Lithuanian Journal of Statistics
Lietuvos statistikos darbai
Vol. 62, 2023, pages 1–19
https://doi.org/10.15388/LJS.2023.33905

The determinants of severe food insecurity in Africa using the longitudinal generalized Poisson mixed model

Adusei Bofa, Temesgen Zewotir

School of Mathematics, Statistics, and Computer Science,
University of KwaZulu Natal Westville Campus
Durban, 4041, South Africa

E-mail: 221119873@stu.ukzn.ac.za; zewotir@ukzn.ac.za

Received November 22, 2022; revised June 18, 2023; published online December 29, 2023

Abstract. Food insecurity is a multifaceted issue (challenge) that affects health care, policies, agriculture output leadership, the environment, the food system, and the politics of global commerce in the food industry. Our aim was to get the relevant components of food security and nutrition concerning Africa holistically and use these identified components to discover the most informative correlates that affect the number of severe food insecure individuals in Africa with its population as an offset. Principal Component Analysis (PCA) was used to detect the relevant components of Africa's food security and nutrition. The Poisson Generalized Linear Mixed Model (GLMM) was employed to identify the significant components. Generalized estimating equations were then applied to account for the overdispersion associated with the Poisson distribution. To make the interpretation of the results more meaningful, 10 PCA components were selected. They explained 74.6% of the variation within the data. The GLMM analysis remarkably identified Nutrient Intake, Average Food Supplied, Child Care, Dietary Supply Adequacy, and Feeding Practices Among Infants to be significantly associated with the Rate of Severe Food Insecure Individuals (p-value < 0.05). A better improvement in the average food supply in Africa is likely to yield an improvement in food security and nutrition. Our findings provide insight concerning Africa which will help policymakers create targeted plans for Africa that will address issues with food security and nutrition, and this will fuel the achievement of Sustainable Development Goal 2.

Keywords: principal component analysis; Poisson generalized linear mixed model; food security; nutrition

1 Introduction

The availability of food and people's capacity to get it, as well as the stability of availability and access through time, all contribute to what is known as "food security", a fundamental human right. The world has pledged to eradicate all facets of food and nutrition insecurity under the Sustainable Development Goals (SDG). By attaining Goal 2 of the SDG, we might help end hunger, guarantee food security, improve nutrition, and boost sustainable agriculture. It is only possible to attain food and nutrition security if all members of a society have physical, social, and economic accessibility to a sufficient quantity of food that is safe (contaminant-free) and nutritious and satisfies their needs for a healthy life [14]. Food has turned into a serious worldwide issue with strong roots in Africa throughout the years. The situation regarding food security has not been good in Africa over the past years, and as of 2019, there were 250.3 million malnourished individuals, with 15.6 million of them living in Northern Africa and 234.7 million in Sub-Saharan Africa [14].

Food insecurity and nutrition is a multifaceted issue (challenge) that impacts health care, policies, agriculture out-put leadership, the environment (which includes mother nature), the food system, and the politics of global commerce in the food industry [11]. These factors might have helped the Food and Agriculture Organization of the United Nations (FAO) introduce the incidence of severe and moderate food insecurity in 2017. Global hunger levels are still alarmingly high. The Global Report on Food Crises 2022 indicates that in 2021, they broke records, having nearly 193 million people being affected by severe food insecurity-roughly 40 million more people than the historical average attained in 2020.

Analyzing the population's existing food and nutrition security conditions is crucial before creating food security policies or programs. This analysis clarifies the conditions of those who are vulnerable to hunger and/or malnutrition in terms of numbers, quality, location, and potential causes of insecurity, a fact that the Africa Regional Assessment of Food Security and Nutrition 2021 report demonstrates. Despite some efforts made against food insecurity in the previous years among countries in Africa, Africa remains a tremendously food-insecure continent [29]. This shows that in terms of tackling food security and nutrition, Africa has not yet lived up to expectations.

To prevent and create an intervention strategy for Africa's food and nutrition insecurity, it is necessary to identify issues relevant to Africa. Additionally, some earlier investigations have shown the over-reliance on food availability and accessibility indicators, while only a small number of authors included utilization indicators in their publications in addition to availability and accessibility indicators. Nicholson *et al.* [25] reviewed works done concerning household (91 papers) and regional (26 papers) food security models for almost the past decade. According to their assessment, few works employed statistical models to determine the components of food security, and all papers reviewed focused on the three pillars of access, availability, and utilization indicators. Once more, they emphasized that few studies were done on regional bases whereas the majority of the research they analyzed focused on households.

It was revealed by Smith and Meade [31] that across 134 countries, low levels of household income, unemployment, weak social networks, and social capital were significantly correlated with household food insecurity as measured by the Food Insecurity Experience Scale (FIES) using multilevel linear probability models. In a cross-country investigation, Yunusa *et al.* [37] employed the Global Food Security Index (GFSI) as a response variable and discovered that population and the availability of water resources were insufficient indicators of national food security. The socio-demographic correlates of food insecurity in Middle Eastern and North African nations were examined by Omidvar *et al.* [26] using household-level FIES data. Additionally, some previous research on Africa has revealed that issues including conflict, economic slump, and climate change are connected to food security [9, 28].

Nicholson *et al.* [25] made four recommendations: Avoid confusing "food availability" and "food security", combine food access indicators, evaluate stable outcomes for food security indicators, and create empirical data tying results from agricultural systems models to outcomes related to food access. This suggests the need for more variables in measuring food security and nutrition, especially for Africa. With a large number of variables within the FAO dataset concerning food and nutrition security, the dispersion matrix may be too large to study and interpret properly. As a result, many researchers are unable to use all of the relevant metrics (variables) provided by the FAO for food and nutrition security. Some metrics are removed, which causes information to be omitted in the data required to identify variations. The concern now is how to use all the metrics (indicators) of food and nutrition security provided by the FAO without losing most or all of the variation within the data. Additionally, the plausible correlation between the observations that are recorded yearly within the same countries (subjects), as well as certain potential heterogeneous variances among observations that are recorded concerning the same countries (subjects) that relate to food and nutrition security in Africa, were not taken into consideration in any of these studies.

In this context, researchers have been applying a focused and critical lens to earlier studies and activities related to food and nutrition security to promote more inclusive and effective approaches. As a result, there is still a critical need to comprehensively identify significant aspects of recent data on food security and nutrition in Africa to comprehend the continent's common dynamics without sacrificing one concept over the other and develop effective coping mecha-nisms. The metrics used by previous works (Nicholson *et al.* [25]) do not show the complete picture of Africa's food security and nutrition holistically.

