

HIERARCHICAL REGRESSION MODELING OF SOME FACTORS AFFECTING WEIGHT OF CHILD AT BIRTH

ABSTRACT

According to World Health Organization (WHO), normal weight of baby at term delivery is (2.5 – 4.2) kilograms. Every child's birth weight below 2.5 kilograms, regardless of gestational age, is regarded as Low birth weight (LBW). WHO estimates that globally, over 20 million LBW babies are born annually and nearly 95.6% of them in developing countries. Half of all perinatal and 1/3rd of all infant deaths occur in babies with LBW. It is therefore, essential to study some of the factors that causes LBW. Hierarchical Multiple Regression analysis was used to study the effects of mother's weight, age and height above and beyond mother's education level in predicting the weight of the child at birth. The results showed that mother's education level explains about 6.1% of the unexplained variations in the weight of the child at birth in block 1 and mother's age, weight and height explained about 3.9% above and beyond mother's education level. It was concluded that mother's education level plays a vital role in predicting the weight of the child at birth because it has a causal effect on the use of prenatal care and improves marriage prospects.

Keywords: Hierarchical Multiple Regression models, Hierarchical linear models, Low birth weight, unexplained variation, infant deaths.

1. Introduction

Birth weight refers to the weight of a baby or young animal at birth. The birth weight of an infant is a reliable index of intrauterine growth and also a sensitive predictor of newborn's chances of survival, growth and long term physical and psychosocial development. The average birth weight in babies of European heritage is 3.5 kilograms (7.7 lb), but the range of normal weight at term delivery as prescribed by World Health Organization (WHO) is 2.5 – 4.2 kilograms (5 pounds 8 ounces – 9 pounds 4 ounces). Low birth weight (LBW) has considerable short and long-term outcomes and may result to high medication costs to the individual and the society at large. Every child's birth weight below the lower bound of the WHO standard (that is less than 2.5 kg) regardless of gestational age [1] is regarded as Low birth weight. There are several determinants of low birth weight (LBW). One of the most relevant is maternal social status, which has a close and direct association with maternal education level. Even in developed countries, mothers in unfavorable socioeconomic status and with low education level present greater vulnerability to having LBW children [2].

Maternal education is a measure of a mother's education level. It affects birth weight by improving the probability and/or productivity of health investment. Additionally, maternal

education improves the financial resources available to the child directly and indirectly through the choice of partner, timing of fertility, and number of offspring (the quantity/quality trade off). The causal effect of education is identified for individuals with low level of education rather than at the upper end of the education distribution. A better understanding how maternal education affects child health may help shed some light on the complexity of the factors involved in a child's well-being. Schultz [3] contended that mothers' education may affect child health in at least five different ways: (1) education may impart better utilization of health inputs in the production of a healthier child; (2) mothers who are more educated may change their perceptions regarding how best to allocate resources for the betterment of children's health; (3) educated mothers may enhance family wealth status—even though many times they do not participate in the labor force but engage in positive assortative mating, marrying wealthier men; (4) schooling may incline parents' preferences for fewer but healthier children; and (5) more educated mothers may ascribe a higher value to their time, particularly when they work outside the home.

A World Health Organization (WHO) multi-country study based on data from 29 African, Asian, Latin American, and Middle Eastern countries found that adolescent mothers (10 - 19 years) were prone to an increased risk of adverse birth outcomes when compared with adult mothers (20 - 24 years) after controlling for covariates [4]. Another study using Demographic Health Survey (DHS) data from 55 low- and middle-income countries (LMICs) found that first-born children, aged 12 - 60 months, of mothers aged younger than 27 years (compared to 27 - 29 years) had higher risk ratios for infant mortality (B12 months) and child stunting, underweight, diarrhea, and anemia among children, after adjusting for parental, household, and social factors [5]. A further study of birth cohorts in five countries found that children born to adolescent mothers aged 19 years or younger (compared to those 20 - 24 years) had an increased risk of low birth weight, preterm birth, and child stunting at 2 years of age, after adjusting for covariates, and in addition, the mothers were at risk of not completing their secondary education [6].

Lobl, Welcher and Mellits [7] have also reported positive relationships between maternal age and infant birth weight. However, when other measures of newborn health status are considered, the children of adolescent mothers appear to be at little or no disadvantage compared to those of older mothers. Although Lester [8] originally found that teenagers' babies scored lower than those of older mothers on the Brazelton scale, when controls for obstetrical and perinatal risks were introduced the differences disappeared.

WHO estimates that globally, out of 139 million live births, more than 20 million LBW babies are born each year, consisting 15.5% of all live births, nearly 95.6% of them in

developing countries [9]. Infants who weigh less than 2.5 kg at birth represent about 26% of all live births in India and more than half of these are born at term. LBW infants are 40 times more likely to die within first four weeks of life than normal birth weight infants. Half of all perinatal and 1/3rd of all infant deaths occur in babies with LBW [1]. LBW is partially a consequence of the choices made by the mother pre and during pregnancy. Thus policies affecting these choices could have large returns as LBW lead to the transmission of inequality between generations [10]. Therefore, it is very crucial to study the factors that causes Low birth weighted babies, identify these factors and recommend ways to cushion their effects in order to reduce deaths amongst new born babies.

Aside the mother's level of education, other variables such as mother's age, mother's weight and mother's height have been found to affect infant birth weight. This study aims to build Hierarchical Multiple Regression model for the prediction of weight of child at birth in Nigeria. The objectives of the study include:

- i) To find the degree of association between weight of child at birth and mother's level of education, height, weight and age.
- ii) To find the amount of variations in weight of child at birth that is accounted for by the mother's level of education
- iii) To find the amount of variations in weight of child at birth is accounted for by the mother's height, weight and age.
- iv) To find the total amount of variations in the weight of child at birth that is accounted for by mother's education level, height, weight and age.

