Journal of Statistical Software

May 2005, Volume 14, Issue 2. *http://www.jstatsoft.org/*

SAS Macro BDM for Fitting the Dale Regression Model to Bivariate Ordinal Response Data

Garnett McMillan
Behavioral Health Research Center of the Southwest

Timothy Hanson
University of Minnesota

Abstract

A SAS macro for fitting an extension of the Dale (1986) regression model to bivariate ordinal data is provided. The macro is described in detail and examples from Dale (1986) and McMillan, Hanson, Bedrick, and Lapham (2005) are discussed.

Keywords: contingency table, ordinal regression, SAS macro.

1. Motivating example and objectives

The BDM SAS macro was developed to fit the Bivariate Dale Model (BDM) to bivariate ordinal level data. The motivating problem was modeling alcohol use frequency and quantity. One data collection instrument that alcohol researchers and clinicians generally rely on is called the Quantity-Frequency survey. Quantity-frequency surveys usually consist of two items that query the respondent on average frequency of drinking events and the average quantity consumed per such event. The Alcohol Use Disorders Identification Test (AUDIT), for example, asks, "On average, how often do you drink alcohol?", which corresponds to the frequency measure. The respondent is then asked, "On days that you drink, on average how much alcohol do you consume?", which corresponds to the quantity measure. Response levels for the frequency and quantity measures used in the AUDIT questionnaire are shown in Table 1.

An appropriate statistical methodology is necessary for quantifying risk factors of alcohol consumption pattern and intervention effect sizes using Quantity-Frequency survey data, which have certain particular features that must be addressed. First, the frequency item includes the response "I Never Drink" for individuals who are teetotalers. This frequency level makes the quantity measure (drinks per drinking occasion) meaningless since one cannot specify the number of drinks per drinking occasion if one never has any drinking occasions! If abstainers are in the sample considered, then a complete analysis requires a two-part analysis of (a) the event at which one drinks at all, and (b) alcohol consumption pattern given that one is not an

Frequency of alcohol use	**Quantity of alcohol use**
On average, how often do you drink alcohol?	On days that you drink, on average how much alcohol do you consume?
(Never)	1-2 drinks
Monthly or less	3-4 drinks
2-4 times per month	5-6 drinks
2-3 times per week	7-9 drinks
4 or more times per week	10 drinks or more

Table 1: Quantity-frequency response scales from the Alcohol Use Disorders Identification Test instrument.

abstainer. This is not a particularly complex problem and is analogous to two-part analyses commonly used in medical-cost analysis (Lachenbruch (2001)).

Second, quantity-frequency measures are usually expressed on an ordinal scale. There is no interval or ratio scale difference between "Monthly" and "2-3 times per week" on the AUDIT frequency of consumption items except to say that the former is less frequent than the latter. Alcohol quantity-frequency modeling therefore requires methods suitable for ordinal data. Finally, clinicians and alcohol epidemiologists agree that alcohol frequency and quantity of consumption are not independent behaviors, and the degree of association between quantity and frequency likely varies among sub-populations (Makela (1996)). For example, alcoholics drink frequently to mitigate the somatic and psychological effects of alcohol withdrawal, but physiological tolerance, which increases the rate of alcohol metabolism and reduces the intoxicating effect of individual doses, requires increased quantities of alcohol per consumption event to obtain the desired "high." Alcohol quantity-frequency modeling requires statistical methods for the bivariate ordinal nature of the quantity-frequency data, while also allowing for sub-population variability in the degree of association between quantity and frequency of alcohol consumption.

We suggest using the Bivariate Dale Model (Dale (1984), Dale (1985), Dale (1986)) to model the quantity and frequency of alcohol consumption and to estimate risk factors for alcohol consumption patterns. The BDM allows us to model the joint distribution of the quantity and frequency of alcohol consumption when recorded on an ordinal scale. BDM parameter estimates are expressed in log-odds ratios and are interpreted in exactly the same manner as ordinal logistic regression results. The BDM also allows one to infer the correlation between quantity and frequency of alcohol consumption, and to model variation in this association as a function of covariates. While the motivating example pertains to alcohol use, the `BDM` macro can be used to fit the Bivariate Dale Model to any data set containing bivariate ordinal response data.

