Generalized Linear Models for Rainfall Patterns In The Presence Of Temperature Trends in Different Climatic Zones in Kenya

PETER MURAGE[1,] JOSEPH MUNG'ATU[2], GEORGE ORWA[3]

1. Phd (Applied Statistics) Student Jomo Kenyatta University of Agriculture and Technology. P.o Box 6200-00200 Nairob-Kenya.Cell phone +254-0710507266

[*]E-mail: gitmurage@gmail.Com or githinji.peter@rocketmail.com

2. Department of Statistics and Actuarial Science Jomo Kenyatta University of Agriculture And Technology Kisumu Campus P.o Box 6200-00200 Nairobi-Kenya cell phone +254-0720824708

E-mail: kmungatu@yahoo.com or j.mungatu@fsc.jkuat.ac.ke

3. Chairman Department of Statistics and Actuarial Science Jomo Kenyatta University of Agriculture and Technology P.o Box 6200-00200 Nairobi-Kenya

Cell phone +254-0722940181

E-mail:orwagoti@yahoo.com or barendvanhoek@yahoo.com

Abstract

Rainfall and temperature series impact heavily on the performance of a country's agricultural production as a component of the economy especially in a developing country like Kenya. These series were analyzed to study the evolution of their mean variability. In particular, this study sought to model the Temporal Rainfall Patterns in different Zones in Kenya considering temperature trends as indicators of the rainfall variations. In achieving this objective, a broad statistical approach was used, based on inference on their entire series to predict the mean amount of rainfall using the Maximum and Minimum temperatures processes. Data on different counties in Kenya regarding rainfall and temperature was obtained from the respective weather stations. The study then came up with zones dependent on rainfall using cluster Analysis. Models specific to each created zone were fitted and included the Poisson, Quasi-Poisson and Negative Binomial models which belong to the broad class of Generalized Linear Models (GLMs). Using the AIC criterion, we identified Negative Binomial model as the best model for the variability of rainfall patterns in these Zones as maximum and minimum temperatures changes, and the discussion about them made thereof.

Keywords: Cluster analysis, Poisson, Quasi-Poisson, Negative binomial, Precipitation, Temperature.

Introduction

Cluster analysis is a major technique for classifying a 'mountain' of information into manageable meaningful piles. It is a data reduction tool that creates subgroups that are more manageable than individual datum. According to Aldenderfer (1985), it examines the full complement of inter-relationships between variables. We used Hierarchical cluster analysis (k-means) method to come up with seven zones. This is the major statistical method for finding relatively homogeneous clusters of cases based on measured characteristics Everitt (2001), Dolnicar (2009), Kaufman (2005). It starts with each case as a separate cluster, i.e. there are as many clusters as cases, and then combines the clusters sequentially, reducing the number of clusters at each step until only one cluster is left. The clustering method uses the dissimilarities or distances between objects when forming the clusters.

Generalized linear models (GLMs) are a technique for modeling introduced by Nelder (1972) and further developed by McCullagah (1989). Methods used in our study included Poisson, Quasi-Poisson and Negative Binomial models. Modeling event counts is important in many fields. Poisson regression model for count data is often of limited use in these disciplines because empirical count data sets, typically exhibit over-dispersion and/or an excess number of zeros. Poisson model assumes the equidispersion of the data Joeh (2005). Modeling count variables is a common task in economics and the social sciences. The quasi-Poisson model and negative binomial model can account for overdispersion Hoef (2007), and both have two parameters. Poisson distribution is widely assumed for modeling the distribution of the observed counts data. It assumes that the mean and variance are equal Hilbe (1994) and Hoffman (2004). However, this restriction is violated in many applications because data are often overdispersed Sileshi (2006). In this case, Poisson distribution underestimates the dispersion of the observed counts. The overdispersion occurs when the single parameter k of Poisson distribution is unable to fully describe event counts. Generally, two sources of overdispersion are determined: heterogeneity of the population. A common way to deal with overdispersion for counts is to use a generalized linear model framework McCullagh and Nelder (1989), where the most common approach is a ‘‘quasi-likelihood,’’ with Poisson-like assumptions (that we call the quasi-Poisson from now on) or a negative binomial model. The Objective of this statistical work is to apply the concepts of Poisson, Quasi-Poisson and Negative Binomial to choose a suitable model that will describe a good relationship between the rainfall, maximum and minimum temperature which can be used to predict and forecast the amount of rainfall expected in a given zone.

