Sampling Design for the Annual Agricultural Sample Survey 2023 (AASS23) of Tanzania

Domains: 31 Regions, i.e. Dodoma, Arusha, Kilimanjaro, Tanga, Morogoro, Pwani, Dar Es Salaam, Lindi, Mtwara, Ruvuma, Iringa, Mbeya, Singida, Tabora, Rukwa, Kigoma, Shinyanga, Kagera, Mwanza, Mara, Manyara, Njombe, Katavi, Simiyu, Geita, Songwe, Kaskazini Unguja, Kusini Unguja, Mjini Magharibi, Kaskazini Pemba, Kusini Pemba.

Frame: list of 90,888 EAs created from the Population and Housing Census 2022 (PHC 2022). Available information for each EA: Region, District, Council, Constituency, Division, Ward, Village, EA, Hamlet, Rural/Urban type, number of households, number of agricultural households, number of households growing crops, number of households rearing livestock, number of households practicing aquaculture.

Frame Limitations*: in the frame provided by NBS (National Bureau of Statistics of Tanzania) the EAs coincide with the Hamlets, hence some of them are highly populated (17.3% of EAs have more than 200 households, hitting the top of 11,660 households). Highly populated hamlets have a higher probability to be extracted in the sample, causing an intense listing operation.*
Since the GIS experts of NBS mentioned that partitioning the large hamlets into smaller EAs by using cartographic material was too time-consuming considering the survey work plan and deadlines, first it was proposed an artificial partition of the larger hamlets before the selection of the sample – i.e. that the hamlets with more than 200/150 households could be split into smaller EAs, with the assumption that both the number of households and the number of agricultural households must be distributed equally among the EAs. If these EAs were extracted in the sample, they would have been demarcated from the main hamlet by using cartographic material, paying attention to split the agricultural households equally between EAs. Albeit sub-optimal, this procedure was considered helpful since the demarcation operation would be conducted just on the sampled high-populated hamlets, not on ***all*** *the large hamlets. However, the procedure could have generated discrepancy between the number of households (and agricultural households) artificially allocated from the hamlet to each EA before the sample extraction and the number found in the field after the demarcation of the hamlet with cartographic material (causing a bias in the sampling weights). In addition, the demarcation based on cartographic material seemed to be difficult for the NBS cartographers.*

Considering the drawbacks of this approach, NBS proposed a dual approach – i.e., to conduct the listing operation in the hamlets with less than 200 households and to use the list of agricultural households as identified during the PHC 2022 in the large hamlets (with more than 200 households), without performing a new listing operation. NBS believes that the PHC 2022 household lists are still updated and they were used also for other surveys conducted in 2023; the sample attrition seems to be quite low and the holdings' contacts seem to be reliable.

However, household lists generally become outdated quickly as people migrate or change their engagement in agriculture. Therefore, it is not advisable to use the same approach for the AASS 2024. ***For the next survey cycles it is fundamental that NBS develops a frame with EAs completely demarcated and usable for statistical operations*** *(all the EAs must contain at most 200 households). Splitting the large EAs in smaller pieces by using cartographic material and ground truth may seem a costly investment, but it is crucial to produce more reliable and precise results from the next surveys, including household and individual surveys. The institutions of Tanzania working on data production and data use can really benefit from it.*

Sampling method: stratified two-stage design.

Sampling units:

1. first stage: Enumeration Areas (Hamlets)
2. second stage: agricultural households, i.e. households growing crops and/or rearing livestock; aquaculture and bee-keeping activities are considered complementary to the first two primary activities. Households practicing just aquaculture or beekeeping are not considered to be part of the sampling units (the reason leading to this decision is that no previous information is available to consider in the computation of the sample size precision requirements of the estimates of these two sectors). A household is considered agricultural only if it has at least 25 square meters of planted land and/or one cattle and/or 5 goats/sheep/pigs and/or 50 chickens/ducks/turkeys.

