

PSTAT 122 Project

Aarti Garaye (aartigaraye)

2024-12-11

Introduction

In the modern, interconnected world shaped by globalization, aviation plays a pivotal role that is often taken for granted. Air transportation is a major industry in its own right and it also provides important inputs into wider economic, political, and social processes. As with most forms of transport, the demand for aviation is derived from the need to achieve other objectives, such as business, tourism, or trade (Talukder et al., 2024). Engineers tasked with designing airplanes bear a profound responsibility for ensuring the safety of millions of lives globally. Despite the critical nature of this work, undergraduate engineering curricula are typically heavily loaded with traditional classroom learning approaches and have a limited number of laboratory-based courses available to students. Yet, research shows that many students benefit greatly from experiential learning opportunities that engage them in active problem-solving and exploration (Hargather et al., 2013).

The fast-evolving technological landscape demands a change from the traditional teaching pathways of our classrooms where memorization and theoretical knowledge for standardized testing is heavily emphasized. Problem-based learning is an effective pedagogical strategy built on this premise. It supports learners in becoming active participants in acquiring new knowledge and applying it in practical contexts (Hmelo-Silver, 2004; Kolodner et al., 2003; Perrenet et al., 2000).

This study aims to engage students in a hands-on experiment using paper airplanes to explore concepts such as weight distribution, dimensional analysis, and scaling in aerospace dynamics. Inspired by Henk Tennekes' book *The Simple Science of Flight: From insects to jumbo jets* by Henk Tennekes this experiment builds on the idea of using paper airplanes as a teaching tool for fundamental principles of flight. In this book, Tennekes explores scaling relationships that are exhibited by flying objects from fruit flies to Boeing jets. While Tennekes illustrates scaling relationships between weight and cruising speed for various flying objects, this study focuses on experimental data collection and analysis to test specific hypotheses.

The research question driving this study is: Does the position of a paperclip on a paper airplane (nose, middle, or rear) significantly affect its flight distance? By addressing this question, we aim to shed light on how weight distribution influences flight efficiency in paper airplanes and explore potential implications for real-world aerodynamics.

This study employs a randomized complete block design (RCBD) to ensure reliability and minimize variability. Each paper airplane serves as a block, while the treatments involve attaching a paperclip to different parts of the plane (nose, middle, rear, or no paperclip as a control). This design allows us to isolate the effect of weight placement on flight distance. It involves building standard paper airplanes, and then adding a paperclip to different parts of the plane to observe how the added weight affects its flight distance.

The experimental setup involves consistent materials and methods to control for extraneous variables. All planes are made from identical paper, and flights are conducted indoors to eliminate wind effects. Throwing technique, launch angle, and measurement methods are standardized to ensure fairness. Flight distances are measured in inches from the launch point to the nose of the airplane, with negative distances avoided, even in cases of backward flight.

The experiment took place on the campus of University of California, Santa Barbara, in the hallway of the education building to minimize the affect of wind. Materials used include a premium 92 bright copy paper sourced from Staples located at the Camino Real Marketplace in Goleta, California, metal paper clips, a 10-foot tape measure, and a laptop to measure the distances in an excel file. Each of the paper airplanes were folded using the website [the art of making things](#) by skipping the optional step.

Ultimately, this paper seeks to determine whether the mean flight distances differ significantly across the treatments. Insights gained from this experiment will not only enhance student learning but also provide a simplified model for understanding fundamental principles in aerospace engineering.

Methods

Data Collection

The experiment was designed using a randomized complete block design (RCBD). The blocks in this experiment are the individual paper airplanes. There are four treatments:

- Control: No paperclip
- Nose: Paper clip attached to the nose of the paper airplane
- Middle: Paper clip attached to the middle of the paper airplane
- Rear: Paper clip attached to the rear end of the paper airplane.

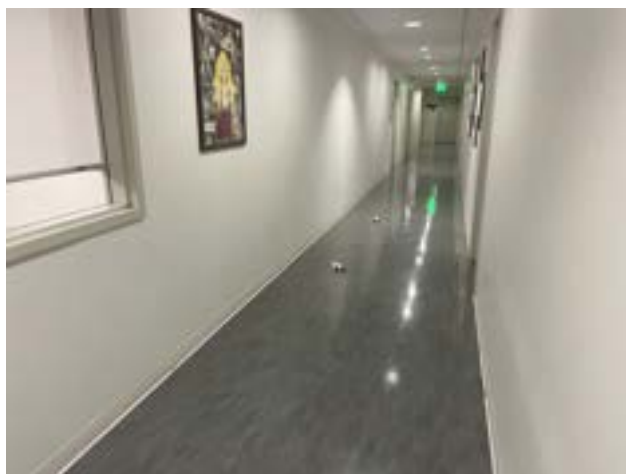
Paper airplanes were constructed using the standard 8.5 x 11 inch printer paper. All paper airplanes were folded using the same method (as mentioned in the introduction) and were thrown by the same person, starting at the same point, at approximately the same launch angle to ensure uniformity. The paperclips used were standard small metal clips (weighing approximately 1 gram). Furthermore, extra paper airplanes were made and thrown before the experiment started to reduce the “first-throw” bias.

I used the sample function to randomize the order of blocks. Then, used the sample function again to randomize the order of the treatment received by each block. Since this is RCBD, we know that each block would the treatment assigned exactly once.

This randomization ensured that no systematic bias influenced the results. Each block was thrown four times, once per treatment, resulting in twenty-eight throws. (Sample size calculations are included in the results). The sequence of throws was predetermined by randomization to control for potential learning effects or fatigue.

Data Measurement

Below is the picture of two blocks thrown in the hallway of the education building when they were both assigned to the control treatment group.



Below is the picture of the tape measure I used to measure the distance of the paper airplane.



