UNIJERPS
Unizik Journal of Educational Research and Policy Studies VOL.17 (1); ***January-June, 2024*** https://unijerps.org
2023 Impact Factor: **6.834;** 2022 ***Impact Factor***: 5.156; ***2021 Impact Factor: 5.641;***
2020 Impact Factor: 5.355. *Indexing and Impact Factor URL:* http://sjifactor.com/passport.php?id=21363

CLUSTER SAMPLING OF HOUSEHOLD INCOME AND EXPENDITURE PATTERNS IN SELECTED STUDY AREAS IN ONDO STATE, NIGERIA

[*1]Akinola Oladiran Adepetun and [2]Elisha Oluwaseun Elijah
[1,2]Department of Statistics, Federal University of Technology, Akure, Nigeria
[*]Corresponding Author's email: aoadepetun@futa.edu.ng

ABSTRACT
The concept of cluster sampling has a lot of applications in survey sampling. Cluster sampling is a method of probability sampling that is often used to study large population, particularly one that is widely geographically dispersed. This study applied cluster sampling technique to estimate the average households' incomes and expenditures with a view to determining the standard of living of the households. In order to achieve this, real data set for the work were collected from 250 households in Rural and Urban areas of Odigbo and Akure South Local Government in Ondo State, Nigeria with the aid of well- structured survey questionnaires. The analyses were carried out using suitable statistical software. The study revealed that the percentage of households that their income could not sustain was more than those that their income could sustain. The results show that the estimated average households' expenditure was lower than the estimated average income which implied that the households did not spend more than their incomes. The percentage of households' standard of living increased more in rural areas while the standard of living in urban areas remains the same.

Keywords: Cluster Sampling Technique, Households, Income, Expenditure, Rural and Urban

Symbols and Notations
N = total number of the population
M = number of elements of the population
n = sample selected from the population
m = number of samples drawn from element of the population
Y_{ij} = j^{th} observation in the i^{th} cluster
Y_i = observations from the population
y_{ij} = j^{th}observations in the i^{th}cluster for the sample
y_i = total of all observations in the i^{th}cluster
N' = total number of cluster population
n' = number of cluster sample
$\sum Y_i$ = total observations from the population
$\sum y_i$ = observations for the sample
$\bar{Y}$ = population mean
$\bar{Y}_i$ = population mean of i[th] cluster
P = proportion mean
P_i = proportion elements in the i[th] cluster of the population
p_i = proportion elements in the i[th] cluster of the sample
$\bar{y}_i$ = sample mean of i[th] cluster
$\bar{y}_{cl}$ = mean of cluster means
S^2 = sampling variance
S_B^2 = mean sum of square between the cluster means in the population
s_b^2 = mean sum of square between the cluster means in the sample
S_w^2 = mean of n mean sum of squares
$\bar{S}_w^2$ = mean sum of square within the cluster means in the population
$\bar{s}_w^2$ = mean sum of square within the cluster means in the sample

S_i^2 = mean sum of square for the i[th] cluster
s_y^2 = variance of the selected clusters of the sample
σ = standard error
ρ = intra – class correlation coefficient
σ^2 = intra - cluster mean of sum of square
σ_w^2 = intra – cluster mean sum of square within the cluster
σ_B^2 = intra – cluster mean sum of square between the cluster
Cluster surveyed is 1 = Agbabu, 2 = Ore II, 3 = Lisa, 4 = Gbogi/Isikan I, and 5 = Gbogi/Isikan II
e = error margin

