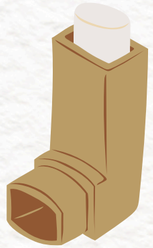
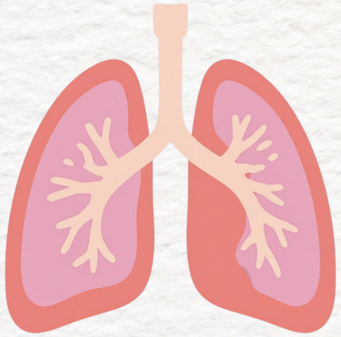


Insights into Asthma Management

Identifying Key Risk Factors and Outcomes

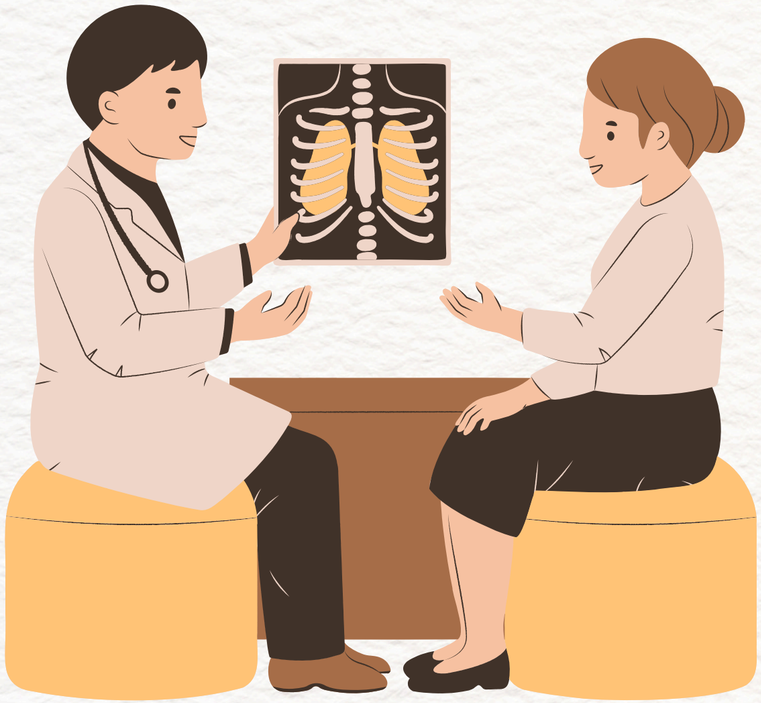


Course: ALY6015 - Intermediate Analytics
Team: Capstone 3

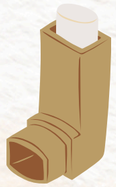
- A. S. Sushmitha Urs
- Lavanya Suresh
- Vaibhavi Bograj

Instructor: Richard He

Introduction



Asthma is a chronic inflammatory airway disease affecting millions worldwide, making it crucial to understand the factors that drive severity, hospital visits, and overall patient outcomes. This project analyzes the Synthetic Asthma Medical Dataset, a realistic dataset of 10,000 patient records with 17 demographic, lifestyle, environmental, and clinical features. It includes variables such as age, BMI, smoking status, physical activity, air pollution levels, FeNO, peak flow, medication adherence, and the number of emergency room (ER) visits. Together, these attributes provide a strong foundation for modeling asthma severity, predicting high-risk patients, and identifying key drivers of healthcare utilization.



Dataset Overview

Variables

The dataset contains 17 key features, including:

- Demographic: Age, Gender, BMI, Family History
- Lifestyle: Smoking Status, Physical Activity
- Environmental: Air Pollution Level
- Clinical: FeNO (airway inflammation marker), Peak Expiratory Flow, Comorbidities
- Asthma Outcomes: Asthma Control Level, Number of ER Visits, Medication Adherence

Data Types

- Numeric: Age, BMI, FeNO, Peak Flow
- Categorical: Gender, Smoking Status, Comorbidities
- Ordered Factors: Asthma Control Level, Air Pollution Level

Why This Dataset Is Useful?

