ESI Preprints Not Peer-reviewed

Assessing Biometric Predictors of Gestational Age Among Ghanaian Women: The Quadratic Classifier Approach

Abukari Alhassan, PhD
Department of Statistics, Faculty of Physical Sciences University for Development Studies, Ghana

Doi: 10.19044/esipreprint.9.2025.p497

Approved: 22 September 2025
Posted: 24 September 2025

Cite As:
Alhassan, A. (2025). *Assessing Biometric Predictors of Gestational Age Among Ghanaian Women: The Quadratic Classifier Approach.* ESI Preprints. https://doi.org/10.19044/esipreprint.9.2025.p497

Abstract

Aims/Objectives: Research has revealed that some illnesses, such as gestational diabetes, might affect a woman's gestation differently. Not all pregnant women who are diagnosed of gestational diabetes, experience gestational variations. Therefore, studying other factors influencing gestational variations amongst women is essential. **Subject/Methods**: The biostatistics department at the War Memorial Hospital in Navrongo, the capital of the Kasena-Nankana municipality in Ghana's Upper East, provided the study's data. It includes information on women and the children who were born to them between January 2014 and January 2017 with a sample size of 1085. The study examined the impact of some characteristics of women on their gestational variation using a quadratic discriminant analysis. **Results:** The main factors influencing the separation of women into different gestational stages in the Municipality are parity, age, and the weight of the unborn child. **Conclusion:** However, the study found that parity was the most important factor in classifying women into below Estimated Date of Confinement, within Estimated Date of Confinement, and above Estimated Date of Confinement of gestation.

Keywords: Estimated Date of Confinement, Gestation, Discriminant Analysis, Quadratic Classification, Preterm Birth

Introduction

The time between conception and birth is referred to as gestation. The gestational age is determined by the final day of the menstrual cycle. 40 weeks from the last day of menstruation, or 280 days, make up the typical gestation period. The due date is referred to medically as the estimated date of confinement (EDC). However, only 4% of women deliver within the EDC (Ohuma E. *et al*., 2023). Women's gestational differences result in preterm and post-term birth, in addition to normal birth. Because of the major contribution of prematurity and post-maturity to morbidity and mortality, the World Health Organization (WHO) put forward some measures in 2010 in order to minimize mortality related to premature deliveries by 50% between 2010 and 2015.

Prematurity, on the other hand, is predicted to be the top cause of mortality in Ghana for the first month of life, with roughly 29,000 newborn deaths each year (UNICEF, 2015). Furthermore, there are various health hazards connected with post-term pregnancy, including lethal macrosomia, according to the American Academy of Family Physicians (AAFP). This has the potential to result in pediatric diabetes, obesity, and metabolic syndrome. When having a large baby, mothers are additionally susceptible to uterine ruptures, genital tract injuries, and excessive bleeding after delivery. A baby that has not yet been born after 42 weeks of gestation - two weeks longer than the typical 40 weeks - is referred to as a post-term pregnancy (WHO, 2021). The birth of a child before 37 weeks of gestation is referred to as preterm birth (PTB) or early birth. Based on the number of gestational weeks, preterm birth is divided into a number of subcategories. Examples include extremely preterm (less than 28 weeks), very preterm (between 28 and 32 weeks), and moderate to late preterm (between 32 and 37 weeks). Studies in the U. S. A. indicated that PTB affects about 9 – 10% of all pregnancies (Martin and Osterman, 2018), and a significant determinant of neonatal mortality and morbidity (March of Dimes Report Card, 2021). Preterm birth (PTB) is also found to be linked with increased risk of maternal mortality and morbidity with adversative health outcomes later in life in both mothers (notably, increased risk of type 2 diabetes and cardiovascular disease), (Henderson et al., 2016; Wu et al., 2018) and their babies (notably, developmental disabilities, asthma, and metabolic syndromes) (Parkinson et al., 2013; Sonnenschein-van der Voort et al., 2014).

