

Paper Airplanes Factorial Experiment

Agneya Poduval

PSTAT 122

2024-12-08

Introduction

Broad Statement

Understanding the principles of aerodynamics allows us to navigate the world effectively. The smallest increases in fuel efficiency or acceleration, for example, can make manufacturers millions of dollars. One example of this would be Airbus, who's introduction of the Airbus A-380 "super-jumbo aircraft" in 19 March 2007 (Simons, 2014) was set to "reduce the market share" of competitor Boeing's 747 by "up to 14.8%" (Irwin and Pavcnik, 2004). However, this decision would also "reduce the market for Airbus's existing wide-bodies", meaning that Airbus would have to balance the introduction of the A380 with the gradual replacement and removal of older commissioned vehicles. Such strategical and economic decisions must be taken carefully and can cause companies such as Airbus to lose or gain billions of dollars and advantages over their competitors.

Studying and understanding how airplanes fly in general is critical when designing more complex systems such as commercial jets and light drones. One such study, conducted by Bing Feng Ng, Qiao Mei Kng, Yin Yin Pey and Jorg Schluter at the School of Mechanical and Aerospace Engineering in Singapore, focuses on constructing "flow visualizations and force measurements" of dart paper airplanes to improve the performance of Micro Air Vehicles (Ng et al. 2009), which may be used in combat or for surveillance (Ifju et al. 2002). These MAVs are low cost, lightweight, and small in size, making them useful in accessing high risk areas such as radioactive zones, along with combat zones and other inaccessible areas. The study discusses how research conducted with paper airplanes was effective in generalizing to Micro Air Vehicles, which are more complex and include motor mechanisms that cannot be accounted for with simple paper airplane designs. This highlights how studying paper airplanes can be useful in designing more complex vehicles for a lower cost.

Specifics

As students, we can try to observe these principles in our daily life or with smaller scale experiments. In this experiment, we will investigate whether attaching paperclips to the front, middle, and back of a single paper airplane impacts its flight distance, showing how added weight and the distribution of that added weight influences the performance and aerodynamics of the plane. The addition of weight to a paper airplane allows us to explore how mass distribution and gravity influence its flight path. The results of this experiment could bear significance on the balance between weight distribution and stability in flight mechanics. Specifically, the the aim of this experiment is to find out whether there are any significant differences in flight distance that can be generated by simply moving added weight around an object in motion. In this case, the added weight is each paperclip, while the object in motion is the paper airplane.

Methods

Data Collection and Factors

I conducted a full factorial design with one paper airplane and three factors of interest, which were having a paper clip attached to the nose, middle, and/or rear of the plane in order to see which combination of these factors would have the most or any impact at all. Each factor had two levels, which were either having the clip on or having it off. With three factors and two levels to each factor, this created $2^3 = 8$ full combinations. Thus, I threw the plane 8 times per replicate. I first decided to conduct a pilot study so that I could determine what sample size I needed to reach 80% power with a high estimate of standard error. After finding that I needed 7 replicates to reach 80% power, I collected 7 new and unique replicates, making sure not to reuse the sample data. The airplanes were not thrown with full power, but instead were flicked forward. This made each of the flights more similar to each other, compared to Lab 2 where throwing the airplanes with full power resulted in wildly inaccurate results. In addition, this allowed me to conduct the experiment indoors, eliminating any issues caused by wind. Finally, note that the values were measured in feet and inches, then converted automatically when entering the data. For example, a value of 5 feet and 2 inches would be inputted into my dataset as $5 + 2 * 0.08 = 5.16$. Similarly, all values were converted to this format for simplicity, which is why many of the data points have similar ending significant digits. This method made the sample size power calculation much easier, as I could use the same standard error values from previous work.

Statistical Method and Assumptions

To analyze the data, I first ran a full interaction model to see whether any of the interaction terms could potentially be significant. After running the interaction model, I decided to compare it to a simpler model without any interaction terms, which did a much better job and actually had two statistically significant effects compared to the original model, which produced no statistically significant effects due to the very low Bonferroni-adjusted significance level that came from having seven total terms. Thus, I continued my analysis with the simpler linear model, rather than the full interaction model.

The assumptions for this simple linear regression are normality of the data, no structure to the data (autocorrelation), and equal variances throughout. Before running the linear model, I thought that these assumptions would be satisfied because I had a large enough sample size as given by my sample size power calculation. After visualizing the data, I noticed from the density plot that the data looked to be normally distributed with some slight skew and some potential outliers. This was slightly surprising, but because I had a large enough sample size, it makes sense that all of the data ended up looking normally distributed due to the Central Limit Theorem. I believe because I was throwing in a controlled airless environment, it makes sense that the distribution of the flight distances could look somewhat normally distributed.

Technical Issues

The biggest technical issue I had was that I was working with a relatively narrow indoor space, meaning that I had to make sure my flights were not going off track. This is why I chose to flick the flights forward rather than throw them, which could have caused them to go off track like in Lab 2. I was able to control my flicking motion and get the flights to go relatively forward, which made the data collection a smooth process. The only other potential issue that may have came up is potential inaccuracies in measurement as I was using a ruler which can only measure up to one foot to measure distances that were longer than that. Thus, some of the distances may not have been perfectly measured. However, since I was throwing in a controlled space, there aren't any wildly inaccurate measurements present in the data. Also, by nature of this factorial design, the same plane was thrown over 50 times. However, I did not notice any creasing or potential issues with the airplane, suggesting that all of the trials were carried out in a similar manner.

Data Collection Documentation

```
knitr::include_graphics("/Users/agnayapoduval/Documents/Classes/PSTAT 122/Lab Files/122Final.jpeg")
```



There is a clear starting point, which is the yellow line in the carpet. From the yellow line, the ruler was used to measure up to the spot where the tip of the paper airplane landed, not where it settled after bouncing off the ground or colliding with other objects. Since this was conducted indoors, there was no concern about wind or other external factors.

