ANOVA

May 7, 2025

1 Dataset Collection

We collected daily close prices for Boeing and Airbus from 2010 to 2025 using Yahoo Finance. Accident records for the same period were obtained from the National Transportation Safety Board (NTSB). Key variables used include: `Date` for aligning accident and stock data; `State` and `Country` for regional grouping; `Make` to identify manufacturer; injury counts such as `FatalInjuryCount`, `SeriousInjuryCount`, and `MinorInjuryCount` to calculate severity scores. Stock price data was linked by `Date` to generate `boeing_Close` and `airbus_Close` time series.

2 Data Processing

The Make column contains 1,542 entries labeled as Boeing and 422 as Airbus, along with 27 missing values. While the missing values represent only 0.11% of the entire dataset (0.0011), they account for approximately 1.37% of the combined Boeing and Airbus entries (0.0137). Given the small proportion and the focus of our analysis on these two manufacturers, it is reasonable to remove the rows with missing Make values without introducing significant bias.

The `State` column contains approximately 19.5% missing values. However, these are not due to random omission but stem from structural differences—nearly all missing entries come from countries where a U.S.-style state designation does not exist (Milliken & Johnson, 2009). Therefore, **this missingness is considered structural rather than stochastic.** In our analysis, we account for this by limiting state-based blocking strategies to data from the United States, where the `State` information is both relevant and largely complete.

3 ANOVA

Model 1 Analysis: Effect of Manufacturer and Region on Injury Severity

Overview. In this model, we investigate whether the severity of aircraft accidents—quantified via a weighted injury score—is associated with the manufacturer (`Make`) or the incident region's activity level (`StateBlock`). The weighted score was defined as: 3×`Fatal + 2`×`Serious + 1`×`Minor`. We limited the analysis to accidents involving Boeing and Airbus, and grouped U.S. states into High / Medium / Low frequency blocks based on incident counts. Given the unbalanced sample sizes across groups, we directly applied a Type III ANOVA, which provides robust main effect estimates by adjusting for all other terms in the model.

Key Analysis Flow

- Cleaned data by removing rows with missing values in `Make`, `State`, and `Country`.
- Created a new variable `StateBlock` from state-wise accident counts.
- Computed `WeightedInjuryScore` using fatal, serious, and minor injury counts.
- Applied Type III ANOVA due to group size imbalance to assess main effects of `Make` and `StateBlock`.
- Diagnosed the model via residual plots, which indicated heteroscedasticity.
- Used Weighted Least Squares (WLS) regression to re-estimate model coefficients.
- Conducted Tukey HSD test on `StateBlock` for pairwise comparison.

Detailed Interpretation

Type III ANOVA. Given the substantial imbalance in group sizes—e.g., over 14,000 incidents in `High` versus only 1,300 in `Low`—we used Type III ANOVA, which evaluates each factor's effect after accounting for all other variables. The result showed that `Make` had no significant effect ($p = 0.965$), whereas `StateBlock` was initially significant ($p = 0.016$).

Residual Diagnostics. The Figure 1 showed a fan-shaped pattern, indicating increasing variance with fitted values. This heteroscedasticity violates ANOVA assumptions and undermines the reliability of F-tests.

