Exploring the Effect of Paperclip Placements on Paper Airplane Flight Distance

Kai Barker

2025-03-12

```r
## import libraries here so we can cache later
library(readxl)
library(ggplot2)
library(dplyr)
library(knitr)
source("power_factorial_23.R")
library(readr)
library(writexl)
```

Introduction

As concerns about climate change only continue to intensify and global efforts to mitigate them struggle to keep pace, any methods of reducing emissions could make a significant difference. According to Our World in Data, aviation contributes to 2.5% of worldwide emissions [8]. While this may seem like a relatively small number, minor improvements in this sector could meaningfully contribute to lowering overall emissions. Furthermore, optimizing airplane emissions, such as limiting fuel consumption, could provide valuable insights for the transportation industry as a whole.
One major contributor to fuel efficiency, and consequentially emissions, is a planes load factor or weight distribution. According to *Airline fuel usage and carbon emissions: Determining factors* [1], published in the *Journal of Air Transportation Management*, one of seven factors that contribute to an airlines total fuel usage, significant at the 1% level, is the airplanes weight or load factor. More specifically, according to emissions calculators from the International Civil Aviation Organization (ICAO) and Sabre Holdings, a database containing information about all flights, freight factor (weight) is a key feature when calculating emissions. While Sabre Holdings attributes freight factor to contributing 20% to emissions in wide body airplanes, 10% in narrow body and only 1% in regional jets, the ICAO places it between 47-88% depending on route and plane body [3]. In either regard, both calculations hold weight as an important contributor to airline emissions. This suggests that optimizing aircraft weight distributions could improve overall fuel efficiency and lower carbon emissions produced by aircraft.

The research question responsible for this study was: does weight placement have a significant impact on the flight distance of an airplane? For the purposes of this study, paper airplanes were chosen as they can provide insight into the aerodynamics and physics of flight, on a smaller scale compared to real aircraft. Additionally, a much more sophisticated study would be required to properly answer this question for real-world aircraft. Such a study would require time, money, and resources well beyond the scope of this research. This experiment allows us to visualize the effects of weight distribution in a more accessible and cost efficient manner. Although paper airplanes and real aircraft differ significantly, especially when considering stabilizers and shifting center of pressure [4, 6], studies such as Goodwill et al. (2022) show that paper airplanes still offer insight into how basic aerodynamic principles can translate to real world aircraft.

For example, while paper airplanes achieve balance though their unique thin, flat wings, real aircraft rely on a tail wing as a stabilizer [6]. While these findings do not directly translate to operational aircraft, such concepts help illustrate the principles of aerodynamics that relate to large scale aviation. Similar studies have explored factors influencing paper airplane flight distance, such as Puspita et al. (2020), which investigates the impact of paper weight, paper length, and the nose of the plane on its mileage (distance). The study found all three factors, including the weight of the plane to be significant to the planes mileage. Another study, *Aerodynamic performances of paper planes* [2] tested different paper airplane designs, including a glider body similar to the one used in our study, to evaluate their aerodynamic efficiency. The glider type produced promising results in terms of its lift coefficient, achieving the highest maximum lift coefficient of all tested designs. However, the glider type was limited when it came to drag, with the plane having at least a 14.3% larger drag coefficient at certain angle of attack regions when compared to the other airplane types. These studies reinforce the basis of the research question, highlighting the connection between paper airplanes and real aircraft, while also determining a significant correlation between weight distribution and carbon emissions.

By examining the impact of weight distribution on the flight distance of paper airplanes, this experiment provides a simplified but intuitive look into key principles of aerodynamics and physics that govern flight. Despite the differences between paper airplanes and real aircraft, studies have shown that factors such as weight distribution or overall airplane weight can play a role in the efficiency of both paper airplanes as well as real aircraft. With further studies finding weight to be a major contributor to aircraft carbon emissions, we can attempt to mitigate these consequences by optimizing aircraft design. Building on this, our experiment will dive into the fundamental concept of how different weight distributions affect the flight distance of paper airplanes.

Methods

Study Design

In this lab, a full factorial design was implemented with the factor of interest being the distance, in inches, that the paper airplane flew for each weight distribution. The factorial design saw a participant throwing a singular paper airplane with paperclips acting as the weights, placed on different parts of the plane according to randomized combinations of the placement factors. The paper airplane was built to specifications to resemble a glider type [5]. The placement factors were nose, middle and, rear, with each factor representing the paperclip (weight) was placed on the nose, middle and, rear of the plane respectively (**Figure 1 and 2**). Each placement factor had levels "yes" and "no", indicating the presence or absence of the paperclip in the described position. It is important to note that standard paperclips (1.3 inch) were used in this experiment. It should be considered as well that the absence of any paperclips "no, no, no", was also considered as a paperclip configuration. The throws were completed by a participant standing at the end of an outstretched measuring tape (**Figure 3**). The paperclips were then readjusted for each individual throw based on the randomized configurations, and the distance the airplane flew was recorded to the nearest quarter inch. The distance flew was determined on where the tip of the airplane landed on first contact with the ground, not the distance when it stopped sliding. Distances were record to the nearest quarter inch due to limitations with the process of observation - to be discussed at the end of this section.

```
knitr::include_graphics("plane1.jpeg")
```

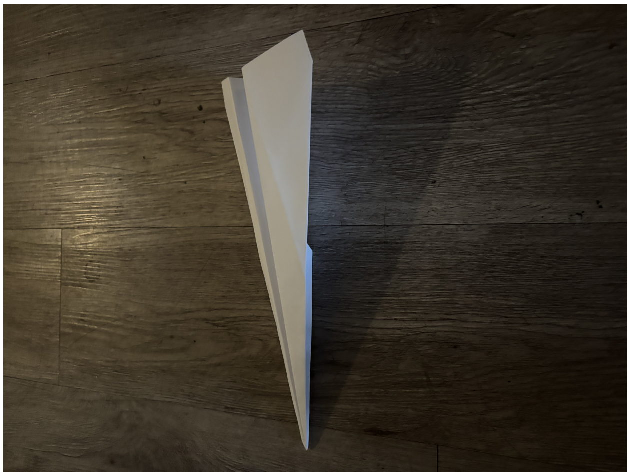

Figure1 : *Depicts the paper airplane as used in the experiment.*

```
knitr::include_graphics("plane2.jpeg")
```