Strata: two to three strata of EAs in each domain created considering three variables: the number of households growing crops, the number of households rearing livestock and the number of households practicing aquaculture as reported in the frame obtained from the PHC2022. This multivariate stratification has been conducted by the k-means algorithm with the R function *kmeans*. The means of the stratifying variables computed in each stratum and domain are given in Table 1. In general, on average, stratum one is the most populated and contains EAs less involved in agriculture, stratum three is the least populated and its EAs include a lot of households involved in crop cultivation, livestock rearing and aquaculture. In some domains the sample size is low, hence just two strata have been identified. The stratification procedure applied in Dodoma is shown in Annex 2.

First stage sampling selection: systematic selection with probability proportional to size (PPS). Within each stratum and domain, the EAs are ordered according to District's and Council's Codes which reflect the geographical proximity, then ordered according to Constituency's, Division's, Ward's, Village's codes. An implicit stratification is also performed, ordering by Urban/Rural type at Ward level. The EAs are selected with probability proportional to size where the measure of size is the number of agricultural households in the EA.

Second stage sampling selection: simple random sampling. In hamlets with more than 200 households, twelve (12) agricultural households are drawn from the PHC 2022 list with a simple random sampling without replacement procedure in each sampled hamlet. In hamlets with 200 households or less, twelve (12) agricultural households will be selected with a simple random sampling without replacement procedure in each sampled hamlet after listing operation.

Subsampling for area measurement. In order to increase the precision of the area-related estimates, a subsample of 5 agricultural households is extracted through a Permanent Random Numbers (PRN) sampling from the 12 agricultural households sampled in each selected EA and all the plots of these subsampled units are measured through GPS. The area-related estimates (crop areas, plot area, agricultural land area) are then corrected through the regression estimator.

Region	Mean HH crops (Stratum 1)	Mean HH livestock (Stratum 1)	Mean HH aquac (Stratum 1)	Mean HH crops (Stratum 2)	Mean HH livestock (Stratum 2)	Mean HH aquac (Stratum 2)	Mean HH crops (Stratum 3)	Mean HH livestock (Stratum 3)	Mean HH aquac (Stratum 3)
Dodoma	52.38	16.49	0.05	151.66	51.70	0.04	341.95	98.56	0.11
Arusha	30.84	18.02	0.06	148.72	129.94	0.14	345.27	302.65	0.16
Kilimanjaro	72.32	37.37	0.09	190.26	101.52	0.21	507.20	194.43	0.59
Tanga	50.00	14.12	0.06	149.55	41.99	0.13	447.79	128.20	0.24
Morogoro	52.56	7.37	0.07	201.88	28.92	0.14	517.65	82.74	0.47
Pwani	47.46	7.09	0.11	193.71	21.67	0.30	663.69	52.00	0.56
Dar es Salaam	6.35	1.52	0.05	22.65	5.04	0.14	NaN	NaN	NaN
Lindi	84.74	4.46	0.04	260.31	13.96	0.17	NaN	NaN	NaN
Mtwara	63.18	3.15	0.07	135.75	5.18	0.07	272.54	8.60	0.05
Ruvuma	64.89	13.92	0.20	143.75	25.04	0.42	398.78	53.67	1.60
Iringa	53.25	8.96	0.06	135.73	21.88	0.12	383.65	45.90	0.18
Mbeya	46.26	20.06	0.11	201.59	72.95	0.29	910.72	401.39	1.00
Singida	72.47	37.59	0.04	172.65	85.01	0.06	435.54	181.82	0.08
Tabora	62.93	31.12	0.03	193.59	77.25	0.10	543.67	189.42	0.35
Rukwa	77.53	26.96	0.08	236.05	90.69	0.18	NaN	NaN	NaN
Kigoma	70.38	9.11	0.05	236.50	30.49	0.15	813.22	150.22	0.46
Shinyanga	46.44	19.36	0.02	115.16	49.80	0.03	302.34	76.52	0.09
Kagera	80.34	15.83	0.07	211.33	38.38	0.11	571.02	95.10	0.39
Mwanza	25.84	9.75	0.04	108.20	40.17	0.06	354.65	59.46	0.25
Mara	42.33	17.61	0.03	109.54	52.94	0.05	228.31	112.19	0.16
Manyara	83.51	55.19	0.03	188.35	123.90	0.15	407.25	256.76	0.10
Njombe	75.53	15.22	0.10	184.29	29.05	0.26	NaN	NaN	NaN
Katavi	70.12	15.61	0.06	232.44	72.92	0.07	610.36	179.55	0.17
Simiyu	61.07	31.70	0.01	111.33	55.99	0.02	199.79	79.96	0.02
Geita	96.53	28.51	0.05	424.99	59.39	0.16	1862.91	167.64	0.73
Songwe	66.09	25.44	0.26	193.07	74.57	0.26	498.67	151.00	0.20
Kaskazini Unguja	21.92	4.39	0.19	47.25	8.29	0.16	77.89	18.05	0.41
Kusini Unguja	42.31	12.24	0.24	88.73	26.79	0.65	NaN	NaN	NaN
Mjini Magharibi	9.27	1.67	0.08	36.09	8.16	0.26	NaN	NaN	NaN
Kaskazini Pemba	33.05	9.02	0.14	61.12	20.43	0.27	90.68	33.85	0.43
Kusini Pemba	32.53	8.54	0.16	61.21	16.71	0.47	NaN	NaN	NaN

