A Statistical Comparison of Fatal Accident Rates in Boeing and Airbus Aircraft

Qasim Salahuddin

Why this question

Aviation incidents are often a cause of controversy, and discussions frequently arise regarding the relative safety of aircraft manufacturers. In this short paper we aim to use likelihood-based methods to determine whether there is a statistically significant difference between fatal accident rates of specifically Boeing and Airbus aircraft.

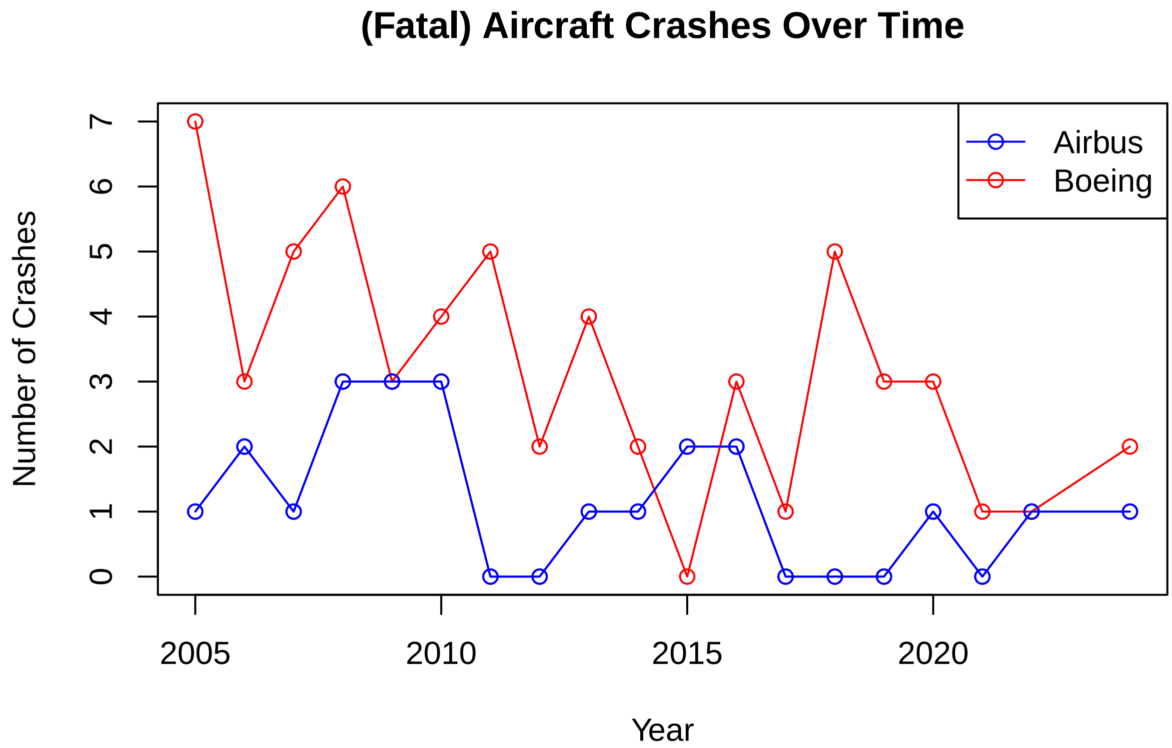
Analysis of this question often does not account for factors such as fleet sizes, number of flights, or other differences in exposure. By considering a statistical approach which estimates accident rates rather than counts, factors overlooked by naive approaches can be taken into account. Public data sets exist with detailed information on aircraft accidents, making this an ideal topic to explore statistically.