2. Statistical development

The model is comprised of two components. Let d be the binary indicator that an individual drinks ($d = 1$ denotes drinking, and $d = 0$ denotes abstinence) and, conditional on the event that a person drinks (i.e., $d = 1$), let f and q denote the frequency and quantity that an individual drinks on the r and c level ordinal scales. For example, $r = 4$ and $c = 5$ in

the AUDIT example shown in Table 1. We model the probability that an individual with covariate vector $\mathbf{x}$ drinks at all using logistic regression as

$$\text{logit}\{P(d = 1)\} = \mathbf{x}'\boldsymbol{\beta}_1.$$

$\boldsymbol{\beta}_1$ is a vector of regression coefficients for the log odds that a person with covariate combination $\mathbf{x}$ drinks, and is interpreted in the usual way (Collett (1991)). Conditional on $d = 1$ we model the bivariate ordinal vector $(f, q)'$ using a BDM (Dale (1986), Molenberghs and Lesaffre (1994)). The marginal probabilities $P(f \leq i)$, $i = 1, \ldots, r-1$, and $P(q \leq j)$, $j = 1, \ldots, c-1$, are modeled using ordinal logistic regression:

$$\text{logit}\{P(f \leq i)\} = \theta_{f,i} + \mathbf{x}'\boldsymbol{\beta}_2,$$

$$\text{logit}\{P(q \leq j)\} = \theta_{q,j} + \mathbf{x}'\boldsymbol{\beta}_3.$$

Note that $\mathbf{x}$ defines covariates applicable to each of these models, but does not necessarily overlap among models. For example, age might be important in predicting the probability that one drinks or frequency of drinking given that one drinks, but might not be included in the model of the quantity that one drinks per drinking occasion. Each $\beta_{2,j}$ parameter expresses the log-odds of drinking at or below frequency level i for an individual with covariate value x_j relative to one with covariate value $x_j - 1$. A similar interpretation holds for the $\boldsymbol{\beta}_3$ parameters, but with respect to quantity consumed. The $\{\theta_{f,1}, \ldots, \theta_{f,r-1}; \theta_{q,1}, \ldots, \theta_{q,c-1}\}$ terms are intercepts expressing the log odds of drinking at or below frequency level i or quantity level j. In this particular parameterization of the ordinal logit model, the intercept terms increase from the lowest to the highest levels on each ordinal scale. Note that the highest level category does not have an intercept term. In particular, $P(f \leq r) = P(q \leq c) = 1$, thus the highest levels need not be considered in the specification of the model. The interpretation of regression coefficients for the ordinal logistic model appears in most introductory texts on categorical data analysis (e.g. Everitt (1994), Agresti (2002)).

Possible dependence between f and q is modeled using a global cross-ratio (GCR) model (Dale (1986), Molenberghs and Lesaffre (1994)). The GCR is a useful measure of association for contingency tables in which the row and column responses are ordered variables with greater than two levels each (Dale, 1984). The GCR is defined for a pair of "cutpoints" (i, j) on the quantity and frequency scales (e.g. Figure 1, Lesaffre and Molenberghs (1991)). Cutpoints refer to particular levels on the quantity and frequency scales about which the level of association between the two is measured. The GCR is equal to the cross-product ratio

$$\psi_{ij} = \frac{P(f \leq i, q \leq j)P(f > i, q > j)}{P(f > i, q \leq j)P(f \leq i, q > j)}$$

for a table dichotomized at cutpoints (i, j). This is the odds ratio of cumulative quantity and frequency levels, and is interpreted as the ratio of the odds of f being at or under category i given q is at or under category j to the odds of f being at or under category i given q is *greater* than category j. Details of the GCR, and its relation to other measures of association are in Dale (1984). We assume the log-linear model

$$\log(\psi_{ij}) = \Delta + \alpha_i + \gamma_j + \mathbf{x}'\boldsymbol{\beta}_4.$$

The GCR is modeled as a function of the frequency and quantity cutpoints (i, j), as well as covariate vector $\mathbf{x}$. Note that $\psi_{r,c}$ is undefined, and that a $\alpha_{r-1} = \gamma_{c-1} = 0$ for modeling

purposes so that $\log(\psi_{r-1,c-1}) = \Delta + \mathbf{x}'\boldsymbol{\beta}_4$. Thus, there are $r-2$ $\{\alpha_i\}$ terms and $c-2$ $\{\gamma_j\}$ terms in the model. When ψ_{ij} does not depend on cutpoints (i,j), then the constant GCR model ($= e^{\Delta+\mathbf{x}'\boldsymbol{\beta}_4}$) obtains for any given covariate vector $\mathbf{x}$ over the entire table (Dale, 1986).

