effect 'A'
contrast 'B' B 1 -1;	Contrast: main effect 'B'
contrast 'inter' A*B 0.5 -0.5 -0.5 0.5;	Contrast: interaction

Exhibit 8.2*a*: Orthogonal Contrasts for Example 4	**Comment**
options pagesize=59 linesize=80 nocenter;	
libname xx '[own.dir]';	
proc genmod data=xx.Ex4 order=data;	No design matrix, only serial
class A B;	numbers
make 'obstats' out = outdata;	
model c= /dist=poisson ✱	Same models as above
link = log offset = ln	
type1	Type 3 analysis; Pearsons' χ^2
type3	log-likelihood ratio statistic G^2
contrast 'B' B 1 -1;	Contrast between categories,
contrast 'A1' A 3 -1 -1 -1;	*within* variables
contrast 'A2' A 0 2 -1 -1;	
contrast 'A3' A 0 0 1 -1;	Same results as Type 3 /E WALD

These contrasts yield same results as Type 3 analysis using Wald statistics and as the SWOV-program WPM

Exhibit 8.3*a*: Creating SAS-file for BAG-data

```
libname xx    '[own.dir]';
filename invoer '[own.dir]BAGw';
data xx.BAGw;
    infile invoer;
    input No Weight Freq Year Sexe BAG ;
    ln = log(Weight) ;
length  No ln Freq Year Sexe BAG 3.;
proc contents;
run;
```

Comment

BAGw: data include weights
SAS-data set (weighted)

'No' is serial number for cells
Variables are Freq, Year, Sexe, and BAG; 'ln' is offset variable, for downweighting Freq;
'No' is needed for interaction contrasts.

Exhibit 8.3*b*: Sequential Analysis BAG-data

```
options  pagesize=59 linesize=80 nocenter;
libname  xx   '[own.dir]';
proc  genmod  data=xx.BAGw ;
  class  No ;
  model  Freq=No / dist = poisson
  link = log
  offset = ln;  run;
```

Comment

Sequential analysis,
No contrasts specified.

Exhibit 8.3*c*: Type 3 Analysis + Contrasts, BAG-data

```
options pagesize=59 linesize=80 nocenter
libname  xx       '[own.dir]';
proc  genmod    data=xx.BAGw;
  class  Year  Sexe  BAG ;
model Freq = Year|Sexe|BAG  /dist = poisson
  link = log  offset = ln   type3 ;

contrast 'BAG1 1 VS 2,3,4'   BAG  -3  1  1  1 / E ;
contrast 'BAG2 2 VS 3,4'     BAG   0 -2  1  1 / E ;
contrast 'BAG3 3 VS 4'       BAG   0  0 -1  1 / E ;
contrast 'BAG1 to BAG3'•     BAG -3 1 1 1, BAG 0 -2 1 1, BAG 0 0 -1 1 /E;  •Combination

contrast 'BAG4 1,2 VS 3,4'   BAG  -1 -1  1  1 / E ;
contrast 'BAG5 1,2'          BAG  -1  1  0  0 / E ;
contrast 'BAG6 3,4'          BAG   0  0 -1  1 / E ;
contrast 'BAG4 to BAG6'•     BAG -1 -1 1 1, BAG -1 1 0 0, BAG 0 0 -1 1 / E; •Combin.
run;
```

Comment

Illustration of forming interaction contrasts and combining them into complex contrast statements
All possible effects.

