Data Analysis | Deliverable 3 | Final Report

Aubrey Reichard-Eline

You must submit a Word document with your final report and an Excel file with your dataset and all your analyses. Make sure to include all relevant tables and charts in your Word report. This final report accounts for 10% of your final grade in the class.

The final report must contain the following sections; use section headers:

- **Introduction:** The introductory narrative from your Deliverable 1. Add one paragraph summarizing your main findings.
- **Descriptive Statistics:** The descriptive statistics and the charts from Deliverable 2. Make sure to include the charts in the Word document (you can copy them from Excel and paste them as images in Word), and to discuss and interpret your charts and calculations.
- **One-sample hypothesis testing:** Identify one variable of interest for which you can build a research hypothesis regarding its mean or a certain proportion. Test this hypothesis using appropriate statistical methods.
- **Two-sample hypothesis testing:** Identify two variables of interest that are directly comparable or one variable split in two groups (female vs male, democrats vs republicans, young vs elderly, depending on the context of your research) and conduct a two-sample test of means (or proportions, depending the case). Make sure to use the appropriate test that corresponds to the nature of your data (independent or dependent samples, known or unknown variance).
- **Correlation:** Conduct at least 2 correlation analyses between your continuous variables. Report and interpret the results.
- **Regression analysis:** Conduct one multiple regression analysis (Review slides for Weeks 13 and 14) with one dependent variable and at least 2 independent variables. Report and interpret the results.
- **Conclusions (500 words):** What are the main insights you have learned from analyzing these data?
- **Final reflection (500 words):** Is data analysis useful for public administrators? Why or why not? How has this class help you reinforce, or change, your understanding of data analysis?

Introduction:

The topic of my research is the prevalence of childhood cancer in the United States. This is a critical area of research with profound implications for many areas such as public health, healthcare policy, and clinical practice. Childhood cancer remains the number one death by disease in children, this warrants more investigation into the causes, diagnosis, treatments, outcomes, and survivorship. Some research questions in this area include clarifying if there are underlying genetic and environmental factors contributing to the

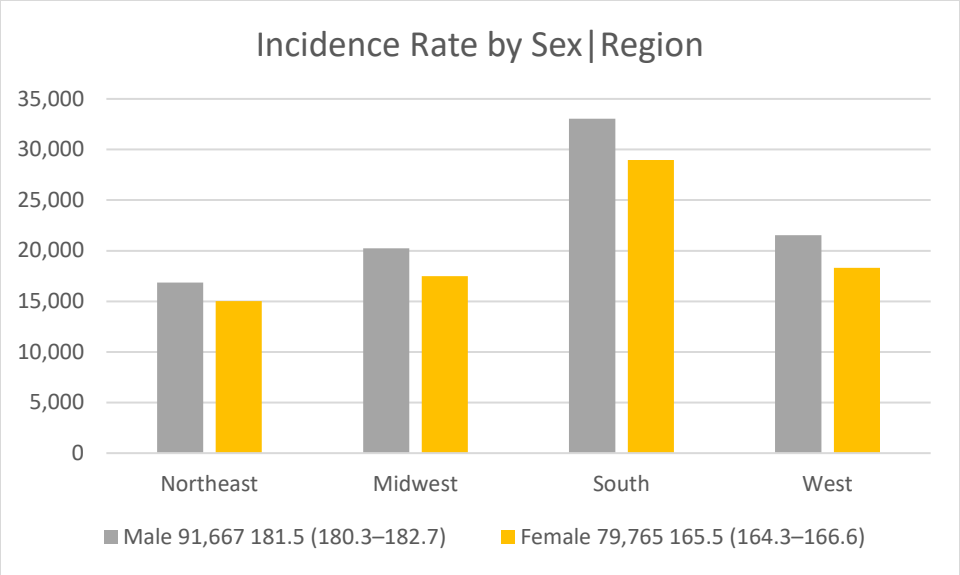
development of childhood cancers, exploring disparities in incidence and outcomes among different demographic groups, assessing protocols to facilitate early detection strategies, and identifying novel therapeutic approaches to improve survival rates and quality of life for children affected by this disease.