Indeed, the selection of variables in the context of food security and nutrition research in Africa is often guided by technical criteria. However, there is a tendency to focus primarily on availability indicators, leading to an overemphasis on certain aspects of food security while neglecting other important components. Also, in both traditional regression and more complex models, it is desirable to have a parsimonious model with as few parameters as possible. While having more variables in the dataset allows for the fitting of more parameters. Multicollinearity is another important consideration when including multiple variables in a model. When variables are highly correlated with each other, it can lead to instability in the parameter estimates and make it difficult to interpret the individual effects of each variable. Moreover, multicollinearity can affect the reliability and precision of the model estimates.

Therefore, it is crucial to carefully select components (variables) that are relevant, informative, and not correlated with each other. A thoughtful component derivation process principal component analysis (PCA) was used to alleviate these challenges. By focusing on the most influential and meaningful components of the data, we can avoid unnecessarily complex models and enhance our understanding of the underlying relationships in the data concerning food security and nutrition in Africa.

Again, we utilized the generalized linear mixed model for a repeated measure to account for the likely correlation between the observed data and any potential variability which is associated with longitudinal data. To better under-stand how food and nutrition are influenced in Africa, this study's goal is to explore these influences (variables) to iden-tify key drivers of food security and nutrition in Africa. Here, we expand on the existing literature by creating a solid framework to look into the correlates of food security and nutrition in Africa.

2 Data

In this section, we provide additional background information about the data used in the study and explain why we chose the specific explanatory factors. The Food and Agriculture Organization (FAO) of the United Nations is responsible for providing data and information to monitor and achieve Sustainable Development Goal 2, which aims to eliminate hunger, food insecurity and malnutrition worldwide. The entire food and nutrition security dataset is available online at the FAO website, along with metadata for Africa that includes variable definitions, sources, data years and units. We handled missing values using the missForest method, which is a type of random forest algorithm [7]. The study period spanned 20 years, specifically from 2000 to 2019, and focused on examining food security and nutrition in Africa. A total of 1080 observations were collected, encompassing data from 54 countries within the continent. The data consisted of 42 indicators (variables) specifically related to various aspects of food security and nutrition as defined by the FAO.

2.1 Response (dependent) variable

Using the FIES, FAO quantifies food insecurity. To ensure cross-country comparability, the FIES Survey Module is administered to samples of the adult population that are nationally representative. National-level results are then calibrated to a global reference scale [1]. Fao *et al.* [14] indicated that severe food insecurity is considerably high in Africa, hence a good indicator for measuring food security and nutrition in a regional context is the metric number of severely food-insecure individuals.

2.1.1 Severely food insecure people

The response metric used is the number of severely food-insecure individuals. Severe food insecurity occurs when people are at serious risk of running out of food, experiencing hunger, and, in the most extreme cases, going days without eating [14].

2.2 Explanatory variables: country features

Data from Africa are gathered by the FAO on a wide range of variables related to food security and nutrition. To identify the crucial variables that correspond to food security and nutrition, Allee *et al.*, [4] reported the usefulness of applying the convergence of evidence strategy across many metrics. The 40 original variables from the FAO data were therefore subjected to Principal Components Analysis (PCA) to prevent information loss and multicollinearity that existed in the data. Finally, ten factors—nutrient intake, average food supply, consumption status, childcare, caloric losses, environment, undernourishment, food or nutritional stability, adequate dietary supply, and newborn feeding practices—were chosen as the explanatory variables which account for 74.6% overall variance in the data.

3 Methods

3.1 Principal Components Analysis (PCA)

One approach to effectively analyze the multitude of variables within the food security domain is through Principal Component Analysis (PCA). PCA enables the summarization and reduction of the variables into a smaller set of uncorrelated components, known as principal components. PCA reduces a large set of variables to a smaller set while retaining the majority of the data from the larger set [20]. The data set's greatest degree of variance is accounted for by the first principle component.

3.2 The PCA procedure

Given that X is a random vector

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

using a matrix of population variance-covariance then the linear combination is

$$Y_1 = \vartheta_{11}X_1 + \vartheta_{12}X_2 + \cdots + \vartheta_{1p}X_p, \tag{1}$$

$$Y_2 = \vartheta_{21}X_1 + \vartheta_{22}X_2 + \cdots + \vartheta_{2p}X_p, \tag{2}$$

$$\vdots$$

$$Y_p = \vartheta_{p1}X_1 + \vartheta_{p2}X_2 + \cdots + \vartheta_{pp}X_p. \tag{3}$$

Equations (1) through (3) can all be considered to be linear regressions predicting predicting Y_i from $X_1, X_2, \ldots X_p$ with no intercept. Whereas the regression coefficients can also be thought of as $\vartheta_{i1}, \vartheta_{i2}, \ldots, \vartheta_{ip}$. Y_i has a population variance

$$var(Y_i) = \sum_{k=1}^{p}\sum_{j}^{p} \vartheta_{ik}\vartheta_{il}\vartheta_{kl} = \vartheta_i' \sum \vartheta_i.$$

Likewise for Y_i and Y_j their population variance is

$$cov(Y_iY_j) = \sum_{k=1}^{p}\sum_{l}^{p} \vartheta_{ik}\vartheta_{il}\vartheta_{kl} = \vartheta_i' \sum \vartheta_j.$$

Storing the coefficients of ϑ_{ij} into a vector

$$\vartheta_i = \begin{pmatrix} \vartheta_{i1} \\ \vartheta_{i2} \\ \vdots \\ \vartheta_{ip} \end{pmatrix}.$$

The linear combination of the x-variables with the largest variation makes up the first principal component (among all linear combinations). The data variance is as fully accounted for as possible. The first Principal component (Y_1) selects $\vartheta_{i1}, \vartheta_{i2}, \ldots, \vartheta_{ip}$ that maximizes $var(Y_i) = \sum_{k=1}^{p} \sum_{j}^{p} \vartheta_{ik}\vartheta_{il}\vartheta_{kl} = \vartheta_i' \sum \vartheta_i$, subject to $\vartheta_1'\vartheta_1 \sum_{j=1}^{p} \vartheta_{1j}^2 = 1$. With the restriction that there is no correlation between the first and second components, as much of the remaining variation as feasible is accounted for by the second principal component, which is also a linear combination of x-variables. This process is repeated until a total of p principal components – equal to the initial number of variables – have been determined.

3.3 Generalized linear mixed model (GLMM) with repeated measures

To analyze the associations between several sets of explanatory variables and the response variable quantitatively, we employed GLMM regression. The longitudinal generalized mixed model employed accounted for the time effect terms for each country, which are correlated (correlated errors) since several observations were made for the same countries from the data described in Section 2. This type of correlation is often modeled using random effects or latent variable models, rather than the time series techniques that are used to address autocorrelation [17].

