HENOK FASIL TELILA
henokfasil@gmail.com
Rome, Italy

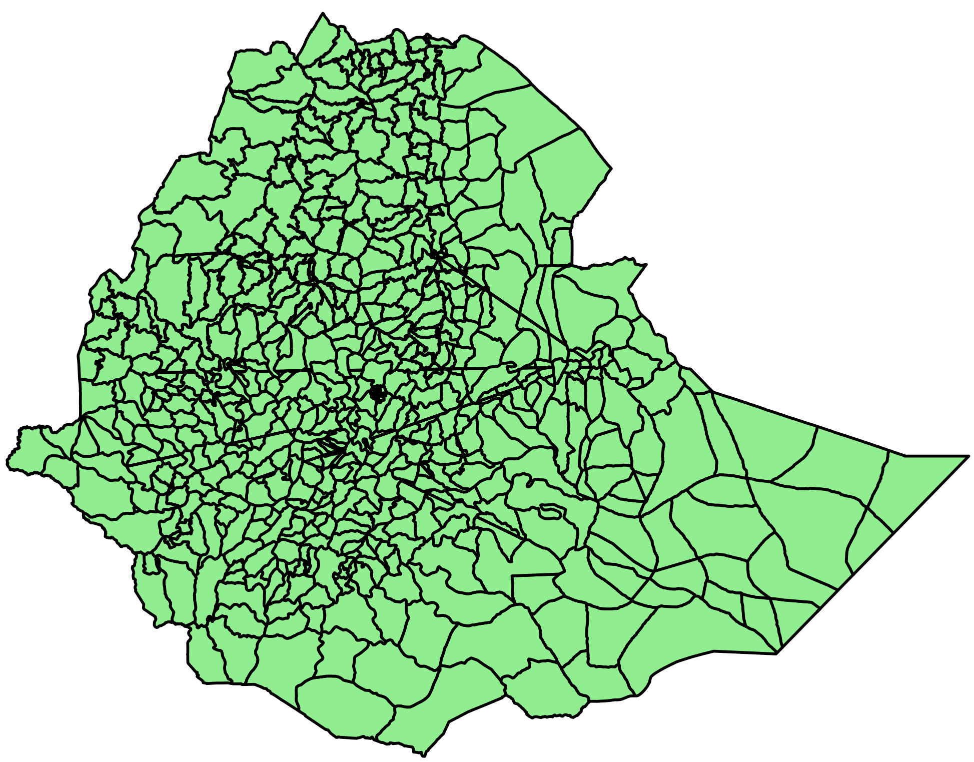
Analysis of Food security in Ethiopia

1) Glimpse of the dataset

1.1) The Administrative map of Ethiopia at wereda level

```
path1 = "C:/Users/Asus2/Desktop/FA_EVAL" ;setwd(path1)
hhshap <- shapefile("ETH_adm3.shp") ;hhdata <- read.csv("Ethiopia_2011.csv")
      hhdata<-hhdata%>%rename("latitude"="lat_mod_11","longitude"="lon_mod_11")
          hhshap_f <- tidy(hhshap, region = "NAME_3")

ggplot() + geom_polygon(data = hhshap_f, aes(x = long, y = lat, group = id),
          colour = "black", fill = "lightgreen") + theme_void() +
            geom_sf(aes(fill = )) + ggtitle("Adminstrative map of Ethiopia ")
```

Adminstrative map of Ethiopia



```
#ggsave("admin_map_Ethiopia.pdf") #helps to save to any file
```

1.2) Summary for Household self-reported that food was not enough during (last 12 months)

