EXPLORING MULTIFACETED DIMENSIONS OF OBESITY

MINI PROJECT

SUBMITTED BY

ANNAMALAI CT (23MST001)

ARISTO T (23MST003)

JANARTHAN R (23MST010)

RAMADEVI N (23MST022)

RANJITH K (23MST023)

SREE SARATHY M (23MST031)

ACKNOWLEDGEMENT

We deem it a pleasure to acknowledge our sense of gratitude to our project guide Dr. E. Sakthivel under whom we have carried out the project work. His incisive and objective guidance and timely advice encouraged us with constant flow of energy to continue the work.

We wish to reciprocate in full measure the kindness shown by Dr. D. Senthilkumar (H.O.D, Department of Statistics) who inspired us with his valuable suggestions in successfully completing the project work.

Finally, we must say that no height is ever achieved without some sacrifices made at some end and it is here where we owe our special debt to our parents and our friends for showing their generous love and care throughout the entire period of time.

CHAPTER I
INTRODUCTION

Obesity is a medical condition defined as abnormal or excessive accumulations of body fat that can put a person's health at risk.

Global estimates suggest that almost 2.3 billion children and adults are living with overweight and obesity.

If current trends continue, 2.7 billion adults could be living with overweight or obesity by 2025.

The prevalence of obesity across the world continues to rise, and this is now recognised as one of the most important public health problems facing the world today.

Obesity is a major public health challenge that affects almost every country in the world. Globally, obesity rates have been increasing steadily over the last three decades. In almost every region, there are now more people living with obesity than underweight, and if current numbers continue to rise, as many as 1 billion adults, or 12% of the world population, will be living with obesity by 2025.

The World Obesity Atlas 2022, published by the World Obesity Federation, predicts that one billion people globally, including 1 in 5 women and 1 in 7 men, will be living with obesity by 2030.

Recent studies have reported that globally, more than 1.9 billion adults are overweight and 650 million are obese. Approximately 2.8 million deaths are reported as a result of being overweight or obese. Due to the consumption of energy dense food (i.e. unhealthy food habits), sedentary life style, lack of health care services and financial support, the developing countries are facing high risk of obesity and their adverse consequences (i.e. diabetes, ischemic heart disease, etc).

In India, more than 135 million individuals were affected by obesity. The prevalence of obesity in India varies due to age, gender, geographical environment, socio-economic status, etc. According to ICMRINDIAB study 2015, prevalence rate of obesity and central obesity are varies from 11.8% to 31.3% and 16.9%-36.3% respectively. In India, abdominal obesity is one of the major risk factors for cardiovascular disease (CVDs). Various studies have shown that the prevalence of obesity among women were significantly higher as compared to men. Obesity is one of the main medical and financial burdens for the government. This problem of obesity can be preventable by spreading public awareness about obesity and its health consequences. Governmental agencies should promote the benefits of healthy life style, food habits and physical activity.

CHAPTER II
AIM & OBJECTIVE

AIM:

In this mini statistical project, we aim to analyze and interpret a dataset related to obesity, focusing on the influence of various factors on individuals such as Height, Weight, BMI, Smoking, Alcohol, H2O, and Physical Activity etc., are employed using a range of statistical tools. This investigation seeks to elucidate the multifaceted nature of obesity, exploring its prevalence, contributing factors, and associated health outcomes. Through rigorous statistical analysis, the aim is to derive meaningful insights that contribute to a deeper understanding of obesity's complexities and inform potential interventions for its prevention and management.

OBJECTIVE:

The primary objective of this project is to investigate the relationships between obesity and key factors such as age, gender, BMI, Smoking, Alcohol, H2O, and Physical Activity status.

By employing statistical methods, we seek to uncover patterns, trends, and potential risk factors contributing to the prevalence of obesity in the studied population.

Also we determine the number of abstinence among the obesity people by employing with different kind of treatments.

Moreover, we interpret the hazard and survival function of different factors.

CHAPTER III
REVIEW OF LITERATURE

Smith, Elizabeth, in her article Healthy Lifestyles sum ups not merely the enormous medical circumstances that can have an result on overweight children (cataract. Liver disease, diabetes, high cholesterol & blood pressure bone and joint problems, sleep apnea, etc.), but lots of the renowned sources. Genetically, there is enough numerical corroboration to illustrate that overweight and obese adults are further expected to have obese children, and that explicit racial/ethnic settings are more have an outcome on than others. In the fetal neighboring, elevated folic acid eating can demonstrate the way to health dilemmas in a fetus and infant, and denied nutrition by the mother as well destructively have an result on the fetus. In accumulation, the social setting in which the infant is raised, in exacting in expressions of work out and nutrition, will produce patterns go behind by the child for mainly in all probability their whole life.

As per to Brownwell (1982), if "cure" from obesity is described as decrease to ultimate weight and maintenance of that weight for five years, an individual is additional expected to recuperate from the majority types of cancer than from obesity. According to Krieshok & Karpowitz (1988), Even though there isn't a cure, there are things that can be done to avert several from being obese.

Egger (2010) in his article 5 myths about hunger in America concentrates on how hunger cycle in children affects learning ability which tend to increases, reduces employment rates, school dropout rates thus diminishing national economic security while promoting a life of crime. He inter connects hunger and obesity again by centering on how youthful adults are too obese to work with the military, as a result decreasing the numeral of military qualified citizens and consequently decreasing national security. Obese children are in addition additional likely to be turn into obese as adults, thus increasing their enduring risk of grave health problems for example heart disease and stroke.

CHAPTER IV
RESEARCH METHDOLOGY

1) SOURCE AND STUDY AREA

The source of data is secondary and the study area is from Kaggle data source.

2) DATA COLLECTION

Data consists of 498 samples which comprises the variables of Gender, Age, Height, Weight, BMI, Family History with Obesity, Intake of Caloric Food, Smoke, Alcohol, H2O, Physical Activities, Transportation and BMI Categorisation.

Out of 498 samples, 61 samples(only obesity) were separated for further study which comprises of treatments they had pursued and their New BMI and weight after the treatment and the respective factors which affect obesity and finally the outcome of the obesity people which includes Absistent or Non Absistent from the problem.

REVIEW OF DATASET:

GENDER:

It consists of Male and Female.

BMI:

It is calculated by using the formula, Weight (in kilograms)/ Height^2 (in meters).

AGE, HEIGHT AND WEIGHT:

It is recorded as per the Dataset.

FAMILY HISTORY WITH OBESITY:

It consists of two categories,

- Yes – Family History with Obesity
- No – Family History with No Obesity

INTAKE OF CALORIC FOOD:

It consists of two categories,

- Yes – They Intake a Caloric Food
- No – They doesn't Intake a Caloric Food

SMOKE:

It consists of two categories,

- Yes – They have a habit of Smoking
- No – They doesn't have a habit of Smoking

ALCOHOL:

It consists of two categories,

- Yes – They have a habit of consuming Alcohol
- No – They doesn't have a habit of consuming Alcohol

H2O:

The water level the respondents had consumed per day consists of three categories,

- Less than 1 litre
- Between 1-2 litre
- More than 2 litre

PHYSICAL ACTIVITIES:

It consists of four categories

- Don't have an habit of doing physical activity
- 1 or 2 days
- 2 or 4 days
- 4 or 5 days

TRANSPORTATION:

It consists of five categories which the respondents used as their transportation,

- Walking
- Cycle
- Bike
- Automobile
- Public Transportation

BMI CATEGORIZATION:

It consists of four categories, and it is categorized as per the following case

- Under Weight : <18.5

- Normal Weight : 18.5 – 24.9
- Over Weight : 25 – 29.9
- Obesity : >=30

TREATMENT:

It consists of five categories

- Calorie Diet
- Self Physical Activities
- Medicine and Supplements
- Weight Loss Program
- Surgery

MONTHS:

It represents the number of months the obesity people had intake the treatments.

OUTCOME:

It consists of two categories,

- Abstinent – They had recovered from the problem of Obesity.
- Non Abstinent – They hadn't recovered from the problem of Obesity.

3) RESEARCH TOOLS :

The collected data were processed with SPSS software and the following statistical tools were used in tune with the objective of the study.

❖ BAR GRAPH:

"As an essential visual tool in our research, bar graphs serve as a dynamic means of presenting and comparing categorical data. In this study, we aim to convey a clear representation of the distribution and relationships within our dataset, specifically focusing on vital factors . Bar graphs are particularly effective for illustrating the differences in frequencies or quantities among distinct categories. By employing bar graphs, our research endeavors to provide a visual narrative that enhances the accessibility and interpretability of our findings. This graphical representation aids in conveying patterns, trends, and disparities, contributing to a comprehensive understanding of the key aspects explored in our study. The

integration of bar graphs not only enhances the visual appeal of our results but also serves as a valuable tool for communicating complex categorical information in a straightforward manner, facilitating a more accessible interpretation of our research outcomes."

- **CORRELATION:**

"As a fundamental component of our research methodology, correlation analysis plays a pivotal role in unveiling the relationships between variables under investigation. In our study, we seek to discern the degree and direction of association between [variable A] and [variable B]. Correlation analysis allows us to quantitatively measure the strength and nature of these associations, shedding light on the interdependence or independence of the variables in our dataset. By employing correlation analysis, we aim to uncover patterns and trends, providing a nuanced understanding of how changes in one variable may correspond to changes in another. This statistical approach not only enhances the depth of our analysis but also serves as a valuable tool for identifying potential explanatory factors and guiding the exploration of intricate relationships within our research domain. The utilization of correlation analysis contributes to the robustness of our findings and facilitates a more comprehensive interpretation of the dynamics at play in our study."

- **CHI SQUARE:**

"In our research project, the chi-square test emerges as a pivotal statistical tool for examining the association between categorical variables. As we delve into the intricate relationships within our dataset, understanding the patterns and dependencies among variables becomes paramount. The chi-square test is particularly suited for analyzing categorical data, allowing us to assess whether observed frequencies significantly differ from expected frequencies. This method is integral to our study as it enables us to explore and quantify associations, providing statistical evidence for the relationships we aim to investigate. By leveraging the chi-square test, we seek to unravel the underlying connections between smoke, alcohol, intake of caloric food, family history with overweight, physical activity, Transportation Hydrated level with the variable obesity contributing to a comprehensive understanding of the dynamics at play within our research domain. The application of this statistical technique enhances the rigor of our analyses, offering valuable insights into the categorical dependencies inherent in our dataset."

❖ ONE WAY ANOVA:

“In this project, we employ Analysis of Variance (ANOVA), a statistical technique widely used in research, to investigate the potential variations in New weight after treatment among different groups. ANOVA allows us to assess whether there are significant differences in the means of multiple groups, providing valuable insights into the factors that may influence [New weight as dependent variable]. By conducting ANOVA, we aim to identify and understand the sources of variability within our data, contributing to a comprehensive analysis of the relationships and patterns that may exist among the variables under consideration. This statistical method is particularly pertinent to our study as it enables a robust examination of group differences, laying the foundation for more nuanced interpretations of our research findings.”

❖ LOGISTIC REGRESSION:

“In this project, logistic regression serves as a pivotal statistical method to analyze the relationship between one or more independent variables and a binary outcome. Our study delves into understanding the factors influencing the likelihood of outcome, and logistic regression proves instrumental in this analysis. By employing logistic regression, we aim to quantify the impact of various predictor variables on the probability of observing the outcome. This method allows us to model complex relationships within our data, providing a robust framework to explore and interpret the factors contributing to the binary nature of our dependent variable(Outcome). The application of logistic regression enhances the depth and accuracy of our analyses, offering valuable insights into the determinants of overall outcome.

❖ PAIRED T TEST:

“As an integral part of our research methodology, we employ paired tests to examine the significance of changes or differences within paired observations. Paired tests are particularly valuable when investigating the impact of an intervention, treatment, or any paired experimental design. In our study, we are interested in exploring the before-and-after effects, or paired relationships, between Weight before treatment and New weight after treatment. The utilization of paired tests allows us to determine whether there is a statistically significant difference between the paired observations, enhancing our ability to draw meaningful

conclusions about the effectiveness or impact of the studied variables. This statistical approach not only provides a nuanced understanding of within-group dynamics but also serves as a robust tool for evaluating the specific changes of interest in our research context."

- **KAPLAN MEIER:**

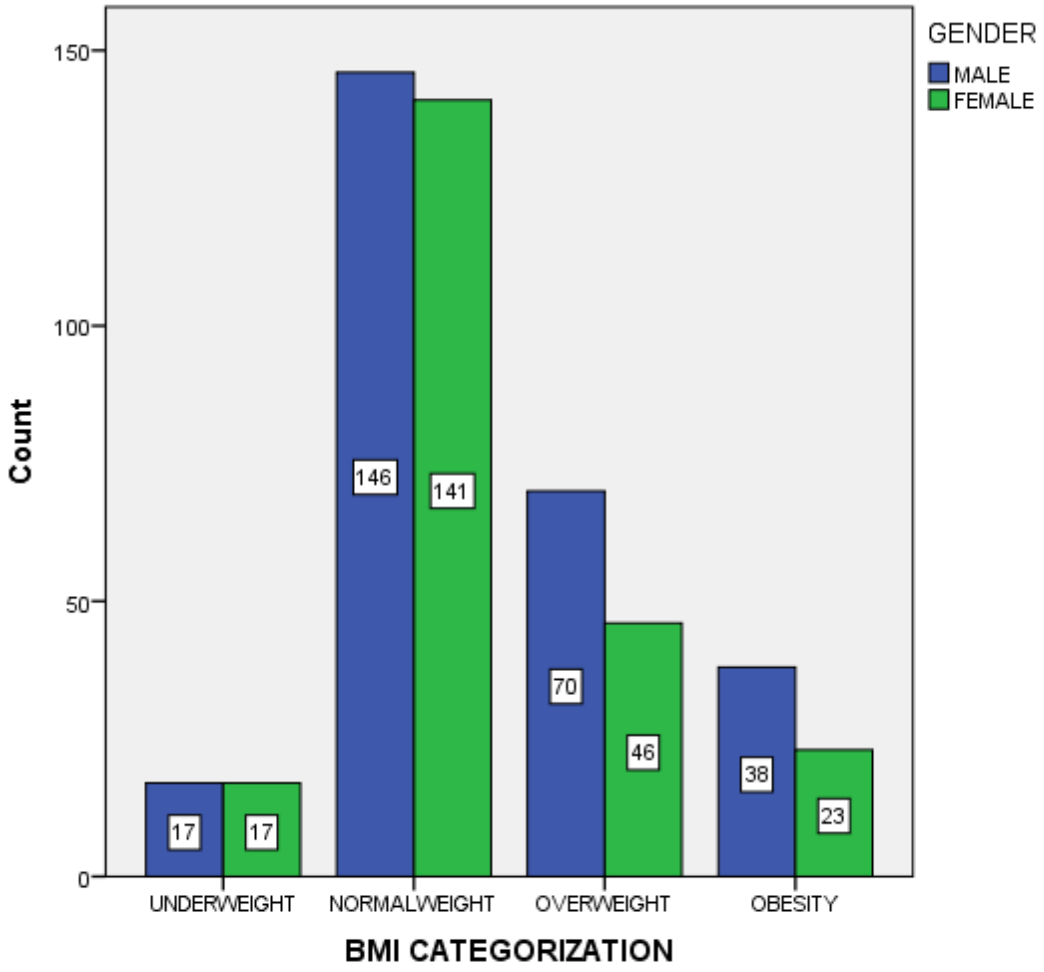
"As a crucial statistical method Survival employed in our research, the Kaplan-Meier estimator provides a powerful tool for analyzing time-to-event data, such as survival or failure times. In our study, we aim to explore and quantify the probability of an event occurring over time, particularly in the context of Survival probability of Obesity. The Kaplan-Meier estimator allows us to construct survival curves, offering a visual representation of the probability of an event's occurrence at different time points. This method is particularly apt for handling censored data, where not all subjects may experience the event of interest during the study period. By utilizing the Kaplan-Meier estimator, our research seeks to provide a comprehensive understanding of the temporal dynamics inherent in smoking, alcohol, family history with overweight and Treatment contributing valuable insights into the survival patterns and durations relevant to our research objectives. The application of this statistical approach enhances the precision and depth of our time-to-event analyses, providing a robust framework for studying the progression of events over time."

