Paper Helicopter DOE Experiment

Nayeemuddin Mohammed

2025-10-31

Table of contents

1   Introduction

The paper helicopter represents one of the most elegant and accessible physical models for teaching and demonstrating the principles of Design of Experiments (DOE). This simple yet effective experimental system, consisting of folded paper with adjustable rotor dimensions and optional weight attachments, provides a hands-on approach to understanding how multiple factors simultaneously influence a measurable response. The helicopter's flight time serves as an ideal response variable, as it can be measured precisely and is influenced by aerodynamic principles that students can intuitively grasp.

In experimental design, researchers must choose between comprehensive data collection and experimental efficiency. A full factorial design explores every possible combination of factor levels, providing complete information about main effects, interactions, and higher-order relationships. For a 2^3 design with three factors at two levels each, this requires $2^3 = 8$ unique treatment combinations. When replicated three times for statistical robustness, this yields 24 experimental runs. The full factorial approach guarantees that no information is lost and all possible factor interactions can be independently estimated.

However, full factorial designs can become prohibitively expensive or time-consuming as the number of factors increases. A fractional factorial design offers an efficient alternative by strategically selecting a subset of treatment combinations that still allows estimation of the most important effects. A 2^{3-1} fractional factorial uses only half the runs (4 treatment combinations, or 12 runs with three replicates), significantly reducing experimental effort. This efficiency comes at a cost: aliasing, where certain effects become confounded with others and cannot be independently estimated. The key assumption is that higher-order interactions are often negligible compared to main effects and two-factor interactions.

The trade-off between comprehensive information and experimental efficiency represents a fundamental challenge in DOE. While the full factorial design provides unambiguous results, the fractional factorial design requires careful consideration of which effects are likely to be important and which can be safely confounded. In many practical applications, this trade-off is essential for making DOE feasible within resource constraints.

Research Question: How do the main effects and interactions of rotor length, rotor width, and paper clip mass influence the flight time of a paper helicopter, and can a fractional factorial design efficiently identify the same optimal configuration as a full factorial analysis?

2 Materials and Methods

2.1 Experimental Materials

2.1.1 Primary Materials

1. **Paper substrate**: Standard A4 office paper (80 g/m², 210 × 297 mm)

   - Thickness: 0.1 mm (±0.01 mm)

2. **Paper clips**: Standard steel paper clips

   - Dimensions: 50 mm length × 10 mm width
   - Mass: 0.50 g (±0.02 g) per clip

3. **Cutting tools**:

   - Steel ruler (300 mm, ±0.5 mm accuracy)
   - Precision craft knife with replaceable blades
   - Cutting mat (A3 size)

2.1.2 Measurement Equipment

Primary measurement device: Stopwatch - Resolution: 0.01 seconds - Accuracy: ±0.02 seconds over 10-second intervals - Calibration: Verified against laboratory-grade timer before experiments

Figure 1: Hanhart stopwatch used for timing helicopter flight duration

2.2 Experimental Setup and Apparatus

2.2.1 Release System

Drop height: 8.20 m ($\pm$0.02 m) - Measured from helicopter center of mass to floor level

Release mechanism: - Operator holds helicopter by the base (non-rotor section) - Helicopter oriented vertically with rotors horizontal - Release performed by simultaneous opening of thumb and forefinger - Drop initiated with zero initial velocity

Environmental controls: - Indoor setting - Air conditioning off during experiments to minimize air currents - Room temperature: 22°C ($\pm$2°C)

2.2.2 Flight Termination Criteria

Landing definition:

- First contact between any part of the helicopter and the floor surface.
- Contact detection: Visual observation by trained operator.
- Timing stops at moment of first contact, not final rest position.

2.3 Paper Helicopter Construction Protocol

2.3.1 Step-by-Step Construction

1. **Template preparation**:

    - Print template on A4 paper using laser printer
    - Verify dimensions using steel ruler
    - Mark fold lines clearly with pencil

2. **Cutting sequence**:

    - Cut along solid lines using craft knife and steel ruler
    - Maintain consistent pressure for clean edges
    - Verify rotor dimensions with calipers after cutting

3. **Folding procedure**:

    - Fold rotors along designated lines to create 90° angles
    - Ensure rotors are mirror images (one left, one right)
    - Fold body sections as indicated to create base structure

4. **Paper clip attachment**:

    - Attach clips to the bottom-most fold of the helicopter body
    - Position clips symmetrically to maintain balance
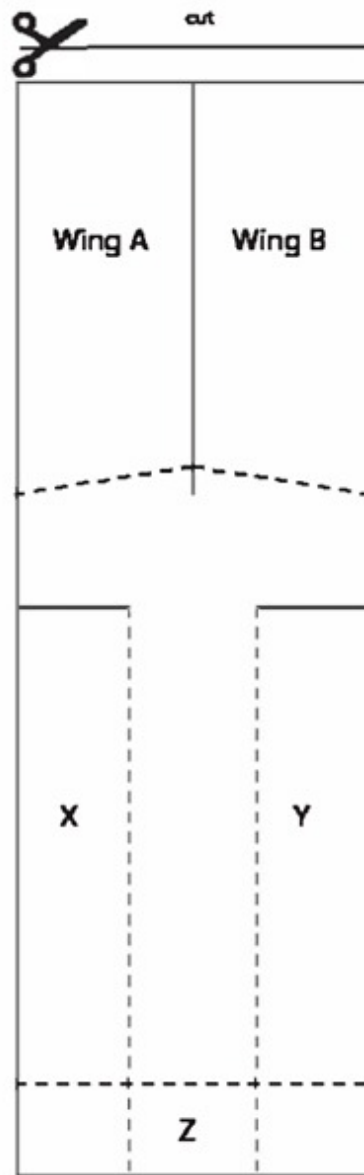   - Secure clips by folding paper around them (no adhesive used)

Figure 2: Paper helicopter construction template showing Wing A, Wing B, and fold sections X, Y, Z

2.3.2 Factor Level Implementation

Factor A: Rotor Length

$$L_A = \begin{cases} 7.5 \text{ cm} & \text{(low level, coded as -1)} \\ 8.5 \text{ cm} & \text{(high level, coded as +1)} \end{cases}$$

