Analyzing Health Data for a Community Health Program

K.K.D. Sandali Navoda Gunathilaka – s15355

Table of Contents

Table of Figure

List of Tables

1.0 Introduction

In today's dynamic healthcare landscape, community health initiatives play a pivotal role in addressing the well-being of residents in specific neighborhoods. This case study delves into a community health program designed to improve the overall health and well-being of residents within a targeted neighborhood. The program's objective was to assess its effectiveness and utilize the gathered insights to make informed decisions for future interventions.

In response to the growing concern over the health disparities and well-being of residents in a particular neighborhood, a dedicated public health team initiated a comprehensive community health program. Over a period of 12 months, the team implemented a range of initiatives aimed at enhancing the overall health and quality of life for the residents. These initiatives included health education workshops, free health screenings, and improved access to nutritious food options.

The primary goals of the community health program were as follows:

1. Reduction of Health Risk Factors: The program aimed to reduce the prevalence of specific health risk factors that were prevalent in the community

2. Improving Health Awareness: The program aimed to enhance health awareness and knowledge among residents. This included promoting healthy lifestyle choices, preventive healthcare measures, and early detection of health issues.

To evaluate the program's effectiveness, data was collected through pre-program and post-program health assessments conducted within the community. The dataset comprises a wealth of information, encompassing participants' demographics, health risk factors, and health outcomes.

By examining the program's successes, challenges, and lessons learned, it offers essential insights for future community health interventions. Furthermore, it highlights the importance of data-driven decision-making in public health, as well as the potential to create sustainable, positive changes in residents lives.

2.0 Objectives of the study

1) Object1

Assess the changes in health risk factors (obesity and hypertension) before and after the program

2) Object 2

Determine if there are any significant differences between various demographic groups in terms of health outcomes.

3) Object 3

Identify correlations between health risk factors and health outcomes.

3.0 Data Description

Attributed Information

CaseStudy.xlsx (Health Data for a Community Health Program)

Define Variables

Table 1:Variable Define Table

Qualitative/ Quantitative	**Variable name**	**Description**
Qualitative	ID	Unique identifier for each participant
	Gender	Gender of the participant (Male/Female/Other)
	Ethnicity	Ethnic background of the participant. (Asia, Africa, Caucasian, Hispanic)
	Pre-Hypertension	Whether the participant had hypertension before the program (Yes/No)
	Pre-Obesity	Whether the participant was obese before the program (Yes/No)
	Post-Hypertension	Whether the participant had hypertension after the program (Yes/No)
	Post-Obesity	Whether the participant was obese after the program (Yes/No)
Quantitative	Age	Age of the participant.
	Pre-Weight (kg)	Weight of the participant before the program.
	Pre-Height (cm)	Height of the participant before the program.
	Pre-BMI	Body Mass Index before the program.
	Post-Weight (kg)	Weight of the participant after the program.
	Post-Height (cm)	Height of the participant after the program.
	Post-BMI	Body Mass Index after the program.

Data Cleaning

1) Checking the variables

```
> colnames(Health_Data)
 [1] "ID"                 "Age"             "Gender"          "Ethnicity"
 [5] "Pre-Weight"         "Pre-Height"      "Pre-BMI"         "Pre-Post-Hypertension"
 [9] "Pre-Obesity"        "Post weight"     "Post- height"    "Post BMI"
[13] "Post-Hypertension"  "Post-Obesity"
```

Figure 1: Column names in Case Study data set

Data set given variables names are not good to data calculations in R studio. Because (-) symbol includes in some variables and Pre-Post-Hypertension variable not define. Change variable names (-) symbol to the underscore (_) and "Pre-Post-Hypertension" variable name change to the "Pre-Hypertension" variable name.

```
> colnames(Health_Data)
 [1] "ID"           "Age"          "Gender"            "Ethnicity"      "Pre_Weight"
 [6] "Pre_Height"   "Pre_BMI"      "Pre_Hypertension"  "Pre_Obesity"    "Post_Weight"
[11] "Post_Height"  "Post_BMI"     "Post_Hypertension" "Post_Obesity"
```

Figure 2: Column Names after the correcting and changing names using RStudio

2) Checking the duplicate values and missing values in data set

```
> #Check for missing values
> sum(is.na(Health_Data))
[1] 0
> #Check Duplicates
> sum(duplicated(Health_Data))
[1] 0
```

No any Missing values or

Duplicate values in dataset

Figure 3: Sum of Missing values and Duplicate values

3) Checking Unique values in Categorical variables

```
> #Unique Values
> unique(Health_Data$Gender)
[1] Male   Female Other
Levels: Female Male Other
> unique(Health_Data$Ethnicity)
[1] Caucasian African   Asian     Hispanic
Levels: African Asian Caucasian Hispanic
> unique(Health_Data$Pre_Hypertension)
[1] No  Yes
Levels: No Yes
> unique(Health_Data$Pre_Obesity)
[1] Yes No
Levels: No Yes
> unique(Health_Data$Post_Hypertension)
[1] No  Yes
Levels: No Yes
> unique(Health_Data$Post_Obesity)
[1] No  Yes
Levels: No Yes
```

Figure 4: Unique values for categorical variables

By using above data in categorical variable,

- Gender has unique levels Male, Female and Other
- Ethnicity have unique levels African, Asian, Hispanic, Caucasian
- Pre-Obesity have Yes and No unique levels
- Post-Obesity have Yes and No unique levels
- Pre-Hypertension have Yes and No unique levels
- Post-Hypertension have Yes and No unique levels

Summary of the Health Data Variables

```
> summary(Health_Data)
       ID              Age            Gender        Ethnicity     Pre_Weight     Pre_Height        Pre_BMI     
 Min.   :  1.00   Min.   :23.00   Female:128   African  :77   Min.   :68.0   Min.   :154.0   Min.   :24.14  
 1st Qu.: 73.25   1st Qu.:35.00   Male  :145   Asian    :62   1st Qu.:75.0   1st Qu.:160.0   1st Qu.:26.15  
 Median :145.50   Median :41.00   Other : 17   Caucasian:77   Median :79.0   Median :168.0   Median :28.22  
 Mean   :145.50   Mean   :42.14                Hispanic :74   Mean   :81.1   Mean   :168.5   Mean   :28.64  
 3rd Qu.:217.75   3rd Qu.:49.00                               3rd Qu.:88.0   3rd Qu.:175.8   3rd Qu.:30.47  
 Max.   :290.00   Max.   :66.00                               Max.   :96.0   Max.   :183.0   Max.   :35.94  
 Pre_Hypertension Pre_Obesity  Post_Weight     Post_Height       Post_BMI     Post_Hypertension Post_Obesity
 No :121          No :158     Min.   :63.00   Min.   :154.0   Min.   :22.86   No :162           No :162     
 Yes:169          Yes:132     1st Qu.:70.00   1st Qu.:160.0   1st Qu.:24.49   Yes:128           Yes:128     
                              Median :75.00   Median :168.0   Median :26.27                                 
                              Mean   :76.32   Mean   :168.5   Mean   :26.96                                 
                              3rd Qu.:83.00   3rd Qu.:175.8   3rd Qu.:29.02                                 
                              Max.   :91.00   Max.   :183.0   Max.   :33.98                                 
```

Figure 5: Summary of the Health Data

Important observations of summary of Health Data :

- No of variables: 14
- Total Observations: 290 Observations
- In categorical variables: *Gender variable have 128 Females, 145 Males and 17 others

  *Ethnicity variable have 77 African, 62 Asian, 77 Caucasian and 74 Hispanic

  * Pre-Hypertension variable have 121 No and 169 Yes

  * Pre-Obesity variable have 158 No and 132 Yes

  * Post-Hypertension variable have 162 No and 128 Yes

  * Post-Obesity variable have 162 No and 128 Yes.

4.0 Data analysis and Interpretation

Check parametric assumptions are Valid or Invalid to the variables

The parametric assumptions are as below,

Normality: Data have a normal distribution.

Homogeneity of Variances: Data from multiple groups have the same variance.

Linearity: Data have a linear relationship.

Independence: Data are independent.

Check Normality assumption are valid or not to the data

Ho : Data are Normal Distributed

H1: Data are Not Normal Distributed

```
> qqnorm(Health_Data$Pre_BMI)
> qqline(Health_Data$Pre_BMI)
> shapiro.test(Health_Data$Pre_BMI)

        Shapiro-Wilk normality test

data:  Health_Data$Pre_BMI
W = 0.94117, p-value = 2.384e-09

> qqnorm(Health_Data$Post_BMI)
> qqline(Health_Data$Post_BMI)
> shapiro.test(Health_Data$Post_BMI)

        Shapiro-Wilk normality test

data:  Health_Data$Post_BMI
W = 0.93482, p-value = 5.464e-10

> qqnorm(Health_Data$Pre_Height)
> qqline(Health_Data$Pre_Height)
> shapiro.test(Health_Data$Pre_Height)

        Shapiro-Wilk normality test

data:  Health_Data$Pre_Height
W = 0.93219, p-value = 3.047e-10
```

For Pre-BMI

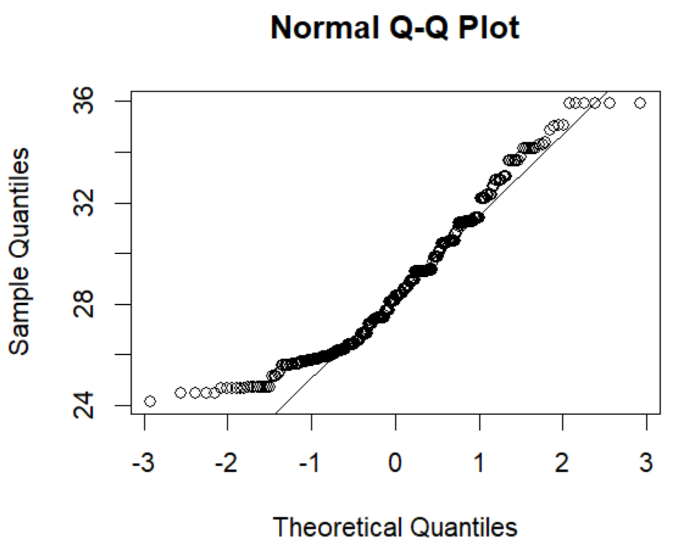


Figure 6: Normal Q-Q Plot for Pre-BMI variable

Figure 7:Check normality for continuous variables

In here considering Shapiro-Wilk Normality Test If p value greater than significant level (alfa=5%) we do not reject Ho and data are normal distributed. By considering Normal Q-Q plot can determine that if the data is normally distributed, the points will fall on the 45-degree reference line.

Firstly, That way Check Normality assumptions for Pre-BMI.

In Shapiro Wilk normality test p value is 2.384×10^{-9} . Therefore, It's less than significant level 0.05. Then reject Ho. So, Data are not normally distributed

In above Normal Q-Q plot for Pre-BMI show data are deviated from the 45-degree reference line. It's indicated also data are not normally distributed. Therefore, Pre-BMI variable data are not normally distributed.

