
Logit models for household food insecurity classification

Abraham Yeyo Owino^{1,2,*}, Leonard Kibojana Atuhaire^{1,2}, Ronald Wesonga^{1,2},
Fabian Nabugoomu^{1,2}, Elijah Muwanga-Zaake^{1,2}

¹Department of Statistics and Actuarial Sciences, College of Business and Management Sciences, Kampala, Uganda

²Makerere University, Kampala, Uganda

Email address:

ayowino@yahoo.co.uk (A. Y. Owino), latuhaire@isae.mak.ac.ug (L. K. Atuhaire), wesonga@wesonga.com (R. Wesonga),
fnabugoomu@yahoo.com (F. Nabugoomu), muwangazake@hotmail.com (E. M. Zaake)

To cite this article:

Abraham Yeyo Owino, Leonard Kibojana Atuhaire, Ronald Wesonga, Fabian Nabugoomu, Elijah Muwanga-Zaake. Logit Models for Household Food Insecurity Classification. *American Journal of Theoretical and Applied Statistics*. Vol. 3, No. 2, 2014, pp. 49-54.
doi: 10.11648/j.ajtas.20140302.14

Abstract: Micro-level measurement of food insecurity is a necessary approach towards a more feasible solution to the global problem for proper classification of households by food insecurity status. Measurement of food insecurity is a challenge because it is a multi-faceted latent and continuous phenomenon explained by a wide range of both quantitative and qualitative variables. In this paper, we examined the quantitative variables and applied exploratory factor analysis to identify which of them significantly influence household food insecurity. Logit models were then developed using the variables identified. Further, empirical data obtained from Tororo and Busia rural households in Uganda were used to fit the models. Four logit models based on four scenarios were developed and compared. The key findings pointed to the fact that if households were to be correctly analyzed and classified into the right food security category, a hybrid dependent variable that represents as many aspects of food insecurity as possible should be used. The model correctly classified 90 % of the combined households for two districts. However, when fitted for separate districts, it was established that 99% of households in Busia and 96% in Tororo district respectively, were found to be food insecure.

Keywords: Hybrid Dependent Variable, Latent Variable, Factor Analysis, Logit Model

1. Introduction

Food insecurity is a situation in which individuals do not have physical or economic access to the nourishment they need and no access to cash or resources for producing food. Measurement of food insecurity is a challenge because it is a multi-faceted latent and continuous phenomenon explained by a wide range of both quantitative and qualitative variables. Poverty and food insecurity are strong correlates[1] which are experienced by a significant proportion of the Ugandan population. Government efforts to ensure food security and adequate nutrition for all people[2] in Uganda meets a number of challenges because food insecurity is a multi-faceted latent phenomenon explained by many different variables which contribute to the phenomenon's different proportions of influence. This paper proposes a quantitative method for measurement of food security that incorporate many measurable variables of all food security components – food availability, food access, food utilization and coping strategies as possible. A list of

indicator variables for each of the components was therefore generated and analyzed [3, 4].

2. Materials and Methodology

The study area was Tororo and Busia districts in Eastern Uganda which largely survive on crop farming, animal farming and fishing for those who live near rivers or by the shores of Lake Victoria. They experience two rainy seasons March/April and August/October. Typical food crops grown are millet, sorghum, maize, sweet potatoes, cassava, rice, beans, and ground nuts. Fruits are not common and are grown by a few households and they include Jackfruit, oranges, passion fruits and some sweet bananas. The soils are generally poor in most parts of the districts as they have been over used and yields are low unless manure is applied[5, 6].

A two-stage stratified sampling design was used where districts were strata, first stage units were local council ones (LC1's) and second stage units were households[7]. The overall sample size was 1332 households with a permissible error $\ell = 8\%$ for the district and 4% for overall estimates.

The design effect used to cater for clustering was $D = 2$ and the non response rate was assumed to be 10%. Personal interviews using structured questionnaires were used to collect primary data from household heads or their spouses. Focus group discussions and Key informants interviews were conducted to validate information on prices of food crops, local measures used for crop sales like basket sizes, tin sizes and their corresponding prices, existing markets and the food situation in the area. Information on quantities of food and non food items were recorded in local measures then converted to standard measures of shillings, kilograms and liters before analysis. EPIDATA software application program was used for data entry and cleaning.

