Part Two

Anto Lourdu Xavier Raj Arockia Selvarathinam

Grand Valley State University

STA 616 01 - TR - Statistical Programming (W24)

Gerald Shoultz

April 26, 2024

**Part Two**

**Personal Reflection**

Working on this project provided me with valuable insights into the complexities of analyzing health disparities and the importance of addressing them. It was both challenging and rewarding to delve into the data and uncover meaningful patterns and trends related to heart disease mortality rates.

One of the most significant learnings from this project was the importance of data cleaning and preparation. I gained a deeper understanding of how critical it is to ensure data integrity and relevance to the research questions. Identifying missing values, inconsistencies, and outliers required careful attention to detail, but it was essential for maintaining the accuracy of the analysis.

The project also highlighted the significance of collaboration and communication, even when working individually. Although I worked on the project alone, I constantly sought feedback from peers and instructors to validate my approach and interpretations. Their insights helped me refine my analysis and strengthen my conclusions.

Throughout the project, I encountered various challenges, particularly in interpreting the results and drawing meaningful conclusions. Analyzing disparities in heart disease mortality rates across gender and ethnicity required a nuanced understanding of social determinants of health and healthcare disparities. It was a complex topic that demanded critical thinking and interdisciplinary knowledge.

Despite the challenges, working on this project was immensely rewarding. It allowed me to apply statistical programming techniques in a real-world context and contribute to the growing body of research on health disparities. Moreover, it underscored the importance of leveraging data-driven approaches to inform public health interventions and policies aimed at reducing cardiovascular health inequities.

In terms of group dynamics, even though I worked individually, I made an effort to engage with my peers and offer assistance whenever possible. I shared my insights and experiences with data cleaning and analysis during class discussions and collaborated with classmates on related assignments. Additionally, I actively participated in peer review sessions, providing constructive feedback to help improve the quality of my peers' work.

Overall, this project provided me with a valuable opportunity to deepen my understanding of health disparities and refine my skills in statistical programming and data analysis. It reinforced the importance of interdisciplinary collaboration and data-driven decision-making in addressing complex public health challenges.

**SAS Code**

/\* Import the dataset \*/

FILENAME REFFILE '/home/u63739604/Project.csv';

/\* Import the dataset \*/

**PROC** **IMPORT** DATAFILE=REFFILE

DBMS=CSV

OUT=WORK.IMPORT

REPLACE; /\* Add REPLACE option to overwrite the dataset if it already exists \*/

GUESSINGROWS=MAX;

GETNAMES=YES;

**RUN**;

/\* Remove rows with missing values for Data\_Value variable \*/

/\* Remove rows where Stratification1 and Stratification 2 is "Overall" \*/

**DATA** WORK.IMPORT\_CLEAN;

SET WORK.IMPORT;

IF NOT MISSING(Data\_Value) AND Stratification2 ne 'Overall' AND Stratification1 ne 'Overall'; /\* Exclude rows where Data\_Value is missing or Stratification2 is 'Overall' or Stratification1 is 'Overall' \*/

/\* Rename variables \*/

RENAME LocationAbbr = State

Stratification1 = Gender

Stratification2 = Ethnicity

Data\_Value = Mortality\_Rate;

**RUN**;

/\* Use PROC MEANS to calculate descriptive statistics for Mortality\_Rate variable \*/

TITLE 'Descriptive Statistics of Mortality\_Rate Variable (Excluding Missing Values)';

**PROC** **MEANS** DATA=WORK.IMPORT\_CLEAN maxdec=**1**;

VAR Mortality\_Rate; /\* Specify the variable \*/

**RUN**;

/\* Additional output \*/

TITLE 'Checking Missing Values in the Dataset';

**PROC** **MEANS** DATA=WORK.IMPORT\_CLEAN NMISS;

**RUN**;

TITLE 'Contents of the Dataset';

**PROC** **CONTENTS** DATA=WORK.IMPORT\_CLEAN VARNUM;

**RUN**;

/\* Calculate descriptive statistics for Mortality Rate by state \*/

**PROC** **MEANS** DATA=WORK.IMPORT\_CLEAN NWAY noprint maxdec=**1**;

CLASS State;

VAR Mortality\_Rate;

OUTPUT OUT=State\_DescriptiveStats

MEAN=Mean\_Mortality\_Rate

MEDIAN=Median\_Mortality\_Rate

STD=Std\_Mortality\_Rate;

**RUN**;

/\* Display the aggregated descriptive statistics in a single table \*/

**PROC** **PRINT** DATA=State\_DescriptiveStats;

TITLE 'Aggregated Descriptive Statistics of Mortality Rate by State (Rounded to 1 Decimal Place)';

VAR State Mean\_Mortality\_Rate Median\_Mortality\_Rate Std\_Mortality\_Rate;

FORMAT Mean\_Mortality\_Rate Median\_Mortality\_Rate Std\_Mortality\_Rate **5.1**; /\* Format variables to display 1 decimal point \*/

**RUN**;

/\* Calculate average mortality rates for each ethnicity \*/

**PROC** **MEANS** DATA=WORK.IMPORT\_CLEAN NOPRINT maxdec=**1**;

CLASS Ethnicity;

VAR Mortality\_Rate;

OUTPUT OUT=Ethnicity\_Avg\_Mortality

MEAN=Avg\_Mortality\_Rate;

**RUN**;

/\* Use PROC MEANS to calculate summary statistics for Mortality\_Rate variable by Gender \*/

TITLE 'Summary Statistics of Heart Disease Mortality Rates by Gender';

