

# Plane Crash Analysis: Does your seat matter?

## Inferential Statistics 2025/2026

Barbato Alberto, Naldoni Martina

2026-01-14

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Vocabulary . . . . .	2
1.2	Literature . . . . .	2
1.3	Perspective . . . . .	3
1.4	Data Gathering Process . . . . .	4
<b>2</b>	<b>Data Description</b>	<b>4</b>
2.1	Data Points and Variables . . . . .	5
2.2	Mortality Rates . . . . .	6
<b>3</b>	<b>Data Analysis</b>	<b>7</b>
3.1	Preliminary analysis . . . . .	7
3.2	Statistical Analysis of differences in mortality rates between sections . . . . .	11
3.3	Modeling the mortality rates . . . . .	12
<b>4</b>	<b>Analysis</b>	<b>17</b>
<b>5</b>	<b>Results</b>	<b>17</b>
<b>6</b>	<b>Conclusions</b>	<b>17</b>
<b>7</b>	<b>Links</b>	<b>17</b>

## 1 Introduction

The Aviation industry is one of the safest in the world. The systems that it has put in place to learn lessons from accidents and make sure they never happen again is well-established and highly regarded. This has made traveling by plane (the safest way to travel)<sup>1</sup>.

---

<sup>1</sup>[<https://flyfright.com/plane-crash-statistics/>]

But perhaps due to these high safety standards, when an aircraft accident happens it makes headlines all over the world. Anxious passengers fear it's going to happen to them as well and they can't help but ask themselves if there is something they could do to have a safer flight.

One of the most frequent question that gets asked is : is there a part of the plane that is "safer" than other parts? Can the seating location make a difference in an accident?

We are going to try to answer this question using statistics.

## 1.1 Vocabulary

Before starting the analysis, the meaning of some terms must be clarified

- Flight : A flight a trip made by an aircraft that connects two airports. It is identified by a number. (Example : AirFrance 225 is the name of the regular service from New Deli to Paris )
- Accident : An Accident is an occurrence where a person is fatally or seriously injured, the aircraft sustains significant structural damage, or the aircraft goes missing.
- Incident : An incident is a dangerous situation where no one is seriously hurt and the plane isn't badly damaged.
- Crash : A crash is a type of accident where an aircraft strikes the ground, water, or an obstacle with enough force to cause severe damage or total destruction.
- Accident flight : when a serious accident happens on a flight, usually the companies save the flight number to refer to that flight, and change the flight number for new flights. (Example : Air France 447 was the name of the connection from Rio to Paris, after the 2009 accident the flight became Air France 445)
- Serious injury : Any injury requiring more than 48 hours of hospitalization or involving broken bones (other than fingers/toes)
- Survivability : an accident is defined as survivable if the forces are within human limits and if the structure remained substantially intact
- CREEP factors : factors that influence the survivability of an accident. They are
  - Container : amount of cabin deformation
  - Restraints : analysis of the seating structure
  - Environment : presence of lethal contact points
  - Energy Absorption : G-load mitigation by fuselage or landing gear
  - Post Crash : Other factors (We considered : airport proximity, presence of fire, daytime, phase of flight)

## 1.2 Literature

While we found no scientific article about this issue, there are many news articles that claim to have conducted "statistical analysis" on aircraft accident data to find which area is the safest in case of a crash. We now list some of their findings :

- *Time Article*<sup>2</sup> : "Statistics show that the middle seats in the rear of an aircraft historically have the highest survival rates. [...] The analysis found that the seats in the back third of the aircraft had a 32% fatality rate, compared with 39% in the middle third and 38% in the front third.". This article was written in June 2015.

---

<sup>2</sup>[<https://time.com/3934663/safest-seat-airplane/>]

- *Reuters Article*<sup>3</sup> : “Data reveals civil aviation’s most astonishing, exceptional survivals—and shows no seat is reliably safe.”
- *Aeroclass Article*<sup>4</sup> : “An analysis of several accidents reveals that: The seats in the back third of the aircraft have a fatality rate of 32%. The seats in the middle third of the plane have a fatality rate of 39%. The seats in the front third have a fatality rate of 38%. [...] The crash data indicates that the front third and middle third of the plane have higher fatality rates than the back third of the plane.” This article was written in November 2017.
- *Wired Article*<sup>5</sup> : “While no part of the plane may generally be the safest [...] Each airline emergency plays out differently, affecting different seats more than others each time.”
- *Allianz Report - Popular Mechanics*<sup>6</sup> : Popular Mechanics also examined 20 accidents and calculated the survival rate in each of four sections of the aircraft. Its results found that in 11 of the 20 crashes, passengers in the rear of the aircraft had a better chance of survival. In seven of those 11 crashes favoring the rear of the aircraft they found the rear section was the only section with survivors. In five accidents the first class and business class section fared the best with a 49% survivability rate. In three out of the 20 crashes no one location had an advantage.
- *Allianz Report - University of Greenwich*<sup>7</sup> : the University of Greenwich studied 105 airline accidents worldwide, and this study concluded that the safest seat on an aircraft is in the one on the aisle nearest the exit, in the front of the aircraft. This seat has a survivability rate of 65% whereas a passenger seated in the rear section only has a 53% survivability rate. Additionally, any seat in the aisle near an exit offers a greater chance of survivability. When seated more than six rows from an exit “the chances of perishing far outweigh those of surviving”

### 1.3 Perspective

We are studying if the seating in an airplane has an effect on the survivability of an aircraft accident. To study this, we need to look at all aircraft accidents, then rule out the accidents where every passenger survived, and every accident in which every passenger died. This is an extremely narrow data set of accidents. Of these accidents, we gathered the seatings arrangements and survivor seating maps for 47 crashes.