According to a worldwide action report on preterm delivery, 13.4 million newborns are delivered too soon per annum (5-18% of all births, WHO, 2021). Preterm birth affects more than one out of every ten babies worldwide. According to the survey, the rate of preterm deliveries increased in several countries between the 1990s and 2010. Nonetheless, preterm births make up 11.1% of all live deliveries globally, with South Asia and Sub-Saharan Africa accounting for 60% of these cases. In these circumstances (28 weeks), more than 90% of extremely preterm neonates die within the first few days of life, and approximately, 7% of all pregnancies are post-term (Ohuma E. *et al*., 2023). Bakhteyor *et al*. (2012) used a logistic regression model to assess factors that affect premature labor in women, and results showed that preterm mothers had a history of obstetric problems, low birth weight, stillbirth, and abortion. According to the findings, the frequency of preterm labor among mothers under the age of 20 was 5.83 times greater than in mothers between the ages of 20 and 35.

According to a 2016 study on maternal height, premature births worldwide are presumably caused in part by maternal short height, probably due to anatomical restrictions. Furthermore, according to a UNICEF and WHO (2021) publication on facts of life, women between the ages of 15 and 18 are more likely to have premature babies, whereas women over 35 are more likely to have post-term birth. Yamoah (2014) used binary quantile and logistic regressions to determine the causes of preterm birth and the resulting cause and effect. It was discovered that 336 out of a total of 711 newborns were prematurely delivered; this means that nearly 4 out of every 9 babies are delivered prematurely in Ghana's Ahafo Ano South District in the Brong-Ahafo Region. The research found that the newborn's weight, the mom's age, intermittent preventive care, along with the number of conceptions were major determinants of preterm birth using binary logistic regression. Evans K. A., et at., (2021), studied the determinants of preterm survival in a tertiary hospital in Ghana from 2010 to 20119 at the Cape Coast Regional Hospital, and reported a prevalence of live preterm babies over the ten-year period of 4.7% with an increased trend in prevalence observed in 2019 recording the highest at 9%.

According to the background research, the deployment of control strategies to reduce the incidence of these gestational disparities depends on identifying the causes of gestational variances in women within a group. The current study aims to use discriminant analysis to classify the gestational period of mothers in the Navrongo municipality of Ghana's Upper East Region based on various characteristics of the pregnant women and their respective neonates.

Methods

Sample

The Biostatistics Unit at the War Memorial Hospital in Navrongo provided data for this study. The data includes information on women and their children from January 2014 to January 2017, for a total sample size of

1085. The information gathered included the mothers' age, height, parity, and gestational period, as well as pregnancy complications and birthweight**.**

Study Area

The War Memorial Hospital in Navrongo, Kasena-Nankana Municipality, was used for the study, where records of women who had given birth were collected. The municipality is located in the Upper East Region. The region has thirteen (13) administrative districts and municipalities in total. The municipality has a population of roughly 27,306 (Ghana Statistical Service, 2012). The districts have a relatively high level of agricultural outputs, along with the raising of goats and cattle, as the primary and oldest occupations of the locals. The only hospital in the community is the War Memorial hospital, serving as a referral hub for smaller clinics in the area, the facility also provides ready access to primary healthcare consultation.

Discriminant Analysis

Discriminant analysis (DA) is a multivariate statistical technique used to determine which variables discriminate between two or more naturally occurring groups. Through DA, one may classify women into two or more mutually exclusive and exhaustive groups on the basis of a set of independent variables.

