Identification of Best Fit Probability Distribution for Infant Birth Weight

Ilesanmi A.O[1]., Ogunboyo Ojo Femi[2]., Odukoya E.A[3]., Aladejana A.E[4]., Ajewole K.P[5]., Popoola O.E.[6].

[1,3,4]Department Of Statistics, Ekiti State University, Ado-Ekiti, Ekiti State, Nigeria.

[2]Epidemiology And Biostatistics Department, University of Medical Sciences Ondo, Ondo State, Nigeria

[5]Department Of Physical Sciences, Mathematics Program, Landmark University, Omu- Aran.

[6]Mathematical Sciences Department, Bamidele Olumilua University of Education, Science and Technology, Ikere.

DOI: https://doi.org/10.51584/IJRIAS.2024.908053

Received: 19 July 2024; Revised: 08 August 2024; Accepted: 13 August 2024; Published: 17 September 2024

ABSTRACT

Birth weight is the first weight recorded for a newborn, taken within 0 to 30minutesofdelivery. It is an indicator of a newborn's chances for survival and growth. This study identified the best fit probability distribution for newborn birth weight in Ekiti State Nigeria. In the year 2021 there were 349 births recorded, with 163 males and 186 females. Also, in 2022 the number of births increased slightly to 353, comprising 179 males and 174 females. The selection of the best fit probability distribution was determined by Goodness of fit such as AIC, BIC, CAIC. Birth weights recorded within 0 to 30 minutes of delivery were obtained from Federal Teaching Hospital Ido-Ekiti and it comprise of both male and female newborn babies between 2021 and 2022.The analysis reveals that the mean birth weight is consistently higher for males compared to females in both 2021 and 2022. The results also showed that Weibull distribution is the least suitable model, with the highest AIC (1073.675), BIC (1077.800), and CAIC (1075.887) values, indicating a poor fit for the birth weight data follow by Normal and Gumbel distributions while the Log-normal distribution has the lowest AIC (737.110), BIC (741.234) and CAIC (739.321). This indicates that Log-Normal distribution is the most appropriate among the four probability distributions considered in this study.

Keywords: Newborn, Birth weight, Maximum likelihood estimation, Goodness of fit.

INTRODUCTION

The birth weight of a new born baby is the first weight recorded after birth, ideally measured within the first hour after birth. World Health Organization (WHO) defined birth weight as the body weight of a baby at their birth. Birth weight is a very important indicator of child survival, future physical growth and mental development. Data on birth weight is essential for monitoring and evaluating the progress towards achieving national goals for lowering neonatal and infant morbidity and mortality (Najmi 2000). Zang *et al* (2020) **also** mentioned that infant birth weight is highly sensitive in two critical aspects; firstly, it is firmly dependent on the health and the nutritional status of the mother. Secondly, the birth weight can determine the chances of survival of a new born as well as healthy growth and development. Subramanyam *et al* (2010) obtained data on birth weight from birth certificates and the third Indian National Family Health Survey which included data on birth weight of children obtained from health cards and maternal recall. They described the population that these data represent and compares the birth weight obtained from health cards with maternal recall data in terms of its socioeconomic patterning and as a risk factor for childhood growth failure, the results suggest that in the absence of other sources, the data on birth weight in the third Indian National Family Health Survey is valuable for epidemiologic research. The WHO (1992) defined birth weight less than 2,500 grams (2.5 kilograms) as low birth weight. The LBW subdivisions include very low birth weight, which is less than 1500 g, and extremely low birth weight, which is less than 1000g. Identifying new born birth weight less than 2500g is critical since below this value infant mortality begin to rise rapidly. Chang (2003) observed that infants weighing less than 2.5 kilograms are approximately 25 to 30 times more likely to die than infants with birth weight exceeding 2.5kg.

