

## Statistics: Summative Assessment

Name: Itumeleng Fanampe

### Question 1

- gender1
- age1

Gender (Gender1) could be a male variable. A typical example would be identifying the likelihood of Gender1 being a non-farm worker

Age1 can also be a variable. A typical example would for what age are females likely to be warm workers.

### Question 2

For each household in the survey, the data was collected randomly. If the data is not collected randomly this will induce a bias in the data collected in that there could many questions why a data collector did not provide surveys to certain households, e.g. due to religious beliefs of the data collector or surveyor and this is a selection bias. Depending on the interviewer and how they understood the interview should be taken, there could be an element of observer bias. The instructions on how interviews should be conducted was attempt by Yale eliminate any bias during the interviews. The population is the entire agricultural households however instructions were provided to sample depending on the number of districts available for a certain country, i.e. the samples are taken evenly across the districts which is the appropriate method to use for sampling for this particular survey.

### Question 3

1. Crop productions – Season s2 provides more yield for planting crop type c1 in Africa
2. Livestock and studies – Temperature rise due to climate change has an adverse on crops

### Question 4

- Season s2 is does not provide more yield for planting crop type c1 in Africa
- Temperature rise due to climate change has no adverse crops

Alternative Hypothesis - Season s2 provides more yield for planting crop type c1 in Africa  
- Temperature rise due to climate change has an adverse on crops

## Types of hypothesis tests

For study 1. : Use linear regression test – this study finds a relationship between an independent variable (Season) and a dependent variable (crop type)

For study 2. : Use the paired t test – There is a relation between temperature rise and the crops yield and a difference of before (before increase in temperature) and after (increase in temperature due to climate change)

For both studies above, in order to determine the significance (p-value) needs to be determined.

#### **Question 5**

For study 1. : Check that the p-value is lower than a predefined significance level. A low value will indicate that a season has significant effect on the on the yield

For study 2. : Check the p-value is low and that will suggest the null hypothesis should be rejected and vice versa for a high p-value.

#### **Question 6**

Discuss at least three types of visualizations that you may use for the unique/different types of variables identified in the dataset provided.

- Scatter plot: This can be used to check the relationship between the temperature rise and yield
- A bar graph: This can be used to understand the effects that males and females have on the crop yields if they work certain or similar farm activities
- Cartographic visualization: display yield productivity zone predictions on a map

#### **Question 7**

This can be important as it can answer a number of question and whether there is wastage and more demand on fertilizer due to various factors such a soil composition or rainfall frequency and quantities. Areas of which seem or are outliers can be identified and corrective measures or improvement can be made. Areas of focus will areas which are more than 2 standard deviations from the mean assuming the data follows a normal distribution.