Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.4, No.12, 2014

Bayesian Hierarchical Spatial Modeling and Mapping of Adult Illiteracy in Kenya

Purity Mukami Kinyamu[1*], Oscar Ngesa[2], George Otieno Orwa[1] and Henry Murray Humphreys[1]

1. Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, P.O Box 62000, Nairobi.
2. Ministry of Devolution and Planning, Kenya, P.O Box 30005-00100 Treasury Building, Nairobi.
3. School of Mathematics, Statistics and Computer Science, University Of Kwazulu Natal, Private Bag X01, Pietermaritzburg, South Africa

*Email of the Corresponding author: pmukami09@gmail.com

Abstract

Regional disparity in literacy levels must be addressed if Kenya is to achieve its international goals such as Education for All (EFA) and Millennium Development Goals (MDG). Literacy level in Kenya has been on the rise. However, the 2007 Kenya National Literacy Survey crude rates showed that on average 38.5 per cent (7.8 million) of Kenya's adult population was illiterate with significant regional and gender variation.

Bayesian binary logistic models (with and without CAR spatial and unstructured random effects) are applied to the Kenya National Adult Literacy Survey (2007) data that was obtained from sampled 18000 households, 4782 in urban and 10914 in rural areas, to investigate spatial variation of illiteracy levels in Kenya. There were 15734 successful interviews that were comprised of 6493 were male and 9241 female

The best fitted model was found to be the CAR model with age, sex, disability and awareness of adult literacy programs as the significant explanatory variables. Smoothed map of illiteracy from the best fitted model was then produced together with its corresponding confidence interval maps for regional variation in Kenya, in order to capture visual uncertainty in estimation. These maps can be used by policy makers to identify the pattern and tailor make programs appropriate for each region.

Keywords: Illiteracy, Bayesian Hierarchical Models, Spatial modeling

1. Introduction

Literacy level in Kenya has been on the rise. Many programs and campaigns have been initiated with the aim of improving the literacy levels. The introduction of the free primary education program in 2003 has also seen the rise of literacy. This program enabled schools to buy teaching materials and books. A 20 percent increase in primary school enrollment was achieved, from 6.0 million in 2002 to 7.2 million pupils in 2003. It further increased to 8.6 million in 2008. Over the same period, Gross Enrollment Rate (GER) increased from 88.2 percent to 107.6 percent in 2008. In 2008, the government of Kenya introduced Free Secondary Education which showed improvement in enrollment rates, UNESCO (2010).

However, KNALS (2007) found out that on average 38.5 percent (7.8 million) of the Kenyan adult population was illiterate. The survey also showed women were poorer in reading and numeracy than men. Regional variations were also noted, Nairobi, the capital city, had an adult literacy rate of 87.1 percent while North Eastern province had an adult literacy of 9.1 percent. The urban areas recorded higher rates than rural areas.

The main feature of literacy in Kenya which still remains consistent over years is the clear disparity in literacy levels among regions, gender and other categories. Regional disparity in literacy levels must be addressed if Kenya is to meet its international goals such as Education for All (EFA), Millennium Development Goals (MDG) and the UN Literacy Decade (2003-2012). If the regions with high illiteracy are not sensitized and given extra attention then the literacy level of the whole country will be dragged down; this will in turn affect negatively the social economics of the whole nation, Carron et al. (1989). A nation with a tainted social economic development record is also prone to political instability, Iftikhar (2011).

Understanding these regional variations is one step in addressing them, Ahmed (2009). Several studies have been done on illiteracy levels in Kenya. All of them have been based on crude rates of measures which are highly affected by the sample size collected per region. The country has now been divided into counties and accurate county statistics are needed for regional planning and fund allocation in Kenyan government programs and thus their importance cannot be overemphasized.

In this study, we statistically modeled illiteracy levels in Kenya using a Bayesian technique which ensures smoothing of estimates and hence overcoming the sample size problem associated with crude rates. Spatial variation was also captured using a convolution model.

