

Bayesian Sequential Estimation of Proportion of Orthopaedic Surgery Among Different Age Groups: A Case Study of National Orthopaedic Hospital, Igbobi-Nigeria

Rotimi K. Ogundeji¹, Ademola J. Adewara^{2*}, Tajudeen S. Nurudeen³

¹Department of Mathematics, Faculty of Science, University of Lagos, Lagos, Nigeria

²Distance Learning Institute, University of Lagos, Lagos, Nigeria

³Department of Mathematics, Lagos State Polytechnic, Ikorodu, Lagos, Nigeria

Abstract Bayesian methods provide more intuitive and meaningful inferences than likelihood-only based inferences. This is simply because Bayesian approach includes prior information as well as likelihood. In empirical Bayes (EB) methodology, we use data to help determine the prior through estimation of the so-called hyperparameters. In this paper, a Bayesian model called Beta-binomial conjugate model is employed using Bayesian sequential estimation method to estimate the proportion of different age groups attended to at the National Orthopaedic hospital, Igbobi, Nigeria. Over the years results show that the highest number of patients at the hospital is within the age group 15 to 44 years but with the smallest proportion of orthopaedic surgeries. Similarly, smallest the numbers of patients are among the age group less than one year and greater than 64 years but with highest proportion of orthopaedic surgeries. Also, overall EB proportion of patients admitted for orthopaedic surgeries in the hospital across the age groups increased steadily. Finally, the results of the comparative analysis of the sample and EB proportions show that the EB estimators are better estimators on the basis of efficiency and consistency.

Keywords Empirical Bayes, Beta-binomial model, Hyperparameters, Population Proportion, National Orthopaedic Hospital, Igbobi-Nigeria

1. Introduction

The considerable impact of Empirical Bayes (EB) on statistical applications continues to receive increasing popularity since its introduction about four decades ago. The EB structure combines information from several but similar sources. The primary interest in EB analysis is in the hyperparameters (η) rather than the parameters from individual studies (θ). More generally, the hierarchical structure allows for assessment of heterogeneity both within and between groups [1]. Thus Bayesian approach to parameter estimation, when conditions of data allow, is to estimate the posterior distribution of the parameter(s) in question (θ) so that inference on θ is then based on the posterior distribution. A prior distribution for θ is needed to derive the posterior and, in some cases, the prior may have its own parameters, called hyperparameter(s) (η). Quite often, η is unknown to the analyst, in which case the prior

is not completely specified. One way of resolving this problem is through empirical Bayes (EB) analysis. More importantly, EB can lead to more precise estimates than sampling theory approaches [2]. EB analysis also provides a more dependable ranking of parameters and aids in the identification of extreme values in the group[3]. These properties of EB derive from the fact that it uses related supplementary data which frequentist inference ignores; (for more on this see[4]. The EB concept was first proposed by[5] in a non-parametric setting. Some notable works in this regard include[6],[7],[8],[9],[10],[11],[12],[13],[14] and[4]. More recent works and applications include [15],[16],[17],[18],[19], and[20], and[21].

The national orthopedic hospital, Igbobi, Nigeria; was specifically commissioned to provide professional orthopaedic surgeries among health institutions in Nigeria. By orthopaedic surgeries or services, the study refers to the branch of surgery concerned with conditions involving the musculoskeletal system (i.e. the body's bones (the skeleton, muscles, cartilage, tendons, ligaments, joints, and other connective tissue that supports and binds tissues and organs together). This study identifies other category of treatments or services rendered by the orthopaedic hospital as non-orthopaedic (e.g. minor injuries resulting from motor

* Corresponding author:

adewaraja@yahoo.com (Ademola J. Adewara)

Published online at <http://journal.sapub.org/statistics>

Copyright © 2012 Scientific & Academic Publishing. All Rights Reserved

vehicle and motorcycle accidents, domestic and industrial accidents, gunshots, sport injury e.t.c.) because such cases can be easily handled by the general hospitals.

In an attempt to determine the proportion of orthopaedic cases handled by the hospital especially among different age group, we employed a Bayesian model called Beta-binomial conjugate model using Bayesian sequential estimation method. From the hospital record, data were collected and collated over a period of three years (2009, 2010 and 2011). The study also compares the computed sample and EB estimates over the three year period and also the variances of the computed estimates.

