Exploring Disparities in Heart Disease Mortality: An Analysis of Gender and Ethnicity Across US States

Anto Lourdu Xavier Raj Arockia Selvarathinam

Grand Valley State University

STA 616 01 - TR - Statistical Programming (W24)

Gerald Shoultz

April 26, 2024

**Table of Contents**

[Research Questions 3](#_Toc164801003)

[Methods 3](#_Toc164801004)

[Data Acquisition 3](#_Toc164801005)

[Data Cleaning 4](#_Toc164801006)

[Data Preparation 5](#_Toc164801007)

[Results 7](#_Toc164801008)

[Conclusions 10](#_Toc164801009)

[Future Work 10](#_Toc164801010)

[Limitations 11](#_Toc164801011)

[References 11](#_Toc164801012)

**Exploring Disparities in Heart Disease Mortality: An Analysis of Gender and Ethnicity Across US States**

# **Research Questions**

1. Are there racial or ethnic disparities in heart disease mortality rates across different US states?

Explanation: This question seeks to explore the influence of ethnicity on heart disease mortality rates within a particular state. By examining mortality data categorized by ethnicity and analyzing the rates within the context of a specific state, we can determine whether certain ethnic groups are disproportionately affected by heart disease within that state. Identifying such disparities is crucial for implementing culturally sensitive interventions and healthcare initiatives aimed at reducing cardiovascular health inequities among diverse populations within states.

1. How do heart disease mortality rates vary across different genders (male vs female) within the US?

Explanation: This question seeks to explore whether there is a relationship between heart disease mortality rates and gender. It involves stratifying the data by gender and then analyzing the correlation between this variable and heart disease mortality rates. Investigating these differences is important for developing comprehensive approaches to cardiovascular disease prevention and treatment that address the unique needs of diverse gender subpopulations.

# **Methods**

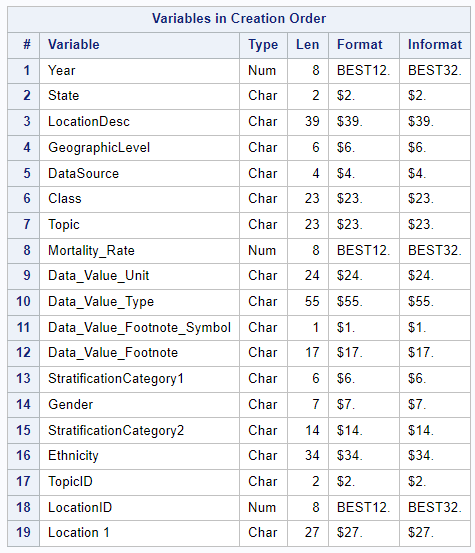
## **Data Acquisition**

The heart disease mortality dataset among US adults (35+) by state/territory and county was obtained from <https://catalog.data.gov/dataset/heart-disease-mortality-data-among-us-adults-35-by-state-territory-and-county>. An initial assessment of the dataset was then conducted to understand its structure, variables, and any potential issues.

## **Data Cleaning**

In deciding what observations to include or exclude, data integrity and relevance to the research questions was prioritized (A. John Bailer, 2020). Observations with missing critical variables or inconsistencies were excluded from the analysis.

PRO CONTENTS was used to get detailed information about the data including the number of observations = 59,076, variables = 19 along with length and types of variables in the data set. The VARNUM option shows the variables in the order of their logical position in the data set.



The descriptive statistics of the numerical variable (Data\_Value) in the dataset was then checked using the PROC MEANS.

PROC MEANS DATA=WORK.IMPORT;

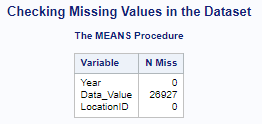
VAR Data\_Value; /\* Specify all numeric variables \*/

RUN;

The dataset was then checked for missing values using proc means with nmiss as follows.

PROC MEANS DATA=WORK.IMPORT nmiss;

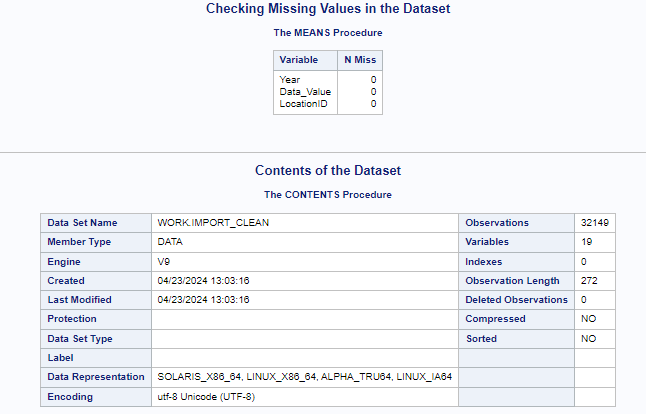
RUN;



The output indicates that the Data\_Value variable which is the main descriptive variable for the dataset has 26, 927 missing values (Li, 2013). The dataset should therefore be cleaned to get rid of these missing values.