2. Theoretical foundation

2.1 Hierarchical Linear Models

Hierarchical linear models and multilevel models are variant terms for what are broadly called linear mixed models (LMM). These models handle data where observations are not independent, correctly modeling correlated error. Uncorrelated error is an important but often violated assumption of statistical procedures in the general linear model family, which includes analysis of variance, correlation, regression, and factor analysis. Violations occur when error terms are not independent but instead cluster by one or more grouping variables. For instance, predicted student test scores and errors in predicting them may cluster by classroom, school, and municipality. When clustering occurs due to a grouping factor (this is the rule, not the exception), then the standard errors computed for prediction parameters will be wrong (ex., wrong β coefficients in regression). Moreover, effects of predictor

variables may be misinterpreted, not only in magnitude but even in direction. Linear mixed modeling, including hierarchical linear modeling, can lead to substantially different conclusions compared to conventional regression analysis.

Linear mixed models in some disciplines are called “random effects” or “mixed effects” models. In statistics, the term “covariance components models” is often used, alluding to the fact that in linear mixed models one may decompose the covariance into components attributable to within-groups versus between-groups effects. In spite of many different labels, the commonality is that all adjust observation-level predictions based on the clustering of measures at some higher level or by some grouping variable. The “linear” in linear mixed modeling has a meaning similar to that in regression: There is an assumption that the predictor terms on the right-hand side of the estimation equation are linearly related to the target term on the left-hand side. Of course, nonlinear terms such as power or log functions may be added to the predictor side (ex., time and time-squared in longitudinal studies). Also, the target variable may be transformed in a nonlinear way (ex., logit link functions).

2.2 Hierarchical Linear Regression Models

Hierarchical Linear Models on the other hand are a type of multilevel linear regression models in which the observations fall into hierarchical or completely nested levels. It is the idea of building new linear regression models, each by adding more predictors in the preceding linear regression model. Hierarchical regression is a way to show if the independent variables of interest explain a statistically significant amount of variance in the Dependent Variable (DV) after accounting for all other variables. This is a framework for model comparison rather than a statistical method.

In this framework, several regression models are built by adding variables to a previous model at each step such that later models always include smaller models in previous steps. In many cases, our interest is to determine whether newly added variables show a significant improvement in R^2 (the proportion of explained variance in DV by the new model). In hierarchical linear regression, models are fitted to a dataset predicting a single outcome variable (usually); where each model is constructed by adding variables to an initial equation, and computing a deviation R-square (R^2) which is the difference between an initial model (or previous model in the sequence) R^2 and the new model R^2 . This might be done 3 or 4 times, as blocks of variables are added incrementally to an initial block, and their impact assessed on predictive accuracy using the R^2 magnitudes.

For example, a researcher might be interested in the incremental predictive accuracy gained from initially predicting the weight of the child at birth (Y_i) using level of education of the mother, (X_1), then the extra accuracy created by including weight (X_2), age (X_3) and height of the mother (X_4) to predict the same weight of the child at birth.

Model 1 (m_1)

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 X_1 \quad (1.1)$$

Model 1 has $R_{\text{squared}} = R_{m1}^2$

Model 2 (m_2)

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 \quad (1.2)$$

Model 2 has $R_{\text{squared}} = R_{m2}^2$ with $\Delta R_{m_2 - m_1}^2 = R_{m2}^2 - R_{m1}^2$

Conventionally, each model's incremental fit (R^2) over the previous model is tested for statistical significance. This is implemented using an ANOVA approach by [11]

$$F_{n-K}^H = \frac{RSS_{\text{smaller}} - \left(\frac{RSS_{\text{larger}}}{H} \right)}{\left[\frac{RSS_{\text{larger}}}{n-K} \right]} \quad (1.3)$$

where,

F_{n-K}^H = F distribution statistics with H and $n-K$ degree of freedom

RSS_{smaller} = The residual sum of squares for the fewer parameters regression model

RSS_{larger} = The residual sum of squares for the greater parameters regression model

H = The number of parameters for the smaller (fewer parameters) regression model

K = The number of parameters for the larger (greater number of parameters) regression model

n = The total number of cases.

2.3 Hierarchical Linear Modeling vs. Hierarchical Regression

When conducting statistical data analysis, one may be faced with the challenge of running either a "hierarchical regression" or a "hierarchical linear model". At a glance, it may seem like these two terms refer to the same kind of analysis. However, "hierarchical linear modeling" and "hierarchical regression" are actually two very different types of analyses that

are used with different types of data and to answer different types of questions. So, what is the difference between the two?

Hierarchical linear modeling is sometimes referred to as “multi-level modeling” and falls under the family of analyses known as “mixed effects modeling” (or simply “mixed models”). This type of analysis is most commonly used when the cases in the data have a nested structure. For example, collecting data from students where the students in the study come from a few different blood groups. The data so collected is said to consist of students nested within blood groups. The students in study that come from the same blood group will share some common variance associated with being in the same blood group, so those cases cannot be treated as truly independent of one another. Since a conventional multiple linear regression analysis assumes that all cases are independent of each other, a different kind of analysis is required when dealing with nested data in this instance. Hierarchical linear modeling allows one to model nested data more appropriately than a regular multiple linear regression.

Hierarchical regression, on the other hand, deals with how predictor (independent) variables are selected and entered into the model. Specifically, hierarchical regression refers to the process of adding or removing predictor variables from the regression model in steps or blocks. For instance, one may want to predict the weight of the child at birth (Y_i) (the dependent variable) based on weight of the mother (X_2), age of the mother (X_3) and height of the mother (X_4) (the independent variables) while controlling for demographic factors (i.e., covariates like the level of education of the mother (X_1)). For this analysis, one might want to enter the demographic factor (X_1) into the model in the first block, and then enter weight of the mother (X_2), age of the mother (X_3) and height of the mother (X_4) into the model in the second block. This would enable one to see the predictive power that the independent variables add to the model above and beyond the demographic factor. Hierarchical regression also includes forward, backward, and stepwise regression, in which predictors are automatically added or removed from the regression model in blocks based on statistical algorithms. These forms of hierarchical regression are useful if the interest is to determine (statistically) which variables have the most predictive power when a very large number of potential predictor are being considered. In a nutshell, hierarchical linear modeling is used when the data is nested while hierarchical regression is used to add or remove variables from the model in multiple steps. Knowing the difference between these two seemingly similar terms is important as it helps to determine the most appropriate analysis for a given study.

