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 .

cor(working\_data, use = "pairwise.complete.obs") %>% knitr::kable()

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Household Size | Age of Respondents | Marital Status | Monthly Income | Vulnerability Index |
| Household Size | 1.0000000 | -0.1530090 | 0.0376277 | -0.0865989 | -0.1163042 |
| Age of Respondents | -0.1530090 | 1.0000000 | 0.1323256 | 0.0372830 | -0.0981772 |
| Marital Status | 0.0376277 | 0.1323256 | 1.0000000 | 0.0363501 | 0.0047085 |
| Monthly Income | -0.0865989 | 0.0372830 | 0.0363501 | 1.0000000 | 0.1646097 |
| Vulnerability Index | -0.1163042 | -0.0981772 | 0.0047085 | 0.1646097 | 1.0000000 |

correlation <- corr.test(working\_data)  
print(correlation, short = F)

## Call:corr.test(x = working\_data)  
## Correlation matrix   
## Household Size Age of Respondents Marital Status  
## Household Size 1.00 -0.15 0.04  
## Age of Respondents -0.15 1.00 0.13  
## Marital Status 0.04 0.13 1.00  
## Monthly Income -0.09 0.04 0.04  
## Vulnerability Index -0.12 -0.10 0.00  
## Monthly Income Vulnerability Index  
## Household Size -0.09 -0.12  
## Age of Respondents 0.04 -0.10  
## Marital Status 0.04 0.00  
## Monthly Income 1.00 0.16  
## Vulnerability Index 0.16 1.00  
## Sample Size   
## Household Size Age of Respondents Marital Status  
## Household Size 8522 8522 8522  
## Age of Respondents 8522 8522 8522  
## Marital Status 8522 8522 8522  
## Monthly Income 7906 7906 7906  
## Vulnerability Index 8522 8522 8522  
## Monthly Income Vulnerability Index  
## Household Size 7906 8522  
## Age of Respondents 7906 8522  
## Marital Status 7906 8522  
## Monthly Income 7906 7906  
## Vulnerability Index 7906 8522  
## Probability values (Entries above the diagonal are adjusted for multiple tests.)   
## Household Size Age of Respondents Marital Status  
## Household Size 0 0 0.00  
## Age of Respondents 0 0 0.00  
## Marital Status 0 0 0.00  
## Monthly Income 0 0 0.00  
## Vulnerability Index 0 0 0.66  
## Monthly Income Vulnerability Index  
## Household Size 0 0.00  
## Age of Respondents 0 0.00  
## Marital Status 0 0.66  
## Monthly Income 0 0.00  
## Vulnerability Index 0 0.00  
##   
## Confidence intervals based upon normal theory. To get bootstrapped values, try cor.ci  
## raw.lower raw.r raw.upper raw.p lower.adj upper.adj  
## HshlS-AgofR -0.17 -0.15 -0.13 0.00 -0.18 -0.12  
## HshlS-MrtlS 0.02 0.04 0.06 0.00 0.01 0.06  
## HshlS-MnthI -0.11 -0.09 -0.06 0.00 -0.12 -0.06  
## HshlS-VlnrI -0.14 -0.12 -0.10 0.00 -0.14 -0.09  
## AgofR-MrtlS 0.11 0.13 0.15 0.00 0.10 0.16  
## AgofR-MnthI 0.02 0.04 0.06 0.00 0.01 0.06  
## AgofR-VlnrI -0.12 -0.10 -0.08 0.00 -0.13 -0.07  
## MrtlS-MnthI 0.01 0.04 0.06 0.00 0.01 0.06  
## MrtlS-VlnrI -0.02 0.00 0.03 0.66 -0.02 0.03  
## MnthI-VlnrI 0.14 0.16 0.19 0.00 0.13 0.20

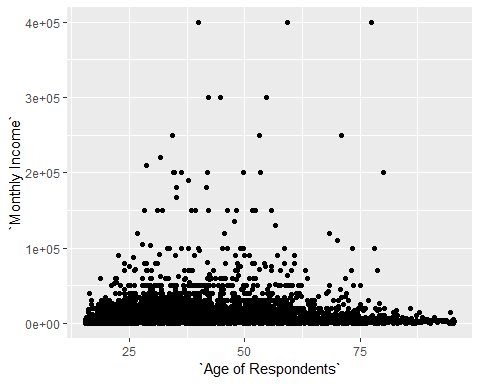
The table above shows the correlation between the variables.