Smoothed maps of county illiteracy levels were produced and their corresponding 95% credible interval maps produced to capture visual uncertainty in estimation. The maps can give clear guidance to policy makers on which regions to target the limited resources available and also tailor make the literacy alleviation programs and campaigns to suit the cultural and social-economic practices of the people in that county.

2. Materials and Methods

2.1 Approaches used to estimate illiteracy

In Cotterell (1975), it is stated that, increasingly the world is becoming aware of adult illiteracy. Its extent may still surprise many people—in 1970 there were 783 million illiterate adults throughout the world, at least one million of whom were living in the United Kingdom—but what should really give us reason to pause are the social and economic implications of the problem.

Most studies on adult illiteracy are not model based and so cannot be generalized. According to Kirsch et al. (2002), adult literacy rate in America was calculated by weighting where full sample and replicate weights were calculated for each record in order to facilitate the calculation of unbiased estimates and their standard errors. The statistical comparisons in this report were based on the t statistic. NAAL (2003), measures the English literacy of America's adults (people age 16 and older living in households or prisons). The findings are presented in percentages and tables.

Mitra and Singh (2008), in their article to highlight the differences in literacy and schooling attainment among the scheduled tribe women in India used the proportions for each region to report their findings and draw conclusions. These results cannot be generalized to other regions.

Mazumdar (2005), in his article developed from considering data from 56 countries around the world whose human development index was below 0.800 in 1999, to determine social-economic factors determining adult literacy in developing countries uses Ordinary Least Squares (OLS) multiple regression equation to estimate the rate of adult literacy of 1999. Backward elimination procedure was tried in this study since all of the eight predictor variables did not have significant coefficients for any of the samples. Accordingly, for different samples they ended up with different sets of variables having all coefficients significant. In all the three samples, regression equations were used as well correlation among variables to explain the relationship. The highest R^2 is 0.8212 and some are as low as 0.29. Therefore there was unexplained random effect of the determination of adult literacy in developing countries.

In Kenya, a study of national adult literacy was conducted in 2006/2007 and the results were based on proportions. The survey's sample sizes in the small areas are not large enough to support direct estimates of adequate precision.

2.2 Spatial approach in estimation of illiteracy

Often data from diverse areas such as climatology, ecology, environmental monitoring and health sciences are spatially correlated. Hence the dependence structure underlying the data is a function of the location information.

Typically, observations located closer will be more similar than those further apart; thus, data exhibit spatial variability which when ignored when working with regression models, the estimates of variation will be biased resulting in inefficient statistical inference, Cressie (1993). Hence, to accurately assess the association between response and covariates it is important to allow for spatial dependence when developing regression models for data that may have spatial correlation.

Spatial data are typically classified into one of three types: point-referenced data, point pattern data or areal data.

Point- referenced data is often referred to as geocoded or geostatistical data. The response $Y(s)$ is a random variable observed at location $s \in R^r$ where s varies continuously over some fixed region $D \subseteq R^r$ and the index set D contains an r-dimensional rectangle of positive volume, Li (2008). The response $Y(s^*)$ at some unobserved location s^* is predicted based on the observed value at a fixed set of locations $Y(s_1), Y(s_2), Y(s_3),\cdots Y(s_n)$.

The dependence between $Y(s_i)$ and $Y(s_{i'})$ is modeled through a covariance model

$$\boldsymbol{COV}[\boldsymbol{Y}(\boldsymbol{s_i}),\boldsymbol{Y}(\boldsymbol{s_{i'}})] = \boldsymbol{c}(\boldsymbol{s_i},\boldsymbol{s_{i'}}) = \boldsymbol{c}(\boldsymbol{d_{i''}}) \qquad (1)$$

Where $d_{i''}$ is the distance between s_i and $s_{i'}$.

