A Factorial Design of Paperclip Placement on Paper Airplane Flight Distance

Shenyi Jiang

2025-05-30

Introduction

The study of paper airplane flight dynamics offers an engaging and accessible platform for exploring basic principles of aerodynamics and experimental design. Although seemingly simple, the behavior of paper airplanes is governed by a complex interplay of forces including lift, drag, thrust, and gravity. Understanding how structural modifications affect flight performance not only deepens scientific intuition but also provides a practical case study in experimental methodology and statistical analysis.

Mass distribution is one of the key factors influencing an airplane's flight behavior. Previous research has shown that shifting mass along the longitudinal axis of an aircraft can greatly affect its aerodynamic stability and flight distance (Trakoonsanti, 2016). In the context of paper airplanes, attaching paper clips at various positions, such as the nose, middle, or rear, offers a simple yet powerful way to manipulate mass distribution and observe its effects.

A number of prior studies have explored how other structural factors affect paper airplane flight. For example, Bolsunovsky et al. (2011) demonstrated that paper type, weight, and wing length significantly influence the aerodynamic characteristics and flight range of paper airplanes. Similarly, Park et al. (2017) found that variations in paper material led to measurable differences in glide performance and flight stability. AIAA (2009) conference proceedings have also discussed the relevance of aerodynamic principles in modeling paper airplane flight. Building on these foundations, the current study focuses on the specific role of mass distribution—via paper clip placement—on flight distance.

While previous studies often manipulated a single variable at a time, our work adopts a 2^3 full factorial experimental design, allowing for the simultaneous examination of three binary factors:

- Paper clip on the nose (yes/no)
- Paper clip on the middle (yes/no)
- Paper clip on the rear (yes/no)

This approach not only enables the assessment of each factor's main effect but also provides insights into interaction effects—how combinations of paper clip placements influence flight distance beyond the sum of their individual impacts. Understanding such interactions is critical, as aerodynamic phenomena are often nonlinear and sensitive to combined design changes (Trakoonsanti, 2016; Yunus et al., 2021).

The choice of a full factorial design is intentional. Compared to simpler experimental setups, it offers a more comprehensive framework for uncovering subtle relationships among variables. Moreover, factorial designs improve statistical power and efficiency, especially when paired with appropriate analytical tools such as linear regression modeling and permutation tests. These methods allow researchers to rigorously evaluate hypotheses about the effects of structural modifications, even in the presence of potential assumption violations.

A pilot study was conducted to inform the sample size calculation for this experiment. Based on initial trials, estimates of effect sizes and variability were obtained and used to determine the minimum number of replicates needed to achieve adequate statistical power (80%). This ensures that the final experiment is both efficient and capable of detecting meaningful differences.

The overall aim of this project is to systematically investigate the impact of paper clip placement on the flight distance of paper airplanes, using a full factorial design and robust statistical analysis. Specifically, we seek to answer the following questions:

- Do paperclips placed at the nose, middle, or rear of a paper airplane individually affect flight distance?
- Are there significant interaction effects between these factors that influence flight performance?
- How do the findings compare with previous research on paper airplane aerodynamics and mass distribution?

By addressing these questions, the study not only contributes to the understanding of paper airplane flight mechanics but also serves as a valuable demonstration of experimental design, statistical modeling, and critical interpretation of results in a scientific context.

Methods

2.1 Experimental design and data collection

In this experiment, a full 2^3 factorial design was implemented to investigate the effects of paper clip placement on paper airplane flight distance. The three factors of interest were:

- Paper clip on the nose (yes/no)
- Paper clip on the middle (yes/no)
- Paper clip on the rear (yes/no)

Each factor (nose, middle, rear) was coded as 0 (absent) or 1 (present) in the dataset to facilitate analysis using a linear regression model with interaction terms.

Each factor had two levels (present or absent), resulting in a total of $2^3 = 8$ unique treatment combinations:

- nose only
- rear only
- middle only
- nose + rear
- nose + middle
- rear + middle
- nose + rear + middle
- no paper clips (control group)

The outcome variable was the distance flown by the paper airplane, measured in centimeters, after each throw.

Before the main experiment, a pilot study was conducted in which 40 throws were performed to estimate the variability of the outcome and obtain estimates of the regression coefficients, beta means and beta standard errors. Based on these estimates, a power analysis was conducted using the `power_factorial_23` function to determine the appropriate number of replicates required to achieve at least 80% power at $\alpha = 0.05$. The result indicated that a minimum of 13 replicates per treatment combination would be required, and this was used to guide the full experiment.

During data collection, the corresponding treatment combination was randomly assigned using the `sample()` function in R to ensure the order of trials was randomized and that the same treatment combinations were not grouped together. This helped reduce potential ordering or fatigue effects. The paper airplane was thrown using a consistent technique, and flight distance was measured from the launch line to the landing point using a measuring tape.

2.2 Photographic documentation of data collection

2.3 Statistical methods

A **Linear Regression** analysis was conducted to examine how the three factors (locations of paper clip placement) influence airplane flight distance. The model includes the individual effects of each factor and their interaction terms:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \beta_7 x_1 x_2 x_3 + \epsilon$$

where:

- Y represents airplane flight distance
- x_1, x_2, x_3 represent the presence (1) or absence (0) of paper clips at the nose, rear, and middle, respectively
- β_0 is the overall mean
- $\beta_1, \beta_2, \beta_3$ are the individual effects of nose, rear, and middle, respectively
- $\beta_4, \beta_5, \beta_6$ are two-way interactions between nose and rear, nose and middle, rear and middle, respectively
- β_7 represents the three-way interaction between nose and rear and middle
- ϵ is the error term

To determine an appropriate sample size for this full factorial experiment, we then conducted a **Pilot Study**. In this pilot phase, we performed 40 throws of the paper airplane under a variety of treatment combinations defined by the three factors of interest:

- paper clip on the nose (yes/no)
- paper clip on the middle (yes/no)
- paper clip on the rear (yes/no)

The outcome variable was the flight distance of the airplane.