Results

Sample Size Calculations

```
# Initial Sample
flight_sample <- data.frame(
  Rear = c("0", "0", "0", "0", "1", "1", "1", "1"),
  Middle = c("0", "0", "1", "1", "0", "0", "1", "1"),
  Nose = c("0", "1", "0", "1", "0", "1", "0", "1"),
  Distance = c("3.42", "5.58", "3.50", "4.83", "2.33", "2.33", "6.67", "6.00")
)
flight_sample$Nose <- as.factor(flight_sample$Nose)
flight_sample$Middle <- as.factor(flight_sample$Middle)
flight_sample$Rear <- as.factor(flight_sample$Rear)

# Linear Model
flight_sample_model <- lm(Distance ~ Nose * Middle * Rear, data = flight_sample)
beta_mean <- as.vector(flight_sample_model$coefficients)

# Sample Size Calculation
source("/Users/agnepapoduval/Documents/Classes/PSTAT 122/Scripts/power_factorial_23.R")

replicates <- 2:10

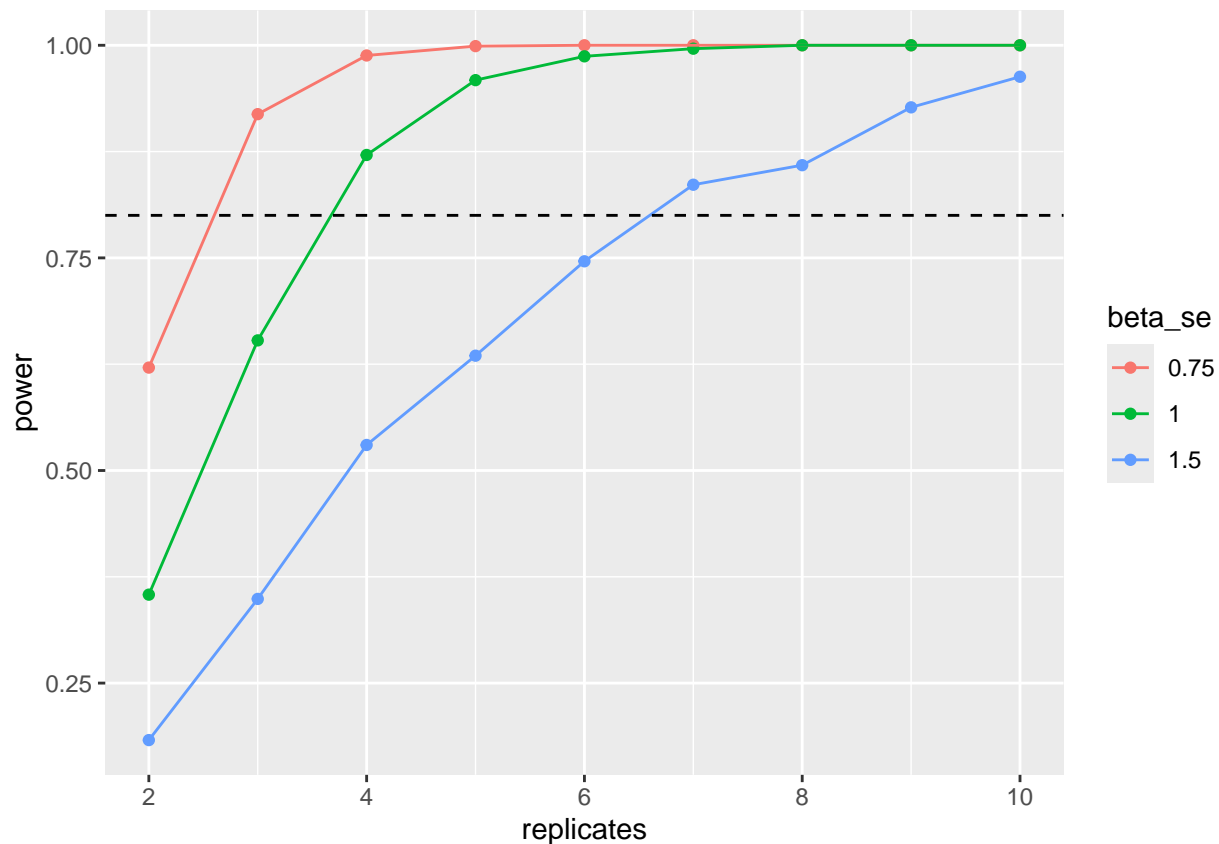
beta_se <- rep(0.75,8)
power1 <- NA
for(i in 1:length(replicates))
{
  power1[i] <- power_factorial_23(beta_mean, beta_se, replicates[i])
}

beta_se <- rep(1,8)
power2 <- NA
for(i in 1:length(replicates))
{
  power2[i] <- power_factorial_23(beta_mean, beta_se, replicates[i])
}

beta_se <- rep(1.5,8)
power3 <- NA
for(i in 1:length(replicates))
{
  power3[i] <- power_factorial_23(beta_mean, beta_se, replicates[i])
}

all_power <- data.frame(
  power = c(power1, power2, power3),
  beta_se = c(rep("0.75", length(power1)),
               rep("1", length(power2)),
               rep("1.5", length(power3))),
  replicates = rep(replicates, 3)
)
```

```
ggplot(data = all_power, mapping = aes(x = replicates, y = power, group = beta_se, color = beta_se)) +
  geom_point() + geom_line() + geom_hline(yintercept = 0.8, linetype = "dashed")
```



I tested for required sample size before conducting the experiment so I would know beforehand how many replicates would be needed. Due to my data being collected in feet, I thought that a standard error of around 1 seemed reasonable. To test “good” and “bad” values for standard error, I used 0.75 and 1.5. Based on the graph, for a “good” estimate of standard error, I could achieve a power above 0.8 with three replicates, giving a power of 0.92. For my baseline standard error estimate of 1 foot, I could achieve a power of 0.8 with four replicates, giving a power of 0.87. Finally, for the “bad” estimate of standard error, more replicates are required. Specifically, I would need 7 replicates to achieve a power of 0.8, giving an approximate power of 0.84. Based on the sample size calculations, I decided to collect 7 full replicates of my data, giving me 7 replicates * 8 flights per replicate = 56 total data points. The power for the experiment is predicted to be around 0.84, based on the sample size calculation graph and the number of replicates I chose to collect data on.

Data Collection and Statistical Summary

```
flightdatarep1 <- data.frame(
  Rear = c("0", "0", "0", "0", "1", "1", "1", "1"),
  Middle = c("0", "0", "1", "1", "0", "0", "1", "1"),
  Nose = c("0", "1", "0", "1", "0", "1", "0", "1"),
  Distance = c("5.08", "2.83", "5.00", "4.75", "2.42", "5.16", "2.58", "4.33")
)
```

```

flightdatarep2 <- data.frame(
  Rear = c("0", "0", "0", "0", "1", "1", "1", "1"),
  Middle = c("0", "0", "1", "1", "0", "0", "1", "1"),
  Nose = c("0", "1", "0", "1", "0", "1", "0", "1"),
  Distance = c("4.58", "2.42", "5.00", "2.83", "2.00", "2.75", "3.33", "3.50")
)

flightdatarep3 <- data.frame(
  Rear = c("0", "0", "0", "0", "1", "1", "1", "1"),
  Middle = c("0", "0", "1", "1", "0", "0", "1", "1"),
  Nose = c("0", "1", "0", "1", "0", "1", "0", "1"),
  Distance = c("3.25", "2.58", "5.42", "4.25", "4.33", "1.08", "4.50", "1.42")
)

flightdatarep4 <- data.frame(
  Rear = c("0", "0", "0", "0", "1", "1", "1", "1"),
  Middle = c("0", "0", "1", "1", "0", "0", "1", "1"),
  Nose = c("0", "1", "0", "1", "0", "1", "0", "1"),
  Distance = c("2.50", "3.92", "5.25", "2.08", "3.83", "1.50", "4.50", "3.08")
)

flightdatarep5 <- data.frame(
  Rear = c("0", "0", "0", "0", "1", "1", "1", "1"),
  Middle = c("0", "0", "1", "1", "0", "0", "1", "1"),
  Nose = c("0", "1", "0", "1", "0", "1", "0", "1"),
  Distance = c("7.08", "2.67", "5.67", "3.16", "1.50", "1.50", "3.08", "2.92")
)

flightdatarep6 <- data.frame(
  Rear = c("0", "0", "0", "0", "1", "1", "1", "1"),
  Middle = c("0", "0", "1", "1", "0", "0", "1", "1"),
  Nose = c("0", "1", "0", "1", "0", "1", "0", "1"),
  Distance = c("4.58", "2.17", "4.25", "5.42", "4.42", "3.67", "3.75", "2.42")
)

flightdatarep7 <- data.frame(
  Rear = c("0", "0", "0", "0", "1", "1", "1", "1"),
  Middle = c("0", "0", "1", "1", "0", "0", "1", "1"),
  Nose = c("0", "1", "0", "1", "0", "1", "0", "1"),
  Distance = c("3.16", "3.58", "4.58", "4.92", "6.25", "3.67", "3.42", "3.84")
)

flight <- rbind(flightdatarep1, flightdatarep2, flightdatarep3, flightdatarep4, flightdatarep5, flightdatarep6, flightdatarep7)
flight$Nose <- as.factor(flight$Nose)
flight$Middle <- as.factor(flight$Middle)
flight$Rear <- as.factor(flight$Rear)
flight$Distance <- as.numeric(flight$Distance)

sum_stats <- flight %>%
  group_by(Rear, Middle, Nose) %>%
  summarize(
    `Sample Mean` = round(mean(Distance), 3),
    `Sample Standard Deviation` = round(sd(Distance), 3),
  )