The first phase of the analysis was identification of variables that significantly influence food availability, food access, food utilization and coping strategies[8]. Then using exploratory factor analysis, we determined the factors that each of the identified variables was uniquely associated with. The main purpose of using factor analysis was data reduction. Factor extraction was done using Principle Component Analysis technique. Exploratory factor analysis was used instead of confirmatory factor analysis because there was no prior information available on the data [9]. In this study, 43 explanatory variables of food insecurity were analyzed using exploratory factor analysis after placing them into three major groupings: 10 for food availability, 22 for food access/coping strategy variables and 11 for food utilization. The data set used was 2009 food security data collected from 1175 households. The 43 variables were reduced to 26 (60%), 5 for food availability, 14 for food access/coping strategies and 7 for food utilization. The 26 variables distinctly associated with 7 factors. The variables were then used to develop logit models for classification of household food security.

Logistic regression analysis was initially done separately for the four variable categories – food availability, food access, food utilization and coping strategy but this did not yield plausible results hence the decision to combine all the variables. All except total land area were categorical variables. Logistic Regression was then used for analysis to correctly predict the category of insecurity for a household.

Model Estimation:

The independent variables X_1, X_2, \dots, X_n can take on any form since logistic regression makes no assumptions about the distribution of the independent variables. The dependent variable Y is dichotomous taking on a value of 1 with a probability of success θ or 0 with a probability of failure $(1 - \theta)$. The relationship between the dependent variable and the independent variables is a logit transformation of θ given as:

$$Y = \text{Logit}[\theta(x)] = \log \left[\frac{\theta(x)}{1 - \theta(x)} \right] \\ = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i + \dots + \beta_k X_k \quad (1)$$

$$\text{Where: } \theta = \frac{e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}{1 + e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}$$

β_i = the coefficient of the i^{th} independent variable; for $i = 1, 2, \dots, k$

α = the constant of the equation

Food security status was taken as a dichotomous dependent variable with a household being either food secure or food insecure. The predictor variables as obtained through factor analysis were mixed – continuous and categorical. A logistic regression model was constructed using the predictor variables with Food security status (Y_f) as a logit function. The model is expressed as:

$$Y_f = \alpha + \sum_{i=1}^k \beta_i X_i + \varepsilon \quad (2)$$

Where: Y_f = A Logit function – Food security status; α was the constant – food security status when the effects of the independent variables are zero.

$B_i, i = 1, \dots, k$ Coefficients

$X_i, i = 1, \dots, k$ Independent variables

ε = The error term

The model was built stepwise using both forward linear regression and backward linear regression procedures. Testing coefficients of independent variables was done using the Wald Test Statistic and the Likelihood Ratio Test Statistic. For the Wald Test, when the p-value of the Wald statistic was less or equal to 0.05, the variable was considered significant and was retained in the model. When p-value was higher than 0.05, it implied the variable did not make a significant contribution to the model and was therefore excluded from the model. The Wald test is usually more reliable when samples are sufficiently large. It was used in this study as there was no danger of bias given the big sample of 1175 households used. The Likelihood Ratio test yields more reliable estimates than the Wald test when samples are small. It is obtained as the ratio of the maximized value of the full model likelihood function L_1 to the maximized value of the simple model likelihood function L_0 .

The likelihood ratio-test statistic $= -2 \log \left(\frac{L_0}{L_1} \right)$

$$= -2[\log(L_0) - \log(L_1)] = -2[L_0 - L_1] \quad (3).$$

This log transformation yields a χ^2 statistic which is preferred in backward stepwise regression analysis. When the p-value of the likelihood-ratio test statistic was less or equal to 0.05, the variable was considered significant and retained in the model. When the p-value was higher than 0.05, it implied the variable did not make a significant contribution to the model and was therefore excluded from the model.

The Hosmer-Lemshow test statistic was used to assess the

goodness of fit of the model to the data. The null hypothesis here was that there was no difference between the predicted values using the model and the actual values of the dependent variable. If the p-value of the Hosmer and Lemeshow goodness-of-fit statistic was less or equal to 0.05, we rejected the null hypothesis. If it was greater than 0.05, we failed to reject the null hypothesis and concluded that the model estimates did fit the data well and explained much of the variance in the dependent variable. The higher the value of the test statistic the better the model fit.

