Example Output with annotation Annotation

# Using the output files

Before we go into detail, these country-level files can be helpful in several ways:

1. The output provides candidate sets of variables (about) equivalent to their ability to group (cluster) subjects by SES. We rank these sets by our measure of cluster performance, called the average silhouette width (ASW), where bigger means better clusters (more similarity among households in the same cluster, more significant differences among households in different clusters). If you are designing your study, these variable sets can be used, informed by your knowledge of the local context, to select which SES/asset variables to include on a future survey where determining SES status is of interest.
2. We also report which variables are consistently chosen among the variables that result in the best-performing clusters. This can also aid in selecting a variable list for a future study.
3. Provide the asset variable patterns in each cluster but only among the top-performing variable set. These can help interpret and/or rank the clusters from lowest to highest SES.
4. We show the statistical relationship of how we rank the clusters (based on the average number of assets in the cluster) to the validation variables and provide a formal inference of this relationship. The greater the magnitude of the association (in the expected direction) and the smaller the p-value, the stronger the evidence for the proposed SES ranking of clusters.

# Example Output: Cameroon

Below, we provide each section of the output and some annotations to help interpret and use it.

## Header Info

Economic Clustering Summary Report: Cameroon 2018

**Data Summary**

**Country Code-year:** CM18 (DHS code)

**Number of observations:** 11710 (total number of households in original survey data)

**Number of variables used:** 32 (number of potential asset variables available for clustering households)

**Variables used in the algorithm:** hv206, hv207, hv208, hv209, hv211, hv225, hv227, hv230b, hv232, hv232y, hv237, hv243a, hv243b, hv244, hv246, hv247, hv252, sh121h, sh121j, sh121k, sh121l, sh121m, sh121o, sh121s, sh122h, sh123a, water, toilet, floor, roof, cookfuel, wall (the DHS variable names of those 32 potential clustering variables).

## Top 10 Variable Sets (Sorted by ASW)

Below are the 10 asset variable sets that result in the 10 most optimal clustering configurations. There is no meaningful ordering of the variables within each set. The ASW column reports the measure of cluster performance that goes from -1 (worst) to 1 (perfect).

| Set | ASW | Var.1 | Var.2 | Var.3 | Var.4 |
| --- | --- | --- | --- | --- | --- |
| 1 | 0.9476 | has television | has mobile telephone | mixer | cable |
| 2 | 0.9381 | has television | has mobile telephone | cable | laptop computer |
| 3 | 0.9319 | has electricity | has television | mixer | cable |
| 4 | 0.9301 | has television | mixer | cable | roof |
| 5 | 0.9245 | has electricity | has television | cable | laptop computer |
| 6 | 0.9222 | has electricity | has mobile telephone | mixer | cable |
| 7 | 0.9217 | has television | has mobile telephone | has bank account | cable |
| 8 | 0.9202 | has television | cable | laptop computer | roof |
| 9 | 0.9184 | has electricity | has television | has mobile telephone | cable |
| 10 | 0.9183 | has television | has refrigerator | has mobile telephone | cable |

## Distributions of the SES variables chosen in top 10 cluster configuration sets and the proportion of times they were included in these top configurations (NOT SURE ABOUT THIS SO CORRECT)

This table provides some basic information on the marginal distribution of the levels of each variable, as well as the % of time they were included in the top cluster configuration sets (a measure of how important the variable is to clustering households).

| Variable | Description | % in Top Sets | Distribution |
| --- | --- | --- | --- |
| hv208 | has television | 90% | Binary, 52% (1/yes) |
| hv206 | has electricity | 40% | Binary, 64% (1/yes) |
| hv243a | has mobile telephone | 60% | Binary, 85.6% (1/yes) |
| sh121m | mixer | 40% | Binary, 12.7% (1/yes) |
| sh121o | cable | 100% | Binary, 38.3% (1/yes) |
| hv209 | has refrigerator | 10% | Binary, 22.3% (1/yes) |
| hv247 | has bank account | 10% | Binary, 16.3% (1/yes) |
| sh122h | laptop computer | 30% | Binary, 12.3% (1/yes) |
| roof | roof | 20% | Categorical, 0 (finished) = 84.1%,   1 (natural) = 13.1%,   2 (other) = 0%,   3 (rudimentary) = 2.8% |

## Patterns that define clusters

# Cluster Configuration (Set #1)

| Cluster Group | has television | has mobile telephone | mixer | cable | Proportion (%) |
| --- | --- | --- | --- | --- | --- |
| 1 | 0 | 1 | 1 | 1 | 11.32 |
| 1 | 1 | 0 | 1 | 1 |  |
| 1 | 1 | 1 | 1 | 1 |  |
| 2 | 1 | 0 | 0 | 0 | 13.97 |
| 2 | 1 | 0 | 1 | 0 |  |
| 2 | 1 | 1 | 0 | 0 |  |
| 2 | 1 | 1 | 1 | 0 |  |
| 3 | 1 | 0 | 0 | 1 | 26.71 |
| 3 | 1 | 1 | 0 | 1 |  |
| 4 | 0 | 1 | 0 | 0 | 35.68 |
| 4 | 0 | 1 | 0 | 1 |  |
| 4 | 0 | 1 | 1 | 0 |  |
| 5 | 0 | 0 | 0 | 0 | 12.32 |
| 5 | 0 | 0 | 1 | 0 |  |

## Relationship of cluster order and validation variables.

First, we only provide these results for the top cluster configuration set (that with the highest ASW). There are three tables per country based on other validation variables, but we will illustrate how to interpret these tables with child mortality. As one goes lower in the rows, the clusters are estimated to be poorer, and as one goes farther to the right in columns, the outcomes are worse (higher % of household child mortality). The p-value relates to a chi-square association test (IS THIS TRUE OR IS THIS BASED ON THE ORDERED LOGISTIC REGRESSION?). In this case, one can see a clear dose response in that the clusters with lower estimated asset wealth have higher child mortality. The p-value indicates that this relationship is significant.

**Using Children Deceased (Sorted by proportion of 0%)**

(higher SES as one goes down)

| Cluster ID/Children Deceased | 0% | 1-33% | 34-66% | 67+% |
| --- | --- | --- | --- | --- |
| 1\* | 1,638 (89.8%) | 119 (6.5%) | 55 (3.0%) | 12 (0.7%) |
| 3 | 3,364 (85.9%) | 386 (9.9%) | 132 (3.4%) | 34 (0.9%) |
| 2 | 1,833 (80.3%) | 321 (14.1%) | 105 (4.6%) | 25 (1.1%) |
| 4 | 3,966 (75.0%) | 954 (18.1%) | 288 (5.4%) | 77 (1.5%) |
| 5 | 926 (67.7%) | 261 (19.1%) | 135 (9.9%) | 46 (3.4%) |
| Total | 11,727 (79.9%) | 2,041 (13.9%) | 715 (4.9%) | 194 (1.3%) |
| \*The chi-squared p-value is 0 | | | | |