Asian Journal of Probability and Statistics

6(1): 55-63, 2020; Article no.AJPAS.53734

ISSN: 2582-0230

Prediction of the Effect of Demographic Characteristics on Parity Using Poisson Regression Model

C. M. Gatwiri[1*], M. M. Muraya[2] and L. K. Gitonga[3]

[1]Department of Physical Sciences, Chuka University, P.O.Box 109-60400, Chuka, Kenya.
[2]Department of Plant Science, Chuka University, P.O.Box 109-60400, Chuka, Kenya.
[3]Department of Nursing, Chuka University, P.O.Box 109-60400, Chuka, Kenya.

Authors' contributions

This work was carried out in collaboration among all authors. Author CMG designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Authors MMM and LKG managed the analyses of the study. Author MMM managed the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2020/v6i130154
Editor(s):
(1) Dr. Patricia J. Y. Wong, Nanyang Technological University, School of Electrical and Electronic Engineering, Singapore.
(2) Dr. Manuel Alberto M. Ferreira, Retired Professor, Department of Mathematics, ISTA-School of Technology and Architecture, Lisbon University, Portugal.
Reviewers:
(1) Anonymous, China.
(2) M. Bhanu Sridhar, G.V.P College of Engineering for Women, India.
Complete Peer review History: http://www.sdiarticle4.com/review-history/53734

Original Research Article

Received: 05 November 2019
Accepted: 10 January 2020
Published: 16 January 2020

Abstract

There is growing interest among the public in demography since demographic change has become the subject of political debates in many countries. Statistics on demography are used to support policy-making and monitor demographic behaviour of political, economic, social and cultural perspectives. Most studies have used descriptive statistics to study demographic characteristics. Moreover, most of these studies investigate effects of individual character at a time. Therefore, there is a need to come up with more robust statistical methods, such as predictive models for demographic studies. The objective of this study was to predict the effect of demographic characteristics on parity using Poisson regression model. Secondary data on parity, age, marital status and education level was collected from Chuka and Embu hospital maternal units from 2013 to 2017. The data was analysed using R-statistical software. Three Poisson regression models (PRMs) were fitted. The likelihood ratio test of all the Poisson regression models had p-values < 0.05 indicating that all the models were statistically significant. Deviance test and Akaike Information Criterion (AIC) were used to assess the fit of Poisson regression models. The overall

**Corresponding author: E-mail: ncastygatwiri20@gmail.com;*

Poisson model had residual deviance of 184.23, which was the lowest of all other fitted PRM models, suggesting that it was the best fit. The AIC of the PRM with both education and marital status as the predictors had the lowest AIC value of 2078.620, indicating that it was the best fitted model. The dispersion test proved that PRM was not over-dispersed, confirming the model as a good fit of the data. The improved model can be used in prediction of population growth rates.

Keywords: Prediction model; parity; demographic characteristics; Poisson regression model.

1 Introduction

Demography is the mathematical way of modelling and statistical analysis of populations [1]. Demographic modelling has been of great concern in empirical statistics, econometrics, and dynamic population studies [2]. Population dynamics modelling is based on demographic components, i.e., life expectancy, migration, fertility, and mortality. Statistics on population structure are increasingly used to support policy-making and monitor demographic behaviour from political, economic, social, and cultural perspectives [3,4]. Demographic data is sophisticated and require robust statistical methods for analysis [5,6,7]. These data take different forms, i.e., nominal, ordinal, interval, or even ratio. Studies have applied different models to study demographic characteristics from time to time. In most of demographic studies, the type of data and the response variable dictates the type of modelling to be carried out. For example, Erkan, et al. [8] compared Quasi Poisson and Conway-Maxwell-Poisson (COM) regression models in determining the factors affecting the number of babies born alive in multiple pregnancies. The model selection based on Akaike Information Criterion values revealed that COM Poisson model outperformed the Quasi Poisson model. Barakat [9] modelled parity using generalised count distributions, i.e., Conway-Maxwell-Poisson and gamma count model. The results indicated that generalised count distributions offered an improved fit compared to customary Poisson and negative binomial models in the presence of under-dispersion and over-dispersion. Moreover, generalised count distributions were more accurate in examining fertility that involves completed parity as an outcome. A Poisson regression model is used when the dependent variable is an observed count that follows the Poisson distribution. Poisson regression has the advantage of fitting non-linear models over the linear regression models including situations involving the number of occurrences (counts) of an event. For example, a study on current and predicted fertility found out that Poisson regression model was an applicable tool for predicting number of children a woman was expected to have in Nigeria [10]. Tejada, et al. (2017) used PRM to evaluate socioeconomic, geographic, reproductive, behavioural, demographic, and chronic disease variables on the number of children born to a woman (parity). The results obtained showed a positive association between the studied characteristics and the number of children born to a woman, rendering the model effective.

