Journal of Al Azhar University-Gaza *(Natural Sciences)*, **2015, 17** : **103-128**

Using Multinomial Logistic Regression and Artificial Neural Networks in the Classification of Living Standard in Palestine

Mo'omen El- Hanjouri

Faculty of Economics and Administrative Sciences;
Al – Azhar University-Gaza
Moamin2000@hotmail.com

Received 9/7/2015 *Accepted* 13/10/2015

ملخص:

يشهد العالم الحالي العديد من القضايا والاضطرابات السياسية والاجتماعية والاقتصادية، حيث يعتبر تحسين مستوى المعيشة للأسر وتقليص التفاوت في دخلها والقضاء على الفقر من أبرز الاهتمامات التي تسعى إليها دول العالم لاسيما فلسطين.

في هذا البحث تم مناقشة أسلوبين إحصائيين للتصنيف وهما الانحدار اللوجيستي المتعدد (MLR)والشبكات العصبية الاصطناعية (ANN) لتصنيف بيانات مسح الإنفاق و الاستهلاك (2011) في فلسطين، و تمت المقارنة بين الأسلوبين بناءً على دقة التصنيف، واستخدام المساحة تحت منحنى ROC ، و معياري التقييم الـ AIC و الـ BIC .

يحتوي البحث على جزأين: الجزء النظري و الجزء التطبيقي الذي استخدم فيه بيانات مسح الإنفاق و الاستهلاك (2011) . حيث تألفت البيانات من (4317) أسرة في الضفة الغربية وقطاع غزة، واشتملت البيانات على (12) متغيراً، بحيث كان المتغير التابع يمثل المستوى المعيشي للأسرة و هو متغير رتبي يتكون من ثلاث فئات (مستوى معيشي مرتفع ، متوسط ، منخفض)، بالإضافة إلى 11 متغيراً مستقلاً.

يهدف البحث إلى اختيار أفضل نموذج إحصائي لبيانات المستوى المعيشي للأسر الفلسطينية، حيث تم اختبار نموذجين لكل أسلوب من أسلوبي التصنيف باستخدام الاختبارات الإحصائية.

أظهرت النتائج أن أسلوب الشبكات العصبية الاصطناعية (ANN)كان أفضل من أسلوب الانحدار اللوجيستي المتعدد (MLR)، حيث أشارت النتائج أن نسبة التصنيف الصحيحة بلغت 90.1% لأسلوب الشبكات العصبية الاصطناعية مقارنة بـ 77.7% لأسلوب الانحدار اللوجيستي المتعدد، و بلغت المساحة تحت منحنى ROC باستخدام الشبكات العصبية الاصطناعية 97.2%

في حين بلغت تلك المساحة 89.9% باستخدام الانحدار اللوجيستي المتعدد، حيث تمت المقارنة باستخدام معياري التقييم الـ AIC و الـ BIC .

Abstract:

The world today is encountering many global issues: political, social and economic. Whereas, improving the standard of living and reducing income inequality; especially, in Palestine are the basic concerns to be solved nowadays.

In this study, Multinomial Logistic Regression (MLR) and Artificial Neural Network (ANN) were discussed on Expenditure and Consumption Survey Data in 2011. The two methods were compared according to accuracy assessment, ROC curve, AIC and BIC assessment criterion.

The study contained two parts, theoretical part and application part in which data was used from Expenditure and Consumption Survey (2011).Data consisted of (4317) households from the West Bank and Gaza Strip. Data contained (12) variables, where the dependent variable was the standard of living, which was ordinal variable and contained of three categories (High, Middle and low standard of living) and (11) other independent variables.

The study aims to choose the best statistical model for Palestinian standard of living data. Two models for each method were tested by group of statistical tests to define the best model. The results showed that Artificial Neural Network method was better than Multinomial Logistic Regression method, where Artificial Neural Network ratio reached 90.1% compared to Multinomial Logistic Regression ratio, which reached 77.7%.Area under ROC curve for Artificial Neural Network analysis reached 97.2%,while (AUC) area under curve analysis model reached 89.9% for Multinomial Logistic Regression. While comparing was done using AIC and BIC assessment criterion.

Key words: Classification, Multinomial Logistic Regression, Artificial Neural Network, ROC curve.

1. Introduction

Living standard severely affects the progress of any country of the world, and the drop in this standard is deemed as a warning to officials and planners to reconsider their future plans and find appropriate solutions. Thus, some countries care of providing the best living standard for their people by providing whatever meets their needs of health and education services.

The economic development of any country in the world is linked to the progress of education and health and the high level of their services. This is what is known as the standard of living, and the standard of living is based on the concept that human beings are the real wealth of nations, as well as expanding human choices. Then the concept of the standard of living expanded than what it was; it has become how much the person obtains goods and services, and whenever the individual is able to get more of those goods and services, the standard of living raises, well-being increases then development is realized. The standard of living has many indicators, including: education, health and income level, (Human rights,2010).

The objective of this study is to determine the best method among two classification techniques (Artificial Neural Network and Multinomial Logistic Regression) to be used for classifying Palestinian households by their standard of living (high , middle ,or low standard of living). This has been done by using some assessment techniques (such as classification table, ROC curve , and AIC and BIC assessment criterion) bearing in mind that the majority of the independent variables are numeric and continuous. Moreover, we are also interested to determine the most influential variables that can classify and predict the standard of living for Palestinian households .

Several researchers considered the problem of comparing classification techniques, Nassar (2013) compared three methods of statistical classification (ANN, LR and LDA) on Palestine market data for classification and prediction of financial failure of the firms using three evaluations statistics (Bootstrapping, Cross Validation and ROC Curve) . The results showed the superiority of ANN to other classification methods.