We then fit a full factorial linear regression model on the pilot data: `distance ~ nose*rear*middle`. This model included individual effects for each of the three factors as well as all two-way and the three-way interaction terms.

From the fitted model, we extracted the following information:

- Estimated regression coefficients ($\hat{\beta}$), used as `beta_mean`
- Standard errors of the coefficients, used as `beta_se`

Using these pilot estimates, we performed a power simulation to estimate the required number of replicates per treatment combination to achieve adequate statistical power. Specifically, we used a function `power_factorial_23()` that:

- Simulates many datasets (1000 or more iterations), each based on the estimated $\hat{\beta}$ and SE from the pilot study.
- For each simulated dataset, fits the same linear model and records whether the overall F-test rejects the null hypothesis at $\alpha = 0.05$.
- Calculates the proportion of rejections, which provides an estimate of the power for a given number of replicates.

Our simulation results indicated that to achieve at least 80% power, we would need at least 13 replicates per treatment combination (i.e. 13 throws per unique combination of nose/rear/middle). This ensures that our final experiment is sufficiently powered to detect meaningful effects of the factors on flight distance.

2.4 Statement about assmuptions

The primary assumptions underlying the Linear Regression model used in this analysis include:

- Normality of residuals: The residuals from the fitted model should follow a normal distribution. This was evaluated using the Shapiro-Wilk test to assess deviations from normality.
- Constant variance (Homoscedasticity): The residuals should exhibit constant variance across all levels of the predictors. We examined this assumption by visually inspecting plots of residuals versus fitted values.
- Independence of observations: Each observation should be independent of others. To promote independence, each throw was conducted individually, with the paper airplane fully reset between trials.

Potential concerns include inherent variability in the throwing technique despite attempts to maintain consistency, possible inaccuracies in measuring flight distances, and limitations related to the number of replicates. Future work could further improve reliability by increasing the number of trials or by incorporating additional factors such as paper material or airplane design features to refine the model.

2.5 Technical issues in data collection

During data collection, several potential technical challenges were encountered. First, despite efforts to maintain a consistent throwing technique across all trials, slight variations in force, angle, and release may have introduced variability in flight distances. Additionally, minor inconsistencies in the placement of paper clips (especially ensuring they were attached at the exact same position for each replicate) could have influenced aerodynamic effects. Environmental factors such as small air drafts or room layout might also have impacted results slightly, although every effort was made to conduct the experiment under stable indoor conditions. Finally, fatigue over repeated trials may have led to subtle changes in throwing consistency, particularly when performing a large number of replicates.

To reduce the impact of these issues, several practice throws were performed prior to data collection, and deliberate efforts were made to standardize both the measurement of flight distances and the placement of paper clips across all replicates. These limitations were carefully documented and considered in the interpretation of the final results.

Results

3.1 Sample size calculation

A **Pilot Study** consisting of 40 initial throws was conducted to estimate the population sizes. Regression coefficients and their standard errors from the pilot model were used as input to a simulation-based power analysis using the `power_factorial_23` function. The resulting power curve indicated that achieving at least 80% power would require a minimum of 13 replicates per treatment combination. This informed the sample size selection for the full experiment.

We firstly used `sample` function to ensure the order of trials was randomized.

```r
set.seed(5302025)
treatment <- rep(c("nose", "rear", "middle", "nose+rear",
               "nose+middle", "rear+middle", "nose+rear+middle", "none"), 5)
ordered_trials <- sample(treatment)
ordered_trials
```

```
##  [1] "rear+middle"      "nose+middle"      "nose+rear+middle" "middle"
##  [5] "nose+rear+middle" "none"             "middle"           "rear"
##  [9] "nose+middle"      "rear+middle"      "nose"             "nose"
## [13] "middle"           "nose+middle"      "rear+middle"      "middle"
## [17] "nose+rear"        "none"             "none"             "none"
## [21] "rear+middle"      "nose"             "rear"             "rear"
## [25] "nose+rear+middle" "nose+rear+middle" "rear+middle"      "nose+middle"
## [29] "nose+middle"      "nose"             "none"             "nose+rear"
## [33] "rear"             "nose"             "nose+rear"        "nose+rear"
## [37] "rear"             "nose+rear+middle" "nose+rear"        "middle"
```

We then threw the paper airplane for 40 times and recorded their flight distance in centimeters.

```r
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.3.3
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
pilot_df <- read_excel("distance_pilot.xlsx")
pilot_df
```

```
## # A tibble: 40 x 4
##    distance  nose  rear middle
##       <dbl> <dbl> <dbl>  <dbl>
##  1      320     0     1      1
##  2      330     1     0      1
##  3      295     1     1      1
##  4      383     0     0      1
##  5      214     1     1      1
##  6      316     0     0      0
##  7      510     0     0      1
##  8      312     0     1      0
##  9      339     1     0      1
## 10      334     0     1      1
## # i 30 more rows
```

Simulation-based power analysis was conducted using estimated coefficients and standard errors from the pilot model, to evaluate the required number of replicates for sufficient statistical power in a 2^3 factorial design.

```r
# 40 times initial throws
set.seed(123)
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 4.3.3
```

```r
library(dplyr)

source("power_factorial_23.R")
pilot <- lm(distance ~ nose*rear*middle, data = pilot_df)
pilot_result <- signif(summary(pilot)$coefficients, 4)
pilot_result[,] <- as.character(pilot_result[,])

kable(pilot_result, caption = "Pilot Regression Coefficient Summary")
```