Understanding the incidence rates of childhood cancer is relevant on multiple fronts. From one perspective, this would provide insights into the complex correlation of genetic predisposition, environmental exposures, and developmental biology in cancer. This knowledge can inform the development of targeted prevention and intervention strategies. Secondly, from a policy perspective, awareness of trends and patterns in childhood cancer incidence can guide the allocation of resources, public health initiatives, and healthcare planning to better address the needs of those affected. Investigating disparities in incidence rates can highlight imbalances in access to care services and inform ways to promote health equity. Research in this area has the potential to significantly impact clinical practice, public health policy, and the lives of children and families affected by cancer. The knowledge of the geographic incidence rates of childhood cancer may then be used to enhance awareness for providers, increase treatment capacities and facilitate better survivorship care and cancer surveillance overall.

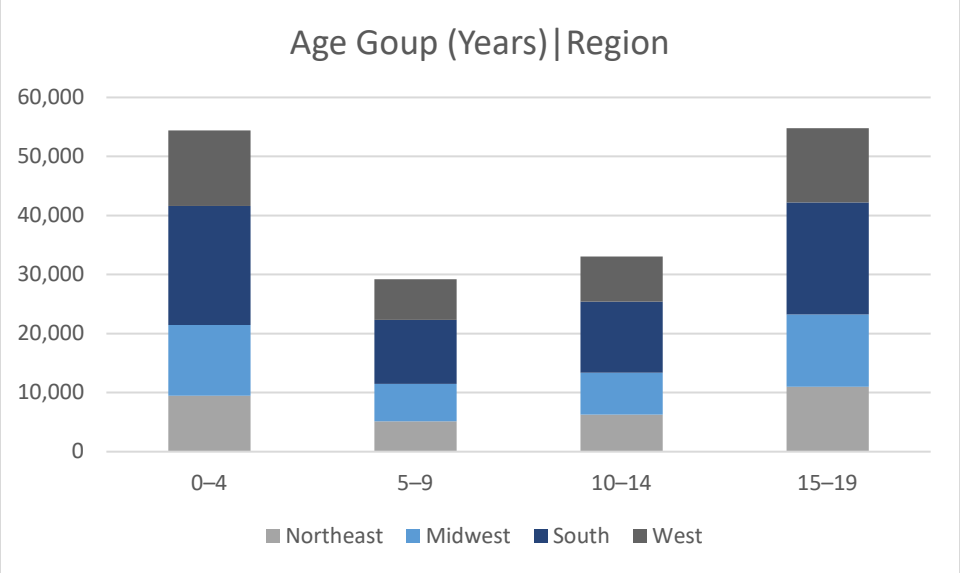
After performing and reviewing the various data statistics there are additional findings from the sample tests, correlation analysis and regression testing. From the sample testing there is evidence that the number of incidences by metropolitan population level, more urban versus non-urban, differs by region. Other additional findings included a test regarding if the average of two regions incidences is the same or different. The data analysis revealed that the answer is there is regional differences. There is also a moderate positive correlation between the two regions tested, South and West, that indicated the link between regional incidence rates and population density as proven by the excel analytics.

Descriptive Statistics:

Overall, these statistics help in summarizing the geographic incidence rates of pediatric cancer in the United States. This analysis provides insights into the central tendency, variability, and distribution of the data across different regions.