As we are dealing with food security issues, I summarize at least one major variable that shows those households self-reported that food was not enough during (last 12 months) during the mentioned study period. For those households self-reported that food was not enough during (last 12 months), the average age and size are 46.48 and 5.07, respectively. Meanwhile, for those households that that food was enough, the average age and size are 44.53 and 5.02, respectively.

```
food_not_enough <- hhdata %>% filter(food_not_enough12m == 1) #food not enough
summary(food_not_enough[,c(15,17,19,36)])
```

```
##     agehead          hhsize          educhead        mean_rain_lt
##  Min.   :15.00   Min.   : 1.000   Min.   : 0.000   Min.   : 3653
##  1st Qu.:35.00   1st Qu.: 3.333   1st Qu.: 0.000   1st Qu.:23133
##  Median :44.00   Median : 5.000   Median : 0.000   Median :28448
##  Mean   :46.48   Mean   : 5.076   Mean   : 1.449   Mean   :29465
##  3rd Qu.:59.00   3rd Qu.: 6.667   3rd Qu.: 2.000   3rd Qu.:35392
##  Max.   :97.00   Max.   :14.333   Max.   :15.000   Max.   :54826
```

1.3) Summary for Household that food was enough

```
free_enough <- hhdata %>% filter(food_not_enough12m == 0) #food enough
summary(free_enough[,c(15,17,19,36)])
```

```
##     agehead           hhsize          educhead        mean_rain_lt
##  Min.   : 15.00   Min.   : 1.000   Min.   : 0.000   Min.   : 3653
##  1st Qu.: 31.00   1st Qu.: 3.333   1st Qu.: 0.000   1st Qu.:21207
##  Median : 40.00   Median : 5.000   Median : 0.000   Median :28030
##  Mean   : 43.57   Mean   : 5.023   Mean   : 2.246   Mean   :29143
##  3rd Qu.: 54.00   3rd Qu.: 6.667   3rd Qu.: 4.000   3rd Qu.:37868
##  Max.   :100.00   Max.   :14.000   Max.   :15.000   Max.   :54826
```

Rstudio (Allaire 2012) is used for computation. Several packages have been implemented, including Rmarkdown (Baumer and Udwin 2015).

(A)

This solution shows the descriptive summary grouped by sex. Particularly, the following table (non-weighted) shows the female-headed household's descriptive summary. There are 893 female-headed households (FHH) and 2659 male-headed households (MHH) included in the survey. The average age for FHH and MHH are 47.66 and 43.32, respectively. However, the maximum age for FHH and MHH are 98 and 100, respectively.

```
setwd(path1)

fhead <- hhdata %>% filter(femhead == 1) #Female headed HH

describeBy(hhlabor+agehead+hhsize+hhszae+educhead+educave15_60+area_gps
+landown+TLU_total+dist_market+nmonth_food_insecure+covrain_lt
+mean_rain_lt ~ femhead, data = fhead) #descriptive for female headed
```

```
##
##  Descriptive statistics by group
## femhead: 1
##                        vars   n    mean       sd   median  trimmed      mad
## hhlabor                   1 893    1.66     1.22     1.00     1.55     1.48
```

```
## agehead                2 893    47.66    15.57    46.00    47.14    16.31
## hhsize                 3 893     3.64     1.92     3.33     3.51     1.98
## hhszae                 4 893     2.87     1.59     2.62     2.69     1.48
## educhead               5 893     1.13     2.91     0.00     0.28     0.00
## educave15_60           6 893     2.37     3.23     0.50     1.73     0.74
## area_gps               7 893     3.97     3.82     2.02     3.83     2.95
## landown                8 893     1.18     2.01     0.21     0.71     0.31
## TLU_total              9 893     1.29     2.12     0.30     0.82     0.44
## dist_market           10 893    65.00    47.28    56.50    58.63    41.66
## nmonth_food_insecure  11 893     1.19     2.09     0.00     0.75     0.00
## covrain_lt            12 893     0.14     0.05     0.13     0.14     0.05
## mean_rain_lt          13 893 28727.73 11067.69 27276.57 28454.41 10666.35
##                          min      max    range skew kurtosis     se
## hhlabor                 0.00     7.00     7.00 0.90     0.85   0.04
## agehead                15.00    98.00    83.00 0.33    -0.48   0.52
## hhsize                  1.00    14.00    13.00 0.85     1.34   0.06
## hhszae                  1.00     9.46     8.46 0.97     0.93   0.05
## educhead                0.00    15.00    15.00 2.94     8.18   0.10
## educave15_60            0.00    15.00    15.00 1.53     1.90   0.11
## area_gps                0.01     9.04     9.04 0.42    -1.63   0.13
## landown                 0.00    15.28    15.28 2.74     9.58   0.07
## TLU_total               0.00    15.35    15.35 2.89    10.92   0.07
## dist_market             2.10   283.30   281.20 1.46     2.85   1.58
## nmonth_food_insecure    0.00    12.00    12.00 2.36     7.20   0.07
## covrain_lt              0.07     0.41     0.34 1.53     3.65   0.00
## mean_rain_lt         3652.63 54826.14 51173.51 0.21    -0.54 370.37
```

