ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

CLASSIFICATION OF ANEMIC PALESTINIAN CHILDREN USING THE MULTINOMIAL LOGISTIC REGRESSION MODEL.

Mahmoud K. Okasha, and Mohammed A. M. Shehada.

Department of Applied Statistics, Al-Azhar University - Gaza, Palestine.

Manuscript Info

Manuscript History:

Received: 14 December 2015
Final Accepted: 19 January 2016
Published Online: February 2016

Key words:
classification, maximum likelihood estimation,multinomiallogistic regression,oddsratio,ridge egression risk factors ,ROC curve.

****Corresponding Author***

Mahmoud K. Okasha.

Abstract

Various estimation methods and optimization algorithms may be used to estimate the parameters of the multinomial logistic regression model. Most of them require justification of the assumptions about the underlying class densities. Therefore, failure to justify these assumptions may result in a great loss of performance. This paper aims to assess the performance of the main estimation methods and algorithms for building reliable multinomial logistic regression models. Seven estimation methods and algorithms are compared using different assessment techniques to arrive at a reliable multinomial logistic regression model for a given dataset. The result is that the ridge multinomial regression method proves to be the most reliable method with the highest area under the receiver operating characteristic (ROC), or ROC curve, and the lowest error rate for classifying children and identifying significant risk factors on anemia status among all other methods. A detailed description of the results of applying this method to a real dataset from a survey, conducted by the Palestinian Bureau of Statistics to classify children of less than five years of age (2010–2011) according to their anemia status, is illustrated. Ten independent variables from the survey are selected and used to classify children according to their anemia status (normal child, mild anemia, moderate anemia and severe anemia), a reliable multinomial regression model is built, and important risk factors of these anemia statuses are identified.

1. Introduction:-

Multinomial logistic regression (MLR) is a classification method which is useful for classifying subjects based on the values of a set of predictors. The MLR model can be used to predict the probabilities of the categorical response variables, and it belongs to the family of generalized linear models (GLM) and often uses the maximum likelihood estimation method to estimate the model parameters. The usage of this model is less restricted in its assumptions than the ordinary linear regression, which uses the least squares method for the estimation of parameters and in which the response variables must be numeric. Several algorithms are available to estimate the parameters of the MLR models. The results of those are not always the same. They usually produce completely different results for the same datasets. The research problem of this study involves finding the best estimation method and numerical optimization algorithm for estimating the parameters of MLR that produces the most reliable results when applied to a real dataset. The methods are applied to an anemic children dataset in order to identify the most important determinants and risk factors of the anemic status among Palestinian children.

In this paper, we review the estimation methods and numeric algorithms that can be used in the estimation of the parameters of the MLR model. The use of the MLR model in classification and predictive modeling is discussed and assessed using a real dataset. The data of this study was obtained from the Palestinian Family Survey of 2010 that was administered by the Palestinian Central Bureau of Statistics (PCBS) in 2010–2011. The survey includes many variables about the Palestinian families. One of the important variables that we are interested in for this paper is the percentage of hemoglobin (Hb) in the blood in grams per deciliter (g\dL). This variable exists in the different categories of Hb levels in the blood (g\dL). The classes of this response variable (anemia) are the following: normal case, mild anemia, moderate anemia, and severe anemia. Furthermore, in this study, we aim to look at the risk factors that affect the phenomenon through the inclusion of significant independent variables in the model that best classify Palestinian children below five years of age according to their status of anemia.

2. Literature Review:-

A study by Al-Sabbah (2009) investigated the health expenses of some biomarkers of births, deaths, and illnesses of children over a two-year period of time (2004–2006), constructed a model for establishing a health rate through the application of statistical models, and investigated performance using a logistic regression model. Pai (2009) determined the efficiency of mathematical programming classification models, or more specifically, linear programming (LP) methods vis-à-vis statistical approaches such as discriminant analysis and logistic regression, neural networks, and a non-parametric technique (e.g., k-nearest neighbor, or k-NN, for four-group classification problems). Furthermore, the study extended an existing two group LP model (BAL et al., 2006) based on the work of Lam and Moy (1996) and applied it to four-group classification problems. Okasha and Abu-Saada (2014) applied the MLR model on the classification of cases and identification of risk factors on violence against women in the Palestinian society. Rashwan and El-Dereny (2012) used Bayesian and maximum likelihood methods to model binary data for prostate cancer. The empirical results showed that the estimates of coefficients and mean square errors for the Bayesian approach are not different from the maximum likelihood approach, but the percentage of correct classifications when using the Bayesian approach was a little greater than that of the maximum likelihood approach. In this paper, we investigate methods of estimating the MLR model and apply the results to the classification of the anemia status among Palestinian children younger than five years of age (2010–2011).

Chen and Kuo (2001) showed that the MLR model can be used to fit a set of loglinear Poisson regression models, where the responses are the number of cases in each category. These models are parameterized with regression coefficients and an intercept parameter for each covariate class. Hedeker (2003) studied the mixed-effects multinomial logistic regression model. The result presented illustrates the usefulness of the mixed-effects approach for longitudinal categorical data. Kneib et al. (2007) validated semi-parametric multinomial logit models against parametric models. Bull et al. (2007) presented methods to construct confidence intervals for multinomial logistic regression parameters that perform better than standard methods in sparse datasets. Rashid (2008) dealt with logistic regression models and their variations. Simulation study ensured that, when sample size increases, the estimated parameters converge to their true values and follow approximately normal distributions. Wang et al. (2013) compared the accuracy of parametric bootstrap and jackknife methods and evaluated the performance and ability of these estimation methods. Musa (2014) obtained a single unbiased estimator that represents the multiple values sufficiently and efficiently. The maximum likelihood estimators and the probability density function are used to capture the uncertainty. Tektas and Gunay (2008) proposed a Bayesian method to model categorical response data. The study refers to the maximum likelihood and Bayesian methods, and the logit and probit models give very similar results.

