Paper Airplanes Full Factorial Design

Jimmy Wu

2025-06-14

Introduction

Design optimization is a foundamental concept across engineering, manufacturing, and even hobbyist domains, where small structural changes can lead to significant differences in performance. In aerodynamics, the key criterion of a successful glider is that the center of mass must be in the 'just right' place (Ristroph & Wang, 2022), as the distribution of weight and shape symmetry are particularly critical for flight stability and distance (Swatton, P.J. 2000). If the center of gravity is too far forward, the airplane may dive; if it's too far back, it may stall (Learning Corner, n.d.). Therefore, it is crucial to make sure that airplane has the exact desired weight distribution before it departs to avoid fatal accidents.

The importance of weight distribution on modern airplanes can then be generalized to many area of interests, such as military, vehicle cargo transportation, and public transportation, to which many critical factors such as safety, profits, and national security are involved. Particularly, such principles are used in aircraft and unmanned aerial vehicle (UAV) design, where the placement of cargo, fuel, and structural components can impact lift (for aircraft), balance, and range (Zhang, Zhang, & Chen. 2020). It can even be a pilot study of many more sophisticated aerodynamic weight-carrying efficiency circumstances (Ismail & Ali. 2021), such as which location should the aircraft carries missiles so that impacts on flight stability can be minimized when the aircraft drops the missiles.

Even in informal educational settings, understanding how balance and mass affect flight can deepen a student's understanding of physical forces as the core question of this topic is rooted in real-world physics. The topic not only help student understand the basics physic ideas and formulas taught in books, but also envisions the realistic application of the subject, potentially motivates prospective students to participate in the discourse in the future.

Paper airplanes, in this case, offer a simplified model through which aerodynamic principles can be explored, especially for educational or experimental purposes. This project investigates how the placement of small weights — specifically, paper clips — affects the flight distance of a paper airplane. Using a full factorial design with three binary factors (presence or absence of a paper clip on the nose, middle, or rear), we aim to determine which configurations impacts flight performance.

Methods

In this experiment, we investigated how the placement of paper clips on a paper airplane affects its flight distance. We consider three binary factors:

- **Nose**: Whether a paper clip is attached to the nose of the plane (0 = No, 1 = Yes)
- **Middle**: Whether a paper clip is attached to the middle (0 = No, 1 = Yes)
- **Rear**: Whether a paper clip is attached to the rear (0 = No, 1 = Yes)

Each factor has 2 levels, resulting in a total of $2^3 = 8$ unique treatment combinations. After a thorough sample size calculation using `power_factorial_23()` function and interactive plotting, we concluded that the minimum replicates for each treatment combination should be 4 times. Thus, to accommodate errors caused by the informal experimental environment, we applied 5 replicates for each combination, yielding 40 total observations.The response variable of interest was flight distance, measured in inches.

All flights were conducted indoors along a hallway to minimize environmental interference such as wind. The plane was thrown from a fixed starting line each time, and a measuring tape was laid out along the floor to mark where the plane first landed, whether or not the airplanes hits any object (i.e. wall, etc). Each throw was made by the same individual in order to reduce variation due to differences in throwing technique.

Besides ensuring identical throwing protocol for every replicate, the design of paper airplanes were also kept the same. The paper plane was folded identically for all trials using standard printer paper. Paper clips were attached securely and consistently in each assigned location. The order in which treatment combinations were flown was randomized using the `sample()` function in R with a fixed `set.seed()` to ensure reproducibility and reduce potential biases caused by learning, fatigue, or other temporal effects.

However, due to the limitation of the experimental environment, the paper airplanes hits the wall more often than the previous conducted studies. While ensuring that distance-measuring protocols are properly followed (i.e. only the distance where the paper airplanes first landed will be measured), we found that the throwing angel has a significant impact on the paper airplanes' horizontal direction and therefore we improved throwing techniques along the way. The chance for paper airplanes to hit the wall greatly reduced after only a few trials.

Our data were analyzed using a series of Linear Models that included all main effects and interaction terms. This approach allows us to assess not only the individual influence of each clip placement, but also whether combinations of placements interact to affect flight performance. For linear model test results to establish, there are three assumptions on the Linear Model Test that have to be met:

1. **Normality of the Data**
2. **No structure to the Data**
3. **Equal Variances across Fitted Values**

All of the three assumptions are proven to be met after conducting model checking section, indicating that our derived conclusion based on the linear model test is statistically valid and a permutation test is not necessarily required.

Results

Sample Size Calculation

We now conduct a pilot study, using the first two full replicates of our experiment as our pilot data. We start by running a linear model test on them.

```r
# We take out the first two full replicates of the experiment.
a_data <- read.csv("Paper_Airplane_Data.csv")

first_2_reps <- a_data %>%
  filter(Replicate %in% c(1, 2)) %>%
  arrange(Replicate)

kable(first_2_reps)
```

Replicate	Nose	Middle	Rear	Distance
1	0	0	0	164
1	0	0	1	138
1	0	1	0	141
1	0	1	1	126
1	1	0	0	147
1	1	0	1	121
1	1	1	0	120
1	1	1	1	100
2	0	0	0	159
2	0	0	1	153
2	0	1	0	141
2	0	1	1	122
2	1	0	0	133
2	1	0	1	111
2	1	1	0	140
2	1	1	1	112

```r
# Run a Linear Model Test on the first two full replicates
model0 <- lm(Distance ~ Nose + Middle + Rear + Nose*Middle +
               Nose*Rear + Middle*Rear + Nose*Middle*Rear, data = first_2_reps)

# Coefficient Summary
coef_df_0 <- as.data.frame(summary(model0)$coefficients)
colnames(coef_df_0) <- c("Estimate", "Std. Error", "t value", "Pr(>|t|)")

kable(signif(coef_df_0, 4), digits = 4, caption = "LM Coefficient Summary on the First Two Full Replica
```

Table 2: LM Coefficient Summary on the First Two Full Replicates

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	161.5	5.879	27.4700	0.0000
Nose	-21.5	8.314	-2.5860	0.0323
Middle	-20.5	8.314	-2.4660	0.0390
Rear	-16.0	8.314	-1.9240	0.0905
Nose:Middle	10.5	11.760	0.8930	0.3979
Nose:Rear	-8.0	11.760	-0.6804	0.5155
Middle:Rear	-1.0	11.760	-0.0850	0.9343
Nose:Middle:Rear	1.0	16.630	0.0601	0.9535