Figure 2: *Depicts the paper airplane with only a middle placed paperclip.*

```
knitr::include_graphics("mehallway1.jpeg")
```

Figure 3: *Participant performing the experiment.*

Randomization and Sample Size

To determine where the weight was placed on the paper airplane (eg. weight on nose and middle while no weight on the rear), all possible combinations of weight placements were pooled together and appropriately sampled to populate the desired number of replicates. The randomization was performed in R, and written to a CSV file for optimal readability and data collection. Further details on this randomization code can be observed in the following "Results" section. The weight placements were randomized to mitigate bias and confounding variables in the experiment. Had the order of the paperclips not been randomized (say twenty throws in a row with no paperclips present), the participant throwing the planes could get used to the weight of the plane and be able to throw it farther purely off of repetition, rather than the factor of weight placement. Additionally, should the plane be damaged, such as the wings or nose getting bent after many repetitions, and with no randomization, the later paperclip combinations would be more influenced by any impact this could have on the planes flight distance. For this experiment a sample size of ten replicates was used, with each replicate containing eight throws for a total of eighty throws. Reasoning for the sample size and its possible benefits and limitations are to be discussed in "Results" and "Discussion".

Statistical Methods

We wanted to determine if different paperclip placements on a paper airplane have any significant impact on the distance the airplane will fly. To test this, we conducted a hypothesis test and fit a linear model to the experimental data. In our model, the response variable was the distance flown and the predictors were the different combinations of paperclip placements. The hypothesis test assessed whether at least one combination of paperclip placements or its interaction with another had an impact on the flight distance.

This method assumes that the residuals of the model are approximately normally distributed, that there is no underlying structure to the data, and that the variance of the residuals in constant across all values of predictors, meaning that the spread of data points in either direction does not drastically vary with different values of predictor variables. For clarification, assuming that there is no structure to the data assumes that there is no observable pattern to the data over time, that we are not able to observe a trend upward, downward, or otherwise.

Without running appropriate statistical tests to determine integrity of the study design, we can only make initial assumptions. The randomization of the paperclip placement should validate the structure of data assumption, as it will eliminate bias and confounding variables. Further, the relatively large (>30) sample size should lead to validity of the normality assumption due to the Central Limit Theorem. Equality of variances is more challenging to determine without proper testing but considering no strong evidence against it, we will assume it holds as well.

Limitations and Technical Errors

Several factors may have impacted the legitimacy of the findings of this experiment. One limitation of the experiment was the measurement of flight distance. Distance was measured to the nearest quarter inch based solely on visual observation. Since the same individual participating in the experiment was also responsible for recording the distances, this was necessary. Still, slight measurement errors may have occurred due to the difficulty of accurately identifying the landing point. Another issue was a limitation of space. Due to weather conditions outdoors, the experiment took place in a hallway with poor lighting and visibility. While the experiment location provided plenty of distance for the plane to fly, the hallways narrow structure lead to multiple collisions which resulted in voided throws. Furthermore, fatigue of the paper airplane was observed during the trials as a result of the narrow space and from the repetitive impact with the ground. After a multitude of throws, damage to the airplanes structural integrity was observable in both the wings and the nose of the aircraft. While not proven, this could have had some impact on the distance flown, regardless of the paperclip configuration.

Results

Pilot Study:
In order to determine the appropriate number of replicates or sample for the experiment, a pilot study was conducted in order to create an estimate for the necessary sample size. Each replicate contains eight throws, indicative of the eight paperclip combinations. To properly perform the pilot study, the weight placement for the throws needed to be randomized. To do this, a random sample was taken from all the possible weight placement combinations.

```
set.seed(03122025)
num_replicates <- 2
sample_per_rep <- 8

pilot_data <- data.frame(replicate = NA,
                         nose = NA,
                         middle = NA,
                         rear = NA)
pilot_data <- na.omit(pilot_data)

pilot_options <- c("no,no,no", "no,no,yes", "no,yes,no", "no,yes,yes", "yes,no,no", "yes,no,yes", "yes,y

for (i in 1:num_replicates) {
  rep = i
  for(i in 1:sample_per_rep) {
    #had to look up the unlist function, not sure if theres an        easier way to do this
    pilot_sample <- sample(pilot_options, 1, replace=TRUE)
    split_values <- unlist(strsplit(pilot_sample, ","))
    new_row <- data.frame(replicate = rep,
                          nose = split_values[1],
                          middle = split_values[2],
                          rear = split_values[3])
    pilot_data <- rbind(pilot_data, new_row)
  }
}

#only needed to get run 1 time
#added "replicate" # in the excel file
#write_csv(pilot_data, "pilot_study.csv")
```

```
## input the data
pilot_study_data <- read.csv("pilot_study.csv", stringsAsFactors=TRUE)

pilot_study_data$nose <- factor(pilot_study_data$nose)
pilot_study_data$middle <- factor(pilot_study_data$middle)
pilot_study_data$rear <- factor(pilot_study_data$rear)

## run a lm on the pilot data to obtain the beta mean, beta se
pilot_lm <- lm(distance ~ nose + middle + rear + nose*middle + nose*rear + middle*rear + nose*middle*rea
pilot_summary <- summary(pilot_lm)$coefficients
kable(pilot_summary, caption="Pilot Study Summary Table")
```

Table 1: Pilot Study Summary Table

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	184.350	16.46607	11.1957477	0.0000036
noseyes	-28.350	30.80520	-0.9202991	0.3843274
middleyes	-57.850	40.33348	-1.4342924	0.1893987
rearyes	-77.350	40.33348	-1.9177618	0.0914306
noseyes:middleyes	53.100	60.50022	0.8776828	0.4057019
noseyes:rearyes	48.725	54.61179	0.8922067	0.3983234
middleyes:rearyes	78.975	60.50022	1.3053672	0.2280570
noseyes:middleyes:rearyes	-62.100	83.96083	-0.7396306	0.4806622