The contribution of an individual who abstains from drinking to the overall likelihood is simply $p(d|\boldsymbol{\beta}_1)$. The contribution from an individual who drinks is given by the product of mass functions $p(d|\boldsymbol{\beta}_1)p(f,q|d=1,\boldsymbol{\tau})$, where $\boldsymbol{\tau}$ is the vector of parameters associated with the BDM. For the motivating example $\boldsymbol{\tau}$ is

$$\boldsymbol{\tau} = (\boldsymbol{\beta}_2', \boldsymbol{\beta}_3', \boldsymbol{\beta}_4', \theta_{f,1}, \theta_{f,2}, \theta_{f,3}, \theta_{q,1}, \theta_{q,2}, \theta_{q,3}, \theta_{q,4}, \Delta, \alpha_1, \alpha_2, \gamma_1, \gamma_2, \gamma_3)'.$$

$p(f,q|d=1,\boldsymbol{\tau})$ has been defined (Dale (1986), Molenberghs and Lesaffre (1994)). The full likelihood function factors into two separate functions of $\boldsymbol{\beta}_1$ and $\boldsymbol{\tau}$, the former based on all subjects and the latter based on only those subjects who drink. Because the likelihood factors, the two separate models (logistic for whether someone drinks and BDM to model the alcohol consumption pattern in those that drink) are fit to the data, but the resulting inferences can be interpreted simultaneously. We can perform a formal statistical test of the null hypothesis of no association between quantity and frequency of alcohol consumption. If the null hypothesis of no association is rejected, then results of simple univariate ordinal regression analysis that assume independence between the quantity and frequency of alcohol use should be regarded with skepticism. The log-likelihoods of independent ordinal logit models fit individually to the quantity and frequency measures are additive. The drop in deviance from the sum of these independent models to the GCR model has a chi-square distribution with $k+1$ degrees of freedom, where k is the number of predictors in the GCR model described above. A statistically significant drop in the deviance after incorporating the GCR into the analysis indicates that the association between the quantity and frequency of alcohol consumption is important and should be considered during statistical analysis.

3. Code description

The `BDM` macro models $P(d=1)$ using `PROC LOGISTIC`, and bivariate ordinal responses are modeled using the `PROC NLMIXED` procedure in SAS Version 8.2. We chose to write the program using SAS/STAT procedures as they are widely used by epidemiologists and biostatisticians. SAS/IML or `PROC NLP` in the SAS/OR module could accomplish the same thing given a reasonable amount of programming experience. The `BDM` macro allows one to control the predictor set for each part of the model, including the cutpoint effects for the GCR model. Predictors can be on either quantitative or categorical scales, but categorical predictors must be coded as dummy variates. Procedure output includes parameter estimates, standard errors, confidence intervals, and p-values for the hypothesis test of no difference from zero. The drop in deviance from the marginal-only models to the model including the GCR portion of the model is also shown to test the improvement in fit when the association between ordinal outcomes is included in the model. The macro generates a table of observed and predicted counts for each covariate combination included in the model, and outputs raw and Pearson residuals for model criticism.

Output datasets include:

- `BDM_BIN_EST` = Covariance matrix for the Bernoulli part (if applicable).

- BDM_BIN_PARMEST = Parameter estimates for the Bernoulli part (if applicable).
- BDM_SPECS = Model specification for the BDM.
- BDM_FITS = Fit statistics for the BDM.
- BDM_EST = Parameter estimates for the BDM.
- BDM_COV = Covariance matrix for the BDM.
- BDM_FINAL = Observed and predicted counts, with raw and Pearson residuals.

4. Data formatting

Each row of the input dataset must pertain to one level of the row response, one level of the column response, and a unique covariate combination within the row-column combinations. Levels of each dependent variable are ordered from 1 to #, with 1 being the lowest and # levels being the highest. The data may be in summarized or unsummarized from. It is advisable to limit the variable name lengths to 8 characters or less, and variable names should not end with digits. This causes instability the algorithm. For example, Dale (1986) provides the following data table (Table 1, p. 910) of patient pain level and medication requirements as a function of operation type.