A descriptive summary for male headed household.

```
#Male headed households descriptive summary
mhead <- hhdata %>% filter(femhead == 0) #Male headed HH

describeBy(hhlabor+agehead+hhsize+hhszae+educhead+educave15_60+area_gps
+landown+TLU_total+dist_market+nmonth_food_insecure+covrain_lt
+mean_rain_lt ~ femhead, data = mhead) #descriptive for female headed
```

```
##
##  Descriptive statistics by group
## femhead: 0
##                      vars    n     mean       sd   median  trimmed      mad
## hhlabor                 1 2659     2.54     1.24     2.00     2.43     0.00
## agehead                 2 2659    43.32    15.36    40.00    42.02    14.83
## hhsize                  3 2659     5.51     2.10     5.33     5.46     1.98
## hhszae                  4 2659     4.13     1.86     4.00     3.99     2.05
## educhead                5 2659     2.32     3.60     0.00     1.54     0.00
## educave15_60            6 2659     2.49     2.98     1.50     1.94     2.22
## area_gps                7 2659     3.66     3.21     2.56     3.44     2.95
## landown                 8 2659     1.97     2.39     1.22     1.54     1.81
## TLU_total               9 2659     2.34     3.34     1.54     1.77     2.28
## dist_market            10 2659    67.98    51.09    58.90    60.92    45.96
## nmonth_food_insecure   11 2659     0.77     1.56     0.00     0.41     0.00
## covrain_lt             12 2659     0.14     0.05     0.13     0.13     0.04
## mean_rain_lt           13 2659 29405.00 10973.76 28448.11 29242.51 10694.45
##                          min      max    range skew kurtosis     se
## hhlabor                 0.00    10.00    10.00 1.24     2.51   0.02
## agehead                15.00   100.00    85.00 0.72    -0.10   0.30
## hhsize                  1.00    14.33    13.33 0.26     0.05   0.04
## hhszae                  1.00    11.86    10.86 0.58    -0.04   0.04
```

```
## educhead                 0.00    15.00    15.00 1.64     2.01   0.07
## educave15_60             0.00    15.00    15.00 1.62     2.74   0.06
## area_gps                 0.01     9.04     9.04 0.66    -1.02   0.06
## landown                  0.00    17.77    17.77 1.87     4.76   0.05
## TLU_total                0.00    66.60    66.60 5.96    73.40   0.06
## dist_market              0.50   283.30   282.80 1.28     1.79   0.99
## nmonth_food_insecure     0.00    12.00    12.00 2.67     9.94   0.03
## covrain_lt               0.07     0.41     0.34 2.08     7.01   0.00
## mean_rain_lt          3652.63 54826.14 51173.51 0.14    -0.46 212.81
```

The weighted descriptive statistics are summarized based on the filter condition where the dummy is 1 for FHH and 0, otherwise. The average and the deviation from the mean statistics are included. FHH_mean, MHH_mean, FHH_sd, and MHH_sd stands for variables average for female-headed HH, the average for male-headed HH, the standard deviation (sd) for female-headed HH, and the sd for male-headed HH, respectively. Accordingly, similar to the result of the non-weighted descriptive summary, the average age is 48.52 and 43.59 for FHH and MHH, respectively. The average household size for female-headed (3.81) is smaller than the male-headed HH (5.58). The male-headed HH market (65.3 km) is farther than the female-headed HH (61.90km). The long-term mean precipitation for both is similar.