Exhibit 8.3*d*: Contrasts as in WPM for BAG data

Comment: Complete orthogonal contrasts analysis of BAG-data.

```
(...)
proc genmod   data=xx.BAGw ;
  class No ;
  model Freq = No   / dist = poisson
  link = log   offset = ln   type3 ;

contrast 'Year 1975 vs 1977'     No 1 1 1 1 1 1 1 1 -1 -1 -1 -1 -1 -1 -1 -1 / E Wald;

contrast 'SEXE'                  No 1 1 1 1 -1 -1 -1 -1 1 1 1 1 -1 -1 -1 -1 / E Wald;

contrast 'BAG1 1 vs 2,3,4'       No 3 -1 -1 -1 3 -1 -1 -1 3 -1 -1 -1 3 -1 -1 -1  / E  Wald;
contrast 'BAG2 2 vs 3,4'         No 0 2 -1 -1 0 2 -1 -1 0 2 -1 -1 0 2 -1 -1 /E Wald;
contrast 'BAG3 3 vs 4'           No 0 0 1 -1 0 0 1 -1 0 0 1 -1 0 0 1 -1 /E Wald;

contrast 'BAG1-4'•               No 3 -1 -1 -1 3 -1 -1 -1 3 -1 -1 -1 3 -1 -1 -1,
                                 No 0 2 -1 -1 0 2 -1 -1 0 2 -1 -1 0 2 -1 -1,
                                 No 0 0 1 -1 0 0 1 -1 0 0 1 -1 0 0 1 -1/ E Wald; •Combin.

contrast 'Year*Sexe'             No 1 1 1 1 -1 -1 -1 -1 -1 -1 -1 -1 1 1 1 1 /E Wald;

contrast 'Sexe*BAG1'             No 3 -1 -1 -1 -3 1 1 1 3 -1 -1 -1 -3 1 1 1 / E Wald;
contrast 'Sexe*BAG2'             No 0 2 -1 -1 0 -2 1 1 0 2 -1 -1 0 -2 1 1 / E Wald;
contrast 'Sexe*BAG3'             No 0 0 1 -1 0 0 -1 1 0 0 1 -1 0 0 -1 1 /E Wald;

contrast 'Year*BAG1'             No 3 -1 -1 -1 3 -1 -1 -1 -3 1 1 1 -3 1 1 1 / E Wald;
contrast 'Year*BAG2'             No 0 2 -1 -1 0 2 -1 -1 0 -2 1 1 0 -2 1 1 / E Wald;
contrast 'Year*BAG3'             No 0 0 1 -1 0 0 1 -1 0 0 -1 1 0 0 -1 1 / E Wald;

contrast 'Year*BAG1-3' •         No 3 -1 -1 -1 3 -1 -1 -1 -3 1 1 1 -3 1 1 1 ,
                                 No 0 2 -1 -1 0 2 -1 -1 0 -2 1 1 0 -2 1 1,
                                 No 0 0 1 -1 0 0 1 -1 0 0 -1 1 0 0 -1 1 /E Wald;  •Combin.

contrast 'Sexe*BAG1-3' •         No 3 -1 -1 -1 -3 1 1 1 3 -1 -1 -1 -3 1 1 1 ,
                                 No 0 2 -1 -1 0 -2 1 1 0 2 -1 -1 0 -2 1 1 ,
                                 No 0 0 1 -1 0 0 -1 1 0 0 1 -1 0 0 -1 1 /E Wald;  •Combin.

contrast 'Year*Sexe*BAG1'        No 3 -1 -1 -1 -3 1 1 1 -3 1 1 1 3 -1 -1 -1/E Wald;
contrast 'Year*Sexe*BAG2'        No 0 2 -1 -1 0 -2 1 1 0 -2 1 1 0 2 -1 -1 /E Wald;
contrast 'Year*Sexe*BAG3'        No 0 0 1 -1 0 0 -1 1 0 0 -1 1 0 0 1 -1 / E Wald;

contrast 'Year*Sexe*BAG1-3'     •No 3 -1 -1 -1 -3 1 1 1 -3 1 1 1 3 -1 -1 -1 ,
                                 No 0 2 -1 -1 0 -2 1 1 0 -2 1 1 0 2 -1 -1 ,
                                 No 0 0 1 -1 0 0 -1 1 0 0 -1 1 0 0 1 -1 /E Wald;  •Combin.
run;
```