Many of demographical studies use descriptive statistics to explain the relationship among factors studied [11,12,13]. Such studies are limited in their power to make inferences on effect of demographic characteristics on the variable of interest. Moreover, these studies look at individual factor effect on variable of interest. Most multivariate analysis techniques assume normality, linearity, and homoscedasticity [14]. This limit their use in the cases where the data being modelled does not obey this assumption making the generalised linear models a good fit.

Finer and Zolna [15] studied age and marital status in the planning status of births and found these factors to be the central stratifying factors in many countries. This study, therefore, focused on age and marital status as the primary independent variables. Moreover, age and marital status are key factors in determining the normative context for childbearing. In addition, the study considered the role of parity in population structure, distinguishing between first, second, and higher-order births. Particular attention was paid to the education of the mother given the substantial diversity in childbearing behaviour and planning status across communities. To achieve the goal of this study, the study used Poisson Regression Model to predict the extent to which these factors affect parity. Consequently, PRM was used to determine the significant factors affecting parity. The analysis of data in this study was done using R Statistical software (R core Team 2018).

2 Methodology

The target population of this study consisted of women aged 15 - 49 years (WHO's reproductive age) in the records for delivery from Chuka and Embu hospitals in Kenya for the years 2013 to 2017. The study area was selected purposively. A mixed method research design was employed in this study. This research design is used when both qualitative and quantitative approaches are applied in a single study when one of the approaches is not complete in itself [16]. This study used two research designs, i.e., causal-comparative research design and survey design. Simple random sampling was employed in selection of the subjects from the hospital records. The method was advantageous since the data to be obtained from the selected representative sample was a fair reflection of the characteristics of the entire population, and every respondent had an equal chance of being selected. Secondary data for the years 2013 to 2017 was collected on the level of education, age, marital status and parity of the selected mothers from the maternal files of Embu and Chuka hospitals in the form of hard copies using a checklist. The independent variables included level of education, age and marital status and the dependent variable was parity. The sample size was calculated according to Mugenda and Mugenda [17] yielding a sample size of 384 mothers from each hospital. This was done to realise a large enough sample for the use of modelling based on the recommendations by Vanvoorhis and Morgan [18]. The final realised sample included a total of 768 mothers.

Table 1 provides a summary of the educational attainment and marital status of the sampled mothers. The sample was slightly dominated by secondary graduates (40.49%) followed by primary graduates (38.41%). Approximately 77.34% of mothers were married with only 22.14% as single.

Table 1. Social-demographic profile of the mothers aged 15-49 years at Chuka and Embu hospitals from 2013 to 2017

Education	**Frequency**	**Percentage**
None	24	3.12
Primary	295	38.41
Secondary	311	40.49
Tertiary	138	17.97
Marital status		
Divorced	1	0.13
Married	594	77.34
Separated	2	0.26
Single	170	22.14
Widowed	1	0.13

The data was compiled using excel and "read excel" package was then used to import data into R. The "data. table" package in R Statistical software was used to store the "read" data and also for data manipulation. Parity was regressed with maternal education and marital status using Poisson regression model. Age was used as an offset variable which acted as a measure of exposure in the context of Poisson regression model. Education was at four levels, i.e., none, primary, secondary, and tertiary with no education level being the reference. Marital status had five statuses, i.e., married, single, separated, divorced, and widowed with single being the reference. However, widowed, divorced and separated were negligible in the study with proportions of mothers under these categories being less than 0.5% thus leading to their omission in modelling. These variables were outliers and it would be misleading to infer from such small proportions. The parameters in the model were estimated using the method of maximum likelihood and the relative ratios for the response variable were calculated from the parameters of the fitted models. Likelihood ratio test was used to test hypothesis. Akaike Information Criterion and Deviance Information Criterion were used to compare the nested Poisson models to assess the model fit. Dispersion test was used to check whether the Poisson regression model was over-dispersed.