In point pattern data, it is the index set that is random and gives the locations of random events. When an event is equally likely to occur at any point of observation area regardless of the location of other events, it is termed as complete spatial randomness, characterized by the stochastic process, the homogeneous Poisson process. The two alternatives of complete spatial randomness are spatial clustering which implies that the events tend to be spatially close to the other points, and spatial regularity which implies that events points space themselves out as much as possible, Li (2008).

For areal data, responses are observed on a regular or irregular lattice consisting of geographical areas with well-defined boundaries. The spatial structure underlying the observations is often summarized through an adjacency matrix W, whose entries code, in some sense, the connectivity (neighborhood structure) of the underlying map. The typical definition for the adjacency matrix, spatially connecting units i and j is

$$\boldsymbol{w} = \begin{cases} \mathbf{0} & \boldsymbol{if\ i = j} \\ \mathbf{0} & \boldsymbol{if\ i,\ j\ are\ not\ neighbours} \\ \boldsymbol{c_{ij}} & \boldsymbol{if\ i,\ j\ are\ neighbours} \end{cases} \quad (2)$$

Hence $\boldsymbol{c_{ij}} > 0$ is the strength of the neighbor relationship areal units i and j. The most commonly used connectivity weight is

$$c_{ij} = \begin{cases} 1 & if\ i, j\ are\ neighbours \\ 0 & if\ i,\ j\ are\ not\ neighbours \end{cases} \quad (3)$$

To model spatial dependence in the data $Y_1,\ Y_2,\ \cdots Y_n$, the joint distribution $f(y_1,\ y_2,\ \cdots y_n)$ is constructed through the specification of a set of simple full conditional distributions $f(y_i \backslash y_j, i \neq j), i = 1,2,3, \cdots n$. In this context, it is clear that given a joint distribution, the full conditional distributions are always uniquely determined; however the converse is not always true. In general, we say that the set of full conditional distributions is compatible if they determine a unique and valid joint distribution, Li (2008).

The conditional distributions will depend on neighborhood structure underlying the map, W, and spatial dependence is thus built into the joint model specification, Li (2008). The full conditional distribution of Y_i can be written as:

$$\boldsymbol{f(y_i \backslash y_j, i \neq j) = f(y_i \backslash y_j, j \in \partial_i), i = 1, 2, 3, \cdots n} \quad (4)$$

Where $\partial_i = \{j, W_{ij} \neq 0\}$ denote the neighbors for region. We use a set of simple local specifications that depend only on lattice adjacencies to develop a spatial dependence structure. This sort of specification is referred to as a Markov random field (MRF) Besag (1974).

A conditionally autoregressive model for modeling areal data $y = [y_2,\ \cdots y_n]^T$, Besag (1974) is, in a sense, the simplest non-trivial special case of Markov random field(MRF) that can be used to model spatial data.

This model is specified through the set of full conditional distributions

$$[y_i | \ y_j, i \neq j] \sim N\left(\sum_j \frac{W_{ij}}{W_{i+}}, \frac{\sigma^2}{W_{i+}}\right), \quad i = 1,2,3, \cdots n \quad (5)$$

Where, $W_{i+} = \sum_{j=1}^{n} W_{ij}$ denotes the sum of the i^{th} row and σ^2 is the variance component, Li (2008).

In our case we shall use binomial distribution,

$$y_{ij} | \ p_{ij} \sim Bin\ (p_{ij})$$

$$\log(p_{ij}) = X_{ij}^T \beta + b_i \quad (6)$$

$$b = (b_1 \cdots b_n)^T \sim CAR\ (\sigma^2)$$

This kind of a model is referred to as generalized linear mixed spatial model. Fitting such a model using standard maximum likelihood techniques and using the likelihood function is difficult and therefore we shall use a Bayesian approach since it does not rely on asymptotic but the inference is based on computing the posterior distributions of the unknowns given the data.

Since the spatial random effects are additive, we shall use the Conditional Auto-Regressive (CAR) models to account for spatial heterogeneity in the data, Osei (2010).