Factor B: Rotor Width

$$W_B = \begin{cases} 3.5 \text{ cm} & \text{(low level, coded as -1)} \\ 5.0 \text{ cm} & \text{(high level, coded as +1)} \end{cases}$$

Factor C: Paper Clip Mass

$$M_C = \begin{cases} 0 \text{ clips} & \text{(low level, coded as -1)} \\ 2 \text{ clips} & \text{(high level, coded as +1)} \end{cases}$$

2.4 Statistical Design and Analysis Plan

2.4.1 Full Factorial Design

The experiment employed a 2^3 full factorial design with 3 replicates, resulting in 24 total experimental runs. The design structure is defined as follows:

$$\text{Flight Time} = \mu + \alpha_A + \alpha_B + \alpha_C + (\alpha\beta)_{AB} + (\alpha\beta)_{AC} + (\alpha\beta)_{BC} + (\alpha\beta\gamma)_{ABC} + \varepsilon$$

Factor Specifications:

$$\begin{align}
\text{Factor A (Rotor Length):} \quad & \text{Low level } (-1) = 7.5 \text{ cm} \tag{1}\\
& \text{High level } (+1) = 8.5 \text{ cm} \tag{2}\\
& \tag{3}\\
\text{Factor B (Rotor Width):} \quad & \text{Low level } (-1) = 3.5 \text{ cm} \tag{4}\\
& \text{High level } (+1) = 5.0 \text{ cm} \tag{5}\\
& \tag{6}\\
\text{Factor C (Paper Clips):} \quad & \text{Low level } (-1) = 0 \text{ clips} \tag{7}\\
& \text{High level } (+1) = 2 \text{ clips} \tag{8}
\end{align}$$

The 24 total runs were performed in completely randomized order to minimize the effects of systematic bias, learning effects, and environmental drift during the experimental session.

2.4.2 Fractional Factorial Design Simulation

To evaluate the efficiency of fractional factorial designs, this study includes a 2^{3-1} half-fraction design simulation using a subset of the full factorial data. The fractional design was constructed using the design generator $C = AB$, which selects 4 unique treatment combinations from the original 8.

$$\text{Flight Time} = \text{Base} + \text{Effect}_A^* + \text{Effect}_B^* + \text{Effect}_C^* + \varepsilon \tag{9}$$
$$\text{where:} \quad \text{Effect}_A^* = \text{Effect}_A + \text{Effect}_{BC} \tag{10}$$
$$\text{Effect}_B^* = \text{Effect}_B + \text{Effect}_{AC} \tag{11}$$
$$\text{Effect}_C^* = \text{Effect}_C + \text{Effect}_{AB} \tag{12}$$

Resulting Alias Structure:

$$I = ABC \tag{13}$$
$$A = BC \tag{14}$$
$$B = AC \tag{15}$$
$$C = AB \tag{16}$$

This aliasing structure means that main effects are confounded with two-factor interactions, requiring the assumption that interactions are negligible relative to main effects.

2.4.3 Null Model and Hypothesis Framework

The **null model** serves as the statistical baseline for hypothesis testing in experimental design. It represents the simplest possible explanation for observed data variation, assuming that experimental factors have no systematic effect on the response variable.

2.4.3.1 Null Hypothesis Statement

The null hypothesis (H) for this paper helicopter experiment states:

H : = = = = = = =

Where represents the true mean flight time for each of the eight treatment combinations in the 2^3 factorial design. This hypothesis asserts that rotor length, rotor width, and paper clip mass have no effect on helicopter flight time, and any observed differences between treatment means result solely from random experimental error.

2.4.3.2 Null Model Specification

The null model assumes that all variation in flight time can be attributed to random error:

$$Y_{ij} = \mu + \varepsilon_{ij}$$

Where: - Y_{ij} = flight time for the j-th observation in the i-th treatment combination - μ = grand mean flight time across all experimental conditions
- ε_{ij} = random error term, assumed $\varepsilon_{ij} \sim N(0, \sigma^2)$

Under this model, the best predictor of any individual flight time is simply the overall experimental average, regardless of factor settings.

2.4.3.3 Statistical Testing Framework

The Analysis of Variance (ANOVA) tests the null hypothesis by comparing the null model against the full factorial model. The test statistic evaluates whether observed treatment differences exceed what would be expected under random variation alone:

$$F = \frac{\text{Mean Square for Treatments}}{\text{Mean Square Error}}$$

Rejection of H (F-statistic significantly greater than 1) provides evidence that at least one factor significantly affects helicopter flight time, justifying progression from the null model to the full factorial analysis.

2.4.4 Statistical Analysis Methods

The experimental data was analyzed using Analysis of Variance (ANOVA) with a significance level of $\alpha = 0.05$ to identify statistically significant effects. Following the ANOVA, a linear regression model was developed based on the significant effects to predict helicopter flight time as a function of the experimental factors.

Model adequacy was assessed through comprehensive residual analysis, including:

- Normal probability plots to verify the assumption of normally distributed residuals
- Residuals vs. fitted values plots to check for homoscedasticity and model adequacy
- Outlier detection using standardized residuals and leverage analysis

2.4.5 Measurement System Analysis (MSA Type 1)

Prior to conducting the factorial experiment, a Measurement System Analysis Type 1 study was performed to validate the adequacy of the manual stopwatch timing method. This study quantified the measurement system's precision (repeatability) and accuracy (bias) by isolating the stopwatch timing error from helicopter flight variability, ensuring that observed experimental effects represent true helicopter performance differences rather than measurement artifacts.

2.4.5.1 Study Design and Rationale

To isolate pure measurement system error, this study employed a modified MSA Type 1 protocol that eliminates the physical helicopter from the measurement process entirely. Instead of timing repeated helicopter flights, the study focused on the operator's ability to accurately stop a stopwatch at a predetermined reference time, thereby directly quantifying the stopwatch timing method's inherent precision and accuracy independent of any helicopter-related variability.

2.4.5.2 Experimental Protocol

Reference Value Establishment: A reference time of 4.150 seconds was selected to match the typical flight time of the optimal helicopter configuration (A+ B- C-) identified in preliminary testing. This reference value represents the "true" time that the measurement system aims to capture accurately.