For the statistical significance test between the group means, I have considered the variable 'educave15_60' that explains the average HH education of members in the working age. Since the level of education might be related to the sex (who leads the household?) of the household leader, it could be reasonable. The result shows that there is a significant difference between the two groups (male-headed and female-headed household) means demonstrated by the significant p-values.

```
#Weighted descriptive statistics by sex
weightedmean <- function(x){ #weighted mean for Female HH
      w.mean(x, fhead$weight)}

          weightedsd <- function(x){ #Weighted sd for Female HH
                w.sd(x, fhead$weight)}

weightedmeanM <- function(x){ #weighted mean for Male HH
      w.mean(x, mhead$weight)}

          weightedsdM <- function(x){ #weighted sd for Male HH
                w.sd(x, mhead$weight)}

fmeans <- as.data.frame(sapply(fhead[,15:36], weightedmean))
  colnames(fmeans) <- c( "FHH_mean") #Female Headed Household mean
mmeans <- as.data.frame(sapply(mhead[,15:36], weightedmeanM))
  colnames(mmeans) <- c( "MHH_mean") #Male Headed Household mean
fstd   <- as.data.frame(sapply(fhead[,15:36], weightedsd))
    colnames(fstd) <- c( "FHH_sd") #Female Headed Household standar deviation
mstd   <- as.data.frame(sapply(mhead[,15:36], weightedsdM))
    colnames(mstd) <- c( "MHH_sd") #Male Headed Household standard deviation

weighted_obs <- as.data.frame(cbind(fmeans,mmeans,fstd,mstd))
round(weighted_obs,digits = 2)
```

Weighted descriptive statistics by sex

```
##                        FHH_mean MHH_mean  FHH_sd  MHH_sd
## agehead                   48.52    43.59   14.75   15.34
## hhlabor                    1.77     2.58    1.22    1.21
## hhsize                     3.81     5.58    1.84    2.01
```

```
## hhszae                   3.05     4.23    1.53    1.81
## educhead                 0.98     2.14    2.58    3.21
## educave15_60             2.38     2.37    3.02    2.64
## credit                   0.23     0.30    0.42    0.46
## extension_advice         0.23     0.31    0.42    0.46
## safewater                0.83     0.80    0.37    0.40
## agwealth_index           0.17     0.26    0.16    0.15
## area_gps                 3.33     3.62    3.42    3.06
## landown                  1.82     2.43    2.50    2.69
## TLU_total                1.79     2.39    2.34    2.74
## dist_market             61.90    65.30   41.48   45.60
## improved_maize           0.10     0.14    0.30    0.35
## fert_inorg_use           0.40     0.50    0.49    0.50
## pest_use                 0.06     0.08    0.24    0.27
## free_food                0.08     0.05    0.27    0.23
## food_not_enough12m       0.37     0.31    0.48    0.46
## nmonth_food_insecure     1.19     0.88    1.90    1.54
## covrain_lt               0.13     0.13    0.04    0.04
## mean_rain_lt         30576.31 31382.95 9715.63 9064.85
```

```
#Statistical significance test of the two group means
t.test(mhead$educave15_60,fhead$educave15_60)
```

Statistical significance test of the two group means

```
##
##  Welch Two Sample t-test
##
## data:  mhead$educave15_60 and fhead$educave15_60
## t = 0.98306, df = 1436.6, p-value = 0.3257
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1197987  0.3604979
## sample estimates:
## mean of x mean of y
##  2.485494  2.365144
```