3. Results

3.1. Descriptive Statistics

A set of 10 food availability variables, 22 food access and coping strategy variables, and 11 food utilization proxy variables were identified[10]. The specific variables in each category can be found in the Paper by [3].

3.2. Summary of Results on Correlations, Determinants, Singularity and Multi-Collinearity

The variables were subjected to a series of tests. For the variables that were eventually selected for factor analysis, the determinants of the R-matrices were all greater than the threshold of 0.00001 implying there was no multi-collinearity. There were no correlations greater than 0.9 which meant there was no problem of singularity. Bartlett's tests were all significant implying original matrices were not identity matrices and implying relationships existed between the retained variables. A summary of the test results can be found in Table 1 in the paper by [3]. The variables that passed the tests were

subjected to factor analysis with the following results. Food Availability Factors: Factor 1 was *land area and food harvested* (variables included: Total land owned, total area cultivated and total food harvested). Factor 2 was *Food Production Trends and Expectations* (variables included: Production trends in past 5 years and whether production expectation was achieved). Food Access/Coping Mechanisms Factors: Factor 1 was *Inability to access food* (variables included: Worried food would run out, food did not last, running out of food and money, did not have balanced meals. Factor 2 was *coping by reducing amount of food eaten* (variables included: Cutting meal size, eating less, hungry but not eating). Factor 3 was *coping by skipping or missing meals* (variables included: Not eating whole day children, adults, Skipping meals and weight loss). Food Utilization Factors: Factor 1 was Poor, cheap low quality or less foods (variables included: No balanced meals, Frequency of meals, eating less, losing weight, low cost foods). Factor 2 was commonly eaten foods for lunch or supper (variables included: Lunch, Supper for below 5 years and Lunch, Supper for household members above 5 years).

3.3. Logistic Regression Results

Three scenarios were considered. In the first, difficulty accessing food was taken as the proxy variable for food security. In the second, food stored was taken to be the food security proxy variable and in the third, food harvested was used. The aim here was to establish which one was more representative and best in correctly classifying the households as food insecurity status. Scenario 1: Difficulty having food access

Table 1. Model: Dependent – Difficulty having food access.

Variable	B	S.E.	Wald	df	Sig.
Children skipping meals because there wasn't enough	-1.28	0.206	38.45	1	0.00
Worried whether our food would run out	-0.69	0.259	7.03	1	0.01
Commonly eaten Lunch for household members over 5 years	2.91	1.32	4.89	1	0.03
Food production trends for last 5 years.	-1.28	0.58	4.82	1	0.03
Couldn't feed children a balanced meal	0.54	0.27	3.92	1	0.05
Children ever not eaten for a whole day?	-0.45	0.23	3.90	1	0.05

Scenario 2: Food Storage as dependent for Food Security
Variables in the model are shown in Table 2. The Hosmer and Lemeshow chi-square was 2.75 with p-value = 0.949 showing the model passes the goodness-of-Fit test. Those

who did not store food were 55% and those who stored food were 45%. The model correctly classified 78.1% of the households.

Table 2. Model: Food Stored as Dependent variable.

Variable	B	S.E.	Wald	df	Sig.
Children skipping meals	-1.07	0.25	17.8	1	0.00
Relying on few kinds of low-cost food to feed our children	-1.32	0.41	10.18	1	0.00
Losing weight	0.69	0.23	8.86	1	0.00
Ever cut the size of children's meals	1.17	0.52	5.05	1	0.03
Couldn't feed our children a balanced meal	0.78	0.39	3.94	1	0.05

Scenario 3: Food Harvested as Dependent variable
Food harvested was used as the dependent variable with households having harvested no food or food worth less than

1,000 Uganda shillings (40 US Cents).

Variables in the model are shown in Table 3. The Hosmer and Lemeshow chi-square was 2.269 with p-value = 0.972

showing the model passes the Goodness-of-Fit test. Those who reported harvesting food worth less than 40,000 shillings (US\$16) were 70% and those who harvested more food were only 30%. This model correctly classified 82.1% of the households. Therefore the third model based on food

harvested as the dependent variable classified more households (82.1%) correctly compared to the model with difficulty accessing food (75.8%) and the one with food storage (78.1%).

