Special Issue on Family Demography, Canadian Studies in Population
Vol. 27(1), 2000, pp. 209-229

Log-linear Rate Modeling of Cross-classified Divorce Data

Teresa S. J. Abada

Department of Sociology and
Population Research Laboratory
University of Alberta
Edmonton, Alberta

Abstract

Analysis of tabular data based on censuses or vital statistics concerning demographic events such as births, deaths, marriages, divorce and migration can be easily executed with log-linear methods. This paper shows how published cross-classified data on divorce can be analyzed with the use of log-linear rate modeling, which is a special case of the general log-linear model. While there exists other introductions to this topic, they tend to be quite technical and assume that readers are familiar with the rudiments and computational aspects of this method. The main objective is to provide a rudimentary exposition of the log-rate model and its application to demographic data analysis. The illustrative analysis uses published vital statistics tabular data on divorces by duration of marriage for the Canadian provinces and the territories for 1971, 1981, 1991, and 1995.

Résumé

L'analyse de données tabulaires issues de recensements ou de statistiques démographiques (naissances, décès, mariages, divorces et migrations) s'exécute aisément grâce aux méthodes log-linéaires. Le présent article montre comment les données de recoupement publiées sur le divorce se prêtent à ce type de modélisation et propose un exemple de modèle log-linéaire général. Bien qu'il existe d'autres exposés sur le sujet, ils sont généralement ardus et s'adressent à des lecteurs avertis. Le présent article, par contre, a pour objectif de présenter les rudiments du modèle et son application à l'analyse des données démographiques. L'exemple traite de données tabulaires publiées portant sur les divorces d'après la durée du mariage dans les provinces et les territoires canadiens pour 1971, 1981, 1991 et 1995.

Key words: log-linear model, log-rate model, cross-classified data, divorce

Introduction

The general log-linear model for frequencies (Feinberg 1977; Bishop, Feinberg and Holland 1975) can be modified to model the expected rate of some event as a function of predictor variables. The usefulness of this type of modeling for demographic analysis is demonstrated in this research note with an application to geographic differences in divorce in Canada between 1971and 1995, using data from Vital Statistics publications[1]. Two illustrative examples are developed.

Log-Linear and Log-Rate Models

Log-linear analysis encompasses a variety of related statistical techniques, including general models of multi-way contingency tables, logit models, and logistic regression (Rodgers 1995; Agresti 1990; Agresti and Finlay 1984; Walsh 1987; Knoke and Burke 1980). Unlike linear regression, general log-linear models do not always require a distinction between the independent and dependent variables. In such models, the criterion to be analyzed is the expected cell frequencies, F_{ij}'s under a fitted model.

Log-linear rate models (also referred to as log-rate models) require the specification of a dependent variable, expressed as a rate; hence information on those experiencing the event of interest (e.g. divorce) is related to the set of corresponding populations exposed to the risk of a given event (e.g. marriages). Unlike the general log-linear model, the log-rate model specifies a dependent variable that is presumed to depend on some set of explanatory factors. Table 1 shows computed log-frequencies and log of rates based on 1971 and 1995 information for divorces and marriages cross-classified by two regions and two duration of marriage categories.[2]

Adoption of the log-rate model implies asking the question: "What is the expected rate of divorce given the set of predictor variables?"[3] The saturated log-frequency model can be expressed as:

$$\log(\mathrm{Fijk}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$$

As indicated, the predicted log frequency cell values ($\log(F_{ijk})$) are expressed as a function of an average effect λ, plus the main effects of predictor variables A, B, C, and their interactions. Similarly, the log-rate equation,

$$\log(\mathrm{Fijk/Fijk}) = \delta + \delta_i^A + \delta_j^B + \delta_k^C + \delta_{ij}^{AB} + \delta_{jk}^{BC} + \delta_{ik}^{AC} + \delta_{ijk}^{ABC}$$

states the logarithm of the expected rate of divorce, is a function of an average effect δ , main effects of A, B, and C (δ_i^A, δ_j^B ,and δ_k^C) and their interactions (δ_{ij}^{AB}, δ_{jk}^{BC} ,δ_{ik}^{AC} and δ_{ijk}^{ABC}). The parameters in this model measure the extent to which a term adds to, or subtracts from log (F_{ijk}/N_{ijk}), when each of the respective variables are set at level 1.

Application of the log-rate model is appropriate when the event of interest for analysis is statistically rare and its rate is small (below .10). An appropriate underlying distribution in such cases is the Poisson (Agresti, 1990; Clogg and Eliason, 1987). Laird and Olivier (1981) have shown that Poisson regression models can be analyzed with log-linear techniques (i.e. log-rate modeling) using Maximum Likelihood to estimate parameters.[4]

Illustrative Analysis

The illustrative analysis proceeds in two steps: first, log-frequency and log-rate models are fitted to a 2 x 2 x 2 table in order to demonstrate how parameters for the saturated model are computed; and secondly, the focus shifts to a larger version of this table, to provide a realistic example of log-rate modeling. The data for both analyses were taken from Vital Statistics publications (Statistics Canada 1995a, 1995b, 1991a, 1991b, 1981, 1971). Table 2 shows calculated

Table 1
Divorce Data by Year, Region and Duration of Marriage for Canada: 1971 and 1995