## Graph showing monthly income across age groups

ggplot(working\_data, aes(x = `Age of Respondents`, y = `Monthly Income`)) + geom\_jitter()

## Don't know how to automatically pick scale for object of type haven\_labelled. Defaulting to continuous.

## Warning: Removed 616 rows containing missing values (geom\_point).



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Below 2,000 | 2,000-5,000 | 5,001-10,000 | 10,001-40,000 | 40,001-100,000 | Above 100,000 |
| Below 20 | 510 | 186 | 105 | 48 | 1 | 0 |
| 20-25 | 199 | 303 | 263 | 198 | 13 | 0 |
| 26-30 | 197 | 340 | 325 | 299 | 21 | 5 |
| 31-35 | 140 | 251 | 282 | 251 | 24 | 8 |
| 36-40 | 128 | 237 | 225 | 222 | 29 | 5 |
| Above 40 | 687 | 984 | 650 | 519 | 99 | 24 |

## Marital Status change

Otieno changed marital status to a dummy variable where 1 was Married and 0 otherwise. Using the new variable, establish whether monthly income is the same regardless of your marital status.

working\_data$`Marital Status` <- as\_factor(working\_data$`Marital Status`)  
new\_working\_data <- working\_data %>% filter(`Marital Status` %in% c("Single/Never Married","Divorced/separated,Widowed", "Married/Living with partner" ))%>% mutate(new\_status = if\_else(`Marital Status`== "Married/Living with partner", 1,0))

To establish whether monthly income changes with the marriage status we run a linear model.

model <- lm(`Monthly Income`~factor(new\_status), new\_working\_data)  
summary(model)

##   
## Call:  
## lm(formula = `Monthly Income` ~ factor(new\_status), data = new\_working\_data)  
##   
## Residuals:  
## <Labelled double>: B3H. Monthly Income (KSh)  
## Min 1Q Median 3Q Max   
## -11619 -7619 -5269 381 392281   
##   
## Labels:  
## value label  
## -1 Don't Know  
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7719.0 440.7 17.517 < 2e-16 \*\*\*  
## factor(new\_status)1 3900.0 527.0 7.401 1.52e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 19650 on 6608 degrees of freedom  
## (511 observations deleted due to missingness)  
## Multiple R-squared: 0.008221, Adjusted R-squared: 0.00807   
## F-statistic: 54.77 on 1 and 6608 DF, p-value: 1.522e-13

## Regression analysis

Joy postulates that household size, respondents age and vulnerability index are covariates of monthly income. This would best be answered using regression where we seek 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).

Our null hypotheses would be;

1. There is no significant relationship between monthly income and household size.
2. There is no significant relationship between monthly income and respondents age.
3. There is no significant relationship between monthly income and vulnerability index.

The linear model is as follows;-

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

##   
## Call:  
## lm(formula = `Monthly Income` ~ `Household Size` + `Age of Respondents` +   
## `Vulnerability Index`, data = working\_data)  
##   
## Residuals:  
## <Labelled double>: B3H. Monthly Income (KSh)  
## Min 1Q Median 3Q Max   
## -15536 -7531 -3699 1605 387154   
##   
## Labels:  
## value label  
## -1 Don't Know  
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2236.60 1159.30 -1.929 0.053734 .   
## `Household Size` -491.02 90.61 -5.419 6.16e-08 \*\*\*  
## `Age of Respondents` 47.81 12.37 3.863 0.000113 \*\*\*  
## `Vulnerability Index` 4717.40 327.14 14.420 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 18340 on 7902 degrees of freedom  
## (616 observations deleted due to missingness)  
## Multiple R-squared: 0.03353, Adjusted R-squared: 0.03316   
## F-statistic: 91.38 on 3 and 7902 DF, p-value: < 2.2e-16