- **COX REGRESSION:**

"As a pivotal component of our research methodology, Cox regression analysis plays a central role in examining the impact of multiple variables on the time until a particular event occurs. In our study, we aim to investigate the factors influencing the survival or duration of Obesity patients .The Cox regression model allows us to assess the hazard rates associated with different covariates, providing a nuanced understanding of their individual and collective contributions to the time-to-event outcome. This statistical approach is particularly well-suited for analyzing censored data and accommodating varying follow-up times among subjects. By leveraging Cox regression analysis, our research seeks to uncover significant predictors, quantify their effects, and elucidate the intricate relationships affecting the time dynamics of Survival and Hazard ratio of Obesity. The application of this method enhances the depth and precision of our survival analysis, offering valuable insights into the factors shaping the duration until the occurrence of the event under investigation."

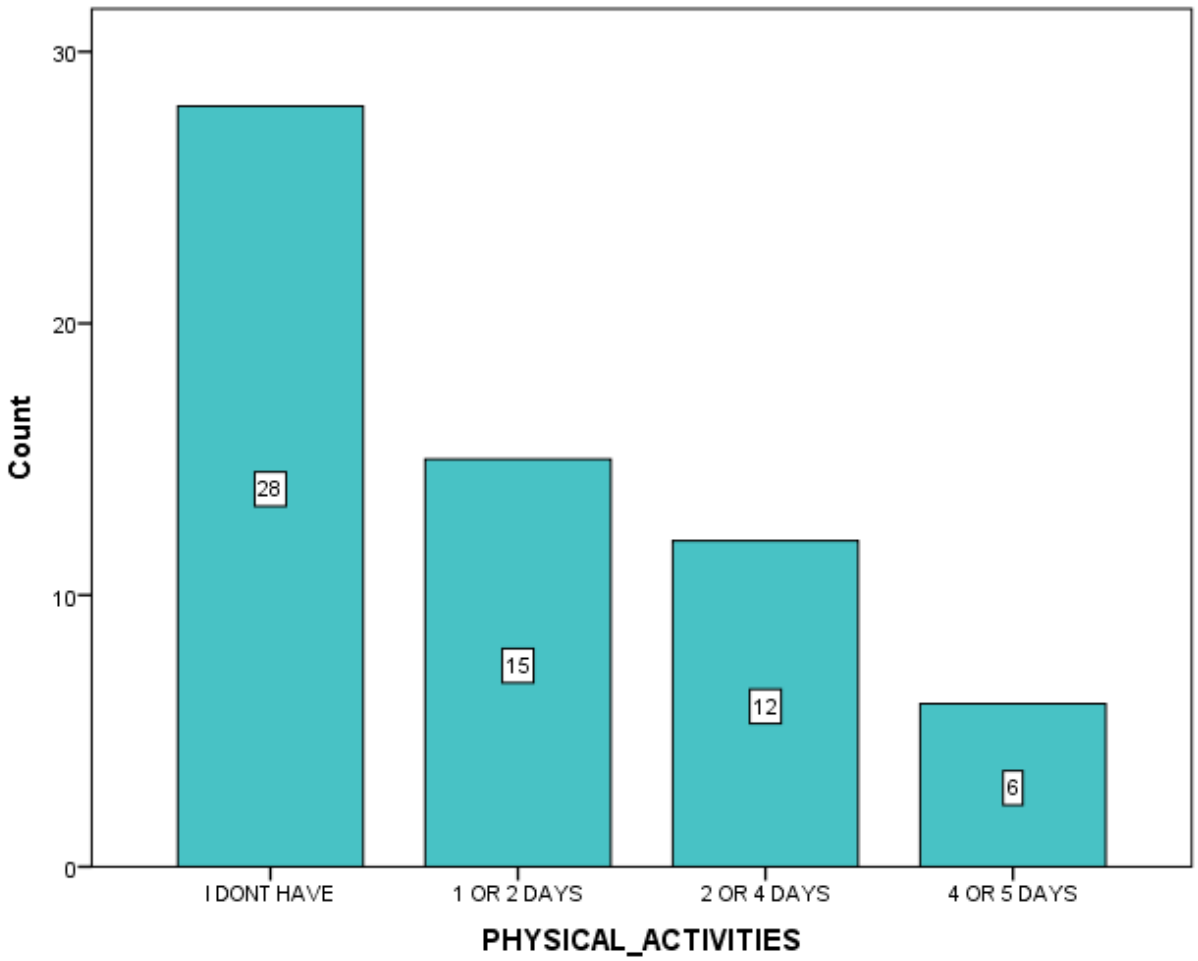
CHAPTER V
ANALYSIS AND INTERPRETATION

BAR DIAGRAM



INTERPRETATION:

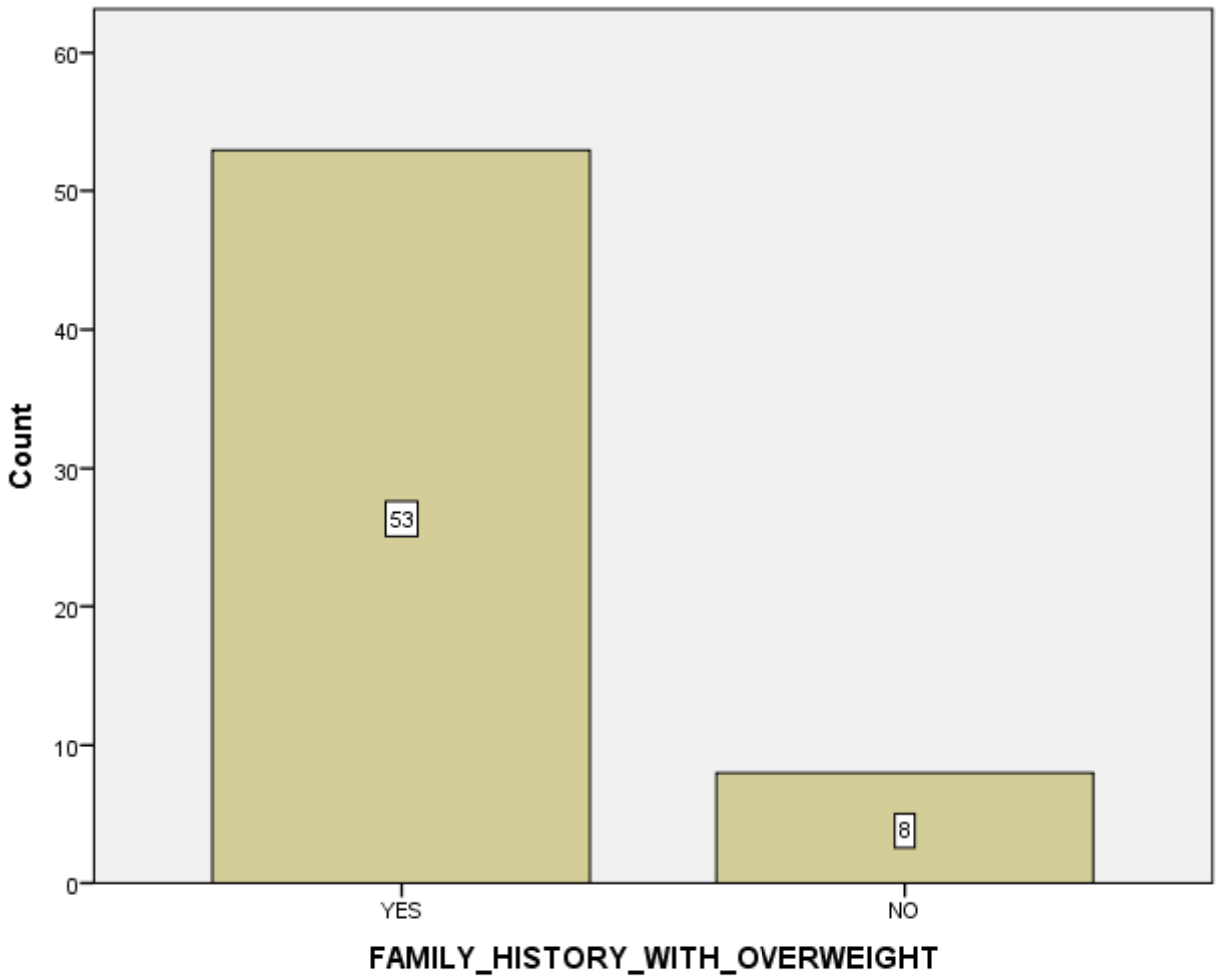
- This Bar diagram illustrates the distribution of individuals across different BMI categories based on gender.
- Each BMI category has two bars representing the counts for males and females. i.e., Under weight - 34(17 Male, 17 Female) ; Normal weight - 287(146 Male, 141 Female) ; Overweight - 116(70 Male, 46 Female) ; Obesity – 61(38 Male, 23 Female)
- This visual representation helps analyze the gender-specific patterns in BMI categories among individuals affected by obesity.



INTERPRETATION:

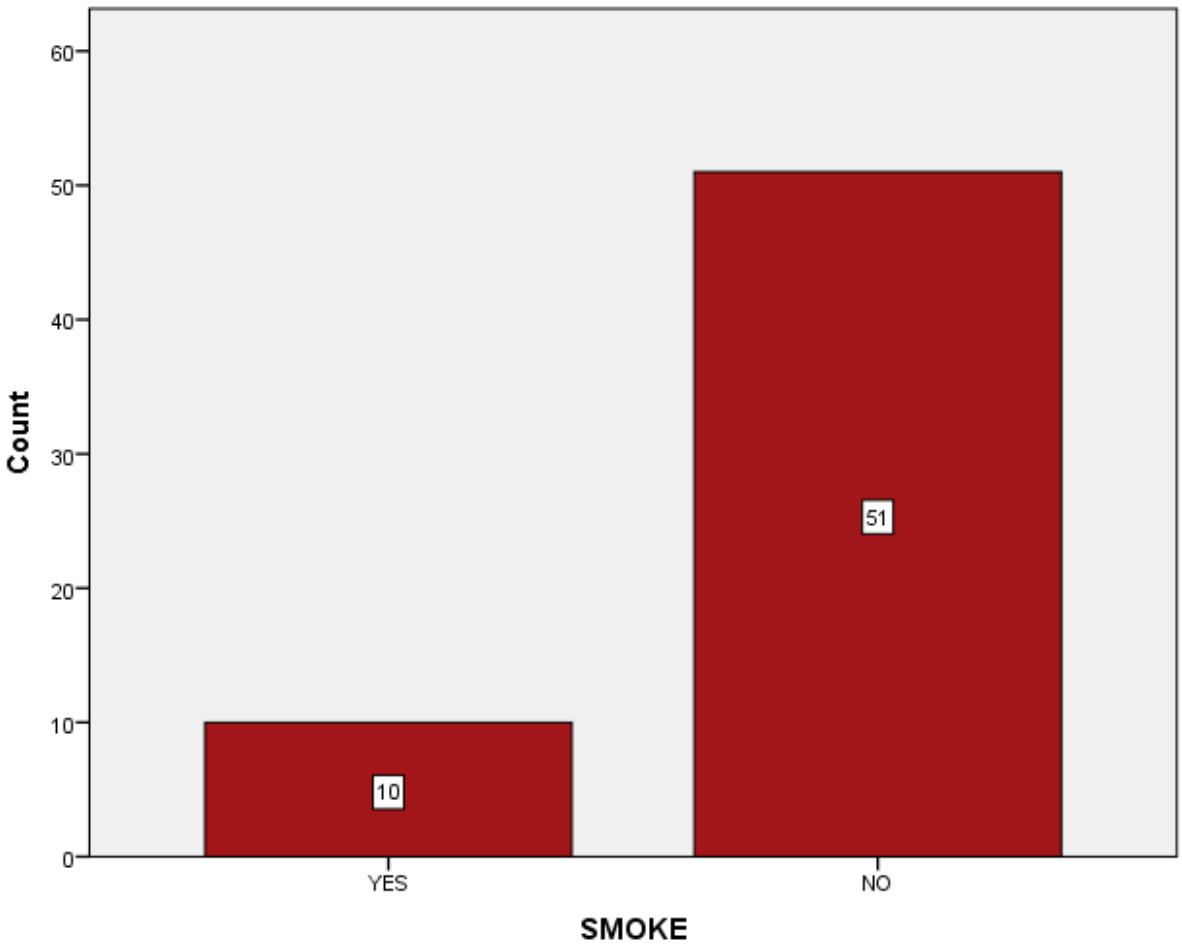
For the physical activity levels among 61 people affected by obesity:

- The bar diagram would illustrate different categories such as "I don't have," "1 or 2 days," "2 or 4 days," and "4 or 5 days."
- The counts for each category are as follows: "I don't have" (28), "1 or 2 days" (15), "2 or 4 days" (12), and "4 or 5 days" (6).
- The majority, 28 individuals, reported not engaging in physical activity.
- There's a range of physical activity levels among those who do engage, with varying frequencies from 1 or 2 days to 4 or 5 days.
- This information suggests a diversity in physical activity habits among individuals affected by obesity in this dataset.



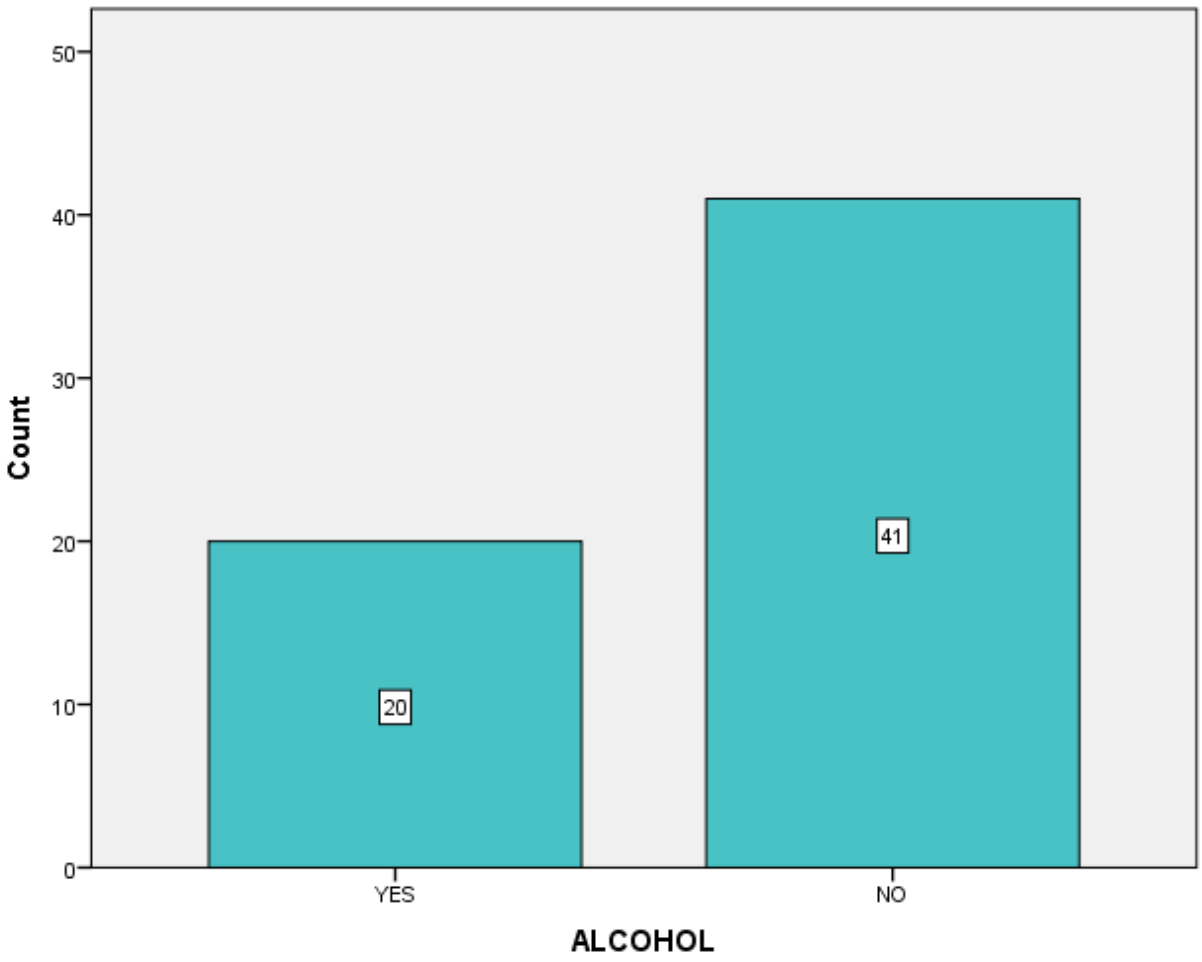
INTERPRETATION:

- For the given data on family history with obesity among 61 people affected by obesity: - The bar diagram would consist of two bars, one for "Yes" (53) and another for "No" (8). - The overwhelming majority, 53 out of 61, have a family history of obesity.
- This suggests a strong correlation between family history and obesity in this group, highlighting the potential genetic or environmental factors contributing to obesity.