According to 2018 Nigerian Demographic and Health survey (NDHS) 2018, the prevalence of low birth weight is about 7%, which ranges form 6.9% in the urban to 7.5% in the rural settings (Fayehun, 2020). Previously, in the NDHS 2008 the estimates of the incidence of birth weight less than 2.5 kilogram in Nigeria to be 14 percent (655 per 1,000), which however varies considerably across social and geographic areas (NPC and ORC Macro, 2009). Najmi (2000) determined the distribution of birth weight among newborns and relationship to specific socio-demographic and medical factors, the results of the study showed that 59% cases of low birth weight were associated with prim parity and grand multiparty and other causes of low birth weights of hospital born Pakistani infants includes twin pregnancy, PROM and severe preeclampsia and eclampsia. Most research on birth weight in Nigeria have focused mainly on risk factors for low birth weight in Nigeria and there are few studies on the most appropriate probability models for newborn birth weight. Identifying the best model for birth weight data, which is an important tool for demographic statistics, is essential. Therefore, this work aims at identifying the most suitable probability distribution that best fit for the newborn birth weights in Ekiti State, Nigeria. Onah e*t al*. (2021) examined the influence of socioeconomic status on birth weight across different regions of Nigeria. The study found that lower socioeconomic status, marked by limited access to healthcare and inadequate maternal nutrition, was significantly associated with lower birth weights. The research highlighted the disparities between urban and rural settings, with urban areas showing slightly higher birth weights due to better healthcare infrastructure. Akinyemi *et al.* (2020) explored the relationship between maternal health services utilization and birth weight in Sub-Saharan Africa, focusing on Nigeria, Ghana, and Kenya. The research found that regular antenatal care visits and skilled birth attendance were positively correlated with higher birth weights. The findings suggest that improving access to maternal health services in these regions could significantly reduce the incidence of low birth weight and associated neonatal complications. Patel *et al*. (2019) investigated the combined effects of environmental and genetic factors on birth weight. The study revealed that maternal exposure to pollutants, particularly in urban areas, coupled with genetic predispositions, significantly affected birth weight outcomes. This research highlights the complex interplay between environmental and genetic factors, suggesting that both need to be considered in public health strategies aimed at improving birth outcomes. A study by Lima *et al*. (2022) analyzed trends in birth weight and maternal health indicators in Brazil over the past decade. The study found that improvements in maternal health services, particularly in prenatal care, were associated with a gradual increase in average birth weights. However, regional disparities persisted, with lower birth weights observed in underdeveloped regions. In Ethiopia, a study by Gebre *et al.* (2021) assessed the impact of maternal nutrition during pregnancy on birth weight. The research found that maternal malnutrition, particularly during the third trimester, was a significant predictor of low birth weight. The study emphasized the critical role of nutritional interventions during pregnancy in improving birth outcomes and reducing neonatal mortality rates in resource-limited settings.

2.1 Data and Methodology

This study used birth weight recorded within 0 to 30 minutes of delivery from Federal Teaching Hospital Ido-Ekiti, Ekiti State, Nigeria. The data comprises of both male and female newborn babies weight between 2021 and 2022.The primary objective was to identify the best-fit probability distribution for these birth weights. Maximum likelihood was used to estimate the parameters of Gumbel, Normal, Log-normal and Weibull. These models were chosen due to their widespread use in data analysis and their ability to capture diverse data patterns. Similarly, the selection of the best fit probability distribution is determined by Goodness of fit such as AIC, BIC, CAIC.

2.1.2 Gumbel Distribution

The probability density function of Gumbel distribution is given as

$$f(x) = \frac{1}{\sigma} exp^{-\left(\frac{x-\mu}{\sigma}\right)} exp^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}} \qquad x > 0, \ \sigma > 0 \qquad (1)$$

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VIII August 2024

where μ is the location parameter and σ is the scale parameter.

2.1.3 Normal Distribution

The probability density function of Normal Distribution is given as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad x > 0, \ \ \sigma > 0 \tag{2}$$

where μ is the location parameter and σ is the scale parameter.

2.1.4 Log Normal Distribution

The probability density function of Log Normal Distribution is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma x}} exp^{-\frac{(lnx-\mu)^2}{2\sigma^2}} \qquad x > 0, \ \ \sigma > 0 \tag{3}$$

where μ is the location parameter and σ is the scale parameter.

2.1.5 Weibull Distribution

The probability density function of Weibull distribution is given as

$$f(x) \ = \frac{\alpha}{\sigma}\left(-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha-1} exp^{-\left(-\left(\frac{x-\mu}{\sigma}\right)\right)^{\alpha}}\right) \qquad x > 0, \qquad \sigma \ > 0 \tag{4}$$

where μ is the location parameter, α is the shape and σ is the scale parameter

2.2 Parameter Estimation

The Maximum Likelihood Estimation (MLE) method was employed to estimate the parameters of the various probability distributions. MLE is a powerful statistical tool used to find the parameter values that maximize the likelihood function, given the observed data. In this study, MLE was applied to estimate the parameters of the Gumbel, Normal, Log-normal, and Weibull distributions. This method was used because it provides efficient and consistent parameter estimates, making it well-suited for the analysis of birth weight data.