And below is the screenshot of my excel file after the measurements. Note that all distances are measured in inches and that the order of blocks and the order of treatment within each block is randomized.

	A	B	C	D
Block	Treatment	Distance		
1	7 control	218.88		
2	7 nose	188.28		
3	7 middle	188.88		
4	7 rear	205.88		
5	8 nose	153.28		
6	8 middle	213.28		
7	8 rear	218.77		
8	9 control	198.28		
9	9 middle	183.28		
10	9 rear	183.88		
11	10 control	188.27		
12	10 nose	188.88		
13	11 control	188.11		
14	11 rear	178.78		
15	12 middle	178.88		
16	12 nose	188.03		
17	13 rear	218.13		
18	14 nose	188.13		
19	14 control	188.78		
20	14 middle	177.77		
21	15 control	188.88		
22	15 nose	157.13		
23	15 rear	188.88		
24	16 middle	188.11		
25	16 nose	188.88		
26	17 control	188.88		
27	17 middle	178.88		
28	17 rear	178.78		

The distance flown by the airplanes was measured in inches from the launch point to the nose of the airplane where it landed. Negative distances were not recorded, even if the airplane flew backward or looped. Each throw was conducted indoors in the hallway of the Education Building on campus to minimize environmental effects such as wind. Other variables, such as the throwing technique, launch angle, and force, were kept as consistent as possible across all trials by having the same individual perform all throws.

Statistical Methods

To analyze the data and test the differences in mean distances, we performed an Analysis of Variance (ANOVA) for RCBD. ANOVA compares the means of three or more groups simultaneously, evaluating if there is a significant difference between them based on the variance within and between groups. ANOVA partitions the total variance in the data into components attributable to different sources (between groups and within groups), then calculates an F-statistic to test the null hypothesis that all group means are equal. However, it only holds true if the assumptions of normality, homoscedasticity, and independence are met. Otherwise, we would have to conduct a permutation test to check whether the difference between means across the different treatment groups is statistically significant or not. This approach allowed us to account for variability between blocks (airplanes) while testing for significant differences in the mean distances across the treatments.

The three things important for experiments are replication, blocking, and randomization. Let's see where each is happening in RCBD:

- Replication: There is more than one paper airplane being thrown. Thus, each block is a replicate.
- Blocking: Each paper airplane trying the four different treatment is a block. Each block gets assigned each treatment once.
- Randomization: The order of block and the order of treatment within each block is randomized using the sample function in R.

Assumptions

As briefly mentioned above, ANOVA only works when the following assumptions are met.

- Normality: The residuals should follow a normal distribution. This could be checked by looking at the histogram of the residuals after we fit the model, use the qqnorm plot, or we could use the Shapiro-Wilk Test. The null hypothesis for the Shapiro-Wilk test says that the residuals follow a normal distribution. Thus, we usually are looking for a high p-value.
- Homogeneity of Variance: The variance across the treatment groups should be equal. We check this by plotting the residuals across the fitted values in the order of which the data was collected. We want to see equal variance across the whole range of fitted values. We do not want to observe a funnel like shape taking place, that would be a cause for concern.
- Independence: The observations should be independent to each other. In other words, there shouldn't be any structure in which the data was collected. We check this by plotting the residuals across the fitted values in the order of which the data was collected. We don't want to see any pattern in the data collected order in order to meet this assumption.

If any of these assumptions aren't met, we have to move to performing the permutation test where these assumptions aren't a pre-requisite. Although, a con of performing a permutation test is that we have to hard code it from scratch since it's not a built-in test in R.

Technical Issues

Some of the technical issues this experiment might face are as follows:

- Variability in throwing techniques: Despite my best effort to be consistent in throwing the paper airplanes, minor differences in throwing force or angle may have influenced flight distance.

- Paper airplane wear and tear: Some airplanes showed slight structural degradation (e.g., creased wings, fractured nose, etc) after multiple throws. This was mitigated by gently reshaping the airplanes between trials.
- Narrowness of the hallway: To avoid the distance being influence by outside factors such as wind, I chose to do this experiment in an elongated hallway of an empty building. However, it wasn't too wide so some of the paper airplane were blocked by the side walls restricting the paper airplanes from going their full potential distance.
- Measurement Precision: Measuring the exact landing point was occasionally challenging, particularly when the airplane did not land straight.
- Taking breaks to record the distance measure: The breaks broke the streak of throwing techniques to be consistent. This might have a slight impact on the throwing techniques. However, it is important to note that everything that was possible was incorporated to make the experiment fair and minimize bias.

Results

Sample size calculation

There are three main approaches to determining the sample size: Literature review, pilot study, and a post-hoc. Since we have conducted a similar experiment in lab 2 of this class, we would like to use that as our pilot study.

To calculate the required sample size for an experiment we need:

- The effect size: This is the differences in means that are either actually true or what you're interested in. It's the differences in means of interest. How big of a difference does it make when you switch your methods. As effect size increases, the required n goes down. How big of an impact does it have if compared to other treatment groups.
- The variance: As variance go up, the required n will go up. We are talking about the variance within the groups. If the data is spread out, it becomes harder to test whether that difference was statistically significant so to test the data with a lot of variance we need a bigger n.
- The alpha level: As alpha increases, the required n should go down. Alpha is the probability of making a type I error. So if we want to increases the probability of rejecting the null, we need our n to go down.
- The desired power: As you increase your desired power, your sample size n needs to go up. Power is the probability of correctly reject the null so to increase the power you need a larger n. Meaning power is a good thing so need to make sure our n is big. Generally speaking, we would want a power of 80% or higher.

To determine the sample size calculation of how many paper airplanes are needed, we need to know the mean distances and the standard deviation of the data we collected in lab 2. Below is the table of data measurements from lab 2.

