American Journal of Theoretical and Applied Statistics
2020; 9(6): 267-271
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.20200906.12
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

Measuring the Socio-economic Status of Adopters of Indigenous Chicken in Mwala and Machakos Central, Kenya: Application of Principal Component Analysis

Ngetich Titus[1], Karanjah Anthony[1], Cheruiyot Kipkoech[2]

[1]Department of Mathematics, Multimedia University of Kenya, Nairobi, Kenya

[2]Department of Mathematics, Technical University of Kenya, Nairobi, Kenya

Email address:

titkib@gmail.com (N. Titus), akaranjah@mmu.ac.ke (K. Anthony), cherukipkoech@gmail.com (C. Kipkoech)

To cite this article:

Ngetich Titus, Karanjah Anthony, Cheruiyot Kipkoech. Measuring the Socio-economic Status of Adopters of Indigenous Chicken in Mwala and Machakos Central, Kenya: Application of Principal Component Analysis. *American Journal of Theoretical and Applied Statistics*. Vol. 9, No. 6, 2020, pp. 267-271. doi: 10.11648/j.ajtas.20200906.12

Received: September 27, 2020; **Accepted**: October 24, 2020; **Published**: November 4, 2020

Abstract: This study focused on impact assessment of indigenous Chicken (KALRO Improved Chicken) in terms of the Socio-economic Status of the beneficiaries. Data analyzed comprised of household assets owned and housing characteristics. Studies have been done to assess the impact of new agricultural technologies to the beneficiaries, however, the measurement of the impact indicator (Socio-economic Status) has been a challenge. Studies rely on monetary data (reported income and expenditure), however the collection of high quality (precise and accurate) income data and expenditure is difficult and requires more resources particularly for household surveys, this approach is usually affected by unreliable reportage and measurement error, high-quality income data and expenditure will still produce biased estimates of household socio-economic status because they measure economic flows which are stochastic and include temporary income shocks. This study used principal component analysis model (PCA) to create an asset index to measure Socio-economic status. It was concluded that PCA is reliable in creating an asset index for measuring Socio-economic status, the results showed that about 40% of the households in Machakos County were poor which implies a small decline compared to 42.6% reported on [11] conducted by Kenya National Bureau of Statistics.

Keywords: Principal Component Analysis, Socio-economic Status, KALRO Improved Chicken

1. Introduction

Kenya Agricultural and Livestock Research Organization (KALRO) bred and came up with an improved indigenous chicken herein referred to as KIC which started in the year 2010. Under similar management practices, this breed has faster growth (5-6 months), lays more eggs (200-250 annually) and has reduced broodiness with the ability to scavenge for feed unlike layers and broilers, making it a suitable enterprise in poverty alleviation in terms of food security and income in the Arid and Semi-Arid lands of Kenya (ASALs).

The information regarding the impact of the technology in terms of how the adopters of the technology (KIC) vary by Socio-economic Status has been the main challenge [2] and is of interest. Studies rely on monetary data (reported income and expenditure), however the collection of high quality income data and expenditure is difficult and requires more resources particularly for household surveys [16, 20] this approach is affected by unreliable reportage and measurement error [14], high-quality income data and expenditure will still produce biased estimates of household socio-economic status because they measure economic flows which are stochastic and include temporary income shocks [13]. Using income as an indicator is difficult [9], since income information does not consider the fact that poor people may derive their income from agriculture (crops and livestock) which could be difficult to account due to the variation in seasonality, therefore measuring income is difficult for the self-employed especially agricultural field due to accounting and seasonality [12, 18]. An alternative to the income or consumption and expenditure is the asset-index approach where respondents are asked to list the type of assets they own, this approach is less likely to be affected by recall measurement error [6] since the interviewer can easily verify these assets physically, also assets can be a long term indicator of living standards compared to income or consumption and expenditure which could be affected by temporary shocks. This approach collects information on the type of assets owned which range from durable assets such as television, radio, mobile, bicycle, car, ox plough and cart to housing characteristics which include materials of floor, wall, roof, toilet and basic services like drinking water and electricity. This study employed the use of principal component analysis model (PCA) by constructing a linear index from asset ownership indicators, to derive weights [12].

2. Methodology

2.1. The Data

The study used household survey data from adopters of indigenous chicken in areas of Machakos County. Machakos was purposely selected based on County prioritized Agricultural Product Value Chains and distribution of a KALRO Improved Chicken. The [8] formula in (1) below was used in selecting the representative sample.

$$n = \frac{Z^2(1-p)p}{e^2} \tag{1}$$

From (1) a sample size of 339 was obtained, where n represents the sample size, Z is the standard score for 95% confidence level, p is the population parameter (proportion) and 'e' is the standard error for estimating the population parameter *(p).*

Multistage sampling approach was applied in selecting the respondents, two sub-counties were selected randomly (Mwala and Machakos central), next three wards selected randomly in each of the two sub-counties and lastly at least 30 respondents randomly selected in each ward.