Table 1. Means of the stratifying variables in each stratum and domain.

Sample size: The total sample size is calculated considering precision requirements of the estimates within each domain. The reference variables used to calculate the sample size are the planted area and the number of Tropical Livestock Units (TLUs) as collected in the Agricultural Sample Census 2019/2020[1] (AC20). Details on the choice of these variables are given in Annex 1.

Variables related to aquaculture and beekeeping have not been considered for the sample size calculation because very few sampling units of the AC20 reported information on these sectors, with the consequence that for some domains they cannot be computed because the holdings sampled in them did not practice aquaculture or beekeeping; for other domains the CVs of these variables are too high, increasing too much cv^2_{ACd} in Formula (1) and, consequently, the sample size required for these domains. Therefore, the survey does not aim at producing reliable estimates for the aquaculture and beekeeping sectors. However, the questionnaire of AASS 2023 will contain a module on aquaculture and beekeeping to collect some information on these sectors and allow to compute a sample size for AASS 2024 that allows to produce reliable estimates for aquaculture and beekeeping in the next survey round.

[1] In the parameters' computation using AC20 data the agricultural households that don't respect the minimum thresholds (0.00617763 for area planted or 1 for cattle or 5 for small ruminants or 50 for small animals) and those presenting outliers for the variables of interest have been excluded.

In each domain the required sample size is computed according to the following formula:

$$\tilde{n}_d = \frac{1}{g_d} n_{ACd} \frac{cv_{ACd}^2}{cv_d^{*2}} \quad , \quad (1)$$

where g_d is the expected response rate in domain *d* (it varies from 0.8 to 0.9), n_{ACd} is the sample size used in domain *d* in AC20, cv_{ACd} is usually the maximum between the CV of the average planted area (cv_{PAd}) and the CV of the average number of TLUs (cv_{TLUd}) as estimated using AC20 data and cv_d^* is the maximum acceptable CV for agricultural estimates in domain *d* for the AASS23. The latter term is set equal to 10% for the regions that contribute most to the agricultural sector in terms of planted area and TLUs. In some regions it is set equal to 15%, 20% or 25% according to the contribution of the domain to the agricultural sector. The value of cv_{ACd} in some regions can be equal to cv_{PAd} or cv_{TLUd} or also to their average, depending on the contribution of crop cultivation or livestock production to the agricultural sector within the domain. The sample sizes as obtained through this procedure are domain are shown in Table 2. In order to find the number of EAs to sample in each region, it is enough to take the upper integer part of the ratio between the domain sample size and the number of agricultural households that will be selected in each EA and that is set equal to 12. The total number of EAs is then calculated to be equal to 1,277. After a careful revision of these theoretical sample sizes which implied also simulations conducted from the AC20 data and after consultations with NBS and OCGS, it was decided to decrease the EAs upper bound to 80 (instead of 120) and to increase the lower bound to 20 (instead of 4) in Mainland and 13 in Zanzibar, for which a separate budget is available, so that it was possible to increase the overall sample size in its 5 Regions (Kaskazini Unguja, Kusini Unguja, Mjini Magharibi, Kaskazini Pemba, Kusini Pemba). The total sample size was then adjusted to be **16,224** agricultural holdings in **1,352 EAs**.