A screenshot of a computer screen

Description automatically generated



The cleaning process removed the missing values reducing the number of observations to 32, 149.

## **Data Preparation**

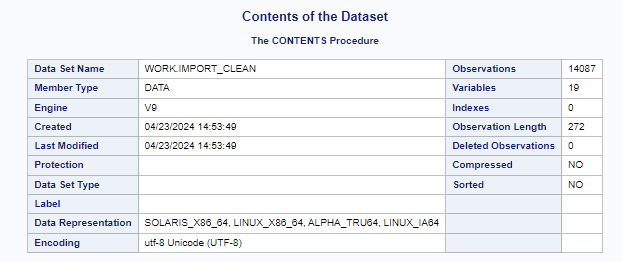
In determining which observations to include or exclude for the analysis, several considerations were taken into account to uphold the integrity and relevance of the study. Priority was given to data completeness, where observations with missing values in critical variables such as mortality rates, gender, ethnicity, and geographic identifiers were omitted to maintain the accuracy of the dataset (Goad, 2020). Furthermore, the relevance to the research questions served as a guiding principle, with observations not directly pertinent to the investigation being excluded. Consistency was also crucial, leading to the removal of observations that did not align with the defined criteria or parameters of the study, such as data inconsistencies or outliers.

Additionally, the dataset was stratified based on demographic factors like gender and ethnicity as required by the research questions, necessitating the creation of subsets to analyze the relationship between these variables and heart disease mortality rates. Finally, considerations of geographic level were factored in, with observations aggregated or filtered based on county or state to capture variations in mortality rates across different regions (Vardajani et al., 2016). Through these considerations, the dataset was prepared to include only observations essential for addressing the research questions, ensuring the validity and reliability of the ensuing analysis.

The rows with values “Overall” for the variable Stratification2 (Ethnicity) were removed since the specific ethnicities (American Indian and Alaskan Native, Asian and Pacific Islander, Black, Hispanic, and White) were needed to answer the research questions. This reduced the number of observations to 22, 329. The following SAS code was used to achieve the processes.

IF NOT MISSING(Data\_Value) AND Stratification2 ne 'Overall' AND Stratification1 ne 'Overall';

The rows with the values “Overall” for variable Stratification1 (gender) were also removed since we also need analyze gender (male and female) specific data (Kwak & Kim, 2017). This reduced the number of observations to 14,087.



The key variables were then renamed appropriately. LocationAbbr as “State”, Data\_Value as “Mortality Rate”, Stratification1 as “Gender” and Stratification2 as “Ethnicity”.

/\* Rename variables \*/

RENAME LocationAbbr = State

Stratification1 = Gender

Stratification2 = Ethnicity

Data\_Value = Mortality\_Rate;