```

```

`Sample Size` = n()) %>%
  arrange(desc(`Sample Mean`))

kable(sum_stats, caption = "Summary Stats for Factor Experiment")

```

Table 1: Summary Stats for Factor Experiment

Rear	Middle	Nose	Sample Mean	Sample Standard Deviation	Sample Size
0	1	0	5.024	0.487	7
0	0	0	4.319	1.535	7
0	1	1	3.916	1.238	7
1	1	0	3.594	0.714	7
1	0	0	3.536	1.664	7
1	1	1	3.073	0.960	7
0	0	1	2.881	0.636	7
1	0	1	2.761	1.495	7

From the initial summary statistics for the data, we can see that the rear clip off, middle clip on, and nose clip off group has a much larger sample mean than the other groups with a particularly low standard deviation as well. The configuration with just the rear clip on has the highest sample standard deviation, which makes sense because loading all of the weight onto the back of the plane would cause it to become unstable. Finally, the configuration with rear clip and nose clip on has the lowest sample mean, but it is closely followed by just having the nose clip on. Both travel below 3 feet on average.

Visualizations

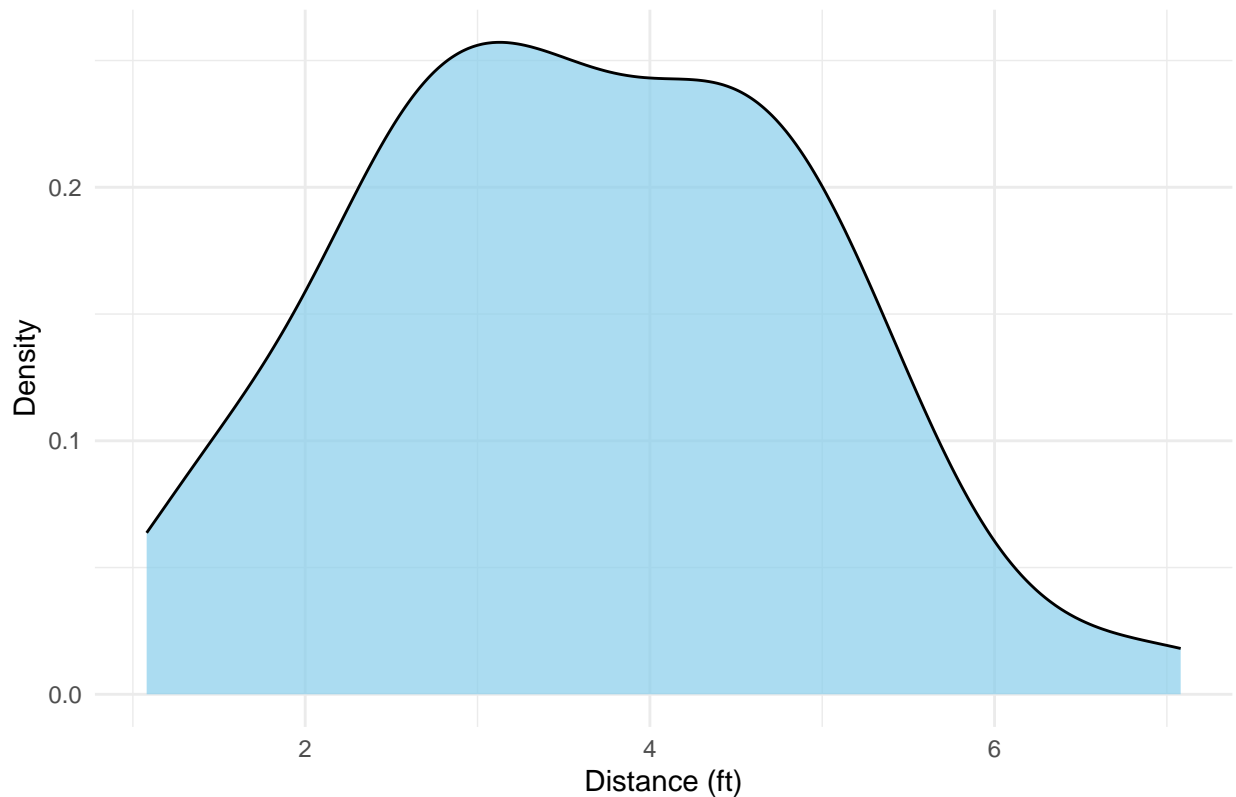
Density Plot

```

# Overall Density Plot
ggplot(flight, aes(x = Distance)) +
  geom_density(fill = "skyblue", alpha = 0.7) +
  labs(title = "Density Plot of Flight Distances", x = "Distance (ft)", y = "Density") +
  theme_minimal()