3 Results and Discussion

This paper aims to predict the effect of demographic characteristics on parity using Poisson regression model (PRM). The literature survey has covered many papers concerned with this idea, from 2000 to 2018. Data was taken from two hospitals and it was imported to "data.table" package in R. PRM was implemented on that data and results are displayed adequately.

3.1 Results of fitted Poisson regression model with education

Relative ratios calculated from the parameters (e^{β_i}) were used to explain the relationship between parity and education. The parity of women with primary, secondary, and tertiary education was significantly different from parity of women with no education. The relative ratio of women with primary education was 0.7006 implying that a woman with primary education was 29.94% less likely to have a higher parity as compared to a woman with no education (Table 2). The relative ratio for secondary education 0.5270 indicates that a woman with secondary education is 47.30% less likely to have a higher parity compared to those with no education. In addition, women with tertiary education are 52.58% less likely to have a higher parity compared to women with no education. This implies that the level of education and parity are indirectly proportional since the relative ratios decrease with an increase in the level of education. This concurs with the findings of Impicciatore and Dalla Zuanna [19] who reported that there exists a negative relationship between various levels of education in women and their total fertility rates in a population.

The results show that all the education levels; primary, secondary, and tertiary education were significant in explaining parity since their p-values were <0.05. According to Anker and Knowles [20], female education in Kenya was insignificantly related to fertility at macro level but significantly (negatively) related to fertility at micro level analysis.

The fitted PRM equation with education is given by;

$$\ln \lambda = -2.0723 - 0.3558\, P.E - 0.6406\, S.E - 0.7461\, T.E \tag{1}$$

3.2 Results of fitted Poisson regression model with marital status

For married women, the relative ratio 1.3781 (Table 3) implied that a married woman is 37.81% more likely to have a higher parity as compared to a single woman. This was in agreement with the study by Phillips and Sweeney [21]. They found out that women who dissolve their marriages lose a period of exposure to childbearing when moving from one marriage to another. These women are, therefore, likely to end up having a lower parity as compared to married women. The results also showed that being married was significant in explaining parity since p-value was < 0.05. This is consistent with Shapiro and Gebreselassie [22] study on the relationship between marriage and fertility transitions in Sub-Saharan Africa. The study showed that entering into marriage had some effect on fertility while being divorced, widowed, non-married, and polygyny had negligible effects on fertility changes.

The fitted PRM equation with marital status is given by,

$$\ln \lambda = -2.8446 + 0.3207M \tag{2}$$

3.3 Results of fitted Poisson regression model with both education and marital status

The relative ratio for primary education against no education was 0.7120 (Table 4), indicating that women with primary education have 28.8% less children as compared to those with no education. Those women with secondary education have 45.02% less children as compared to those with no education, with a relative ratio of 0.5498. Those with tertiary education have 50.84% less children than those with no education, with a relative ratio of 0.4916. All the education levels were significant in explaining parity since their p-values were <0.05, implying that as the level of education increased parity decreased for the levels of education but the decrease in parity was highest at the tertiary level. These results concurred with Cygan-Rehm and Maeder [23] study which found out that women who have attained college education levels tend to have fewer children when compared to those with high school levels or lower levels.

