Project 1 – Descriptive Epidemiology and Population Characterization

Overview of project:

This project focuses on descriptive epidemiology to characterize the study population using key demographic and clinical variables such as age, sex, cancer type, and treatment patterns. The goal is to generate summary statistics, explore data distributions, and create visualizations to understand the patient population better.

Objectives:

Perform data cleaning and preparation to ensure dataset integrity

Compute descriptive statistics (mean, median, standard deviation, proportions)

Analyze demographic patterns (age distribution, sex ratios)

Examine treatment regimens across different cancer types

Generate interactive visualizations to summarize findings

Dataset: This is a fake dataset created by me using Microsoft excel sheet.

Name: RWE\_Oncology\_Epidemiology\_Study

Key Variables:

Age: Continuous (in years)

Sex: Categorical (Male, Female)

Cancer\_Type: Categorical (e.g., Breast, Lung, Colorectal)

Treatment\_Type: Categorical (e.g., Chemotherapy, Immunotherapy, Surgery)

Methods & Analysis

Data Cleaning & Preparation

Remove duplicate records and handle missing data

Standardize categorical values (e.g., treatment names)

Convert variables to appropriate data types (e.g., Age as numeric)

Descriptive Statistics

Central Tendency: Mean, median, mode

Dispersion: Standard deviation, interquartile range

Frequency Distribution: Cancer type, treatment type

Data visualization

Histogram: Age Distribution of Patients

Bar Chat: Cancer type frequencies

Pie chart: Proportion of treatment types

Box plot: Age distribution by treatment group

Project 2 – Survival Analysis and Treatment Patterns

**Overview**

This project applies survival analysis techniques to assess the effectiveness of different treatment regimens (e.g., chemotherapy vs. immunotherapy) in cancer patients. Using Kaplan-Meier survival curves and Cox proportional hazards models, the study evaluates survival outcomes while accounting for censored data.

**Objectives**

* Estimate survival probabilities using Kaplan-Meier curves
* Compare treatment groups based on survival time distributions
* Model hazard ratios using Cox regression to identify significant predictors of survival
* Visualize survival patterns across treatment types

**Key Variables:**

* Time\_of\_Death: Survival time in days
* Censor: Indicator for censored data (1 = censored, 0 = event occurred)
* Treatment\_Type: Categorical (Chemotherapy, Immunotherapy, Surgery)
* Cancer\_Type: Categorical (e.g., Breast, Lung, Colorectal)
* Age: Continuous (years)

Project 3 – Comparative Effectiveness Research

**Overview**

This project conducts comparative effectiveness research (CER) to evaluate the impact of different treatment options on cancer outcomes. By applying logistic regression and hypothesis testing, the study identifies whether chemotherapy or immunotherapy leads to better patient outcomes.

**Objectives**

* Compare treatment outcomes using statistical models
* Assess treatment effectiveness with logistic and linear regression
* Test hypotheses using t-tests, ANOVA, and chi-square tests
* Generate insights to inform clinical decision-making