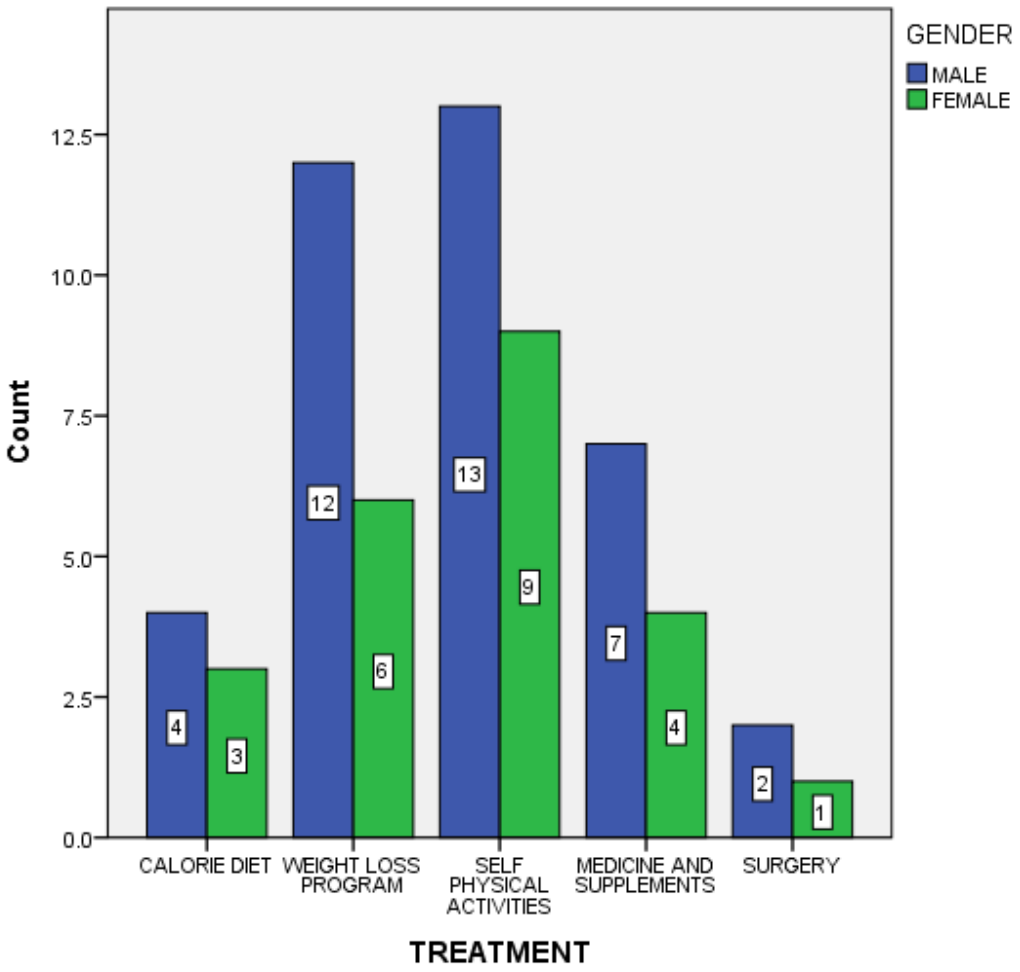
INTERPRETATION:

- Out of 61 people affected by obesity, 10 are smokers, and 51 are non-smokers. The bar diagram would show two bars, one representing "Yes" (smokers) and the other "No" (nonsmokers).
- This suggests that a relatively small percentage of people with obesity in this group are smokers.



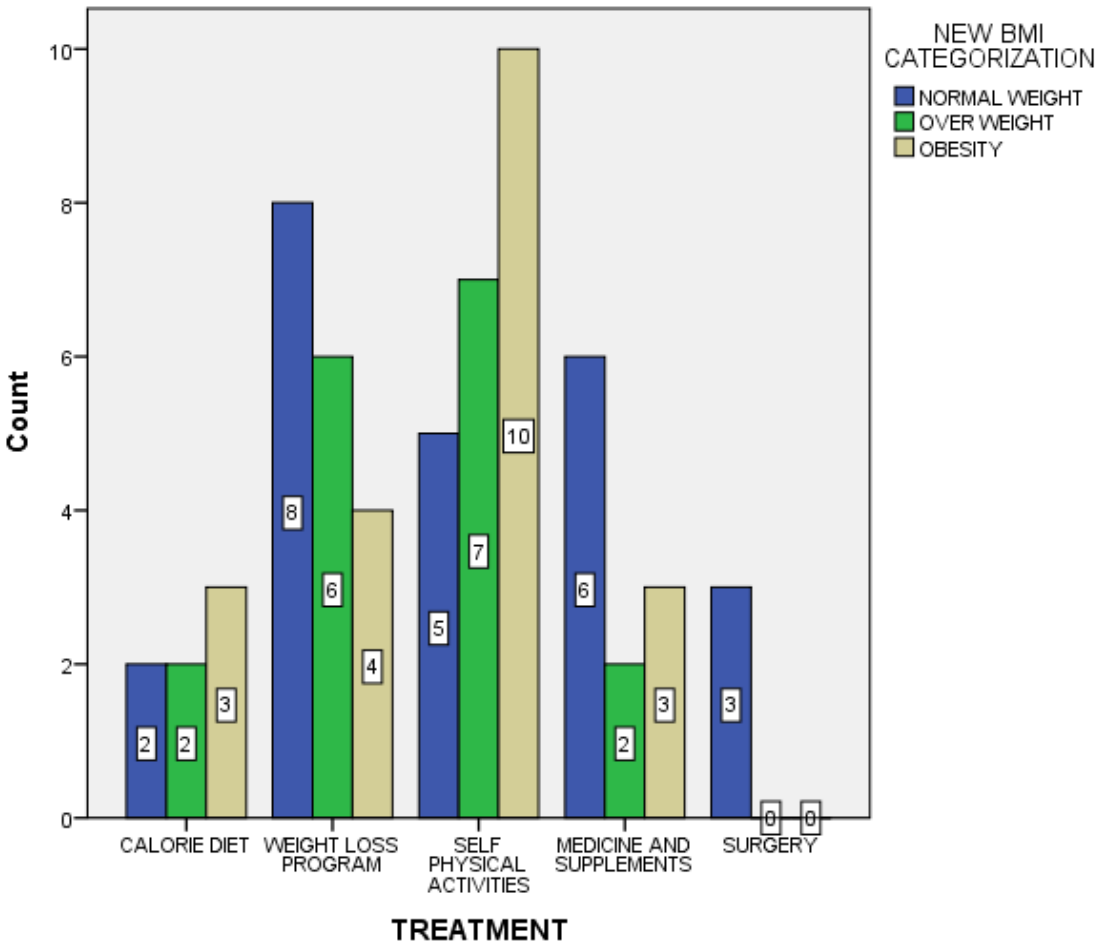
INTERPRETATION:

- Among the same 61 people affected by obesity, 20 consume alcohol, while 41 do not. The bar diagram for this would also have two bars, one for "Yes" (drinkers) and one for "No" (nondrinkers).
- This indicates that a larger percentage of people with obesity in this group consume alcohol. In interpretation, you may want to consider the potential implications of these habits on obesity and health.



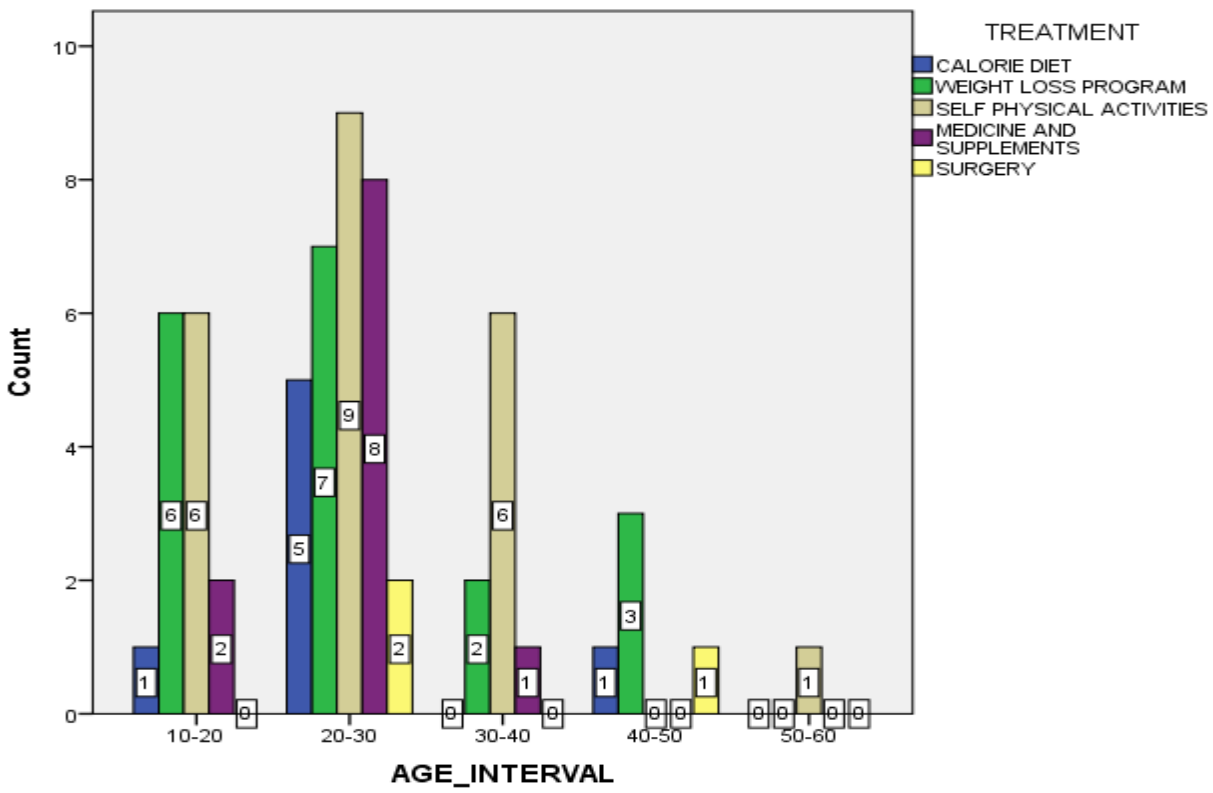
INTERPRETATION:

- This diagram illustrates the distribution of treatments across different genders. Each treatment method has two bars representing the counts for males and females. It provides a visual comparison of the treatments received by individuals affected by obesity based on their gender.
- There are 7 for calorie diet(3 female, 4 male), 18 for weight loss program(6 female, 12 male), 22 for self physical activities(9 female, 13 male), 11 for medicine and supplements(4 female, 7 male) and 3 for surgery(1 female, 2 male).
- This visual representation helps analyze the distribution of treatments based on gender among individuals affected by obesity. It indicates the gender-specific patterns in the types of treatments received.



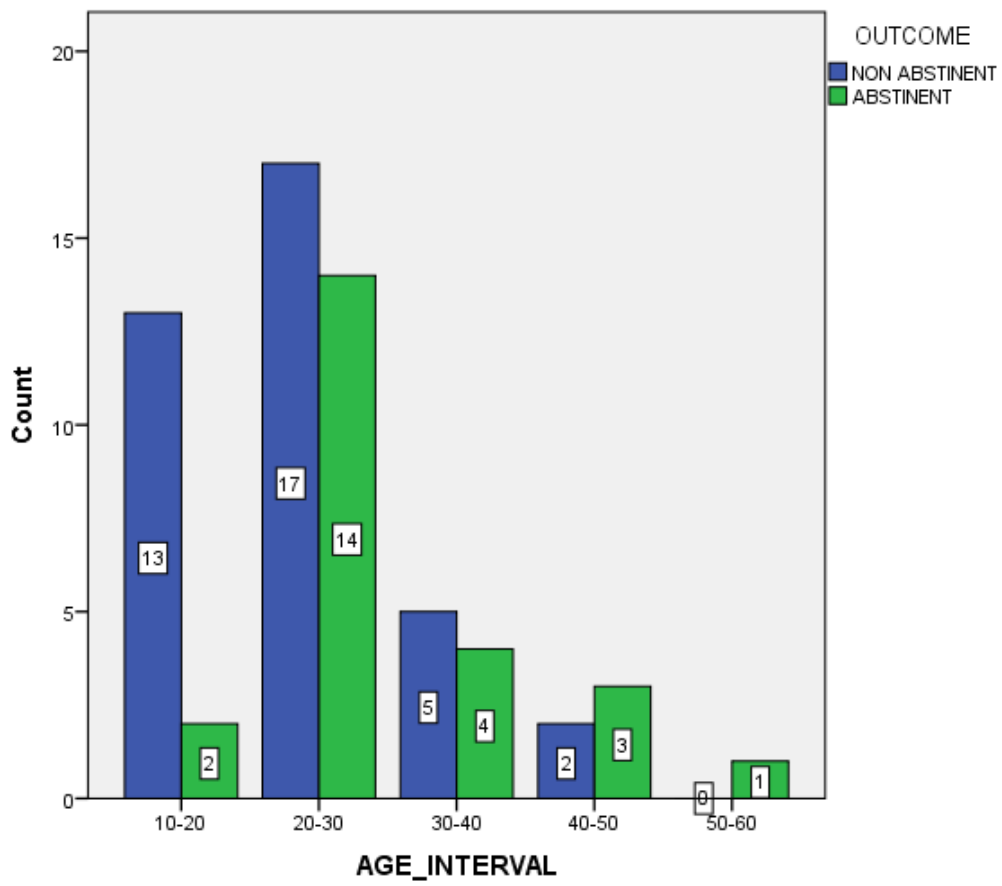
INTERPRETATION:

- Out of 61 people affected by obesity, there are 7 for calorie diet, 18 for weight loss program, 22 for self physical activities, 11 for medicine and supplements and 3 for surgery.
- This diagram illustrates the distribution of treatments across different BMI categories. Each treatment method has three bars representing the counts for normal weight, overweight, and obesity. For surgery, there are no cases reported for overweight or obese individuals.
- This visual representation helps analyze the distribution of treatments based on BMI categories among individuals affected by obesity. It also indicates that surgery is reported only for individuals with normal weight in this dataset.



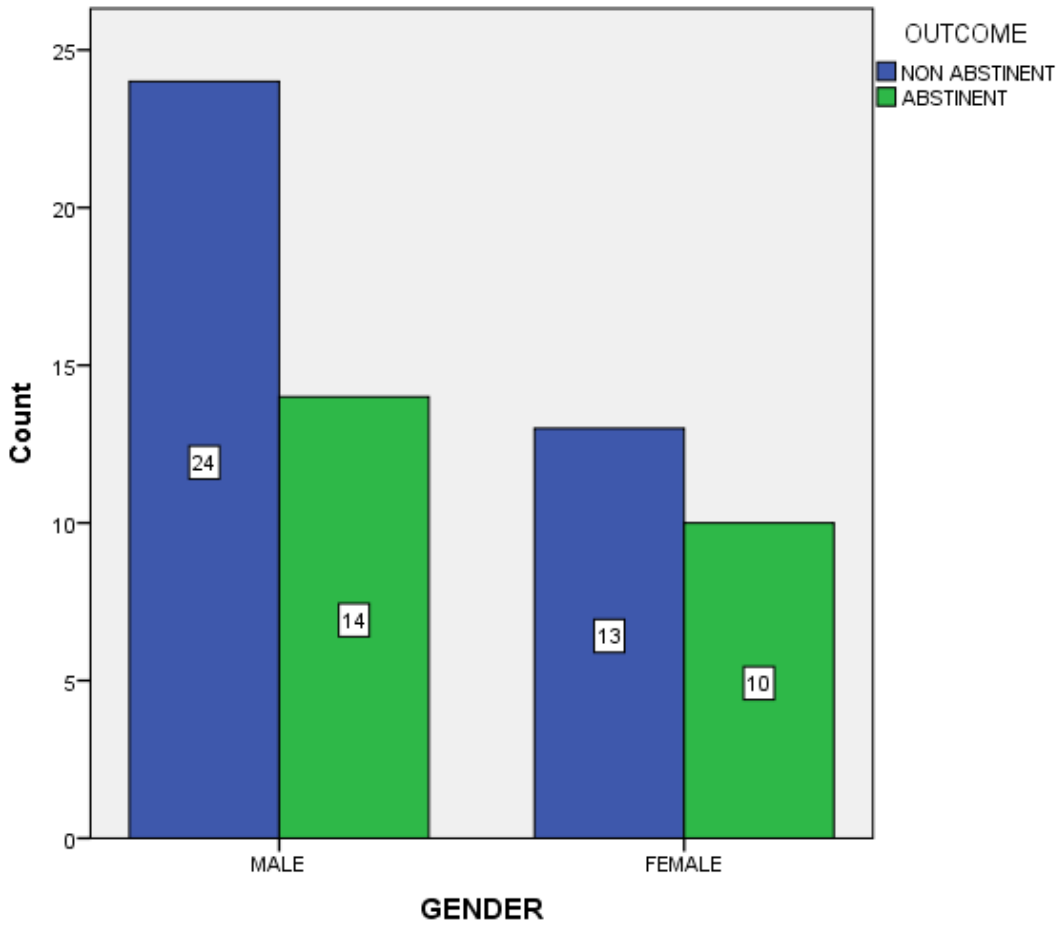
INTERPRETATION:

- This bar diagram illustrates the distribution of different treatments across different age intervals. Each age interval has bars representing the counts for each treatment category.
- It provides a visual comparison of treatment patterns among individuals affected by obesity based on their age.
- It shows how the prevalence of different treatments varies across different age groups.