We obtained a series of estimates and standard errors from the Linear Model. For the sake of power and therefore sample size simulation, we define the initial `beta_mean` values as the estimates we obtained by running `model0`, and the initial `beta_se` values as the standard errors we obtained by, again, running `model0`. That is, `beta_mean` and `beta_se` is our pilot statistics for the first two full experimental replicates.

```r
set.seed(5192025)

source("power_factorial_23.R")
beta_mean <- c(161.5, -21.5, -20.5, -16.0, 10.5, -8.0, -1.0, 1.0)
beta_se <- c(5.879, 8.314, 8.314, 8.314, 11.760, 11.760, 11.760, 16.630)
```

Now, we use the `power_factorial_23` function to test on the number of replicates on each treatment group we need based on our pilot statistics.

```
l3 <- power_factorial_23(beta_mean, beta_se, 3)
l4 <- power_factorial_23(beta_mean, beta_se, 4)
l5 <- power_factorial_23(beta_mean, beta_se, 5)

power_df <- data.frame(
  replicates = c(3,4,5),
  power = c(l3,l4,l5)
)

kable(power_df)
```

replicates	power
3	0.784
4	0.945
5	0.989

If we repeat each treatment group by 3 times, this gives us a power of 0.784. However, if we repeat each treatment group by 4 or 5 times, the power, 0.945 and 0.989 accordingly, will be significantly higher than the traditional 0.8 power level. At this point, this suggests that we need to do a minimum of 4 replicates on each treatment group to ensure the 0.8 traditional power level is met.

The lowest beta value we get by the model is β_1 = -21.5, which is worth-investigating as there is a chance that the true beta value it may not be (as extreme as) this value. Therefore, we now conduct a power simulation on this particular value using our minimum full replicates of 4 times for each treatment group.

```
set.seed(5192025)

beta1 <- seq(0,-21.5, length.out=11)
power <- matrix(NA, nrow=1, ncol=11)

for(i in 1:length(beta1)){
  beta_mean[2] <- beta1[i]
  power[i] <- power_factorial_23(beta_mean, beta_se,
                                 replicates=4)
}

colnames(power) <- beta1
kable(power)
```

0	-2.15	-4.3	-6.45	-8.6	-10.75	-12.9	-15.05	-17.2	-19.35	-21.5
0.793	0.761	0.818	0.784	0.826	0.827	0.866	0.889	0.902	0.919	0.944

This shows that even if the effect size of β_1 was as small as -4.3, we will still have a power of 0.818, given that every other thing holds equal. Therefore, repeating each treatment group by at least 4 times should be a fair amount of replication to ensure the power level.

The second-to-lowest beta value we get by the model is β_2 = -20.5, and so we also conduct a power simulation on the variation of it, again using our minimum full replicates of 4 times for each treatment group.

```r
set.seed(5212025)

beta2 <- seq(0,-20.5, length.out=11)
power2 <- matrix(NA, nrow=1, ncol=11)

for(i in 1:length(beta2)){
  beta_mean[3] <- beta2[i]
  power2[i] <- power_factorial_23(beta_mean, beta_se, replicates=4)
}

colnames(power2) <- beta2
kable(power2)
```

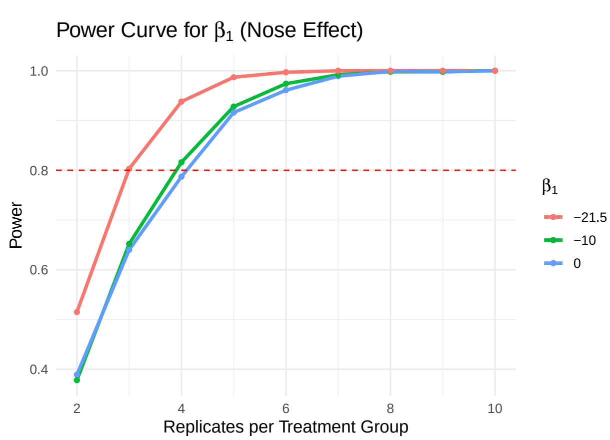
0	-2.05	-4.1	-6.15	-8.2	-10.25	-12.3	-14.35	-16.4	-18.45	-20.5
0.889	0.879	0.867	0.873	0.863	0.876	0.87	0.896	0.916	0.924	0.952

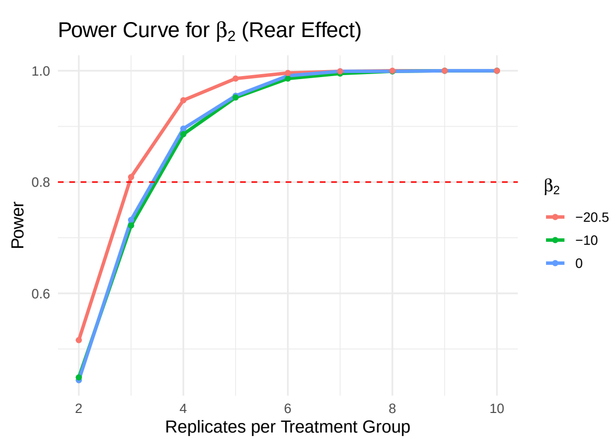
This shows that even if the effect size of β_2 was as small as 0, we will still have a power of 0.889, given that every other thing holds equal. Additionally, among all the results given the variating β_2 value, none of the power level falls below 0.8.

To better understand the impact of different β_1 and β_2 level on the power, we created two side-by-side plots.

```r
source("power_factorial_23.R")

replicates <- 2:10

beta_se <- c(5.879, 8.314, 8.314, 8.314, 11.760, 11.760, 11.760, 16.630)

# Beta 1 power curve
beta1_vals <- c(0, -10, -21.5)
beta1_df <- data.frame()

for (b1 in beta1_vals) {
  for (r in replicates) {
    beta_vec <- c(161.5, b1, -20.5, -16.0, 10.5, -8.0, -1.0, 1.0)
    pwr <- power_factorial_23(beta_vec, beta_se, replicates = r)
    beta1_df <- rbind(beta1_df, data.frame(replicates = r, power = pwr, beta1 = b1))
  }
}


library(ggplot2)
ggplot(beta1_df, aes(x = replicates, y = power, color = factor(beta1))) +
  geom_line(linewidth = 1.2) +
  geom_point() +
  geom_hline(yintercept = 0.8, linetype = "dashed", color = "red") +
  labs(
    title = expression("Power Curve for " * beta[1] * " (Nose Effect)"),
    x = "Replicates per Treatment Group",
    y = "Power",
    color = expression(beta[1])
  ) +
```

```
  theme_minimal(base_size = 13)


# Beta 2 power curve
beta2_vals <- c(0, -10, -20.5)
beta2_df <- data.frame()

for (b2 in beta2_vals) {
  for (r in replicates) {
    beta_vec <- c(161.5, -21.5, b2, -16.0, 10.5, -8.0, -1.0, 1.0)
    pwr <- power_factorial_23(beta_vec, beta_se, replicates = r)
    beta2_df <- rbind(beta2_df, data.frame(replicates = r, power = pwr, beta2 = b2))
  }
}

# Plot for beta2
ggplot(beta2_df, aes(x = replicates, y = power, color = factor(beta2))) +
  geom_line(linewidth = 1.2) +
  geom_point() +
  geom_hline(yintercept = 0.8, linetype = "dashed", color = "red") +
  labs(
    title = expression("Power Curve for " * beta[2] * " (Rear Effect)"),
    x = "Replicates per Treatment Group",
    y = "Power",
    color = expression(beta[2])
  ) +
  theme_minimal(base_size = 13)
```