(B)

Before merging, I have done what we call 'rastering' (by the famous package called 'Raster'). The CHIRPS precipitation data are organized as a tiff file and downloaded with a resolution of (0.05, 0.05) latitude and longitude. The following shows how a single operation is conducted. It is a sample file called 'chirps-v2.0.2011.04.tif' representing the seasonal rainfall for April 2011.

```
a1 = 'chirps-v2.0.2011.04.tif' #the sesonal rainfall for April 2011
raster(a1)
```

```
## class      : RasterLayer
## dimensions : 1600, 1500, 2400000  (nrow, ncol, ncell)
## resolution : 0.05, 0.05  (x, y)
## extent     : -20, 55, -40, 40  (xmin, xmax, ymin, ymax)
## crs        : +proj=longlat +datum=WGS84 +no_defs
## source     : C:/Users/Asus2/Desktop/FA_EVAL/chirps-v2.0.2011.04.tif
## names      : chirps.v2.0.2011.04
```

The author's multmerge function (see below:) gives me a data frame consisting of the coordinates and the seasonal precipitation values. Averaging provides a monthly seasonal mean at an African level.

```r
#This code aims to read the downloaded tiff data, raster and convert to dataframe
multmerge = function(mypath){
  #read the files
  filenames=list.files(path=mypath, full.names=TRUE)
  #raster the tif file that is downloaded from the website
  datalist = lapply(filenames, function(x){raster(x)})
  #convert in to values of the X,Y coordinates
  datalistx = lapply(datalist, function(x){rasterToPoints(x, spatial = TRUE)})
  #Convert in to dataframe
  finaldata = lapply(datalistx, function(x){data.frame(x, xy = TRUE)})
  #Merge the files
  Reduce(function(x,y) {merge(x,y)}, finaldata)
}

#mydata = multmerge("C:/Users/Asus2/Desktop/FA_EVAL/new")
#write.csv(mydata, file="africa_rainfall.csv")

mydata <- read.csv("africa_rainfall.csv"); mydata <- mydata[,-1]
#Seasonal mean
seasonal_mean = cbind(mydata[,1:2],as.data.frame(rowMeans(mydata[,5:13])))
colnames(seasonal_mean) <- c("latitude","longtiude","Seasonal_mean")
#write.csv(seasonal_mean, file = "africa_seasonalmean.csv")
#View(seasonal_mean)

seasonal_mean$Seasonal_mean[seasonal_mean$Seasonal_mean==-9999]<-NA
seasonal_meanx <- na.omit(seasonal_mean)
head(seasonal_meanx)
```

```
##       latitude longtiude Seasonal_mean
## 841 -0.0249997    10.025      116.1583
## 842 -0.0249997    10.075      116.6661
## 843 -0.0249997    10.125      114.9995
## 844 -0.0249997    10.175      113.3753
## 845 -0.0249997    10.225      110.2869
## 846 -0.0249997    10.275      109.0759
```

Merging the African seasonal mean data with the ERSS dataset

For this task, since the coordinates (latitude and longitude) in both (the existing data 'Ethiopia_2011.csv' and the downloaded & rastered data 'seasonal_mean' data frame) data are different even for a small decimal number, I compute it by considering K-Nearest Neighbours with a certain threshold (km). The data frame 'lastmerge' shows the final dataset containing the seasonal mean of precipitations at an African level.

```r
#Merging operation

# substitute seasonal mean of -9999 with NA
seasonal_mean$Seasonal_mean[seasonal_mean$Seasonal_mean==-9999]<-NA

# Household ID in eth is not perfectly formatted.
# You can open it in csv and see that IDs are repeated.
# So I'll create a new one assuming that each row is a unique household.
hhdata$id=1:nrow(hhdata)

# Sample afr (otherwise it's too huge)
# You can increase sample_size
sample_size=200000
set.seed(100)
```

