Analysis of Women's Family Size Preferences in Saki Area, South – Western Nigeria: A Log Linear Model Approach

Fasasi, S. K. & Osuntoki N. B.
Department of Statistics,
The Oke-Ogun Polytechnic,
Saki
shafykay2012@gmail.com

Bisiriyu, L. A.
Department of Demography and Social Statistics,
Obafemi Awolowo University,
Ile-Ife

ABSTRACT
This study focuses on women's family size preferences in Saki – West Local Government Area. A log-linear model is constructed to capture socioeconomic and demographic factors. The model was estimated using primary data obtained through questionnaires administration and five variables considered are women's number of preferred children, women's education, employment status, level of income and the use of contraceptive. The result from log linear model analysis shows that all the main effects and the two – way interaction are significant $(p < 0.05)$ *in obtaining the expected cell count of women's family size preferences. The three – way interaction, four – way interaction and the five – way interaction are found not to be significant*$(p > 0.05)$*. The estimated results show that socio-economic and demographic factors have significant impact on women's family size preferences. The results show that improvement in socio-economic factors and demographic behaviors have direct impact on women's family size preferences. The estimated model is expected to be used for policy experiments by promoting reduction in the family size preferences.*

Keywords*: Log-linear, model, socioeconomic, demographic, family size, preferences, interaction, policy, experiment.*

INTRODUCTION

Family size is an important population, socioeconomic and reproductive health issue globally. The total fertility rate of various countries vary, while low figures predominant in industrialized countries, higher figures are noted in developing countries of the world. The global total fertility rate fell from 5 children per woman to 2.7 children in 2005. Nigeria has a total fertility rate of 5.7 births per woman, a figure that is not only higher than that recorded for sub-Saharan Africa 5.2, but is also indeed among the highest globally. In industrialized countries, both the crude birth rate (the number of live births per 1000 population) and the overall fertility rate (the number of live births per 1000 women aged 15 – 44 years) have dropped with women having fewer children (DHS, 2008). There is evidence that women are delaying childbearing globally due to the need to attain high educational levels. Decision regarding family size should be a shared responsibity strengthened by mutual spousal communication. Willingness of husbands to adopt or allow their wives to use contraceptive determines the place of fertility reduction.

The study of fertility is very important for fertility determines the rate of population growth

in a country. In developing countries, like Nigeria, where there is a high population growth rate, rapid economic development is being hindered and it is due to fertility level. It has been said that not only biological factors determine the level of fertility, but also socio – economic and cultural factors. Within the limit of established physiological factors, a multiplicity of economic, social and cultural factors are the ultimate determinants of fertility levels and their variation in different societies (Isiugo–Abanihe, 1996).

The study intends to fit and develop an appropriate model for the women's family size preferences. Categorical data are commonly analyzed by the use of log linear and logistic models. Categorical data are sub – divided into un – ordered data called nominal data and the ordered data referred to as ordinal data. Log linear modeling is essentially a discrete multivariate statistical technique that is designed specifically for analyzing data when both the independent and dependent variables are categorical or nominal.

LITERATURE REVIEW

The analysis of contingency tables with multi – way classifications originates from the historical development of statistical inference with 2×2 tables. In the initial extension to the case of $2 \times 2 \times k$ tables, Bartlett (1935) discussed testing for three – way interaction and derived an estimate of the common odds ratio suggested by R.A. Fisher (1925). Norton (1945) and Simpson (1951) supplied interpretations of varied interactions which led to the well – known Simpson's paradox (Blyth 1992). Roy and Kastenbaum (1956) showed that Bartlett's procedure is an implicit Maximum Likelihood Estimation (MLE), conditioned upon the fixed margins of each 2×2 table. The celebrated Analysis of Variance (Fisher 1925) inspired discussions of partitioning chi – square within the contingency tables, notable by Lancaster (1951), Mood (1950) and Claringbold (1961) among others. In related research in biostatistics, Cochran (1954), Woolf (1955) and Mantel and Haenszel (1959) developed chi – square tests for no association between two variables across levels of the third variable. Those early studies led to further analysis of three way tables which included estimating the common odds ratio, testing zero interaction and testing no association across strata for examples, Kullback (1959), Gart (1962), Darroch (1962), Lewix (1962), Plackett (1962) and Birch (1963, 1964).