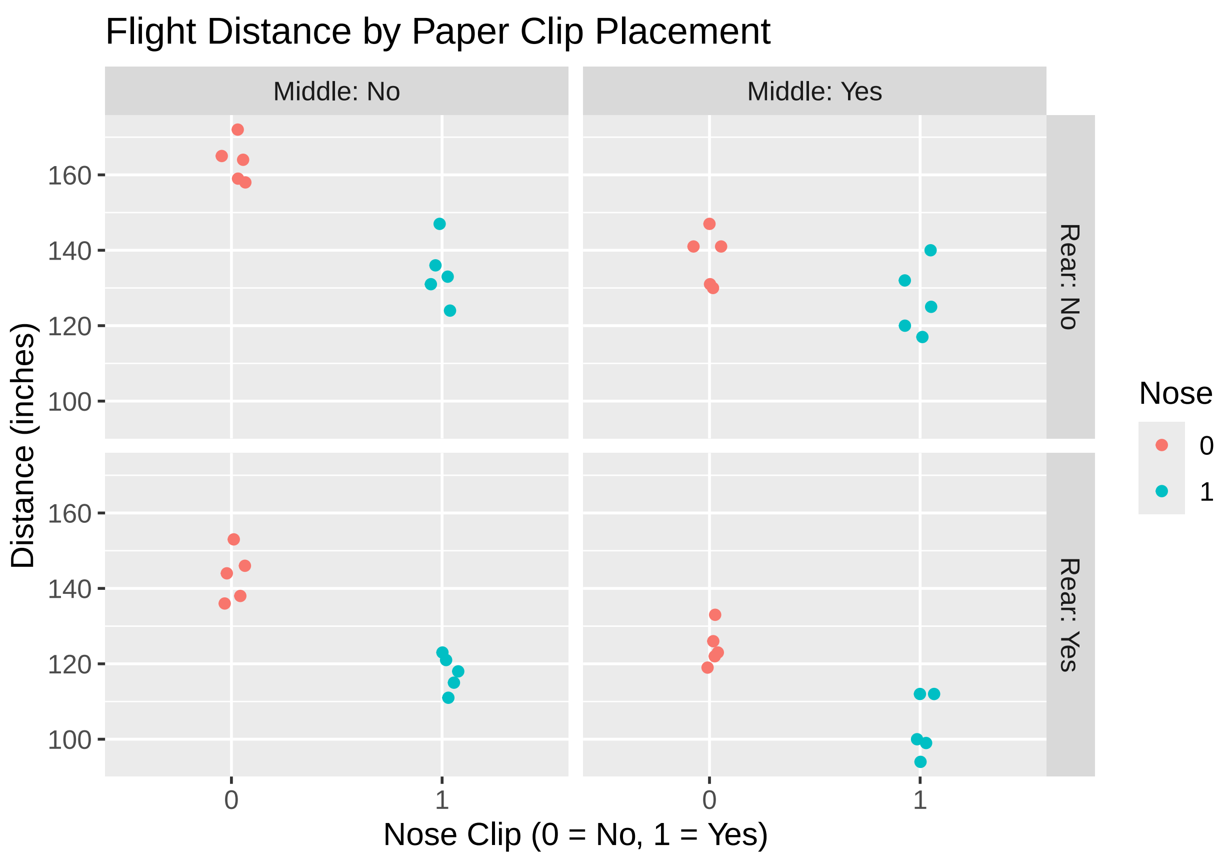
The power curve also shows consistent trend. That is, when the full replicates is or is over 4 times, the power of our experiment will be over 0.8 for both three levels of β_1 and β_2.Therefore, again, repeating each treatment group by at least 4 times should be a fair amount of replication to ensure the power level.

Since the experimental environment is technically informal, we increase the number of replicates to 5 for each treatment group to accommodate any potential inaccuracies or effects caused by environmental factors.

Data Visualization

We start by creating a jittered scatter plot that represents the interactions between each factor of interest (i.e. placing or not placing paper clip on the airplane's nose, middle and rear) after read in the data.

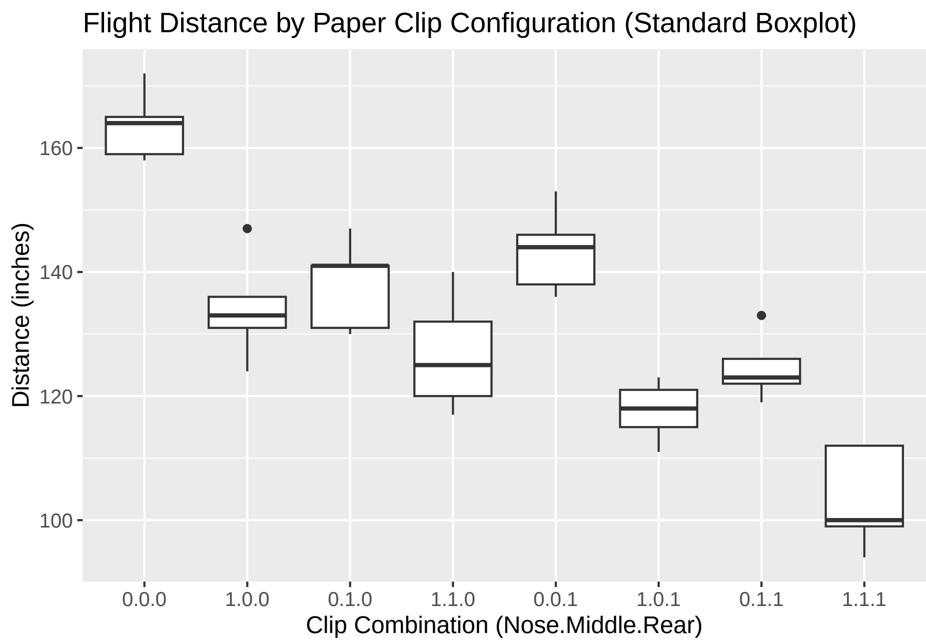
```r
a_data$Nose <- as.factor(a_data$Nose)
a_data$Middle <- as.factor(a_data$Middle)
a_data$Rear <- as.factor(a_data$Rear)

# Update theme for clarity
theme_update(text = element_text(size = 12))

# Create jittered scatter plot faceted by Middle and Rear, colored by Nose
ggplot(data = a_data, mapping = aes(x = Nose, y = Distance, color = Nose)) +
  geom_jitter(width = 0.08, height = 0) +
  facet_grid(Rear ~ Middle, labeller = labeller(
    Rear = c("0" = "Rear: No", "1" = "Rear: Yes"),
    Middle = c("0" = "Middle: No", "1" = "Middle: Yes")
  )) +
  ggtitle("Flight Distance by Paper Clip Placement") +
  xlab("Nose Clip (0 = No, 1 = Yes)") +
  ylab("Distance (inches)")
```