- Clean and structured, similar to real clinical data
- Includes both predictors and outcomes required for severity modeling and risk prediction
- Supports exploratory analysis, correlation assessment, ordinal regression, logistic modeling, and evaluating environmental or behavioral impacts on asthma

Research Questions

1. What demographic, lifestyle, and environmental factors influence asthma severity?
2. Can we predict likelihood of repeated ER visits (proxy for readmission)?
3. How do medication adherence and air pollution jointly affect patient outcomes?

Dataset Characteristics - EDA

Key Patient Characteristics:

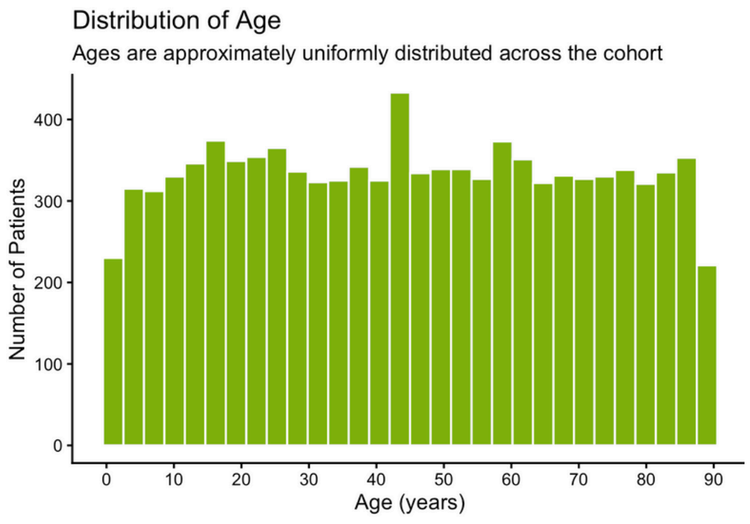
- Age: 1-89 years (peak at 40-50)
- Gender: Balanced (48% F, 48% M, 4% Other)
- BMI: 15-45 kg/m² (mean 25)

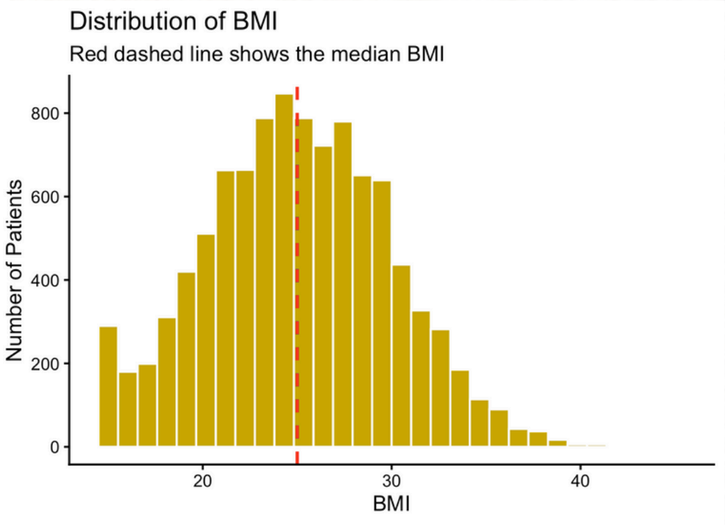
Asthma Prevalence & Control:

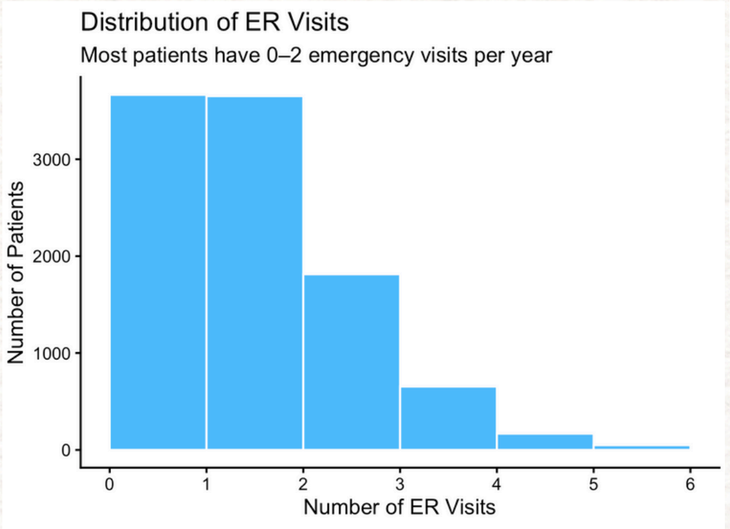
- 24.3% have asthma (n=2,433)
- Only 3.5% well-controlled
- 96.5% poorly/not controlled
- Average 1 ER visit annually