The remainder of the paper is organized as follows: Bayesian Sequential methodology is described in section 2. Beta- Binomial model is presented in section 3. Section 4 presents the results of the application of Beta- Binomial model and Bayesian Sequential method along with comparative analyses of results. Section 5 concludes the paper.

2. The Bayesian Sequential Methodology

If a set of observations $x_1, x_2, x_3, \dots, x_n$ generates a posterior distribution and, in a similar situation, additional data are collected beyond these observations, then the posterior distribution found with earlier observations becomes the new prior distribution and the additional observations give a new posterior distribution and inference can be made from the second posterior distribution. This procedure can continue with even more observations. In other words, the second posterior becomes the new prior, and the next set of observations give the next posterior from which the inference can be made [22]. This is the principle of Bayesian sequential methodology that we applied to estimate the proportion of counted data obtained from the hospital.

Based on the Bayesian approach described above, data were collected monthly and collated yearly for three years (2009, 2010 & 2012) from the hospital records. The population proportion of patients admitted for orthopaedic surgery is denoted by P_0 while the proportion of patients admitted for orthopaedic surgery in age group j is P_j ($j = 1, 2, \dots, 5$). X_{ij} represents a random outcome of patient i examined in age group j .

$$Y_{ijk} = \begin{cases} 1 & \text{if } i\text{th Patient is admitted} \\ & \text{for orthopaedic surgery in} \\ & \text{age group } j \text{ and in year } k \\ 0 & \text{Otherwise} \end{cases}$$

$$f(P_{jk} | Y_{jk}, \eta) = \frac{\binom{n_{jk}}{Y_{jk}} P_{jk}^{Y_{jk}} (1-P_{jk})^{n_{jk}-Y_{jk}} \frac{1}{B(r,s)} P_{jk}^{r-1} (1-P_{jk})^{s-1}}{\int P_{jk}^{Y_{jk}} (1-P_{jk})^{n_{jk}-Y_{jk}} \frac{1}{B(r,s)} P_{jk}^{r-1} (1-P_{jk})^{s-1} dP_{jk}}, \quad \eta = (r,s) \quad (2)$$

$$Y_{jk} = \sum_{i=1}^{n_{jk}} Y_{ijk} = \text{the total number of patients admitted}$$

for orthopaedic surgery in age group j in year k .

n_{jk} = the total number of patients admitted for treatments (both orthopaedic and non-orthopaedic surgeries) in the hospital in age group j in year k .

$$P_{jk} = \frac{Y_{jk}}{n_{jk}} = \text{the proportion of patients admitted for}$$

orthopaedic surgery in age group j and year k .

For each year in each age group, we computed sample proportions P_{jk} as follows:

$$\text{In 2009 and age group } j: P_{j1} = \frac{Y_{j1}}{n_{j1}}$$

$$\text{In 2010 and age group } j: P_{j2} = \frac{Y_{j2}}{n_{j2}}$$

$$\text{In 2011 and age group } j: P_{j3} = \frac{Y_{j3}}{n_{j3}}$$

$$\text{Estimators of sample proportions: } \hat{P}_{jk} = \frac{Y_{jk}}{n_{jk}} \quad \text{and}$$

$$\text{Var}(\hat{P}) = \frac{\hat{P}_{jk}(1-\hat{P}_{jk})}{n_{jk}}.$$

3. The Beta-Binomial Model

The EB model to be applied is a conjugate beta-binomial model where the binomial distribution represents the likelihood of the observed data likelihood and the beta distribution serves as the prior distribution of the binomial parameter. The posterior mean is

$$\tilde{P}_{jk} = \int P_{jk} f(P_{jk} | Y_{jk}, \eta) dP_{jk} \quad (1)$$

A key component of this integral is $f(P_{jk} | Y_{jk}, \eta)$, The posterior distribution of which is P_{jk} . Under the general Bayesian framework and using the beta conjugate prior plus the binomial likelihood, the posterior distribution of P_{jk} is:

There is need to estimate the hyperparameters r and s of the beta distribution in order to completely specify the prior. This can be achieved easily through reparameterisation of $f(P_{jk} | \eta)$, and using moment estimation[16]. Letting

$$P_o = \frac{r}{r+s}; M = r+s \text{ and using the prior distribution of } P_{jk}; E(P_{jk}) = P_o \text{ and } Var(P_{jk}) = \frac{rs}{(r+s+1)(r+s)^2} = \frac{P_o(1-P_o)}{M+1}.$$