Alternatively, we also visualized the data with a standard boxplot and a table of summary statistics for a more convenient reference.

```r
# Boxplot
ggplot(a_data, aes(x = interaction(Nose, Middle, Rear), y = Distance), fill = Nose) +
  geom_boxplot() +
  xlab("Clip Combination (Nose.Middle.Rear)") +
```

```r
  ylab("Distance (inches)") +
  ggtitle("Flight Distance by Paper Clip Configuration (Standard Boxplot)")
```




```r
# Summary Statistics
grouped_a_data <- a_data %>%
  group_by(Nose, Middle, Rear) %>%
  summarize(sample_size = n(),
            sample_mean = mean(Distance),
            sample_sd = sd(Distance),
            .groups = "drop" # This is just to remove the warning message
            )

kable(grouped_a_data, caption = "Summary Statistics of Flight Distance")
```

Table 6: Summary Statistics of Flight Distance

Nose	Middle	Rear	sample_size	sample_mean	sample_sd
0	0	0	5	163.6	5.594640
0	0	1	5	143.4	6.767570
0	1	0	5	138.0	7.280110
0	1	1	5	124.6	5.319774
1	0	0	5	134.2	8.408329
1	0	1	5	117.6	4.774935
1	1	0	5	126.8	9.311284
1	1	1	5	103.4	8.173127

The visualization reveals clear trends in how paper clip placement affects flight distance. In the jittered scatterplot, planes without a paper clip on the nose (Nose = 0) consistently fly farther than those with a clip on the nose (Nose = 1) in nearly all panels, suggesting a strong negative impact of nose placement on distance. The effect of rear and middle clips appears more complex. For instance, when Rear = 0 and Middle

= 0, distances are generally higher than when either or both are set to 1. This suggests that adding clips to the middle or rear may also reduce flight distance, although the reduction is not as dramatic as with the nose placement.

The trend is also evidence in the boxplot and our summary statistics table.The combination with no paper clips at all (0.0.0) consistently results in the longest flight distance of around 163.6 inches. This is evident in both the boxplot, which shows the highest range for this group, and the bar chart, where it has the greatest mean distance among all combinations. As more clips are added, especially in multiple positions, the average distance flown by the paper airplane tends to decrease. For example, the combination (1.1.1), where clips are attached to all three positions, shows the lowest average flight distance of 103.4 inches.

Overall, the plots supports the idea that clip position impacts flight performance, particularly the nose placement, and interactions among factors may be influencing the results as well. We will investigate whether the impacts are statistically significant or not in the following sections.

Statistical Tests

With the doubts in mind, we then proceeds to conduct the appropriate statistical tests. In this case, the null hypothesis is that none of the clip placements, whether individually or in combination, has**no** impact on the paper airplanes' flight distance, denoted as H_0: $\beta_{Nose} = \beta_{Middle} = \beta_{Rear} = \beta_{Nose:Middle} = \beta_{Nose:Rear} = \beta_{Middle:Rear} = \beta_{Nose:Middle:Rear} = 0$, where β_i is the effects on flight distance for the i paper clip configuration. The alternative hypothesis is that **at least one** effects on flight distance is **not**, denoted as H_A: At least one $\beta_i \neq 0$.

$$H_0 : \beta_{Nose} = \beta_{Middle} = \beta_{Rear} = \beta_{Nose:Middle} = \beta_{Nose:Rear} = \beta_{Middle:Rear} = \beta_{Nose:Middle:Rear} = 0$$
$$H_A : \text{At least one } \beta_i \neq 0$$

Overall Linear Model Test

To examine the interactions between the three variables and to provide a reference full model for the later main effect checking, we now conduct a linear model test.

```r
model1 <- lm(Distance ~ Nose + Middle + Rear + Nose*Middle +
             Nose*Rear + Middle*Rear + Nose*Middle*Rear, data = a_data)

# Coefficient Summary
coef_df <- as.data.frame(summary(model1)$coefficients)
colnames(coef_df) <- c("Estimate", "Std. Error", "t value", "Pr(>|t|)")

kable(signif(coef_df, 4), digits = 4, caption = "Linear Model Coefficient Summary")
```

Table 7: Linear Model Coefficient Summary

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	163.6	3.184	51.3900	0.0000
Nose1	-29.4	4.502	-6.5300	0.0000
Middle1	-25.6	4.502	-5.6860	0.0000
Rear1	-20.2	4.502	-4.4870	0.0001
Nose1:Middle1	18.2	6.367	2.8580	0.0074
Nose1:Rear1	3.6	6.367	0.5654	0.5757
Middle1:Rear1	6.8	6.367	1.0680	0.2935
Nose1:Middle1:Rear1	-13.6	9.004	-1.5100	0.1408

```r
# Summary Statistics

rse <- summary(model1)$sigma
df <- model1$df.residual
r2 <- summary(model1)$r.squared
adj_r2 <- summary(model1)$adj.r.squared
f_val <- summary(model1)$fstatistic[1]
f_df1 <- summary(model1)$fstatistic[2]
f_df2 <- summary(model1)$fstatistic[3]
p_val <- pf(f_val, f_df1, f_df2, lower.tail = FALSE)

formatted_p_val <- formatC(p_val, format = "e", digits = 3)  # scientific notation
formatted_rse <- signif(rse, 5)
formatted_r2 <- signif(r2, 5)
formatted_adj_r2 <- signif(adj_r2, 5)
formatted_f_val <- signif(f_val, 5)

model_summary_df <- data.frame(
  Statistics = c(
    "Residual Standard Error",
    "Degrees of Freedom (Residual)",
    "Multiple R-squared",
    "Adjusted R-squared",
    "F-statistic",
    "F-statistic DF (numerator)",
    "F-statistic DF (denominator)",
    "Overall Model p-value"
  ),
  Value = c(
    formatted_rse,
    df,
    formatted_r2,
    formatted_adj_r2,
    formatted_f_val,
    f_df1,
    f_df2,
    formatted_p_val
  )
)

kable(model_summary_df, caption = "Summary Statistics for Linear Model")
```