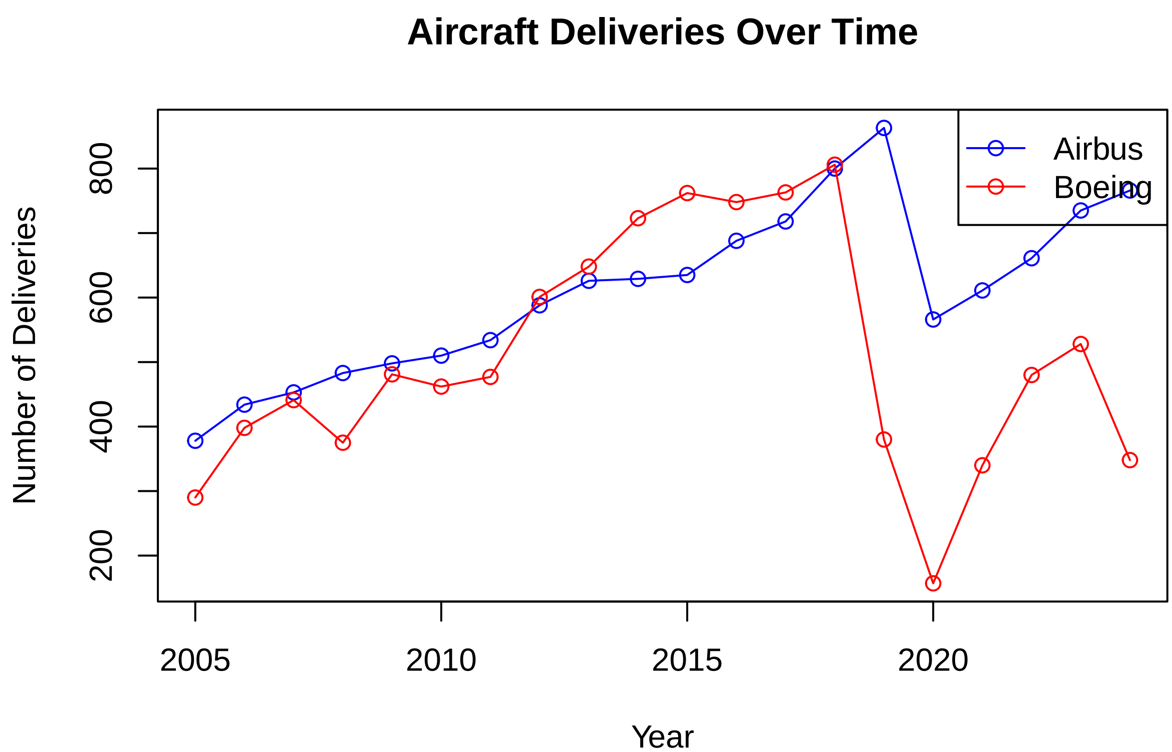
Data sources and preprocessing

In this paper, we will only consider aircraft accidents which resulted in 1 or more fatalities, for simplification. A large data frame is available on Hugging Face[1], which contains extensive fatal accident data from the early 20th century, storing various information about each flight. This data set first needs some processing - this was handled externally via R, where the set was shortened to Boeing and Airbus data from 2005 till 2024. Additionally, we are only interested in the date of the accidents, as well as the flight numbers (for some identifier of each individual entry, if needed). A time series of this data is shown on the next page.

Now, this data alone will not lead us to any meaningful conclusions - we need a way to also weigh up the differences in exposure of the 2 companies. There are many ways to do this - ideally we would want data regarding the total amount of flight hours each aircraft produced by each company, per year. However this type of data is not freely available, so we have opted to use the amount of aircraft delivered to airlines each year as an indication for each companies relative exposure. This is far from ideal, and there are a number of flaws with this method (a detailed discussion of improvements can be found on page 6), but for the purposes of this short paper, this will suffice. Such a data set can be obtained from Wikipedia[2]. A time series of this data can also be viewed on the next page.

