STATISTICS SUMMATIVE ASSESSMENT

Q1

I.   Dependent variable, **wagehrm**, cell [AGG,5] = 750(for household code 10010604) is dependent on the number of days worked across all activities.

II. Independent variable, **landhrm1,** cell [RA,5] = 0(for household code 10010604) is the number of days worked on land preparation, during season 1.
The wage earned is dependent on the number of days worked across all activities on the farm.

Q2

I.   Observation and parallel group design.

II. The observation design entailed collecting historical data from the respondents with the aim of trying to determine if a country's production was correlated to its farming system and climatic variations.
The parallel group design is an experimental study design aimed at finding an impact of climatic variations and adaptation to these variations by African farmers, a causal relationship between climatic changes and a country's production.

III. Multistage stratified sampling method. 1st stage involved selecting 11 countries in Africa who had indicated an interest in the experiment to represent the 4 sub-regions of the continent – North, South, East and West Africa. 2nd stage involved selecting districts from each country based on a pre-defined classification. 3rd stage involved random selection of households within villages in the selected districts which were representative of the of the agro-ecological zones in that country.

IV. Yes, it was appropriately used.

V.   Yes, there are sources of bias.

Q3

I.   Does the use of heavy machinery on farmlands impact average crop yield?

II. Do adaptation strategies to climatic variations impact on the net income of a household on both farm and non-farm activities?

Q4

I.   State the hypotheses, determine the significance level, use chi-square test for independence to determine significance.

II. Null: The use of heavy machinery does not cause a higher crop yield.
Alternative: The use of heavy machinery causes a higher crop yield.

III. Chi-square test for independence and simple linear regression. The former will help me determine if average crop yield is independent on the use of heavy machinery in a farmland. The latter will help me determine if there is a statistically significant relationship between the use of heavy machines on the farm and average crop yield.

IV. They are used to explore the relationship between two variables to determine if there is a statistical relationship between them.
V. Using the p-value.

Q5

I. This means there is a relationship between the use of heavy machinery on a farm and average crop yield and the result was not due to chance.
II. This means there is no relationship between the use of heavy machinery on a farm and average crop yield.
III. Farmlands using heavy machinery experience a higher average crop yield.

Q6

I. Contingency table to show a summary of the distribution of the frequency of observations for owned, born, lost and purchased livestock, poultry and other farm animals (lsv1-8, Num,born,lost)
II. Bar plot to show average yield of 5 principal crops in a year (pc1-5 & nyield1-5). this is used to compare each data set in the given period.
III. Pie chart to represent taxes paid over a year (taxincome, taxprop, taxsales, taxother). This visualization will show proportion of the taxes in percentage.

Q7

Yes, it is important for research as these variables are input costs to the estimation of net farm revenue. Anomalies in the quantities reported for these variables will distort the calculation of the annual cost expended and subsequently, distort the result of the research in calculating net farm revenue which is to be used to explain the variation in net farm revenue across several districts, agro-climatic regions and countries in Africa.