Linear Discriminant/Classification Model ($\boldsymbol{\Sigma i = \Sigma j = \Sigma}$)

Assume that the two populations $\pi_1$and π_2 have multivariate normal densities $X' = [x_1, x_2, \ldots, x_p]$ and that their respective mean vectors and covariance matrices are, μ_1, Σ_1 and μ_2, Σ_2 correspondingly given by

$$f_i(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(x - \mu_i)'\Sigma^{-1}(x - \mu_i)\right] \; for \; i = 1, 2 \tag{1}$$

The allocation rule that minimizes the expected cost of misclassification (ECM) is given by: Allocate x_0 to π_1 if:

$$(\mu_1 - \mu_2)'\Sigma^{-1}x_0 - \frac{1}{2}(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 + \mu_2) \geq \ln\left[\left(\frac{c\left(\frac{1}{2}\right)}{c\left(\frac{2}{1}\right)}\right)\left(\frac{p_2}{p_1}\right)\right] \tag{2}$$

Allocate x_0 to π_2 otherwise (Johnson and Wichern 2007).

The population parameters in (2) can be replaced by its sample estimates; $\bar{x}_1, \bar{x}_2$ and S_{pooled}. Given a special case when there are equal prior probabilities and equal misclassification cost, then we assign x_0 to π_1 if:

$$(\bar{x}_1 - \bar{x}_2)' \, S_{pooled}{}^{-1} \, x - \frac{1}{2}(\bar{x}_1 - \bar{x}_2)' \, S_{pooled}{}^{-1} \, (\bar{x}_1 + \bar{x}_2)' \geq 0 \tag{3}$$

We can estimate additional discriminant functions, such as the one shown above, when there are more than two groups. When there are three groups, for instance, we could estimate two functions: one to distinguish between group 1 and the combination of groups 2 and 3, and another to distinguish between group 2 and group 3. We could, for instance, have a function that distinguishes between high school graduates who attend college and those who do not (rather, go on to get a job or attend a professional school), and another function that distinguishes between graduates who attend professional schools and those who do not. The coefficients in those discriminant functions could then be interpreted as before.

The Quadratic Classification Model ($\boldsymbol{\Sigma i \neq \Sigma j}$)

The density ratio serves as the decision boundary or the minimum estimated cost of misclassification $f_1(x)/f_2(x)$. Substituting multivariate normal densities with different covariance matrices into (1) after taking natural logarithms and simplifying, the resulting classification regions are:

$$R_1: -\frac{1}{2}x'(\Sigma_1^{-1} - \Sigma_2^{-1})x + (\mu'_1\Sigma_1^{-1} - \mu'_2\Sigma_2^{-1})x - K \geq \ln\left[\left(\frac{c(1/2)}{c(2/1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

$$R_2: -\frac{1}{2}x'(\Sigma_1^{-1} - \Sigma_2^{-1})x + (\mu'_1\Sigma_1^{-1} - \mu'_2\Sigma_2^{-1})x - K \geq \ln\left[\left(\frac{c\left(\frac{1}{2}\right)}{c\left(\frac{2}{1}\right)}\right)\left(\frac{p_2}{p_1}\right)\right] \tag{4}$$

By substituting sample estimates for population parameters, the allocation function that reduces the expected cost of misclassification is obtained, and the minimal ECM is given as follows:
Allocate x_0 to π_1 if:

$$-\frac{1}{2}x'_0(S_1^{-1} - S_2^{-1})x_0 + (\overline{x'}_1 S_1^{-1} - S_2^{-1}\overline{x'}_2)x_0 - K \geq \ln\left[\left(\frac{c\left(\frac{1}{2}\right)}{c\left(\frac{2}{1}\right)}\right)\left(\frac{p_2}{p_1}\right)\right] \tag{5}$$

Allocate x_0 to π_2 otherwise (Johnson and Wichern 2007).
Where,

$$K = \frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + \frac{1}{2}(\overline{x'}_1 S_1^{-1}\overline{x_1} - \overline{x'}_2 S_2^{-1}\overline{x_2}) \tag{6}$$

$\left(\frac{c(1/2)}{c(2/1)}\right)$ is the expected cost ratio and $\left(\frac{p_2}{p_1}\right)$ is the prior probability ratio.