In that way showing other variables to the Shapiro Wilk Normality Test and drawing Normal Q-Q plot can get following data.

Table 2:Check Normality Assumption for the continuous variables

Variable	Shapiro Wilk Test P-Value	Normal Q-Q Plot	Data are Normally Distributed or not
Pre-BMI	2.384×10^{-9}	Data deviated from the 45-degree reference line.	Not Normally distributed
Pre-Height	3.047×10^{-10}	Data deviated from the 45-degree reference line.	Not Normally distributed
Pre-Weight	1.228×10^{-8}	Data deviated from the 45-degree reference line.	Not Normally distributed
Post-BMI	5.464×10^{-10}	Data deviated from the 45-degree reference line.	Not Normally distributed
Post-Height	3.047×10^{-10}	Data deviated from the 45-degree reference line.	Not Normally distributed
Post-Weight	1.586×10^{-8}	Data deviated from the 45-degree reference line.	Not Normally distributed
Age	2.984×10^{-5}	Data deviated from the 45-degree reference line.	Not Normally distributed

This value shows Pre-BMI, Pre-Height, Pre-Weight, Post-BMI, Post-Height, Post-Weight, Age are not Normally distributed and these variables are Normality assumptions violated.

So, variables do not meet the assumptions required for parametric testing; therefore, **Non-Parametric Tests** are the appropriate choice.

Object 1 :Assess the changes in health risk factors (obesity and hypertension) before and after the program.

Considering Count of the Obesity and Hypertension Changes

In our analysis of health data, we categorized the variable 'Obesity' into two distinct states: Pre-Obesity (Yes or No) and Post-Obesity (Yes or No) as categorical variables. To assess changes in obesity status, we compared these two variables.

Specifically, if an individual's Pre-Obesity status was 'Yes,' and their Post-Obesity status became 'No,' we classified this as 'Obesity Decreasing.' Conversely, if an individual's Pre-Obesity status was 'No,' and their Post-Obesity status changed to 'Yes,' we categorized this as 'Obesity Increased.' Finally, when both Pre-Obesity and Post-Obesity statuses were 'No,' we labeled it as 'No Changes.' Our analysis will include a count of each type of change (Increased, Decreased, or No Changes) to provide valuable insights into the impact of our intervention program on obesity trends within our study population.

After that calculated Increased, Decreased and No Change counts can examine they are as follows.

```
> Count_Increased_Obe = sum(Change_Obesity == "Increased")
> Count_Increased_Obe
[1] 0
> Count_Decreased_Obe = sum(Change_Obesity == "Decreased")
> Count_Decreased_Obe
[1] 4
> Count_No_Change_Obe= sum(Change_Obesity == "No Change")
> Count_No_Change_Obe
[1] 286
```

Figure 8: Counts of changes in Obesity

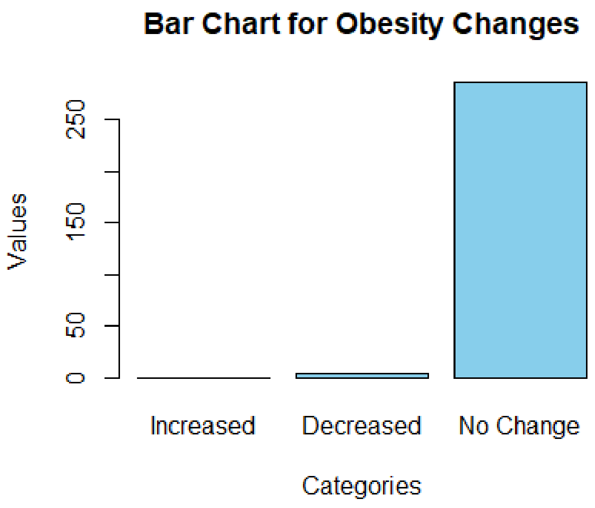
Pre-Obesity “No” and Post-Obesity “Yes” = “Obesity Increase” observation count is 0

Pre-Obesity “Yes” and Post-Obesity “No” = “Obesity Decrease” observation count is 4

(Pre-Obesity “Yes” and Post-Obesity “Yes”) or (Pre-Obesity “No” and Post-Obesity “No”) = “Obesity No Change” observation count is 286.

Therefore, considering these observations it can conclude that Most of the Participants Obesity is not changes after the program. But 4 peoples Obesity is decreasing. It seems 4 peoples are improving their health awareness after the program. This suggests that these four participants made positive lifestyle changes as a result of the program.

To enhance the clarity of our findings, we can visualize the results using a bar chart as follows.



The utilization of a bar chart further enhances the clarity of our results. Then finally we can conclude that health program is to some extent access to the change in obesity.

Figure 9: Bar Chart for the counts of change of obesity

Certainly, same as the Obesity when evaluating health risk factors such as Hypertension, we also examine two categorical variables: Pre-Hypertension (status before the program) and Post-Hypertension (status after the program). These variables can take on two distinct values, namely "Yes" and "No." Similar to our approach with Obesity, we analyze the differences in Hypertension outcomes, specifically categorizing them into three groups: "Increasing," "Decreasing," and "No Change."

```
> Count_Increased_Hyp = sum(Change_Hypertension == "Increased")
> Count_Increased_Hyp
[1] 0
> Count_Decreased_Hyp = sum(Change_Hypertension == "Decreased")
> Count_Decreased_Hyp
[1] 41
> Count_No_Change_Hyp = sum(Change_Hypertension == "No Change")
> Count_No_Change_Hyp
[1] 249
```

Figure 10: Counts for changes in Hypertension

Pre-Hypertension “No” and Post- Hypertension “Yes” = “Hypertension Increase” observation count is 0

Pre- Hypertension “Yes” and Post- Hypertension “No” = “Hypertension Decrease” observation count is 41

(Pre- Hypertension “Yes” and Post- Hypertension “Yes”) or (Pre- Hypertension “No” and Post- Hypertension “No”) = “Hypertension No Change” observation count is 249.

To enhance the clarity of our findings, we can visualize the results using a bar chart as follows, Therefore, considering these observations it can conclude that Most of the Participants Hypertension is not changes after the program. But 41 peoples Hypertension is decreasing. It seems 41 peoples are improving their health awareness after the program. This suggests that these 41 participants made positive lifestyle changes as a result of the program

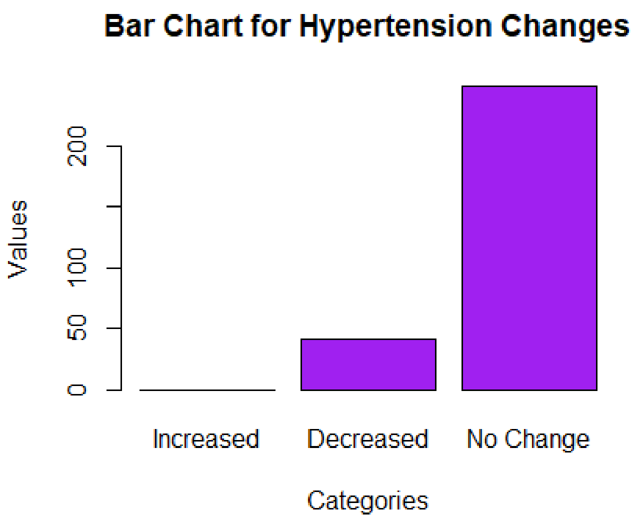


Figure 11: Bar Chart for the Hypertension Changes

Using a bar chart provides additional visual clarity to our findings. In conclusion, we can infer that the health program has a discernible impact on Hypertension level.

Considering Wilcoxon's Sign Rank Test for Macheted Pairs test For Weight

Firstly Considering Two-Sided Test

Hypothesis

$H_0 : M_{(Pre\ weight - Post\ weight)} = 0 \qquad H_1 : M_{(Pre\ weight - Post\ weight)} \neq 0$

$M_{(Pre\ weight - Post\ weight)}$ is Median difference between weight in the before and after the health program

Test Statistic

Conducting a two-sided hypothesis test using Wilcoxon's signed-rank test for matched pairs yields a p-value that assesses the statistical significance of the outcome. This succinctly captures the essence of the analysis without unnecessary details.

```
> #wilcox.test(Health_Data$Pre_Weight,Health_Data$Post_Weight,paired=TRUE)
> wilcox.test(Health_Data$Pre_Weight,Health_Data$Post_Weight,mu=0, alternative ="two.sided",paired = T,exact=F,con
f.int = T,conf.level = 0.95)

        Wilcoxon signed rank test with continuity correction

data:  Health_Data$Pre_Weight and Health_Data$Post_Weight
V = 42195, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 4.999977 4.999984
sample estimates:
(pseudo)median
      4.999954
```

Figure 12:Wilcoxon's Sign Rank Test for matched Paired Two sided in Pre weight and Post Weight

In here, P value $=2.2\times 10^{-16}$

Critical Value

Chosen Significance level is 0.05

Decision Rule

The p-value in the output represents the significance level of the test. Typically compare this p-value to our chosen significance level (0.05) to determine whether the difference between the paired samples is statistically significant. If the p-value is less than or equal our chosen significance level, reject Ho and it can conclude that there is a significant difference between the paired samples.

Conclusion

Therefore, In here P value (2.2×10^{-16}) < Significance value (0.05). Then reject Ho at 5% significance level. Therefore, we have enough evidence to conclude that Median value of Pre and Post Weights are significantly difference at 5% significance level.

One sided test

Hypothesis

$H_0 : M_{(Pre\ weight - Post\ weight)} \leq 0 \qquad H_1 : M_{(Pre\ weight - Post\ weight)} > 0$

$M_{(Pre\ weight - Post\ weight)}$ is Median difference between weight in the before and after the health program

Test Statistic

Conducting a One-sided hypothesis test using Wilcoxon's signed-rank test for matched pairs yields a p-value that assesses the statistical significance of the outcome. This succinctly captures the essence of the analysis without unnecessary details.

```
> #~~~~~~~~~ Ho:M(Pre-Post)=<0  H1:M(Pre-Post) > 0
> wilcox.test(Health_Data$Pre_Weight,Health_Data$Post_Weight,mu=0, alternative ="greater",paired = T,exact=F,conf.i
nt = T,conf.level = 0.95)

        Wilcoxon signed rank test with continuity correction

data:  Health_Data$Pre_Weight and Health_Data$Post_Weight
V = 42195, p-value < 2.2e-16
alternative hypothesis: true location shift is greater than 0
95 percent confidence interval:
 4.99996      Inf
sample estimates:
(pseudo)median
      4.999954
```

Figure 13:Wilcoxon's Matched paired test for One sided in Pre and Post weight

In here, P value $=2.2\times 10^{-16}$

Significance level Value

Chosen Significance level is 0.05

Decision Rule

The p-value in the output represents the significance level of the test. Typically compare this p-value to our chosen significance level (0.05) to determine whether the difference between the paired samples is statistically significant. If the p-value is less than or equal our chosen significance level, reject Ho and it can conclude that there is a significant difference between the paired samples.

Conclusion

Therefore, In here P value (2.2×10^{-16}) < Significance value (0.05). Then reject Ho at 5% significance level. Therefore, we have enough evidence to conclude that Median value of Pre Weight is greater than Median value of the Post Weights at 5% significance level.