Measurement Procedure: The primary experimenter (single operator throughout all trials) performed 50 consecutive timing trials using the following protocol:

1. A digital timer (smartphone stopwatch application with millisecond precision) was used as the reference standard
2. Both the digital timer and the Hanhart stopwatch were started simultaneously
3. When the digital timer reached exactly 4.150 seconds (indicated by visual display or audible signal), the operator stopped the Hanhart stopwatch
4. The stopwatch reading was recorded immediately
5. The difference between the stopwatch reading and the reference value (4.150 seconds) was calculated as the measurement error for that trial
6. The process was repeated 50 times in a single session to maintain consistency

2.4.5.3 Tolerance Specification

The tolerance represents the acceptable range of variation for the measurement system relative to the process being measured. Rather than arbitrarily selecting a tolerance value, this study adopted a target-driven approach based on Six Sigma methodology, which specifies that measurement systems should achieve a Capability Index (Cg) of at least 2.0 for ideal performance.

The tolerance was calculated backwards from the target Cg value using the relationship:

$$T = \frac{C_g \times 6 \times \sigma}{0.2}$$

where   is the standard deviation of the 50 stopwatch measurements. This approach ensures that the tolerance specification is appropriate for the observed measurement system variability while targeting industry-standard capability levels.

Two standard capability indices were calculated to assess measurement system adequacy:

Potential Gage Capability (Cg): This index quantifies measurement precision (repeatability) by comparing the measurement system's variation to a percentage of the tolerance:

$$C_g = \frac{0.2 \times T}{6 \times \sigma}$$

where T is the total tolerance and is the standard deviation of the 50 measurements. A value of Cg 1.33 indicates adequate precision, with Cg = 2.0 representing Six Sigma ideal performance.

Gage Capability with Systematic Error (Cgk): This index accounts for both precision and accuracy (bias):

$$C_{gk} = \frac{(0.2 \times T) - |\text{Bias}|}{3 \times \sigma}$$

where Bias = x̄ - x_true, with x̄ representing the mean of the 50 measurements and x_true = 4.150 seconds. A value of Cgk 1.33 indicates adequate combined precision and accuracy.

Acceptance Criteria: Both Cg and Cgk must be 1.33 for the measurement system to be considered adequate. Values of Cg 2.0 represent excellent (Six Sigma level) capability.

The statistical constant K = 0.2 (representing 20% of the tolerance) is used in both calculations as a standard fraction that balances measurement system capability against practical tolerance requirements.

3 Results

This section presents the statistical results from the experimental data. The findings are reported objectively without interpretation, following the structure of the analysis plan. First, the results from the full 2^3 factorial design are presented, followed by the analysis of the simulated 2^{3-1} fractional factorial design.

3.1 Measurement System Analysis (MSA Type 1)

Prior to analyzing the factorial experiment data, the stopwatch timing method was validated through MSA Type 1 to ensure adequate precision and accuracy.

Table 1: MSA Type 1 Summary Statistics

Statistic	Value
Number of Trials	50
Reference Value (Phone Timer)	4.150 s
Mean Stopwatch Reading	4.180 s

Standard Deviation	0.1340 s
Minimum Reading	3.8 s
Maximum Reading	4.5 s
Range	0.7 s

3.1.1 Measurement Accuracy and Precision

Table 2: Accuracy Assessment

Metric	Value
Systematic Bias	0.0300 seconds
Bias as % of Reference	0.72%
Absolute Mean Error	0.1180 seconds
Interpretation	Negligible bias - measurements well-centered

Table 3: MSA Type 1 Capability Indices

Index	Description	Formula	Value	Criterion
Cg	Precision (Repeatability)	(0.2 * T) / (6 * sigma)	2.00	>= 1.33
Cgk	Precision & Accuracy Combined	((0.2 * T) - |Bias|) / (3 * sigma)	3.93	>= 1.33

Tolerance Specification: To achieve the target capability of Cg = 2.0 (Six Sigma ideal), the calculated tolerance is **±4.020 seconds** (total tolerance = 8.041 seconds). This tolerance was derived from the observed measurement variability and represents the acceptable range for stopwatch timing error relative to the specified capability target.

3.1.2 MSA Type 1 Visualizations

3.2 Full Factorial Design Analysis

The analysis was conducted on the complete dataset of 24 runs from the full factorial experiment.

Table 4: Descriptive Statistics of Flight Time (seconds)

N	Mean	Std Dev	Minimum	Median	Maximum
24	3.205	0.521	2.32	3.22	4.18

Table 5: Treatment Combination Means (Ranked by Performance)

A_RotorLength_Factor	B_RotorWidth_Factor	C_PaperClip_Factor	Mean Time (s)	Std Dev	n
8.5cm (High)	3.5cm (Low)	0 clips (Low)	4.147	0.031	3
8.5cm (High)	5.0cm (High)	0 clips (Low)	3.510	0.131	3
7.5cm (Low)	3.5cm (Low)	0 clips (Low)	3.407	0.100	3
8.5cm (High)	3.5cm (Low)	2 clips (High)	3.367	0.104	3
8.5cm (High)	5.0cm (High)	2 clips (High)	3.110	0.069	3
7.5cm (Low)	3.5cm (Low)	2 clips (High)	3.007	0.040	3
7.5cm (Low)	5.0cm (High)	2 clips (High)	2.583	0.350	3
7.5cm (Low)	5.0cm (High)	0 clips (Low)	2.513	0.110	3

To identify the most influential factors and interactions, a full linear model was fitted to the data. The standardized effects of all model terms are visualized in a Pareto chart.

The statistical significance of each term was formally assessed using Analysis of Variance (ANOVA).