```


Density Plot of Flight Distances

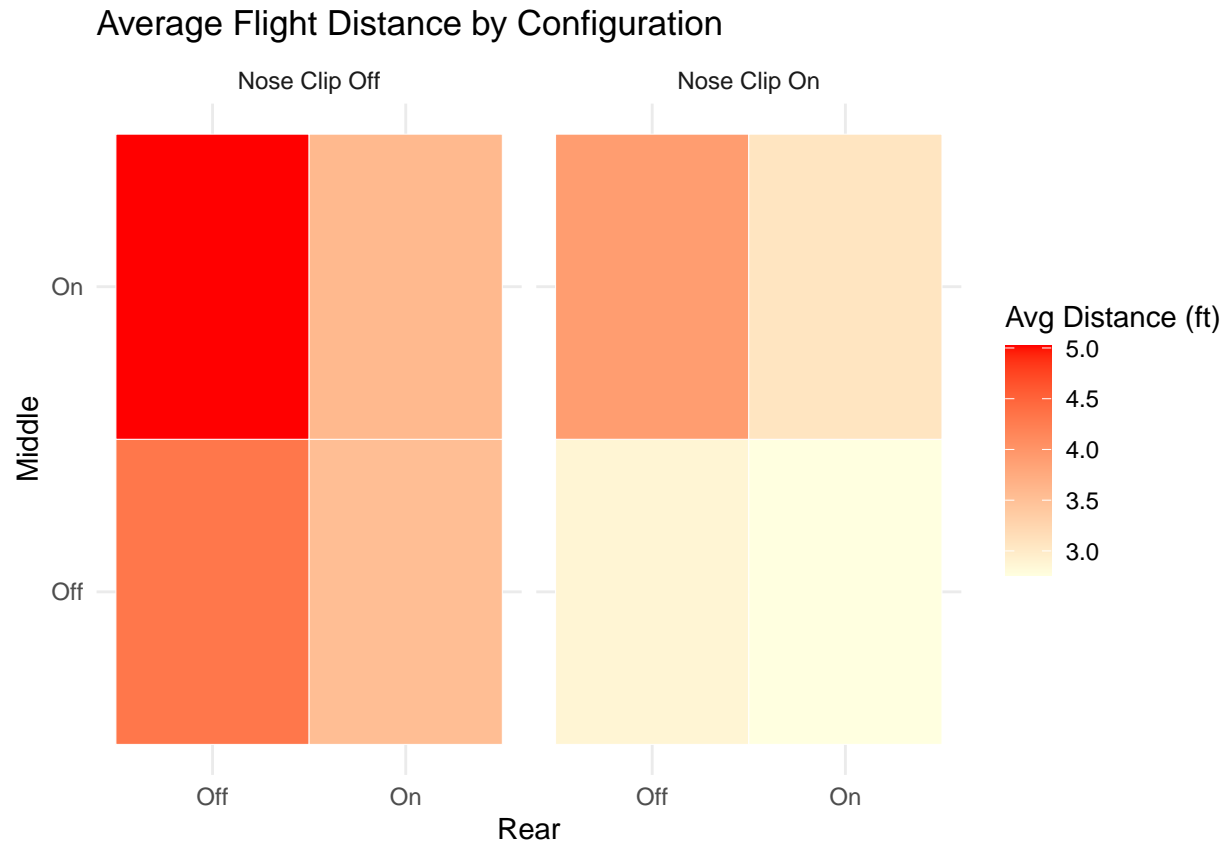


We can see from the density plot of the flight distances that the data seems to be relatively normal distributed, but more of the data is clustered to the left side of the graph. In addition, there seems to be a tail to the right side, creating a right skew. The normality assumption of the data will be tested in the **Model Checking** section of this report.

Heatmap

```
# Summarize average distance by configuration
heatmap_data <- flight %>%
  group_by(Rear, Middle, Nose) %>%
  summarize(Avg_Distance = mean(Distance)) %>%
  ungroup()

ggplot(heatmap_data, aes(x = Rear, y = Middle, fill = Avg_Distance)) +
  geom_tile(color = "white") +
  facet_wrap(~ Nose, labeller = labeller(
    Nose = c("0" = "Nose Clip Off", "1" = "Nose Clip On")
  )) +
  scale_fill_gradient(low = "lightyellow", high = "red") +
  labs(title = "Average Flight Distance by Configuration", x = "Rear", y = "Middle", fill = "Avg Distance") +
  scale_x_discrete(labels = c("Off", "On")) +
  scale_y_discrete(labels = c("Off", "On")) +
  scale_color_discrete(labels = c("Off", "On")) +
  theme_minimal()
```

From the heatmap, we can see the average flight distance for all eight configurations used in this experiment. Based on the graph, the average distance when the rear clip is off, the nose clip is off, and the middle clip is on seems to be the highest among all the groups. On the other hand, the average distance when the rear clip is on, the nose clip is on, and the middle clip is off, which is the opposite configuration of the group that produced the highest average distance, seems to be the lowest. In addition, it seems that when the nose clip is off, the average distance is higher.

Boxplot Comparisons of Distance

```
# Boxplot for Nose
plot1 <- ggplot(flight, aes(x = Nose, y = Distance)) +
  geom_boxplot(fill = "skyblue") +
  labs(title = "Distance (Nose Clip)", x = "Nose", y = "Distance (ft)") +
  scale_x_discrete(labels = c("Off", "On")) +
  theme_minimal()

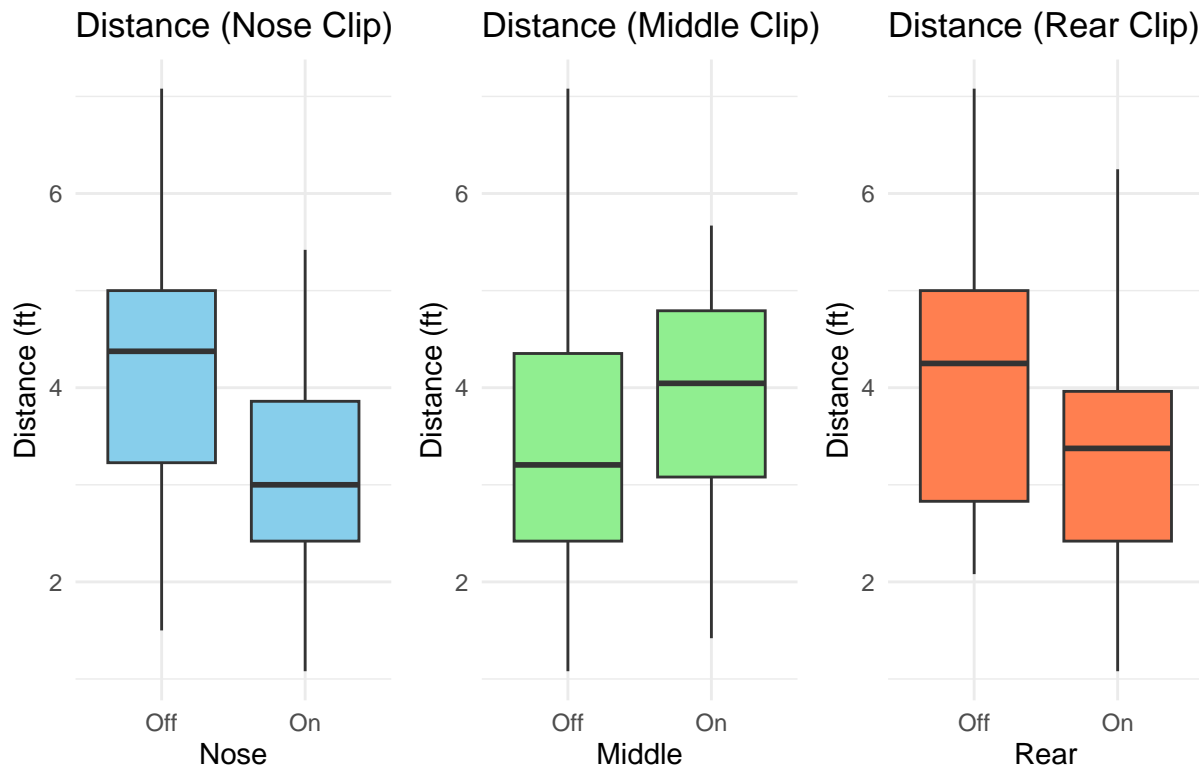
# Boxplot for Middle
plot2 <- ggplot(flight, aes(x = Middle, y = Distance)) +
  geom_boxplot(fill = "lightgreen") +
  labs(title = "Distance (Middle Clip)", x = "Middle", y = "Distance (ft)") +
  scale_x_discrete(labels = c("Off", "On")) +
  theme_minimal()

# Boxplot for Rear
plot3 <- ggplot(flight, aes(x = Rear, y = Distance)) +
```

```
geom_boxplot(fill = "coral") +
labs(title = "Distance (Rear Clip)", x = "Rear", y = "Distance (ft)") +
scale_x_discrete(labels = c("Off", "On")) +
theme_minimal()

(plot1 | plot2 | plot3) +
plot_annotation(title = "Comparison of Flight Distance in Feet by Clips Used")
```

Comparison of Flight Distance in Feet by Clips Used



When comparing the flight distance for each clips on vs off, we can see that the means have noticeable differences for each of the plots. For the boxplot comparison of having the nose clip on or off, we can see that the mean distance traveled when the nose clip is off seems to be higher by about 1.5 feet compared to when the nose clip is on. For the boxplot comparison of having the middle clip on or off, we can see that the mean distance traveled when the middle clip is off seems to be lower by about 1 foot compared to when the middle clip is on. For the boxplot comparison of having the rear clip on or off, we can see that the mean distance traveled when the rear clip is off seems to be higher by about 1 foot compared to when the rear clip is on.