Region	Duration of Marriage			
	1971		1995	
Region 1	<10	10+	<10	10+
Divorced	6943	11874	24382	29468
Marriages	1166519	2533221	1175428	3702440
Rate	0.005952	0.004687	0.020743	0.007959
log (freq)	8.845489	9.382106	10.1016	10.29106
log(rate)	-5.1241	-5.3629	-3.87554	-4.83344
Region 2				
Divorced	4920	5923	11292	11609
Marriages	432803	966066	564701	1487965
Rate	0.011368	0.006131	0.019996	0.007802
log(freq)	8.501064	8.501064	9.33185	9.359536
log(rate)	-4.47697	-5.09439	-3.9122	-4.85338

by Region and Duration of Marriage for Canada: 1971 and 1995

Duration of Marriage	Region 1	Region 2
Duration = <10		
Divorce	31325	16212
Marriages	2341947	997504
Rate	0.013376	0.016253
log(freq)	10.35217	9.693507
log(rate)	-4.31432	-4.1195
Duration = 10+		
Divorce	41342	17532
Marriages	6235661	2454031
Rate	0.00663	0.007144
log(freq)	10.62963	9.771783
log(rate)	-5.01616	-4.94146

Table 1 (Continued)
by Year and Duration of Marriage for Canada: 1971 and 1995

Duration of Marriage	1971	1995
Duration = < 10		
Divorce	11863	35674
Marriages	1599322	1740129
Rate	0.007418	0.020501
log(freq)	9.38118	10.48218
log(rate)	-4.90391	-3.88729
Duration = 10+	1971	1995
Divorce	17797	41077
Marriages	3499287	5190405
Rate	0.005086	0.007914
log(freq)	9.786785	10.6232
log(rate)	-5.28128	-4.83912

by Year and Region for Canada: 1971 and 1995

Region	1971	1995
Region 1		
Divorce	18817	53850
Marriages	3699740	4877868
Rate	0.005086	0.01104
log(freq)	9.842516	10.89396
log(rate)	-5.28126	-4.50626
Region 2		
Divorce	10843	22901
Marriages	1398869	2052666
Rate	0.007751	0.011157
log(freq)	9.291275	10.03894
log(rate)	-4.8599	-4.49571

Table 2
Log-Frequency and Log-Rate Models Applied to the Data in Table 1

Parameter Value	Saturated Log-Frequency $\boxed{\lambda \text{ values}}$	Saturated Log-Rate $\boxed{\delta \text{ values}}$
Average	9.3124	-4.6916
Year 1971	-0.4586	-0.323
Year 1995	0.4586	0.323
Region 1	0.3427	-0.1074
Region 2	-0.3427	0.1074
Duration <10	-0.1174	0.3444
Duration 10+	0.1174	-0.3444
Year1971 x Region1	-0.0827	-0.1215
Year 1995 x Region2	0.0827	0.1215
Year 1971 x Duration <10	-0.0631	-0.1304
Year 1995 x Duration 10+	0.0631	0.1304
Region1 x Duration <10	-0.0641	-0.0452
Region2 x Duration 10+	0.0641	0.0452
Year 1971 x Region1 x Duration <10	-0.0237	-0.0494
Year 1995 x Region2 x Duration 10+	0.0237	0.0494

Note: All parameters from each model had Z-values greater than 1.96 and are therefore statistically significant.

parameters for log-frequency and log-rate saturated models (refer to appendix for a detailed presentation of these computations). Unlike the log-frequency model, the log-rate model accounts for the population at risk as the denominator (yearly marriage cohorts). For example, the parameter log-frequency value for Region 1 is .3427; once the population at risk is accounted for in the log-rate model, the parameter changes to -.1074, indicating that Region 1 has a lower probability of divorcing, given those exposed to the risk of marital dissolution.

Below, the analysis is based on provincial/territorial differences in duration of marriage during four points in time: 1971, 1981, 1991 and 1995. Here, duration of marriage consists of 15 levels (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10-14, 15-19, 20-24, 25-29, 30+ years); and province/territory has 12 categories (Newfoundland, Prince Edward Island, New Brunswick, Nova Scotia, Quebec, Ontario, Manitoba, Saskatchewan, Alberta, British Columbia, Yukon, and Northwest Territories). The population at risk of divorce are the yearly marriage cohorts corresponding to each category of marital duration across province/region and time period. Thus for those divorces which had a duration of zero in a given year i, in which divorces were contracted, the corresponding marriage cohort is those couples that married in the same year i (where i is the current calendar year in which divorce took place); and for duration 1, the appropriate marital cohort would be the marriages contracted in year i-1. Furthermore, divorces contracted in year i of duration 2 are assumed to have originated within the marriages that took place in year i-2, and so forth. One of the fitted models in this analysis is,

$$\log(\mathrm{Fijk/Fijk}) = \delta + \delta_i^Y + \delta_j^P + \delta_k^D + \delta_{ij}^{YP} + \delta_{ik}^{YD} + \delta_{jk}^{PD}$$

with constraints

$$\Sigma\delta_i^Y = \Sigma\delta_j^P = \Sigma\delta_j^D = \Sigma\delta_{ij}^{YP} = \Sigma\delta_{ik}^{YD} = \Sigma\delta_{jk}^{PD} = 0$$

where:

F_{ijk} = the expected number of divorces in year i, province/territory j, and duration of marriage k;

N_{ijk} = the number of marriages associated with duration k.

