Economic Clustering Summary Report: Kenya 2020

Nolan Gunter

17 September, 2023

# Summary of Data

**Country Code-year:** KE20

**Number of observations:** 7952

**Number of clusters:** 5

**Number of variables used:** 29

**Distance used:** Hamming

**Variables used in the algorithm:** hv206, hv207, hv208, hv210, hv211, hv225, hv227, hv243a, hv243b, hv244, hv246, hv246a, hv246e, hv247, sh116g, sh116h, sh116i, sh116j, sh116l, sh116m, sh116o, sh119, sh119a, water, toilet, floor, roof, cookfuel, wall

# Top Variable Groupings (Sorted by ASW)

| Group | ASW | Var. 1 | Var. 2 | Var. 3 | Var. 4 |
| --- | --- | --- | --- | --- | --- |
| 1 | 0.9178 | has television | has mobile telephone | has dvd player | roof |
| 2 | 0.9161 | has motorcycle/scooter | has table | has chair | roof |
| 3 | 0.9153 | has table | has chair | has dvd player | roof |
| 4 | 0.9141 | has table | has clock | has dvd player | roof |
| 5 | 0.9135 | has motorcycle/scooter | has table | member use phone for financial transactions | roof |
| 6 | 0.9128 | has table | has dvd player | member use phone for financial transactions | roof |
| 7 | 0.9128 | has mobile telephone | has table | has chair | has dvd player |
| 8 | 0.9125 | has table | has chair | has clock | roof |
| 9 | 0.9125 | has chair | has clock | has dvd player | roof |
| 10 | 0.9122 | has motorcycle/scooter | has chair | member use phone for financial transactions | roof |

# Marginal Distributions

| Variable | Description | % Time in Top Clusters | Distribution |
| --- | --- | --- | --- |
| hv208 | has television | 10% | Binary, 42.4% 1s (or yes) |
| hv211 | has motorcycle/scooter | 30% | Binary, 12.8% 1s (or yes) |
| sh116h | has table | 70% | Binary, 84.3% 1s (or yes) |
| hv243a | has mobile telephone | 20% | Binary, 88.6% 1s (or yes) |
| sh116i | has chair | 60% | Binary, 88.1% 1s (or yes) |
| sh116m | has clock | 30% | Binary, 16.1% 1s (or yes) |
| sh116o | has dvd player | 60% | Binary, 12.6% 1s (or yes) |
| sh119 | member use phone for financial transactions | 30% | Binary, 85.4% 1s (or yes) |
| roof | roof | 90% | Categorical, 0 (finished) = 93%,   1 (natural) = 5.1%,   2 (other) = 1%,   3 (rudimentary) = 0.9% |

# Cluster #1 Configuration

| Cluster Group | has television | has mobile telephone | has dvd player | roof | Proportion (%) |
| --- | --- | --- | --- | --- | --- |
| 1 | 0 | 0 | 1 | finished | 12.55 |
| 1 | 0 | 0 | 1 | natural |  |
| 1 | 0 | 1 | 1 | finished |  |
| 1 | 0 | 1 | 1 | natural |  |
| 1 | 1 | 0 | 1 | finished |  |
| 1 | 1 | 1 | 1 | finished |  |
| 1 | 1 | 1 | 1 | natural |  |
| 1 | 1 | 1 | 1 | rudimentary |  |
| 2 | 0 | 1 | 0 | natural | 3.45 |
| 2 | 1 | 1 | 0 | natural |  |
| 3 | 0 | 0 | 0 | finished | 11.05 |
| 3 | 0 | 0 | 0 | natural |  |
| 3 | 0 | 0 | 0 | other |  |
| 3 | 0 | 0 | 0 | rudimentary |  |
| 3 | 1 | 0 | 0 | finished |  |
| 3 | 1 | 0 | 0 | natural |  |
| 3 | 1 | 0 | 0 | rudimentary |  |
| 4 | 0 | 1 | 0 | finished | 43.61 |
| 4 | 0 | 1 | 0 | other |  |
| 4 | 0 | 1 | 0 | rudimentary |  |
| 4 | 1 | 1 | 0 | rudimentary |  |
| 5 | 1 | 1 | 0 | finished | 29.34 |

# Validation Tables

## a.1) Using Children Deceased (Sorted by proportion of 0%)

| Cluster ID/Children Deceased | 0% | 1-33% | 34-66% | 67+% |
| --- | --- | --- | --- | --- |
| 1\* | 819 (93.5%) | 45 (5.1%) | 9 (1.0%) | 3 (0.3%) |
| 3 | 2,178 (91.8%) | 147 (6.2%) | 44 (1.9%) | 3 (0.1%) |
| 2 | 197 (90.0%) | 20 (9.1%) | 2 (0.9%) | 0 (0.0%) |
| 5 | 2,561 (89.0%) | 229 (8.0%) | 71 (2.5%) | 15 (0.5%) |
| 4 | 380 (88.8%) | 36 (8.4%) | 10 (2.3%) | 2 (0.5%) |
| Total | 6,135 (90.6%) | 477 (7.0%) | 136 (2.0%) | 23 (0.3%) |
| \*The chi-squared p-value is 0.002 | | | | |

## a.2) Aggregating proportions greater than 0%

| Cluster ID/Children Deceased | 0% | >0% |
| --- | --- | --- |
| 1\* | 819 (93.5%) | 57 (6.5%) |
| 3 | 2,178 (91.8%) | 194 (8.2%) |
| 2 | 197 (90.0%) | 22 (10.0%) |
| 5 | 2,561 (89.0%) | 315 (11.0%) |
| 4 | 380 (88.8%) | 48 (11.2%) |
| Total | 6,135 (90.6%) | 636 (9.4%) |
| \*The chi-squared p-value is 0 | | |

## b) Using Individual Education Level Attained (Sorted by weighted average by row)

| Cluster ID/Education | 0 | 1 | 2 | 3 | 4 | 5 | W. Avg. |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1\*a | 6 (0.7%) | 70 (8.0%) | 150 (17.1%) | 175 (20.0%) | 214 (24.4%) | 261 (29.8%) | 3.49 |
| 3 | 84 (3.5%) | 352 (14.8%) | 464 (19.6%) | 547 (23.1%) | 500 (21.1%) | 425 (17.9%) | 2.97 |
| 5 | 268 (9.3%) | 731 (25.4%) | 699 (24.3%) | 596 (20.7%) | 394 (13.7%) | 188 (6.5%) | 2.24 |
| 4 | 99 (23.1%) | 155 (36.2%) | 75 (17.5%) | 66 (15.4%) | 26 (6.1%) | 7 (1.6%) | 1.50 |
| 2 | 106 (48.4%) | 54 (24.7%) | 25 (11.4%) | 17 (7.8%) | 16 (7.3%) | 1 (0.5%) | 1.02 |
| Total | 563 (8.3%) | 1,362 (20.1%) | 1,413 (20.9%) | 1,401 (20.7%) | 1,150 (17.0%) | 882 (13.0%) | 2.57 |
| \*The chi-squared p-value is 0 | | | | | | | |
| a0=none,1=incomplete primary, 2=primary, 3=incomplete secondary, 4=secondary, 5=higher | | | | | | | |

## c) Using Primary Healthcare Source (Sorted by % enrolled in public healthcare [ascending order])