

Impact of Paperclip Positions on Final Distance

Hannah Nguyen

2025-06-11

Introduction

Weight distribution plays an important role in airplane design, affecting the stability of the flight and its fuel efficiency (Boyd et al., 2016, Nogoud et al., 2023, Liu et al., 2018). For example, increasing the weight of an aircraft will result in longer takeoff and landing distances (FAA 2007, FAA 2008, as cited in Boyd et al., 2016). Additionally, exceeding the center of gravity limits (CG) will result in issues related to plane stability and makes recovery from an aerodynamic stall impossible due to the loss of elevator control authority (FAA 2007, as cited in Boyd et al., 2016). This shows that strategic distribution of the weight is crucial in maintaining airplane stability for safety. From a study, it was found that high fraction of fatal accidents (50-80% depending on the phase of flight) was related to over-loading or exceeding the CG limits of the aircraft (Nogoud et al., 2023). Determining the optimal weight distribution will help in preventing these fatal accidents.

In addition to safety, the weight distribution of the airplane in regards to its aerodynamics is important in maximizing fuel efficiency, since there are many concerns related to climate change and global warming with the overuse of fuel in the environment (Nogoud et al., 2023). Modifying the plane's design can help reduce over consumption of fuel and reduce the risks of harming the environment. From a different study, they found that relocating the CG from 14% of the Mean Aerodynamic Chord (MAC) to 39.5% of the MAC will save approximately 5164.5L of fuel over a range of 13,500 km (Nogoud et al., 2023). It is important to create a strategic design of the distribution of weight to ensure safety and minimize the harm done to the environment.

The position of the weight can significantly change the aerodynamics of the plane. For example, from one study, they found that a more aft CG position produces less drag, which results in less fuel consumption (Liu et al., 2018). The location of choice can influence how much drag the plane produces and as a result, how far of a distance the plane can go.

These studies highlight the importance of optimizing weight distribution in aircrafts. This study aims to test if, on a smaller scale, the location of the paperclips significantly impacts the final distance of the paper airplane. It was hypothesized that at least one combination of the paperclips on the paper airplane will significantly impact the final distance. The experiment is a factorial design with three factors for the paperclip position to test this hypothesis.

Methods

Study Design

The experiment consisted of three factors, nose, middle, and rear, with two levels each, yes or no, making this a 2^3 factorial experiment. Each factor described the position of a paperclip on the paper airplane, where the plane could have anywhere from zero to three paperclips placed in their respective positions. Paperclips were placed in the same positions consistently throughout the trials, with a distinct colored paperclip assigned to each position to prevent bias involving inconsistencies across paperclips. The paper

airplane was created using an online tutorial by Livingston (2019) and the same plane was used across all trials to ensure consistency. Trials were randomized to reduce any bias from outside factors, including damage to the plane in later throws and exhaustion from the thrower.

Data was obtained by setting up a measuring tape, a Stanley brand, twelve feet, on the ground, and standing behind the line before throwing the paper airplane forward. In the case that the paper airplane flew over twelve feet or 144 inches, the distance was roughly estimated by eye. The distance was measured from the zero inch mark to the first impact with the ground. In the event that the airplane landed behind the zero inch mark, it was recorded with zero distance. The paper airplanes were thrown by the same person with roughly the same effort each time. In the occasional case of headwinds, the person waited for it to end before proceeding with their throw. This ensures that each toss was performed in the same conditions to prevent bias within the data.

Sample Size

A pilot study was performed for sample size calculations, since the materials and resources were inexpensive. As a starting point, we will be performing eight replicates per treatment group. There are three factors with two levels, so there are $2^3 = 8$ different combinations, bringing the trial number to $8 \times 8 = 64$ total trials. We will be randomizing the order of the throws to reduce bias and variability within the data.