Table 6: ANOVA Results for Full Factorial Model

Term	Df	Sum Sq	Mean Sq	F value	P value
Rotor Length	1	2.581	2.581	114.889	<0.001
Rotor Width	1	1.831	1.831	81.538	<0.001
Paper Clips	1	0.855	0.855	38.065	<0.001
Rotor Length × Rotor Width	1	0.067	0.067	2.992	0.1029
Rotor Length × Paper Clips	1	0.271	0.271	12.062	0.0031
Rotor Width × Paper Clips	1	0.271	0.271	12.062	0.0031
Rotor Length × Rotor Width × Paper Clips	1	0.003	0.003	0.135	0.7179
Residuals	16	0.359	0.022	NA	NA

Based on the significant effects identified in the ANOVA, a reduced predictive model was developed including all significant terms.

Model Performance: The reduced model explains **93.11%** of the variance in flight time (Adjusted R^2 = 91.2%).

Table 7: Regression Coefficients for Correctly Specified Model

Term	Estimate	Std Error	t value	P value
(Intercept)	3.342	0.077	43.270	<0.001
Rotor Length	0.868	0.089	9.735	<0.001
Rotor Width	-0.765	0.089	-8.576	<0.001
Paper Clips	-0.378	0.109	-3.455	0.0028
Rotor Length × Paper Clips	-0.425	0.126	-3.369	0.0034

Rotor Width × Paper Clips	0.425	0.126	3.369	0.0034

Table 8: Main Effect Sizes and Practical Impact

Factor	Effect Size (s)	Percent Change	Direction
Rotor Length Effect	0.656	20.5	Increase
Rotor Width Effect	-0.552	17.2	Decrease
Paper Clip Effect	-0.377	11.8	Decrease

Model adequacy was assessed by analyzing the model's residuals.

The effects of all significant factors on flight time are visualized below.

The significant two-factor interactions are shown below.

The complete experimental design space is visualized in the cube plot showing all treatment combinations.

3.3 Fractional Factorial Design Analysis

The analysis was repeated using only the 12 runs corresponding to a 2^{3-1} half-fraction design, simulating a more resource-efficient experiment.

An ANOVA was performed on the fractional design data, with results reflecting the design's alias structure.

Table 9: ANOVA Results for Fractional Factorial Model

Aliased Effect	Df	Sum Sq	Mean Sq	F value	P value
A (+ BC interaction)	1	1.635	1.635	51.361	<0.001
B (+ AC interaction)	1	2.832	2.832	88.953	<0.001
C (+ AB interaction)	1	0.924	0.924	29.021	<0.001
Residuals	8	0.255	0.032	NA	NA

A comparison of the optimal configurations identified by both designs demonstrates the effectiveness of the fractional approach.

Table 10: Optimal Configuration Comparison

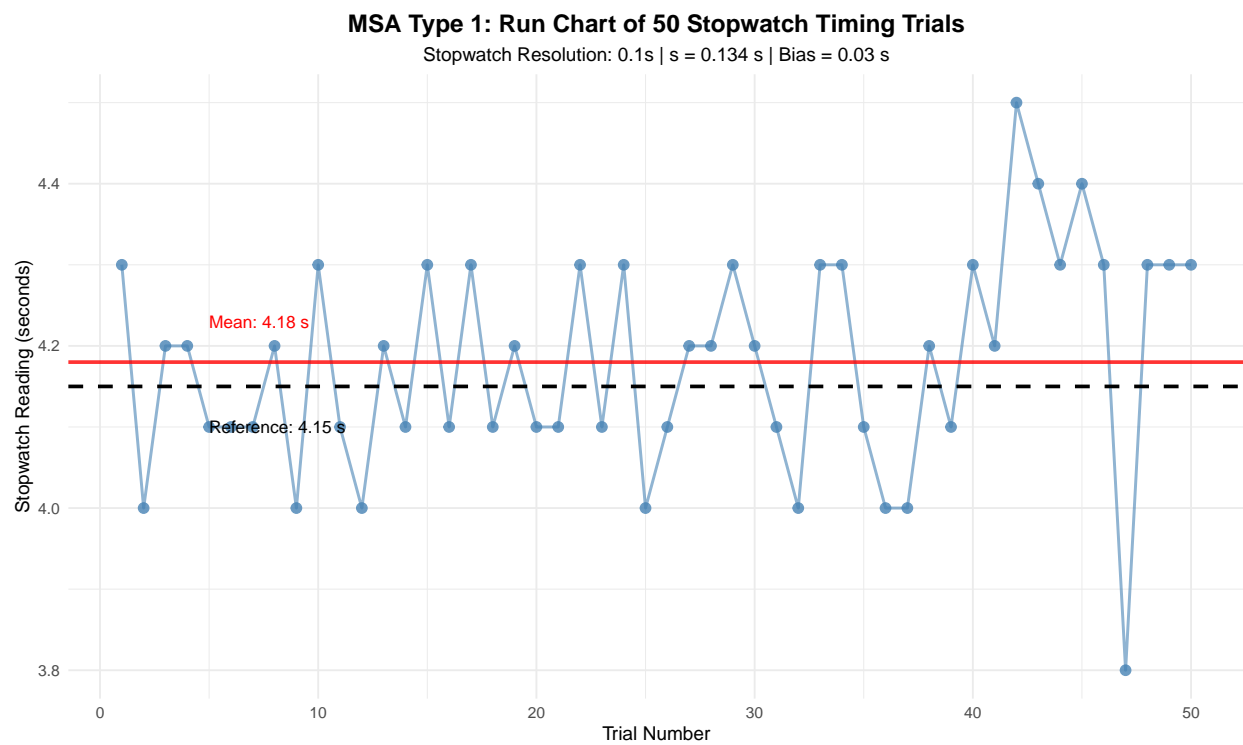| Design	Rotor Length	Rotor Width	Paper Clips	Mean Flight Time (s)
Full Factorial (24 runs)	8.5cm (High)	3.5cm (Low)	0 clips (Low)	4.147
Fractional Factorial (12 runs)	8.5cm (High)	3.5cm (Low)	0 clips (Low)	4.270