If we assume that there are equal prior probability and misclassification costs for each population, the allocation rule becomes,

$$-\frac{1}{2}x'_0(S_1^{-1} - S_2^{-1})x_0 + (\overline{x'}_1 S_1^{-1} - S_2^{-1}\overline{x'}_2)x_0 - K \geq 1 \tag{7}$$

Error Rate Estimation

The holdout or cross-validation approach was used to assess the performance of the classification function. This method usually holds one observation and classifies the hold-out observation. The process is repeated until all observations are classified producing unbiased estimates of the misclassification probabilities (Lachenbruch and Mickey 1968).

Organization of Data

Three categories; below EDC, within EDC, and above EDC, were used to categorize the gestational differences of mothers. Conceptually, the various gestational categories were viewed as follows: Within EDC is defined as delivery between 38 to 40 weeks of gestation, above EDC is defined as delivery after 41 weeks of gestation or above, while below EDC is defined as delivery in exactly 37 weeks of gestation or less. Some characteristics of the mothers and their neonates were examined quantitatively as the study's independent variables. These factors were as follows: the weight of the infant, the mother's height, her parity, and her age.

Results and Discussion

Table 1 displays the general characteristics of mothers and their newborns. According to the findings, women who give birth below the EDC have a mean age of 25, with a standard deviation of 5.91, those who give birth within the EDC have a mean age of 26 with a standard deviation of 6.64 and those who give birth above the EDC have a mean age of 31 with a standard deviation of 6.63. The result did not show much variance in mean maternal height between the various categories of time differences of birth. As compared to the parity of 1 for the two categories of women who give birth below and within the EDC of gestational differences, the results demonstrated a higher parity of 2 for women who deliver above EDC. Additionally, the results showed that babies born below EDC had a mean weight of 2.76 kg, kids born within EDC had a mean weight of 2.98 kg, and babies delivered above EDC had a mean weight of 2.95 kg. This demonstrates that babies born within the EDC are generally heavier than those born above or below the EDC.

Table 1: The Descriptive Statistics for Selected Variables

Variables	Below EDC		Within EDC		Above EDC	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Maternal Age	25.370	5.9149	26.791	6.6394	31.634	6.6345
Maternal Height	160.120	3.0584	160.670	3.4716	160.656	3.3281
Parity	1.295	1.1457	1.597	1.2031	2.527	1.2648
Baby's Weight	2.763	0.4251	2.980	0.3943	2.951	0.4252

Quadratic Discriminant Analysis

The Box M test of equality of population covariance matrices was initially run in order to test the three groups under consideration for equal covariance matrices. The log determinant of the groups was as illustrated in Table 2. The Box M test was found to be significant at 1% level under the null hypothesis of equal covariance matrices, showing a violation of the assumption of equal covariance matrices.

Table 2: Test for Equality of Population Covariance Matrices

Gestation	Rank	Log Determinant	Chi Square	df	P value
Below EDC	4	3.224	50.365523	20	0.0002*
Within EDC	4	3.508			
Above EDC	4	3.200			
Pooled	4	3.431			

*Significant at 1%

A diagnostic test for multicollinearity also revealed that there was no multicollinearity among the variables because the variance inflation (VIF) values of the independent variables ranged from 1 to 10, as shown in Table 3. Violations of the normality assumption are typically not "fatal" as long as it is cause by skewness and not outliers (Tabachnick and Fidell, 1996). Linearity assumption in discriminant analysis is frequently ignored, unless transformed variables are used as new predictor variables (Clelok, 2017).

Table 3: Test for Multicollinearity

Statistic	Baby's Weight	Maternal Height	Parity	Age
Tolerance	0.9255	0.9227	0.2942	0.2939
VIF	1.0806	1.0838	3.3985	3.4030

The data was then fitted with a quadratic classification function. The quadratic classifier's results demonstrated a significant performance at 1% significant level (Table 4).

