**INFO 5770: FINAL REPORT**

**TITLE: "A COMPARATIVE ANALYSIS OF RESPIRATORY DISEASE CONCENTRATING ON ASTHMA AND INFLUENZA"**

**CONTENTS**

|  |  |
| --- | --- |
| Introduction | 3 |
| Causes of the disease | 4 |
| Treatment | 4 |
| Influence factors | 4 |
| Research questions | 5 |
| Related work | 5 |
| Methods | 6 |
| Data Preprocessing | 8 |
| Descriptive Statistics | 11 |
| Demographics | 12 |
| Treatment | 30 |
| Normal Distribution | 35 |
| Results | 41 |
| Discussion | 42 |
| Reflection | 43 |
| Conclusion | 44 |
| References | 45 |
|  |  |

**INTRODUCTION**

|  |  |
| --- | --- |
| **TITLE** | "A Comparative Analysis Of Respiratory Disease Concentrating On Asthma And Influenza" |
| **MEDICAL CONDITION** | Influenza (ICD-10: J11 & J9) and Asthma (ICD-10: J45) |
| **TEAM MEMBERS** | **ARJUN PALNISWAMY:** Addressed research questions, created visualizations and removed outliers.  **AAVULA USHA:** Processed and cleaned the data, removed outliers, Drafted, prepared variables for analysis.  **KAUSHIK DINAKARAN**: Conducted Chi-square and Mann-Whitney U tests, analyzed correlations.  **ROHITH SINGH THAKUR:** Data cleaning**,** Designed the PowerPoint presentation, Mann-Whitney U tests, finalized the report. |

Individuals throughout the planet become impacted by two recognized viruses that result in major physical and economical implications: influenza (ICD-10: J11) and asthma (ICD-10: J45). Though influenza is a temporary virus it circulates across the colder months and can result in devastating disease consequence, asthma is a passed on and atmospheric disorders which calls for regular care. Both diseases emphasize the need of resource allocation and healthcare planning.

**INFLUENZA (ICD-10 CODE J11)**

**CAUSES:**

* Infection caused by any influenza virus (type A, B, or C).
* Coughing, sneezing, or even chatting by an infected person produces respiratory droplets.
* Transmission occurs through contaminated surfaces, followed by contact with the face.

**TREATMENT:**  
Medication for influenza contains antiviral drugs such as baloxavir, zanamivir, or oseltamivir, which work best when started after 48 hours of the beginning of symptoms. Rest, fluids, and over-the-counter fever and pain drugs can all help with controlling symptoms. Prevention is essential, which includes receiving an annual flu vaccine, practicing excellent sanitation of the hands, wearing masks, avoiding sharing space with ill people.

**INFLUENTIAL FACTORS:**

Age is a vital factor, both children and older persons more prone. Long-term diseases such as heart disease and diabetes, as well as a weakened immune system, raise the chance of severe illness. Outside factors such as cramped living areas and insufficient ventilation let the virus spread even more quickly.

**ASTHMA (ICD-10: J45)**

**CAUSES**:

* Genetic are the main cause for the asthma as well as allergies.
* Natural allergies include pollen grains, fungi, mites and dust, and fur from pets.
* pollutants which includes nicotine, polluted air, and unpleasant smells.
* Workplace dust or exposure to chemicals, as well as viral illnesses.

**TREATMENT:**The medication is a quick-relief treatment for severe signs of asthma, whereas inhaling corticosteroids (such fluticasone) and combinations inhalers can be utilized for long-term control. For serious cases, corticosteroids taken by mouth may be required. Asthma treatment strategies, risk avoiding them, and periodic exams with doctors are beneficial to patients.

**INFLUENCE FACTORS:**

Natural factors that worsen asthma are irritants and polluted air. Health management is subject to financial hurdles, such as limiting ability to obtain healthcare. Although lifestyle choices like tobacco use and being overweight, as well as psychological stress, can aggravate signs and symptoms, complete therapy is needed for positive outcomes.