GLMM combines the strengths of mixed models and generalized linear models to create a potent class of statistical models. In addition to addressing population heterogeneity, this can resolve the modeling issue of data over-dispersion [36]. This makes it a crucial tool for longitudinal (repeated) data analysis, especially in the field of public health. To incorporate both random and fixed effects (hence mixed models), the model has the following general form (written in matrix notation): $y = X\beta + Zy + \varepsilon$. Two random components are present in the Generalized Linear Mixed Model, the "R-side" and the "G-side" random effects. The variance-covariance matrix for random effects is G whereas the R is the variance-covariance matrix of the residual effect (residual effects). With a mean of 0 and a variance of R, the distribution of the errors ε is normal. One provides the columns of the Z matrix and the structure of G when modeling with G-side effects. The covariance structure of the R matrix must be explicitly specified when modeling with R-side effects [8].

3.3.1 Model specification

In this context, we have observed data Y as an $n \times 1$ matrix, along with Y a $q \times 1$ and an $n \times p$ design matrix X for the fixed effect, as well as an $n \times q$ design matrix Z for the random effect. The linear predictor, denoted as η, is the combination of the fixed and random effects without residuals, expressed as $\eta = X\beta + Zy$ from the linear mixed model perspective. The relationship between the distribution of the response variable y and the linear predictor η is established through the concept of the link function. The link function, denoted as $g(\cdot)$, connects the expected value of y, denoted as μ, with the linear predictor η. This is represented by the equation $g(u) = \eta$, where $g(\cdot)$ is a carefully selected one-to-one and differentiable function.

In the context of our marginal (R-side random effects) $g(E[Y]) = g(\mu)\mathbf{X}\beta$. For Poisson data, the variance of the response variable, denoted as $var[Y]$, is equal to its expected value, which is μ. In our model, we represent the variance functions as a diagonal matrix $\mathbf{A}$, then the variance matrix can be expressed as $var[Y] = A^{1/2}RA^{1/2}$.

This formulation accounts for the random variation in the response variable by incorporating the R-side random components through the variance matrix. By considering the appropriate variance functions and covariance structure, we modeled and estimated the variability in the response variable within our marginal model framework.

We employ the probability mass function and the log link function for our count variable. These are $\log(\mu) = \eta$, $Pr(X = k) = \frac{\mu^k e^{-\mu}}{k!}$. Finally, the design matrix X is the 10 components derived from the PCA (Fig. 3).

Since the number of people with severe food insecurity is analyzed using discrete count data, overdispersion was a key notion in our model. Ignoring overdispersion will lead to a type 1 error there making a wrong inference. Altinisik [5] indicated that to properly handle the problem of overdispersion, it is crucial to adopt models that incorporate the proper correlation structure. Because the anticipated covariance structure could be incorrect, the covariance matrix of the parameter estimations should not be based on the model alone. Hence we employed generalized estimating equations (GEE) [22] which results in the marginal model. It uses empirical ("sandwich") estimators to produce results that are resistant to changes in the chosen working covariance structure while accounting for overdispersion [33].

3.3.2 Covariance Structure

The search for the best covariance structure to suit the data is one of the difficulties in repeated measures analysis, but once the random effects are determined one can choose the covariance structure that fits the data. If the model is very complex, it may compromise test power and the effectiveness of testing for fixed effects. Zhang *et al.* [39] highlighted the importance of accurately specifying a covariance model for conducting a valid analysis, even though the true covariance structure is often unknown.

The covariance structure method for repeated measures allows for the incorporation of various covariance structures, including time series and heterogeneous structures, in addition to compound symmetry and unstructured covariance. This flexibility enables researchers to capture the complex relationships and dependencies present in repeated (longitudinal) data more accurately. By considering different covariance structures, we can account for the specific patterns and correlations within the data, leading to more robust and comprehensive statistical analyses. Below are the covariance Structures considered.

AUTOREGRESSIVE (1) The AR(1) structure, also known as the autoregressive structure of order 1, assumes that the variances of measurements are constant and the correlations between measurements decline exponentially with distance. In our case, this implies that the variability in a specific measurement, such as the number of severely food-insecure individuals, remains the same regardless of the time at which it is measured. Additionally, two measurements taken close together in time will have a relatively high correlation (depending on the value of ρ), while measurements that are further apart will have lower correlations. This structure accounts for the dependence between successive measurements and allows us to model the changing correlations over time.

COMPOUND SYMMETRY (CS): The CS structure, or compound symmetry structure, assumes a constant correlation between any two separate measurements, regardless of the time interval between them. In other words, it assumes that the correlation between measurements remains the same irrespective of the distance in time between the repeated measures. This structure implies that there is a consistent level of correlation between time points within subjects, and this correlation is assumed to be uniform across all sets of times. It allows for the modeling of correlated errors within subjects while assuming that the correlation remains constant over time.

TOEPLITZ (TOEP): This structure is similar to the AR(1) structure in that it assumes a correlation pattern between measurements. In the TOEP structure, measurements that are adjacent to each other have the same correlation, measurements that are two-time points apart have the same correlation (which may be different from the first), measurements that are three-time points apart have the same correlation (which may be different from the first two), and so on. However, unlike the AR(1) structure, the correlations in the TOEP structure do not necessarily follow a specific pattern. The TOEP structure allows for more flexibility in capturing the correlation between measurements at different time lags. It is a more general form of correlation structure that encompasses the AR(1) structure as a special case.

UNSTRUCTURED (UN): The UN covariance structure is the most flexible and allows for each pairwise combination of measurements to have a unique correlation. This means that every correlation between any two measurements is estimated separately, resulting in the need to estimate the highest number of parameters compared to other covariance structures. The UN structure requires fitting $t(t+1)/2$ parameters, where t represents the number of time points or measurements in the data. it provides the greatest flexibility in capturing the correlations between measurements, allowing for potentially different correlation patterns across the time points.

AR(1) and TOEP IN HETEROGENEOUS VERSIONS: The heterogeneous versions of the covariance structures mentioned earlier, namely AR(1) and Toeplitz, are the extension that allows for variation in the variances along the diagonal of the covariance matrix. In other words, each measurement can have its own unique variance. This extension acknowledges that the variability of measurements may differ across time points or individuals. By allowing for heterogeneity in the variances, we introduce additional parameters that need to be estimated. Specifically, we estimate a separate variance parameter for each measurement, resulting in an increased number of parameters compared to the homogeneous versions of the covariance structures that were also considered. The TOEP structure allows for heterogeneity in both variances and correlations, while the Heterogeneous Autoregressive of Order 1 structure assumes a constant correlation across all time intervals.