These are known as prior mean and variance respectively. Consequently,

$$f(P_i | Y_{jk}, \hat{\mu}, \hat{M}) = \frac{1}{B(\alpha, \beta)} P_{jk}^{\alpha-1} (1-P_{jk})^{\beta-1} \quad (3)$$

Where;

$$\begin{aligned} \hat{\alpha} &= Y_{jk} + \hat{M}\hat{P}_o; \\ \hat{\beta} &= n_i - Y_{jk} + \hat{M}(1 - \hat{P}_o), \\ \hat{P}_o &= \frac{\sum Y_{jk}}{\sum n_{jk}} \end{aligned}$$

and

$$\hat{M} = \frac{\hat{P}_o(1-\hat{P}_o) - S_p^2}{\hat{P}_o(1-\hat{P}_o) \sum \frac{1}{n_{jk}} - S_p^2 - \frac{N}{N}}$$

where

$$S_p^2 = \frac{N \sum n_{jk} (\hat{P}_{jk} - \hat{P}_o)^2}{(N-1) \sum n_i}.$$

With M and P_o estimated, then;

$$\begin{aligned} \tilde{P}_{EB} &= E(P_{jk} | Y_{jk}, \hat{P}_o, \hat{M}) \\ &= \frac{\alpha}{\alpha + \beta} = \frac{Y_{jk} + \hat{M}\hat{P}_o}{n_{jk} + \hat{M}} \\ &= \left(\frac{\hat{M}}{n_{jk} + \hat{M}} \right) P_o + \left(\frac{n_{jk}}{n_{jk} + \hat{M}} \right) \frac{Y_{jk}}{n_{jk}} \\ Var(\tilde{P}_{EB}) &= \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}. \end{aligned} \quad (4)$$

Consequently, $\hat{\lambda} = \frac{\hat{M}}{n_{jk} + \hat{M}}$ and it can be readily seen

that where $\hat{M}$ (the scale factor) is large relative to n_{jk} ,

$\hat{\lambda}$ is large and $\hat{P}_o$ receives a larger weight than $\frac{Y_{jk}}{n_{jk}}$. But

large $\hat{M}$ implies small prior variance. Thus, the estimate which is associated with smaller variance receives larger weight in determining the posterior mean $\tilde{P}_{EB}$. On the other hand, if $\hat{M}$ is small relative to n_{jk} , the sample mean receives more weight. Note that the posterior density for the overall age group proportion P_o is obtained by replacing Y_{jk} and n_{jk} in equation (3) with Y and N , respectively.

Under conjugacy, the EB estimator of a proportion $\hat{P}_{ij}$ is a weighted mean of two estimators, the mean of the prior density P_o and the sample proportion estimator $\hat{P}_{ij}$. Thus,

$$\tilde{P}_{EB} = \lambda P_o + (1 - \lambda) \hat{P}_{ij} \quad (6)$$

$\tilde{P}_{EB}$ is the empirical Bayes Estimators with λ as the shrinkage factor. λ is a function of the prior and sample estimator variances such that, if variance of sample estimator is large, the weight of $\hat{P}_o$ (i.e. λ) will be large and $\tilde{P}_{EB}$ will shrink towards $\hat{P}_o$. Two components of the above model λ and $\hat{P}_o$ are derived from the EB process,[14].