Methodology

Cluster analysis (zoning)

The objective of this cluster analysis is to identify in Kenya, regions that are very similar with regard to the relationship between rainfalls, maximum and minimum temperature in those regions. According to, Anderberg (1973), Ball (1967), Dolnicar S. (2003) and Romesburg(2004), Cluster analysis (CA) in our case called zones, is an exploratory data analysis tool for organizing observed data into meaningful zones or clusters, based on combinations of independent variables, which maximizes the similarity of cases within each cluster (zone) while maximizing the dissimilarity between groups that are initially unknown. In this sense, CA creates new groupings without any preconceived notion of what clusters (zones) may arise. Each cluster (zone) thus describes, in terms of the data collected, the class to which its members belong. Area(s) in each cluster (zone) are similar in some ways to each other and dissimilar to those in other clusters (zones). Cluster analysis reduces the number of observations or cases by grouping them into a smaller set of clusters. The collected data of the amount of rain, amount of maximum temperature and minimum temperature were grouped together and using r-package for Hierarchical and k-means clustering methodology, seven zones were created and discussions about them made thereof.

GENERALIZED LINEAR MODELS (GLMs)

GLMs represents a class of regression models that allow us to generalize the regression approach to accommodate many types of response variables including Count, binary, proportions and positive valued continuous distributions. According to Hilbe (1994), Nelder (1972) and Hoffman (2004) generalized linear models include three components.

A random component which is the response and associated probability distribution, a systematic component which includes explanatory variables and relationship among them (e.g. interaction terms), A link function which specifies the relationship between the systematic component or linear predictor and the mean of the response. It is the link function that allows generalization of the linear models for count data, binomial and percent data thus ensuring linearity and constraining the predictions to be within a range of possible values Guisan (2002).

Stated formally the three components of GLMs are:

$$g(\mu) = \mu$$

$$\mu = \beta_0 + \beta_i x_i + \beta_{i+1} x_{i+1} + ...\qquad 1$$

Where μ is the mean of response in the equation 1 above

g (μ) shows that the mean of the response is linked to the structural model, thus

g (μ) in equation 1 above is the link function and η is the structural component, with X_i

denoting explanatory Variables and β_i the parameters to be estimated

Poisson GLM model

Y_i is Poisson distributed with mean μ_i. By definition of this distribution, the variance of

Y_i is also equal to μ_i. The systematic part is given by equation 2 shown below

$$(x_{i1}, x_{i2}, ...x_{iq}) = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_q x_{iq} \qquad 2$$

There is a logarithmic link between the mean of Y_i and the predictor function $\eta(x_{i1}, x_{i2} ....x_{iq})$

The logarithmic link (also called a log link) in equation 2 above, ensures that the fitted values are always non-negative. As a result, the following is obtained. $Y_i \sim P(\mu_i)$

E (Y_i) = μ_i and var (Y_i) = μ_i

log $(\mu_i) = (x_{i1}, x_{i2} ....x_{iq})$ = or $\mu_i = e^{(X_{i1}, X_{i2} ....X_{iiq})}$

Overdispersion

According to Anderson (2008) statistical methods are often based on the iid assumption: independent and identically distributed data. This assumption is nearly always made in application (time series and spatial models are exceptions). However in reality, data are often Somewhat dependent and not identically distributed. These conditions fall under the concept of overdispersion. Count data (zero and positive integers stemming from some count) are often said to be over dispersed. Data are over dispersed if the conditional variance exceeds the conditional mean. The assumption is that there is no structural lack of fit, thus lack of fit can only be blamed on overdispersion. If there is a reason to suspect over dispersion, then an over dispersion parameter θ can be estimated by $\hat{\theta} = \frac{x^2}{p}$

Where x^2 is the usual goodness of fit test statistic based on the global model and p is the degrees of freedom for the test. The over dispersion parameter θ is also called a variance inflation factor. Under the iid $\theta \equiv 1$

The models used for overdispersed data are Quasi-Poisson and Negative Binomial glm models.

Quasi-Poisson GLM

This model uses the mean regression function and the variance function from the Poisson Glm but leaves the dispersion parameter θ unrestricted. Thus θ is not assumed to be fixed at 1 but is estimated from the data. This strategy leads to the same coefficients estimates as the standard Poisson model but inference is adjusted for overdispersion. Consequently quasipoisson adopt the estimating function view of the Poisson model. For a quasipoisson, the variance is assumed to be a linear function of the mean. If Y is a random variable, then E (Y) = μ and var. (Y) = vpoi (μ) = $\theta\mu$ where E (Y) is the expectation of Y. var (Y) is the Variance of Yi and $\mu > 0$, $\theta > 1$.E (Y) is also known as the mean of the distribution. θ is an over dispersion parameter. According to Lee [2000], the quasipoisson model formulation has the advantage of leaving parameters in a normal, interpretable state and allows standard model diagnostic without a loss of efficient fitting algorithms. Quasi-likelihood approach does not attempt to make any specification of the sampling distribution of Yi. i.e. there is no specification of a quasipoisson distribution.