Table 1: Matrix of data measurements

	with	without	difference
Hilary	186	143	43
Xinyue	237	182	55

	with	without	difference
SiddharthC	236	200	36
Will	190	256	-66
Aarti	212	164	48
SiddharthS	260	95	165

In this lab there were only two treatments throwing the paper airplane without the paper clip and with the paper clip attached to the nose. A group of six people took turns throwing the same paper airplane with and without the paper clip in a randomly generated order.

To use this as our pilot study for the sample size calculation, we need to find the total mean, standard deviations, and variances across each group. Below is the data frame of the summary statistics of our pilot study.

Table 2: Summary Statistics of with and without paperclip

	n	mean	var	sd
with	6	220.1667	852.9667	29.20559
without	6	173.3333	2952.6667	54.33845

Since we are interested in the variance of within group, we have two scenario here. The best case scenario would be that the within group variance would be 852.9667 and the worst case scenario would be that it is 2952.6667.

We should conduct the power ANOVA test in both cases and see what n it gives us in each case. Only then can we proceed and decide how many blocks we shall have to make.

```
groupmeans <- c(220.1667, 173.3333)

#best case scenario where within var is 852.9667
best_case <- power.anova.test(groups = length(groupmeans),
                              between.var = var(groupmeans),
                              within.var = 853,
                              power = 0.8, sig.level = 0.05, n=NULL)

best_case_df <- data.frame(
  Groups = best_case$groups,
  Between_Var = var(groupmeans),
  Within_Var = 853,
  Power = best_case$power,
  Sig_Level = best_case$sig.level,
  Required_n = best_case$n
)

kable(best_case_df, caption = "Required n for the best case scenario")
```

Table 3: Required n for the best case scenario

Groups	Between_Var	Within_Var	Power	Sig_Level	Required_n
2	1096.684	853	0.8	0.05	7.204477

```
#Worst case scenario where within var is 2952.6667
worst_case <- power.anova.test(groups = length(groupmeans),
  between.var = var(groupmeans),
  within.var = 2953,
  power = 0.8, sig.level = 0.05, n=NULL)

worst_case_df <- data.frame(
  Groups = worst_case$groups,
  Between_var = var(groupmeans),
  Within_var = 2953,
  Power = worst_case$power,
  Sig_Level = worst_case$sig.level,
  Required_n = worst_case$n
)

kable(worst_case_df, caption = "Required n for the worst case scenario")
```

Table 4: Required n for the worst case scenario

Groups	Between_var	Within_var	Power	Sig_Level	Required_n
2	1096.684	2953	0.8	0.05	22.137

Since we are doing RCBD, each block would be getting assigned to each treatment. Going with the best case scenario seems more feasible. This means in total we would have $7 \times 4 = 28$ throws. Furthermore, this design already reduces variability because of the comparisons made between treatments is within each blocks.

Let's look at the graphical representation of the appropriate sample size when within group variances is changed.

```
library(ggplot2)

groupmeans <- c(220.1667, 173.3333)
between.var <- seq(500, 1200, by=30)

nvar <- nvar2 <- rep(NA, length(between.var))

for(i in 1:length(between.var)){
  nvar[i] <- power.anova.test(groups = 2,
    between.var = between.var[i],
    within.var = 853,
    power = 0.8, sig.level = 0.05, n = NULL)$n
}

for(i in 1:length(between.var)){
  nvar2[i] <- power.anova.test(groups = 2,
```

```

        between.var = between.var[i],
        within.var = 2953,
        power = 0.8, sig.level = 0.05, n = NULL)$n
}

sample_sizes_df <- data.frame(
  n = c(nvar, nvar2),
  between.var = rep(between.var, 2),
  within.var = c(rep("853", length(nvar)),
                 rep("2953", length(nvar2)))
)

sample_sizes_df$within.var <- factor(sample_sizes_df$within.var, levels = c("853", "2953"))

ggplot(data = sample_sizes_df, mapping = aes(x=between.var, y=n,
                                             group = within.var, color = within.var)) +
  geom_point() + geom_line() + theme_minimal()

```

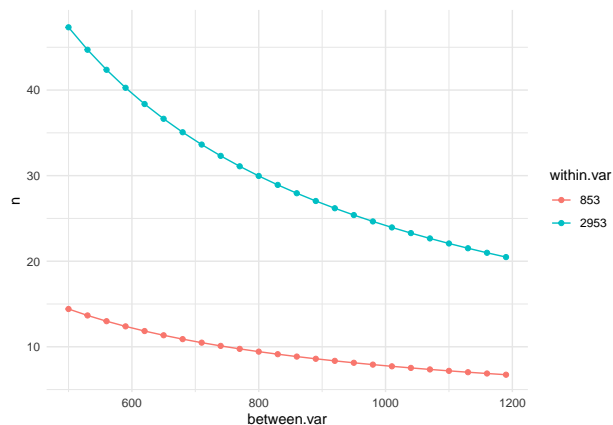


Figure 1: This graph shows the appropriate n when within group variances is changed. Since we will be working with the best case scenario we will be looking at the red curve.

Notice how 853 and 2953 are very further apart. To make it visually more appealing and seeing the middle group in between both of these, let's try to include a within group variance of 1903 which is the middle point of these two.

```

nvar <- nvar2 <- nvar3 <- rep(NA, length(between.var))

for(i in 1:length(between.var)){
  nvar[i] <- power.anova.test(groups = 3,
    between.var = between.var[i],
    within.var = 853,
    power = 0.8, sig.level = 0.05, n = NULL)$n
}