The Compound Symmetry (CS) covariance structure was chosen for our marginal model since it was the only structure that successfully converged, hence the most suitable choice as it provided a reasonable approximation of the correlation between measurements. By using the CS structure, we are accounting for the presence of correlation between measurements taken at different time points within subjects. It allows us to model the dependency between these measurements by assuming a constant correlation throughout the study period, regardless of the time intervals between them. This approach simplified the covariance structure and facilitated the estimation of the model parameters. Our model employed an estimation procedure well known as restricted pseudo-likelihood (RPL) [35]. Figure 1 below gives a summary of the methodology. All methods for our marginal model were programmed in SAS using GLIMMIX.

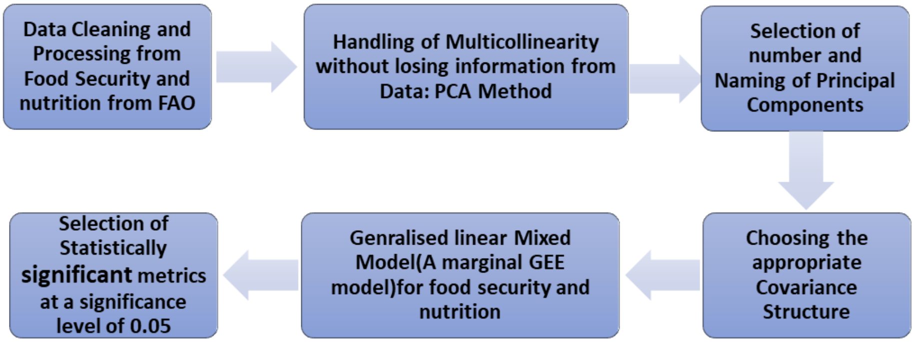

Fig. 1. Flow chart of the methodology.

4 Results

As specified by Li *et al.* [22] it always necessary to check the correlation between repeated measures before any statistical analysis is been done. And so after investigating the correlation structure for our response variable, it was revealed that the repeated observations for the various measures were strongly positively correlated (see Fig. 5 in the appendix). Food security and nutrition data were collected on a large number of variables from countries concerning Africa's single population. To make the data more meaningful our PCA analysis selected the first 10 components that explained 74.6% of the variation within the data (Fig. 1). The results of Barlett's Sphericity test and the Kaiser-Meyer-Olkin values (p. value = 0.00 and KMO = 0.729, respectively) (Table 1) supported the suitability of PCA [6]. An eigenvalue larger than one was considered for principal component extraction [18]. To ensure that each major component provided unique information, a VariMax-orthogonal rotation method was used [18]. Fig. 3 gives a summary of the 10 principal components regarding food security and nutrition in Africa

Ten of the original variables have a substantial correlation with the first principal component (PC1). This component comprises Number of people undernourished, Number of children under 5 years affected by wasting, Percentage of children under 5 years of age who are stunted, Number of children under 5 years of age who are overweight, Number of obese adults (18 years and older), Prevalence of anemia among women of reproductive age (15–49 years), Prevalence of low birth weight, Number of newborns with low birth weight, Minimum dietary energy requirement, and Average dietary energy requirement. We concluded that this principal component (1) which explained 19.26% of the variation is essentially a measure of nutrient intake based on a correlation of 0.9.

Table 1. KMO and Bartlett's test.

Kaiser-Meyer-Olkin		0.729
Bartlett's Test	Chi-Square	54597.015
	df	861
	Sig.	0.000

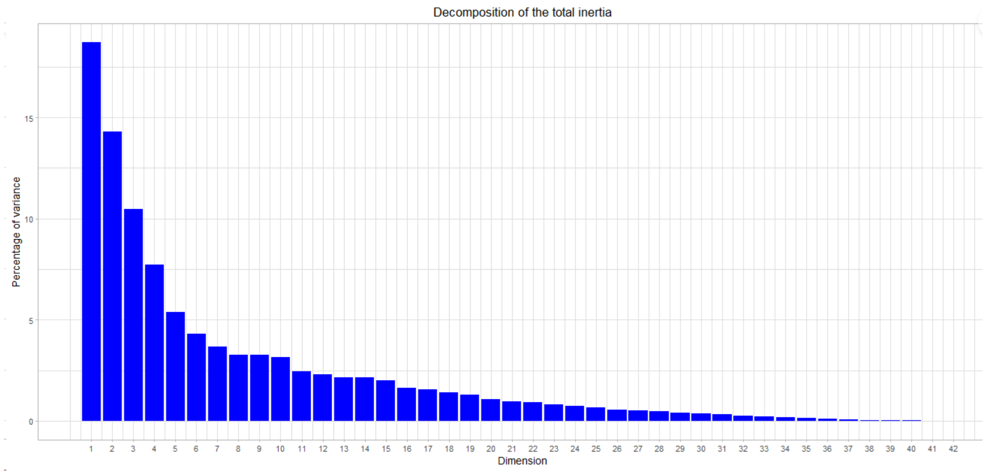

Fig. 2. Distribution of the percentage of variation explained by each component.

The second principal component (PC2) explained 14.96% of the variation that correlates with Average food supply, Average value of food production, Dietary energy supply, Share of dietary energy supply, Average protein supply, Average supply of protein of animal origin, Cereal import dependency ratio, and Average fat supply. With a correlation of 0.9, PC2 measures the Average Food Supplied in Africa

The third principal component (PC3) explained 10.96% of the variation and measures the Consumption Status of Africans based on a correlation of 0.8. PC3 relates mainly to Gross domestic product per capita, Percentage of children under 5 years of age who are stunted, Percentage of children under 5 years of age who are overweight, Prevalence of obesity in the adult population, Minimum dietary energy requirement and Percentage of population using at least basic drinking water services.

Only 5.71% of the total variation was explained by the fourth component (PC4) with the name child care which correlates Average dietary energy requirement, Number of children under 5 years of age who are stunted, and number of women of reproductive age (15–49 years) affected by anaemia (millions) at a correlation coefficient of 0.8. PC5, which is the fifth component, explains 5.51% of the total variation in the data and is correlated with Caloric losses at a measure of 0.6. Similarly, PC6, the sixth component, accounts for 4.47% of the variability and is associated with the environment at a correlation of 0.6. PC7 measures undernourishment and has a correlation coefficient of 0.8, explaining 3.72% of the total variation. PC8, which has a correlation coefficient of 0.7, measures food stability and accounts for 3.43% of the variation. Another component, PC9, accounts for 3.39% of the variability and is associated with Dietary Supply Adequacy with a correlation coefficient of 0.8. Finally, PC10 measures Feeding Practices Among Infants and is correlated with a value of 0.8, explaining 3.19% of the variation in the data.

Generalized linear mixed Poisson analyses results: Stage One

Musunuru *et al.* [24] pointed out that the Pearson Chi-Square/DF ratio which serves as a measure of residual variability should be approximately 1.0 when modeling count data with a Poisson distribution. In contrast, our ratio for the generalized chi-

Fig. 3. Selected principal components concerning Africa.

