

Bayesian Estimation of Above-Average Performance in Tertiary Institutions: A Case Study of University of Lagos, Akoka, Nigeria.

¹Rotimi Ogundeji, ²Ganiyu Lawal and ³Olatunji Arowolo.

¹Department of Mathematics, University of Lagos, Akoka, Nigeria, ²Department of Statistics, Yaba College of Technology, Lagos, Nigeria and ³Department of Mathematics, Lagos State Polytechnic, Ikorodu, Lagos, Nigeria.

Abstract: Bayesian approach for parameter estimation has the capacity to yield more precise estimates than methods based on sampling theory. There are several common Bayesian models; in this study we applied Empirical Bayes (EB) model called Beta-binomial model. The study is motivated by the need to beam searchlight on universities, faculties or fields of study with graduates who may not be eligible for further educational pursuits. This study provides means of assessment or a basis of evaluation of students' performances among faculties or fields of study and overall performance of a university. This study uses Bayesian methods of inference to estimate the proportion of above-average performance of graduates from the various faculties in University of Lagos. The model adopted generated results which are of smaller variances compared with variances of sample Proportions, showing that the posterior proportions generated are more efficient estimators. This is further evidenced in narrow widths of the computed confidence intervals. The overall result shows that the proportion of above-average performance of graduates of University of Lagos, who are eligible for further educational pursuits (i.e. higher degrees), is approximately 72% of the university graduates.

Keywords: Empirical Bayes Models, Beta-Binomial Model, Cumulative Grade Point Average, Posterior Proportion, Sample Proportion, University of Lagos-Nigeria.

I. Introduction

A university graduate can be said to have performed above average if his/her performance at the undergraduate level is acceptable for further university education i.e. admissible into master and doctoral degree programmes. For example, admission into any postgraduate programme at University of Lagos will require a minimum of second class lower i.e. a minimum CGPA of 2.40. In education, a grade (or mark) is a teacher's standardized evaluation of a student's work. A cumulative grade point average (CGPA) is the mean of grade point average (GPA) from all semesters whereas GPA may only refer to a single semester.

There is need to beam searchlight on universities, faculties or fields of study with graduates who may not be eligible for further educational pursuits (i.e. fields of study that may result in brain drain within a university system). This study provides means of assessment or a basis of evaluation of students' performances among faculties or fields of study and overall performance of a university.

This study uses Bayesian methods of inference to estimate the proportion of above-average performance of graduates of University of Lagos (UNILAG). It derives motivation from Rubin's (1980) illustrations of the Bayesian method as an extremely powerful tool for the applied statistician especially in the way it could provide sensible answers in a straight forward manner in problems where sampling theories approaches appear awkward. Further support is derived from Link and Hahn (1996) who argued that empirical Bayes methods provide alternative approaches that incorporate the structural advantages of Bayesian models while requiring less stringent specification of prior knowledge.

Empirical Bayes methods have been applied in a number of contexts. Some examples include applications to problems in forest science (Burk et al., 1982); to monitoring of air pollution (Suggs et al., 1983), and in a variety of medical applications (Stijnen et al., 1990). More applications include using an empirical Bayes model to estimate currency exchange, (Okafor, 1999), analysis of missing data using empirical Bayes methods (John Brandel, 2004); estimating population sizes from survey data (Johnson, 1989) and identification of extremes in collections of parameter estimates, Link et al (1996). More recent works and applications include (Natarajan et al., 2000), (Brandel, 2004), (Theobald et al., 2006), (Kazembe et al., 2008), (Ogundeji et al., 2010), (Okafor et al., 2010), and (Okafor et al., 2011).

Link & Hahn (1996) advanced that a Bayesian analysis can be thought of as a combination of existing knowledge with new knowledge (sample data), the two being synthesized in such a way as to account for the amount of confidence that is placed in each source of knowledge. Usually, the existing knowledge base and updated knowledge base are summarized by probability distributions describing the likely range of values for

each unknown parameter which are referred to as the prior and posterior distributions, respectively (Carlin et al., 2000).

University Of Lagos, Nigeria (Unilag).

The University of Lagos founded in 1962 is made up of two campuses, the main campus at Akoka, Yaba and the College of Medicine in Idi-Araba, Surulere. From a modest intake of 131 students in 1962, enrolment in the university has now grown to over 39,000. It has total staff strength of 3,365 made up of 1,386 Administrative and Technical Staff, 1,164 Junior and 813 Academic Staff.