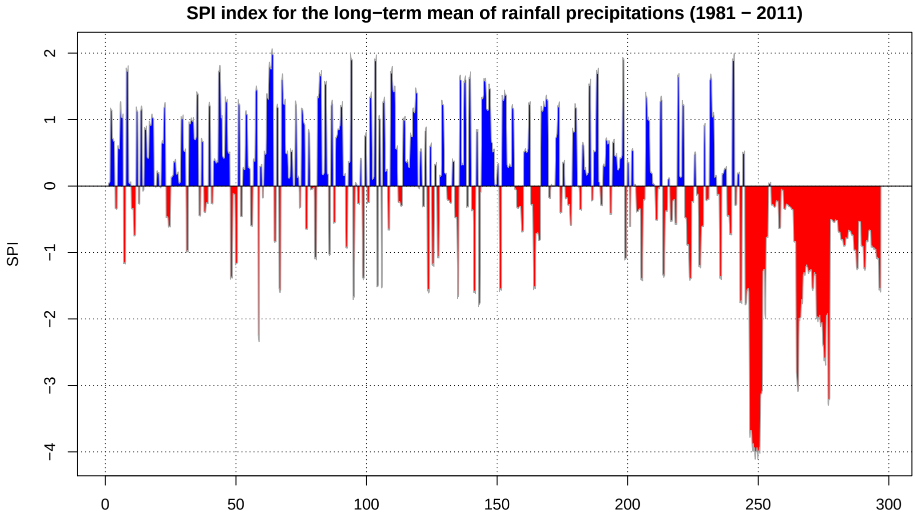
```
afr_sample=seasonal_mean[sample(nrow(seasonal_mean),size=sample_size),]

# Settings: where to look for the seasonal mean
# maximum distance in km for the point to be even
# considered a neighbor
dist_max=20
# up to how many nearest neighbors to consider
k=3

# merge data
last_merge <- hhdata %>%
  geo_left_join(afr_sample,
                distance_col = "distance",
                max_dist = dist_max)%>%
  group_by(id)%>% # do next steps for each household id
  arrange(distance)%>% #sort by distance so that smaller values are first
  mutate(closeness_rank=row_number())%>% # closeness rank is 1 for the closests,
                                    # 2 for the 2nd closests, etc.
  filter(closeness_rank<=k)%>% # select only k nearest neighbors
  mutate(Seasonal_mean=mean(Seasonal_mean,na.rm=TRUE))%>% #find their average
                                #seasonal mean
  filter(closeness_rank==1)%>% # keep 1 row per household
  select(-distance,-latitude.y,-longitude.y,-closeness_rank)
    # get rid of extra columns

  # write merged data to csv
  #write.csv(last_merge,"eth_merged.csv")
```

(C)

Index creation to identify the presence of drought: Following (McKee et al. 1993), I applied the Standardized Precipitation Index (SPI) used to define drought periods. Requiring only precipitation as input, the SPI covers a variety of timescales and allows comparison of drought severity both between periods in time and between different locations. McKee et al. (1993) defined the following four drought categories: Mild drought (SPI between 0 and -0.99, occurring 24 % of the time), moderate drought (SPI between -1.00 and -1.49, occurring 9.2 % of the time), severe drought (-1.50 to -1.99, occurring 4.4 % of the time), and extreme drought (SPI -2.00 or less, occurring 2.3 % of the time). A drought event may then be defined as a period during which the SPI is continuously negative and reaches a value of -1 or less at one or more time steps. Drought begins when the SPI first falls below zero and ends with the first positive value. Although there are alternatives, however, the SPI index exploits the gamma distribution (Gebrehiwot and Veen 2015).

```
last_merge <- read.csv("eth_merged.csv")[-1]
spi_mean = spi(last_merge$mean_rain_lt, scale = 3, distribution = 'Gamma')
plot(spi_mean,
main = "SPI index for the long-term mean of rainfall precipitations (1981 - 2011)")
```

