Professional Geographer, 41(2), 1989, pp. 190–198

ANALYSIS OF COUNT DATA USING POISSON REGRESSION*

Andrew Lovett and Robin Flowerdew

University of Lancaster

In geographical research the data of interest are often in the form of counts. Standard regression analysis is inappropriate for such data, but if certain assumptions are met, a form of regression based on the Poisson distribution can be used. This paper illustrates the use of Poisson regression in the computer package GLIM with an example from historical geography. Apprentice migration to Edinburgh is regressed on a combination of categorical, count, and continuous explanatory variables. **Key Words: Poisson regression, count data, migration, GLIM, Edinburgh.**

Many geographers receive a basic training in statistical analysis but lack the opportunity or desire to keep up to date with developments in statistical methods that may be useful in contexts reasonably common in geographical research. This paper describes one such development, Poisson regression analysis, in sufficient detail that interested researchers with moderate knowledge of statistical applications in geography can learn to use it for themselves.

Regression analysis constructs an explanatory or predictive model of a dependent or response variable (usually denoted Y) on the basis of one or more independent or explanatory variables (denoted X_1, X_2, etc.). Poisson regression is appropriate when the dependent variable is a count, such as the number of times an event occurs or the number of people in a certain category. It may be particularly useful if some observations have very low values. The method is still not very widely known in geography, or indeed the social or environmental sciences generally. It does not yet appear, for example, in any textbook of statistical methods in geography. A growing number of applications exist in the journal literature, however. Dependent variables which have been analyzed using Poisson regression include the number of migrants from one city to another (Constantine and Gower 1982; Flowerdew and Aitkin 1982; Flowerdew and Lovett 1988), airline passengers from one city to another (Fotheringham and Williams 1983), origins of US college students (Pickles 1985), interregional industrial movements in Britain (Twomey 1986), radiation exposure categories (Darby et al. 1985), plant species on islands (Vincent and Haworth 1983), family planning clinics in Nigeria (Senior 1987), shopping trips (Guy 1987), children immunized against whooping cough (Gatrell 1986), heart disease mortality (Lovett et al. 1986), spina bifida (Lovett and Gatrell 1988), looted stores in New York City and national newspapers of the world (Flowerdew and Lovett 1984). Lovett (1984) has produced a detailed guide to the technique as implemented in the GLIM statistical package.

In the first section of this paper, the characteristics of Poisson regression are introduced by contrasting them with the standard ordinary least squares (OLS) form of regression analysis. An outline of the commands needed to carry out Poisson

* This work was done under a grant from the Economic and Social Research Council. The manuscript was prepared while the second author was working with the Institute for Market and Social Analysis, Toronto. We would like to thank Ian Whyte for supplying the data, and present and former members of the Centre for Applied Statistics, University of Lancaster, especially Murray Aitkin, Brian Francis, John Hinde, and Andrew Wood for statistical help.

regression in GLIM follows. Finally an illustration of how these can be linked together is provided in the third section through an analysis of a data set from historical geography.

Poisson Regression and the Generalized Linear Model

Linear regression analysis assumes that the dependent variable has a linear relationship with the independent variables. If there is only one independent variable X, the predicted value of the dependent variable Y for case i is given by the equation

$$\hat{y}_i = \beta_0 + \beta_1 x_i \tag{1}$$

where $\hat{y}_i$ is the predicted value of Y for case i, x_i is the value of X for case i, β_0 is the intercept—the value of $\hat{y}$ when X is zero, and β_1 is the slope—the amount by which $\hat{y}$ changes as X increases by one unit. The observed value y_i corresponding to x_i can be envisaged as a realization of a random variable Y_i whose mean is estimated to be $\hat{y}_i$. In OLS regression, the Y_i variables are assumed to have normal distributions and a constant variance.

OLS regression has been used to analyze count dependent variables, but its assumptions are not valid in this context. The values of the y_is are counts and so must be non-negative integers; therefore the random variables generating them cannot have a normal distribution, which is continuous and can take on negative values. If the counts are large, the normal distribution may give a reasonable approximation to the true discrete distribution; but when the counts have small values, the use of OLS regression is inappropriate and the results obtained are unreliable.