Operation	Pain level	Medication Requirements			
		Never	Seldom	Occ.	Reg.
VP	None	170	7	8	0
VP	Slight	18	5	8	3
VP	Significant	7	0	4	14
VA	None	170	7	5	2
VA	Slight	22	7	8	1
VA	Significant	8	1	8	9
VH	None	176	8	5	2
VH	Slight	26	6	5	5
VH	Significant	14	3	2	9
RA	None	181	6	6	2
RA	Slight	17	12	7	3
RA	Significant	10	2	3	11

These data needs to be modified to appear as follows:

pain	VP	VA	VH	RA	MED	CT
1	1	0	0	0	1	170
1	1	0	0	0	2	7
1	1	0	0	0	3	8
1	1	0	0	0	4	0
2	1	0	0	0	1	18
2	1	0	0	0	2	5

et cetera...

5. Inputs

The macro is called as:

```
%BDM(   dat=,
Cond=,
Condpred=,
Rowvar=,
Colvar=,
Ct=,
Rowpred=,
Colpred=,
Gcrpred=,
Gcrrow=,
Gcrcol=);
```

Inputs are defined as follows.

- `dat` = Name of input dataset. See below for data formatting.
- `Cond` = Name of outcome measure for Bernoulli portion of two-part model. This is left blank for standard BDM model fitting.
- `Condpred` = Predictors used for the Bernoulli piece, separated by blanks.
- `Rowvar` = Name of the ordinal response for the row variable.
- `Colvar` = Name of the ordinal response for the column variable.
- `Ct` = Name of the cell count variable.
- `Rowpred` = Predictors used for the row variable marginal model, separated by blanks. Leaving this blank will only model the Row response with intercept terms.
- `Colpred` = Predictors used for the column variable marginal model, separated by blanks. Leaving this blank will only model the Column response with intercept terms.
- `Gcrpred` = Predictors used for the GCR model.
- `Gcrrow` = 'row' indicates the GCR model will contain row level terms. Leaving this blank will omit the row level in the GCR model.
- `Gcrcol` = 'col' indicates the GCR model will contain column level terms. Leaving this blank will omit the column level in the GCR model.

6. Examples

6.1. Example 1

This example is a replication of an analysis in Dale (1986). In this example, Dale fits the BDM to self-reported pain level and medication requirements, each of which is measured on an ordinal scale. The data are shown in Table 1, above. The following SAS code reads the table, formats it for use with the BDM, and calls the `BDM` macro to fit the BDM. The fitted model includes no predictors on either ordinal response scales, but has operation type 'VH' in the GCR model, along with terms that vary the association between pain level and medication requirement over levels of the medication requirements. Data and code for this example are included with the `BDM` macro.

```
*****   Example: Dale (1986), Table 3, P.913 **********;

Data dale1986;
    input OPERATION $ pain NEVER SELDOM OCC REG;

IF OPERATION = 'VP' THEN VP = 1; ELSE VP = 0;
IF OPERATION = 'VA' THEN VA = 1; ELSE VA = 0;
IF OPERATION = 'VH' THEN VH = 1; ELSE VH = 0;
IF OPERATION = 'RA' THEN RA = 1; ELSE RA = 0;

    MED = 1; CT = NEVER;   OUTPUT;
    MED = 2; CT = SELDOM;  OUTPUT;
    MED = 3; CT = OCC;     OUTPUT;
    MED = 4; CT = REG;     OUTPUT;

DROP  NEVER SELDOM OCC REG operation;
DATALINES;
VP  1   170 7   8   0
VP  2   18  5   8   3
VP  3   7   0   4   14
VA  1   170 7   5   2
VA  2   22  7   8   1
VA  3   8   1   8   9
VH  1   176 8   5   2
VH  2   26  6   5   5
VH  3   14  3   2   9
RA  1   181 6   6   2
RA  2   17  12  7   3
RA  3   10  2   3   11
;
RUN;

%BDM(dat=dale1986,
rowvar=PAIN,
colvar=MED,
    ct=ct,
gcrpred=VH,
gcrcol=col);
```