Therefore, Considering above results it shows After the health program, as Median Weight is decreases. If Weight is decreasing then Obesity is also decreasing practically. Remember that even losing what seems like a small amount of weight it can significantly reduce your risk of developing obesity-related complications like Hypertension. It can be concluded that the program effectively addresses the health risk factor of obesity and Hypertension.

Object 2 : Determine if there are any significant differences between various demographic groups in terms of health outcomes.

Demographic groups are categories of people who share certain characteristics or traits based on factors such as age, gender, race, ethnicity, income, education, marital status, and more. In this data set considering demographic group as Age, Gender and Ethnicity. Gender groups divide individuals into categories such as male, female and other. Ethnicity group divided into Asia, Africa, Caucasian, Hispanic. Analyzing these groups helps identify and address health disparities, develop targeted interventions, and promote equitable access to healthcare services to improve overall public health.

The health outcome variables (Pre and Post BMI, Weight, Height) in this dataset do not follow a normal distribution. Therefore, we should apply non-parametric tests to analyze these variables. Our objective is to examine potential significant differences in health outcomes among various demographic groups, considering each health outcome individually.

Testing significance difference between Ethnicity groups in Health outcomes

In here, four distinct ethnic groups in this dataset, and our aim to assess if there are significant differences among these groups in terms of health outcomes. To do so, we will employ the Kruskal-Wallis Test as an appropriate statistical method.

By the way, Considering Health outcomes individually (Pre-BMI, Post-BMI, Pre-Weight, Post-Weight, Pre-Height, Post-Height)

Create a data frame as follows,

Pre_Weight_DataFrame

African	Asian	Caucasian	Hispanic
72	76	85	160
80	75	92	180
78	75	88	155
88	73	90	175
78	75	95	158
70	78	84	155
80	75	82	180
78	73	92	162
88	76	88	175
78	75	93	165
72	75	95	178
80	73	85	155
78	76	88	180
88	75	92	162
78	75	93	175
72	73	95	165
80	76	85	155
78	92	85	180
88	90	88	162
78	75	92	175
84	75	93	165
80	73	95	165
75	90	85	158
79	75	72	180
88	75	68	155
78	73	84	180
88	90	75	162
78	75	92	175
79	88	93	165
88	73	95	180
78	77	85	155
88	72	75	180
78	76	92	162
79	86	93	175

Pre_BMI_DataFrame

African	Asian	Caucasian	Hispanic
27.44	26.30	27.76	27.34
25.25	24.49	32.63	25.31
30.47	29.30	32.32	28.32
26.54	26.81	29.39	29.87
30.47	24.49	29.32	28.85
27.34	26.98	25.93	28.31
24.14	29.30	25.85	25.93
31.24	26.81	33.82	27.44
27.77	24.73	34.38	29.87
30.47	26.01	30.37	25.71
28.13	29.30	29.32	26.57
24.69	26.81	26.23	28.40
31.24	24.73	32.30	25.93
27.77	26.01	35.94	27.44
30.47	29.30	30.37	29.87
28.13	26.81	29.32	25.71
24.69	24.73	26.23	28.40
31.24	35.94	26.84	25.93
27.77	33.05	32.30	27.44
30.47	26.01	35.94	29.87
26.57	29.30	30.37	25.71
24.69	26.81	29.32	32.30
26.01	33.05	26.23	31.24
25.80	26.01	28.13	24.69
28.73	29.30	28.40	28.31
31.24	26.81	25.93	25.93
27.77	33.05	29.30	27.44
30.47	26.01	35.94	29.87
25.80	32.66	30.37	25.71
28.73	25.87	29.32	24.69
31.24	30.85	26.23	28.31
27.77	26.47	29.30	25.93
30.47	30.37	35.94	27.44

Figure 14:Ethnicity group in Pre-BMI Data Frame

Figure 15:Ethnicity groups in Pre-Weight Data Frame

Hypothesis

H_0: There is no difference in location among the health outcome populations from which the Asian, Africa, Caucasian, Hispanic samples have been drawn

H_1: There is a difference in location in at least one health outcome population from which the Asian, African, Caucasian and Hispanic samples have been drawn

(In here, we have to consider firstly, hypothesis in health outcome= Pre-BMI / secondly, health outcome=Post-BMI / thirdly, health outcome=Pre-Weight such that Post-Weight, Pre-Height, Post-Height also separately)

<u>Test Statistic</u>

The p-value in a Kruskal-Wallis test helps you make an inference about whether there are differences among the Ethnicity groups being compared.

Therefore, Calculate P-values for Kruskal Wallis test as follows,

```
> # Testing significance difference between Ethnicity groups in Pre_BMI
> Kruskal_Ethnicity_Pre_BMI=kruskal.test(Health_Data$Pre_BMI~Health_Data$Ethnicity)
> Kruskal_Ethnicity_Pre_BMI

        Kruskal-Wallis rank sum test

data:  Health_Data$Pre_BMI by Health_Data$Ethnicity
Kruskal-Wallis chi-squared = 32.764, df = 3, p-value = 3.612e-07

> # Testing significance difference between Ethnicity groups in Post_BMI
> Kruskal_Ethnicity_Post_BMI=kruskal.test(Health_Data$Post_BMI~Health_Data$Ethnicity)
> Kruskal_Ethnicity_Post_BMI

        Kruskal-Wallis rank sum test

data:  Health_Data$Post_BMI by Health_Data$Ethnicity
Kruskal-Wallis chi-squared = 35.326, df = 3, p-value = 1.04e-07

> # Testing significance difference between Ethnicity groups in Pre_Weight
> Kruskal_Ethnicity_Pre_Weight=kruskal.test(Health_Data$Pre_Weight~Health_Data$Ethnicity)
> Kruskal_Ethnicity_Pre_Weight

        Kruskal-Wallis rank sum test

data:  Health_Data$Pre_Weight by Health_Data$Ethnicity
Kruskal-Wallis chi-squared = 72.157, df = 3, p-value = 1.473e-15

> # Testing significance difference between Ethnicity groups in Post_Weight
> Kruskal_Ethnicity_Post_Weight=kruskal.test(Health_Data$Post_Weight~Health_Data$Ethnicity)
> Kruskal_Ethnicity_Post_Weight

        Kruskal-Wallis rank sum test

data:  Health_Data$Post_Weight by Health_Data$Ethnicity
Kruskal-Wallis chi-squared = 67.277, df = 3, p-value = 1.634e-14

> # Testing significance difference between Ethnicity groups in Pre_Height
> Kruskal_Ethnicity_Pre_Height=kruskal.test(Health_Data$Pre_Height~Health_Data$Ethnicity)
> Kruskal_Ethnicity_Pre_Height

        Kruskal-Wallis rank sum test

data:  Health_Data$Pre_Height by Health_Data$Ethnicity
Kruskal-Wallis chi-squared = 6.8226, df = 3, p-value = 0.07777

> # Testing significance difference between Ethnicity groups in Post_Height
> Kruskal_Ethnicity_Post_Height=kruskal.test(Health_Data$Post_Height~Health_Data$Ethnicity)
> Kruskal_Ethnicity_Post_Height

        Kruskal-Wallis rank sum test

data:  Health_Data$Post_Height by Health_Data$Ethnicity
Kruskal-Wallis chi-squared = 6.8226, df = 3, p-value = 0.07777
```

Figure 16: Kruskal Wallis Test for define Significance difference in Ethnicity groups in health outcomes

In this way calculate for P-value for all Ethnicity in health outcomes as follows,

Table 3: P values for Ethnicity groups in Health Outcomes

Ethnicity group in health outcomes	P-Value	Reject Ho / Accept Ho
Pre-BMI	3.612×10^{-7}	Reject Ho
Post-BMI	1.04×10^{-7}	Reject Ho
Pre-Weight	1.473×10^{-15}	Reject Ho
Post-Weight	1.634×10^{-14}	Reject Ho
Pre-Height	0.07777	Accept Ho
Post-Height	0.07777	Accept Ho

<u>Decision Rule</u>

If the p-value is less than or equal to the chosen significance level (α=0.05), reject the null hypothesis.

<u>Conclusion</u>

In here, Pre-BMI, Post-BMI, Pre-Weight, Post-Weight health outcomes P-values are less than chosen significance level (α=0.05). Then reject Ho in these variables. But, Pre-Height and Post-Height variables P-values are greater than chosen significance level (α=0.05). Then Pre-Height and Post-Height variables are do not reject Ho. Therefore, we have enough evidence to conclude that,

There is a difference in location in at least one Pre-Weight population from which the Asian, African, Caucasian and Hispanic samples have been drawn

There is a difference in location in at least one Post-Weight population from which the Asian, African, Caucasian and Hispanic samples have been drawn

There is a difference in location in at least one Pre-BMI population from which the Asian, African, Caucasian and Hispanic samples have been drawn

There is a difference in location in at least one Post-BMI population from which the Asian, African, Caucasian and Hispanic samples have been drawn

There is no difference in location in Pre-Height one or more populations from which the Asian, Africa, Caucasian, Hispanic samples have been drawn

There is no difference in location in Post-Height one or more populations from which the Asian, Africa, Caucasian, Hispanic samples have been drawn

At 5% level of significance.

Testing significance difference between Gender groups in Health outcomes

Pre_Weight_DataFrame_b

Female	Male	Other
72	85	70
92	76	68
88	80	68
78	75	68
75	82	68
68	90	68
73	88	68
78	95	68
72	92	70
70	84	68
92	82	68
88	75	68
78	80	68
75	78	68
72	84	68
73	93	68
78	88	68
70	95	NA
72	92	NA
88	85	NA
92	84	NA
78	76	NA
75	80	NA
72	75	NA
73	84	NA
78	93	NA
70	88	NA
72	95	NA
88	92	NA
92	85	NA
78	85	NA
75	76	NA
72	80	NA

Figure 17: Data frame for the Gender group in Pre-Weight

Pre_BMI_DataFrame_a

Female	Male	Other
27.44	27.76	27.34
32.63	26.30	28.31
32.32	25.25	28.40
30.47	24.49	28.40
29.30	25.31	28.40
28.32	29.39	28.31
26.81	26.54	28.31
30.47	29.32	28.31
28.85	29.87	27.34
27.34	25.93	26.56
33.82	25.85	26.56
34.38	24.49	27.24
31.24	24.14	27.24
29.30	26.98	27.24
27.44	25.93	27.24
26.81	30.37	27.24
30.47	27.77	27.24
25.71	29.32	NA
28.13	29.87	NA
32.30	26.23	NA
35.94	26.57	NA
31.24	24.73	NA
29.30	24.69	NA
27.44	26.01	NA
26.81	25.93	NA
30.47	30.37	NA
25.71	27.77	NA
28.13	29.32	NA
32.30	29.87	NA
35.94	26.23	NA
31.24	26.84	NA

Figure 18:Data frame for the Gender groups in Pre-BMI

Hypothesis

H_0: There is no difference in location among the health outcome populations from which the Male, Female, Other samples have been drawn

H_1: There is a difference in location in at least one health outcome population from which Male, Female, Other samples have been drawn

(In here, we have to consider firstly, hypothesis in health outcome= Pre-BMI / secondly, health outcome=Post-BMI / thirdly, health outcome=Pre-Weight such that Post-Weight, Pre-Height, Post-Height also separately)

Test Statistic

The p-value in a Kruskal-Wallis test helps you make an inference about whether there are differences among the Gender groups being compared.