Appendix 9. Goodman's (1970) Data: Contrast Vectors

Ex4ibit 9. Orthogonal Contrasts for Goodman data

Comment

```
(....)
proc   genmod data=xx Goodhalf;
class No;  model Freq = No
         / dist = poisson   link = log
           offset = ln    obstats   type3 ;

contrast 'Newsp1 VS Newsp0'
  No 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 -1 -1 -1 -1 -1 -1 -1 -1  -1 -1 -1 -1 -1 -1 -1 -1 / E Wald;
contrast 'Lect1 VS Lect0'
  No 1 1 1 1 1 1 1 1 -1 -1 -1 -1 -1 -1 -1 -1 1 1 1 1 1 1 1 1 -1 -1 -1 -1 -1 -1 -1 -1 / E Wald;
contrast 'Radio1 VS Radio0'
  No 1 1 1 1 -1 -1 -1 -1 1 1 1 1 -1 -1 -1 -1 1 1 1 1 -1 -1 -1 -1 1 1 1 1 -1 -1 -1 -1 / E Wald;
contrast 'Solid1 VS Solid0'
  No 1 1 -1 -1 1 1 -1 -1 1 1 -1 -1 1 1 -1 -1 1 1 -1 -1 1 1 -1 -1  1 1 -1 -1 1 1 -1 -1 / E Wald;
contrast 'Knowl1 VS Knowl0'
  No 1 -1 1 -1 1 -1 1 -1 1 -1 1 -1 1 -1 1 -1 1 -1 1 -1 1 -1 1 -1  1 -1 1 -1 1 -1 1 -1 / E Wald;
contrast 'Knowl*Solid'
  No 1 -1 -1 1 1 -1 -1 1 1 -1 -1 1 1 -1 -1 1 1 -1 -1 1 1 -1 -1 1  1 -1 -1 1 1 -1 -1 1 / E WAld;
run;
```

Data is here $10(m_{ij} + 0.5)$,
since GENMOD doesn't accept +0.5

offset⇒dividing cells by ln(10)

χ^2 and z values for Goodman data with SAS and WPM
'T3' means Type 3 analysis in SAS; 'W'(Wald) yields Pearson's χ^2
'OC' means orthogonal contrast vectors, adding 0.5, offset: ln(10)

Source of Interaction	χ^2	z	df	p-value
Newsp1 vs Newsp0				
WPM	48.2321	6.9445	1	0.000
SAS-OC/T3/W	48.2321	±6.9445	1	0.000
Goodman	48.2321	6.9445	1	0.000
Lect1 vs Lect0				
WPM	389.4305	-19.7340	1	0.000
SAS-OC/T3/W	389.4305	±19.7340	1	0.000
Goodman	389.4305	-19.7340	1	0.000
Radio1 vs Radio0				
WPM	52.2795	-7.2305	1	0.000
SAS-OC/T3/W	52.2795	±7.2305	1	0.000
Goodman	52.2795	-7.2305	1	0.000
Solid1 vs Solid0				
WPM	1.9194	1.3854	1	0.000
SAS-OC/T3/W	1.9194	±1.3854	1	0.000
Goodman	1.9194	1.3854	1	0.000
Knowl1 vs Knowl0				
WPM	5.6269	-2.3726	1	0.000
SAS-OC/T3/W	5.6269	±2.3726	1	0.000
Goodman	5.6269	-2.3726	1	0.000
Knowl*Solid				
WPM	5.6824	2.3838	1	0.000
SAS-OC/T3/W	5.6824	±2.3838	1	0.000
Goodman	5.6824	2.3838	1	0.000

Analyses. Goodman (1970); WPM: ML; Orthogonal contrast vectors; adds 0.5 to each cell; uses Pearson's χ^2 statistic.
SAS-PROC GENMOD: Type 3 analysis (partialized effects) Orthogonal contrast vectors, Wald: Pearson's χ^2.