2.3 Bayesian Analysis

Statistical inference concerns the learning of some unknown aspect of the population from which the data was drawn, Li (2008). Bayesian inference fits a probability model to observed data and summarizes the results through a probability distribution on the unknown parameters and / or unobserved data we are interested in.

The Bayesian method offers potentially attractive advantages over the frequency statistical approach for modeling spatial data as described in Banerjee et al. (2004). First, the Bayesian approach allows one to induce specific spatial correlation among random effects by use of prior distributions. Second, computational challenges associated with computing the posterior distributions can be overcome by use of Markov chain Monte Carlo methods. Third, the hierarchical Bayesian models provide mechanisms to specify a complicated non-Gaussian data through several layers, each of which can be easily understood and computed. Finally the Bayesian method explicitly acknowledges the uncertainty of the model and parameters.

The parameters in the model will be estimated from the posterior distribution which will be obtained by updating the prior distribution with the observed data through the likelihood function of the observed data.

The likelihood for the data $\{y_i\}$, $i = 1,2,3, \cdots n$ is defined as

$$L(x|\theta) = \pi_{i=1}^{n} f(x_i|\theta) \tag{7}$$

The data are assumed to be conditionally independent.

A prior distribution is a distribution assigned to the parameter before seeing the data. The sample provides additional ' data ' for a problem and so they can be used to improve estimation or identification of parameters, Lawson (2009). These distributions of parameters also have parameters controlling their forms since parameters in models are regarded as stochastic (and thereby have probability distributions governing their behavior), then these parameters also have distributions. These distributions are known as hyper-priors and their parameters are known as hyper parameters.

The posterior distribution is the product of the likelihood and the prior distributions. It describes the behavior of the parameters after the data are observed and prior assumptions are made, Lawson (2009). The posterior distribution is the conditional distribution of the parameters after observing the data and it integrates the prior and the sample information. For a probability distribution whose population parameter is a random variable, then the joint probability function can be written as $f(\theta, X_1, \cdots X_n) = f(X_1, \cdots X_n|\theta)\pi(\theta)$. The marginal probability density of X will be given by $f(X_1, \cdots X_n)$. Then the conditional distribution becomes;

$$f(\theta|x) = \frac{f(\theta, X_1, \cdots X_n)}{f(X_1, \cdots X_n)} = \frac{L(X_1, \cdots X_n|\theta)\pi(\theta)}{\sum L(X_1, \cdots X_n|\theta)\pi(\theta)} \tag{8}$$

$\pi(\theta)$ represent the prior distribution, $L(X_1, \cdots X_n|\theta)$ is the probability of the data given our prior belief and it is referred to as the likelihood, and the updated $f(\theta|x)$ is the posterior.

2.4 Linear Models

As stated in Li (2008), regression models are commonly used to assess the relationship between response variables and explanatory variables. They can be used for prediction, inference and hypothesis testing. A standard regression specification for continuous data is the linear model

$$\begin{aligned} Y &= X\beta + \varepsilon, \\ \varepsilon &= MVN(0, \delta^2 I) \end{aligned} \tag{9}$$

Where $Y = (Y_1, Y_2, \cdots)^T$ is a vector of response variables, X is a $n \times p$ matrix of covariates and β is the p - vector of regression coefficients. The vector ε is of dimension n and the elements of ε are assumed to be independent and identically distributed based on Gaussian specification $\varepsilon \sim N(0, \delta^2)$. This model is appropriate when the response data is from normal distribution, however if normality assumption does not hold, a generalized model proposed by Nelder and Wedderburn (1972) is more appropriate for the exponential family of distributions. Examples of distributions that belong to the exponential family include Bernoulli, binomial, Poisson, beta, gamma and negative binomial distributions. Li (2008) also explains that the generalized linear models are used to analyze data under independence assumptions.

From Uppal et al. (2012), for n trials in each of which there are only two outcomes (failure and success) with probabilities f and s respectively, the probability $p(y)$ of exactly y successes is given by

$$p(y) = \binom{n}{y} f^{n-y} s^{y}, \text{ y=0, 1, 2, ……} \tag{20}$$

The area of study will be divided into n disjoint regions. The regions under consideration will be the 47 counties in the country.