```



```

for(i in 1:length(between.var)){
  nvar2[i] <- power.anova.test(groups = 3,
                              between.var = between.var[i],
                              within.var = 1903,
                              power = 0.8, sig.level = 0.05, n = NULL)$n
}

for(i in 1:length(between.var)){
  nvar3[i] <- power.anova.test(groups = 3,
                              between.var = between.var[i],
                              within.var = 2953,
                              power = 0.8, sig.level = 0.05, n = NULL)$n
}

sample_sizes_df <- data.frame(
  n = c(nvar, nvar2, nvar3),
  between.var = rep(between.var, 3),
  within.var = c(rep("853", length(nvar)),
                 rep("1903", length(nvar2)),
                 rep("2953", length(nvar3)))
)

sample_sizes_df$within.var <- factor(sample_sizes_df$within.var, levels = c("853", "1903", "2953"))

ggplot(data = sample_sizes_df, mapping = aes(x=between.var, y=n,
                                             group = within.var, color = within.var)) +
  geom_point() + geom_line() + theme_minimal()

```

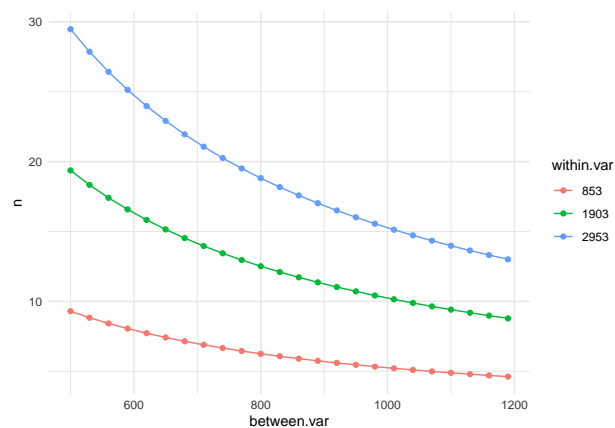


Figure 2: The graph shows the appropriate sample size when within group variance is 853, 1903, and 2953.

Since we are choosing to go with the best case scenario, and we know the between group variance is somewhere around 800, our n comes to be somewhere between 6-9. This is exactly what the power anova test gave me.

Both of the tests and the graphical representation gives us the same number for n which is 7. Thus, we proceed with making seven identical paper airplanes as our blocks.

Summary Statistics

After measuring the distances of paper airplane we should put it in a rectangular data frame where each row is an observation and each column is my variable.

The raw data after collecting looks like this data frame.

Table 5: Rectangular dataframe of distances each block (airplane) went with each treatment group.

Blocks/Treatments	control	nose	middle	rear
1	208.04	188.03	174.39	178.76
2	192.11	199.96	257.30	187.07
3	196.25	153.29	212.18	216.70
4	230.79	199.23	177.77	219.13
5	198.56	227.33	200.11	209.24
6	188.22	196.08	192.36	155.36
7	218.85	186.26	169.65	230.91

Before we turn this into a rectangular data frame, we should look at the summary statistics for our measured data and see whether there appears to be a difference in the means of different treatment groups.

Table 6: Summary statistics of the distance of seven airplanes across four different treatments.

	n	mean	sd	var
Control	7	204.6886	15.44488	238.5443
Paperclip on Nose	7	192.8829	22.06886	487.0344
Paperclip in middle	7	197.6800	30.34116	920.5858
Paperclip at rear	7	199.5957	26.75681	715.9268

Before we move on, we need to make the rectangular data frame so that each observation is in a row and every variable is my column.

Table 7: rectangular dataframe for the distances the airplane flies.

distance	treatment	block
208.04	control	airplane1
188.03	nose	airplane1
174.39	middle	airplane1
178.76	rear	airplane1
192.11	control	airplane2
199.96	nose	airplane2
257.30	middle	airplane2
187.07	rear	airplane2
196.25	control	airplane3
153.29	nose	airplane3
212.18	middle	airplane3
216.70	rear	airplane3
230.79	control	airplane4
199.23	nose	airplane4

distance	treatment	block
177.77	middle	airplane4
219.13	rear	airplane4
198.56	control	airplane5
227.33	nose	airplane5
200.11	middle	airplane5
209.24	rear	airplane5
188.22	control	airplane6
196.08	nose	airplane6
192.36	middle	airplane6
155.36	rear	airplane6
218.85	control	airplane7
186.26	nose	airplane7
169.65	middle	airplane7
230.91	rear	airplane7

Graphical Representation

As we can see there appears to be some difference between the means, however, there isn't concrete evidence yet to comment on whether these difference is statistically significant or not. To judge these better it would help to look at it in some graphical sense. It would give us some sense on the magnitude of difference between the mean distances across different treatment groups. Starting with a side-by-side boxplot.

```
ggplot(data = airplane_df, mapping = aes(x=treatment, y=distance)) +
  geom_boxplot() + theme_bw() +
  ggtitle("Distances travelled by the paper airplanes across different treatments.")
```

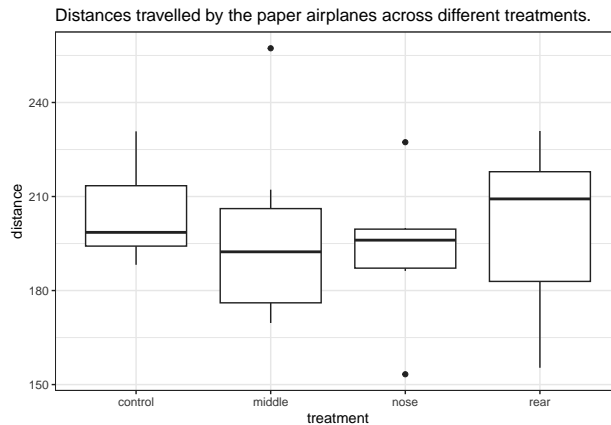


Figure 3: A side-by-side boxplot showing the difference between mean distances across different treatments. As we can see there seems to be some difference between treatments, however, it is not very noticeable. The highest median distance seems to be when the paperclip is attached to the rear end of the paper airplane.