Introduction

Household income and expenditure are indicators of the living standard of household. They are used to measure the purchasing power of the population. The income and expenditure of the studied Local Governments in Ondo State deserve attention, as well as the notion of household expenditure and income in Akure South Local Government and Odigbo Local Government. Generally, consumer behaviour reflects a nation's degree of economic well-being and poverty. To know the changing patterns in household consumption spending, it is critical to measure consumption expenditure over time in every economy. There is a direct and significant relationship between household consumption expenditure and income, while other variables, except inflation, exhibit the same relationship. Inflation shows an indirect relationship since there was a long-term relationship between household expenditure and income Tokoya *et al* (2022). According to one definition, household income is the "total pay earned by every family member age 15 or older living in the same household (Shuani, 2016). Cluster sampling involves obtaining a random sample of clusters from the population, with all members of each selected cluster invited to participate (Sedgwick, 2014). In cluster sampling, we assume that the population contains N clusters with M elements each, n cluster are selected from N cluster by simple random sampling without replacement and all the elements in the selected clusters are included in the sample (Etorti *et al*, 2022). Parmar *et al*. (2015) conducted household survey using two – stage cluster sampling in five states in eastern Myanmar to determine the health and human rights in eastern Myanmar after the political transition. Data were collected from 6, 178 households on demographics, mortality, health outcomes, water and sanitation, food security and nutrition, malaria, and human rights violations (HRV). Papamichail *et al*. (2017) applied cluster sampling to obtain reliable quantitative estimates of vaccination coverage of the Greek Roma child population at a national level, and identify risk factors regarding Roma settlement characteristics associated with low immunization coverage. Rennberger *et al*. (2018) applied cluster sampling to conduct a survey on the occurrence of Foliar Pathogens of Watermelon on Commercial Farms in South Carolina Estimated with Stratified Cluster Sampling. Lafta *et al*. (2015) used cluster sampling survey to characterize injuries, deaths, and

disabilities arising during 11 years of conflict in Baghdad. Thirty (30) clusters were selected, proportionate to population size estimates. Interviews were conducted in 900 households containing 5148 persons. Baquero *et al*. (2018) used a two-stage cluster sampling design to estimate the population sizes of owned dogs and cats in Pinhais, Brazil. The study population was based on the urban households of Pinhais identified in the last census, from 2010, in which 36,149 households grouped in 136 census tracts and 117,008 persons were counted. Auplish *et al*. (2017) applied intra-cluster correlation coefficient for evaluating an educational intervention program to improve rabies awareness and dog bite prevention among children in Sikkim in India. A two-stage cluster sampling was used and was based on the 2011 population and housing census Enumeration Area (EAs) and the data were collected in twelve months. The approach gives a precise result and also replicate and improvement in precision on the subject matters of tobacco and alcohol use in 2015/2016 Namibia Household income and expenditure survey (Nghipandulwa and Mukong 2023). The application of void and cluster sampling of large scattered trajectories proposed data reduction technique of scattered data based on statistical sampling and these used efficient and iterative approach (Rapp *et al* 2019). The new work framework incorporates a design component that accounts for the variability induced on the estimator by the treatment assignment mechanism, when the number of clusters in the sample is a nonnegative fraction of the number of clusters in the population, convectional clustered (Abadie *et al* 2023).

This study, therefore, applied Cluster Sampling technique to estimate the average Household income and their expenditure patterns in selected areas of Ondo State, Nigeria.

Methodology

Methodology according to (Creswell, 2008) is a process of steps used to collect and analyze information to increase our understanding of a topic or issue. It consists of three steps: pose a question, collect data to answer the question and present an answer to the question. The study was carried out to estimate the average household incomes and their expenditures of the residents in the selected study areas in Ondo State, Nigeria. This aspect of the study comprises procedures that were used for the entire study. It consists of the research design, population and sample to be studied, types and sources of data, tools to be used in the collection of data, the statistical tools to be employed for data presentation and analysis. The primary data for this study was collected through the administration of well-designed questionnaire to individual households in the selected areas of Ondo State, Nigeria. For this study, the analysis was based on the questionnaires retrieved from the selected households that constituted the needed clusters for analysis with the use of suitable software. The appropriate sample size was computed using the formula proposed by Yamane (1967) which provided a simplified formula to calculate sample sizes. Random number table was also used for selecting suitable sample from the population for the purpose of determining parameters of the whole population. Since