Appendix 10. Comparison of WPM & SAS: BAG-Data

Exhibit 10*a*: Contrasts as in WPM for BAG-data	**Comment**
(.....)	Complete orthogonal contrasts analysis of BAG-data

```
(.....)
proc genmod    data=xx.BAGw ;
  class No ;
  model Freq = No   / dist = poisson
  link = log   offset = ln   type3 ;

contrast 'Year 1975 vs 1977'      No 1 1 1 1 1 1 1 1 -1 -1 -1 -1 -1 -1 -1 -1 / E Wald;

contrast 'SEXE'                   No 1 1 1 1 -1 -1 -1 -1 1 1 1 1 -1 -1 -1 -1 / E Wald;

contrast 'BAG1 1 vs 2,3,4'        No 3 -1 -1 -1 3 -1 -1 -1 3 -1 -1 -1 3 -1 -1 -1 / E Wald;
contrast 'BAG2 2 vs 3,4'          No 0 2 -1 -1 0 2 -1 -1 0 2 -1 -1 0 2 -1 -1 /E Wald;
contrast 'BAG3 3 vs 4'            No 0 0 1 -1 0 0 1 -1 0 0 1 -1 0 0 1 -1 /E Wald;

contrast 'BAG1-4' •               No 3 -1 -1 -1 3 -1 -1 -1 3 -1 -1 -1 3 -1 -1 -1,
                                  No 0 2 -1 -1 0 2 -1 -1 0 2 -1 -1 0 2 -1 -1,
                                  No 0 0 1 -1 0 0 1 -1 0 0 1 -1 0 0 1 -1/ E Wald;  •Combin.

contrast 'Year*Sexe'              No 1 1 1 1 -1 -1 -1 -1 -1 -1 -1 -1 1 1 1 1 /E Wald;

contrast 'Sexe*BAG1'              No 3 -1 -1 -1 -3 1 1 1 3 -1 -1 -1 -3 1 1 1 / E Wald;
contrast 'Sexe*BAG2'              No 0 2 -1 -1 0 -2 1 1 0 2 -1 -1 0 -2 1 1 / E Wald;
contrast 'Sexe*BAG3'              No 0 0 1 -1 0 0 -1 1 0 0 1 -1 0 0 -1 1 /E Wald;

contrast 'Year*BAG1'              No 3 -1 -1 -1 3 -1 -1 -1 -3 1 1 1 -3 1 1 1 / E Wald;
contrast 'Year*BAG2'              No 0 2 -1 -1 0 2 -1 -1 0 -2 1 1 0 -2 1 1 / E Wald;
contrast 'Year*BAG3'              No 0 0 1 -1 0 0 1 -1 0 0 -1 1 0 0 -1 1 / E Wald;

contrast 'Year*BAG1-3' •          No 3 -1 -1 -1 3 -1 -1 -1 -3 1 1 1 -3 1 1 1 ,
                                  No 0 2 -1 -1 0 2 -1 -1 0 -2 1 1 0 -2 1 1,
                                  No 0 0 1 -1 0 0 1 -1 0 0 -1 1 0 0 -1 1 /E Wald;  •Combin.

contrast 'Sexe*BAG1-3' •          No 3 -1 -1 -1 -3 1 1 1 3 -1 -1 -1 -3 1 1 1 ,
                                  No 0 2 -1 -1 0 -2 1 1 0 2 -1 -1 0 -2 1 1 ,
                                  No 0 0 1 -1 0 0 -1 1 0 0 1 -1 0 0 -1 1 /E Wald;  •Combin.

contrast 'Year*Sexe*BAG1'         No 3 -1 -1 -1 -3 1 1 1 -3 1 1 1 3 -1 -1 -1/E Wald;
contrast 'Year*Sexe*BAG2'         No 0 2 -1 -1 0 -2 1 1 0 -2 1 1 0 2 -1 -1 /E Wald;
contrast 'Year*Sexe*BAG3'         No 0 0 1 -1 0 0 -1 1 0 0 -1 1 0 0 1 -1 / E Wald;


contrast 'Year*Sexe*BAG1-3'       • No 3 -1 -1 -1 -3 1 1 1 -3 1 1 1 3 -1 -1 -1 ,
                                    No 0 2 -1 -1 0 -2 1 1 0 -2 1 1 0 2 -1 -1 ,
                                    No 0 0 1 -1 0 0 -1 1 0 0 -1 1 0 0 1 -1 /E Wald;  •Combin.
run;
```