Statistical Analyses

Methods of statistical analysis used included the lm function, since this is a factorial experiment, so using the aov function will give incorrect p-values due to the presence of interaction terms. The assumptions for this model include normality of the residuals, no structure to the data, equal variances, and no extreme outliers. We think that for the normality and structure assumptions, they may be violated, since the data collection was split into two days, so the conditions were not consistent throughout the whole data collection process. Equal variances may have been violated as well, so for day 2, the replicates were performed as closely as possible to day 1's data. There are likely no extreme outliers as while the conditions may not be exactly, we attempted to perform each throw consistently, so no large outliers were found.

Technical Issues

Some technical issues included my initial misunderstanding of the experiment, where we resulted in starting 6 replicates on one day and added 4 more replicates the next day. The conditions for both days were not exactly the same, which can cause inconsistencies and bias within my data. Additionally, the tape measure was not long enough, so plane throws that flew over the 12 feet mark had their final displacement guessed as accurately as possible by eye. A small portion of the throws in the first 6 replicates went over 12 feet, but an even smaller portion in the last 4 replicates went over 12 feet, which could indicate inconsistencies with the throws over all 10 replicates.



Figure 1: Plane, Paperclip positions, and materials



Figure 2: Day 1 (June 1, 2025) Weather and Setup

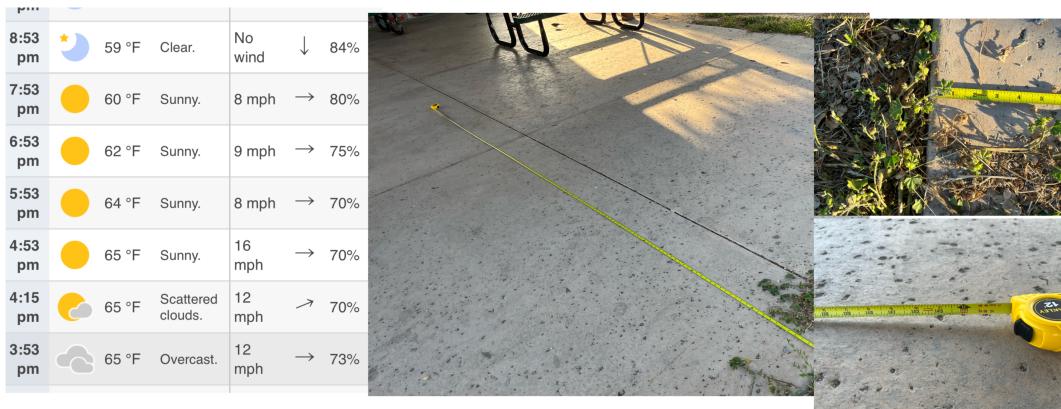


Figure 3: Day 2 (June 3, 2025) Weather and Setup

Results

Sample Size Calculations

We first started by conducting a pilot study with 8 replicates.

```
# Randomizing Order
set.seed(06012025)
combinations <- c("Nose", "Middle", "Rear", "Nose_Middle",
                  "Middle_Rear", "Nose_Rear", "All", "None")
trials <- rep(combinations, each = 8)
ordered_trials <- sample(trials)
```

Once we've obtained the data, we inputted it into R and performed linear regression analysis on it.

```
# Sample Size Calculations

source("power_factorial_23.R")
airplane_data_pilot <- read_csv("airplanes_pilot_study2.csv")

airplane_dataframe_pilot <- data.frame(distance = airplane_data_pilot$Distance,
                                         nose = airplane_data_pilot$Nose,
                                         middle = airplane_data_pilot$Middle,
                                         rear = airplane_data_pilot$Rear)

# Changing columns to factors
airplane_dataframe_pilot$nose <- as.factor(airplane_dataframe_pilot$nose)
airplane_dataframe_pilot$middle <- as.factor(airplane_dataframe_pilot$middle)
airplane_dataframe_pilot$rear <- as.factor(airplane_dataframe_pilot$rear)

# Using a linear model for regression analysis
model0 <- lm(distance ~ nose + middle + rear + nose*middle + middle*rear + nose*middle*rear,
               data = airplane_dataframe_pilot)
kable(summary(model0)$coefficients)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	131.000	8.850144	14.8020184	0.0000000
nose1	-11.125	12.515994	-0.8888627	0.3778792
middle1	-7.125	12.515994	-0.5692716	0.5714483
rear1	12.875	12.515994	1.0286838	0.3080511
nose1:middle1	-8.500	17.700289	-0.4802182	0.6329429
middle1:rear1	-10.250	17.700289	-0.5790866	0.5648529
nose1:rear1	0.250	17.700289	0.0141241	0.9887812
nose1:middle1:rear1	24.625	25.031989	0.9837413	0.3294740