Negative Binomial GLM

The Negative Binomial (NB) glm is given by:-

Y_i is negative binomial distributed with mean μ_i and parameter k. The systematic part is given by

$$(x_{i1},x_{i2}\ldots\ldots x_{iq}) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots\ldots + \beta_q x_{iq} \qquad 3$$

There is a logarithmic link between the mean of Y_i and the predictor function _ (x_{i1},x_{i2}……………………….x_{iq}) in equation 3 above

The logarithmic link (also called log link) ensures that the fitted values are always non-negative. The three steps yields Yi ~ NB (μ_i; k) E (Yi) = μ_i and var. (Yi) = $\mu_i + \frac{\mu^2}{k}$

$\log\ (\mu_i = \eta(x_{i1}, x_{i2}...............x_{iq})$ or $\mu_i = e^{(X_{i1}, X_{i2}, ...X_{iq})}$

To estimate the regression parameters, we need to specify the likelihood criterion and obtain the first order and second order derivatives. Recalling that pdf of a Negative binomial distribution is as shown in equation 4 below;

$$f(y_i; k, \mu_i) = \frac{\Gamma(y_i + k)}{\Gamma(k) \times \Gamma(y_i + 1)} \times \left(\frac{k}{\mu + k_i}\right)^k \times \left(1 - \frac{k}{\mu_i + k}\right)^y \qquad 4$$

Then this probability function is then used in the log likelihood criterion shown in equation 5 below

$$\log(L) = \sum_i \log(f(y_i; k, \mu_i)) \qquad 5$$

substituting equation 4 into equation 5 leads to

$$\log(L) = \sum \left\{ k \times \log\left(\frac{k}{\mu_i + k}\right) + y_i \times \log\left(\frac{\mu_i}{\mu_i + k}\right) + \log(\Gamma(y_i + k) - \log\Gamma(k) - \log(\Gamma(y_i + 1))) \right\}$$

which is the equation used to determine the parameters of the Negative binomial through first derivative and second derivative as