**RESEARCH QUESTIONS**

1. Do diagnoses for influenza and asthma differ by gender?

2. What is the difference in the average age of diagnosis between influenza and asthma?

3. What racial/ethnic trends exist in the frequency of influenza and asthma?

4. Do cases of influenza and asthma change by region?

5. Do the groups with influenza and asthma have significantly different medical expenses?

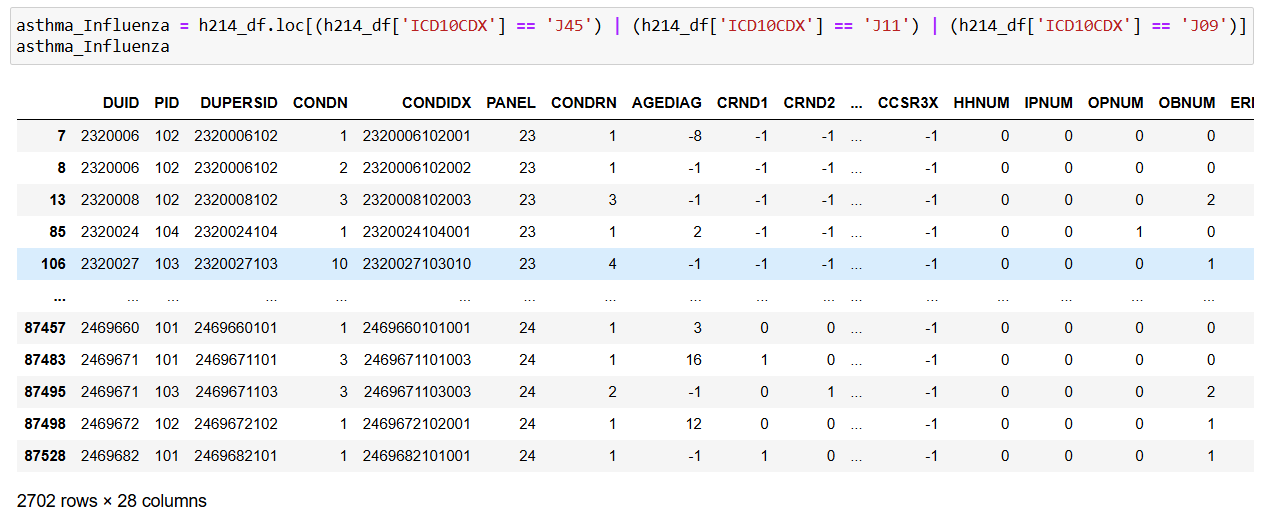
**RELATED WORK**

"Impact of Influenza on Asthma Exacerbations: A Systematic Review" is one of the well-known studies that relates influenza infections to asthma attacks. It draws attention to the fact that influenza significantly raises the likelihood of a serious asthma attack; therefore, the rates of hospitalization for patients with asthma increase together with those for emergency care. By prioritizing it as a precautionary measure, it reduces the chances of symptoms and reduces the financial burden on patients. Therefore, vaccination of asthma sufferers against influenza is important each year. The comprehensive management approaches for the concomitant respiratory diseases should thus be implemented.

**METHODS**

**DATA PREPROCESSING**

**STEP 1:** SELECTION OF DATASET



**STEP 2:** DATA LOADING AND DATA FILTERING.

A screenshot of a computer

Description automatically generated

**STEP 3:** MERGED h214.csv WITH h216.csv.

* DUPERSID: Unique identifier for each survey respondent.
* SEX: Gender of the respondent (1 = male, 2 = female).
* DOBMM/DOBYY: Month (DOBMM) and year (DOBYY) of birth.
* AGEDIAG: Age at diagnosis of a specific condition, with missing values coded as negative or null.
* MARRY31X: Marital status (e.g., 1 = married, 2 = widowed, etc.).
* RACEV2X: Racial/ethnic background (e.g., 1 = White, 2 = Black, etc.).
* REGION42: Region of U.S. residence (1 = Northeast, 2 = Midwest, etc.).
* TOTEXP19: Total annual healthcare expenditures in 2019.
* ICD10CDX: ICD-10 diagnosis code for health conditions (e.g., J45 = asthma).
* ADAPPT42: No. medical appointments attended during the last year.
* ADRISK42: Risk assessment score for asthma or related health conditions.
* ASATAK31: Indicator of asthma attack in the last year (1 = Yes, 0 = No).
* ASDALY31 Use daily preventive asthma medication. (1 = Yes, 2 = No)
* ASSTIL31 Currently have asthma. (1 = Yes, 0 = No)
* ASTHEP31 Time since last asthma attack - in days, weeks, or months.
* ASACUT31 Used quick-relief inhaler in past three months. (1 = Yes, 0 = No).
* ASPREV31 Ever used preventive asthma medication. (1 = Yes, 0 = No).
* ASPKFL31: Has peak flow meter to monitor asthma (1 = Yes, 0 = No)
* MCDOC19: Visited a Medicaid or doctor in 2019 tracking access or frequency.
* ADHECR42: Adherence to health care recommendations
* Current\_Age: Respondent's current age.
* Flu shot taken: Has taken a flu shot in the past year (1 = Yes, 0 = No).

**STEP 4:** CHECK FOR MISSING VALUES BY MASKING VALUES GREATER THAN 0.

**STEP 5:** SEGREGATING CATEGORICAL AND NUMERICAL VARIABLES.

A screenshot of a computer code

Description automatically generated

* Categorical columns: Index(['SEX', 'DOBMM', 'MARRY31X', 'RACEV2X', 'REGION42', 'ICD10CDX', 'ADAPPT42', 'ADRISK42', 'ASATAK31', 'ASDALY31', 'ASSTIL31', 'ASACUT31','ASPREV31', 'ASPKFL31', 'MCDOC19', 'DSFL1853', 'DSFL1953', 'DSFL2053'], dtype='object')
* Numerical columns: Index(['DUPERSID', 'DOBYY', 'AGEDIAG', 'TOTEXP19', 'ADHECR42'], dtype='object')

**STEP 6:** IQR COMPUTATION FOR ALL NUMERICAL VARIABLES AND REMOVAL OF OUTLIERS WITH A THRESHOLD OF 3 TIMES OF IQR.

**STEP 7:** GENERATE A CORRELATION MATRIX FOR NUMERICAL VARIABLES TO SEE WHICH COULD BE REDUNDANT.

**STEP 8:** CALCULATE AND VISUALIZE THE CORRELATION MATRIX FOR NUMERICAL VARIABLES ('AGEDIAG', 'TOTEXP19', 'ADAPPT42', 'ADRISK42', 'Current\_Age').

A screenshot of a computer code

Description automatically generated

A screenshot of a graph

Description automatically generated

**STEP 9:** CHI-SQUARE VALUES FOR THE SAME VARIABLES.

**STEP 10:** PERFORM MANN-WHITNEY U TEST TO COMPARE MEDICAL EXPENDITURES BETWEEN ASTHMA AND INFLUENZA GROUPS.

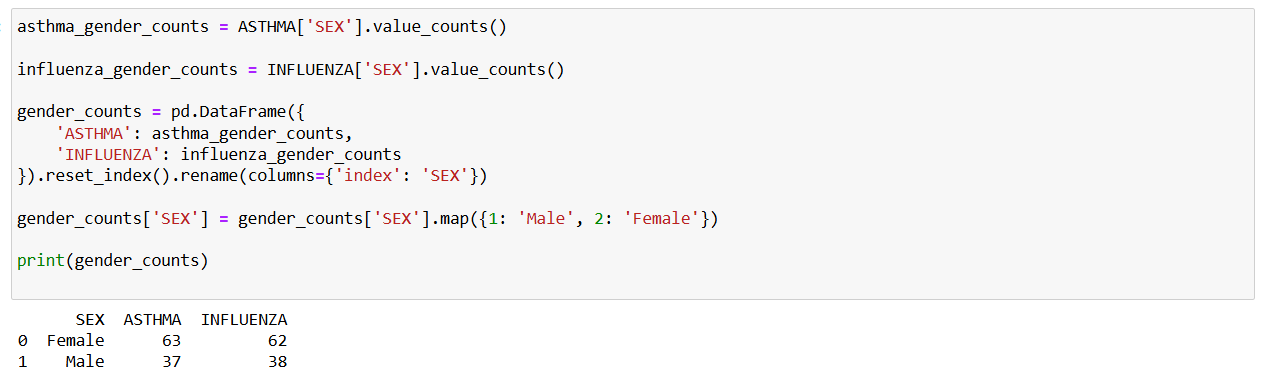
**DESCRIPTIVE ANALYSIS**

The aim of this descriptive statistics is to explore the medical and social factors affecting the expenditure of patients diagnosed with **Influenza and Asthma**, focusing on the following aspects:

1. Investigating gender differences in diagnoses.
2. Examining differences in the average age at diagnosis.
3. Identifying racial/ethnic patterns in prevalence.
4. Analyzing regional differences cases.
5. Assessing the variation in medical expenditures among both patients.

**DEMOGRAPHICS**

**1. DO DIAGNOSES FOR INFLUENZA AND ASTHMA DIFFER BY GENDER?**



The code compares the gender distribution in asthma and influenza diagnoses. It calculates the frequency of males and females for each condition and then combines data into a single table. The gender values are mapped to "Male" and "Female," and the data is visualized using a stacked bar chart. This chart displays the gender differences between asthma and influenza in order to analyze patterns.

A graph of two people

Description automatically generated with medium confidence

Description of the Visualization - The bar chart shows the asthma and influenza conditions among Female and Male genders. For this, the x-axis represents the gender category; the y-axis represents the occurrences. Each bar is further subdivided into two divisions to represent asthma in blue color and influenza in orange color. To differentiate between the conditions, there exists a legend.

**Results:**

**1. Female Distribution:**

* The total frequency in females is more than 120.
* Asthma cases contribution (blue segment) is something around 60.
* Influenza (orange segment) is rather higher, suggesting that more females are affected by influenza than asthma.

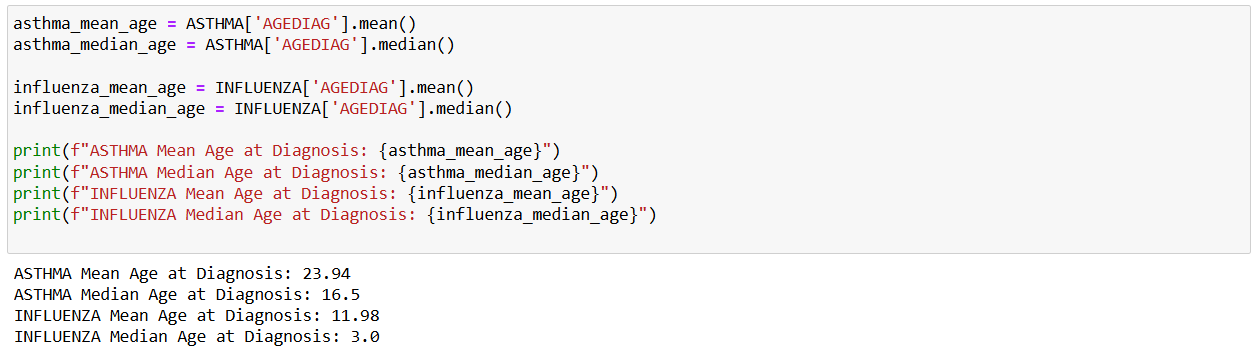
**2. Male Distribution:**

* The total frequency in males is about 80.
* The conditions seem to appear with about the same frequency, but influenza cases are a little higher compared to asthma.

**3. Comparison:**

* In general, the total frequency of females in both conditions is greater than that of males.
* Influenza has been more prevalent than asthma in both genders, but the prevalence in females is much higher.
* The visualization really captures the prevalence of asthma and influenza based on gender, with influenza, especially among females, having a higher impact.

**2. WHAT IS THE DIFFERENCE IN THE AVERAGE AGE OF DIAGNOSIS BETWEEN INFLUENZA AND ASTHMA?**

****

The code calculates the average (mean) and middle (median) ages at diagnosis for asthma and influenza patients. It uses the AGEDIAG column from each dataset to compute these statistics separately for both conditions. The results are printed to show the central tendencies of age at diagnosis for asthma and influenza. This helps compare how the average and typical ages differ between the two conditions.

A close-up of a computer screen

Description automatically generated

The code will generate histograms of the age at diagnosis for both asthma and influenza patients. Plots are generated separately for each condition to show the frequency of diagnosis at different age groups, hence highlighting patterns or differences in age distributions between the two conditions.

A comparison of different colored bars

Description automatically generated with medium confidence

The graphic shows two histograms side by side to demonstrate the age distribution at diagnosis for both asthma and influenza. The prevalence of asthma diagnoses across different age groups is shown by the asthma histogram. • The y-axis displays the frequency of instances, while the x-axis displays the age at diagnosis, which ranges from 0 to 70. The influenza histogram shows how common influenza diagnoses are across different age groups.

• The x-axis, which ranges from 0 to 70, is the same as in the asthma chart.

• The frequency is shown on the y-axis.

**ANALYSIS:**

**Age distribution of Asthma:**

* The age group with the most diagnoses of asthma includes more than 25 in the youngest age group (0-10 years).
* Notice a decrease in frequency as age increases; the smaller peak is around the age of 40-50 years.