Table 2. Fit statistics.

-2 Res Log Pseudo-Likelihood	4215.04
Generalized Chi-Square	1969.12
Gener. Chi-Square/DF	1.84

square statistic for our model was 1.84 (Table 2) which gives us an indication that there was over-dispersion in our data. Even though Musunuru *et al.* [24] permitted that a Pearson Chi-Square/DF ratio of more than 2 requires remedial action, we employed the GEE to account for over-dispersion [3] using the Compound Symmetry co-variance structure.

The type III fixed effect test (Table 3) indicates that Nutrients Intake, Average Food Supplied, Child Care, Dietary Supply Adequacy, and Feeding Practices Among Infants were found to be significantly associated with the Rate of Severe Food Insecure Individuals (Food Security and Nutrition) in Africa (Fig. 4). Contrary to expectations, consumption status, caloric losses, environment, undernourishment, and food stability were not associated with food insecurity in Africa (Table 3).

Table 4 gives us the point estimates for our marginal model with its respective expected count associated with nutrient intake, average food supplied, consumption status, child care, caloric losses, environment, undernourishment, food stability, and

Table 3. Type III tests of fixed effects for gee mixed effect model.

Effect	Num DF	F Value	Pr > F
Nutrients Intake	1	16.9	<.0001
Average Food Supplied	1	9.17	0.0025
Consumption Status	1	2.94	0.0866
Child Care	1	24.91	<.0001
Caloric Losses	1	0.09	0.7625
Environment	1	0.65	0.4195
Undernourishment	1	0.33	0.5652
Food Stability	1	0.41	0.5209
Dietary Supply Adequacy	1	33.63	<.0001
Feeding Practices Among Infants	1	5.06	0.0248

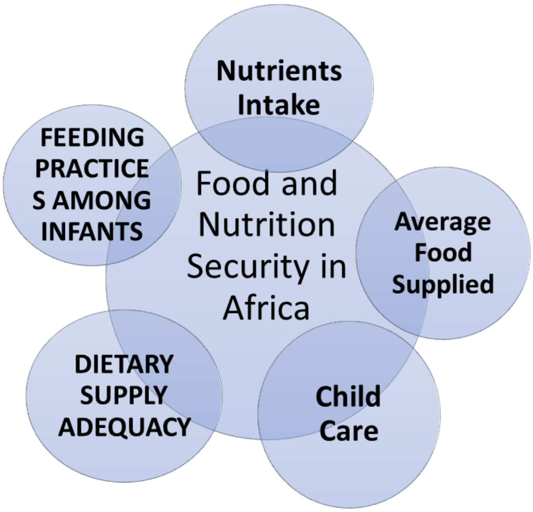

Fig. 4. Significant factors of food security in Africa.

feeding practice among infants. We observed that as nutrient intake increases the average number of severely food insecure individuals changes by a factor of exp(1.106) with a 95% CI of $(1.054, 1.160)$ which is also significant (Table 4). Table 4 indicates a significant negative trend $(\exp(-0.1723) = 0.842)$ for the average food supply, which shows that as the average food supply increases the rate of severely food insecure individuals decreases by just a little above half. As can be seen in Table 4 child care is significantly affecting the rate of severe food insecurity positively by a factor of $\exp(= 0.2652 = 1.304)$ with a 95% CI $(1.175, 1.447)$, thus affecting the growth rate of severe food insecurity by 30.4%. Dietary supply adequacy significantly affects the rate of severe food insecurity by 63.6% by a factor of 1.636 with a 95% CI $(1.385, 1.932)$. As feeding practice among infants increases, the rate of severe food insecurity increases by exp(0.1022) with a 95% CI $(1.013, 1.211)$ with a significant positive trend.

Significant variables validation: stage two (final Model)

In order to ensure the validity of our model, it was important to verify if the exclusion of certain variables that were found to be statistically non-significant in

Table 4. Parameter estimates for effect on severely food insecure individualS using GLMM.

Covariates	Point estimate	SE	p-value	Exp(point estimate)	95% CI		Exp(CI)	
Intercept	−4.226	0.1202	<.0001	0.015	−4.467	3.985	0.011	0.019
Nutrients Intake	0.1006	0.02448	<.0001	1.106	0.053	0.149	1.054	1.16
Average Food Supplied	−0.1723	0.05689	0.0025	0.842	−0.284	0.061	0.753	0.941
Consumption Status	0.1789	0.1043	0.0866	1.196	−0.026	0.384	0.975	1.467
Child Care	0.2652	0.05314	<.0001	1.304	0.161	0.37	1.175	1.447
Caloric Losses	0.001799	0.00595	0.7625	1.002	−0.01	0.013	0.99	1.014
Environment	−0.09333	0.1156	0.4195	0.911	−0.32	0.133	0.726	1.143
Undernourishment	0.0347	0.06031	0.5652	1.035	−0.084	0.153	0.92	1.165
Food Stability	0.0438	0.0682	0.5209	1.045	−0.09	0.178	0.914	1.194
Dietary Supply Adequacy	0.4922	0.08488	<.0001	1.636	0.326	0.659	1.385	1.932
Feeding Practices Among Infants	0.1022	0.04545	0.0248	1.108	0.013	0.191	1.013	1.211

Table 5. Type III tests of fixed Effects for the model (without the variables “ENVIRONMENT” and “UNDERNOURISHMENT”).

Effect	Num DF	F Value	Pr > F
Nutrients Intake	1	14.93	0.0001
Average Food Supplied	1	9.15	0.0025
Consumption Status	1	2.55	0.1105
Child Care	1	20.01	<.0001
CALORIC LOSSES	1	0.00	0.984
Food Stability	1	0.87	0.35
Dietary Supply Adequacy	1	36.73	<.0001
Feeding Practices Among Infants	1	5.86	0.0156

the presented model would also result in their non-significance in other models. For instance, we examined the model without the variables “ENVIRONMENT” and “UNDERNOURISHMENT”. The results of this analysis, as shown in Table 5, indicate that even when these non-significant variables are excluded, the variables “Consumption Status”, “Caloric Losses”, and “Food Stability” remain non-significant in the model. This suggests that these variables do not have a significant effect on the outcome variable, regardless of the inclusion or exclusion of other variables. By conducting this analysis, we can gain further confidence in the robustness and consistency of our findings, as it demonstrates that the non-significant variables do not significantly impact the results of the model, even when other variables are adjusted for.

Table 6 presents the results of the final model, which only includes significant variables (with p-values $\leqslant 0.05$) identified in stage one, to assess if dropping non-significant variables alters the p-values of the remaining variables or renders some of the significant variables non-significant. From the results in Table 6, all five significant variables (Nutrient Intake, Average Food Supplied, Child Care, Dietary Supply Adequacy, and Feeding Practices Among Infants) identified in the initial model (stage one) remained statistically significant even after the non-significant variables were excluded from the model.

