Loglinear Analysis of Family Planning in Fiji

Jared Cummings

December 2021

Abstract

Abstract will go here.

Contents

Introduction

Analyses of the cell counts of contingency tables are a prominent application of generalized linear models (GLMs). Such analyses are done with a log link function and a Poisson response, modeling the expected count as a function of the categorical variables that uniquely identify that cell. In this report a loglinear model is fit on the responses from part of the Fijian stage of the World Fertility Survey (WFS), the Fiji Fertility Survey, that was run in 1974 and reported by Little and Pullum (1979). By modeling the cell counts, inference on cell probabilities of the unknown population's contingency table is made accessible. It is this population demographics that are the quantity of interest in this analysis. Statistical significance will be assessed at the 0.05 level.

Understanding contraceptive use in a population is important for projecting the fertility rate of a group. This has obvious applications for city and national planning, projecting taxable funds, draft response rates, and other critical civic measures. This information can also inform policy decisions. It may be ill-advised to invest heavily in expanding childhood education in an area that is seeing a drop in the birthrate.

Data

Little and Pullum (1979) reported some of the summarized responses of the Fijian stage of the WFS, the Fiji Fertility Survey, run in 1974. Because there are nontrivial behavioral disparities between ethnic groups in Fiji, these responses are specific to Indian women living in the country. As such, conclusions drawn from the data should be understood in this specific context. Of interest to this analysis is the part of the survey where the women reported ever having used any efficient contraceptive method. The data as reported by Little and Pullum is given in Table 1.

Table 1: Data as reported by Little and Pullum. Age is a factor with four levels, education, desire for children (wants_more) and contraceptive usage (reported here in the not_using and using columns) all have two levels. The contingency table that the loglinear model will fit on will then be a 4x2x2x2 table.

Age	Education	Desires more children	Never used contraceptives	Used contraceptives
<25	low	yes	53	6
<25	low	no	10	4
<25	high	yes	212	52
<25	high	no	50	10
25-29	low	yes	60	14
25-29	low	no	19	10
25-29	high	yes	155	54
25-29	high	no	65	27
30-39	low	yes	112	33
30-39	low	no	77	80
30-39	high	yes	118	46
30-39	high	no	68	78
40-49	low	yes	35	6
40-49	low	no	46	48
40-49	high	yes	8	8
40-49	high	no	12	31

The full contingency table is depicted graphically in Figure 1. All of the variables visibly have an effect on count, so it is expected that a well-fitting model will have to incorporate some interaction terms. This figure also tells the story of public education access in Fiji. The women in the oldest age bucket came of age right before secondary education began to be formally organized. The women in the next age bucket down, the 30-39 group, came of age as secondary education was becoming widespread. Those in the youngest two buckets grew up with easy access to secondary school. It can be observed that the youngest two age buckets are dominated by highly educated women, the third by a transition period, and the last has a majority with low education.

A second phenomenon seen in Figure 1 is that women who are older seem more prone to having used efficient contraceptives at some point in their comparatively longer lives than younger women. Lines associated with not having used efficient contraceptives start at the top on the left-hand side of the plot and end at the bottom of the right-hand side. The same is true of wanting more children. Older women seem less prone than their younger counterparts to be interested in expanding their family size.