```
# inputting beta_mean and beta_se values from the lm model
```

```
beta_mean <- c(131.00, -11.125, -7.125, 12.875, -8.5, -10.250, 0.250, 24.625)
beta_se <- c(8.85, rep(12.516, 3), rep(17.7, 3), 25.032)
```

The beta_mean and beta_se values from the lm model are represented as “Middle” in the graphical representation below.

```

# Graphical Representation

replicates <- 2:10

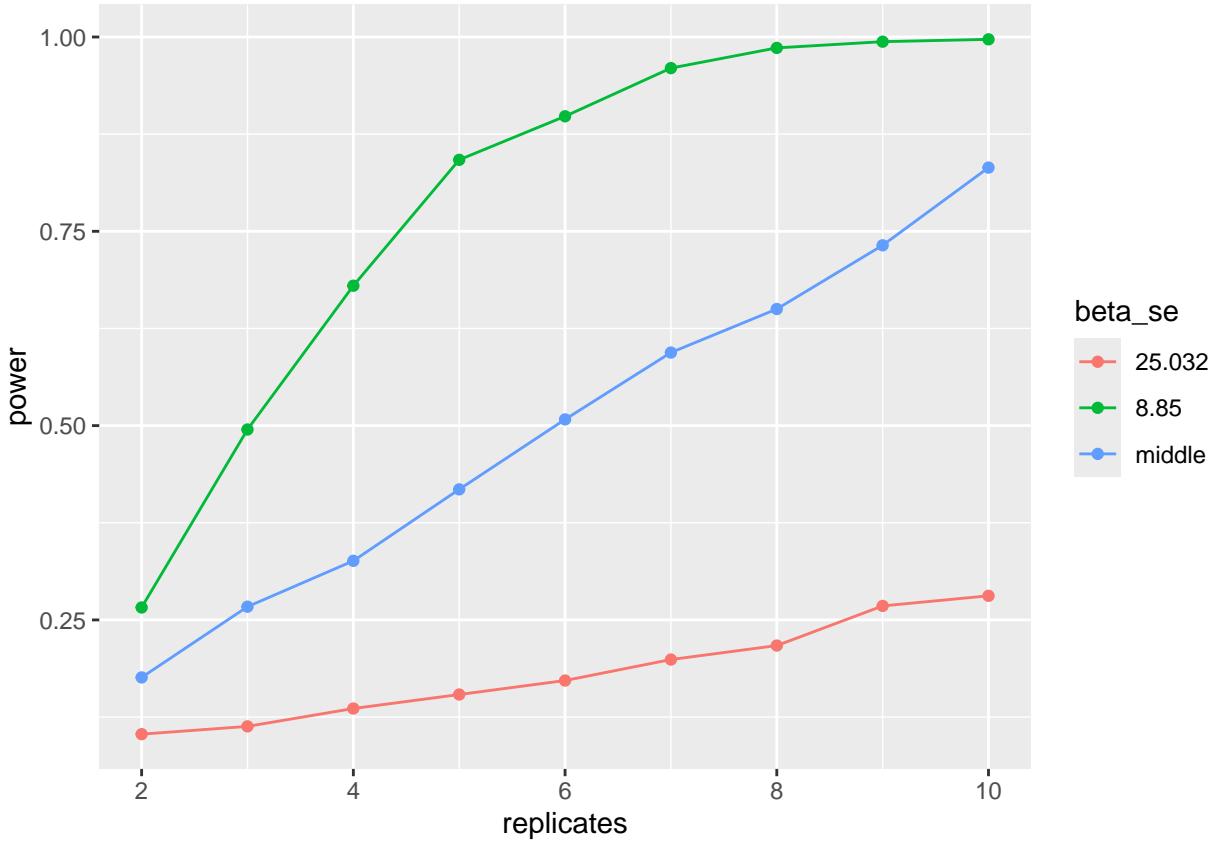
# Lowest SE value
beta_se0 <- c(rep(8.85, 8))
power1 <- NA
for(i in 1:length(replicates)){
  power1[i] <- power_factorial_23(beta_mean, beta_se0, replicates[i]) }