$$E(Y_i) = \mu_i \quad \text{and} \quad Var(Y_i) = \mu_i + \frac{\mu^2}{k}$$

RESULTS AND DISCUSSION

1) ZONING

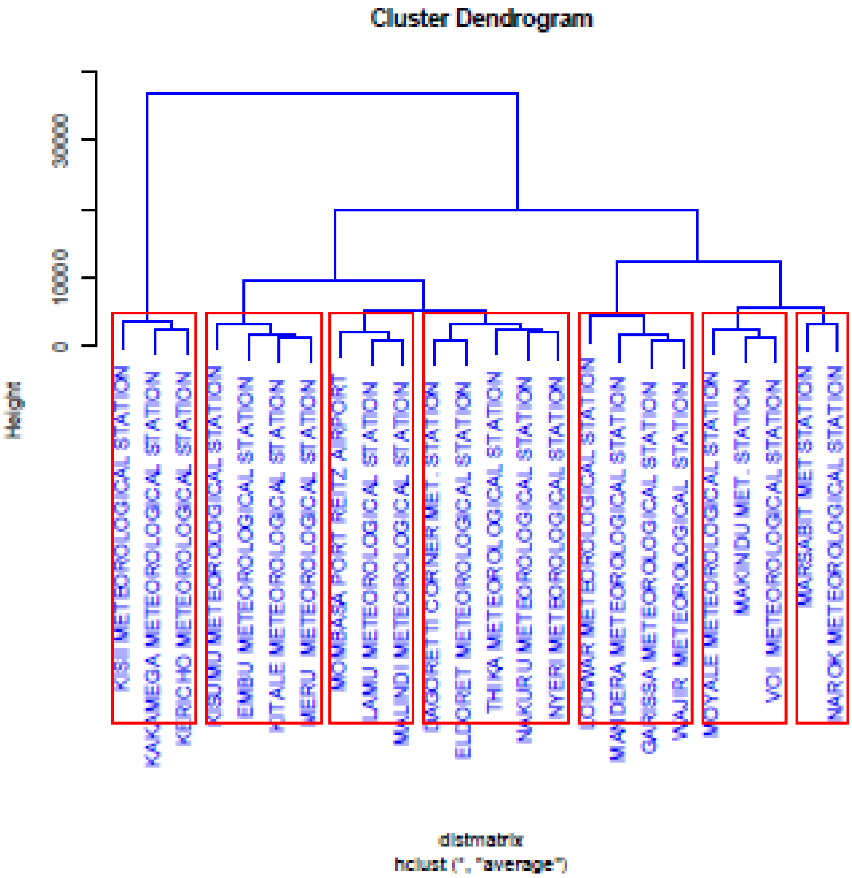

Figure.1 Precipitation, Maximum and Minimum Temperature

The variables used to create these zones in Figure 1, are amount of rainfall resulting as maximum and minimum temperature varies over time. Figure 1 is the result of cluster analysis we carried out using r-software. Zone 1 comprises of the regions represented by Kisii, Kakamega and Kericho meteolorogical weather stations. This is a zone that receives very high amounts of rainfall of above 200mm per year on average. This region experiences low minimum temperatures of between 10-15^{0}c and also low maximum temperatures of between 23–25^{0}c. This is a cool zone. Zone 2 comprises of regions represented by Kisumu, Embu, Kitale and Meru weather stations. The minimum temperatures are slightly higher than those of zone 1 ranging from 12–16^{0}c. The maximum temperature ranges between 23–29^{0}c. This zone is slightly warmer and receives an average rain of approximately 200mm per year which is fairly high. The four regions that make this zone are quite far from one another Kisumu being around lake Victoria, Embu and Meru being around Mt Kenya and Kitale being on the North Rift of Kenya. Zone 3 is made up of regions represented by Mombasa, Lamu and Malindi weather stations. The minimum temperatures in this zone range between 22-23^{0}c while the maximum temperature ranges between 30-32^{0}c. The average amount of rainfall is slightly below 200mm. Zone 4 comprises of regions represented by Dagoretti, Eldoret, Thika, Nakuru and Nyeri weather stations. The minimum temperature in this zone ranged between 9–12^{0}c while maximum temperature ranged between 23–24^{0}c. The average amount of rain experienced in this zone is between 100 –150mm. The frequency of the rain in this zone is high. This makes the regions to be wet. Zone 5 is made by regions represented by Lodwar, Mandera, Garissa and Wajir weather stations whereby the minimum temperature ranged between 22–24^{0}c while maximum temperature ranged between 32 -33^{0}c. The rainfall experienced in this region was 30mm on average which is extremely low. This is a very hot and dry zone. The regions of this zone are found on the northern part of Kenya. Zone 6 is made by regions represented by Moyale, Makindu and Voi weather stations whose minimum temperature ranged between 17 -19^{0}c while maximum temperature ranged between 27–29^{0}c. The rainfall experienced in this zone was between 50mm and 60mm which is very low. This was a very hot and the second most dry zone. Geographically, Moyale is in the northern Kenya while Makindu and Voi are on southern part of Kenya. Zone 7 comprised of regions represented by Marsabit and Narok weather stations, whose minimum temperature ranged between 10 -16^{0}c while maximum temperature was about 24^{0}c. The rainfall experienced in this zone was about 80mm which is fairly low. This was a moderately warm and fairly dry zone. Marsabit lies in the northern Kenya while Narok is in the southern part of Kenya.

We note here that the naming of the zone as number 1 to 7 in the figure 1 above, is arbitrary and has no bearing to the amount of rainfall experienced by the zone.

2) GLMs

The data collected, was a count data. Several count models were investigated to see which would best fit the data for each zone? The investigation was based on how well Poisson, Quasi-Poisson and Negative Binomial Generalized Linear Models (GLMs) would explain the relationship between rainfall, maximum temperature and minimum temperature in all our zones above. Then the Akaike Information Criterion (AIC) was used to select the best model that would help explain our problem. The best model is the one that have the lowest AIC. In writing the glm model, x_1 represented Maximum temperature and x_2 represented Minimum temperature while μ represented the mean amount of rainfall.

The following are the results of models using the criterion of Precipitation, Maximum And Minimum Temperature.

Zone 1; Precipitation, Maximum and Minimum temperature

Poisson distribution Coefficients

	Estimate	standard error	p.value
Intercept	7.422	0.03167	$2e^{-16}$
Temperature maximum	-0.1367	0.00150	$2e^{-16}$
Temperature minimum	0.08663	0.00129	$2e^{-16}$

$$AIC = inf.$$

.