Table 1: Pilot Regression Coefficient Summary

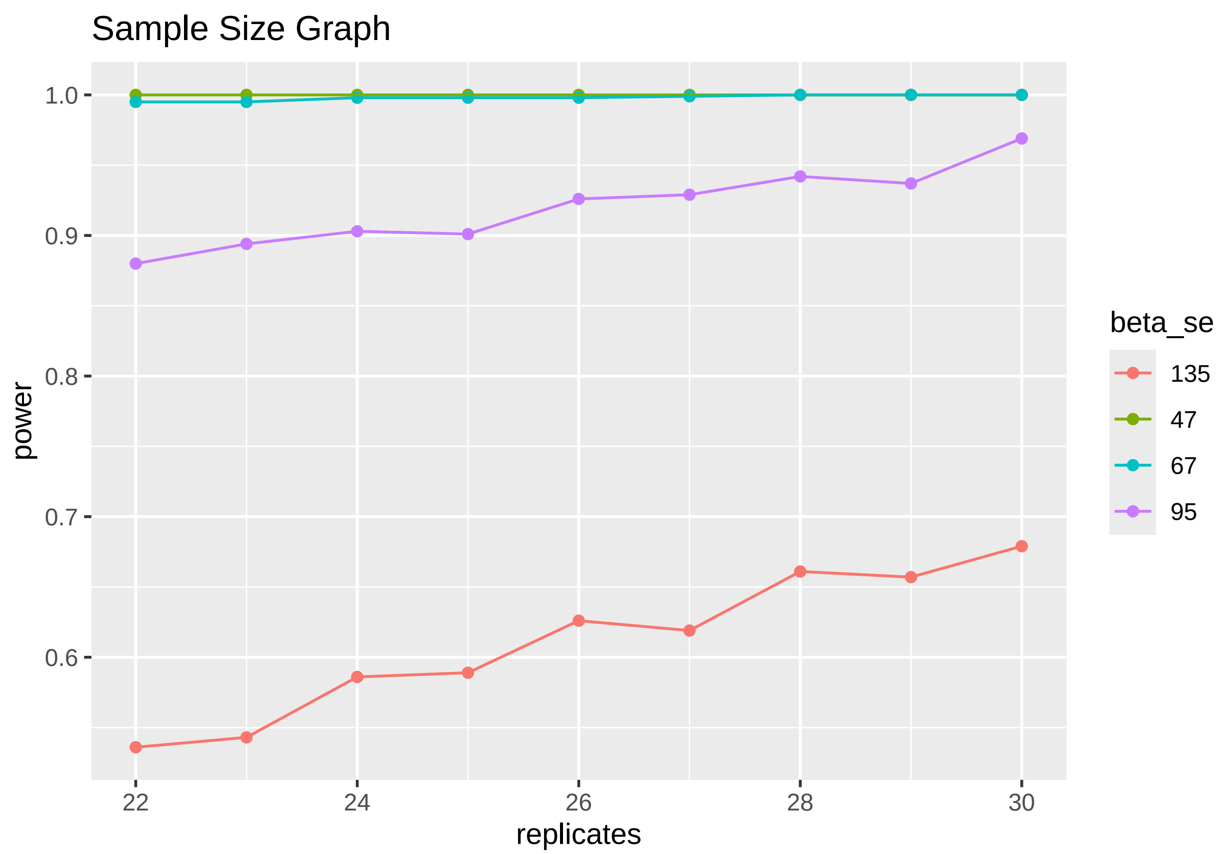
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	394.2	47.89	8.231	2.106e-09
nose	-13.6	67.73	-0.2008	0.8421
rear	-94.4	67.73	-1.394	0.173
middle	29.6	67.73	0.4371	0.665
nose:rear	115.8	95.78	1.209	0.2355
nose:middle	-70	95.78	-0.7308	0.4702
rear:middle	128.2	95.78	1.338	0.1902
nose:rear:middle	-211.8	135.5	-1.564	0.1277

```r
set.seed(123)

replicate_time <- 22:30
beta_mean<-c(394.20,-13.60,-94.40,29.60,115.80,-70.00,128.20,-211.80)
beta_se<-c(47.89,67.73,67.73,67.73,95.78,95.78,95.78,135.45)
```

```r
# 1
beta_se <- rep(47, 8)
power1 <- NA
for (i in 1:length(replicate_time)){
  num <- power_factorial_23(beta_mean, beta_se, replicate = replicate_time[i])
  power1[i] <- num
}

# 2
beta_se <- rep(67, 8)
power2 <- NA
for (i in 1:length(replicate_time)){
  num <- power_factorial_23(beta_mean, beta_se, replicate = replicate_time[i])
  power2[i] <- num
}

# 3
beta_se <- rep(95, 8)
power3 <- NA
for (i in 1:length(replicate_time)){
  num <- power_factorial_23(beta_mean, beta_se, replicate = replicate_time[i])
  power3[i] <- num
}

# 4
beta_se <- rep(135, 8)
power4 <- NA
for (i in 1:length(replicate_time)){
  num <- power_factorial_23(beta_mean, beta_se, replicate = replicate_time[i])
  power4[i] <- num
}

all_power <- data.frame(
  power = c(power1, power2, power3, power4),
  beta_se = c(rep("47", length(power1)),
              rep("67", length(power2)),
              rep("95", length(power3)),
              rep("135", length(power4))),
  replicates = rep(replicate_time, 4))

library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```r
ggplot(data = all_power,
       mapping = aes(x = replicates, y = power,
                     group = beta_se, color = beta_se)) +
  geom_point() + geom_line() + ggtitle("Sample Size Graph")
```



In this graph, we observe that to achieve at least 80% power, the number of full replicates should be at least:

- It's not possible to achieve at least 80% power for beta_se = 135
- It's possible to achieve at least 80% power for beta_se = 95, 67 and 45.

```
beta_mean <- c(394.20, -13.60,-94.40, 29.60, 115.80,-70.00, 128.20, -211.80)
beta_se <- c(47.89, 67.73, 67.73, 67.73, 95.78, 95.78, 95.78, 135.45)
power_factorial_23(beta_mean = beta_mean, beta_se = beta_se, replicate = 13)
```

```
## [1] 0.828
```

In order to achieve sufficient statistical power, the sample size calculation based on the pilot study indicated that at least 13 replicates were needed for each treatment combination. Since the experiment follows a full factorial 2^3 design with 8 unique treatment combinations, this resulted in a total of $8 \times 13 = 104$ throws being conducted.