2.2.1 Maximum Likelihood Estimation for Gumbel Distribution

The location and scale parameters of Gumbel distribution given in equ (1) is obtained as follows:

$$L(\mu, \sigma) = f\left(x_{1,} x_{2,} \dots, x_{n,} \ / \ \mu, \sigma\right) = \prod_{i=1}^{n} f(x_i \ / \ \mu, \sigma)$$

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma} exp^{-\left(\frac{x_i-\mu}{\sigma}\right)} exp^{-exp^{-\left(\frac{x_i-\mu}{\sigma}\right)}} \tag{5}$$

$$L = \left(\frac{1}{\sigma}\right)^n exp^{-\Sigma\frac{x_i-\mu}{\sigma}} exp^{-\Sigma exp^{-\left(\frac{x_i-\mu}{\sigma}\right)}} \tag{6}$$

The log-likelihood function is given as

$$In(L) = -nIn(\sigma) - \sum \frac{x_i-\mu}{\sigma} - \sum exp^{-\left(\frac{x_i-\mu}{\sigma}\right)} \tag{7}$$

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VIII August 2024

Differentiate w.r.t to the μ and σ parameters individually and set the derivatives equal to zero

$$\frac{\partial l}{\partial \mu} = \frac{n}{\sigma} - \frac{1}{\sigma}\sum exp^{-\left(\frac{x_i-\mu}{\sigma}\right)} = 0 \quad (8)$$

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^2}\sum(x_i - \mu) - \sum\left(\frac{x_i-\mu}{\sigma^2}\right) exp^{-\left(\frac{x_i-\mu}{\sigma}\right)} = 0 \quad (9)$$

Equ (8) and (9) will be solved by numerical approach to obtain the estimates of the location and scale parameters

2.2.2 Maximum Likelihood Estimation for Normal Distribution

The likelihood and log-likelihood function of Normal distribution are similarly obtained as follows:

$L(\mu, \sigma)$ =$f\left(x_{1,}x_{2,} \dots, x_{n,} / \mu, \sigma\right)$=$\prod_{i=1}^{n} f(x_i / \mu, \sigma)$

$$L\ (\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} exp^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad (10)$$

$$l = -nln\sigma\sqrt{2\pi} - \sum\frac{-(x-\mu)^2}{\sigma^2} \quad (11)$$

Differentiate with respect to μ and σ^2individually and setting the derivatives equal to zero

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2}\sum(x - \mu) = 0 \quad (12)$$

$$\frac{\partial l}{\partial \sigma} = \frac{-n}{2\sigma^2} + \sum\frac{-(x-\mu)^2}{2(\sigma^2)^2} = 0 \quad (13)$$

Solve Simultaneously

$$\hat{\mu} = \bar{x} \quad (14)$$

$$\widehat{\sigma^2} = \sum\frac{(x-\mu)^2}{\sigma^2} \quad (15)$$

2.2.3 Maximum Likelihood Estimation for the Log-Normal distribution

From eq (9) the resulting likelihood and the log-likelihood functions can be obtained as

$$L\ (\mu, \sigma) = \ (2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}}\prod_{i=1}^{n}\left(\frac{1}{x}\right) exp^{\sum\frac{-(lnx-\mu)^2}{2\sigma^2}}$$

$$\ln(L) = \ -\frac{n}{2}ln\pi - \ \frac{n}{2}ln\sigma^2 + \sum lnx - \frac{1}{2\sigma^2}\sum(lnx - \mu)^2$$

Differentiate with respect to $\mu\ and \sigma^2$ individually and setting the derivatives equal to zero

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma}\sum(lnx - \mu)\ = 0 \quad (16)$$

$$\frac{\partial l}{\partial \sigma} = \ -\frac{n}{2\sigma^2} + \ \frac{1}{2\sigma^4}\sum(lnx - \mu)^2 = 0 \quad (17)$$