Sanchez and Villardon (2013) used penalized ridge estimation of the logistic model parameters in order to avoid the problems produced by separation that makes the estimators undefined. Zahid and Tutz (2009) included ridge regression for the multinomial logit model with symmetric side constraints, which yields parameter estimates that are independent of the reference category. Bentz and Merunka (2000) showed that the neural network can be used as a diagnostic and specification tool for a logit model, which provides interpretable coefficients and significance statistics. The application of the maximum likelihood function to nonlinear least squares was completed by McCullough and Renfro (2000), who explained the numerical sources of the inaccuracy of the nonlinear solver. Different software packages dealt with nonlinear regression and maximum likelihood estimation differently. The maximum likelihood equations are derived from the probability distribution of the dependent variables and solved using the Newton-Raphson method for multinomial logistic regression methods (Czepiel, 2012). An iteratively reweighted least squares method for maximum likelihood estimation of the model parameters were proposed by Nelder and Wedderburn (1972). Seven different algorithms were reviewed for finding the maximum likelihood estimate. Iterative scaling has been shown to apply under weaker conditions than usually assumed which is equivalent to the algorithm of Collins et al. (2002). The best performers in terms of running time are the line search algorithms and Newton-type algorithms, which far outstrip iterative scaling (Minka, 2003).

3. The Multinomial Logistic Regression Model:-

MLR modeling is a classification technique that generalizes a binary logistic regression model to a multiclass problem. An MLR model can be used to predict the probabilities of different possible outcomes of a categorical response variable, given a set of independent variables. MLR is often considered an attractive model for analysis because it does not assume normality, linearity, or homoscedasticity. Indeed, MLR does have assumptions, such as the assumption of independence among the dependent variable choices. Furthermore, MLR also assumes non-perfect separation. If the groups of the response variable are perfectly separated by the predictor(s), then unrealistic coefficients will be estimated and the effect of size will be greatly exaggerated.

There are different parameter estimation methods based on the inferential goals of MLR analysis. The generalized linear modeling technique of MLR can be used to model unordered categorical response variables. This model can be understood as a simple extension of logistic regression that allows each category of an unordered response variable to be compared to an arbitrary reference category providing a number of logit regression models. This procedure outputs a number of logistic regression models that make specific comparisons of the response categories. When there are j categories of the response variable, the model consists of j-1 logit equations which fit simultaneously.

MLR model classifies d-dimensional real-valued input vectors x ∈ Rd into one of the k outcomes $c \in \{0, \ldots, k-1\}$ using $k - 1$ parameter vectors $\beta_0, \ldots, \beta_{k-2} \in R^d$, (Carpenter, 2008). Let the response variable $Y \in \{1, \ldots, k\}$ have k possible values (categories). A generic form of the MLR model is given by

$$p(Y = r \mid x) = \frac{\exp(x^T \beta_r)}{\sum_{s=1}^{k} \exp(x^T \beta_s)} = \frac{\exp(\eta_r)}{\sum_{s=1}^{k} \exp(\eta_s)} \tag{1}$$

where $\beta_r^T = (\beta_{r0}, \cdots, \beta_{rp})$. It is obvious that one has to specify some additional constraints since the parameters $\beta_1^T, \cdots, \beta_k^T$ are not identifiable. An often used side constraint is based on choosing a reference category (RSC). When category k is chosen, one sets $\beta_k^T = (0, \cdots, 0)$ yielding $\eta_k = 0$. Of course, any of the response categories can be chosen as a reference. Fitting a model using a reference category k, the corresponding model is

$$p(Y = r \mid x) = \frac{\exp(x^T \beta_r)}{1 + \sum_{s=1}^{q} \exp(x^T \beta_s)} \qquad for \quad r = 1, \cdots, q \quad . \tag{2}$$

4. Estimation of the Parameters of the Model:-

4.1 The Maximum Likelihood Estimators:-

Given a sequence of n data points $D = (x_i, c_j)_{j<n}$, with $x_j \in R^d$ and $c_j \in \{0, \ldots, k-1\}$, assuming that the stochastic part e of the utility function is distributed with a double exponential distribution, the likelihood of observing actual choices, given input vector X and model parameter vector B, can be expressed by

$$L(Y \mid X, B) = \prod_{j<n} p(c_j \mid x_j, \beta) = \prod_{j=1}^{n} \left(\frac{\exp(\beta_{cj} . x_j)}{1 + \sum_{c'<k-1} \exp(\beta_{c'} . x_j)} \right) \quad . \tag{3}$$

The log likelihood of the data in the model with parameter matrix β is:-

$$\begin{aligned} \log l(\beta) &= \log P(D \mid \beta) = \log \prod_{j<n} p(c_j \mid x_j, \beta) \\ &= \sum_{j<n} \left\{ \beta_{cj} . x_j - \log(1 + \sum_{c'<k-1} \exp(\beta_{c'} . x_j)) \right\} \end{aligned} \quad . \tag{4}$$

The maximum likelihood estimate (MLE) $\hat{\beta}$ is the value of β that maximizes the likelihood of the data D:

$$\hat{\beta}_{MLE} = \arg\max_{\beta} \log l(\beta) = \arg\max_{\beta} \log P(D \mid \beta) \quad . \tag{5}$$

Carpenter (2008) used gradient descent on β as

$$\frac{\partial}{\partial \beta_{c,i}} \log l(\beta) = \sum_{j<n} \frac{\partial}{\partial \beta_{c,i}} \log p(c_j \mid x_j, \beta) = \sum_{j<n} \left\{ x_{j,i} \left[I(c = c_j) - p(c \mid x_j, \beta) \right] \right\} .$$

When $\frac{\partial}{\partial \beta_{c,i}} \log l(\beta) = 0$, we have,

$$\sum_{j<n} \left\{ x_{j,i} \left[1 - p(c \mid x_j, \beta) \right] \right\} = 0$$

$$\sum_{j<n} \left\{ x_{j,i} \left[p(c \mid x_j, \beta) \right] \right\} = 0$$

where $I(c = c_j) = \begin{cases} 1 & \text{if } c_j = c \\ 0 & \text{if } c_j \neq c \end{cases}$.