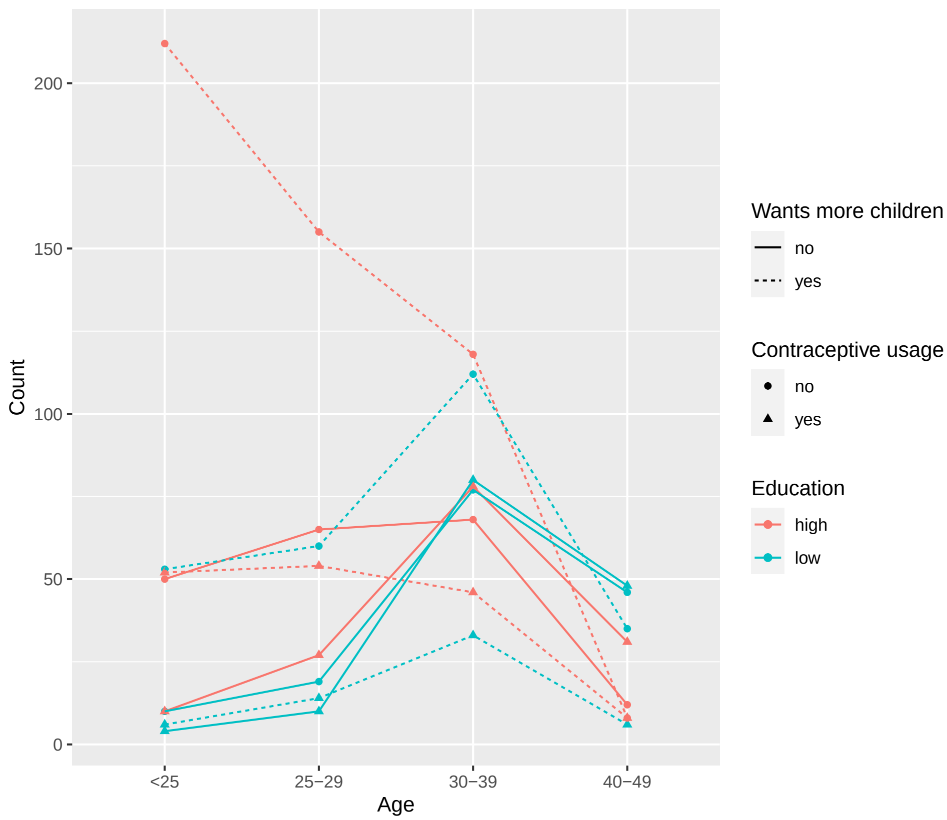

Figure 1: A graphical depiction of the 4x2x2x2 contingency table. On display are the effects of each variable on cell count. There are visible interactions at play between the factors. A well-fitting model will have to incorpoate terms that capture these interactions.

Methodology

As described in the introduction, a loglinear model is a special application of Poisson regression. Observed cell counts, n_{cell}, are assumed to be Poisson distributed. The expected cell count, μ_{cell} is modeled with the log link function–giving rise to the name "loglinear." Though other link functions exist in Poisson regression, such as the square-root transformation recommended by the Box-Cox proceedure, those are presupposed to provide less-optimal fits by using loglinear methods. Formally, the model can be defined

$$\begin{aligned} n_{\text{cell}} &\overset{ind.}{\sim} Pois(mu_{\text{cell}}) \\ \log mu_{\text{cell}} &= X\beta \end{aligned} \tag{1}$$

where $X\beta$ is a linear regression function of the factors that define the contingency table. For the Fiji WFS data these factors are `age`, `education`, `desire` for more children, and past `usage` of contraceptives. Since the explanatory variables are all factors in the loglinear model terms beyond the first-order are exclusively interactions.

This Poisson regression model defined in (1) makes fairly minimal assumptions about the data. Firstly, the data is assumed to follow the stated Poisson distribution. Secondly, the observed counts are assumed to be independent. Thirdly, monotonicity is assumed between the log expected counts and the regression equation. Some would include a fourth assumption here, that the cell counts in the contingency table being modeled are sufficiently large, though this is assumption only concerns itself with assessing statistical significance and not the propriety of fitting the model.

Results

Through the process of manual backward selection, beginning at a saturated model and walking backward, the linear regression component of the loglinear model chosen to model this data relationship was

$$\begin{aligned}
\log \mu_{ijkl} \quad = \quad & \beta_0 + \beta_{1,1} \cdot \mathbf{1}(\texttt{age}_k =' 25-29') + \beta_{1,2} \cdot \mathbf{1}(\texttt{age}_k =' 30-39') + \beta_{1,3} \cdot \mathbf{1}(\texttt{age}_k =' 40-49') + \\
& \beta_2 \cdot \mathbf{1}(\texttt{usage}_j = \text{using}) + \beta_3 \cdot \mathbf{1}(\texttt{education}_i = \text{low}) + \beta_4 \cdot \mathbf{1}(\texttt{desire}_l = \text{yes}) + \\
& \alpha_1 \cdot \mathbf{1}(\texttt{age}_k =' 25-29' \wedge \texttt{usage}_j = \text{yes}) + \alpha_2 \cdot \mathbf{1}(\texttt{age}_k =' 30-39' \texttt{usage}_j = \text{yes}) + \\
& \alpha_3 \cdot \mathbf{1}(\texttt{age}_k =' 40-49' \wedge \texttt{usage}_j = \text{yes}) + \alpha_4 \cdot \mathbf{1}(\texttt{age}_k =' 25-29' \wedge \texttt{education}_i = \text{low}) + \\
& \alpha_5 \cdot \mathbf{1}(\texttt{age}_k =' 30-39' \wedge \texttt{education}_i = \text{low}) + \alpha_6 \cdot \mathbf{1}(\texttt{age}_k =' 40-49' \wedge \texttt{education}_i = \text{low}) + \\
& \alpha_7 \cdot \mathbf{1}(\texttt{age}_k =' 25-29' \wedge \texttt{desire}_l = \text{yes}) + \alpha_8 \cdot \mathbf{1}(\texttt{age}_k =' 30-39' \wedge \texttt{desire}_l = \text{yes}) + \\
& \alpha_9 \cdot \mathbf{1}(\texttt{age}_k =' 40-49' \wedge \texttt{desire}_l = \text{yes}) + \alpha_{10} \cdot \mathbf{1}(\texttt{usage}_j = \text{yes} \wedge \texttt{education}_i = \text{low}) + \\
& \alpha_{11} \cdot \mathbf{1}(\texttt{usage}_j = \text{yes} \wedge \texttt{desire}_l = \text{yes})
\end{aligned} \tag{2}$$