Table 1: *Summary from linear model fitted to pilot study data.*

This linear model fitted to the pilot study data was then used to create a sample size graph to determine the appropriate number of replicates, or samples, needed for sufficient statistical power. From the linear model, values for beta mean and beta standard error (beta_se) were obtained in order to plot the necessary sample size for our experiment.

```r
set.seed(03122025)
## create the sample size graph
beta_mean <- c(184.350, -28.350, -57.850, -77.350, 53.100, 48.725, 78.975, -62.100)
beta_se <- c(16.46607, 30.80520, 40.33348, 40.33348, 60.50022, 54.61179, 60.50022, 83.96083)

sd_beta_se <- sd(beta_se)
mean_beta_se <- mean(beta_se)
replicates <- 2:20

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

#deviate the standard error by the sd of the original beta_se
subtracted <- mean_beta_se - sd_beta_se
beta_se2 <- rep(subtracted, 8)
power2 <- NA
for (i in 1:length(replicates)){
power2[i] <- power_factorial_23(beta_mean, beta_se2, replicates[i])
}

#deviate the standard error by the sd of the original beta_se
added <- mean_beta_se + sd_beta_se
beta_se3 <- rep(added, 8)
power3 <- NA
for (i in 1:length(replicates)){
power3[i] <- power_factorial_23(beta_mean, beta_se3, replicates[i])
}


all_power <- data.frame(power = c(power2, power1, power3),
                        beta_se =  as.factor(c(rep("27.5", length(power2)), rep("pilot_se", length(powe
                        replicates = rep(replicates, 3))
```

```r
ggplot(data=all_power, mapping=aes(x=replicates, y=power, color=beta_se, group=beta_se)) +
  geom_point() +
  geom_line() +
  scale_color_discrete(name="Beta SE") +
  labs(x = "Replicates", y = "Statistical Power") +
  labs(title="Power Analysis for Determining Sample Size")
```

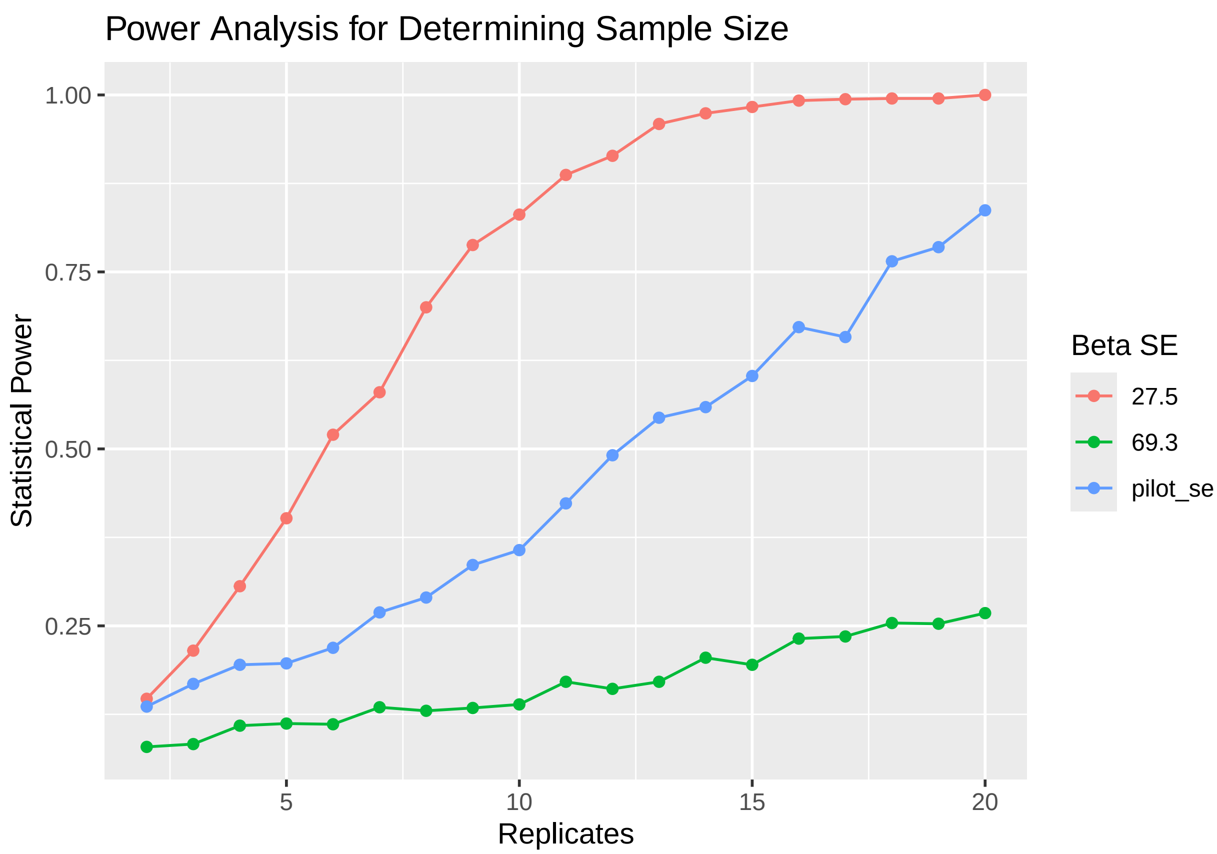

Figure 4: *Sample size graph plotting replicates against statistical power for three levels of standard error. The blue line is standard error as produced by the linear model fitted to the pilot study data. The red and green lines represent lower and upper limits for the standard error respectively.*

Three levels for the standard error of beta were used to calculate appropriate sample sizes. The middle level (blue) was a baseline standard error obtained from the linear model fitted to the pilot study data (**Figure 4**). The low standard error (red) was obtained by taking the mean of the pilot study standard error and subtracting its standard deviation, and the high standard error (green) was obtained by adding its standard deviation. Since standard deviation measures the variation of values around the mean, this method gives us a rough estimate of the range in which standard error data points are likely to fall based on the variation observed in the pilot study.
The number of replicates needed to achieve an appropriate statistical power of 80% is approximately twenty replicates assuming our standard error is exactly equal to that of the pilot study (**Figure 4**). If all our standard errors are approximately equal to the low standard error of (27.5464387), we would need around eight to ten replicates. At the high of (69.3313838), we would need over twenty replicates to just achieve a statistical power of 25%.