If 47 crashes look like a lot, we should consider the greater perspective of air travel safety in general.

We can get a clear picture of just how rare an accident is by looking into the US General aviation data between 2015 and 2020<sup>8</sup> :

- $\frac{1}{260256}$  : chance of boarding any flight and it being an accident flight
- $\frac{1}{6,864,250}$  : chance of being on a plane involved in an accident that results in at least 1 fatality (possible case study of this study)
- $\frac{1}{816,545,929}$  chance of you specifically, dying in a plane crash

---

<sup>3</sup>[<https://www.reuters.com/graphics/AVIATION-SAFETY/lgpdaagabvo/>]

<sup>4</sup>[<https://www.aeroclass.org/the-safest-place-to-sit-on-a-plane/>]

<sup>5</sup>[<https://www.wired.com/story/whats-the-safest-seat-on-an-airplane/>]

<sup>6</sup>[[https://www.allianz.com/content/dam/onemarketing/azcom/Allianz\\_com/migration/media/press/document/other/AGCS-Global-Aviation-Safety-Study-2014.pdf](https://www.allianz.com/content/dam/onemarketing/azcom/Allianz_com/migration/media/press/document/other/AGCS-Global-Aviation-Safety-Study-2014.pdf)]

<sup>7</sup>[[https://www.allianz.com/content/dam/onemarketing/azcom/Allianz\\_com/migration/media/press/document/other/AGCS-Global-Aviation-Safety-Study-2014.pdf](https://www.allianz.com/content/dam/onemarketing/azcom/Allianz_com/migration/media/press/document/other/AGCS-Global-Aviation-Safety-Study-2014.pdf)]

<sup>8</sup>[<https://flyfright.com/plane-crash-statistics/#tve-jump-18c020d9166>]

## 1.4 Data Gathering Process

For this experiment, we found that there weren't already available datasets to get the accident seating maps.

//talk about the fact that in all major accidents either all die or all survive -> wikipedia list

//not all incidents feature a map. Not all incidents where people both survived and died were crashes (11A)  
// argue that we have a significative dataset

We decided that we would gather the information from the final reports that are written as outcomes of all investigations into aircraft accidents. We also gathered data from Wikipedia after checking that the source was indeed the final report of the investigation into the accident.

//What are FR and that they are official ; redacted by apposite agencies - only official document

//manually counted all seats, so the data is ours. PILOTS AND FA not included as they are not passengers

We decided that to be able to get some insight on the safest part of the aircraft, we should not use single seats but rather chunks of the aircraft. Not knowing if dividing into thirds or halves, we inserted both these information in our dataset. //not all aircraft have the same seating arrangements, little data -> 3 and 2

## 2 Data Description

Let's take a look at the structure and meaning of the data we gathered:

```
accident_data <- read.csv("AllCREEP_cleaned_eng.csv")
str(accident_data)
```

```
## 'data.frame': 47 obs. of 26 variables:
## $ Airline      : chr "singapore airlines" "british airtours" "british midland" "china airlines"
## $ FlightNum    : int 6 28 92 120 123 129 140 148 191 204 ...
## $ X1.third.minor : int 17 36 0 0 0 4 0 0 0 3 ...
## $ X1.third.major : int 2 0 11 0 0 0 0 0 0 0 ...
## $ X1.third.dead  : int 15 0 22 0 136 14 18 16 55 33 ...
## $ X2.third.minor : int 1 30 4 5 0 5 7 1 0 25 ...
## $ X2.third.major : int 15 0 30 0 0 0 0 0 8 0 ...
## $ X2.third.dead  : int 64 16 13 8 214 60 139 34 51 1 ...
## $ X3.third.minor : int 26 10 0 5 36 24 0 8 10 29 ...
## $ X3.third.major : int 17 0 27 0 0 0 0 0 7 0 ...
## $ X3.third.dead  : int 0 36 11 21 109 43 91 30 16 0 ...
## $ X1.half.minor  : int 17 56 32 3 0 7 7 1 0 7 ...
## $ X1.half.major  : int 3 0 0 0 0 0 0 1 0 ...
## $ X1.half.dead   : int 41 8 34 6 226 23 95 36 79 35 ...
## $ X2.half.minor  : int 26 20 39 7 4 26 0 7 10 46 ...
## $ X2.half.major  : int 31 0 0 0 0 0 0 14 0 ...
## $ X2.half.dead   : int 34 44 13 23 225 90 145 46 48 0 ...
## $ DataOrigin     : chr "w" "w" "w" "w" ...
## $ PhaseOfFlight  : chr "takeoff" "takeoff" "landing" "landing" ...
## $ Time          : chr "night" "day" "night" "day" ...
## $ Place          : chr "airport" "outside" "outside" "airport" ...
## $ HasFire        : chr "fire" "fire" "fire" "fire" ...
## $ CrushedFuselage: int 1 1 1 1 1 1 1 1 1 ...
## $ RestraintIntact: int 0 0 0 1 0 0 0 0 0 ...
## $ Environment    : chr "dangerous" "dangerous" "dangerous" "clear" ...
## $ EnergyAbsorption: chr "nogear" "nogear" "gear" "nogear" ...
```

```

#
#> str(accident_data)
#'data.frame': 47 obs. of 26 variables:
# $ Airline      : chr "singapore airlines" "british airtours" "british midland" "china airlines"
# $ FlightNum   : int 6 28 92 120 123 129 140 148 191 204 ...
# $ X1.third.minor : int 17 36 0 0 0 4 0 0 0 3 ...
# $ X1.third.major : int 2 0 11 0 0 0 0 0 0 0 ...
# $ X1.third.dead : int 15 0 22 0 136 14 18 16 55 33 ...
# $ X2.third.minor : int 1 30 4 5 0 5 7 1 0 25 ...
# $ X2.third.major : int 15 0 30 0 0 0 0 0 0 8 0 ...
# $ X2.third.dead : int 64 16 13 8 214 60 139 34 51 1 ...
# $ X3.third.minor : int 26 10 0 5 36 24 0 8 10 29 ...
# $ X3.third.major : int 17 0 27 0 0 0 0 0 0 7 0 ...
# $ X3.third.dead : int 0 36 11 21 109 43 91 30 16 0 ...
# $ X1.half.minor : int 17 56 32 3 0 7 7 1 0 7 ...
# $ X1.half.major : int 3 0 0 0 0 0 0 0 0 1 0 ...
# $ X1.half.dead : int 41 8 34 6 226 23 95 36 79 35 ...
# $ X2.half.minor : int 26 20 39 7 4 26 0 7 10 46 ...
# $ X2.half.major : int 31 0 0 0 0 0 0 0 14 0 ...
# $ X2.half.dead : int 34 44 13 23 225 90 145 46 48 0 ...
# $ DataOrigin    : chr "w" "w" "w" "w" ...
# $ PhaseOffFlight: chr "takeoff" "takeoff" "landing" "landing" ...
# $ Time          : chr "night" "day" "night" "day" ...
# $ Place         : chr "airport" "outside" "outside" "airport" ...
# $ HasFire       : chr "fire" "fire" "fire" "fire" ...
# $ CrushedFuselage: int 1 1 1 1 1 1 1 1 1 ...
# $ RestraintIntact: int 0 0 0 1 0 0 0 0 0 0 ...
# $ Environment   : chr "dangerous" "dangerous" "dangerous" "clear" ...
# $ EnergyAbsorption: chr "nogear" "nogear" "gear" "nogear" ...
#>
#

