**Understanding Diabetes Mortality Trends in 2020: Implications for Public Health**

**Captain Corgu**

**4/18/2024**

**Introduction**

This paper delves into the analysis of Provisional Diabetes Death Counts for 2020, aiming to uncover vital insights into the trends and implications of diabetes-related mortality within the context of the COVID-19 pandemic. Understanding diabetes mortality rates is crucial for informing effective public health interventions and optimizing healthcare resource allocation. By scrutinizing the intersection between diabetes, COVID-19, and other comorbidities, this analysis seeks to provide actionable intelligence for mitigating the burden of diabetes-related mortality and strengthening healthcare systems.

**Data Collection**

The data for this analysis was sourced from the National Center for Health Statistics (NCHS), a reputable institution known for its comprehensive health-related data collection and analysis. The NCHS provides reliable and standardized mortality statistics, making it an ideal source for studying diabetes mortality trends. The dataset includes several key features relevant to our analysis, such as deaths with COVID-19 as either the underlying or contributing cause (C19PlusDiabetes), and deaths where both COVID-19 and diabetes are listed as causes (C19PlusDiabetes). The target variable for our analysis is the Provisional Diabetes Death Counts for the year 2020, encompassing all deaths attributed to diabetes as either the underlying cause or a contributing cause. This dataset provides a comprehensive foundation for understanding the dynamics of diabetes mortality and its relationship with COVID-19 and other comorbidities.

**Data Preprocessing**

The data preprocessing phase commenced with the initialization of necessary libraries for data manipulation and file operations. Subsequently, the script navigated to the directory containing the dataset file and imported the data into a Pandas DataFrame named DiabetesData. To adhere to best practices, a new DataFrame, df, was created to preserve the original dataset for reference. Redundant columns such as 'Data as of', 'Date\_Of\_Death\_Year', 'Date\_Of\_Death\_Month', 'Diabetes.uc', and 'Diabetes.mc' were systematically eliminated to streamline the dataset for analysis.

To ensure data integrity, a systematic approach was employed to identify and remove rows with redundant or conflicting age groups. This meticulous process aimed to maintain the accuracy and reliability of the dataset. Following data cleaning, the script meticulously calculated the mean values of various death counts associated with diabetes and comorbidities. This involved extracting the relevant death counts from the DataFrame and leveraging statistical functions to compute the mean values. Rows with missing values in specific columns were omitted to ensure data accuracy before computing the mean for each category of deaths.

In the final step of data preprocessing, the script presented the calculated mean values for each category, offering valuable insights into the average mortality rates attributed to different comorbidities associated with diabetes. This methodical data preprocessing procedure adhered to professional standards, ensuring the reliability and accuracy of subsequent analyses conducted on the Diabetes Dataset.

**Exploratory Data Analysis (EDA):**

To provide insights into the diabetes death count data for 2020, summary statistics were computed. These statistics offer a snapshot of the distribution and central tendencies of the data. Additionally, visualizations such as pie graphs and correlation matrices were utilized to further explore patterns and trends in diabetes mortality rates.

Summary statistics revealed the following mean values for death counts associated with various comorbidities in conjunction with diabetes:

* COVID-19 and Diabetes: 279.92
* COVID-19 and Hypertensive Diseases: 370.14
* COVID-19 and Major Cardiovascular Diseases: 785.59
* COVID-19, Hypertensive Diseases, and Major Cardiovascular Diseases: 785.59
* COVID-19 and Chronic Lower Respiratory Disease: 144.82
* COVID-19 and Kidney Disease: 148.47
* COVID-19 and Chronic Liver Disease/Cirrhosis: 14.09
* COVID-19 and Obesity: 59.50

These mean values provide a quantitative understanding of the average mortality rates associated with each comorbidity in relation to diabetes.

Moreover, visualizations such as pie graphs and correlation matrices were employed to uncover additional insights into the relationships between diabetes mortality rates and various comorbidities. Pie graphs illustrate the proportion of diabetes-related deaths attributed to different comorbidities, offering a visual representation of their relative contributions. Correlation matrices, on the other hand, quantify the strength and direction of associations between diabetes mortality rates and other health conditions, providing valuable insights into potential risk factors and interactions. (*Introduction to The Correlation Matrix | Built In*, n.d.)

Model Development: Difference of means

The "difference of means" concept refers to the disparity or variation between the average values (means) of two different groups or categories, quantifying how much one group's mean value differs from another group's mean value. In the context of this analysis, the differences of means offer valuable insights into the comparative averages across various mortality rates of comorbidities of diabetes.

For instance, when comparing "C19PlusDiabetes" to "C19PlusHypertensiveDiseases," the difference of means indicates that individuals with diabetes tend to have approximately 90.23 units lower values than those with hypertensive diseases. Similarly, the comparison between "C19PlusDiabetes" and "C19PlusMajorCardiovascularDiseases" reveals a substantial difference, with individuals with major cardiovascular diseases showing mean values approximately 505.68 units lower than those with diabetes. These comparisons extend across different pairs of health conditions, shedding light on the relative severity or impact of each condition in relation to others.

For example, individuals with chronic liver disease and cirrhosis exhibit mean values approximately 265.83 units higher than those with diabetes, indicating a potentially heightened health burden associated with this condition. Overall, these differences of means provide a quantitative understanding of how various health conditions compare in terms of their average manifestations, facilitating informed decision-making in healthcare policy, resource allocation, and clinical management.