Log linear models are used to determine whether there are any significant relationships in multiway contingency tables that have three or more categorical variables and to determine if the distribution of the counts among the cells of a table can be explained by a simpler, underlying structure (restricted model). The saturated model contains all the variables being analyzed and all possible interactions between the variables Agresti (2007). In the 1980's the analysis of cross classified data changed quite dramatically with the publication of series of papers on log linear models by Darroch (1962), Birch (1963), Bishop et al (1975) and Goodman (1981).

Demaris (1990) stated that the logistic regression or logit model predicts group membership for the dependent variable and measuring the rate of change in the probability of occurrence of an event with change in a given predictor. Logistic regression is a more flexible instrument than log linear modeling for analyzing a mixed set of nominal / ordinal and interval variables (Hosmer and Lemeshow, 1989).

DATA AND METHOD

Primary data were used for this study. The target population was married women who were within the reproductive ages in the last 12 months. Reproduction age is defined as ages 15 to 49 years following the classification of Nigeria Demographic and Health Survey (2008). The study adopted the multistage sampling technique. Saki West Local Government consists of eleven wards. Five wards were selected using simple random technique. In addition, five

streets were selected from each ward and out of each of the selected street, ten households were selected using systematic sampling technique and one eligible respondent was selected from each house. Two hundred and fifty questionnaires were administered, 239 were recovered and used for the analysis. Five variables considered were women's number of preferred children, women's education, employment status, level of income and the use of contraceptive. Statistical Package for Social Sciences (SPSS) was used to fit log – linear model. Notations of the model term are defined below:

$$\begin{aligned} logM_{ijklm} = \mu &+ u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + u_{5(m)} + u_{12(ij)} + u_{13(ik)} + u_{14(il)} \\ &+ u_{15(im)} + u_{23(jk)} + u_{24(jl)} + u_{25(jm)} + u_{34(kl)} + u_{35(km)} + u_{45(lm)} \\ &+ u_{123(ijk)} + u_{124(ijl)} + u_{125(ijm)} + u_{134(ikl)} + u_{135(ikm)} + u_{145(ilm)} \\ &+ u_{234(jkl)} + u_{235(jkm)} + u_{245(jlm)} + u_{345(klm)} + u_{1234(ijkl)} + u_{1235(ijkm)} \\ &+ u_{1245(ijlm)} + u_{1345(iklm)} + u_{2345(jklm)} + u_{12345(ijklm)} \end{aligned}$$

Where,

$logM_{ijklm}$ is the log of the expected cell frequency in the contingency table

μ is the overall mean of the natural log of the expected frequencies

u is the effect which the variables have on the cell frequencies

i,j,k,l and m refer to the categories within the variables

$u_{1(i)}$ is the main effect

$u_{12(ij)}$ is the two way interaction effect

$u_{123(ijk)}$ is the three way interaction effect

$u_{1234(ijkl)}$ is the four way interaction effect

$u_{12345(ijklm)}$ is the five way interaction effect

RESULTS

Hierarchical log – linear analysis was conducted, using a backward elimination of effects, starting with a saturated model, including all main – effects and interaction terms. The backward elimination process indicated that third – order effects, fourth – order effects and fifth – order effects could be eliminated.

From the analysis, the result of log linear model is shown in table 1 below. It was found that all the main effects are significant (p-value < 0.05) that is, (No. of preferred children, education, employment, income and contraceptive). Also on the two – ways interaction, nine are found to be significant, these two – way interactions are: (education*employment, employment*income, no. of children*income, no. of children*employment, employment*contraceptive, education*income, no. of children*education, income*contraceptive, no. of children*contraceptive). The three – way interaction, four – way interaction and the five – way interaction are found not to be significant.