```

## 2.1 Data Points and Variables

The data consists of 47 observations (aircraft accidents) and 26 variables. They are the focus of our analysis.

We want to remark that the data consists only of aircraft accidents where there was at least 1 survivor and at least 1 fatality, because the final goal was to compare mortality rates.

// ...

The variables consist of:

- **Airline & FlightNum** : identifier of the aircraft accident. Useful to retrieve more information about the accident if needed.
- **X variables** : these variables represent the number and the type of injury (minor-none / serious / fatal) in each section of the airplane (Each third and each half). //... The format is the following : X{section}.{part}.{type of injury}, where section is 1,2,3 for the thirds and 1,2 for the halves, part is “third” or “half”, and type of injury is “minor”, “major” or “dead”.
- **DataOrigin** : source of the data, either Wikipedia (W) or directly from the final report (FR) or other sources (empty). // Remove Column in the data
- **PhaseOffFlight** : phase of flight when the accident happened, The possible values are “takeoff” or “landing”, landing includes emergency landings.

- **Time** : time of the day when the accident happened, possible values are “day” or “night”. This is intended to signal if natural light was present or not.
- **Place** : location where the accident happened. Possible values are “airport” or “outside”. “Airport” means that the accident happened within airport boundaries (where the airport fire rescue service arrive in less than 3 minutes), “outside” means that the accident happened outside an airport.
- **HasFire** : indicates if there was an uncontained fire after the crash that spread to the cabin. Possible values are “fire” or “no-fire”.
- **CrushedFuselage** : indicates if the fuselage was crushed, so deformed with significant loss of volume, during the accident. Possible values are 1 (yes) or 0 (no).
- **RestraintIntact** : if the seating structure remained intact and the seat belt system functioned correctly, the Restraint system is considered to have functioned correctly. Possible values are 1 (yes) or 0 (no).
- **Environment** : indicates the state of the cabin after the crash. Possible values are “clear” or “dangerous”. It is considered dangerous if it was determined that parts of the cabin like overhead bins failed and debris was scattered in the cabin.
- **EnergyAbsorption** : indicates if the landing gear was extended and if it absorbed a significant amount of impact forces before collapsing. If not, the airplane either hit the ground with the gear retracted or hit the ground nose-first or tail-first, causing the impact forces to be absorbed by the passenger cabin. Possible values are “gear” or “nogear”

```
// fix nogear -> no-gear
```

## 2.2 Mortality Rates

To start the analysis, let's prepare the data by adding some useful variables, and transform some variables into a more manageable format.

```
//...
```

```
# add a column of # of passengers for each airplane section
accident_data <- within(accident_data, {
  X2.half.total <- X2.half.minor + X2.half.major + X2.half.dead
  X1.half.total <- X1.half.minor + X1.half.major + X1.half.dead
  X3.third.total <- X3.third.minor + X3.third.major + X3.third.dead
  X2.third.total <- X2.third.minor + X2.third.major + X2.third.dead
  X1.third.total <- X1.third.minor + X1.third.major + X1.third.dead
})

# now we transform the variables that are categorical into factors
factor_cols <- c("PhaseOfFlight", "Time", "Place", "HasFire", "Environment",
                  "EnergyAbsorption", "CrushedFuselage", "RestraintIntact")

accident_data[factor_cols] <- lapply(accident_data[factor_cols], as.factor)
```

Now we can add a new variable that will be the center of our analysis : the mortality rate. It is measured for each section of the airplane. The mortality rate is defined as the number of deaths divided by the total number of passengers in the section, for each section. Note that aircraft don't always fly on a full load of passengers, so there might be sections with very few passengers or possibly empty. That's why we are considering rates per total passengers instead of per number of seats.

```

accident_data <- within(accident_data, {
  X1.third.mortality.rate <- X1.third.dead / X1.third.total
  X2.third.mortality.rate <- X2.third.dead / X2.third.total
  X3.third.mortality.rate <- X3.third.dead / X3.third.total
  X1.half.mortality.rate <- X1.half.dead / X1.half.total
  X2.half.mortality.rate <- X2.half.dead / X2.half.total
})

```

What we see in these graphs is very interesting and significant.