Solve simultaneously

$$\hat{\mu} = \frac{\sum lnx}{n} \quad (18)$$

$$\widehat{\sigma^2} = \frac{\sum(lnx-\mu)^2}{n} \quad (19$$

2.2.4 Maximum Likelihood Estimation for the Weibull distribution

From eq (4), the resulting likelihood and the log-likelihood functions can be obtained as

$$L(\mu,\sigma,\alpha) = \alpha^n \sigma^{-n} \prod_{i=1}^{n} \left(\frac{x-\mu}{\sigma}\right)^{\alpha-1} exp^{\sum -\left(\frac{x-\mu}{\sigma}\right)^{\alpha}} \quad (20)$$

$$\ln(L) = -nln(\alpha) - nln\sigma + (\alpha-1)\sum ln\left(\frac{x-\mu}{\sigma}\right) + \sum - \left(\frac{x-\mu}{\sigma}\right)^{\alpha} \quad (21)$$

Similarly, the likelihood function is also maximized w.r.t μ, σ *and* α to obtain the following system of log-likelihood equations

$$\frac{\partial l}{\partial \mu} = -\frac{(\alpha-1)}{\sigma}\sum\left(\frac{x-\mu}{\sigma}\right)^{-1} + \frac{\alpha}{\sigma}\sum\left(\frac{x-\mu}{\sigma}\right)^{-\alpha-1} = 0 \quad (22)$$

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} - \frac{(\alpha-1)}{\sigma}\sum\left(\frac{x-\mu}{\sigma}\right)^{-1}\left(\frac{x-\mu}{\sigma}\right) + \frac{\alpha}{\sigma}\sum\left(\frac{x-\mu}{\sigma}\right)^{\alpha} = 0 \quad (23)$$

$$\frac{\partial l}{\partial \alpha} = -\frac{n}{\sigma} + \sum\left(\frac{x-\mu}{\sigma}\right)^{\alpha} - \sum\left(\frac{x-\mu}{\sigma}\right)^{\alpha} \ln\left(\frac{x-\mu}{\sigma}\right)^{\alpha} = 0 \quad (24)$$

Eq (22), (23) and (24) will be solved by numerical approach to obtain the estimates of the location, scale and shape parameters.

GOODNESS OF FIT TEST

Goodness of fit test is important for identifying how adequate the probability distributions is to the series of the recorded birth weight data. It is a test that helps to determine the best distribution that fits the data. In this work, three different criterions are used to compare the probability distributions given in the study, these criteria are

1. The Akaike information criterion (AIC) can be calculated from this formula:

   $$\mathrm{AIC} = -2\log(\mathrm{L}) + 2 \times \mathrm{k}$$

   Where k is the number of estimated parameters

2. Bayesian Information Criterion (BIC) its formula is very close to the Akaike information criterion (AIC). It can be defined as:

   $$\mathrm{BIC} = -2\log(\mathrm{L}) + \mathrm{k} \times \log(\mathrm{n})$$

   Where n is the number of observations (sample size) and k is the number of estimated parameters

3. Consistent Akaike Information Criterion (CAIC). It can be defined as:

   $$\mathrm{CAIC} = -2\log(\mathrm{L}) + \mathrm{k}\left(1 + \log\left(\frac{\mathrm{n}}{\mathrm{k}}\right)\right)$$

where k is the number of estimated parameters for probability.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue VIII August 2024

ANALYSIS AND DISCUSION OF RESULTS

Table 1: The Birth Weight Statistics by Year and Gender

Year	Gender	Total Number of Births	Mean (SD)	Median (Range)	Min (kg)
2021	Male	163	2.98(0.82)	3.1(3.72)	0.98
	Female	186	2.8(0.84)	2.96(4.72)	0.58
2022	Male	179	2.98(0.86)	3.05(4.3)	0.5
	Female	174	2.79(0.85)	2.85(4.4)	0.5

Interpretation: It was observed from Table 1 above that in 2021, there were 349 births recorded, with 163 males and 186 females. Also, in 2022 the number of births increased slightly to 353, comprising 179 males and 174 females. The mean birth weight is higher for males compared to females in both years, and the medians follow a similar trend. There is a slight increase in the variability for male birth weights from 2021 to 2022, while female birth weights show consistent variability over the two years. Both genders exhibit a wide range in birth weights, with an increase in the range for males from 2021 to 2022, whereas a slight decrease is observed for females.