Results of this analysis are shown below:

```
     Specifications for Bivariate Dale Model                          [1]
          Response variables PAIN MED

   Descr                                  Value

   Data Set                               WORK.BDM_D1
   Dependent Variable                     ll
   Distribution for Dependent Variable    General
   Optimization Technique                 Trust Region
   Integration Method                     None

   Convergence Status of Bivariate Dale Model                       [2]

                  Reason
   NOTE: GCONV convergence criterion satisfied.

Fit Statistics for Bivariate Dale Model and Marginal Models         [3]

                              Sum -2 LogL     Drop in
                   Value     for Marginal    Deviance
Description       for BDM    ONLY Models      w/GCR      DF    P-value

-2 Log Likelihood  2622.9       2913.56       290.645     4    <.0001

Parameter Estimates for Bivariate Dale Model                        [4]

             Log Odds    Standard
Parameter     Ratio       Error   P-value    Lower CL    Upper CL
```

MED_int1	1.4349	0.07951	<.0001	1.2708	1.5990
MED_int2	1.9060	0.09324	<.0001	1.7136	2.0985
MED_int3	2.7289	0.1303	<.0001	2.4600	2.9978
PAIN_int1	1.0858	0.07132	<.0001	0.9386	1.2330
PAIN_int2	2.1677	0.1014	<.0001	1.9584	2.3770
GCRdelta	3.8333	0.3048	<.0001	3.2042	4.4625
GCRMED1	-1.0806	0.2793	0.0007	-1.6571	-0.5041
GCRMED2	-0.7872	0.2529	0.0047	-1.3091	-0.2653
GCR_VH	-0.7617	0.3511	0.0402	-1.4864	-0.03698

Note that regression coefficients for the marginal model indicate the log odds ratio of being at or below a particular response level. Negative coefficients indicate reduced probability of being at or BELOW a particular level.

Observed & Predicted Values from Bivariate Dale Model
Dataset = work.final [5]

pain	MED	VH	Predicted count	ct	Raw Residual (O-E)	Pearson Residual
1	1	0	521.727	521	-0.72748	-0.05756
1	2	0	22.059	20	-2.05851	-0.44487
1	3	0	14.904	19	4.09593	1.07163
1	4	0	3.497	4	0.50268	0.26942
2	1	0	63.025	57	-6.02465	-0.79284
2	2	0	18.766	24	5.23430	1.22367
2	3	0	22.706	23	0.29354	0.06255
2	4	0	8.091	7	-1.09076	-0.38555
3	1	0	22.615	25	2.38521	0.50929
3	2	0	6.480	3	-3.47989	-1.37297
3	3	0	13.629	15	1.37144	0.37491
3	4	0	34.502	34	-0.50181	-0.08746
1	1	1	175.019	176	0.98071	0.12916
1	2	1	9.732	8	-1.73150	-0.56570
1	3	1	8.033	5	-3.03306	-1.08700
1	4	1	2.337	2	-0.33704	-0.22146
2	1	1	24.082	26	1.91757	0.41013
2	2	1	4.659	6	1.34143	0.62712
2	3	1	6.529	5	-1.52864	-0.60589
2	4	1	3.807	5	1.19338	0.61617
3	1	1	11.700	14	2.30017	0.68807
3	2	1	2.028	3	0.97203	0.68524
3	3	1	3.222	2	-1.22208	-0.68506
3	4	1	9.853	9	-0.85297	-0.27702
				====		
				1013		

Each result section is described below:

`[1]` Model description information.

`[2]` Convergence status of the BDM model.

`[3]` Results of the test of the null hypothesis of no association between the ordinal responses after model adjustment. The drop in deviance that occurs after adding the GCR portion of the model is highly statistically significant, indicating a marked improvement in model fit once the association between pain level and medication requirements is built into the analysis.