A type of regression analysis more appropriate for small counts is one in which each Y_i variable is assumed to have a Poisson distribution (Frome et al. 1973). The Poisson distribution describes the probability that an event occurs k times in a fixed period given that each occurrence is independent and has a constant probability. The shape of a Poisson distribution depends on the value of its mean (which is equal to its variance). If the mean is close to zero, then the distribution is strongly skewed; if the mean is larger, the peak occurs further from the vertical axis. If the mean is very large, the Poisson distribution can be approximated by the normal.

In a Poisson regression model the predicted value of the dependent variable for case i is the maximum likelihood estimate $\hat{\lambda}_i$ of the mean of a Poisson-distributed variable Y_i. The natural logarithm of this estimate is equal to a linear combination of the corresponding values of the independent variables; if there is only one independent variable the relationship has the form

$$\ln(\hat{\lambda}_i) = \beta_0 + \beta_1 x_i \tag{2}$$

This equation can also be written

$$\hat{\lambda}_i = \exp(\beta_0 + \beta_1 x_i) \tag{3}$$

This is the Poisson regression equivalent of the OLS regression equation (1).

Two characteristics of the Poisson regression model are worthy of further discussion. Unlike OLS regression, the Poisson model does not assume that the data are homoscedastic. Indeed, the variance of each case is equal to the corresponding predicted value. Consequently the variances associated with the values of Y_i cannot be the same and cannot be normally distributed. Second, the observed value of Y_i is a count of independent events generated by a Poisson distribution with parameter λ_i. This means that the model is not appropriate where the occurrence of one event increases the probability of others. For example, modeling the incidence of a contagious disease would not be appropriate.

The OLS and Poisson regression models can be considered as variants of what Nelder and Wedderburn (1972) termed generalized linear models, explained more fully in Appendix A of the GLIM User's Guide (Payne 1985) and by McCullagh and Nelder (1983). Generalized linear models can be used whenever an analyst wants to estimate the value of a dependent variable Y on the basis of a linear predictor (a linear combination of explanatory variables X_1, X_2, etc.). An observed value y_i is regarded as the sum of a systematic component (μ_i) and a random (or error) component (ϵ_i):

$$y_i = \mu_i + \epsilon_i \qquad (4)$$

where y_i is the observed value of Y for case i, μ_i is the predicted value of Y for case i, and ϵ_i is a randomly distributed error term. Types of generalized linear models are differentiated on the basis of the probability distribution that Y_i is assumed to have and the link function g that associates the mean of Y_i (μ_i) with the linear predictor

$$g(\mu_i) = \sum_{j=1}^{m} \beta_j x_{ij} \qquad (5)$$

where m is the number of parameters, β_j is the parameter for the j^{th} independent variable, and x_{ij} is the value of the i^{th} observation on the j^{th} independent variable. Equation (4) can be rewritten

$$y_i = g^{-1}\left(\sum_{j=1}^{m} \beta_j x_{ij}\right) + \epsilon_i \qquad (6)$$

A range of linear models can be constructed by specifying different types of error term ϵ_i and link function g. Options for the link function include identity, logarithmic, reciprocal, and square root, among others. The error distribution can be any member of the exponential family of probability distributions, including such distributions as the normal, binomial, Poisson, and gamma.

In OLS regression the error distribution is normal. The predicted value $\hat{y}_i$ is equal to $\sum_{j=1}^{m} \beta_j x_{ij}$, the linear predictor, so the link is simply the identity function. In Poisson regression, the error distribution is Poisson, but it is the natural logarithm of the predicted value, not $\hat{\lambda}_i$ itself, which is equal to the linear predictor, so the link function is logarithmic. Log-linear modeling in GLIM (Bowlby and Silk 1982; Wrigley 1985) also uses the Poisson probability model and logarithmic link. Indeed, it can be regarded as a special case of Poisson regression when all the independent variables are categorical.