Let y_{ij} be the literacy status of the j^{th} individual in the i^{th} county. This variable is binary since y_{ij} is discrete count data, so the Bernoulli distribution can be used to model the data counts.

$$y_{ij} \sim Bern(p_{ij}) \tag{31}$$

Where p_{ij} is the probability that the j^{th} adult in the i^{th} county is illiterate. Then the mass function of y_{ij} can be expressed as

$$f(y_{ij} \mid p_{ij}) = (p_{ij})^{y_{ij}} (1 - p_{ij})^{1-y_{ij}} \quad (42)$$

According to Li (2008), we model the expected value, p_{ij}, of y_{ij} as a linear function of the covariates x_{ij} after transformation by

$$\eta_{ij} = g(p_{ij}) = x_{ij}{}^{T}\beta \quad (53)$$

where $g(.)$ is the link function which is usually a nonlinear monotonically increasing function, transforming p_{ij} to the real line.

When, $y_{ij} \sim Bern(p_{ij})$, the Bernoulli logistic regression is obtained through the logit link

$$g(p_{ij}) = \log \frac{p_{ij}}{1 - p_{ij}} = x_{ij}{}^{T}\beta \quad (64)$$

that constrains p_{ij} between 0 and 1.

2.5 Statistical Modeling of adult illiteracy in Kenya

The covariates that were considered included gender, disability, awareness of literacy programs and age. The variable y_{ij} is the illiteracy status of the j^{th} individual in the i^{th} county. y_{ij} being discrete , i.e count data. The PMF is

$$f(y_{ij} \mid p_{ij}) = (p_{ij})^{y_{ij}} (1 - p_{ij})^{(1-y_{ij})} \quad (75)$$

The likelihood function of the observed data is then given by

$$L(y_{ij} \mid p_{ij}) = \prod_{i=1}^{n} (p_{ij})^{y_{ij}} (1 - p_{ij})^{(1-y_{ij})} \quad (86)$$

The posterior distribution of the parameter, p_{ij}, was computed as

$$\pi(p_{ij} \mid y_{ij}) = \frac{L(y_{ij} \mid p_{ij}) \times \pi(p_{ij})}{g(y_{ij})} \quad (97)$$

Where $g(y_{ij}) = \sum_{p_{ij}} f(y_{ij} \mid p_{ij}) \times \pi(p_{ij})$ is the marginal distribution of Y_{ij} of the i^{th} county.

Four models were compared. The probability was linked to other parameters and co-variates via a linear predictor. Individual level co-variates, X_{ij}, related to illiteracy were introduced into the model as follows;

$$logit(p_{ij}) = \boldsymbol{X}_{ij}^{T}\boldsymbol{\beta} \quad (108)$$

Where the β parameters are assumed to have prior distributions so that

$$P(\beta, \tau_{\beta} \mid y_{ij}) = L(y_{ij} \mid \beta, \tau_{\beta}) \times f(\beta \mid \tau_{\beta}) \quad (119)$$

Where $f(\beta \mid \tau_{\beta})$ is the joint distribution of the co-variates parameters conditional on the hyper-parameter vector τ_{β}. Regarding these parameters as independent,

$$f(\beta \mid \tau_{\beta}) = \prod_{j=1}^{p} f_{j}(\beta_{j} \mid \tau_{\beta_{j}}) \quad (20)$$

Assigning to the co-variate parameters a Gaussian distribution and assuming that they correlate $\beta \sim N_{p}(0, \Sigma_{\beta})$, so that

$$f(\beta \mid \tau_{\beta}) = \prod_{j=1}^{p} N(0, \tau_{\beta_{j}}) \quad (21)$$