Appendix 10. Comparison of WPM & SAS: BAG-Data
(continued)

Interaction Sum of Squares: WPM vs SAS

SAS-GENMOD Procedure with
- TYPE 3 instead of TYPE 1 analysis (default);
- Orthogonal Contrast Vectors;
- Pearson's χ^2 instead of LR-statistic (default),

DATA: BAG-data Blood-Alcohol-Level data '75 - '77
S. OPPE, 1993, SWOV: D-93-11

Classification Variables:
- YEAR (1975, 1977);
- SEX (m, f);
- BAG (4 classes)

FREQ (per cell): Response variable
WEIGHTS: the data are corrected for exposition

Interaction SS for the BAG-data using SAS and WPM
'T3' means Type 3 analysis in SAS, 'W': (WAld) yields Pearson's χ^2

Source of Interaction	χ^2	df	p-value
Year×Sex			
WPM	0.21	1	
SAS-T3/W	0.23	1	0.633
Year×BAG			
WPM	1.70	3	
SAS-T3/W	1.72	3	0.633
Sex×BAG			
WPM	102.71	3	
SAS-T3/W	104.94	3	0.000
Year×Sex×BAG			
WPM	4.04	3	
SAS-T3/W	4.02	3	0.259

Analyses: WPM: Minimum Chi-Squared Method (Oppe, 1993);
Orthonormal contrast vectors; interactions only;
adds 0.5 to each cell; uses Pearson's χ^2-statistic.

SAS-GENMOD: Type 3 analysis (partialized effects)
Contrast Vectors, Wald: Pearson's χ^2-statistic

Appendix 10. Comparison of WPM & SAS: BAG-Data (continued)

SAS-GENMOD	T3 (=Type 3) vs T1 (=Type 1) analysis ; LR: Likelihood-Ratio χ^2 statistic; WALD (W): Pearson's χ^2 test statistic, Overall test (all variables in analysis);
WPM	Program for Weighted Poisson Analysis; Uses Pearson's χ^2 test statistic and design matrix: orthonormal contrast vectors
BAG-data	Blood-Alcohol-Level data '75-'79 (Oppe, 1993).

Comparison of Results for BAG-data using SAS and WPM

Source of Interaction	χ^2	df	p-value	SS	Test Statistic
Year × Sex					
WPM	0.21	1			
SAS-T3/W	0.228	1	0.633	Type 3	Wald
SAS-T3/LR	0.228	1	0.633	Type 3	LR
SAS-T1/LR	1.439	1	0.077	Type 1	LR
Year × BAG					
WPM	1.70	3			
SAS-T3/W	1.720	3	0.633	Type 3	Wald
SAS-T3/LR	1.745	3	0.627	Type 3	LR
SAS-T1/LR	2.709	3	0.439	Type 1	LR
Sex × BAG					
WPM	102.7	3			
SAS-T3/W	104.942	3	0.000	Type 3	Wald
SAS-T3/LR	142.539	3	0.000	Type 3	LR
SAS-T1/LR	144.738	3	0.000	Type 1	LR
Year × Sex × BAG					
WPM	4.04	3			
SAS-T3/W	4.023	3	0.259	Type 3	Wald
SAS-T3/LR	3.997	3	0.264	Type 3	LR
SAS-T1/LR	3.977	3	0.264	Type 1	LR

In comparing the results of the analyses above, we see that
- WPM and SAS-T3/W produce comparable output (and equal to Goodman's);
- SAS-T3/LR and SAS-T1/LR can produce quite different results (e.g., Sex × BAG, Year × Sex, and Year × BAG).
- SAS-T1/LR, sequential analysis, yields most often (and largest) deviant results.