Introductory Econometrics II - Assignment

Kenny

August 10, 2019

## Abstract

This document to shed some light on monthly income, household size, Age of respondent, vulnerability index and marital Status as variables from the Kenya FinAccess Survey (FinAccess), 2018 to answer the following questions. The variables for 47 counties except Turkana County.

## load packages

library(tidyverse)  
library(haven)

The haven package helps us with tools for unpacking the document in stata or spss format. Tidyverse will help in transformoing and interragating the data.

## Load data

data <- read\_sav("2018 Finaccess Data.sav")

The above code show how we read the data into our analysis software.

## Extract the variables

working\_data <- data %>% select(a1, a10, a13, a17, b3h1, vul\_index) %>% filter(a1 != 23) %>% select(-a1)

The above code is how we extract the six variables i.e. including the county names. Note that we have removed Turkana county which is county number 023. Then we remove the county variable because we only need five variables.

## renaming the variables and viewind the first four entries

working\_data <- working\_data %>% rename( "Household Size" = a10, "Age of Respondents"= a13,"Marital Status" = a17,"Monthly Income" = b3h1, "Vulnerability Index" = vul\_index)  
  
head(working\_data,4) %>% knitr::kable()

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Household Size | Age of Respondents | Marital Status | Monthly Income | Vulnerability Index |
| 3 | 69 | 2 | 2000 | 1 |
| 1 | 21 | 2 | 500 | 3 |
| 1 | 38 | 2 | 5000 | 3 |
| 2 | 20 | 4 | 1000 | 3 |

## Correlation

The main aim of this correlation is to estimate the strength of relationship between Monthly income, Household Size, Respondents Age and Vulnerability Index .

correlation <- cor(working\_data, use = "complete.obs")  
round(correlation, 2) %>% knitr::kable()

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Household Size | Age of Respondents | Marital Status | Monthly Income | Vulnerability Index |
| Household Size | 1.00 | -0.15 | 0.04 | -0.09 | -0.12 |
| Age of Respondents | -0.15 | 1.00 | 0.13 | 0.04 | -0.10 |
| Marital Status | 0.04 | 0.13 | 1.00 | 0.04 | 0.01 |
| Monthly Income | -0.09 | 0.04 | 0.04 | 1.00 | 0.16 |
| Vulnerability Index | -0.12 | -0.10 | 0.01 | 0.16 | 1.00 |

The table above shows the correlation between the variables. It is seem that all the variables show are very weak relationship between each other. Other the hand,

## Regression analysis

The main aim of this regression is to estimate the how Monthly income as the dependent variable is likely to be influenced by Household Size, Respondents Age and Vulnerability Index (as the independent variables).

model <- lm(formula = `Monthly Income` ~ `Household Size` + `Age of Respondents` + `Vulnerability Index`, working\_data)