Table 4: Test of Model Adequacy

Test Statistic	Value	F Value	Num DF	Den DF	P Value
Wilks' Lambda	0.87260139	19.02	8	2158	<.0001*
Pillai's Trace	0.13040666	18.83	8	2160	<.0001*
Hotelling-Lawley Trace	0.14255142	19.22	8	1539.1	<.0001*
Roy's Greatest Root	0.11168620	30.16	4	1080	<.0001*

* Significant at 1%

Table 5 presents the result of classification and misclassification rates. 32.98 % of the women were correctly classified as Below EDC of gestation with a misclassification rate of 67.02%. However, 14.94% of women Within EDC of gestation were misclassified and 85.06% correct classification was achieved. The results further indicated that for women above EDC of gestation, 7.53% were correctly classified whiles 16.13% and 76.34% were misclassified into below EDC and within EDC respectively. Consequently, an overall error rate of 0.3963 was achieved under the classification model. Further, the cross-validation option provides better assessment of classification accuracy. For this data, 84.42% of women who gave birth Within EDC were classified correctly with a misclassification rate of 15.58% into the Below EDC category. From the result it can be observed that approximately 60.37% (1–0.3963) correct classification of gestation was achieved under classification with QDF as well as 60.00% (1–0.4000) correct classification rate under the cross validated results.

Table 5: Quadratic Function Classification Results

	Classified			
	Below EDC	**Within EDC**	**Above EDC**	**Total**
True/Original				
Below EDC	124	252	0	376
Percent	32.98	67.02	0.00	100.00
Within EDC	92	524	0	616
Percent	14.94	85.06	0.00	100.00
Above EDC	15	71	7	93
Percent	16.13	76.34	7.53	100.00
Total	231	847	7	1085

Percent	21.29	78.06	0.65	100.00
Error Rate	0.6702	0.1494	0.9247	0.3963
Priors	0.3465	0.5677	0.0857	
Cross Validation				
Below EDC	124	252	0	376
Percent	32.98	67.02	0.00	100.00
Within EDC	96	520	0	616
Percent	15.58	84.42	0.00	100.00
Above EDC	15	71	7	93
Percent	16.13	76.34	7.53	100.00
Total	231	847	7	1085
Percent	21.29	78.06	0.65	100.00
Error Rate	0.6702	0.1558	0.9247	0.4000

The eigenvalue and canonical correlation coefficient were also used to examine the performance of the discriminant function. The canonical correlation's strength indicates how well the discriminant function can distinguish between different groups. According to Johnson and Wichern (2007), the total structure coefficient is deemed beneficial if it is equal to or higher than 0.30. The eigenvalue and canonical correlation coefficient in Table 6 demonstrate a well-defined model. The hypothesis that the canonical correlation in the current row and all that follows are zero indicated significance at 5 % level of significance which showed that QDF was correctly specified.

Table 6: Test of Canonical Correlation

	Can. Corr.	**Adj. Can. Corr.**	**Approx. SE**	**Square Can. Corr.**	**Eigenvalue**
Function 1	0.316963	0.310859	0.027321	0.100466	0.1117
Function 2	0.173035	0.169916	0.029463	0.029941	0.0309
Test	**Likelihood Ratio**	**F Value**	**Df 1**	**Df 2**	**P – Value**
Function 1	0.87260139	19.02	8	2158	<.0001*
Function 2	0.97005892	11.11	3	1080	<.0001*

* Significant at 5%

The univariate test of class means (Table 7) reveals the minimum number of variables necessary for discrimination as well as the relevance of each variable in discrimination. The findings show that parity, age, and baby's weight were all significant at 1% ($P < 0.01$). While maternal height was significant at 5% ($P < 0.05$). The R-square and the adjusted R-square values shows the amount of variation explained by each discriminating variable. Parity, Age and baby's weight explained large proportions of the variability (7.45%, 6.63% and 6.28%) among the classes and hence indicating their level of contribution to the group separation (Table 7). In contrast to a previous study on maternal height by Derraik et al. (2016), which found that globally, idiopathic preterm births are likely influenced by maternal small stature, partly because of anatomical limitations., the results of this study showed that maternal height was not a contributing variable to the group separation. However, the findings showed that a woman's age had an impact on the gestational differences, which supported a previous publication by UNICEF, WHO, UNESCO, UNFPA, UNDP, UNAIDS, WFP, and the World Bank (2010) on facts of life, which stated that women between the ages of 15 and 18 are more likely to give birth prematurely, while those over 35 are more likely to have post-term birth.