The University is comprised of nine Faculties and a College of Medicine. The Faculties offer a total of 117 programmes in Arts, Social Sciences, Environmental Sciences, Pharmacy, Law, Engineering, Sciences, Business Administration and Education. It also offers Master’s and Doctorate degrees in most of the aforementioned programmes. Recently, University of Lagos emerged the “first and the best” in the global Webometric ranking in Nigeria.

II. Empirical Bayes (Eb) Model: Beta-Binomial Model

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

parameter. The posterior mean is
$$\tilde{P}_i = \int P_i f(P_i | X_i, \eta) dP_i \tag{1}$$

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

$$f(P_i | X_i, \eta) = \frac{\binom{n_i}{X_i} P_i^{X_i} (1 - P_i)^{n_i - X_i} \frac{1}{B(r, s)} P_i^{r-1} (1 - P_i)^{s-1}}{\int P_i^{X_i} (1 - P_i)^{n_i - X_i} \frac{1}{B(r, s)} P_i^{r-1} (1 - P_i)^{s-1} dP_i}, \quad \eta = (r, s) \tag{2}$$

There is need to the estimate the hyperparameters r and s of the beta distribution in order to completely specify the prior. This can be achieved easily through by the reparameterisation scheme of $f(P_i | \eta)$, and using

moment estimation (Brandel, 2004). Letting $P_o = \frac{r}{r + s}$; $M = r + s$ and using the prior distribution of P_i ;

$E(P_i) = P_o$ while $Var(P_i) = \frac{rs}{(r + s + 1)(r + s)^2} = \frac{P_o(1 - P_o)}{M + 1}$. These are known as prior mean and

variance respectively. Consequently,

$$f(P_{ij} | X_i, \hat{\mu}, \hat{M}) = \frac{1}{B(\alpha, \beta)} P_i^{\alpha-1} (1 - P_i)^{\beta-1} \tag{3}$$

Where; $\hat{\alpha} = X_i + \hat{M}\hat{P}_o$; $\hat{\beta} = n_i - X_i + \hat{M}(1 - \hat{P}_o)$, $\hat{P}_o = \frac{\sum X_i}{\sum n_i}$ and

$$\hat{M} = \frac{\hat{P}_o(1 - \hat{P}_o) - S_p^2}{S_p^2 - \frac{\hat{P}_o(1 - \hat{P}_o) \sum \frac{1}{n_i}}{N}} \quad \text{where } S_p^2 = \frac{N \sum n_{ij} (\hat{P}_{ij} - \hat{P}_o)^2}{(N - 1) \sum n_i}.$$

$\hat{\alpha}$ and $\hat{\beta}$ are the sample estimates of r and s respectively.

With M and P_o estimated, then;

$$\tilde{P}_{EB} = E(P_i / X_i, \hat{P}_o, \hat{M}) = \frac{\alpha}{\alpha + \beta} = \frac{X_i + \hat{M}\hat{P}_o}{n_i + \hat{M}} = \left(\frac{\hat{M}}{n_i + \hat{M}} \right) P_o + \left(\frac{n_i}{n_i + \hat{M}} \right) \frac{X_i}{n_i} \tag{4}$$

$$\text{Var}(P_{EB}) = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2} \tag{5}$$

Consequently, $\hat{\lambda} = \frac{\hat{M}}{n_i + \hat{M}}$ and it can be readily seen that where \hat{M} (the scale factor) is large relative

to n_i , λ is large and \hat{P}_o receives a larger weight than $\frac{X_i}{n_i}$. 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_{ij} , the sample mean receives more weight. Note that

the posterior density for the overall University proportion P_o is obtained by replacing X_i and n_i in equation (4) with X and N, respectively. Under conjugacy, the EB estimator of a faculty proportion \tilde{P}_{EB} is a weighted mean of two estimators, the mean of the prior density P_o and the sample proportion estimator \hat{P}_i . Thus,

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

\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.

III. Data And Method

The data used for this study is the graduating Cumulative Grade Point Average (CGPA) of randomly selected graduates of the University of Lagos. The EB model is applied to the graduating CGPA of the 2007/2008 graduates of the institution. These data are as obtained from the records of results approved by the senate of University of Lagos as final list of graduates for the 2007/2008 convocation. The data covers 48 departments in 8 faculties of the University of Lagos.