**Age distribution of influenza:**

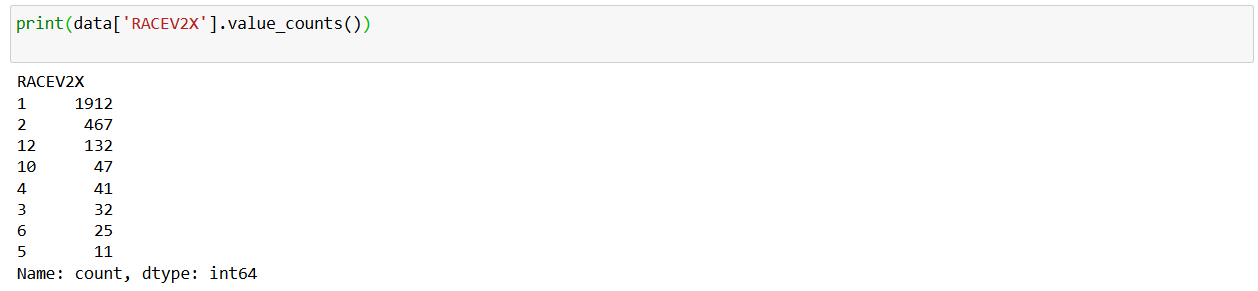
* More than 50 cases of influenza diagnoses occur within the youngest age group, 0-10 years.
* There is a significant reduction in frequency with increasing age, with very few cases beyond 30 years.

**Comparison:**

* Both conditions have the highest frequency of diagnoses in the youngest age group.
* Influenza has a much higher frequency in the 0-10 age group compared to asthma.
* Asthma has a secondary peak in middle age, 40-50, whereas influenza does not.

The visualization shows that both conditions are most prevalent among young children, with a higher prevalence of influenza in this population. Asthma also affects the middle-aged to a significant level.

**3. WHAT RACIAL/ETHNIC TRENDS EXIST IN THE FREQUENCY OF INFLUENZA AND ASTHMA?**

****

**A computer screen shot of a computer code

Description automatically generatedA screenshot of a computer program

Description automatically generated**

This Python code will analyze data about patients to compare the ratio of asthma-ICD10 code 45 and influenza-ICD10 code 11 and 9-between different races. It takes race codes and maps them against their descriptive labels, performs stratified sampling of up to 100 records in each stratum to balance the set, counts the number of asthma and influenza cases based on race, and visualizes using a bar chart. This makes the chart a clear comparison by race for each condition. The script provides an appropriate message instead of generating a plot when no data meets the criteria.

A graph of different colored columns

Description automatically generated

The bar graph entitled "Number of Asthma and Influenza Patients by Race" depicts the breakdown of patients with asthma and influenza within various racial categories.

**KEY OBSERVATIONS:**

**Overall Trends:**

* Asthma patients consistently outnumber influenza patients in every racial category.
* For both conditions, the majority of patients belong to the "White" and "Multiple Races" categories.

**Specific Racial Groups:**

* White: This race has the highest number of patients for both asthma, about 70, and influenza, about 30, showing a high prevalence of the conditions.
* Multiple Races: Like the White race, there is a high number of asthma patients, about 70, and a relatively high number of influenza patients, about 25.
* Black: This race represents a moderate number of patients, with about 30 for asthma and 20 for influenza.
* Asian Indian: Asthma patients are a fair amount, about 20, while influenza patients are approximately 5.
* American Indian, Chinese, and Filipino: Patients in this category are relatively fewer, from 5 to 15 asthmatic patients, while the influenza cases are less than these.
* Others: This has a somewhat fair distribution where the number of asthma patients reaches up to about 20 and less than 15 in case of influenza.

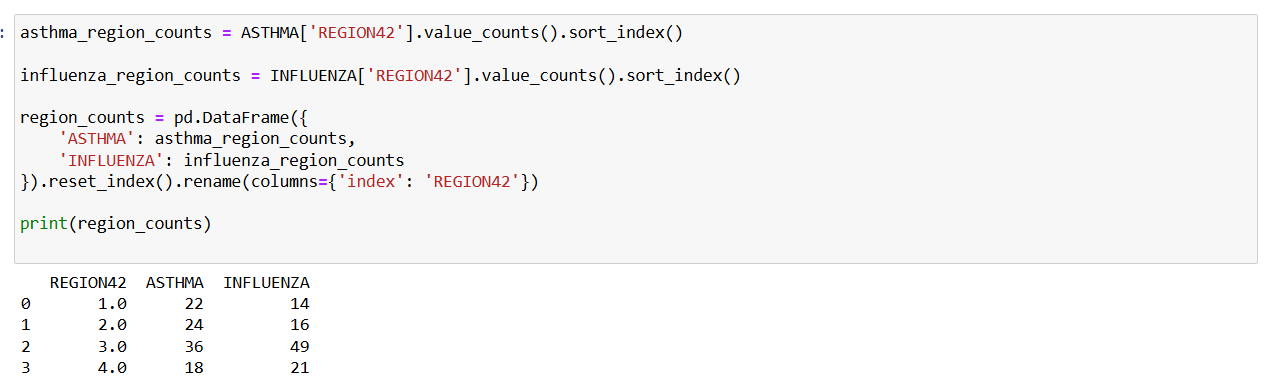
**Conditions Comparison:**

* Asthma appears to be more prevalent compared to influenza across the board.
* There is the widest gap in the "White" and "Multiple Races" asthmatic to flu patient ratio.

**Conclusion:**

We are able to see considerable disparities in asthma and flu from this visualization. Whites, with multi-race individuals taking second place, are overly afflicted with asthma. This is significant for health planning and health resources planning towards the disparities.

**4. DO CASES OF INFLUENZA AND ASTHMA CHANGE BY REGION?**

****

****

The code visualizes the geographic distribution of asthma and influenza across the regions using a heat map. First, it reshapes the data in the format that can plot a "REGION42" indicating a region, "Condition" for health conditions, and "Count" for cases. Then, it visualizes a heatmap using seaborn, which will provide values of each cell and provides more color contrast. The plot displays the case distribution across regions through which the trend and regional variations can easily be observed. The chart has a descriptive title and axis labels for clarity and understandability.A chart with different colors

Description automatically generated

The "Geographic Distribution: ASTHMA vs. INFLUENZA" heatmap below shows the amount of patients suffering from influenza and asthma in each of the four different regions.