# Highest SE value
beta_se0 <- c(rep(25.032, 8))
power2 <- NA
for(i in 1:length(replicates)){
  power2[i] <- power_factorial_23(beta_mean, beta_se0, replicates[i]) }

# Middle SE value
beta_se0 <- c(8.85, rep(12.516, 3), rep(17.7, 3), 25.032)
power3 <- NA
for(i in 1:length(replicates)){
  power3[i] <- power_factorial_23(beta_mean, beta_se0, replicates[i]) }

# Creating the dataframe of power values
all_power <- data.frame(
  power = c(power1, power2, power3),
  beta_se = c(rep("8.85", length(power1)),
             rep("25.032", length(power2)),
             rep("middle", length(power3))),
  replicates = replicates)
ggplot(data = all_power, mapping = aes(x = replicates, y = power,
                                         group = beta_se, color = beta_se)) +
  geom_point() + geom_line()

```



Based on the graph, when β_{se} is 8.85, power is at least 80% when there are at least 5 replicates. When β_{se} is "middle", power is at least 80% when there are at least 10 replicates. When β_{se} is 25.032, power doesn't reach 80% within 10 replicates. So the ideal β_{se} would be 8.85 or 12.516, since they both reach 80% power within 10 replicates, while $\beta_{se} = 25.032$ doesn't.

Collecting the data

Based on the sample size calculations, we will need 10 replicates for our experiment. Due to unforeseen circumstances, the first 6 replicates were done on day 1 (June 1, 2025), while the last 4 replicates were done on day 2 (June 3, 2025).

```
# Randomizing the order of the trials
trials2 <- rep(combinations, each = 6)
ordered_trials <- sample(trials2)
```

```
# Adding 4 more replicates a day later
trials3 <- rep(combinations, each = 4)
ordered_trials <- sample(trials3)
```

```
# Read in the data
airplane_data <- read_csv("airplanes_data.csv")

airplane_dataframe <- data.frame(distance = airplane_data$Distance,
                                  nose = airplane_data$Nose,
                                  middle = airplane_data$Middle,
```

```

    rear = airplane_data$Rear)

# Changing columns to factors
airplane_dataframe$nose <- as.factor(airplane_dataframe$nose)
airplane_dataframe$middle <- as.factor(airplane_dataframe$middle)
airplane_dataframe$rear <- as.factor(airplane_dataframe$rear)

```

Visualization