Facet Grid

```
theme_update(text = element_text(size = 13))
ggplot(data = flight, mapping = aes(x = Rear, y = Distance, color = Middle, shape = Nose)) + geom_jitter()
  facet_grid(Middle ~ Nose,
             labeller = labeller(
               Middle = c("0" = "Middle Clip Off", "1" = "Middle Clip On"),
```

```

    Nose = c("0" = "Nose Clip Off", "1" = "Nose Clip On")
  )) +
  ggtitle("Flight Distance Based on Attached Clips") +
  scale_x_discrete(name = "Rear Clip", labels = c("Rear Clip Off", "Rear Clip On")) +
  ylab("Flight Distance (ft)")

```



The facet grid accounts for all three factors and allows us to visualize all of the data separated into their respective groups. We can clearly see that having the rear clip on decreases the average flight distance compared to having it off no matter the configuration of the other clips. In addition, we can see that the rear clip off, middle clip off, nose clip configuration has very low standard deviation, but adding the rear clip destabilizes the plane and creates a larger spread of values. Finally, there is no visual confirmation based on this graph that adding the middle clip creates any significant difference.

Statistical Analysis

```

# Initial Interaction Model
interactionmodel <- lm(Distance ~ Nose * Middle * Rear, data = flight)

output <- signif(summary(interactionmodel)$coefficients, 4)
output[,] <- as.character(output[,])
overall_p1 <- signif(pf(summary(interactionmodel)$fstatistic[1],
  summary(interactionmodel)$fstatistic[2],
  summary(interactionmodel)$fstatistic[3],
  lower.tail = FALSE), 4)

```

```

r_squared <- signif(summary(interactionmodel)$r.squared, 4)
adj_r_squared1 <- signif(summary(interactionmodel)$adj.r.squared, 4)
output <- rbind(output, "Overall P-Value" = c("", "", "", overall_p1))
output <- rbind(output, "R-Squared" = c("", "", "", r_squared))
output <- rbind(output, "Adjusted R-Squared" = c("", "", "", adj_r_squared1))
knitr::kable(output, caption = "Interaction Model of Flight Distances by Clips Used")

```

Table 2: Interaction Model of Flight Distances by Clips Used

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.319	0.4425	9.759	5.596e-13
Nose1	-1.437	0.6258	-2.296	0.02606
Middle1	0.7057	0.6258	1.128	0.2651
Rear1	-0.7829	0.6258	-1.251	0.217
Nose1:Middle1	0.3286	0.885	0.3712	0.7121
Nose1:Rear1	0.6629	0.885	0.749	0.4575
Middle1:Rear1	-0.6471	0.885	-0.7312	0.4682
Nose1:Middle1:Rear1	-0.07571	1.252	-0.06049	0.952
Overall P-Value				0.01041
R-Squared				0.3055
Adjusted R-Squared				0.2042

We can see that the overall model is statistically significant at an $\alpha = 0.05$ level of significance because $0.01 < 0.05$.

The Bonferroni-corrected α level is $\frac{0.05}{7} = 0.0071$. Thus at this adjusted significance level, none of the individual effects are actually significant. However, we can see that the main effect for having the Nose clip on vs off has a very low p-value of 0.025, which would pass the original α significance level of 0.05. We know that the Bonferroni adjustment is known to be conservative, causing an increase in Type II error rate.

Two appropriate courses of action are to run the experiment again but with a larger sample size, or to remove all of the interaction terms from the model. Because I already collected the data based on the initial sample size power calculation, I will try rerunning the model without the interaction terms instead.

Improved Model

```

# Linear Model
flightmodel <- lm(Distance ~ Nose + Middle + Rear, data = flight)

output <- signif(summary(flightmodel)$coefficients, 4)
output[,] <- as.character(output[,])
overall_p2 <- signif(pf(summary(flightmodel)$fstatistic[1],
                        summary(flightmodel)$fstatistic[2],
                        summary(flightmodel)$fstatistic[3],
                        lower.tail = FALSE), 4)
r_squared <- signif(summary(flightmodel)$r.squared, 4)
adj_r_squared2 <- signif(summary(flightmodel)$adj.r.squared, 4)
output <- rbind(output, "Overall P-Value" = c("", "", "", overall_p2))
output <- rbind(output, "R-Squared" = c("", "", "", r_squared))
output <- rbind(output, "Adjusted R-Squared" = c("", "", "", adj_r_squared2))
knitr::kable(output, caption = "Linear Model of Flight Distances by Clips Used")

```

Table 3: Linear Model of Flight Distances by Clips Used

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.251	0.3081	13.8	5.132e-19
Nose1	-0.9604	0.3081	-3.117	0.002974
Middle1	0.5275	0.3081	1.712	0.09284
Rear1	-0.7939	0.3081	-2.577	0.01285
Overall P-Value				0.0008718
R-Squared				0.2706
Adjusted R-Squared				0.2285

We can see that this new model without the interaction terms is still statistically significant at an $\alpha = 0.05$ level of significance because $0.001 < 0.05$.

The new Bonferonni-corrected α level is $\frac{0.05}{3} = 0.017$. Thus at this new adjusted significance level, the main effect for having the Nose clip on vs off is now significant because $0.003 < 0.017$.

We can also see that the main effect for having the Rear clip on vs off is also now significant at the new Bonferonni-corrected α level because $0.013 < 0.017$.

Overall, this new model without the interaction terms is a much better overall model than the full interaction model. The adjusted R-squared of 0.228 is higher than the previous model's adjusted R-squared of 0.204, meaning that more of the variation in the data is explained by the new model after adjusting for the amount of predictors present. In addition, the new p-value of 0.001 is less than the previous p-value of 0.01, meaning that the model is more likely to produce statistically significant results and is less likely to cause a Type II error. Thus, the following model checks will be performed on the simpler model rather than the full interaction model.