Model Evaluation with ANOVA

The ANOVA test serves as a robust statistical tool to validate the difference of means by assessing the significance of variations between groups. By comparing the variation between groups with the variation within groups, ANOVA determines whether the observed differences in means are statistically significant. In the context of this analysis, the ANOVA results demonstrated significant differences between the groups, indicating substantial variations in mean mortality rates across various comorbidities associated with diabetes. This validation underscores the reliability of the difference of means in quantifying the disparities in mortality rates between different health conditions. Additionally, the identification of significant differences between specific pairs of comorbidities through post-hoc analyses further strengthens the validity of the difference of means concept, providing valuable insights into the comparative severity of each condition in relation to others. (*ANOVA (Analysis of Variance)*, n.d.)

To validate the difference of means and assess the significance of variations between groups, Analysis of Variance (ANOVA) was employed. ANOVA results demonstrated significant differences between the groups (F(8, 2025) = 61.64, p < 0.001), indicating substantial variations in mean mortality rates across various comorbidities associated with diabetes.

Post-hoc analyses using Tukey's Honestly Significant Difference (HSD) test further elucidated the differences between specific groups. The results revealed significant differences between several pairs of comorbidities, including Hypertensive Diseases versus Chronic Liver Disease and Cirrhosis (p = 0.0079), Hypertensive Diseases and Major Cardiovascular Diseases versus Chronic Liver Disease and Cirrhosis (p < 0.001), Hypertensive Diseases and Major Cardiovascular Diseases versus Chronic Lower Respiratory Disease (p < 0.001), Hypertensive Diseases, Major Cardiovascular Diseases, and Obesity versus Chronic Lower Respiratory Disease (p < 0.001), and Hypertensive Diseases, Major Cardiovascular Diseases, Obesity, and COVID-19 versus Chronic Lower Respiratory Disease (p < 0.001). (*13.6*, 2021)

These findings underscore the significance of understanding the variations in mortality rates across different comorbidities associated with diabetes. Such insights are crucial for informing healthcare policy, resource allocation, and clinical management strategies aimed at mitigating the impact of these comorbidities on diabetes-related mortality.

Model Interpretation

The analysis of differences in means provides a nuanced insight into the comparative mortality rates associated with various comorbidities of diabetes. Negative differences, observed between C19PlusDiabetes and conditions such as C19PlusHypertensiveDiseases, C19PlusMajorCardiovascularDiseases, and C19PlusHypertensiveDiseasesAndMCVD, signify that, on average, individuals with diabetes experience lower mortality rates than those afflicted with these specific comorbidities. This suggests a potentially protective effect of diabetes against mortality when compared to these cardiovascular-related conditions. However, we know this is erroneous.

Conversely, positive differences, exemplified by the contrast between C19PlusDiabetes and C19PlusChronicLowerRespiratoryDisease, highlight higher mortality rates associated with chronic lower respiratory disease relative to diabetes. These disparities underscore the diverse impact of different comorbidities on mortality outcomes. Additionally, instances where the difference of means approaches zero, as seen between C19PlusMajorCardiovascularDiseases and C19PlusHypertensiveDiseasesAndMCVD, suggest a negligible difference in mortality rates between these conditions.

The magnitude of differences further elucidates the varying degrees of influence that each comorbidity exerts on mortality rates, with larger differences indicating a more significant impact. Moreover, considering the direction of differences is crucial, as positive differences imply higher mortality rates for the first comorbidity, while negative differences indicate the opposite.

This comprehensive interpretation not only aids in understanding the relative severity of different comorbidities but also informs targeted healthcare interventions aimed at reducing mortality rates among individuals with diabetes. By recognizing the differential impact of various conditions, healthcare practitioners can tailor strategies to address specific health concerns effectively, thereby optimizing patient outcomes and resource allocation in diabetes care.

Conclusion

In conclusion, this study offers valuable insights into the complex dynamics of diabetes-related mortality, uncovering significant variations in mortality rates among different comorbidities conditions but also highlights the heightened risk associated with specific respiratory ailments. These findings have important implications for public health interventions and policy development.

The observed differences in mortality rates underscore the need for tailored healthcare interventions and targeted policy initiatives to address the diverse needs of individuals with diabetes. By understanding the varying risks associated with different comorbidities, policymakers and healthcare practitioners can develop more effective strategies for prevention, early detection, and management of diabetes-related complications. Moreover, addressing socio-economic and demographic factors that contribute to health disparities is crucial for ensuring equitable access to healthcare services and improving outcomes for vulnerable populations.

Furthermore, future research efforts should focus on elucidating the underlying mechanisms driving the observed differences in mortality rates. Longitudinal studies tracking changes in mortality patterns over time and across different populations can provide valuable insights into the evolving nature of diabetes-related mortality. Additionally, exploring the impact of lifestyle factors, genetic predispositions, and healthcare disparities on mortality outcomes can inform the development of targeted interventions to mitigate the burden of diabetes on public health.

By prioritizing research in these areas and translating findings into actionable policies and interventions, stakeholders can work collaboratively to reduce diabetes-related mortality rates, improve patient outcomes, and enhance the overall quality of life for individuals living with diabetes.

**Reference**

National Center for Health Statistics. AH Provisional Diabetes Death Counts for 2020. Date accessed [Last accessed date]. Available from <https://data.cdc.gov/d/qdcb-uzft>.

*13.6: Post‐hoc Analysis – Tukey’s Honestly Significant Difference (HSD) Test85*. (2021, July 15). Statistics LibreTexts. <https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Inferential_Statistics_and_Probability_-_A_Holistic_Approach_(Geraghty)/13%3A_One_Factor_Analysis_of_Variance_(ANOVA)/13.06%3A_Posthoc_Analysis__Tukeys_Honestly_Significant_Difference_(HSD)_Test85>

*ANOVA (Analysis of Variance)*. (n.d.). Statistics Solutions. Retrieved April 18, 2024, from <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/anova/>

*Introduction to The Correlation Matrix | Built In*. (n.d.). Retrieved April 18, 2024, from <https://builtin.com/data-science/correlation-matrix>