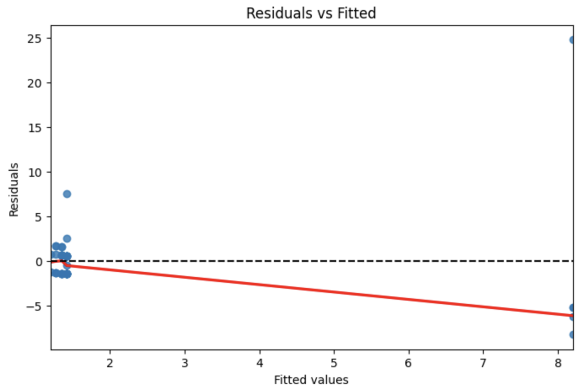

Figure 1: Residuals vs Fitted Values Plot for Model 1 Showing Heteroscedasticity

WLS Regression. To correct for heteroscedasticity, we applied WLS regression using the inverse of squared residuals as weights. As shown in **??**. The updated model showed poor fit ($R^2 = 0.027$), and neither `Make` nor `StateBlock` remained significant ($p > 0.4$ for all).

```
                            WLS Regression Results
==============================================================================
Dep. Variable:     WeightedInjuryScore   R-squared:                       0.027
Model:                             WLS   Adj. R-squared:                 -0.042
Method:                  Least Squares   F-statistic:                    0.3956
Date:                 Wed, 07 May 2025   Prob (F-statistic):              0.757
Time:                         15:22:40   Log-Likelihood:                -75.437
No. Observations:                   46   AIC:                             158.9
Df Residuals:                       42   BIC:                             166.2
Df Model:                            3
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.7258      0.221      7.795      0.000       1.279       2.173
x1            -0.2001      0.271     -0.740      0.464      -0.746       0.346
x2             1.2867      2.852      0.451      0.654      -4.468       7.042
x3            -0.2644      0.401     -0.659      0.514      -1.074       0.545
==============================================================================
Omnibus:                      196.811   Durbin-Watson:                   1.547
Prob(Omnibus):                  0.000   Jarque-Bera (JB):                5.984
Skew:                          -0.035   Prob(JB):                       0.0502
Kurtosis:                       1.234   Cond. No.                         27.0
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 2: WLS Regression Summary Table for Model 1

Post Hoc Comparison (Tukey's HSD). To identify which levels of `StateBlock` differ significantly, we applied Tukey's Honest Significant Difference (HSD) test.

The results showed that:

- `High` vs. `Low` states differ significantly ($p = 0.009$).
- `Low` vs. `Medium` states also show significant difference ($p = 0.028$).
- However, `High` vs. `Medium` states did not differ significantly ($p = 0.997$).

This indicates that the `Low`-incident states have systematically different injury scores compared to both `High` and `Medium` blocks, suggesting that lower exposure (or fewer cases) might be associated with higher injury severity per case.

Conclusion

Although the initial Type III ANOVA suggested that `StateBlock` had a significant effect, residual diagnostics revealed clear heteroscedasticity, potentially undermining the validity of the F-tests. After applying Weighted Least Squares (WLS) to address variance inequality, no predictors remained significant. However, post-hoc Tukey's HSD tests revealed that `Low` frequency states significantly differ from both `High` and `Medium` blocks. This suggests potential regional disparities in injury severity that warrant further investigation. Meanwhile, `Make` (Airbus vs. Boeing) consistently showed no association with injury severity.

Model 2: One-Way ANOVA on Standardized Prices (Airbus vs Boeing)

Overview. Initial analysis on raw close prices indicated a significant mean difference between Airbus and Boeing stock prices. However, this difference likely stemmed from inherent scale discrepancies. To address this, we standardized closing prices within each company using z-score normalization and then applied one-way ANOVA.

- The dependent variable is the standardized closing price.
- The independent variable is the company label (Airbus vs Boeing).
- Timeframe was aligned from 2010 onward to ensure comparable periods.

Rationale. Raw stock prices are influenced by company-specific factors such as initial offering, historical valuation, and currency denomination. Standardization removes scale effects and allows fair comparison of relative fluctuations and trends between companies.