The CGPA for a student is obtained by multiplying the grade point obtained by the student in each course by the number of units assigned to the course to arrive at the weighted scores thus, adding together the weighted for all the courses, taken up to the final year of study and dividing the weighted scores by the total unit taken. The graduating class of students in the University of Lagos is based on the CGPA obtained by each student, see first and second columns of table 1. The last two columns of table 1 show conversion of each graduating class into binary code with the corresponding description.

Table 1: Graduating Class with CGPA Range and Conversion into Binary Codes

Class Category	CGPA Range	Binary Code	Description
First Class	4.5 – 5.0	1	Above-average Performance
Second Class Upper	3.5 – 4.49	1	Above-average Performance
Second Class Lower	2.4 – 3.49	1	Above-average Performance
Third Class	1.5 – 2.39	0	Below-average Performance
Pass	1.0 – 1.49	0	Below-average Performance

By above-average performance, we mean graduates who may be eligible for further educational pursuits (i.e. admissible into master and doctoral degree programmes.), otherwise below-average performance. For the purpose of this analysis, the graduating CGPA are recoded as follows:

x_{ji} represents a random outcome of CGPA of a graduate j ($j = 1, 2, \dots, n$) in faculty i , ($i = 1, 2, \dots, 8$).

$$x_{ij} = \begin{cases} 1 & \text{Graduate } j \text{ in faculty } i \text{ with CGPA equal \& above } 2.40 \\ 0 & \text{Otherwise} \end{cases}$$

n_i = number of graduates examined in faculty i .

$$P_i = \frac{\sum_{j=1}^n x_{ij}}{n_i}$$

is the sample proportion of graduates who are classified to have “Above-average

Performance” in faculty i. $x_i \stackrel{iid}{\sim} Bi(n_i, P_i)$

IV. Results

Using sample information, sample proportions P_i for each faculty is computed and the hyperparameters $\hat{\mu} = 0.717118$ and $\hat{M} = 35.1727536$ are also obtained. These were subsequently used to estimate the expected posterior proportions and corresponding posterior variances. The results are shown in Table 2, Table 3 below.

Table 2: Estimated Sample Proportions and Expected Posterior Proportions of Above -Average Graduates.

S/n	Faculty	n _i	x _i	Sample Proportions P _i	Shrinkage Factors		Posterior Proportions P _{EB}	Ranking
					M/(n+M)	n/(n+M)		
1	Law	228	128	0.561404	0.133649	0.866351	0.582215	8
2	Arts	571	419	0.733800	0.058024	0.941976	0.732832	4
3	Education	854	591	0.692037	0.039557	0.960443	0.693030	5
4	Business Administration	564	371	0.657801	0.058702	0.941298	0.661283	6
5	Environmental Sc.	381	317	0.832021	0.084515	0.915485	0.822310	1
6	Engineering	552	449	0.813406	0.059902	0.940098	0.807638	2
7	Sciences	592	388	0.655405	0.056081	0.943919	0.658866	7
8	Social Sc.	242	194	0.801653	0.126898	0.873102	0.790926	3
	UNILAG	3984	2857	0.717118	0.008751	0.991249	0.717118	

Table 3: Variances of Estimated Sample and Posterior Proportions with Respective Differences.

S/n	Faculty	Sample Proportions P _i	Posterior P _{EB}	Sample Variance (σ ²)	P _{EB} Variance(σ ²)	Variance Differences
1	Law	0.561404	0.582215	0.001080	0.000921	0.000159
2	Arts	0.733800	0.732832	0.000342	0.000322	0.000020
3	Education	0.692037	0.693030	0.000250	0.000239	0.000011
4	Business Admin.	0.657801	0.661283	0.000399	0.000373	0.000026
5	Environmental Sc.	0.832021	0.822310	0.000367	0.000350	0.000017
6	Engineering	0.813406	0.807638	0.000275	0.000264	0.000011
7	Sciences	0.655405	0.658866	0.000382	0.000358	0.000024
8	Social Sc.	0.801653	0.790926	0.000657	0.000594	0.000063
	UNILAG	0.717118	0.717118	0.000051	0.000050	0.000001

In Figure 1, each bar of the Bar chart represents sample proportion of graduates with above-average performance in each faculty (i.e. graduates whose CGPA are equal and above 2.4). For instance, 81.34% represents the sample proportion of above average graduates in the faculty of engineering and 56.14% for faculty of Law while 71.71% for the overall University of Lagos.