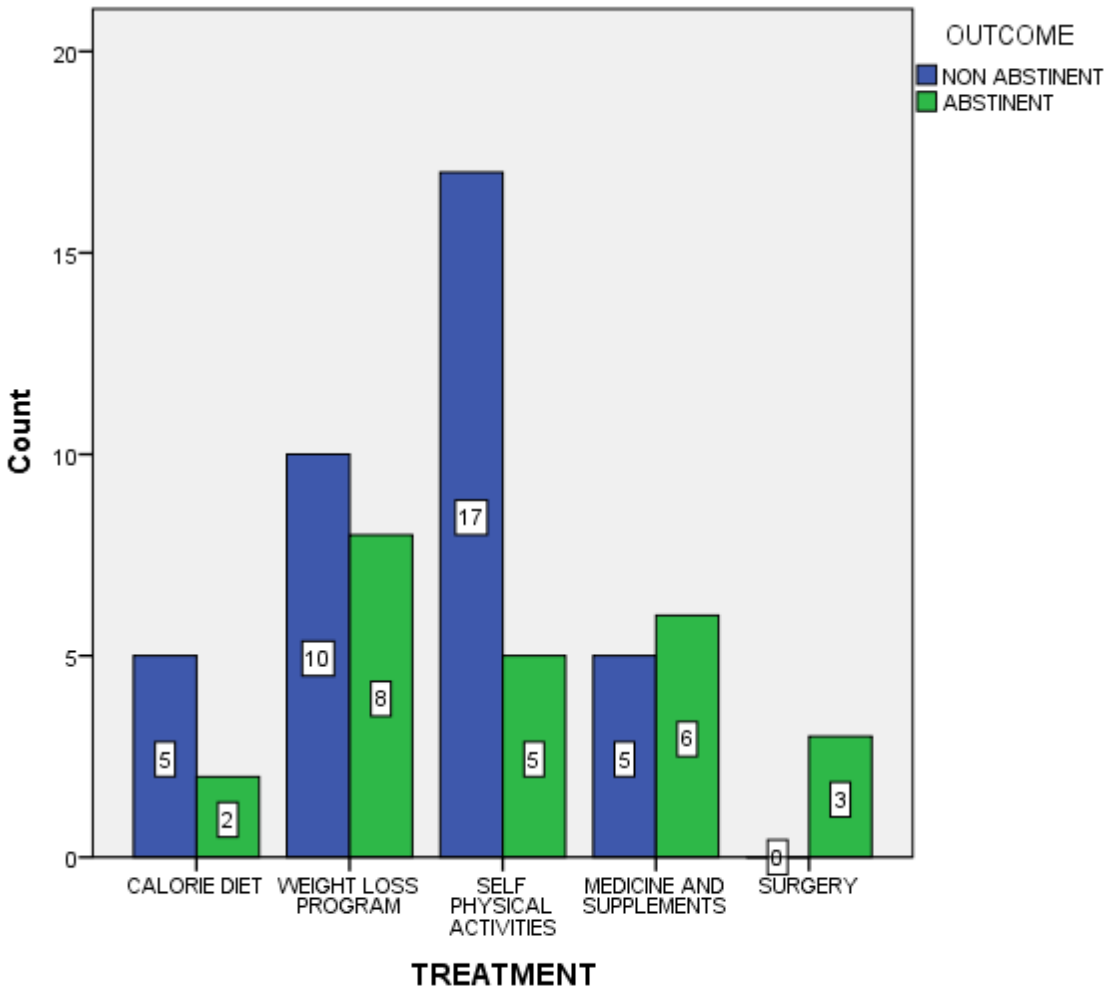
INTERPRETATION:

- This bar diagram illustrates the distribution of abstinent and non-abstinent individuals across different age intervals for 60 obesity people.
- Each age interval has two bars representing the counts for abstinent and non-abstinent individuals. It provides a visual comparison of abstinence patterns among individuals affected by obesity based on their age.
- There are 15 people in 10-20 (2 Absistent, 13 Non Absistent), 31 people in 20-30 (14 Absistent, 17 Non Absistent), 7 people in 30-40 (4 Absistent, 5 Non Absistent), 5 people in 40-50 (3 Absistent, 2 Non Absistent) and 1 in 50-60 (1 Absistent, 0 Non Absistent).
- This visual representation helps analyze the distribution of abstinence based on age intervals among individuals affected by obesity. It shows how the prevalence of abstinence varies across different age groups.



INTERPRETATION:

- Out of 61 people affected by obesity, there are 38 males and 23 females.
- The bar diagram would have two sets of bars for each gender, distinguishing between abstinent and non abstinent individuals. It shows the distribution of genders based on abstinence status.



INTERPRETATION:

- Out of 61 people affected by obesity, there are people counts of 7 for calorie diet, 18 for weight loss program, 22 for self physical activities, 11 for medicine and supplements and 3 for surgery.
- This bar diagram shows the distribution of treatment methods among abstinent and non-abstinent individuals. Each treatment method has two bars representing the respective abstinent and non abstinent counts.
- This visual representation helps compare the distribution of genders and treatment methods within the abstinent and non-abstinent groups among individuals affected by obesity.

CORRELATION

CORRELATION BETWEEN AGE AND NEW BMI

Correlations

		AGE	NEW BMI
AGE	Pearson Correlation	1	-.141
	Sig. (2-tailed)		.278
	N	61	61
NEW BMI	Pearson Correlation	-.141	1
	Sig. (2-tailed)	.278	
	N	61	61

HYPOTHESIS:

NULL HYPOTHESIS: There is no correlation between Age and New BMI ($\rho = 0$).

ALTERNATIVE HYPOTHESIS: There is a correlation between Age and New BMI ($\rho \neq 0$).

INTERPRETATION:

The Pearson correlation coefficient between Age and New BMI is -0.141, indicating a weak negative correlation. Since the p-value is 0.278 (greater than 0.05), this correlation is no statistically significant. In practical terms, it suggests that there is no strong linear relationship between Age and New BMI in the analyzed data.

CORRELATION BETWEEN NEW BMI AND ALCOHOL

Correlations

		NEW BMI	ALCOHOL
	Pearson Correlation	1	.212
	Sig. (2-tailed)		.101
	N	61	
NEW BMI	Pearson Correlation	.212	61 1
	Pearson Correlation	-.141	1
	Sig. (2-tailed)	.101	
ALCOHOL	N	61	61

HYPOTHESIS:

NULL HYPOTHESIS: There is no correlation between New BMI and Alcohol ($\rho = 0$).

ALTERNATIVE HYPOTHESIS: There is a correlation between New BMI and Alcohol ($\rho \neq 0$).

INTERPRETATION:

The Pearson correlation coefficient between New BMI and Alcohol is 0.212, indicating a positive correlation. Since the p-value is 0.101 (greater than 0.05), this correlation is not statistically significant. In practical terms, it suggests that there is no strong linear relationship between New BMI and Alcohol in the analyzed data.

CORRELATION BETWEEN SMOKE AND NEW BMI

Correlations

		SMOKE	NEW BMI
SMOKE	Pearson Correlation	1	-.065
	Sig. (2-tailed)		.621
	N	61	61
NEW BMI	Pearson Correlation	-.065	1
	Sig. (2-tailed)	.621	
	N	61	61

HYPOTHESIS:

NULL HYPOTHESIS: There is no correlation between Smoke and New BMI ($\rho = 0$).

ALTERNATIVE HYPOTHESIS: There is a correlation between Smoke and New BMI ($\rho \neq 0$).

INTERPRETATION:

The Pearson correlation coefficient between Age and New BMI is -0.065, indicating a very weak negative correlation. Since the p-value is 0.621(greater than 0.05), this correlation is not statistically significant. In practical terms, it suggests that there is no strong linear relationship between Age and New BMI in the analyzed data.

CORRELATION BETWEEN NEW BMI AND INTAKE OF CALORIC FOOD

Correlations

		NEW BMI	INTAKE_OF_CALORIC_FOOD
NEW BMI	Pearson Correlation	1	.024
	Sig. (2-tailed)		.857
	N	61	61
INTAKE_OF_CALORIC_FOOD	Pearson Correlation	.024	1
	Sig. (2-tailed)	.857	
	N	61	61

HYPOTHESIS:

NULL HYPOTHESIS: There is no correlation between New BMI and Intake of Caloric Food($\rho = 0$).

ALTERNATIVE HYPOTHESIS: There is a correlation between New BMI and Intake of Caloric Food ($\rho \neq 0$).

INTERPRETATION:

The Pearson correlation coefficient between New BMI and Intake of Caloric Food is 0.024, indicating a weak positive correlation. Since the p-value is 0.857 (greater than 0.05), this correlation is not statistically significant. In practical terms, it suggests that there is no strong linear relationship between New BMI and Intake of Caloric Food in the analyzed data.

CORRELATION BETWEEN NEW BMI AND FAMILY HISTORY WITH OVERWEIGHT

Correlations

		NEW BMI	FAMILY HISTORY WITH OVERWEIGHT
NEW BMI	Pearson Correlation	1	-.137
	Sig. (2-tailed)		.292
	N	61	61
FAMILY_HISTORY_WITH_ OVERWEIGHT	Pearson Correlation	-.137	1
	Sig. (2-tailed)	.292	
	N	61	61

HYPOTHESIS:

NULL HYPOTHESIS: There is no correlation between New BMI and Family History with Overweight ($\rho = 0$).

ALTERNATIVE HYPOTHESIS: There is a correlation between New BMI and Family History with Overweight
($\rho \neq 0$).

INTERPRETATION:

The Pearson correlation coefficient between New BMI and Family History with Overweight is -0.137, indicating a weak negative correlation. Since the p-value is 0.292 (greater than 0.05), this correlation is not statistically significant. In practical terms, it suggests that there is no strong linear relationship between New BMI and Family History with Overweight in the analyzed data.

CORRELATION BETWEEN NEW BMI AND OUTCOME

Correlations

		NEW BMI	OUTCOME
NEW BMI	Pearson Correlation	1	-.669**
	Sig. (2-tailed)		.000
	N	61	61
OUTCOME	Pearson Correlation	-.669**	1
	Sig. (2-tailed)	.000	
	N	61	61

HYPOTHESIS:

NULL HYPOTHESIS: There is no correlation between New BMI and Outcome ($\rho = 0$). ALTERNATIVE HYPOTHESIS: There is a correlation between New BMI and Outcome ($\rho \neq 0$).

INTERPRETATION:

The Pearson correlation coefficient between New BMI and Outcome is -0.669, indicating a strong negative correlation. Since the p-value is 0.001(less than 0.05), this correlation is statistically significant. In practical terms, it suggests that there is a strong linear relationship between New BMI and Outcome in the analyzed data.

CORRELATION BETWEEN AGE AND OUTCOME

Correlations

		AGE	OUTCOME
AGE	Pearson Correlation	1	.239
	Sig. (2-tailed)		.064
	N	61	61
OUTCOME	Pearson Correlation	.239	1
	Sig. (2-tailed)	.064	
	N	61	61

HYPOTHESIS:

NULL HYPOTHESIS: There is no correlation between Age and Outcome ($\rho = 0$).

ALTERNATIVE HYPOTHESIS: There is a correlation between Age and Outcome ($\rho \neq 0$).

INTERPRETATION:

The Pearson correlation coefficient between Age and Outcome is 0.239, indicating a positive correlation. Since the p-value is 0.064 (greater than 0.05), this correlation is not statistically significant. In practical terms, it suggests that there is a strong linear relationship between Age and Outcome in the analyzed data.

CHI SQUARE

SMOKING*OBESITY

SMOKING	OBESITY	NON-OBESITY
YES	10	22
NO	51	415

Test Statistics

	SMOKING	OBESITY
Chi-Square	378.225[a]	283.888[a]
df	1	1
Asymp. Sig.	.000	.000

HYPOTHESIS:

NULL HYPOTHESIS : There is no association between groups.

ALTERNATIVE HYPOTHESIS : There is a association between groups.

INTERPRETATION:

- The value of the test statistic for smoking is 378.225.
- Since the p-value is less than our chosen significance level ($\alpha = 0.05$), we reject the null hypothesis. Rather, we conclude that there is an enough evidence to suggest an association between Smoking and Obesity.
- Based on the results, we can state that there is a association between Smoking and Obesity ($X^2(2) > = 378.225$, $p = 0.000$).

ALCOHOL*OBESITY

ALCOHOL	OBESITY	NON-OBESITY
SOMETIMES	30	242
FREQUENTLY	9	37
NO	22	158

Test Statistics

	ALCOHOL	OBESITY
Chi-Square	155.614[a]	283.888[b]
df	2	1
Asymp. Sig.	.000	.000

HYPOTHESIS:

NULL HYPOTHESIS : There is no association between groups.

ALTERNATIVE HYPOTHESIS : There is a association between groups.

INTERPRETATION:

- The value of the test statistic for Alcohol is 155.614
- Since the p-value is less than our chosen significance level ($\alpha = 0.05$), we reject the null hypothesis. Rather, we conclude that there is an enough evidence to suggest an association between Alcohol and Obesity.
- Based on the results, we can state that there is a association between Alcohol and Obesity ($X^2(2) > = 155.614$, $p = 0.000$).

PHYSICAL ACTIVITY*OBESITY

PHYSICAL ACTIVITY	OBESITY	NON-OBESITY
NO	28	134
1 OR 2 DAYS	15	143
2 OR 4 DAYS	12	101
4 OR 5 DAYS	6	59

Test Statistics

	PHYSICAL ACTIVITY	OBESITY
Chi-Square	49.807[a]	283.888[b]
df	3	1
Asymp. Sig.	.000	.000

HYPOTHESIS:

NULL HYPOTHESIS : There is no association between groups.

ALTERNATIVE HYPOTHESIS : There is a association between groups.

INTERPRETATION:

- The value of the test statistic for Physical Activity is 49.807.
- Since the p-value is less than our chosen significance level ($\alpha = 0.05$), we reject the null hypothesis. Rather, we conclude that there is an enough evidence to suggest an association between Physical Activity and Obesity.
- Based on the results, we can state that there is a association between Physical Activity and Obesity ($X^2(2) > = 49.807$, $p = 0.000$).

FAMILY HISTROY WITH OVERWEIGHT*OBESITY

FAMILY HISTORY	OBESITY	NON-OBESITY
YES	53	247
NO	8	190

Test Statistics

	FAMILY_HISTORY	OBESITY
Chi-Square	20.892[a]	283.888[a]
df	1	1
Asymp. Sig.	.000	.000

HYPOTHESIS:

NULL HYPOTHESIS : There is no association between groups.

ALTERNATIVE HYPOTHESIS : There is a association between groups.

INTERPRETATION:

- The value of the test statistic for Family History With Overweight is 20.892.
- Since the p-value is less than our chosen significance level ($\alpha = 0.05$), we reject the null hypothesis. Rather, we conclude that there is an enough evidence to suggest an association between Family History With Overweight and Obesity.
- Based on the results, we can state that there is a association between Family History With Overweight and Obesity ($X^2(2) > = 20.892$, $p= 0.000$).