2.4 Review of previous works

Quite a number of works had been done to investigate some of the factors that influence child birth weights in the world. These include [12] which supports the view that the effect of education on health is causal, [13] and [14] tested the causality for developing countries while [15] was the first paper that investigated the issue for a developed country. Using 30 years of Vital Statistics natality, a register of all the births in the United States, [15] used college proximity as an instrument for maternal education. They found that one year of maternal education reduced both the probabilities of low birth weight and premature birth by 1 percentage point. In the UK two studies used change in school leaving age to assess the effect of maternal education on child health. [16] used the 1997-2002 Health Survey of England to identify the effects of parental education and income on self-reported child health. [17] estimated the effects of parental education on several health outcomes by discontinuity design, but found little support for a causal effect. Many investigators have examined the relationship between maternal age and infants' health at birth. Some studies, for example, have found that women who bear children during adolescence are at increased risk of adverse fetal outcome compared to women who delay childbearing until a later age [18, 19].

Hierarchical regression had been used by some researchers in the behavioural sciences to predict factors responsible for some behavioral patterns of humans. [20] used it to generate prediction equations for all of the calculated WASI-II and WAIS-IV indexes. [21] used it to predict depression and self-esteem, their results showed that trait perfectionism dimensions were significant predictors of unique variance after other personality trait measures had been entered. In both instances, the concern over mistakes subscale predicted unique variance beyond the variance attributable to neuroticism and conscientiousness. [22] used it to study the believe by researchers that there are both cognitive (thinking) and affective (feeling) components to attitudes. The results of the study indicate that, for any particular behavior, either cognition or affect will make a statistically significant unique contribution – a finding that should not occur regularly if attitude does not have these two components. [23] also used hierarchical regression analyses to examine the contribution of PASS processes to reading and mathematics at the end of kindergarten and grade 1. [24] conducted multiple hierarchical regression analyses with discrimination and ERI also included in the equations to predict subjective sleep quality and quantity as well as objective sleep duration and wake minutes after sleep onset, respectively. Other works in this regard include [25, 26, 27, 28]. [29] assessed statistical mediation using hierarchical regression model.

Whereas the works reviewed investigated the effect of one or two predictors namely maternal level of education and maternal age on the dependent variable, infants' birth weight, this study investigated the effect of the maternal level of education, maternal age, maternal weight and maternal height on the infants' birth weight.

3. Methodology

The hierarchical Regression modeling uses the Ordinary Least Squares method in each block while explanatory variables are added.

3.1 The Model

The Simple Linear Regression model can be stated as:

$$y_i = \beta_o + \beta_1 x_i + e_i ; i = 1, 2, \dots, n \quad (1)$$

This statement implies a set of n linear equations:

$$\begin{aligned} y_1 &= \beta_o + \beta_1 x_1 + e_1 \\ y_2 &= \beta_o + \beta_1 x_2 + e_2 \\ &\cdot \quad \cdot \quad \cdot \\ &\cdot \quad \cdot \quad \cdot \\ y_n &= \beta_o + \beta_1 x_n + e_n \end{aligned} \quad (2)$$

where each equation represents the y value of a given individual in terms of the parameters of the model, the individual's x values, and an error component. Representing (2) in a matrix form, we have

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix}_{n \times 1} \quad x = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{bmatrix}_{n \times 2} \quad \beta = \begin{bmatrix} \beta_o \\ \beta_1 \end{bmatrix}_{2 \times 1} \quad e = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_n \end{bmatrix}$$

$$y = x\beta + e \quad (3)$$

A typical multiple regression model can therefore be written (in non-matrix form) as

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + e \quad (4)$$

The matrix form can simply be expressed as

$$\underline{y} = \underline{x}\underline{\beta} + \underline{e} \quad (5)$$

where

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \underline{x} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \underline{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

The Normal equation in matrix form is:

Given $y = x\beta + e$ for two independent variables (say x_1, x_2)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e \quad (6)$$

In matrix form, this becomes

$$y = x\beta + e \quad (7)$$

where

$$\underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad \underline{x} = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{bmatrix}$$

The sample estimate of β can be written as

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \quad x'x = \begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ \sum x_{i2} & \sum x_{i1}x_{i2} & \sum x_{i2}^2 \end{bmatrix}$$

$$x'x\hat{\beta} = \begin{bmatrix} n\hat{\beta}_0 & + \hat{\beta}_1 \sum x_{i1} & + \hat{\beta}_2 \sum x_{i2} \\ \hat{\beta}_0 \sum x_{i1} & + \hat{\beta}_1 \sum x_{i1}^2 & + \hat{\beta}_2 \sum x_{i1}x_{i2} \\ \hat{\beta}_0 \sum x_{i2} & + \hat{\beta}_1 \sum x_{i1}x_{i2} & + \hat{\beta}_2 \sum x_{i2}^2 \end{bmatrix} \quad x'y = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \end{bmatrix}$$

The normal equation can be expressed as

$$x'x\hat{\beta} = x'y \quad (8)$$

Hence,

$$\begin{bmatrix} n\hat{\beta}_0 & + \hat{\beta}_1 \sum x_{i1} & + \hat{\beta}_2 \sum x_{i2} \\ \hat{\beta}_0 \sum x_{i1} & + \hat{\beta}_1 \sum x_{i1}^2 & + \hat{\beta}_2 \sum x_{i1}x_{i2} \\ \hat{\beta}_0 \sum x_{i2} & + \hat{\beta}_1 \sum x_{i1}x_{i2} & + \hat{\beta}_2 \sum x_{i2}^2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \end{bmatrix}$$

These are the normal equations and we can solve for three parameter estimates as a system of linear equations. By differentiating the sum of squares of the residual of the normal equation with respect to β , that is

$$\begin{aligned} e'e &= (y - x\beta)'(y - x\beta) \\ &= y'y - 2\beta'x'y + \beta'x'x\beta \\ \frac{\partial(e'e)}{\partial\beta} &= -2x'y + 2\beta x'x = 0 \\ \Rightarrow \hat{\beta} &= (x'x)^{-1}x'y \quad \text{provided that } |x'x| \neq 0 \end{aligned}$$

The Hierarchical Multiple Regression was conducted in two blocks (block 1 containing the response variable and maternal education level, block 2 containing the response variable with addition of mother's age, weight and height to the model). Each model's incremental fit (R^2) over the previous model was tested for statistical significance using ANOVA approach by [11].