Therefore, Calculate P-values for Kruskal Wallis test as follows,

```
> # Testing significance difference between Gender groups in Pre_BMI
> Kruskal_Gender_Pre_BMI=kruskal.test(Health_Data$Pre_BMI~Health_Data$Gender)
> Kruskal_Gender_Pre_BMI

        Kruskal-Wallis rank sum test

data:  Health_Data$Pre_BMI by Health_Data$Gender
Kruskal-Wallis chi-squared = 82.034, df = 2, p-value < 2.2e-16

> # Testing significance difference between Gender groups in Post_BMI
> Kruskal_Gender_Post_BMI=kruskal.test(Health_Data$Post_BMI~Health_Data$Gender)
> Kruskal_Gender_Post_BMI

        Kruskal-Wallis rank sum test

data:  Health_Data$Post_BMI by Health_Data$Gender
Kruskal-Wallis chi-squared = 72.262, df = 2, p-value < 2.2e-16

> # Testing significance difference between Gender group in Pre_Weight
> Kruskal_Gender_Pre_Weight=kruskal.test(Health_Data$Pre_Weight~Health_Data$Gender)
> Kruskal_Gender_Pre_Weight

        Kruskal-Wallis rank sum test

data:  Health_Data$Pre_Weight by Health_Data$Gender
Kruskal-Wallis chi-squared = 82.42, df = 2, p-value < 2.2e-16

> # Testing significance difference between Gender group in Post_Weight
> Kruskal_Gender_Post_Weight=kruskal.test(Health_Data$Post_Weight~Health_Data$Gender)
> Kruskal_Gender_Post_Weight

        Kruskal-Wallis rank sum test

data:  Health_Data$Post_Weight by Health_Data$Gender
Kruskal-Wallis chi-squared = 80.913, df = 2, p-value < 2.2e-16
```

```
> # Testing significance difference between Gender groups in Pre_Height
> Kruskal_Gender_Pre_Height=kruskal.test(Health_Data$Pre_Height~Health_Data$Gender)
> Kruskal_Gender_Pre_Height

        Kruskal-Wallis rank sum test

data:  Health_Data$Pre_Height by Health_Data$Gender
Kruskal-Wallis chi-squared = 223.34, df = 2, p-value < 2.2e-16

> # Testing significance difference between Gender groups in Post_Height
> Kruskal_Gender_Post_Height=kruskal.test(Health_Data$Post_Height~Health_Data$Gender)
> Kruskal_Gender_Post_Height

        Kruskal-Wallis rank sum test

data:  Health_Data$Post_Height by Health_Data$Gender
Kruskal-Wallis chi-squared = 223.34, df = 2, p-value < 2.2e-16
```

Figure 19: Kruskal Wallis test for define Gender group in health outcomes

In this way calculate for P-value for all Gender in health outcomes as follows,

Table 4: Kruskal Walis test p values for define Gender groups in health outcomes

Gender group in health outcomes	P-Value	Reject Ho / Accept Ho
Pre-BMI	2.2×10^{-16}	Reject Ho
Post-BMI	2.2×10^{-16}	Reject Ho
Pre-Weight	2.2×10^{-16}	Reject Ho
Post-Weight	2.2×10^{-16}	Reject Ho
Pre-Height	2.2×10^{-16}	Reject Ho
Post-Height	2.2×10^{-16}	Reject Ho

Decision Rule

If the p-value is less than or equal to the chosen significance level (α=0.05), reject the null hypothesis.

Conclusion

In here, Pre-BMI, Post-BMI, Pre-Weight, Post-Weight, Pre-Height, Post-Height all health outcomes P-values are less than chosen significance level (α=0.05). Then reject Ho in these variables at 5% level of significance. Therefore, we have enough evidence to conclude that.

There is a difference in location in at least one Pre-Weight population from which the Male, Female and Other samples have been drawn

There is a difference in location in at least one Post-Weight population from which the Male, Female and Other samples have been drawn

There is a difference in location in at least one Pre-BMI population from which the Male, Female and Other samples have been drawn

There is a difference in location in at least one Post-BMI population from which the Male, Female and Other samples have been drawn

There is a difference in location in at least one Pre-Height population from which the Male, Female and Other samples have been drawn

There is a difference in location in at least one Post-Height population from which the Male, Female and Other samples have been drawn

At 5 % level of significance.

Testing significance difference between Age groups in Health outcomes

Hypothesis

H_0: There is no difference in location among the health outcome populations from which the k samples have been drawn

H_1: There is a difference in location health outcome one or more populations from which k samples have been drawn

(In here, we have to consider firstly, hypothesis in health outcome= Pre-BMI / secondly, health outcome=Post-BMI / thirdly, health outcome=Pre-Weight such that Post-Weight, Pre-Height, Post-Height also separately)

k= age in 32, 41, 28, 57, 31, 49, 36, 44, 29, 50, 39, 33, 45, 52, 37, 55, 42, 30, 47, 25, 62, 48, 35, 27, 51, 40, 34, 53, 54, 43, 38, 24, 61, 26, 46, 56, 23, 60, 63, 58, 66, 59

```
> unique(Health_Data$Age)
 [1] 32 41 28 57 31 49 36 44 29 50 39 33 45 52 37 55 42 30 47 25 62 48 35 27 51 40 34 53 54 43 38 24 61 26 46
[36] 56 23 60 63 58 66 59
```

Figure 20: Unique values for Age variable

Test Statistic

The p-value in a Kruskal-Wallis test helps you make an inference about whether there are differences among the Age being compared. Therefore, Calculate P-values for Kruskal Wallis test as follows,

```
> # Testing significance difference between Age in Pre_Weight
> Kruskal_Age_Pre_Weight=kruskal.test(Health_Data$Pre_Weight~Health_Data$Age)
> Kruskal_Age_Pre_Weight

        Kruskal-Wallis rank sum test

data:  Health_Data$Pre_Weight by Health_Data$Age
Kruskal-Wallis chi-squared = 199.2, df = 41, p-value < 2.2e-16

> # Testing significance difference between Age in Post_Weight
> Kruskal_Age_Post_Weight=kruskal.test(Health_Data$Post_Weight~Health_Data$Age)
> Kruskal_Age_Post_Weight

        Kruskal-Wallis rank sum test

data:  Health_Data$Post_Weight by Health_Data$Age
Kruskal-Wallis chi-squared = 200.54, df = 41, p-value < 2.2e-16

> # Testing significance difference between Age in Pre_BMI
> Kruskal_Age_Pre_BMI=kruskal.test(Health_Data$Pre_BMI~Health_Data$Age)
> Kruskal_Age_Pre_BMI

        Kruskal-Wallis rank sum test

data:  Health_Data$Pre_BMI by Health_Data$Age
Kruskal-Wallis chi-squared = 165.92, df = 41, p-value < 2.2e-16

> # Testing significance difference between Age in Post_BMI
> Kruskal_Age_Post_BMI=kruskal.test(Health_Data$Post_BMI~Health_Data$Age)
> Kruskal_Age_Post_BMI

        Kruskal-Wallis rank sum test

data:  Health_Data$Post_BMI by Health_Data$Age
Kruskal-Wallis chi-squared = 170.63, df = 41, p-value < 2.2e-16
```

```
> # Testing significance difference between Age in Pre_Height
> Kruskal_Age_Pre_Height=kruskal.test(Health_Data$Pre_Height~Health_Data$Age)
> Kruskal_Age_Pre_Height

        Kruskal-Wallis rank sum test

data:  Health_Data$Pre_Height by Health_Data$Age
Kruskal-Wallis chi-squared = 120.3, df = 41, p-value = 1.011e-09

> # Testing significance difference between Age in Post_Height
> Kruskal_Age_Post_Height=kruskal.test(Health_Data$Post_Height~Health_Data$Age)
> Kruskal_Age_Post_Height

        Kruskal-Wallis rank sum test

data:  Health_Data$Post_Height by Health_Data$Age
Kruskal-Wallis chi-squared = 120.3, df = 41, p-value = 1.011e-09
```

Figure 21:Kruskal Wallis test for significance difference in Age group in health outcomes

In this way calculate for P-value for all Age in health outcomes as follows,

Table 5: P values for Kruskal Wallis test Age group in health outcomes

Age group in health outcomes	P-Value	Reject Ho / Accept Ho
Pre-BMI	2.2×10^{-16}	Reject Ho
Post-BMI	2.2×10^{-16}	Reject Ho
Pre-Weight	2.2×10^{-16}	Reject Ho
Post-Weight	2.2×10^{-16}	Reject Ho
Pre-Height	2.2×10^{-16}	Reject Ho
Post-Height	2.2×10^{-16}	Reject Ho

<u>Decision Rule</u>

If the p-value is less than or equal to the chosen significance level (α=0.05), reject the null hypothesis.

<u>Conclusion</u>

In here, Pre-BMI, Post-BMI, Pre-Weight, Post-Weight, Pre-Height, Post-Height all health outcomes P-values are less than chosen significance level (α=0.05). Then reject Ho in these variables at 5% level of significance. Therefore, we have enough evidence to conclude that.

There is a difference in location in Pre-Weight one or more populations from which the Age k samples have been drawn

There is a difference in location in Post-Weight one or more populations from which the Age k samples have been drawn

There is a difference in location in Pre-BMI one or more populations from which the Age k samples have been drawn

There is a difference in location in Post-BMI one or more populations from which the Age k samples have been drawn

There is a difference in location in Pre-Height one or more populations from which the Age k samples have been drawn

There is a difference in location in Post-Height one or more populations from which the Age k samples have been drawn

At 5 % level of significance.

Therefore, Overall data considering can determine There have significant difference between various demographic groups of the health outcomes except Ethnicity group in Pre Height and Post Height.

Object 3 : Identify correlations between health risk factors and health outcomes.

Identifying correlations between health risk factors and health outcomes is a fundamental aspect of public health research. This process involves systematically studying and analyzing data to understand the relationships between specific factors that contribute to health problems and the resulting health outcomes. Identifying correlations between health risk factors and health outcomes helps us pinpoint the root causes of health problems and enabling us to reduce the risk.

To assess the correlation between health risk factors and health outcomes, we typically employ the Pearson correlation coefficient test. However, it's crucial to note that the Pearson correlation test is specifically designed for measuring the strength and direction of linear relationships between pairs of continuous variables. In our case, the health risk factors, such as Obesity and Hypertension, are categorical variables. Therefore, we need to transform them into continuous variables by creating dummy variables for Obesity and Hypertension. This transformation allows us to appropriately apply the Pearson correlation coefficient test and examine potential associations between these categorical risk factors and health outcomes.