The concept of a generalized linear model is the basis of the computer package GLIM which provides a relatively easy means of performing Poisson regression. (It is also available in other packages, including GENSTAT and LIMDEP.) Poisson regression is not available in most of the other major statistical packages, such as SAS, SPSS-X, SYSTAT, or BMDP. This paper, therefore, focuses on the implementation of Poisson regression in GLIM.

Poisson Regression in GLIM

GLIM was developed by the Numerical Algorithms Group in conjunction with the Working Party on Statistical Computing of the Royal Statistical Society (O'Brien 1987). In addition to the mainframe version, a microcomputer version is available relatively cheaply from the Numerical Algorithms Group in Oxford or in Downers Grove, IL. Payne (1985) provides a detailed guide, and several articles have been published describing how to use GLIM for particular purposes. Examples by geographers include Bowlby and Silk (1982) and O'Brien (1983).

GLIM allows the user to specify the error distribution and link function to be used in the regression. Once the data have been entered, any necessary transformations can be done on the variables, and the data can be plotted. The example in the next section illustrates these operations. The main process involved in fitting the model, however, is the selection of independent variables for inclusion.

The results of fitting a model can be evaluated according to the value of scaled deviance, which is a measure of the overall difference between the observed values of Y and those predicted by the model. Deviance is computed in different ways according to the probability distribution in use, and is based on the log-likelihood ratio. In OLS regression, deviance is equal to the error (or residual) sum of squares.

In Poisson regression (where it is sometimes denoted G^2), the scaled deviance (d) is

$$d = 2 \sum_{i=1}^{n} y_i \ln(y_i/\hat{\lambda}_i) \qquad (7)$$

The deviance is large when the correspondence between observed and estimated values is poor, and small when it is close.

Because the deviance of a Poisson model has a distribution approximating to chi-square, it is possible to assess whether the estimated and observed values of Y are significantly different by comparing the calculated deviance with the critical value of chi-square for the appropriate degrees of freedom and significance level. The number of degrees of freedom is equal to the number of cases minus the number of parameters fitted. If the deviance exceeds the critical value, the null hypothesis that the model could have generated the data should be rejected, and the model should not be regarded as providing an adequate explanation (at least in statistical terms) of the variation in the values of the dependent variable. However, little is known about how good the approximation to chi-square is for small sets of data (Payne 1985, 111), and this test should be regarded only as a general guide in assessing goodness of fit.

A regression model can be evaluated from several different perspectives. One of these is goodness-of-fit, already mentioned. Another aspect is the significance of the parameters, which can be assessed using t values, obtained by dividing each parameter estimate by its standard error. A third consideration is interpretability, which can best be evaluated by seeing if the sign and magnitude of parameter estimates accord with theoretical expectations. Ideally a regression model (Poisson or otherwise) should have a good fit, significant parameters, and a clear interpretation. In practice, however, these are not always attainable simultaneously, and the best compromise must be chosen. Therefore it is sensible to fit models according to a systematic strategy, particularly because the significance of individual parameters is likely to vary between models depending on the other variables included (especially if they are collinear).

One strategy begins by assessing the overall variation in the dependent variable by fitting a null model, in which the only parameter is an intercept term (equal to the mean of Y). The deviance of this model can be used as a benchmark to assess the effect of adding independent variables. The importance of each variable added in turn to the null model can be judged by comparing the reduction in deviance with the critical value of chi-square. Introducing numerical variables involves the loss of only one degree of freedom, and so a decline in deviance of more than 3.84 indicates that the variable is significant at the 0.05 level. The same procedure can also be used to test the significance of factors and interaction terms, though here the degrees of freedom lost will depend on the number of categories of the factor(s) concerned.

This procedure can be used to find which variable is the most successful in reducing the deviance. Others are then added one by one in the order defined by the magnitude of their effect on deviance. At each stage the estimates and standard errors are examined to remove any variables whose parameters are insignificant. New variables added to the model can alter the significance of others, so that variables which initially seemed significant may drop out, and others, not significant originally, may become so. This procedure is complete when a significant reduction in deviance can no longer be obtained. At this stage, those variables and interaction terms not in the model should be reintroduced to test whether they are now significant. Once this process is completed, the estimates and residuals can be displayed and the overall adequacy of the model evaluated.