At the next level of hierarchy, a prior distribution of $\tau_{\beta_{j}}$ was taken as

$$\tau_{\beta_{j}} \sim Gamma(\alpha, \beta) \quad (22)$$

The probability was then linked to other parameters and co-variate and random effects via a linear predictor such as, Lawson (2009). Therefore,

$$logit(p_{ij}) = \boldsymbol{X}_{ij}^{T}\boldsymbol{\beta} + \mathbf{U}_{ij}^{T}\mathbf{l} \quad (23)$$

Assuming that a logistic link was appropriate for the probability and that the random effect at the individual county was included, then

$$p_{ij} = \frac{\exp(\boldsymbol{X}_{ij}^{T}\boldsymbol{\beta} + \mathbf{U}_{ij}^{T}\mathbf{l})}{1 + \exp(\boldsymbol{X}_{ij}^{T}\boldsymbol{\beta} + \mathbf{U}_{ij}^{T}\mathbf{l})} \quad (24)$$

Hence the hierarchy is as follows;

$$
\begin{aligned}
& y_{ij} \sim Bern(p_{ij}) \\
& logit(p_{ij}) = \boldsymbol{X}_{ij}^{T}\boldsymbol{\beta} + \mathbf{U}_{ij}^{T}\mathbf{I} \\
& \beta_{ij} \mid \tau_{\beta_{ij}} \sim N(0, \tau_{\beta_{ij}}) \\
& \tau_{\beta_{ij}} \sim Gamma(\alpha, \beta) \\
& u_{ij} \mid u_{ji} \sim N(u_{u_{ij}}, \delta_{u_{ij}}^{2})
\end{aligned}
\tag{25}
$$

Secondly, incorporating unstructured heterogeneity in the model in addition to the co-variates,

$$logit(p_{ij}) = \boldsymbol{X}_{ij}^{T}\boldsymbol{\beta} + \boldsymbol{v_i} \tag{26}$$

Where $\boldsymbol{v_i}$ was modeled using a Gaussian process with mean zero.

Thirdly, for correlated heterogeneity,

$$logit(p_{ij}) = \boldsymbol{X}_{ij}^{T}\boldsymbol{\beta} + \boldsymbol{u_i} \tag{27}$$

Where $\boldsymbol{u_i}$ is the correlated component.

To capture both spatial variation and unstructured heterogeneity the following model was used;

$$logit(p_{ij}) = \boldsymbol{X}_{ij}^{T}\boldsymbol{\beta} + \boldsymbol{v_i} + \boldsymbol{u_i} \tag{28}$$

To assess the quality of the estimators, we used a loss function, $L(p_i, a_i)$ that measures the loss incurred by using $a_{ij} = \hat{p}_{ij}$ as an estimate of p_{ij} .The "best" estimate is chosen so as to minimize the Expected loss, $E(L(p_i, \hat{p}_i))$, where the expectation is taken over p_i with respect to the posterior distribution $f(y_i \mid p_i)$.

Samples were obtained from the posterior distribution using the Markov Chain Monte Carlo simulation technique. Parameters were estimated using sample quantiles from the posterior distribution. The estimate of the marginal distribution of the parameter p_i was obtained from the empirical distribution of the sample values, Lawson (2009).

Since the modeling was done under the Bayesian platform, each of the parameters in β was treated as a random variable, hence was given a prior distribution. This allowed adjustment of the uncertainty parameters by assigning prior distributions to the parameters, Osei (2010).