CALORIC FOOD*OBESITY

CALORIC	OBESITY	NON-OBESITY
YES	44	304
NO	17	133

Test Statistics

	CALORIC FOOD	OBESITY
Chi-Square	78.723[a]	283.888[a]
df	1	1
Asymp. Sig.	.000	.000

HYPOTHESIS:

NULL HYPOTHESIS : There is no association between groups.

ALTERNATIVE HYPOTHESIS : There is a association between groups.

INTERPRETATION:

- The value of the test statistic for Caloric Food is 78.723.
- Since the p-value is less than our chosen significance level ($\alpha = 0.05$), we reject the null hypothesis. Rather, we conclude that there is an enough evidence to suggest an association between Caloric Food and Obesity.
- Based on the results, we can state that there is a association between Caloric Food and Obesity ($X^2(2) > = 78.723$, $p = 0.000$).

H2O*OBESITY

H2O	OBESITY	NON-OBESITY
LESS THAN 1 LIT	17	118
BETWEEN 1 TO2 LIT	25	241
MORE THAN 2 LIT	241	78

Test Statistics

	H2O	OBESITY
Chi-Square	94.711[a]	283.888[b]
df	2	1
Asymp. Sig.	.000	.000

HYPOTHESIS:

NULL HYPOTHESIS : There is no association between groups.

ALTERNATIVE HYPOTHESIS : There is a association between groups.

INTERPRETATION:

- The value of the test statistic for H2O is 94.711.
- Since the p-value is less than our chosen significance level ($\alpha = 0.05$), we reject the null hypothesis. Rather, we conclude that there is an enough evidence to suggest an association between H2O and Obesity.
- Based on the results, we can state that there is a association between H2O and Obesity ($X^2(2) > = 94.711$, $p = 0.000$).

TRANSPORTATION*OBESITY

TRANSPORTATION	OBESITY	NON-OBESITY
WALKING	3	52
CYCLING	3	8
BIKE	1	6
AUTOMOBILE	19	80
PUBLIC TRANSPORTATION	35	291

Test Statistics

	TRANSPORTATION	OBESITY
Chi-Square	699.510[a]	283.888[b]
Df	4	1
Asymp. Sig.	.000	.000

HYPOTHESIS:

NULL HYPOTHESIS : There is no association between groups.

ALTERNATIVE HYPOTHESIS : There is a association between groups.

INTERPRETATION:

- The value of the test statistic for Transportation is 699.510
- Since the p-value is less than our chosen significance level ($\alpha = 0.05$), we reject the null hypothesis. Rather, we conclude that there is an enough evidence to suggest an association between Transportation and Obesity.
- Based on the results, we can state that there is a association between Transportation and Obesity ($X^2(2) > = 699.510$, $p = 0.000$).

ANOVA(ONE WAY)

ANOVA FOR NEW WEIGHT AND TREATMENTS

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1053.145	4	263.286	.854	.497
Within Groups	17260.494	56	308.223		
Total	18313.639	60			

HYPOTHESIS:

NULL HYPOTHESIS: There is no significant difference in the mean new weights among the treatment groups.

ALTERNATIVE HYPOTHESIS: There is a significant difference in the mean new weights among the treatment groups.

INTERPRETATION:

The P-Value is greater than the significant value (0.497>0.05) which indicates an enough evidence to suggest that there is no significant difference in the mean new weights among the treatment groups. So we can conclude that there is no significant difference between the mean new weights among the treatment groups.

LOGISTIC REGRESSION

LOGISTIC REGRESSION ON AGE, HEIGHT, WEIGHT WITH OUTCOME

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Age	.074	.042	3.105	1	.078	1.077
	Height	.163	.073	5.050	1	.025	1.177
	Weight	-.163	.057	8.116	1	.004	.849
	Constant	-14.483	8.190	3.127	1	.077	.000

a. Variable(s) entered on step 1: AGE, HEIGHT, WEIGHT.

INTERPRETATION:

Y=-14.483+0.074(AGE)+0.163(HEIGHT)-0.163(WEIGHT)

In age and height there is a value greater that 1 which indicates higher odds of the event happening and each unit increase in the increased in the predictor variable multiplies the odds by the odds ratio.

While in weight there is a value less than 1 which suggests that lower odds of the event happening

In age and height there is a positive coefficient which increases the log-odd of the dependent variable.

On the other hand there is a negative coefficient for weight which suggests decreased log-odd of the dependent variable.

LOGISTIC REGRESSION ON FAMILY HISTORY WITH OVERWEIGHT, INTAKE OF CALORIC FOOD, SMOKE, ALCOHOL WITH OUTCOME

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Family_History_with_Overweight	.034	.812	.002	1	.967	1.034
	Intake_of_Caloric_Food	.199	.612	.106	1	.745	1.220
	Smoke	.464	.783	.351	1	.554	1.590
	Alcohol	-1.402	.606	5.352	1	.021	.246
	Constant	.737	1.902	.150	1	.698	2.091

a. Variable(s) entered on step 1: FAMILY_HISTORY_WITH_OVERWEIGHT, INTAKE_OF_CALORIC_FOOD, SMOKE, ALCOHOL.

INTERPRETATION:

Y=0.737-1.402(alcohol)+0.464(smoking)+0.199(intake of caloric food)+0.034(family history with overweight)

In Family history of over height , Intake of caloric food and smoking has a value greater that 1 which indicates higher odds of the event happening and each unit increase in the increased in the predictor variable multiplies the odds by the odds ratio.

While in Alcohol there is a value less than 1 which suggests that lower odds of the event happening.

In Family history with overweight , intake of caloric food and smoking has a positive coefficient which increases the log-odd of the dependent variable. On the other hand there is a negative coefficient for Alcohol which suggests decreased log-odd of the dependent variable.

PAIRED T TEST

PAIRED T TEST ON OLD WEIGHT AND NEW WEIGHT

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	WEIGHT - NEW WEIGHT	17.6148	5.8917	.7543	16.1058	19.1237	23.351	60	.000

HYPOTHESIS:

NULL HYPOTHESIS : There is no significant difference between the paired observations of two weights which is obtained while before and after the treatment (mean difference is zero).

ALTERNATIVE HYPOTHESIS : There is a significant difference between the paired observations of two weights which is obtained while before and after the treatment.

INTERPRETATION:

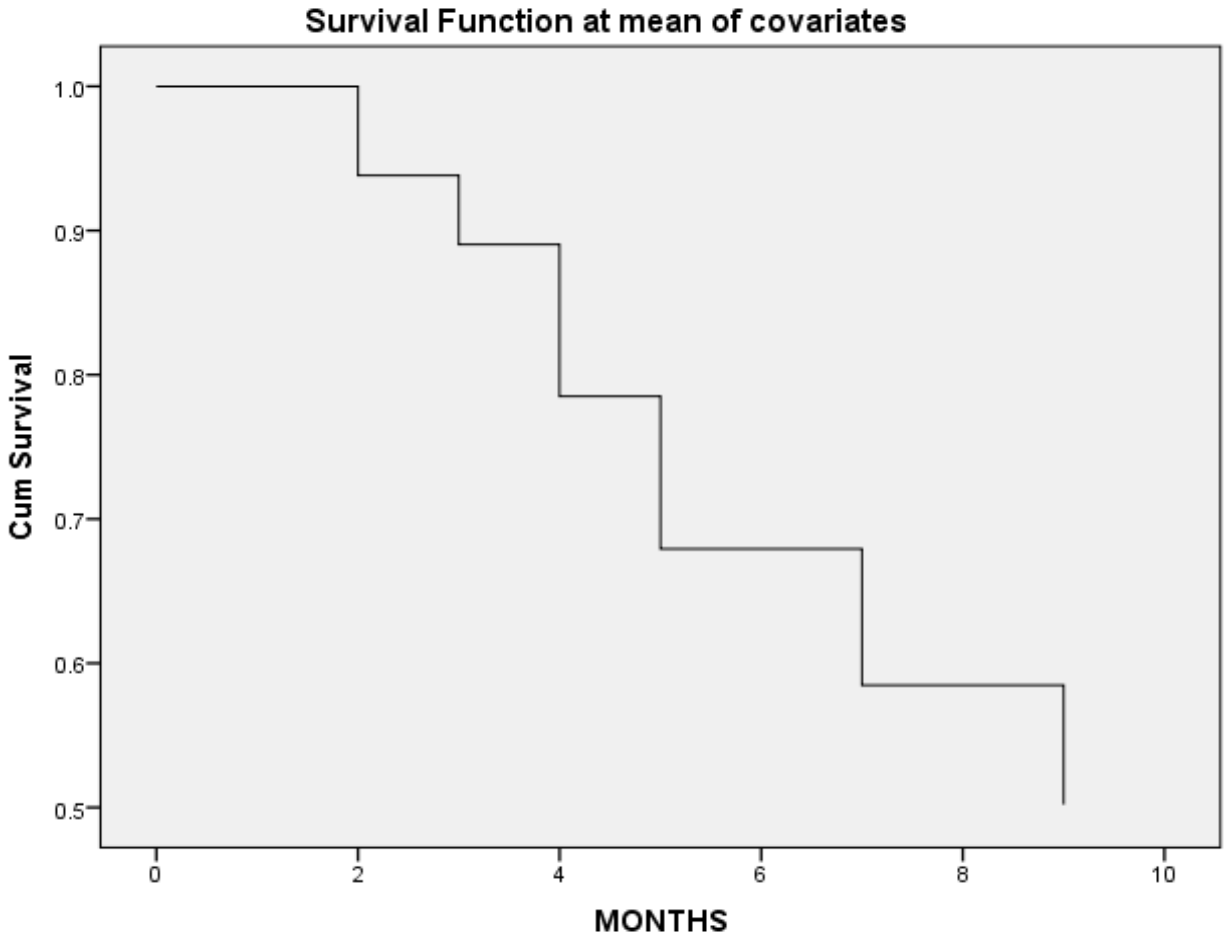
The P-Value is less than the significant value ($0.000 < 0.05$) which indicates, there is no enough evidence to suggest as there is no significant difference between the paired observations of two weights which is obtained while before and after the treatment.

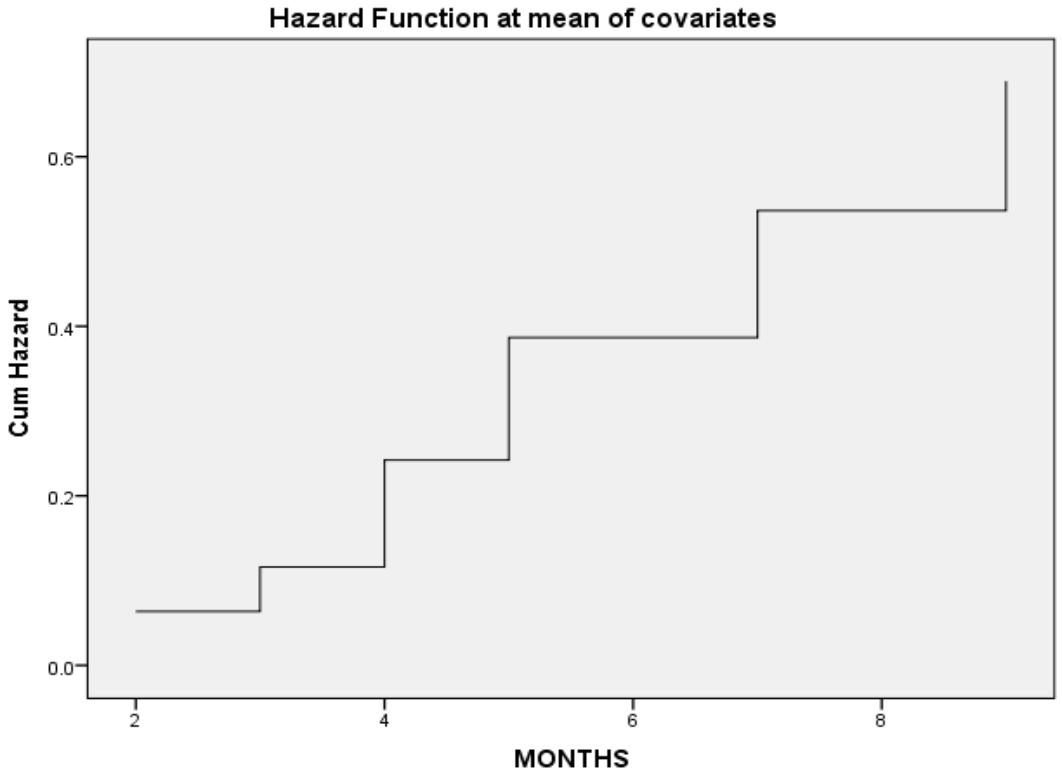
So we can conclude that there is a significant difference between the two weights which is obtained while before and after the treatment.

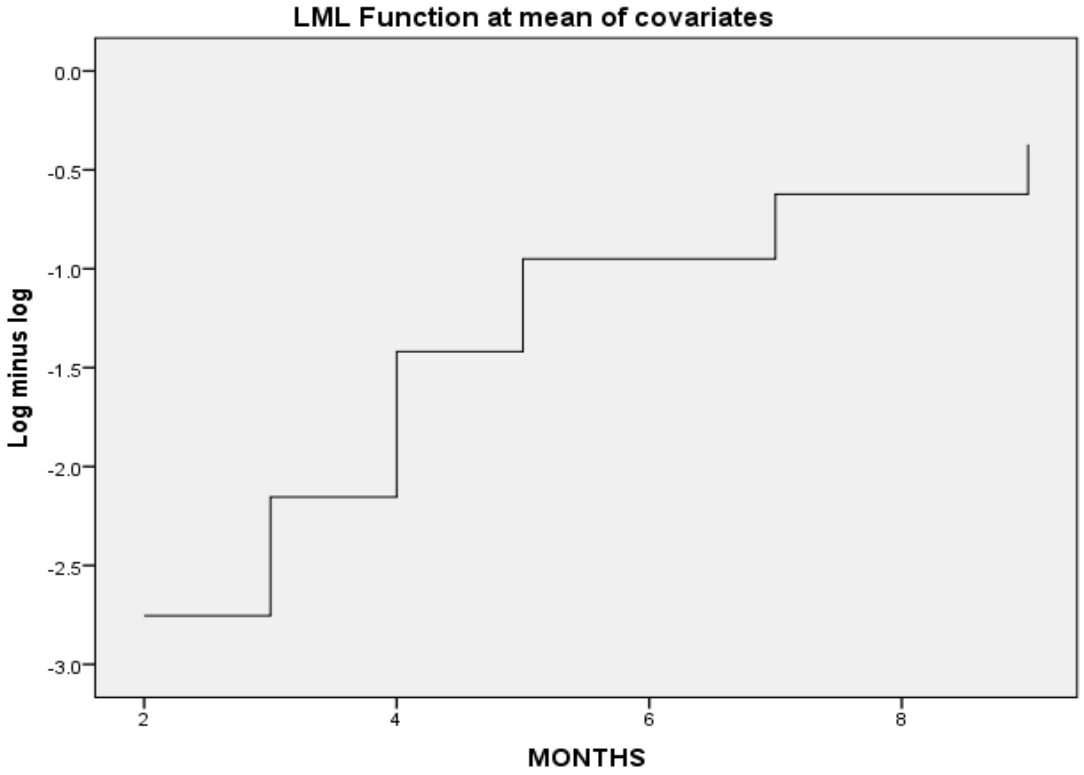
COX REGRESSION

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
FAMILY_HISTORY_WITH_OVERWEIGHT	.096	.631	.023	1	.879	1.101
INTAKE_OF_CALORIC_FOOD	.408	.525	.604	1	.437	1.503
SMOKE	-.084	.560	.022	1	.881	.920
H2O	-.087	.318	.075	1	.784	.917
PHYSICAL_ACTIVITIES	-.418	.252	2.750	1	.097	.659




Hazard Function at mean of covariates
Cum Hazard
0.0
0.2
0.4
0.6
2
4
6
8
MONTHS


LML Function at mean of covariates
Log minus log
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
2
4
6
8
MONTHS

HYPOTHESIS:

NULL HYPOTHESIS: The effect of the predictor variable is not statistically significant.
ALTERNATIVE HYPOTHESIS: The effect of the predictor variable is statistically significant.

