International Journal of Financial, Administrative, and Economic Sciences
(IJFAES)
Vol. (4), No. (4)

IJFAES
April 2025

المجلة الدولية للعلوم المالية والإدارية والاقتصادية

الإصدار (4)، العدد (4)

Using Multinomial Logistic Regression to Identify Key Factors Affecting the Standard of Living in Sinnar, Sudan

Abdalrahim Ahmed Gissmalla Mustafa
Dept. of Information Management, Riyadh, Maximus KSA Company, Saudia Arabia
Ph.D. in Applied Statistics, Statistics, Sinnar University, Sudan
Abdalrhim192@gmail.com

Abstract

This study aims to identify and analyze the key determinants of household living standards in Sinnar State, Sudan, by using Multinomial Logistic Regression (MLR) to assess the impact of economic, demographic, and health-related factors. Given the disparities in living standards across the region, understanding the underlying causes is essential for guiding policy interventions. The study classifies households into three categories—high, middle, and low standard of living—based on variables such as income, education, access to healthcare, household size, and occupation. Primary data were collected through a detailed questionnaire, and secondary data were used to complement the findings. The sampling method involved a two-stage cluster sampling approach, resulting in a sample size of 800 households, and data analysis was conducted using SPSS software.

The study reveals that significant predictors of living standards include monthly household income, place of residence (urban vs. rural), occupation type, ownership of assets (such as cars and smart screens), and access to healthcare. The MLR model showed a classification accuracy of 84.1%, with higher accuracy for low-standard households (89.5%) compared to middle (81.2%) and high-standard households (75.4%). Key factors such as professional occupation and sufficient income were

https://doi.org/10.59992/IJFAES.2025.v4n4p11 E-ISSN 2977-1498

found to increase the likelihood of higher living standards, while rural residence and insufficient income were linked to lower living standards.

The research highlights the importance of these determinants for informed policymaking and the need for targeted interventions to improve living standards. Recommendations for future studies include expanding the set of variables considered and applying advanced statistical techniques for better classification accuracy. Additionally, the study calls for job creation initiatives to improve income levels and reduce socio-economic disparities. This research provides a understanding the socio-economic factors that influence living standards in Sinnar State.

Keywords: Multinomial Logistic Regression, Household Living Standards, Economic Factors, SPSS, Sinnar State, Sudan.

Introduction

Accurately identifying the determinants of living standards is a critical component of socio-economic research, particularly in regions where development efforts are ongoing. In the context of Sinnar State, Sudan, understanding the multifaceted nature of the standard of living is essential for informed decision-making and policy development. This study adopts Multinomial Logistic Regression (MLR) as the primary analytical framework to explore and quantify the most influential economic, demographic, and health-related factors that affect household living standards.

Multinomial Logistic Regression is a widely used multivariate statistical technique for modeling outcomes with more than two categorical responses. It enables the estimation of the probability of membership in each category of a dependent variable, based on one or more independent variables. Unlike linear models that assume constant variance and normality, MLR is particularly well-suited for analyzing unordered categorical outcomes without imposing restrictive assumptions on the distribution of predictors or the structure of the response variable.

The application of MLR in this research allows for the classification of households in Sinnar into distinct standard-of-living categories, based on variables such as income, education level, access to health services, household size, and income sources. This method provides not only the ability to identify statistically significant predictors but also to interpret the relative importance of each factor in determining the likelihood of a household falling into a particular living standard category.

By focusing on Sinnar State, this study provides localized insights that can support targeted policy interventions. The findings are expected to contribute to a broader understanding of socio-economic disparities and offer practical guidance for enhancing living conditions through data-driven planning.

Research Problem

Sinnar State faces significant disparities in household living standards, influenced by a range of economic and social factors. However, there is limited understanding of which factors most strongly predict these differences. Traditional methods often fall short in accurately classifying households into living standard levels. This research addresses the problem by using Multinomial Logistic Regression to determine the most import factors affecting the standard of living and identify the key predictors and classify families into high, middle, or low living standard categories, supporting better-informed policy decisions.

Research Questions

1. What are the most important economic factors influencing the standard of living for families, as identified through multinomial logistic regression?
2. Which economic variables play a statistically significant role in predicting a family’s standard of living category?
3. Does the Multinomial Logistic Regression (MLR) model effectively classify households in Sinnar according to their standard of living?

Research Objectives

1. To identify and analyze the most important economic factors affecting the standard of living for families in Sinnar using Multinomial Logistic Regression (MLR).
2. To determine which economic variables significantly contribute to predicting a family's classification into high, middle, or low standard of living categories using MLR.
3. To assess the effectiveness of the Multinomial Logistic Regression (MLR) model in classifying households into different standard of living categories (high, middle, or low).

Research Importance

This research is crucial for providing valuable scientific evidence to planners, policymakers, and the key factors influencing household living standards in Sinnar State. By identifying the most significant economic, social, and health determinants, the study aims to enhance the classification of households into high, middle, and low living standard categories and identify the most important factors thar affecting the standard of living. Through the application of Multinomial Logistic Regression (MLR),

Research Methodology

Sampling Methods:

A two-stage cluster sample, known as the "double stage sample," was used to select samples from households in which the paterfamilias of Sinnar state. Firstly, the locality was considered as a cluster, and all 23 administrative units of the state were included in the study. In the second stage of sampling, from each cluster (administrative unit), households were selected using simple random sampling.

Sample Size:

$$n_0 = \frac{z^2 pq}{d^2} \quad n = \frac{1.96^2(0.50)(0.50)}{0.05^2} \approx 400 * 2 = 800$$

n_0: Primary sample size

p : Population parameter

: (and we selected p = 50% because they took the largest sample size possible).

d : Statistical adjustment on both ends of p (taken here 5%).