Randomization for the Experiment:
The same randomization procedure utilized in the pilot study was used to randomize the order of the weight

distribution for the experiment. After randomizing the paperclip placements, the data frame containing all trials was written to a CSV where the distance column was added and as trials were run, and distances were recorded.

```
set.seed(03132025)
study_num_replicates <- 10
study_sample_per_rep <- 8

study_random_df <- data.frame(replicate = NA,
                         nose = NA,
                         middle = NA,
                         rear = NA)
study_random_df <- na.omit(study_random_df)

weight_options <- c("no,no,no", "no,no,yes", "no,yes,no", "no,yes,yes", "yes,no,no", "yes,no,yes", "yes

for (i in 1:study_num_replicates) {
  rep = i
  for(i in 1:study_sample_per_rep) {
    #had to look up the unlist function, not sure if theres an        easier way to do this
    study_sample <- sample(weight_options, replace=TRUE)
    split_values <- unlist(strsplit(study_sample, ","))
    new_row <- data.frame(replicate = rep,
                          nose = split_values[1],
                          middle = split_values[2],
                          rear = split_values[3])
    study_random_df <- rbind(study_random_df, new_row)
  }
}

#only need to run this line one time
#write_csv(study_random_df, "airplane_study_data.csv")
```

After completing all trails, the CSV file was read back in for statistical analysis.

```
airplane <- read.csv("airplane_study_data.csv", stringsAsFactors=TRUE)
```

Summaries of Data from the Experiment:
After performing the experiment and collecting the data, we can analyze it through multiple methods of visualization. First, we can examine the general summary statistics for our experiment data. This includes analyzing the number of observations, mean and standard deviation for flight distance over all paperclip combinations.

```
## create a summary statistics table
sum_stats <- airplane %>%
  group_by(nose, middle, rear) %>%
  summarize(n=n(),
  mean=mean(distance),
  sd=sd(distance))
```

```
## 'summarise()' has grouped output by 'nose', 'middle'. You can override using
## the '.groups' argument.
```

```r
sum_matrix <- as.matrix(sum_stats)
colnames(sum_matrix) <- c("Nose", "Middle", "Rear", "n", "mean", "sd")
kable(sum_matrix,, caption="Summary Statistics for observations on Flight Distance")
```

Table 2: Summary Statistics for observations on Flight Distance

Nose	Middle	Rear	n	mean	sd
no	no	no	8	196.3750	18.79400
no	no	yes	7	117.9286	31.58808
no	yes	no	11	186.2500	18.20268
no	yes	yes	12	100.8542	33.94087
yes	no	no	13	170.6731	15.43278
yes	no	yes	9	139.0556	41.41912
yes	yes	no	10	160.9000	12.96909
yes	yes	yes	10	151.5000	28.12275

Table 2: *Observations (n), mean, and standard deviation for all paperclip configurations.*

We can observe that the largest mean of value of 196.3750 corresponds when no paperclips were present on the airplane, while the smallest mean value of 100.8542 occurred when paperclips were placed on the middle and rear of the plane (**Table 2**). Additionally the smallest standard deviation corresponds to the throws with paperclips on the nose and middle of the plane, and the largest standard deviation was observed when the plane had a paperclip on both the nose and rear. All eight groups had at least 7 observations with the least observations for only rear with 7 and the greatest for only nose with 13.

Additionally, we can visualize the data through a factored box plot and a jittered scatter plot. A jittered scatter plot slightly randomizes or moves data point positions to prevent multiple points from overlapping, improving visual analysis.

```r
## create a factored boxplot
theme_update(text = element_text(size = 12))

ggplot(airplane, aes(x=nose, y=distance, fill=middle)) +
geom_boxplot() +
facet_grid(cols = vars(rear),
labeller=labeller(rear = c("no" = "Rear No", "yes" = "Rear Yes"))) +
scale_fill_discrete(name="Middle", labels=c("No", "Yes")) +
labs(x = "Nose", y = "Distance (in)") +
ggtitle("Boxplot of Flight Distance Based on Paperclip Placement") +
scale_x_discrete(name="Nose", labels=c("No", "Yes"))
```

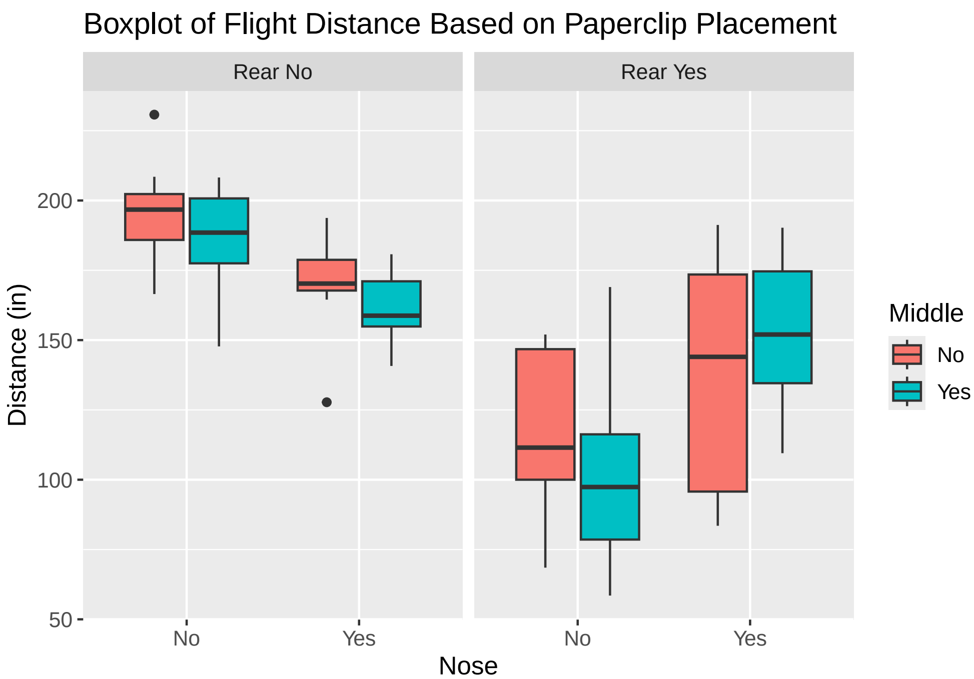

Figure 5: *Box plot with x-axis for the nose condition ("no", "yes"), y-axis for distance flown (inches), filled and colored by middle condition ("no", "yes"), and faceted by rear condition ("no", "yes").*