Region	Area Planted (%)	TLU (%)	Average contribution area&TLU	CV (%) Area Planted	CV (%) - TLU	Realized sample	max CV (%)	CV* (%)	Response rate	sample size (ag. households)	sample size (EA)	sample size (EA) adjusted	sample size (ag. households) adjusted
Dodoma	10.0%	6.7%	8.4%	4.0	7.5	1,209	7.5	10.0	0.90	755	63	63	756
Arusha	1.5%	4.8%	3.1%	7.8	12.2	1,339	12.2	15.0	0.90	977	81	81	972
Kilimanjaro	1.6%	2.5%	2.1%	4.4	13.3	1,166	13.3	20.0	0.90	574	48	48	576
Tanga	4.4%	3.7%	4.0%	5.6	11.5	1,461	11.5	15.0	0.90	958	80	80	960
Morogoro	4.9%	2.7%	3.8%	3.9	15.5	1,619	9.7	15.0	0.85	796	66	66	792
Pwani	2.2%	2.1%	2.1%	4.8	35.8	1,081	4.8	15.0	0.85	128	11	25	300
Dar Es Salaam	2.2%	0.1%	1.1%	9.6	15.9	356	9.6	15.0	0.70	211	18	20	240
Lindi	4.1%	0.8%	2.5%	5.3	9.0	702	5.3	15.0	0.90	97	8	20	240
Mtwara	4.7%	0.6%	2.6%	3.1	7.3	1,140	3.1	15.0	0.90	53	4	25	300
Ruvuma	6.3%	1.8%	4.0%	4.0	6.9	1,338	6.9	15.0	0.90	312	26	40	480
Iringa	2.4%	2.0%	2.2%	5.8	17.6	570	17.6	20.0	0.80	555	46	46	552
Mbeya	4.0%	2.8%	3.4%	5.5	11.1	1,382	11.1	15.0	0.90	837	70	70	840
Singida	3.9%	5.7%	4.8%	4.3	7.0	969	7.0	15.0	0.90	237	20	25	300
Tabora	7.6%	7.2%	7.4%	4.2	9.5	1,277	9.5	10.0	0.80	1,440	120	80	960
Rukwa	3.5%	2.6%	3.1%	4.5	9.8	798	9.8	20.0	0.85	226	19	20	240
Kigoma	3.6%	2.8%	3.2%	5.9	8.6	853	8.6	20.0	0.90	175	15	25	300
Shinyanga	3.9%	5.6%	4.7%	3.5	5.6	1,193	5.6	15.0	0.90	186	16	30	360
Kagera	4.2%	3.1%	3.6%	4.0	11.3	992	11.3	20.0	0.85	371	31	35	420
Mwanza	3.7%	8.3%	6.0%	4.9	8.8	1,278	8.8	10.0	0.85	1,172	98	90	1,080
Mara	2.1%	6.7%	4.4%	3.8	9.5	1,139	9.5	15.0	0.85	543	45	45	540
Manyara	3.7%	7.5%	5.6%	7.2	8.7	1,167	8.7	10.0	0.90	977	81	80	960
Njombe	1.9%	1.2%	1.6%	4.9	7.9	846	7.9	20.0	0.90	146	12	20	240
Katavi	1.3%	1.9%	1.6%	6.5	13.5	647	13.5	20.0	0.80	371	31	31	372
Simiyu	3.5%	6.4%	5.0%	4.6	8.9	892	8.9	15.0	0.90	349	29	29	348
Geita	4.1%	6.6%	5.3%	5.8	10.2	831	10.2	10.0	0.80	1,074	90	80	960
Songwe	3.6%	3.1%	3.3%	7.6	24.4	672	24.4	20.0	0.80	1,247	104	60	720
Kaskazini Unguja	0.2%	0.1%	0.2%	5.2	14.1	427	14.1	25.0	0.90	152	13	35	420
Kusini Unguja	0.1%	0.1%	0.1%	5.2	12.3	275	12.3	25.0	0.90	75	6	17	204
Mjini Magharibi	0.2%	0.2%	0.2%	5.4	12.1	271	12.1	25.0	0.90	71	6	16	192
Kaskazini Pemba	0.3%	0.2%	0.2%	5.1	16.1	342	16.1	25.0	0.90	159	13	37	444
Kusini Pemba	0.5%	0.2%	0.3%	4.2	9.4	345	9.4	20.0	0.90	86	7	13	156
Tanzania	100.0%	100.0%		1.3	3.19972	28,577				15,310	1,277	**1,352**	**16,224**
Zanzibar												**118**	**1,416**
Mainland												**1,234**	**14,808**

