AIIP 2019 Statistics Summative Assessment

An agricultural survey for more than 9,500 African households

Tauriq Ally

17.02.2019

Statistics summative assessment 2019

1. List at least two variable types and provide unique examples of the corresponding variable in the dataset

1. Nominal - an example would be “gender” (1.2.1) from the supplied dataset.

2. Continuous data - an example would be “age” (1.2.2) from the supplied dataset

2. Identify and evaluate the study design for this agricultural survey. How was the population sampled? Is the method used appropriately for the application? Are there any sources of bias?

A combination of simple random sampling, multistage sampling and stratified sampling was used to select sample households. First the continent was divided into four subregions, North, South, East and West Africa (Stratified sampling). The countries within the four subregions were then selected based on their interest in the survey, the countries were then subdivided into regions and then districts based on agro-climatic conditions (multistage sampling). The final selection of farms to be sampled was done using simple random sampling. The selection of the participating countries was based on their eagerness and interest in the survey. This is a form of self selection bias, countries that were excluded due to lack of interest/eagerness could of had substantial agri-economies which would influence the survey outcome.

3. Develop 2 key research questions that can be studied based on the dataset provided.

1. What percentage of households surveyed are relying only on natural water sources for irrigation? (4.13.2 - 3 - groundwater, 4 - rain-fed)
2. Do farm managers have sufficient years of +experience and knowledge to adapt to the effects of climate change?

4. Develop a plan for a statistical analysis of the data in response to one of your 2 key questions. What are your null and alternative hypotheses? (Clearly state the two), Which tests would you use and why?, Why are they appropriate?, How would you determine significance?

- Definition of natural water sources - water sources that occur naturally without

human intervention .ie counts of instances (4.13.2 - 3 - groundwater, 4 - rain-fed)
- These counts will be use to compare how many farmers make use of only natural water sources vs farmers who rely on man made water sources.
- Null hypotheses : Only a very small percentage (<10%) of farmers rely solely on natural water sources for irrigation.
- Alternate hypotheses: A large percentage (>30%) of farmers rely solely on natural water sources for irrigation.
- The repeated measures ANOVA method will be used for statistical analysis because we are comparing differences within the same dataset subject.

5. How would you interpret your results if: 1) a significant effect resulted or 2) the results were not significant? What conclusion can you make about the population and your research question?

1. A significant result would reveal that a large percentage of farmers would be negatively impacted by the effects of climate change wrt water security.
2. An insignificant result would suggest that only a small percentage would be negatively impacted by the effects of climate change wrt water security.
3. This information can be used to prioritize water security education programs for farmers using only natural water sources. For governments, this information can assist with infrastructure planning for man made water security measures such as dams,reservoirs and water delivery systems.

6. Discuss at least three types of visualizations that you may use for the unique/different types of variables identified in the dataset provided.

An important indicator of economic viability is labour costs, a boxplot can quickly establish average labour rates ie (wagehrm and wagehrf) in a given district, country or region, the geographical information can be easily compared using bar plots.

For determining correlations between variables, a scatter plot can be useful. Given the dataset, we can attempt to check correlation between annual yield of a specific crop (pc1 and nyieldc1) vs Access and Extension services (extc). This correlation test can expose whether or not crop assistance services make a meaningful impact.

The histogram method of data visualization can be successfully applied to “System of

farming", (fsystem1 and fsystem2). The farming system categories can be binned, by doing this the audience will easily be able to see which farming systems are most popular among farmers. This information can be useful when grouping together regions and fertilizer/hectare irrigated.

7. During the research, data on fertilizer use, pesticide use and the irrigated area was compared to national average values from the FAO and the World Bank. Is this important for research? Discuss the reasons as to why or why not.

Yes this is an important indicator. It can be used to represent growth in the sector. A steady increase in the usage of mineral fertilizers indicates the farmer's willingness to increase the the input cost of his/her operation due to demand for his/her crop. The use of fertilizer statistics also represents a portion of the farmer's input cost per irrigated /arable hectare, this information is can be used for funding purposes. The suppliers of fertilizer and pesticides can also use this information to predict production quantities for the new year as well as optimizing logistical processes.