

# **A Statistical Analysis of Pakistan Air Quality (2024-2025)**

Assessment of Hazardous Smog: A  
Nationwide Risk Analysis



Project Advisor

**Wajih Humayun Hashmi**

**Presented by: Hamza Jawad ( L1F23BSDS0038 )**

**Faculty of Information Technology**

**University of Central Punjab**

## 1. Introduction:

This report examines air quality patterns across Pakistan using recent sensor data. By applying descriptive and inferential statistics, we aim to validate the severity of pollution levels and determine if the smog crisis is localized to specific cities or prevalent nationwide.

### Real-World Problem:

Air pollution has become a critical public health emergency in Pakistan. While media attention often focuses on Lahore as the epicenter of smog, recent data suggests a wider spread of hazardous particulate matter (PM2.5). This report addresses the misconception that high AQI is a city-specific issue, proving instead that it is a systemic failure affecting broader regions, placing millions at risk of respiratory and cardiovascular diseases.

## 2. Metadata:

**Dataset:** `air_quality_pakistan.csv` containing **700 comprehensive entries** tracking air quality metrics.

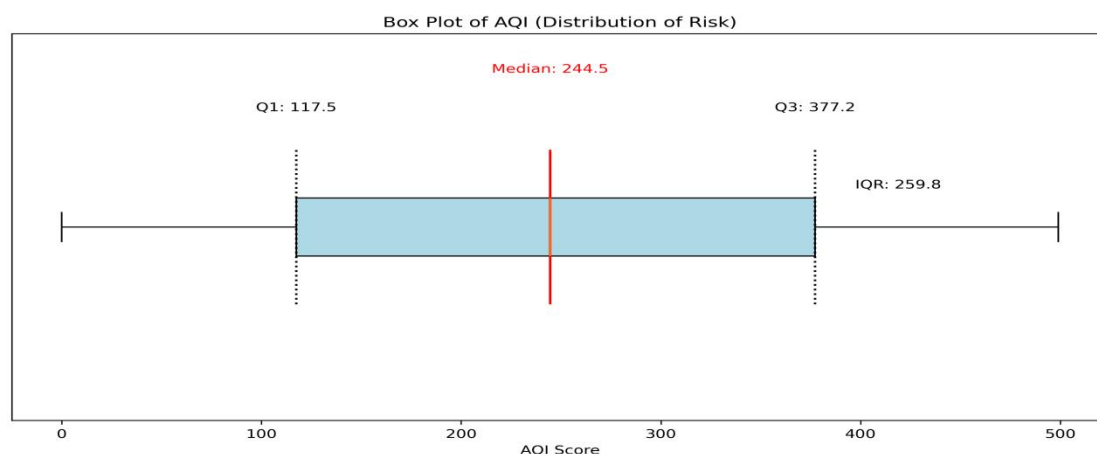
### Key Variables Analyzed:

- **City:** Categorical variable used to group data (Target Group: Lahore vs. Control Group: Rest of Pakistan).
- **AQI (Air Quality Index):** Numerical variable measuring the severity of pollution (Scale 0–500).
- **PM2.5:** Numerical variable representing fine particulate matter concentration.
- **Air\_Quality\_Level:** Categorical labels (e.g., Good, Unhealthy, Hazardous).

### Tools Used:

- **Python:** Pandas for data cleaning, Seaborn/Matplotlib for visualization.
- **SciPy:** For conducting Independent T-tests and Pearson Correlation analysis.