```

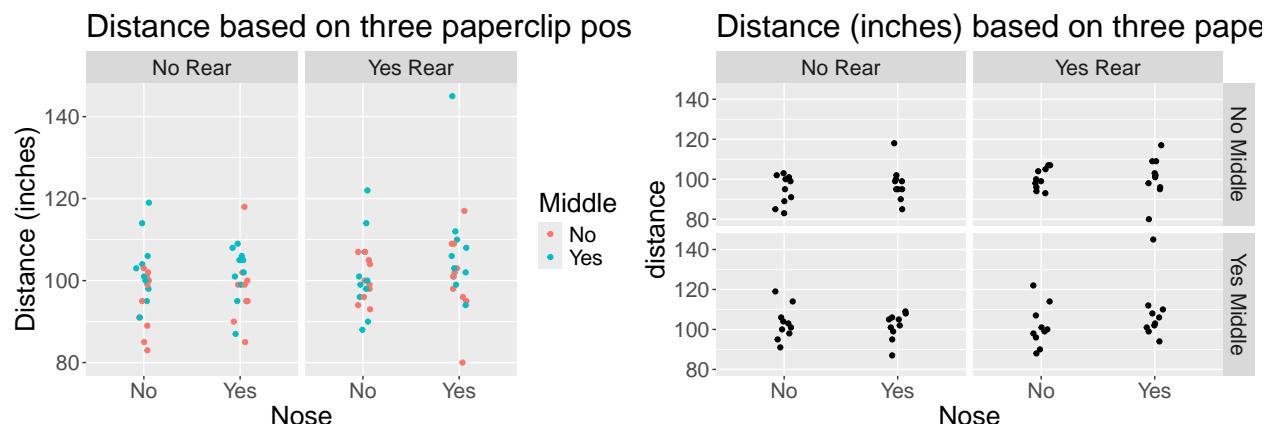
# 2 Visual Representations of the data

library(ggplot2)
theme_update(text = element_text(size = 20))

ggplot(data = airplane_dataframe, mapping = aes(x = nose, y = distance, color = middle)) +
  geom_jitter(width = 0.08, height = 0) +
  facet_grid(cols = vars(rear),
             labeller = labeller(rear = c("0" = "No Rear",
                                           "1" = "Yes Rear")))) +
  ggtitle("Distance based on three paperclip positions") +
  scale_x_discrete(name = "Nose", labels = c("No", "Yes")) +
  scale_color_discrete(name = "Middle", labels = c("No", "Yes")) +
  ylab("Distance (inches)")

ggplot(data = airplane_dataframe, mapping = aes(x = nose, y = distance)) +
  geom_jitter(width = 0.08, height = 0) +
  facet_grid(middle ~ rear,
             labeller = labeller(middle = c("0" = "No Middle",
                                             "1" = "Yes Middle"),
                                 rear = c("0" = "No Rear",
                                         "1" = "Yes Rear")))) +
  scale_x_discrete(name = "Nose", labels = c("No", "Yes")) +
  ggtitle("Distance (inches) based on three paperclip positions")

```



Based on the left plot, when no paperclip is on the rear, the distances tend to be slightly further when there is a paperclip on the middle oppose to not having one. Having a paperclip on the nose did not seem to affect the distribution of the distances. When there is a paperclip on the rear, the distribution seems to be more

spread out. Whether or not there was a paperclip on the middle, the data points were scattered and it did not seem to have an effect within this group. Having a paperclip on the nose did not seem to affect the distribution. Comparing the distributions of “No Rear” and “Yes Rear”, there doesn’t seem to be much differences, so the paperclip being on the rear may not have much effect on the final distance.

Based on the right plot, within “No Rear” and “No Middle”, there is little difference in distribution between “No Nose” and “Yes Nose”. Within “Yes Rear” and “No Middle”, the distributions are similar, but “No Nose” has closer data points, while “Yes Nose” is more spread out. Within “No Rear” and “Yes Middle”, “No Nose” has a slightly higher mean than “Yes Nose” and both distributions still remain similar to each other. Lastly, within “Yes Rear” and “Yes Middle”, “No Nose” has a larger spread than “Yes Nose”, but both have distributions.

Both plots have similar distributions across all three factors, so this might suggest that the factors don’t have much of an effect on the final distance of the paper airplanes.

Statistical Analysis

```
# Performing appropriate statistical analysis
model1 <- lm(distance ~ nose + middle + rear +
               nose*middle + middle*rear + nose*rear +
               nose*middle*rear, data = airplane_dataframe)
kable(summary(model1)$coefficients)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	94.8	2.910422	32.5725973	0.0000000
nose1	3.0	4.115958	0.7288704	0.4684452
middle1	8.3	4.115958	2.0165414	0.0474741
rear1	5.5	4.115958	1.3362624	0.1856714
nose1:middle1	-4.4	5.820844	-0.7559042	0.4521727
middle1:rear1	-7.1	5.820844	-1.2197544	0.2265385
nose1:rear1	-2.3	5.820844	-0.3951317	0.6939139
nose1:middle1:rear1	10.2	8.231916	1.2390797	0.2193401

The Bonferroni correction is $\frac{\alpha=0.05}{\text{numberoftest}=7} = 0.00714$.

nose1:

At $\alpha_{\text{Bonferroni}} = 0.00714$, we fail to reject the null hypothesis, since the p-value, 0.4684, is greater than $\alpha_{\text{Bonferroni}} = 0.00714$. We conclude that the impact of the nose factor is not statistically significant after adjusting for the middle factor, the rear factor, and their interactions.

middle1:

At $\alpha_{\text{Bonferroni}} = 0.00714$, we fail to reject the null hypothesis, since the p-value, 0.0475, is greater than $\alpha_{\text{Bonferroni}} = 0.00714$. We conclude that the impact of the middle factor is not statistically significant after adjusting for the nose factor, the rear factor, and their interactions.

rear1:

At $\alpha_{\text{Bonferroni}} = 0.00714$, we fail to reject the null hypothesis, since the p-value, 0.1857, is greater than $\alpha_{\text{Bonferroni}} = 0.00714$. We conclude that the impact of the rear factor is not statistically significant after adjusting for the nose factor, the middle factor, and their interactions.

nose1:middle1:

At $\alpha_{Bonferroni} = 0.00714$, we fail to reject the null hypothesis, since the p-value, 0.4522, is greater than $\alpha_{Bonferroni} = 0.00714$. We conclude that the impact of the nose and middle interaction is not statistically significant after adjusting for the three way interaction between the nose, middle, and rear factors.

nose1:middle1:rear1:

At $\alpha_{Bonferroni} = 0.00714$, we fail to reject the null hypothesis, since the p-value, 0.2265, is greater than $\alpha_{Bonferroni} = 0.00714$. We conclude that the impact of the middle and rear interaction is not statistically significant after adjusting for the three way interaction between the nose, middle, and rear factors.

nose1:rear1:

At $\alpha_{Bonferroni} = 0.00714$, we fail to reject the null hypothesis, since the p-value, 0.6939, is greater than $\alpha_{Bonferroni} = 0.00714$. We conclude that the impact of the nose and rear interaction is not statistically significant after adjusting for the three way interaction between the nose, middle, and rear factors.

nose1:middle1:rear1:

At $\alpha_{Bonferroni} = 0.00714$, we fail to reject the null hypothesis, since the p-value, 0.2193, is greater than $\alpha_{Bonferroni} = 0.00714$. We conclude that the impact of the nose, middle, and rear interaction is not statistically significant.

Model Checking

Factorial experiments have the same assumptions as ANOVA assumptions, which include checking for normality of the data, structure to the data, and equal variances. If one of these assumptions fail, we will need to perform a permutation test on it to obtain an accurate p-value.

```
par(mfrow = c(2, 2))
# Normality of Data
hist(model1$residuals, xlab = "residuals", main = "Paper Airplane Distances Residuals")

qqnorm(model1$residuals)
qqline(model1$residuals)

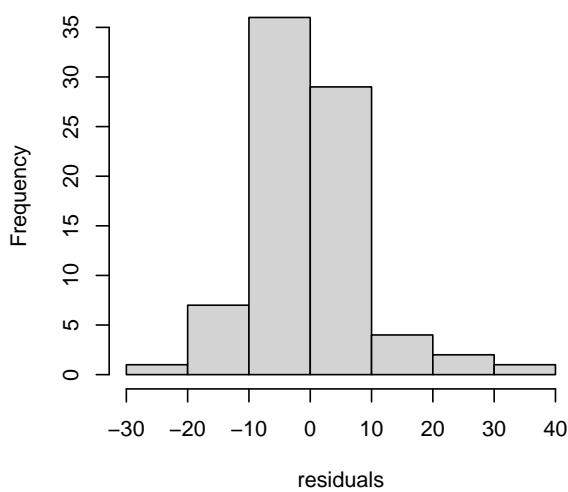
shapiro_result <- shapiro.test(model1$residuals)
shapiro_data <- data.frame(W = shapiro_result$statistic,
                             p_value = shapiro_result$p.value)
kable(shapiro_data)
```