In the first and second thirds we can see that in most cases, either most die or most survive. In the third third the distribution is much more even, except for the relatively high 0% mortality rate frequency. This is already a clue that the diversity is great and therefore

## 3 Data Analysis

### 3.1 Preliminary analysis

#### 3.1.1 Plotting

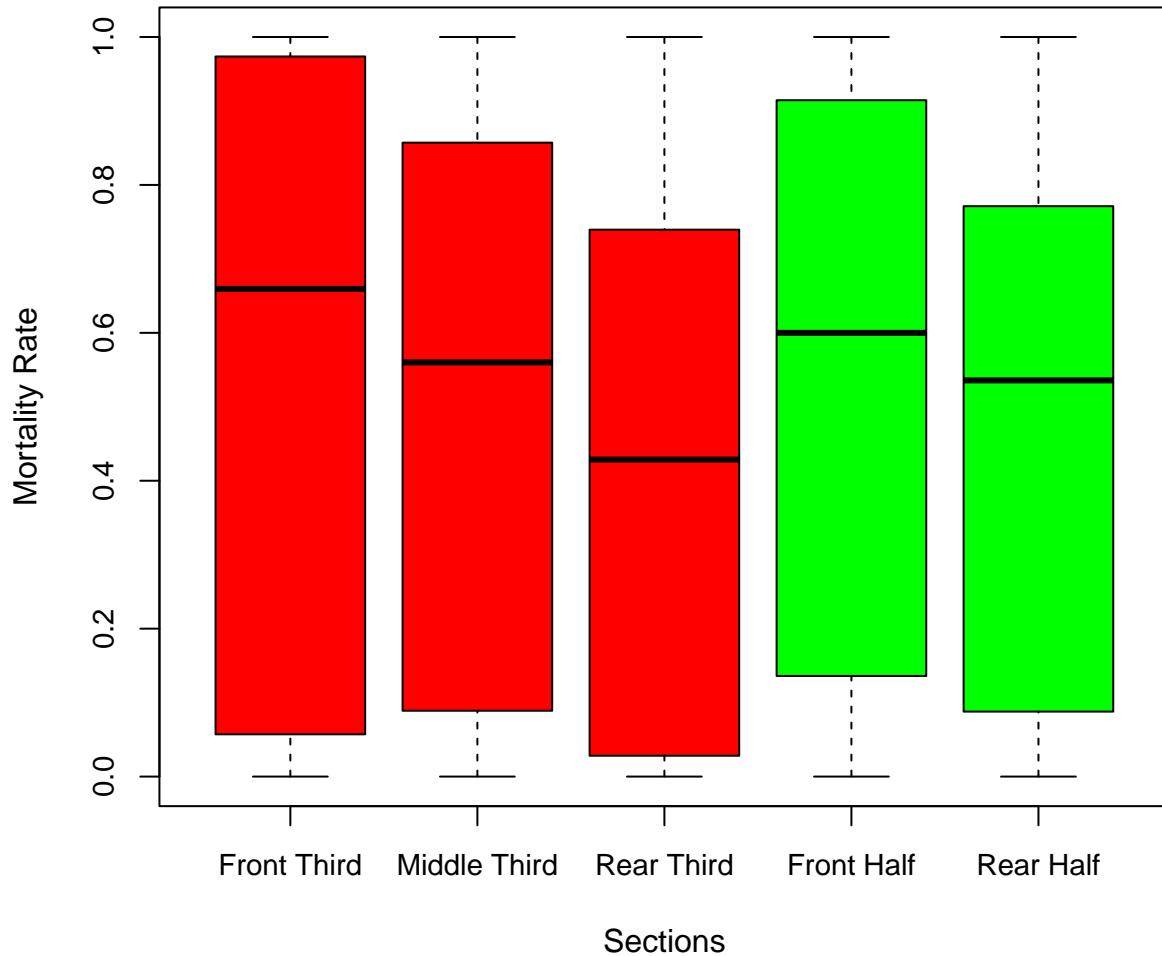
The first step we can do is plotting some feature of the data to have a visual understanding of it. We do this to get some useful insights and also check for possible outliers or anomalies.

```

with(accident_data, {
  boxplot(X1.third.mortality.rate,
          X2.third.mortality.rate,
          X3.third.mortality.rate,
          X1.half.mortality.rate,
          X2.half.mortality.rate,
          names = c("Front Third", "Middle Third", "Rear Third", "Front Half", "Rear Half"),
          main = "Mortality Rates by Section",
          ylab = "Mortality Rate",
          xlab = "Sections",
          col = c("red", "red", "red", "green", "green"),
          cex.axis = 0.9)
})

```

## Mortality Rates by Section



Interestingly, there seems to be a general trend of lowering mortality rates the more rear the section is. This is seen in both the third and half divisions of the airplane.

### 3.1.2 Distribution

Now we can try to test the kind of population distribution of the mortality rates, to see if they follow a normal distribution or not. This will help us choose the right statistical tests for our analysis.