δ = the constant term (the average of log of expected divorce rates);[5]

δ_i^Y = parameters for year (1=1971, 2 = 1981, 3 = 1991, 4 =1995);

δ_j^P = parameters for province/territory (1=Newfoundland, 2 = Prince Edward Island, 3 = Nova Scotia, 4 = New Brunswick, 5 = Quebec, 6 = Ontario, 7 = Manitoba, 8 = Saskatchewan, 9 = Alberta, 10 = British Columbia, 11= Yukon, 12 = Northwest Territories);

δ_k^D = parameters for marriage duration category (0=0, 1=1, 2=2,...10=9, 11=10-14,...,14=25-29, and 15=30+);

δ_{ij}^{YP} = parameters for the interaction of year and province;

δ_{ik}^{YD} = parameters for the interaction of year and duration;

δ_{jk}^{PD} = parameters for the interaction of duration and province.

Unlike the log-linear frequency model, the dependent variable in this case is the logarithm of a rate, $\log(F_{ijk}/N_{ijk})$. Note that the populations at risk N_{ijk} are not estimated by this model; rather they serve as cell-specific offsets, used to generate parameter values in the Maximum Likelihood procedures (Agresti, 1990; Clogg and Eliason, 1987; Laird and Olivier, 1981). Although Maximum Likelihood Estimates (MLE) of log-rate parameters can be computed with a number of programs, in the present case SPSS LOGLINEAR was used.

Assessing Goodness of Fit

The goodness of fit of alternate models can be assessed on the basis of the fact that log-linear models are nested; that is, selecting a parsimonious model that also provides a good fit to the data can begin with the saturated model, and then proceeding to less complicated ones, given that every model fitted is a subset of a larger and more complex model. A saturated model would include in the equation the three way interaction term, $\delta_{ijk.}$ A less complicated model would include only the two-way interactions while leaving out the three-way term. The main effects model is even less complicated: $\delta + \delta_i^Y + \delta_j^P + \delta_k^D$. A model containing only the intercept assumes that the probability of divorce is uniform across all i-j-k intersections.[6]

The goodness of fit of models can be evaluated using the Likelihood Ratio Chi-square Statistic (L^2). If the difference in L^2 between a simpler model and a more complex one is statistically significant under the Chi-Square distribution, then we would choose the more complex model. If the difference is not statistically significant, the simpler model would be chosen since both would be essentially equivalent. A good-fitting log-linear model is one that produces a minimum difference between observed and expected cell frequencies. When this difference is small, the Log-Likelihood Chi-square statistic L^2 will be statistically insignificant under the Chi-Square probability distribution. The

Table 3. Log-Rate Models for Divorce

Model Fitted		L2	DF	R2 analog
Model 12	[Year, Prov] [Year, Dur] [Prov, Dur]	1360.73	462	99.00%
Model 11	[Year, Dur] [Prov, Dur]	6911.54	495	94.34%
Model 10	[Year, Prov] [Prov, Dur]	5278.8	504	96.13%
Model 9	[Year, Prov] [Year, Dur]	2769.29	616	97.97%
Model 8	[Prov, Dur]	10519.99	537	92.29%
Model 7	[Year, Dur]	8342.37	649	93.89%
Model 6	[Year, Prov]	6551.56	658	95.20%
Model 5	[Year] [Prov] [Dur]	11845.92	691	91.32%
Model 4	[Dur]	33584.26	705	75.39%
Model 3	[Prov]	125138.79	708	8.30%
Model 2	[Year]	123538.08	716	9.47%
Model 1	(Equiprobability)	136459.08	719	

Note: These models are specified by their highest order terms, implying the presence of the corresponding lower-order terms. For example, Models 12, 11, 10, 9, 8, 7, and 6 all include the main effects, year, province and duration.

drawback of L^2 is that it is proportional to the size of the cell values in a table. Therefore, with large cell frequencies, L^2 values will seldom be insignificant. In such cases, model selection can be evaluated by employing the R^2 analog: $1-(L_m^2/L_b^2)$, where $L_m^2 = L^2$ for an alternative model, $L_b^2 = L^2$ for the baseline model.

Table 3 presents models fitted to the divorce data, their respective L^2 values, degrees of freedom, and R^2 analog. The baseline (equiprobability) model gives a poor fit to the data with an L^2 value of 136459.08. Therefore, the hypothesis that all duration-specific divorce rates are equally likely in all provinces for all four time periods can be rejected. Models 2, 3 and 4 account for 9.47%, 8.30% and 75.39% of the variation in divorce rates, suggesting that divorce is strongly related to marital duration and weakly related to province/territory.