Dispersion parameter for Poisson family was taken to be 1

Quasi-Poisson distribution Coefficients

	Estimate	standard error	p.value
Intercept	7.422	0.195	$2e^{-16}$
Temperature maximum	-0.1367	0.00927	$2e^{-16}$
Temperature minimum	0.0866	0.00794	$2e^{-16}$

$$AIC = N/A$$

Dispersion parameter was taken to be 37.93

Negative Binomial Coefficients

	Estimate	standard error	p.value
Intercept	7.699	0.2043	$2e^{-16}$
Temperature maximum	-0.1617	0.00953	$2e^{-16}$
Temperature minimum	0.113	0.00889	$2e^{-16}$

$$AIC = 12243$$

Dispersion parameter for negative binomial(3.7441) was taken to be 1
Theta is 3.744
standard error 0.163

Table 1; Kisii, Kakamega and Kericho

Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.5, No.3, 2015

Zone 2;Precipitation, Maximum and Minimum temperature

Poisson distribution Coefficients

	Estimates	standard error	p.value
Intercepts	5.0355	0.02285	$2e^{-16}$
Temperature maximum	-0.0038096	0.000921	$2e^{-16}$
Temperature minimum	0.04130	0.000628	$2e^{-16}$

$$AIC = inf.$$

Dispersion parameter for Poisson family was taken to be 1

Quasi-Poisson distribution Coefficients

	Estimates	standard error	p.value
Intercepts	5.03545	0.22185	$2e^{-16}$
Temperature maximum	-0.0381	0.00895	$2.2e^{-05}$
Temperature minimum	0.0413	0.0061	$1.84e^{-11}$

$$AIC = N/A$$

Dispersion parameter for Quasi-Poisson family was taken to be 94.3028

Negative Binomial distribution Coefficients

	Estimates	standard error	p.value
Intercepts	5.0028	0.2501	$2e^{-16}$
Temperature maximum	-0.05454	0.01015	$7.74e^{-08}$
Temperature minimum	0.07128	0.00794	$2e^{-16}$

$$AIC = 17162$$

Dispersion parameter for Negative Binomial(0.904) was taken to be 1
Theta 0.904
standard error 0.0302

Table 2; Kisumu, Embu, Meru and Kitale

Zone3; Precipitation, Maximum and Minimum temperature

Poisson GLM Coefficients

	Estimates	standard error	p.value
Intercepts	7.06656	0.0191356	$2e^{-16}$
Temperature Maximum	-0.1650170	0.0010564	$2e^{-16}$
Temperature Minimum	0.1051609	0.0006899	$2e^{-16}$

$$AIC = inf.$$

Dispersion parameter for Poisson family was taken to be 1

Quasi-Poisson GLM Coefficients

	Estimates	standard error	p.value
Intercepts	7.066559	0.175697	$2e^{-16}$
Temperature Maximum	-0.165017	0.009700	$2e^{-16}$
Temperature Minimum	-0.105161	0.006335	$2e^{-16}$

$$AIC = N/A$$

Dispersion parameter for Quasi-Poisson was taken to be 84.30281

Negative Binomial GLM Coefficients

	Estimates	standard error	p.value
Intercepts	8.092429	0.198923	$2e^{-16}$
Temperature Maximum	-0.217046	0.010306	$2e^{-16}$
Temperature Minimum	0.126329	0.006458	$2e^{-16}$

$$AIC = 30998$$

Dispersion parameter for Negative Binomial(0.8243) was taken to be 1
Theta 0.843
standard error 0.0203

Table 3; Mombasa, Lamu and Malindi

Zone 4; Precipitation, Maximum and Minimum temperature

Poisson GLM Coefficients

	Estimates	standard error	p.value
Intercepts	3.856176	0.063895	$2e^{-16}$
Temperature Maximum	-0.222177	0.002570	$2e^{-16}$
Temperature Minimum	0.351388	0.002223	$2e^{-16}$

$$AIC = inf.$$

Dispersion parameter for Poisson family was taken to be 1

Quasi-Poisson GLM Coefficients

	Estimates	standard error	p.value
Intercepts	3.85618	0.53348	$1.19e^{-12}$
Temperature Maximum	-0.22218	0.02146	$2e^{-16}$
Temperature Minimum	0.35139	0.01856	$2e^{-16}$

$$AIC = N/A$$

Dispersion parameter for Quasi-Poisson family was taken to be 69.71252

Negative Binomial GLM Coefficients

	Estimates	standard error	p.value
Intercepts	0.90442	0.74472	0.225
Temperature Maximum	-0.43587	0.03562	$2e^{-16}$
Temperature Minimum	0.84096	0.03867	$2e^{-16}$

$$AIC = 6533.6$$

Dispersion parameter for Negative Binomial(0.4657) was taken to be 1
Theta 0.4657
standard error 0.0236

