Economic Clustering Summary Report: Uganda 2016

Andy Kim

02 May, 2024

# Summary of Data

**Country Code-year:** UG16

**Number of observations:** 19519

**Number of clusters:** 5

**Number of variables used:** 29

**Distance used:** Hamming

**Variables used in the algorithm:** hv206, hv207, hv208, hv210, hv211, hv225, hv227, hv237, hv243a, hv243b, hv244, hv246, hv247, sh120a, sh121g, sh121h, sh121i, sh121j, sh121k, sh121l, sh121m, sh125, sh144b, water, toilet, floor, roof, cookfuel, wall

# Top Variable Groupings (Sorted by ASW)

| Group | ASW | Var. 1 | Var. 2 | Var. 3 | Var. 4 |
| --- | --- | --- | --- | --- | --- |
| 1 | 0.9252 | has electricity | has television | has cassette/cd/dvd player | has bed |
| 2 | 0.9247 | has television | has cassette/cd/dvd player | has bed | dwelling sprayed in the past 6 months |
| 3 | 0.9231 | has electricity | has television | has cassette/cd/dvd player | dwelling sprayed in the past 6 months |
| 4 | 0.9217 | has electricity | has television | has mobile telephone | has cassette/cd/dvd player |
| 5 | 0.9125 | has television | has cassette/cd/dvd player | floor | roof |
| 6 | 0.9107 | has television | has mobile telephone | has bank account | has cassette/cd/dvd player |
| 7 | 0.9081 | has electricity | has television | has cassette/cd/dvd player | roof |
| 8 | 0.9073 | has electricity | has television | has bed | dwelling sprayed in the past 6 months |
| 9 | 0.9068 | has television | has cassette/cd/dvd player | has sofa set | has bed |
| 10 | 0.9057 | has electricity | has cassette/cd/dvd player | has bed | dwelling sprayed in the past 6 months |

# Marginal Distributions

| Variable | Description | % Time in Top Clusters | Distribution |
| --- | --- | --- | --- |
| hv206 | has electricity | 60% | Binary, 26.2% 1s (or yes) |
| hv208 | has television | 90% | Binary, 14.7% 1s (or yes) |
| sh121g | has cassette/cd/dvd player | 90% | Binary, 10.7% 1s (or yes) |
| hv243a | has mobile telephone | 20% | Binary, 72% 1s (or yes) |
| sh121k | has bed | 50% | Binary, 79.7% 1s (or yes) |
| floor | floor | 10% | Categorical, 0 (37) = 0.1%,   1 (finished) = 33.3%,   2 (natural) = 65.8%,   3 (other) = 0.2%,   4 (rudimentary) = 0.6% |
| hv247 | has bank account | 10% | Binary, 46.5% 1s (or yes) |
| sh121j | has sofa set | 10% | Binary, 23.4% 1s (or yes) |
| sh125 | dwelling sprayed in the past 6 months | 40% | Binary, 10.1% 1s (or yes) |
| roof | roof | 20% | Categorical, 0 (finished) = 67.3%,   1 (natural) = 31.9%,   2 (other) = 0.1%,   3 (rudimentary) = 0.7% |

# Cluster #1 Configuration

| Cluster Group | has electricity | has television | has cassette/cd/dvd player | has bed | Proportion (%) |
| --- | --- | --- | --- | --- | --- |
| 1 | 1 | 1 | 0 | 0 | 5.09 |
| 1 | 1 | 1 | 0 | 1 |  |
| 2 | 0 | 1 | 1 | 0 | 9.80 |
| 2 | 0 | 1 | 1 | 1 |  |
| 2 | 1 | 0 | 1 | 0 |  |
| 2 | 1 | 0 | 1 | 1 |  |
| 2 | 1 | 1 | 1 | 0 |  |
| 2 | 1 | 1 | 1 | 1 |  |
| 3 | 0 | 0 | 0 | 1 | 54.57 |
| 3 | 0 | 0 | 1 | 1 |  |
| 3 | 0 | 1 | 0 | 1 |  |
| 4 | 1 | 0 | 0 | 1 | 10.55 |
| 5 | 0 | 0 | 0 | 0 | 19.99 |
| 5 | 0 | 0 | 1 | 0 |  |
| 5 | 0 | 1 | 0 | 0 |  |
| 5 | 1 | 0 | 0 | 0 |  |