	W	p_value
W	0.9436091	0.001511

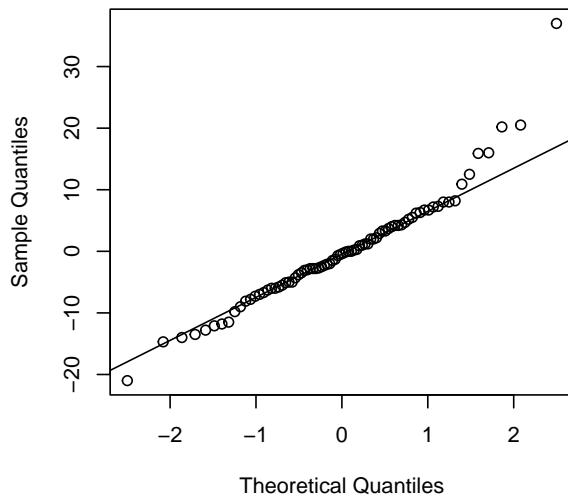
```
# Structure of Data
x <- 1:length(model1$residuals)
plot(model1$residuals ~ x, ylab = "residuals", cex.lab = 2,
     main = "Residuals vs. Order of Data", cex.main = 2)

# Equality of Variances
plot(model1$residuals ~ model1$fitted.values,
      xlab = "fitted values", ylab = "residuals", cex.lab = 2,
      main = "Residuals vs. fitted values", cex.main = 2)
```

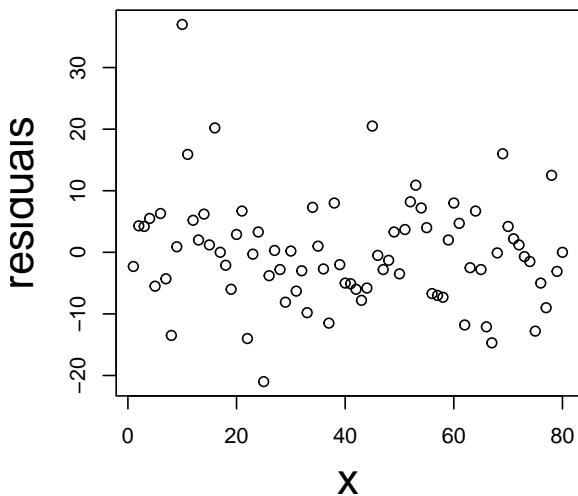
Paper Airplane Distances Residuals



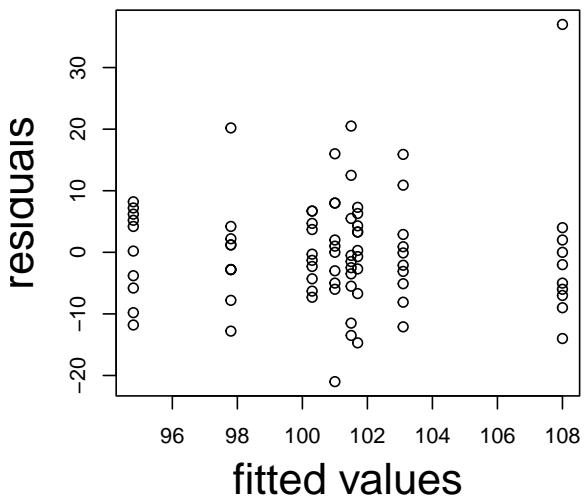
Normal Q-Q Plot



Residuals vs. Order of Data



Residuals vs. fitted values



Normality of Data: At $\alpha = 0.05$, we reject the null hypothesis, since the p-value, 0.001511, is less than $\alpha = 0.05$. We conclude that we found statistically significant evidence that the residuals do not follow a normal distribution. Since one of the assumptions of the factorial experiment was violated, the p-value obtained from the linear model may not be accurate.

Structure of Data: Based on the graph, there are no noticeable patterns when x increases. While there are a few outliers, most of the data points follow a horizontal path centered at 0. We can conclude that there is no structure to the data.

Equality of Variances: There are no noticeable patterns when fitted values increase. They appear to be within specific values and there are no shapes to the distribution. We can conclude that there is equal variances.