3.2 Assumptions of Multiple Linear Regression

Assumption 1: The relationship between the independent variables and the dependent variable is linear. This assumption was tested by producing scatterplot for each of the independent variables and the dependent variable.

Assumption 2: There is no multicollinearity in the data.

This is essentially the assumption that the predictors are not highly correlated with one another. This assumption was tested by checking the Correlations table. Correlations of more than 0.8 was considered problematic. We could also use Variance Inflation Factor (VIF) and Tolerance statistics to assess this assumption. For the assumption to be met we want VIF scores to be well below 10, and tolerance scores to be above 0.2.

Variance Inflation Factor is given by

$$V.I.F = \frac{1}{1 - R_j^2} \quad (9)$$

where R_j^2 is the multiple R – squared for the regression of X_j on the other covariates (a regression that does not involve the response variable).

Tolerance is given by

$$\text{Tolerance} = \frac{1}{V.I.F} \quad (10)$$

Assumption 3: The values of the residuals are independent (or uncorrelated). Here, the Durbin-Watson statistic was used to test the assumption. This statistic can vary from 0 to 4. To meet this assumption, we want this value to be close to 2. Values below 1 and above 3 are cause for concern and may render the analysis invalid.

Assumption 4: The variance of the residuals is constant (i.e. homoscedasticity). This is the assumption that the variation in the residuals (or amount of error in the model) is similar at each point across the model. In other words, the spread of the residuals should be fairly constant at each point of the predictor variables (or across the linear model). To test this assumption, we plotted the standardised values our model would predict, against the standardised residuals obtained. As the predicted values increase (along the X-axis), the variation in the residuals should be roughly similar. If everything is ok, this should look like a random array of dots. If the graph looks like a funnel shape, then it is likely that this assumption has been violated.

Assumption 5: The values of the residuals are normally distributed.

This assumption was tested by looking at the distribution of residuals (i.e. by looking at the P-p plot for the model). The closer the dots lie to the diagonal line, the closer to normal the residuals are distributed.

Assumption 6: There are no influential cases biasing the model.

Significant outliers and influential data points can place undue influence on the model, making it less representative of the data as a whole. Cook's Distance could be used to test for outliers. Values above 1 are likely to be significant outliers and should be removed before the analysis is rerun. In this study however, the box and whiskers plot was used to remove outliers.

3.3 Method of data analysis

The data used were extracted from the 2018 Nigeria Demographic and Health Survey data. The independent variables of interest are maternal weight, maternal height, maternal age and maternal education levels while the dependent variable is the child's birth weight. The

data collection was naturally stratified by maternal education levels (1 = Higher Education, 2 = Secondary Education, 3 = Primary Education and 4 = No Education). We used the equal sample allocation technique, which assigns equal samples sizes ($n_i = 200$ for the i^{th} maternal education level in this case) to all strata irrespective of the stratum population size, stratum variability or cost per unit. However, after removing the rows with possible outliers (in child's birth weight) detected by using the box and whiskers plot, the following sample sizes were used (Higher Education level = 193, Secondary Education = 180, Primary Education = 169 and No Education = 193) giving a total sample size of 735 valid cases. Due to the large volume of data used, we could not include them as appendix in this paper for want of space. The statistical analysis was done using Statistical Package for the Social Sciences (SPSS) version 23 for windows.

4. Results

Table 1: Descriptive Statistics

	Mean	Std. Deviation	Skewness		Kurtosis		N
			Statistic	Std. Error	Statistic	Std. Error	
Weight of Child at Birth	3.1971	.58667	.202	.090	-.019	.180	735
Highest Education Level	2.4925	1.14108	-	-	-	-	735
Mother's Weight (kg)	62.2359	14.10518	1.022	.090	1.091	.180	735
Mother's Age	29.91	6.261	.290	.090	-.263	.180	735
Mother's Height (cm)	158.7490	6.38408	-1.237	.090	9.093	.180	735

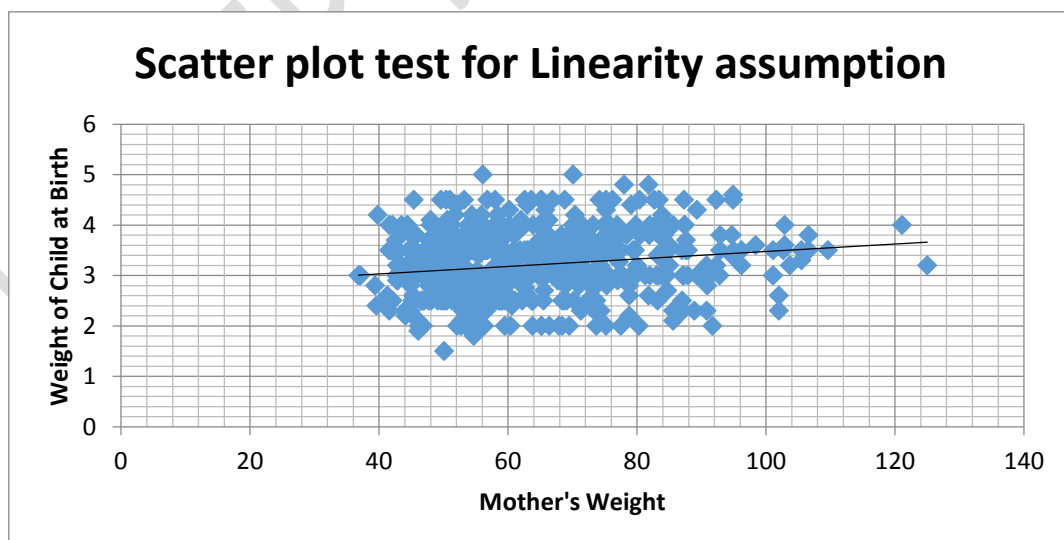


Figure 1: Linearity assumption between mother's weight and weight of child at birth