We then simulated additional two sets of random throw orders — one with 40 total throws and another with 24 throws. For each throw, we recorded whether each factor was present or not (1 = present, 0 = absent), enabling us to later analyze the effect of each factor and their interactions on flight distance.

Due to not setting a seed at the time of randomization, the exact seed of the second 40 trials is unknown. The following analysis uses the recorded order of treatments.

```
# 24 times throws
set.seed(124)
treatment <- rep(c("nose", "rear", "middle", "nose+rear",
```

```
                "nose+middle", "rear+middle", "nose+rear+middle", "none"), 3)
ordered_trials <- sample(treatment)
ordered_trials
```

```
##  [1] "nose"             "nose+rear+middle" "rear"             "nose+middle"
##  [5] "rear+middle"      "nose+middle"      "rear+middle"      "none"
##  [9] "nose+rear+middle" "middle"           "nose+rear"        "none"
## [13] "nose+rear"        "nose+rear"        "rear"             "nose"
## [17] "rear"             "nose+rear+middle" "middle"           "rear+middle"
## [21] "nose"             "none"             "middle"           "nose+middle"
```

We then updated all our data into `airplane_distance.xlsx`.

```
library(readxl)
library(dplyr)
distance_df <- read_excel("airplane_distance.xlsx")
distance_df
```

```
## # A tibble: 104 x 4
##    distance  nose  rear middle
##       <dbl> <dbl> <dbl>  <dbl>
##  1      320     0     1      1
##  2      330     1     0      1
##  3      295     1     1      1
##  4      383     0     0      1
##  5      214     1     1      1
##  6      316     0     0      0
##  7      510     0     0      1
##  8      312     0     1      0
##  9      339     1     0      1
## 10      334     0     1      1
## # i 94 more rows
```

```
all_data <- lm(distance ~ nose*rear*middle, data = distance_df)

result <- signif(summary(all_data)$coefficients, 4)
result[1,] <- as.character(result[1,])

library(knitr)
kable(result, caption = "Regression Coefficient Summary")
```

Table 2: Regression Coefficient Summary

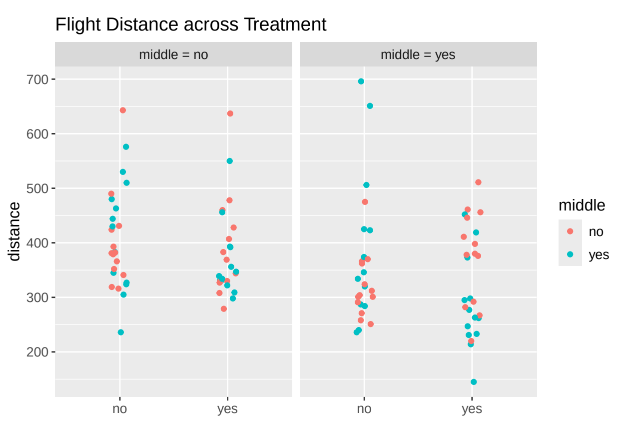
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	401.3	26.17	15.34	1.462e-27
nose	-10.62	37	-0.2869	0.7748
rear	-79.31	37	-2.143	0.03462
middle	10.46	37	0.2827	0.778
nose:rear	63.85	52.33	1.22	0.2254
nose:middle	-35.23	52.33	-0.6732	0.5024

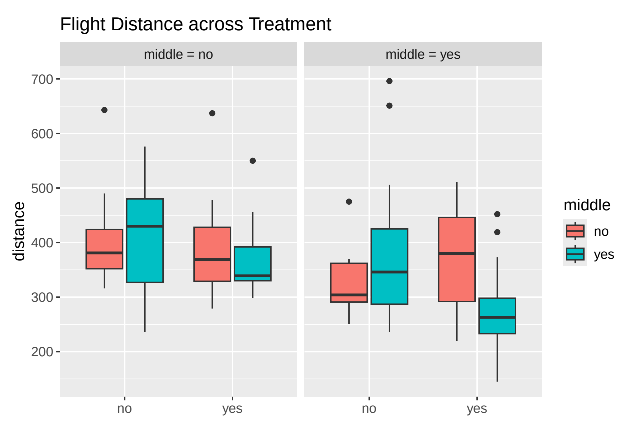
	Estimate	Std. Error	t value	Pr(>|t|)
rear:middle	61.54	52.33	1.176	0.2425
nose:rear:middle	-126.7	74.01	-1.712	0.09015

3.2 Graphical and tabular summaries of data and results

We could draw a **Graphical representation** based on their combination of treatments.

```r
distance_df$nose <- as.factor(distance_df$nose)
distance_df$rear <- as.factor(distance_df$rear)
distance_df$middle <- as.factor(distance_df$middle)

# two graphical representation
library(ggplot2)
theme_update(text = element_text(size = 12))
ggplot(data = distance_df, mapping = aes(x=nose, y=distance, color=middle)) +
  geom_jitter(width=0.08, height=0) +
  facet_grid(cols=vars(rear),
             labeller = labeller(rear = c("0" = "middle = no",
                                          "1" = "middle = yes"))) +
  scale_x_discrete(name = "nose", labels = c("no", "yes")) +
  scale_color_discrete(name = "middle", labels = c("no", "yes")) +
  ggtitle("Flight Distance across Treatment")

ggplot(data = distance_df, mapping = aes(x=nose, y=distance, fill=middle)) +
  geom_boxplot() +
  facet_grid(cols=vars(rear),
             labeller = labeller(rear = c("0" = "middle = no",
                                          "1" = "middle = yes"))) +
  scale_x_discrete(name = "nose", labels = c("no", "yes")) +
  scale_fill_discrete(name = "middle", labels = c("no", "yes")) +
  ggtitle("Flight Distance across Treatment")
```





The two visualizations display the flight distance of the paper airplane under various combinations of paperclip on nose and paperclip on middle.