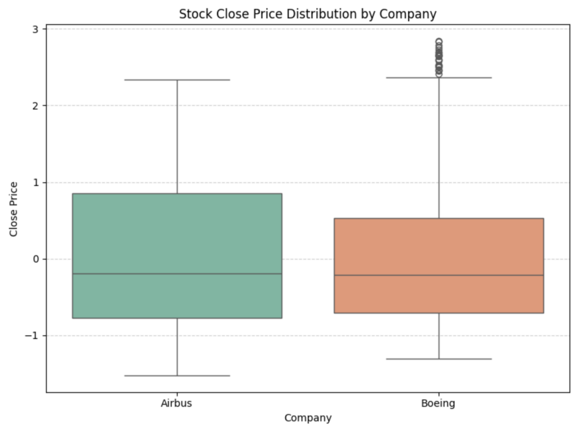

Figure 3: Boxplot of Standardized Close Prices (2010–2025) by Company

Boxplot Visualization. **??** shows that after standardization, the distributions for Airbus and Boeing stock prices appear similar in range and central tendency.

ANOVA and Post-Hoc Results. The one-way ANOVA yielded an F-statistic near 0 with $p = 1.0$, indicating no significant difference between company means. Tukey's HSD post-hoc test confirmed this: the mean difference was 0.0, with adjusted $p = 1.0$.

Conclusion. Although the raw prices showed a significant difference, standardizing the data revealed that Airbus and Boeing stock prices behave similarly in relative terms. This underscores the importance of addressing scale effects before conducting comparative statistical tests.

Model 3: WLS ANOVA with Interaction – Do Airbus and Boeing React Differently to Fatal Accidents?

Research Goal and Methodology

This model investigates whether Airbus and Boeing stocks react differently to fatal accident events. Specifically, we examine whether stock prices within a short window around fatal accidents significantly differ from normal periods, and whether this response varies by company. This follows prior event study designs, which often define event windows of ± 3 days to capture immediate market reactions (MacKinlay, 1997).

To test this, we use a Weighted Least Squares (WLS) ANOVA model with:

- **Dependent variable:** Daily closing price
- **Fixed effect:** `PeriodType` (FatalAccidentPeriod vs. NormalPeriod)

- **Block:** `Month` (1–12) to control for seasonal variation
- **Grouping factor:** `Company` (Airbus vs. Boeing)
- **Interaction term:** `PeriodType ×Company`

We first label each date as `FatalAccidentPeriod` if it falls within ±3 days of any fatal accident involving the same company. Daily stock data are merged with the accident event list by `Date` and `Company`. Month is extracted as a blocking variable to account for seasonal patterns.

Separate Company Analyses

To assess company-specific responses, we fit separate WLS ANOVA models for Airbus and Boeing using `PeriodType` and `Month`. Residual-based weights from initial OLS fits are applied.

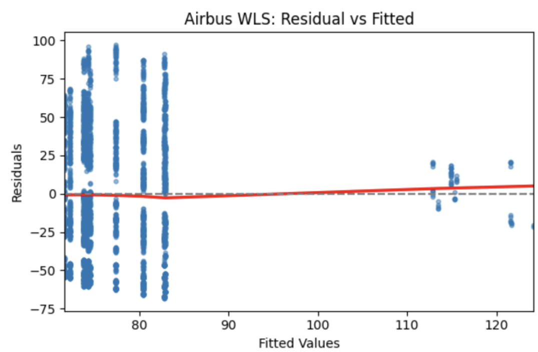

Figure 4: Airbus WLS: Residuals vs Fitted Values

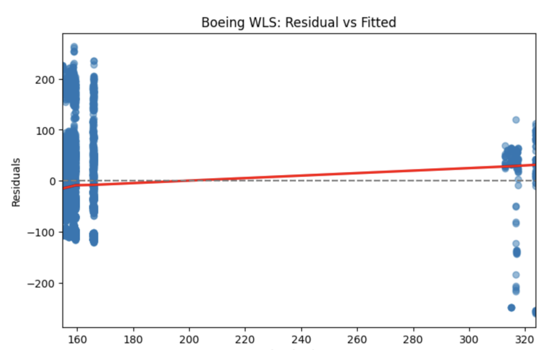

Figure 5: Boeing WLS: Residuals vs Fitted Values

Both models show strong statistical significance ($p < 0.0001$) for `PeriodType`. Residual plots reveal persistent variance instability in low fitted regions, indicating that time-varying volatility may still exist despite WLS correction. This motivates potential use of GARCH models in follow-up analysis.

Interaction Analysis

We then fit a full model with an interaction term `PeriodType ×Company` to test whether the impact of fatal accidents differs between Airbus and Boeing. The interaction effect is highly significant ($F = 1039.3$, $p < 0.0001$).

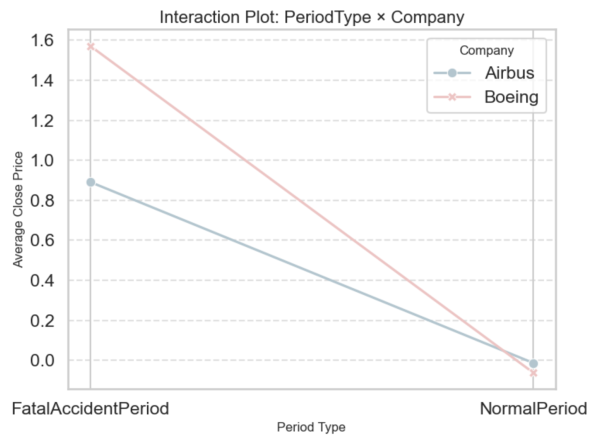

Figure 6: Interaction Plot: PeriodType × Company Effects on Average Close Price

The interaction plot shows that while both companies experience price declines during fatal accident periods, Boeing's drop is more pronounced. The non-parallel trend lines visually support the statistical significance of the interaction.

Conclusion and Implications

These results suggest that:

- `PeriodType` significantly affects stock prices for both companies.
- The `Company` moderates this effect — the market response to fatal accidents is more severe for Boeing.
- Volatility remains an issue even after WLS correction, supporting the case for GARCH modeling in future stages.

This model highlights that fatal accidents not only impact prices but do so differently across firms, emphasizing the importance of including interaction terms and accounting for volatility.

References

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature*, *35*(1), 13–39.

Milliken, G. A., & Johnson, D. E. (2009). *Analysis of messy data volume 1: Designed experiments*. Chapman; Hall/CRC.