2.2. Principal Component Analysis

The first principal component was applied in constructing the asset index for measuring the SES, the asset index is internally coherent, robust and a comparable indicator of SES [4].The study comprised of data on asset ownership which include durable assets and housing characteristics. These variables were corded as binary that is; (1) for a household who own a particular asset and (0) for a household that don't own the asset, similarly for the housing characteristics.

PCA is a data reduction technique it reduces data and detects underlying variables in a data set, it is a multivariate technique [10, 17] applicable to data that have a uniform scale in each original variable, it reduces the number of variables with no loss of much information in the process and transforms the set of correlated original variables into a set of linearly uncorrelated variables (principal components) [1, 15, 19].The new uncorrelated variables explains most of the variation in the original data, it explains the variance covariance structure of a set of variables through a linear combinations of the original variables [5].

2.2.1. Principal Component Analysis Algebra

For a data matrix with n observations on p correlated random variables, PCA looks for a transformation of the X_i into p new variables Y_i , where the p principal components Y_1 , $Y_2, \ldots, Y_p$ are linear combinations (uncorrelated) of the original variable, $X_1, X_2, \ldots, X_p$, given as;

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p$$

$$Y_2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p$$

$$\vdots$$

$$Y_p = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p$$

The equations are expressed as $Y = ax$, where Y=Y_1,Y_2,…, Y_p, X= (X_1, X_2,..., X_p) and a *is* the matrix of coefficients below.

$$\begin{bmatrix} a_{11} & a_{12} \ldots & a_{1p} \\ a_{21} & a_{22} \ldots & a_{2p} \\ a_{p1} & a_{p2} & a_{pp} \end{bmatrix} \tag{2}$$

The 1st principal component Y_1 is the linear combination of X such that;

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \tag{3}$$

which has the greatest sample variance compared to the other linear combinations.

The first principal component Y_1 coefficients (a_{11}, $a_{12\ldots}$ a_{1p} denoted as $a_{1)}$ could be increased without limits, therefore a restriction must be put in the coefficient such that;

$$\acute{a}_1 = a_1 \tag{4}$$

The second principal component Y_2 is put to the two conditions such that;

$$\acute{a}_2\, a_2 = 1 \tag{5}$$

$$\acute{a}_2\, a_1 = 0$$

This ensures that Y_1 and Y_2 are uncorrelated. Also, the j^{th}Principal component is put in to the conditions;

$$\acute{a}_p\, a_q = 1$$

$$\acute{a}_p\, a_q = 0 \text{ (p<q)} \tag{6}$$

For the vector X, with covariance matrix Σ and eigenvalue, eigenvector pair

$$(\lambda_1, e_1, \lambda_2, e_2 \ldots \ldots \ldots . \lambda_p, e_p)$$

The i^{th} principal component,

$$Y_i = e`X = e`_{i1}X1 + e`_{i2}X2 + \cdots + \cdots + e`_{ip}Xp \text{ (7)}, i = 1,2, \ldots . p \tag{7}$$

Where; eigenvalue computed by solving the characteristic equation;

$$det\ (\Sigma - \lambda I) = 0 \quad (8)$$

and;

Eigenvector computed by solving the equation;

$$\Sigma\, e_i = \lambda_i\, e_i \quad (9)$$

The first principal component has a variance λ_1 which is the largest eigenvalue of the covariance matrix, the variance of the second principal component is λ_2 and is uncorrelated with the first principal component, it explains additional but less variation in the original variable compared to the first principal component. Further subsequent principal component are also defined in a similar way, each principal component is not correlated with all the others, and principal component analysis involves decomposition of eigenvalue / eigenvectors of the covariance matrix.

In assessment of how well a subset of principal components Y_i summarize the original variables X_i The percent of total population variation (p_j) due to the K^{th} principal component was computed as

$$p_j = \frac{\lambda_k}{\sum_{i=1}^{p} \lambda_i} * 100 \quad (10)$$

Where λ_i is the eigenvalues of a and $\sum_{i=1}^{p} \lambda i = Trace(a) = p = 18$

In some instance where a large proportion of the variance is attributed to some few principal components, the original variables can be replaced with the new few principal components without losing much information.

PCA is constructed in a manner that the first principal component explains the most variation in the data than the subsequent components. Some statisticians recommend using all eigenvectors with eigenvalues greater than one; others suggest the "scree test". For this study, the 1^{st} principal component was applied for constructing the asset index for measuring the socio-economic status.

2.2.2. Asset Index

The creation of a reliable asset index was based on the first principle component in (3) PCA assigns weights to the variables of interest where assets found in all the household is given a weight of zero [7] while those which varies most across the household given large weight.