We can observe that when there is no rear placed paperclip, the spread of the flight distances are roughly similar, with the smallest spread correlating to a paperclip placed on the nose but not in the middle (**Figure 5**). The spread was the largest when the configuration was "No, Yes, No" (nose, middle, rear respectively). When a rear placed paperclip is present, the smallest spread corresponds to the "No, Yes, Yes" group of throws and the largest to the "Yes, No, Yes" group. The large spread indicates a high level or variability in the "No, Yes, No" and "Yes, No, Yes" throws, with a wider range of values, while the small spread indicates low variability, or relatively similar flight distances for the "Yes, No, No" and "No, Yes, Yes" groups. Further, the highest median flight distance of approximately 200 inches occurred when there was no paperclips in any position on the plane, and the lowest median flight distance of approximately 100 inches occurred in the "No, Yes, Yes" group. The medians reflect the tendency of the data, with half of the "No, No, No" throws landing at or beyond approximately 200 inches and half of the "No, Yes, Yes" throws landing at or below 100 inches.

The large spread indicates a high level or variability, with a wider range of values, while the small spread indicates low variability, or relatively similar flight distances. Further, the highest median flight distance of approximately 200 inches occurred when there was no paperclips in any position on the plane, and the lowest median flight distance of approximately 100 inches occurred in the "No, Yes, Yes" group (**Figure 5**). The medians reflect the tendency of the data, with half of the "No, No, No" throws landing at or beyond approximately 200 inches and half of the "No, Yes, Yes" throws landing at or below 100 inches.

```
theme_update(text = element_text(size = 12))

## create a jitter plot
ggplot(data=airplane, mapping=aes(x=nose, y=distance, color = middle)) +
geom_jitter(width=0.08, height=0) +
facet_grid(cols = vars(rear),
labeller=labeller(rear = c("no" = "Rear No", "yes" = "Rear Yes"))) +
scale_color_discrete(name="Middle", labels=c("No", "Yes")) +
labs(x = "Nose", y = "Distance (in)") +
```

```
ggtitle("Distance Flown based on Paperclip Placement") +
scale_x_discrete(name="Nose", labels=c("No", "Yes"))
```

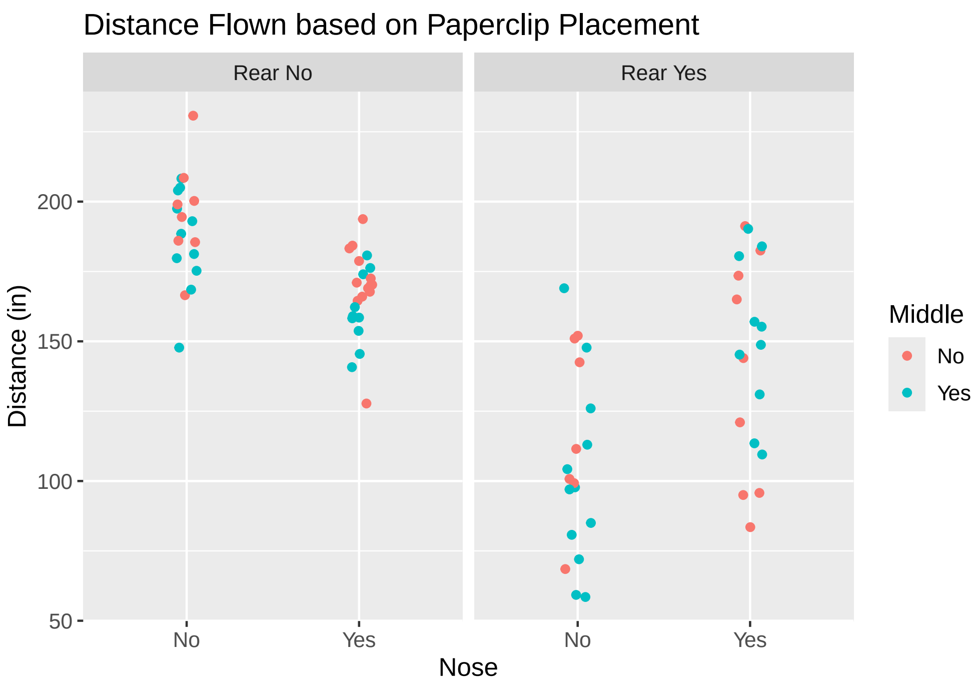

Figure 6: *Jittered scatter plot with x-axis for the nose condition ("no", "yes"), y-axis for distance flown (inches), colored by middle condition ("no", "yes"), and faceted by rear condition ("no", "yes").*

Additionally, all throws without a paperclip present on the nose or rear, regardless of the presence of a middle paperclip, seem to be tightly clustered between approximate flight distances of 150 and 225 inches (**Figure 6**). Throws with a paperclip on the nose but no paperclip on the rear, with either condition of middle, again appear tightly clustered between approximate flight distances of 125 and 200 inches. Considering both middle conditions once again and observing throws with a rear "Yes", those combined with a nose "No" appear to have a much larger spread than previously discussed throws with a rear "No". Throws in this category appear to be spread between approximate distances of 55 and 175 inches while throws with "Yes, No/Yes, Yes" appear spread between approximate distances of 80 and 195 inches. Similarities between points with either condition of middle suggest that it is more consistent given other conditions present in the paperclip configuration.

Model Checking

As discussed in the methods section, several assumptions were made in order to proceed with the with the statistical methods used in this experiment. The assumptions were: residuals of the model are approximately normally distributed, there is no underlying structure to the residuals, and that the variance of the residuals in constant across all values of predictors.

```
model1 <- lm(distance ~ nose*middle*rear, data=airplane)
## check normality
# histogram
hist(model1$residuals, xlab="Residuals", main="Paper Airplane Flight Residuals")
```

