1. **VARIABLES**

Farmtype: This variable is an example of ordinal data showing the type of farm entity of the households ranked primarily in small, medium and large sizes, based on different metrics. This data is ordinal as the sizes are defined specifically for each country and depend on the self-assessment of the households.

Farmingexperience: This variable is a representation of nominal data which shows the frequency of the named category, in this case the number of years of farming experience.

2. **STUDY DESIGN**

This study is specifically designed to target the key agro-climatic zones and farming systems in Africa. A combination of three sampling methods (i.e. Multi-stage, stratified and random sampling) is used to cover this population across 11 African countries.

The survey design is carefully considered as described in the Survey Manual, and the questionnaire results and dataset have been verified by many researchers as a useful tool. Their research include important topics such as the likely impact of climate change on agriculture in Africa.

The data is subject to some non-sampling errors during data collection such as non-response and measurement errors which could lead to bias. Although control methods are put in place to reduce convenience bias (e.g. household head availability), it could not be completely avoided. An example is the cost restriction affecting interviewers from conducting multiple visit surveys.

3. **RESEARCH QUESTION**

- Which country in West Africa has the highest net income from farming activity in the last 12 months?
- Do households with more years of farming experience earn more net income from farming activity?

4. **STATISTICAL ANALYSIS PLAN**

Question: Is there a relationship between number of years of farming experience ('farmexperience') and net income of the household from farming activities (incfarm)?

(i) The **null hypotheses** states that the variables 'farmexperience' and 'incfarm' are independent. The **alternative hypotheses** states that 'farmexperience' and 'incfarm' are not independent of each other.

(ii) In this example, the significance level chosen can be 0.05 and the test used to analyze the sample will be the Chi-squared test for independence. This test is relevant as the variables are categorical, independent, data will be randomly sampled and when tabulated contains a cell count of at least 5.

(iii) Significance is determined using a large sample size to calculate the degrees of freedom, expected frequencies, test statistics and importantly p-value. The statistical significance is gotten by comparing the derived p-value with the chosen significance level. Practical significance is more dependent on experts in the field.

5. **RESULT INTEPRETATION**

After choosing a large sample size and calculating the p-value of the sample, we compare with the significance level of 0.05 chosen for this test. One of two results can be achieved:

- A significant effect occurs if p-value is < 0.05 and the null hypotheses is rejected.
- A non-significant result occurs if p-value > 0.05 and the null hypotheses is accepted.

If a significant effect occurs we can consider the alternative hypotheses of there being a relationship between the variables being compared.

6. **VISUALIZATION METHODS**

- **Scatterplots** for the strength of a relationship such as positive or negative correlation between net income and farming experience for instance.
- **Histograms** to show the frequency distribution of the sampled data and could possibly reveal peaks, extremes and possible gaps in data for instance the average age of a sample population can be visualized.
- **Contingency Tables** can be used to show the frequency of data such as the number of households per country in the conducted survey.

7. **DATA COMPARISON**

The comparison to a general population data from FAO and World Bank is important to better understand the data, considering the limitations and biases in the survey. For example pesticide consumption between the household survey and FAO data shows disparities which could possibly point to difference in methods of data collection. This helps to provide context and accuracy to our research.