Additional ANOVA Comparison Between Models

```
model11 <- lm(Distance ~ Nose * Middle * Rear, data = flight)
model12 <- lm(Distance ~ Nose + Middle + Rear, data = flight)

anova_df <- as.data.frame(anova(model11, model12))
anova_df <- signif(anova_df, 4)
rownames(anova_df) <- c("Linear Model", "Interaction Model")

kable(anova_df, caption = "ANOVA Comparison Between Models")
```

Table 4: ANOVA Comparison Between Models

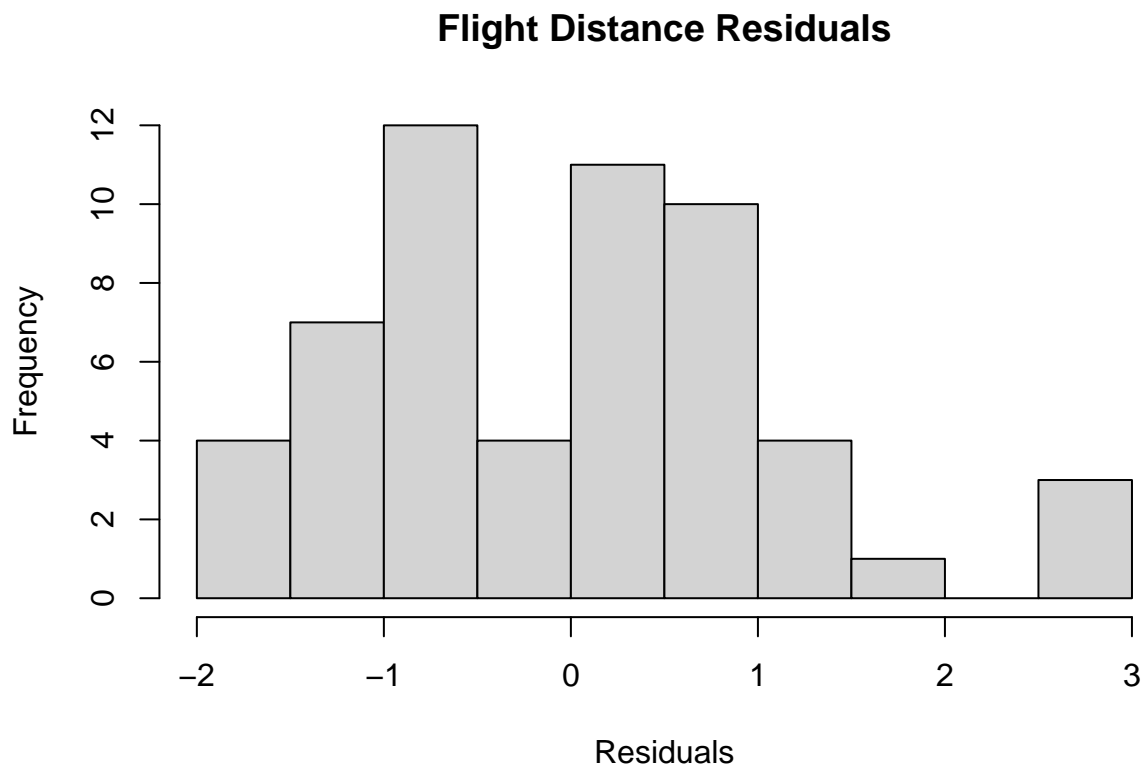
	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
Linear Model	48	65.80	NA	NA	NA	NA
Interaction Model	52	69.11	-4	-3.31	0.6037	0.6618

To fully make sure that the interaction should not be included, we can see from this ANOVA comparison that the p-value for the interaction effect by itself of 0.662 is not statistically significant at any Bonferonni-adjusted significance level. Thus, it is ok to move forward without the interaction term.

Model Checking

Normality of Residuals

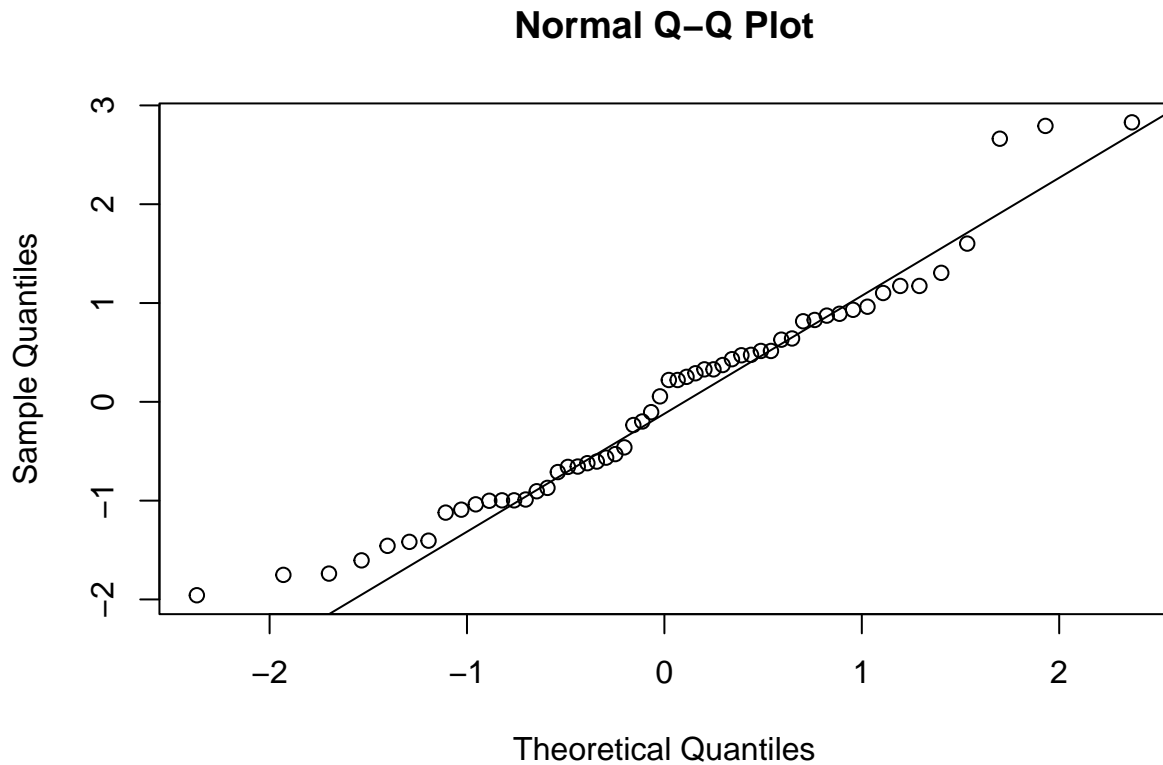
```
hist(flightmodel$residuals, xlab="Residuals", main="Flight Distance Residuals")
```



The flight distance residuals seem to be normally distributed, with no extreme skewed values or other issues.

Normality Assumption

```
qqnorm(flightmodel$residuals)  
qqline(flightmodel$residuals)
```



The normal Q-Q plot does show some deviation off the line near the higher quantiles. However, the rest of the values seem to follow the trend line closely. Thus, a Shapiro Test is needed to see if the normality assumption was violated.