Z: The significant level of z distribution so that it is equal to 0.05

With a design effect of (2) for the multistage nature of cluster sampling, accordingly, the sample size for the study was (800) households.

Data Sources:

- **Primary Data:**

  In this research, primary data were collected through a questionnaire administered to households. This method provided direct insights from respondents and served as the foundation for analyzing their perspectives and experiences.

- **Secondary Data:**

  Secondary data were sourced from scientific books, reputable websites, and other relevant academic literature. This data complemented the primary data by offering broader context and supporting evidence from existing research.

Multinomial Logistic Regression (MLR):

Multinomial logistic regression analysis is extension of analysis dichotomous variables, this model can be easily modified to handle the case where the outcome variable is nominal with more than two levels. For example, consider a study of choice of a health plan from among three plans offered to the employees of a large corporation. The outcome variable has three levels indicating which plan, A, B or C is chosen. Possible covariates might include gender, age, income, family size, and others. The goal is to estimate the probability of choosing each of the three plans as well as to estimate the odds of plan choice as a function of the covariates and to express the results in terms of odds ratios for choice of different plans. McFadden (1974) proposed a modification of the logistic regression model and called it a discrete choice model. As a result, the model frequently goes by that name in the business and econometric literature while it is called the multinomial, polychotomous, or polytomous logistic regression model in the health and life sciences.

Here the terms dichotomous and polytomous used to refer to logistic regression models, and the terms binominal and multinomial used to refer to logit models. (David W. Hosmer, JR. Stanley Lemeshow, 2000)

To develop the model, assume we have p covariates and a constant term, denoted by the vector x, of length p+1, where x_0= 1. We denote the two logit functions as

$$g_1 (X_1) = \text{in} \left[\frac{pr(Y=1|X}{\Pr\,(Y=0|X)}\right]$$

$$\beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \cdots \beta_{p1}x_p$$

$$\boldsymbol{x'\beta_1}$$

and

$$g_2 (X_2) = \text{in} \left[\frac{pr(Y=2|X}{\Pr\,(Y=0|X)}\right]$$

$$\beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \cdots \beta_{2p}x_p$$

$$\boldsymbol{x'\beta_2}$$

It follows that the conditional probabilities of each outcome category given the covariate vector are

$$Pr\ (Y = 0|x) = \frac{1}{1+e^{g1(x)}+e^{g2(x)}}$$

International Journal of Financial, Administrative and Economic Sciences, London Vol (4), No (4), 2025
https://doi.org/10.59992/IJFAES.2025.v4n4p11 E-ISSN 2977-1498

$$Pr\ (Y = 1|x) = \frac{e^{g1(x)}}{1 + e^{g1(x)} + e^{g2(x)}}$$

$$Pr\ (Y = 2|x) = \frac{e^{g2(x)}}{1 + e^{g1(x)} + e^{g2(x)}}$$

Following the convention for the binary model, we let π_i (**x**) = pr (Y = j |**x**) for j =0, 1, 2, each probability is a function of the vector of 2(p+1) parameters $\boldsymbol{\beta}' = (\boldsymbol{\beta}'_1 . \boldsymbol{\beta}'_1)$

A general expression for the conditional probability in the three-category model is

$$\pi\ j\ (x) = Pr(Y\ =\ j|x) = \frac{e^{gj(x)}}{\sum_{k=0}^{2} e^{gk(x)}}$$

Where the vector β_0 =0 and g_0 **(x)** = 0.

Maxiumum Likelihood Estimation:

Generalizing to a multinomial dependent variable requires us to make some notational adaptations. Let **J** represent the number of discrete categories of the dependent variable, where **J** ≥ 2. Now, consider random variable **Z** that can take on one of J possible values. If each observation is independent, then each **Z**i is a multinomial random variable. Once again, we aggregate the data into populations each of which represents one unique combination of independent variable settings. As with the binomial logistic regression model, the column vector n contains elements $\mathbf{n_i}$ which represent the number of observations in population i, and such that $\sum_{i-1}^{n} M$ the total sample size.

Since each observation records one of ***J*** possible values for the dependent variable, Z, let y be a matrix with *N* rows (one for each population) and

$\mathbf{J - 1}$ columns. Note that if $\mathbf{J = 2}$ this reduces to the column vector used in the

International Journal of Financial, Administrative, and Economic Sciences (IJFAES)
Vol. (4), No. (4)
IJFAES
April 2025
المجلة الدولية للعلوم المالية والإدارية والاقتصادية
الإصدار (4)، العدد (4)

binomial logistic regression model. For each population, $\mathbf{y_{ij}}$ represents the observed counts of the $j\ th$ value of **Zi**. Similarly, π is a matrix of the same dimensions as y where each element π_{ij} is the probability of observing the $j\ th$ value of the dependent variable for any given observation in the $i\ th$ population.