where β coefficients are first-order effects of the variables and the α coefficients are first order interactions. Table 2 reports the estimated coefficients, their standard errors, and the associated z scores and p-values for the model defined in (2). Significance is determined at the 0.05 level. `R` code needed to reproduce the model and associated analysis is linked in the appendix.

Table 2: Parameter estimates, standard errors, z scores and p-values from the model defined in (2). Significance ('Sig.') is determined at a 0.05 cutoff. Of possible two-way interactions, all but an interaction between education and desire for more children were included in the model. The significance of interactions with age is not surprising after the discussion given in the Data section.

	Coeficient	Estimate	Std. Error	z score	Pr(>|z|)	Sig.
beta 0	Intercept	3.74	0.13	29.31	0.00	Y
beta 1,1	Age 25-29	0.27	0.16	1.63	0.10	N
beta 1,2	Age 30-39	0.57	0.15	3.77	0.00	Y
beta 1,3	Age 40-49	-0.89	0.21	-4.25	0.00	Y
beta 2	Used contraceptives	-0.82	0.16	-5.17	0.00	Y
beta 3	Low education	-1.44	0.13	-10.99	0.00	Y
beta 4	Wants more kids	1.65	0.13	12.38	0.00	Y
alpha 1	Age 25-29 : Used contraceptives	0.39	0.18	2.21	0.03	Y
alpha 2	Age 30-39 : Used contraceptives	0.90	0.16	5.48	0.00	Y
alpha 3	Age 40-49 : Used contraceptives	1.18	0.21	5.50	0.00	Y
alpha 4	Age 25-29 : Low education	0.44	0.17	2.55	0.01	Y
alpha 5	Age 30-39 : Low education	1.53	0.16	9.84	0.00	Y

	Coeficient	Estimate	Std. Error	z score	Pr(>|z|)	Sig.
alpha 6	Age 40-49 : Low education	2.42	0.21	11.63	0.00	Y
alpha 7	Age 25-29 : Wants more kids	-0.57	0.17	-3.32	0.00	Y
alpha 8	Age 30-39 : Wants more kids	-1.32	0.15	-8.51	0.00	Y
alpha 9	Age 40-49 : Wants more kids	-2.17	0.21	-10.49	0.00	Y
alpha 10	Used contraceptives : Low education	-0.30	0.12	-2.47	0.01	Y
alpha 11	Used contraceptives : Wants more kids	-0.82	0.12	-7.04	0.00	Y

Model fit

The model is a strong fit to the data. The Pearson χ^2 statistic was found to be 2.668 and the Neyman χ^2 statistic was found to be 2.674, both with 14 degrees of freedom. Both statistics support the conclusion that the model in (2) is not significantly different than the saturated model with both p-values practically equal to 1. Since both statistics support the same conclusion and are relatively similar in value, the cell counts in the contingency table being modeled are sufficiently large for statistical significance of the model and parameter estimates to be considered valid.

The assumptions of the model all appear to hold. The data being modeled is indeed Poisson counts. The survey design and execution does not prompt doubt about the independence assumption. The model in (2) does not contain any estimates of slope, only intercepts for various combinations of the factors in question, so the mononicity assumption is never in question.