Table 8: Summary Statistics for Linear Model

Statistics	Value
Residual Standard Error	7.1186
Degrees of Freedom (Residual)	32
Multiple R-squared	0.87518
Adjusted R-squared	0.84788
F-statistic	32.054
F-statistic DF (numerator)	7
F-statistic DF (denominator)	32

Statistics	Value
Overall Model p-value	1.012e-12

The overall linear model included all three binary factors — Nose, Middle, and Rear — along with their two-way and three-way interactions. This model aimed to comprehensively evaluate how paperclip placements at different positions on a paper airplane influence flight distance. The model's F-statistic was 32.0538445, with numerator and denominator degrees of freedom of 7 and 32, respectively.The model captures substantial variation in flight distance (with an adjusted R^2 of 0.8478802, confirming that our factorial design effectively detected true differences in performance across treatment configurations.

Given that we are testing eight hypotheses (three main effects, three two-way interactions, one three-way interaction, and the intercept), we applied a Bonferroni-adjusted significance level of $\alpha = 0.05 / 8 = 0.0063$ to control for Type I error across multiple comparisons. Even with this stricter threshold, the overall model p-value of 0 is highly significant, indicating that the combination of predictors provides a strong fit to the data and that at least one of the effects — whether a main effect or an interaction — plays a significant role in explaining variation in flight distance. Namely, we reject the null hypothesis and conclude that at least one effect is statistically significant.

This result validates the importance of investigating individual and interactive effect of the three treatment combinations.

Main and Interaction Effect Linear Model Test

To further investigate the effect of each factor as well as their combination on the flight distance, we now conduct several additional linear model tests. This approach isolates each term's effect by evaluating how much it contributes to the model's explanatory power when omitted.

```r
# Nose main effect
no_nose_model <- lm(Distance ~ Middle + Rear + Middle*Rear, data = a_data)
nose_result <- anova(no_nose_model, model1)

kable(nose_result, digits = 10, caption = "Nose Main Effect ANOVA Table")
```

Table 9: Nose Main Effect ANOVA Table

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
36	6883.8	NA	NA	NA	NA
32	1621.6	4	5262.2	25.96053	1.2e-09

```r
p_val_nose <- formatC(nose_result[2, "Pr(>F)"], format = "e", digits = 1)
```

The p-value we derived for the main effect of Nose is 1.2e-09.

```r
# Middle main effect
no_middle_model <- lm(Distance ~ Nose + Rear +  Nose*Rear, data = a_data)
middle_result <- anova(no_middle_model, model1)

kable(middle_result, digits = 10, caption = "Middle Main Effect ANOVA Table")
```

Table 10: Middle Main Effect ANOVA Table

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
36	4784.6	NA	NA	NA	NA
32	1621.6	4	3163	15.60434	3.509e-07

```
p_val_middle <- formatC(middle_result[2, "Pr(>F)"], format = "e", digits = 1)
```

The p-value we derived for the main effect of Middle is 3.5e-07.

```
# Rear main effect
no_rear_model <- lm(Distance ~ Nose + Middle + Nose*Middle, data = a_data)
rear_result <- anova(no_rear_model, model1)

kable(rear_result, digits = 10, caption = "Rear Main Effect ANOVA Table")
```

Table 11: Rear Main Effect ANOVA Table

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
36	5148.4	NA	NA	NA	NA
32	1621.6	4	3526.8	17.39911	1.122e-07

```
p_val_rear <- formatC(rear_result[2, "Pr(>F)"], format = "e", digits = 1)
```

The p-value we derived for the main effect of Rear is 1.1e-07.

```
# Interaction Effect: Nose * Middle
no_nose_middle_model <- lm(Distance ~ Nose + Middle + Rear + Nose*Rear + Middle*Rear, data = a_data)
nose_middle_result <- anova(no_nose_middle_model, model1)

kable(nose_middle_result, digits = 10, caption = "Nose and Middle Interaction Effect ANOVA Table")
```

Table 12: Nose and Middle Interaction Effect ANOVA Table

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
34	2062.1	NA	NA	NA	NA
32	1621.6	2	440.5	4.346325	0.02138672

```
p_val_nose_middle <- formatC(nose_middle_result[2, "Pr(>F)"], format = "e", digits = 1)
```

The p-value we derived for the interaction effect of Nose and Middle is 2.1e-02.

```
# Interaction Effect: Nose * Rear
no_nose_rear_model <- lm(Distance ~ Nose + Middle + Rear + Nose*Middle + Middle*Rear, data = a_data)
nose_rear_result <- anova(no_nose_rear_model, model1)

kable(nose_rear_result, digits = 10, caption = "Nose and Rear Interaction Effect ANOVA Table")
```

Table 13: Nose and Rear Interaction Effect ANOVA Table

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
34	1762.8	NA	NA	NA	NA
32	1621.6	2	141.2	1.393192	0.2629366

```
p_val_nose_rear <- formatC(nose_rear_result[2, "Pr(>F)"], format = "e", digits = 1)
```

The p-value we derived for the interaction effect of Nose and Rear is 2.6e-01.

```
# Interaction Effect: Middle * Rear
no_middle_rear_model <- lm(Distance ~ Nose + Middle + Rear + Nose*Middle + Nose*Rear, data = a_data)
middle_rear_result <- anova(no_middle_rear_model, model1)

kable(middle_rear_result, digits = 10, caption = "Middle and Rear Interaction Effect ANOVA Table")
```

Table 14: Middle and Rear Interaction Effect ANOVA Table

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
34	1737.2	NA	NA	NA	NA
32	1621.6	2	115.6	1.140602	0.3322788

```
p_val_middle_rear <- formatC(middle_rear_result[2, "Pr(>F)"], format = "e", digits = 1)
```

The p-value we derived for the interaction effect of Middle and Rear is 3.3e-01.

```
# Three-Way Interaction Effect: Nose * Middle * Rear
no_threeway_model <- lm(Distance ~ Nose + Middle + Rear + Nose*Middle + Nose*Rear + Middle*Rear, data =
threeway_result <- anova(no_threeway_model, model1)

kable(threeway_result, digits = 10, caption = "Three-Way Interaction Effect ANOVA Table")
```

Table 15: Three-Way Interaction Effect ANOVA Table

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
33	1737.2	NA	NA	NA	NA
32	1621.6	1	115.6	2.281204	0.1407627

```
p_val_threeway <- formatC(threeway_result[2, "Pr(>F)"], format = "e", digits = 1)
```

The p-value we derived for the three-way interaction effect of Nose, Middle, and Rear is 1.4e-01.

