THA4

**Research Question:**

My overall hypothesis for my final is that Out Migration in Nepal and in the Chitwan Valley in Nepal in particular will decrease the overall agricultural productivity/output in the region.

In order to test that hypothesis, I am investigating my smaller research question. **Does outmigration in Nepal affect the production and cultivation of rice in the region?** Rice is one of the major crops produced in the country and the region. This research question will help us see whether an important regional crop like rice was affected by outmigration in this valley. .

**Data Source**

The Data source for this project comes from the Chitwan Valley [Nepal] Family Study: Labor Outmigration, Agricultural Productivity and Food Security Household Agriculture and Migration Survey.

You can access the data on the google drive [here](https://https//drive.google.com/drive/u/1/folders/14zC51PKe4KMkbua1cRBjRBtB0_ThbpEw)

I have saved a copy of the .dta file on my google drive for access.

**Variables**

**Independent variable:** Migration from each household in Chitwan valley from (2006-2016)

**Dependent variable:** 1) Amount of Rice Planted in Chitwan Valley from (2006-2016) (B11)

2)Amount of land area used for rice cultivation in Chitwan Valley from (2006-2016) B11A

3)Production of Rice in KG (B11B)

3) Amount of Rice Sold in KG in Chitwan Valley from (2006-2016)B11C

**Workflow- Do File- Stata**

* There were several variables in the data set that was coded as ‘household does not exist’ or ‘household merged’ those were coded as 999 or 998. Those values were not necessary for the purpose of this research question, so those values were dropped and coded as .

**Recoded New Variables**

foreach x of varlist A\* B\* MIG REM {

recode `x' (999=.)

recode `x' (998=.)

^The code above removes households that did not exist, which is code 999

^ This code 998 is household merge which will drop missing values since the household combined and one household does not exist anymore

* Same variable also had a lot responses that said planting by other and planted on rent and those were coded as 2 or 3, whereas planting Yes or No were coded as 0 or 1, since planting by anybody is measure as production, we recoded that as 1.

**Recoded New Variables**

recode B11\_ (2=1)

recode B11\_ (3=1)

^ The code above codes creates a yes answer =1 and a no answer = 0. B11 stand for rice production in data set. Yes for rice planting = 1

^B11 also had households who rented land out for planting. This was coded as = 3. This counts as rice planting in the household even though the land was rented out, so this was also coded as a yes and therefore = 1.

* We also wanted to total production of rice in proportion for the land that was used for the production. In order to do so we need to first change the measurement of the land from katha (Nepali measurement) to acre. 1katha= 11.97295 acre.

**Replaced Old Variable with New Measurements**

replace B11A\_ = B11A\_/11.97295

replace B13A\_= B13A\_/11.97295

B11A=Rice planting land area (in Katha)

B13A=Wheat planting land area (Katha)

^The code above changes the land used and the land that the crop was produced on from Nepali Katha to acres.

* I then generated total rice production as also looking in a proportional manner. How much production was taking place compared to the land it was being produced on. So we generated a new variable and coded rice production.

**Generated New Variables**

generate rice\_prod= B11B\_/B11A\_

^The code above gives the total rice production by calculating B13B/B13A (production of wheat/land area used for wheat production)

^The code above gives the total rice production by calculating B11B/B11A (production of wheat/land area used for wheat production)

**Testing**

Please see the log book attached in Git Hub for testing and Results

