

Agricultural Survey of more than 9500 households

Statistics Summative Assessment - Jackline Atsango

During the growing seasons between 2002 - 2004, surveys were conducted across 9,500 households in eleven African countries. The information collected contains various farming characteristics that can assist to infer the influence of climate change and extreme weather events on agricultural production in the different countries. The data contains many nominal variables with the country being an initial distinction but there are other recorded attributes such as gender, marital status and religion. In terms of quantitative information, interval data such as distance or time it would take to the nearest market in order to sell produce is captured, as well as continuous variables such as the average wage.

In terms of the survey conduction, a random sample of households within a range of districts in a country that met certain criteria such as climatic region, farm size and type among other conditions, were selected for participation. The method was appropriate for the application because the research question was targeted towards very specific criteria and the variables that might have an influence on it. Although the selection the country and district may be biased, they were selected in a manner that would represent different climatic regions and diverse agricultural types.

Further smaller analyses can be investigated from the dataset given the array of variables collected. A possible analysis could be the influence that education has on the net revenue of the household stemming from farming activities. The null hypothesis in this case would be that the years of education does not have an influence on the proportion of revenue that is contributed from farming activities, whereas the alternative would be that less revenue is contributed from farming activities as the number of years of education within a household increases. This can be tested by looking at the relationship between the two attributes using a correlation coefficient because the variables don't necessarily depend on one another. The correlation ranging between -1 and 1 would indicate the relationship between the two attributes while a P value < 0.05 would determine a statistically significant observation.

Otherwise, some possible visualisations of this dataset could be a histogram showing the net income of the various households in the different countries, a pie chart looking at the overall gender distribution of the dataset and a horizontal bar chart to show the Primary occupation of the head of the household.

Further data on fertiliser and pesticide use, as well as the irrigation of an area would be significant information to use in the research. These are attributes that may have an influence on the yield

produced, the time spent on the farm and the overall revenue of the pertaining household. It can also be used to infer further distinctions in the different agro-climatic regions for example if one region requires more pesticides than another or if there is any deterioration in the soil quality which would therefore require more fertiliser use.