4. Results

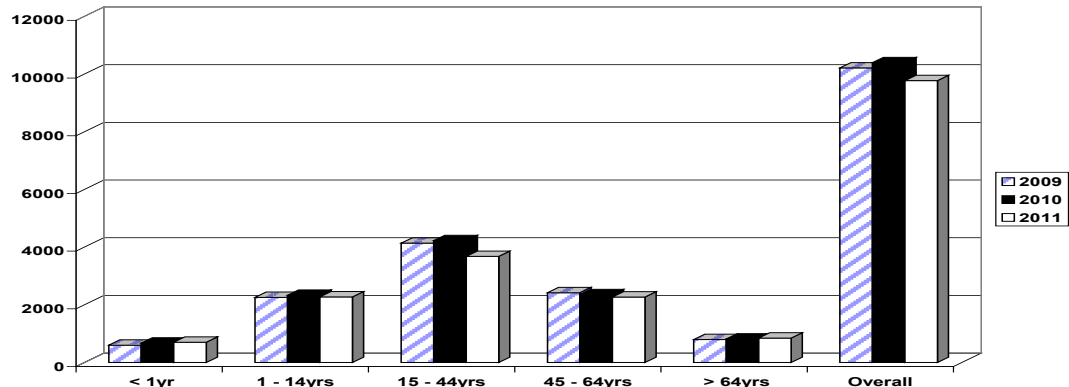
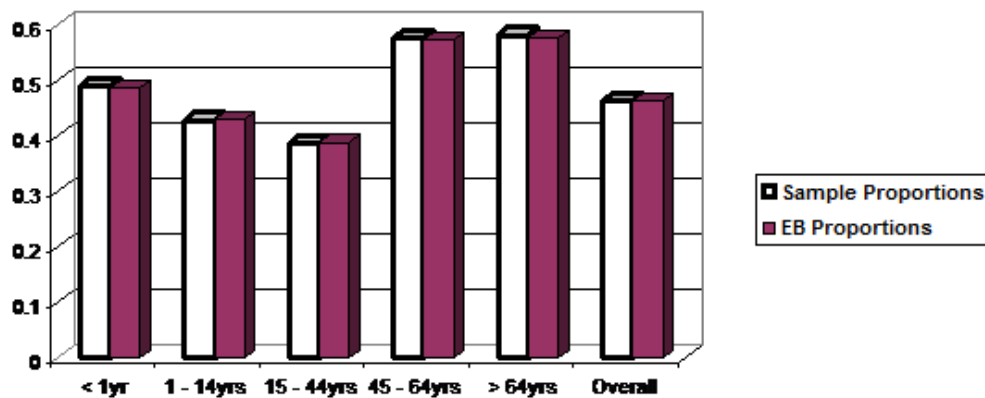
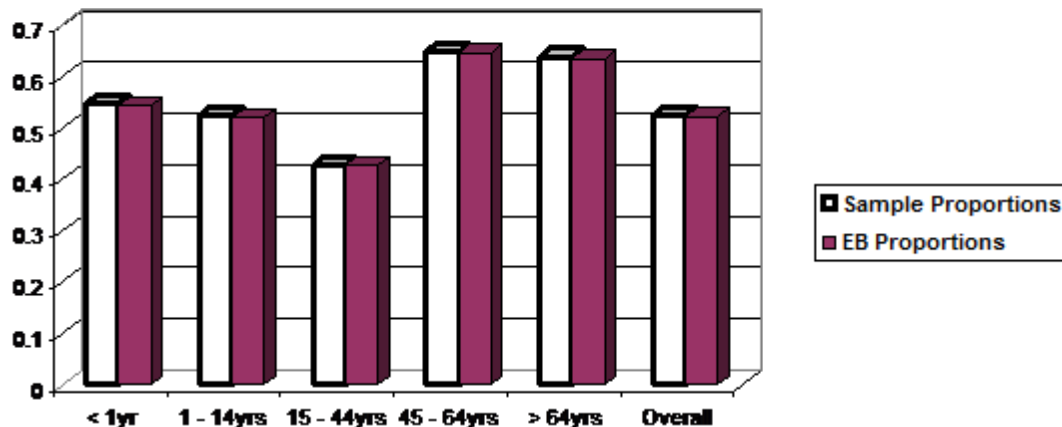
The results of the application of Beta-Binomial model and Bayesian sequential methods to the data of different age group patients for the three years (2009, 2010 and 2011) observations are presented in Table 1 & 2 below. The hyperparameters μ and M are estimated using sample information. These were subsequently used to determine the parameters of the posterior distributions α and β thereby completely specifying them. In our analyses, we compare on yearly basis estimated sample proportions and EB proportions as well as variances of estimated sample proportions and EB proportions.

Table 1. Comparative Analysis of Estimated Sample Proportions and EB Proportions

Year:	2009		2010		2011	
Age group	Pj1	PEB	Pj2	PEB	Pj3	PEB
< 1yr	0.4900332	0.4885043	0.547619	0.5430286	0.4608819	0.470019
1 - 14yrs	0.4288879	0.4294206	0.52383	0.5215909	0.5281195	0.527877
15 - 44yrs	0.3881102	0.3887539	0.4268812	0.4263775	0.4754009	0.474256
45 - 64yrs	0.5764463	0.574785	0.6481088	0.6463872	0.6910533	0.689386
> 64yrs	0.5823899	0.5772227	0.6392638	0.6352365	0.7375887	0.727948
Overall	0.4628868	0.4628868	0.5231361	0.52281	0.5593983	0.559072

Table 2. Comparative Analysis of Variances of Estimated Sample Proportions and EB Proportions