Table 2. Sample size computed by domain.

Sample allocations to strata: the EAs allocations to the strata are computed according to a multivariate criterion which follows the formula below:

$$m_{hd} = m_d \frac{\frac{M_{hd}}{M_d} V_{hd}}{\sum_{g=1}^{H} \frac{M_{gd}}{M_d} V_{gd}}, \qquad (2)$$

where m_{hd} is the sample size allocated to stratum h in domain d, m_d is the sample size in domain d as computed in the last column of Table 1, M_{hd} is the number of EAs in stratum h in domain d, M_d is the number of EAs in domain d, H is the total number of strata and V_{hd} is a convex combination of the single variables' variances in stratum h and domain d, i.e. $V_{hd}^2 = \sum_{k=1}^{K} q_{kd} S_{hkd}^2$, with S_{hkd}^2, q_{kd} and K indicating respectively the variance of variable k in stratum h and domain d, the coefficient of importance of the variable k and the total number of variables considered in the allocation procedure. In our case K is equal to three and the stratifying variables are the number of households growing crops ($k = 1$), number of households rearing livestock ($k = 2$) and number of households practicing aquaculture ($k = 3$), whose coefficients qs are respectively 0.4, 0.4 and 0.2 in each domain and stratum. In some domains the number of strata is set equal to two because of the low sample size. The final EAs allocations by domain are shown in Table 3. An example of strata allocations in Dodoma is given in Annex 2.

	Allocations – EAs			
Region	Stratum1	Stratum2	Stratum3	Total sample size
Dodoma	27	23	13	63
Arusha	27	38	16	81
Kilimanjaro	25	18	5	48
Tanga	44	29	7	80
Morogoro	34	22	10	66
Pwani	15	7	3	25
Dar es Salaam	13	7	NA	20
Lindi	13	7	NA	20
Mtwara	11	8	6	25
Ruvuma	23	14	3	40
Iringa	23	18	5	46
Mbeya	46	19	5	70
Singida	13	9	3	25
Tabora	48	23	9	80
Rukwa	11	9	NA	20
Kigoma	15	7	3	25
Shinyanga	15	11	4	30
Kagera	18	13	4	35
Mwanza	39	40	11	90
Mara	17	19	9	45
Manyara	37	29	14	80
Njombe	13	7	NA	20
Katavi	16	8	7	31
Simiyu	13	10	6	29
Geita	57	19	4	80
Songwe	30	20	10	60
Kaskazini Unguja	11	15	9	35
Kusini Unguja	10	7	NA	17
Mjini Magharibi	11	5	NA	16
Kaskazini Pemba	12	15	10	37
Kusini Pemba	7	6	NA	13

Table 3. EAs allocations to sampling strata in each domain.