Yarmohammadi et al., (2004) designed an algorithmic model based on the logistic regression model (LRM) and a non-algorithmic model based on the ANN. The ability of these models was compared together in clinical application to differentiate malignant from being breast tumor in a study group of (161) patients' records. Each patient’s record consisted of 6 subjective features extracted from MRI appearance. Results of the study showed that ANN and LRM prove the relationship between extracted morphological features and biopsy results. Using statistically significant variables reduced LRM outperformed ANN with remarkable specificity while high sensitivity is achieved.

Lee (2010) proposed a method for multi-way classification problems using ensembles of MLR models. The multinomial logit model showed that it can be applied to each mutually exclusive subset of the feature space without variable selection. By combining multiple models the proposed method could handle a huge database without a parametric constraint needed for analyzing high-dimensional data, and the random partition could improve the prediction accuracy by reducing the correlation among base classifiers. The proposed model showed a substantial improvement in overall prediction accuracy over a multinomial logit model.

The rest of the paper is organized as follow, Sections 2 and 3 introduce the theoretical framework of the classification methods, MLR and ANN, respectively.

Section 4 reviews the methods of validation and evaluation of classification model. The description of the standard of living data is described in Section 5. The MLR and ANNs classification models are obtained in Section 6 and 7, respectively. Section 8 presents the comparison between the evaluation of MLR and ANN models and Section 9 draws the conclusions.

2. Multinomial Logistic Regression

Logistic regression analysis (or simply logistic regression) is part of a category of generalized linear models. It is a type of multivariate regression that has a predictive model that can be used when the target variable is a categorical variable. The technique aims to modeling the relationship between a set of independent variables and the probability that a case is a member of one of the categories of the dependent variables (Hosmer and Lemeshow, 2000) .

Logistic regression has many uses, It is used to predict a dependent variable on the basis of continuous and/or categorical independents, to determine the percentage of variance in the dependent variable explained by the independents; to rank the relative importance of independents, to assess interaction effects, and to understand the impact of covariate control variable. (Kutner et al., 2004) .

logistic regression makes no assumption about the distribution of the independent variables. They do not have to be normally distributed, linearly related or of equal variance within each group. The relationship between the predictor and response variables is not a linear function, instead, the logistic regression function use the logit transformation of the probability of for a positive response,π. The logistic regression model can be written as:

$$\pi = \frac{e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}}{1+e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}}. \quad (1)$$

Model (1) can be written in terms of the log of the odds and so-called the logit model as given in (2)

$$\log\left(\frac{\pi}{1-\pi}\right) = logit(\pi) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k. \quad (2)$$

where $x_1,\ldots,x_k$ are continuous measurements corresponding to covariates and/or dummy variables corresponding to factor levels and $\beta_1,\ldots,\beta_k$ are the parameters. The range of values that the left-hand of (2) side can potentially take is between $-\infty$ and ∞, which is the same range as that of the right-hand side. Now we have a linear model on the logit scale. This is the most common form of the logistic regression model. An alternative and equivalent way of writing the logistic regression model in (1) is in terms of the odds (Agresti,2002).

$$\frac{\pi}{1-\pi} = e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}. \quad (3)$$

This model is very widely used for analyzing data involving binary or binomial responses and several explanatory variables. Estimates for the parameters and response probabilities are typically obtained by the

method of maximum likelihood. These estimates will be computed and returned by the software package.
The advantages of logistic regression come at a cost: it requires much more data to achieve stable, meaningful results (Nassar, 2013).

3. Artificial Neural Networks

Neural networks, also called artificial neural networks (ANNs), are models for classification and pattern recognition capabilities. ANNs were designed to model the functioning of human brain, where neurons are inter-connected and learn from experience. We use ANNs for our research for two reasons. First, ANNs have the ability to decipher and solve nonlinear relationship problems. Second, research over last two decades indicates that neural networks may achieve better classification and prediction compared to standard statistical methods (Sharda, 1994). This has been corroborated by a number of successful ANN applications, such asviolence against women prediction (Abu Saada, 2013),high hypertension diseases prediction (Abu Samra, 2012), bankruptcy prediction (Odom & Sharda, 1990), bank failure prediction (Tam & Kiang 1990), and market segmentation (Fish et al., 1995). Neural networks structure captures complex relationships between the predictor variables and the response variable through a layer of neurons. Some have one layer (single layer neural networks) and some have more (multilayer neural networks). While various neural networks architectures have been reviewed in the literature, the most successful applications in classification and prediction have been multilayer feed forward networks. The layer where input patterns are applied is the input layer. The layer from which an output response is desired is the output layer. In the case of a binary outcome, the network has only one output node. Layers between the input and output layers are known as hidden or transfer layers, because their outputs are not readily observable. ANNs are flexible models useful for discrimination and classification, and they are implanted by a computerized "black-box" trained by a training data set (Dodge, 2003).
There are two functions governing the behavior of a unit in a particular layer, which normally are the same for all units within the whole ANN, i.e. the input function, and the output/activation function.

Input into a node is a weighted sum of outputs from nodes connected to it. The input function is normally given by equation (4) as follows:

$$net_i = \Sigma w_{ij} x_j + u_i \qquad (4)$$

where *net* describes the result of the net inputs x_i (weighted by the weights w_{ij}) impacting on unit *i*. Also, w_{ij} ware weights connecting neuron *j* to neuron *i*,x_j is output from unit *j* and u_i is a threshold for neuron *i*. Threshold term is baseline input to a node in absence of any other inputs. If a weight w_{ij} is negative, it is termed inhibitory because it decreases net input, otherwise it is called excitatory, for more information on activation function see (Herz et al., 1991).