Application of the Model

An analysis of apprentice migration to Edinburgh between 1775 and 1799 is chosen to demonstrate the discussion above with a real and reasonably small data set

TABLE 1
DATA USED IN THE APPRENTICE MIGRATION EXAMPLE

D	A	POP	U	R	County
21	225	56 000	18.8	3	Midlothian
24	22	18 000	37.9	2	West Lothian
33	44	30 000	43.4	3	East Lothian
33	3	7000	30.3	1	Kinross
36	41	94 000	41.3	1	Fife
41	9	9000	29.3	3	Peebles
41	2	11 000	47.4	1	Clackmannan
52	5	5000	41.9	3	Selkirk
54	23	147 000	68.1	2	Lanark
56	11	31 000	15.2	3	Berwick
67	9	34 000	31.8	3	Roxburgh
71	13	51 000	31.1	2	Stirling
78	26	126 000	14.4	1	Perth
79	0	21 000	27.3	2	Dunbarton
85	5	99 000	55.3	1	Angus
86	3	55 000	25.9	3	Dumfries
92	1	78 000	69.9	2	Renfrew
110	2	84 000	26.4	2	Ayr
110	0	29 000	11.3	3	Kirkcud-bright
125	1	26 000	12.3	1	Kincardine
132	0	12 000	43.6	2	Bute
156	3	123 000	23.1	1	Aberdeen
157	0	22 000	23.2	3	Wigtown
159	4	36 000	12.9	1	Banff
174	2	27 000	27.6	1	Moray
175	0	8000	45.3	1	Nairn
179	1	72 000	12.7	2	Argyll
234	7	74 000	10.8	1	Inverness
274	4	55 000	10.7	1	Ross
274	1	23 000	28.3	1	Caithness
283	0	23 000	10.3	1	Sutherland
366	0	29 000	9.0	1	Orkney
491	1	22 000	7.7	1	Shetland

D = distance, A = number of apprentices, U = degree of urbanization, R = categorical variable (direction of county from Edinburgh). The data set should not include column headings when it is read into GLIM. Source: Lovett 1984.

(Lovett et al. 1985; Whyte and Whyte 1986). The dependent variable (A) is the number of apprentices originating from each of the 33 Scottish counties registered in Edinburgh between 1775 and 1799. Apprentice migration from each county was expected to be positively related to its population (POP), as recorded in the 1801 census, and negatively related to its distance (D) and its degree of urbanization (U), measured by the percentage of the county population living in urban settlements. These relationships might vary regionally because of differences in the ease of travel and the nature of intervening opportunities. A three-level categorical variable (R) was therefore included, defined according to whether the county was north, west or south of Edinburgh.

The following is an annotated transcript of a Poisson regression analysis of the data in GLIM 3.77 on a VAX 11/780 running under VMS. Input is indicated by upper case, output by lower case. The following commands read the data from a file called EDIN.DAT (Table 1) on channel (or unit) 10.

```
$UNITS 33 $
$DATA D A POP U R $
$FORMAT (F4.0,F4.0,F7.0,F5.1,F2.0) $
$DINPUT 10 $
File name? EDIN.DAT
```

Population and distance variables are logarithmically transformed, according to usual gravity model practice. The logarithmic link function used in Poisson regression makes it unnecessary to transform the number of migrants.

```
$CALC LP = %LOG(POP) $
$CALC LD = %LOG(D) $
```

The regional variable R must be specified as a factor, and A must be identified as the dependent variable.

```
$FACTOR R 3 $
$YVAR A $
$ERROR POISSON $
$LINK LOG $
```

(The commands $ERROR NORMAL $ and $LINK IDENTITY $ would be used at this point if OLS regression was to be done.) Now the null model can be fitted, followed by each independent variable in turn.

```
$FIT $
scaled deviance = 1350.4 at cycle 5
            d.f. =   32
$FIT LP $
scaled deviance = 1217.8 at cycle 5
            d.f. =   31
```

```
$FIT LD $
scaled deviance =  393.96 at cycle 4
            d.f. =   31
$FIT U $
scaled deviance = 1348.9 at cycle 6
            d.f. =   31
$FIT R $
scaled deviance = 1054.1 at cycle 5
            d.f. =   30
```

The cycle number indicates how many iterations were required for the solution.