Calculation of sampling weights. The sampling weight assigned to the agricultural households *j* in EA *i* in domain *d* and stratum *h* is calculated as follows:

$$w_{jihd} = 1/[\left(m_{hd}\frac{F_{ihd}}{F_{hd}}\right) * \left(\frac{n_{ihd}}{N_{ihd}}\right)], \quad (3)$$

where m_{hd} is the number of EAs selected in stratum *h* in domain *d,* F_{ihd} is the total number of agricultural households in the *i-th* EA and stratum *h* in domain *d* as listed in the sampling frame, F_{hd} is the total number of agricultural households in stratum *h* and domain *d* as listed in the sampling frame, n_{ihd} and N_{ihd} are respectively the number of agricultural households sampled and found after the listing operation in the *i-th* EA in stratum *h* and domain *d*.

Estimation. In order to estimate the total $\widehat{Y}$ of a variable Y, it's enough to apply the following formula:

$$\sum_{d=1}^{31}\sum_{h=1}^{H}\sum_{i=1}^{m_{hd}}\sum_{j=1}^{n_{ihd}} w_{jihd}y_{jihd}, \quad (4)$$

Where y_{jihd} is the value of the variable *Y* for unit *j* in the *i-th EA* in stratum *h* and domain *d* and *H* is the total number of strata in domain *d*.

An approximation of the variance of the total $\widehat{Y}$ considering only the PSU component is given by:

$$V(\widehat{Y}) = \sum_{d=1}^{25}\sum_{h=1}^{H}\frac{{M_{hd}}^2}{m_{hd}(m_{hd}-1)}\sum_{i=1}^{m_{hd}}\left(\widehat{Y}_{ihd} - \frac{1}{m_{hd}}\sum_{i=1}^{m_{hd}}\widehat{Y}_{ihd}\right)^2, \quad (5)$$

where M_{hd} is the total number of EAs found in stratum *h* and domain *d* and $\widehat{Y}_{ihd}$ is the estimate of the total amount of *Y* in the *i-th* PSU in stratum *h* and domain *d.*

Annex 1. Key variables for the calculation of the sample size.

In order to compute the sample size for the Annual Agricultural Sample Survey 2023, we need to consider the targeted and realized precision of few key variables correlated to the overall set of variables that are going to be estimated using AASS23 data. It's not easy to select few representative variables for all the hundreds estimates that are usually produced through a survey. For the agricultural surveys, the choice usually falls on the *planted area* for the crop-related estimates and on the *number of TLUs* for the livestock-related estimates. The derivative variables *crop production* and *livestock production* are sometimes used, even though their variability (e.g. CV) is usually higher because they are obtained through quite

complex operations on several original variables. It is also preferable to focus on the planted area than on the harvested area because the latter is usually collected on less holdings and, consequently, it may be less representative. If the agricultural survey aims at producing important estimates on fishery and aquaculture, it is necessary to include key variables for these sectors in the computation of the sample size, such as the area devoted to aquaculture activities or the number of fishes harvested. For the household surveys, the choice of the key variables usually falls on the household's income or on variables related to poverty indexes usually identified by the analyst, such as the access to electricity.

For the Annual Agricultural Sample Survey 2023 the key variables selected for the computation of the sample size were the *planted area* and the *number of TLUs.* With *planted area* is intended the operated land under temporary, permanent crops, planted trees and fallow (categories 5.1.1., 5.1.2., 5.1.3., 5.1.4., 5.1.5., 5.1.8 and 5.1.10 of Question 5.1. in the AC20 questionnaire). Since the planted area by crop is collected both for the short rainy season and the long rainy season it was preferred to choose the variable generated by the more general Question 5.1. as reference for the calculation of the sample size. Using the AC20 data, it has been shown that the holding's planted area (computed summing up the planted area by crop) is extremely correlated with the holding's harvested area (computed summing up the harvested area by crop), with a correlation greater than 90%.