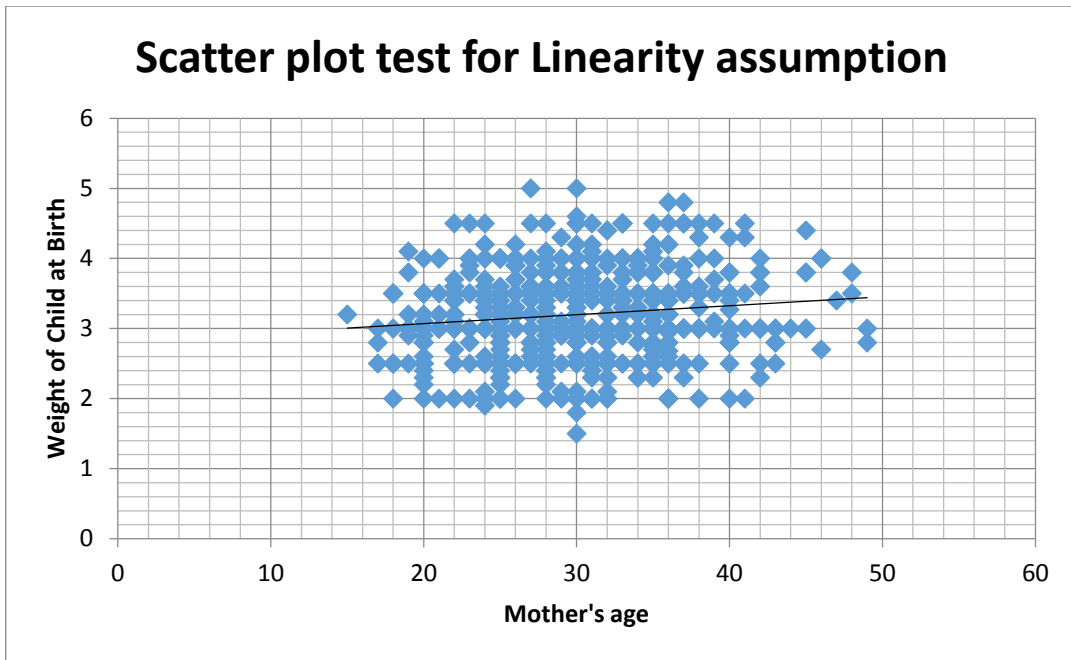


Figure 2: Linearity assumption between mother's age and weight of child at birth

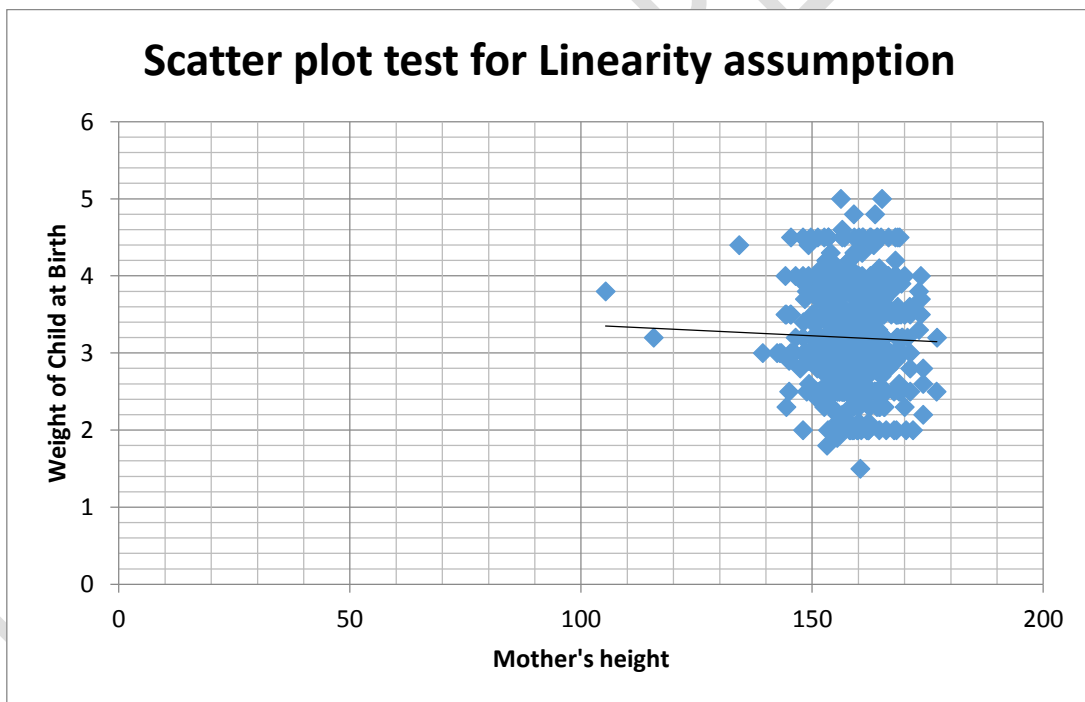


Figure 3: Linearity assumption between mother's height and weight of child at birth

Table 2: Correlations

		Weight of Child at Birth	Maternal Education Level	Mother's Weight (kg)	Mother's Age	Mother's Height (cm)
Pearson Correlation	Weight of Child at Birth	1.000	-.247	.178	.136	-.031
	Maternal Education Level	-.247	1.000	-.249	-.035	-.219
	Mother's Weight (kg)	.178	-.249	1.000	.281	.340
	Mother's Age	.136	-.035	.281	1.000	.019
	Mother's Height (cm)	-.031	-.219	.340	.019	1.000
Sig. (1-tailed)	Weight of Child at Birth	.	.000**	.000**	.000**	.201
	Highest Education Level	.000**	.	.000**	.172	.000**
	Mother's Weight (kg)	.000**	.000**	.	.000**	.000**
	Mother's Age	.000**	.172	.000**	.	.301
	Mother's Height (cm)	.201	.000	.000	.301	.

Footnote: ** sig at 5%.