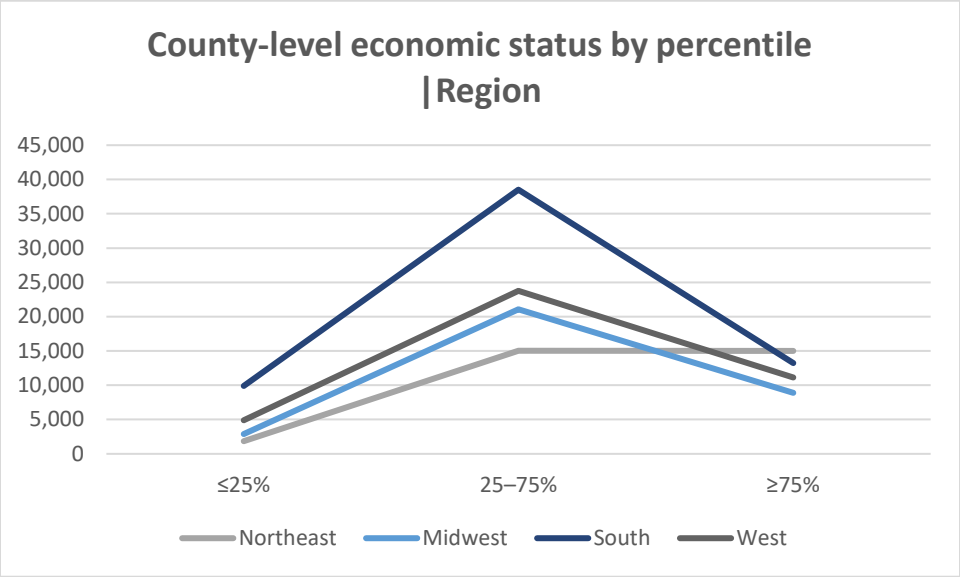


This bar graph of childhood cancer incidence rates by geographic area, broken out by male or female, provides a visual of how these cancer incidence rates vary across different regions for males and for females separately. This is an easy way of displaying the data. Using the breakout by sex, allows us to identify areas that are notably higher and therefore need more targeted reviews and interventions. This graph clearly visualizes that the incidence rates are higher in males in every region and that the South warrants a deeper dive to attempt to identify things that may be contributing to this.

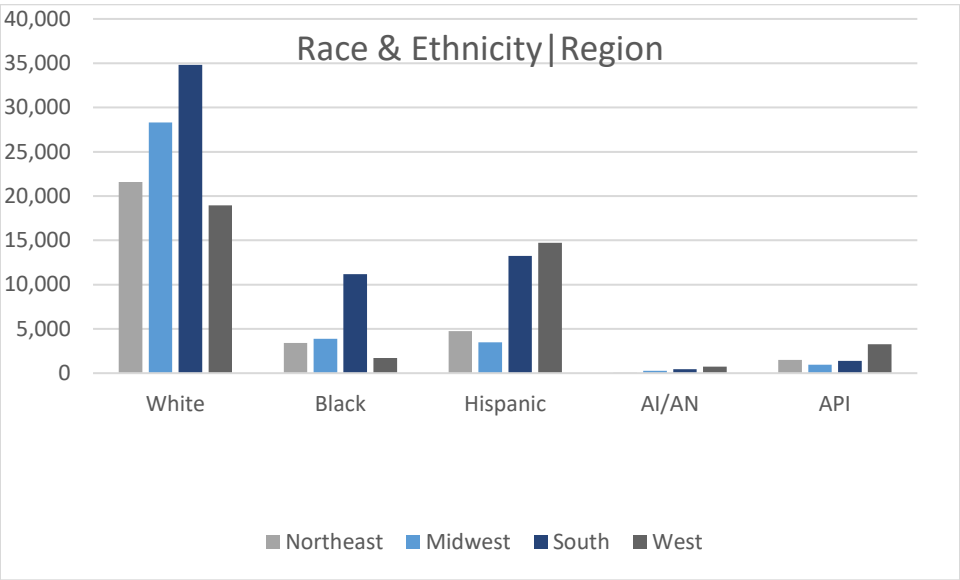


In this type of graph it allows for easy visual comparison, within each age range, of the childhood cancer incidence rates across different racial and ethnic groups. Again, we can see that the South warrants additional investigation as they have a higher rate within each age set, and in particular the youngest range and the oldest range. This graph is highlighting the disparities and variations