Year:	2009		2010		2011	
Age group	Var(Pj1)	Var(PEB)	Var(Pj2)	Age group	Var(Pj1)	Var(PEB)
< 1yr	0.00042	0.00039	0.00037	< 1yr	0.00035	0.00031
1 - 14yrs	0.00011	0.00011	0.00011	1 - 14yrs	0.00011	0.00011
15 - 44yrs	0.00006	0.00006	0.00006	15 - 44yrs	0.00007	0.00007
45 - 64yrs	0.00010	0.00010	0.00010	45 - 64yrs	0.00009	0.00009
> 64yrs	0.00031	0.00029	0.00028	> 64yrs	0.00023	0.00021
Overall	0.00002	0.00002	0.00002	Overall	0.00003	0.00002

Figure 1: Comparative number of Patients at the Hospital across the different age groups.**Figure 1.** Comparative number of Patients at the Hospital across the different age groups**Comparative Proportions Among Different Age Groups in 2009.****Figure 2.** Comparative Proportions Among different age groups in 2009**Comparative Proportions Among Different Age Groups in 2010.****Figure 3.** Comparative Proportions Among different age groups in 2010

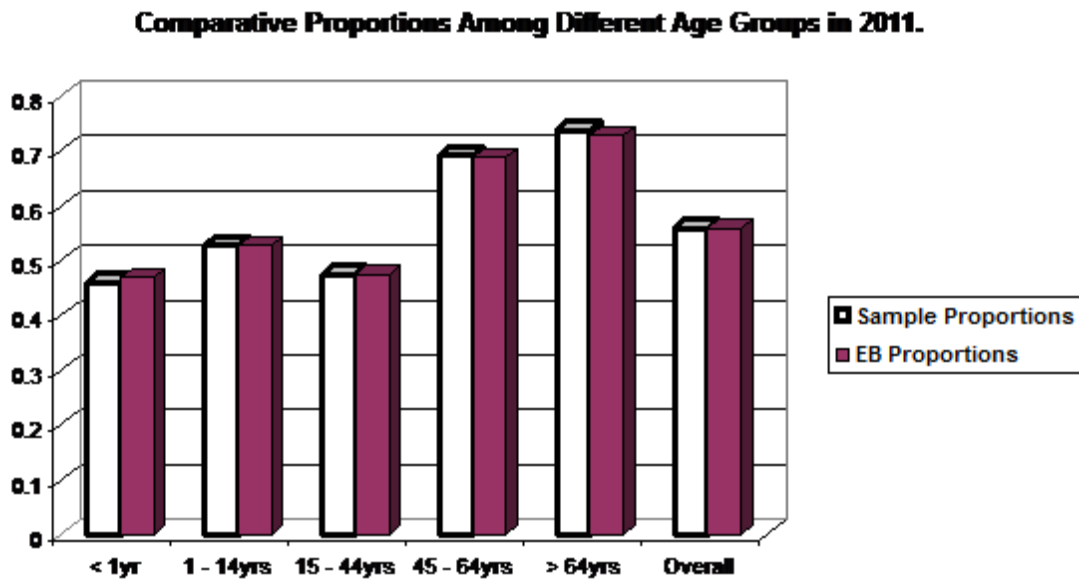


Figure 4. Comparative Proportions Among different age groups in 2011

Over the years under consideration, the results show that the highest number of patients attended to at the hospital is among the age group 15 to 44 years (see figure 1) but in figures 2, 3, 4 this age group also has the smallest proportion of orthopaedic surgeries. Similarly, the smallest number of patients can be found among the two age groups ‘less than one year’ and ‘greater than 64’. Also the highest proportion of orthopaedic surgeries is among patients whose ages are greater than 64 years followed by those who are less than one year old.

The overall EB proportion of patients admitted for orthopaedic surgeries in the hospital across the age groups increased steadily (0.4629, 0.5228 and 0.5591) over the respective three years of study.

Furthermore, the computed variances of the sample and EB estimates are smallest among the age groups ‘15 – 44 years’ and ‘44 – 64 years’ while highest variances are noted in age group ‘less than one’ and ‘greater than 64 years’ (see Table 2).