Annex 2. Example: calculation of sample size and strata allocations in Dodoma.

During the Agricultural Sample Census 2019/2020, 1,209 agricultural households were sampled in Dodoma ($n_{ACdodoma} = 1{,}209$) whose contribution to the national agricultural sector is 10% in terms of area planted and 6.7% in terms of TLUs. Given the high contribution both in terms of crop cultivation and livestock, cv_d^* is set equal to 10% and $cv_{ACdodoma}$ is the maximum between $cv_{PAdodoma}$ (=4%) and $cv_{TLUdodoma}$ (=7.49%), i.e. $cv_{ACdodoma}$=7.49%.

Hence, from Formula (1) the sample size in Dodoma is calculated to be:

$$\tilde{n}_d = \frac{1}{g_{dodoma}} n_{ACdodoma} \frac{cv_{ACdodoma}^2}{cv_{dodoma}^{*2}} = \tilde{n}_d = \frac{1}{0.9} 1{,}209 \frac{7.49^2}{10^2} = 755$$

To compute the number of EAs to be sampled in Dodoma (m_{dodoma}), it is enough to divide $\tilde{n}_{dodoma}$ by 12 and take the integer part, i.e.:

$$m_{dodoma} = \left\lfloor \frac{\tilde{n}_{dodoma}}{12} \right\rfloor = \left\lfloor \frac{755}{12} \right\rfloor = 63$$

The EAs in Dodoma have been stratified in three strata according to the number of households growing crops, rearing livestock and practicing aquaculture. From Table 1 and Figure 1 it is possible to see that the third stratum is composed by EAs which includes on average 341.95 households cultivating crops, 98.56 households rearing livestock, 0.11 households practicing aquaculture. In the other two strata EAs contain on average less

agricultural households, particularly stratum one is formed by EAs where households are less intensively engaged in crop cultivation, livestock rearing and aquaculture.