Table 6. Type III tests of fixed effects for the final model (only significant variables).

Effect	Num DF	F Value	Pr > F
Nutrients Intake	1	7.31	0.0070
Average Food Supplied	1	9.36	0.0023
Child Care	1	17.63	<.0001
DiEtary Supply Adequacy	1	36.73	<.0001
Feeding Practices Among Infants	1	5.86	0.0156

Table 7. Type III tests of fixed effects for significant variable cross validation (2018 and 2019 data).

Effect	Num DF	F Value	Pr > F
Nutrients Intake	1	5.57	0.0223
Average Food Supplied	1	8.25	0.0060
Child Care	1	48.33	<.0001
DiEtary Supply Adequacy	1	16.73	0.0002
Feeding Practices Among Infants	1	0.09	0.7687

To ensure the robustness and generalizability of our findings, we conducted cross-validation to verify if the five significant determinants of food security identified in the final model remained important in more recent years (validated on 2018 and 2019 data). Table 7 presents the results of this validation process. After validation, Nutrient Intake, Average Food Supplied, Child Care, and Dietary Supply Adequacy were found to remain important in more recent years, whereas Feeding Practice Among Infants was no longer significant

5 Discussion

Researchers often have access to a plethora of explanatory variables related to food security and nutrition, provided by organizations such as FAO. However, selecting the appropriate variables involves technical criteria, and the presence of multicollinearity may restrict the use of a large number of relevant variables, which can hinder obtaining a comprehensive view of food security and nutrition in Africa. Additionally, the response variable is often a count that is repeatedly measured, making it unsuitable to use ordinary least squares (OLS) regression. Hence we used the application of PCA and GLMM to alleviate the problem.

Accordingly, our main important aim was to use PCA to get the relevant component of food security and nutrition concerning Africa holistically and use this identified component to discover the most informative correlates that affect the number of severely food insecure individuals (food security and nutrition) for Africa with its population as an offset, while accounting for the plausible correlation between the repeated measures. The Compound Symmetry was selected based on the convergence criteria. This criterion has been pointed out in previous work as acting like a lack-of-fit test [27, 38]. Again, the PCA method used in this work corroborates previous work [21, 32]. The GEE used to account for overdispersion agrees with previous work (see [2, 10]).

The Poisson GLMM analyses found that nutrient intake significantly increases the rate of severe food insecurity (food insecurity and nutrition) [12] in Africa. The nutrient intake in Africa can be modelled using Number of children under 5 years affected by wasting, Percentage of children under 5 years of age who are stunted, Number of children under 5 years of age who are overweight, Number of obese adults (18 years and older), Prevalence of anemia among women of reproductive age (15–49 years), Prevalence of low birth weight, Number of newborns with low birth weight, Minimum dietary energy requirement, and Average dietary energy requirement. All these listed components give us evidence that nutrient intake has an adverse effect on health which will affect food production in Africa. This finding agrees with previous work [16, 13, 30].

Our marginal model also revealed that the average food supply has a significant effect on food insecurity as the increase in average food supply decreases the rate of severe food insecurity in Africa. This indicates that in Africa we do not meet the average food supply to avoid severe food insecurity [11]. On the contrary, as the average food supply increases the rate of severely food insecure individuals decreases by just a little over half ($\exp(-0.1723) = 0.842$). The components of average food supply are Average value of food production, Dietary energy supply, Share of dietary energy supply, Average protein supply, Average supply of protein of animal origin, Cereal import dependency ratio, and Average fat supply. This finding is in line with what Grote *et al.* [15] and Morales *et al.* [23] reported.

Notably, Child care was also found to be related to the rate of severe food insecurity, which highlights the importance of caring for children in the efforts to significantly reduce food insecurity rates [11]. Additionally, our study found that Dietary supply adequacy was a significant factor in increasing the rate of severely food insecure individuals, with a factor of 1.636 and a 95% CI $(1.385, 1.932)$, representing a 63% rate of increment. Finally, Feeding practice among infants and Dietary supply adequacy were also found to have a positive impact on severe food insecurity in Africa, which is consistent with previous studies [11, 19, 34].

6 Conclusion

Based on our analysis using a marginal model from the generalized linear mixed model, we have identified five significant determinants of food security and nutrition in Africa. These determinants include Nutrient intake, Average food supply, Child care, Dietary supply adequacy, and Feeding practices among infants. We hypothesize that children in Africa are the most vulnerable to severe food insecurity and malnutrition, as evidenced by the fact that two of the most informative factors identified – Feeding practices among infants and Child care – are directly related to children. Nutrient intake, Average food supplied, Child care, and Dietary supply adequacy were found to remain important in more recent years. Furthermore, a better improvement in the average food supply in Africa is likely to yield an improvement in food security and nutrition. Our results provide policymakers with important information that can be used to create targeted plans for Africa aimed at addressing issues related to food security and nutrition and ultimately achieving sustainable development goal 2. We recommend that stakeholders in Africa and its partners allocate more resources to improve the welfare of children and increase the average food supply.

Appendix

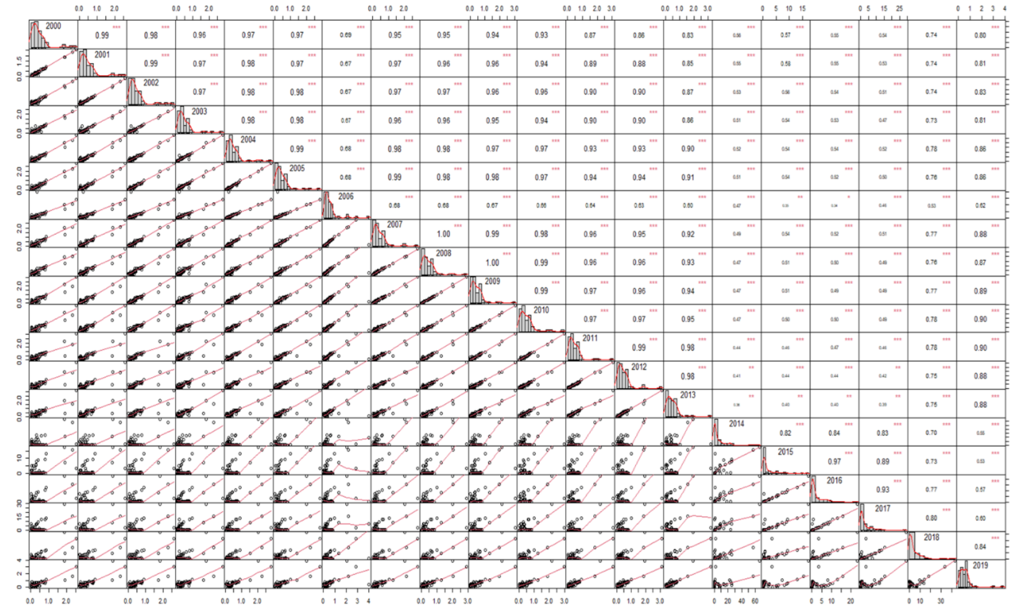

Fig. 5. Correlation between repeated measures for the dependent variable.