The design matrix of independent variables**,** **X**, remains the same—it contains **N** rows and **K + 1** columns where **K** is the number of independent variables and the first element of each row, $x_{i0} = 1$, the intercept. Let β be a matrix with **K + 1** rows and **J − 1** columns, such that each element $\boldsymbol{\beta_{ki}}$ contains the parameter estimate for the k th covariate and the j th value of the dependent variable. For the multinomial logistic regression model, we equate the linear component to the log of the odds of a j th observation compared to the J th observation. That is, we will consider the J th category to be the omitted or baseline category, where logits of the first **J − 1** categories are constructed with the baseline category in the denominator. (Scott A. Gzepiel, 2002 citation)

$$\text{Log}\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \log\left(\frac{\pi_{ij}}{1-\sum_{j-1}^{1-J}\pi_{ij}}\right) = \sum_{k=0}^{k}\pi_{ij}\beta\pi_{ij} \quad i = 1, 2, \ldots.\text{N}$$

$$j = 1, 2, \ldots\ldots. \text{J-1}$$

Solving for π_{ij}, we have:

$$\pi_{ij} = \frac{e\sum_{K=0}^{K} X_{i;}}{1+\sum_{I-1}^{J-1} e^{\sum_{k=0}^{k} x_{ik}\beta_{ki}}} \quad i < J$$

$$\pi_{ij} = \frac{1}{1+\sum_{I-1}^{J-1} e^{\sum_{k=0}^{k} x_{ik}\beta_{ki}}}$$

Parameter Estimation:

The goal of logistic regression is to estimate the **K+1** unknown parameters **β** in EqThis is done with maximum likelihood estimation which entails finding the set of parameters for which the probability of the observed data is greatest. The maximum

https://doi.org/10.59992/IJFAES.2025.v4n4p11 E-ISSN 2977-1498

likelihood equation is derived from the probability distribution of the dependent variable. Since each **y_i** represents a binomial count in the $i\ th$ population, the joint probability density function of **Y** is:

$$f(y|\beta) = \prod_{j=1}^{N} \frac{n_i!}{y_{i!}(n_i - y_i)} \ \ \pi^{y_i}(1 - \ \pi_i)^{n_i - y_i}$$

For each population, there are $\binom{n_i}{y_i}$ different ways to arrange successes from among $n_i$trials. Since the probability of a success for any one of the $n_i$trials isπ_i the probability of accesses is $\pi_i{}^{y_i}$_i Likewise, the probability of **n_i-y_i** failures is $(1 - \pi_i)^{\ n_i - y_i}$

For each population, the dependent variable follows a multinomial distribution with J levels. Thus, the joint probability density function is:

$$f(y|\beta) = \prod_{I=1}^{N} \left[\frac{n_i!}{\prod_{i=1}^{J} y_{ij}!} . \prod_{j=1}^{J} \pi_{ij}{}^{y_{ij}} \right]$$

When **J = 2**, this reduces to Eq (3-18). The likelihood function is algebraically equivalent to Eq. (3-19), the only difference being that the likelihood function expresses the unknown values of in terms of known fixed constant values for y**.** Since we want to maximize Eq. (3-19) with respect to, the factorial terms that do not contain any of the terms can be treated as constants. Thus, the kernel of the log likelihood function for multinomial logistic regression models is:

$$L\ (y|\beta) = \prod_{i=1}^{N} \prod_{j=1}^{J-1} \pi_{ij}{}^{y_{ij}}$$

Replacing the $i\ th$ terms, Eq. (3-19) becomes:

International Journal of Financial, Administrative and Economic Sciences, London Vol (4), No (4), 2025
https://doi.org/10.59992/IJFAES.2025.v4n4p11 E-ISSN 2977-1498

International Journal of Financial, Administrative, and Economic Sciences (IJFAES)
Vol. (4), No. (4)
IJFAES
April 2025
المجلة الدولية للعلوم المالية والإدارية والاقتصادية
الإصدار (4)، العدد (4)

$$= \prod_{i=1}^{N} \prod_{j=1}^{J-1} {\pi_{ij}}^{y_{ij}} \, . \, {\pi_{ij}}^{n_i - \sum_{j=1}^{J-1} y_{ij}}$$

$$= \prod_{i=1}^{N} \prod_{j=1}^{J-1} {\pi_{ij}}^{y_{ij}} \, . \, {\pi_{ij}}^{\frac{{\pi_{ij}}^{n_i}}{n_i - \sum_{j=1}^{J-1} y_{ij}}}$$

$$= \prod_{i=1}^{N} \prod_{j=1}^{J-1} {\pi_{ij}}^{y_{ij}} . \frac{\pi_i \, J^{n_i}}{\prod_{j=1}^{J-1} \pi_i \, J^{y_{ij}}}$$

Since a^{x+y}=$a^x a^y$, the sum in the exponent in the denominator of the last term becomes a product over the first J-1 terms of j. Continue by grouping together the terms that are raised to the $\mathbf{y_{ij}}$ power for each j up to **J-1**:

$$= \prod_{i=1}^{N} \prod_{j=1}^{J-1} \left(\frac{\pi_{ij}}{\pi_{iJ}}\right)^{y_{ij}} . \, {\pi_{iJ}}^{n_i}.$$

$$\prod_{i=1}^{N} \prod_{j=1}^{J-1} \left(e^{\sum_{k=0}^{k} x_{ik}\beta_{ki}}\right)^{y_{ij}} . \left(\frac{1}{1+\sum_{j=1}^{J-1} e^{\sum_{k=0}^{k} x_{ik}\beta_{ki}}}\right)^{n_i}$$

$$= \prod_{i=1}^{N} \prod_{j=1}^{J-1} \left(e^{y_{ij}\sum_{k=0}^{k} x_{ik}\beta_{ki}}\right)^{y_{ij}} . \left(1 + \sum_{j=1}^{J-1} e^{\sum_{k=0}^{k} x_{ik}\beta_{ki}}\right)^{-n_i}$$

