
Statistics Summative Assessment

1. Types of variables

Qualitative

- a. Ordinal data = farm type
- b. Nominal data = crop type

Quantitative

- c. Interval data = number of cattle born
- d. Continuous data = distance to market

2. The study was done using experimental design whereby sampling took place using multi-stage stratified random sampling. This method of sampling is appropriate for this kind of study because at first respondents were grouped based on shared characteristics like the four sub-regions of Africa - North, East, South and West. Secondly respondents were then further broken down into districts across different agro-climatic conditions (stratified sampling) and finally respondents were randomly selected within a district and grouped based on small, medium or large farm types.

Clustering samples around villages could introduce some form of convenience bias. Other forms of bias could have been introduced by geo-political issues in Zimbabwe as well as the land reform program in South Africa.

3. Research Questions

- a. What economic impact does climate change have on agriculture in Africa?
- b. Does the relative amount of fertilizer usage differ for the types of farm entity's – small, medium and large and what influence does this have on yields?

4. Analysis would be done by measuring central tendency using the median value for the test conducted. To verify that there are no unusual outcomes further testing will be done by using hypothesis testing for question b.

H_0 - There is no difference in fertilizer usage between the three different farm types

H_a - There is a difference in fertilizer usage between the three different farm types

To perform hypothesis testing we will be using one-way anova (analysis of variance) testing which is suitable for testing the difference between 3 or more samples (farm types – categorical groups) with one variable (fertilizer usage – continuous data)

5. To determine variability between groups we will determine the F – statistic which is a measure of variability between the groups tested relatively to variance inside each group. A large F-value indicates more variance within groups. To quantify the significance of this value, the p-value is considered.

If there is a significant difference where the probability is less than 5%, or $p < 0.05$ we will reject the null hypothesis and consider the alternative hypothesis.

6. Visualizations to be used
 - a. Box plots to compare the three groups of farming scales and the fertilizer usage
 - b. Scatter plot to visualize the relation between the different data sets and show outliers
 - c. Pie-chart could be used to visualize the use of fertilizer per category
7. A histogram will be a great way to visually compare costs for the categories as per Q4.20 where each of the 5 categories will be represented by one bar versus the total farming expenditure.
8. - Average household size = Sum of all household sizes / number of households = 7.28
 - Mean of total in-kind payments =

$$\begin{aligned}
 & (\text{inkindhmr}/\text{no payments} + \text{inkindhrrf}/\text{no payments} + \text{inhindhrc}/\text{no of payments})/3 \\
 & = 2626 + 1902 + 1112 / 3 \\
 & = 1880
 \end{aligned}$$

- Median sale price of breeding bulls = 0
 - o Median value is the middle value of the dataset which for the below series would be

$$\begin{aligned}
 & 5+5/2 = 5 \\
 & 1,1,1,2,2,3,3,4,4,5,5,5,6,6,7,7,7,8,8,9
 \end{aligned}$$

- Main mode for transport is the most frequently occurring value which is 4. Truck or other motorized vehicle