**OVERVIEW OF DISTRIBUTION:**

* The greatest number of patients with both influenza and asthma in Region 3 is 49 for influenza patients and 36 for asthma patients.
* Compared to asthma, influenza cases appear to be concentrated in Region 3.

**Regional Analysis:**

* **Region 1:** This region has a somewhat greater prevalence of asthma, with 22 cases and 14 influenza cases.
* **Region 2:** There are 16 influenza cases and 24 asthma cases, suggesting a similar pattern of higher asthma cases.
* With the highest numbers 36 for asthma and 49 for influenza—Region 3 stands out as having a serious public health issue.
* **Region 4:** Compared to other regions, this one has the fewest asthma patients (18), but it has slightly more influenza cases (21).

**Comparing Different Conditions:**

* Influenza cases are much higher in Region 3, making it a regional hotspot for the illness; asthma cases are fairly evenly distributed between regions, with a modest peak in Region 3.
* **Color Coding:** Darker hues indicate more patients, and the heatmap's color intensity corresponds with the number of instances. The darkest region, region 3, emphasizes its importance under both circumstances.

**Conclusion:**

The visualization of data demonstrates that Region 3 represents the highest recorded cases in both asthma and influenza and is, therefore, a significant cause for concern. Targeted healthcare strategies and resource allocation may hence be required in this region to meet the higher prevalence of these conditions. Other regions have a consistent pattern where asthma cases are higher compared to influenza, except for Region 4, which has a slightly higher count of influenza.

**CORRELATION ANALYSIS OF KEY NUMERICAL VARIABLES**

**A screenshot of a computer code

Description automatically generated**

This code calculates the Pearson's correlation coefficients between selected numeric variables, such as age at diagnosis (AGEDIAG), total expenditure (TOTEXP19), number of appointments (ADAPPT42), risk assessment (ADRISK42), and current age (Current\_Age). The correlation matrix presents the relationship between these variables, with values ranging from -1 for strong negative correlation to 1 for strong positive correlation. The output heatmap is developed to represent these correlations graphically, employing a variation in color to emphasize the strength and direction of each relationship. This visualization helps identify strong dependencies or trends in the dataset.

A screenshot of a graph

Description automatically generated

The Pearson correlation matrix indicates a number of important relationships among the numerical variables in this dataset. The correlation of 0.73 between `AGEDIAG` and `Current\_Age` is strong and positive, as one might expect, because older participants would have had to be older at the time of diagnosis. The positive, moderate relations exist between TOTEXP19 and Current age 0.32 and Age at diagnosis 0.23; thus, the older the person, or the later the diagnosis of the disease, the slightly higher the costs are. In addition, a weak positive correlation, 0.16 between the risk scores, ADRISK42, and healthcare expenditures, would indicate that those people who perceive their health risks higher may have marginally higher costs.

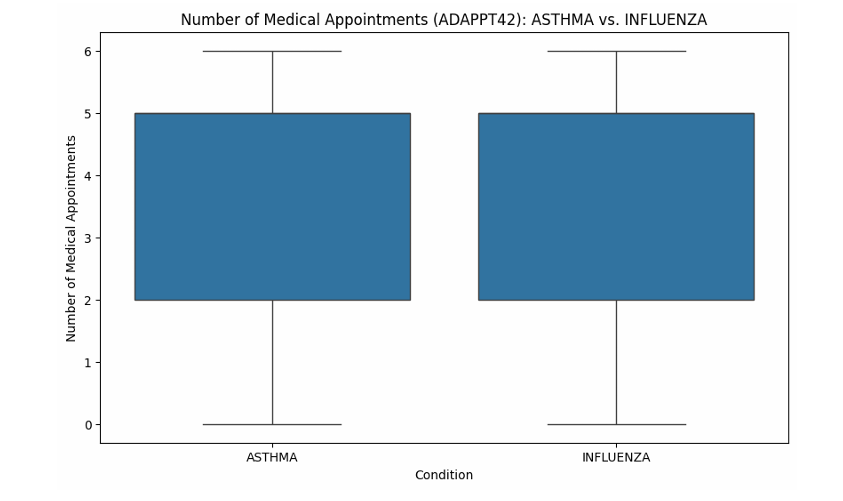
On the other hand, medical appointments (`ADAPPT42`) are inversely related to several factors, with the most considerable negative associations observed for risk score (-0.34) and current age (-0.15), suggesting that the higher the level of risk and the older one is, the fewer the numbers of medical appointments, reflecting disparities in health care access or different healthy behaviors. The negative correlation coefficient between `ADAPPT42` and age at diagnosis is -0.06, thus meaning that appointment frequency is rather weakly but inversely related to when a condition was diagnosed. These correlations underpin the key demographic and health care expenditure trends and can hence be useful for further analysis of asthma and influenza patients in a more behavioral and characteristic way.