4.2 Maximum A Posteriori Probability (MAP) Estimates:-

In Bayesian statistics, an MAP estimation is the value of the parameter that maximizes the entire posterior distribution, which is calculated using the likelihood. A MAP estimate is the mode of the posterior distribution (Huberty, 1994; Ripley, 1996). Note that there is no difference between the MLE and the MAP estimate if the prior distribution we were assuming was a constant. The MAP can be used to obtain a point estimate of an unobserved quantity on the basis of empirical data. It is closely related to Fisher's method of MLE, but employs an augmented optimization objective which incorporates a prior distribution over the quantity one wants to estimate. Sparacino et al. (2000) showed that MAP estimation can be seen as a regularization of ML estimation as follows.

$$\hat{\beta}_{MAP} = \arg\max_{\beta} \log P(\beta \mid D) \tag{6}$$

$$= \arg\max_{\beta} \{\log P(D \mid \beta) + \log P(\beta)\} \tag{7}$$

MAP estimates can be computed via a numerical optimization algorithm such as the conjugate gradient method or Newton's method. This usually requires first or second derivatives, which have to be evaluated analytically or numerically (Steele, Patterson, & Redmond, 2003).

4.3 Ridge Estimation with Symmetric Side Constraints and the Iteratively Reweighted Least Square Procedure:-

Ridge regression, one of the oldest penalization methods for linear models, was extended to GLM type models by Nyquist (1991) although a definition of a ridge estimator for the logistic regression model, which is a particular case of GLM, was suggested by Segerstedt (1992). Many alternative penalization/shrinkage methods were proposed for univariate GLMs. Krishnapuram et al. (2005) considered MLR with lasso type estimates. Zhu and Hastie (2004) used ridge type penalization, and Friedman et al. (2010) used the penalties L1 (the lasso), L2 (ridge regression), and a mixture of the two (the elastic net). In multinomial logit models, the identifiability of parameter estimates is typically obtained by side constraints that specify one of the response categories as a reference category. When parameters are penalized, shrinkage of estimates should not depend on the reference category. Ridge regression for the MLR model with symmetric side constraints yields parameter estimates that are independent of the reference category. When the number of predictors is large, as compared to the number of observations, the MLR model suffers from problems such as complete separation, the estimates of parameters are not uniquely defined (some are infinite), and/or the maximum of log likelihood is achieved at 0. The use of regularization methods can help to overcome such problems. Regularization methods based on penalization typically maximize a penalized log likelihood.

Again, let the response variable Y ∈ {1, . . . , k} have k possible values (categories). A generic form of the MLR model is given by Eq. (1) and (2). An alternative side constraint that is more appropriate when defining regularization terms as the symmetric side constraint (SSC) is given by

$$\sum_{s=1}^{k} \beta_s^* = 0 \quad . \tag{8}$$

With β_r^* denoting the corresponding parameters, the MLR model in (1) becomes:

$$p(Y = r \mid x) = \frac{\exp(x^T \beta_r^*)}{\sum_{s=1}^{k} \exp(x^T \beta_s^*)} = \frac{\exp(\eta_r^*)}{\sum_{s=1}^{k} \exp(\eta_s^*)} \qquad \text{for } r = 1, \cdots, q \quad . \tag{9}$$

Although the model's parameters for SSC are different from parameters with a reference category, they consequently have different interpretations. In the case of SSC, i.e., $\sum_{s=1}^{k} \beta_s^* = 0$ the median response can be viewed as the reference category, and it is defined by the geometric mean. Then, from Eq. (9), one can obtain

$$\frac{p(Y = r \mid x)}{GM(x)} = \frac{\exp(\eta_r^*)}{\sqrt[k]{\prod_{s=1}^{k} p(Y = s \mid x)}} .$$

$$\text{and} \qquad \log\left(\frac{p(Y = r \mid x)}{GM(x)}\right) = x^T \beta_r^*$$

Therefore, β_r^* reflects the effects of **x** on the logits when P(Y = r |**x**) is compared to the median response GM(**x**). It should be noted that whatever side constraint is used, the log odds between two response probabilities and the corresponding weights are easily computed by

$$\log\left(\frac{p(Y = r \mid x)}{p(Y = s \mid x)}\right) = x^T \left(\beta_r^* - \beta_s^*\right)$$

which follows from Eq. (7) and (9) for any choice of response categories $r,\ s \in \{1, 2,\ \ldots,\ k\}$.

Let the following $\beta^T = (\beta_1^T, \cdots, \beta_q^T)$ and $\beta^{*T} = (\beta_1^{*T}, \cdots, \beta_q^{*T})$ denote the parameter vectors for the MLR model under the two situations, i.e., reference category side constraint (**β_k=0**) and symmetric side constraint ($\sum_{s=1}^{k} \beta_s^* = 0$).

For a model with an intercept and p covariates, logits are given by

$$\log\left(\frac{\pi_r}{\pi_3}\right) = x^T \beta_r \qquad r = 1, 2$$

$$\log\left(\frac{\pi_r^*}{\pi_3^*}\right) = x^T \beta_r^* \qquad r = 1, 2 \quad .$$

Equating the logits for these two cases, we get **2(p + 1)** equations which can easily be solved to get the result

$$\beta^{*} = \left(T\,\beta^{T}\right)^{T} \quad , \quad \beta = (T^{-1}\beta^{*^{T}})^{T}$$

where T and T^{-1} are the 2 × 2 matrices from above and $\beta^{*^{T}} = (\beta^{*}_{10} \quad \beta^{*}_{20})$ and $\beta^{T} = (\beta_{10} \quad \beta_{20})$ are (p + 1) × 2 matrices composed of parameter vectors with RSC and SSC, respectively.