3. Results

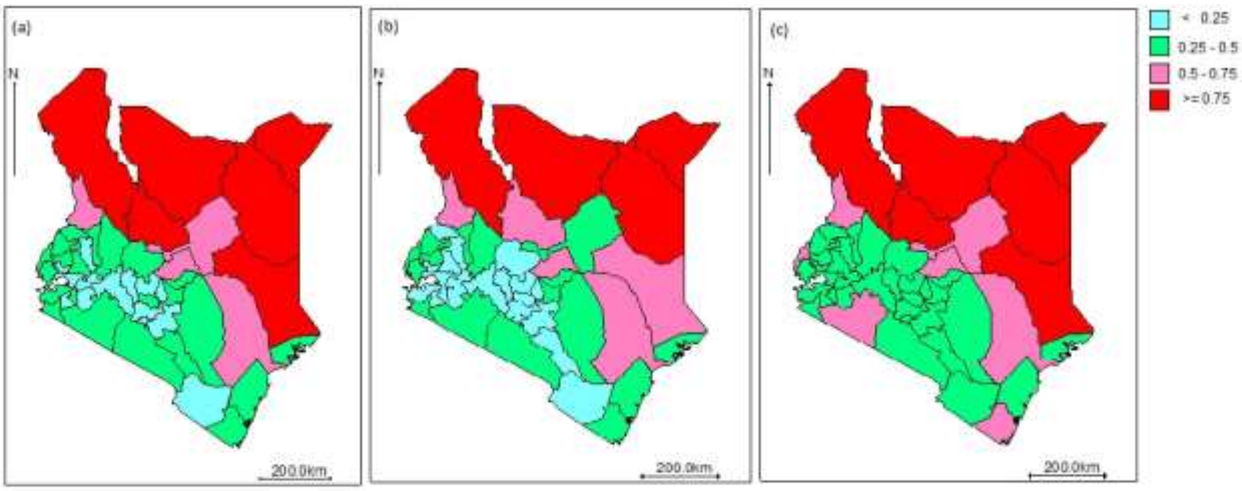
The four models were applied and the best model based on DIC identified. The significant covariates that explain variation in illiteracy are reported. Finally the smooth maps are presented. All this was done using WINBUGS software.

Table of results:

Parameter	Model 1	Model 2	Model 3	Model 4
Intercept	-2.601(-2.723,- 2.470)	-2.852(-3.197,- 2.439)	-2.201(-2.353,- 1.900)	-2.251(-2.323,- 1.970)
aware	1.193(1.104,1.287)	0.867(0.790,0.948)	0.865(0.785,0.946)	0.865(0.785,0.946)
disab	1.335(1.235,1.433)	1.275(1.169,1.396)	1.271(1.160,1.390)	1.274(1.163,1.391)
Sex	0.537(0.495,0.581)	0.437(0.400,0.476)	0.437(0.399,0.478)	0.437(0.340,0.479)
Age	0.050(0.048,0.053)	1.070(1.067,1.073)	0.068(0.064,0.071)	0.067(0.064,0.071)
DIC	17219	13882.7	13881.3	13884.7

The best fitting model is Model 3 in which the CAR spatial random effects have been incorporated. There is a moderate association between illiteracy and disability as well as with awareness of the literacy programs. For every one unit increase in age, the expected change in log odds is 0.068. The smoothed maps are as follows.

Smoothed Map:



4. Conclusion and Recommendations

The factors considered in this study reflect the individual differences at aggregate level. Evidently, the illiteracy level is highest in the Northern part of the country. This could be due to the low level of development, culture and economic activities.

The maps give clear guidance to policy makers on which regions to target the limited resources available and also tailor make the illiteracy alleviation programs and campaigns to suit the cultural and social-economic practices of the people in that county.

If the regions with high illiteracy are not sensitized and given extra attention then the literacy level of the whole country will be dragged down; this will in turn affect negatively the social economic status of the whole nation as Carron et al. (1989). Consequently a nation with a tainted social economic development record is also prone to political instability Iftikhar (2011).

This approach is not only applicable to data on illiteracy; it can be used to model other aspects in the counties such as epidemiology, climatology and accidents among others.

Illiteracy levels at any region will vary from time to time. As such, an extension of our model, to include illiterate individuals in county i at time t, and a vector of time dependent covariates incorporated in the logit model would make it possible to examine the time-lagged associations of adult illiteracy over time. Indeed, levels of adult illiteracy evolve over time and allowing the regression parameters to evolve over time is a worthy avenue of future research.