Shapiro Test

```
shapiro_flight <- shapiro.test(flightmodel$residuals)

shapiro_table <- data.frame(
  Statistic = round(shapiro_flight$statistic, 3),
  P_Value = round(shapiro_flight$p.value, 3)
)

kable(shapiro_table, caption = "Shapiro-Wilk Test for Normality of Residuals")
```

Table 5: Shapiro-Wilk Test for Normality of Residuals

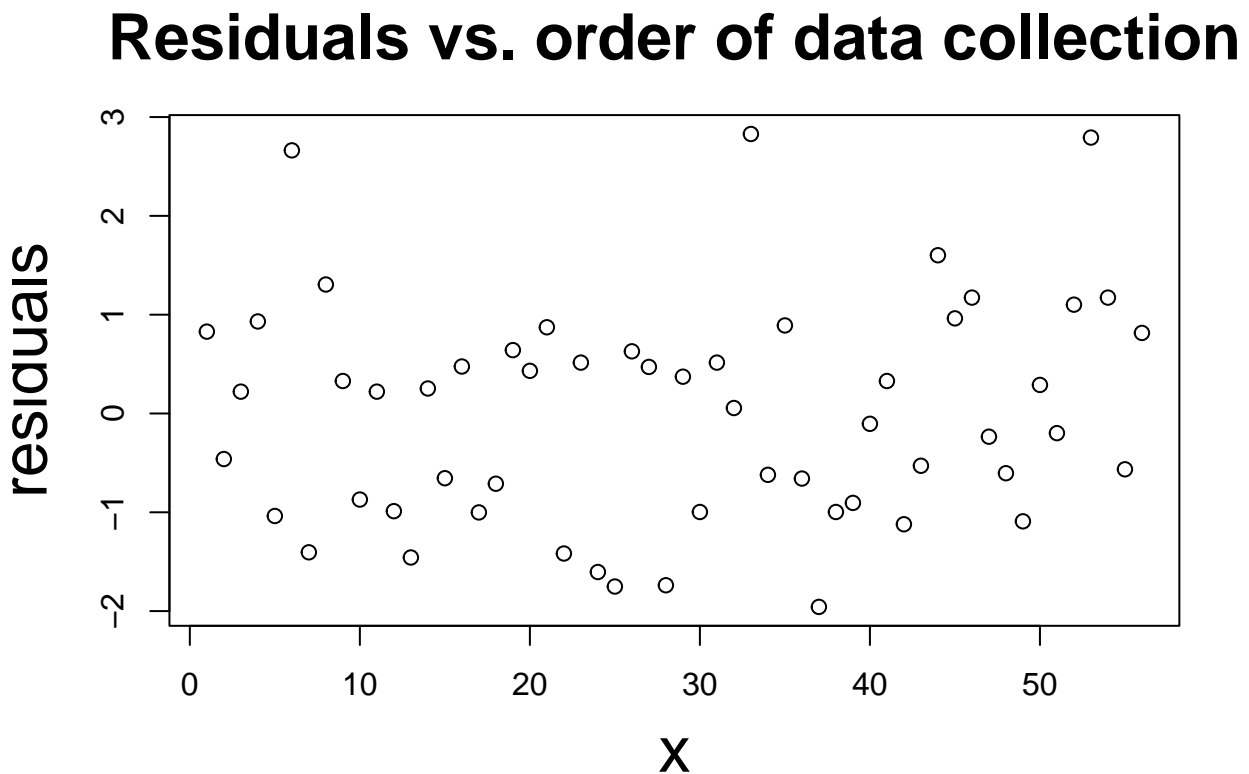
	Statistic	P_Value
W	0.96	0.059

The Shapiro-Test checks whether the residuals are normally distributed. At an $\alpha = 0.05$ level of significance, we FAIL to reject the null hypothesis that the residuals are normally distributed and conclude that the residuals DO follow a normal distribution. In other words, the normality assumption is valid for this model.

However, the p-value is dangerously close to the α significance level, so this model may not be reliable without a larger sample size.

Structure to the Data

```
x <- 1:length(flightmodel$residuals)
plot(flightmodel$residuals ~ x, ylab="residuals", cex.lab=2,
main="Residuals vs. order of data collection", cex.main=2)
```



There does not seem to be any concern to the structure of the data. In addition, I know that I randomized the order of data collection for each replicate, so this should not be a problem. Either way, we can investigate this further with the Durbin-Watson test for autocorrelation.

Durbin-Watson Test

```
dw_flight <- dwtest(flightmodel)

dw_table <- data.frame(
  Statistic = round(dw_flight$statistic, 3),
  P_Value = round(dw_flight$p.value, 3)
)

kable(dw_table, caption = "Durbin-Watson Test for Autocorrelation")
```

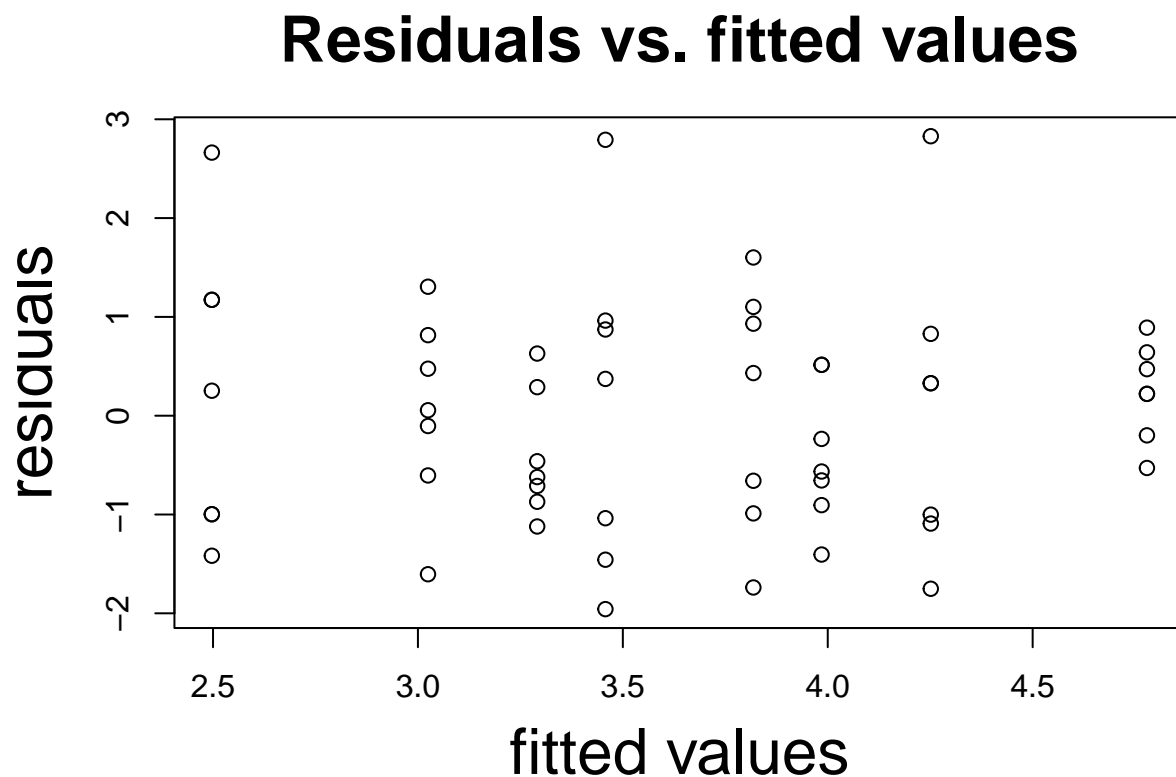
Table 6: Durbin-Watson Test for Autocorrelation

	Statistic	P_Value
DW	2.075	0.63

The Durbin-Watson test for autocorrelation checks for a trend in the order of the data collection. At an $\alpha = 0.05$ level of significance, we FAIL to reject the null hypothesis that there is no autocorrelation in the data. Thus we can definitively conclude that there is no structure to the data collection, just like we observed with the residuals vs order of data collection graph.

