**Public Health: National Nutritional Health Project**

Author: Alatise Monsurah Bisola (@AlatiseMB)

Github Repo: <https://github.com/AlatiseMB/Hackbio-Module-9-Project.git>

**TASK**

I installed the required libraries – ggplot2 and tidyverse, Inspected the data (total observations, rows, types of data and did the required analysis for the task)

1. The dataset contains 5000 observations (rows) of 32 variables. It contains 33991 NA (null values). Deleting all NA reduces the dataset to 17 observations (rows) of 32 variables.So, I just replaced the NA with 0
2. I created a new variable for weight in pounds (by multiplying by 2.2) and Plotted histograms forthe 4 different variables in a 2 \* 2 grid.
3. The mean pulse rate for all participants in the data is 63.06
4. I checked the range using min and max of the diastolic blood pressure variables and got 0 & 116
5. The variance on the income variable is 1264147754 and standard deviation is 35554.86
6. I created 3 different data plots visualizing the relationships across gender, diabetes and smoking status
7. The results of the t-test between the variables using a significant p-value <= 0.05

* Age and Gender

NULL HYPOTHESIS, H0 = The mean age of individuals is the same across the gender groups

ALTERNATIVE HYPOTHESIS, H1 = The mean age of individuals is different across genders

After independent t-test, p-value = 0.08022, therefore we fail to reject the null hypothesis

* BMI and Diabetes

H0 = The BMI of people with diabetes is the same as those without diabetes

H1 = The mean BMI of those with diabetes is different from those without diabetes

After independent t-test, p-value = 0.00000000000000022, therefore we reject the null hypothesis. This means there is significant difference in BMI between individuals with and without diabetes.

* Alcohol Year and Relationship Status

H0 = There is no difference between the amount of alcohol consumed and the relationship status (single or committed)

H1 = There amount of alcohol consumed is different across relationship status

After independent t-test, p-value = 0.0000000609, therefore, we reject the null hypothesis.