Table 4; Dagoretti, Eldoret, Thika, Nakuru and Nyeri

ZONE 5: Precipitation, Maximum and Minimum temperature

Poisson GLM Coefficients

	Estimates	standard error	p.value
Intercepts	6.217790	0.046432	$2e^{-16}$
Temperature Maximum	-0.122618	0.002049	$2e^{-16}$
Temperature Minimum	0.066511	0.001286	$2e^{-16}$

$$AIC = inf.$$

Dispersion parameter for Poisson family taken to be 1

Quasi-Poisson GLM Coefficients

	Estimates	standard error	p.value
Intercepts	6.21779	0.43742	$2e^{-16}$
Temperature Maximum	-0.12262	0.01930	$3.15e^{-10}$
Temperature Minimum	0.06651	0.01211	$5.01e^{-08}$

$$AIC = N/A$$

Dispersion parameter for Quasi-Poisson family taken to be 88.7472

Negative Binomial GLM Coefficients

	Estimates	standard error	p.value
Intercepts	6.70110	0.42906	$2e^{-16}$
Temperature Maximum	-0.13273	0.01841	$5.66e^{-13}$
Temperature Minimum	0.05144	0.01143	$6.83e^{-06}$

$$AIC = 10229$$

Dispersion parameter for Negative Binomial(0.6043) was taken to be 1

Theta 0.6043

standard error 0.0242

Table 5;Lodwar, Mandera, Garissa and Wajir

Zone 6; Precipitation, maximum and minimum temperature

Poisson GLM Coefficients

	Estimates	standard error	p.value
Intercepts	17.89414	0.46791	$2e^{-16}$
Temperature Maximum	-0.40833	0.01043	$2e^{-16}$
Temperature Minimum	-0.03456	0.01216	0.00449

$$AIC = inf.$$

Dispersion parameter for Poisson family taken to be 1

Quasi-Poisson GLM Coefficients

	Estimates	standard error	p.value
Intercepts	17.89414	2.82534	$7.01e^{-10}$
Temperature Maximum	-0.40833	0.06297	$2.89e^{-10}$
Temperature Minimum	-0.03456	0.07344	0.638

$$AIC = N/A$$

Dispersion parameter for Quasi-Poisson family taken to be 36.46082

Negative Binomial GLM Coefficients

	Estimates	standard error	p.value
Intercepts	13.79769	3.51149	$8.52e^{-05}$
Temperature Maximum	-0.38025	0.07779	$1.02e^{-06}$
Temperature Minimum	0.09480	0.07983	0.235

$$AIC = 2384.8$$

Dispersion parameter for Negative Binomial(0.279) taken to be 1
Theta 0.279
standard error 0.6043

Table 6; Moyale, Makindu and Voi

Zone 7; Precipitation, Maximum and Minimum temperature

Poisson GLM Coefficients

	Estimates	standard error	p.value
Intercepts	3.987714	0.094365	$2e^{-16}$
Temperature Maximum	-0.269421	0.002657	$2e^{-16}$
Temperature Minimum	0.360687	0.003570	$2e^{-16}$

$$AIC = inf.$$

Dispersion parameter for Poisson family taken to be 1

Quasi-Poisson GLM Coefficients

	Estimates	standard error	p.value
Intercepts	3.98771	0.80207	$7.66e^{-07}$
Temperature Maximum	-0.26942	0.02254	$2e^{-16}$
Temperature Minimum	0.36069	0.03034	$2e^{-16}$

$$AIC = N/A$$

Dispersion parameter for Quasi-Poisson family taken to be 72.24358

Negative Binomial GLM Coefficients

	Estimates	standard error	p.value
Intercepts	2.60916	1.10488	0.0182
Temperature Maximum	-0.41794	0.03550	$2e^{-16}$
Temperature Minimum	0.63157	0.0745	$2e^{-16}$

$$AIC = 8236.9$$

Dispersion parameter for Negative Binomial(0.3145) taken to be1
Theta 0.3145
standard error 0.0133

Table 7; Marsabit and Narok

Models

Table 1 represent the r-software output of the analysis for zone 1.The best model relating precipitation, maximum and minimum temperature in zone 1 had an AIC of 12243. The model was

$$In\mu = 7.699 - 0.1617x_1 + 0.113x_2 \qquad 7$$

The mean amount of rainfall obtained as a result of maximum and minimum temperature was

$$\mu = e^{7.699 - 0.1617x_1 + 0.113x_2}$$

The model (equation 7) showed that for every increase of 1^0c in maximum temperature, the mean amount of rainfall decreased by 1.1755mm. At the same time an increase of 1^0c in minimum temperature caused an increase in the amount of rainfall by an amount of 1.1196mm. The net effect was a decrease in the amount of rainfall with an amount of 0.1504mm.The dispersion parameter of 3.744 whose standard error is 0.163 shows that the data from this zone was highly dispersed. The coefficients of the maximum and minimum temperatures were by far less than 0.05. Hence these variables were significant in the variability of rainfall.