To convert the categorical health risk factors (Obesity and Hypertension) into dummy variables (0 and 1), we will assign "1" to represent "Yes" and "0" to represent "No" for each variable. Once after completed this conversion, then can proceed to examine the correlation between these transformed health risk factors and the health outcomes.

```
#@@@@@@@@@@@@@@@@@@@ Assign Categorical variables to Dummy variable @@@@@@@@@@@@@@@@@@@
Health_Data$Pre_Obesity=ifelse(Health_Data$Pre_Obesity == "Yes", 1, 0)
Health_Data$Post_Obesity <- ifelse(Health_Data$Post_Obesity == "Yes", 1, 0)
Health_Data$Pre_Hypertension <- ifelse(Health_Data$Pre_Hypertension == "Yes", 1, 0)
Health_Data$Post_Hypertension <- ifelse(Health_Data$Post_Hypertension == "Yes", 1, 0)
```

Figure 22:Assign Categorical variables to the dummy variables (0 and 1)

Pearson Correlation Coefficient Test

Hypothesis

H_0 : ρ = 0 (The population correlation coefficient is zero; No association between two Variables)

H_1 : ρ ≠ 0 (The population correlation coefficient is not zero; There is an association between two Variables)

First considering Correlation between Pre-Obesity and Health outcomes separately

Test Statistic

```
> #~~~~~~~~~~~~~~~~~~~~~~~~~~ Pearson Correlation Test for Pre_Obesity and Health outcomes
> cor.test(Health_Data$Pre_Obesity,Health_Data$Pre_Weight,method="pearson")

        Pearson's product-moment correlation

data:  Health_Data$Pre_Obesity and Health_Data$Pre_Weight
t = 10.426, df = 288, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4344878 0.6023338
sample estimates:
      cor
0.5234708

> cor.test(Health_Data$Pre_Obesity,Health_Data$Pre_Height,method="pearson")

        Pearson's product-moment correlation

data:  Health_Data$Pre_Obesity and Health_Data$Pre_Height
t = -5.7287, df = 288, p-value = 2.547e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.4195587 -0.2124830
sample estimates:
       cor
-0.3198351
```

Figure 23: Pearson Correlation test for correlation between Health Pre-Obesity and Health outcomes (Pre-Weight, Pre-Height)

Like this can find all correlation values ,

Table 6: Correlation coefficients between Pre-Obesity and Health outcomes

Pre-Obesity Vs Health Outcome variable	**Correlation Value**	**There's an Association (Reject Ho)/ No Association (Do not reject Ho)**	**Correlation Type**
Pre-Weight	0.5234708	There's an Association (Reject Ho)	Moderately positive correlation.
Pre-Height	-0.3198351	There's an Association (Reject Ho)	Low Negative Correlation
Pre-BMI	0.8153709	There's an Association (Reject Ho)	High Positive Correlation
Post-Weight	0.5358848	There's an Association (Reject Ho)	Moderately positive correlation.
Post-Height	-0.3198351	There's an Association (Reject Ho)	Low Negative Correlation
Post-BMI	0.8229475	There's an Association (Reject Ho)	High Positive Correlation

<u>Conclusion</u>

Therefore, we have enough evidence to conclude that there is an association between Pre-Obesity and health outcomes (Pre-Weight, Post-Weight, Pre-Height, Post-Height, Pre-BMI, Post-BMI) at 5% level of significance. Correlation type is showing in the above table.

In here, Pre-Obesity Vs Pre and Post Weights are having Moderately Positive Correlation.

Pre-Obesity Vs Pre and Post Heights are having Low Negative Correlation.

Pre-Obesity Vs Pre and Post BMI are having High Positive Correlation.

Considering Correlation between Post-Obesity and Health outcomes separately

Test Statistic

```
> #~~~~~~~~~~~~~~~~~~~~~~~~~~ Pearson Correlation Test for Post_Obesity and Health outcomes~-
~
> cor.test(Health_Data$Post_Obesity,Health_Data$Pre_Weight,method="pearson")

        Pearson's product-moment correlation

data:  Health_Data$Post_Obesity and Health_Data$Pre_Weight
t = 11.004, df = 288, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4575571 0.6203680
sample estimates:
      cor
0.5440638

> cor.test(Health_Data$Post_Obesity,Health_Data$Pre_Height,method="pearson")

        Pearson's product-moment correlation

data:  Health_Data$Post_Obesity and Health_Data$Pre_Height
t = -5.6028, df = 288, p-value = 4.93e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.4137429 -0.2057532
sample estimates:
       cor
-0.3135032
```

Figure 24:Pearson Correlation test for Post-Obesity and Health outcomes (Pre-Weight, Pre-Height)

Like this can find all correlation values ,

Table 7: Correlation values for Post Obesity and Health outcomes

Post-Obesity Vs Health Outcome variable	**Correlation Value**	**There's an Association (Reject Ho)/ No Association (Do not reject Ho)**	**Correlation Type**
Pre-Weight	0.5440638	There's an Association (Reject Ho)	Moderately positive correlation.
Pre-Height	-0.3135032	There's an Association (Reject Ho)	Low Negative Correlation
Pre-BMI	0.8300645	There's an Association (Reject Ho)	High Positive Correlation
Post-Weight	0.5594242	There's an Association (Reject Ho)	Moderately positive correlation.
Post-Height	-0.3135032	There's an Association (Reject Ho)	Low Negative Correlation

Post-BMI	0.8414126	There's an Association (Reject Ho)	High Positive Correlation

Conclusion

Therefore, we have enough evidence to conclude that there is an association between Post-Obesity and health outcomes (Pre-Weight, Post-Weight, Pre-Height, Post-Height, Pre-BMI, Post-BMI) at 5% level of significance. Correlation type is showing in the above table.

Considering Correlation between Post-Obesity and health outcomes,

Post-Obesity and Pre and Post Weights have Moderately positive correlations.

Post-Obesity and Pre and Post Height have Low Negative Correlation.

Post-Obesity and Pre and Post BMI have High Positive Correlation.

Considering Correlation between Pre-Hypertension and Health outcomes separately

Test Statistic

```
> #~~~~~~~~~~~~~~~~~~~~~~~~~~~ Pearson Correlation Test for Pre_Hypertension and Health outcomes~
~~~~~
> cor.test(Health_Data$Pre_Hypertension,Health_Data$Pre_Weight,method="pearson")

        Pearson's product-moment correlation

data:  Health_Data$Pre_Hypertension and Health_Data$Pre_Weight
t = 3.292, df = 288, p-value = 0.001119
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.0769422 0.2990542
sample estimates:
      cor
0.1904341

> cor.test(Health_Data$Pre_Hypertension,Health_Data$Pre_Height,method="pearson")

        Pearson's product-moment correlation

data:  Health_Data$Pre_Hypertension and Health_Data$Pre_Height
t = -12.784, df = 288, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.6704166 -0.5227479
sample estimates:
       cor
-0.6016992
```

Figure 25: Pearson Correlation test for Pre-Hypertension and Health outcomes (Pre-Weight, Pre-Height)

Like this can find all correlation values ,

Table 8: Correlation values for Pre-Hypertension and Health outcomes

Pre-Hypertension Vs Health Outcome variable	**Correlation Value**	**There's an Association (Reject Ho)/ No Association (Do not reject Ho)**	**Correlation Type**
Pre-Weight	0.1904341	There's an Association (Reject Ho)	Low Positive Correlation
Pre-Height	-0.6016992	There's an Association (Reject Ho)	Moderately negative Correlation
Pre-BMI	0.7600659	There's an Association (Reject Ho)	High Positive Correlation
Post-Weight	0.2120427	There's an Association (Reject Ho)	Low Positive correlation.
Post-Height	-0.6016992	There's an Association (Reject Ho)	Moderately negative Correlation
Post-BMI	0.7529998	There's an Association (Reject Ho)	High Positive Correlation

Conclusion

Therefore, we have enough evidence to conclude that there is an association between Pre-Hypotension and health outcomes (Pre-Weight, Post-Weight, Pre-Height, Post-Height, Pre-BMI, Post-BMI) at 5% level of significance. Correlation type is showing in the above table.

Considering Correlation between Pre-Hypertension and health outcomes,

Pre-Hypertension and Pre and Post Weights have Low Positive correlations.

Pre-Hypertension and Pre and Post Height have Moderately Negative Correlation.

Pre-Hypertension and Pre and Post BMI have High Positive Correlation.

Considering Correlation between Post-Hypertension and Health outcomes separately

Test Statistic

```
> #~~~~~~~~~~~~~~~~~~~~~~~~~ Pearson Correlation Test for Post_Hypertension and Health outcomes~
~~~~~~
> cor.test(Health_Data$Post_Hypertension,Health_Data$Pre_Weight,method="pearson")

        Pearson's product-moment correlation

data:  Health_Data$Post_Hypertension and Health_Data$Pre_Weight
t = 11.004, df = 288, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4575571 0.6203680
sample estimates:
      cor
0.5440638

> cor.test(Health_Data$Post_Hypertension,Health_Data$Pre_Height,method="pearson")

        Pearson's product-moment correlation

data:  Health_Data$Post_Hypertension and Health_Data$Pre_Height
t = -5.6028, df = 288, p-value = 4.93e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.4137429 -0.2057532
sample estimates:
       cor
-0.3135032
```

Figure 26:Pearson correlation test for Post- Hypertension and health outcomes (Pre-Weight, Pre-Height)

Like this can find all correlation values ,

Table 9: Correlation values for Post-Hypertension and Health outcomes

Post-Hypertension Vs Health Outcome variable	**Correlation Value**	**There's an Association (Reject Ho)/ No Association (Do not reject Ho)**	**Correlation Type**
Pre-Weight	0.5440638	There's an Association (Reject Ho)	Moderately Positive Correlation
Pre-Height	-0.3135032	There's an Association (Reject Ho)	Low negative Correlation
Pre-BMI	0.8300645	There's an Association (Reject Ho)	High Positive Correlation
Post-Weight	0.5594242	There's an Association (Reject Ho)	Moderately Positive correlation.
Post-Height	-0.3135032	There's an Association (Reject Ho)	Low negative Correlation

Post-BMI	0.8414126	There's an Association (Reject Ho)	High Positive Correlation

Conclusion

Therefore, we have enough evidence to conclude that there is an association between Post-Hypotension and health outcomes (Pre-Weight, Post-Weight, Pre-Height, Post-Height, Pre-BMI, Post-BMI) at 5% level of significance. Correlation type is showing in the above table.

Considering Correlation between Post-Hypertension and health outcomes,

Post-Hypertension and Pre and Post Weights have Moderately Positive correlations.

Post-Hypertension and Pre and Post Height have Low Negative Correlation.

Post-Hypertension and Pre and Post BMI have High Positive Correlation.

Over all Considering It's seems to like between All risk factors (Obesity , Hypotension) and All Health outcomes (Pre and Post Weight, Height, BMI) have some kind of the correlation. It's range from -1 (Perfect Negative Correlation) to +1 (Perfect Positive Correlation) .