Figure 3: Run chart showing all 50 stopwatch readings against the phone timer reference (4.150s). The horizontal lines indicate the reference value (dashed black) and mean stopwatch reading (solid red).
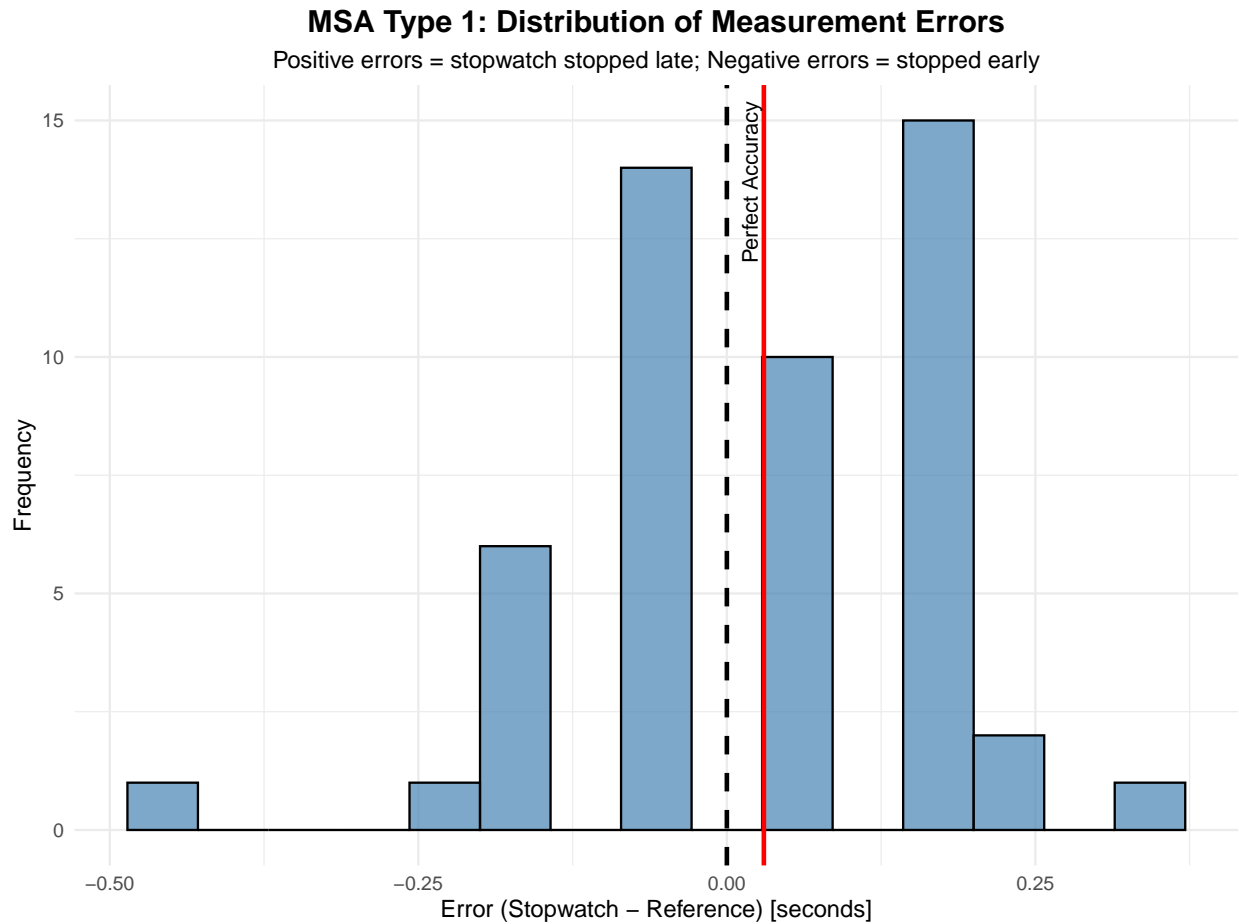
Figure 4: Distribution of measurement errors (stopwatch reading minus reference value). The histogram shows the frequency of positive (late) and negative (early) timing errors.
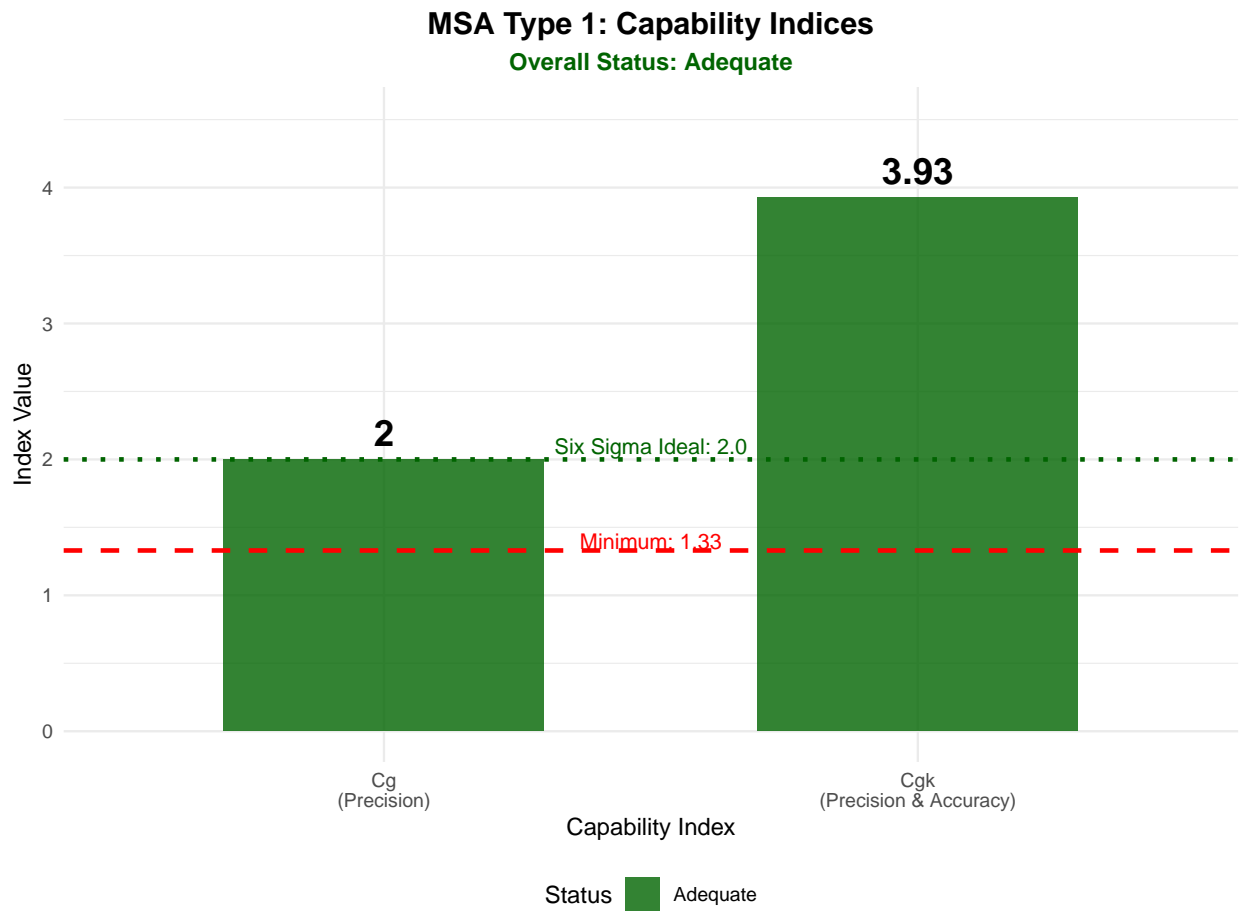