In zone 2, the best model describing the relationship between Precipitation, maximum and minimum temperature was extracted from table 2

$$In\mu = 5.0028 - 0.0544x_1 + 0.017128x_2 \quad 8$$

With an AIC of 17162. The mean minimum amount of rainfall experienced in this zone was

$$\mu = e^{5.0028 - 0.0544X_1 + 0.017128X_2}$$

For every 1^0c increase in maximum temperature, there was a decrease of the mean rainfall of 1.056 mm while an increase of 1^0c in minimum temperature caused an increase in rainfall of 1.0174mm. The overall effect was a decrease in the rainfall by 0.039mm. as deduced from the model (equation 8). Data from this zone was overdispersed by a factor of 0.904 with standard error of 0.0302. Low p-values for the data from this zone justify the significance of the intercept β_0, coefficient β_1 and β_2 describing the variation of rainfall in this zone.

The best model to describe a good relationship between precipitation, maximum temperature and minimum temperature in zone 3 was extracted from r-output table 3

$$In\mu = 8.09249 - 0.21704x_1 + 0.126329x_2 \quad 9$$

While the mean amount of rainfall obtained due to the effect of maximum and minimum temperature in this zone was $\mu = e^{8.09249 - 0.126329X_1 + 0.126329X_2}$

With an AIC of 30998. The dispersion parameter of data collected in this zone was 0.843 with a standard error of 0.0203. In this zone, an increase of 1^0c in maximum temperature caused a decrease of 1.2424mm of rainfall while it caused an increase of 1.13465mm of rainfall as deduced from the model (equation 9). The net effect is a decrease in the amount of rainfall experienced in this zone. The zone was becoming warm and drier.

From r-software output table 4, the best model to describe the relationship between precipitation, maximum and minimum temperature in zone 4 was;

$$In = 0.90442 - 0.43887x_1 + 0.84096x_2 \quad 10$$

While the mean amount of rainfall in this zone was $\mu = e^{0.90442 - 0.43887x_1 + 0.84096x_2}$

With an AIC of 6533.6. Using the model (equation 10), An increase of 1^0c in maximum and minimum temperature caused a 1.546mm decrease in the amount of rainfall and an increase of 2.319mm of rainfall respectively. The net effect is an increase in the amount of rainfall in this zone of about 0.773mm. The intercept and the coefficients of maximum and minimum temperatures were very significant going by their very small p-value.

From the r-software output table5, the best model to describe the relationship between the precipitation, maximum temperature and minimum temperature in zone 5 was

$$In\mu = 6.70110 - 0.13273x_1 + 0.05144x_2 \quad 11$$

The mean amount of rainfall experienced in this zone due to combined effect of maximum and minimum temperature was $\mu = e^{6.70110-0.13273x_1+0.05144x_2}$ with an AIC of 10229. The data dispersion from this zone was 0.6043 with a standard error of 0.0242. This model (equation 11) indicates that a rise of 1^0c of maximum temperature, the rainfall decreased by 1.1419mm while for the same while for the same rise in minimum temperature causes an increase of 1.05278mm. The net effect is a decrease in the amount of rainfall by 0.0891mm.

From the r-software output table 6, the best model to describe the relationship between the precipitation, maximum temperature and minimum temperature in zone 6 was

$$In\mu = 13.7969 - 0.38025x_1 + 0.9480x_2 \quad 12$$

The mean amount of rainfall experienced in this zone due to combined effect of maximum and minimum temperature was $\mu = e^{13.7969-0.38025x_1+0.9480x_2}$ with an AIC of 2384.8. The data dispersion from this zone was 0.279 with a standard error of 0.06043. This model (equation 12), indicates that a rise of 1^0c of maximum temperature, the rainfall decreased by 1.46265mm while for the same while for the same rise in minimum temperature causes an increase of 2.5805mm. The net effect is an increase in the amount of rainfall by 1.1178mm.