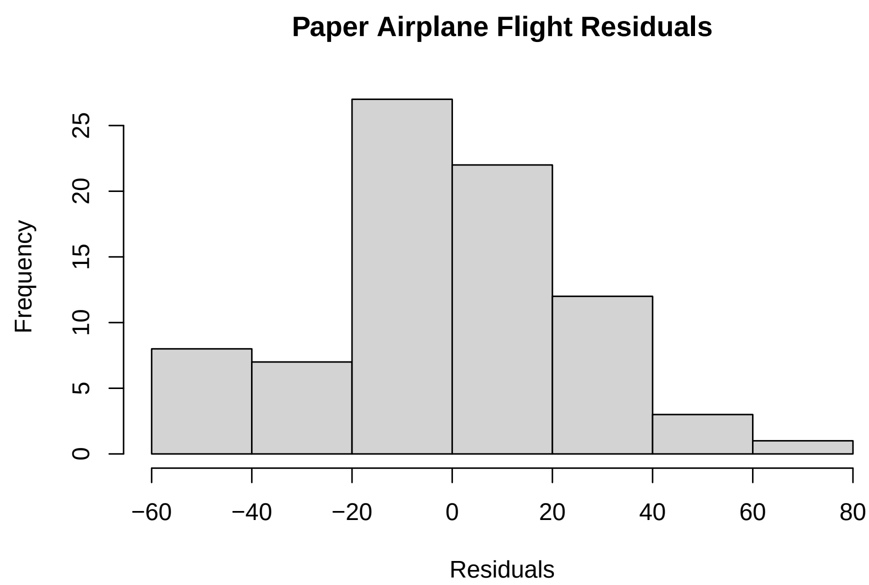

Figure 7: *Histogram plotting residuals against frequency.*

```
# qqplot
qqnorm(model1$residuals)
qqline(model1$residuals)
```

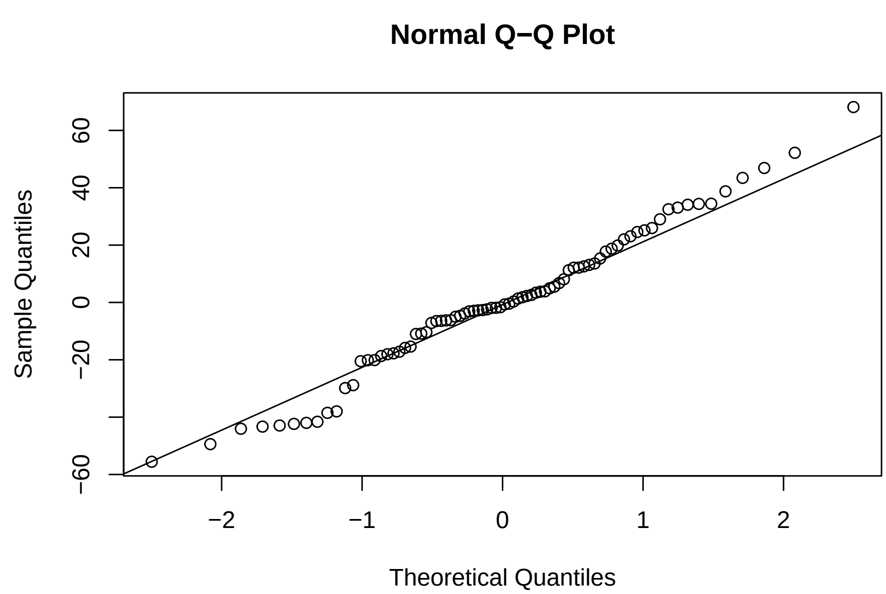

Figure 8: *Q-Q plot plotting residuals against a line of normality.*

```
# shapiro-wilk
shap_test <- shapiro.test(model1$residuals)
shap_p_val <- shap_test$p.value
shap_matrix <- matrix(c(shap_test$statistic, shap_test$p.value), ncol=2)
```

```r
colnames(shap_matrix) <- c("Data", "p-value")
knitr::kable(shap_matrix, caption="Shapiro-Wilk Test for Model Residuals")
```

Table 3: Shapiro-Wilk Test for Model Residuals

Data	p-value
0.9832963	0.38223

Table 3: *Summary of Shapiro-Wilk test on residuals from linear model fitted airplane flight distance data.*

The residuals appear to follow a rough bell curve, suggesting normality, where a larger sample size would provide more confidence (**Figure 7**). The residuals also appear normally distributed in the Q-Q plot, as most points closely follow the normal line (**Figure 8**). While the histogram and Q-Q plot suggest normality, the strongest evidence that the residuals are approximately normally distributed comes from the Shapiro-Wilk test, with a resulting p-value of 0.38223, indicating that at $\alpha = 0.05$, we fail to reject the null hypothesis of normality (**Table 3**). Thus there is not enough evidence that the residuals are not normally distributed. From all our checks on normality, we can conclude the residuals are at least approximately normally distributed.

```r
## check structure
x <- 1:length(model1$residuals)
plot(model1$residuals ~ x, ylab="residuals", cex.lab=1,
main="Residuals vs. order of data collection")
```

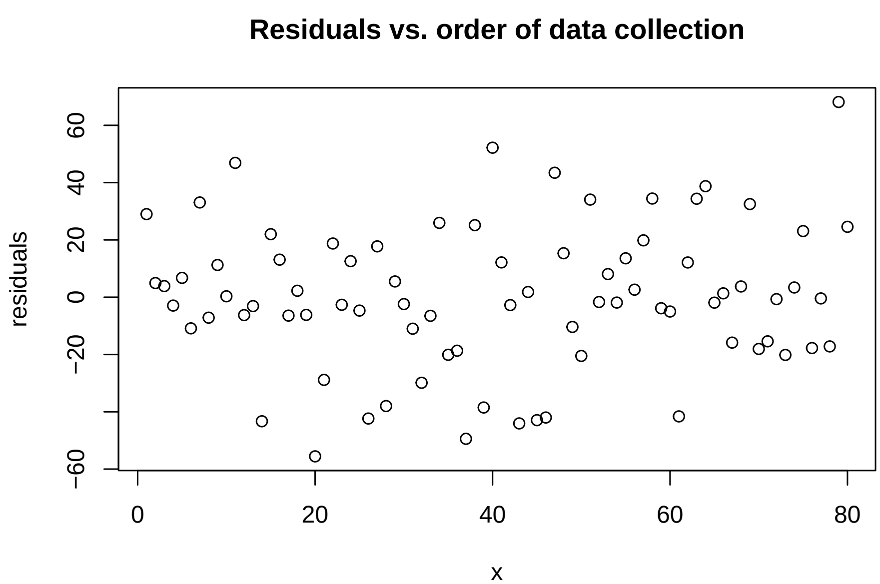

Figure 9: *Plot of residuals over time.*

The residuals also do not appear to have any observable structure over time (**Figure 9**). Since the data was randomized and then throws were completed one by one, the order of the indices in the data frame is the order the data was collected. The residuals are scattered randomly around zero, with no patterns appearing over the order of data collection. This supports the initial assumption that the residuals are independent.

```r
## check equal variance
plot(model1$residuals ~ model1$fitted.values,
xlab="fitted values", ylab="residuals", cex.lab=1,
main="Residuals vs. fitted values")
```