We will use the Shapiro-Wilk test for normality.

```
# shapiro test for normality, p-value extracted and labeled
shapiro_results <- data.frame(
  Section = c("Front Third", "Middle Third", "Rear Third", "Front Half", "Rear Half"),
  P_Value = c(
    shapiro.test(accident_data$X1.third.mortality.rate)$p.value,
    shapiro.test(accident_data$X2.third.mortality.rate)$p.value,
    shapiro.test(accident_data$X3.third.mortality.rate)$p.value,
```

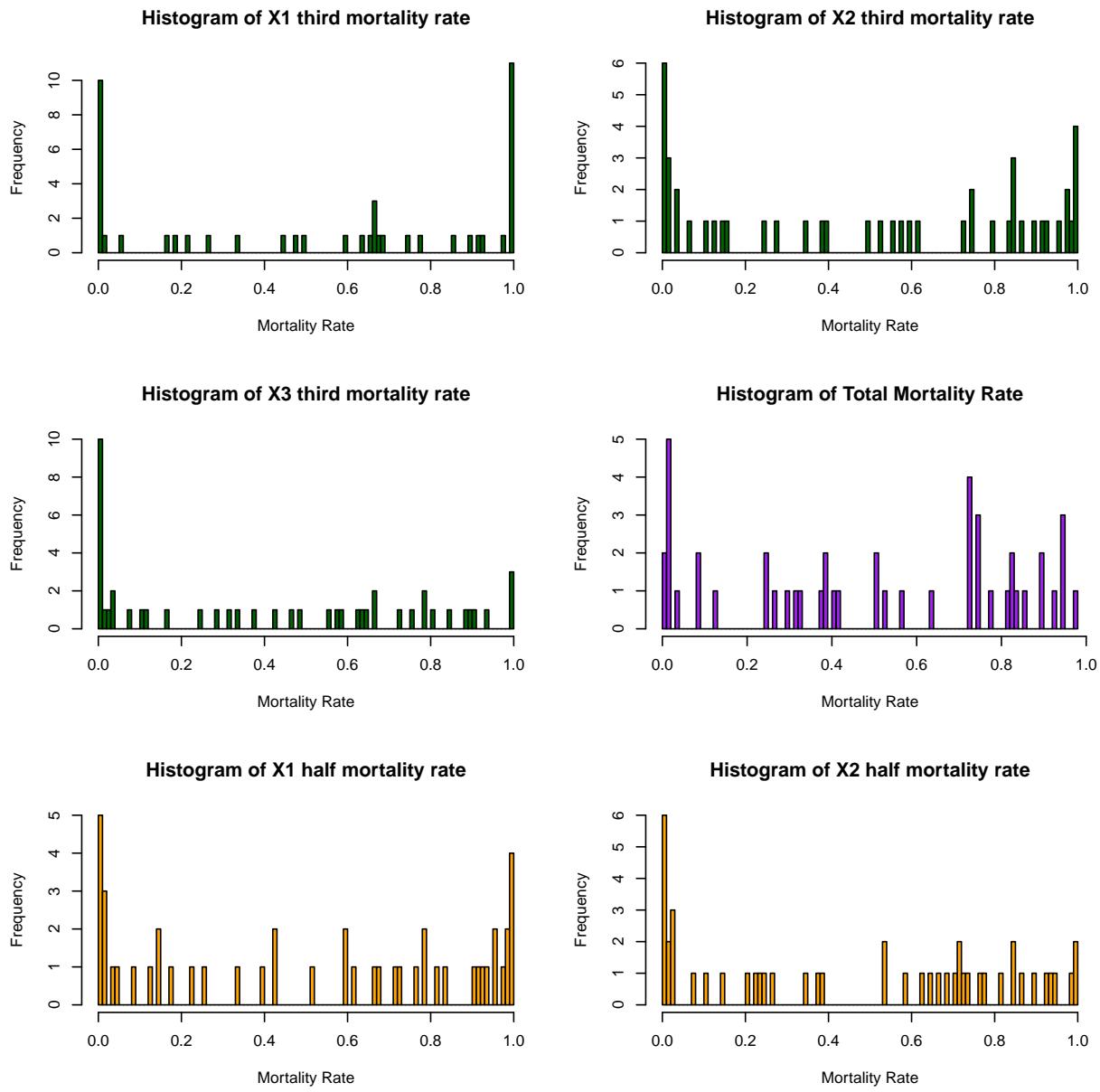
```

    shapiro.test(accident_data$X1.half.mortality.rate)$p.value,
    shapiro.test(accident_data$X2.half.mortality.rate)$p.value
)
)
shapiro_results

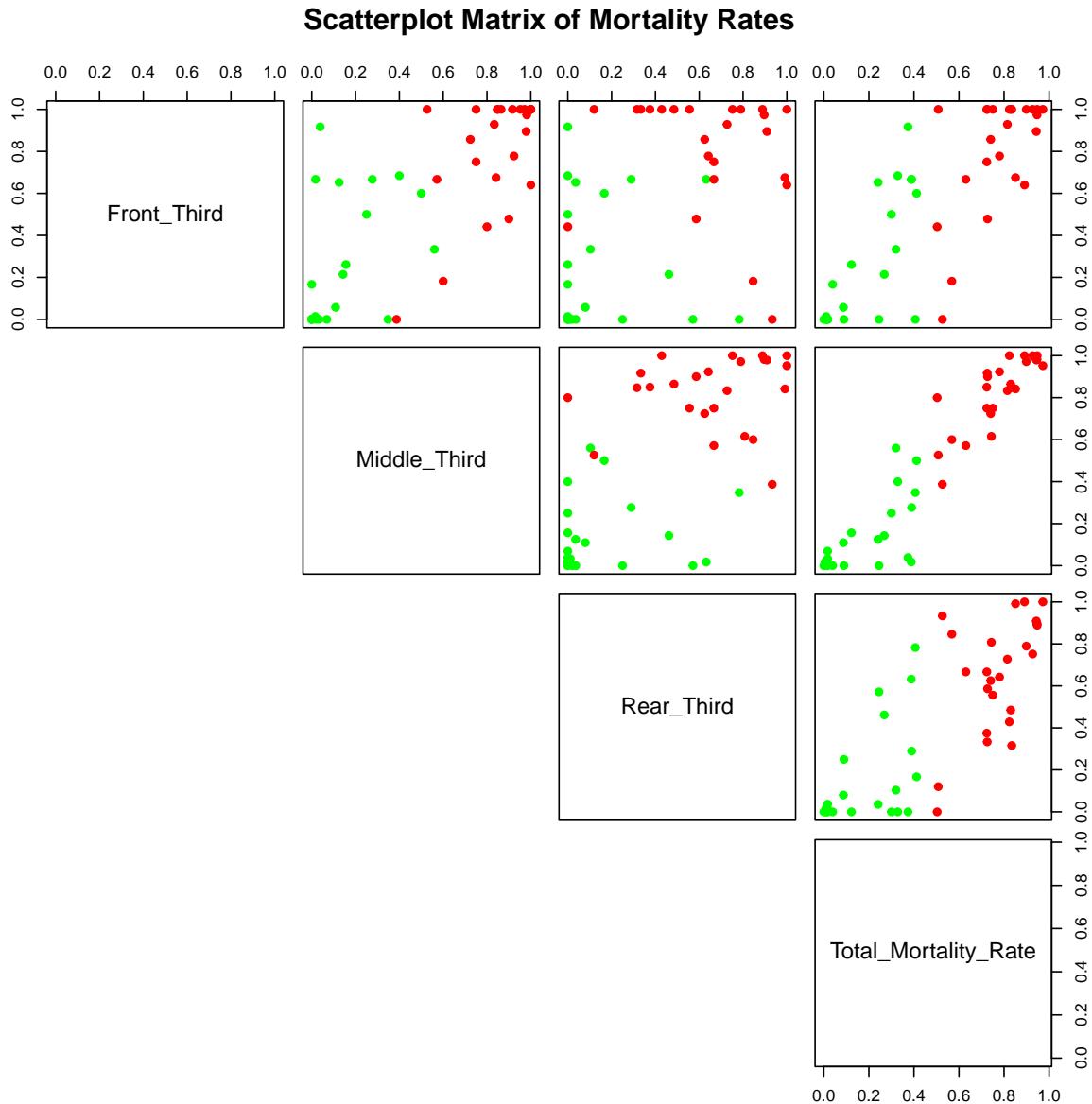
##           Section      P_Value
## 1   Front Third 1.873378e-05
## 2 Middle Third 5.727713e-05
## 3   Rear Third 1.964064e-04
## 4   Front Half 8.609941e-05
## 5   Rear Half 2.491057e-04