The null model deviance is fairly large, and log distance is the best explanatory variable, followed by region. They can be tried together.

```
$FIT LD + R $
scaled deviance = 329.60 at cycle 5
            d.f. = 29
```

The deviance is reduced by 64 for the loss of two degrees of freedom, so it is worth retaining LD and R together in the model. Next the other variables can be added to the model. In the following commands, the + indicates that the new model consists of those variables in the previous model plus the additional one specified.

```
$FIT + LP $
scaled deviance =    97.91 at cycle 4
        (change = -231.7)
            d.f. =   28
        (change =   -1)
$FIT + U $
scaled deviance =   80.524 at cycle 4
        (change =  -17.390)
            d.f. =   27
        (change =   -1)
```

Both variables produce significant deviance reductions, but the overall deviance is still well over 40.11, which is the critical chi-square value at 0.05 for 27 degrees of freedom.

Before interaction terms are introduced into the model, the parameter estimates and standard errors can be examined.

```
$DISPLAY E $
```

	estimate	s.e.	parameter
1	0.7346	0.9885	1
2	−2.055	0.09349	LD
3	−0.1016	0.1726	R(2)
4	0.5044	0.1477	R(3)
5	0.9822	0.07816	LP
6	−0.01705	0.004141	U

scale parameter taken as 1.000

The low standard errors of LD, LP and U confirm their significance in accounting for variation in A. The parameter estimate for LP is near 1, as the gravity model would suggest. The value for LD is negative as expected, as is that for urbanization. The values shown for R(2) and R(3) are differences from the intercept; thus region 3 (the south) actually has the highest parameter value, with region 1 slightly in excess of region 2. Wrigley (1985) provides more details about interpreting this type of GLIM output.

Now the interaction of region with the other variables can be investigated. Adding the interaction between a factor (R) and a continuous variable (LP) is equivalent to estimating a separate parameter for each level of the factor.

```
$FIT + R.LP $
scaled deviance = 76.951 at cycle 4
        (change = -3.57)
            d.f. =   25
        (change = -2)
```

This is not significant, so it can be omitted from the model and the other interaction terms fitted.

```
$FIT - R.LP + R.LD $
scaled deviance =  51.836 at cycle 4
        (change = -25.11)
            d.f. =  25
        (change =  -0)
$FIT + R.U $
scaled deviance =  40.919 at cycle 4
        (change = -10.92)
            d.f. =  23
        (change =  -2)
```

TABLE 2
PARAMETER ESTIMATES AND STANDARD ERRORS DISPLAYED BY GLIM

	Estimate	SE	Parameter[a]
1	−1.623	1.949	1
2	−1.549	0.2013	LD
3	−3.035	3.837	R(2)
4	8.657	3.164	R(3)
5	1.032	0.1386	LP
6	−0.02619	0.008928	U
7	−2.049	0.8342	LD.R(2)
8	−1.316	0.2999	LD.R(3)
9	1.144	0.6820	LP.R(2)
10	−0.3891	0.2305	LP.R(3)
11	−0.02834	0.02682	U.R(2)
12	0.03025	0.01206	U.R(3)

[a] Scale parameter taken as 1.000.

Both terms produce significant reductions in deviance when added to the model, and the overall value is approaching an acceptable fit (the critical value of chi-square at 0.01 is 41.64).

Although the interaction term R.LP was not significant earlier, it may be now that the other interaction terms have been added.

```
$FIT + R.LP $
scaled deviance = 32.642 at cycle 4
        (change = −8.277)
            d.f. = 21
        (change = −2)
```

This term is now significant, and the deviance is now below the critical value of chi-square at 0.05. This model could therefore have generated the observed data; the null hypothesis that apprentice migration is determined by these variables cannot be rejected on the basis of this data set. The current set of parameter estimates can now be requested with the $DISPLAY E $ command (Table 2). Apprentice migration to Edinburgh is positively associated with origin population, and negatively with distance and urbanization. There are also regional variations; for example, the distance-decay gradient is shallower to the north and steeper to the west.