[1] https://huggingface.co/datasets/haor/AirSafe_DB
[2] https://commons.wikimedia.org/wiki/Data:Orders_for_and_deliveries_of_Airbus_and_Boeing_aircraft.tab


(Fatal) Aircraft Crashes Over Time
Number of Crashes
Year
Airbus
Boeing
0
1
2
3
4
5
6
7
2005
2010
2015
2020


Aircraft Deliveries Over Time
Number of Deliveries
Year
Airbus
Boeing
200
400
600
800
2005
2010
2015
2020

Model

Let $X_{i,t}$ denote the number of fatal accidents involving aircraft manufactured by company $i \in \{B, A\}$ for each year $t \in \{2005, ..., 2024\}$. Accidents are modelled as rare events, and we assume

$$X_{i,t} \sim \text{Poisson}(\lambda_i E_{i,t}),$$

where λ_i denotes the accident rate for company i and $E_{i,t}$ is the number of deliveries to company i in year t.

A Poisson model seems suitable as the events in question (aircraft accidents) are independent, rare, and occur in a fixed interval. The only problematic assumption is the modelling of the rate parameter λ_i as being homogeneous across all aircraft produced by each company - we will discuss this later in the limitations section.

Maximum Likelihood Estimation

Noting that $X_{i,t}$ are independent (but not i.i.d), we have:

$$L(\lambda_i) = \prod_{t=2005}^{2024} \frac{(\lambda_i E_{i,t})^{X_{i,t}} \, e^{-\lambda_i E_{i,t}}}{X_{i,t}!}$$

where $L(\lambda_i)$ is the likelihood function.

Taking logs:

$$\begin{aligned}
\log L(\lambda_i) &= \sum_{t=2005}^{2024} \Big(X_{i,t} \log(\lambda_i E_{i,t}) - \lambda_i E_{i,t} - \log(X_{i,t}!) \Big) && (1) \\
&= \sum_{t=2005}^{2024} \Big(X_{i,t} \log(\lambda_i) + X_{i,t} \log(E_{i,t}) - \lambda_i E_{i,t} - \log(X_{i,t}!) \Big) && (2)
\end{aligned}$$

Differentiating with respect to λ_i:

$$\begin{aligned}
\frac{d}{d\lambda_i} \log L(\lambda_i) &= \sum_{t=2005}^{2024} \left(\frac{X_{i,t}}{\lambda_i} - E_{i,t} \right) && (3) \\
&= \frac{1}{\lambda_i} \sum_{t=2005}^{2024} X_{i,t} - \sum_{t=2005}^{2024} E_{i,t}. && (4)
\end{aligned}$$

Now setting (4) to 0 to obtain the maximum likelihood estimator:

$$\begin{aligned}
\frac{1}{\hat{\lambda}_i} \sum_{t=2005}^{2024} X_{i,t} &= \sum_{t=2005}^{2024} E_{i,t} && (5) \\
\hat{\lambda}_i &= \frac{\sum_{t=2005}^{2024} X_{i,t}}{\sum_{t=2005}^{2024} E_{i,t}}. && (6)
\end{aligned}$$

(Upon evaluating the second derivative we see that this is indeed a maximum.) Intuitively $\hat{\lambda}_i$ is the total number of accidents divided by the total number of deliveries.

Variance of MLE

Noting that the variance of a Poisson distribution with parameter λ is also λ, we have:

$$\begin{aligned}
\mathrm{Var}(\hat{\lambda}_i) &= \mathrm{Var}\left(\frac{\sum_{t=2005}^{2024} X_{i,t}}{\sum_{t=2005}^{2024} E_{i,t}}\right) &(7)\\
&= \frac{1}{\left(\sum_{t=2005}^{2024} E_{i,t}\right)^2} \mathrm{Var}\left(\sum_{t=2005}^{2024} X_{i,t}\right) &(8)\\
&= \frac{1}{\left(\sum_{t=2005}^{2024} E_{i,t}\right)^2} \sum_{t=2005}^{2024} \mathrm{Var}(X_{i,t}) &(9)\\
&= \frac{1}{\left(\sum_{t=2005}^{2024} E_{i,t}\right)^2} \sum_{t=2005}^{2024} \lambda_i E_{i,t} &(10)\\
&= \frac{\lambda_i \sum_{t=2005}^{2024} E_{i,t}}{\left(\sum_{t=2005}^{2024} E_{i,t}\right)^2} &(11)\\
&= \frac{\lambda_i}{\sum_{t=2005}^{2024} E_{i,t}} &(12)
\end{aligned}$$