4. Methods of validation and Evaluation of Classification Model

One way to avoid bias is to split the sample into two parts, The "training" sample and the "validation" sample. Then, the classification rule is created using the training sample and the apparent error rate is determined using the validation sample. Training the algorithm and evaluating its statistical performance on the same data yields an overoptimistic result. In most real applications, only a limited amount of data is available, which leads to the idea of splitting the data: Part of data (the training sample) is used for training the algorithm, and the remaining data (the validation sample) are used for evaluating the performance of the algorithm. The validation sample can play the role of new data (Arlot and Celisse, 2010).

4.1 Cross-Validation Method

Schneider (1997) indicated that Cross-validation is a model evaluation method that is they do not give an indication of how well the learner will do when it is asked to make new predictions for data it has not already seen. Cross-validation, sometimes called rotation estimation, is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. The ideal procedure of Cross Validation: Divide data into three sets, training, validation and test sets. Find the optimal model on the training set, and use the test set to check its predictive capability. See how well the model can predict the test set. The validation error gives an unbiased estimate of the predictive power of a model .This type of validation is, of course, more expensive computationally, but useful when the most accurate estimate of a classifier's error rate is required. A popular choice for the value of k is , yielding the well-known leave-one-out method.

4.2 Evaluation of a Classification Model

Some classification models produce continuous scores as final results. For two classes' problems, there is an important graphical technique for evaluating classification models that produce continuous scores.

4.2.1 ROC Curve

The ROC curve (receiver operating characteristic) , is a technique for visualizing, organizing, improving and selecting classifiers based on their performance. They facilitate our conception of classifiers and is therefore useful in research and in result presentation. Such representation also has a practical meaning, since we are able to construct a binary classifier for each point on the convex hull. Each straight segment of a convex hull is defined with two endpoints that correspond to two classifiers. We will label the first one A and the other one B. A new (combined) classifier C can be defined. For a given value of parameter a Î (0, 1) we can combine the predictions of classifiers A and B. We take the prediction of A with probability α and the prediction of B with probability 1−α. The combined classifier C corresponds to the point on the straight segment and by varying the parameter α we cover the whole straight segment between A and B. If the original ROC graph corresponds to probabilistic classifier, its convex hull also corresponds to a probabilistic classifier that is always at least as good as the original one. Convex hull is just one approach for constructing a ROC curve from a given set of points (Centor, 1991).

ROC graph is defined by X axis presents FP rate(t) and Y axis presents TP rate(t) where FP and TP are the False Positive rate and the True Positive rate , respectively.

4.2.2 Area under the ROC curve :

Area under the ROC curve is often used as a measure of quality of a probabilistic classifier. It is close to the perception of classification quality that most people have. Area under ROC curve is computed using the following formula:

$$A_{ROC} = \int_{0}^{1} \frac{TP}{P} d\frac{FP}{N} = \frac{1}{NP}\int_{0}^{N} TP\, dFP \qquad (5)$$

P= TP + FN and N=TN + FP

where FN and TN are the False Negative rate and the True Negative rate, respectively.

A random classifier (e.g. classifying by tossing up a coin) has an area under curve 0.5, while a perfect classifier has 1. Classifiers used in practice should therefore be somewhere in between, preferably close to 1 (Vuk and Curk , 2006).

5. Standard Living Data

Data were obtained by the Palestinian Central Bureau of Statistics (PCBS) Survey of Expenditure and Consumption in 2011. The survey data was made available to us for research through direct communication with the officials of the PCBS according to a special agreement between the PCBS and Al-Azhar University - Gaza .

The data consists of 12 variables (standard of living , monthly income , assistance , Animals holdings , agricultural land ,imputed rent , total expenditure , total consumption , non-consumption expenditure , remittances , taxes and children number).The dependent variable is the standard of living which takes the values (1 for high Standard of living , 2 for Middle standard of living , and 3 for Low standard of living) . such that:

Monthly Income: Cash or in kind revenues for individual or household within a period of month

Assistances: Includes data about cash and in kind assistances (assistance value, assistance source), also collecting data about household situation, and the procedures to cover expenses.

Animal holdings : The household have animal holdings (Cattle, Sheep and Goats, Poultry, Horses and Mules, Beehives).

Agricultural land : The household have agricultural land.

Imputed rent : Value of imputed rent for house hold in Israeli Shekel

Total Expenditure: It refers to the amount of Cash spent on purchase of goods and services for living purposes, and the value of goods and services payments or part of payments received from the employer, and Cash expenditure spent as taxes (non-commercial or non-industrial), gifts, contributions, interests on debts and other non-consumption items.

Total Consumption : It refers to the amount of Cash spent on purchase of goods and services for living purposes, and The value of goods and service payments or part of payments received from the employer, and own-produced goods and food, including consumed quantities during the recording period, and Imputed rent for own housing.

Non-consumption expenditure: Interests on loans , fees and taxes.

Remittances: Includes data about cash and in kind remittances (remittances value, remittances source), also collecting data about household situation, and the procedures to cover expenses.

Taxes : Includes data about cash and in kind taxes which paid from household .

Children number: Represents the number of male and female Sons in the household.

Investigating the Standard of living for Palestinian household, we now illustrate some descriptive statistics for the important variables in the study.

Table (1) illustrates that (16.28 %) of household in the sample belong to a high standard of living، (51.1 %) of household belong to a Middle standard of living, and (32.62%)household belong to a Low standard of living.

Table (1) : Frequencies for Standard of living

Categories of Living Standard	Frequency	Valid Percent
High standard of living	703	16.28 %
Middle standard of living	2206	51.10 %
Low standard of living	1408	32.62 %
Total	***4317***	***100.00***

Table (2) presents some descriptive statistics of the numeric variables, associated with the one way ANOVA test. The results in Table 2 show that there are significant differences between the means of the dependent variable categories according to standard of living, this means that the numeric independent variables which used in the analysis can be distinguish and differentiate between categories of the dependent variable.