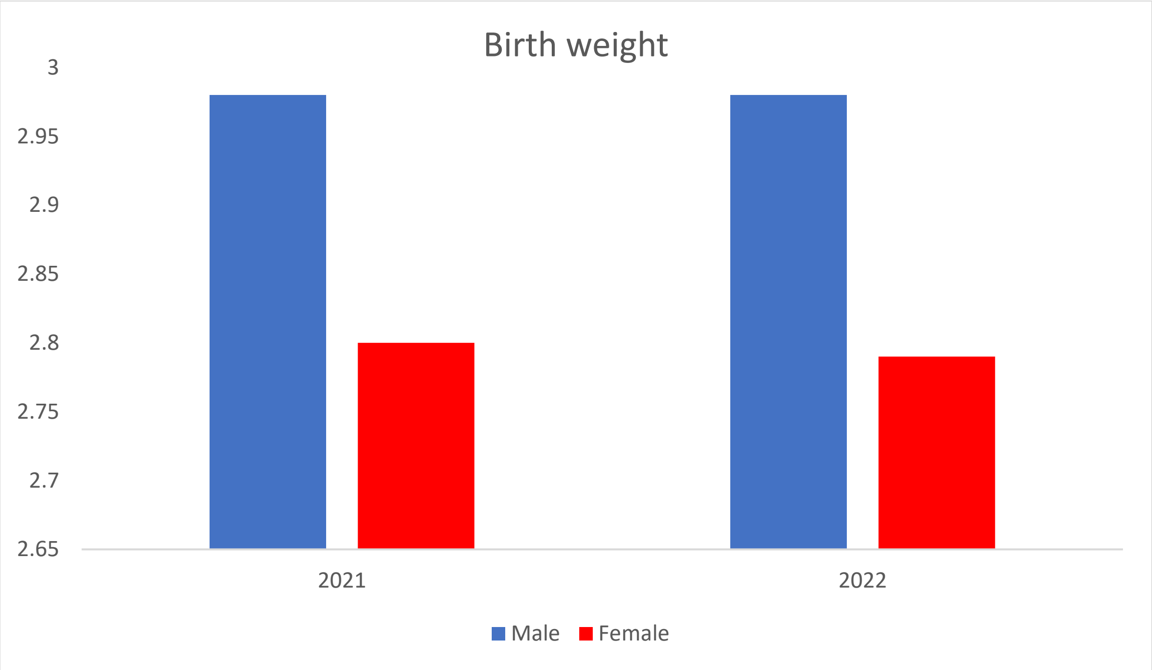

Fig. 1: The chart above illustrated the mean birth weights for males and females between 2021 and 2022.

Table 2: Maximum likelihood estimate for Gumbel, Normal, Log-normal and Weibull models

Parameters	μ	σ	α
Gumbel	1.19	0.78	
Normal	2.942	0.432964	
Log-normal	3.076	0.353	
Weibull	3.153	1.733	1.934

Table 3: Results of the AIC, BIC and CAIC

	AIC	BIC	CAIC
Gumbel	765.034	768.792	767.287
Normal	759.452	763.577	751.664
Log-normal	737.110	741.234	739.321
Weibull	1073.675	1077.800	1075.887

Interpretation: The results from Table 3 showed that Log-normal distribution has the lowest Akaike's Information Criterion, Bayesian Information Criterion and Consistent Akaike Information Criterion value.

DISCUSSION OF RESULTS

In this study, we identified the best-fitting probability distribution for newborn birth weights in Ekiti State, Nigeria, using the birth weight data collected from the Federal Teaching Hospital Ido-Ekiti for both male and female between 2021 and 2022.In 2021, there were 349 births recorded, with 163 males and 186 females. Also, in 2022 the number of births increased slightly to 353, comprising 179 males and 174 females. Four probability distributions were considered which include the Gumbel, Normal, Log-normal, and Weibull distributions. The best fitting probability distribution was selected using goodness-of-fit criteria: AIC, BIC, and CAIC. It was observed from Table 1 that male newborns had higher mean birth weights than females, with males averaging 2.98 kg in both years. The median birth weights follow a similar pattern, with males having slightly higher medians. The results from the Table 2 shows the estimate of the parameters using maximum likelihood and the analysis of the goodness-of-fit criteria in the Table 3 indicates that the Log-normal distribution is the best fit for the birth weight data, with the lowest AIC (737.110), BIC (741.234), and CAIC (739.321) values among the four probability distributions considered. This suggests that the distribution of birth weights is positively skewed, with most newborns having lower weights and a few having much higher weights. The Normal distribution, with AIC (759.452), BIC (763.577), and CAIC (751.664) values, is a close contender but slightly less suitable. The Gumbel distribution, having higher AIC (765.034), BIC (768.792), and CAIC (767.287) values, is even less appropriate, indicating that extreme values in birth weights are not as significant as the Gumbel model would suggest.