Table 1 below shows that all main effects are significant. The interaction between $u_{123(ijk)}$, $u_{124(ijl)}$, $u_{125(ijm)}$, $u_{134(ikl)}$, $u_{135(ikm)}$, $u_{145(ilm)}$, $u_{234(jkl)}$, $u_{235(jkm)}$, $u_{245(jlm)}$, $u_{345(klm)}$, $u_{1234(ijkl)}$, $u_{1235(ijkm)}$, $u_{1245(ijlm)}$, $and\ u_{12345(ijklm)}$ were not significant at the 0.05 level of significance.

Therefore, the model being recommended on the basis of the result in table 1 is expressed as follow:

$$\begin{aligned} logM_{ijklm} = \mu &+ u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + u_{5(m)} + u_{12(ij)} + u_{13(ik)} + u_{14(il)} \\ &+ u_{15(im)} + u_{23(jk)} + u_{24(jl)} + u_{34(kl)} + u_{35(km)} + u_{45(lm)}. \end{aligned}$$

From the table 1 the result of analysis using the hierarchical modeling, it shows that education*employment ($u_{23(jk)}$), employment*income ($u_{34(kl)}$), no. of children*income ($u_{14(il)}$), education*income ($u_{24(jl)}$), no. of children*employment ($u_{13(ik)}$),

employment*contraceptive ($u_{35(km)}$), no. of children*education ($u_{12(ij)}$), income*contraceptive ($u_{45(lm)}$) and no. of children*contraceptive ($u_{15(im)}$) are significant.

TABLE 1 RESULT OF ANALYSIS FOR HIERARCHICAL MODELING USING BACKWARD ELIMINATION METHODS

Step[a]	Effects	Chi–square[c]	Df	Sig.
Generating Class[b]	Education*Employment	49.951	6	0.000
	Employment*Income	40.976	6	0.000
	No. of Children*Income	60.012	12	0.000
	No. of Children*Employment	57.433	8	0.000
	Employment*Contraceptive	25.365	2	0.000
	Education*Income	58.623	9	0.000
	Income*Contraceptive	33.428	3	0.000
	No. of Children*Education	1119.312	12	0.000
	No. of children*Contraceptive	122.443	4	0.000
	Education*Contraceptive	1.872	3	0.599
	No. of children*Education*Contraceptive	4.384	12	0.975
	Education*Income*Contraceptive	3.901	9	0.918
	No. of children*Employment*Contraceptive	2.442	8	0.964

a. At each step, the effect with the largest significance level for the Likelihood Ratio Change is deleted, provided the significance level is larger than 0.05.
b. Statistics are displayed for the best model at each step after step 0.
c. For 'Deleted Effect', this is the change in the Chi-Square after the effect is deleted from the model.

TABLE 1 RESULT OF ANALYSIS FOR HIERARCHICAL MODELING USING BACKWARD ELIMINATION METHODS

Step[a]	Effects	Chi–square[c]	Df	Sig.
	No. of children*Employment*Income	1.338	24	1.000
	No. of children*Income*Contraceptive	0.485	12	1.000
	Education*Employment*Income	0.306	18	1.000
	Employment*Income*Contraceptive	0.000	6	1.000
	Education*Employment*Contraceptive	0.000	6	1.000
	No. of children*Education*Income	1.843	36	1.000
	No. of children*Education*Employment	0.006	24	1.000
	Education*Employment*Income*Contraceptive	0.000	18	1.000
	No. of children*Employment*Income*Contraceptive	0.000	24	1.000
	No. of children*Education*Income*Contraceptive	0.000	36	1.000
	No. of children*Education*Employment*Contraceptive	0.000	24	1.000
	No. of children*Education*Employment*Income	0.000	72	1.000
	No. of children*Education*Employment*Income*Contrac.	0.000	72	1.000

a. At each step, the effect with the largest significance level for the Likelihood Ratio Change is deleted, provided the significance level is larger than 0.05.
b. Statistics are displayed for the best model at each step after step 0.
c. For 'Deleted Effect', this is the change in the Chi-Square after the effect is deleted

from the model.