Table (2) : Descriptive statistics of the Data (Numerical Variables)

Variable name	Level	Min.	Max.	Mean	S.D.	F	sig.
Monthly Income	High	0.0	38995.7	4939.87	3966.95	150.047	0.00
	Middle	0.0	45645.17	4353.94	3832.56		
	Low	0.0	18010.2	2656.48	1945.44		
Assistances	High	0.0	105000	98.58	534.87	12.399	0.00
	Middle	0.0	47500	217.66	1900.17		
	Low	0.0	31963	417.15	1116.27		
Total Expenditure	High	341.17	73219.3	6306.39	5402.82	108.374	0.00
	Middle	298.5	56036.17	4930.35	4336.14		
	Low	351.5	28826	3613.47	2600.53		
Total Consumption	High	555	71329.9	6777.80	4927.68	152.847	0.00
	Middle	383	55971.3	5143.54	4168.65		
	Low	651.5	21841.5	3772.58	2146.54		
Imputed Rent	High	0.0	9100	918.18	936.34	269.669	0.00
	Middle	0.0	4160	462.71	309.43		
	Low	0.0	5100	499.38	265.11		
Remittances	High	0.0	4000	181.35	402.32	29.838	0.00
	Middle	0.0	13632.5	172.24	540.63		
	Low	0.0	2750	65.34	209.09		
Taxes	High	0.0	1000	78.21	130.08	424.58	0.00
	Middle	0.0	702.5	13.63	30.65		
	Low	0.0	962.5	1.66	26.81		
Non-consumption expenditure	High	0.0	38293	204.19	1471.03	8.853	0.00
	Middle	0.0	11237	184.25	454.27		
	Low	0.0	21160	316.41	1139.28		
Children Number	High	1	17	5.26	2.43	74.497	0.00
	Middle	1	22	5.83	2.71		
	Low	1	28	6.67	2.79		

Table (3) shows that the 1.08 % of sample have animals holding and belong to the high standard of living , 9.66 % of sample have animals holding and belong to the middle standard of living , and 6.88 % of sample have animals holding and belong to the low standard of living class. Furthermore, there are 1.32% of sample have agricultural land and belong to the high standard of living, 14.55% of sample have agricultural land and belong to the middle standard of living, and 2.52% of sample have agricultural land and belong to the Low standard of living.

Moreover, the relationship between the type of standard of living and agricultural Land can be tested using test of independence. Results of the Chi-Square test are given in Table (3) and show that the Agricultural Land is dependent of standard of living as the p-value is close to (0.00). This result indicates that there is a significant association between types of standard of living of household and having agricultural land. Furthermore, Table (3) shows that the animals holding is dependent of standard of living as the p-value is close to (0.00). This result indicates that there is a significant association between types of standard of living of household and having animals.

Table (3): Descriptive statistics of the Data (Categorical Variables)

Variable	Categories	Standard of living			Total	Pearson Chi-Square (sig)
		High	Middle	Low		
Animals holding	Yes (%)	47 (1.08)	417 (9.66)	297 (6.88)	761 (17.62)	305.152 (0.00)
	No (%)	656 (15.2)	1789 (51.1)	1111 (32.62)	3556 (82.38)	
	Total %	***16.28%***	***51.1%***	***32.62%***	***100 %***	
Agricultural land	Yes (%)	57 (1.32)	628 (14.55)	109 (2.52)	794 (18.39)	72.086 (0.00)
	No (%)	646 (14.96)	1578 (36.55)	1299 (30.1)	3523 (81.61)	
	Total %	***16.28%***	***51.1%***	***32.62***	***100 %***	

6. Statistical Analysis of Standard of living Using Multinomial Logistic Regression

6.1 Building MLR Model

In this subsection MLR analysis was performed on the standard of living data set. The presence of a relationship between the dependent variable and combination of independent variables is based on the statistical significance of the final model chi-square. According to the results shown in Table (4), it is seen that -2 log likelihood value of basic model only with intercept term was 8668.980, this value decreased to 5066.781 with the independent variables appearance in the model. In this analysis, the probability of the model chi-square (3602.199) was (0.00), less than the level of significance (0.05). The null hypothesis that there was no difference between the model without independent variables and the model with independent variables was rejected. That is an evidence of existence of a relationship between the independent variables and the dependent variable.
Used in statistics and statistical modeling to compare an anticipated frequency to an actual frequency.

Goodness-of-fit tests are often used in business decision making. In order to calculate a chi-square goodness-of-fit, it is necessary to state the null hypothesis and the alternative hypothesis, choose a significance level (such as $\alpha = 0.05$) and determine the critical value.

Table (4) : Model Fitting Information

Model	Model Fitting Criteria			Liklehood Ratio Tests		
	AIC	BIC	-2Log Liklehood	Chi-Square	Df	Sig.
Intercept Only	8672.98	8685.72	8668.98	3602.199	24	0.00
Final	5118.781	5284.409	5066.781			

The results of the goodness of fit test of the final model are given in Table (5). This shows that the model provides a significant fit to the data as the p-value of Pearson goodness of fit test less than 0.05.

Table 5 : Goodness-of-Fit

Test	Chi-Square	D.f	Sig.
Pearson	5.6×10^7	8608	0.000
Deviance	5066.781	8608	1.000

As Pseudo R^2 values presented in Table 6 are examined, explanation ratios of dependent variables upon independent variables are given. Nagelkerke R^2 value is the modified form of Cox & Snell coefficient. According to the results shown in Table (8), it is seen that independent variables define 56.6 % of the variance in dependent variables (the proportion of variance of the response variable explained by the predictors) according to Cox & Snell R^2 value, 65.4 % according to Nagelkerke R^2 value, and 41.6 % according to McFadden value.