Covariance Structures used in this study

AUTOREGRESSIVE(1)

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho^2 & 1 \end{bmatrix}.$$

COMPOUND SYMMETRY

$$\begin{bmatrix} \sigma^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2+\sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2+\sigma_1^2 \end{bmatrix}.$$

TOEPLITZ UNSTRUCTURED

$$\begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 \end{bmatrix}.$$

Conflicts of Interest

The authors declare that they have no conflict of interests.

References

[1] G.S. Afridi, A. Jabbar, S. Khan, N. Akmal. An analysis of food insecurity in Pakistan: Prevalence of Undernourishment (PoU) and Food Insecurity Experience Scale (FIES). *J. Appl. Econ. Bus. Stud.*, **5**(1):175–190, 2021. https://doi.org/10.34260/jaebs.519.

[2] H.T.K. Akdur. Gee-based bell model for longitudinal count outcomes. *Commun. Stat. Theory Methods*, **51**(13):4602–4616, 2022. https://doi.org/10.1080/03610926.2022.2056751.

[3] S. Aliabadi, P. Anyanwu, E. Beech, E. Jauneikaite, P. Wilson, R. Hope, A. Majeed, B. Muller-Pebody, C. Costelloe. Effect of antibiotic stewardship interventions in primary care on antimicrobial resistance of Escherichia coli bacteraemia in England (2013–18): a quasi-experimental, ecological, data linkage study. *Lancet Infect. Dis.*, **21**(12):1689–1700, 2021. https://doi.org/10.1016/S1473-3099(21)00069-4.

[4] A. Allee, L.R. Lynd, V. Vaze. Cross-national analysis of food security drivers: comparing results based on the Food Insecurity Experience Scale and Global Food Security Index. *Food Secur.*, **13**(5):1245–1261, 2021.

[5] Y. Altinisik. Addressing overdispersion and zero-inflation for clustered count data via new multilevel heterogenous hurdle models. *J. Appl. Stat.*, **50**(2):408–433, 2022.

[6] M.S. Bartlett. A note on the multiplying factors for various χ^2 approximations. *J. R. Stat. Soc., B: Stat. Methodol.*, **16**(2):296–298, 1954.

[7] A. Bofa, T. Zewotir. Filling the gap in food and nutrition security data: what imputation method is best for Africa's food and nutrition security? *Lith. J. Stat.*, **61**:16–31, 2022.

[8] W. Chen, M.G. Genton, Y. Sun. Space-time covariance structures and models. *Annu. Rev. Stat. Appl.*, **8**:191–215, 2021. https://doi.org/10.1146/annurev-statistics-042720-115603.

[9] N.O. Co. *A Discourse on African Food Insecurity*. 2021. https://doi.org/10.2139/ssrn.3841573.

[10] R. Coelho, P. Infante, M.N. Santos. Comparing GLM, GLMM, and GEE modeling approaches for catch rates of bycatch species: a case study of blue shark fisheries in the South Atlantic. *Fish. Oceanogr.*, **29**(2):169–184, 2020. https://doi.org/10.1111/fog.12462.

[11] M.K. Dodo. Understanding africa's food security challenges. In B. Mahmoud(Ed.), *Food Security in Africa*, 2020.

[12] H.A. Eicher-Miller. A review of the food security, diet and health outcomes of food pantry clients and the potential for their improvement through food pantry interventions in the United States. *Fish. Oceanogr.*, **220**:112871, 2020. https://doi.org/10.1016/j.physbeh.2020.112871.

[13] FAO, IFAD, UNICEF, WFP, WHO. The State of Food Security and Nutrition in the World 2020. Transforming food systems for affordable healthy diets, 2020.

[14] FAO and ECA and AUC. *Africa Regional Overview of Food Security and Nutrition 2019*. Accra, Ghana, 2020. ISBN 978-92-5-132051-8.

[15] U. Grote, A. Fasse, T.T. Nguyen, O. Erenstein. Food security and the dynamics of wheat and maize value chains in Africa and Asia. *Front. Sustain. Food Syst.*, **4**:617009, 2021. https://doi.org/10.3389/fsufs.2020.617009.

[16] S. Gupta, N. Sunder, P.L. Pingali. Market access, production diversity, and diet diversity: evidence from India. *Food Nut. Bull.*, **41**(2):167–185, 2020.

[17] A.R. Ives. Random errors are neither: on the interpretation of correlated data. *Methods Ecol. Evol.*, **13**(10):2092–2105, 2022.

[18] H.F. Kaiser. The application of electronic computers to factor analysis. *Educ. Psychol. Meas.*, **20**:141–151, 1960.

[19] M.A. Kamenya, S.L. Hendriks, C. Gandidzanwa, J. Ulimwengu, S. Odjo. Public agriculture investment and food security in ecowas. *Food Policy*, **113**:102349, 2022.

[20] F. Kherif, A. Latypova. Principal component analysis. In *Mach. Learn.*, pp. 209–225. Elsevier, 2020.

[21] W. Li, Y. an Quan, J. Wang, Z. He, F. Qin, Z. Wang, M. Zeng, J. Chen. Effects of ten vegetable oils on heterocyclic amine profiles in roasted beef patties using UPLC-MS/MS combined with principal component analysis. *Food Chem.*, **347**:128996, 2021.

[22] K.-Y. Liang, S.L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, **73**(1):13–22, 1986.

[23] D.X. Morales, S.A. Morales, T.F. Beltran. Racial/ethnic disparities in household food insecurity during the covid-19 pandemic: a nationally representative study. *J. Racial Ethn. Health*, **8**(5):1300–1314, 2021. https://doi.org/10.1007/s40615-020-00892-7.

[24] A. Musunuru, D. Proffitt, R. Ewing, W.H. Greene. Poisson and negative binomial regression analysis. In *Advanced Quantitative Research Methods for Urban Planners*, pp. 74–94. Routledge, 2020.

[25] C.F. Nicholson, E.C. Stephens, A.D. Jones, B. Kopainsky, D. Parsons, J. Garrett. Food security outcomes in agricultural systems models: current status and recommended improvements. *Agric. Syst.*, **188**:103028, 2021. https://doi.org/10.1016/j.agsy.2020.103028.