```

All of the p-values are far below the standard threshold of 0.05, indicating that we can reject the null hypothesis of normality for all sections. This means that the mortality rates do not follow a normal distribution, let's try to plot a histogram for some of the sections to visualize an approximation of the distribution.



```
## corrplot 0.95 loaded
```



It seems that there is no clear distribution pattern. What we can see is that the data is more concentrated at both ends of the mortality rate values, especially for the front third section.

### 3.2 Statistical Analysis of differences in mortality rates between sections

Now that we have a better understanding of the data, we can proceed with the statistical analysis to test if there are significant differences in mortality rates between the different sections of the airplane. Also, we will try to see if there is some correlation between the mortality rates and the other variables in the dataset.

Since the distribution of the mortality rates is not normal, we will not use tests that assume normality, like t-tests or ANOVA. We will instead use the Kruskal-Wallis test, which is a non-parametric test to compare the distribution of two or more groups. The null hypothesis of the Kruskal-Wallis test is that all the populations have the same distribution.

```

# Kruskal-Wallis test for differences in mortality rates between sections (only thirds)
mortality_data_third <- data.frame(
  SectionThird = rep(c("Front Third", "Middle Third", "Rear Third"), each = nrow(accident_data)),
  MortalityRateThird = c(accident_data$X1.third.mortality.rate,
                        accident_data$X2.third.mortality.rate,
                        accident_data$X3.third.mortality.rate)
)

kruskal_result_third <- kruskal.test(MortalityRateThird ~ SectionThird, data = mortality_data_third)
kruskal_result_third

##
##  Kruskal-Wallis rank sum test
##
## data: MortalityRateThird by SectionThird
## Kruskal-Wallis chi-squared = 2.9755, df = 2, p-value = 0.2259

# Kruskal-Wallis test for differences in mortality rates between sections (only halves)
mortality_data_half <- data.frame(
  SectionHalf = rep(c("Front Half", "Rear Half"), each = nrow(accident_data)),
  MortalityRateHalf = c(accident_data$X1.half.mortality.rate,
                        accident_data$X2.half.mortality.rate)
)

kruskal_result_half <- kruskal.test(MortalityRateHalf ~ SectionHalf, data = mortality_data_half)
kruskal_result_half

##
##  Kruskal-Wallis rank sum test
##
## data: MortalityRateHalf by SectionHalf
## Kruskal-Wallis chi-squared = 0.65567, df = 1, p-value = 0.4181

```

[Interpretation of results to be added here, OR later in the report]

### 3.3 Modeling the mortality rates

Now we will try to fit a model to possibly discover the effect of the other variables on the mortality rates of the different sections. The focus is not specifically on prediction, but rather on trying to find a significant difference between the sections. Now the questions are:

- Are there variables that significantly affect the mortality rate?
- If so, do they affect differently the various sections of the airplane?