Table (6) : Pseudo R-Square measurements

Measurements	R^2 values
Cox and Snell	0.566
Nagelkerke	0.654
McFadden	0.416

Table (7) demonstrates the likelihood ratio test evaluates the overall relationship between an independent variable and the dependent variable. we checked the same point with all explanatory variables used to build model separately. The result was referred that the existence of a relationship between each of the explanatory variables and the response variable was supported.

According to the results shown in Table (7), it is seen that there is a statistically significant relationship between all the independent variables and the dependent variable.

Table (7) : Likelihood Ratio Tests

Effect	**Model Fitting Criteria**	**Likelihood Ratio Tests**		
	-2 Log Likelihood of Reduced Model	**Chi-Square**	**D.F**	**Sig.**
Intercept	5066.781	0.00	0	.
Monthly Income	5110.606	43.825	2	0.00
Assistances	5569.243	502.462	2	0.00
Expenditure	5084.680	17.899	2	0.00
Consumption	5104.725	37.944	2	0.00
Imputed Rent	5249.714	182.932	2	0.00
Remittances	5080.774	13.992	2	0.01
Taxes	5560.315	493.534	2	0.00
Non Consumption Expenditure	5533.231	466.449	2	0.00
Children Number	5109.912	43.131	2	0.00
Agricultural Land	5195.060	128.279	2	0.00
Animals Holding	5100.867	34.086	2	0.00

6.2 Estimates MLR Models

Since the dependent variable has three categories, we have two models. Table (8) shows the odds ratios (OR) for middle standard living (Model 1). The odds ratio is ‘the increase (or decrease) in odds of being in one outcome category when the value of the predictor increases by one unit.

As we can see in the Table 8 the variable monthly income is the first variable in the model with a p-value equal (0.00), and odds ratio equals (2.94) .This means the more the Monthly Income increases by one shekel with firmness other factors , the more the chance of household transfer of Low standard living to Middle increases in 3 times.

The second significant variable in the model is the Assistances with a p-value equal (0.00), and odds ratio equals (0.004) .This means the more the Assistances increases in one shekel with firmness other factors , the more the chance of household transfer of Middle standard living to Low decreases in 0.004 times.

The third significant variable is Consumption with a p-value equal (0.045), and odds ratio equals (10.77) .Next variable is Imputed Rent with a p-value equal (0.00), and odds ratio equals

(7.3×10^{-9}), Next variable is Remittances with a p-value equal (0.00), and odds ratio (1.024) ,Next variable is Taxes with a p-value equal (0.00), and odds ratio (1.717) ,Next variable is Non Consumption-Expenditure with a p-value equal (0.00), and odds ratio as (0.317), Other significant variables include Children Number with a p-value equal (0.00), and odds ratio equals (0.907), Next variable is Animals Holding =1 with a p-value equal (0.00), and odds ratio as (0.515) , which means that the household that have Animals Holding have the chance to be classified as higher standard of living approximately one half chance as much as household haven't Animals Holding.

But the Intercept and the variable Expenditure are not significant because there Sig are larger than ($\alpha = 0.05$).

Table (8): MLR Model (1) Middle standard of living Data and their standard errors, odds ratios, and p-values.

	Estimate (β)	Std. Error	Wald(Z value)	SIG	EXP(Odds ratio)
Intercept	3.606	3.701	0.949	0.330	.
Monthly Income	1.078	0.165	42.576	0.00	2.94
Assistances	-5.59	.275	413.054	0.00	0.004
Expenditure	-0.92	3.98	0.053	0.82	0.399
Consumption	2.377	1.185	4.03	0.045	10.77
Imputed Rent	-18.74	1.888	98.54	0.00	7.3×10^{-9}
Remittances	0.024	0.007	12.9	0.00	1.024
Taxes	0.54	0.037	210.08	0.00	1.717
Non Consum. Expenditure	-1.15	0.067	291.778	0.00	0.317
Children Number	-0.097	0.021	21.18	0.00	0.907
Animals Holding1	-0.663	0.14	22.434	0.00	0.515

As we can see in the Table (9), the variable monthly income is the first variable in the model with a p-value close to (0.03), and odds ratio equals (1.758) . This means the more the Monthly Income increases by one shekel with firmness other factors , the more the chance of household transfer of Middle standard living to High increases in 2 times.

The second significant variable in the model is the Assistances with a p-value equals (0.00), and odds ratio equals

(0.003) ,The third variable is the Expenditure with a p-value equals (0.00), and odds ratio equals (1.9×10^{-9}). This means the more the Expenditure increases in one shekel with firmness other factors , the more the chance of household transfer of High standard living to Middle decreases in 1.9×10^{-9} times.

The third variable is Consumption with a p-value equals (0.00), and odds ratio equals (1.67×10^{4}), Next variable is Imputed Rent with a p-value equals (0.02), and odds ratio equals (0.004),Next variable is Remittances with a p-value equals (0.002), and odds ratio (1.026) ,Next variable is Taxes with a p-value equals (0.00), and odds ratio as (1.862), Next variable is Non Consumption Expenditure with a p-value equals (0.00), and odds ratio equals (0.229), the last significant variables include Children Number with a p-value equals (0.00), and odds ratio equals (0.830).

While the Intercept , and Agricultural Land are not significant because there Sig are larger than ($\alpha = 0.05$) .