The significance of interaction effects can be discerned by contrasting the fit of a model that includes a given interaction term with one that excludes it. From the comparisons of the main effects model (model 5) and the more complex ones (i.e., 6, 7 and 8) the equation containing the interaction of province with year (model 6) produces the highest R^2 analog value (95.20%). Comparing model 6 to the next set of more complex ones (9, 10, and 11) model 9 shows the largest improvement in R^2 (97.97%), suggesting model 9 gives a very good fit to the data. From this, it appears that the probability of divorce by province/territory has changed significantly over time and the effects of duration on divorce also vary over time. The very slight improvement in the R^2 for model 8 indicates that the effects of duration on divorce do not vary significantly by province. Model 12 includes the interaction of duration with province, but adding this interaction would not significantly improve the fit. Model 9 is the accepted model.

Table 4 presents the MLE parameters derived from model 9 in Table 3.[7] The independent variables were deviation coded, whereby the reference category of each predictor was coded as –1; thus, parameter values for the reference categories are the negative sum of the estimated parameters for a particular variable. For example, in Table 4, the effect for 1995 time period on divorce risk is the negative sum of parameters for 1971, 1981 and 1991 (=.3493). As mentioned earlier, the parameters in the log-rate equation indicate how much a given covariate adds to or subtracts from the overall expected divorce rate. The exponent of parameter values derives the extent to which each term multiplies the overall expected divorce rate.

The main effects of province/territory show that five of the twelve geographic areas are associated with increased risks (Quebec, Ontario, Alberta, BC and Yukon), while the remaining areas show reduced effects on divorce. The table also shows that the distribution of divorce follows a curvilinear pattern across duration of marriage, with the risk being lowest in the first two years of marriage and highest at durations 4 and 5.[8]

Table 4
Maximum Likelihood Parameter Estimates for Year, Province and Duration Effect for Canada (Model 9 in Table 3)

Predictor	MLE	Z	e^{λ}
1971	-0.6316	-30.06	0.5317
1981	0.1191	7.65	1.1263
1991	0.1633	10.57	1.1774
1995 (R)	0.3491	NA	1.4178
Province			
Newfoundland	-0.7042	-28.01	0.4945
Prince Edward Island	-0.3838	-9.56	0.6813
Nova Scotia	-0.0706	-4.72	0.9318
New Brunswick	-0.3004	-17.43	0.7405
Quebec	0.1780	17.76	1.1948
Ontario	0.1755	18.22	1.1918
Manitoba	-0.0361	-2.69	0.9646
Saskatchewan	-0.2171	-14.65	0.8048
Alberta	0.4217	39.13	1.5245
British Columbia	0.4771	45.77	1.6113
Yukon	0.4525	8.06	1.5722
Northwest Territories(R)	0.0075	NA	1.0075
Duration			
0	-2.5630	-58.68	0.0771
1	-0.5095	-32.09	0.6008
2	0.2939	26.23	1.3417
3	0.5692	57.21	1.7668
4	0.7448	81.61	2.1059
5	0.7624	84.34	2.1435
6	0.7075	76.56	2.0289
7	0.6443	67.48	1.9047
8	0.6015	61.82	1.8249
9	0.5121	50.24	1.6689
10-14	0.3067	50.67	1.3589
15-19	0.0023	0.34	1.0023
20-24	-0.2242	-30.49	0.7991
25-29	-0.5879	-67.01	0.5555
30+(R)	-1.2599	NA	0.2837
λ	-4.7834		0.00837

Note: Likelihood Ratio $\chi 2 = 2769.28$; DF = 616; $R^2_A = 97.97\%$; Z values below 2.00 are considered statistically insignificant. (R) indicates omitted category.
NA - not computed.

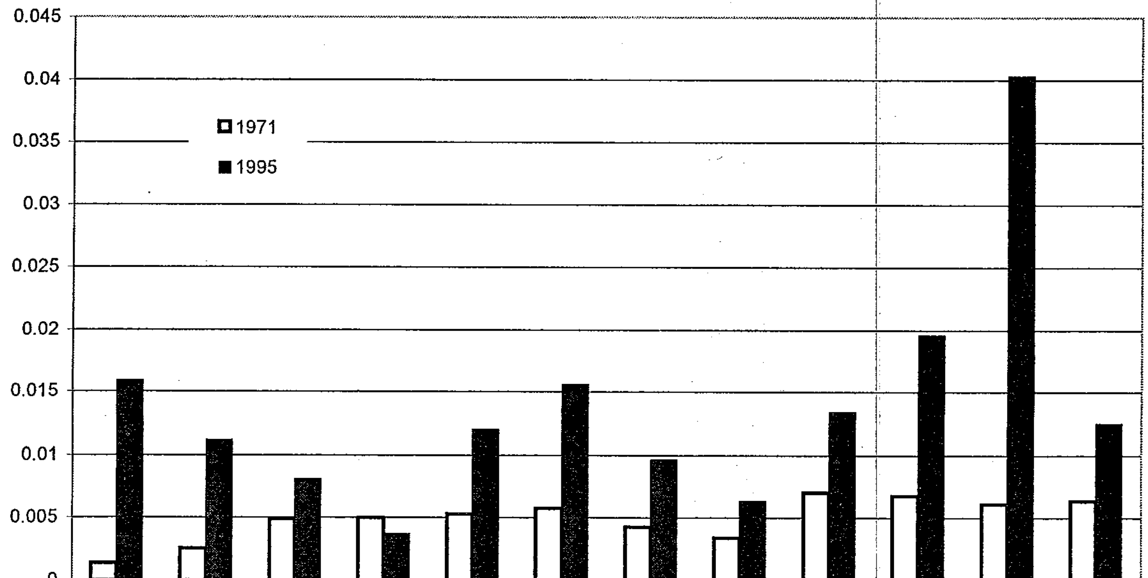
Figure 1 shows the estimated provincial/territory divorce rates for 1971 and 1995. The rates for each province/territory are higher in 1995 than in 1971, particularly in the cases of Newfoundland and the Yukon territories. PEI, Quebec, Ontario, Alberta and BC also show notable increases. The 1995 year effect was .3491 and applies to all provinces and territories. The interaction effect 'province by year' is also taken into account when calculating the predicted rates. The magnitude of its positive effect is greatest for Newfoundland and the Yukon territory (.99330 and .77039 respectively). In contrast, New Brunswick shows a decline in 1995 due to the negative interaction effect of .87393.