Behavioral & Environmental:

- 60.7% never-smokers
- 49% moderate air pollution
- Mean medication adherence: 50%
- 30% have family history







Research Question 1

What demographic, lifestyle, and environmental factors influence asthma severity?

Method Used: Ordinal Logistic Regression (POLR)

- Asthma Control Level has three ordered categories

(Well Controlled → Poorly Controlled → Not Controlled).

- Ordinal Logistic Regression is the correct method when the outcome is ordered but not numeric.
- Captures changes in severity across levels and estimates the direction & strength of predictors.

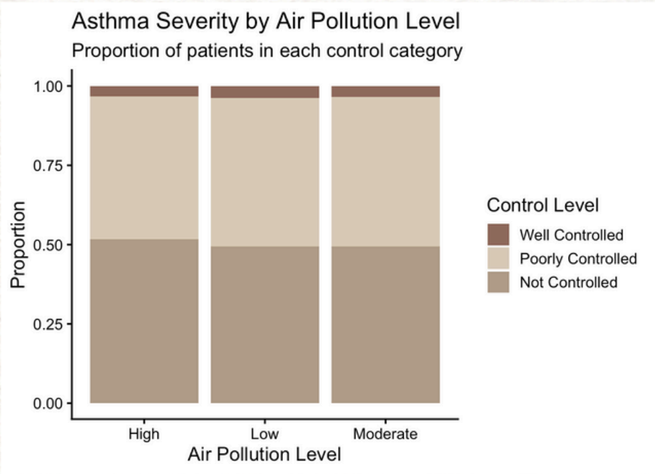
Key Results (Outputs)

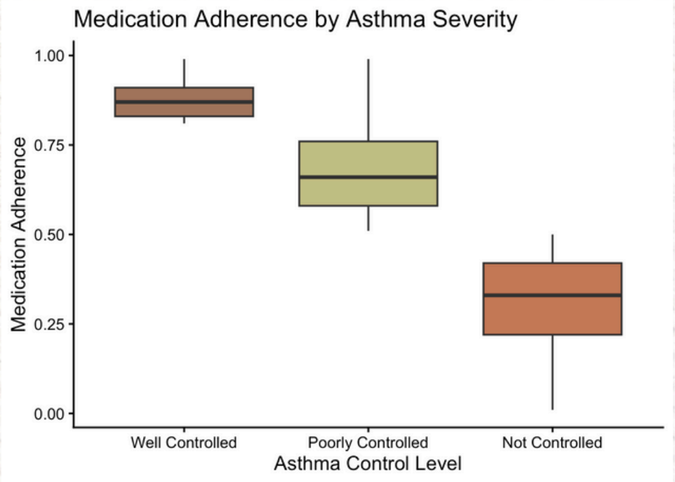
Significant predictors:

- Higher Air Pollution → higher odds of poor control
- Lower Medication Adherence → worse severity
- Smoking Status, BMI, Age, and Comorbidities also significant

Interpretation:

Patients with low adherence and high pollution exposure are much more likely to fall into poorly or not-controlled asthma categories.