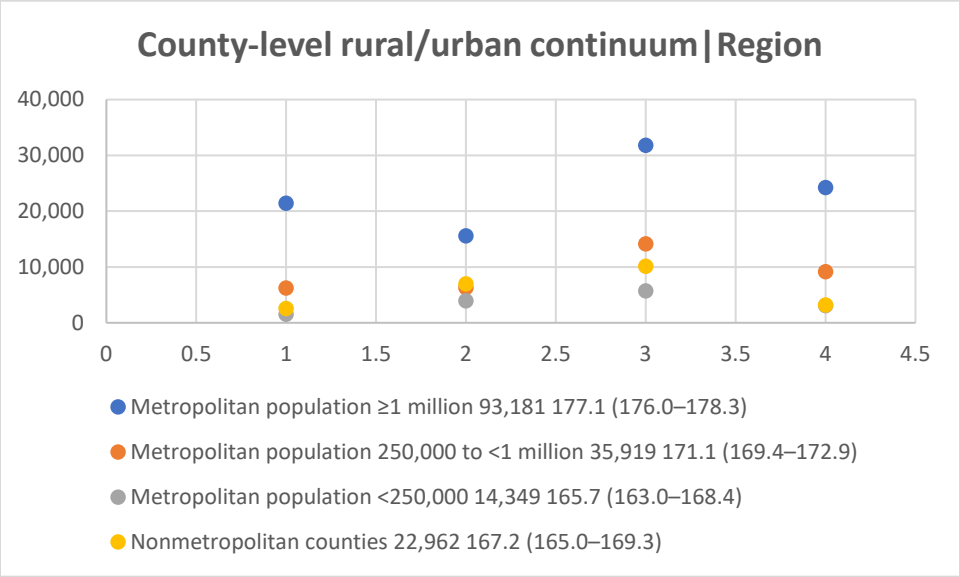
in cancer rates among these age groups. Then we can potentially inform public health efforts, guide research into risk factors and disparities that may be influencing the rates.



Here we look at the data for county level economic status as it relates to childhood cancer incidence rates by geographic region. This indicates the relationship between economic status and incidence rates across different counties. This represents trends across the regions for each economic range where we can see a similar trend across all. This indicates a peak in all areas in the 25-75% percentile and then a decrease in incidence rates in all areas when in the greater than 75% rate. The lines slope upward as economic status increases, suggesting a positive correlation between economic status and childhood cancer incidence rates across all regions. Yet a significant decrease in rates occurs across all areas as economic status areas reaches the highest range for this graph. This could support the theory that access to healthcare and economic factors play a role in occurrence rates.



Here we represent this data set through a bar graph – where you can clearly see the highest incidence rates are in the white race. This also again exemplifies how the south is higher, in incidence rates in whites and blacks, whereas the west incidence rates are driven by Hispanic and API occurrences. We can now identify any variations in cancer rates among different groups within specific geographic regions, which will benefit public health planning, allocating resources, and, I think most importantly, allow for targeted interventions to address disparities in childhood cancer incidence.



Two-sample hypothesis testing:

See excel file.

Correlation:

See excel file.

Regression analysis:

See excel file.

Conclusions (500 words):

The main insights learned from analyzing this childhood cancer data on incidence rates by geographic location as well as ethnicity include insights into disparities, identifying potentially high-risk populations, susceptibility, potential environmental factors as well as access to care. Identification of geographic regions with higher or lower incidence rates of childhood cancer can highlight disparities in access to healthcare, environmental factors, or genetic predispositions. Understanding these regional differences can guide resource allocation and public health interventions to address disparities and improve outcomes. Another part of the analysis is the examination of childhood cancer incidence rates by ethnicity which can reveal disparities in the

cancer burden among different racial and ethnic groups. This analysis may uncover socio-economic factors, cultural differences, or genetic predispositions that contribute to variations in cancer diagnosis and outcomes. This analysis can facilitate addressing these disparities through targeted interventions aimed at reducing barriers to healthcare access, improving detection efforts, and addressing social determinants of health. Identifying populations that are at higher risk for cancer and certain types of cancer due to genetic predisposition or environmental factors. Understanding these patterns, because of analysis like this, can inform preventive strategies, screening programs, accessibility to care and genetic counseling efforts to mitigate risk and improve outcomes. Clusters of childhood cancer cases in specific geographic areas may be identified through this analysis which would allow for additional analysis to identify and then mitigate and address this to improve outcomes. Examining childhood cancer incidence rates by ethnicity may reveal genetic susceptibility patterns that influence cancer risk. Certain ethnic groups may have an increased predisposition to specific types of cancer. Understanding these genetic factors can facilitate personalized risk assessment, genetic counseling, and targeted screening strategies for at-risk populations. This leads to the topic of the analysis as it relates to disparities in childhood cancer incidence rates by geographic location and ethnicity may reflect differences in healthcare access, quality, and utilization. Areas with limited access to healthcare services or lower socioeconomic status may experience higher cancer incidence rates due to delayed diagnosis, inadequate treatment, or accessibility to care for various factors. Addressing healthcare disparities is essential for ensuring equitable access to cancer diagnosis, and treatment services for all children.