Figure 1
Chart of Computed Sample Proportions for each Faculty in University of Lagos

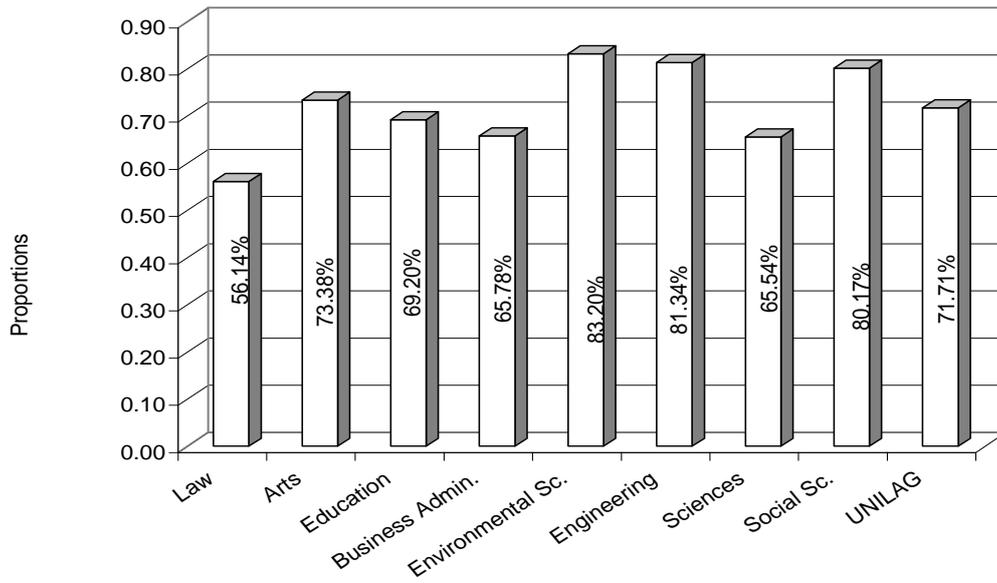


Figure 2
Chart of Estimated Posterior (PEB) of each Faculty in University of Lagos.