The next step in the analysis might be to display the residuals, and to examine them to see if any observations are out of line with the others. As with other forms of regression analysis, the spatial pattern of residuals can often give the investigator ideas for other variables which may be of relevance.

In general, the analysis has been satisfactory (not always the case!). An acceptable fit has been achieved, and all the parameter values are in accordance with theoretical expectations. The $STOP command concludes the GLIM run.

Conclusion

This paper has presented the basics of Poisson regression analysis from a practical point of view. The example shows how a researcher may investigate a series of models with speed and flexibility. Not all analyses will produce a model with an adequate fit, especially if data for important explanatory variables are not available. Even if a good fit is not achieved, however, the results will allow an evaluation of which variables contribute the most to reducing the deviance. It must be remembered, of course, that any variable or any model, regardless of its goodness of fit, may have no causal link at all to the dependent variable, and may be just a surrogate for other variables.

The applicability of Poisson regression analysis depends on the assumption that the counts which constitute the dependent variable are generated by a Poisson distribution, which is true only if the individual events making up the counts are independent of one another. In our example, we assume that apprentices acted independently in making their moves to Edinburgh. This assumption is probably not entirely true, as apprentices may have been preceded or accompanied by relatives or friends from the same origin. In other cases, such as overall interurban migration (Flowerdew and Aitkin 1982), the independence assumption is certainly not true, for families are likely to migrate together. In such a case, some form of generalized or compound Poisson model may be appropriate (Flowerdew and Lovett 1989; Hinde 1982); fitting such a model is less straightforward and cannot be done in GLIM without special macro procedures. If a Poisson regression of a compound or generalized Poisson dependent variable is conducted, the deviance can be expected to be considerably greater than for a Poisson variable. However, Davies and Guy (1987) have shown that, for a large class of compound and generalized Poisson distributions, the parameters derived from Poisson regression are unbiased estimates of the parameters of the compound or generalized process.

Poisson regression analysis is a method developed fairly recently and as yet is little known in geography. Its use is appropriate in many contexts where the dependent variable is a count. The analysis is reasonably easy to apply, given suitable computer software such as GLIM, and geographers should make more use of it in the future.

Literature Cited

Bowlby, S., and J. Silk. 1982. Analysis of qualititative data using GLIM: Two examples based on shopping survey data. *The Professional Geographer* 34:80–90.

Constantine, A. G., and J. C. Gower. 1982. Models for the analysis of interregional migration. *Environment and Planning A* 14:477–97.

Darby, S., E. Nakashima, and H. Kato. 1985. A parallel analysis of cancer mortality among atomic bomb survivors and patients with ankylosing spondylitis given X-ray therapy. *Journal of the National Cancer Institute* 75:1–21.

Davies, R. B., and C. M. Guy. 1987. The statistical modeling of flow data when the Poisson assumption is violated. *Geographical Analysis* 19: 300–14.

Flowerdew, R., and M. Aitkin. 1982. A method of fitting the gravity model based on the Poisson distribution. *Journal of Regional Science* 22:191-202.

Flowerdew, R., and A. Lovett. 1984. Poisson regression models for the analysis of geographical data. End-of-grant Report to the Economic and Social Research Council, London.

Flowerdew, R., and A. Lovett. 1988. Fitting constrained Poisson regression models to inter-urban migration flows. *Geographical Analysis* 20:297-307.

Flowerdew, R., and A. Lovett. 1989. Compound and generalised Poisson models for inter-urban migration. In *Advances in Regional Demography: Forecasts, Information, Models,* ed. P. Congdon and P. Batey, 246–56. London: Belhaven Press.

Fotheringham, A. S., and P. A. Williams. 1983. Further discussion on the Poisson interaction model. *Geographical Analysis* 15:343–47.

Frome, E. L., M. H. Kutner, and J. J. Beauchamp. 1973. Regression analysis of Poisson-distribution data. *Journal of the American Statistical Association* 68:935–40.

Gatrell, A. C. 1986. Immunization against whooping cough in Salford: A spatial analysis. *Social Science and Medicine* 23:1027–32.