Since we've conducted 7 anova test, our corresponding significance level should be $\alpha = 0.05\ /\ 7 = 0.0071$ by Bonferroni-adjusting.

The main effect of placing a clip on the Nose of the airplane yielded a p-value of 1.2e-09, well below the Bonferroni-adjusted threshold. This result suggests that having a clip on the nose significantly affects the flight distance, likely due to its strong impact on the airplane's center of gravity and initial trajectory. Similarly, placing a clip in the Middle resulted in a p-value of 3.5e-07, and placing it in the Rear gave a p-value of 1.1e-07, both indicating statistically significant effects. Thus, placing a clip on each of the three locations on a paper airplane will have a statistically significant impact on its flight distance.

Beyond individual placements, we examined the two-way interactions. The interaction between Nose and Middle clips yielded a p-value of 2.1e-02, suggesting that the combined effect of these two clips is statistically significant to the flight distance of paper airplane. The Nose:Rear interaction with a p-value of 2.6e-01 and Middle:Rear interaction with a p-value of 3.3e-01 does not met the Bonferroni threshold, confirming that combinations of two clips in different positions has no statistically significant impact on the flight distance.

The three-way interaction Nose:Middle:Rear had a p-value of 1.4e-01, which is greater than the Bonferroni threshold.Therefore, the interaction between nose, middle, and rear clip does not have a statistically significant impact on the paper airplane flight distance.

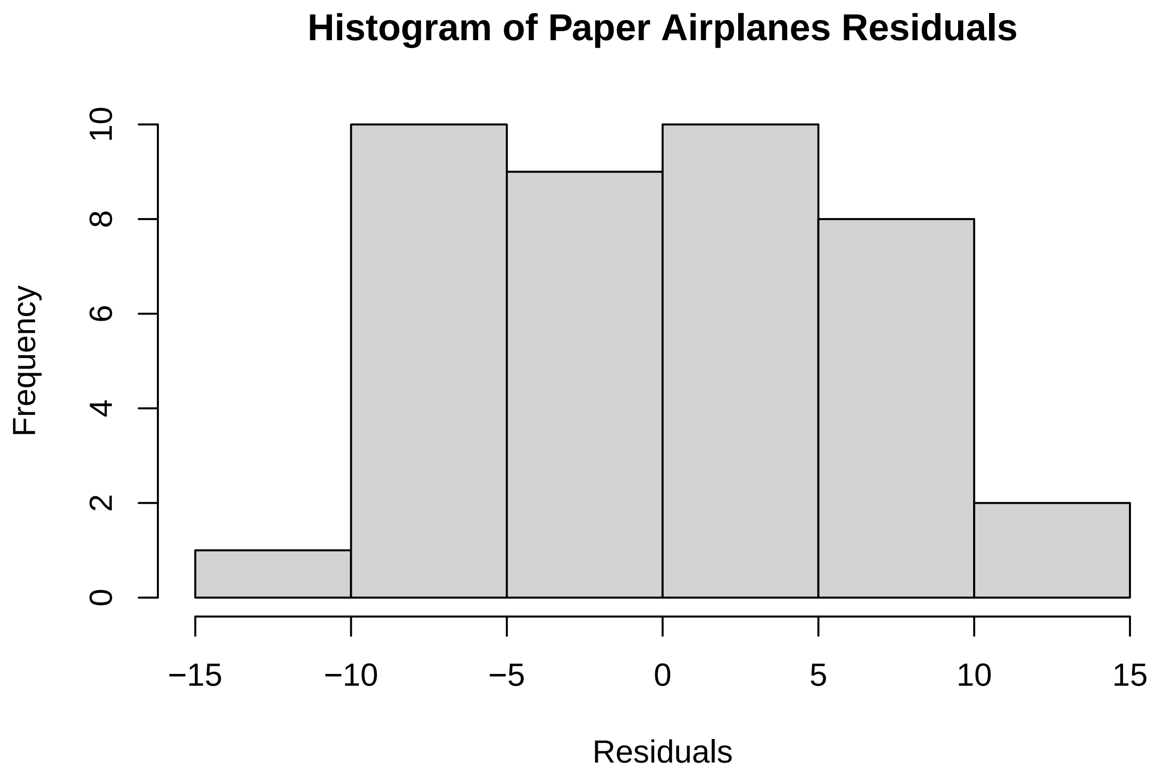
Model Checking

To ensure that our test results are reliable, we now conduct a model checking on the following three assumptions on linear model:

1. **Normality of the Data**
2. **No structure to the Data**
3. **Equal Variances across Fitted Values**