REFERENCES

Ahmed, M.(2009). The state and development of adult learning and education in Asia and Pacific Regional synthesis report. Technical report, Hamburg: UNESCO: Institute for Life-long Learning.

Banerjee, S, Carlin, P. B., and Gelfand, E. A. (2004). Hierarchical Modeling and Analysis for Spatial Data. New York: CRC.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society, B 36:192–236.

Besag, J.,York, J., and Mollie, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). Ann Inst Stat Math, 43:1–59.

Carron, Mwiria, K., and Righa, G. (1989). The functioning and effects of the Kenya literacy program: A view from the local level (research report no.77). Technical report, Paris:International Institute for Educational Planning.

Cotterell, A. (1975). Adult illiteracy: an urgent problem. EMERALD, Vol. 17 Iss: 1/2:7 –8.

Cressie, N. A. (1993). Statistics for Spatial Data. New York: Wiley.

Dobson, A. J. (1990). An Introduction to Generalized Linear Models. London,Chapman and hall.

Gamerman, D. and Lopes, H. F. (2006). Markov Chain Monte Carlo. New York: CRC.

Gill, J. (2000). Generalised linear models : A unified approach. Papers Series on Quantitative Applications in Social Science.32

Iftikhar, H. (2011). Spatial analysis of rural illiteracy pattern in punjab: Variations in literacy patterns in punjab and the factors accompanying this variation.

Kirsch, I. S., Jungeblut, A., Jenkins, L., and Kolstad, A. (2002). Adult literacy in america. Technical report, U.S. Department of Education Office of Educational Research and Improvement.

KNALS (2007). Kenya national adult literacy survey report. Technical report, Kenya National Bureau of Statistics.

Koch, K. R. (2007). Introduction to Bayesian Statistics. Springer.

Kothari, C. R. (1990). Research Methodology. NEW AGE INTERNATIONAL (P) LIMITED PUBLISHERS.

Lawson, A., Browne, W., Rodeiro, C., and Wiley, J. (2003). Disease mapping with Win-BUGS and MLwiN.

Lawson, A. B. (2009). BAYESIAN DISEASE MAPPING: Hierarchical Modeling in Spatial Epidemiology. Chapman and Hall/CRC.

Li, H. (2008). Bayesian hierarchical models for spatial count data with application to fire frequency in British Columbia. Master's thesis, University of Victoria.

Mazumdar, K. (2005). Socio-economic factors determining adult literacy in developing countries. International Journal of Social Economics, Vol. 32 Iss: 1:98 – 120.

Mitra, A. and Singh, P. (2008). Trends in literacy rates and schooling among the scheduled tribe women in India. International Journal of Social Economics, 35 Iss:1: 99 – 110.

Mollie., A. (1996). Bayesian Mapping of Diseases.

NAAL (2003). A first look at the literacy of Americas adults in the 21st century. Technical report, U.S. Department of Education, Institute of Education Sciences NCES 2006-470.33

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. Journal of the Royal Statistical Society., A 135:370–384.

Oliveira, V. D. and Song, J. J. (2008). Bayesian analysis of simultaneous autoregressive models. The Indian Journal of Statistics, Volume 70-B, Part 2:323–350.

Osei, P. P. (2010). Statistical methods for disease mapping. Master's thesis.

Raudenbush, S. W., Cheong, Y. F., and Fotiu, R. P. (1996). Social inequality, social segregation, and their relationship to reading literacy in 22 countries. Collected papers from the IEA Reading Literacy Study, 4:5–53.

Tsokos, C. P. and M., R. K. (2009). Mathematical Statistics with Applications. Academic Press.

UNESCO (2010). UNESCO Country Programming Document Kenya.

Uppal, M. S., Otieno, O. R., Murray, H. H., and Otieno, O. G. (2012). Probability Distribution Of a Single Random Variable. New Age International Publishers.

You, Y. and Zhou, Q. M. (2011). Hierarchical bayes small area estimation under a spatial model with application to health survey data. Technical report, Statistics C