We will try to fit a Generalized Linear Model (GLM) with binomial family, since the mortality rate is a proportion (number of deaths / total number of passengers). Unfortunately, because of the nature of the data, the assumptions of the binomial distribution are not met, since the passenger's deaths are not independent events.

To account for this, we will use the quasibinomial family, which is a more flexible version of the binomial family that allows for overdispersion and corrects the standard errors and p-values accordingly.

```

### test glm with binomial family to see the effect of other variables on mortality rate for each section

## also print the varius confidence intervals for the coefficients, and trasform them into odds ratios
glm_third_1 <- glm(cbind(accident_data$X1.third.dead, accident_data$X1.third.total - accident_data$X1.third.total,
                           #accident_data$PhaseOfFlight +
                           #accident_data$Time +
                           #accident_data$Place +
                           accident_data$HasFire +
                           #accident_data$Environment +
                           accident_data$EnergyAbsorption +
                           accident_data$CrushedFuselage +
                           accident_data$RestraintIntact,
                           family = quasibinomial(link = "logit"))

glm_third_2 <- glm(cbind(accident_data$X2.third.dead, accident_data$X2.third.total - accident_data$X2.third.total,
                           #accident_data$PhaseOfFlight +
                           #accident_data$Time +
                           #accident_data$Place +
                           accident_data$HasFire +
                           #accident_data$Environment +
                           accident_data$EnergyAbsorption +
                           accident_data$CrushedFuselage +
                           accident_data$RestraintIntact,
                           family = quasibinomial(link = "logit"))

glm_third_3 <- glm(cbind(accident_data$X3.third.dead, accident_data$X3.third.total - accident_data$X3.third.total,
                           #accident_data$PhaseOfFlight +
                           #accident_data$Time +
                           #accident_data$Place +
                           accident_data$HasFire +
                           #accident_data$Environment +
                           accident_data$EnergyAbsorption +
                           accident_data$CrushedFuselage +
                           accident_data$RestraintIntact,
                           family = quasibinomial(link = "logit"))

# now try to model the total mortality rate

glm_total <- glm(cbind(accident_data$X1.third.dead + accident_data$X2.third.dead + accident_data$X3.third.dead,
                         accident_data$X1.third.total + accident_data$X2.third.total + accident_data$X3.third.total,
                         (accident_data$X1.third.dead + accident_data$X2.third.dead + accident_data$X3.third.dead),
                         #accident_data$PhaseOfFlight +
                         #accident_data$Time +
                         #accident_data$Place +
                         accident_data$HasFire +
                         #accident_data$Environment +
                         accident_data$EnergyAbsorption +
                         accident_data$CrushedFuselage +
                         accident_data$RestraintIntact,
                         family = quasibinomial(link = "logit"))

```

```

summary(glm_third_1)

## Warning in summary.glm(glm_third_1): observations with zero weight not used for
## calculating dispersion

##
## Call:
## glm(formula = cbind(accident_data$X1.third.dead, accident_data$X1.third.total -
##   accident_data$X1.third.dead) ~ accident_data$HasFire + accident_data$EnergyAbsorption +
##   accident_data$CrushedFuselage + accident_data$RestraintIntact,
##   family = quasibinomial(link = "logit"))
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -1.9872    2.7344  -0.727  0.4715
## accident_data$HasFireno-fire   -1.2185    0.7003  -1.740  0.0894 .
## accident_data$EnergyAbsorptionnogear  0.9057    0.5688   1.592  0.1190
## accident_data$CrushedFuselage1     2.4226    2.6803   0.904  0.3714
## accident_data$RestraintIntact1    -2.2228    1.1712  -1.898  0.0648 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 18.04978)
##
## Null deviance: 1354.33 on 45 degrees of freedom
## Residual deviance: 849.89 on 41 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5

```

```
summary(glm_third_2)
```

```

##
## Call:
## glm(formula = cbind(accident_data$X2.third.dead, accident_data$X2.third.total -
##   accident_data$X2.third.dead) ~ accident_data$HasFire + accident_data$EnergyAbsorption +
##   accident_data$CrushedFuselage + accident_data$RestraintIntact,
##   family = quasibinomial(link = "logit"))
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -1.8743    2.2596  -0.829  0.41151
## accident_data$HasFireno-fire   -1.0375    0.7351  -1.411  0.16553
## accident_data$EnergyAbsorptionnogear  1.6822    0.6219   2.705  0.00982 **
## accident_data$CrushedFuselage1     1.9259    2.2170   0.869  0.38995
## accident_data$RestraintIntact1    -2.2223    1.2069  -1.841  0.07263 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 32.73325)
##
## Null deviance: 2358.0 on 46 degrees of freedom