this study adopted cluster sampling technique, it is considered relevant to present the following statistical expressions used in the work. Consequently, the estimates of population total, population variance, proportion, proportion variance, intra class correlation and relative efficiency are respectively defined in equations 1-7 below:

$$\hat{Y} = \frac{N}{n}\sum_{i=1}^{5} M\bar{y}_i \quad (1)$$

$$\bar{y}_{cl} = \frac{1}{n'}\sum_{i=1}^{n'} \bar{y}_i \quad (2)$$

$$\widehat{Var}(\bar{y}_{cl}) = \frac{N-n}{Nn} S_b^2 \quad (3)$$

$$\hat{P}cl = \frac{1}{n'}\sum_{i=1}^{n'} P_i \quad (4)$$

$$\widehat{Var}(\hat{P}_{cl}) = \frac{(N-n)\left[n\hat{P}_{cl}\hat{Q}cl - \sum_{i=1}^{N} PiQi\right]}{Nn\,(n'-1)} \quad (5)$$

$$\rho = \frac{S_b^2 - \frac{\sigma_w^2}{M-1}}{\sigma^2} \quad (6)$$

$$\hat{E} = \frac{1}{M} + \left(\frac{M-1}{M}\right)\frac{\bar{s}_w^2}{Ms_b^2} \quad (7)$$

Results and Discussion

The result revealed that the mean and variance of the income was estimated in table 1 as described as average (mean) as 228.3 in thousand naira while the variance is 206.8 in thousand naira which implies the variance is biased across all the clustered. Meanwhile Gbogi/Isikan II has the highest mean 268.8 in thousand naira while Agbabu has the least mean of 179.4 in thousand naira. Table 2; mean and variance of the expenditure was estimated and describe average (mean) as 98.2 in thousand naira while the variance is 4.82 in thousand naira which implies the variance is also biased likewise, Gbogi/Isikan II also has the highest average 126.3 in thousand naira while Agbabu has the least average 66.3 in thousand naira as expenditure. The average household rate of expenditure is 0.06 per thousand naira which implies household spent 60% of their income in a month as shown in table 3

Table 1: Estimation of Average and Variance of Household Income in each Cluster

Cluster	M_i	m_i	Y_{ij}	y_{ij}	$\bar{Y}_i$	$\bar{y}_i$	$\bar{Y}$	$\bar{y}$	$(\bar{Y}_i - \bar{Y})^2$	$(Y_{ij} - \bar{Y}_i)^2$	$(y_{ij} - \bar{y}_i)^2$	$(\bar{y}_i - \bar{y})^2$
1	50	44	8649.3	7895.1	173.0	179.4	236.8	228.3	4070.4	71847661.7	59532026.5	2391.2
2	50	44	10123.5	8698.1	202.5	197.7	236.8	228.3	1176.5	98426241	72256800.2	936.4
3	50	44	12293.2	10533.8	245.9	239.4	236.8	228.3	82.8	145137437.3	105974671.4	123.2
4	50	44	14834.9	11283.3	296.7	256.4	236.8	228.3	3588	211359259.2	121592523.6	789.6
5	50	44	13302.9	11828	266.1	268.8	236.8	228.3	858.5	169958154.2	113615104.6	1640.3
Total					**1184.2**	**1141.7**			**9776.2**	**696728753.4**	**472971126.3**	**5880.7**

The table above shows the summation of average income for the population and sample in each cluster, the sum of squares of income between the cluster means in the population and the sample, sum of squares of income within the cluster means in the population and sample.