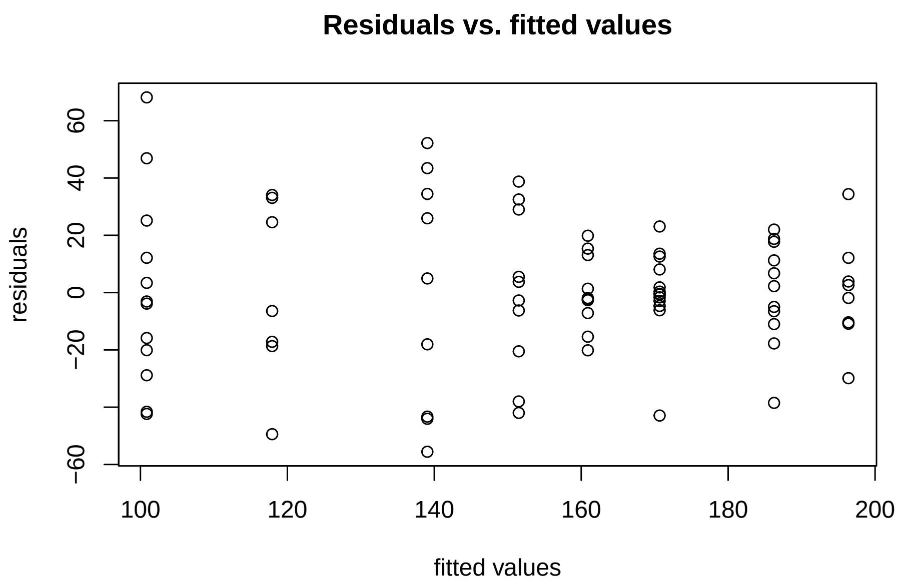

Figure 10: *Plot of fitted values against residuals.*

Further, the fitted values against residuals appear roughly constant over all values, providing no evidence against the last of our model assumptions of constant variance. While some values deviate slightly more than others, all fitted values have approximately similar spreads of residuals (**Figure 10**).

Considering model assumptions appear to be satisfied, we can confidently fit a linear model on our experiment data and proceed with conducting a hypothesis test. Had the data not met the assumptions of our statistical method, a permutation test could serve as an alternative method, as shuffling and sampling from a large number of permutations should mitigate the assumption violations.

Hypothesis Test

We can proceed with a hypothesis on the fitted linear model. The hypothesis test will asses whether the placement or presence of paperclips significantly impact the flight distance of the paper airplane. The null hypothesis (H_0) is that the paperclip placement has no effect on flight distance, while the alternative hypothesis (H_A) states that at least one paperclip position or interaction term has a significant effect on the flight distance. The null hypothesis asserts that all predictors ($\beta_1, \beta_2, \beta_3$) and interaction terms ($\beta_{12}, \beta_{13}, \beta_{23}, \beta_{123}$) in the model are equal to zero, while the alternative asserts that at least one predictor or interaction is not equal to zero.

To test this hypothesis, we will examine p-values from the linear model at significance level $\alpha = 0.05$. If an individual predictor or interactions p-value is less than our significant level, we will have sufficient evidence to reject the null hypothesis in favor of the alternative, otherwise we will fail to reject the null hypothesis.

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_{12} = \beta_{13} = \beta_{23} = \beta_{123} = 0$$

$$H_A : At\ least\ one\ \beta \neq 0$$

```r
model1 <- lm(distance ~ nose*middle*rear, data=airplane)
model_summary <- summary(model1)$coefficients
kable(model_summary, caption="Summary of Linear Model for Paper Airplane Experiment")
```

Table 4: Summary of Linear Model for Paper Airplane Experiment

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	196.3750000	9.285096	21.1494853	0.0000000
noseyes	-25.7019231	11.801151	-2.1779166	0.0326889
middleyes	-10.1250000	12.203007	-0.8297135	0.4094433
rearyes	-78.4464286	13.591989	-5.7715195	0.0000002
noseyes:middleyes	0.3519231	16.460193	0.0213803	0.9830015
noseyes:rearyes	46.8289072	17.732177	2.6409001	0.0101340
middleyes:rearyes	-6.9494048	17.461897	-0.3979754	0.6918259
noseyes:middleyes:rearyes	29.1669261	23.927946	1.2189482	0.2268425

Table 4: *Summary table of the linear model fitted to the experimental data. Response variable is flight distance with predictor variables including the presence (or absence) of paperclips at the nose, middle, and rear, as well as their interactions.*

```r
#use the formula from lectures
f <- summary(model1)$fstatistic
calc_pval <- pf(f[1], df1=f[2], df2=f[3], lower.tail=FALSE)

#calcualte overall p-value from f-statistic
ovr_p_val <- round(calc_pval[[1]], 14)
p_matrix <- matrix(c(f[[1]], ovr_p_val), ncol = 2)
colnames(p_matrix) <- c("F-statistic", "p-value")

kable(p_matrix, digits = 14, caption="Overall Model Statistics")
```

Table 5: Overall Model Statistics

F-statistic	p-value
15.51565	3.23e-12

```r
#calculate bonferroni adjusted significance level
bonferroni <- round((0.05/7), 3)
```

Table 5: *Model F-statistic and overall p-value for the linear model calculated from the F-statistic.*

With an overall p-value of 3.23×10^{-12}, far below significance level $\alpha = 0.05$, there is sufficient evidence to reject the null hypothesis in favor of the alternative. Thus, there is statistical evidence that at least one predictor or interaction term is significantly different from zero and has a significant impact on flight distance.