In general, $\beta^{T}_{.J} = (\beta_{1j},\cdots,\beta_{k-1j})$, $\beta^{*T}_{.J} = (\beta^{*}_{1j},\cdots,\beta^{*}_{k-1j})$, $j = 0,\ldots,p$ collect parameter vectors for single variables with reference category k or SSCs, respectively. Then one obtains the transformation

$$\beta^{*}_{.j} = T\,\beta_{.j} \quad , \quad \beta = T^{-1}\beta^{*} \quad for \quad j = 0,\ldots,p. \tag{10}$$

T^{-1} is a (q × q)-matrix with diagonal entries. The same transformation holds for ML estimates. Estimates of the parameters with SSC can be computed by transforming estimates with reference category side constraint and vice versa. With $\pi^{T}_{i} = (\pi_{i1},\cdots,\pi_{iq})$, $q = k-1$ denoting the (q×1)-vector of probabilities with $\pi_{ir} = P(Y = r|x_i)$**,** the MLR model has the form

$$\pi_i = h(X_i\,\beta) = h(\eta_i) \tag{11}$$

where h is a vector-valued response function, X_i is a (q × [p + 1])-design matrix and $\beta^{T} = (\beta^{T}_{1},\cdots,\beta^{T}_{q})$ is the vector of unknown parameters of length (q × [p + 1]). The MLR model is given by

$$\pi_{ir} = \frac{\exp(x^{T}\beta_r)}{1+\sum_{s=1}^{q}\exp(x^{T}\beta_s)} \qquad for \quad r = 1,\cdots,q$$

which for side constraint with reference category **k** yields

$$\log\left(\frac{p(Y = r\,|\,x)}{p(Y = k\,|\,x}\right) = x^{T}\beta_r \quad for \quad r = 1,\cdots,q \quad . \tag{12}$$

The log-odds compares $\pi_r = P(Y = r|x)$ to the probability $\pi_k = P(Y = k|x)$ of the reference category k. The q logits

$$\log\left(\frac{p(Y = 1\,|\,x)}{p(Y = k\,|\,x}\right),\ldots\ldots\ldots\ldots,\log\left(\frac{p(Y = q\,|\,x)}{p(Y = k\,|\,x}\right)$$

given by (12) determine the response probabilities $p(Y = 1\,|\,x),\ldots\ldots\ldots p(Y = k\,|\,x)$ uniquely since the constraint $\sum_{r=1}^{k} p(Y = r|x) = 1$ holds. Therefore, only q = k−1 response categories and parameter vectors have to be specified. The representation of the MLR model in (11) and the corresponding response function h depend distinctly on the choice of the reference category. Since the parameters β^{*} with SSC may be obtained by reparameterization of the parameters β with RSC, the numerical computation of ML estimates of β^{*} makes use of a transformation of the design matrix X. The transformed design matrix for SSC has the form

$$X^{*} = XT^{*}$$

where X is the total design matrix of order q(n × (p + 1)) with X_i, and it is a q × q(p + 1) matrix (composed of $\mathbf{x}_i$). T^{*} is a q((p + 1) ×(p + 1)) matrix composed of the elements of T^{-1} in order to satisfy 0, $\beta^{*}_{.j} = T\,\beta_{.j}$, $\beta = T^{-1}\beta^{*}$ for $j = 0,\ldots,p$.

In matrix notation, Zahid and Tutz (2009) showed that

$$s(\beta^*) = X^{*T} D(\beta^*) \sum^{-1}(\beta^*) \left[y - h(\eta^*) \right],$$

$$y^T = (y_1^T, \cdots, y_n^T), \qquad h(\eta^*)^T = (h(\eta_1^*)^T, \cdots h(\eta_n^*)^T),$$

$$\sum(\beta^*) = diag(\sum_i^{-1}(\beta^*)), \quad W(\beta^*) = diag(\sum_i^{-1}(\beta^*)) \text{ and } D(\beta^*) = diag(D_i(\beta^*))$$

Then, the Fisher scoring iteration, which can also be viewed as an iteratively reweighted least square procedure, has the form

$$\hat{\beta}^{*(k+1)} = \hat{\beta}^{*(k)} + (X^{*T} W(\hat{\beta}^{*(k)}) X^*)^{-1} s(\hat{\beta}^{*(k)}).$$

4.4 Neural Networks and the Multinomial Logit:-

Neural network modeling has recently received increasing attention and has been applied to an array of marketing problems such as market response or segmentation. A feed forward neural network with softmax output units and shared weights can be viewed as a generalization of the MLR model. The main difference between the two approaches lies in the ability of neural networks to model non-linear preferences with few, if any, a priori assumptions about the nature of the underlying utility function, while the MLR model can suffer from a specification bias (Bentz & Merunka, 2000). Being complementary, these approaches are combined into a single framework. A neural network is used as a diagnostic and specification tool for the logit model, which will provide interpretable coefficients and significance statistics.

