Analysing household agricultural data for eleven african countries

An analysis has been done on the agricultural dataset collected by researchers on more the 9500 households in growing seasons 2002/2003 or 2003/2004 in eleven African countries: Burkina Faso, Cameroon, Ghana, Niger and Senegal in western Africa; Egypt in northern Africa; Ethiopia and Kenya in eastern Africa; South Africa, Zambia and Zimbabwe in southern Africa.

The dataset includes variables with the following types; Categorical type for example Age of occupant, and Continuous type for example Education (in number of years)

In study design for this survey, the population was sampled using a multi-level random selection. Countries were select first. Within each selected country, districts were selected to capture representative farms. The third level of the selection was villages within districts and surveys were conducted at in-village farms selected randomly from a list of farmers had prepared with the assistance of respective district level agricultural authorities. We can see a form of selection bias from the last step if some farms were forgotten by the district level agricultural authorities and are not listed. These farms can have common characteristics, for example they maybe the least socially active farms or the farthest from the village center, etc.

A few key research questions that could be studied based on the dataset provided include but are not limited to;

- How can farming system characteristics help inform about the importance of each system for a country’s agricultural production and its ability to cope with short-term and long-term climate changes or extreme weather events?
- What impact do noticed long term shifts in the mean rainfall impact have on the farm productivity and the motivation of farmers?

To find answers to one of the research questions, a plan for a statistical analysis of the data was developed.

Responding to question: What impact do noticed long term shifts in the mean rainfall impact have on the farm productivity and the motivation of farmers?

- Null hypothesis (Ho): there is no significant difference in the farm productivity in result of a noticed long term shifts in the mean rainfall.
- Alternative hypothesis (H1): there is a significant difference in the farm productivity in result of a noticed long term shifts in the mean rainfall (one-tailed test).

The starting point was a t-test and followed by a linear regression to include control variables.

The t-test was a good start since there was a difference in productivity mean between farmers who noticed a long term shifts in the mean rainfall and those who didn't.

The linear regression was necessary to include control variables that may also explain the dependent variable (farm productivity) but also could be correlated with the main independent variable (noticed long term shifts in the mean rainfall). This was particularly useful for this research question since it was not a randomized control trial.

Given the test significance, the null hypothesis Ho was rejected in favor of the alternative H1 with a p-value less that 5%.

The conclusions that could be made about the population and the research question with regard to significance were;

- If a significant effect resulted, the Null hypothesis would be rejected. So in the context of the countries included in the survey, there is significant difference in the farm productivity in result of a noticed long term shifts in the mean rainfall.
- If the results were not significant, the Null hypothesis would be accepted. So in the context of the countries included in the survey, it was clear that there was a significant difference in the farm productivity as a result of a noticed long term shift in the mean rainfall.

The unique types of variables in the dataset could be represented using different types of visualizations;

- A bar graph: this could be used to visualise the counts of discrete data e.g. mean age of households
- Pie chart: using this can visualise what distribution of amounts in Amount consumed, sold, lost, value of crops sold, amount of seeds for food crops
- Histogram: could be used to plot frequencies of number of years of being a farmer or the total household net income.

Comparing the survey data to national average values from the FAO and the World Bank data on fertilizer use, pesticide use and the irrigated area is important because;

- correlating these values to see whether they are the same is insightful
- use of technology: fertilizer, pesticide, irrigation can mitigate the impact of climate change and have to be included as control variables.