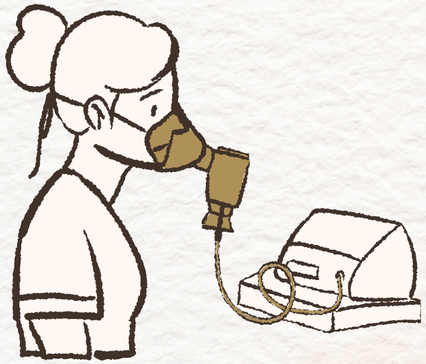


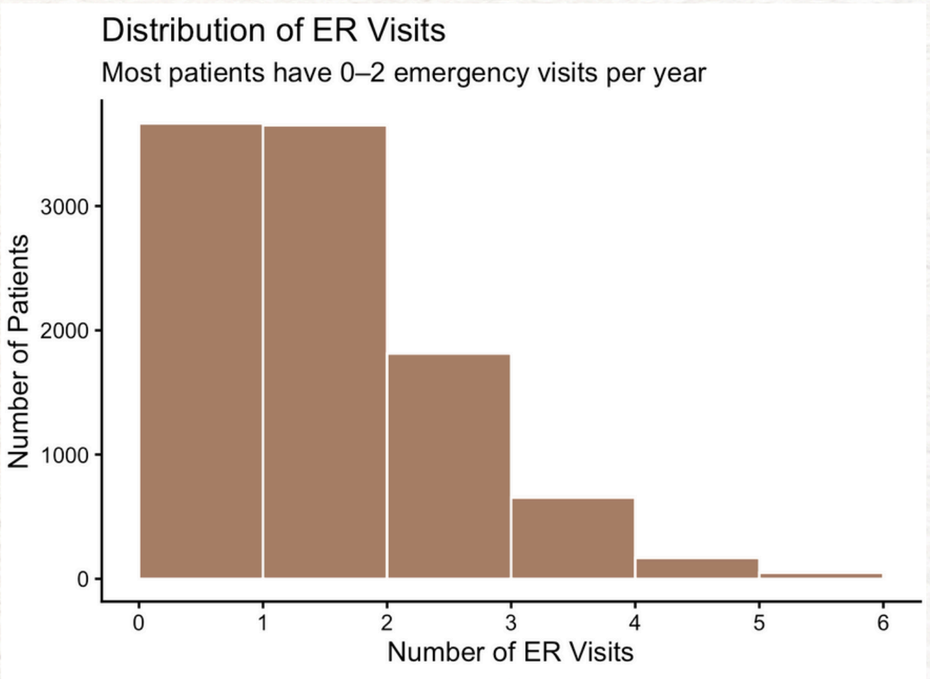
Research Question 2

Can we predict likelihood of repeated ER visits (proxy for readmission)?

Method Used: Binary Logistic Regression

- Outcome variable: Multiple ER Visits (2+ per year)
- Binary outcome makes logistic regression the correct method
- Estimates probability of high healthcare utilization