where the third equality is justified by independence of $X_{i,t}$. We can also estimate λ_i with $\hat{\lambda}_i$, giving:

$$\mathrm{Var}(\hat{\lambda}_i) \approx \frac{\hat{\lambda}_i}{\sum_{t=2005}^{2024} E_{i,t}}$$

and

$$\widehat{\mathrm{SE}}(\hat{\lambda}_i) \approx \sqrt{\frac{\hat{\lambda}_i}{\sum_{t=2005}^{2024} E_{i,t}}}$$

Confidence Interval

By an asymptotic property of MLE's[3]:

$$\frac{\hat{\lambda}_i - \lambda_i}{\sqrt{\lambda_i / \sum_t E_{i,t}}} \xrightarrow{d} \mathcal{N}(0,1),$$

as the total exposure $\sum_t E_{i,t}$ increases. Let $Z \sim \mathcal{N}(0,1)$ and let $z_{\alpha/2}$ denote the $(1-\alpha/2)$ quantile of the standard normal distribution, defined by

$$\mathbb{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha.$$

For $\alpha = 0.05$, $z_{0.025} = 1.96$, yielding the approximate confidence statement

$$\mathbb{P}\Big(\hat{\lambda}_i - 1.96\,\widehat{\mathrm{SE}}(\hat{\lambda}_i) \leq \lambda_i \leq \hat{\lambda}_i + 1.96\,\widehat{\mathrm{SE}}(\hat{\lambda}_i)\Big) \approx 0.95.$$

[3] Statistical Modelling 1 Lecture Notes, Theorem 5, Dr Riccardo Passeggeri, Imperial College London

Hypothesis Test

To assess whether there is a statistically significant difference in fatal accident rates between Boeing and Airbus aircraft, we consider the hypotheses

$$H_0: \ \lambda_B = \lambda_A \qquad \text{against} \qquad H_1: \ \lambda_B \neq \lambda_A,$$

where λ_B and λ_A denote the fatal accident rates per delivered aircraft for Boeing and Airbus respectively.

Under the Poisson model, the maximum likelihood estimators $\hat{\lambda}_B$ and $\hat{\lambda}_A$ are asymptotically normal and independent, with

$$\hat{\lambda}_i \approx \mathcal{N}\left(\lambda_i, \ \frac{\lambda_i}{\sum_{t=2005}^{2024} E_{i,t}}\right), \qquad i \in \{B, A\}.$$

Consequently, under the null hypothesis H_0, the difference $\hat{\lambda}_B - \hat{\lambda}_A$ satisfies

$$\frac{\hat{\lambda}_B - \hat{\lambda}_A}{\sqrt{\frac{\lambda_B}{\sum_{t=2005}^{2024} E_{B,t}} + \frac{\lambda_A}{\sum_{t=2005}^{2024} E_{A,t}}}} \xrightarrow{d} \mathcal{N}(0,1).$$

Substituting the unknown parameters by their maximum likelihood estimates yields the Wald test statistic

$$Z = \frac{\hat{\lambda}_B - \hat{\lambda}_A}{\sqrt{\frac{\hat{\lambda}_B}{\sum_{t=2005}^{2024} E_{B,t}} + \frac{\hat{\lambda}_A}{\sum_{t=2005}^{2024} E_{A,t}}}}.$$

We now compute this statistic with our data (time series are shown on page 2):

$$\sum_{t=2005}^{2024} X_{A,t} = 22, \qquad \sum_{t=2005}^{2024} X_{B,t} = 60$$

$$\sum_{t=2005}^{2024} E_{A,t} = 12176, \qquad \sum_{t=2005}^{2024} E_{B,t} = 10208$$

Implying

$$\hat{\lambda}_A \approx 0.001806 \qquad \hat{\lambda}_B \approx 0.005878$$

Finally, we have

$$|Z| \approx 4.7851$$

For a significance level $\alpha = 0.05$, the null hypothesis is rejected if $|Z| > 1.96$. Hence the results are statistically significant - there is sufficient evidence to warrant the rejection of H_0; the evidence favours H_1, specifically implying $\lambda_B > \lambda_A$.

