

Statistics Summative Assessment

1. Two variable types are:
 - Explanatory variable, example is household size
 - Response variable, example is household roster
2. The population was sampled using multi-stage stratified random sampling. The method was appropriate as the population was first sorted into the different agro-economic zones. This ensures that all agro-economic zones are represented and data is not biased to any particular zone. Sources of bias include clustering the sampled households in a chosen district (convenience bias); this may not be a true representation of the population within that district.
3. Two key research questions can be:
 - Percentage of income from non-farm related activities versus the subsidies received per family
 - Number of years being a farmer and the average yield per unit area
4. The null hypothesis for key research question 2 is that there is no significant relationship between the number of years being a farmer and the average yield per unit area. The alternative hypothesis states that there is a significant relationship between number of years being a farmer and the average yield per unit area. I would use the Chi square test for independence. This is appropriate as it will show whether or not the 2 variables are independent or are dependent.
5. Significance is determined by determining the p-value. If the p-value is higher than alpha then we fail to reject the null hypothesis but if p value is less than alpha then the null hypothesis may be rejected. If the probability is greater than 0.05 then we fail to reject the null hypothesis that there is no significant relationship between the number of years being a farmer and the average yield per unit area.
6. 3 unique visualizations that may be used are:
 - Histogram – these are used to show the frequency of binned categories. One may use a histogram to show the number of households per administrative unit (AU) for each of the 11 countries.
 - Contingency table – these are used to tabulate the frequency of several variables. One may use a contingency table to show the total income of 3 crops (maize, sorghum, millet) in 3 different countries (Kenya, Zimbabwe, South Africa)
 - Pie Chart – these are used to show a percentage of a specific characteristic out of a whole. One may use a pie chart to show that of the number of animals owned, how many were born, lost, stolen, killed or purchased
7. Data on fertilizer use, pesticide use and irrigation area are important as these have an effect on the yield produced. It is important to compare them to other data to see if there are any outliers or if there are any changes in trends observed. This may point to errors in the raw data obtained or may show a trend that is worth further researching into.