Table 3. Model: Harvest as dependent variable.

Variable	B	S.E.	Wald	df	Sig.
Total area cultivated	0.29	0.07	16.35	1	0.00
Difficulty accessing food	-0.96	0.22	18.92	1	0.00
Total food stored	-1.60	0.21	60.56	1	0.00
Total land area	0.35	0.13	7.01	1	0.01
Ever not eaten for a whole day	-0.68	0.26	7.07	1	0.01
Frequency of borrowing from friends or relatives?	-1.83	0.70	6.83	1	0.01
Total combined Family income category for past one month?			21.64	9	0.01

Table 4. Summary Table for the three Models.

Variable	Model: Dependent – Difficulty having food access.	Model: Dependent Total Food Harvested	Model: Dependent - Whether or not Food is Stored
Commonly eaten Lunch for household members over 5 years	AAAAAAAAAAAAAAAA		
Children skipping meals	AAAAAAAAAAAAAAAA		SSSSSSSSSSSSSSSSSS
Couldn't feed children a balanced meal	AAAAAAAAAAAAAAAA		SSSSSSSSSSSSSSSSSS
Food production trends for last 5 years.	AAAAAAAAAAAAAAAA		
Worried whether our food would run out	AAAAAAAAAAAAAAAA		
Children ever not eaten for a whole day?	AAAAAAAAAAAAAAAA		
Ever not eaten for a whole day		HHHHHHHHHHHHHHHH	
Difficulty accessing food		HHHHHHHHHHHHHHHH	
Frequency of borrowing from friends or relatives?		HHHHHHHHHHHHHHHH	
Total combined Family income category for past one month?		HHHHHHHHHHHHHHHH	
Total land area		HHHHHHHHHHHHHHHH	
Total area cultivated		HHHHHHHHHHHHHHHH	
Total food stored		HHHHHHHHHHHHHHHH	
Losing weight			SSSSSSSSSSSSSSSSSS
Ever cut the size of children's meals			SSSSSSSSSSSSSSSSSS
Relying on few kinds of low-cost food to feed our children			SSSSSSSSSSSSSSSSSS

Table 4 shows that there is hardly an overlap of variables for the three models except for 2 variables “*Children skipping meals*” and “*Couldn't feed children a balanced meal*” which were found in the first model with “difficulty having food access” as dependent variable and third model with “whether or not food is stored” as dependent variable. This implies that although each of the variables is important in explanation of food insecurity, the models discriminate on which ones they consider significant in explaining a given food insecurity component.

A more wholesome picture of food security which takes into consideration all the variables in the three models is therefore one with a hybrid dependent variable with the explanatory variables of all three models combined. Therefore, this paper proposes an additive model as explained in scenario 4.

Scenario 4: Creating a hybrid dependent variable

The three variables used as dependent variables in the earlier models each reflected an aspect of food security. Table 4 shows that there was hardly any overlap between independent variables of the three models and yet they all assess some aspect of food security. In order to bring together the effect of all the independent variables of the three models, an additive dependent variable combining the three dependent variables was obtained. The rationale here was that, in reality if one was food insecure, it would mean he or she had difficulty accessing food, he/she may not have had adequate food stored or may not have had a good harvest. Combining the three aspects of access difficulty, storage and food harvest to come up with a hybrid dependent variable catered for the multi-faceted nature of food security.

Table 5. Compound Food Security Dependent Variable Model with all Households.