Table 3: Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	Maternal Education Level	.	Enter
2	Mother's Age, Mother's Height (cm), Mother's Weight (kg)	.	Enter

Table 4: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.247	.061	.060	.56890	.061	47.562	1	733	0.000**	
2	.316	.100	.095	.55812	.039	10.527	3	730	0.000**	1.749

Table 5: ANOVA for Regression

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	15.393	1	15.393	47.562	.000
	Residual	237.234	733	.324		
	Total	252.627	734			
2	Regression	25.230	4	6.308	20.249	.000
	Residual	227.397	730	.312		
	Total	252.627	734			

Table 6: Regression coefficients and t – test for significance of individual regression coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	p-value	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	3.513	.050		69.655	.000		
	Maternal Education Level	-.127	.018	-.247	-6.897	.000	1.000	1.000
2	(Constant)	4.813	.553		8.704	.000		
	Maternal Education Level	-.123	.019	-.238	-6.497	.000	.917	1.091
	Mother's Weight (kg)	.006	.002	.138	3.467	.001	.780	1.281
	Mother's Age	.009	.003	.091	2.489	.013	.914	1.094
	Mother's Height (cm)	-.012	.003	-.132	-3.482	.001	.860	1.163

Table 7: Excluded Variables in the first regression block

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	Mother's Weight (kg)	.124	3.387	.001	.124	.938	1.066	.938
	Mother's Age	.127	3.588	.000	.131	.999	1.001	.999
	Mother's Height (cm)	-.089	-2.447	.015	-.090	.952	1.051	.952

Table 8: Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	2.7782	3.7334	3.1971	.18540	735
Std. Predicted Value	-2.259	2.893	.000	1.000	735
Adjusted Predicted Value	2.7768	3.7439	3.1971	.18554	735
Residual	-1.42365	1.75709	.00000	.55660	735
Std. Residual	-2.551	3.148	.000	.997	735
Cook's Distance	.000	.016	.001	.002	735

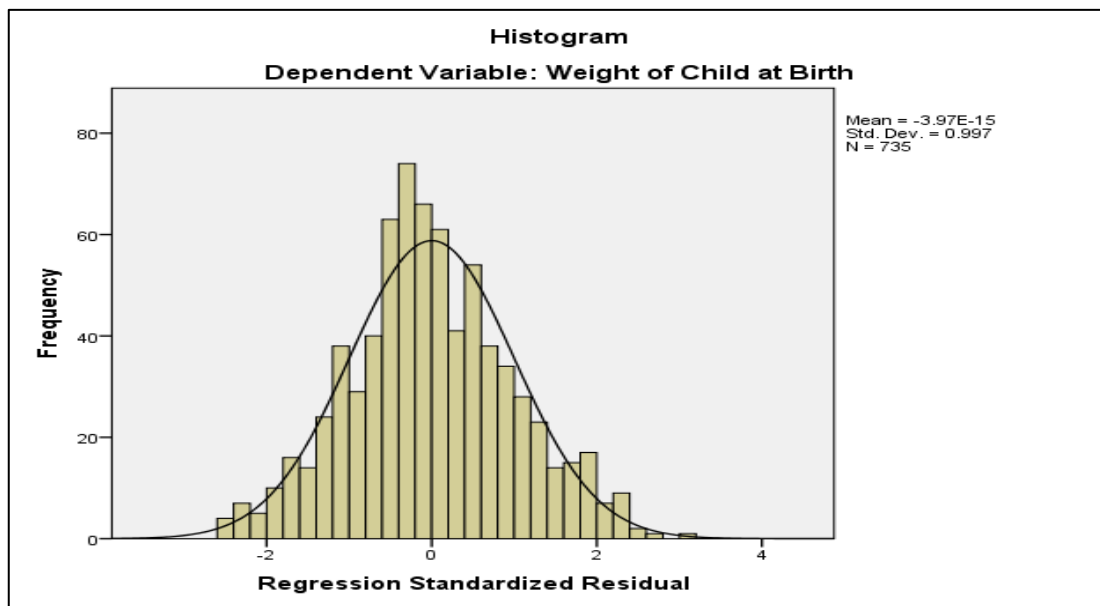


Figure 4: Histogram plot of the standardized regression residuals

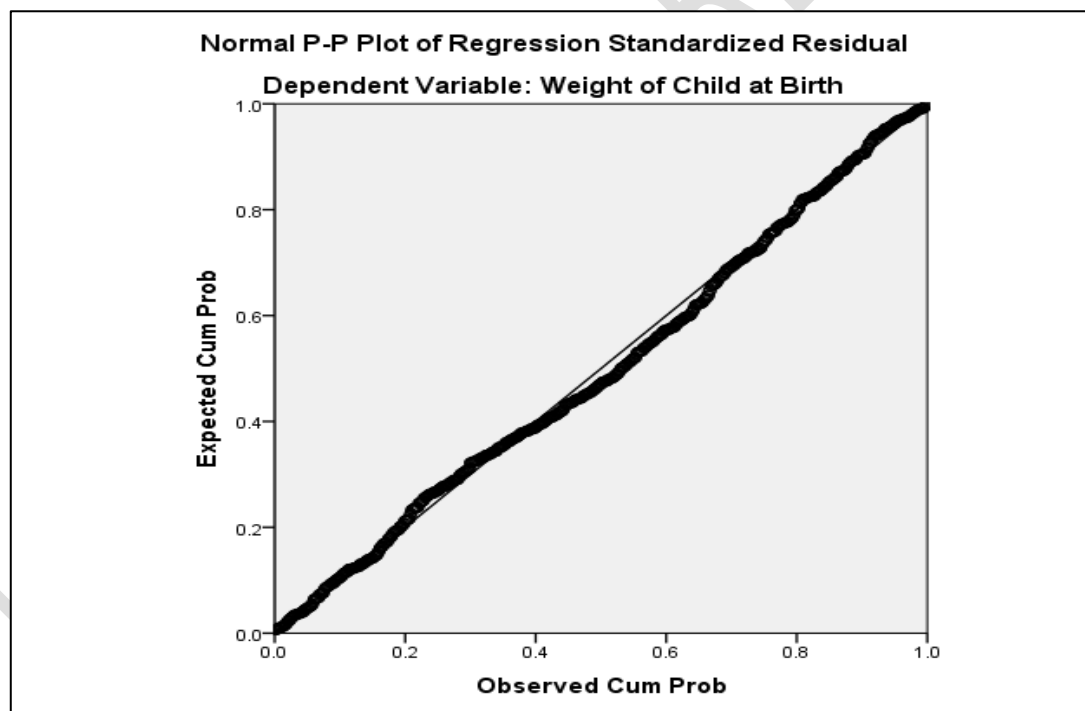


Figure 5: Normality Probability Plot of regression standardized errors



Figure 6: Constant error variance test

5. Discussion

A symmetrical dataset has skewness equal to 0 and kurtosis of 3.0 as a measure of fit for the normal distribution. The continuous independent variables do not appear to have come from normal distribution (skewness: 0.202, 1.022, 0.290 and -1.237, kurtosis: -0.19, 1.091, -0.263 and 9.093). However, our assumption bothers on normality of errors, so we can continue with the analysis. The relationship between the response variable (weight of child at birth) and the independent variables appears to be linear (Figures 1, 2 and 3). While there is a positive linear relationship between weight of child at birth and mother's weight and age (the linear line went from bottom side of the Y -axis towards the top right corner), the relationship between weight of child at birth and mother's height is negative. The implication is that aged mothers that weigh more and are shorter tend to give birth to heavy weighted babies.