**PROC** **MEANS** DATA=WORK.IMPORT\_CLEAN MEAN MEDIAN STD MIN MAX maxdec=**1**;

CLASS Gender;

VAR Mortality\_Rate;

**RUN**;

/\* Use PROC MEANS to calculate summary statistics for Mortality\_Rate variable by Ethnicity \*/

TITLE 'Summary Statistics of Heart Disease Mortality Rates by Ethnicity';

**PROC** **MEANS** DATA=WORK.IMPORT\_CLEAN MEAN MEDIAN STD MIN MAX maxdec=**1**;

CLASS Ethnicity;

VAR Mortality\_Rate;

**RUN**;

ods graphics / reset;

**proc** **template**;

define statgraph SASStudio.Pie;

begingraph;

entrytitle "Average Heart Disease Mortality Rates by Gender" /

textattrs=(size=**14**);

layout region;

piechart category=Gender response=Avg\_Mortality\_Rate /;

endlayout;

endgraph;

end;

**run**;

ods graphics / reset width=**6.4**in height=**4.8**in imagemap;

**proc** **sgrender** template=SASStudio.Pie data=WORK.AVG\_MORTALITY\_RATE\_BY\_GENDER;

**run**;

ods graphics / reset;

/\* Create bar charts for average heart disease mortality rates by racial/ethnic groups \*/

ods graphics / reset width=**6.4**in height=**4.8**in imagemap;

**PROC** **SGPLOT** DATA=Ethnicity\_Avg\_Mortality;

VBAR Ethnicity / RESPONSE=Avg\_Mortality\_Rate GROUPORDER=DESCENDING fillattrs=(color=CXa0f5e5);

yaxis grid;

TITLE 'Average Heart Disease Mortality Rates by Racial/Ethnic Groups';

YAXIS LABEL='Average Mortality Rate (per 100,000 population)';

LABEL Ethnicity = "Ethnicity"; /\* Assign longer labels to Ethnicity variable \*/

**RUN**;

ods graphics / reset;

/\* Calculate average mortality rate by state and ethnicity \*/

**PROC** **MEANS** DATA=WORK.IMPORT\_CLEAN NWAY NOPRINT maxdec=**1**;

CLASS State Ethnicity;

VAR Mortality\_Rate;

OUTPUT OUT=State\_Ethnicity\_Avg\_Mortality

MEAN=Avg\_Mortality\_Rate;

**RUN**;

/\* Box plots of average heart disease mortality rates by racial/ethnic groups within states \*/

ods graphics / reset width=**6.4**in height=**4.8**in imagemap;

**PROC** **SGPLOT** DATA=State\_Ethnicity\_Avg\_Mortality;

VBOX Avg\_Mortality\_Rate / CATEGORY=Ethnicity fillattrs=(color=CXe6cadf);

yaxis grid;

TITLE 'Box Plots of Average Heart Disease Mortality Rates by Racial/Ethnic Groups';

YAXIS LABEL='Average Mortality Rate (per 100,000 population)';

**RUN**;

ods graphics / reset;

/\* Calculate average mortality rate by Gender \*/

**PROC** **MEANS** DATA=WORK.IMPORT\_CLEAN MEAN noprint maxdec=**1**;

CLASS Gender;

VAR Mortality\_Rate;

OUTPUT OUT=Avg\_Mortality\_Rate\_By\_Gender MEAN=Avg\_Mortality\_Rate;

**RUN**;

/\* Create bar chart for average heart disease mortality rates by Gender \*/

TITLE 'Average Heart Disease Mortality Rates by Gender';

ods graphics / reset width=**6.4**in height=**4.8**in imagemap;

**PROC** **SGPLOT** DATA=Avg\_Mortality\_Rate\_By\_Gender;

VBAR Gender / RESPONSE=Avg\_Mortality\_Rate GROUPORDER=DATA fillattrs=(color=CXf5a0f2);

yaxis grid;

YAXIS LABEL='Average Mortality Rate';

XAXIS LABEL='Gender';

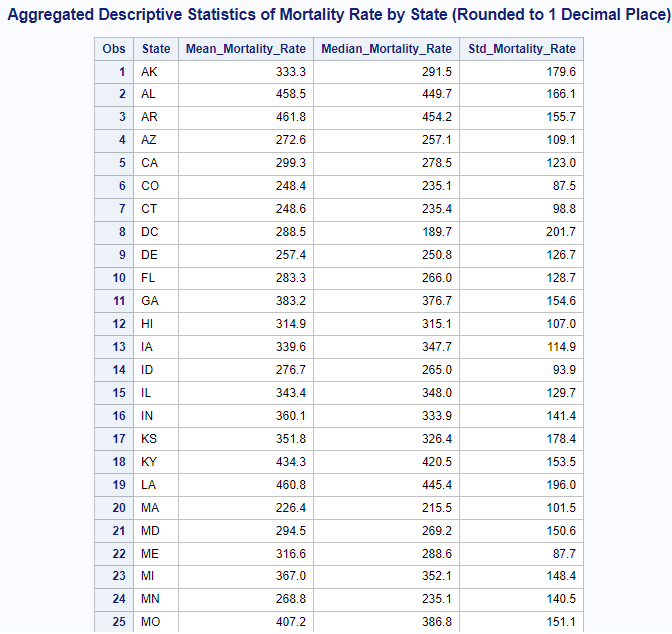
**RUN**;

**Dataset Link:** <https://catalog.data.gov/dataset/heart-disease-mortality-data-among-us-adults-35-by-state-territory-and-county>.

**Appendix**

**A screenshot of a data analysis

Description automatically generated**

****

**A screenshot of a data table

Description automatically generated**