Table (9) : MLR Model (2) High standard of living Data and their standard errors, odds ratios, and p-values

	Estimate(β)	Std. Error	Wald(Z value)	SIG	EXP(Odds ratio)
Intercept	6.45	5.205	1.535	0.215	.
Monthly income	0.56	0.191	8.740	0.03	1.758
Assistances	-5.68	0.417	185.501	0.00	0.003
Expenditure	-20.078	5.615	12.787	0.00	1.9×10^{-9}
Consumption	9.729	1.663	34.207	0.00	1.67×10^{4}
Imputed Rent	-5.579	2.4	5.402	0.02	0.004
Remittances	0.026	0.008	9.672	0.002	1.026
Taxes	0.621	0.039	257.001	0.00	1.862
Non Consum. Expenditure	-1.474	0.080	339.127	0.00	0.229
Children Number	-0.186	0.029	41.177	0.00	0.830
Agricultural Land	-0.305	0.221	1.896	0.169	0.737

6.3 Classification results of Standard of living Data Set

Table (10) presents the Classification results of applying the initial model (between parentheses) and the final model of MLR. The initial model (with no independent variables included) can predict 51.1% of Standard of living correctly.

Table (10) : Initial and final MLR Model and Classification Using MLR Method

Observed	Predicted			
	High Standard of living	**Middle Standard of living**	**Low Standard of living**	**Percent Correct**
High Standard of living	0 (308)	703 (346)	0 (49)	0.0 % (43.8%)
Middle Standard of living	0 (91)	2206 (1875)	0 (240)	100 % (85%)
Low Standard of living	0 (8)	1408 (229)	0 (1171)	0.0 %(83.2%)
Overall Percentage	0.0 % (9.4 %)	100 % (56.8%)	0.0 % (33.8 %)	51.1 % (77.7%)

The percentages for the final model are given between parentheses.

Results in Table (10) show that the model with independent variables is better than the model without independent variables, where the model can correctly predict 85% of those belong to the middle standard of living category, 83.2% of those belong to the low standard of living has percentage accuracy of (83.2 %), and 43.8% of those belong to the high standard of living category. The final model can predict 77.7% of the correct standard of living.

7. Statistical Analysis of Standard of living Using Artificial Neural Networks (ANN):

Neural networks have several advantages. Most important is the ability to learn from data and thus potential to generalize, i.e. produce an acceptable output for previously unseen input data (important in prediction tasks). This even holds to a certain extent when input series contain low-quality or missing data. Another valuable quality is the non-linear nature of a neural network. Potentially a vast amount of problems may be solved . Furthermore no expert system (typically a programmer coding rules in a computer program) is needed which makes the network extremely flexible to changes in the environment. One only has to retrain the system. In summary , neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. This expert can then be used to provide projections given new situations of interest and answer "what if" questions.

7.1 The application of Artificial Neural Networks :

In artificial neural network model we included all the independent variables and the size of the resulting artificial neural network classification, was a three-layer structure with 10 input layers,8 hidden layers and 1 output layer.

The “*nnet*” library in R software had been used . We defined the hidden layer to be 4 , the initial random weights = 0.01 , the parameter for weight decay = 5×10^{-4} , the maximum number of iterations is 261 time and the convergence if the maximum number of iterations was reached set to 1 and otherwise set to 0 .

We examined all the available independent variable to be involved in the final model . The resulting neural network with model appear in table (11) below.

Table (11) : Classification Using ANN Method

Sample	Observation	Predicted Group Membership			
		Low living standard	Middle living standard	High living standard	Percent correct
Training	Low living standard	299	197	0	60.3%
	Middle living standard	64	1485	0	95.9%
	High living standard	0	3	974	99.7%
	Overall Percent	12.0%	55.8%	32.2%	91.3%
Testing	Low living standard	125	82	0	60.4%
	Middle living standard	45	612	0	93.2%
	High living standard	0	1	430	99.8%
	Overall Percent	13.1%	53.7%	33.2%	90.1%

Table (11) indicates that the model can correctly predict 60.3% of the Low living standard, 95.9% of middle living standard, and 99.7% of the high living standard. The model can predict 91.3% of the Palestinian families for in the training set, and can predict 90.1% of the Palestinian families for the tested set.

7.2 The ROC curve Analysis of Artificial Neural Network Model :

The ROC curve has been drown for different categories of the dependent variable separately in the same graph as given in Figure (1) and the area under the ROC curves has been computed for different categories of standard living, the area under the curve of the low standard living is (94.8%), the area under the middle standard living " is (79.1%) and the area under the curve of third category "high standard living is (99.9%). Finally, the area under curve of Artificial Neural Network model (multi-class area under the curve) is (97.2%).

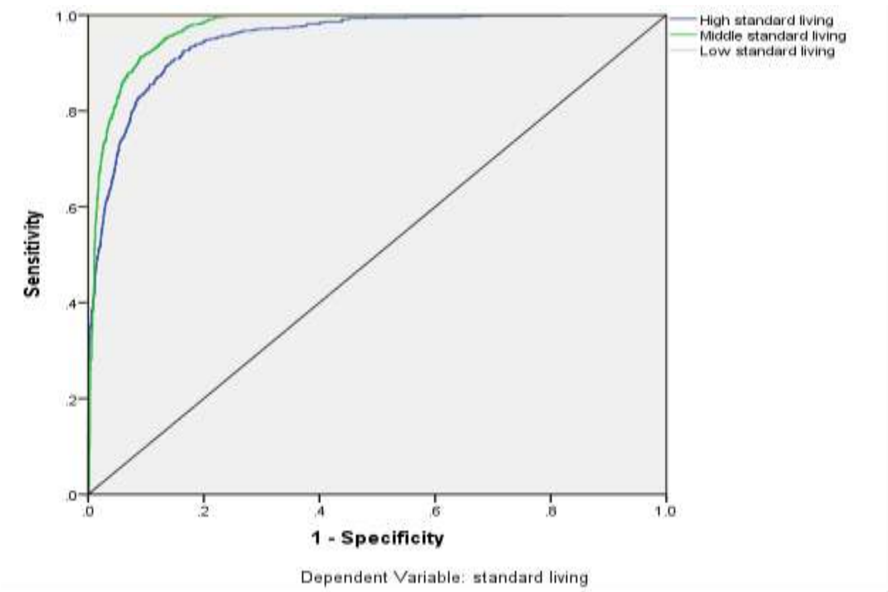


Figure (1) ROC curve of the model Neural Networks

8. Comparison Between Classification Methods

In this section both of MLR and ANN classification models of the standard of living for Palestinian households are compared by using several criteria.