5. Numerical Algorithms for Estimating the Model's Parameters:-

Logistic regression is workhorse of statistics and is closely related to methods used in machine learning, including the perceptron and the support vector machine. Various algorithms for computing an ML and a MAP estimate are available (Minka, 2003). The derivation of iterative scaling applies more generally than the conventional one. Another useful algorithm is the modified iterative scaling algorithm of Collins et al. (2002). The majority of the algorithms operate in the primal space, but they can also work in dual space. Different numerical algorithms can be used for finding the maximum likelihood method to estimate the model's coefficient. The Newton-Raphson algorithm is the most popular for estimation. In general, statistics software such as SPSS, R, Minitab, SAS, and Systat use different algorithms in the estimation process of an MLR model's coefficients. In this paper, we used different possible algorithms in R software in the analysis of a real dataset on the anemia status among Palestinian children and compared the results. We mainly used the following algorithms which are defined in R software as functions, and each one is available within a specific library:

1. glmnet (in glmnet library): Fits a generalized linear model via penalized maximum likelihood. The regularization path is computed for the lasso or elastic net penalty at a grid of values for the regularization parameter lambda (Tibshirani et al., 2012).
2. mlogit (in mlogit library): Estimates the multinomial logit model by maximum likelihood, with alternative-specific and/or individual specific variables (Train, 2004).
3. multinom (in nnet library): Fits multinomial log-linear models via neural networks. It fits single hidden layer neural network, possibly with skip-layer connections. (Venables & Ripley, 2002).
4. polr (in MASS library): This model is what Agresti (2002) calls a cumulative link model. It fits a logistic or probit regression model to an ordered factor response. The default logistic case is proportional odds logistic regression, after which the function is named.
5. vglm (in VGAM library): A vector generalized linear model (VGLM) is loosely defined as a statistical model that is a function of M linear predictors (Yee, 2008, 2010).
6. RidgeMultinomialRegression (in NominalLogisticBiplot library): It calculates an object with the fitted multinomial logistic regression for a nominal variable. It compares with the null model, so that we will be able to compare which model fits the variable better (Heinze & Schemper, 2002).
7. regmlogit (in reglogit library): It is a wrapper around the Gibbs sampler inside reglogit, invoking C-1 linked chains for C classes, extending the polychotomous regression scheme outlined by Holmes and Held (2006).

6. The Data:-

In this section, we present an analysis of a real dataset as an application of various MLR estimation methods that have been discussed earlier and conduct a comparison between those methods. We start with describing the variables in the data. Then, we illustrate some of important features in the data, and we build the statistical model and estimate its parameters using all possible methods. Further, we test the significance of the parameters and identify the best subset of independent variables for the MLR model. The main target is to conduct a comparison between different methods of analysis of categorical data using MLR models and to identify the most important risk factors for the anemia status among Palestinian children.

The response variable is the anemia status among Palestinian children which has four categories. The sample is composed of 2,406 children of which 79.3% are normal children, 12.7% suffer from mild anemia, 7.8% suffer from moderate anemia, and only 0.2% suffer from severe anemia. Ten independent variables have been selected to be included in the MLR model among many variables included in the survey as possible risk factors. Table 1 describes those variables. All the independent variables are categorical except two numeric variables which are cage and body mass index (BMI) variables (PCBS, 2010).

Table 1:
Description of the Independent Variables:-

Variable Name	Description	Value Label
BF1	Child has been breastfed	1. Yes 2. No 3. Do not know
IM18	Child has been given Vitamin A dose within last 6 months	1. Yes 2. No 3. Do not know
PIM2	Child has received an iron syrup constantly after 6 months and for 1 years	1. Yes 2. No 3. Do not know
cage	Age (Months)	1–59 months
HH6	Area	1. Urban 2. Rural 3. Camps
HL4	Sex	1. Male 2. Female
melevel	Mother's level of education	1. None 2. Primary 3. Secondary+
region	Region	1. West Bank 2. Gaza
windex5	Wealth index quantiles	1. Poorest 2. Second 3. Middle 4. Fourth 5. Richest
BMI	Body Mass Index	6–42

7. Comparison Between Different Algorithms Using the Anemia Status Among Palestinian Children's Data:-

In this section, we fit the MLR model for a real dataset on the anemia status among Palestinian children under 5 years of age using seven different algorithms. This is to compare results of different algorithms used in R, and other statistical software packages, in terms of their correct classification rates and ROC curves. Later, we give the full analysis of the data using the algorithm that produces the best results with the highest correct classification rate.

Table 2 displays the correct classification rates for the seven algorithms. The tables shows that the highest correct classification rate is 92.44% which has been produced by the ridge regression method while all other methods produced approximately similar correct classification rates of about 79.3%.

Table 2: Correct Classification Rates for All Algorithms

Method	Correct Classification Rates
mlogit : iterative maximum likelihood	79.26%
glmnet : penalized log likelihood	79.44%
regmlogit : maximum a posteriori probability (MAP)	79.44%
multinom : neural networks	79.26%
polr : maximum likelihood	79.30%
vglm : iteratively reweighted least squares	79.26%
RidgeMultinomialRegression : ridge regression method	92.44%

The ROC curves for all algorithms are exhibited in Figure 1, and the area under the curve has been computed for different classes of the dependent variable for the seven methods. Table 3 shows the area under the ROC curve for different algorithms. We can easily observe that the area under the ROC curve for the ridge regression method is the highest for all categories of anemia status particularly for the normal children and for severe anemia cases.

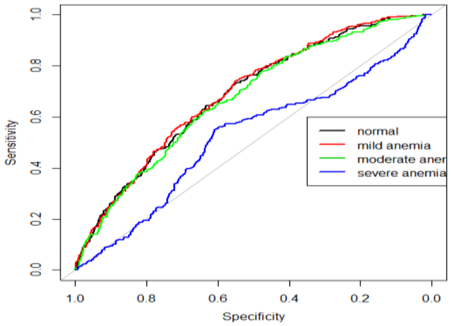
Table 3: ROC Curve Analysis: Area Under the ROC Curves for All Algorithms

Method	**Normal**	**Mild Anemia**	**Moderate Anemia**	**Severe Anemia**
mlogit : iterative maximum likelihood	0.685	0.690	0.672	0.523
glmnet : penalized log likelihood	0.685	0.690	0.672	0.521
regmlogit : maximum a posteriori probability (MAP)	0.697	0.707	0.658	0.559
multinom : neural networks	0.685	0.690	0.672	0.521
polr : maximum likelihood	0.684	0.684	0.684	0.684
vglm : iteratively reweighted least squares	0.685	0.690	0.672	0.523
RidgeMultinomialRegression : ridge regression method	0.972	0.976	0.634	0.742