Limitations, Improvements and Critical Remarks

There are a few key points to consider after looking at our final results:

- When defining the Poisson model, we chose to use a single λ_i to represent a homogeneous accident rate across all aircraft within each company - this would be far from true in reality. For example, the Boeing 737 Max 8 was involved in 2 major accidents shortly after its introduction, while the Boeing 787 has an almost flawless safety record. In an ideal model, each aircraft type should be considered separately, each with their own Poisson parameter $\lambda_{i,j}$, where j would represent the different aircraft manufactured by each company.

  The model also does not account for improvements in aircraft safety over time - while older aircraft may be differentiated from newer aircraft with better safety records via our individual model approach, it would still be difficult to account for improvements in regulations and maintenance cycles. Aviation procedures are constantly improving, and one would expect over time that fatal accidents would decline. A Poisson model taking this into account would perhaps need further tinkering with its parameter, or it may be the case that a different distribution altogether would be a better model.

- Perhaps the single biggest flaw in our whole model was the use of aircraft deliveries as our exposure measure. To recap, after taking into account all the different fatal accidents by each liner, another source of data was needed to factor in the differences in fleet size between the 2 companies. Aircraft delivery data was chosen as it was readily available, however it is in no way the best method to capture exposure. Aircraft deliveries tell us that an airliner placed an order for a specific aircraft (even this would be a few years before when the delivery was recorded). A higher number of aircraft deliveries in a year then does not translate into a higher number of flights taken with aircraft produced by that manufacturer in that year. In reality, one aircraft may have flown many times the number of flights another aircraft flew in the same year. It also does not account for existing fleet sizes.

  A better indicator of exposure would be the total number of flights flown by Airbus or Boeing aircraft in a given year. The total number flight hours are not even necessary - a 2 hour flight still undergoes the same critical phases of flight as a 16 hour flight (takeoff, landing, pressurising and depressurising the cabin), and it would be unrealistic to say that the latter has an 8x higher risk of an accident. The only drawback of this type of data is that it is hard to obtain.

  If total number of flight hours had been used, it could have major effects on our final hypothesis test. There is a good chance that there would not have been enough evidence to reject H_0.

- It is worth mentioning that in all our statistical analysis, we are only attempting to describe a correlation or a comparison between data - we cannot make a causal statement in our conclusion. In fact making a causal statement regarding such a question is extremely difficult for the following reason.

  When a fatal accident occurs, and investigative teams perform their analysis, the final reports will not always state the cause of the accident as a manufacturing issue. Weather, pilot error, maintenance issues, or a combination of all 3 are all common causes of incidents. There is also no guarantee that the cause of the incident in the final crash report is decisive - for example flight MH370 is still speculative regarding the reason for its disappearance. There is also almost always controversy surrounding investigative reports, with some questioning cover ups or diverting blame.

This means that it would be almost impossible to make a statement such as 'Airbus aircraft are more safer than Boeing aircraft', and certainly cannot be made with statistics alone.

Critical Use of AI

AI played a number of roles in this project. The source of the fatal accidents data set claim that AI had been used to scrape the data and process it, before it was further processed for the purposes of this paper. AI has also been used to assist with the R code to generate plots, as well as with general assistance with LaTeX. The statistical method was also developed with assistance from AI, where, for example, it recommended using an exposure variable within the Poisson parameter. It also assisted with general R Markdown issues (where this paper was written in).

Conclusion

To conclude, we have showed that with our model and data, for a significance level $\alpha = 0.05$, there is evidence to suggest that Boeing aircraft have a higher accident rate than Airbus aircraft.

However, our model was heavily simplified via multiple assumptions, and was hindered by not having access to better data. If this project was done again, the model should be enhanced to factor in time, variations in aircraft, as well as including a more precise exposure estimate.