The three assumptions needed for the factorial experiment were normality of residuals, no structure to the data, and equal variances. Since the first assumption, normality of residuals, was violated, the linear model may not provide an accurate p-value, so we will need to perform the permutation test on it to obtain an accurate p-value.

Permutation Test

```
perm_f <- NA
reps <- 10000

for (i in 1:reps){
  perm_data <- airplane_dataframe
  perm_data$distance <- sample(perm_data$distance)

  perm_model <- lm(distance ~ nose + middle + rear +
    nose*middle + middle*rear + nose*rear +
    nose*middle*rear, data = perm_data)
  perm_f[i] <- summary(perm_model)$fstatistic[1]
}

# Comparing data value to the permuted null distribution
F <- summary(model1)$fstatistic[1]

perm_p_value <- sum(perm_f >= F) / reps
```

At $\alpha = 0.05$, we fail to reject the null hypothesis, since we obtain a permuted p-value of 0.1129. We conclude that there is no combination of factors that has a statistically significant effect on the final distance of the paper airplane.

Discussion

This study analyzed the effects of different combinations of paperclips on a paper airplane and its effect on the final distance of the plane. The three factors were the positions of the paperclip on the plane, the nose, middle, and rear, where we examined the effects of all 8 combinations of these paperclip positions. The experiment was conducted using 10 replicates of the 8 combinations, with randomization of the trials within the first 6 replicates, while 4 additional replicates were added later on. From our results, we found that no combination of these paperclips had a significant effect on the final distance of the plane. Our results did not support the hypothesis that some combination of the weights on the plane would make a significant difference in its final distance with regards to the aerodynamics of its design.

Overall, these results highlight the impact of weight distribution on airplane flight distances. The results show that the paperclips' weight was not significant enough in relation to the plane's weight, so its effects were minimal. This suggests that the weights need to be heavier to have a more noticeable impact on the flight distance. While this experiment was done on a smaller scale, this idea extends to larger aircrafts, where an emphasis of heavier weights relative to the plane's own weight can impact its flight, as suggested by earlier studies (Boyd et al., 2016, Nogoud et al., 2023, Liu et al., 2018). Optimizing the weight, as well as determining the ideal center of gravity, can lead to longer distance flights and efficient use of resources.

Various issues with the experiment included the first six replicates being done on one day, while the last four replicates were done on another day. Since both conditions weren't the same, as well as the full ten replicates not being completely randomized, there was bias introduced into the dataset. The first six replicates originally met all three assumptions of the linear model test, however after adding the last four

replicates, the data failed the normality assumption, which prompted us to use the permutation test to obtain an accurate p-value. This indicates some bias within our data. Another issue included the length and positioning of the measuring tape. Since the tape was only twelve feet long, any distance beyond that point would have to be estimated roughly, which leaves more room for error within the dataset. The measuring tape was set on the ground to be as straight as possible, with the throws closely following the same direction. Additionally, the design of the experiment indicated that the final distance was recorded by the first impact with the ground, where any position behind the measuring tape was a zero, and if the plane traveled further sideways, only the vertical distance from the start to the first impact would be recorded. This was done for simplicity in data collection, but it would be difficult to generalize this data to a broader context, since the distance recorded was only the positive vertical distance from the starting point. For future studies, it would be best to account for all directions when recording distances to ensure a more accurate representation of the study question at hand.

References

- Boyd, D. D. (2016). General aviation accidents related to exceedance of airplane weight/center of gravity limits. *Accident Analysis & Prevention*, 91, 19-23.
- Liu, Y., Yang, Z., Deng, J., & Zhu, J. (2018, March). Investigation of fuel savings for an aircraft due to optimization of the center of gravity. In IOP Conference Series: Materials Science and Engineering (Vol. 322, No. 7, p. 072018). IOP Publishing.
- Livingston, S. (2019, December 3). Making a Paper Airplane. The Art of Making Things.
<https://artofmakingthings.com/articles/making-a-paper-airplane>
- Nogoud, Y. A., Mohamed, O., KamalAlden, M., & Abuelnuor, A. A. (2023). Reducing the amount of fuel consumed by adjusting the location of the center of gravity. *Results in Engineering*, 20, 101482.