Conclusion

Log-rate models are particularly useful in studying social demographic phenomena such as births, deaths, marriages, divorces, and migration that are assumed to be Poisson distributed. As demonstrated in this study, an important feature of the log-rate method is that it can be readily applied to published cross-tabular data. Although not elaborated in the present work, the log-rate method can also be simplified by transforming, where appropriate, categorical predictors into linear covariates. This would reduce the number of parameters to be estimated. For example, in the preceding analysis, duration of marriage had 15 categories (and therefore 14 parameters were estimated); linear transformation would reduce the number of parameters to one. One may also include in the log-rate model "external" linear covariates, taken from other relevant cross-classifications. For instance, in the present analysis one could append to the data matrix province-period specific linear covariates thought to be important in the explanation of geographic differentials in divorce (e.g. female labour force participation rate, male unemployment rate, in-migration rate, etc.).

Log-rate models are linked to survival models. Laird and Olivier (1981) specified this important link. They developed a piecewise exponential survival model based on the same assumptions and procedures associated with the log-linear Poisson model (i.e. log-rate model), that yielded parameter estimates analogous to Cox's (1972) proportional hazards.[9] Further extensions of the log-rate model have been presented. Larson (1984) developed a piecewise exponential Poisson model for independent competing risks (e.g. multiple causes of death), and Vermunt (1996) provides a comprehensive overview of the log-rate model and its variants in the context of event history analysis. One area worthy of further exploration is the possible incorporation of log-rate models of demographic rates (e.g. death rate, fertility rate, divorce rate, etc.) in the context of multilevel analysis (Bryk and Raudenbush, 1992; Goldstein, 1995).

Figure 1. Estimated Divorce Rates for Canada: 1971 and 1995*



	Nfld	PEI	NS	NB	Que	Ont	Man	Sask	Alta	BC	Yukon	NWT
1971	0.0013	0.0025	0.0048	0.0049	0.0052	0.0057	0.0042	0.0033	0.0069	0.0067	0.0060	0.0063
1995	0.0158	0.0111	0.0080	0.0036	0.0120	0.0155	0.0096	0.0063	0.0132	0.0195	0.0403	0.0125

*Estimated rates were derived by adding relevant main and period interaction terms to the overall divorce rate in Table 4 and then exponentiating the sum. For example, the Quebec effect for 1971 is calculated as: exp(-4.7834 - .6316 + .1780 - 0.0116) = 0.00525.

Acknowledgements:

A note of thanks is extended to the anonymous reviewers and the Co-Editors of this special issue of Canadian Studies in Population for having provided insightful comments and suggestions for revision. Any errors or omissions are the sole responsibility of the authors.

Endnotes:

1. Studies of divorce differentials have relied on the application of various statistical techniques, including OLS regression (Breault and Kposowa 1980; Cherlin 1977; Trent and South 1989), as well as methods suited to the analysis of dichotomous dependent variables, such as proportional hazards models, probit and logit regression (Waite and Lillard 1991; South and Spitze 1986; Teachman 1982). For the most part, these statistical methods have been applied to micro level survey data drawn from sample surveys.

2. For this part of the study Region 1 includes Ontario, Quebec and the Maritimes, while the Region 2 includes Manitoba, Saskatchewan, Alberta, BC and the Territories.

3. Logit models can also be used to analyze the relationship between a dichotomous dependent variable and independent variables. The dependent variable is expressed in terms of the log odds ratio, defined as the proportion of the number of cases in one category to the number of cases in another category. Adoption of the logit model is predicated on the question for analysis being "What are the odds of divorce given a set of predictor variables?"

4. In situations involving large sample cases, and where the probability of occurrence of the event of interest is small relative to the population at risk, the expected distribution of the event is Poisson which also assumes that observations are independent random events with equal mean and variance. A random discrete event has a Poisson distribution if its probability distribution is given by:

$$P(x;\lambda) = \frac{\lambda^{x} e^{-x}}{x!} \quad \text{for } x = 0,\ 1,\ 2\ldots$$

   Clogg and Eliason (1987) note that when the rate of some event, π is small, the corresponding odds of some event ω approximate almost perfectly π. For example, if $\pi = .05$, then $\omega = .0526$; if $\pi = .08$, then $\omega = .0869$. If however, $\pi > .10$, then ω is no longer approximately equal to π.

