Heart Health Analysis: Exploring Factors Influencing Cardiovascular Health and Mortality

## Project Objectives:

1. **Identify Key Risk Factors**: Determine which variables are significantly associated with the incidence of coronary heart disease (CHD) and other heart-related issues.
2. **Assess Mortality Causes**: Explore the distribution of causes of death and identify trends based on demographic factors.
3. **Examine Relationships Among Variables**: Investigate correlations and relationships among key variables like age, cholesterol, blood pressure, smoking, and weight.
4. **Make Recommendations**: Provide recommendations for improving heart health based on the analysis.

## Data Preparation:

### Data Import and Cleaning:

1. Load the sashelp.heart dataset into the SAS environment.
2. Check for missing values and impute or handle them as needed.
3. Convert categorical variables to appropriate formats for analysis.

### Data Exploration:

1. Examine the distribution of key variables using summary statistics and visualisation techniques (e.g., histograms, box plots, bar charts).
2. Identify potential outliers and examine their impact on analysis.

## Statistical Analyses:

### Descriptive Statistics:

1. Calculate means, medians, standard deviations, and ranges for numeric variables.
2. Summarise categorical variables with frequencies and percentages.

### Correlation Analysis:

1. Examine correlations between numerical variables to understand their relationships (e.g., correlation matrix).
2. Use scatter plots to visualise correlations between variables like age, cholesterol, blood pressure, etc.

### Comparative Analysis:

1. Use t-tests or ANOVA to compare means across groups (e.g., comparing cholesterol levels between smokers and non-smokers).
2. Use chi-square tests for categorical comparisons (e.g., comparing smoking rates between genders).

### Survival Analysis:

1. Conduct survival analysis to explore factors affecting mortality.
2. Use Kaplan-Meier plots to visualise survival curves based on different factors like gender, smoking status, and cholesterol levels.
3. Perform Cox proportional hazards regression to identify significant predictors of survival.

### Logistic Regression:

1. Build logistic regression models to predict the probability of developing CHD or other heart-related conditions based on significant predictors (e.g., age, blood pressure, smoking, cholesterol).

## Visualisation and Reporting:

### Visualisation:

1. Use graphs and charts to illustrate key findings, such as bar plots for categorical variables and scatter plots for correlations.
2. Create a heatmap for the correlation matrix to visualise relationships between variables.
3. Generate survival curves to depict differences in survival rates across groups.

### Reporting:

1. Prepare a summary report that outlines the key findings from the analyses.
2. Discuss any significant predictors or correlations and what they might indicate about heart health.
3. Include visualisations to support the findings and provide insights into trends or patterns.

## Conclusion and Recommendations:

Based on the analyses, summarise key conclusions, focusing on:

1. Factors that are strongly associated with increased risk of heart disease or mortality.
2. Recommendations for improving heart health based on the findings (e.g., lifestyle changes, monitoring certain health metrics).
3. Suggestions for further research or data collection to deepen understanding of heart health.