The most frequently used criterion for comparison between the two methods is classification error. Table (12) shows the classification results of the two statistical methods (MLR and ANN) for dependent variable standard of living. Results in Table (12) indicates that the MLR model can correctly classify 77.7% compare to 90.1% of households which were correctly classified by the obtained ANN.

Table (12) : Classification Result of Two Statistical Methods

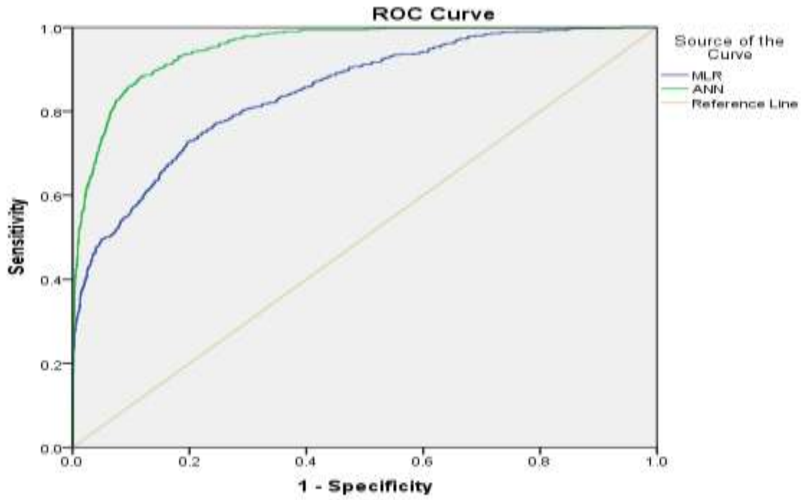
	MLR	**ANN**
High standard of living	43.8%	60.4%
Middle standard of living	85%	93.2%
Low standard of living	83.2%	99.8%
Classification of models	77.7%	90.1%

Table (13) presents the area under the ROC curve for all categories for MLR and ANN methods .

Table (13) : Area under the curve for Two Models

	Area (MLR)	**Area (ANN)**
High standard of living	87.8 %	94.8%
Middle standard of living	88.2 %	97.1%
Low standard of living	93.8 %	99.9%
Area of models	89.9 %	97.2%

The ROC curve of all the two statistical models have been drawn in the same graph. The area under the ROC curve has also been computed for each category. Figures (2) to (4) can clearly show that the ROC curve of the MLR is always higher than the other model. This indicates The MLR model provides classification for all categories with higher accuracy than ANN model when the two methods are applied at Expenditure and consumption (2011) data. The areas under the ROC curves are computed and presented in Table (15). The area under curve for MLR model is 89.9 % compared to that of ANN is (97.2 %) . Results of the ANN proved to be much better than MLR .



Figure(2) : ROC curve for the first category (High Standard of living)

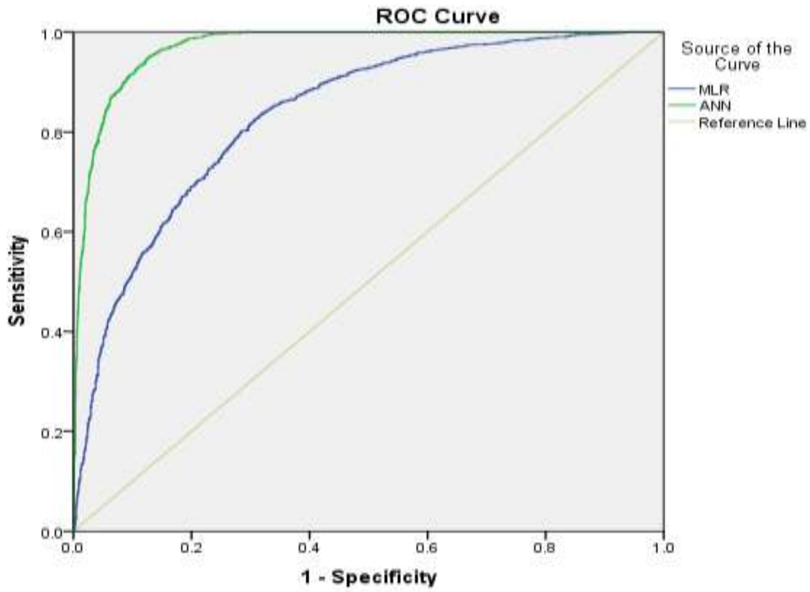
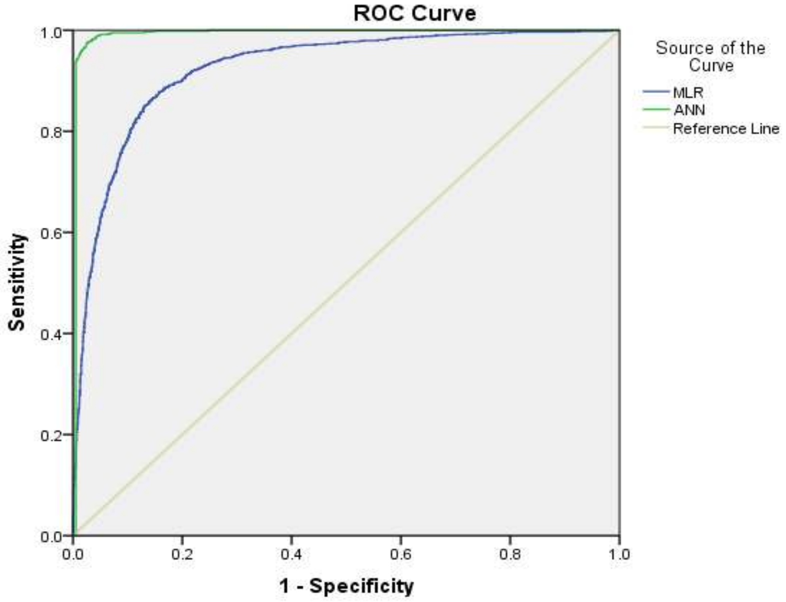