INTERPRETATION:

HAZARD RATIO:

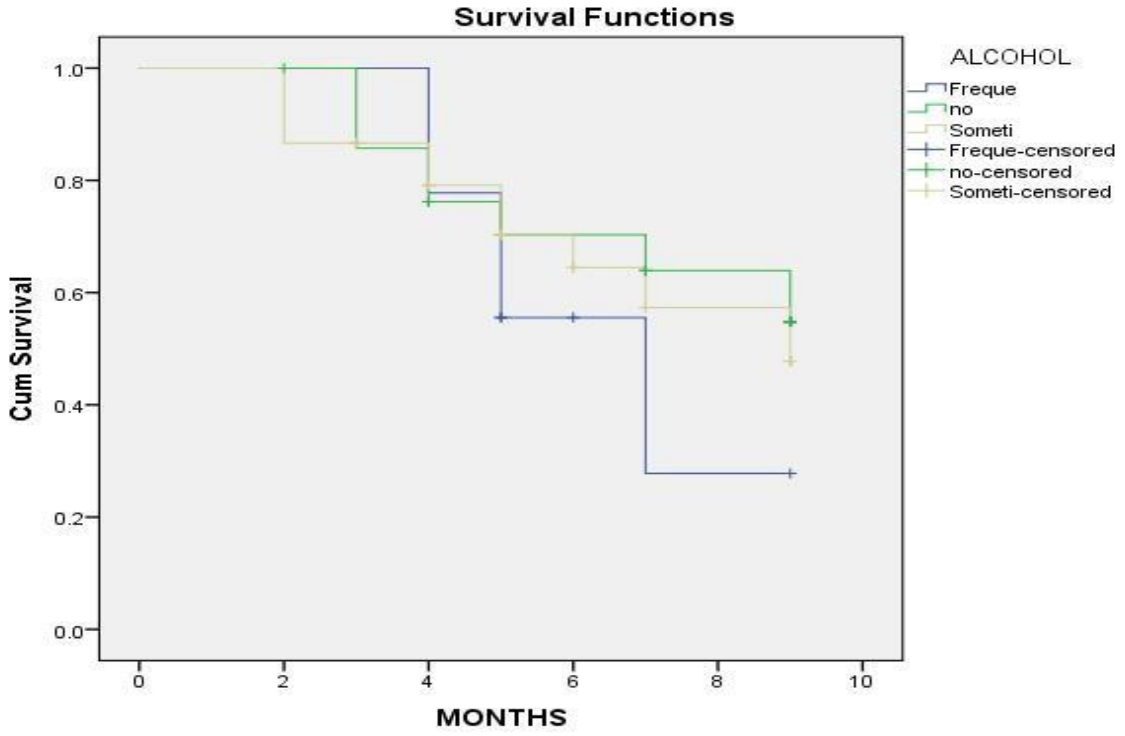
In analysing the survival tool of cox regression analysis, it is evident that the Hazard Ratio,

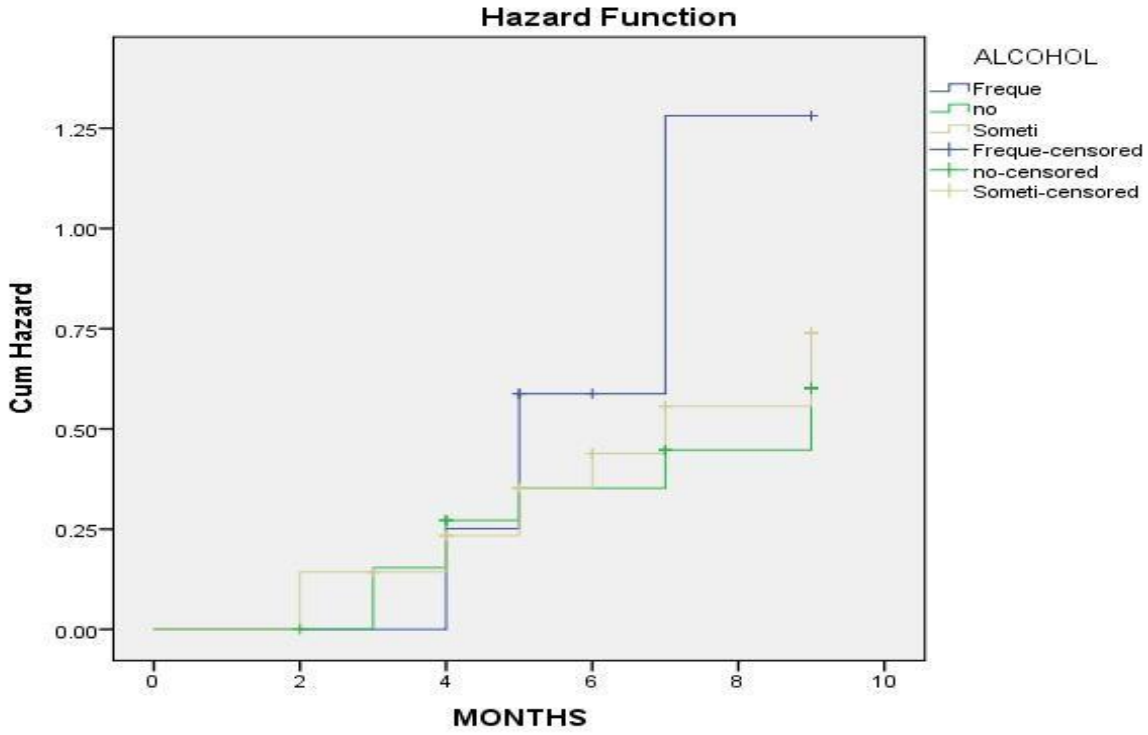
- Family history with overweight is greater than 1 which indicates the risk of affecting with obesity.
- Intake of caloric food is also greater than 1 which indicates the risk of affecting with obesity.
- Smoking factor has hazard ratio closer to one but less than one which suggests that it has lower risk of affecting with obesity
- H2O factor has hazard ratio less than one which suggests that it has lower risk of affecting with obesity
- Physical activity has hazard ratio less than one which suggests that it has lower risk of affecting with obesity

P Value is greater than the significance level. So we accept the Null Hypothesis and conclude that the effect of all predictor variable is not statistically significant for all the taken factors.
The Survival function at mean of covariates shows a gradual decrease in the probability of the factors. The log minus log graph indicates that in initial four months, there is a greater impact of covariates on survival probability and then the horizontal line gradually remains which suggests that there is a constant hazard rates so we can conclude after 5 months the risk is consistent.

KAPLAN MEIER

KAPLAN MEIER ANALYSIS ON ALCOHOL





Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	.818	2	.664

Test of equality of survival distributions for the different levels of Alcohol .

HYPOTHESIS:

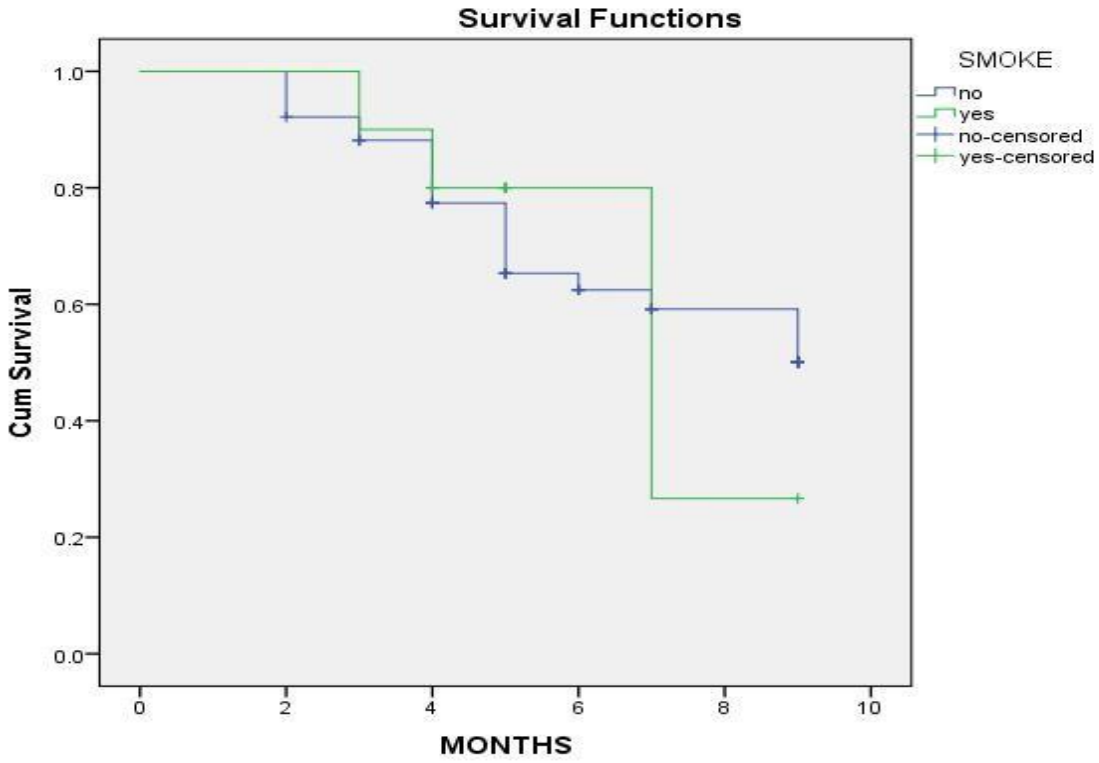
NULL HYPOTHESIS: There is no difference in survival between the groups under investigation.

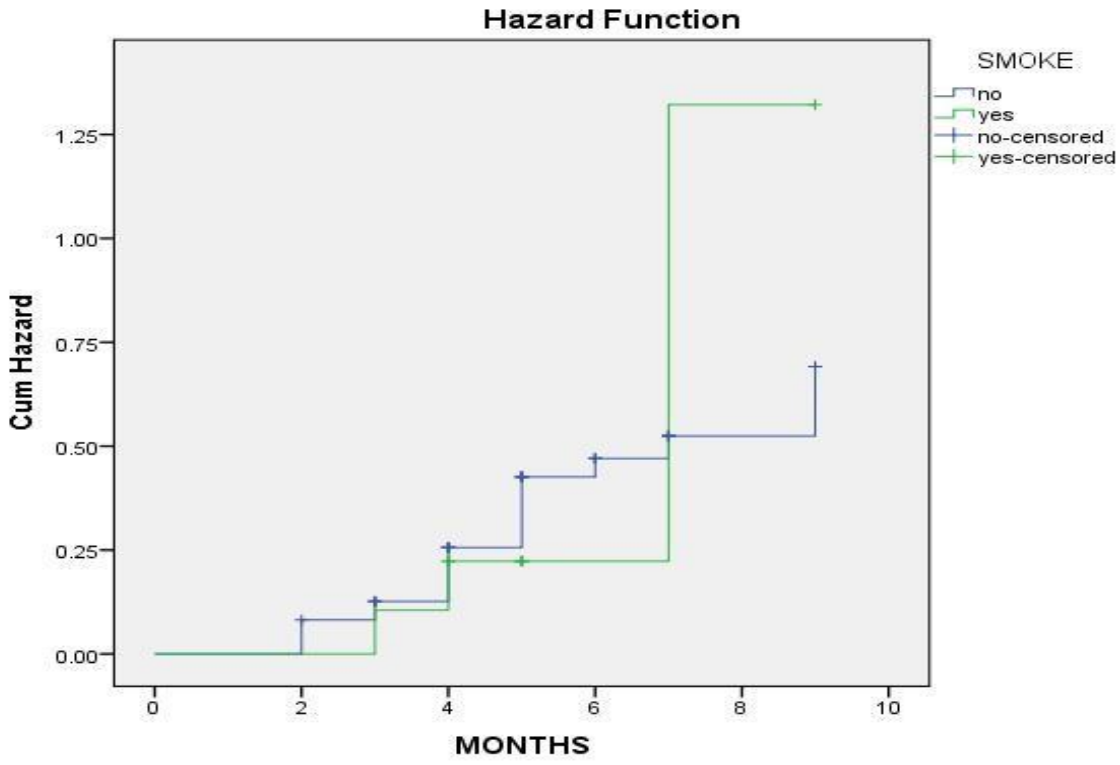
ALTERNATIVE HYPOTHESIS: There is a significant difference in survival between the groups under investigation.

INTERPRETATION:

- P Value is greater than the significance level. So we accept the Null Hypothesis and conclude that there is no difference in survival between the groups under investigation.
- Kaplan-Meier curves provide a visual representation of survival probabilities of alcohol consumptions where the survival probability of frequently consuming alcohol is less compared with sometimes and no alcohol consumption.
- From the Hazard function curve we can conclude that no alcohol consumption has least hazard ratio compared to other two factors which indicates that there is a lower risk of obesity for those who does not consume alcohol.

KAPLAN MEIER ANALYSIS ON SMOKE





Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	.026	1	.873

Test of equality of survival distributions for the different levels of Smoke.

HYPOTHESIS:

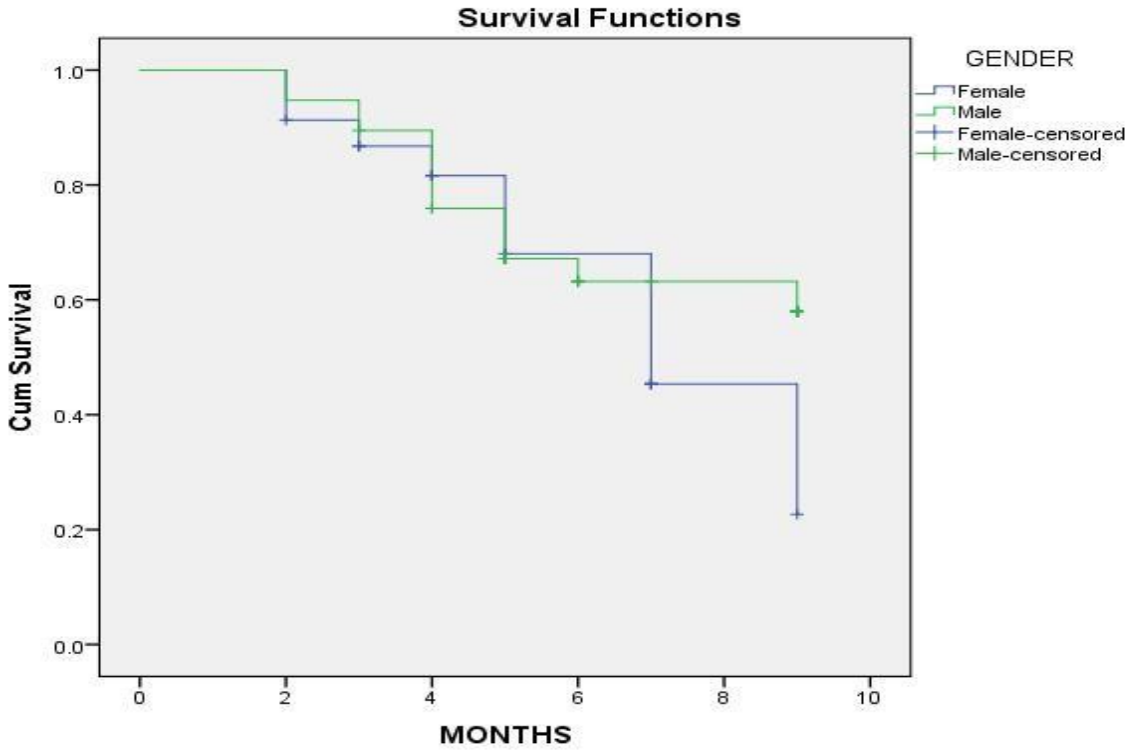
NULL HYPOTHESIS: There is no difference in survival between the groups under investigation.