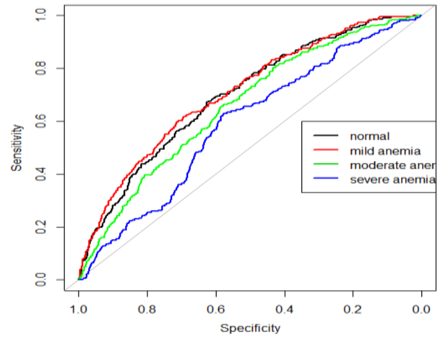
Figure 1 clearly shows that the ROC curve of the ridge regression method is always higher than those of the other algorithms. This indicates that the ridge regression method produces classification results for all categories with much higher accuracy than all other methods. Thus, in the next section, we give a detailed discussion of the results of the application of the ridge regression method on the dataset for the anemia statuses among Palestinian children under 5 years of age.

8. Application of the Ridge Multinomial Regression on Classifying Anemic Palestinian Children:-

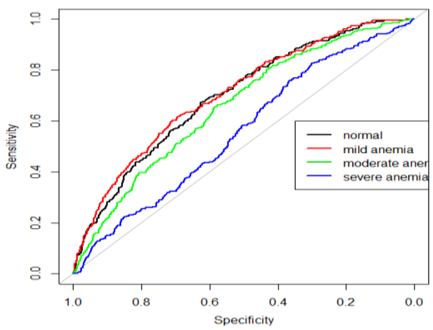
The ridge multinomial regression model fitting method has been described in the previous section in detail. To apply this method, the ridge multinomial regression function in R software, which calculates an object with the fitted multinomial logistic regression for a nominal variable, was used. It was compared with the null model, so that we could determine which model fit the variables better. In this section, we built a MLR model for the anemic Palestinian children's dataset using the ridge multinomial regression method. The independent variables that were found to have a significant effect on the anemic status of the Palestinian children were BF1, region, cage, melevel, HL4, and windex5. Table 4 represents the significant risk factors and estimates of the coefficients of the three equations of the MLR model as well as their indicators of the significance of each estimate.



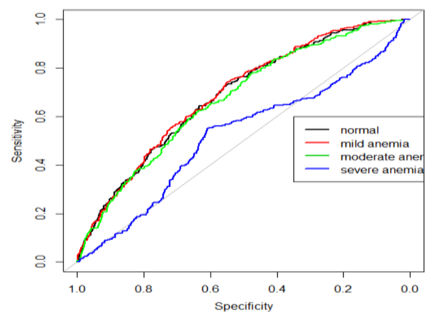
(a) mlogit



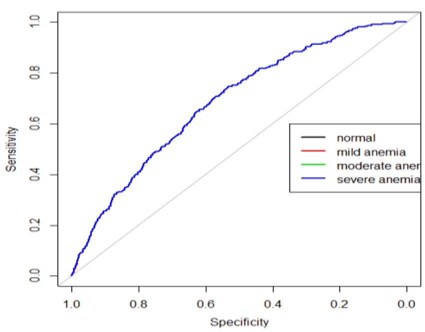
(b) glmnet



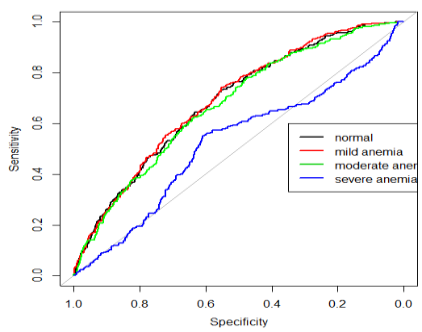
(c) regmlogit



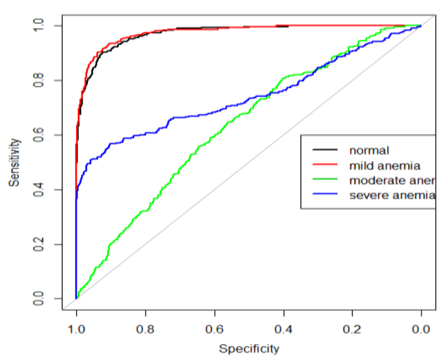
(d) multinom



(e) polr



(f) vglm



(g) RidgeMultinomialRegression

Figure 1. The ROC curve for all algorithms.

Table 4: The estimates of the Parameters of the Three Equations of the Final Ridge Multinomial Regression Model (Reference Category: Normal).

First Model: Mild Anemia

Intercept	BF1	region	cage	melevel	HL4	windex5
-4.36106	-5.91264	2.036081	-4.12358	-0.70928	-1.00526	-3.925590

Second Model: Moderate Anemia

Intercept	BF1	region	cage	Melevel	HL4	windex5
-5.034404	-6.182310	-1.851448	2.299793	1.2857029	0.2938031	-2.968174

Third Model: Severe Anemia

Intercept	BF1	region	cage	melevel	HL4	windex5
-7.629087	-1.859191	1.434944	2.398387	-0.92157	0.2053777	-1.092213

The fitting information shown in Table 5 for the assessment of the model's fit shows that the -2 log likelihood, which equals 3165.393 for the null model and 951.7937 for the final model, and the Akaike information criterion (AIC) (993.7937 for the final model), as well as the high chi-square test value (χ^2 (18) = 2213.6 with $p < .001$), indicates a highly significant and reliable model. This indicates that our final model is a reliable one.

Table 5**:** Model Fitting Information of the Final Ridge Multinomial Regression Model

Model	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC	BIC	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only			3165.393			
Final	993.7937	1115.294	951.7937	2213.6	18	.000

The results also indicate that the proportion of variance in the response variable explained by the predictors is 60.15%, according to both the Cox and Snell R^2 and Nagelkerke R^2 values. The McFadden R^2 value equals 69.93%. The remaining proportion of total variations in the dependent variable may be explained by other variables which are not available in the dataset. Correct classification rates from the final model are shown in Table 6. The overall rate of correct classification of the model as indicated in the table is 92.44%.