5. Schoen (1970) has demonstrated that the geometric mean is actually a standardized measure of risk, thus disregarding the need for conventional standardization procedures as in the case of direct standardization of rates to remove confounding effects of age composition or some other compositional effect(s).

6. A model worth considering is $\ln(F_{ijk}) = \delta + \delta_{\ln(N_{ijk})}$ to test whether the effects of the denominators on the log-frequency are statistically significant. An insignificant effect would dictate reverting to the simpler log-frequency model.

7. Inclusion of the two-way interactions, "Province x Year", and "Duration x Year" into the model tests the hypothesis that the effects of province and duration on divorce risk remain constant across time periods (Table available on request).

8. Duration still has a positive effect on divorce after 5 years of marriage, although the magnitude of its effect declines until reaching durations 20 and beyond, where the effects become negative. The duration effects on divorce also follow this general pattern across the three time periods. Duration did not vary significantly across province/territory. (Table available upon request).

9. Laird and Olivier (1981) divide the time axis into I mutually exclusive and exhaustive intervals, $\Omega_1, \ldots, \Omega_I$ with a constant hazard function (h_i)within each interval, $h(t, X) = h_i e^{x^T\beta}$ for $t \in \Omega_i$. As i becomes larger, each interval length also becomes arbitrarily small, resulting in a non-parametric model, yielding an estimate of β equivalent to Cox's (1972) estimate. A log-linear hazard function can be written with duration representing the time axis, such that: $\ln h(t, X) = \ln h_i + X^T \beta$, $t \in \Omega_i$ where X is represented by categorical covariates and β is a vector of unknown parameters denoting the effect of covariates.

References:

Agresti, A. 1990. *Categorical Data Analysis*. John Wyley and Son: New York.

Agresti, A. and B. Finlay. 1984. *Statistical Methods for the Social Sciences*. 2nd Edition. San Francisco, CA: Dellen Publishing Company.

Bishop, Y.M.M., Feinberg, S.E. and P.W. Holland. 1975. *Discrete Multivariate Analysis: Theory and Practices*. Cambridge, Massachusetts: M.I.T. Press.

Breault, K.D. and A.J. Kposowa. 1987. "Explaining divorce in the United States: A study of 3111 countries, 1980," *Journal of Marriage and the Family* 49: 549-558.

Bryk, Anthony S. and S.W. Raudenbush. 1992. *Hierarchical Linear Models.* Newbury Park, CA: Sage Publications.

Cherlin, A. 1992. Marriage, Divorce and Remarriage. Cambridge, Massachusetts: Harvard University Press.

Cherlin, A. 1977. "The effect of children on marital disruption," *Demography* 14: 265- 84.

Clogg, C. and S. Eliason. 1987. "Some common problems in log-linear analysis," *Sociological Methods and Research* 16 (1): 8-44.

Cox, D.R. 1972. "Regression models and life tables," *Journal of the Royal Statistical Society* Ser. V, 34: 187-220.

Fienberg, S.E. 1977. *The Analysis of Cross-Classified Data.* Cambridge, Massachusetts: M.I.T. Press.

Goldstein, Harvey. 1995. *Multilevel Statistical Models.* London: Edward Arnold.

Knoke, D. and P.J. Burke. 1980. *Log-linear Models.* Beverly Hills, CA: Sage.

Laird, N. M. and D. Olivier. 1981. "Covariance analysis of censored survival data using log-linear analysis techniques," *Journal of the American Statistical Association* 56: 223-224.

Larson, M. G. 1984. "Covariate analysis of competing-risks data with log-linear models," *Biometrics* 40: 459-469.

Norusis, M. 1993. *SPSS for Windows: Professional Statistics, Release 6.0.* Chicago: SPSS INC.

Rodgers, W. 1995. "Analysis of cross-classified data," In L.G. Grimm and P.R. Yarnold (eds.) *Reading and Understanding Multivariate Statistics.* Washington, DC: American Psychological Association.

Ruggles, S. 1997. "The rise of divorce and separation in the United States, 1880-1990," *Demography* 34 (4): 455-466.

Schoen, R. 1970. "The geometric mean for the age-specific death rates as a summary index of mortality," *Demography* 7 (3): 317-324.

South, S. and G. Spitze. 1986. "Determinants of divorce over the marital life course," *American Sociological Review* 51: 583-590.

Statistics Canada 1995a. *Divorces*. Catalogue No. 84-213.

Statistics Canada 1995b. *Marriages*. Catalogue No. 84-212.

Statistics Canada 1991a. *Divorces*. Catalogue No. 84-213.

Statistics Canada 1991b. *Marriages*. Catalogue No. 84-212.

Statistics Canada 1981. *Marriages and Divorces*. Catalogue No. 84-205.

Statistics Canada 1971. *Marriages and Divorces*. Catalogue No. 84-205.

Teachman, J. D. 1982. "Methodological issues in the analysis of family formation and dissolution," *Journal of Marriage and the Family* Nov. 1037-1053.

Trent, K. and S. South. 1989. "Structural determinants of the divorce rate: A cross-societal analysis," *Journal of Marriage and the Family* 51: 391-404.

Vermunt, Jeroen K. 1996. *Log-Linear Models for Event Histories*. Thousand Oaks, CA: Sage Publications.