CONCLUSION

The result from log linear model analysis shows that all the main effects (number of preferred children, level of education, employment status, level of income and contraceptive use), the two ways interaction (education*employment, employment*income, no. of children*income, no. of children*employment, employment*contraceptive, education*income, no. of children*education, income*contraceptive, no. of children*contraceptive) are significant in obtaining the expected cell count of women's fertility preferences.

It can be concluded that the level of education, employment status, level of income and contraceptive use play an important role in the women's fertility preferences. It is found out the level of education and employment status are significant in the log linear model.

It is therefore, recommended that researchers should use the techniques to analyze such categorical data sets to ensure that their statistical results are robust.

REFERENCES:

Agresti, A. (2007): *An Introduction to Categorical Data Analysis (2nd ed.).* Hoboken, NJ: Wiley Inter – Science. P. 212. ISBN 978-0-471-22618-5.

Birch, M.W. (1963): Maximum Likelihood in three – way contingency tables. *J. Roy. Statist. Soc.* B 25, 220 – 233.

Birch, M.W. (1964): The detection of partial association. *J. Roy. Statist. Soc.* B 26, 313 – 324.

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975): *Discrete Multivariate Analysis.* MIT Press.

Blyth, C.R. (1992): On Simpson's paradox and the sure thing principle, *J. American Statistical Association* 67, 364 – 366.

Bartlett, M.S. (1935): Contingency table interactions. *J. Roy. Statist. Soc.* B 2, 248 – 252.
Claringbold, P.J. (1961): The use of orthogonal polynomials in partition of chi – square tests. *Australia J. Statist.* 3, 48 – 63.

Cochran, W.G. (1954): Some methods for strengthening the common chi – square tests. *Biometrics* 24, 315 – 327.

Darroch, J.N. (1962): Interactions in multifactor contingency tables. *J. Roy. Statist. Soc.* B 24, 251 – 263.

Demaris, A. (1990): Interpreting logistic regression results: A critical commentary. *Journal of Marriage and Family,* 52: 271 – 277.

Fisher, R.A. (1925): Statistical methods for Research Workers. Oliver and Bold.

Gart, J.J. (1962): On the combination of relative risks. Biometrics 18, 601 – 610.

Goodman, L.A. (1981): On partitioning χ^2 and detecting partial association in three – way contingency tables. J. Roy. Statist. Soc. B 31, 486 – 498.

Hosmer, D. & Lemeshow S. (1989): Applied logistic regression. New York: John Wiley.

Knoke, D. and Burke, P.J. (1980): *Log – Linear Models.* Sage Publications, Inc. Newberry Park, California, USA.

Krishnan, P. (1984): Interrelation between fertility and non-economic characteristics in India: results from a log-linear model. *Population Research Laboratory,* 1984. P. 23

Kullback, S. (1959): Information theory and statistics. John Wiley & Sons, Inc. New York, USA.

Lancaster, H.O. (1951): Complex contingency tables treated by the partition of chi – square. *J. Roy. Statist. Soc.* B 13, 242 – 249.

Lewis, B.N. (1962): On the analysis of interaction in multi – dimensional tables. *J. Roy. Statist. Soc.* A 125, 88 – 117.

Mood, A.M. (1950): Introduction to the Theory of Statistics. Mc Graw Hill.

Mantel and Haenzel, W. (1959): Statistics aspect of the analysis of data from retrospective studies of diseases. J. National Cancer Institute 22, 719 – 748.

Norton, H. W. (1945): Calculation of chi – square for complex contingency tables. J. American Statistical Association. 40, 251 – 258.

Plackatt, R.L. (1962): A note on interactions in contingency tables. *J. Roy. Statist. Soc.* B 24, 162 – 166.

Simpson, E.H. (1975): The interpretation of interaction in contingency tables. *J. Roy. Statist. Soc.* B 13, 238 – 241.

Tabachnick, B.G. and Fidell, L.S. (1996): *Using Multivariate Statistics.* 3rd Edition. Harper Collins. New York, USA.

Woolf, B. (1955): On estimating the relation between blood group and disease. *An Human Genetics.* 19, 251 – 253.