Correlation Matrix in Summary of Correlation values,

It provides a quick summary of how each variable relates to every other variable, showing the strength and direction (positive or negative) of these relationships. Correlation values range from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no linear correlation.

```
> round(correlation_matrix,2)
                              Health_Data.Pre_Hypertension Health_Data.Post_Hypertension
Health_Data.Pre_Hypertension                          1.00                          0.75
Health_Data.Post_Hypertension                         0.75                          1.00
Health_Data.Pre_Obesity                               0.72                          0.97
Health_Data.Post_Obesity                              0.75                          1.00
Health_Data.Pre_Weight                                0.19                          0.54
Health_Data.Pre_Height                               -0.60                         -0.31
Health_Data.Pre_BMI                                   0.76                          0.83
Health_Data.Post_Weight                               0.21                          0.56
Health_Data.Post_Height                              -0.60                         -0.31
Health_Data.Post_BMI                                  0.75                          0.84
                              Health_Data.Pre_Obesity Health_Data.Post_Obesity Health_Data.Pre_Weight
Health_Data.Pre_Hypertension                     0.72                     0.75                   0.19
Health_Data.Post_Hypertension                    0.97                     1.00                   0.54
Health_Data.Pre_Obesity                          1.00                     0.97                   0.52
Health_Data.Post_Obesity                         0.97                     1.00                   0.54
Health_Data.Pre_Weight                           0.52                     0.54                   1.00
Health_Data.Pre_Height                          -0.32                    -0.31                   0.48
Health_Data.Pre_BMI                              0.82                     0.83                   0.48
Health_Data.Post_Weight                          0.54                     0.56                   0.99
Health_Data.Post_Height                         -0.32                    -0.31                   0.48
Health_Data.Post_BMI                             0.82                     0.84                   0.53
```

```
                                  Health_Data.Pre_Height Health_Data.Pre_BMI Health_Data.Post_Weight
Health_Data.Pre_Hypertension                       -0.60                0.76                    0.21
Health_Data.Post_Hypertension                      -0.31                0.83                    0.56
Health_Data.Pre_Obesity                            -0.32                0.82                    0.54
Health_Data.Post_Obesity                           -0.31                0.83                    0.56
Health_Data.Pre_Weight                              0.48                0.48                    0.99
Health_Data.Pre_Height                              1.00               -0.54                    0.47
Health_Data.Pre_BMI                                -0.54                1.00                    0.49
Health_Data.Post_Weight                             0.47                0.49                    1.00
Health_Data.Post_Height                             1.00               -0.54                    0.47
Health_Data.Post_BMI                               -0.48                0.99                    0.55
                                  Health_Data.Post_Height Health_Data.Post_BMI
Health_Data.Pre_Hypertension                        -0.60                 0.75
Health_Data.Post_Hypertension                       -0.31                 0.84
Health_Data.Pre_Obesity                             -0.32                 0.82
Health_Data.Post_Obesity                            -0.31                 0.84
Health_Data.Pre_Weight                               0.48                 0.53
Health_Data.Pre_Height                               1.00                -0.48
Health_Data.Pre_BMI                                 -0.54                 0.99
Health_Data.Post_Weight                              0.47                 0.55
Health_Data.Post_Height                              1.00                -0.48
Health_Data.Post_BMI                                -0.48                 1.00
```

Figure 27: Correlation matrix for the relationship between variables

Correlation Heat Map for Interpret the Correlation value Strength

A correlation heatmap is a visual representation of a correlation matrix, typically using colors to indicate the strength and direction of relationships between variables. It helps identify patterns and associations in data, making it easier to understand how variables relate to each other. In here, Positive correlations are often represented by warm colors (Orange), while negative correlations use cool colors (White).

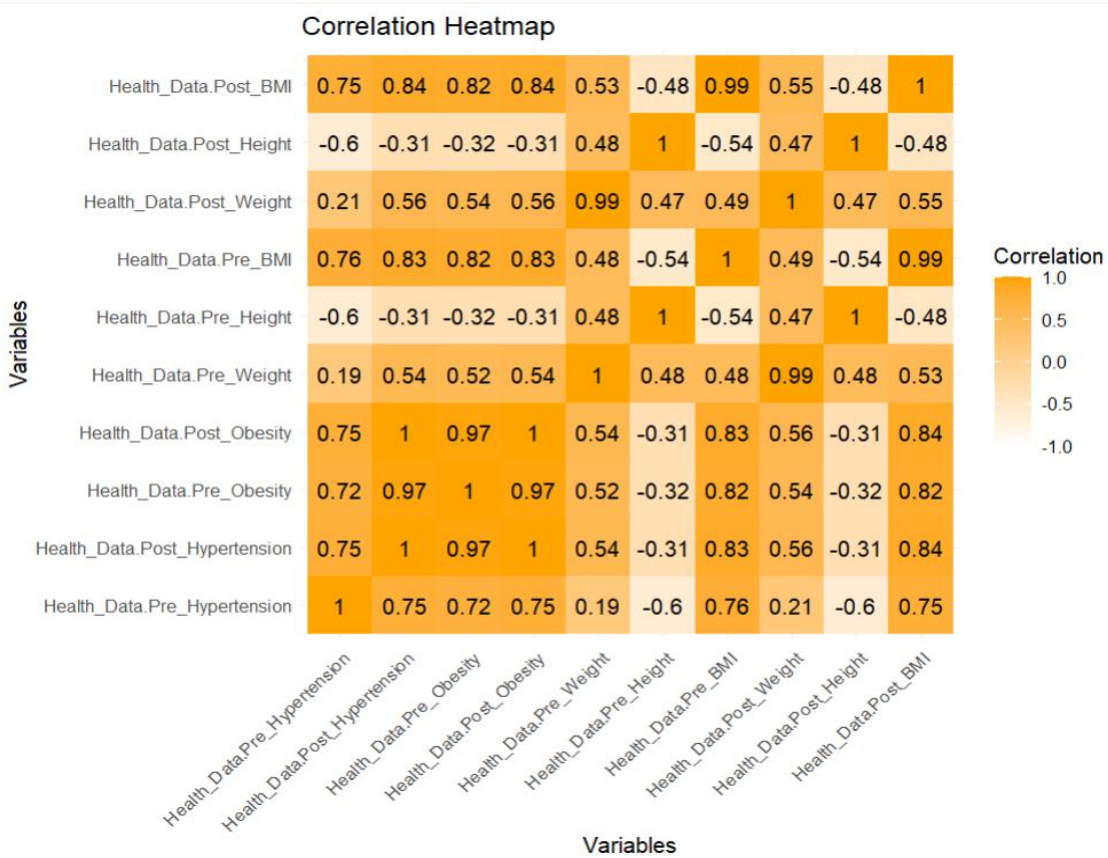


Figure 28: Correlation Heatmap for the relationship between variables

5.0 Discussion / Conclusion

In conclusion our analysis revealed notable, Changes in health risk factors (Obesity and Hypertension), before and after the implementation of the health program we observed that the program had a positive impact on reducing Hypertension and Obesity levels among participants.

Additionally, our study identified there are significant difference in all demographic groups (Ethnicity, Gender, Age) with all health outcomes except Ethnicity in (Pre Height, Post Height) health outcomes. It shows that health outcomes vary across these demographic categories.

Furthermore, we found correlations between health risk factors and health outcomes. It shows Obesity and Weight have moderately positive correlation and then weight is increasing obesity also increasing and weight is decreasing obesity also decreasing. It suggesting that reducing Obesity may lead to improved health outcomes. These findings underscore the program's effectiveness in addressing key health risk factors and promoting better health outcomes.

As we move forward, it is crucial to consider these results in the context of program planning and future interventions. Further research and targeted strategies may be needed to address health disparities among different demographic groups and to continue improving health outcomes for all program participants. Finally, It can be conclude that this results are successfully affect to the our goal reduced the prevalence of certain health risk factors (Obesity and hypertension) and improve overall health awareness conducting this health program.

Finally, our results affirm the successful achievement of our goal. Our objective, which was to reduce the prevalence of specific health risk factors (Obesity and Hypertension) and enhance overall health awareness through this health program, has been effectively realized.

6.0 Appendix: R codes

```
setwd("C:\\Users\\sanda\\OneDrive\\Desktop\\ST3010_CaseStudy")

#install.packages("readxl")

library(readxl)

Health_Data=read_xlsx("CaseStudy.xlsx")

getwd()

#attach(CaseStudy)

summary(CaseStudy)

#Explore the Data

head(Health_Data)

summary(Health_Data)

str(Health_Data)

colnames(Health_Data)

#Change Pre-Post-Hypertension column name to Pre-Hypertension

colnames(Health_Data)[colnames(Health_Data) == "Pre-Weight"] <- "Pre_Weight"

colnames(Health_Data)[colnames(Health_Data) == "Pre-Height"] <- "Pre_Height"

colnames(Health_Data)[colnames(Health_Data) == "Pre-BMI"] <- "Pre_BMI"

colnames(Health_Data)[colnames(Health_Data) == "Pre-Post-Hypertension"] <- "Pre_Hypertension"

colnames(Health_Data)[colnames(Health_Data) == "Pre-Obesity"] <- "Pre_Obesity"

colnames(Health_Data)[colnames(Health_Data) == "Post weight"] <- "Post_Weight"

colnames(Health_Data)[colnames(Health_Data) == "Post- height"] <- "Post_Height"

colnames(Health_Data)[colnames(Health_Data) == "Post BMI"] <- "Post_BMI"

colnames(Health_Data)[colnames(Health_Data) == "Post-Hypertension"] <- "Post_Hypertension"

colnames(Health_Data)[colnames(Health_Data) == "Post-Obesity"] <- "Post_Obesity"

colnames(Health_Data)

#Categorical Variable

Health_Data$Gender=as.factor(Health_Data$Gender)

Health_Data$Ethnicity=as.factor(Health_Data$Ethnicity)

Health_Data$Pre_Hypertension=as.factor(Health_Data$Pre_Hypertension)

Health_Data$Pre_Obesity=as.factor(Health_Data$Pre_Obesity)

Health_Data$Post_Hypertension=as.factor(Health_Data$Post_Hypertension)
```

Health_Data$Post_Obesity=as.factor(Health_Data$Post_Obesity)

summary(Health_Data)

#Check for missing values

sum(is.na(Health_Data))

#Check Duplicates

sum(duplicated(Health_Data))

#Unique Values

unique(Health_Data$Gender)

unique(Health_Data$Ethnicity)

unique(Health_Data$Pre_Hypertension)

unique(Health_Data$Pre_Obesity)

unique(Health_Data$Post_Hypertension)

unique(Health_Data$Post_Obesity)

#____Object 1____

#Assess Changes in Health Risk Factors (Obesity and Hypertension) Before and After the Program:

#~~~~~~~~~~~~~~~~~~~~~~~~~ Find Using Calculating Hypothesis Testing ~~~~~~

#@@@@@@@@@@@@@@@@@@@@@@@@@@@@ Find Parametric assumptions are valid or invalid for Variables

qqnorm(Health_Data$Pre_BMI)

qqline(Health_Data$Pre_BMI)

shapiro.test(Health_Data$Pre_BMI)