Waite, L. J. and L. Lillard 1991. "Children and marital disruption," *American Journal of Sociology* 96 (4): 930-53.

Walsh A. 1987. "Teaching understanding and interpretation of logit regression," *Teaching Sociology* 15: 178-183.

Received February, 1999; Revised January, 2000

Appendix
Computation of Parameters for the Log-Frequency and Log-Rate Saturated Models

The computations in Table 1 are described in this Appendix. Each cell in the three-way table is denoted by an appropriate combination of subscripts, the first representing year, the second region, and the third marital duration. To illustrate, $freq_{111}$ in the log-frequency model symbolizes the frequency count in the intersection of Year 1971, Region1 and Duration of Marriage <10; and $rate_{111,}$ in the log-rate model represents the corresponding rate for this intersection.

Log-frequency

$\lambda = \log(freq_{111}) + \log(freq_{112}) + \log(freq_{121}) + \log(freq_{122})$
$+ \log(freq_{211}) + \log(freq_{212}) + \log(freq_{221}) + \log(freq_{222})/rc$
(where r = rows, c = columns; in this case rc = 2x2x2=8.
$= [\log(6943) + \log(11874) + \log(4920) + \log(5923) + \log(24382)$
$+ \log(11292) + \log(29468) + \log(11609)]/8 = [8.845489$
$+ 9.382106 + 8.501064 + 8.6865984 + 10.1016$
$+ 9.33185 + 10.29106 + 9.359536]/8$
$= 9.3124$

$\lambda^{1971} = [\log(freq_{111}) + \log(freq_{112}) + \log(freq_{121}) + \log(freq_{122})]/4$
$- \lambda = [8.845489 + 9.382106 + 8.501064$
$+ 8.6865984]/4 - 9.3124$
$= -0.4586$

$\lambda^{1995} = 0.4586$

$\lambda^{\text{Region 1}} = [\log(freq_{111}) + \log(freq_{112}) + \log(freq_{211}) + \log(freq_{212})]/4$
$- \lambda = [8.845489 + 9.382106 + 10.1016 + 10.29106]/4$
$- 9.3124$
$= 0.3427$

$\lambda^{\text{Region 2}} = -0.3427$

$\lambda^{\text{Duration <10}} = [\log(freq_{111}) + \log(freq_{121}) + \log(freq_{211}) + \log(freq_{221})]/4$
$- \lambda = [8.845489 + 8.501064 + 10.1016$
$+ 9.33185]/4 - 9.3124$
$= -0.1174$

$\lambda^{\text{Duration 10+}} = 0.1174$

Appendix (continued)

$$\lambda^{1971 \times \text{Region } 1} = [\log(freq_{111}) + \log(freq_{112})]/2 - [(\log(freq_{111}) + \log(freq_{112}) + \log(freq_{121}) + \log(freq_{122})]/4 - [(\log(freq_{111}) + \log(freq_{112}) + \log(freq_{211}) + \log(freq_{212})]/4 + \lambda = [8.845489 + 9.382106]/2 - [8.845489 + 9.382106 + 8.501064 + 8.6865984]/4 - [8.845489 + 9.382106 + 10.1016 + 10.29106]/4 + 9.3124 = -0.0827$$

$$\lambda^{1995 \times \text{Region } 2} = 0.0827$$

$$\lambda^{1971 \times \text{Duration} <10} = [\log(freq_{111}) + \log(freq_{121})]/2 - [\log(freq_{111}) + \log(freq_{112}) + \log(freq_{121}) + \log(freq_{122})]/4 - [(\log(freq_{111}) + \log(freq_{112}) + \log(freq_{211}) + \log(freq_{212})]/4 + \lambda = [8.845489 + 8.501064]/2 - [8.845489 + 9.382106 + 8.501064 + 8.6865984]/4 - [8.845489 + 8.501064 + 10.1016 + 9.33185]/4 + 9.3124 = -0.0631$$

$$\lambda^{1995 \times \text{Duration } 10+} = 0.0631$$

$$\lambda^{\text{Region } 1 \times \text{Duration} <10} = [\log(freq_{111}) + \log(freq_{211})]/2 - [\log(freq_{111}) + \log(freq_{112}) + \log(freq_{211}) + \log(freq_{212})]/4 - [(\log(freq_{111}) + \log(freq_{121}) + \log(freq_{211}) + \log(freq_{221})]/4 + \lambda = [8.845489 + 10.1016]/2 - [8.845489 + 9.382106 + 10.1016 + 10.29106]/4 - [8.845489 + 8.501064 + 10.1016 + 9.33185]/4 + 9.3124 = -0.0641$$

$$\lambda^{\text{Region } 2 \times \text{Duration } 10+} = 0.0641$$

$$\lambda^{1971 \times \text{Region } 1 \times \text{Duration} <10}$$

$$= \log(freq_{111}) - [\lambda + \lambda^{Y} + \lambda^{R} + \lambda^{D} + \lambda^{YR} + \lambda^{RD} + \lambda^{YD}] = 8.845489 - [9.3124 + -0.4586 + 0.3427 + -0.1174 + -0.0827 + -0.0631 + -0.0641] = -0.0237$$

$$\lambda^{1995 \times \text{Region } 2 \times \text{Duration } 10+} = 0.0237$$

Appendix (continued)