The side-by-side boxplot provides a visual summary of the distribution of flight distances for each treatment group (control, nose, middle, rear). This graph allows us to assess the central tendency, spread, and presence of any outliers within each treatment group. From the boxplot, we can observe that most of the distances across the four treatment groups appear to be centered around similar values, with medians roughly between 190 and 210. However, there is some variation in the spread, with certain treatments (e.g., "middle") showing

a larger range of distances. While the boxplot does not explicitly indicate the significance of differences, it provides a preliminary idea that the treatments may not differ drastically in terms of central tendency.

To explore this further let's look at the distances traveled by each block across each treatment group.

```
airplane_df$treatment <- factor(airplane_df$treatment, levels = c("control", "nose", "middle", "rear"))

ggplot(data=airplane_df, aes(x=treatment, y=distance,
                             group = block, colour = block)) +
  geom_point() +
  geom_line() +
  scale_x_discrete(limits=c("control", "nose", "middle", "rear")) +
  theme_bw() +
  ggtitle("Distances travelled across different treatments")
```

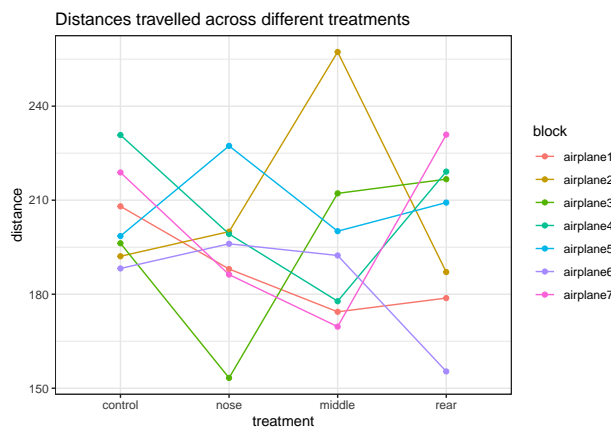


Figure 4: Different airplanes, blocks, distance across each treatment group. Although this graph looks a bit chaotic at first, we can follow the colors to see the distance traveled by each block. It, however, doesn't give us much information about the mean distances across the treatment groups.

The line graph, showing the distances traveled by each airplane (block) across the four treatments, offers an in-depth look at the interaction between blocks and treatments. Each colored line corresponds to an individual airplane, and the patterns help us understand variability both within and between the blocks. Some airplanes show a consistent trend across treatments, while others exhibit more fluctuation, suggesting that the effect of treatment might depend on the airplane. For example, an airplane with a steep rise or drop in the graph indicates that its performance varied significantly between treatments. On the other hand, flat or nearly parallel lines suggest minimal treatment effect. This graph reveals that while the treatment effect might be small overall, variability between blocks is substantial, emphasizing the importance of accounting for this variability in the RCBBD analysis.

Now let's look at a scatterplot to see the points across each treatment. It should tell us where most of the data is centered. Furthermore, let's look at the mean distance comparisons across the treatment groups.

```
ggplot(data = airplane_df, mapping = aes(x=treatment, y=distance)) +
  geom_point() + theme_bw() +
  ggtitle("Distances travelled by the paper airplanes across different treatments.")
```

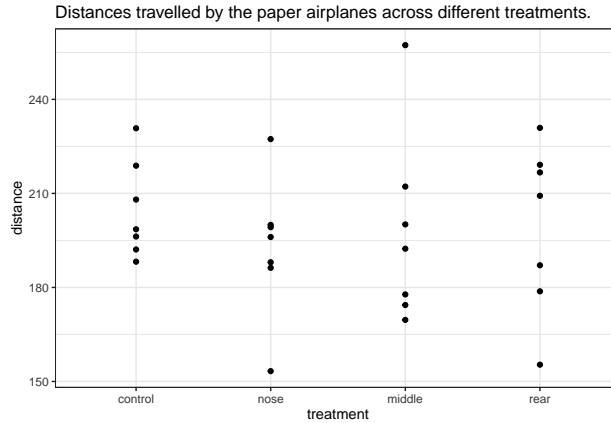


Figure 5: Across the four treatment groups, the data points seem to be mostly centered around 200-210. There doesn't seem to be too much of difference in the distance, however, we cannot concretely comment on the difference unless we perform the hypothesis test and construct the model to decide whether there is actually any difference between these or not.

The scatterplot focuses on the raw distances for all throws, plotted by treatment group, without connecting lines or grouping by block. This visualization is valuable for identifying patterns across treatments without the potential bias introduced by block-specific characteristics. From the scatterplot, it becomes clear that the distances for all treatment groups are mostly centered around the same range, roughly between 180 and 220. This suggests a lack of substantial differences between the groups. However, the scatterplot also highlights how much variability exists within each treatment group, as the spread of points is quite similar across all treatments. This reinforces the need for statistical hypothesis testing to determine whether these differences are significant or just due to random variability.

```
library(dplyr)

airplane_temp <- airplane_df %>%
  group_by(treatment) %>%
  summarize(mean_dist=mean(distance))

ggplot(data=airplane_temp, aes(y=mean_dist, x=treatment, group = 1)) +
  geom_point() + geom_line() +
  theme_bw() +
  ggtitle("Mean Distance Comparisons") + ylab("Mean Distance")
```

The mean distance graph provides a concise summary of the average performance of the four treatments, aggregating the data across all airplanes. This plot helps focus on the overarching trends rather than individual variability. From the graph, we observe that the mean distances for the four treatments are not as close as the other graphs showed us. This warrants for further investigation. Once we do the ANOVA testing, it would be clear whether these differences are statistically significant or not.

These graphs also provide valuable feedback about the experiment itself. For example, the large degree of overlap in distances across treatments, as shown in the scatterplot and boxplot, suggests that the treatment effect might be minimal relative to random variation. This aligns with the mean distance graph, where the averages across treatments are close. On the other hand, the line graph demonstrates significant between-block variability, which highlights the importance of including blocks in the analysis. This variability could stem from small differences in airplane construction, paperclip placement, or throwing technique. Overall, these graphs validate the use of a Randomized Complete Block Design (RCBD) to control for block-level variability while isolating treatment effects.