Table 2. Relative ratios and 95% confidence intervals following a poisson regression

	Estimate	Std. error	z value	Pr(>|z|)	R.R	2.5 %	97.5 %
(Intercept)	-2.0723	0.1066	-19.44	0.0000	0.1259	0.1014	0.1541
Primary Education (P.E)	-0.3558	0.1132	-3.144	0.0017	0.7006	0.5646	0.8804
Secondary Education (S.E)	-0.6406	0.1157	-5.535	0.0000	0.5270	0.4224	0.6653
Tertiary education (T.E)	-0.7461	0.1265	-5.899	0.0000	0.4742	0.3716	0.6105

Table 3. Relative ratios and 95% confidence intervals following a poisson regression model

	Estimate	Std. error	z value	Pr(>|z|)	R.R	2.5 %	97.5 %
(Intercept)	-2.8446	0.0682	-41.71	0	0.0582	0.0507	0.0663
Married (M)	0.3207	0.0737	4.3490	0	1.3781	1.1955	1.5965

Table 4. Relative ratios and 95% confidence intervals following a poisson regression model

	Estimate	Std. error	z value	Pr(>|z|)	Odds	2.5 %	97.5 %
(Intercept)	-2.3149	0.1279	-18.09	0.0000	0.0988	0.0764	0.1262
Married (M)	0.2573	0.0745	3.455	0.0006	1.2934	1.1204	1.5004
Primary Education (P.E)	-0.3397	0.1133	-2.999	0.0027	0.7120	0.5737	0.8949
Secondary Education (S.E)	-0.5982	0.1163	-5.144	0.000	0.5498	0.4402	0.6949
Tertiary Education (T.E)	-0.7100	0.1268	-5.598	0.0000	0.4916	0.3850	0.6333

The study shows that married women have 1.2934 times more children than single women, which is 29.34% (Table 4). The married status was significant in explaining parity since the p-value was <0.05. The results showed that those women who are married have more children as compared to those who are single. This is consistent with Torche and Rich (2017) study which found out that married couples are more likely to want children than unmarried people. It can be concluded that all the education levels and married were significant in explaining parity since their p-values were < 0.05.

The fitted PRM equation with both education and marital status is written as,

$$ln\,\lambda = -2.3149 + 0.2573\,M - 0.3397P.E - 0.5982\,S.E - 0.7100\,T.E \quad (3)$$

3.4 Model selection among the Poisson regression models

The study fitted three nested PRM whose significance in predicting parity using demographic characteristics was evaluated using a likelihood ratio test. Deviance test was then performed on each of the three fitted models to determine the model that included the most relevant predictor variables. Further, an AIC approach was used to come up with the model that had the highest prediction ability. The selected PRM based on AIC approach was tested for over-dispersion and finally confirmed significant using a p-value test.

Table 5 shows that the chi-square values of the fitted Poisson models with their associated p-values. The results indicate that all the models had p-values less than 0.05. This implies that all the levels of education and marital status have a statistically significant effect on parity at a 5% level of significance. This is in agreement with four other studies that found out that changes in fertility rates are as a result of changes in maternal education, marital status among other factors [24-27].

Table 6 presents the Poisson models in terms of the power of prediction. According to Spiegelhalter, et al. [28], null deviance shows how well the response variable is predicted by a model and it governs the model that includes only the intercept. When the model includes predictor variables then the deviance is residual. A lower value of residual deviance indicates that the model has become better after including the predictor variables. There are three Poisson regression models with the same null deviance but different residual deviance (Table 6). Therefore, the best model is the one with the lowest residual deviance, which is the model with education and marital status.

Table 7 provides an analysis of the Akaike's Information Criterion (AIC) for the studied Poisson models. AIC provides a method for assessing the quality of a model through comparison of related models [29,30,31]. This criterion prevents studies from including irrelevant predictors. The results of backward selection were used to select the best model (Table 7). In this method, the study starts with saturated model and removes independent variables one by one until the best model fit model is arrived at. Akaike's information criterion was used to select the best model. The saturated model which had education, marital status, and the interaction between education and marital status had an AIC value of 2083.974. The model with marital status and education only with no interaction had an AIC value of 2078.620. The models with education and marital status only had AIC values of 2089.268 and 2121.694, respectively. The fitted model with education and marital status only with no interaction was the best model since it had the smallest AIC value.