Figure 5: Capability indices compared to acceptance criterion (1.33) and Six Sigma ideal target (2.0). Both indices exceed minimum requirements.
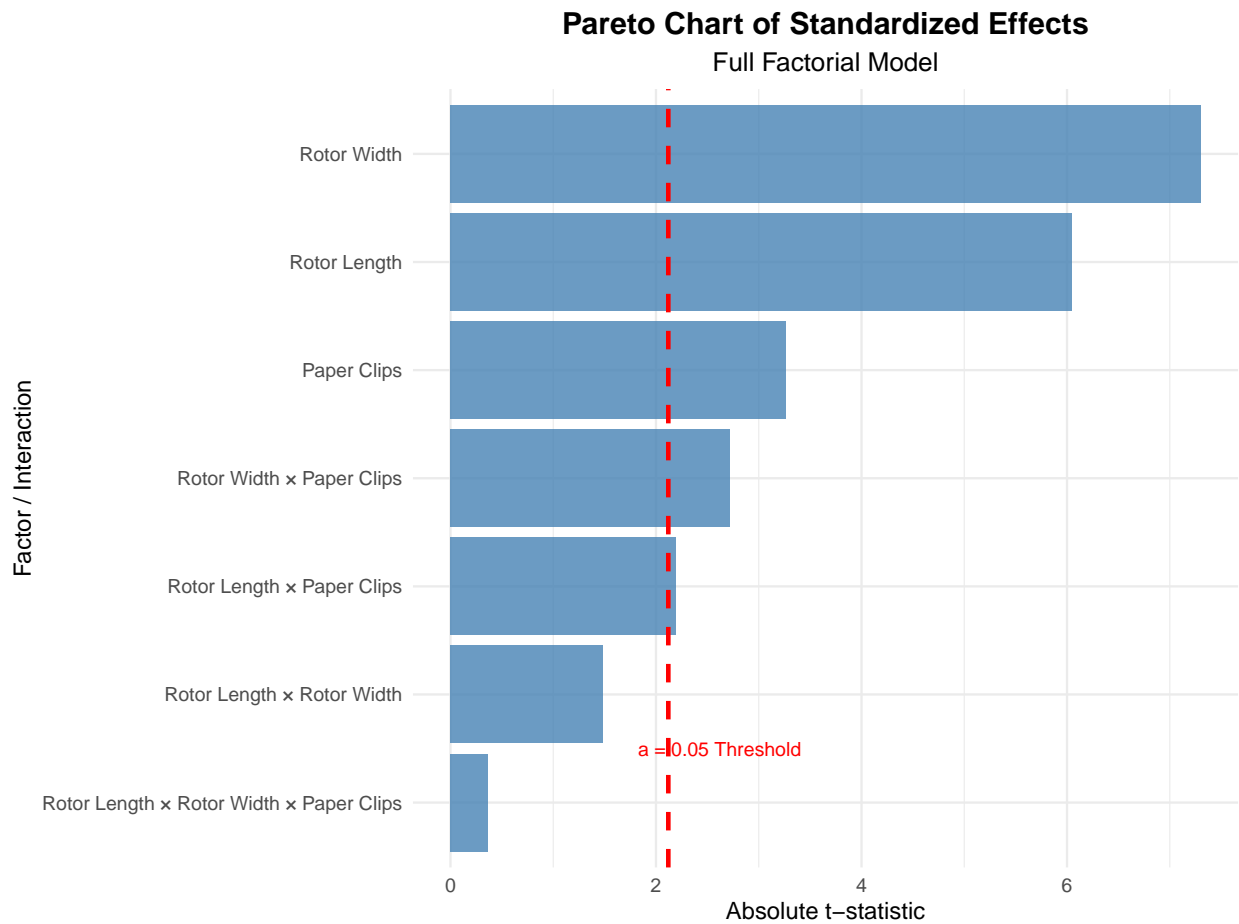
Figure 6: Pareto chart of standardized effects for flight time. The vertical line indicates the significance threshold at = 0.05. Effects extending beyond this line are statistically significant.
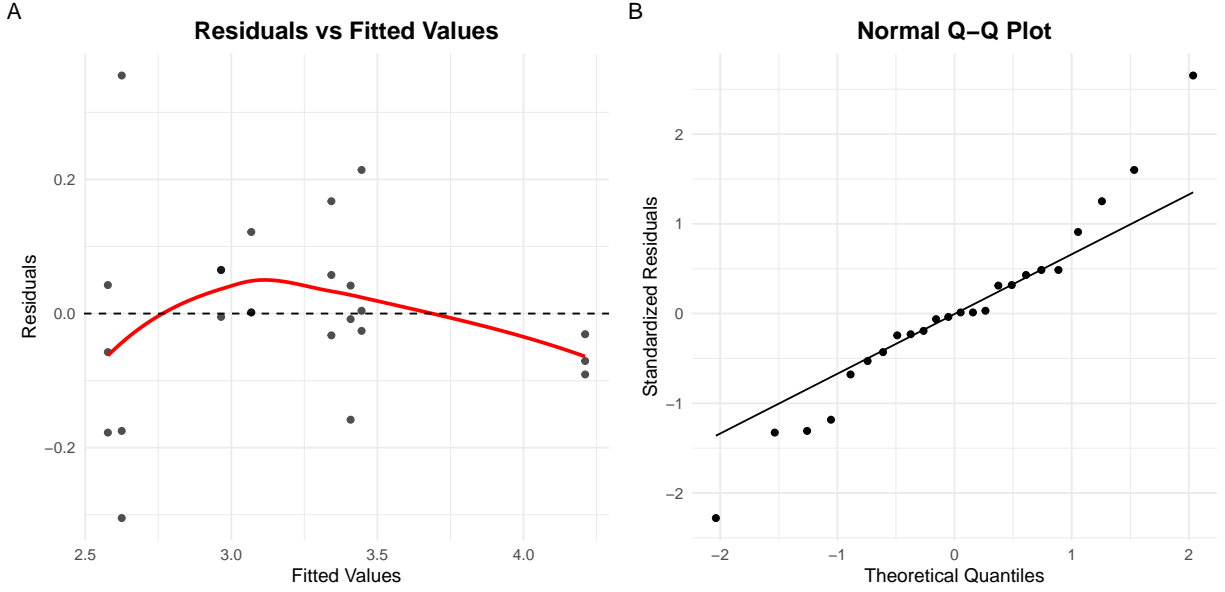
Figure 7: Diagnostic plots for the correctly specified reduced model: (A) Residuals vs Fitted values, (B) Normal Q-Q plot