#p-value<0.05 we reject Ho -> Pre_BMI variable is not normally distributed

qqnorm(Health_Data$Post_BMI)

qqline(Health_Data$Post_BMI)

shapiro.test(Health_Data$Post_BMI)

#p-value<0.05 we reject Ho -> Post_BMI variable is not normally distributed

qqnorm(Health_Data$Pre_Height)

qqline(Health_Data$Pre_Height)

shapiro.test(Health_Data$Pre_Height)

#p-value<0.05 we reject Ho -> Pre_Height variable is not normally distributed

```
qqnorm(Health_Data$Post_Height)

qqline(Health_Data$Post_Height)

shapiro.test(Health_Data$Post_Height)

#p-value<0.05 we reject Ho -> Post_Height variable is not normally distributed

qqnorm(Health_Data$Pre_Weight)

qqline(Health_Data$Pre_Weight)

shapiro.test(Health_Data$Pre_Weight)

#p-value<0.05 we reject Ho -> Pre_Weight variable is not normally distributed

qqnorm(Health_Data$Post_Weight)

qqline(Health_Data$Post_Weight)

shapiro.test(Health_Data$Post_Weight)

#p-value<0.05 we reject Ho -> Post_Weight variable is not normally distributed

qqnorm(Health_Data$Age)

qqline(Health_Data$Age)

shapiro.test(Health_Data$Age)

#p-value<0.05 we reject Ho -> Post_Weight variable is not normally distributed

##### Then normality assumptions are violated in here. So, we use Non Parametric test

#@@@@@@@@@@@@@@@@@@@@@@@@@@@ Health Risk Changes Calculating By Pre, Post Weight

#============== Wilcox's Signed Rank Test for Matched Pairs for Pre and Post Weight

#...........Ho:M(Pre-Post)=0  H1:M(Pre-Post) =/ 0

wilcox.test(Health_Data$Pre_Weight,Health_Data$Post_Weight,mu=0, alternative ="two.sided",paired =
T,exact=F,conf.int = T,conf.level = 0.95)

#Reject Ho

#............Ho:M(Pre-Post)=<0  H1:M(Pre-Post) > 0

wilcox.test(Health_Data$Pre_Weight,Health_Data$Post_Weight,mu=0, alternative ="greater",paired =
T,exact=F,conf.int = T,conf.level = 0.95)

#Reject Ho

##### Therefore Pre Weight is higher than Post we can determine Obesity Risk is reduced then can assume
Hypertension is also reduced.

#~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ Find Using Pre and Post Change Count Considering in Health Risk Factors

#install.packages("dplyr")

library(dplyr)
```

```
# Assuming you have a data frame named 'health_data' with 'Pre_Obesity' and 'Post_Obesity' columns

# Create a new variable 'Change_in_Obesity' with custom labels

# @@@@@@@@@@@@@@@@@@@@@@@ For Obesity Find Count Number of Changes

Change_Obesity = ifelse(Health_Data$Pre_Obesity == "No" & Health_Data$Post_Obesity == "Yes", "Increased",

                        ifelse(Health_Data$Pre_Obesity == "Yes" & Health_Data$Post_Obesity == "No", "Decreased",
"No Change"))

Change_Obesity

Count_Increased_Obe = sum(Change_Obesity == "Increased")

Count_Increased_Obe

Count_Decreased_Obe = sum(Change_Obesity == "Decreased")

Count_Decreased_Obe

Count_No_Change_Obe= sum(Change_Obesity == "No Change")

Count_No_Change_Obe

# Create the bar chart for Change in Obesity

Count_Data_Obe=c(0,4,286)

Label_Count_Data_Obe=c("Increased","Decreased","No Change")

Lable_Count_Data_Value_Obe=c("0","4","286")

bar_heights=barplot(Count_Data_Obe,

    names.arg = Label_Count_Data_Obe,   # Add labels to the bars

    col = "skyblue",      # Customize the bar color

    main = "Bar Chart for Obesity Changes",  # Add a title

    xlab = "Categories",  # Label for the x-axis

    ylab = "Values"       # Label for the y-axis

)

#@@@@@@@@@@@@@@@@@@@@@ For Hypertension Find Number of Changes

Change_Hypertension = ifelse(Health_Data$Pre_Hypertension == "No" & Health_Data$Post_Hypertension == "Yes",
"Increased",

                ifelse(Health_Data$Pre_Hypertension == "Yes" & Health_Data$Post_Hypertension == "No",
"Decreased", "No Change"))

Change_Hypertension

Count_Increased_Hyp = sum(Change_Hypertension == "Increased")

Count_Increased_Hyp

Count_Decreased_Hyp = sum(Change_Hypertension == "Decreased")
```

```
Count_Decreased_Hyp

Count_No_Change_Hyp = sum(Change_Hypertension == "No Change")

Count_No_Change_Hyp

# Create the bar chart for Change in Hypertension

Count_Data_Obe=c(0,41,249)

Label_Count_Data_Obe=c("Increased","Decreased","No Change")

Lable_Count_Data_Value_Obe=c("0","41","249")

bar_heights=barplot(Count_Data_Obe,

        names.arg = Label_Count_Data_Obe,   # Add labels to the bars

        col = "purple",      # Customize the bar color

        main = "Bar Chart for Hypertension Changes",  # Add a title

        xlab = "Categories",  # Label for the x-axis

        ylab = "Values"       # Label for the y-axis

)

# Display the updated data frame

head(health_data)

#___________________________________________________________ Object 2
_________________________________________________________________

#Determine if there are any significant differences between various demographic groups in terms of health outcomes:

unique(Health_Data$Ethnicity)

#*******************************Perform Kruskal Wallis Test for Demography groups in health outcomes:

#@@@@@@@@@@@@@@@@@@@@@@@ Create a data frame for Understanding testing significance

African_Data <- subset(Health_Data,Health_Data$Ethnicity == "African")

Asian_Data <- subset(Health_Data,Health_Data$Ethnicity == "Asian")

Caucasian_Data <- subset(Health_Data,Health_Data$Ethnicity == "Caucasian")

Hispanic_Data <- subset(Health_Data,Health_Data$Ethnicity == "Hispanic")

# Data frame Pre_Height data for Ethnicity

Pre_Height_African <- African_Data$Pre_Height

Pre_Height_African

Pre_Height_Asian <- Asian_Data$Pre_Height

Pre_Height_Asian
```

```
Pre_Height_Caucasian <- Caucasian_Data$Pre_Height

Pre_Height_Caucasian

Pre_Height_Hispanic <- Hispanic_Data$Pre_Height

Pre_Height_Hispanic

max_length=max(c(length(Pre_Height_African),length(Pre_Height_Asian),length(Pre_Height_Caucasian),length(Pre_
Height_Hispanic)))

max_length

Pre_Height_DataFrame=data.frame(African=c(Pre_Height_African,rep(NA,max_length-
length(Pre_Height_African))),Asian=c(Pre_Height_Asian,rep(NA,max_length-
length(Pre_Height_Asian))),Caucasian=c(Pre_Height_Caucasian,rep(NA,max_length-
length(Pre_Height_Caucasian))),Hispanic=c(Pre_Height_Hispanic,rep(NA,max_length-length(Pre_Height_Hispanic))))

Pre_Height_DataFrame

# Data frame Pre_Weight data for Ethnicity levels

Pre_Weight_African <- African_Data$Pre_Weight

Pre_Weight_African

Pre_Weight_Asian <- Asian_Data$Pre_Weight

Pre_Weight_Asian

Pre_Weight_Caucasian <- Caucasian_Data$Pre_Weight

Pre_Weight_Caucasian

Pre_Weight_Hispanic <- Hispanic_Data$Pre_Weight

Pre_Weight_Hispanic

max_length1=max(c(length(Pre_Weight_African),length(Pre_Weight_Asian),length(Pre_Weight_Caucasian),length(Pr
e_Weight_Hispanic)))

max_length1

Pre_Weight_DataFrame=data.frame(African=c(Pre_Weight_African,rep(NA,max_length1-
length(Pre_Weight_African))),Asian=c(Pre_Weight_Asian,rep(NA,max_length1-
length(Pre_Weight_Asian))),Caucasian=c(Pre_Weight_Caucasian,rep(NA,max_length1-
length(Pre_Weight_Caucasian))),Hispanic=c(Pre_Height_Hispanic,rep(NA,max_length1-
length(Pre_Weight_Hispanic))))

Pre_Weight_DataFrame

# Data frame Pre_BMI data for Ethnicity

Pre_BMI_African <- African_Data$Pre_BMI

Pre_BMI_African

Pre_BMI_Asian <- Asian_Data$Pre_BMI

Pre_BMI_Asian
```

```
Pre_BMI_Caucasian <- Caucasian_Data$Pre_BMI

Pre_BMI_Caucasian

Pre_BMI_Hispanic <- Hispanic_Data$Pre_BMI

Pre_BMI_Hispanic

max_length2=max(c(length(Pre_BMI_African),length(Pre_BMI_Asian),length(Pre_BMI_Caucasian),length(Pre_BMI_H
ispanic)))

max_length2

Pre_BMI_DataFrame=data.frame(African=c(Pre_BMI_African,rep(NA,max_length2-
length(Pre_BMI_African))),Asian=c(Pre_BMI_Asian,rep(NA,max_length2-
length(Pre_BMI_Asian))),Caucasian=c(Pre_BMI_Caucasian,rep(NA,max_length2-
length(Pre_BMI_Caucasian))),Hispanic=c(Pre_BMI_Hispanic,rep(NA,max_length2-length(Pre_BMI_Hispanic))))

Pre_BMI_DataFrame

# Data frame Pre_BMI data for Gender

Female_Data <- subset(Health_Data, Health_Data$Gender == "Female")

Male_Data <- subset(Health_Data, Health_Data$Gender == "Male")

Other_Data <- subset(Health_Data, Health_Data$Gender == "Other")

Pre_BMI_Female <- Female_Data$Pre_BMI

Pre_BMI_Female

Pre_BMI_Male <- Male_Data$Pre_BMI

Pre_BMI_Male

Pre_BMI_Other <- Other_Data$Pre_BMI

Pre_BMI_Other

max_length_a=max(c(length(Pre_BMI_Female),length(Pre_BMI_Male),length(Pre_BMI_Other)))

max_length_a

Pre_BMI_DataFrame_a=data.frame(Female=c(Pre_BMI_Female,rep(NA,max_length_a-
length(Pre_BMI_Female))),Male=c(Pre_BMI_Male,rep(NA,max_length_a-
length(Pre_BMI_Male))),Other=c(Pre_BMI_Other,rep(NA,max_length_a-length(Pre_BMI_Other))))

Pre_BMI_DataFrame_a

# Data frame Pre_Weight data for Gender

Pre_Weight_Female <- Female_Data$Pre_Weight

Pre_Weight_Female

Pre_Weight_Male <- Male_Data$Pre_Weight

Pre_Weight_Male

Pre_Weight_Other <- Other_Data$Pre_Weight
```

```
Pre_Weight_Other

max_length_b=max(c(length(Pre_Weight_Female),length(Pre_Weight_Male),length(Pre_Weight_Other)))

max_length_b

Pre_Weight_DataFrame_b=data.frame(Female=c(Pre_Weight_Female,rep(NA,max_length_b-
length(Pre_Weight_Female))),Male=c(Pre_Weight_Male,rep(NA,max_length_b-
length(Pre_Weight_Male))),Other=c(Pre_Weight_Other,rep(NA,max_length_b-length(Pre_Weight_Other))))

Pre_Weight_DataFrame_b

# Data frame Pre_Height data for Gender

Pre_Height_Female <- Female_Data$Pre_Height

Pre_Height_Female

Pre_Height_Male <- Male_Data$Pre_Height

Pre_Height_Male

Pre_Height_Other <- Other_Data$Pre_Height

Pre_Height_Other

max_length_c=max(c(length(Pre_Height_Female),length(Pre_Height_Male),length(Pre_Height_Other)))

max_length_c

Pre_Height_DataFrame_c=data.frame(Female=c(Pre_Height_Female,rep(NA,max_length_c-
length(Pre_Height_Female))),Male=c(Pre_Height_Male,rep(NA,max_length_c-
length(Pre_Height_Male))),Other=c(Pre_Height_Other,rep(NA,max_length_c-length(Pre_Height_Other))))

Pre_Height_DataFrame_c

#~~~~~~~~~~~~~~~~~~~~~~~~~~ Testing Significance Using Non Parametric Tests

#@@@@@@@@@@@@@@@@@@@@@@ Using Kruskal Wallis Test Ethnicity Vs Health Outcomes

# Testing significance difference between Ethnicity groups in Pre_BMI

Kruskal_Ethnicity_Pre_BMI=kruskal.test(Health_Data$Pre_BMI~Health_Data$Ethnicity)

Kruskal_Ethnicity_Pre_BMI

# Testing significance difference between Ethnicity groups in Post_BMI

Kruskal_Ethnicity_Post_BMI=kruskal.test(Health_Data$Post_BMI~Health_Data$Ethnicity)

Kruskal_Ethnicity_Post_BMI

# Testing significance difference between Ethnicity groups in Pre_Weight

Kruskal_Ethnicity_Pre_Weight=kruskal.test(Health_Data$Pre_Weight~Health_Data$Ethnicity)

Kruskal_Ethnicity_Pre_Weight
```

```
# Testing significance difference between Ethnicity groups in Post_Weight
Kruskal_Ethnicity_Post_Weight=kruskal.test(Health_Data$Post_Weight~Health_Data$Ethnicity)
Kruskal_Ethnicity_Post_Weight

# Testing significance difference between Ethnicity groups in Pre_Height
Kruskal_Ethnicity_Pre_Height=kruskal.test(Health_Data$Pre_Height~Health_Data$Ethnicity)
Kruskal_Ethnicity_Pre_Height

# Testing significance difference between Ethnicity groups in Post_Height
Kruskal_Ethnicity_Post_Height=kruskal.test(Health_Data$Post_Height~Health_Data$Ethnicity)
Kruskal_Ethnicity_Post_Height

#@@@@@@@@@@@@@@@@@@@@@@@ Using Kruskal Wallis Test Gender Vs Health Outcomes

# Testing significance difference between Gender groups in Pre_BMI
Kruskal_Gender_Pre_BMI=kruskal.test(Health_Data$Pre_BMI~Health_Data$Gender)
Kruskal_Gender_Pre_BMI

# Testing significance difference between Gender groups in Post_BMI
Kruskal_Gender_Post_BMI=kruskal.test(Health_Data$Post_BMI~Health_Data$Gender)
Kruskal_Gender_Post_BMI

# Testing significance difference between Gender group in Pre_Weight
Kruskal_Gender_Pre_Weight=kruskal.test(Health_Data$Pre_Weight~Health_Data$Gender)
Kruskal_Gender_Pre_Weight

# Testing significance difference between Gender group in Post_Weight
Kruskal_Gender_Post_Weight=kruskal.test(Health_Data$Post_Weight~Health_Data$Gender)
Kruskal_Gender_Post_Weight

# Testing significance difference between Gender groups in Pre_Height
Kruskal_Gender_Pre_Height=kruskal.test(Health_Data$Pre_Height~Health_Data$Gender)
```

```
Kruskal_Gender_Pre_Height

# Testing significance difference between Gender groups in Post_Height
Kruskal_Gender_Post_Height=kruskal.test(Health_Data$Post_Height~Health_Data$Gender)
Kruskal_Gender_Post_Height

#@@@@@@@@@@@@@@@@@@@@@@@@@ Using Kruskal Wallis Test Age Vs Health Outcomes

unique(Health_Data$Age)

# Testing significance difference between Age in Pre_Weight
Kruskal_Age_Pre_Weight=kruskal.test(Health_Data$Pre_Weight~Health_Data$Age)
Kruskal_Age_Pre_Weight

# Testing significance difference between Age in Post_Weight
Kruskal_Age_Post_Weight=kruskal.test(Health_Data$Post_Weight~Health_Data$Age)
Kruskal_Age_Post_Weight

# Testing significance difference between Age in Pre_BMI
Kruskal_Age_Pre_BMI=kruskal.test(Health_Data$Pre_BMI~Health_Data$Age)
Kruskal_Age_Pre_BMI

# Testing significance difference between Age in Post_BMI
Kruskal_Age_Post_BMI=kruskal.test(Health_Data$Post_BMI~Health_Data$Age)
Kruskal_Age_Post_BMI

# Testing significance difference between Age in Pre_Height
Kruskal_Age_Pre_Height=kruskal.test(Health_Data$Pre_Height~Health_Data$Age)
Kruskal_Age_Pre_Height

# Testing significance difference between Age in Post_Height
Kruskal_Age_Post_Height=kruskal.test(Health_Data$Post_Height~Health_Data$Age)
```