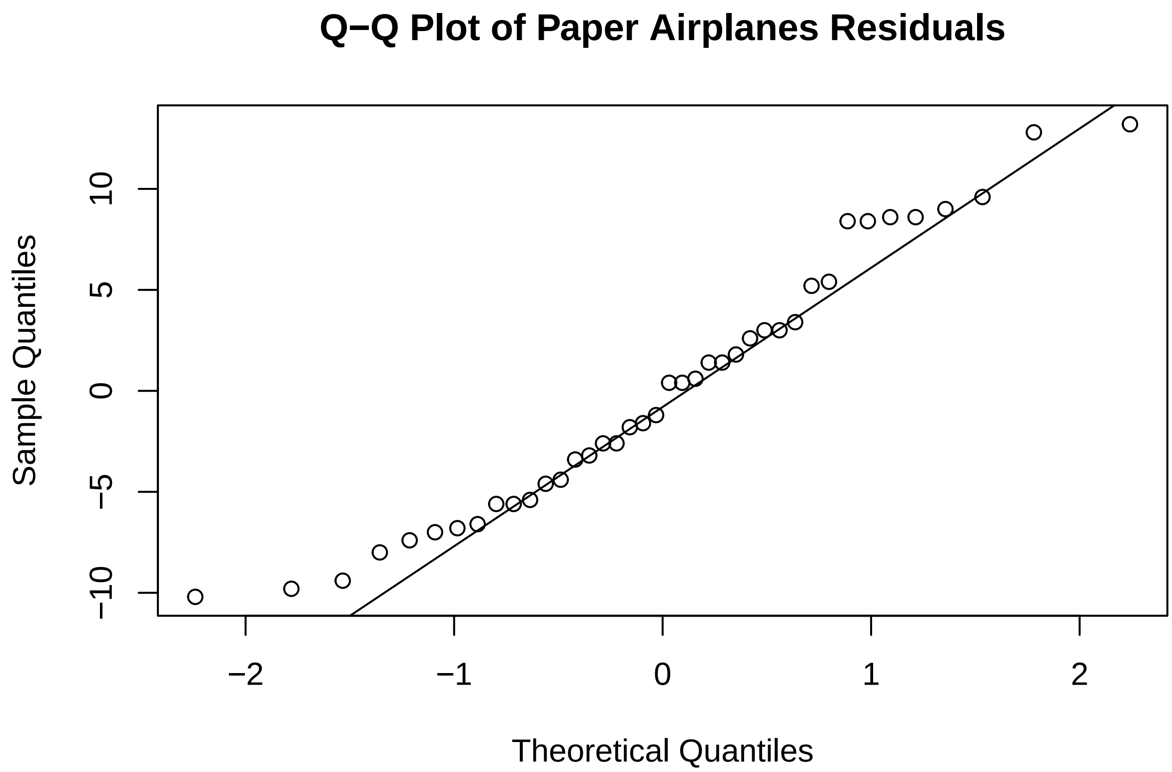
```r
# Histogram of the Residuals
hist(model1$residuals, main = "Histogram of Paper Airplanes Residuals", xlab = "Residuals")
```

Normality of the Data



According to the histogram, the residuals of the linear model looks approximately bell-shaped, with slightly more data clustered between the middle (residual value pf -10 ~ 10) of the distribution. Therefore, there should be little to no concerns about the normality of the data.

```
# Q-Q Plot
qqnorm(model1$residuals, main = "Q-Q Plot of Paper Airplanes Residuals")
qqline(model1$residuals)
```



According to the Q-Q plot, the residual data points lies close to the regression line, with a relatively slightly greater variations before -1 theoretical quantiles and after 1 theoretical quantiles. Therefore, again, there should be little to no concern about the normality of the data.

```r
# Shapiro-Wilk Normality Test
shapiro_result <- shapiro.test(model1$residuals)

shapiro_df <- data.frame(
  Statistic = signif(shapiro_result$statistic, 5),
  P_Value = signif(shapiro_result$p.value, 5)
)

kable(shapiro_df, caption = "Shapiro-Wilk Normality Test for Model Residuals")
```

Table 16: Shapiro-Wilk Normality Test for Model Residuals

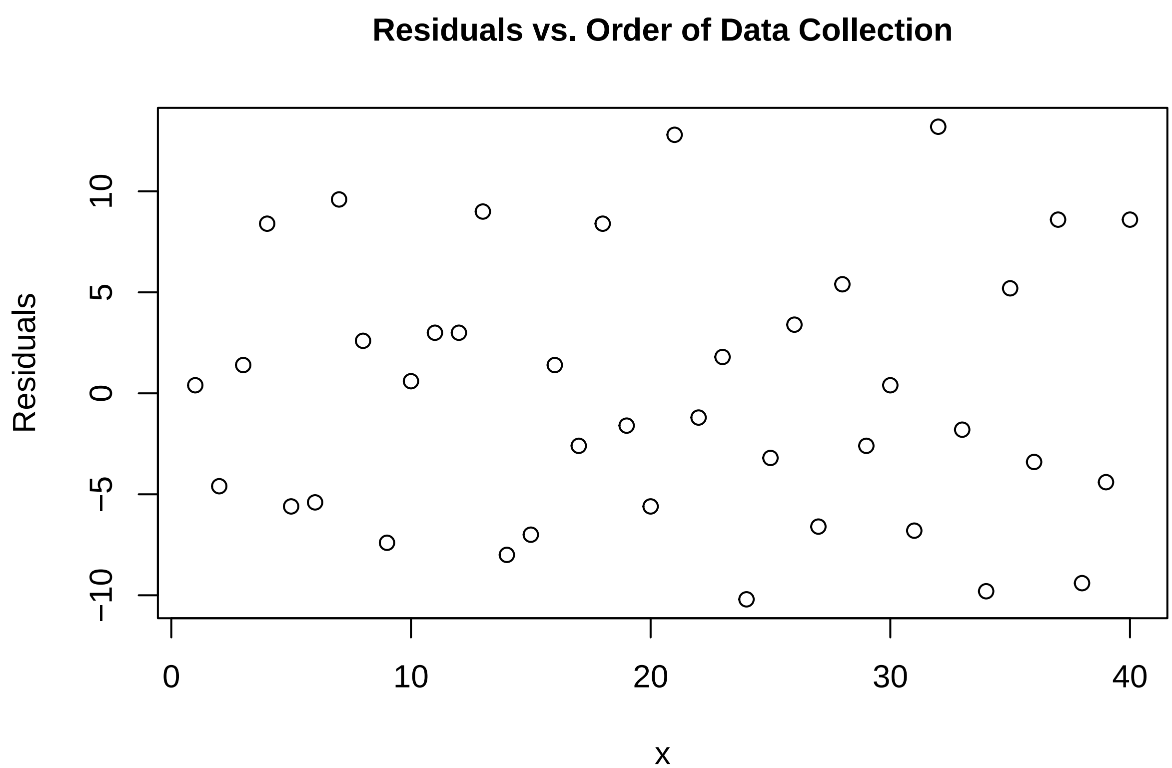
	Statistic	P_Value
W	0.95975	0.16428

According to the Shapiro-Wilk Normality Test, we derive a p-value of 0.16428 at $\alpha = 0.05$. Therefore, we fail to reject the null hypothesis of the test that our data follows a normal distribution and conclude that our data follows a normal distribution.