Log-rate

$\delta =$ $\log(\text{rate}_{111}) + \log(\text{rate}_{112}) + \log(\text{rate}_{121}) + \log(\text{rate}_{122})$
$+ \log(\text{rate}_{211}) + \log(\text{rate}_{212}) + \log(\text{rate}_{221}) + \log(\text{rate}_{222})/rc$
(where r = rows, c = columns; in this case rc = 2x2x2=8.
$= [\log(6943/1166519) + \log(11874/2533221)$
$+ \log(4920/432803) + \log(5923/966066)$
$+ \log(24382/1175428) + \log(11292/564701)$
$+ \log(29468/3702440) + \log(11609/1487965)]/8$
$= [-5.1241 + -5.3629 + -4.47697 + -5.09439$
$+ -3.87554 + -3.9122 + -4.83344 + -4.85338]/8$
$= -4.6916$

$\delta^{\text{Year 1971}} = [(\log(\text{rate}_{111}) + \log(\text{rate}_{112}) + \log(\text{rate}_{121}) + \log(\text{rate}_{122})]/4$
$- \delta = [-5.12405 + -5.3629 + -4.47697 + -5.09439]/4$
$- -4.6916$
$= -0.3230$

$\delta^{\text{Year 1995}} = 0.3230$

$\delta^{\text{Region 1}} = [(\log(\text{rate}_{111}) + \log(\text{rate}_{112}) + \log(\text{rate}_{211}) + \log(\text{rate}_{212})]/4$
$- \delta = [-5.12405 +-5.3629 + -3.87554 +-4.83344]/4$
$- -4.691609$
$= -0.1074$

$\delta^{\text{Region 2}} = 0.1074$

$\delta^{\text{Duration <10}}$ $= [\log(\text{rate}_{111}) + \log(\text{rate}_{121}) + \log(\text{rate}_{211})$
$+ \log(\text{rate}_{221})]/4 - \delta$
$= [-5.12405 + -4.47697 + -3.87554 + -3.9122]/4$
$- -4.691609$
$= 0.3444$

$\delta^{\text{Duration 10+}} = -3444$

$\delta^{1971 + \text{Region 1}}$ $= [\log(\text{rate}_{111}) + \log(\text{rate}_{112})]/2 - [(\log(\text{rate}_{111})$
$+ \log(\text{rate}_{112}) + \log(\text{rate}_{121}) + \log(\text{rate}_{122})]/4$
$- [(\log(\text{rate}_{111}) + \log(\text{rate}_{112}) + \log(\text{rate}_{211})$
$+ \log(\text{rate}_{212})]/4 + \delta$
$= [-5.12405 + -5.3629]/2 - [-5.12405$
$+ -5.3629 + -4.47697 + -5.09439]/4$
$- [-5.12405 + -5.3629 + -3.87554 + -4.83344]/4$
$+ -4.691609$
$= -0.1215$

Appendix (continued)

$\delta^{1995 \times \text{Region } 2} = 0.1215$

$\delta^{1971 \times \text{Duration} <10} = [\log(\text{rate}_{111}) + \log(\text{rate}_{121})]/2 - [\log(\text{rate}_{111})$
$+ \log(\text{rate}_{112}) + \log(\text{rate}_{121}) + \log(\text{rate}_{122})]/4$
$- [\log(\text{rate}_{111}) + \log(\text{rate}_{121}) + \log(\text{rate}_{211})$
$+ \log(\text{rate}_{221})]/4 + \delta$
$= [-5.12405 + -4.47697]/2 - [-5.12405$
$+ -5.3629 + -4.47697 + -5.09439]/4$
$- [-5.12405 + -4.47697 + -3.87554 + -3.9122]/4$
$+ -4.691609$
$= -0.1304$

$\delta^{1995 \times \text{Duration } 10+} = 0.1304$

$\delta^{\text{Region } 1 \times \text{Duration} <10} = [\log(\text{rate}_{111}) + \log(\text{rate}_{211})]/2 - [\log(\text{rate}_{111})$
$+ \log(\text{rate}_{112}) + \log(\text{rate}_{211}) + \log(\text{rate}_{212})]/4$
$- (\log(\text{rate}_{111}) + \log(\text{rate}_{121}) + \log(\text{rate}_{211})$
$+ \log(\text{rate}_{221})]/4 + \delta = [-5.12405 + -3.87554]/2$
$- [-5.12405 + -5.3629 + -3.87554 + -4.83344]/4$
$- [-5.12405 + -4.47697 + -3.87554 + -3.9122]/4$
$+ -4.691609$
$= -0.0452$

$\delta^{\text{Region } 2 \times \text{Duration } 10+} = 0.0452$

$\delta^{1971 \times \text{Region } 1 \times \text{Duration} <10}$

$= \log(\text{rate}_{111})/1 - [\delta + \delta^{Y} + \delta^{R} + \delta^{D} + \delta^{YR} + \delta^{RD} + \delta^{YD}]$
$= -5.12405 - [-4.691609 + -2.50686 + -0.10072 + -2.09729$
$+ -5.12945 + -4.23506 + -8.59002]$
$= -0.0494$

$\delta^{1995 \times \text{Region } 2 \times \text{Duration } 10+} = 0.0494$