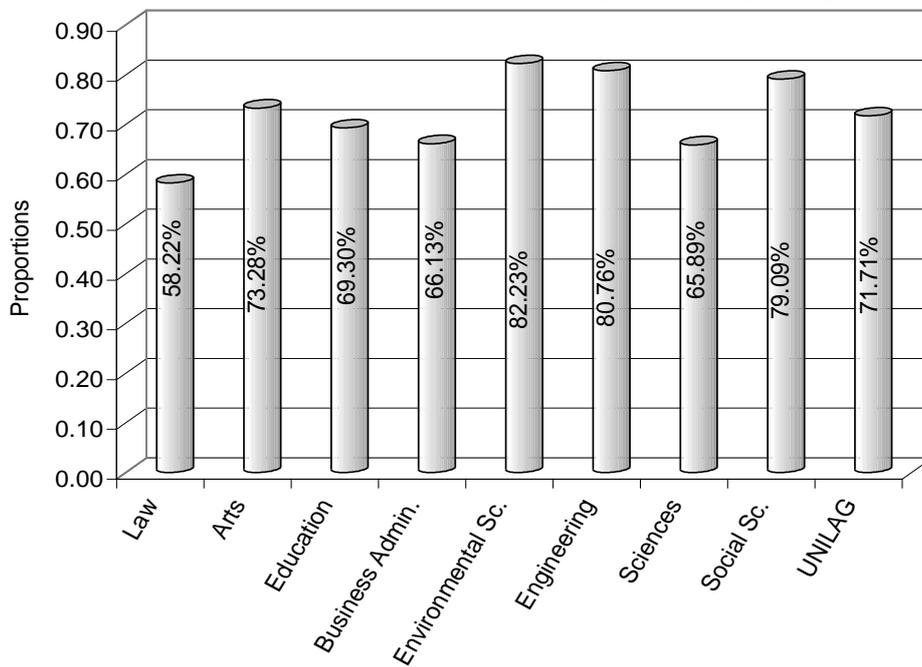


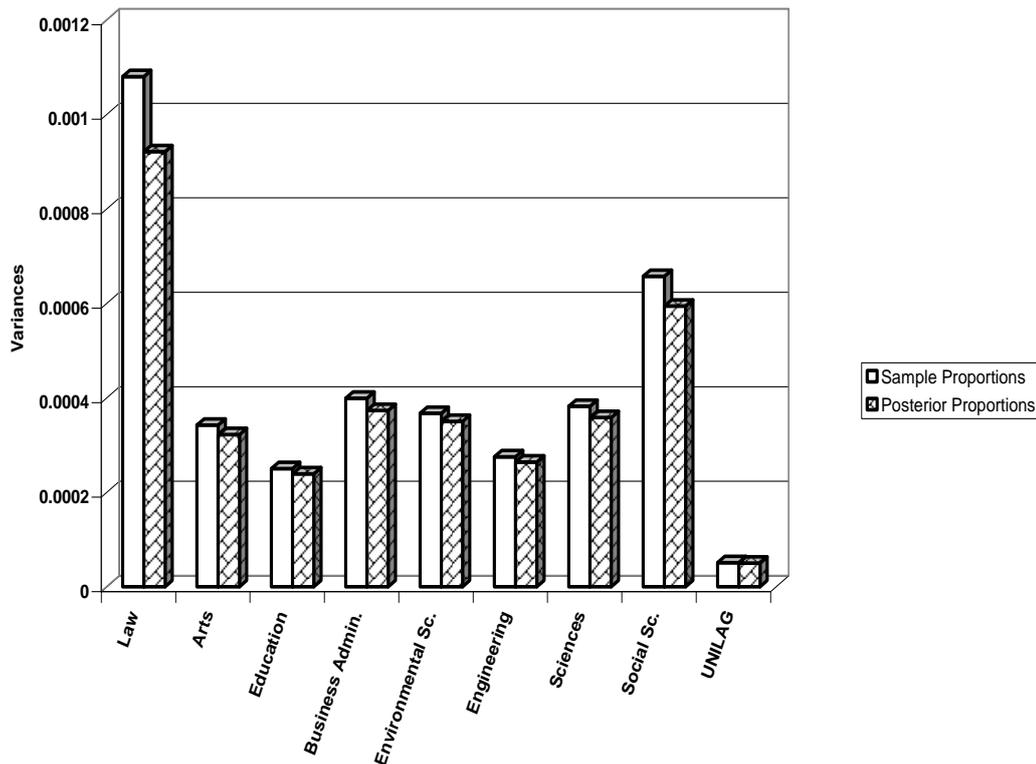
Table 4: 95% Confidence Intervals (CI) for Posterior Proportions (P_{EB})

S/n	Faculty	n_i	x_i	P_{EB}	P_{EB} $Var(\sigma^2)$	LCI	UCI	Width of CI
1	Law	228	128	0.582215	0.000921	0.501929	0.620878	0.12
2	Arts	571	419	0.732832	0.000322	0.698604	0.768996	0.07
3	Education	854	591	0.693030	0.000239	0.661737	0.722338	0.06
4	Business Admin.	564	371	0.661283	0.000373	0.619937	0.695666	0.08
5	Environmental Sc.	381	317	0.822310	0.000350	0.795339	0.868703	0.07
6	Engineering	552	449	0.807638	0.000264	0.781551	0.845260	0.06
7	Sciences	592	388	0.658866	0.000358	0.618331	0.692480	0.07
8	Social Sc.	242	194	0.790926	0.000594	0.753865	0.849441	0.10
	UNILAG	3984	2857	0.717118	0.000050	0.703196	0.731041	0.03

Table 5: 95% Confidence Intervals (CI) for Sample Proportions (P_i)

S/n	Faculty	n_i	x_i	Sample Proportion (P_i)	Sample Variance (σ^2)	LCI	UCI	Width of CI
1	Law	228	128	0.561404	0.001080	0.496993	0.625814	0.13
2	Arts	571	419	0.733800	0.000342	0.697548	0.770052	0.07
3	Education	854	591	0.692037	0.000250	0.661075	0.723000	0.06
4	Business Admin.	564	371	0.657801	0.000399	0.618645	0.696958	0.08
5	Environmental Sc.	381	317	0.832021	0.000367	0.794482	0.869560	0.08
6	Engineering	552	449	0.813406	0.000275	0.780905	0.845906	0.07
7	Sciences	592	388	0.655405	0.000382	0.617123	0.693688	0.08
8	Social Sc.	242	194	0.801653	0.000657	0.751412	0.851893	0.10
	UNILAG	3984	2857	0.717118	0.000051	0.703132	0.731104	0.03

Figure 3: Comparison of the Variances of Estimated Sample and Posterior Proportions.