Table 2: Estimation of Average and Variance of Household Expenditure in each Cluster

Cluster	M_i	m_i	Y_{ij}	y_{ij}	$\bar{Y}_i$	$\bar{y}_i$	$\bar{Y}$	$\bar{y}$	$(\bar{Y}_i - \bar{Y})^2$	$(Y_{ij} - \bar{Y}_i)^2$	$(y_{ij} - \bar{y}_i)^2$	$(\bar{y}_i - \bar{y})^2$
1	50	44	3355.3	2919.8	67.1	66.3	102.9	98.2	1281.6	10812259.2	8142462.3	1017.6
2	50	44	4498.2	3913.9	90.0	89.0	102.9	98.2	166.4	19432227.2	14629860.0	84.6
3	50	44	5085.8	4115.4	101.7	93.5	102.9	98.2	1.4	24841252.8	16175679.6	22.1
4	50	44	6331.4	5090.4	126.6	115.7	102.9	98.2	561.7	38499543.0	24747640.1	306.3
5	50	44	6455.2	5555.2	129.1	126.3	102.9	98.2	686.4	40019541.2	29472955.2	789.6
Total					**514.5**	**490.8**			**2697.5**	**98954823.4**	**93168597.2**	**2220.2**

The table above shows the summation of average expenditure for the population and sample in each cluster, the sum of squares of expenditure between the cluster means in the population and the sample, sum of squares of e within the cluster means in the population and sample.

Table 3: Estimation of Average and Variance of Household Rate Expenditure in each Cluster

Cluster	M_i	m_i	Y_{ij}	y_{ij}	$\bar{Y}_i$	$\bar{y}_i$	$\bar{Y}$	$\bar{y}$	$(\bar{Y}_i - \bar{Y})^2$	$(Y_{ij} - \bar{Y}_i)^2$	$(y_{ij} - \bar{y}_i)^2$	$(\bar{y}_i - \bar{y})^2$
1	50	44	3.74	3.31	0.0748	0.0752	0.0614	0.0617	0.0002	13.43	10.46	0.0002
2	50	44	3.68	3.21	0.0736	0.0729	0.0614	0.0617	0.0001	13.01	9.84	0.0001
3	50	44	2.75	2.36	0.055	0.0236	0.0614	0.0617	0.00004	7.26	5.46	0.0015
4	50	44	2.25	2.00	0.045	0.0455	0.0614	0.0617	0.0003	4.86	3.82	0.0003
5	50	44	2.94	2.71	0.0588	0.0616	0.0614	0.0617	0.000007	8.30	7.01	0.00000001
Total					**0.3072**	**0.2788**			**0.0006**	**46.86**	**36.59**	**0.0021**

The table above shows the summation of average rate expenditure for the population and sample in each cluster, the sum of squares of rate of expenditure between the cluster means in the population and the sample, sum of squares of rate of expenditure within the cluster means in the population and sample

Table 4 and 5: The percentage of households that their income could not sustain was more than those that their income could sustain in the study areas. These implied that the income of the populace for both Odigbo and Akure South Local Governments of Ondo State was not sufficient for their well-being. As for Table 6 and 7, standard of living in the study areas remains the same during the study period (July to September, 2023) while there was higher increase in Odigbo study area. The standard of living remains the same in Akure South because many of the citizens are salary earners and small-scale business owners in which inflation of basic items had serious effect while Odigbo Local Government study area is widely populated with farmers. Inflation in price of cocoa and Kolanut which are the major planted crops had little effect on their income.

Table 4: Percentage distribution of Household Income sustained (Odigbo)

Income sustained	Frequency	Percentage
Yes	37	37
No	63	63

Table 4 shows that the income could not sustain 63% of household for a month while it could only sustain 37%.

Table 5: Percentage distribution of Household Income sustained (Akure South)

Income sustained	Frequency	Percentage
Yes	60	40
No	90	60

Table 5 shows that the income could not sustain 60% of household for a month while it could only sustain 40%.

Table 6: Percentage distribution of standard of living (Odigbo)

Standard of Living	Frequency	Percentage
Decreased	14	14
Stayed the same	58	58
Increased	28	28

Table 6 shows that 58% of the household standard of living remains the same, while 28% are increased and 14% are decreased.