A generalized linear model can sometimes give a good summary of the data since both the linear predictor and the distribution are correctly chosen, but the fit of the full model may be poor. One possible reason for this may be an over-dispersion. Over-dispersion occurs when the variance of the response is larger than would be expected for the chosen distribution. For example, if we use a Poisson distribution to model the data, we would expect the variance to be equal to the mean value: $\mu=\sigma$

This test is used to test whether the mean and variance in the Poisson regression model are equal [32]. If the two are not equal, then there is an overdispersion and an additional model needs to be used for the data to fit the model; otherwise, there is no overdispersion and the model is fit. Under the null hypothesis, the constant c is assumed to be 0. When it is not 0, there is an overdispersion which is the alternate hypothesis. The dispersion test is given by,

$$Var(y) = E(y) + c(f(y))$$

The hypothesis to be tested is,

$H0_1$: The mean and the variance of the model are equal
$H0_2$: The mean and the variance of the model are not equal

The results from dispersion test give the following $z, p, and\ \alpha$ values.

$$z = -28.251,$$

$$p - value = 1$$

$$\alpha = -0.7236723$$

Since the $p - value > 0.05$ the level of significance, the study fails to reject the null hypothesis that the model does not have an overdispersion and so the model fits the data.

To test the significance of the model, the hypothesis tested is,

H_0: Education and marital status are not significant in predicting parity
H_1: Education and marital status are significant in predicting parity

The p-value for this hypothesis was $2.942091e - 14 < 0.05$. The study thus rejected the null hypothesis and concluded that the fitted model was significant.

Table 5. Likelihood ratio test of the Poisson models

Model	Df	Loglik	Df	Chisq	Pr(>chisq)
Parity ~education:					
saturated model	1	-1069.0			
Fitted model	4	-1040.6	3	56.814	<0.05
Parity~ marital status:					
Saturated model	1	-1069.0			
Fitted model	2	-1058.8	1	20.388	<0.05
Parity~ education*marital status					
Saturated model	1	-1069			
Fitted model	8	-1034	7	70.109	<0.05
Parity~ education +marital status					
Saturated model	1	-1069			
Fitted model	5	-1034.3	4	69.463	<0.05

Table 6. Deviance information analysis for the Poisson regression models

Model	Null deviance	d.f	Residual deviance	d.f
parity~education	254.34	763	197.52	760
Parity~marital status	254.34	763	232.95	762
Parity~education+marital status	254.34	763	184.23	756

Table 7. Stepwise model selection using AIC

Model	AIC
Parity~education	2089.268
Parity~ marital status	2121.694
Parity~ education*marital status	2083.974
Parity~education+marital status	2078.620

4 Conclusion

The parity of women at all the education levels was significantly different from that of women without any education. Thus education is a key variable that affect parity levels. This shows that educating women can be used as a mean of controlling birth rates. The study also revealed that the marital status was significant in predicting parity. This is supported by the fact that the best Poisson model out of the three fitted models was the model with both education and marital status. Notably, women's status is a significant factor of economic growth in many nations. The opportunity cost between education access and motherhood among women affects their economic opportunities and socio-economic role in many countries. Therefore, countries should invest more in women education to increase the social and economic benefit among women and boost their economic growth.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Diekmann O, Heesterbeek JAP. Mathematical epidemiology of infectious diseases: Model building, Analysis and Interpretation, John Wiley & Sons. 2000;5.

[2] Weegman MD, Bearhop S, Fox AD, Hilton GM, Walsh AJ, Mcdonald JL, Hodgson DJ. Integrated population modelling reveals a perceived source to be a cryptic sink. Journal of Animal Ecology. 2016;85(2):467–475.

[3] Dahlgren G, Whitehead M. Policies and strategies to promote social equity in health. Background document to WHO-Strategy Paper for Europe (No. 2007: 20114) Stockholm: Institute for Future Studies; 1991.

[4] Boyden J. Childhood and the policy makers: A comparative perspective on the globalization of childhood. In Constructing and Reconstructing Childhood Routledge. 2015;185-219.

[5] Brantingham PJ, Brantingham PL. Patterns in Crime. New York: Macmillan; 1984.

[6] Chesnais JC. The demographic transition: Stages, patterns and economic implications. Oxford University Press; 1992.

[7] Noland RB, Oh L. The effect of infrastructure and demographic change on traffic-related fatalities and crashes: A case study of illinois county-level data. Accident Analysis & Prevention. 2004; 36(4):525-532.

[8] Erkan G, Evkaya O, Türkan S. Determination of the affecting factors of the number of babies born alive in multiple pregnancies with poisson models. Türkiye Klinikleri Biyoistatistik. 2017;9(3):222-229.