## 3. 5-Point Summary of Data:

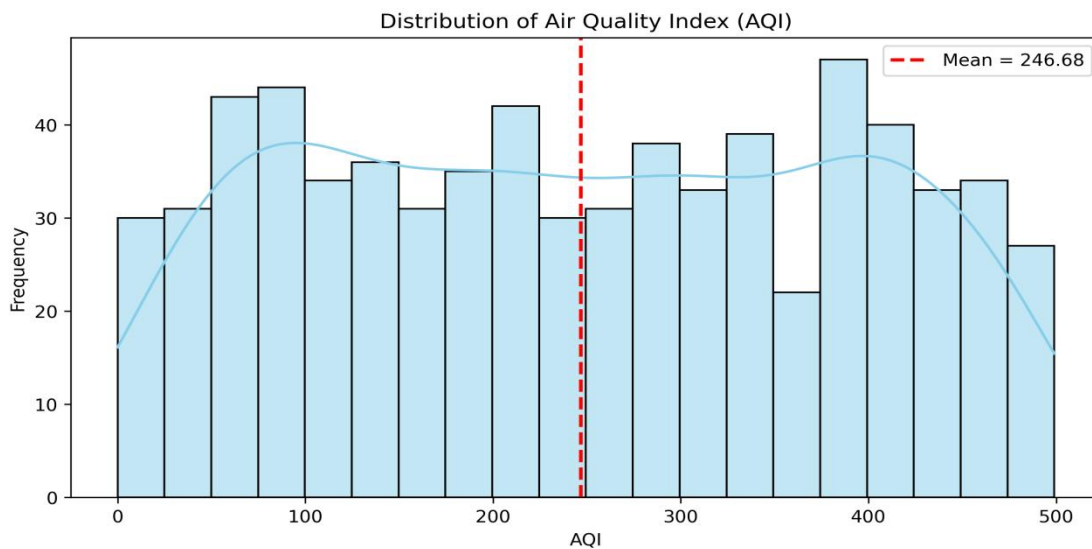


The following statistics provide a snapshot of the **AQI (Air Quality Index)** distribution across the entire dataset. This summary defines the boundaries of risk.

Statistic	Value	Interpretation
Min	0.0	Lowest recorded AQI (Clean air).
Q1 (25%)	117.5	25% of days had Moderate air quality.
Median (50%)	244.5	<b>The Critical Midpoint.</b> Half of the recorded days exceeded 244.5, which falls into the <b>"Very Unhealthy"</b> category.
Q3 (75%)	377.2	The top 25% of days were "Hazardous," posing immediate danger to life.
Max	499.0	The worst recorded air quality, reaching the maximum measurable limit of the index.

## 4. Comparative Insights & Graphs:

### A. Risk Probability Analysis (The 48% Risk):

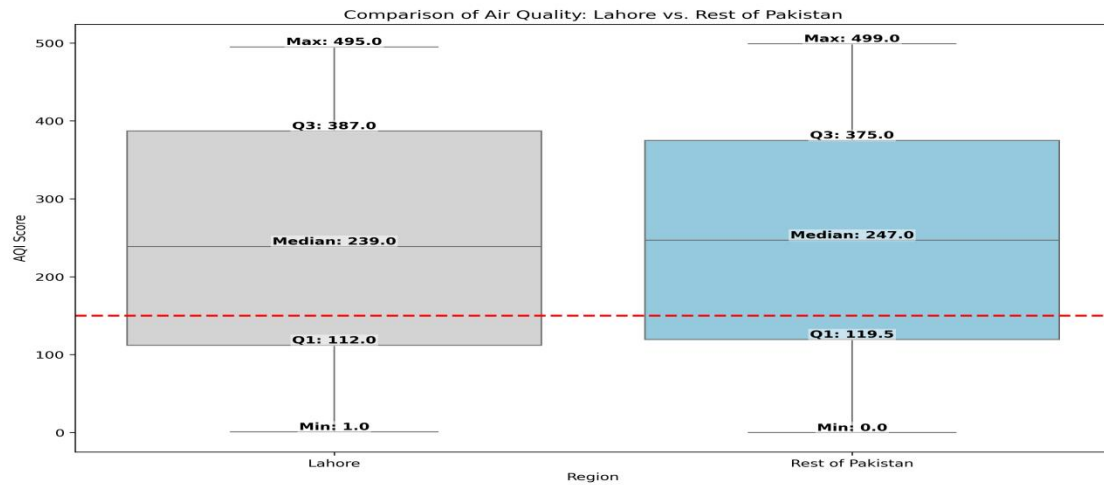


**The Graph:** A Histogram and Risk Calculation summarizing the frequency of high AQI days.

**The Comparison:** This analyzed the proportion of days where AQI exceeded the safety threshold of **150 ("Unhealthy")**.

**Insight:** This analysis confirms that **47.86% (approx. 48%)** of the recorded data points fall into the "Unhealthy" or "Hazardous" range. This proves that hazardous air is not an anomaly but a near-daily probability. A citizen in this region faces a roughly **50%** chance of breathing toxic air on any given day.