**5. DO THE GROUPS WITH INFLUENZA AND ASTHMA HAVE SIGNIFICANTLY DIFFERENT MEDICAL EXPENSES?**

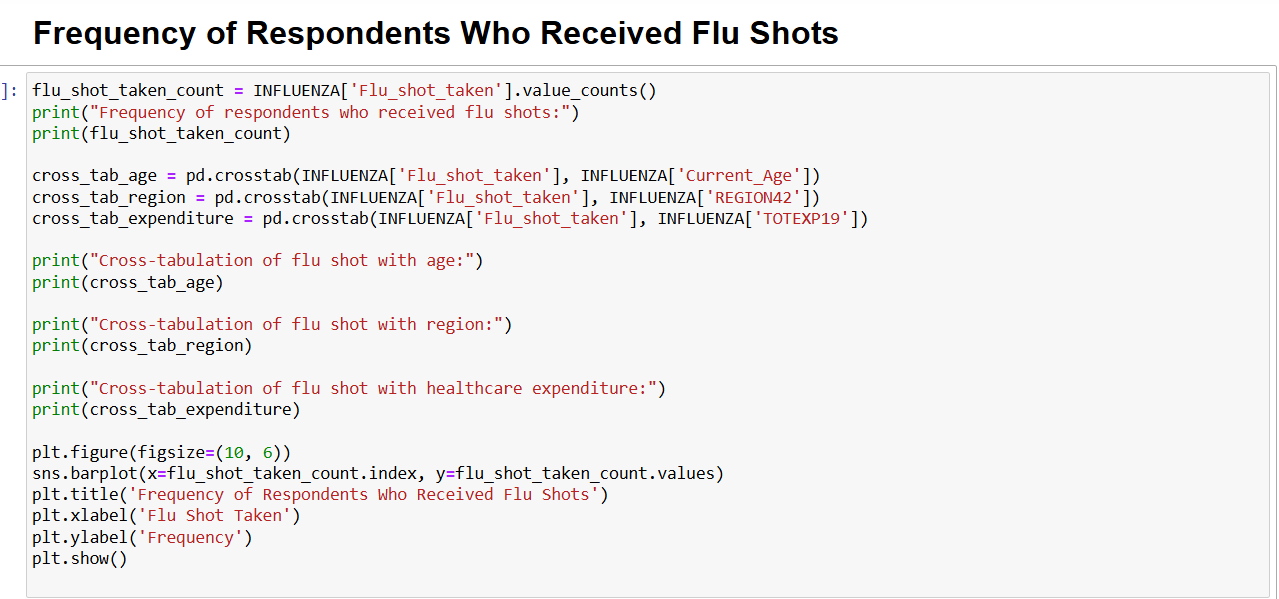
**A screenshot of a computer program

Description automatically generated**

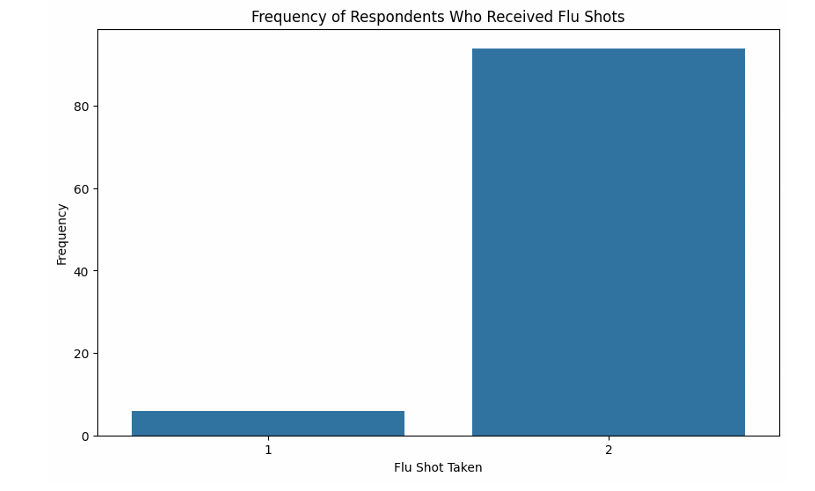
This code calculates and compares the mean number of medical appointments (ADAPPT42) for asthma and influenza patients. The mean values are printed to provide a quantitative overview of appointments for each condition. Then, it combines data into one dataset with a condition label for visualization purposes. A boxplot is created to display the distribution of medical appointments for asthma and influenza, highlighting variability and central tendencies. The plot has a clear title and labeled axes, providing a good way to compare the frequency of appointments across the two conditions.



**TREATMENT**



This code analyses the uptake of flu shots amongst influenza patients. It calculates the frequency of the respondents who have taken flu shots, then plots this in a bar chart to clearly visualize it. Cross-tabulations are created to identify the relationship of the uptake with demographic factors like age, geographical region, and total health care expenditure. These cross-tabs detail more profound flu shot adoption trends among the analyzed groups. Whereas the bar chart presents a high-level look at overall frequencies, these cross-tabs target detailed analyses by the factors influencing this trend.



A close-up of a computer code

Description automatically generated

INFLUENZA Mean Total Expenditure (TOTEXP19): 9843.76

INFLUENZA Mean Number of Appointments (ADAPPT42): 3.94

A screenshot of a computer code

Description automatically generated

This code calculates and compares the major statistics in healthcare expenditure TOTEXP19 for both asthma and influenza patients. It further calculates the key measures, that is, mean, median, minimum, maximum, and IQR for each of the two conditions. These measures give an overall description of the expenditure distribution by focusing on central tendency and dispersion. Each result is then printed separately for asthma and influenza to ensure comparability in health care costs across the two conditions. The IQR further assists in finding the variability of expenditure and possible outliers. This provides important information about the economic burdens associated with each condition.

**ASTHMA GROUP:**

* By having a higher mean and median, this indicates that the majority of the patients within this group are consistently spending high on medical costs, showing that their treatments are ongoing.
* With the expenditure range smaller and the IQR, a majority of the expenditure is identical among the patients to reflect reasonable and stable costs within the group.

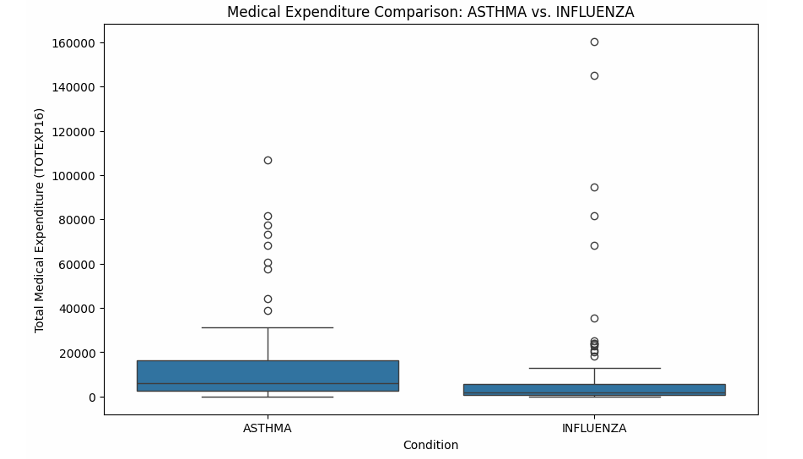
**INFLUENZA GROUP:**

* The largest range and greater maximum expense indicates that many patients would have low expense, although there are the cases who may have extraordinarily high expense. This suggests variabilities in the severity of the flu and duration in the flu cases.
* The lower mean and median expenditures suggest that, on average, treatments for influenza may not be as expensive as treatments for asthma, but the presence of outliers increases overall variability.