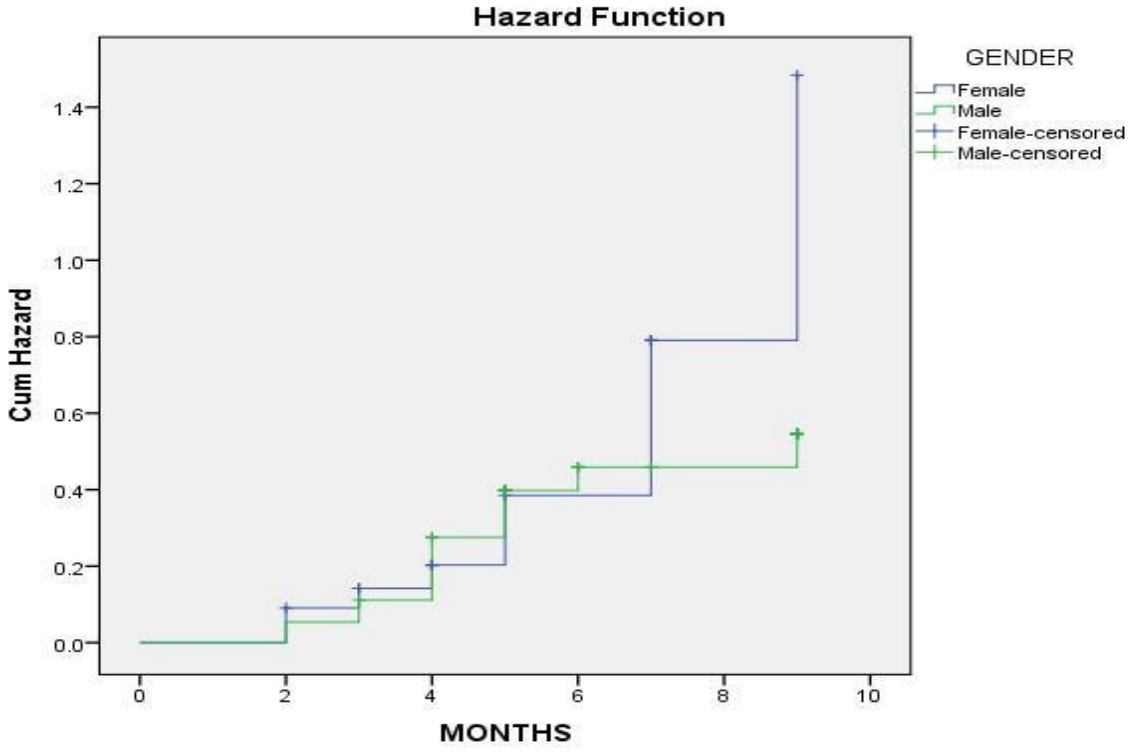
ALTERNATIVE HYPOTHESIS: There is a significant difference in survival between the groups under investigation.

INTERPRETATION:

- P Value is greater than the significance level. So we accept the Null Hypothesis and conclude that there is no difference in survival between the groups under investigation.
- Kaplan-Meier curves provide a visual representation of survival probabilities of Smoking, where the survival probability of Non-smokers is greater compared with smokers which implies that smoke is not the main factor for the cause of obesity.
- From the Hazard function curve we can conclude that Smokers has least hazard ratio compared to Non smokers which shows that they have greater risk and we can conclude smoke is not the main factor for the cause of obesity.

KAPLAN MEIER ANALYSIS ON GENDER





Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	.780	1	.377

Test of equality of survival distributions for the different levels of Gender.

HYPOTHESIS:

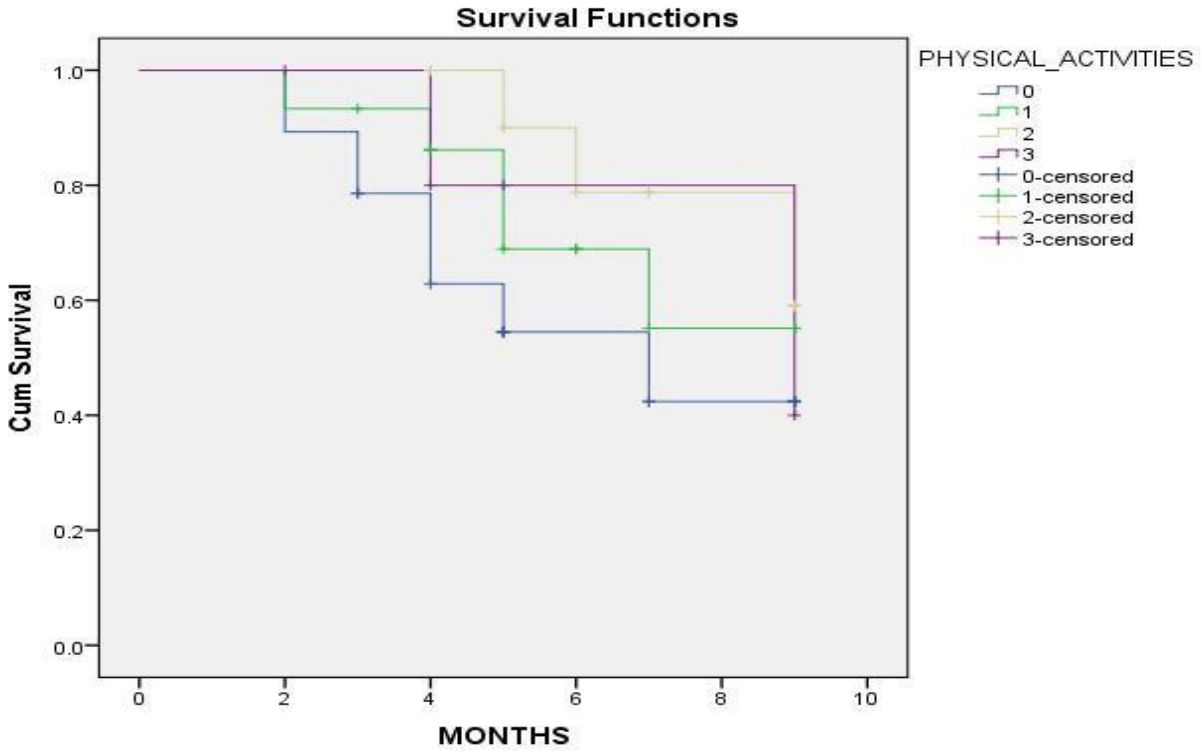
NULL HYPOTHESIS: There is no difference in survival between the groups under investigation.

ALTERNATIVE HYPOTHESIS: There is a significant difference in survival between the groups under investigation.

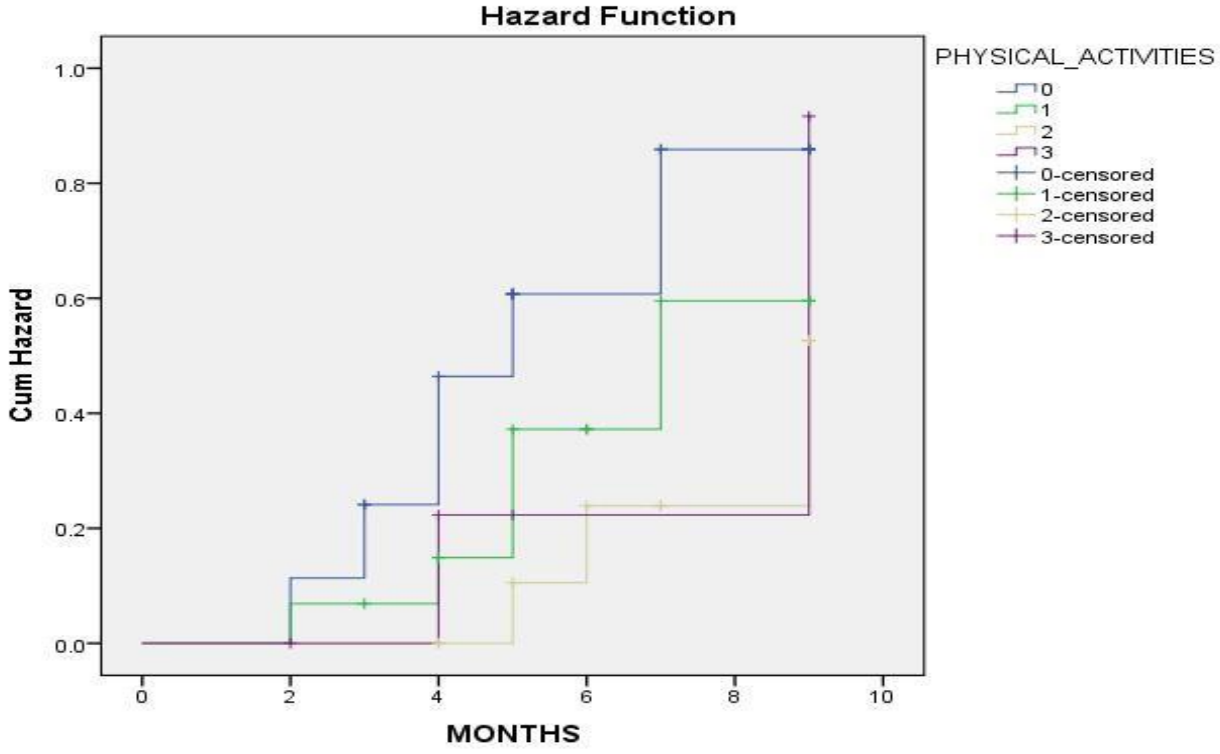
INTERPRETATION:

- P Value is greater than the significance level. So we accept the Null Hypothesis and conclude that there is no difference in survival between the groups under investigation
- Kaplan-Meier curves provide a visual representation of survival probabilities of male is greater compared to female.
- From the Hazard function curve we can conclude that male has least hazard ratio compared to male which indicates that there is a lower risk of obesity for male than females.

KAPLAN MEIER ANALYSIS ON PHYSICAL ACTIVITIES



\



Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	3.688	3	.297

Test of equality of survival distributions for the different levels of Physical Activities.

HYPOTHESIS:

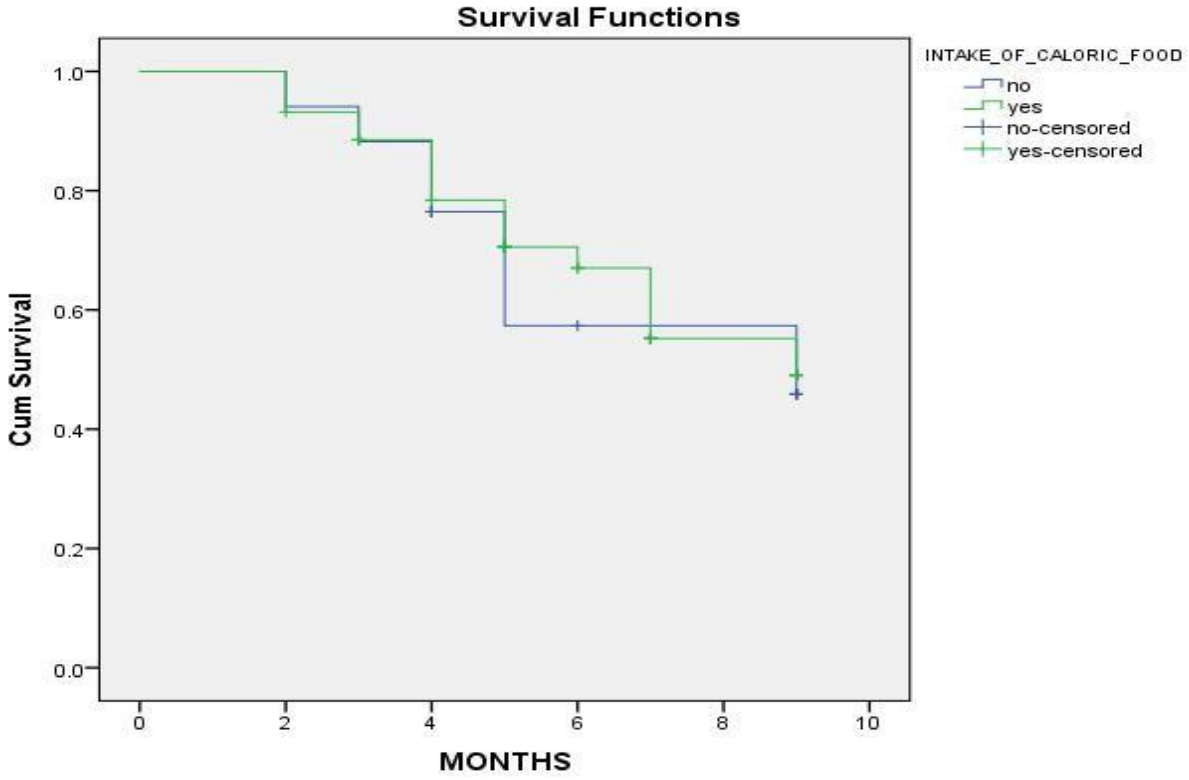
NULL HYPOTHESIS: There is no difference in survival between the groups under investigation.

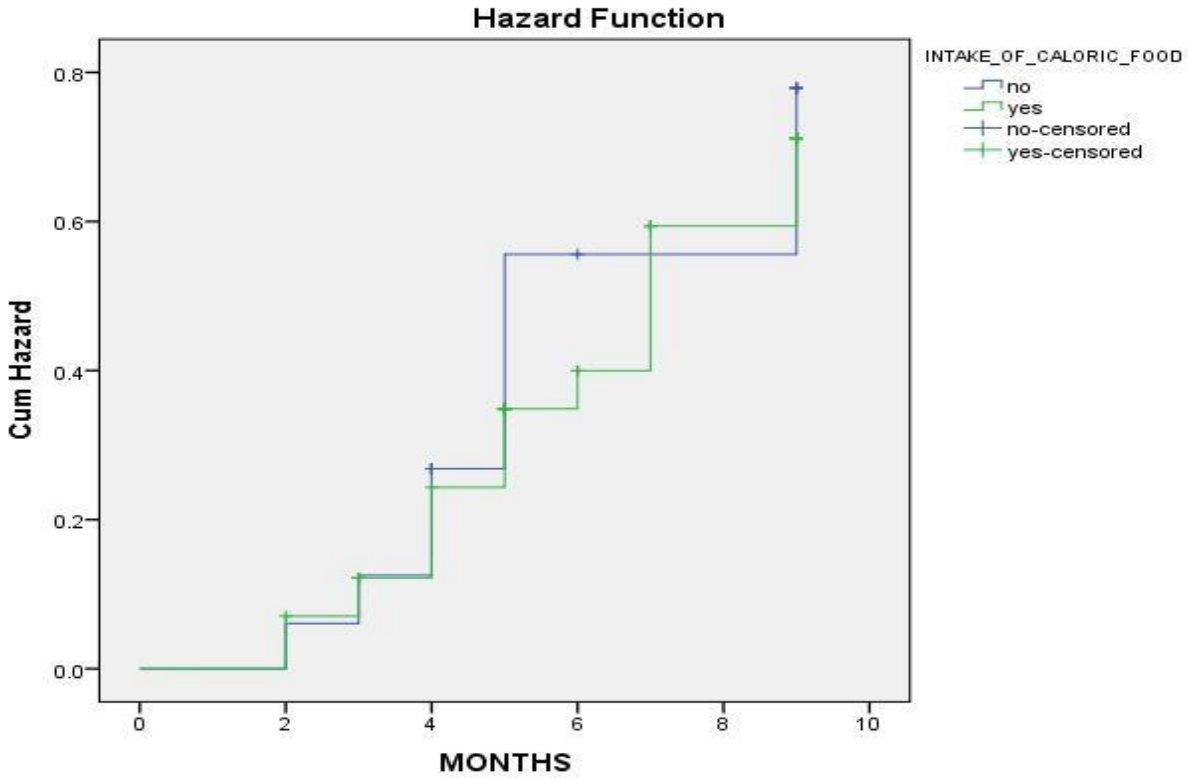
ALTERNATIVE HYPOTHESIS: There is a significant difference in survival between the groups under investigation.

INTERPRETATION:

- P Value is greater than the significance level. So we accept the Null Hypothesis and conclude that there is no difference in survival between different level of physical activities
- Kaplan-Meier curves provide a visual representation of survival probabilities of physical activities, where the survival probability of 2 or days physical activity is greater compared with other factors in physical activity.
- From the Hazard function curve we can conclude that 2 or 4 days physical activity has least hazard ratio compared to other three factors which indicates that there is a lower risk of obesity for those who has 2 or 4 days physical activity.

KAPLAN MEIER ANALYSIS ON INTAKE OF CALORIC FOOD





Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	.070	1	.791

Test of equality of survival distributions for the different levels of Intake of Caloric Food.