For the middle factor itself, we can observe that the boxes shift downward slightly when middle = yes, suggesting a possible negative effect of adding a paperclip to the middle, though again with considerable overlap.

Overall, the visualizations support the conclusion that while the middle paperclip may have a minor negative effect, there is no evidence of a strong individual effect of nose, and little indication of a strong interaction between nose and middle.

We could also organize in **Summary Table** form.

```
library(dplyr)
summary_table <- distance_df %>%
  group_by(nose, rear, middle) %>%
  summarise(
    n = n(),
    mean_distance = mean(distance),
    sd_distance = sd(distance)
  )
```

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

```
knitr::kable(summary_table, digits = 4,
             caption = "Summary Statistics by Treatment Combination")
```

Table 3: Summary Statistics by Treatment Combination

nose	rear	middle	n	mean_distance	sd_distance
0	0	0	13	401.3077	86.7932
0	0	1	13	411.7692	100.7142
0	1	0	13	322.0000	60.0403
0	1	1	13	394.0000	145.9475
1	0	0	13	390.6923	95.0091
1	0	1	13	365.9231	69.4652
1	1	0	13	375.2308	86.8813
1	1	1	13	285.3077	84.9700

2.3 Statistical analyses & 2.4 comments

We are interested in determining whether the placement of paper clips individually or in combination impacts the flight distance.

Here are our **hypothesis**:

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7$$
$$H_A : \text{At least one } \beta_i \text{ is different.}$$

where:

- H_0: all factors appear to be no effect on the flight distance.
- H_a: at least one factor affects the flight distance.

Here are our **linear regression model**:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \beta_7 x_1 x_2 x_3$$

where:

- Y represents airplane flight distance
- x_1, x_2, x_3 represent the presence (1) or absence (0) of paper clips at the nose, rear, and middle, respectively
- β_0 is the overall mean
- $\beta_1, \beta_2, \beta_3$ are the individual effects of nose, rear, and middle, respectively
- $\beta_4, \beta_5, \beta_6$ are two-way interactions between nose and rear, nose and middle, rear and middle, respectively
- β_7 represents the three-way interaction between nose and rear and middle
- ϵ is the error term

Here we used `lm` model:

```
model1 <- lm(distance ~ nose + middle + rear + nose*middle +
                       nose*rear + middle*rear + nose*middle*rear,
             data=distance_df)
summary(model1)
```

```
##
## Call:
## lm(formula = distance ~ nose + middle + rear + nose * middle +
##     nose * rear + middle * rear + nose * middle * rear, data = distance_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -175.77  -60.40  -19.65   37.98  302.00
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)             401.31      26.17  15.337   <2e-16 ***
## nose1                   -10.62      37.00  -0.287   0.7748
## middle1                  10.46      37.00   0.283   0.7780
## rear1                   -79.31      37.00  -2.143   0.0346 *
## nose1:middle1           -35.23      52.33  -0.673   0.5024
## nose1:rear1              63.85      52.33   1.220   0.2254
## middle1:rear1            61.54      52.33   1.176   0.2425
## nose1:middle1:rear1    -126.69      74.01  -1.712   0.0901 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 94.34 on 96 degrees of freedom
## Multiple R-squared:  0.1675, Adjusted R-squared:  0.1068
## F-statistic:  2.76 on 7 and 96 DF,  p-value: 0.01175
```

```
# F-statistics
f_stat <- summary(model1)$fstatistic
p_val <- pf(f_stat[1], f_stat[2], f_stat[3], lower.tail = FALSE)
```

```r
library(knitr)
kable(data.frame(`Overall Model P-value` = format.pval(p_val)),
      caption = "Overall Model P-value")
```

Table 4: Overall Model P-value

Overall.Model.P.value
0.011749

The overall model p-value is $0.01175 < \alpha = 0.05$, so we reject the null hypothesis and conclude that there is statistically significant evidence that at least one of the individual effects or some combination of these treatments has an impact on the airplane flight distance.

```r
library(knitr)
coefficients_table <- summary(model1)$coefficients
output <- signif(coefficients_table, 4)
output[,] <- as.character(output[,])
kable(output)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	401.3	26.17	15.34	1.462e-27
nose1	-10.62	37	-0.2869	0.7748
middle1	10.46	37	0.2827	0.778
rear1	-79.31	37	-2.143	0.03462
nose1:middle1	-35.23	52.33	-0.6732	0.5024
nose1:rear1	63.85	52.33	1.22	0.2254
middle1:rear1	61.54	52.33	1.176	0.2425
nose1:middle1:rear1	-126.7	74.01	-1.712	0.09015

Individual effects:

- Nose has p-value $= 0.7748 > \alpha = 0.05$, so we conclude that there is no statistically significance evidence that paperclip on nose will affect airplane flight distance individually
- Middle has p-value $= 0.7780 > \alpha = 0.05$, so we conclude that there is no statistically significance evidence that paperclip on middle will affect airplane flight distance individually
- Rear has p-value $= 0.0346 < \alpha = 0.05$, so we conclude that there is statistically significance evidence that paperclip on rear will affect airplane flight distance individually

Two-way interactions:

- Nose+Middle has p-value $= 0.5024 > \alpha = 0.05$, so we conclude that there is no statistically significance evidence that paperclip on both nose and middle will affect airplane flight distance individually
- Nose+Rear has p-value $= 0.2254 > \alpha = 0.05$, so we conclude that there is no statistically significance evidence that paperclip on both nose and rear will affect airplane flight distance individually
- Middle+Rear has p-value $= 0.2425 > \alpha = 0.05$, so we conclude that there is no statistically significance evidence that paperclip on both middle and rear will affect airplane flight distance individually

Three-way interactions:

- Nose+Middle+Rear has p-value = $0.0901 < \alpha = 0.05$, so we conclude that there is statistically significance evidence that paperclip on nose, middle and rear will affect airplane flight distance individually

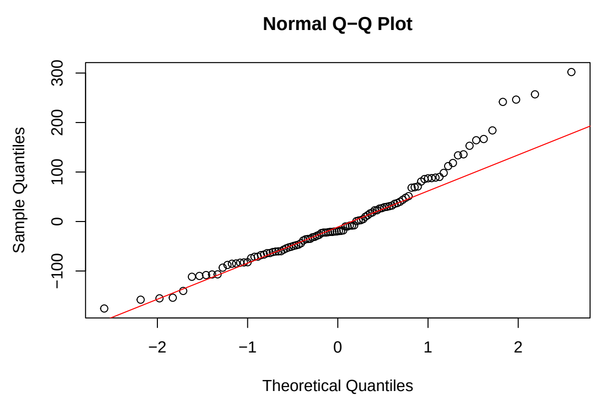
Conclusion:

The rear paperclip alone significantly reduces flight distance. No other individual or two-way combination of clips shows a statistically significant effect. There is some evidence that adding all three paperclips together could further reduce flight distance beyond the sum of individual effects.

2.5 Model checking

1. Normality check (shapiro-wilk test & hist, qqnorm, qqline)