[9] Barakat B. Generalised count distributions for modelling parity. Demographic Research. 2017;36: 745-758.

[10] Fagbamigbe AF, Adebowale AS. Current and predicted fertility using Poisson regression model: Evidence from 2008 Nigerian Demographic Health survey. African Journal of Reproductive Health. 2014;18(1):71-83.

[11] López PO, Bréart G. Sociodemographic characteristics of mother's population and risk of preterm birth in Chile. Reproductive Health. 2013;10(1):26.

[12] Emelumadu OF, Ukegbu AU, Ezeama NN, Kanu OO, Ifeadike CO, Onyeonoro UU. Socio-demographic determinants of maternal health-care service utilization among rural women in Anambra State, South East Nigeria. Ann Med Health Sci Res. 2014;4(3):374–382.

[13] Tarekegn SM, Lieberman LS, Giedraitis V. Determinants of maternal health service utilization in Ethiopia: Analysis of the 2011 Ethiopian demographic and health survey. BMC Pregnancy Childbirth. 2014;14(1):161-161.

[14] Darlington RB, Hayes AF. Regression analysis and linear models: Concepts, applications and implementation. Guilford Publications; 2016.

[15] Finer LB, Zolna MR. Declines in unintended pregnancy in the United States, 2008–2011. New England Journal of Medicine. 2016;374(9):843-852.

[16] Creswell JW, Creswell JD. Research design: Qualitative, quantitative and mixed methods approaches. Sage Publications; 2017.

[17] Mugenda O, Mugenda A. Research methods: Quantitative and qualitative approaches. Nairobi, Kenya; 2003.

[18] Vanvoorhis CW, Morgan BL. Understanding power and rules of thumb for determining sample sizes. Tutorials in Quantitative Methods for Psychology. 2007;3(2):43-50.

[19] Impicciatore R, Dalla Zuanna G. The impact of education on fertility in Italy. Changes across cohorts and South–north Differences. Quality & Quantity. 2017;51(5):2293-2317.

[20] Anker R, Knowles JC. Fertility determinants in developing countries: A case study of Kenya; 1982.

[21] Phillips JA, Sweeney MM. Premarital cohabitation and marital disruption among white, black and Mexican American women. Journal of Marriage and Family. 2005;67(2):296-314.

[22] Shapiro D, Gebreselassie T. Fertility transition in Sub-Saharan Africa: Falling and stalling. African Population Studies. 2013;23(1):3-23.

[23] Cygan-Rehm K, Maeder M. The effect of education on fertility: Evidence from a compulsory schooling reform. Labour Economics. 2013;25:35-48.

[24] Rindfuss RR, Parnell AM. The varying connection between marital status and childbearing in the United States. Population and Development Review. 1989;447-470.

[25] Murithi MJ. W The Effect of Female Education on Fertility in Kenya; 1998.
Available:https://wol.iza.org/uploads/articles/228/female-education-and-its-impact-on-fertility

[26] Andersen AMN, Wohlfahrt J, Christens P, Olsen J, Melbye M. Maternal age and fetal loss: Population-based register linkage study. British Medical Journal. 2000;320(7251):1708-1712.

[27] Stephen EH, Chandra A. Declining estimates of infertility in the United States: 1982–2002. Fertility and Sterility. 2006;86(3):516-523.

[28] Spiegelhalter DJ, Best NG, Carlin BP, Linde A. The deviance information criterion: 12 years on. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2014;76(3):485-493.

[29] Koehler AB, Murphree ES. A comparison of the Akaike and Schwarz criteria for selecting model order. Applied Statistics. 1988;187-195.

[30] Pan W. Akaike's information criterion in generalized estimating equations. Biometrics. 2001;57(1): 120-125.

[31] Symonds MR, Moussalli A. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's Information Criterion. Behavioural Ecology and Socio-biology. 2011;65(1):13-21.

[32] Patience EO, Osagie AM. Modelling the prevalence of malaria in Niger State: An application of Poisson regression and negative binomial regression models. International Journal of Physical Sciences. 2014;4:061-068.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sdiarticle4.com/review-history/53734