Overall, analyzing childhood cancer data by geographic location and ethnicity provides valuable insights into the complex correlation of genetic, environmental, social, and healthcare-related factors influencing cancer risk and outcomes. These insights can inform actions for earlier interventions, policy decisions, and research priorities aimed at reducing the burden of childhood cancer and improving health equity for all children. Expanding on the policy perspective, the awareness of trends and patterns in childhood cancer incidence facilitates the focus on the needs of this community. This also guides the allocation of resources, public health initiatives, and healthcare planning to better address the needs of those affected. Investigating disparities in incidence rates can highlight imbalances in access to care services and inform ways to promote health equity. These facts, data analysis and continued research in this area has the potential to significantly impact clinical practice, public health policy, and the lives of children and families affected by cancer. The knowledge of the geographic incidence rates of childhood cancer may be used to enhance awareness for providers, increase treatment capacities and facilitate better survivorship care and cancer surveillance overall. Continuing to use these data points and analytics will provide information to continue to highlight and drive for the need to address the burden of childhood cancer.

Final reflection (500 words):

Data analysis is useful for public administrators as this role encounters complex issues and policy decisions that affect communities and populations. Data analysis provides fact-based insights and evidence to inform decision-making processes. Through analyzing data, public administrators, and those in similar roles after completing an MPA, can identify trends, assess the

effectiveness of policies and programs, and make informed choices to address public needs and priorities. Public administrators often face resource constraints and competing demands this is where data analysis helps optimize resource allocation by identifying areas of greatest need, evaluating the impact of existing programs, and directing resources towards interventions with the highest potential for positive outcomes. This ensures that limited resources are used efficiently and effectively to achieve desired objectives. I would like to highlight again that these priorities, as a result of data gathering and analytics, also identify where there are truly the most need in the community and get to work addressing it.

Performance monitoring and evaluation is another aspect of the data analytics that are important to roles in public administration. These roles are responsible for overseeing the implementation of policies and programs. Data analysis enables them to monitor performance, track progress towards goals, and evaluate the effectiveness of interventions. By measuring outcomes and assessing programmatic success, administrators can make adjustments as needed to improve efficiency and effectiveness while focusing efforts on proven areas of need.

In the public sector and societies like what we have here in the United States, public administrators are accountable to the public and elected officials. Data analysis promotes accountability and transparency by providing objective evidence of government actions and outcomes. Through data-driven reporting and analysis, administrators can demonstrate the impact of their decisions, promote trust in government institutions, and enhance public confidence in governance. Data analysis also provides the fact-based support for why certain areas are prioritized and addressed. Which leads into the next factor of evidence-based policy development. Data analysis facilitates evidence-based policymaking by grounding policy decisions in clear evidence and rigorous analysis. By examining data on social, economic, and environmental factors, administrators can identify root causes of problems, forecast future trends, and design targeted interventions to address complex challenges effectively.

This class has shown me the levels of data analysis that are available and achievable for use in decision making and prioritizing areas of focus and need for the community. This class reinforced my understanding of data analysis by offering practical tools, techniques, and methodologies for analyzing data in various contexts. It has deepened my appreciation for the role of data in informing decision making and driving positive change in public administration. This class has been extremely challenging, and I am not sure the level and depth of analysis will be naturally applied by a non-statistician in a public administer role. I am truly amazed at the level of analysis that is available for use in decision making and for that I am thankful for this class showing me this. Additionally, the class has highlighted the importance of critical thinking, interpretation of results, and effective communication of findings to stakeholders. Overall, the class has equipped me with valuable skills and knowledge that are essential for leveraging data to address real-world challenges in public administration.