Variables in the Equation	B	S.E.	Wald	Df	Sig.
You or other adults in your household) ever get food or borrow money for food from friends or relatives?	1.77	0.50	12.54	1.00	0.00
We worried whether our food would run out before we got money to buy more.	-2.64	0.92	8.19	1.00	0.00
Did any of the children ever skip meals because there wasn't enough food or money for food?	-1.41	0.52	7.50	1.00	0.01
Which category represents on average the total combined income of all members of this Family during the	0.22	0.09	5.96	1.00	0.01

Variables in the Equation	B	S.E.	Wald	Df	Sig.
past one month					
Did you ever cut the size of (your child's/any of the children's) meals because there wasn't enough food or money for food?	4.04	1.67	5.87	1.00	0.02
Did you get less or more than expected or as expected from your crops	1.66	0.77	4.62	1.00	0.03

A condition was therefore created where a food secure household was defined as one who either had adequate food harvested; food stored or had no difficulty accessing food. A household would be considered food insecure if it had no food stored, had little or no harvest or had difficulty accessing food. Variables in the model are presented in Table 5. The Hosmer and Lemeshow chi-square was 3.574 with p-value of 0.893 showing that the model passes the goodness-of-fit test. Those who were food insecure according to the combined model were 83.7% and those who were not food insecure were 16.3%. The model correctly classified 89.7% of the households.

The analysis using the hybrid dependent variable was also

done for the separate district datasets. It was established from the analysis that 99.2% of Busia households were found to be food insecure while for Tororo, 96.2% of the households were food insecure.

The model correctly classified 91.5% of the Busia households. Variables in the model are shown in Table 6. When the model was applied to the Tororo dataset, it correctly classified 94.3% of the households. Whereas 89.7% of the households were correctly classified in the combined dataset, when the datasets were separated by districts, higher percentages of households (91.5%) were correctly classified for Busia and 94.3% for Tororo districts.

Table 6. Logit Model for Busia.

Variables in the Equation	B	S.E.	Wald	df	Sig.
Food production trends in your garden over the last five years	-1.58	0.80	3.91	1	0.05
How many meals do you normally have in your home	-3.18	1.42	5.02	1	0.03
Breakfast for children over 5 years and Adults	1.64	0.64	6.54	1	0.01
Did any of the children ever skip meals because there wasn't enough food or money for food	-0.85	0.37	5.36	1	0.02
Was your child/ were the children ever hungry but you just couldn't afford more food	4.50	1.85	5.91	1	0.02
We worried whether our food would run out before we got money to buy more	-2.11	0.52	16.3	1	0.00
Food we harvested or bought just didn't last, and we didn't have money to get more	1.03	0.48	4.68	1	0.03
Which category represents on average the total combined income of all members of this Family during the past one month	3.09	1.06	8.57	1	0.00
What are your major sources of livelihood/income	-4.51	1.71	6.95	1	0.01
Borrowing/Lending arrangements are you engaged with	2.68	0.87	9.6	1	0.00
Total area cultivated	0.	0.09	10.80	1	0.00

5. Discussion

Food security as already mentioned is a latent variable which can only be detected by use of observable variables. Food security is also a multi-faceted variable having different dimensions. Major components commonly agreed on are food availability, food access, food utilization and stability. A number of researches have considered food security as represented by one of these components while some have attempted to combine them by including in the food security analysis independent variables from all these components. The analysis carried out in this study took the later approach initially combining variables of food availability, food access, food utilization as explanatory variables. Statistical tests showed that between 76%, 78% and 82% of households were correctly classified in the process. The dependent variables earlier used separately in three models, namely, difficulty accessing food, harvested food and storage were combined to form a new variable. The resultant dependent variable yielded an improved classification of households. About 90% to 94% of the households could be correctly classified using the new variable depending on which dataset one used. It also

showed that more households were food insecure as compared to the other three separate models. There were more common predictors of food insecurity between the combined model and the model for separate districts. For instance there were three common predictors between Busia model and combined model; Worried whether our food would run out before we got money to buy more, Some of the children ever skipped meals because there was not enough food or money for food and Which category represents on average the total combined income of all members of this Family during the past one month. This is in contrast to Table 4 that shows hardly any commonalities between predictors in the food production/harvested, difficulty accessing food and food storage models.

Therefore, if households are to be correctly analyzed and classified into the right category of food security, it is important to have a hybrid dependent variable that represents as many aspects of food insecurity as possible.