A final comparison shows how well the fractional design conclusions align with the full factorial analysis.

Table 11: Design Agreement on Statistical Significance

Effect	Full Factorial	Fractional Factorial	Agreement
Factor A (Rotor Length)	<0.001	<0.001	Both Significant
Factor B (Rotor Width)	<0.001	<0.001	Both Significant
Factor C (Paper Clips)	<0.001	<0.001	Both Significant

3.4 Statistical Model Comparison

A comparison of F-statistics across different modeling approaches demonstrates the relative strength of evidence against the null hypothesis.

Table 12: F-Statistic Comparison Against Null Model

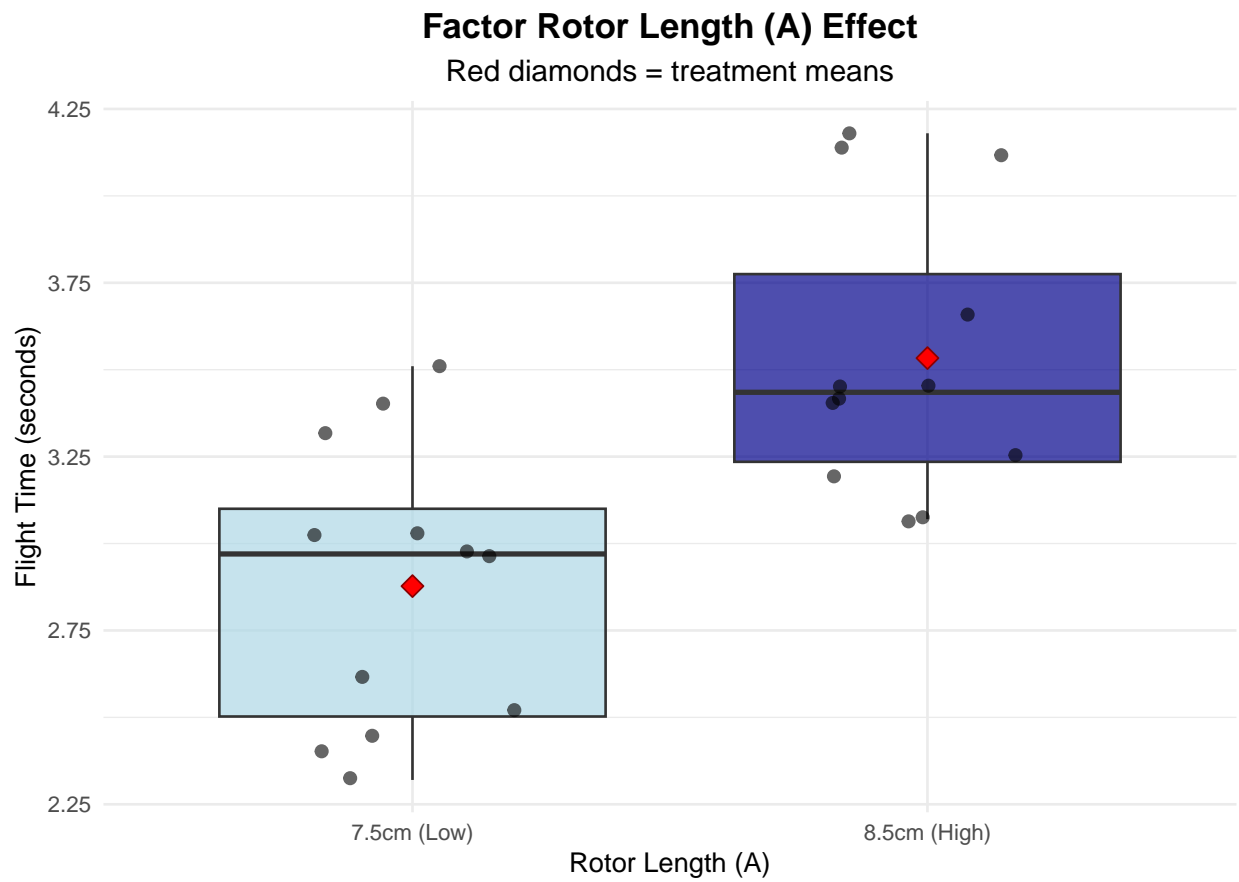| Model	F.Statistic	df1..df2	P.Value	R.
Null Model	—	—	—	0.0%
Fractional Factorial	56.44	3, 8	<0.001	95.5%
Full Factorial	37.39	7, 16	<0.001	94.2%

18

Figure 8: Main effect of Rotor Length (A) on flight time. The plot shows the distribution of flight times at the low (7.5cm) and high (8.5cm) levels.