Table 6: Classification Table for the Final Ridge Multinomial Regression Model:-

Observed	Predicted				
	Normal	Mild Anemia	Moderate Anemia	Severe Anemia	Percent Correct
Normal	1,855	31	22	0	97.2%
Mild Anemia	69	226	10	0	74.1%
Moderate Anemia	38	12	138	0	73.4%
Severe Anemia	0	0	0	5	100%
Overall Percentage	81.5%	11.2%	7.1%	0.2%	92.44%

The above classification table shows how well our model correctly classifies cases. A perfect model would show only values on the diagonal and, thus, correctly classify all cases. Adding across the rows represents the number of cases in each category in the actual data and adding down the columns represents the number of cases in each category as classified by the final model. The key piece of information is the overall percentage in the lower right corner which shows that our model is 92.44% accurate.

The areas under the ROC curve as indicated in Table 3 above are 0.9720366 for normal children, 0.9756006 for mild anemic cases, 0.6341221 for moderate anemic cases, and 0.7418944 for severe anemic children. Moreover, when a multi-class area under the curve is used, the area becomes 0.9238. The ROC curves for the final ridge multinomial regression model are exhibited in Figure 1 above which shows how well the final model using the ridge multinomial regression method compared to other methods.

Examination of the odds ratios (OR) for binary independent variables as risk factors can easily be interpreted. However, since many of the available independent variables are not binary, but either numeric or multi-class categorical variables, they are not easily interpretable. Thus, the ORs of the risk factors of region and HL4, as the only two significant binary independent variables, can easily be interpreted. The interpretation of the OR for the variable region is that, being in the Gaza Strip (X = 2), the child has 7.66 times the chance of suffering from mild anemia and 4.20 times the chance of suffering from severe anemia more than a child in the West Bank (X = 1). However, a child living in the West Bank has 6.37 times greater chance of suffering from moderate anemia than a child living in the Gaza Strip. The ORs of variable sex (HL4) are not so high as to indicate that a female child would have a greater chance of suffering from mild anemia. This means that, the larger the value of the OR, the higher the chance is of suffering from one of the classes of anemia, and the opposite is true for values shown inside the brackets in Table 7 (Note that the definitions and possible values of independent variables are available in Table 1. The majority of ORs in the table seem to be very high, indicating that the risk factors are highly important ones.

Table 7: Odds Ratios of the Three Equations of the Final Ridge Multinomial Regression Model (Reference Category: Normal):-

First Model: Mild Anemia

Intercept	BF1	region	cage	melevel	HL4	windex5
(78.34)	(369.68)	7.66	(61.78)	(2.03)	(2.73)	(50.68)

Second Model: Moderate Anemia

Intercept	BF1	region	cage	melevel	HL4	windex5
(153.61)	(484.11)	(6.37)	9.97	3.62	1.34	(19.46)

Third Model: Severe Anemia

Intercept	BF1	region	cage	melevel	HL4	windex5
(2057.17)	(6.42)	4.20	11.01	(2.51)	1.23	(2.98)

9. Conclusion:-

In the above discussion, we compared different methods and algorithms that may be used for analyzing categorical data using MLR models. Based on this discussion, the following conclusions may be drawn:

1. It is evident that the ridge estimation method of the MLR model can more accurately classify cases than all other methods in the study. This result can be seen clearly from the ROC analysis and the comparison between correct classification rates of different algorithms.
2. The ridge estimation method showed that it is more efficient to apply than all other methods particularly for classifying cases when some categories contain a small number of observations and when some independent variables are cross-correlated. In fact, it was the only method in the study that correctly classified the severe anemia cases in the data.
3. The important risk factors for the anemia among Palestinian children under 5 years of age that have been found to have a significant effect on their anemic status are the following: history of being breastfed (BF1), place of residence (region), age in months (cage), mother's level of education (melevel), sex of the child (HL4), and wealth index quantiles (windex5).
4. Using the MLR model in the classification and discrimination of problems in the medical and health sciences is important because of its ability to provide a good classification and prediction technique and to identify the most important risk factors of a specific disease.
5. For alleviating the problem of anemic children in Palestine, special care should be given to mothers particularly in the Gaza Strip as they are the primary source of healthcare for their children.