The focus group discussions also revealed some two important factors that were not factored in the model but could have also influenced the food security situation. The first one was the sale of food to Southern Sudanese and Kenyans who came to homes and booked the food before harvest or paid for the gardens and collected all the food

once they were ready for harvest. This explains why nearly all the respondents felt food insecure and yet there seemed to be evidence of reasonable volume of production or reasonable area cultivated. The second factor was cultural influence on farming and feeding practices. From the descriptive statistics, most households had at least 3 to 4 acres of land and would have been expected to grow a variety of food crops, vegetables and fruits to feed their families. It was however observed that fruits were not grown by most households and the common crops that were grown in bulk included; cassava, millet and sorghum. Vegetables were also not grown by a substantial number of households or they were not considered crops to be grown. Culturally, the staple food for Busia and Tororo are Cassava and Millet. As far as the communities are concerned, if you do not have enough millet and cassava, you do not have food. It is possible therefore that people felt food insecure and said so even when they had other foods like Maize, rice, ground nuts and others.

It is therefore possible for a household to have enough food to live on but still feel food insecure. It is also possible for a household to produce enough food but use the food to obtain what they consider more important than food for consumption, for example, medical care for a household member or payment of school fees for a child [11]. Cultural mindset, attitude or exposures are also other factors that could significantly influence household food security [12]. Two households with the same resources and capacity may differ in food security status just because of perceptions, attitudes or mindset. The challenge, however, is that these attributes may not be easily measurable in an objective manner.

6. Conclusion

In conclusion, one must always have a hybrid dependent variable for better results in classification of households when assessing food insecurity status of households especially in the developing countries.

Acknowledgements

We would like to acknowledge the support of Carnegie Corporation for the research grant that enabled this study to be conducted. We also acknowledge the support of the Makerere University Graduate School through which we received the grant as part of facilitation for the research as an initial component of PhD research on models for measurement of household food insecurity.

References

- [1] H. C. J. Godfray, J. R. Beddington, I. R. Crute, L. Haddad, D. Lawrence, J. F. Muir, J. Pretty, S. Robinson, S. M. Thomas, and C. Toulmin, "Food security: the challenge of feeding 9 billion people," *science*, vol. 327, pp. 812-818, 2010.
- [2] A. D. Jones, F. M. Ngure, G. Pelto, and S. L. Young, "What are we assessing when we measure food security? A compendium and review of current metrics," *Advances in Nutrition: An International Review Journal*, vol. 4, pp. 481-505, 2013.
- [3] A. Y. Owino, L. K. Atuhairu, R. Wesonga, F. Nabugoomu, and E. S. K. Muwanga-Zaake, "Determining Factors that Influence Household Food Insecurity in Uganda: A case study of Tororo and Busia districts," *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, vol. 14, pp. 394 - 404, 2014.
- [4] G. Bickel, M. Nord, C. Price, W. Hamilton, and J. Cook, "Guide to measuring household food security," *Alexandria. Department of Agriculture Food and Nutrition Service*, 2000.
- [5] L. L. Ching, E. Dano, and H. Jhamtani, "Rethinking agriculture," *Third World Resurgence*, 2010.
- [6] J. G. M. Majaliwa, M. K. Magunda, M. M. Tenywa, and F. Musitwa, "Soil and nutrient losses from major agricultural land-use practices in the Lake Victoria basin," 2012.
- [7] C. Carletto, A. Zezza, and R. Banerjee, "Towards better measurement of household food security: Harmonizing indicators and the role of household surveys," *Global Food Security*, vol. 2, pp. 30-40, 2013.
- [8] J. L. Dzanja, M. Christie, I. Fazey, and T. Hyde, "The role of social capital on rural food security: the case study of Dowa and Lilongwe Districts in Central Malawi," 2013.
- [9] L. R. Fabrigar, D. T. Wegener, R. C. MacCallum, and E. J. Strahan, "Evaluating the use of exploratory factor analysis in psychological research," *Psychological methods*, vol. 4, p. 272, 1999.
- [10] O. Faye, A. Baschieri, J. Falkingham, and K. Muindi, "Hunger and food insecurity in Nairobi's slums: an assessment using IRT models," *Journal of Urban Health*, vol. 88, pp. 235-255, 2011.
- [11] J. Olson and L. Berry, "Land degradation in Uganda: its extent and impact," *available at lada. virtualcentre.org/eims/download.asp*, 2003.
- [12] P. McMichael and M. Schneider, "Food security politics and the Millennium Development Goals," *Third World Quarterly*, vol. 32, pp. 119-139, 2011.