**Outline for presentation**

[Slide 1: Title Slide]

Title: Understanding Heart Disease Mortality Rates in the United States

Subtitle: Analyzing Gender, Ethnicity, and State Impacts

Presenter:

Date:

[Slide 2: Introduction]

Heart disease is a significant public health concern in the United States.

CDC reports show it as the leading cause of death, surpassing cancer and COVID-19.

Mortality rates continue to rise, reaching 20% of all deaths in 2021.

Our study focuses on heart disease mortality rates in individuals aged 35 and above in 2014.

[Slide 3: Dataset Overview]

Dataset includes mortality rates per 100,000 people for each county.

Demographic variables: gender, race, and geographical coordinates.

Our goal: Investigate how gender, race, and location impact heart disease mortality.

[Slide 4: Data Cleaning/Preparation]

Initial dataset contained 59,077 rows and 19 columns.

Removed invalid data and 'overall' tag rows.

Focus on individual-level data for improved analysis.

Outliers identified and removed using interquartile range.

[Slide 5: Exploratory Data Analysis]

Mean heart disease mortality rate: 347 individuals per 100,000 population.

Data exhibits normal distribution, suitable for testing and modeling.

Visualizations reveal distribution patterns and outliers.

[Slide 6: Chi-Square Test Results]

Gender and ethnicity show significant associations with mortality rates.

No association found for county or state (based on chi-square tests).

Additional z-tests confirm significant impact of gender and ethnicity.

[Slide 7: Model Selection]

Models considered: linear regression, logistic regression, clustering, naïve Bayes.

Chose linear regression and clustering for their suitability with our data.

[Slide 8: Hypotheses Testing - Gender]

Hypothesis 1: Significant difference in mortality rates between genders.

Linear regression results support rejection of null hypothesis.

Females have lower mortality rates compared to males.

[Slide 9: Hypotheses Testing - Ethnicity]

Hypothesis 2: Significant difference in mortality rates between ethnicities.

Multilinear regression shows significant impacts for Hispanic and Asian/Pacific Islander.

Results align with z-tests, indicating lower mortality rates for these ethnicities.

[Slide 10: Hypotheses Testing - Gender & Ethnicity Interaction]

Hypothesis 3: Significant interaction effect between gender and ethnicity.

Multilinear regression reveals combined effects on mortality rates.

Findings highlight the complexities of gender and ethnicity interactions.

[Slide 11: Hypotheses Testing - State]

Hypothesis 4: Significant difference in mortality rates between states.

Linear regression results suggest some states have impacts, but further testing needed.

Potential implications for public health policies at state level.

[Slide 12: Clustering Analysis]

Clustering reveals patterns aligned with earlier findings.

Asians show lowest mortality rates, consistent with other analyses.

[Slide 13: Conclusion & Recommendations]

Significant impacts observed within gender and ethnicity categories.

Females, Asian/Pacific Islanders, and Hispanics show lower mortality rates.

Further exploration needed on variables with significant impacts.

Recommendations for future research and public health interventions.

[Slide 14: Thank You]

Questions?

[Slide 15: References]