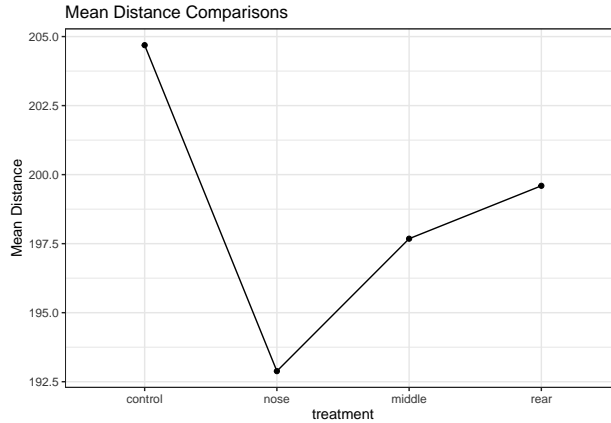


Figure 6: We can see that the mean of control is very different than the mean of distances when there is a paper clip attached to the nose of the paper airplane. It looks much lower than when the paper clip is attached to the middle and the rear. If anything there seems to be a big jump between mean distances between mean distance of control and the mean distance covered when the paper clip is attached to the nose.

The combined information from these graphs underscores the need for hypothesis testing to determine whether the observed differences between treatments are statistically significant. While the mean distance graph and scatterplot suggest that treatment effects are minimal, the line graph and boxplot remind us that variability within and between blocks could mask any true effects. By performing ANOVA, we can partition the total variability into components attributable to treatments and blocks, providing a rigorous statistical framework to assess the significance of treatment effects. These graphs collectively emphasize the complexity of the data and the importance of analyzing it carefully to draw valid conclusions.

ANOVA Testing

Hypotheses

The main objective is to determine whether the placement of the paperclip on the paper airplane significantly affect the distance flown.

Null Hypothesis: $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ There is no significant difference in the mean flight distances across the four treatment groups.

Alternate Hypothesis: $H_A : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$ At least one treatment group has a mean flight distance significantly different from the others.

Where:

- μ_1 : is the mean distance of the first treatment group
- μ_2 : is the mean distance of the second treatment group
- μ_3 : is the mean distance of the third treatment group
- μ_4 : is the mean distance of the fourth treatment group

Testing

```
airplane_model <- aov(distance~treatment+block, data = airplane_df)
kable(summary(airplane_model)[[1]], caption = "Analysis of the Airplane model.")
```

Table 8: Analysis of the Airplane model.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment	3	500.810	166.9367	0.2619533	0.8518579
block	6	2701.572	450.2620	0.7065411	0.6485316
Residuals	18	11470.976	637.2764	NA	NA

From the table above we can see that the p-value of 0.8518579 is greater than the alpha significance level of 0.05. Thus, we do not have sufficient evidence to reject the null hypothesis. The F value of 0.262 suggests that the variation between treatment groups is minimal compared to the residual (unexplained) variation. In other words, the placement of the paperclip does not appear to significantly affect the distance traveled by the paper airplanes.

Neither the treatment nor the block effects are statistically significant. This implies that neither the position of the paperclip nor the variability between airplanes meaningfully explains the differences in distances traveled.

Although the sample size was guided by the pilot study, it is important to note that The small sample size (n=7 blocks and n=28 total observations) may have limited the statistical power of the analysis. Low power increases the likelihood of failing to detect a real treatment effect, if one exists.

Type I and Type II Errors

The type I error means that we reject the null hypothesis when it is actually true. The probability of committing a type I error is alpha, which in this case is 0.05.

Although we do not know whether our null is actually true or not, our sample size calculations use the alpha significance level of 0.05. Therefore, as alpha increases the required n goes down. Which also correlates to if we want to correctly reject our null we want a higher probability of rejecting the null which is given by the alpha level.

The type II error is when we fail to reject the null when the null is actually false. Since our model's p-value is higher than the alpha significance level we could potentially be committing a type II error. During our sample size calculations we set our power to be 80%. The probability of committing a type II error is 1-power. With a low sample size (n=7) meaning only 28 observations, our study may lack the statistical power to detect meaningful differences between group.

Thus, this study risks concluding that treatment (paperclip positions) does not affect flight distance when, in reality, it does. This could lead to overlooking a potentially impactful factor in airplane performance.

Therefore, the high p-values and large residual variability suggest that the study is underpowered to detect small effects. While Type I error is not a concern here, the Type II error rate may be high due to insufficient data or a weak experimental design.

Contextualizing Results

The goal of your analysis was to investigate whether the placement of a paperclip on paper airplanes (treatments: control, nose, middle, rear) affects the distance they travel. After looking at the ANOVA results we can say that there is no statistical difference between the mean distances across the treatments and across the blocks. Although since we are in an RCBD, we don't care about the blocks.

The ANOVA results suggest that the placement of the paperclip does not significantly affect the distance traveled by the airplanes. The large p-value (0.852) indicates that any observed differences in distances across treatments are likely due to random variation, not a systematic effect of the treatment.

Model Assumptions

As mentioned above the three assumptions for ANOVA are normality of the residuals, homogeneity of variances, and independence.

