Covenant Journal of Physical & Life Sciences (CJPL) Vol. 5 No. 2, Dec 2017

An Open Access Journal available online

Multilevel Regression Analysis of Age at First Birth

Adeniyi O. I.[1*], Oyinbo-Oke O. C.[1] & Akinrefon A. A.[2]

[1]University of Ilorin, Kwara State,
[2]Modibbo Adama University of Technology, Yola, Adamawa State
* E-mail: niyikiitan@yahoo.com

***Abstract*:** Knowledge about the factors associated with age at first birth plays a major role in controlling the rate of population growth. This paper presents Hierarchical Linear Modeling known for its robustness not only in dealing with hierarchical data structure but also in its ability to explain the effects of the shared variances present in the study on the variable of interest. Data from 2013 Nigeria Demographic Health Survey (NDHS), collected via a hierarchically clustered sampling scheme were used. It investigated the factors that were thought to be associated with variation in age at first birth among Nigerian women were investigated. The model provided parameter estimates as well as estimates of the random effects variances at all the levels. It was observed that the average age at which a Nigerian woman gives birth to her first child without considering any factor effect is 19 years which is a teenage year. 22% and 18% in the variation of ages at first birth resides in the differences in the states and zones in the country.

Keyword: Hierarchical, Multilevel, Intra-class correlation, Restricted Maximum Likelihood

Introduction

Nigeria, a country with estimated population of over 160 million individuals, is currently experiencing high growth rate even in the absence of commensurate infrastructural growth. Although, there are many reasons but one of the major causes of this high growth rate is the variation in age at first birth among Nigerian women. Knowing fully well that our high population is a major cause of unemployment in the nation, effect of variations in age at first birth (teenage mothers, early mothers and later mothers) among Nigerian women goes beyond increase in population to poor socio-economic status.

When a teenager starts having children the tendency is higher that she will not only attain her desired family size at the end of her reproductive age but will have more children than she would have loved to have if she did not start motherhood as a teenager [1]. Women who are married before the age of 18 tend to have more children than those who marry later in life and many young women get married before age 18 [1,2]. Early marriage remains common in developing countries [3]. Teenage pregnancy can be seen as an early warning sign of inequality in a society [4].

Research has shown that educational level is a risk factor in teenage pregnancy [5]. Teenage pregnancy decreases as educational status increases implying that educational attainment is negatively associated with teenage pregnancy.

The relationship between an individual and the group to which they belong are often of great concern in research. This leads to multilevel or hierarchical data structure where individuals are nested within groups. For example, in educational research we may have a sample of schools, and within each school a sample of pupils. Multilevel models is known in the literature under a variety of names, such as ‘hierarchical linear model' [6, 7], ‘variance component model' [8], and ‘random coefficient model' [9, 10].

Material and Methods

The Multilevel Regression Model

The multilevel regression model also known as mixed linear model, random coefficient model, hierarchical linear model or variance component model in the literature assumes hierarchical data with the response variable measured at lowest level while the explanatory variables can exist at all levels.

Estimating the Parameters

In a two-level multilevel regression model with levels 1 and 2, the dependent variable y_{ij} and the explanatory variables x_{ij} are measured at the lowest level while the explanatory variable z_j's are measured at the second (higher) level. Separate level 1 models are developed for each level 2 units. These models are also called within-units models as they describe the effect in the context of a single group [11]. The separate regression equation for each group is given as

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + e_{ij} \qquad (1)$$

Where y_{ij} is the dependent variable measured for the i^{th} level 1 unit nested within the j^{th} level 2 unit, x_{ij} is the value on the level 1 prediction, β_{0j} is the intercept for the j^{th} level 2 unit, β_{1j} is the regression coefficient associated with x_{ij} for the j^{th} level 2 unit and e_{ij} is the random error associated with the i^{th} level 1 unit nested within the j^{th} level 2 unit.