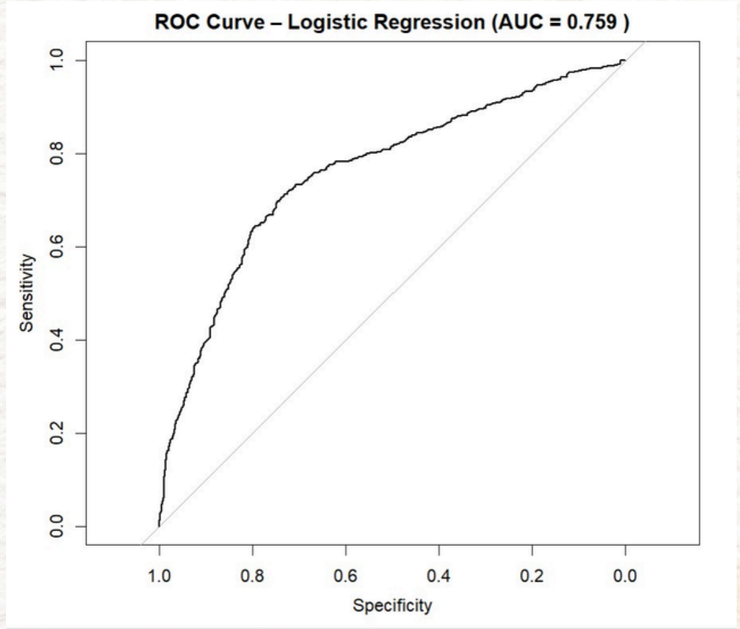


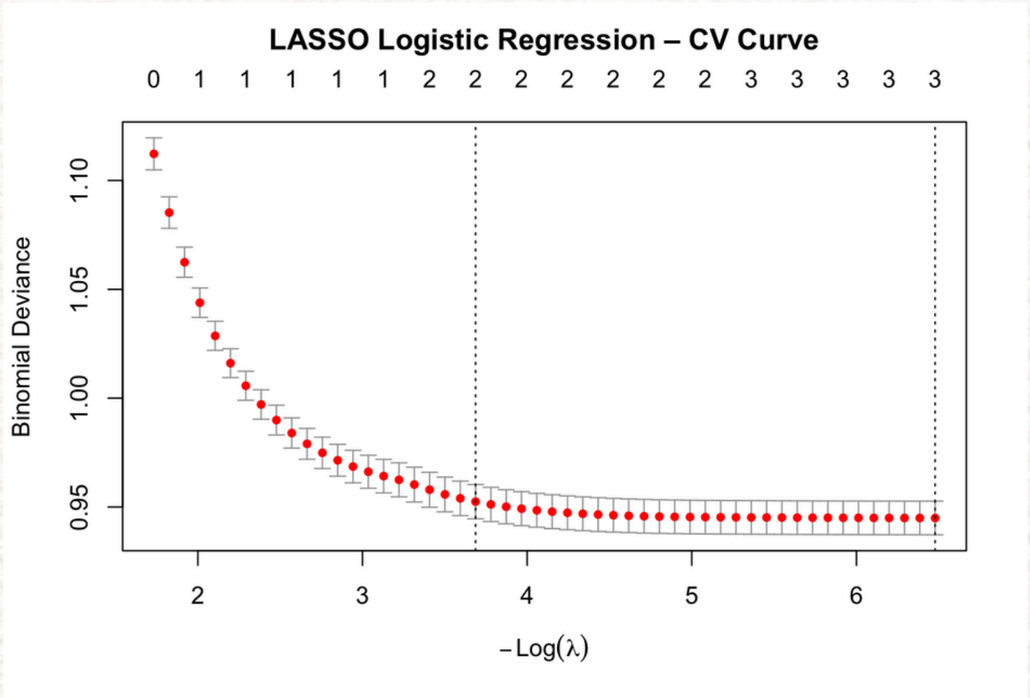


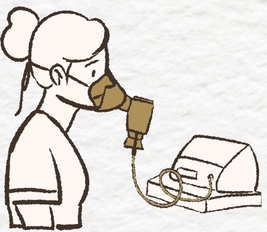
Research Question 2

Key Results:

- High FeNO, low medication adherence, smoking, comorbidities, and high pollution significantly increase the odds of repeated ER visits
- Model performance: AUC ≈ 0.759 (good predictive power)
- LASSO logistic regression was used for variable selection and showed similar AUC.







Research Question 3

3.

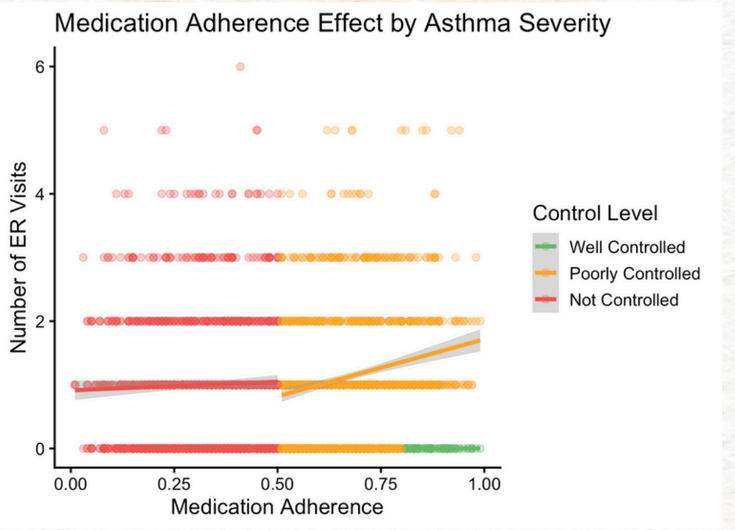
How do medication adherence and air pollution jointly affect patient outcomes?