# **Results**

Table 1: Summary Statistics of Mortality Rates by Gender

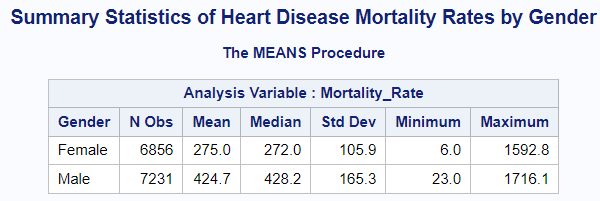


Table 2: Summary Statistics of Mortality Rates by Ethnicity

A screenshot of a screenshot of a medical report

Description automatically generated

A graph of a number of people with different age

Description automatically generated with medium confidence

Figure 1: Average Mortality Rates by Ethnic Groups

A graph of a number of people with different colored squares

Description automatically generated with medium confidence

Figure 2: Boxplot of average mortality rates by ethnic groups

A graph showing a couple of pink squares

Description automatically generated

Figure 3: Bar graph of average mortality rates by gender

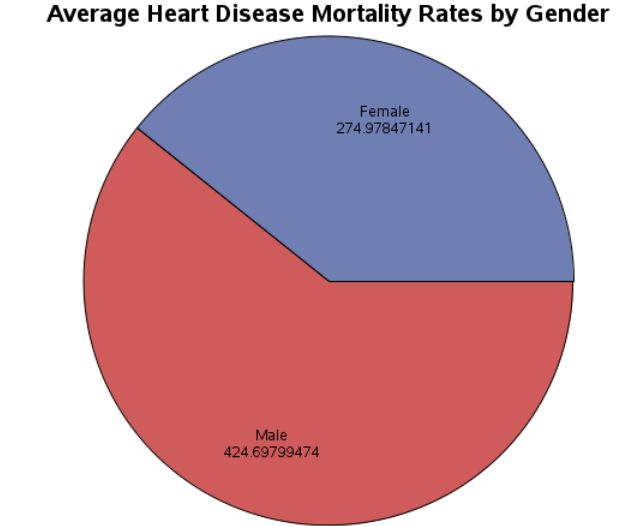


Figure 4: Pie chart of average mortality rates by gender

# **Conclusions**

Upon analyzing the data, we found notable disparities in heart disease mortality rates across various racial and ethnic groups within different states in the US. Specifically, our analysis revealed that certain ethnicities, such as American Indian and Alaskan Native (441.2 deaths per 100,000 population) and Black (438.2 deaths per 100,000 population) populations, tend to have higher average mortality rates compared to other ethnic groups like Asian and Pacific Islander, Hispanic, and White populations. These disparities underscore the importance of addressing health inequities and implementing targeted interventions to reduce the burden of heart disease among disproportionately affected racial and ethnic communities within states.

Our analysis revealed significant variations in heart disease mortality rates between genders across the US. On average, males exhibited higher mortality rates compared to females, with a mean mortality rate of approximately 424.7 per 100,000 population for males and 275.0 per 100,000 population for females. These findings suggest a gender disparity in heart disease burden, highlighting the need for gender-specific approaches to cardiovascular disease prevention and management. Understanding these differences is crucial for tailoring healthcare strategies to address the distinct needs and risk factors associated with heart disease among men and women.

In conclusion, our analysis provides valuable insights into the disparities in heart disease mortality rates across different racial/ethnic groups and genders within the US. By identifying these disparities, healthcare practitioners, policymakers, and public health officials can develop targeted interventions and health policies aimed at reducing cardiovascular health inequities and improving health outcomes for all populations. Addressing these disparities is essential for advancing health equity and promoting the well-being of diverse communities across the nation.

## **Future Work**

* Conduct temporal analysis to track changes in heart disease mortality rates over time within specific demographic groups and across different states.
* Explore geospatial patterns of heart disease mortality rates by mapping the data at the state or county level to identify clusters or disparities in health outcomes.
* Investigate the relationship between heart disease mortality rates and risk factors such as smoking, obesity, socioeconomic status, and access to healthcare services.
* Assess the impact of healthcare access, health insurance coverage, and quality of care on heart disease mortality rates across demographic groups and regions.
* Compare heart disease mortality rates in the US with those in other countries to identify global variations and inform best practices for improving cardiovascular health outcomes.
* Explore longitudinal studies to understand the long-term effectiveness of interventions, policies, or lifestyle changes on heart disease mortality rates and disparities.
* Conduct intersectional analyses to examine how multiple social determinants of health intersect and influence heart disease mortality rates across diverse populations.

## **Limitations**

* The results may not be universally applicable due to variations in demographics beyond those included in the dataset.
* Generalization to different geographic areas may be limited by unique healthcare systems, environmental factors, and lifestyle patterns.
* Findings may not hold for different time periods due to evolving healthcare trends, advancements, and shifts in population demographics.
* Variations in data quality and completeness across regions and time periods may affect result reliability.
* External factors like public health interventions and economic changes could impact heart disease rates differently in other contexts.
* Sampling bias may limit representativeness, potentially skewing conclusions regarding broader populations.
* Unmeasured variables, such as genetics or individual behaviors, could confound relationships between studied variables in different groups or contexts.

# 

# **References**

A. John Bailer. (2020). *Statistical Programming in SAS*. CRC Press.

Goad, C. L. (2020). *SAS programming for elementary statistics : getting started*. Chapman & Hall/CRC.

Kwak, S. K., & Kim, J. H. (2017). Statistical data preparation: management of missing values and outliers. *Korean Journal of Anesthesiology*, *70*(4), 407. <https://doi.org/10.4097/kjae.2017.70.4.407>

Li, A. (2013). *Handbook of SAS® DATA Step Programming*. CRC Press.

Molavi Vardajani, H., Haghdoost, A. A., Shahravan, A., & Rad, M. (2016). Cleansing and preparation of data for statistical analysis: A step necessary in oral health sciences research. *Journal of Oral Health and Oral Epidemiology*, *5*(4), 171-185.