Kruskal_Age_Post_Height

#____ Object 3____

#@@@@@@@@@@@@@@@@@@@ Assign Categorical variables to Dummy variable @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@

Health_Data$Pre_Obesity=ifelse(Health_Data$Pre_Obesity == "Yes", 1, 0)

Health_Data$Post_Obesity <- ifelse(Health_Data$Post_Obesity == "Yes", 1, 0)

Health_Data$Pre_Hypertension <- ifelse(Health_Data$Pre_Hypertension == "Yes", 1, 0)

Health_Data$Post_Hypertension <- ifelse(Health_Data$Post_Hypertension == "Yes", 1, 0)

#~~~~~~~~~~~~~~~~~~~~~~~~~~ Pearson Correlation Test for Pre_Obesity and Health outcomes~~~~~~~~~~~~~~~~~~~~~~~~

cor.test(Health_Data$Pre_Obesity,Health_Data$Pre_Weight,method="pearson")

cor.test(Health_Data$Pre_Obesity,Health_Data$Pre_Height,method="pearson")

cor.test(Health_Data$Pre_Obesity,Health_Data$Pre_BMI,method="pearson")

cor.test(Health_Data$Pre_Obesity,Health_Data$Post_Weight,method="pearson")

cor.test(Health_Data$Pre_Obesity,Health_Data$Post_Height,method="pearson")

cor.test(Health_Data$Pre_Obesity,Health_Data$Post_BMI,method="pearson")

#~~~~~~~~~~~~~~~~~~~~~~~~~~ Pearson Correlation Test for Post_Obesity and Health outcomes~~~~~~~~~~~~~~~~~~~~~~~~

cor.test(Health_Data$Post_Obesity,Health_Data$Pre_Weight,method="pearson")

cor.test(Health_Data$Post_Obesity,Health_Data$Pre_Height,method="pearson")

cor.test(Health_Data$Post_Obesity,Health_Data$Pre_BMI,method="pearson")

cor.test(Health_Data$Post_Obesity,Health_Data$Post_Weight,method="pearson")

cor.test(Health_Data$Post_Obesity,Health_Data$Post_Height,method="pearson")

cor.test(Health_Data$Post_Obesity,Health_Data$Post_BMI,method="pearson")

#~~~~~~~~~~~~~~~~~~~~~~~~~~ Pearson Correlation Test for Pre_Hypertension and Health outcomes~~~~~~~~~~~~~~~~~~~~~~~~

cor.test(Health_Data$Pre_Hypertension,Health_Data$Pre_Weight,method="pearson")

```
cor.test(Health_Data$Pre_Hypertension,Health_Data$Pre_Height,method="pearson")

cor.test(Health_Data$Pre_Hypertension,Health_Data$Pre_BMI,method="pearson")

cor.test(Health_Data$Pre_Hypertension,Health_Data$Post_Weight,method="pearson")

cor.test(Health_Data$Pre_Hypertension,Health_Data$Post_Height,method="pearson")

cor.test(Health_Data$Pre_Hypertension,Health_Data$Post_BMI,method="pearson")



#~~~~~~~~~~~~~~~~~~~~~~~~~~ Pearson Correlation Test for Post_Hypertension and Health
outcomes~~~~~~~~~~~~~~~~~~~~~~~~

cor.test(Health_Data$Post_Hypertension,Health_Data$Pre_Weight,method="pearson")

cor.test(Health_Data$Post_Hypertension,Health_Data$Pre_Height,method="pearson")

cor.test(Health_Data$Post_Hypertension,Health_Data$Pre_BMI,method="pearson")

cor.test(Health_Data$Post_Hypertension,Health_Data$Post_Weight,method="pearson")

cor.test(Health_Data$Post_Hypertension,Health_Data$Post_Height,method="pearson")

cor.test(Health_Data$Post_Hypertension,Health_Data$Post_BMI,method="pearson")


#~~~~~~~~~~~~~~~~~~~~~~~~~~~ Create a correlation matrix for data ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Person_DataFrame=data.frame(Health_Data$Pre_Hypertension,Health_Data$Post_Hypertension,Health_Data$Pre_
Obesity,Health_Data$Post_Obesity,

                    Health_Data$Pre_Weight,Health_Data$Pre_Height,Health_Data$Pre_BMI,

                    Health_Data$Post_Weight,Health_Data$Post_Height,Health_Data$Post_BMI)

Person_DataFrame

correlation_matrix=cor(Person_DataFrame)

correlation_matrix

round(correlation_matrix,2)


#~~~~~~~~~~~~~~~~~~~~~~~~~~ Create Heat map to the data ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

library(ggplot2)

library(reshape2)


heatmap_data <- melt(correlation_matrix)
```

```
# Create a heatmap plot with correlation values

ggplot(data = heatmap_data, aes(x = Var1, y = Var2, fill = value, label = round(value, 2))) +
  geom_tile() +
  geom_text() +
  scale_fill_gradient(low = "white", high = "orange", limits = c(-1, 1)) +
  theme_minimal() +
  labs(
    title = "Correlation Heatmap",
    x = "Variables",
    y = "Variables",
    fill = "Correlation"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Subject – ST 3010 (Introduction to Health Statistics)

Name – K.K.D. Sandali Navoda Gunathilaka

Registrations Number – 2020s18059

Index Number – s15355

Case Study Name – Analyzing Health Data for a Community Health Program

End