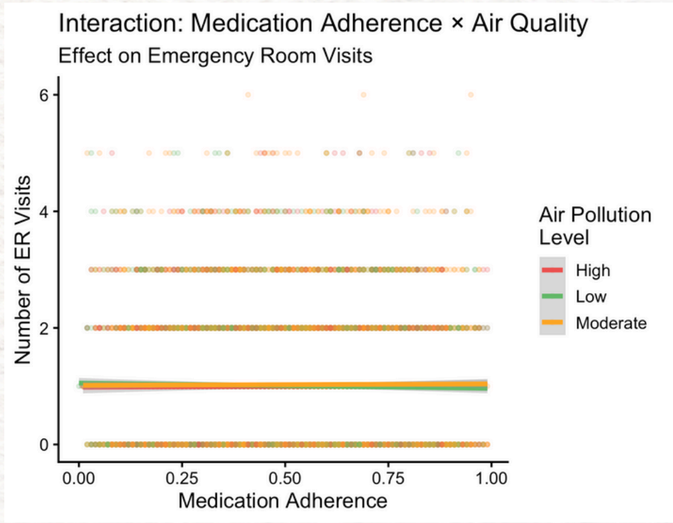
Method Used:

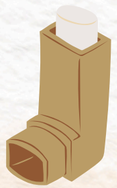
- Linear Regression with Interaction Term
- Poisson Regression for validation (count outcome)

Key Results:

- Interaction term is statistically significant
- High pollution weakens the benefit of medication adherence
- Patients with low adherence under high pollution conditions show the highest ER visit rates







Non-Parametric Methods Overview

Why Non-Parametric Tests?

- Many variables (ER visits, FeNO, BMI) are skewed or ordinal.
- Non-parametric methods do not assume normality and are ideal for clinical data.
- Used to compare groups and check associations when parametric assumptions fail.

Tests Used

1. Kruskal-Wallis Test
   - Compared ER visits across asthma severity groups.
   - Result: Significant differences → worse severity = more ER visits.
2. Pairwise Wilcoxon Tests
   - Compared each severity group pair.
   - All comparisons significant.
3. Wilcoxon Rank-Sum (BMI vs Asthma)
   - Tested BMI differences between asthma and non-asthma patients.
   - Result: Significant difference in BMI across groups.

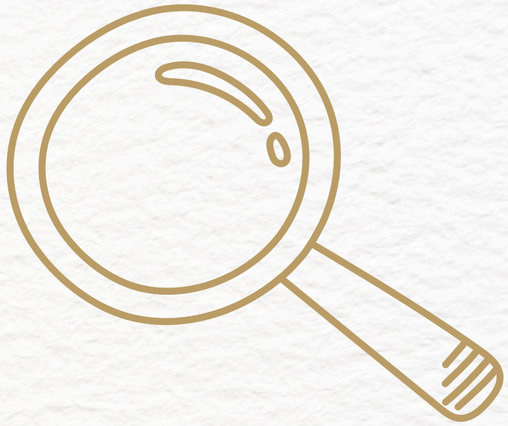
Conclusion

- Key Predictors Identified:
  Air pollution, medication adherence, BMI, smoking status, FeNO levels, and comorbidities were found to significantly influence asthma severity and ER visit frequency.

- Strong Predictive Performance:
  The models effectively identified high-risk patients, and interaction analysis revealed that poor air quality weakens the positive impact of good medication adherence.

- Future Enhancement Opportunities:
  Incorporating time-series data, real hospital readmission records, broader environmental variables, and developing a real-time predictive dashboard can further strengthen asthma risk monitoring and personalized patient management.

References



Synthetic Asthma Dataset (2025). Asthma Management - Identifying Key Risk Factors and Outcomes. Kaggle.

Field, A. (2017). Discovering Statistics Using R. Sage Publications.
Centers for Disease Control and Prevention (CDC). Asthma Data & Surveillance (2024).

Thank You