An asset index (yi) is assigned to each household (i) which is the linear combination in (11) below.

$$yi = \sum_{k=1}^{p} \alpha_k \left(\frac{x_{ki-}\overline{x_k}}{s_k}\right), i = 1,2,3,..n \quad (11)$$

where $\overline{x_k}$ is the mean of asset,

$$x_k \text{ and } \overline{x_k} = \frac{1}{n}\sum_{i=1}^{n} x_{ki} \quad (12)$$

x_{ki} is asset k for i^{th} household, s_k standard deviation and α_k is the weight for the k^{th} asset.

The asset index may have positive, negative or both values, in this study, the estimated asset index was based on the 287 respondents. The asset scores for each household was created in the pooled data set through PCA using (11), they are standardized values that has a normal distribution of N (0, 1).

2.2.3. Socio Economic Status

In order to assess the SES of households, the asset index was ranked in ascending order, wealth quintiles computed and households divided into quintiles from one (lowest) to five the (highest). Households within the first and second quintile were categorized as poor (coded as 0), while those in the 3^{rd} 4th and 5^{th} categorized "not poor "(coded as 1). The proportion of poor and "not poor "was computed to give the SES of adopters of indigenous chicken in study area.

3. Results

3.1. First Principal Component Results

The results of the first principal component is shown in table 1 below, the descriptive statistics generated comprised of mean, standard deviation and factor score. The factor score/weights are the coefficients in (3).

Table 1. *First principal component results.*

Variable	Mean	SD	N	Weight/Factor score
Floor Type				
Cement	.75	.436	287	.935
Mud	.25	.436	287	-.935
Wall Type				
Wooden	.20	.400	287	-.541
Block	.40	.492	287	.368
Mud	.11	.315	287	-.246
Cement	.06	.230	287	.276
Iron sheet	.15	.354	287	.086
Raw brick	.08	.267	287	.079
Roof Type				
Iron sheet	.95	.208	287	.000
Tiles	.02	.155	287	.072
Grass	.02	.143	287	-.078
Durable Goods				
Ox plough	.16	.371	287	.095
Television	.56	.497	287	.293
Motorcycle	.15	.361	287	.145

Variable	Mean	SD	N	Weight/Factor score
Mobile	.75	.432	287	.035
Radio	.70	.459	287	.027
Bicycle	.33	.470	287	.220
Carts	.16	.371	287	.035
Largest Eigen value	2.737			
Proportion of variance explained	15.204			

The factor score has both positive and negative values, variables with negative values are associated with low SES, while those with positive values associated with higher SES. It was observed that almost all the household dwellings were roofed with iron sheet (95%), this variable was assigned very low weight (0.000), and implying having an iron sheet roofed dwelling does not explain the variation in socio economic status of the households. In contrast having cemented floor, block wall, television and motorcycle were weighted more heavily (at least 0.145) implying these variables explains more of the variation and are the important variables for measuring the SES. The largest eigenvalue for the first principal component was 2.737 and explains 15.204 percent variation in the data.

The percent of variance explained by the principal components is illustrated in table 2 below, the proportions were generated using (10).

Table 2. *Eigenvalues and percent of variance for the principal components.*

Component	Eigenvalues		
	Total	% of variance	Cumulative%
1	**2.737**	**15.204**	**15.204**
2	2.124	11.798	27.002
3	1.868	10.377	37.379
4	1.594	8.856	46.235
5	1.264	7.021	53.256
6	1.183	6.573	59.829
7	1.131	6.286	66.115
8	1.010	5.608	71.724
9	.962	5.347	77.070
10	.841	4.674	81.744
11	.792	4.401	86.145
12	.730	4.055	90.200
13	.672	3.735	93.935
14	.578	3.211	97.145
15	.505	2.807	99.952
16	.009	.048	100.000
17	1.143E-16	6.348E-16	100.000
18	-3.740E-17	-2.078E-16	100.000

The percent of variation of the 1st, 2nd to the 8th principal components were 15.204, 11.78, and 5.608 respectively. The results indicate that the largest variance in the original data is explained by the first principal component (15.2%) with the largest eigenvalue of 2.737 compared to subsequent principal components which have decreasing proportion of variance. The original variables could sufficiently be explained by the first eight components (71.724%) with eigenvalues>1 as shown in table 2, but in this study the first principal component had the greatest variance and was applied in creating the asset index for measuring the SES.

3.2. Socio-economic Status of Adopters of Indigenous Chicken

Table 3 below shows the SES of adopters of indigenous chicken in the study area. The asset index for each household was created using (11), the SES was generated as described in section 2.2.3, and the proportion of the two levels of SES were tabulated below.