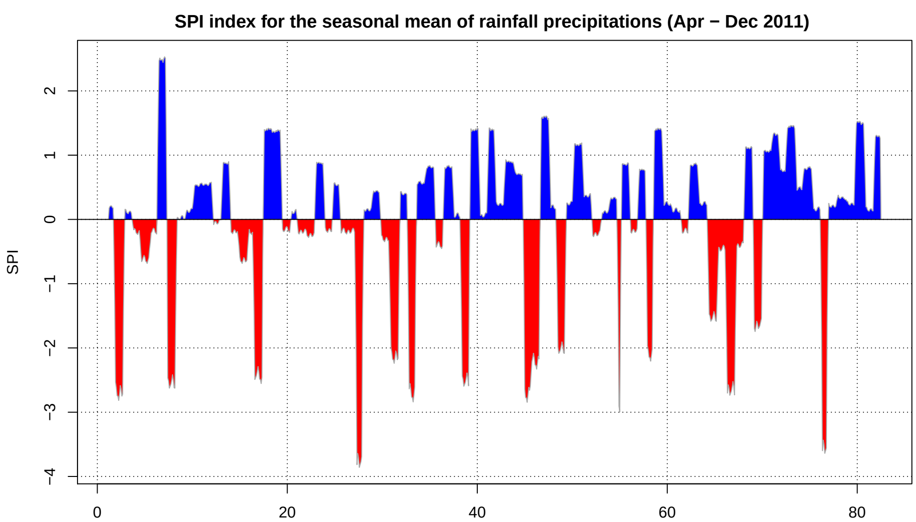
SPI for the long-term mean of rainfall precipitations (1981 - 2011)



The SPI for the long-term mean of rainfall precipitations (1981 - 2011) values are between 0 and 2, indicating a moist condition. However, an area to the right part of the figure shows a value between -4 and 0, indicating an extremely dry situation following the National Climatic Data Center (NCDC) definition.

```
#SPI for the seasonal mean
lastx = na.omit(last_merge)
spi_season = spi(lastx$Seasonal_mean, scale = 3, distribution = 'Gamma')
plot(spi_season, main = "SPI index for the seasonal mean of rainfall precipitations (Apr - Dec 2011) ")
```

SPI for the seasonal mean of rainfall precipitations (Apr - Dec 2011)



Furthermore, the NCDC defines the range of the SPI as follows. Positive SPI values indicate greater than median precipitation, and negative values indicate less than median precipitation. Relatively high negative deviations represent drought periods. Normally, the 'drought' part of the SPI range is arbitrary split into moderately dry (-1.0 > SPI > -1.49), severely dry (-1.5 > SPI > -1.99) and extremely dry conditions (SPI < -2.0).

Most of the SPI (NA values are excluded) values for the seasonal mean of rainfall precipitations (Apr-Dec 2011) are partially between 0 and 2.5 and between 0 and -4. The positive part of the figure (blue shaded area) shows the values in the moist condition. In contrast, the figure's red part shows the drought range of the seasonal mean of the rainfall precipitations.

(D)

Testing the empirical conditional association: As the regression output shows, there is a strong association between food aid and household insecurity shown by the statistically significant p-value. Household self-reported that food was not enough during (last 12 months), the dependent variables in the study, takes the value of 1 if food was not enough and 0 otherwise. With 3552 number of observations, the linear model fit has the following output as follows. HH head age, HH members education age, total land owned, distance to the market, HH uses inorganic fertilizer, HH received free food and the long_term mean rainfall significantly influenced household self-reported that food was not enough during (last 12 months).