Normality For checking the normality assumption, we will be making a histogram of the residuals, analyze the qqnorm plot, and perform the Shapiro-Wilk test.

```
hist(airplane_model$residuals)
```

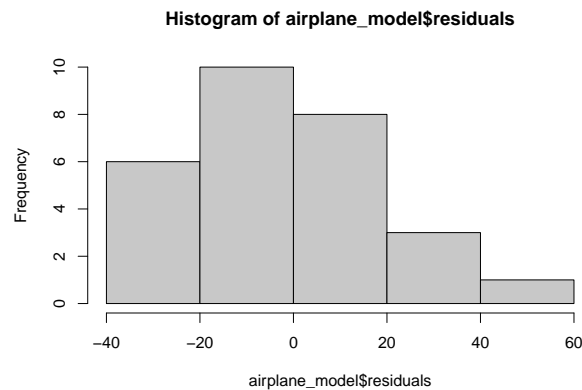


Figure 7: The histogram of residuals. This doesn't exactly show a normal bell-shaped curve, however, we cannot definitely comment on this unless the other factors for checking normality are looked at. The result of the histogram shows that the model residuals somewhat uniform and this is a good indication.

```
qqnorm(airplane_model$residuals)  
qqline(airplane_model$residuals)
```

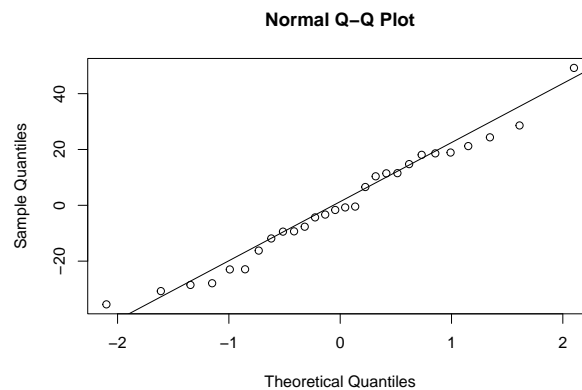


Figure 8: Looking at the qqnorm we can see that the residuals are following a spiral-like structure near the line. This is a good indication of normality. Just by looking at the qqnorm plot we can say that the residuals of our model follow a close to normal distribution. However, it is also important to look at the Shapiro-Wilk test to concretely comment on the normality of the residuals.


```

airplane_shapiro <- shapiro.test(airplane_model$residuals)

airplane_shapiro_df <- data.frame(
  Statistic = airplane_shapiro$statistic,
  Pvalue = airplane_shapiro$p.value
)

kable(airplane_shapiro_df, caption = "Shapiro Wilk Test Results for the residuals of the airplane model")

```

Table 9: Shapiro Wilk Test Results for the residuals of the airplane model.

	Statistic	Pvalue
W	0.9756617	0.7369568

Under the null hypothesis for the Shapiro-Wilk test it is assumed that the residuals follow a normal distribution. Therefore, we generally want a high p-value so that we can confirm that the null is true and that they follow a normal distribution.

As we can see from the table above, our p value of 0.7369568 is much greater than the alpha significance level of 0.05. Therefore, we do not have sufficient evidence to reject the null meaning that the residuals of the airplane model follow a normal distribution.

Homogeneity of Variance The variance across the treatment groups should be equal. We check this by plotting the residuals across the fitted values in the order of which the data was collected. We want to see equal variance across the whole range of fitted values. We do not want to observe a funnel like shape taking place, that would be a cause for concern.

```

plot(airplane_model$residuals~airplane_model$fitted.values)

```

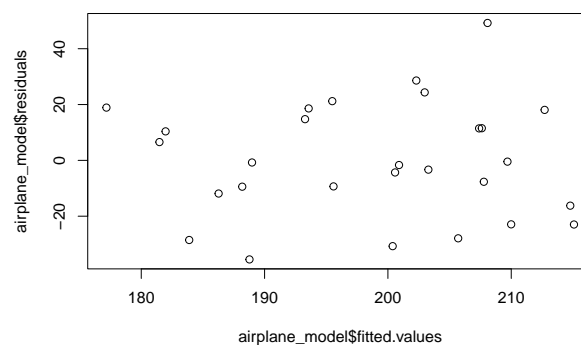


Figure 9: There doesn't appear to be a funnel-like shape which would've been a cause for concern. The variances looks pretty equal across the whole range of the fitted value, especially if we overlook the top most point which might be an outlier or an influential point.

We want to see equal variance across the whole range of fitted values. There seems to be more range in the residuals as we move to the right, however, it's not too concerning. There isn't a distinct funnel like shape. It's not overly dramatic but it's a little bit noticeable. Furthermore, if we overlook the top most point, the spread looks pretty equal. That point might be an outlier, however, we can't just ignore that point.

Structure of Data/Independence The observations should be independent to each other. In other words, there shouldn't be any structure in which the data was collected. We check this by plotting the residuals across the fitted values in the order of which the data was collected. We don't want to see any pattern in the data collected order in order to meet this assumption.

```
y <- 1:28  
plot(airplane_model$residuals~y)
```

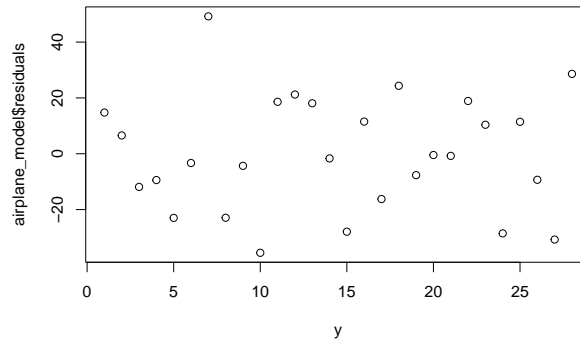


Figure 10: As we can see from the graph there doesn't appear to be any structure to the data in the order of which the data was collected. This shows that the trials or each observation was independent of each other.

Looking at all the graphs and tests we can say that our model is safe. It passes all the assumptions required to make analysis. There doesn't seem any deviations from the assumptions, although, we would like to see a more random scatter when assessing the homoscedasticity of variances. Normality and Independence are proven just by looking at the graph. Despite the histogram of the residuals not giving us much to work with the qqnorm and the Shapiro-Wilk test provides us with enough proof that the normality assumption is met.