V. Comments And Conclusion

This analysis has generated more reliable proportions (P_{EB}) for the various faculties in University of Lagos which are between 0.58 and 0.82. This suggests that across the various faculties of University of Lagos, 58% to 82% of the graduates performed above average and are eligible for further educational pursuits. We ranked in descending order the estimated proportions of the various faculties and noted that faculty of Environmental Sciences had the highest proportion of above-average performance graduands while faculty of Law had the lowest. The study also shows that from a total of 3,984 graduates of University of Lagos in 2007/2008 session, an overall estimated proportion of 0.72, approximately 2,868 graduates of University of Lagos can be said to have performed above average and are eligible for further educational pursuits. The performance of the graduates of faculty of Law, Science and Business Administration (i.e. ranked 8, 7, and 6 respectively) all fall below the overall proportion for graduates of University of Lagos.

We recommend further studies into performances of these faculties using current and more enlarged data set, with a view to improving the evaluation and assessment of performances in tertiary institutions. Also, there is need to encourage, by way of scholarship, graduates who performed above-average towards further educational pursuit, in order to reduce brain drain in the university system.

The differences in variances between the posterior and sample proportions for each faculty are evidence to the fact that the posterior proportions with smaller variances are more efficient (i.e. more reliable) estimators compared to corresponding sample proportions. Thus, the Bayesian approach using the empirical model is justified.

References

- [1]. Burk, T. E., and Ek A. R. (1982): Application of empirical Bayes James Stein procedures to simultaneous estimation problems in forest inventory. *Forest Science* 28:753-771.
- [2]. Carlin B. P. and Louis T. A. (2000): Empirical Bayes: Past, Present and Future. *Journal of the American Statistical Association*, 95(452): 1286-1289.
- [3]. John Brandel, (2004): Empirical Bayes methods for missing data analysis. U.U.D.M. Project Report 2004: 11.
- [4]. Johnson. D.H. (1989): An empirical Bayes approach to analyzing recurring animal surveys. *Ecology* 70:945-952.
- [5]. Kazembe L.N, Chirwa T.F., Simbeye J.S., Namangale J.J. (2008): Applications of Bayesian Approach in modeling risk of Malaria-related Hospital Mortality. *B.M.C. Medical Research Methodology* (2008). 8: 6. <http://www.biomed-central.com/1471-22888/8/6>.
- [6]. Laird, N. M., and Louis T. A. (1987): Empirical Bayes confidence intervals based on bootstrap samples. *Journal of the American Statistical Association* 82:739-757.
- [7]. Link, W. A., and Sauer J. R. (1996). Extremes in ecology: on avoiding the misleading effects of sampling variation in summary analyses. *Ecology* 77:1633-1640.
- [8]. Link W.A. and Hahn .C. D. (1996): Empirical Bayes Estimation of Proportions with Application to Cowbird Parasitism Rates. *Ecology*, 77(8), 1996, pp. 2528-2537
- [9]. Ogundeji, R. K. & Okafor, R. O. (2010): Empirical Bayes Approach to Estimate Mean CGPA of University Graduates, *International Journal of Applied Mathematics & Statistics*, 17(J10): 77-89.
- [10]. Okafor R.O., Opara A.I., Ogundeji R.K., Mbata U.A., & Olalude G. A. (2010): Empirical Bayes Model for Inference on VehicularTraffic Density at the Main Campus of University of Lagos, Lagos, Nigeria, *International Journal of Statistics and Systems*, 5 (1): 53-62.
- [11]. Okafor R.O. & Mbata U.A. (2011): A Bayesian Model for Estimation of Population Proportions. Section on Bayesian Statistical Science: Joint Statistical Meeting 2011, USA.
- [12]. Okafor R. O.(1999): Using an empirical Bayes model to estimate currency exchange rates. *Journal of Applied Statistics*, 26(8): 978 – 983.
- [13]. Rubin, D. (1980): Using empirical Bayes techniques in the law school validity studies. *Journal of the American Statistical Association*, 75: 801-816.
- [14]. Stijnen, T., and Van Houwelingen J. C. (1990): Empirical Bayes methods in clinical trials meta-analysis. *Biometrical Journal* 32:335-346.
- [15]. Suggs, J. C., and Curran.T. C.(1983): An empirical Bayes method for comparing air pollution data to air quality standards. *Atmospheric Environment* 17:837-842.