```

```

## Residual deviance: 1588.4 on 42 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5

summary(glm_third_3)

##
## Call:
## glm(formula = cbind(accident_data$X3.third.dead, accident_data$X3.third.total -
##   accident_data$X3.third.dead) ~ accident_data$HasFire + accident_data$EnergyAbsorption +
##   accident_data$CrushedFuselage + accident_data$RestraintIntact,
##   family = quasibinomial(link = "logit"))
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -1.9307    2.2409  -0.862  0.394
## accident_data$HasFireno-fire   -0.7861    0.6864  -1.145  0.259
## accident_data$EnergyAbsorptionnogear  0.6255    0.5193   1.205  0.235
## accident_data$CrushedFuselage1      1.8392    2.1924   0.839  0.406
## accident_data$RestraintIntact1     -0.8980    0.9222  -0.974  0.336
##
## (Dispersion parameter for quasibinomial family taken to be 26.04673)
##
## Null deviance: 1573.7 on 46 degrees of freedom
## Residual deviance: 1350.8 on 42 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5

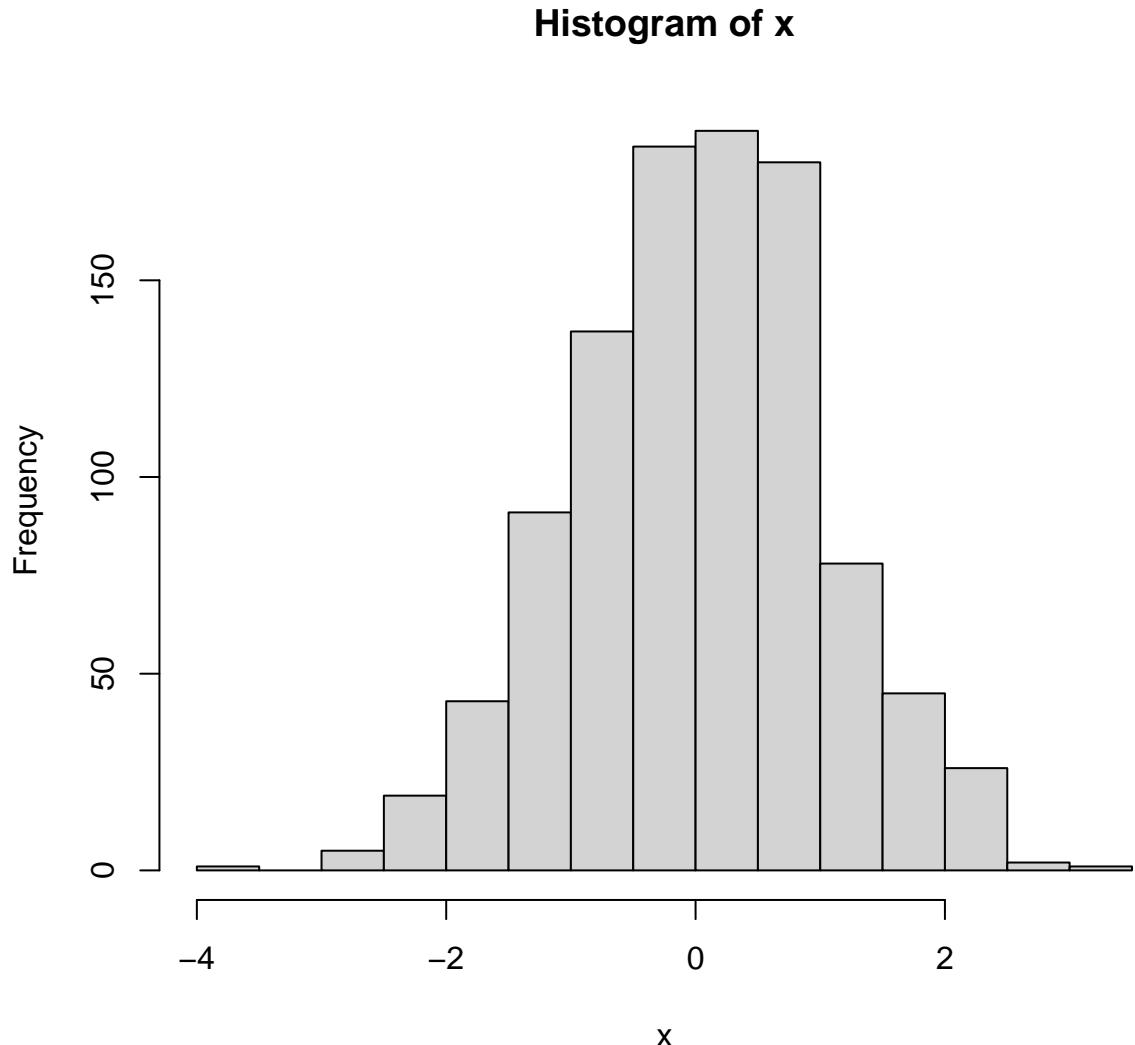
summary(glm_total)

##
## Call:
## glm(formula = cbind(accident_data$X1.third.dead + accident_data$X2.third.dead +
##   accident_data$X3.third.dead, accident_data$X1.third.total +
##   accident_data$X2.third.total + accident_data$X3.third.total -
##   (accident_data$X1.third.dead + accident_data$X2.third.dead +
##     accident_data$X3.third.dead)) ~ accident_data$HasFire +
##   accident_data$EnergyAbsorption + accident_data$CrushedFuselage +
##   accident_data$RestraintIntact, family = quasibinomial(link = "logit"))
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -1.9728    2.0101  -0.981  0.3320
## accident_data$HasFireno-fire   -0.9866    0.6040  -1.634  0.1098
## accident_data$EnergyAbsorptionnogear  1.1101    0.4819   2.304  0.0263 *
## accident_data$CrushedFuselage1      2.0553    1.9685   1.044  0.3024
## accident_data$RestraintIntact1     -1.6568    0.9158  -1.809  0.0776 .
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 57.72175)

```

```
##  
## Null deviance: 4184.2 on 46 degrees of freedom  
## Residual deviance: 2795.4 on 42 degrees of freedom  
## AIC: NA  
##  
## Number of Fisher Scoring iterations: 5
```

```
x <- rnorm(1000)  
hist(x)
```



**4 Analysis**

**5 Results**

**6 Conclusions**

**7 Links**