Taking the natural log of Eq. (3-17) gives us log likelihood function for the multinomial logistic regression model:

$$l(\beta) = \sum_{i=1}^{N} \sum_{j=1}^{J-1} \left(y_{ij} \sum_{K=0}^{K} x_{ik}\beta_{kj}\right) - n_i log\left(1 + \sum_{j=1}^{J-1} e^{\sum_{k=0}^{k} x_{ik}\beta_{kj}}\right)$$

https://doi.org/10.59992/IJFAES.2025.v4n4p11 E-ISSN 2977-1498

And now we want to find the values for **β** which maximize Eq. (3-18). We will do this using the Newton -Raphson method, which involves calculating the first and second derivatives of the log likelihood function. We can take the first derivatives:

$$\frac{\partial l(\beta)}{\partial \beta}=\sum_{i=1}^{N} y_{ij}x_{ik}-n_i.\frac{1}{1+\sum_{j=1}^{J-1}e^{\sum_{k=0}^{k}x_{ik}\beta_{kj}}}.\frac{\partial}{\partial \beta_{kj}}\left(1+\sum_{j=1}^{J-1}e^{\sum_{k=0}^{k}x_{ik}\beta_{kj}}\right)$$

$$=\sum_{i=1}^{N} y_{ij}x_{ik}-n_i.\frac{1}{1+\sum_{j=1}^{J-1}e^{\sum_{k=0}^{k}x_{ik}\beta_{kj}}}.e^{\sum_{k=0}^{k}x_{ik}\beta_{ki}}\frac{\partial}{\partial \beta_{kj}}\sum_{K=0}^{K}x_{ik}\beta_{kj}$$

$$=\sum_{i=1}^{N} y_{ij}x_{ik}-n_i.\frac{1}{1+\sum_{j=1}^{J-1}e^{\sum_{k=0}^{k}x_{ik}\beta_{kj}}}.e^{\sum_{k=0}^{k}x_{ik}\beta_{ki}}.x_{ik}$$

$$=\sum_{i=1}^{N} y_{ij}x_{ik}-n_i\pi_{ij}x_{ik}$$

Note that there are **(J-1). (K+1)** equations in Eq. (3-22) which we want to set equal to zero and solve for each $\boldsymbol{\beta_{kj}}$. Although technically a matrix, we may consider **β** to be a column vector, by appending each of the additional columns below the first. In this way, we can form the matrix of second partial derivatives as a square matrix of order (J-1). **(K+1).** for each $\boldsymbol{\beta_{kj}}$**,** we need to differentiate Eq. (3-30) with respect to every other β_{kj}. We can express the general form of this matrix as:

$$\frac{\partial^2 l(\beta)}{\partial \beta_{kj}\partial \beta_{K'j'}}=\frac{\partial}{\partial \beta_{K'j'}}\sum_{i=1}^{N} y_{ij}x_{ik}-n_i\pi_{ij}x_{ik}$$

$$= \frac{\partial}{\partial \beta_{K'j'}} \sum_{i=1}^{N} -n_i \pi_{ij} x_{ik}$$

$$= -\sum_{i=1}^{N} n_i \pi_{ij} x_{ik} \frac{\partial}{\partial \beta_{K'j'}} \left(\frac{e^{\sum_{k=0}^{k} x_{ik}\beta_{kj}}}{1 + \sum_{j=1}^{J-1} e^{\sum_{k=0}^{k} x_{ik}\beta_{kj}}} \right)$$

Appling the quotient rule of Eq. (3-23). not that the derivates of the numerator and denominator differ depending on whether or not $\boldsymbol{j' = j'}$

$$\left(\frac{f}{g}\right)'(a) = \frac{g(a).f'(a) - f(a).g'(a)}{[g(a)]^2}$$

$$f'(a) = g'(a) = e^{\sum_{k=0}^{k} x_{ik}\beta_{kj}.x_{ik'}} \quad j' = j$$

$$f'(a) = 0 \quad g'(a) = e^{\sum_{k=0}^{k} x_{ik}\beta_{kj}.x_{ik'}} \; j' \neq j$$

Thus, when $j' = j$, the partial derivative in Eq.(3-26) becomes:

$$\frac{\left(1 + \sum_{k=0}^{k} e^{\sum_{k=0}^{k} x_{ik}\beta_{kj}}\right).e^{\sum_{k=0}^{k} x_{ik}\beta_{kj}.x_{ik'}} - e^{\sum_{k=0}^{k} x_{ik}\beta_{kj}}.e^{\sum_{k=0}^{k} x_{ik}\beta_{kj}.\pi_{ik'}}}{\left(1 + \sum_{j=1}^{J-1} e^{\sum_{k=0}^{k} x_{ik}\beta_{kj}}\right)^2}$$

$$= \frac{e^{\sum_{k=0}^{k} x_{ik}\beta_{kj}.\pi_{ik'}} \left(1 + \sum_{k=0}^{k} e^{\sum_{k=0}^{k} x_{ik}\beta_{kj}} - e^{\sum_{k=0}^{k} x_{ik}\beta_{kj}}\right)}{\left(1 + \sum_{j=1}^{J-1} e^{\sum_{k=0}^{k} x_{ik}\beta_{kj}}\right)^2}$$

$$= \pi_{ik} x_{ik'} (1 - \pi_{ik}) \ldots \ldots \ldots.$$

And when $j' \neq j$, they are:

$$\frac{0-\ e^{\sum_{k=0}^{k} x_{ik}\beta_{kj}}.\ e^{\sum_{k=0}^{k} x_{ik}\beta_{kj'}.x_{k_{ik'}}}}{\left(1+\sum_{j=1}^{J-1} e^{\sum_{k=0}^{k} x_{ik}\beta_{kj}}\right)^{2}}$$

$$= -\pi_{ik}x_{ik'}-\pi_{ij'}$$

We can now express the matrix of second partial derivatives for the multinomial logistic regression model as:

$$\frac{\partial^2 l(\beta)}{\partial\beta_{kj}\partial\beta_{K'j'}} = \sum_{i=1}^{N} n_i x_{ik}\pi_{ik}(1-\pi_{ik})\pi_{ik'}\ \ j'=j$$

$$\sum_{i=1}^{N} n_i x_{ik}\pi_{ik}\pi_{ij}x_{ik'}\ \ j'\neq j$$

Data Analysis and Discussion

Table (1): distribution of sample individuals by localities

Localities	Frequency	Percent
Singh	98	12.3%
Sinnar	184	23.0%
Dandar	125	15.6%
Suki	127	15.9%
Dali and Mimmoum	48	6.0%
Sharg Sinnar	140	17.5%
Abu Hagar	78	9.8%
Total	800	100.0%