Table 7: Univariate Test of Class Means

Variable	Total SD	R-Square	Adjusted R-Square	F value	P value
Parity*	1.2313	0.0694	0.0745	40.32	<.0001
Age*	6.5996	0.0622	0.0663	35.86	<.0001
Maternal Height**	3.3289	0.0062	0.0062	3.35	0.0353
Baby's Weight*	0.4200	0.0591	0.0628	33.99	<.0001

* Significant at 1%; ** Significant at 5%

The Receiver Operating Characteristic (ROC) curve is shown in Figure 1. ROC curve is a useful way to interpret sensitivity and specificity levels and to determine related cut scores. The area under the curve (AUC) of a ROC curve represents the overall diagnostic accuracy. The findings of this investigation supported the model's correct specification with an Area Under the Curve (AUC) of 65.4% which was fairly high and significant P value at 5% level.

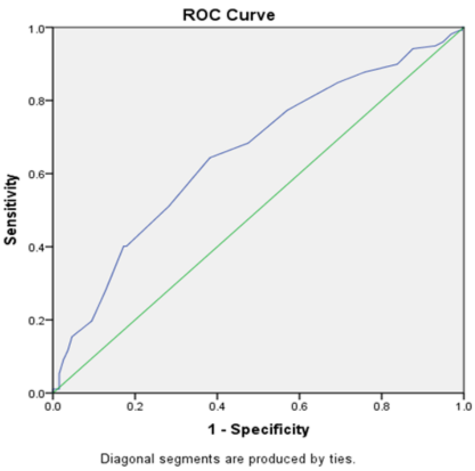

Figure 1. The Receiver Operating Characteristic (ROC)

The structural matrix in Table 8 was used to analyze the significance of each variable in the discriminant function. According to the findings, the first function's main discriminating variables were parity, age and baby's weight, while the second function's only most important discriminating variable is baby's weight. Therefore, these elements are what contribute to birth time discrepancies. However, because parity had the highest structural coefficient of the three factors, it was determined that it was the most important amongst the three variables.

Table 8: Structure Matrix

Variables	Function 1	Function 2
Parity	0.762873	-0.603169
Age	0.710869	-0.616739
Maternal Height	0.216592	0.220065
Baby's Weight	0.645211	0.760002

Table 9 displays the standardized and unstandardized canonical discriminant coefficients of the QDF for below EDC, within EDC, and above EDC of gestational differences in women, with first canonical class means of -0.40, 0.14, and 0.72 and second canonical class means of -0.11, 0.13, and -0.43 respectively. The score functions for the quadratic discriminant analysis are computed using the standardized canonical discriminant function coefficients in the table. From the results it was observed that for both functions (1 and 2), baby's weight had the greatest magnitude amongst the other variables. To classify future observations of pregnant women, the unstandardized canonical discriminant function coefficients of Table 9 can be used. For each function, women are classified as belonging to the class whose canonical coefficient is closest to the class mean.

Table 9: Unstandardized and Standardized Canonical Discriminant Coefficients

Variables	Unstandardized		Standardized	
	Canonical 1	Canonical 2	Canonical 1	Canonical 2
Parity	0.464871473	-0.252724515	0.5724168710	-0.3111909080
Age	0.039490654	-0.056991583	0.2606240095	-0.3761238007
Maternal Height	0.018835953	-0.001096003	0.0627038206	-0.0036485304
Baby's Weight	1.541879715	1.865502707	0.6476329653	0.7835637492

The results of this study supported earlier research on post-term and preterm births that found parity, age, and baby's weight to be important variables influencing the time variations in birth (Kalogiannidis, et al., 2011, UNICEF, WHO, and other organizations 2010). The finding also supports the conclusions made by Marie D. et al., (2017) that, common risk factors underpin changes in the gestational age distribution among women.