`[4]` Parameter estimates. These are interpreted in the usual way for ordinal logit models. Parameters relevant for each response variable are given that variable name as a prefix. For example, the medication requirement intercepts are given the prefix `MED`. The `int` suffix identifies the intercept terms. The `GCR` prefix denotes parameter estimates for the GCR portion of the model. Detailed interpretation of the parameter estimates is given in Dale (1986).

b

		Gender								
		Females					Males			
			Beers / occasion					Beers / occasion		
	Frequency		1	2-3	4-5	6+		1	2-3	4-5	6+
Age $\leq$ 30											
No abuse	Abstainer	36					75				
	Up to 1-2 times / month		20	53	4	0		65	171	42	7
	A few times / month		1	30	12	1		12	130	81	19
	A few times / week		0	4	2	1		1	32	27	14
	Almost daily		0	0	0	0		0	2	2	4
Abuse	Abstainer	12					23				
	Up to 1-2 times / month		4	19	3	0		12	43	7	8
	A few times / month		0	13	7	0		4	49	25	9
	A few times /week		0	0	1	5		1	11	12	7
	Almost daily		0	0	0	0		0	3	2	1
Age > 30											
No abuse	Abstainer	37					64				
	Up to 1-2 times / month		24	33	3	0		62	159	27	6
	A few times / month		2	22	11	2		5	106	59	16
	A few times / week		0	8	4	1		3	50	40	8
	Almost daily		0	0	0	0		0	6	1	9
Abuse	Abstainer	24					16				
	Up to 1-2 times / month		4	25	2	1		15	53	9	1
	A few times / month		0	11	6	2		2	56	22	10
	A few times / week		0	4	3	2		1	28	21	12
	Almost daily		0	0	2	0		0	4	4	14

Table 2: DWI offender alcohol quantity-frequency data by age, gender, and history of physical/sexual abuse.

[5] Table of observed and predicted cell counts for each covariate combination and response level. None of the Pearson's residuals exceed 2, indicating a reasonable model fit to the data.

6.2. Example 2

This example, from McMillan *et al.* (2005), concerns the relationship between the consumption of beer and physical or sexual abuse among DWI offenders. These data include non-beer drinkers, demonstrating the two-part model described in the prequel. See details therein for detailed description of the study sample. The original data is shown in Table 2.

Note that the data are relatively sparse for certain covariate combinations. This poses no problem to the modeling framework proposed here, although complex three-way interaction effects (e.g. age by gender by smoking status) might not be estimable. Such conditions are familiar to epidemiologists working with data in which the outcome measure is strongly

separated by regions in the predictor space. This phenomenon is referred to as "quasi/complete separation" and results in infinite maximum likelihood estimates. Also note that the data appear to be concentrated on the diagonals or upper right areas of each table. This strongly suggests a high degree of association between the quantity and frequency variables. The quantity and frequency variables were re-ordered so that 4 corresponds to the lowest quantity (1 beer per occasion) and lowest frequency (up to 1-2 times per month) categories. This reordering is necessary so that the regression coefficients are expressed in log-odds of drinking at or above each quantity or frequency level. Positive coefficients therefore express greater risk of drinking more frequently or with greater quantity, which is more easily understood by alcohol researchers. The macro code is called up to fit the BDM.

```
%BDM(dat=analysis,
     cond=DRINKER,
     condpred=SEX,
     rowvar=Q,
     colvar=F,
     ct=COUNT,
     rowpred= PHYS_ABUS SEX sexphys,
     colpred=AGECAT PHYS_ABUS SEX ,
     gcrpred= SEX ,
     gcrrow=row,
     gcrcol=);
```

The output is identical to that of the first analysis, with the addition of the logistic regression results for the probability that one drinks.

Logistic regression model parameter estimates of the probability of DRINKER.

Variable	DF	Estimate	StdErr	WaldChiSq	Prob ChiSq
Intercept	1	1.1722	0.1096	114.3707	<.0001
SEX	1	1.0312	0.1351	58.2565	<.0001

Specifications for Bivariate Dale Model
Response variables Q F

Descr	Value
Data Set	WORK.BDM_D1
Dependent Variable	ll
Distribution for Dependent Variable	General
Optimization Technique	Trust Region
Integration Method	None

Convergence Status of Bivariate Dale Model

Reason
NOTE: GCONV convergence criterion satisfied.