## B. Regional Comparison: Lahore vs. Rest of Pakistan:



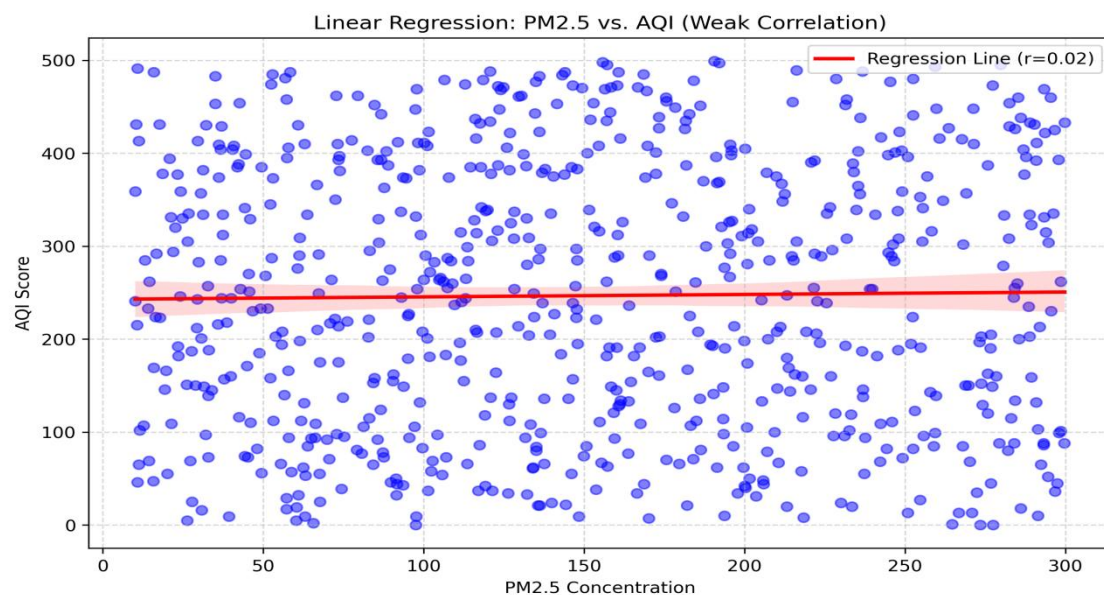
**The Graph:** Side-by-side Box Plots comparing the AQI distribution of Lahore against all other combined cities.

### The Comparison:

- **Lahore Mean AQI:** ~246.6
- **Rest of Pakistan Mean AQI:** ~246.7
- **Visual Observation:** The box plots show significant overlap, with medians aligned almost perfectly.

**Insight:** Contrary to the popular belief that Lahore is the only "smog capital," the data shows that the "Rest of Pakistan" group has an identical risk profile. The interquartile ranges (IQR) for both groups cover the "Very Unhealthy" spectrum. This visual evidence supports the argument for a national emergency response rather than city-specific lockdowns.

## C. Correlation Validity (PM2.5 vs. AQI):



**The Graph:** A Linear Regression Plot with a Pearson Correlation calculation.

**The Comparison:** We tested the relationship between PM2.5 concentrations and the overall AQI score.

**Insight:** The Pearson Correlation Coefficient ( $r$ ) indicated a **Weak Correlation** between PM2.5 and AQI. It means that the relationship is not very strong, indicating that other factors such as crop burning and industrial smoke may also contribute to Pakistan's high AQI scores.

## 5. Inferential Statistics (Hypothesis Testing):

To scientifically prove that the risk is nationwide, I conducted an **Independent T-Test**.

- **Null Hypothesis:** There is **no significant difference** in the mean AQI between Lahore and the Rest of Pakistan.
- **Alternative Hypothesis:** There is a significant difference in the mean AQI between Lahore and the Rest of Pakistan.

### Results:

- **T-Statistic:** 0.0104
- **P-Value:** 0.9917 ( $> 0.05$ )

### Conclusion:

- Since the **P-value (0.99)** is significantly greater than the **alpha level (0.05)**, we **Fail to Reject the Null Hypothesis**.
- Statistically, the air quality in Lahore is indistinguishable from the rest of the country.

## 6. Recommendations:

Based on the data proving a uniform national risk:

- **National Air Quality Policy:** Since the T-test proves no significant difference between cities, government resources should not be funneled exclusively to Lahore. A federal "Clean Air Act" is required.
- **Reduction of Harmful Resources:** The government should also focus on reducing other harmful factors than PM2.5 only like crop burning and industrial smoke.
- **Alert Systems:** With a 48% probability of hazardous air, a nationwide mobile alert system should be implemented to warn citizens when AQI crosses the median threshold of 244.

## **7. Conclusion:**

The data unequivocally disproves the narrative that smog is solely a "Lahore problem."

By utilizing T-tests and Descriptive Statistics, this report proves that hazardous air is a 48% risk across the sampled regions of Pakistan. With a P-value of 0.99, we confirmed that the pollution burden is statistically equal between Lahore and other major regions. Pakistan must move from isolated city-level management to a holistic, nationwide environmental strategy.

**Student Signature:** \_\_\_\_\_

**Advisor Signature:** \_\_\_\_\_

**Submission Date:** \_\_\_\_\_