Table (2): distribution of sample individuals by the Place of residence

Place of residence	Frequency	Percent
Rural	376	47.0%
Urban	424	53.0%
Total	800	100.0%

The table above shows that 53% of the respondents are from urban areas and 47% are from rural areas.

https://doi.org/10.59992/IJFAES.2025.v4n4p11 E-ISSN 2977-1498

Table (3): distribution of sample individuals by Gender

Gender	Frequency	Percent
Male	609	76 %
Female	191	24%
Total	800	100.0%

The table above shows that 76% of the respondents were male, and 24% were female.

Table (4): distribution of sample individuals by Age

Age	Percent	Frequency
20-30	56	7.0%
31-40	166	20.8%
41-50	208	26.0%
51-60	186	23.3%
61-70	146	18.3%
71-80	31	3.9%
81-90	7	.9%
Total	**800**	**100.0%**

The table above shows that (26%) of the sample between (41-50) years, (23.3%) of the respondents between (51-60) years, (20.8%) of the respondents between (31-40) years, (18.3%) of the respondents between (61-70) years, (7%) of the respondents between (20-30) years, (0.9%) of the respondents between (81-90) years, The results show that the majority of the samples in the age groups (41–50), (51–60) The survey's goal was to question paterfamilias, and we can see in our society that the majority of paterfamilias their age are in these groups.

Table (5): distribution of sample individuals by to educational level

Educational level	Frequency	Percent
Illiterate	83	10.4%
Reads and writes	184	23.0%
Basis / Primary	93	11.6%
Intermediate level	110	13.8%
Secondary	175	21.9%
Diploma	88	11.0%
Bachelor	29	3.6%
High Diploma	14	1.8%
Master	15	1.9%

https://doi.org/10.59992/IJFAES.2025.v4n4p11 E-ISSN 2977-1498

PHD	9	1.1%
Total	800	100.0%

The high level of education increases the standard of living of the individual as well as the cultural level in various economic, social, and health aspects, and the respondents have been surveyed on the school grade that they have completed, as shown in table (4-5). The first thing to note is the drop in the percentage of university and postgraduate education to 19.4%. Which has a negative impact on the living situation.

Table (6): distribution of sample individuals according to social status

social status	Frequency	Percent
Married	684	85.5%
Single	40	5.0%
Divorcee	27	3.4%
Widower	49	6.1%
Total	800	100.0%

The table above shows that the majority of the sample is married at 85.5%, while the proportion of unmarried was 5%, and the proportion of divorced and widowed was 9.5%.

Table (7): distribution of sample individuals according occupation

Occupation	Frequency	Percent
Occupational	46	5.8%
business owner	181	22.6%
Employer	157	19.6%
Professional	67	8.4%
Worker	112	14.0%
Policeman / Army	25	3.1%
Farmer	158	19.8%
Other	54	6.8%
Total	800	100.0%

The table above shows that the distribution of sample according to the occupation, 22.6% of the of the sample is business owner.

https://doi.org/10.59992/IJFAES.2025.v4n4p11 E-ISSN 2977-1498

Table (8): distribution of sample individuals by Family type

Family type	Frequency	Percent
Extended family	322	40.3%
Small family	478	59.8%
Total	800	100.0%

The table above shows the distribution of the sample according to family type; approximately 60% of the sample consists of small families with parents and children, and the other 40% are extended families.

Table (9): distribution of sample individuals by standard of living level

what is evaluation of the standard of living for your family type	Frequency	Percent
High	115	14.4%
Middle	324	40.5%
Low	361	45.1%
Total	800	100.0%

The table above shows that 45.1% of households are at a low standard of living, 40.5% of households are at a medium level of living, and only 14.4% of households are at a high standard of living.

Multinomial Logistic Regression (MLR):

The standard of living data set was subjected to multinomial logistic regression analysis in this part. The presence of a link between the dependent variable and a combination of independent variables is determined by the statistical significance of the final model chi-square.

Table (10): Model-Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1602.466	925.880	214	.000
Final	676.586			

The 2-log likelihood value of the basic model only with intercept term was 1602.466, as indicated in the findings in Table above. With the inclusion of independent variables in the model, this value fell to 676.586. The probability of the model chi-

https://doi.org/10.59992/IJFAES.2025.v4n4p11 E-ISSN 2977-1498

square (925.880) in this analysis was (0.000) less than the level of significance (0.05). The null hypothesis, that there was no difference between the models with and without independent variables, was rejected. That demonstrates the presence of a relationship between the independent and dependent variables.

