Full Analysis

Boeing 737/757 flights — JFK, LGA, EWR — March 2013

1) Sample and Baseline Finding (What the Data Shows Before Modeling)

After dataset merges, filtering for Boeing 737/757 aircraft, and removing rows with missing departure delay or visibility, the analytical sample contains:

- N = 5769 flights

Exploratory visualizations (overall boxplot and airport-faceted plots) show that as visibility deteriorates, the entire distribution of departure delays shifts upward, not just extreme values.

Operational interpretation:

This indicates a systemic capacity constraint, rather than isolated disruption events.

2) Main Statistical Findings (Controlled Effect)

2.1 OLS: Departure Delay in Minutes (HC3 Robust Errors)

Model specification:

> dep_delay ~ visib + C(origin) + C(hour) + wind_speed + precip + temp + pressure + humid

Key result:

- Visibility coefficient = −3.099 minutes per mile
- 95% CI: [−4.774, −1.424]
- $p < 0.001$

Interpretation:

An improvement of 1 mile in visibility is associated with approximately 3.1 minutes less departure delay on average, after controlling for airport, time of day, and other weather variables.

2.2 Logistic Model: Probability of Significant Delay (>15 min)

Model specification:

is_delayed_15 ~ visib + C(origin) + C(hour) + controls

Key result:

- Visibility coefficient = −0.156 (log-odds per mile)
- 95% CI: [−0.225, −0.088]
- $p < 0.001$

Interpretation:

Better visibility significantly reduces the probability that a flight departs more than 15 minutes late.

3) Operational Impact (Scenario Analysis)

To translate statistical results into operational terms, scenario predictions were computed within the observed data range:

- High visibility: 10 miles
- Low visibility: 2 miles

All other variables were held at representative observed values.

3.1 Predicted Mean Delay (OLS)

Results:

- Predicted delay @ 10 miles: 6.12 minutes
- Predicted delay @ 2 miles: 30.91 minutes
- Impact: +24.79 minutes

Managerial interpretation:

A deterioration from good to poor visibility increases expected departure delay by roughly 25 minutes.

3.2 Airport-Specific Predictions

Airport	Delay @10 mi	Delay @2 mi	Increase
EWR	8.48	33.28	+24.79
JFK	4.20	28.99	+24.79
LGA	9.61	34.40	+24.79

Important note:

The increase is identical across airports because the model estimates a common visibility effect (no interaction between visibility and airport). Differences between airports appear only in baseline delay levels.

4) Robustness Checks (Ensuring Results Are Not Driven by Outliers)

Departure delay data is heavily right-skewed with extreme values. Additional models were estimated to test robustness.

4.1 Winsorized OLS (1st–99th Percentile Clipping)

- Visibility coefficient = −2.68
- Still highly significant ($p < 0.001$)

Interpretation:

The negative relationship remains strong even after limiting the influence of extreme delays, though the effect size decreases as expected.

4.2 Median (Quantile) Regression

- Visibility coefficient ≈ −0.98 minutes per mile
- $p < 0.001$

Interpretation:

Visibility affects the typical flight (median delay), but the effect is smaller than for the mean. This indicates that poor visibility disproportionately increases the likelihood of severe disruption events, which inflate average delays.