Furthermore, the individual coefficient p-values can be evaluated at a Bonferroni-adjusted significance level of 0.007 to determine their significance. At this adjusted significance level, the "rearyes" predictor was the sole significant predictor as all other coefficient p-values were greater than 0.007, thus

failing to be significant.
It should be noted that neither method of determining significance should be held over the other. However, both the global significance test and the Bonferroni-adjusted partial significance test identified at least one predictor as significant. As a result, the linear model suggests that the weight placement on the paper airplane may have a statistically significant effect on flight distance.

Discussion

This study assessed the potential for different weight distributions to effect the flight of an aircraft. Due to limited resources and time, this study is based on an experiment involving paper airplanes and paperclips as this simplistic model can still provide valuable insight into real world aerodynamics. This study explored if the presence of paperclips in any configuration had some impact on the distance the plane flew. The plane was thrown with some combination of paperclips based on randomized factors representing the presence, or absence of a paperclip placed on the nose, middle or rear of the plane. Each factor had levels “yes” and “no” to indicate this. The throws with no paperclips at all had all factors at the “no” level.

The experiment data consisted of ten replicates of eight samples each. While sample size calculations indicated anywhere from eight to over twenty replicates could be needed, the sample size of eighty proved to have significant statistical power. This was likely due to differences in the pilot study and the actual experiment, which will be discussed further. Additionally, from the sample size of eighty it was observed that some groups such as the rear only group had nearly half of the observations of nose only group. By visualizing the data with a factored box plot and a jittered scatter plot, it was observed that in any combination of nose and rear paperclip placement, both middle “yes” and middle “no” throws seemed to produce similar data points, suggesting it was consistent over both levels, or that it may have less of an impact compared to the other paperclip placements. However, further statistical tests would need to be performed to validate this suggestion.

Results from the linear model fitted to the experimental data supported the alternative hypothesis that at least one of the paperclip placements, either nose, middle, rear, or some combination of the three had a significant effect on the flight distance. However, a Bonferroni-adjusted partial significance test found that the only significant condition was when a paperclip was placed solely on the rear. While the adjusted partial significance test still supports the alternative hypothesis, additional testing could provide more insight into these results. Validity of the model assumptions required to perform the statistical methods were verified prior to conducting the statistical tests as assumptions of normality, structure and variance were met. By analyzing the linear model residuals with a histogram, Q-Q plot, and Shapiro-Wilk test, the residuals appeared to be approximately normally distributed. Further, plots of the residuals against fitted values and the residuals over time verified that residual variance was consistent and the data unstructured. Overall, these results suggest an effect on flight distance as a result of different paperclip placements.
This simplified experiment aligns with real-world studies such as Jardine (2009), Brueckner et al. (2017), and others previously discussed that concluded weight distribution in real-world aircraft was a factor influencing the fuel efficiency and carbon emissions. Understanding how wieght distribution affects flight dynamics in even a simple model such as the experiment conducted in this study reinforces the importance of optimizing weight distribution in aircraft. In doing so, airlines could enhance fuel efficiency and contribute to reducing carbon emissions globally, as well as decrease refueling time and increase profits with reduced fuel consumption. However, in order for us to determine a strong correlation between weight distribution and real aircraft flight, further detailed research involving more real-world experiments would be necessary.

Additionally, it should be noted that while the weight distributions were randomized to eliminate bias, and the appropriate model assumptions were met to proceed with our statistical tests, certain factors could still have impacted the findings of this study. One such factor was the sample size of eighty throws. Assuming a near identical beta standard errors to the those in the pilot study, an appropriate experiment would have included at least twenty replicates per condition, resulting in a sample size of one-hundred and sixty throws. In comparison, the actual sample size was only eighty. The difference in sample size was due to limitations of resources, primarily time constraints. This discrepancy was assumed to reduce the statistical power of the study based on the sample size calculations. However, rejection of the null hypothesis in favor of the alternate suggest the variance in the experiment was smaller or effect size larger than suggested by the pilot study. Considering the limited number of throws in the trial study, and due to factors we will discuss next, the pilot study is most likely based on larger beta standard errors compared to those in the experiment. Another potential factor of error is bias in the throws, despite the randomization. While the randomization should have eliminated bias, the weight difference in the plane caused by the paperclips (approximately one gram each) might not have had a sufficient impact on the participants throws, allowing them to improve their technique after a large number of throws. This does not mean that the paperclips may have had no power to influence the planes flight distance, but rather that the participant may have been able to adapt their throws to the weight difference whether consciously or subconsciously. Given that some treatments were more populated over time, this may have skewed overall results. Future research, which mitigates the effects of these potentially limiting factors, and which includes a more extensive pilot study to determine proper sample size could further determine the validity of this study's results.

References

1. Brueckner JK, Abreu C. Airline fuel usage and carbon emissions: determining factors. J Air Transp Manag. 2017;62:10-17. doi:10.1016/j.jairtraman.2017.01.004

2. Ismail NI, Ali ZM, Ishak IS, Noor RM, Rabilah R. Aerodynamic performances of paper planes. Journal of Advanced Research in Fluid Mechanics and Thermal Sciences. 2021 Apr 23;77(1):124-31.

3. Jardine CN. Calculating the carbon dioxide emissions of flights. University of Oxford; 2009 Feb.

4. Li H, Goodwill T, Jane Wang Z, Ristroph L. Centre of mass location, flight modes, stability and dynamic modelling of gliders. Journal of Fluid Mechanics. 2022;937:A6. doi:10.1017/jfm.2022.89

5. Livingston S. Making a paper airplane: The art of making things. The Art of Making Things [Internet]. 2020 Mar 1 [cited 2025 Mar 17]. Available from: artofmakingthings.com/articles/making-a-paper-airplane

6. NYU Web Communications. For new insights into aerodynamics, scientists turn to paper airplanes. NYU. 2022 Mar 1. Available from: https://www.nyu.edu/about/news-publications/news/2022/march/for-new-insights-into-aerodynamics--scientists-turn-to-paper-air.html

7. Puspita ANG, Ambarwati L, Ludfiyanti E, Tejonugroho DP, Zen Y. Effect of paper weight, paper length, and nose of paper plane on aircraft mileage in paper airplane game. 2020.

8. Ritchie H. What share of global CO2 emissions come from aviation? Our World in Data [Internet]. 2024 [cited 2025 Mar 17]. Available from: https://ourworldindata.org/global-aviation-emissions