

STATISTICAL REPORT: AN AGRICULTURAL SURVEY FOR MORE THAN 9500 AFRICAN HOUSEHOLDS

More than 9500 agricultural households were interviewed in Africa to determine the effect of climate change on agriculture on the continent.

Multi-stage stratified random sampling was used to identify households to survey in this study [1]. The use of this sampling method is appropriate to stratify the respondents into agro-climatic zones and farming systems. This being said, Countries that participated in the survey were those which formally expressed interest [1], this introduces a self-selection bias. Interviews were held face to face to minimise a non-response bias, however the respondents responses might have been affected by the interviewer's presence [1].

Apart from the primary intention of this survey, to determine the effects of climatic change on farming activities in Africa, this dataset can be used to answer a number of other questions such as does household labour vs hired labour have an effect on the farming yields? The train of thought is that household labour would provide greater yields due to a more invested interest in the success of the farming production. The dataset can also be used to study this relationship.

Further, the question of the effect of the level of tax charges on farming yields can be investigated. Do tax breaks incentivise farming households to invest more money into their farming activities, thus securing the regions food security?

In order to quantify whether an outcome of the survey is not unusual from the population, hypothesis testing should be performed.

The the question of the effect of tax levels on farming yields is considered, for regions with lower respective tax levels:

- The null hypothesis would be: $H_0 = \text{Farming yields are unchanged.}$

- The alternative hypothesis would be: H_1 = Farming yields are higher.

The correlation between tax levels and farming yields should be tested based on the Pearson's correlation coefficient as both variables, tax level and farming yield can be considered continuous.

The significance level is to be considered the probability of 0.05, beyond which the null hypothesis can be rejected. This is chosen as this is the most acceptable value, without considering the level of risk that can be tolerated.

Regardless of whether the results are found to be statistically significant or not significant, the practical significance must be evaluated and coupled to the statistical significance to be able to draw meaningful results. This is to be performed in collaboration with subject matter experts to ultimately determine the practical significance of the results.

References:

[1] Waha, K. et al., An agricultural survey for more than 9,500 African households, Sci. Data 3:160020 doi:10.1038/sdata.2016.20, (2016)