References:-

1. Agresti, A. (2002). Categorical data. Second edition. Wiley.
2. Al-Sabbah, Shrook A. S. (2009). Building a modified model of logistic regression for a life of premature infants in Karbala governorate. (Ph.D. Thesis in Statistics). College of Administration and Economics, University of Baghdad, Iraq.
3. Bal, H., Örkcu, H. H., & Çelebioglu, S. (2006). An alternative model to Fisher and linear programming approaches in two-group classification problems: Minimizing deviations from the group median. G.U. Journal of Science, 19(1), 49–55.
4. Bentz, Y., & Merunka, D. (2000). Neural networks and the multinomial logit for brand choice modelling: a hybrid approach. Journal of Forecasting, 19, 177–200.
5. Bull, S. B., Lewinger, J. P., & Lee, S. F. (2007). Confidence intervals for multinomial logistic regression in sparse data. Statistics in Medicine, 26, 903–918.
6. Carpenter, B. (2008). Lazy sparse stochastic gradient descent for regularized multinomial logistic regression. Carpenter, B. (2008). Alias-i, Inc., Tech. Rep, 1–20. Retrieved from https://lingpipe.files.wordpress.com/2008/04/lazysgdregression.pdf
7. Chen, Z., & Kuo, L. (2001). A Note on the estimation of the multinomial logit model with random effects. The American Statistician, 55(2), 89–95.
8. Collins, M., Schapire, R. E., & Singer, Y. (2002), Logistic regression, AdaBoost and Bregman distances. Machine Learning, 48, 253–285.
9. Czepiel, S. A. (2002). Maximum likelihood estimation of logistic regression models: theory and implementation. Available at czep. net/stat/mlelr. pdf. Retrieved from: http://czep.net/stat/mlelr.pdf
10. Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent Journal of Statistical Software, 33(1), 1–22.
11. Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. Statistics in Medicine, 22, 1433–1446.
12. Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. Statistics in Medicine, 21, 2109–2419.
13. Holmes, C., & Held, K. (2006). Bayesian auxiliary variable models for binary and multinomial regression. Bayesian Analysis, 1(1), 145–168.
14. Huberty, C. J. (1994). Applied discriminant analysis (p. XIII). New York: Wiley.
15. Kneib, T., Baumgartner, B., & Steiner, W. J. (2007). Semiparametric multinomial logit models for analysing consumer choice behaviour. AStA Advances in Statistical Analysis, 91(3), 225–244.
16. Krishnapuram, B., Carin, L., Figueiredo, M. A., & Hartemink, A. J. (2005). Sparse multinomial logistic regression: Fast algorithms and generalization bounds. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 27(6), 957–968.
17. Lam, K. F., & Moy, J. W. (1996). Improved linear programming formulations for the multi-group discriminant problem. Journal of Operational Research Society, 47, 1526– 1529.
18. McCullough, B. D., & Renfro, C. G. (2000). Some numerical aspects of nonlinear estimation. Journal of Economic and Social Measurement, 26, 63-77.
19. Minka, T. P. (2003). Algorithms for maximum-likelihood logistic regression. Retrieved from http://research.microsoft.com/en-us/um/people/minka/papers/logreg .
20. Musa, A. B. (2014). Logistic Regression Classification for Uncertain Data. Research Journal of Mathematical and Statistical Sciences, 2(2), 1-6.
21. Nelder, J., & Wedderburn, R. (1972). Generalized linear models. Journal of the Royal Statistical Society, Series A (General), Blackwell Publishing, 135(3), 370–384.
22. Nyquist, H. (1991). Restricted estimation of generalized linear models. Journal of Applied Statistics, 40, 133–141.
23. Okasha, M. K., & Abu-Saada, Abed H. K. (2014). Modeling violence against women in Palestinian society. American International Journal of Contemporary Research. Center for Promoting Ideas, USA. 4(1), 209–220.
24. Pai, D. R. (2009). Determining the efficacy of mathematical programming approaches for multi-group classification. (Ph.D. dissertation). The Graduate School-Newark Rutgers, The State University of New Jersey.
25. Palestinian Central Bureau of Statistics. (2010). Palestinian family survey. Ramallah, Palestine.

26. Rashid, M. (2008). Inference on logistic regression models (Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the Ph.D. requirement). Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=bgsu1214165101
27. Rashwan, N. I., & El-Dereny M. (2012). The comparison between results of application Bayesian and maximum likelihood approaches on logistic regression model for prostate cancer data. Applied Mathematical Sciences, Dept. of Statistics & Mathematics Faculty of Comm., Tanta University, 6(23), 1143–1158.
28. Ripley, B. D. (1996). Pattern recognition and neural networks. Cambridge: Cambridge University Press.
29. Sanchez, J. C. H., & Villardon, J. L. V. (2013) Logistic biplot for nominal data. Spanish Statistical Office and University of Salamanca. Retrieved from http://xxx.tau.ac.il/pdf/1309.5486v1.pdf
30. Segerstedt, B. (1992). On ordinary ridge regression in generalized linear models. Communications in Statistics-Theory and Methods, 21, 2227–2246.
31. Sparacino, G., Tombolato, C., & Cobelli, C. (2000). Maximum-likelihood versus maximum a posteriori parameter estimation of physiological system models: the c-peptide impulse response case study. Biomedical Engineering, IEEE Transactions on 47(6), 801–811.
32. Steele, B. M., Patterson, D. A., & Redmond, R. L. (2003). Toward estimation of map accuracy without a probability test sample. Ecological and Environmental Statistics, 10, 333–356.
33. Tektas, D., & Gunay, S. (2008). A Bayesian approach to parameter estimation in binary logit and probit models. Hacettepe Journal of Mathematics and Statistics, 37(2), 167–176.
34. Tibshirani, R., Bien, J., Friedman, J., Hastie, T., Simon, N., Taylor, J., & Tibshirani, R. (2012). Strong rules for discarding predictors in lasso-type problems. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 74(2), 245–266. Stanford Statistics Technical Report. Retrieved from http://www-stat.stanford.edu/~tibs/ftp/strong.pdf
35. Train, K. (2004). Discrete choice modelling with simulations. Cambridge: Cambridge University Press.
36. Venables, W. N., & Ripley, B. D. (2002). Modern applied statistics with S-PLUS (4th ed.). Springer Science & Business Media.
37. Wang, S., Jiao, H., & Xiang, Y. (2013). A comparison of different reduction methods for bias of maximum likelihood estimator of student ability based on graded response and generalized partial credit models. National Council on Measurement in Education, April 28–30, San Francisco, CA.
38. Yee, T. W. (2008). The VGAM package. R News, 8, 28–39.
39. Yee, T. J. (2010). The VGAM package for categorical data analysis. Journal of Statistical Software, 32(10). Retrieved from http://www.jstatsoft.ord/v32/i10/
40. Zahid, F. M., & Tutz, G. (2009). Ridge estimation for multinomial logit models with symmetric side constraints. Department of Statistics, University of Munich, Technical Report Number 067. Retrieved from http://www.stat.uni-muenchen.de
41. Zhu, J., & Hastie, T. (2004). Classification of gene micro arrays by penalized logistic regression. Biostatistics, 5, 427–443.