The regression coefficients carry a subscript j indicating that they may vary across the level 2. These are modelled by explanatory variables and random residual term at the level 2. Level 2 models are also referred to as between-unit models as they describe the variability across multiple levels [11]. Consider the case of a single level 2 predictor, the model is given as

$$\beta_{0j} = \gamma_{00} + \gamma_{01} z_j + u_{0j} \qquad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} z_j + u_{1j} \qquad (3)$$

Where z_j is value on the level-2 predictor, γ_{00} is the overall mean intercept adjusted for z, γ_{10} is the overall mean intercept adjusted for z, γ_{01} is the regression coefficient associated with z relative to level-1 intercept, γ_{11} is the regression coefficient associated with z relative to level-1 slope, U_{0j} is the random effects of the j^{th} level-2 unit adjusted for z on the intercept and U_{1j} is the random effects of the j^{th} level-2 unit adjusted for z on the slope.

Substituting equations 2 and 3 into equation 1, it gives a single-equation of the multilevel regression model

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}z_j + \gamma_{11}z_jx_{ij} + u_{ij}x_{ij} + u_{0j} + e_{ij} \quad (4)$$

If there are p-explanatory variables at the lowest level and q-explanatory variables at the higher level, then, equation 4 becomes

$$y_{ij} = \gamma_{00} + \sum_p \gamma_{p0}x_{pij} + \sum_q \gamma_{0q}z_{qj} + \sum_q\sum_p \gamma_{pq}z_{qj}x_{pij} + \sum_p u_{pj}x_{pij} + u_{0j} + e_{ij} \quad (5)$$

The γ are the regression coefficients, u's are the residuals at the group level and e is the residual at the lowest level.

The proportion of variance in the population explained by the clustering structure is given by the intra-class correlation ρ which is estimated using the null model, that is, the model with no explanatory variable, called the intercept-only model;

$$y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \quad (6)$$

The intra-class correlation ρ is estimated by the equation:

$$\rho = \frac{\sigma^2_{u0}}{\sigma^2_{u0} + \sigma^2_e} \quad (7)$$

where σ^2_e is the variation at lowest level and σ^2_{u0} is the variation at the second level.

Likelihood estimation

There are two most commonly used approaches to parameter estimation in multilevel regression analysis. The full maximum likelihood (FML) and the restricted maximum likelihood (REML), FML estimation includes both the regression coefficients and the variance components, that is, both the fixed-effect and the random-effects terms in the likelihood function. The FML method treats the fixed effect as unknown quantities when the variance components are estimated, but do not take into account the degrees of freedom lost in generating the fixed effects. This causes the FML estimates to be biased with smaller variance. The REML methods compare two models that are nested in their random effect terms with the same fixed effects design. REML is more realistic because it leads to better estimates of the variance component.

A multilevel model is of the form

$$y = \underbrace{x\beta}_{fixed} + \underbrace{zb}_{random} + \underbrace{\varepsilon}_{error}, \quad (8)$$

where y is the n x 1 response vector and n is the number of observations, x is an n x p fixed effects design matrix, β is a p x 1 fixed effects vector, z is an n x q random design matrix, b is a q x 1 random effect vector and ε is the n x 1 observation error vector. The random-effects vector, b, and the error vector, ε, are assumed to have the prior distributions;

$$b \sim N(0, \sigma^2 D(\theta)), \qquad \varepsilon \sim N(0, \sigma^2 I),$$

where D is a symmetric and positive semi-definite matrix, parameterized by a variance component vector θ, I is an $n \times n$ identity matrix, and σ^2 is the error variance.

In this model, the parameters to estimate are the fixed-effects coefficients β, and the variance components θ and ε.

Restricted Maximum Likelihood (REML)

REML includes only the variance components; that is, the parameter that parameterise the random-effect terms in the linear mixed-effect model. β is estimated in a second step. Assuming a uniform improper prior distribution for β and integrating the likelihood $L(y|\theta,\sigma^2)$ with respect to β results in the restricted likelihood $L(y|\theta,\sigma^2)$.

$$L(y \mid \theta, \sigma^2) = \int L(y \mid \beta, \theta, \sigma^2) L(\beta) d\beta = \int L(y \mid \beta, \theta, \sigma^2) d\beta$$

The algorithm first profiles out $\hat{\sigma}^2{}_R$ and then maximize the remaining objective function with respect to θ to find $\hat{\theta}_R$. The restricted likelihood is then maximized with respect to σ^2 to find $\hat{\sigma}^2{}_R$. Then, it maximizes β by finding its expected value with respect to the posterior distribution.