We want to see some level of association between the response variable and the independent variables, however, we do not want the degree of association to be too obvious to avoid having best linear unbiased estimators that are not significant with larger variance values and confidence intervals. The correlations between weight of child at birth and every other independent variable are significant ($p\text{-value} < \alpha\text{ value } 0.05$) except with mother's height ($p\text{-value } 0.201 > \alpha\text{ level } 0.05$). Since the significant correlation values are not above 0.80, we continued with the Ordinary Least squares regression since multicollinearity

was not a problem otherwise, we could have looked for regularization regression methods such as Partial Least Squares, Ridge Regression, Principal Component regression, [30] instead.

Having met the linearity and correlation assumptions for regression, we proceeded with the Hierarchical Multiple Regression. The Hierarchical Multiple Regression was conducted in two blocks (Table 3: block 1 containing the response variable and maternal education level, block 2 containing the response variable with addition of mother's age, weight and height to the model). The Hierarchical Multiple regression in block 1 showed that the education level of the mother of the child explains about 6.1% (Table 4, R-squared value) of the unexplained variations in the weight of the child at birth while the block 2 showed that mother's age, weight and age explains additional 3.9% (Table 4, R square change value) of the unexplained variations in the weight of the child at birth above and beyond mother's level of education. The models in block 1 and 2 are statistically significant in the linear prediction of weight of the child at birth (Table 4, Sig. f – change values are less than alpha level of 0.05). Further verification of the absence of multicollinearity was confirmed in Collinearity statistics (Table 5: VIF below 10 and Tolerance above 0.2 for all the regression coefficients in block 1 and 2). The values of the errors from block 2 (the final model used for prediction) are independent (Table 4: Durbin-watson 1.749).

Block 1 Prediction Model:

$$\hat{Y} = 3.513 - 0.247 X_1$$

Block 2 Prediction Model:

$$\hat{Y} = 4.813 - 0.238X_1 + 0.138X_2 + 0.091X_3 - 0.132X_4$$

where \hat{Y} is the weight of child at birth, X_1 is mother's education level, X_2 is mother's weight, X_3 is mother's age and X_4 is mother's height.

Block 1 and 2 models are statistically significant (Table 5, p-values $0.000 < \alpha 0.05$). All the independent variables used for prediction in the final model (block 2) are statistically significant (Table 6, p –values less than alpha 0.05). We ensured that there are no influential or wild numbers in the four levels of mother's education level and this is true (Table 8, maximum cook's distance not greater than 1.0). The prediction errors from block 2 model are normally distributed (Figures 4 and 5) and with constant variance (Figure 6).

Based on the findings/results, we provide here the answers to our research objectives:

- i) Weight of child at birth is positively correlated with mother's weight and mother's age but negatively correlated with mother's height and mother's education level. This means that mothers who weigh more and are older are likely to give birth to heavy weighted babies at birth. Also, shorter and uneducated mothers are likely to give birth to heavy weighted babies. These findings agree with Currie and Moretti (2003) because the causal effect of education is identified for individuals with low level of education rather than at the upper end of the education distribution. The findings are also in agreement with [31] which showed that there are lower rates of stillbirth and neonatal mortality among children of younger mothers than among those of older mothers. [7] have also reported positive relationships between maternal age and infant birth weight.
- ii) Mother's education level accounted for about 6.1% of the unexplained variations in the weight of the child at birth.
- iii) Mother's height, weight and age accounted for an additional 3.9% of the unexplained variations in the weight of the child at birth.
- iv) The block 2 model showed that mother's education level, age, height and weight accounted for about 10% of the unexplained variations in the weight of the child.

6. Conclusion

On the basis of the above findings, we concluded that the mother's age, weight and height are significant in predicting the weight of the child above and beyond the mother's education level. Since weight of child at birth is positively correlated with mother's age and weight, aged mothers are advised to watch their diets and nutrition to avoid giving birth to obese children who are at high risk of dying at birth. We also advise mothers to embrace education as education may impart better utilization of health inputs in the production of a healthier child. Mothers who are more educated may change their perceptions regarding how best to allocate resources for the betterment of children's health and this may enhance family wealth status. Again, schooling may incline parents' preferences for fewer but healthier children.

7. References

- [1] Park. K. (2007). "Park's Text Book of Preventive and Social Medicine". Jabalpur: *M/s. Banarsidas Bhanot Publishers* 17: 426–428.