Table 3. *Socio-economic status of adopters of indigenous chicken.*

Socio economic status	Count	%
Not poor	173	60.3
poor	114	39.7
Total	287	100.0

It is observed that about 40% of the household in study area were poor, which was close to 42.6% reported on [11] conducted by Kenya National Bureau of Statistics

The poverty threshold in rural Kenya is measured by an expenditure of less than 1 USD per day per person, the rural poverty in Kenya was 40.1% as reported [3]

4. Conclusion

Application of PCA model is reliable for creating and asset index for measuring the socio economic status of the households, PCA extracts most important information from the large data set without losing much information, it uses the covariance matrix to derive systematic weights (asset index) that can be replicable. From the results, about 40% of the respondents were poor, and this figure was close to 42.6% poverty level in Machakos reported on the [11], the overall rural poverty in Kenya was 40.1% reported on [3]. This implies a small decline in rural poverty in Machakos compared to that reported on the [11] conducted by Kenya National Bureau of Statistics.

References

[1] Alice Constance Mensah, Joseph Dadzie, (2020). Application of Principal Component Analysis on Perceived Barriers to Youth Entrepreneurship. *American Journal of Theoretical and Applied Statistics*. Vol. 9, No. 5, 2020, pp. 201-209. doi: 10.11648/j.ajtas.20200905.13.

[2] Anne M., (2014). Construction of Household Asset-Based Wealth Index for Eastern Region, Kenya.

[3] Basic Report on well-being in Kenya (2015/2016). Annals of Kenya Integrated Household Budget Survey.

[4] Booyen, R., (2002). Using Demographic and Health Surveys to Measure poverty. An Application to South Africa.

[5] Chand, Prem & Kanwal, Vinita. (2019). Principal Component Analysis. In Book: Quantitative Methods for Social Sciences.

[6] Chen; Ravallion; Ellis, (2000). The Construction of an Asset Index Measuring Asset Accumulation in Ecuador.

[7] Chuma, J., & Molyneux, C., (2009). Estimating inequalities in ownership of insecticide treated nets: Does the choice of socioeconomic status measure matter? Health Policy and Planning, 24, pp. 83–93.

[8] Cochran, W. G., (1963) Sampling Techniques, Wiley, New York.

[9] Cope JR, Doocy S, Frattaroli S, McGready J., (2012). Household Expenditures as a Measure of Socioeconomic Status Among Iraqis Displaced in Jordan and Syria. World Health Popul 2012; 14 (1): pp. 19-30. doi: 10.12927/whp.2013.23063.

[10] Denis, Daniel, (2020). Principal Component Analysis. 10.1002/9781119549963.ch10.

[11] Economic survey, (2014). Kenya National Bureau of Statistics.

[12] Filmer D., Pritchett LH. (2001). Estimating wealth effect without expenditure data – or tears: an application to educational enrollments in states of India. Demography 38: pp. 115–32.

[13] Gwatkin DR., Rustein S., Johnson K. et al., (2000a). Socio-economic differences in Brazil. Washington, DC: HNP/Poverty Thematic Group of the World Bank. Accessed 5 January 2004 online at: [http://www.worldbank.org/poverty/health/data/index.htm#lcr].

[14] Hope Michelson, Maria Muñiz & Kyle DeRosa, (2013) Measuring Socioeconomic Status in the Millennium Villages: The Role of Asset Index Choice, The Journal of Development Studies, 49: 7, 917-935, DOI: 10.1080/00220388.2013.785525.

[15] JolliffeI. T., Cadima J., (2016) .Principal component analysis: a review and recent developments. Phil. Trans. R. Soc. A374: 20150202. http://dx.doi.org/10.1098/rsta.2015.0202.

[16] McKenzie DJ. (2003). Measure inequality with asset indicators. BREAD Working Paper No. 042. Cambridge, MA: Bureau for Research and Economic Analysis of Development, Center for International Development, Harvard University.

[17] Mishra, Sidharth & Sarkar, Uttam & Taraphder, Subhash & Datta, Sanjoy & Swain, Devi & Saikhom, Reshma & Panda, Sasmita & Laishram, Menalsh. (2017). Principal Component Analysis. International Journal of Livestock Research. 1. 10.5455/ijlr.20170415115235.

[18] Sahn D., Stifel D., (2003). Exploring alternative measures of welfare in the absence of expenditure data. Review of Income and Wealth 49: pp. 463–89.

[19] Sidharth Prasad Mishra, Uttam Sarkar, et al., (2017). Multivariate Statistical Data Analysis-Principal Component Analysis (PCA). International Journal of Livestock Research eISSN: 2277-1964 NAAS Score -5.36.

[20] Vyas, S. Kumaranayake L. (2006). Constructing socio-economic status Indices. How to use principal components analysis. Health policy plan, 21 (6): pp. 459-68.