On the other hand, the Weibull distribution is the least suitable model, with the highest AIC (1073.675), BIC (1077.800), and CAIC (1075.887) values, indicating a poor fit for the birth weight data. This implies that the characteristics of the Weibull distribution do not align well with the observed data, making it an inappropriate choice for modeling newborn birth weights in this study. Overall, the Log-normal distribution emerges as the most accurate model, effectively capturing the skewed nature of the birth weight data.

CONCLUSIONS

In this paper, we identify the best fit probability distribution for newborn birth weight recorded within 0 to 30minutes. The results from Table 3 showed that Log-normal distribution has the lowest Akaike's Information Criterion, Bayesian Information Criterion and Consistent Akaike Information Criterion value. This indicate that Log-normal distribution is the most appropriate among the four probability distributions considered in this work.

REFRENCES

1. Akinyemi, A. I., Adedini, S. A., & Odimegwu, C. O. (2020). Utilization of Maternal Health Services and Birth Weight in Sub-Saharan Africa: A Comparative Study of Nigeria, Ghana, and Kenya. *Health Care for Women International, 41(6)*, 642-658. https://doi.org/10.1080/07399332.2019.1682083
2. Boutaleb, Y., Lahlou, N., Oudghiri, A., & Mesbahi, M. Le (1982). Poids de naissance dans un pays

Africain [Birth weight in an African country]. *J Gynecol Obstet Biol Reprod (Paris), 11(1)*, 68-72. French. PMID: 7096955.
3. Chang, S., O'Brien, K.O., Nathanson, M.S., Mancini, J., & Witter, F.R. (2003). Hemoglobin concentrations influence birth outcomes in pregnant African-American adolescents. *Journal of Nutrition, 133*, 2348-55.
4. Fayehun, O., & Asa, S. (2020). Abnormal birth weight in urban Nigeria: An examination of related factors. *PLoS One, 15(11)*, e0242796. https://doi.org/10.1371/journal.pone.0242796. PMID: 33232372; PMCID: PMC7685448.
5. Gebre, A., Desalegn, T., & Bekele, B. (2021). Maternal Nutrition and Birth Weight in Ethiopia: The Importance of Nutritional Interventions during Pregnancy. *Journal of Global Health, 11(2)*, 04014. https://doi.org/10.7189/jogh.11.04014
6. Lima, G., Silva, M., & Ribeiro, A. (2022). Trends in Birth Weight and Maternal Health Indicators in Brazil: A Decadal Review. *Brazilian Journal of Maternal and Child Health, 22(3)*, 234-243. https://doi.org/10.1590/0102-311X00043221
7. Najmi, R.S. (2000). Distribution of birthweights of hospital born Pakistani infants. *J Pak Med Assoc, 50(4)*, 121-4. PMID: 10851832.
8. National Population Commission (NPC) (Nigeria) and ORC Macro (2009). Demographic and Health Survey, 2008. Calverton, Maryland: National Population Commission and ORC Macro.
9. Onah, H., Ezeh, O., & Nduka, I. (2021). Socioeconomic Status and Birth Weight Disparities in Nigeria. *BMC Pregnancy and Childbirth, 21(1)*, 132. https://doi.org/10.1186/s12884-021-03642-9
10. Patel, R., Singh, A., & Sharma, K. (2019). Environmental and Genetic Determinants of Birth Weight: Insights from an Indian Cohort. *Journal of Environmental Health, 82(2)*, 54-62. https://doi.org/10.1080/0020739X.2019.1571745
11. Subramanyam, M.A., Ackerson, L.K., & Subramanian, S.V. (2010). Patterning in Birthweight in India: Analysis of Maternal Recall and Health Card Data. *PLoS ONE, 5(7)*, e11424. https://doi.org/10.1371/journal.pone.0011424
12. WHO. (1992). Low birthweight: A tabulation of available information. Geneva: WHO.
13. Zang, X., Liu, Y., Huang, W., & Zhang, X. (2020). Infant birth weight and its critical determinants: Maternal health and nutritional status. *Journal of Pediatrics*, *220*(4), 473-482. https://doi.org/10.1016/j.jpeds.2020.03.014