Equality of Variance

```
plot(flightmodel$residuals ~ flightmodel$fitted.values,
     xlab="fitted values", ylab="residuals", cex.lab=2,
     main="Residuals vs. fitted values", cex.main=2)
```



There is no large concern about inequality of variances across fitted values based on the graph of residuals vs fitted values. We can investigate this further with the Breusch-Pagan test for homoscedasticity.

Breusch-Pagan Test

```
bp_flight <- bptest(flightmodel)

bp_table <- data.frame(
  Statistic = round(bp_flight$statistic, 3),
  P_Value = round(bp_flight$p.value, 3)
)

kable(bp_table, caption = "Breusch-Pagan Test for Homoscedasticity")
```

Table 7: Breusch-Pagan Test for Homoscedasticity

	Statistic	P_Value
BP	5.285	0.152

The Breusch-Pagan test checks for heteroscedasticity with H_0 that the residuals have constant variance, or the assumption of homoscedasticity. If the p-value is small (ex. < 0.05), we would reject the null hypothesis, indicating heteroscedasticity. For our data, at an $\alpha = 0.05$ level of significance, we FAIL to reject the null hypothesis that the residuals have constant variance. Thus we can definitively conclude that there is homoscedasticity in the data, meaning that there is no large concern about inequality of variances across fitted values.

Model Evaluation Conclusion

Because the normality assumption is valid, there is no autocorrelation or structure to the data, and the model shows homoscedasticity, we can conclude that the model is valid. Thus, I did not run a permutation test because the assumptions were met and so the ANOVA/linear model will perform better.

Discussion

Restatement and Summary of Results

From the results section of this report, we found that the created linear model for the effect of three factors (having a clip on the Rear, Middle, and Nose of the plane) showed statistically significant results at an $\alpha = 0.05$ level of significance due to the overall model p-value of 0.001. At the Bonferonni-adjusted α level of $\frac{0.05}{3} = 0.017$ for this model, we also found that the main effect for having the Nose clip on vs off was significant because $0.003 < 0.017$, and that the main effect for having the Rear clip on vs off was also significant because $0.013 < 0.017$. Finally, we saw that all of the ANOVA assumptions were satisfied, meaning that this model could produce reliable results. However, keep in mind that the Shapiro-Wilk test caused a p-value that was very close to the cutoff significance level, but still was not enough to cause concern. For future tests, a larger sample size would be required to make sure that the normality assumption was in fact valid.

In conclusion, we found that the rear clip off, middle clip on, and nose clip off group had a much larger sample mean than the other groups with a particularly low standard deviation. The configuration with just the rear clip on had the highest sample standard deviation, which makes sense because loading all of the weight onto the back of the plane would cause it to become unstable. Finally, the configuration with rear clip and nose clip on had the lowest sample mean, but it is closely followed by just having the nose clip on. Both travel below 3 feet on average. Along with these individual configuration results, we found that the mean distance traveled when the nose clip or rear clip were off was higher than when they were on, while the mean distance traveled when the middle clip was off was lower than when it was on. We can conclude

based on these results that having only the middle clip on is the easiest way to increase a paper airplane's flight distance when compared to the control configuration. We found that having the middle clip attached stabilized the paper airplane and allowed it to fly a further distance in a straight line on average.

Broader Interest and Implications

As discussed previously, the design elements of paper airplanes can be used to generalize to larger aircraft. One example of the similarities between paper airplanes and some of their large scale counterparts is that they share a low Reynolds number. The Reynolds number, otherwise known as the scale effect, indicates the “relative importance of [fluid momentum]... on the airfoil” (Lissaman 1983). It is known as the scale effect because airfoils that have advantages for birds or insects may not be as effective when scaled up to a full size airplane due to their Reynolds number ranges being different. Based on this knowledge, an ideal way to scale up an airfoil design to airplane size would be to find a simple design that also inhabits a similar Reynolds number range. In a study conducted by Jörg U Schlüter in 2014, the author found that paper airplanes indeed occupy the exact same Reynolds number range as Micro Air Vehicles, meaning that their designs can be scaled up and used to predict how MAVs will act in specific flight situations. In fact, one main finding from this study was that the “centerfold of the paper airplane has a modest influence on the lift of the aircraft” (Schlüter, 2014). This means that stabilizing the center crease of the plane and middle section of the plane allowed it to create more lift and thus fly further or more efficiently. In fact, I found the exact same result in my study, indicating that the result of this experiment may in fact be generalizable to vehicles larger than paper airplanes themselves.

Limitations, Issues, and Implications

A major limitation in my experiment was that I was not able to use a yard stick, so I had to settle for ruler measurements instead. This meant that every landing spot of the airplanes may have been measured slightly off by a couple inches or so. In addition, due to a lack of space, I was forced to flick the airplanes forward rather than throw them with full power. While this led to more controlled results and eliminated any concerns about wind, it may not be fully generalizable to other paper airplane experiments. I do not believe this is a concern, as paper airplane construction can vary wildly from person to person, so any individual recreating this experiment would want to construct their airplanes and experiment in a similar manner but not feel forced to recreate my throwing conditions and environment.

In addition, the initial sample may not have been a perfectly random sample of all possible paper airplane flights. Because I calculated the first sample with only one flight per combination, running the sample size power calculation again with any of my established replicates may produce significantly different results, causing the number of replicates needed to produce 80% power to go up or down significantly. Also included in this section was the assumption that the standard error of the flight distances was between 0.75 and 1.5 feet, but the actual population standard error could have been different. Thankfully this was not a concern, as I found slightly lower lower-bound standard errors and slightly higher upper-bound standard errors - the standard errors within each combination varied from 0.487 from 1.664.

Overall, the linear model only had an adjusted R^2 value of 0.228, which means that it explains less than a third of the variation in the data. This could be due to the fact that having certain clips on or off of the plane may not be the sole predictor of flight distance, especially when flight mechanics and the impact of wind can be so dynamic and tough to control for. Thus, it is not a major cause of concern, but it should be noted because this report is not intended to establish a causal link between the weight distribution of a plane and its flight distance.

References, Sorted by Order of Appearance

- Ng, Bing Feng, et al. "On the aerodynamics of paper airplanes." 27th AIAA applied aerodynamics conference. 2009.
- Ifju, P., et al. "Flexible-wing-based micro air vehicles." 40th AIAA aerospace sciences meeting & exhibit. 2002.
- Simons, Graham M. The Airbus A380: A history. Pen and Sword, 2014.
- Irwin, Douglas A., and Nina Pavcnik. "Airbus versus Boeing revisited: international competition in the aircraft market." *Journal of international economics* 64.2 (2004): 223-245.
- Lissaman, P. B. S. "Low-Reynolds-number airfoils." *Annual review of fluid mechanics* 15.1 (1983): 223-239.
- Schlueter, Joerg U. "Aerodynamic study of the dart paper airplane for micro air vehicle application." *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering* 228.4 (2014): 567-576.