All three approaches we used to examine the normality of our data - histogram, Q-Q plot, and shapiro-wilk normality test - indicates the same conclusion: the normality assumption of the test is not violated and therefore **the normality assumption is met**.

```
x <- 1:length(model1$residuals)

plot(model1$residuals ~ x, ylab="Residuals", cex.lab=1,
  main="Residuals vs. Order of Data Collection", cex.main=1)
```

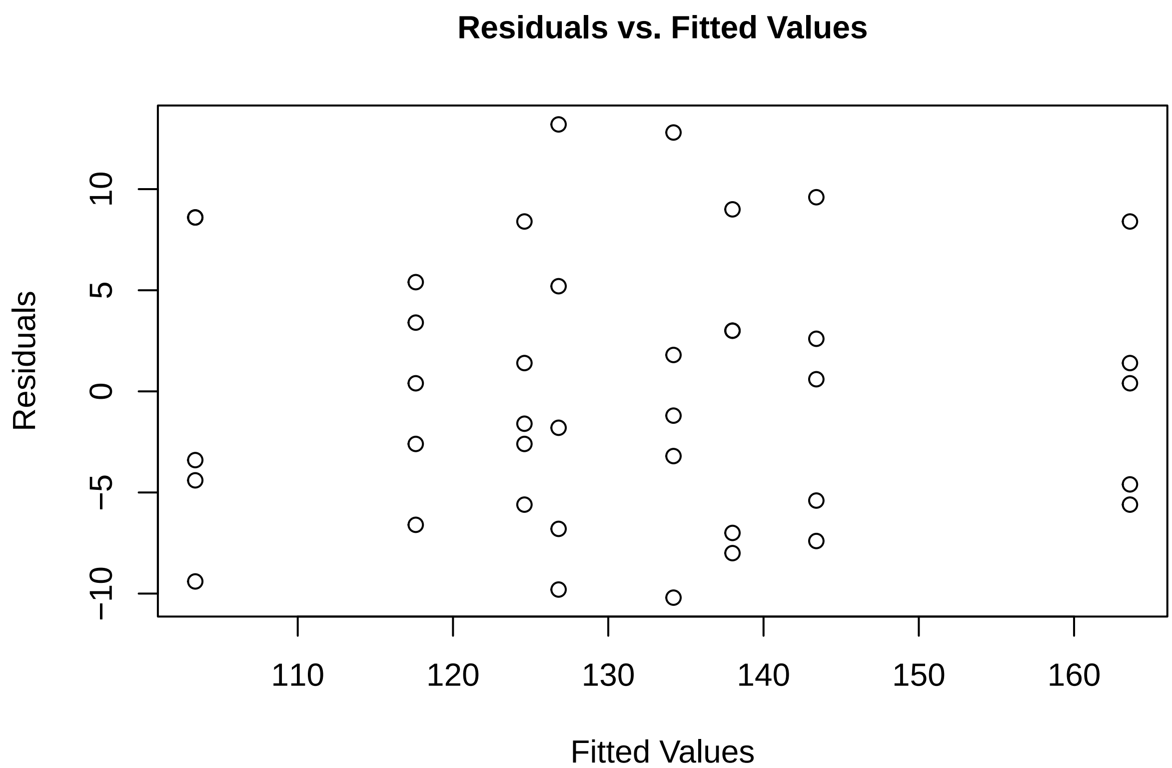
Structure of the Data



The residuals' value, across the order of data collections, does not seems to go up or down. Therefore, there are no apparent patterns in the plot and so the **no structure to the data assumption is met**.

```
plot(model1$residuals ~ model1$fitted.values,
  xlab="Fitted Values", ylab="Residuals", cex.lab=1,
  main="Residuals vs. Fitted Values", cex.main=1)
```

Equal Variances across Fitted Values



According to the plot, variances across different fitted values seems to be roughly the same without significant structure or pattern. Therefore, the **equal variance across fitted value assumption** is met.

Since all three assumptions on the linear model test are satisfied, the test result can be trusted and hence we do not need to conduct a permutation test.

Discussion

This study investigated how the placement of paper clips on a paper airplane — specifically on the nose, middle, and rear — affects its flight distance. Using a full factorial design, we tested all combinations of these three binary factors with five replicates each, resulting in 40 total observations. Our statistical analyses using a linear model confirmed that each clip location — nose, middle, and rear — significantly influences flight distance, with all three main effects being statistically significant at both the traditional $\alpha = 0.05$ level and the Bonferroni-adjusted threshold. Additionally, among the two-way interactions, we observed a significant interaction between nose and middle placements, indicating that the effect of placing clips in those locations has statistically significant impact on the flight distance. However, no significant three-way interaction was found, meaning the combination of all three clip placements did not have a compounded effect beyond the lower-order terms.

The results offer insights with broader relevance beyond paper airplane experiments. The fundamental principle that small changes in weight distribution can significantly impact flight performance reflects its importance in real-world engineering practices such as aircraft design and cargo balancing in aviation. These findings emphasize, again, the importance of weight placement in gliding objects, which has implications in aerospace design, education, and even prototyping for drones or similar lightweight aircraft. By showing how clip placement (as a proxy for mass distribution) directly affects flight stability and distance, this experiment helps contextualize abstract aerodynamic theory in an accessible way.

However, several practical limitations must be acknowledged. Although efforts were made to standardize procedures, such as consistent plane design, controlled indoor conditions, and a fixed throwing technique, the experiment was not conducted in a professionally controlled environment. The hallway setting occasionally caused the airplanes to strike walls, especially early in the data collection process. Improvements in throwing technique reduced this issue, but this learning curve could have introduced minor variability. This experiment was also conducted by a single participant, reducing variability due to thrower technique but also limiting generalizability to all kinds of throwers or throwing forces.

Additionally, since our study sample consisted of a single paper airplane type and thrower, results may not generalize to other plane designs, paper weights, or individuals with different throwing styles. Nonetheless, this minor consequence can be ignored as it fundamentally goes off and contradicts with our subject. Particularly, the subject is to investigate the effect of placing paper clip(s) on paper airplanes using a full factorial design; introducing different airplane designs would, for instance, turn the experiment into a complete block design.

In conclusion, this project confirms that clip placement meaningfully alters the flight distance of paper airplanes. The significant main effects and specific two-way interaction (Nose:Rear) highlight the nuanced ways in which mass distribution affects performance. These results highlight the importance of balance and weight configuration in flight, offering both a valuable learning opportunity and a practical example of factorial experimental design in action.

References

Ismail, N. I., & Ali, Z. M. (2021). *Aerodynamic Performances of Paper Planes.* Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, 77(1), 124–131.

Ristroph, L., & Wang, J. (2022). *Centre of mass location, flight modes, stability and dynamic modelling of gliders.* Journal of Fluid Mechanics, 933, A14.

Swatton, P. J. (2000). *Aircraft Performance Theory for Pilots.* Blackwell Science.

Learning Corner. (n.d.). *Understanding the Center of Gravity and Weight Distribution in Paper Airplanes for Optimal Flight Performance.* Learning Corner.

Zhang, Y., Zhang, X., & Chen, G. (2020). *The aerodynamic optimisation of a low-Reynolds paper plane with adjoint method.* The Aeronautical Journal, 125(1285), 510–524.