- [2] Shi L., Macinko J., Starfield B., Xu J., Regan J., Politzer R., et al. (2004). "Primary care, infant mortality, and low birth weight in the states of the USA". *J Epidemiol Community Health*;58:
- [3] Schultz, T.P. (1984). "Studying the Impact of Household Economic and Community Variables on Child Mortality". *Population and Development Review*, 10(suppl.), 215-235.
- [4] Ganchimeg T., Ota E., Morisaki N., Laopaiboon M., Lumbiganon P., Zhang J. et al (2014). "Pregnancy and childbirth outcomes among adolescent mothers: "A World Health Organization multicountry study". *BJOG*; 121(Suppl 1): 40_8.
- [5] Finlay J.E., Ozaltin E. and Canning D. (2011). "The association of maternal age with infant mortality, child anthropometric failure, diarrhea and anaemia for first births: evidence from 55 low- and middle-income countries". *BMJ Open*; 1: e000226.
- [6] Fall C.H., Sachdev H.S., Osmond C., Restrepo-Mendez M.C., Victora C., Martorell R, et al. (2015). "Association between maternal age at childbirth and child and adult outcomes in the offspring: a prospective study in five low-income and middle-income countries (COHORTS collaboration)". *Lancet Glob Health*; 3: e366_77.
- [7] Lobl M., Welcher D.W. and Mellits E.D. (1971). "Maternal age and intellectual functioning of offspring". *Johns Hopkins Med J*; 128:347-357.
- [8] Lester B.M. (1978). "Relations between teenage pregnancy and neonatal behavior". *Progress Report to NICHD, DHEW*.
- [9] Sachdev H.P.S. (2001). "Low Birth Weight in South Asia". *Int. J. Diab. Dev. Countries* 21: 13–29.
- [10] Currie, J. and E. Moretti (2005). "Biology as destiny? Short and long-run determinants of intergenerational transmission of birth weight", *National Bureau of Economic Research*, 11567.
- [11] Hamilton, L.C. (1992) *Regression with Graphics: A Second Course in Applied Statistics*. Belmont, California: Brooks-Cole (see Eq.3.28, page 80)
- [12] Grossman, M (2006). "Education and non-market outcomes", *Handbook of Economics of Education*, Hanushek E. and F. Welsh (Eds), Elsevier, Holland.
- [13] Desai, S. and Alva S. (1998). "Maternal education and child's health: Is there a strong causal relationship?" *Demography*, 35, 71-81.
- [14] Thomas, D., Strauss, J. and Henriques M.H. (1991), "How does mother's education affect child height?" *Journal of Human Resources*, 26, 1, 183–211.
- [15] Currie, J and Moretti, E. (2003). "Mother's education and the intergenerational transmission of human capital: Evidence from college openings", *Quarterly Journal of Economics*, 118, no. 4, 1495-1532.
- [16] Doyle O., Harmon C. and Walker I. (2005) "The impact of Parental Income and Education on the Health of their Children", *IZA Discussion Paper No. 1832*. <https://ftp.iza.org/dp1832.pdf>
- [17] Lindeboom, Maarten, Llena-Nozal, Ana and Van der Clauw, Bas (2006). *Parental Education and Child Health: Evidence from a Schooling Reform*. IZA Discussion Paper No. 2516. <https://ftp.iza.org/dp2516.pdf>

- [18] Baird D. (1945). "The influence of social and economic factors on stillbirths and neonatal deaths". *Obstet and Gynaec Brit Comm* 1945: 52:217.
- [19] Menken J. (1980). "The health and demographic consequences of adolescent pregnancy and childbearing". IN: Chilman C (ed): Adolescent Pregnancy and Childbearing. *Washington DC: Govt Printing Office*.
- [20] James A. Holdnack, Mike R. Schoenberg, Rael T. Lange, Grant L. Iverson, (2013). Chapter 5 - Predicting Premorbid Ability for WAIS-IV, WMS-IV and WASI-II, WAIS-IV, WMS-IV, and ACS, Academic Press, Pages 217-278. <https://doi.org/10.1016/B978-0-12-386934-0.00005-5>.
- [21] Gordon L. Flett, Paul L. Hewitt, (2015). Chapter 21 - Measures of Perfectionism, Measures of Personality and Social Psychological Constructs, Academic Press, Pages 595-618. <https://doi.org/10.1016/B978-0-12-386915-9.00021-8>.
- [22] David Trafimow, (2004). Attitude Measurement, Encyclopedia of Applied Psychology, Elsevier, Pages 233-243, ISBN 9780126574104, <https://doi.org/10.1016/B0-12-657410-3/00181-1>.
- [23] George K. Georgiou, George Manolitsis, Niki Tziraki, (2015). Chapter 12 - Is Intelligence Relevant in Reading "μάνα" and in Calculating "3 + 5"?, Cognition, Intelligence, and Achievement, Academic Press, Pages 225-243. <https://doi.org/10.1016/B978-0-12-410388-7.00012-9>.
- [24] Yuen Mi Cheon, Pak See Ip, Tiffany Yip, (2019). "Chapter Seven - Adolescent profiles of ethnicity/race and socioeconomic status: Implications for sleep and the role of discrimination and ethnic/racial identity, Advances in Child Development and Behavior, JAI, Volume 57, Pages 195-233. <https://doi.org/10.1016/bs.acdb.2019.04.002>.
- [25] Victoria C.P. Knowland, Michael S.C. Thomas, (2011). Chapter Three - Developmental Trajectories in Genetic Disorders, International Review of Research in Developmental Disabilities, Academic Press, Volume 40, Pages 43-73. <https://doi.org/10.1016/B978-0-12-374478-4.00003-4>.
- [26] Marco Weber, Claudia Harzer, E. Scott Huebner, Kimberly J. Hills, (2015). Chapter 5 - Measures of Life Satisfaction Across the Lifespan, Measures of Personality and Social Psychological Constructs, Academic Press, pages 101-130, <https://doi.org/10.1016/B978-0-12-386915-9.00005-X>.
- [27] Fordyce, M. W. (1977). Development of a program to increase personal happiness. *Journal of Counseling Psychology*, 24, 511-521. doi10.1037/0022-0167.24.6.511.
- [28] Timothy A. Salthouse, Julie L. Earles, (1996). Age, Perceived Health, and Specific and Nonspecific Measures of Processing Speed, *Advances in Psychology*, North-Holland, Volume 114, Pages 315-337. [https://doi.org/10.1016/S0166-4115\(96\)80014-8](https://doi.org/10.1016/S0166-4115(96)80014-8).
- [29] Allen, P. A., Ruthruff, E. Lien, M. C. (2007). Attention, *Encyclopedia of Gerontology* (Second Edition), Elsevier, 2007, Pages 120-129. <https://doi.org/10.1016/B0-12-370870-2/00018-4>.
- [30] Olewuezi, N.P., Onoghojobi, B., & Obite, C. P. (2016). Comparison of Linear Prediction Models in Time Series. *Journal of the Nigerian Association of Mathematical Physics*, vol. (34): 151-156.

[31] Mednick B.R. (1979). "Consequences of family structure and maternal state for child and mother's development". *Progress Reports to NICHD, DHEW*.

UNDER PEER REVIEW