Application to data on age at first birth

Dataset from the 2013 Nigeria Demographic and Health Survey (NDHS) were analysed. Individual data were available for 7810 women. The survey was designed to provide information at regional and state levels for both urban and rural areas. Information on the age at first birth was collected from the women. The data captures the 36 states and Federal Capital territory in Nigeria. As a result, the data is hierarchical in structure with level 1 (individuals) nested within level 2 (States) where states are nested within level 3 (Geo-political zones or regions).

The information collected on woman *i* from state *j* and zone *k* was recorded as (y_{ijk} , (i=1,…,7810), (*j*=1,…, 37), (*k*=1,…,6).

The hierarchical structure of the dataset as used in this study is therefore described as follows

Individual level: The age at first birth of the individual woman is considered the lowest level and the unit of analysis in this study.

State level: Each woman belongs to one of the 37 distinct geographical locations (districts) that represent the states.

Zone level: Each state belongs to one of the 6 geo-political zones (zones) that represent the region. The zones includes; North West, North East, North Central, South West, South East and South South.

Description of variables

The response variable (age at first birth) is a continuous variable measured at the individual level (lowest level) while the independent variables include; place of residence, highest education attained, religion, ethnicity and age at first sexual intercourse.

The multilevel model for the three-hierarchical level model is then written as

$$\begin{aligned}(Age\text{ at first birth})_{ijk} = {} & \gamma_{000} + \beta_1 Firstsex + \beta_2 \Pr imary + \beta_3 Secondary + \beta_4 Higher + \\ & \beta_5 Christian + \beta_6 Yoruba + \beta_7 Igbo + \beta_9 Others + v_{0k(Zone)} + \\ & u_{ojk(State)} + e_{ijk(individual)}\end{aligned} \tag{9}$$

The overall aim is to assess the extent to which the observed factors at various levels (individual, state and zone) affect the age at first birth where $e_{ijk(individual)}$ is nested within $e_{ijk(individual)}$ which is further nested within $v_{0k(zone)}$. The variances $\sigma^2{}_e$, $\sigma^2{}_{uo}$ and $\sigma^2{}_{vo}$ represent the variances of random effects due to individual, state and zone respectively. The higher the value of $\sigma^2{}_e$ the greater the degree of differences in the individual women age at first birth. Also, the higher the values of $\sigma^2{}_{uo}$ and $\sigma^2{}_{vo}$, the greater the degrees of differences induced by state and zone clustering respectively and the higher the degree of similarity among fertility experiences of women within the same state and zone respectively.

Since there is possibility of having some of the individual level (level 1) predictors to vary across either state or region, we carry out a post hoc test to know whether the variables vary and to know whether the Random coefficient model is better than the random intercept model by using the log likelihood ratio test.

Results

Table 1 presents the prevalence of early child birth in Nigeria. It presents the intercept only model to determine the average age at which Nigerian women gives birth. As can be observed from the table, the average age at which the average Nigerian woman gives birth is approximately 19 years. The intercept only model is given as;

(Age at First Birth)$_{ijk}$ = λ_{000} = 19.33, where γ_{000} is the grand mean.

Table 1: Intercept only model

Age at first birth	Coefficient	Std. Error	Z	P>|Z|	95% Confidence Interval	
_cons	19.3391	0.0496	389.77	0.000	19.2417	19.4364

Table 2 presents the Null Model with Random intercept. It does not have variables at all levels to calculate the intra-class correlation coefficient between the states and geo-political zones. The intra-class correlation coefficient represents the percentage variation in age at first birth that is between the state and zone variables. The standard deviation of each individual on each age at first birth from its state mean is estimated as 3.7370 while the standard deviation of each state mean within its zone is 0.8842. Also, the standard deviation of each zone from its grand mean to be 1.7918 and the overall grand mean is 20.1506. To calculate the intra-class correlation coefficient (ICC), we used the method illustrated by [12] which defines the intra-class correlations at the state and region level as:

$$\rho_{state} = \frac{\sigma_{v0}^{2} + \sigma_{u0}^{2}}{\sigma_{v0}^{2} + \sigma_{u0}^{2} + \sigma_{e}^{2}} \qquad (10)$$

And

$$\rho_{zone} = \frac{\sigma_{v0}^{2}}{\sigma_{v0}^{2} + \sigma_{u0}^{2} + \sigma_{e}^{2}} \qquad (11)$$

Using (11) and (12), the intra-class correlations at state and zone levels are estimated as 0.2223 and 0.1788 indicating that 22 per cent of the variation in age at first birth stems from differences among the states and 18 per cent stems from differences among the zones.