Exploring the r-software output table 7, he best model to describe the relationship between the precipitation, maximum temperature and minimum temperature in zone 7 was

$$In\mu = 2.60916 - 0.4179x_1 + 0.63157x_2 \quad 13$$

The mean amount of rainfall experienced in this zone due to combined effect of maximum and minimum temperature was $\mu = e^{2.60916-0.4179x_1+0.63157x_2}$ with an AIC of 8236.9. The data dispersion from this zone was 0.3145 with a standard error of 0.0133. This model (equation 13), indicates that a rise of 1^0c of maximum temperature, the rainfall decreased by 1.5188mm while for the same while for the same rise in minimum temperature causes an increase of 1.0mm. The net effect is a decrease in the amount of rainfall by 0.0891mm.

An important point to note is that, different zones experience different amount of rainfall. This is the amount of rainfall on average for the zone. Each zone is made up of one or more area(s) and all the areas might not be receiving the same amount of rainfall, however the rain seasons occur at the same time for all area(s) of the given zone

CONCLUSION AND RECOMMENDATION

In all the above zones, Negative binomial was the best model to describe the relationship between the rainfall, maximum and minimum temperature. This is the model with the lowest AIC value. The AIC criterion is used whenever there are some competing models and the one with the lowest AIC is chosen. Using the GLMs, we were able to predict and forecast the amount of rainfall expected in different zones. It was evident that for every rise of 1^0c the maximum amount of rainfall decreased, while the minimum amount of rainfall increased. However, the amount of decrease was higher than the amount of increase in all zones. The net effect is a decrease in rainfall experienced in all the zones except zone 6 where the amount of increase of minimum rainfall was higher than the amount of decrease of maximum rainfall.

Authorities should use negative binomial model to predict and forecast the amount of rainfall expected in each zone using the temperature trends. Statisticians should do further research on rainfall in zone 6.

References

Aldenderfer, M. a. (1985). Cluster Analysis. *Sage Publication* .

Aldenderfer, M. (1985). Cluster Analysis. *Sage Publication* .

Anderberg, M. (1973). *Cluster Analysis for application.* New Yolk: Academic Press.

Ball, G. a. (1967). A clustering technique for summarizing multivariate data. *Behaviour Science* , 12:153-155.

Bazira, E. (1997). Predictability of East Africa seasonal rainfall using sea surface temperature. *Research Kenya.org* .

Butterfield, R. (2009). Economic impacts of climate change (Kenya, Rwanda and Burundi. *ICPAC* .

Cameron A.C, P. T. (1988). a macroeconomic model of the demand for health care and health Insurance in Australia. *Review of economic studies* , 86-104.

Dolnicar, S. a. (2009). Challenging "Factor-Cluster segmentation". *J Travel* , Res47(1):63-71.

Everitt, B. L. (2001). *Cluster Analysis.* London: Arnold.

Gurmu, S. a. (1996). Excess zero in count models for recreational trips. *journal of business and economic statistics* , 469-477.

Hilbe, J. (1994). *Generalized Linear Models.* London: Chapman and Hall.

Hoef, V. a. (2007). *Quasi-Poissonvs Negative Binomial Regression:How should we model overdispersed data.* Washington: Ecological society of America.

Hoffman, J. (2004). *Generalized Linear Models.* New York: Springer Verlag.

Joeh, H. a. (2005). Generalized Poisson distribution:The property of mixture of Poisson and Comparison with negative Binomial distribution. *Biometrical journa* , 219-229.

Kaufman, L. a. (2005). *Finding groups in data. An introduction to cluster analysis.* New Yolk: Wiley Hoboken.

McCullagah, P. a. (1989). *Generalized Linear Models.* London: Chapman and Hall.

Nelder, J. a. (1972). Generalized Linear Models. *journal of the royal statistical society* , 370-384.

Romesburg, C. (2004). *Cluster analysis for reseachers.* Manisvilles: Lulu press.

Sileshi, G. (2006). Selecting the right statistical model for analysis of insect count data by using information theoretic measures. *Bulletin of Entomological research* , 479-458.

Toniazzo, T. D. (2008). Biasis and variability in the higam model. *climate change research group* .

Tularam, G. a. (2010). Time series of rainfall and temperature interaction in coastal catchments. *Department of mathematics and Statistics* .

Ward, J. J. (1963). hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* , *58*, 236-244.

Winkelmann, R. (1997). Duration Dependence and Dispersion in count data models. *journal of business and economic statistics* , 467-474.

Youne's, M. a. (2012). Poisson Regression and zero inflated Poisson regression:Application to private health Insurance data. *European Actuarial journal* , 8-14.