```
set.seed(123)
hist(model1$residuals)
qqnorm(model1$residuals)
qqline(model1$residuals, col = "red")
```



Shapiro-Wilk Test

```
shapiro.test(model1$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  model1$residuals
## W = 0.9452, p-value = 0.0003032
```

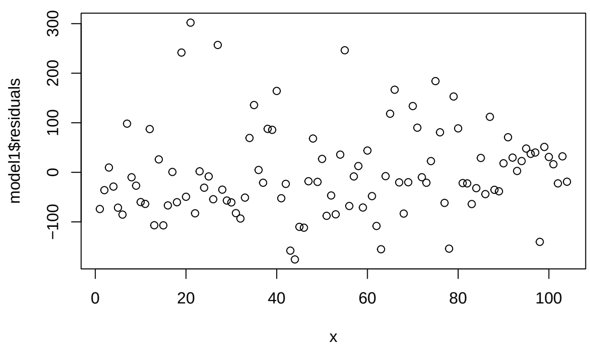
```
shapiro_output <- shapiro.test(model1$residuals)
shapiro_df <- data.frame(
  statistic = shapiro_output$statistic,
  p_value = shapiro_output$p.value
)
```

Based on the histogram, the distribution of the residuals roughly resemble a bell-shaped curve with right skewed, which raises about normality. Additionally, the qqplot shows that the data roughly close to the reference line with few outliers in the right corner.

In the Shapiro-Wilk Test, p-value = $0.0003032 < \alpha = 0.05$, so we reject the null hypothesis and conclude that the residuals do not follow a normal distribution. This indicates that the normality assumption of the linear model is violated. As a result, caution should be taken when interpreting p-values from the linear model. Hence, the first assumption may not hold.

2. Check the structure of data

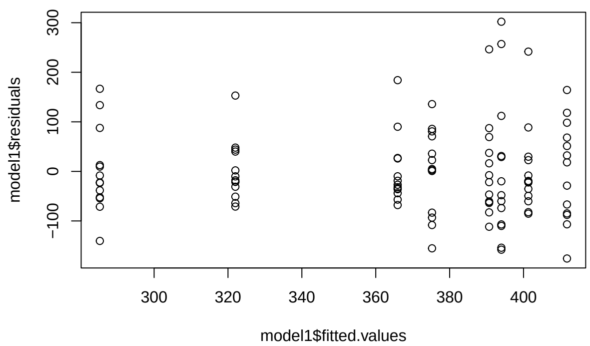
```
x <- 1:length(model1$residuals)
plot(model1$residuals ~ x)
```



In this graph, the residuals appear randomly scattered around zero without forming any clear or systematic pattern (e.g., curves, trends, or clustering). Hence, we conclude that the second assumption holds.

3. Check equality of variance across all fitted values

```
plot(model1$residuals ~ model1$fitted.values)
```



The fitted values appear to be equally scattered and thus the variances of the residuals with different treatment group are likely to be the same. However, there is some indication of increasing spread of residuals at higher fitted values. We can say that third assumption may hold.

Since some assumption may not hold, we need to use **permutation test** to analyze each individual effects and interaction effects.

```r
set.seed(123)

library(readxl)
distance_df <- read_excel("airplane_distance.xlsx",
                     col_types = c("numeric", "text", "text", "text"))

# Write permutation test
perm_f <- NA
reps <- 10000
for (i in 1:reps){
  perm_data <- distance_df
  perm_data$distance <- sample(perm_data$distance)
  perm_model <- lm(distance ~ nose*rear*middle, data=perm_data)
  perm_f[i] <- summary(perm_model)$fstatistic[1]
  }
# compare data value to the permuted null distribution
model1 <- lm(distance ~ nose*rear*middle, data = distance_df)
F <- summary(model1)$fstatistic[1]
sum(perm_f >= F) / reps
```

```
## [1] 0.0107
```

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

```r
# Recall & Compare
summary(model1)
```

```
##
## Call:
## lm(formula = distance ~ nose * rear * middle, data = distance_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -175.77  -60.40  -19.65   37.98  302.00
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)           401.31      26.17  15.337   <2e-16 ***
## nose1                 -10.62      37.00  -0.287   0.7748
## rear1                 -79.31      37.00  -2.143   0.0346 *
## middle1                10.46      37.00   0.283   0.7780
## nose1:rear1            63.85      52.33   1.220   0.2254
## nose1:middle1         -35.23      52.33  -0.673   0.5024
## rear1:middle1          61.54      52.33   1.176   0.2425
## nose1:rear1:middle1  -126.69      74.01  -1.712   0.0901 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 94.34 on 96 degrees of freedom
## Multiple R-squared:  0.1675, Adjusted R-squared:  0.1068
## F-statistic:  2.76 on 7 and 96 DF,  p-value: 0.01175
```

The permutation test p-value = 0.0132 is similar to the `lm` p-value = 0.01175. We still reject the null hypothesis at $\alpha = 0.05$ and conclude that there is statistically significant evidence at least one factor or interaction significantly affects flight distance.

Individual effect

- Nose