Table 2: The Null Model with Random intercept

Age at first birth	**Coefficient**	**Std. Error**	**Z**	**P>|Z|**	**95% Confidence**	**Interval**
cons	20.1506	0.7478	26.95	0.000	18.6850	21.6163

Random-effects Parameter	Estimate	Std. Error	95% Confidence Interval	
Zone: Standard deviation	1.7918	0.5415	0.9909	3.2400
State: Standard deviation	0.8842	0.1257	0.6691	1.1684
Individual: Standard deviation	3.7370	0.0313	3.6761	3.7989

Further discussions of effects of the observed factors on age at first birth and the level variances are therefore based on Model 9.

Table 3 presents the Varying intercept model with the individual level predictors. No variable was declared for the state and zone levels. As observed, the marginal effect of delaying sexual initiation by one year increases the expected age at first birth by 0.862. Christians' age at first birth is 0.32 higher than for Muslims; women with Higher education are 0.315 times older than people with No Education. Age of Yoruba women at first birth is 0.370 times higher than that of Hausa women while Igbo women are 0.441 times older than Hausa women.

Table 3: Varying Intercept model with individual level predictors

	Age at first birth	Coefficient	Std. Error	P>|Z|	95% Confidence Interval	
Age at First Sex	Age at first sex	0.862	0.010	<0.001*	0.842	0.881
Highest Level of Education Attained	Koranic	0.051	0.095	0.588	-0.134	0.237
	Secondary	0.149	0.099	0.132	-0.045	0.343
	Higher	0.315	0.146	0.031*	0.028	0.602
Religion	Christian	0.320	0.104	0.002*	0.116	0.525
Ethnicity	Yoruba	0.370	0.177	0.036*	0.024	0.717
	Igbo	0.441	0.206	0.032*	0.037	0.844
	Others	0.030	0.107	0.780	-0.179	0.239

Constant	4.811	0.229	0.000	4.362	5.259

Note: Asterisk (*) indicate the variables that are significant at 5% level of significance

Random-effects Parameter	Estimate	Std. Error	95% Confidence Interval	
Zone: Standard deviation	0.267	0.120	0.110	0.646
State: Standard deviation	0.341	0.060	0.242	0.482
Individual: Standard deviation	2.606	0.022	2.564	2.650

Table 4 presents the Varying intercept and Varying coefficient model with the inclusion of the state level (level two) predictor (place of residence), the inclusion of place of residence at the state level reported the variable; Christianity to be significant which reduces the age at first birth of women who are Christian to 0.286 when compared to Muslim women as against the 0.32 when the state level predictor was not included while the marginal effect of delaying sexual initiation by one year increases the expected age at first birth by 0.859 which also reduces the age at first birth

Table 4: Varying intercept and varying coefficient model with the inclusion of the state level (level two) predictor (place of residence)

	Age at first birth	Coefficient	Std. Error	P>|Z|	95% Confidence Interval	
Age at First Sex	Age at first sex	0.859	0.010	<0.001*	0.839	0.879
Highest Level of Education Attained	Primary	0.068	0.095	0.472	-0.118	0.255
	Secondary	0.140	0.101	0.166	-0.0579	0.3376
	Higher	0.241	0.149	0.106	-0.0513	0.5337
Religion	Christian	0.286	0.105	0.006*	0.081	0.491
Ethnicity	Yoruba	0.309	0.181	0.088	-0.046	0.663
	Igbo	0.383	0.208	0.066	-0.026	0.791
	Others	0.013	0.107	0.900	-0.196	0.223
	Constant	4.839	0.228	0.000	4.393	5.285

Note: Asterisk (*) indicate the variables that are significant at 5% level of significance

Random-effects Parameter	Estimate	Std. Error	95% Confidence Interval	
Zone: Standard deviation	0.2686	0.1251	0.1079	0.6689
State: Independent				
Standard deviation (Urban)	0.4902	0.1029	0.3248	0.7398

Standard deviation (_cons)	0.2991	0.0655	0.1947	0.4594
Individual: Standard deviation				
(Residual)	2.5984	0.0218	2.5560	2.6416

The test to know whether there is significant difference between varying intercept model and varying coefficient model gives *Prob < 0.05* indicating that there is no statistically significant difference between the two models. Therefore, the result of the varying intercept model with individual level predictors is upheld.