HYPOTHESIS:

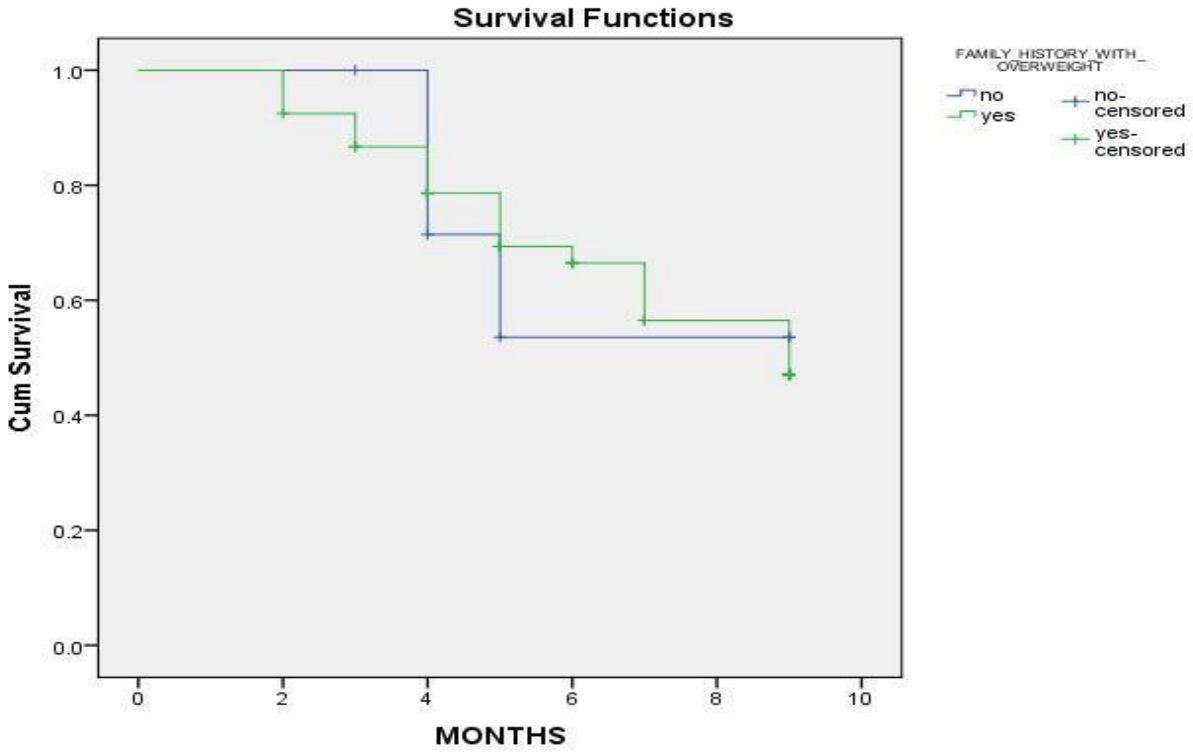
NULL HYPOTHESIS: There is no difference in survival between the groups under investigation.

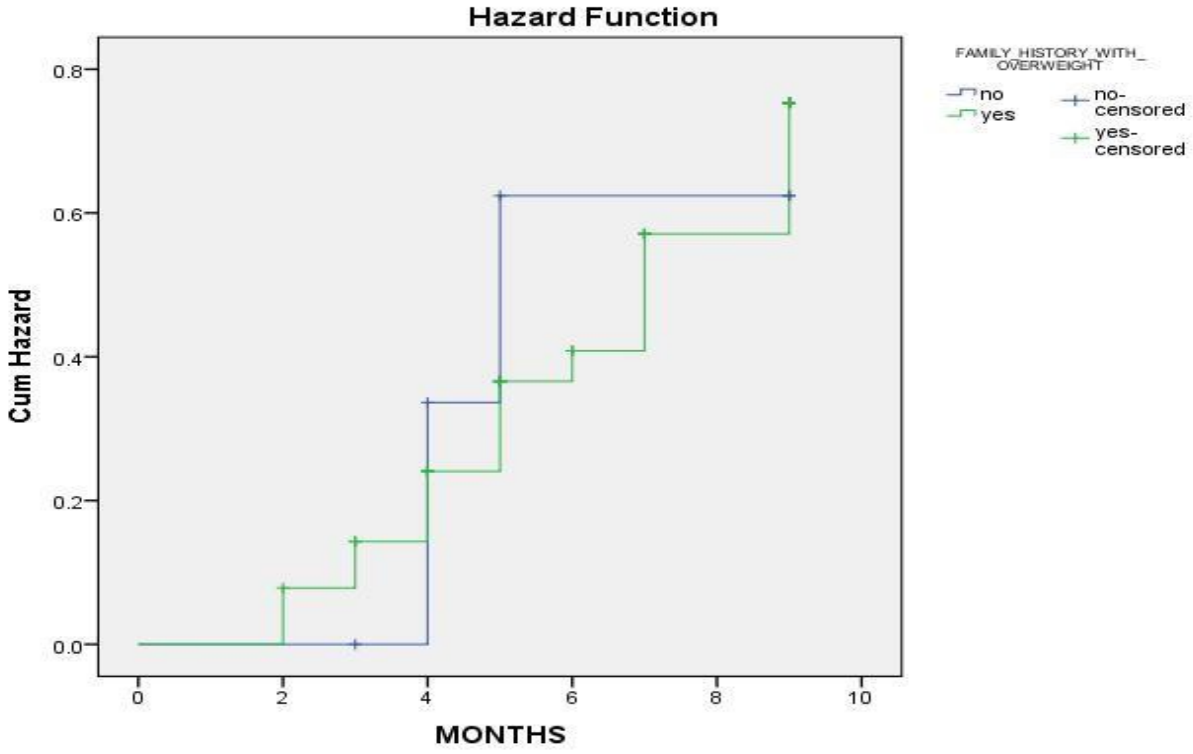
ALTERNATIVE HYPOTHESIS: There is a significant difference in survival between the groups under investigation.

INTERPRETATION:

- P Value is greater than the significance level. So we accept the Null Hypothesis and conclude that there is no difference in survival between the groups under investigation
- Kaplan-Meier curves provide a visual representation of survival probabilities of caloric diet, where the survival probability of caloric diet is greater than the non caloric diet.
- From the Hazard function curve we can conclude that caloric diet has least hazard ratio compared to non caloric diet, which indicates that there is a lower risk of obesity for those who follows caloric diet.

KAPLAN MEIER ANALYSIS ON FAMILY HISTORY WITH OVERWEIGHT





Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	.001	1	.977

Test of equality of survival distributions for the different levels of Family History With Overweight.

HYPOTHESIS:

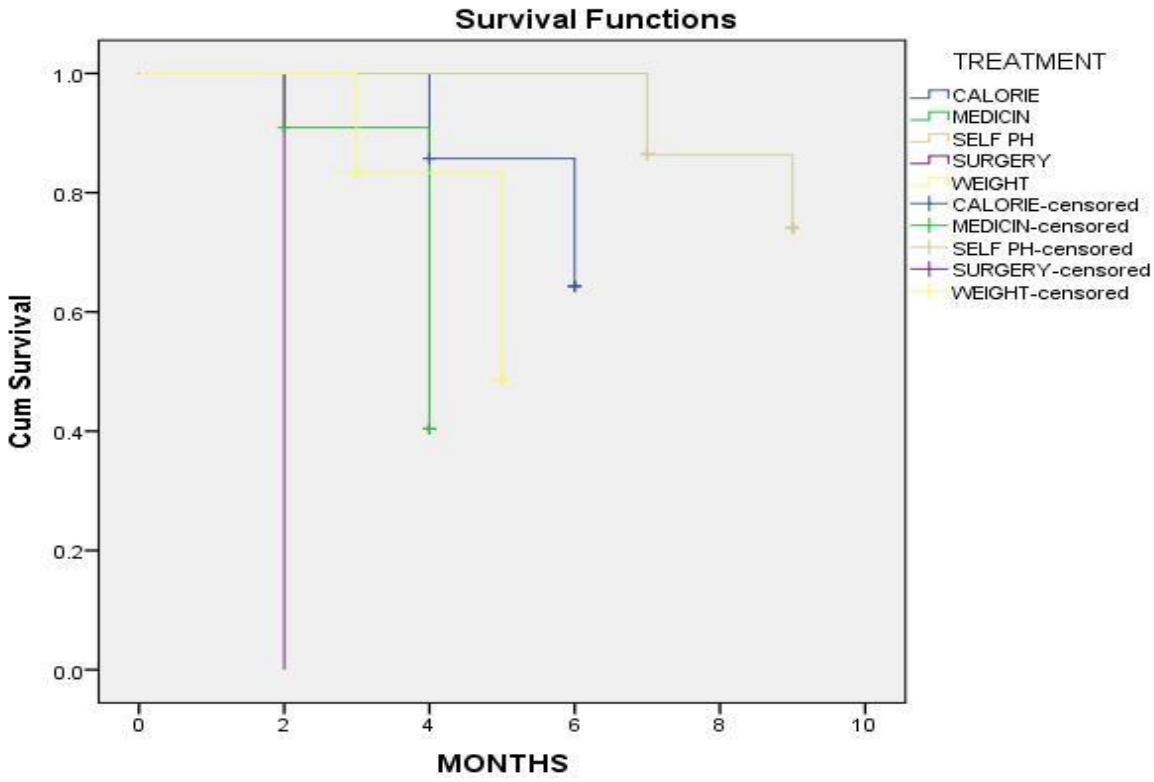
NULL HYPOTHESIS: There is no difference in survival between the groups under investigation.

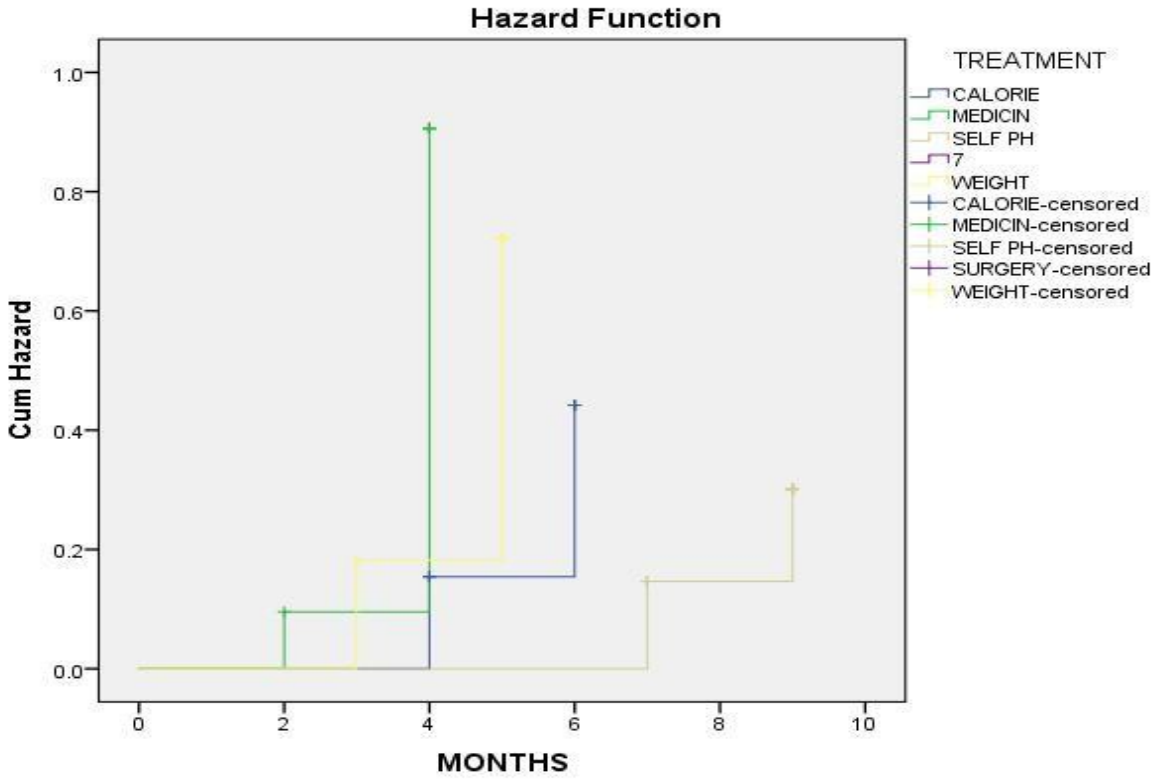
ALTERNATIVE HYPOTHESIS: There is a significant difference in survival between the groups under investigation.

INTERPRETATION:

- P Value is greater than the significance level. So we accept the Null Hypothesis and conclude that there is no difference in survival between the family history with obesity.
- Kaplan-Meier curves provide a visual representation of survival probabilities of family history with obesity where the survival probability of family history with obesity is less compared with family history with no obesity.
- From the Hazard function curve we can conclude that family history with obesity contains greater hazard ratio compared to other, which indicates that there is a higher risk of obesity for those who got family history with obesity.

KAPLAN MEIER ANALYSIS ON TREATMENTS





Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	64.872	4	.000

Test of equality of survival distributions for the different levels of Treatment.

HYPOTHESIS:

NULL HYPOTHESIS: There is no difference in survival between the groups under investigation.

ALTERNATIVE HYPOTHESIS: There is a significant difference in survival between the groups under investigation.

INTERPRETATION:

- P Value is less than the significance level. So we accept the Alternate Hypothesis and conclude that there is a difference in survival between the groups under treatment.
- Kaplan-Meier curves provide a visual representation of survival probabilities of treatments, where the survival probability of self physical activity is greater compared to all other treatments.
- From the Hazard function curve we can conclude that Medicine and supplements has greater hazard ratio compared to other treatments and also we can conclude that medicine and supplements has increased risk and also there is a decreased risk of obesity for those has self physical activity.

CHAPTER VI
CONCLUSION

BAR GRAPH

- Male has more obesity count than female
- No physical activity has more obesity count
- Family history with obesity contains higher count on obesity
- Self physical activity has more number of count than any other treatment and obesity patients also undergone this treatment
- Age interval between 20-30 many have chosen self physical activity as their treatment
- Outcome also has more count in the age interval of 20-30
- Outcome rate of male is greater than female
- The outcome of Non abstinent also has a greater ratio of self physical activity

CORRELATION

From our analysis there is indirect relationship (negative correlation) between Obesity and Age

- Age (-0.141)
- Outcome (-0.669)
- Smoke (-0.065)
- Family History with Overweight (-0.137)

From our analysis there exist direct relationship (positive correlation) between Age and outcome

- Outcome (0.239)
- Alcohol (0.212)
- Intake of Caloric food (0.024)

CHI SQUARE

From analysis on chi square test there is significant association between

- Smoking and Obesity
- Alcohol Consumption and Obesity

- Physical Activity and Obesity
- Family History with Overweight and Obesity
- Caloric Food and Obesity
- H2O and obesity
- Transportation and obesity

ONE WAY ANOVA

There is no Significant difference between the mean new weights among the treatment groups.

LOGISTIC REGRESSION

Family history with overweight, Intake of caloric food, Smoke has higher odds of event happening of Obesity.

PAIRED T TEST

The weight of the patients before and after the treatment has a significant differences.

COX REGRESSION

Family history with overweight and intake of caloric food has increased risk of Obesity and the physical activity has lower risk of Obesity.

KAPLAN MEIER

- Frequent alcohol consumption has less probability of Survival
- The survival probability of male is greater than female
- Family history with overweight has less survival probability and also has a risk of Obesity
- The survival probability of self physical activity is greater compared to all other treatments.

CHAPTER VII
SUGGESTION

- ❖ Consult with a healthcare professional for personalized advice.
- ❖ Adopt a balanced diet with whole foods, fruits, vegetables, and lean proteins.
- ❖ Engage in regular physical activity that you enjoy.
- ❖ Monitor portion sizes and practice mindful eating.
- ❖ Stay hydrated with water instead of sugary drinks.
- ❖ Get enough sleep for overall well-being.
- ❖ Consider seeking support from a healthcare team or a registered dietitian.

REFERENCES:

https://pro.novonordisk.com/disease-area/obesity.html?cid=pse-u19aih75o0&gclid=Cj0KCQiAnfmsBhDfARIsAM7MKi35biFpHwwdm4MJU1NfSe2Vadk-cLQEXZFaTRA0b8FVyf6wEwkWzZAaAqcKEALw_wcB

https://www.ncbi.nlm.nih.gov/books/NBK476331/

https://www.sciencedirect.com/science/article/pii/S0010482521005485

https://www.nature.com/ijo/

https://pro.novonordisk.com/disease-area/obesity.html?cid=pse-u19aih75o0&gclid=Cj0KCQiAnfmsBhDfARIsAM7MKi06cQkbl59IWGWJclXeZnLSwPAXVMKgJOJAIC5AAlsUumzrsoCYRLoaAjM8EALw_wcB

https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(17)32129-3/fulltext

https://nutritionj.biomedcentral.com/articles/10.1186/1475-2891-3-3

https://www.cancer.gov/about-cancer/causes-prevention/risk/obesity/obesity-fact-sheet

https://www.oecd-ilibrary.org/sites/641a2e79-en/index.html?itemId=/content/component/641a2e79-en

https://www.mayoclinic.org/diseases-conditions/obesity/symptoms-causes/syc-20375742