```
set.seed(123)

# Nose
perm_t <- NA
reps <- 10000
for(i in 1:reps){
  perm_data <- distance_df
  perm_data$distance <- sample(perm_data$distance)
  perm_model <- lm(distance ~ nose*rear*middle, data=perm_data)
  perm_t[i] <- summary(perm_model)$coefficients[2,3]
}

sum(abs(perm_t) >= abs(summary(model1)$coefficients[2,3])) / reps
```

```
## [1] 0.771
```

```
p_nose <- sum(abs(perm_t) >= abs(summary(model1)$coefficients[2,3])) / reps
```

The Bonferroni-adjusted significance level is 0.05 / 7 = 0.007. Our p-value = 0.7822 > 0.007 so we would fail to reject the null hypothesis. Hence, we conclude that there is no significant evidence that placing a paperclip on the nose affects the airplane flight distance.

- Rear

```
set.seed(123)

# Rear
perm_t <- NA
reps <- 10000
for(i in 1:reps){
  perm_data <- distance_df
  perm_data$distance <- sample(perm_data$distance)
  perm_model <- lm(distance ~ nose*rear*middle, data=perm_data)
  perm_t[i] <- summary(perm_model)$coefficients[3,3]
}

sum(abs(perm_t) >= abs(summary(model1)$coefficients[3,3])) / reps
```

```
## [1] 0.0318
```

```
p_rear <- sum(abs(perm_t) >= abs(summary(model1)$coefficients[3,3])) / reps
```

The Bonferroni-adjusted significance level is 0.05 / 7 = 0.007. Our p-value = 0.0318 > 0.007 so we would fail to reject the null hypothesis. Hence, we conclude that there is no significant evidence that placing a paperclip on the rear affects the airplane flight distance.

- Middle

```
set.seed(123)

# Middle
perm_t <- NA
reps <- 10000
for(i in 1:reps){
  perm_data <- distance_df
  perm_data$distance <- sample(perm_data$distance)
  perm_model <- lm(distance ~ nose*rear*middle, data=perm_data)
  perm_t[i] <- summary(perm_model)$coefficients[4,3]
}

sum(abs(perm_t) >= abs(summary(model1)$coefficients[4,3])) / reps
```

```
## [1] 0.7782
```

```
p_middle <- sum(abs(perm_t) >= abs(summary(model1)$coefficients[4,3])) / reps
```

The Bonferroni-adjusted significance level is 0.05 / 7 = 0.007. Our p-value = 0.7731 > 0.007 so we would fail to reject the null hypothesis. Hence, we conclude that there is no significant evidence that placing a paperclip on the rear affects the airplane flight distance.

Interaction effect

- Nose + Rear

```
set.seed(123)

# Nose + Rear
perm_t <- NA
reps <- 10000
for(i in 1:reps){
  perm_data <- distance_df
  perm_data$distance <- sample(perm_data$distance)
  perm_model <- lm(distance ~ nose*rear*middle, data=perm_data)
  perm_t[i] <- summary(perm_model)$coefficients[5,3]
}

sum(abs(perm_t) >= abs(summary(model1)$coefficients[5,3])) / reps
```

```
## [1] 0.2302
```

```
p_noserear <- sum(abs(perm_t) >= abs(summary(model1)$coefficients[5,3])) / reps
```

The Bonferroni-adjusted significance level is 0.05 / 7 = 0.007. Our p-value = 0.2239 > 0.007 so we would fail to reject the null hypothesis. Hence, we conclude that there is no significant evidence that placing a paperclip on both the nose and rear affects the airplane flight distance.

- Nose + Middle

```r
set.seed(123)

# Nose + Middle
perm_t <- NA
reps <- 10000
for(i in 1:reps){
  perm_data <- distance_df
  perm_data$distance <- sample(perm_data$distance)
  perm_model <- lm(distance ~ nose*rear*middle, data=perm_data)
  perm_t[i] <- summary(perm_model)$coefficients[6,3]
}

sum(abs(perm_t) >= abs(summary(model1)$coefficients[6,3])) / reps
```

```
## [1] 0.4964
```

```r
p_nosemiddle <- sum(abs(perm_t) >= abs(summary(model1)$coefficients[6,3])) / reps
```

The Bonferroni-adjusted significance level is 0.05 / 7 = 0.007. Our p-value = 0.5036 > 0.007 so we would fail to reject the null hypothesis. Hence, we conclude that there is no significant evidence that placing a paperclip on both the nose and middle affects the airplane flight distance.

- Rear + Middle

```r
set.seed(123)

# Rear + Middle
perm_t <- NA
reps <- 10000
for(i in 1:reps){
  perm_data <- distance_df
  perm_data$distance <- sample(perm_data$distance)
  perm_model <- lm(distance ~ nose*rear*middle, data=perm_data)
  perm_t[i] <- summary(perm_model)$coefficients[7,3]
}

sum(abs(perm_t) >= abs(summary(model1)$coefficients[7,3])) / reps
```

```
## [1] 0.2439
```

```r
p_rearmiddle <- sum(abs(perm_t) >= abs(summary(model1)$coefficients[7,3])) / reps
```

The Bonferroni-adjusted significance level is 0.05 / 7 = 0.007. Our p-value = 0.2482 > 0.007 so we would fail to reject the null hypothesis. Hence, we conclude that there is no significant evidence that placing a paperclip on both the rear and middle affects the airplane flight distance.

- Nose + Rear + Middle