Discussion and Conclusion

A study was carried out on the age at first birth of Nigerian women using dataset from 2013 Nigeria Demographic and Health Survey (NDHS). For the study, a three-level model which account for hierarchical structure of the data was used. 22 per cent of the variation in age at first birth of Nigeria women stems from differences among the states and 18 per cent from differences among the zones, the implication is that, both state and zone of residence of the individual women are has effect on the ages at which they have their first child, most importantly the state of the resident. Results of the post hoc test reveals that varying intercept model is not significantly different from the varying the coefficient model. Delaying the onset of sexual initiation of women increases the age at first birth of Nigeria women. Educated women are more likely to give birth to their first child at a later and more matured age than those women with primary education. Christian women were found to give birth to their first child at a later age compared to Muslim women. Among the ethnic groups, Yoruba and Igbo women give birth to their first child at a matured age than the Hausa women.

The finding which is of great concern is the estimated average age at which Nigeria women gives birth to their first child that falls in the teen age. This confirms the statement of [3] which states that in 2008, there were 16 million births to mothers aged 15 – 19 years, representing 11% of all births worldwide [3]. Almost 95% of these births occur in developing countries. In reality, half of all the births occur in just seven countries, Nigeria, Ethiopia, Democratic Republic of Congo, Bangladesh, Brazil and the United States of America [13].

Recommendation

Young people should be educated on the risk of early sexual activity so as to increase the age at first birth among women. Women education should be encouraged and made mandatory as this will enlighten the women on the risk of early child birth as it affects their health status. Sensitization of the risk of early child birth should be put in place among the ethnic groups through their community and the religious leaders. Nigeria as a country should do all within her power to ensure that her girls are protected against early child birth to control the alarming population growth of the country.

References

1.Ajala, O. A. (2014). Factors associated with teenage pregnancy and fertility in Nigeria. *Journal of Economics and Sustainable Development*, 5(2): 62-70.

2. UNICEF (2005), Early Marriage, a Harmful Traditional Practice: A Statistical Exploration, UNICEF.

3.World Health Organization (2011) Early Marriages, Adolescent and Young Pregnancies. Report by the Secretariat Executive Board 130th session EB130/12.

4. Barcena, A. (2013) Teen pregnancy – early warning sign for inequality Tuesday July 23, 2013 Jamaica Observe: Jamaican News Outline Retrieved from www.jamaicaobserver.com/pfversion/Teen-pregnancy -early-warning-sign for inequality Sept 10 2013.

5. Chau-Kuang, C., Ward, C., Willians, K., Abdullah, A. (2013). Investigating risk factors affecting teenage pregnancy rates in the United States. *European International Journal of Science and Technology*, 2: 41-51.

6. Randenbush, S. W., Bryk, A. S. (1986). A Hierarchical Model for Studying School Effects. *Sociology of Education*, 59: 1-17.

7.Randenbush, S. W., Bryk, A. S. (2002). Hierarchical Linear Models: Applications and Data Analysis Methods. Second edition. Newbury Park, CA: Sage Publications.

8.Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 78: 817-827.

9. De Leeuw, J., Kreft, I. G. G. (1986). Random Coefficient Models for Multilevel Analysis. *Journal of Education Statistics*, 11: 57-85.

10. Longford, N. T. (1993). Random Coefficient Models. Oxford University Press, Oxford, UK.

11 Gill, J. (2003). Hierarchical linear models. Kimberly Kempf-Leonard (Ed.), Encyclopedia of social measurement. New York: Academic Press.

12.Siddiqui, O., Hedeker, D., Flay, B. R., Hu, F. B. (1996). Intraclass correlation estimates in a school-based smoking prevention study: outcome and mediating variables by sex and ethnicity. *Am J Epidemiology*, 4:425–433.

13.World Health Organization (2011). Early marriages, adolescent and young pregnancies. Report by the Secretariat Executive Board 130th session EB130/12