Guy, C. M. 1987. Recent advances in spatial interaction modelling: An application to the forecasting of shopping travel. *Environment and Planning A* 19:173–86.

Hinde, J. 1982. Compound Poisson regression models. In *GLIM 82,* ed. R. Gilchrist, 109–21. New York: Springer-Verlag.

Lovett, A. 1984. Poisson regression using the GLIM package. Lancaster: Department of Geography, University of Lancaster, Computer Package Guide No. 5.

Lovett, A., C. G. Bentham, and R. Flowerdew. 1986. Analysing geographic variations in mortality using Poisson regression: The example of ischaemic heart disease in England and Wales 1969–1973. *Social Science and Medicine* 23:935–43.

Lovett, A., and Gatrell, A. C. 1988. The geography of spina bifida in England and Wales. *Transactions, Institute of British Geographers* n.s. 13:288–302.

Lovett, A., I. D. Whyte, and K. A. Whyte. 1985. Poisson regression analysis and migration fields: The example of the apprenticeship records of Edinburgh in the seventeenth and eighteenth centuries. *Transactions, Institute of British Geographers* n.s. 10:317–32.

McCullagh, P., and J. A. Nelder. 1983. *Generalized Linear Models.* London: Chapman and Hall.

Nelder, J. A., and R. W. M. Wedderburn. 1972. Generalized linear models. *Journal, Royal Statistical Society A* 135:370–84.

O'Brien, L. G. 1983. Generalised linear modelling using the GLIM system. *Area* 15:327–35.

O'Brien, L. G. 1987. GLIM 3.77 (software review). *The Professional Geographer* 39:229-30.

Payne, C. D., ed. 1985. *The GLIM System Release*

3.77: *Manual*. Oxford: Numerical Algorithms Group.
Pickles, A. 1985. *An Introduction to Likelihood Analysis*. Concepts and Techniques in Modern Geography, No. 42. Norwich, UK: Geo Books.
Senior, M. L. 1987. The establishment of family planning clinics in south-west Nigeria by 1970: Analyses using logit and Poisson regression. *Area* 19:237-45.
Twomey, J. 1986. Establishment migration: an analytical framework. *Environment and Planning A* 18:967-79.
Vincent, P. J., and J. M. Haworth. 1983. Poisson regression models of species abundance. *Journal of Biogeography* 10:153-60.
Whyte, I. D., and K. A. Whyte. 1986. Patterns of migration of apprentices into Aberdeen and Inverness during the eighteenth and early nineteenth centuries. *Scottish Geographical Magazine* 102:81-92.
Wrigley, N. 1985. *Categorical Data Analysis for Geographers and Environmental Scientists*. London: Longman.

ANDREW LOVETT (B.A., University College of Wales, Aberystwyth) is Teaching Fellow, Department of Geography, University of Lancaster, Lancaster LA1 4YB, UK. His research interests are in medical geography and industrial linkage studies. ROBIN FLOWERDEW (Ph.D., Northwestern University) is Lecturer in Geography, University of Lancaster. His research interests are in migration modeling and geographical information systems.

OPEN FORUM

Professional Geographer, 41(2), 1989, pp. 198-203

THE HANFORD ENVIRONMENTAL DOSE RECONSTRUCTION PROJECT

Richard Morrill
University of Washington

A five year, $15 million study is underway to estimate radiation doses that people might have received from operations at the Hanford, WA, nuclear reservation. The study is funded by the US Department of Energy (DOE), and the research is conducted by Battelle's Pacific Northwest Laboratory, the current main research contractor at Hanford. The work is directed by an independent Technical Steering Panel (TSP) on which I serve as the demographer. Despite the intensely geographic nature of the problem, the possibility of a geographer as such as a panel member was never raised; we have a long way to go. This report provides some background on the issues surrounding the Hanford controversy, with the hope that readers with expertise in the area or topic will contact me.

The Hanford Reservation

Hanford is one of the nation's major nuclear production and research reservations. As many as nine primary production reactors have existed at various times. Eight water-cooled reactors along the Columbia River have been the source of large