Discussion

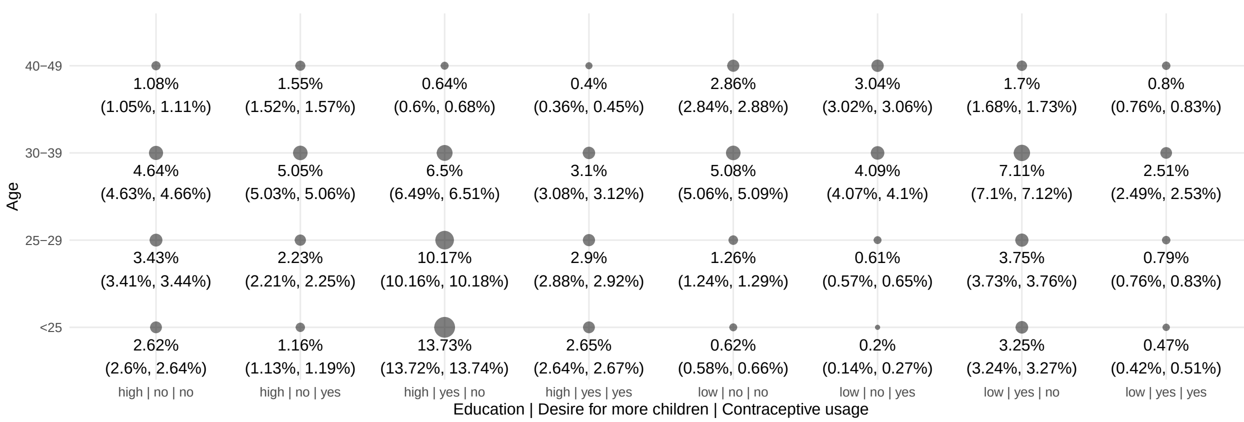

Figure 2: Expected cell count as percent of total count with 95% confidence intervals. A nontrivial segment of Fijian Indian women are both interested in having more children and have never taken any measures toward birth control.

Figure 2 displays the expected percent of Fijian Indian women who fall into each arrangement of factors with 95% confidence intervals. It is estimated that a large percentage of Fijian Indian women both are interested in having more children and have no history of contraceptive use (47.76% summing over age buckets and education levels). Of these women, the majority (66.90%) are highly educated. While the data used in this analysis does not include information on the number of children born to each woman, it is notable that birth control utilization is just shy of 50% in this population.

An opposite scenario are the active antinatalist contingent. These are women who are both not interested in having more children and have used birth control. Summing again over age buckets, we find that only 19.61% of women are in this group. This group breaks down evenly between education levels, with 50.94% being in the high education group.

The remaining third of Fijian Indian women are behaviors that do not align with their interest in family expansion. An estimated 21.69% of the total population do not want children and have never used efficient contraceptives. 54.52% of these women have a high education. This group does offer an indication that there may be a lack of access to birth control in this population. The remaining 11.04% do want more children but have a history of contraceptive use. This is the most imbalanced education segment, 81.97% being highly educated.

The combined percentage of women interested in expanding their family size is 58.80%. This number supports of a belief of a healthy, expanding population period. Women in their twenties constitute 64.13% of these women. Women in their thirties are the account for 32.69%. Women in their forties represent just 3.18%. The Indo-Fijian subpopulation, which at the time of the survey was the literal majority ethnic group in Fiji thanks to the influence of English colonists and heavy imigration, based on this data should continue to grow.

Conclusion

This analysis applies a loglinear model to the Fijian WFS responses. The findings of this model support the belief that the Indo-Fijian population was on a growth trajectory. It might be surprising then that this group has severely cut back their presence in Fiji. Successful coups against Indo-Fijian controlled governments were conducted in 1987 and 2000 were both racially motivated. The diaspora movements kicked off by both of these events dramatically reduced Indo-Fijian representation and was a significant brain-drain for Fiji. Mentioned before, insights need to be interpreted in the context of the population the survey was conducted among. These interpretations, including expectations of population growth, necessarily do not account for then-future political events that altered the course and construction of the Fijian nation.

Bibliography

Little, Roderick JA, and Thomas W Pullum. 1979. “The General Linear Model and Direct Standardization: A Comparison.” *Sociological Methods & Research* 7 (4): 475–501.

Appendix

The data from the WFS used in this analysis can be found here.