5. Conclusions

This paper has been able to show how Bayesian sequential estimation of proportion can be applied to statistical process control for different ages of the patients and years. The result of the analysis was compared both of yearly estimated sample proportions and EB proportions plus the variances. It was found that overall EB proportion of patients admitted for orthopaedic surgeries in the hospital across the age groups increased steadily. Similarly, the overall variances of the proportions tend more to zero over the three years under review. Thus, the results show that the EB estimators are better estimators on the basis of efficiency and consistency properties of good estimators.

REFERENCES

- [1] Carlin, B. P. and Louis, T. A. (2000a), “Empirical Bayes: Past, Present and Future”, *Journal of the American Statistical Association*, Vol. 95, No. 452, pp. 1286-1289.
- [2] Rubin, D. (1980), “Using empirical Bayes techniques in the law school validity studies”, *Journal of the American Statistical Association*, 75, pp. 801-816.
- [3] Link, W. A. and Hahn, D. C. (1996), “Empirical Bayes Estimation of Proportions with Application to Cowbird Parasitism Rates”, *Journal of Ecological Society of America*, 77(8), pp. 2528-2537.
- [4] Okafor, R. (1999), “Using an Empirical Bayes Model to estimate currency exchange rate”, *Journal of Applied Statistics*, Vol. 26, No. 8, 973-983.
- [5] Robbins, H. (1955), “An empirical Bayes approach to Statistics”, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1.
- [6] Deely, J. and Lindley, V. (1981), “Bayes empirical Bayes”, *Journal of the American Statistical Association*, 76, pp. 833-841.
- [7] Hui, S. and Berger, J. (1983), “Empirical Bayes estimation of rates in longitudinal studies”, *Journal of the American Statistical Association*, 78, pp. 753-760.
- [8] Morris, C. (1983), “Parametric empirical Bayes inference: Theory and applications”, *Journal of the American Statistical Association*, Vol. 78, pp. 47-65.
- [9] Wong, G. (1987), “Bayesian models for directed graphs”, *Journal of the American Statistical Association*, 82, pp. 140-148.
- [10] Smith, P. J. (1991), “Bayesian analyses for a multiple

- capture-recapture model”, *Biometrika* 78, pp. 399-407.
- [11] Raghunathan, T. E. (1993), “A Quasi-Empirical Bayes methods for Small Area Estimation”, *Journal of the American Statistical Association*, 88, pp. 1444-1448.
 - [12] Altman, N. S. and Casella, G. (1995), “Nonparametric Empirical Bayes Growth Curve Analysis”, *Journal of the American Statistical Association*, Vol. 90, pp. 508-515.
 - [13] Efron, B. (1996), “Empirical Bayes Methods for Combining Likelihoods (with discussion)”, *Journal of the American Statistical Association*, Vol. 91, pp. 538-565.
 - [14] Carlin, B. P. and Louis, T. A. (2000b), *Bayes and Empirical Bayes Methods for Data Analysis*, Boca Raton, Florida: Chapman and Hall/CRC Press.
 - [15] Natarajan, R. and Kass, R. E. (2000), “Reference Bayesian methods for Generalized Linear mixed models”, *Journal of the American Statistical Association*, 95, pp. 227-237.
 - [16] Brandel, J. (2004), *Empirical Bayes Methods for missing data analysis*, Department of Mathematics Uppsala University, Project Report.
 - [17] Theobald and Wuttke (2006), “Empirical Bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian Procrustes problem”, *The National Academy of Sciences of the USA*, vol. 103 no. 49, pp. 18521-18527.
 - [18] Ogundej, R. K. and Okafor, R. O. (2010), “Empirical Bayes Approach to Estimate Mean CGPA of University Graduates”, *International Journal of Applied Mathematics & Statistics* Vol. 17; No. J10; pp. 77-89.
 - [19] Okafor R.O., Opara A.I., Ogundej R.K., Mbata U.A., and Olalude G. A. (2010), “Empirical Bayes Model for Inference on Vehicular Traffic Density at the Main Campus of University of Lagos, Lagos, Nigeria”, *International Journal of Statistics and Systems*, Vol. 5 No. 1, pp. 53–62.
 - [20] Okafor R.O and Mbata U.A. (2011), “ A Bayesian Model for Estimation of Population Proportions”, *Section on Bayesian Statistical Science – Joint Statistical Meeting 2011, USA*.
 - [21] Hogg R.V., McKean J.W. and Craig A.T. (2001), *Introduction to Mathematical Statistics*, Pearson Prentice Hall, New Jersey 07458.
 - [22] Lee, P. (2004), *Bayesian Statistics An Introduction*, Hodder Arnold, New York.