Table 7: Percentage distribution of standard of living (Akure South)

Standard of Living	Frequency	Percentage
Decreased	40	26.7
Stayed the same	66	44
Increased	44	29.3

Table 7 shows that 44% of the household standard of living stayed the same, while 29.3% are increased and 26.7% are decreased.

Conclusion

The study applied Cluster Sampling technique to estimate the average household income and expenditure in selected areas of Ondo State, Nigeria. The results show that the average household income and expenditure were 228.3 and 98.2 in thousand naira respectively. The rate of household expenditure was 60% per month. It was further revealed in the study that income of households could not sustain them. It was also revealed that high percentage of expenditure was due to high cost of basic needs, of the citizens. Their income could not sustain the households because of the aforementioned inflation rate. Also, their standard of living remains the same while inflation had serious effect on rural area compared to urban area.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, 138(1), 1-35.

Auplish, A., Clarke, A. S., Van Zanten, T., Abel, K., Tham, C., Bhutia, T. N. & Firestone, S. M. (2017). Estimating the intra-cluster correlation coefficient for evaluating an educational intervention program to improve rabies awareness and dog bite prevention among children in Sikkim, India: A pilot study. *Acta tropica*, 169, 62-68.

Baquero, O. S., Marconcin, S., Rocha, A., & Garcia, R. D. (2018). Companion animal demography and population management in Pinhais, Brazil. *Preventive veterinary medicine*, 158, 169-177.

Creswell, J.W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.) Upper Saddle River, NJ: Pearson Education, Inc.

Etorti, I. J., Ferdinand M. E. & Agada, A. N. (2022). Theoretical and Empirical Comparison of Stratified and Cluster Sampling Schemes with Application to Mathematics Students Scores. *AJSTME*, 8(6). 506-513.

Lafta, R., Al-Shatari, S., Cherewick, M., Galway, L., Mock, C., Hagopian, A., & Burnham, G. (2015). Injuries, death, and disability associated with 11 years of conflict in Baghdad, Iraq: a randomized household cluster survey. *PLoS one*, 10(8), e0131834.

Nghipandulwa, M. T., & Mukong, A. K. (2023). Estimating Income-Related Health Inequalities Associated with Tobacco and Alcohol Consumption in Namibia. *International Journal of Environmental Research and Public Health*, 20 (2), 1062.

Parmar, P. K., Barina, C. C., Low, S., Tun, K. T., Otterness, C., Mhote, P. P. & Richards, A. K. (2015). Health and human rights in eastern Myanmar after the political transition: a population-based assessment using multistage

household cluster sampling. *PloS one*, 10(5), e0121212.

Papamichail, D., Petraki, I., Arkoudis, C., Terzidis, A., Smyrnakis, E., Benos, A., & Panagiotopoulos, T. (2017). Low vaccination coverage of Greek Roma children amid economic crisis: national survey using stratified cluster sampling. *The European Journal of Public Health*, 27(2), 318-324.

Rapp, T., Christopher, P., & Carsten, D. (2019). Void-and-cluster sampling data and trajectories. *IEEE transactions on visualization and computer graphics* 26(1), 780-789.

Rennberger, G., Gerard, P., & Keinath, A. P. (2018). Occurrence of foliar pathogens of watermelon on commercial farms in South Carolina estimated with stratified cluster sampling. *Plant disease*, 102(11), 2285-2295.

Sedgwick, P. (2014). Cluster sampling. *BMJ*, 348.

Shuani, F. (2016). Family Income Types: Money, Real and Psychic Income. Retrieved from http://www.yourarticlelibrary.com/family/family-income-types-money-real-and-psychic-income/47908/.

Tokoya, B. O., Aiyeku, O. H., Shonibare, A. A., Ipinjolu, O. A., & Dairo, R. O. (2022). Income and Household Consumption Expenditure in Nigeria (1986–2020). *Journal of Economics, Management and Trade*, 28(3), 30-41.

Yamane, T. (1967). *Statistics: An Introductory Analysis* (2nd ed.). Harper and Row.