**MEDICAL EXPENDITURE COMPARISION: ASTHMA VS INFLUENZA**

A screenshot of a computer code

Description automatically generated



The ASTHMA group likely has higher median medical expenditures and a wider range of expenses, indicating more substantial and varied treatment costs.

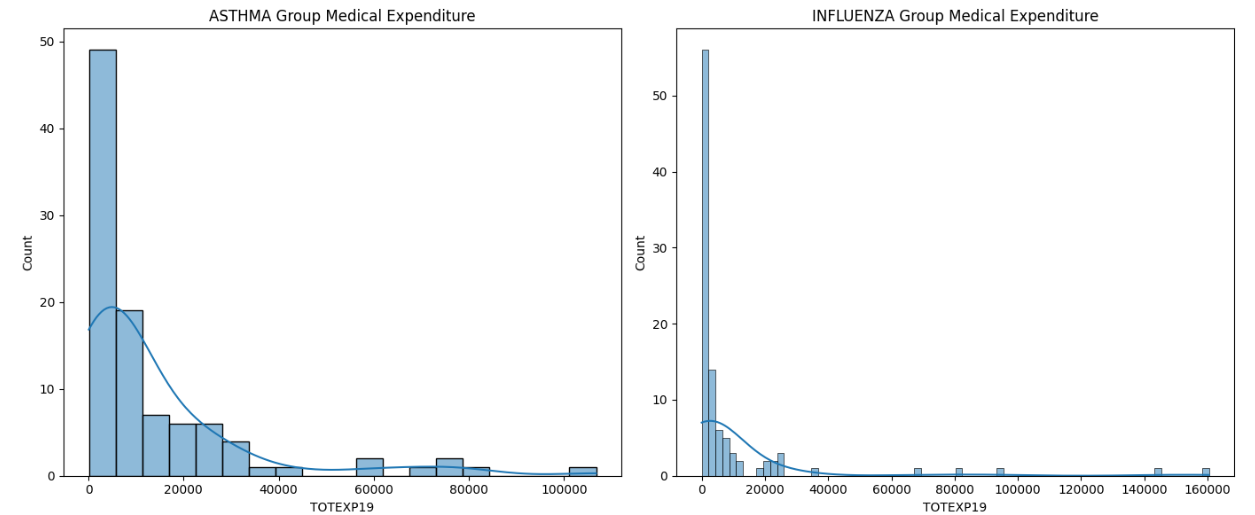
The INFLUENZA group might have lower median expenditures but more variability in costs, as indicated by the presence of outliers.

**DISTRIBUTION ANALYSIS OF MEDICAL EXPENDITURES FOR ASTHMA AND INFLUENZA GROUPS**

A screenshot of a computer program

Description automatically generated

In order to examine the distribution features of medical expenses, TOTEXP19, for the influenza and asthma groups, this code computes the skewness and kurtosis. Kurtosis evaluates the data distribution's sharpness, while skewness gauges asymmetry. To see spending trends, histograms with KDE plots are made for both groups. These metrics and charts offer information about the distribution and form of spending habits, which could be useful in spotting irregularities or trends.



**ASTHMA GROUP:**

* Skewness: 2.57 - This indicates that the distribution is favorably skewed and that the right tail is significantly longer; most medical expenses are low, but some are extremely high.
* The distribution is thought to have heavy tails and a spiked peak, as indicated by the comparatively high kurtosis value of 6.84. There are more extreme outlier data points.

**INFLUENZA GROUP:**

* Skewness: 4.37 - This is even more positively skewed than the ASTHMA group, and it indicates that most of the expenditures are low while there are some very high outliers.
* Kurtosis: 20.01 - The very high value of kurtosis indicates extremely heavy tails and a sharp peak, which means that there is a significant number of outliers and high variability in the expenditures.

**DETERMINING NORMALITY:**

* **Shape:** The two groups' histograms lack the symmetrical bell curve form of a normal distribution because they are right-skewed.
* **Skewness:** An unbalanced distribution is indicated by both groups' positive skewness scores.
* **Kurtosis:** Both groups have high kurtosis values, which suggest the presence of heavy tails and more extreme values than those found in a normal distribution.

**CHI-SQUARE TEST FOR ASSOCIATION BETWEEN DIAGNOSIS AND MEDICAL EXPENDITURE**

**A computer code with many text

Description automatically generated with medium confidenceA white background with black text

Description automatically generated**

This code checks the relation of ICD10CDX medical conditions categories (above or below median) versus TOTEXP19 total medical expenditure categories by a Chi-Square test. A contingency table is developed, and this test is conducted to see whether the observed distribution of cases is significantly different from the distribution that would be expected under chance. It prints the Chi-square, p-value, degrees of freedom, and expected frequencies. The result decides whether there is a statistically significant relation between the type of diagnosis and the level of expenditure, with significance being assessed by the p-value threshold of 0.05.

The results show that there is a significant association between the diagnosis (ICD10CDX) and medical expenditures (TOTEXP19):

Asthma (ICD10CDX 45) is associated with higher expenditures, probably because of chronic management costs.

Flu influenza, ICD10CDX 11 and 9 tends to incur lower expenditure consistent with an acute often less expensive treatment.

**MANN-WHITNEY U TEST FOR COMPARING MEDICAL EXPENDITURES BETWEEN ASTHMA AND INFLUENZA GROUPS**

**REASONING:**

**Variable Types:**

* Categorical Variable: ICD10CDX, which will have two groups: asthma and influenza.
* Numerical Variable: TOTEXP19, representing medical expenditure.