Fit Statistics for Bivariate Dale Model and Marginal Models

Description	Value for BDM	Sum -2 LogL for Marginal ONLY Models	Drop in Deviance w/GCR	DF	P-value

-2 Log Likelihood	8243.3	8643.10	399.759	4	<.0001

Parameter Estimates for Bivariate Dale Model

Parameter	Log Odds Ratio	Standard Error	P-value	Lower CL	Upper CL
F_int1	-4.3642	0.1782	<.0001	-4.7179	-4.0105
F_int2	-2.2853	0.1235	<.0001	-2.5303	-2.0403
F_int3	-0.5471	0.1122	<.0001	-0.7698	-0.3244
F_AGECAT	0.2550	0.07636	0.0012	0.1034	0.4065
F_PHYS_ABUS	0.5172	0.09341	<.0001	0.3319	0.7026
F_SEX	0.5920	0.1122	<.0001	0.3692	0.8147
Q_int1	-3.1507	0.1473	<.0001	-3.4431	-2.8583
Q_int2	-1.5390	0.1288	<.0001	-1.7947	-1.2834
Q_int3	1.3133	0.1282	<.0001	1.0589	1.5676
Q_PHYS_ABUS	0.8813	0.1912	<.0001	0.5019	1.2608
Q_SEX	0.7020	0.1350	<.0001	0.4340	0.9700
Q_sexphys	-0.5309	0.2080	0.0122	-0.9437	-0.1182
GCRdelta	2.9001	0.3019	<.0001	2.3010	3.4993
GCRQ1	-0.4323	0.2505	0.0876	-0.9294	0.06485
GCRQ2	-0.7175	0.2065	0.0008	-1.1272	-0.3078
GCR_SEX	-0.7845	0.2613	0.0034	-1.3031	-0.2659

Note that regression coefficients for the marginal model indicate the log odds ratio of being at or below a particular response level. Negative coefficients indicate reduced probability of being at or BELOW a particular level.

A portion of the residual analysis table is shown below.

SEX	PHYS_ ABUS	sexphys	AGECAT	Predicted count	COUNT	Residual (O-E)	Pearson
1	1	1	1	4.364	14	9.6359	4.65306
1	1	1	0	2.287	8	5.7128	3.79995

Covariate combinations with Pearson's residuals larger than about 3 are poorly fit and might indicate some degree of model inadequacy. In this result, old men who have suffered physical or sexual abuse as a child have substantially lower predicted probabilities of being in the highest quantity and frequency group (observed count = 14, expected = 4.4). Also, younger men who have suffered physical or sexual abuse as a child are more common that expected in the highest quantity / lowest frequency levels of the responses. These subjects are not well fit by the model.

References

Agresti A (2002). *Categorical Data Analysis*. Wiley and Sons Inc., Hoboken, N.J., second edition.

Collett D (1991). *Modelling Binary Data*. Chapman and Hall, London.

Dale J (1984). "Local Versus Global Association for Bivariate Ordered Responses." *Biometrika*, **71**, 507–514.

Dale J (1985). "A Bivariate Discrete Model of Changing Colour in Blackbirds." In D Brillinger, S Fienberg, J Gane, J Hartigan, K Krickeberg (eds.), "Lecture Notes on Statistics: Statistics in Ornithology," volume 29, pp. 25–35. Springer-Verlag.

Dale J (1986). "Global Cross-ratio Models for Bivariate, Discrete, Ordered Responses." *Biometrics*, **42**, 909–917.

Everitt B (1994). *The Analysis of Contingency Tables*. The Guilford Press, New York, second edition.

Lachenbruch P (2001). "Comparisons of Two-part Models with Competitors." *Statistics in Medicine*, **20**, 1215–1234.

Lesaffre E, Molenberghs G (1991). "Multivariate Probit Analysis: A Neglected Procedure in Medical Statistics." *Statistics in Medicine*, **10**, 1391–1403.

Makela K (1996). "How to Describe the Domains of Drinking and Consequences." *Addiction*, **91**, 1447–1449.

McMillan G, Hanson T, Bedrick E, Lapham S (2005). "Using the Bivariate Dale Model to Jointly Estimate Predictors of Frequency and Quantity of Alcohol Use." *Journal of Studies on Alcohol*. In press.

Molenberghs G, Lesaffre E (1994). "Marginal Modeling of Correlated Ordinal Data Using a Multivariate Plackett Distribution." *Journal of the American Statistical Association*, **89**, 633–644.

Affiliation:

Garnett P. McMillan
Behavioral Health Research Center of the Southwest
612 Encino Pl. N.E.
Albuquerque, NM 87102, United States of America
E-mail: GMcMillan@bhrcs.org

Journal of Statistical Software
May 2005, Volume 14, Issue 2.
http://www.jstatsoft.org/

Submitted: 2005-01-07
Accepted: 2005-05-01