Table (11): Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	1479.810	1380	.031
Deviance	676.586	1380	1.000

The goodness of fit of the final model is tested in the table above. The testing results show that the model gives a significant fit to the data since the Pearson goodness of git test P-value less than 0.05

Table (12): Pseudo R-Square

Cox and Snell	.686
Nagelkerke	.793
McFadden	.578

The pseudo R-Square value presented in the table above explains the rations of dependent variables upon independent variables. The Nagelkerke R-Square value is the modified Cox and Snell coefficient. According to the result in the table, it is seen that independent variables define 79% of the variation in dependent variables (the proportion of variance of the response variable explained by the predictors). We depend on Nagelkerke to explain the proportion of variance because sometimes Cox and Snell give a value greater than 1.

https://doi.org/10.59992/IJFAES.2025.v4n4p11 E-ISSN 2977-1498

Table (13): multinomial logistic regression model (1) middle standard of living (The Reference Category is high)

What is evaluation of the standard of living for your family?	B	Std. Error	Wald	df	Sig.	Exp (B)
Intercept	-55.838	13547.81	0	1	0.997	
monthly household income	-0.005	0.003	3.893	1	0.048	0.995
[Place of residence=1]	1.656	0.575	8.285	1	0.004	5.236
[Place of residence=2]	0^b	.	.	0	.	.
[occupation=1]	1.163	1.34	0.753	1	0.386	3.199
[occupation=2]	0.939	1.082	0.752	1	0.386	2.556
[occupation=3]	1.55	1.139	1.853	1	0.173	4.714
[occupation=4]	4.347	1.53	8.074	1	0.004	77.221
[occupation=5]	0.544	1.179	0.213	1	0.645	1.722
[occupation=6]	-0.302	1.705	0.031	1	0.859	0.739
[occupation=7]	0.547	1.121	0.238	1	0.625	1.728
[occupation=8]	0^b	.	.	0	.	.
[Type of housing ownership? =1]	3.727	1.299	8.237	1	0.004	41.571
[Type of housing ownership? =2]	3.196	1.74	3.375	1	0.066	24.441
[Type of housing ownership? =3]	3.073	5.693	0.291	1	0.589	21.616
[Type of housing ownership? =4]	20.357	2398.851	0	1	0.993	693136408
[Type of housing ownership? =5]	0^b	.	.	0	.	.
[car =1]	-1.196	0.519	5.305	1	0.021	0.303
[car =2]	0^b	.	.	0	.	.
[smart screen =1]	1.968	0.716	7.555	1	0.006	7.158
[smart screen =2]	0^b	.	.	0	.	.
[Is the income of household sufficient for living expense=0]	-3.423	1.122	9.3	1	0.002	0.033
[Is the income of household sufficient for living expense =1]	0^b	.	.	0	.	.
[what the main shopping place =1]	0.703	0.951	0.547	1	0.46	2.02
[what the main shopping place =2]	-1.175	0.598	3.856	1	0.05	0.309
[what the main shopping place =3]	0^b	.	.	0	.	.
[If you feel sick, break or injured, where do you g =1]	5.395	1.818	8.809	1	0.003	220.35
[If you feel sick, break or injured, where do you go =2]	3.512	1.836	3.659	1	0.056	33.503
[If you feel sick, break or injured, where do you go?=3]	4.965	1.816	7.475	1	0.006	143.315
[If you feel sick, break or injured, where do you go =4]	1.61	2.028	0.63	1	0.427	5.003
[If you feel sick, break or injured, where do you go =5]	0^b	.	.	0	.	.

According to the table above, the odds rations for each of the significant variable ratios are "the increase or decrease in odds of being in one outcome categories when the value of the predictor increases by one unit, using the first category's high standard of living as the reference category."

- **Monthly Household Income:** If monthly income decreases by one unit, the odd ratio that the family will be at the middle standard of living compared to the family at the high standard of living is (0.995).
- **Place of Resident:** The odd ratio of families who live in rural areas compared to those that live in urban areas is 5.23 times greater to be in middle standard of living compared to a high standard of living.
- **Occupation:** includes 8 categories (occupational, business owner, employer, professional, worker, Policeman / Army, farmer, Housewife)
- If the head of families works as a professional compared to a housewife, the odds ratio of the family being at the middle standard of living is 77.22 times higher than the family at the high standard of living.
- **Car:** The odd ratio of households with a car with those without a car is 0.303 times lower to be in a median level of life versus a high standard of living.
- **Smart Screen:** the odd ratio of households with a smart screen with those without a smart screen is 7.15 times greater to be in a median level of life versus a high standard of living.
- **Sufficient Income** the odd ratio of households with a sufficient income versus those with an insufficient income is 0.033 times lower to be at a median level of life versus a high standard of living.
- **The Main Shopping Places:** the odd ratio of households shopping at the big market **versus** those shopping at stores in their neighborhood or village is 0.309 times lower to be at a median level of standard of living **versus** a high standard of living.

https://doi.org/10.59992/IJFAES.2025.v4n4p11 E-ISSN 2977-1498

International Journal of Financial, Administrative, and Economic Sciences (IJFAES)
Vol. (4), No. (4)
IJFAES
April 2025
المجلة الدولية للعلوم المالية والإدارية والاقتصادية
الإصدار (4)، العدد (4)

- **If You Feel Sick, Break Or Injured, Where Do You Go:** includes 5 categories (hospital, private clinic, Health center, ALbaser, and do not go)
  - **Hospital**: the odd ratio of households that go to the hospital versus those that do not go is 220.35 times greater to be at a median level of life versus a high standard of living.
  - **Private Clinic:** the odd ratio of households that go to the private clinic versus those that do not go is 33.503 times greater to be at a median level of life versus a high standard of living.
  - **Health Center:** the odd ratio of households that go to the Health center versus those that do not go is 143.31 times greater to be at a median level of life versus a high standard of living.