Discussion

In this study we aimed to investigate the impact of paperclip placement on the distance flown by paper airplanes, using a Randomized Complete Block Design (RCBD) to account for potential block effects. The overarching goal of this study was to provide a basis for using this experiment in classrooms to make students learn by creative design building and experiments.

The analysis did not find any statistically significant differences in flight distances across the different paperclip placements (nose, middle, rear, or no paperclip). This suggests that, under the conditions of this experiment, the placement of the paperclip did not have a measurable effect on the aerodynamic performance of the paper airplanes.

The absence of significant findings has important implications. One possible explanation is that the paperclip's effect on the airplane's balance and aerodynamics might be too subtle to influence the overall flight distance in a meaningful way, at least within the constraints of this particular experiment. It is also possible that the experimental design or sample size did not capture the subtle effects of paperclip placement.

In this study there were four treatment levels tested: no paperclip (control), paperclip placed on the nose, paperclip placed in the middle, and paperclip placed on the rear of the airplane. Each airplane served as its own block to control for individual differences in performance, ensuring that any observed variations in flight distances could be attributed to the placement of the paperclip rather than inherent variability between the

airplanes. The analysis revealed a high p-value, indicating no statistically significant differences in the flight distances between the various treatments.

One limitation of this study is the relatively small sample size, which could affect the power of the statistical test and the ability to detect small but meaningful differences. While each airplane acted as its own block, a larger number of airplanes might have provided more robust results. Increasing the sample size could help reduce the variability and improve the chances of detecting a significant difference if one truly exists. Additionally, the distances were measured manually, and slight variations in the measurement process could have introduced error into the data. Future studies could use more precise measurement techniques or automated systems to minimize this potential source of bias.

The study's hypothesis was based on the assumption that small changes in the mass distribution, through the placement of the paperclip, would significantly impact the flight distance of the airplanes. However, the lack of a significant result may suggest that the effect of paperclip placement is not as pronounced as originally expected. It is possible that other factors, such as the airplane's aerodynamics, play a more substantial role in determining flight performance. In a [study by William Francis Ballhaur](#) weight distribution influences the aerodynamics of the airplane from sixteen different angles, depending on the single engine land type aircraft.

Integrating this type of design and study in the coursework can improve understanding on students about topics that are important to the modern world of technology. Although the topic is clearly laid out by several textbooks, this topic is difficult for students to grasp because of the vague idea. To allow students to better learn this material, a simple experiment using [paper airplanes](#) was developed that students can perform to generate data which they then analyze using dimensional analysis. This activity was pioneered with a fluid mechanics class at New Mexico Tech in the 2012 fall semester.

Despite the lack of significant findings, the study provides useful insights into the complexities of experimental design and statistical analysis. It highlights the importance of considering factors such as sample size, measurement precision, and environmental conditions when interpreting the results of an experiment. The study also underscores the challenge of detecting small effects in small-scale experimental settings, where inherent variability can mask subtle differences between treatments.

To make the model between we suggest using a larger sample size, conducting an experiment with even more controlled conditions, ensure as much uniformity as possible in the blocks, and having much more precise measurement practices. In conclusion, while the study did not find evidence of a significant effect of paperclip placement on paper airplane flight distances, it contributes to the broader understanding of experimental design and the potential challenges of detecting small effects in real-world settings. Future studies with larger sample sizes, more precise measurement techniques, and a broader set of variables could provide additional insights into the factors that influence paper airplane flight performance and challenge students when design-based learning is incorporated in classrooms making these concepts stick in their brains for longer.

In conclusion, this study explored the relationship between flying distance and weight distribution of paper airplanes under four different treatment groups. Despite the hypothesis that altering the position of the paperclip would significantly affect flight performance, the results showed no statistically significant differences in the distances flown across the treatment groups. Future research with larger sample sizes, improved measurement techniques, and a broader examination of contributing factors could offer deeper insights into the aerodynamic properties of paper airplanes and the impact of design modifications on their performance which then could be used to place a claim on the design of real airplanes.

References

1. Bluhm, William and Kronholm, Martha M. (1991) "Paper Airplanes: A Critical Thinking Model," Iowa Science Teachers Journal: Vol. 28: No. 1, Article 2. Available at: <https://scholarworks.uni.edu/istj/vol28/iss1/2>

2. Talukder, Tanisha (2024) "IMPACT OF AVIATION ON GLOBALIZATION," Anveshana's International Journal of Research in Engineering and Applied Sciences: Vol. 9, Issue. 6, Available at: <https://publications.anveshanaindia.com/wp-content/uploads/2024/08/IMPACT-OF-AVIATION-ON-GLOBALIZATION.pdf>
3. Hargather, M. J., Hussan, S., Jacomb-Hood, T. W., Francis, Z., Seneca, C., Quinlin, M., & Fernando, R. (2013, June). Fluid dynamics dimensional analysis take-home experiment using paper airplanes. In 2013 ASEE Annual Conference & Exposition (pp. 23-610).
4. Hmelo-Silver, C.E. Problem-Based Learning: What and How Do Students Learn?. Educational Psychology Review 16, 235–266 (2004). <https://doi.org/10.1023/B:EDPR.0000034022.16470.f3>
5. Tennekes, H. The Simple Science of Flight; MIT Press: Cambridge, Massachusetts, 2009.
6. Stephan, E. A.; Bowman, D. R.; Park, W. J.; Sill, B. L.; Ohland, M. W. Thinking like an engineer: An active learning approach; Pearson: New Jersey, 2011.
7. Ballhaus, William Francis (1947) Aerodynamic and Geometric Parameters Affecting Aircraft Weight. Dissertation (Ph.D.), California Institute of Technology. [doi:10.7907/T60T-F120](https://doi.org/10.7907/T60T-F120). <https://resolver.caltech.edu/CaltechETD:etd-10302003-152006>