Figure (3) : ROC curve for the second category (Middle Standard of living)



Figure(4) : ROC curve for the third category (Low Standard of Living)

Akaike Information Criteria (AIC) and Bayesian Information Criterion (BIC) are two methods used to measure of model fit, A model having smaller AIC and BIC is considered the better fitting model.

Table (14) can clearly show that the AIC and BIC of the ANN is less than the MLR model, this indicates that the ANN model provides classification for all categories with much higher accuracy than MLR model when the two methods are applied at Expenditure and consumption (2011) data.

Table (14) : AIC and BIC for Two Models

	MLR models	ANN models
AIC	5118.78	4652.28
BIC	5284.41	3457.53

9. Conclusions and Recommendation

In this study we conducted a comparison between the MLR and ANN, using three different assessment techniques (classification Table, ROC curve and AIC,BIC Criteria) in order to obtain the best model that represents the dataset of standard of living for Palestinian household. The results of evaluation concluded that the ANN is much better than the ANN model.

Therefore, we recommend the use of ANN in classification when the assumptions of the normality and linearity are not satisfied. Furthermore, we invite researchers to consider the inclusion of additional independent variables related to standard of living to achieve a model with higher rate of correct classification and less error rate.

Reference :

Abu Saada A. (2013)," Statistical Methods for predicting violence against women cases in Palestine", Al-Azhar university , Gaza , Palestine , Master thesis.

Abu Samra A. (2012), "A Comparative Study of Classification Methods for Patients with Hypertension in Palestine", Al-Azhar university , Gaza , Palestine , Master thesis.

Agresti A. (2002)," Categorical Data Analysis", 2nd Edition. University of Florida. New York,Wiley.

Arlot S., &Celisse A. (2010),"A Survey Of Cross-Validation Procedures For Model Selection", Statistics Surveys: Vol. 4, PP. 40 –79.

Centor, R.M. Signal detectability ,(1991), " The use of ROC curves and their Analyses ". Med Decision Making 1991;11:102-106.

Dodge, Y. (2003), "The Oxford Dictionary of Statistical Terms", Oxford University, p. 282.

Eğrioğlu, E. &Aladağ, Ç.H. (2005). "Yapay Sinir Ağlarıve Arıma Modellerinin Melez Yaklaşımı İle Zaman Serilerinde Öngörü", VII. Ulusal Ekonometrive İstatistik Sempozyumu, İstanbul, 26-27 .

Fish, K., Barnes J. & Aiken M. (1995), "Artificial neural networks - A new methodology for industrial market segmentation", Industrial Marketing Management, 24 431-438.

Hosmer, D W. &Lemeshow, S. (2000)," Applied Logistic Regression. 3rd edition", New York: Johnson Wiley & Sons.

Human Right (2010) , Electronic Text book "Universal Declaration of Human Rights" Article 25.

Kutner,M. , Nachtsheim, C.,&Neter, J. (2004)."Applied Linear Statistical Models ". Fifth Edition, New York , Americas.

Lee, K. (2010), "A Multi-class Classification using Ensembles of Multinomial Logistic Regression Models", A Ph. D. Thesis, Department of Applied Mathematics and Statistics, Stony Brook University.

Nassar ,M. (2013), " A Comparative study among Logistic Regression, Discriminant Analysis and NeuralNetwork to Predict the Financial Failure ", Al-azhar university , Gaza , Palestine , Master thesis.

Odom, M. & Sharda R. (1990). "A neural network model for bankruptcy, Prediction In Proceedings of the international joint conference on neural , Networks ", 2 163–168. IEEE Press, Alamitos, CA.

Schneider,J. (1997),“ Cross Validation “ , http://www.cs.cmu.edu /~ schneide /tut5/ node42.html " .

Sharda, R. (1994), "Neural networks for the MS/OR analyst An Rioul, O.,Vetterli, M., 1991", Wavelet and signal processing, IEEE Application, Bibliography, Interfaces 24 (2), 116–130.

Tam, K.& Kiang M. (1990), "Predicting bank failures: A neural network approach. Applied Artificial Intelligence". 4(4) 265-282.

Vuk, M. &Curk,T. (2006)." ROC Curve, Lift Chart and Calibration Plot", Metodoloˇskizvezki, Vol. 3, No. 1, 2006, 89-108, Slovenia.

Yarmohammadi, M., Abdolmaleki, P., &Gity., M., (2004), "Comparison of logistic regression and neural network models in predicting the outcome of biopsy in breast cancer from MRI findings", Iran. J. Radiat. Res., 2004; 1 (4): 217-228.