Table (14): multinomial logistic regression model (2) low standard of living (The reference category is high)

Variables	B	Std. Error	Wald	df	Sig.	Exp(B)
Intercept	24.321	8008.5	0.000	1	0.997	
monthly household income	-0.015	0.004	11.642	1	0.000	0.9854
[Place of residence=1]	1.987	0.630	9.953	1	0.001	7.2955
[Place of residence=2]	0^b			0		
[occupation=1]	0.723	1.430	0.256	1	0.613	2.0607
[occupation=2]	-0.467	1.161	0.162	1	0.687	0.6268
[occupation=3]	0.683	1.200	0.324	1	0.569	1.9800
[occupation=4]	4.913	1.607	9.347	1	0.002	135.9896
[occupation=5]	-1.573	1.269	1.535	1	0.215	0.2075
[occupation=6]	0.170	1.731	0.010	1	0.921	1.1856
[occupation=7]	-0.600	1.200	0.251	1	0.616	0.5485
[occupation=8]	0^b			0		
[Type of housing ownership?=1]	4.441	1.682	6.968	1	0.008	84.8228
[Type of housing ownership? =2]	4.507	2.089	4.653	1	0.031	90.6412
[Type of housing ownership? =3]	2.621	5.867	0.200	1	0.655	13.7535
[Type of housing ownership? =4]	17.37	2398.85	0.000	1	0.994	3506097
[Type of housing ownership? =5]	0^b			0		
[car =1]	-1.416	0.629	5.063	1	0.024	0.2428
[car =2]	0^b			0		
[refrigerator =1]	1.043	0.656	2.531	1	0.111	2.8381
[refrigerator =2]	0^b			0		
[air conditioner =1]	-0.990	0.585	2.866	1	0.090	0.3717
[air conditioner =2]	0^b			0		

International Journal of Financial, Administrative, and Economic Sciences (IJFAES)
Vol. (4), No. (4)

IJFAES
April 2025

المجلة الدولية للعلوم المالية والإدارية والاقتصادية
الإصدار (4)، العدد (4)

[freon air conditioner =1]	1.558	0.716	4.732	1	0.029	4.7491
[freon air conditioner =2]	0[b]			0		
[smart screen =1]	2.072	0.878	5.564	1	0.018	7.9379
[smart screen =2]	0[b]			0		
[I-pad =1]	-2.854	1.229	5.396	1	0.020	0.0576
[I-pad =2]	0[b]			0		

Table (15): multinomial logistic regression model (2) low standard of living (The reference category is high)

Variables	B	Std. Error	Wald	df	Sig.	Exp (B)
[Is the income of household sufficient=0]	-6.992	1.159	36.39	1	0.000	0.0009
[Is the income of household sufficient=1]	0[b]			0		
[Do you resort to permanent borrowing to provide the living expenses for the family? =1]	1.669	0.746	5.004	1	0.0253	5.3080
[Do you resort to permanent borrowing to provide the living expenses for the family? =2]	0[b]			0		
[What the main shopping place? =1]	-0.890	1.032	0.744	1	0.3883	0.4105
[What the main shopping place? =2]	-2.120	0.652	10.562	1	0.0012	0.1200
[what the main shopping place?=3]	0[b]			0		
[Do the family use the family planning methods=1]	-1.403	0.667	4.426	1	0.0354	0.2458
[Do the family use the family planning methods=2]	0[b]			0		
[Which kind of methods are used in family planning? =1]	1.645	0.731	5.069	1	0.0244	5.1834
[Which kind of methods are used in family planning? =2]	1.945	4.268	.208	1	.649	6.994
[Which kind of methods are used in family planning? =3]	.281	1.159	.059	1	.809	1.324
[Which kind of methods are used in family planning? =4]	1.223	1.627	.565	1	.452	3.398
[Which kind of methods are used in family planning? =5]	0[b]	.	.	0	.	.
[Which kind of methods are used in family planning? =6]	0[b]	.	.	0	.	.
[If you feel sick, break or injured, where do you go? =1]	4.566	1.985	5.293	1	0.0214	96.1607
[If you feel sick, break or injured, where do you go? =2]	2.110	2.028	1.082	1	0.2983	8.2462
[If you feel sick, break or injured, where do you go? =3]	3.865	2.001	3.729	1	0.0535	47.6801
[If you feel sick, break or injured, where do you go? =4]	1.554	2.277	0.466	1	0.4950	4.7307
[If you feel sick, break or injured, where do you go? =5]	0[b]			0		
[Has health care effected the spending of family in the last 12 months? =1]	1.094	0.527	4.305	1	0.0380	2.9861
[Has health care effected the spending of family in the last 12 months? =2]	0[b]			0		

- **Monthly Household Income:** If monthly income decreases by one unit, the odd ratio that the family will be at the low standard of living compared to the family at the high standard of living is (0.9854).
- **Place of Resident:** The odd ratio of families who live in rural areas compared to those that live in urban areas is 7.2955times greater to be in low standard of living compared to a high standard of living.
- **Occupation:** If the head of a family works as a professional compared to a housewife, the odds ratio of the family being at a low standard of living is 135.9896 times higher than the family at a high standard of living.
- **Type of Housing Ownership:** includes 5 categories (owned house, owned house (moving), rented house (moving), governmental, other)
  - The odd ratio of families who live in owned houses compared to those who live in houses gifted to them (or other) to be at a low standard of living is 84.82 times greater than the families at a high standard of living.
  - The odd ratio of families who live in owned houses(moving) compared to those who live in houses gifted to them (or other) to be at a low standard of living is 90.64 times greater than the families at a high standard of living.
- **Car:** The odd ratio of households with a car compared to those without a car is 0.2428 times lower to be in a low versus a high standard of living.
- **Freon Air Conditioner**: the odd ratio of households with a freon air conditioner versus those without a freon air conditioner is 4.7491 times greater to be at a low level versus a high standard of living.
- **Smart Screen:** the odd ratio of households with a smart screen versus those without a smart screen is 7.937 times greater to be at a low level versus a high standard of living.
- **I-pad:** the odd ratio of households with an I-pad versus those without an I-pad is 0.0576 times greater for being at a low standard of living versus a high standard

https://doi.org/10.59992/IJFAES.2025.v4n4p11 E-ISSN 2977-1498

of living.