Conclusions

In this study, the causes of gestational variations among women in the Navrongo Municipality of Ghana's Upper East Region were examined. The findings indicate that parity, age, and the weight of the new born were the three main factors used to classified women in the research area. The study found that parity was the most important factor in discriminating women in the study area. The study revealed that parity was the most influential discriminating variables either between below EDC, within EDC and above EDC. Research should focus on possible biochemical explanations for this relationship, including the cellular mechanism behind parity and gestational difference. Health monitoring targeted identifiable population risk factors for EDC may provide a useful preterm birth prevention paradigm.

Acknowledgments

We are grateful to the administrator at War Memorial Hospital for authorizing the release of the data used for the study.

Conflict of Interest: The authors reported no conflict of interest.

Data Availability: All data are included in the content of the paper.

Funding Statement: The authors did not obtain any funding for this research.

References:

1. Bakhteyar, K., Lorzadeh, N., Pouria, Y., Birjardi, M., Ebrahimzadeh, F., and Karman, A., (2012). Factors associated with preterm delivery in women admitted to hospitals in Khorramabad: A case control study. *International Journal of Health and Allied Sciences*, 1(3): 147-152.
2. Célok, T. (2017). *Testing the assumptions for discriminant analysis.* Retrieved from :http://www.tankonyvtar.hu/hu/tartalom/tamop425/0049_08_quantitative_information_forming_meth Accessed 5 April 2017.
3. Derraik, J., Lundgren, M., Cutfield, W., & Ahlsson, F. (2016). Maternal height and preterm birth: A Study on 192,432 Swedish Women. *PLos ONE*.
4. Evans K. A., et at., (2021). Determinants of preterm survival in a tertiary hospital in Ghana: A ten-year review. LoS One. 22;16(1). doi: 10.1371/journal.pone.0246005.

5. Ghana Statistical Service (2021). 2010 Population and Housing Census (PHC) Final Results. [Online] Available: www.statsghana.gov.gh/docfiles/2010phc/ (October 28, 2012)
6. Klecka, William R. (1980). *Discriminant analysis*. Quantitative Applications in the Social Sciences Series, No. 19. Thousand Oaks, CA: Sage Publications
7. Lachenbruch PA, Mickey MA (1968). Estimation of error rates in discriminant analysis. Technometrics 10(15):1–11
8. Tabachnick, Barbara, G., & Fidell, L. S. (1996). *Using Multivariate Statistics.* New York,NY: HarperCollins College Publishers.
9. Marie D., Laust M., Ashna D. H., Beatrice B., Mika G., Michael R. K., Jennifer L. R., Paromita D., Jocelyn R., Naho M., Natasha N., Francisco B., Sylvie B., Anne-Marie N. A., and Michael S. K., Jennifer Z. (2017). International variations in the gestational age distribution of births: an ecological study in 34 high-income countries. *The European Journal of Public Health*, Vol. 28, No. 2, 303–309. *doi:10.1093/eurpub/ckx131.*
10. March of Dimes Report Card (2021). March of Dimes Report Card. March of Dimes Report Card
11. March of Dimes, Partnership for Maternal, New born & Child Health (PMNCH), Save the Children, and World Health Organization (WHO), (2021). Born Too Soon: The Global Action Report on Preterm Birth. Eds CP Howson, MV Kinney, JE Lawn.
12. Ohuma E, Moller A-B, Bradley E (2023). National, regional, and worldwide estimates of preterm birth in 2020, with trends from 2010: a systematic analysis. Lancet. (In Press).
13. Yamoah, N. A. (2014). *Using Binary Logistic and Quantile Regressions for Determinants of Preterm.*
14. Zhou, X. H., Obuchowski, N. A., & Obushcowski, D. M. (2002). *Statistical methods diagnostic medicine.* Wiley & Sons: New York.