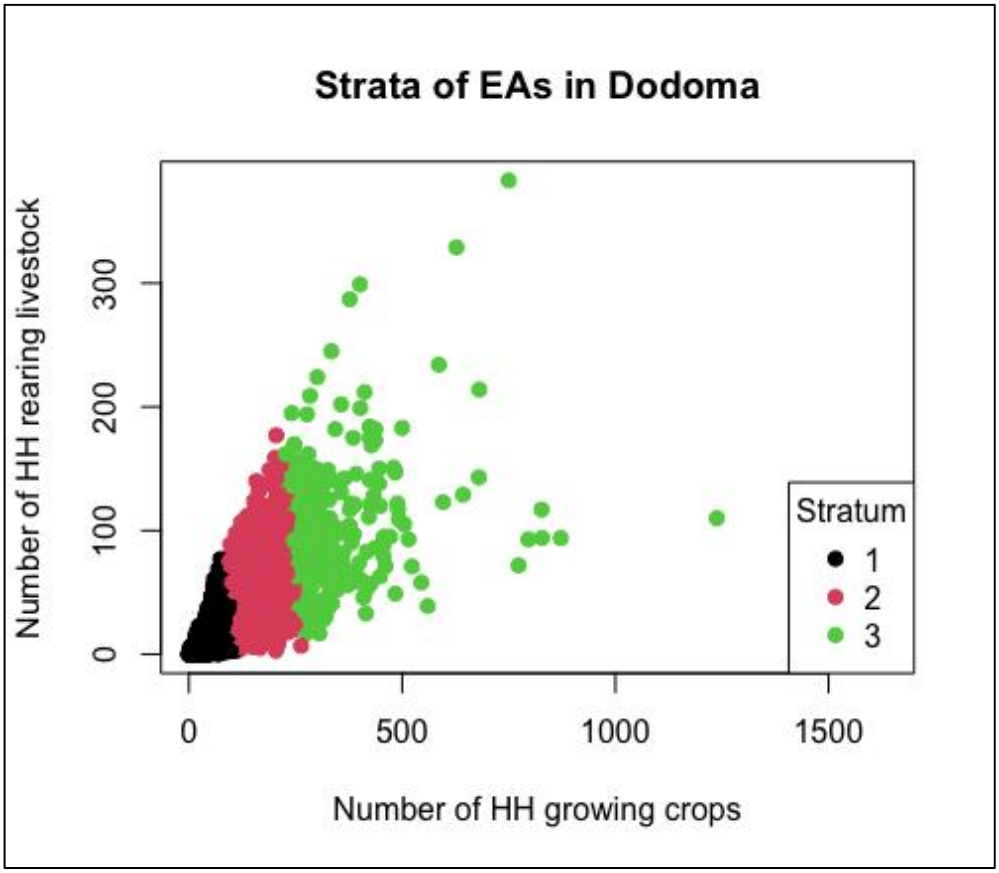

Figure 1. Enumeration Areas stratified according to the number of households growing crops and rearing livestock.

Formula 2 allows to take into consideration both the numerosity of the strata in Dodoma ($M_{1dodoma} = 2755, M_{2dodoma} = 1679, M_{3dodoma} = 291$) , the variances of the three stratifying variables in stratum one ($S^2_{11dodoma}$=853.71, $S^2_{12dodoma}$=211.66, $S^2_{13dodoma}$=0.065), in stratum two ($S^2_{21dodoma}$ =1525.71, $S^2_{22dodoma}$ =585, $S^2_{23dodoma}$ =0.055), in stratum three ($S^2_{31dodoma}$=20059.31, $S^2_{32dodoma}$=2257.48, $S^2_{33dodoma}$=0.123) and the coefficients q_1, q_1 and q_3 set respectively to 0.4, 0.4 and 0.2 in all the domains and strata.

According to Formula 2 and considering $V^2_{1dodoma} = \sum_{k=1}^{3} q_k S^2_{1kdodoma} = 426.163$, $V^2_{2dodoma} = \sum_{k=1}^{3} q_k S^2_{2kdodoma} = 844.32$, $V^2_{3dodoma} = \sum_{k=1}^{3} q_k S^2_{3kdodoma} = 9046.728$, the allocations to stratum 1, stratum 2, stratum 3 are calculated to be:

$$m_{1dodoma} = m_{dodoma} \frac{\frac{M_{1dodoma}}{M_{dodoma}} V_{1dodoma}}{\sum_{g=1}^{H} \frac{M_{gdodoma}}{M_{dodoma}} V_{gdodoma}}$$

$$= 63 * \frac{\frac{2755}{4725} * 20.64}{\frac{2755}{4725} * 20.64 + \frac{1679}{4725} * 29 + \frac{291}{4725} * 95.11} = 26.88,$$

$$m_{2dodoma} = m_{dodoma} \frac{\frac{M_{2dodoma}}{M_{dodoma}} V_{2dodoma}}{\sum_{g=1}^{H} \frac{M_{gdodoma}}{M_{dodoma}} V_{gdodoma}}$$

$$= 63 * \frac{\frac{1679}{4725} * 29}{\frac{2755}{4725} * 20.64 + \frac{1679}{4725} * 29 + \frac{291}{4725} * 95.11} = 23.02,$$

$$m_{3dodoma} = m_{dodoma} \frac{\frac{M_{3dodoma}}{M_{dodoma}} V_{3dodoma}}{\sum_{g=1}^{H} \frac{M_{gdodoma}}{M_{dodoma}} V_{gdodoma}}$$

$$= 63 * \frac{\frac{291}{4725} * 95.11}{\frac{2755}{4725} * 20.64 + \frac{1679}{4725} * 29 + \frac{291}{4725} * 95.11} = 13.08,$$

Rounding the three values to the nearest integer we get exactly the allocations given in Table 3, i.e. 27, 23, 13.