As expected, HH labor age, household received credit, HH received extension service, HH access to safe water, HH average asset wealth index, total land cultivated area (gps acres) and total livestock in TLU negatively influenced the dependent variable.

```
y <- as.matrix(hhdata[,15:32])
#y <- as.matrix(last_merge[,c(15:34,35:37,39)])

food_insecurity <- hhdata$food_not_enough12m
model_1 <- lm(food_insecurity ~ y + hhdata$mean_rain_lt )

summ(model_1,robust = "HC1", pvals = TRUE)
```

```
## MODEL INFO:
## Observations: 3552
## Dependent Variable: food_insecurity
## Type: OLS linear regression
##
## MODEL FIT:
## F(19,3532) = 11.52, p = 0.00
## R² = 0.06
## Adj. R² = 0.05
##
## Standard errors: Robust, type = HC1
## -----------------------------------------------------------
##                                Est.   S.E.   t val.      p
## ------------------------- ------- ------ --------- ------
## (Intercept)                    0.18   0.04     4.01   0.00
## yagehead                       0.00   0.00     3.57   0.00
## yhhlabor                      -0.03   0.01    -3.33   0.00
## yhhsize                        0.01   0.01     1.24   0.21
## yhhszae                        0.02   0.01     1.66   0.10
## yeduchead                      0.00   0.00     0.07   0.94
## yeducave15_60                 -0.02   0.00    -3.86   0.00
## ycredit                        0.08   0.02     4.35   0.00
## yextension_advice             -0.04   0.02    -1.50   0.13
## ysafewater                    -0.04   0.02    -2.04   0.04
## yagwealth_index               -0.35   0.06    -6.32   0.00
## yarea_gps                     -0.01   0.00    -2.24   0.02
## ylandown                       0.00   0.00     0.30   0.77
## yTLU_total                    -0.01   0.00    -2.52   0.01
## ydist_market                   0.00   0.00     4.93   0.00
## yimproved_maize               -0.01   0.03    -0.23   0.82
## yfert_inorg_use                0.06   0.02     2.53   0.01
## ypest_use                     -0.01   0.03    -0.30   0.76
```

```
## yfree_food                    0.09   0.03     2.99   0.00
## hhdata$mean_rain_lt           0.00   0.00     2.96   0.00
## -------------------------------------------------------
```

```
View(last_merge)
```

(E)

Identification strategy: There might be a possible bias arising from the non-random assignment of the food aid program, i.e., the household which is already suffering from food insecurity is more likely to get the aid. Thus, when we estimate our average treatment effect, the result will be biased since the assumption of random assignment is not being met. The treatment effect can be estimated using a simple t-test of means for the household which received aid versus those which did not.

A better alternative would be to run a randomized control trial (RCT). We would cluster the households at the county (wereda) or district level, and then one district would receive the treatment and one would not. After running the RCT, we can estimate, which would give us an unbiased estimate of the treatment's true impact. While this does seem a little bit unethical and would leave out more "deserving" households, the statistical impact will be more accurate.

Trends might also play a part, as households that were food insecure in the past might be more likely to receive future aid so that an RCT would overcome that effect as well.

Henok Fasil Telila
henokfasil@gmail.com
Jan-31-2021
Rome, Italy

References

Allaire, J. 2012. "RStudio: Integrated Development Environment for r." *Boston, MA* 770: 394.

Baumer, Benjamin, and Dana Udwin. 2015. "R Markdown." *Wiley Interdisciplinary Reviews: Computational Statistics* 7 (3): 167–77.

Gebrehiwot, Tagel, and Anne van der Veen. 2015. "Estimating the Impact of a Food Security Program by Propensity-Score Matching." *Journal of Development and Agricultural Economics* 7 (1): 38–47.

McKee, Thomas B, Nolan J Doesken, John Kleist, and others. 1993. "The Relationship of Drought Frequency and Duration to Time Scales." In *Proceedings of the 8th Conference on Applied Climatology*, 17:179–83. 22. Boston.