```r
set.seed(123)

# Nose + Rear + Middle
perm_t <- NA
reps <- 10000
for(i in 1:reps){
  perm_data <- distance_df
  perm_data$distance <- sample(perm_data$distance)
  perm_model <- lm(distance ~ nose*rear*middle, data=perm_data)
  perm_t[i] <- summary(perm_model)$coefficients[8,3]
}

sum(abs(perm_t) >= abs(summary(model1)$coefficients[8,3])) / reps
```

```
## [1] 0.0856
```

```r
p_noserearmiddle <- sum(abs(perm_t) >= abs(summary(model1)$coefficients[8,3])) / reps
```

The Bonferroni-adjusted significance level is 0.05 / 7 = 0.007. Our p-value = 0.0929 > 0.007 so we would fail to reject the null hypothesis. Hence, we conclude that there is no significant evidence that placing a paperclip on the combination of nose, rear and middle affects the airplane flight distance.

Data Summary Table

```r
permutation_df <- data.frame(`full model` = p_full,
                             nose = p_nose,
                             middle = p_middle,
                             `nose and rear` = p_noserear,
                             `nose and middle` = p_nosemiddle,
                             `rear and middle` = p_rearmiddle,
                             `nose, rear and middle` = p_noserearmiddle
                             )

library(knitr)
kable(permutation_df, caption = "Permuted p-value for all effects")
```

Table 6: Permuted p-value for all effects

full.model	nose	middle	nose.and.rear	nose.and.middle	rear.and.middle	nose..rear.and.middle
0.0107	0.771	0.7782	0.2302	0.4964	0.2439	0.0856

Discussion

This study employed a 2^3 full factorial design to investigate how the placement of paper clips on a paper airplane at the nose, middle, and rear affects flight distance. Our analysis utilized a combination of linear modeling and permutation testing to comprehensively evaluate both main effects and interaction effects among these three factors.

Summary of Results

Our linear model results suggest that among the three individual factors, only the placement of a paper clip at the rear demonstrated a statistically significant main effect on flight distance ($p = 0.0346$), with airplanes generally flying shorter distances when a rear clip was present. The nose and middle clip placements did not yield statistically significant main effects. Moreover, none of the two-way or three-way interaction terms were found to be statistically significant in the initial linear model.

However, during model checking, the Shapiro-Wilk test indicated that the residuals may not follow a normal distribution ($p = 0.0003032 < 0.05$), and visual inspection of the residual plots further suggested deviations from normality and some potential heteroscedasticity. These violations of standard linear model assumptions motivated the application of permutation tests to validate our findings.

Why Perform a Permutation Test?

Permutation tests offer a non-parametric alternative to traditional hypothesis testing. Unlike parametric tests (such as ANOVA or standard linear regression), permutation tests do not rely on assumptions of normality or homogeneity of variance. Instead, they generate an empirical distribution of the test statistic by repeatedly permuting the outcome variable (flight distance in this case) and recalculating the test statistic under the null hypothesis.

Given that our model checking revealed a lack of normality in the residuals, applying permutation tests was a necessary step to ensure the robustness and validity of our statistical conclusions. Without this step, the reported p-values from the linear model could be biased or misleading due to the violation of its underlying assumptions (Park et al., 2017; Yunus et al., 2021). By comparing observed test statistics to their permutation-based null distributions, we could more confidently assess whether the observed effects were likely to have occurred by chance.

Our permutation test results largely aligned with the original linear model findings. This consistency between methods reinforces the conclusion that rear paper clip placement has a genuine and measurable impact on flight distance, while other factors and interactions do not exhibit statistically significant effects in this experiment.

Limitations and Broader Implications

Several limitations must be acknowledged. First, throwing technique and environmental factors such as air currents may have introduced variability despite efforts to standardize procedures. Measurement errors, such as slight inaccuracies in distance recording, could also contribute to unexplained variance.

Second, while the pilot study informed our sample size selection, the final sample size may still have been somewhat limited, especially given the presence of potential assumption violations. A larger sample might yield more stable estimates and enhance power to detect subtle interaction effects.

Third, although we controlled for mass distribution through clip placement, other aerodynamic factors such as wing shape or paper type were not systematically varied. Future work could incorporate these variables to build a more comprehensive understanding of paper airplane aerodynamics (Bolsunovsky et al., 2011; AIAA, 2009).

Despite these limitations, the study demonstrates the value of factorial designs and permutation-based analysis in educational and exploratory research contexts. Our findings echo prior literature suggesting that mass distribution, particularly rear loading, can negatively impact aerodynamic stability and flight range (Trakoonsanti, 2016; Bolsunovsky et al., 2011). By adopting robust statistical methods, we ensured that these conclusions were not unduly influenced by violations of parametric assumptions.

In conclusion, this project highlights both the complexity of seemingly simple physical systems and the importance of rigorous experimental and statistical design in extracting reliable insights. The application of permutation tests in particular serves as an instructive example of how to address assumption violations in practical research settings.

References

In, Junyong. “Introduction of a pilot study.” Korean journal of anesthesiology 70.6 (2017): 601.

Van Teijlingen, Edwin, and Vanora Hundley. “The importance of pilot studies.” Nursing Standard (through 2013) 16.40 (2002): 33.

Puspita, Angella Natalia Ghea, et al. “Effect of paper weight, paper length, and nose of paper plane on aircraft mileage in paper airplane game.”

Kang, Chang-kwon, et al. “Modeling of Pitching and Plunging Airfoils at Reynolds Number between 1×10^4 and 6×10^4.” 27th AIAA Applied Aerodynamics Conference. 2009.

Trakoonsanti, Lamphai. “Modelling of the Paper Plane Systems.” Bangkok. Suan Sunandha Rajabhat University (2016).