# Validation Tables

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

| Cluster ID/Children Deceased | 0% | 1-33% | 34-66% | 67+% |
| --- | --- | --- | --- | --- |
| 5\* | 1,942 (89.3%) | 175 (8.0%) | 37 (1.7%) | 20 (0.9%) |
| 3 | 953 (87.9%) | 103 (9.5%) | 22 (2.0%) | 6 (0.6%) |
| 4 | 1,542 (80.0%) | 305 (15.8%) | 63 (3.3%) | 18 (0.9%) |
| 1 | 7,717 (75.9%) | 1,917 (18.9%) | 443 (4.4%) | 91 (0.9%) |
| 2 | 2,291 (72.7%) | 677 (21.5%) | 151 (4.8%) | 33 (1.0%) |
| Total | 14,445 (78.1%) | 3,177 (17.2%) | 716 (3.9%) | 168 (0.9%) |
| \*The chi-squared p-value is 0 | | | | |

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

| Cluster ID/Children Deceased | 0% | >0% |
| --- | --- | --- |
| 5\* | 1,942 (89.3%) | 232 (10.7%) |
| 3 | 953 (87.9%) | 131 (12.1%) |
| 4 | 1,542 (80.0%) | 386 (20.0%) |
| 1 | 7,717 (75.9%) | 2,451 (24.1%) |
| 2 | 2,291 (72.7%) | 861 (27.3%) |
| Total | 14,445 (78.1%) | 4,061 (21.9%) |
| \*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. |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 5\*a | 44 (2.0%) | 426 (19.6%) | 233 (10.7%) | 800 (36.8%) | 120 (5.5%) | 551 (25.3%) | 3.00 |
| 3 | 32 (3.0%) | 234 (21.6%) | 134 (12.4%) | 402 (37.1%) | 56 (5.2%) | 226 (20.8%) | 2.82 |
| 4 | 142 (7.4%) | 781 (40.5%) | 285 (14.8%) | 523 (27.1%) | 21 (1.1%) | 176 (9.1%) | 2.01 |
| 1 | 1,050 (10.3%) | 5,384 (53.0%) | 1,446 (14.2%) | 1,912 (18.8%) | 50 (0.5%) | 326 (3.2%) | 1.56 |
| 2 | 803 (25.5%) | 1,730 (54.9%) | 240 (7.6%) | 322 (10.2%) | 7 (0.2%) | 50 (1.6%) | 1.10 |
| Total | 2,071 (11.2%) | 8,555 (46.2%) | 2,338 (12.6%) | 3,959 (21.4%) | 254 (1.4%) | 1,329 (7.2%) | 1.77 |
| \*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])

| Cluster ID/Primary Healthcare Source | 0 | 1 | 2 |
| --- | --- | --- | --- |
| 5\*a | 259 (39.5%) | 389 (59.4%) | 7 (1.1%) |
| 3 | 129 (39.9%) | 192 (59.4%) | 2 (0.6%) |
| 4 | 338 (59.9%) | 216 (38.3%) | 10 (1.8%) |
| 1 | 1,795 (66.7%) | 876 (32.5%) | 22 (0.8%) |
| 2 | 403 (70.7%) | 162 (28.4%) | 5 (0.9%) |
| Total | 2,924 (60.9%) | 1,835 (38.2%) | 46 (1.0%) |
| \*The chi-squared p-value is 0 | | | |
| a0=public/government, 1=private, 2=other | | | |

# Ordered Logisitic Regression Analysis

## Ordered Logistic Using ordinal::clm

This implements a Test of Trend and extracts the p-values for the validation variables. The p-values come from a one-sided hypothesis test that the coefficient of the rank in the model is positive, so that as the rank increases (become more asset heavy) the validation variable increases (assuming its values imply increasing wealth/SES).

## Warning: package 'ordinal' was built under R version 4.1.2

##   
## Attaching package: 'ordinal'

## The following object is masked from 'package:dplyr':  
##   
## slice

**Cluster scoring:** 0

**Ordinal Logistic Regression P-values**

**P-value for Proportion of Children Deceased** 3.4509018^{-64}

**P-value for Educational Attainment** 0