**Number of Groups:**

* Two independent groups: asthma patients, where ICD10CDX = 45, and influenza patients where ICD10CDX = 11 or 9.

**Type of Samples:**

* Independent samples, given that the expenditures for one group are not related to the other.

**Data Distribution:**

* From earlier analyses, the expenditure data is non-normal (skewed). Hence, a non-parametric test such as the Mann-Whitney U test is appropriate.

**Random Sampling:**

* Random samples from Question 2 ensure unbiased comparison.

**Rationale:**

* The Mann-Whitney U Test is a robust non-parametric test that compares the medians of two independent groups without requiring the assumption of normality.
* This makes it suitable for our skewed data and ensures reliable results.

A screenshot of a computer

Description automatically generated

The Mann-Whitney U Test was applied to see if there is any significant difference in medical expenditure between two groups, a group of asthma patients and influenza patients. What these test results really mean in simple terms are that the test statistic (6710.0) is a number that helps in calculating whether the differences between the two groups are simply due to chance.

The p-value 0.00002954 tells us the probability of observing this large of a difference between the groups, assuming there was no real difference. A very small p-value less than 0.05 implies that this is unlikely to be due to chance.

**What this means:**

* There is a significant difference in medical costs between the two groups.
* While the test doesn't say which group spends more, earlier results showed that asthma patients generally have higher medical costs than influenza patients.
* This could be a difference of either the length or expense of treatment for asthma compared to influenza.
* In one sentence, the two groups don't share the same medical expenses, and this is very likely not due to chance.

**DISSCUSION**

The comparative study on asthma and influenza highlights significant differences in prevalence, treatment costs, and demographic impacts. The report, through comprehensive analysis, has shown that asthma has higher ongoing medical expenses compared to influenza since it requires continuous management and preventive measures. Influenza, because it is seasonal and usually acute, has variability in costs, often influenced by severity and outlier cases. Descriptive analysis gives insight into demographic patterns in both conditions, disproportionately affecting young children, while asthma also peaks among middle-aged adults. Gender-based analysis further reveals a higher prevalence of influenza among females, whereas asthma is almost equally distributed between the genders.

The geographic distribution analysis underlines Region 3 as an area of concern for both diseases and thus requires special healthcare attention and resource allocation. The correlation and expenditure analysis reveal some important trends, including a strong association between age at diagnosis and medical costs, and disparities in healthcare access as reflected by appointment frequency. Such findings underline the economic and social burden of managing respiratory diseases, especially asthma, which requires continuous healthcare resources.

Finally, statistical testing using Mann-Whitney U test produces a significant difference in medical spending between the two groups. It therefore follows that asthma patients have significantly higher healthcare costs, which is well explained by the chronic need for treatment. These not only offer useful insights that may guide healthcare policymakers, but they also set up a useful platform for carrying out future research on affordable healthcare and equitable access.

**REFLECTION**

This report bears testament to the power of data analytics to drive decision-making at all levels in healthcare. Drawing upon statistical tools and visualization, this study develops a layered comprehension of respiratory diseases and their implications. Most importantly, it reflects upon interaction within demographics and health care: for example, where the greater impact of asthma on middle-aged adults brings attention to a gap in preventive health measures during younger years. Furthermore, the regional disparity in cases suggests systemic issues in healthcare infrastructure that need urgent addressing.

The broadening of research on gender, age, and racial/ethnic trends puts greater emphasis on how societal factors interact with health outcomes. The detailed analysis of expenditure, especially the measures of skewness and kurtosis, points out the uncertainty regarding medical costs, particularly for influenza. This uncertainty makes robust health insurance systems all the more important in order to protect patients against financial risks.

This analysis is generally a reflection of how clinical data should be integrated with socioeconomic factors to come up with meaningful insights. It also put forth the need for multifaceted healthcare strategies addressing not just the medical but also the environmental and behavioral aspects of disease prevention and management.

**CONCLUSION**

The comparative analysis of asthma and influenza reveals two different yet interlinked healthcare challenges: one being a chronic condition that necessitates regular medical intervention and the other showing seasonal variability in treatment costs. The study emphasizes the need for targeted healthcare planning, especially in areas with high disease prevalence, such as Region 3. The insights gained from demographic and expenditure patterns show the need for equity in access to healthcare and strategies for cost management. Further research is needed to identify novel strategies that could lower the economic burden of these diseases while improving prevention and the allocation of resources. This report serves to illustrate how data-driven approaches are informing effective health policies and interventions.

**REFERENCES**

1)Veerapandian, R., Snyder, J. D., & Samarasinghe, A. E. (2018). Influenza in asthmatics: For better or for worse? *Frontiers in Immunology*, *9*. <https://doi.org/10.3389/fimmu.2018.01843>

2)Li, K. H., Leong, P.-Y., Tseng, C.-F., Wang, Y. H., & Wei, J. C.-C. (2021). Influenza vaccination is associated with lower incidental asthma risk in patients with atopic dermatitis: A nationwide cohort study. *Frontiers in Immunology*, *12*. <https://doi.org/10.3389/fimmu.2021.729501>

3)Jha, A., Dunning, J., Tunstall, T., Thwaites, R. S., Hoang, L. T., Kon, O. M., Zambon, M. C., Hansel, T. T., & Openshaw, P. J. (2019). Patterns of systemic and local inflammation in patients with asthma hospitalised with influenza. *European Respiratory Journal*, *54*(4), 1900949. <https://doi.org/10.1183/13993003.00949-2019>

4)Wang, Z., Li, Y., Gao, Y., Fu, Y., Lin, J., Lei, X., Zheng, J., & Jiang, M. (2023). Global, regional, and national burden of asthma and its attributable risk factors from 1990 to 2019: A systematic analysis for the global burden of disease study 2019. *Respiratory Research*, *24*(1). <https://doi.org/10.1186/s12931-023-02475-6>