[26] N. Omidvar, D. Ahmadi, K. Sinclair, H. Melgar-Quiñonez. Food security in selected Middle East and North Africa (MENA) countries: an inter-country comparison. *Food Secur.*, **11**:531–540, 2019. https://doi.org/10.1007/s12571-019-00935-w.

[27] V. Pinto, R. Sooriyarachchi. Comparison of methods of estimation for a goodness of fit test – an analytical and simulation study. *J. Stat. Comput. Sim.*, **91**(9):1846–1866, 2021. https://doi.org/10.1080/00949655.2021.1872078.

[28] A. Rahaman, A. Kumari, X.-A. Zeng, I. Khalifa, M.A. Farooq, N. Singh, S. Ali, M. Alee, R.M. Aadil. The increasing hunger concern and current need in the development of sustainable food security in the developing countries. *Trends Food Sci. Technol.*, **113**:423–429, 2021. https://doi.org/10.1016/j.tifs.2021.04.048.

[29] D. Raheem, M. Dayoub, R. Birech, A. Nakiyemba. The contribution of cereal grains to food security and sustainability in Africa: potential application of UAV in Ghana, Nigeria, Uganda, and Namibia. *Urban Sci.*, **5**(1):8, 2021. https://doi.org/10.3390/urbansci5010008.

[30] J. Sarfo, E. Pawelzik, G.B. Keding. Dietary patterns as characterized by food processing levels and their association with the health outcomes of rural women in East Africa. *Nutrients*, **13**(8):2866, 2021. https://doi.org/10.3390/nu13082866.

[31] M.D. Smith, B. Meade. Who are the World's Food Insecure? Identifying the risk factors of Food Insecurity around the World. *Amber Waves: The Economics of Food, Farming, Natural Resources, and Rural America*, 2019.

[32] M.S. Sulaiman, M.M. Abood, S.K. Sinnakaudan, M.R. Shukor, G.Q. You, X.Z. Chung. Assessing and solving multicollinearity in sediment transport prediction models using principal component analysis. *ISH J. Hydraul. Eng.*, **27**(sup1):343–353, 2021. https://doi.org/10.1080/09715010.2019.1653799.

[33] J. Thompson, K. Hemming, A. Forbes, K. Fielding, R. Hayes. Comparison of small-sample standard-error corrections for generalised estimating equations in stepped wedge cluster randomised trials with a binary outcome: a simulation study. *J. Stat. Comput. Simul.*, **30**(2):425–439, 2021. https://doi.org/10.1177/0962280220958735.

[34] K. Waha, F. Accatino, C. Godde, C. Rigolot, J. Bogard, J.P. Domingues, E. Gotor, M. Herrero, G. Martin, D. Mason-D'Croz. The benefits and trade-offs of agricultural diversity for food security in low-and middle-income countries: a review of existing knowledge and evidence. *Glob. Food Sec.*, **33**:100645, 2022.

[35] R. Wolfinger, M. O'connell. Generalized linear mixed models a pseudo-likelihood approach. *J. Stat. Comput. Simul.*, **48**(3–4):233–243, 1993. https://doi.org/10.1080/00949659308811554.

[36] P. Ye, X. Qiao, W. Tang, C. Wang, H. He. Testing latent class of subjects with structural zeros in negative binomial models with applications to gut microbiome data. *Stat. Methods Med. Res.*, **31**(11):2237–2254, 2022. https://doi.org/10.1177/09622802221115881.

[37] I.A. Yunusa, A. Zerihun, M.R. Gibberd. Analysis of the nexus between population, water resources and global food security highlights significance of governance and research investments and policy priorities. *J. Sci. Food Agric.*, **98**(15):5764–5775, 2018. https://doi.org/10.1002/jsfa.9126.

[38] J. Zhang, J. Ding, Y. Yang. Is a classification procedure good enough? — A goodness-of-fit assessment tool for classification learning. *J. Am. Stat. Assoc.*, **118**(542):1115–1125, 2021. https://doi.org/10.1080/01621459.2021.1979010.

[39] M.F. Zhang, J.F. Dawson, R.B. Kline. Evaluating the use of covariance-based structural equation modelling with reflective measurement in organizational and management research: a review and recommendations for best practice. *Br. J. Manag.*, **32**(2):257–272, 2021. https://doi.org/10.1111/1467-8551.12415.

REZIUMĖ

Didelį maisto stygių Afrikoje nusakantys veiksniai naudojant ilgalaikį Puasono apibendrintąjį mišrųjį modelį

A. Bofa, T. Zewotir

Maisto stygius yra daugialypė problema (iššūkis), turinti įtakos sveikatos priežiūrai, politikai, žemės ūkio produkcijos lyderystei, aplinkai, maisto sistemai ir pasaulinės prekybos maisto pramonėje politikai. Mūsų tikslas buvo nustatyti svarbius aprūpinimo maistu ir mitybos veiksnius, kompleksiškai susijusius su Afrika, ir panaudoti šiuos nustatytus veiksnius, kad atrastume informatyviausius koreliatus, kurie įtakoja badaujančiųjų (maisto nepriteklių patiriančių asmenų) skaičių Afrikoje ir jos gyventojus kaip atsvarą. Pagrindinių komponenčių analizė (PKA) buvo naudojama siekiant surasti minėtus svarbius Afrikos aprūpinimo maistu ir mitybos veiksnius. Reikšmingiems veiksniams nustatyti buvo naudojamas Puasono apibendrintasis tiesinis mišrusis modelis (ATMM). Tada buvo taikomos apibendrintosios įvertinimo lygtys, kad būtų atsižvelgta į perteklinę dispersiją, susijusią su Puasono skirstiniu. Kad rezultatų interpretacija būtų prasmingesnė, buvo pasirinktos 10 PKA komponenčių. Jos paaiškino 74,6% duomenų variacijos. ATMM analizė parodė, kad maistinių medžiagų suvartojimas, vidutinis tiekiamas maistas, vaikų priežiūra, mitybos tiekimo pakankamumas ir kūdikių maitinimo praktika yra statistiškai reikšmingai susiję su maisto nepriteklių patiriančių (badaujančių) asmenų skaičiumi (p reikšmė $< 0{,}05$). Tikėtina, kad pagerinus vidutinę maisto pasiūlą Afrikoje, pagerės aprūpinimas maistu ir mityba. Mūsų išvadose pateikiama įžvalgų apie Afriką, kurios padės politikos formuotojams parengti Afrikai skirtus tikslinius planus, kuriuose bus sprendžiami aprūpinimo maistu ir mitybos klausimai, o tai padės siekti 2-ojo darnaus vystymosi tikslo.

Raktiniai žodžiai: pagrindinių komponenčių analizė; Puasono apibendrintasis tiesinis mišrusis modelis; aprūpinimas maistu; mityba