- **Sufficient Income:** the odd ratio of households with a sufficient income versus those with an insufficient income is 0.0009 times lower to be at a low level of life versus a high standard of living.
- **Borrowing to Cover the Family's Living Expenses:** the odd ratio of households that borrow to cover the family's living expenses compared to those that do not borrowing to cover the family's living expenses is 5.3080 times greater being at a low standard of living versus a high standard of living.
- **The Main Shopping Places:** the odd ratio of households shopping at the big market **versus** those shopping at stores in their neighborhood or village is 0.1200 times lower to be at a low standard of living versus a high standard of living.
- **The Family Planning Methods**: the odd ratio of households that use family planning methods versus those that do not is 0.2458 times lower, being at a low versus a high standard of living.
- **Which Kind of Methods are used in Family Planning?** (Contraceptive pills, Helix contraception, injection, Natural organization, Nexplanon Implanon slice, not used).
    - The odd ratio of households that use contraceptive pills compared to those that use Nexplanon Implanon (slice) and do not use it is 5.1834 times greater, being at a low versus a high standard of living.
- **If You Feel Sick, Break or Injured, Where Do You Go:**
    - **Hospital:** the odd ratio of households that go to the hospital versus those that do not go is 90.16 times greater to be at a low level of life versus a high standard of living.

Table (16): Classification

Observed	Predicted			
	High	**Middle**	**low**	**Percent Correct**
High	86	26	2	75.4%
Middle	15	263	46	81.2%
Low	2	36	323	89.5%
Overall Percentage	12.9%	40.7%	46.4%	84.1%

The table shows that the model correctly classifies 84.1% of the standard of living of a household.

The model could be able to classify 89.5% of households that belong to the low standard of living. That means it succeeded in classification in 323 households and filed in 38 households.

The model could be able to classify 81.2% of households that belong to the low standard of living. That means it succeeded in the classification of 263 households and filed in 59 households.

The model could be able to classify 75.4% of households that belong to the low standard of living. That means it succeeded in classification in 86 households and filed in 28 households.

Results

1. The results showed the multinomial logistic regression fits with the data and can be used to predict the levels of standard of living.
2. The results showed that there were 15 variables are (is the income sufficient for living expenses, having a car, and has health care affected family spending in the last 12 months, place of resident) that could predict the families with a low standard of living compared to those with a high standard of living.
3. The results showed that there were 12 variables (is the income sufficient for living expenses, having a car, and has health care affected family spending in the

last 12 months) that could predict families with a middle standard of living compared to families with a high standard of living.

4. The results revealed that the multinomial logistic regression model has a higher rating accuracy of classifications of new observations, with an accuracy of 84.1%.

Recommendations

1. Further research should be conducted that take into consideration 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.
2. Taking advantage of advanced statistical methods from, logistics models, to classify between more than two groups in all Regions in Sudan.
3. Creating new jobs to achieve an abundance in income rather than depleting enterprises, to be able to cover the monthly spending on food, housing, and health.

References

1. Abdelmonem, A. & May, S. (2012). *Practical multivariate analysis* (5th ed). Taylor & Francis Group, LLC. https://doi.org/978-1-4665-0324-3
2. Abdelmonem, A., May, S., Donatello, R. A., & Clark, V. A. (2020). *Practical multivariate analysis* (6th ed). Taylor & Francis Group, LLC. https://doi.org/978-1-138-70222-6
3. Izenman, A. J. (2008). *Modern multivariate statistical techniques: Regression, classification, and manifold learning*. Springer-Verlag New York. https://doi.org/10.1007/978-0-387-78185-8
4. Fadl Al-Mawla, A. A. (2015). *Comparison of discriminant analysis, binary logistic regression, and neural network models in observation classification, Sudan* (Unpublished master's thesis). Khartoum University of Science and Technology.

5. Rencher, A. C. (2002). *Methods applied multivariate analysis*. Springer-Verlag New York, Inc.
6. Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (3rd ed). John Wiley & Sons, Inc.
7. Everitt, B., & Hothorn, T. (2011). *An introduction to applied multivariate analysis with R*. Springer, Verlag New York. https://doi.org/10.1007/978-1-4419-6358-4
8. Czepiel, S. A. (2002). Maximum likelihood estimation of logistic regression models: Theory and implementation. Retrieved from http://czep.net
9. Garson, G. D. (2014). *Logistic regression binary and multinomial*. Statistical Associates Publishing.
10. Hilbe, J. M. (2009). *Logistic regression models*. Jet Propulsion Laboratory, California Institute of Technology, Taylor & Francis Group, LLC.
11. Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). Pearson Prentice Hall.
12. Ho, R. (2006). *Handbook of univariate and multivariate data analysis and interpretation with SPSS*. Taylor & Francis Group, LLC.
13. Karl, W., & Simar, L. (2007). *Applied multivariate statistical analysis*. Springer, Berlin Heidelberg New York.
14. Ahmad, Z., & Ejaz, Z. (2011). Classification of households with respect to poverty by using cluster analysis. *Proceedings of ICCS-11*, Lahore, Pakistan.