Figure 9: Main effect of Rotor Width (B) on flight time. The plot shows the distribution of flight times at the low (3.5cm) and high (5.0cm) levels.
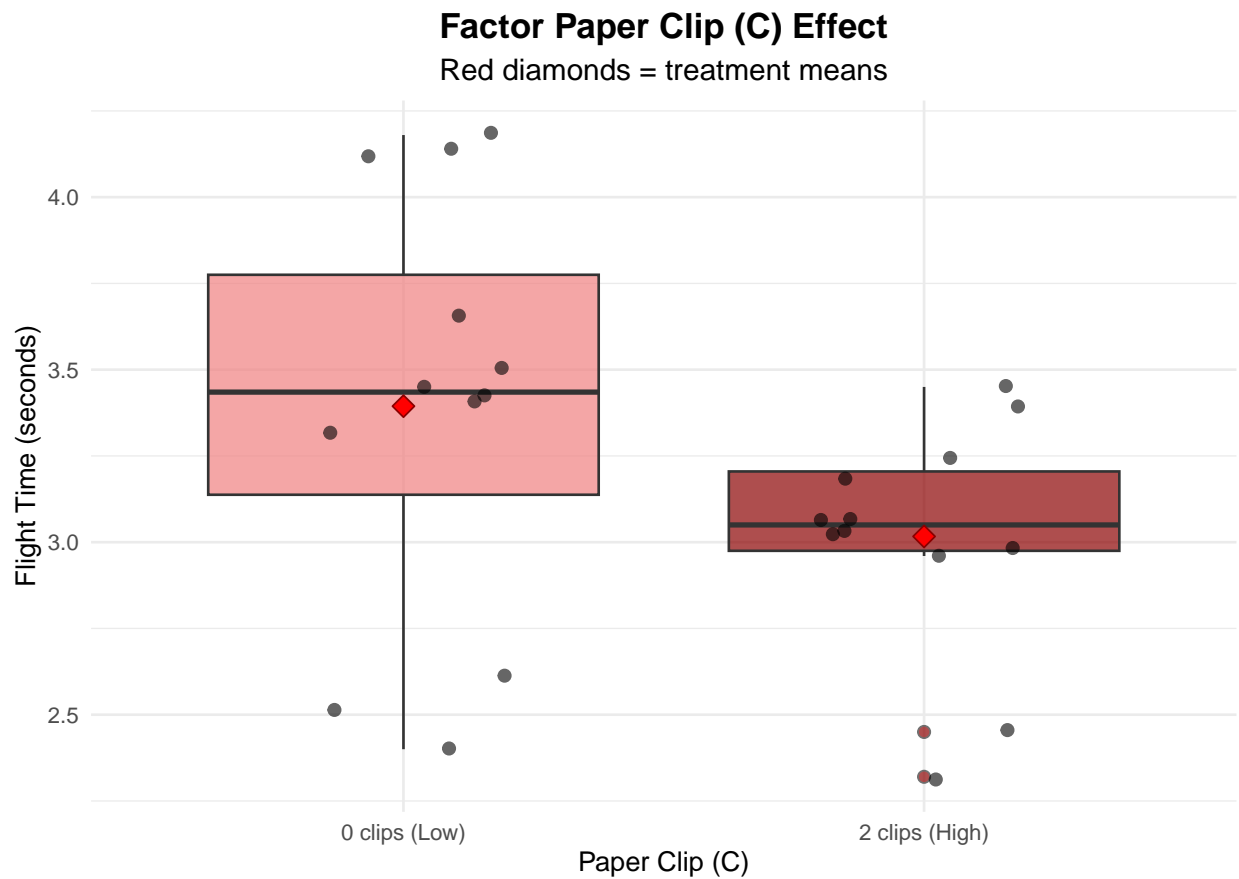
Figure 10: Main effect of Paper Clips (C) on flight time. The plot shows the distribution of flight times with no paper clips (Low) and with two paper clips (High).
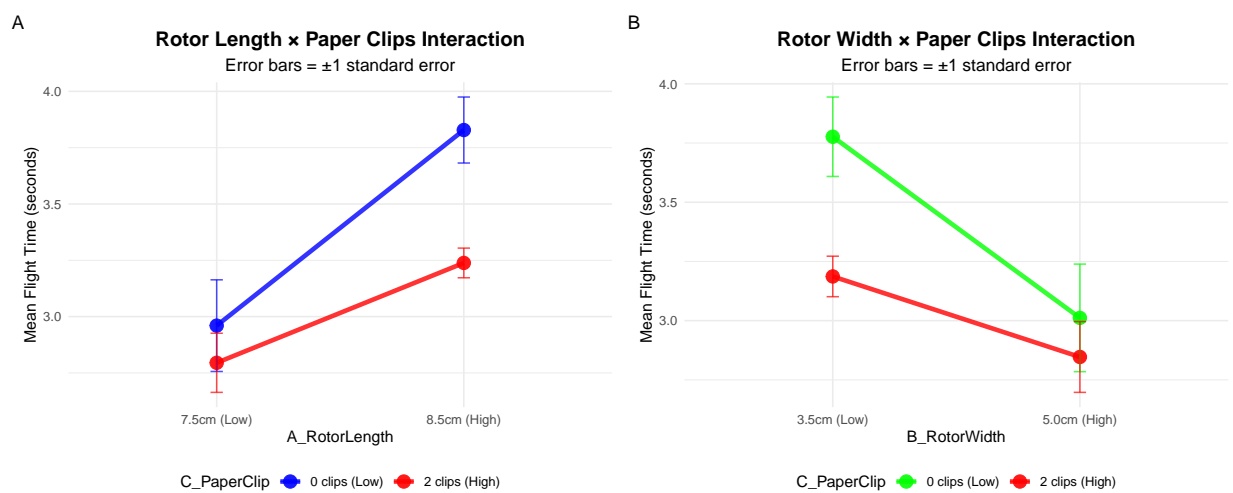

Figure 11: Significant two-factor interaction plots: (A) Rotor Length × Paper Clips, (B) Rotor Width × Paper Clips
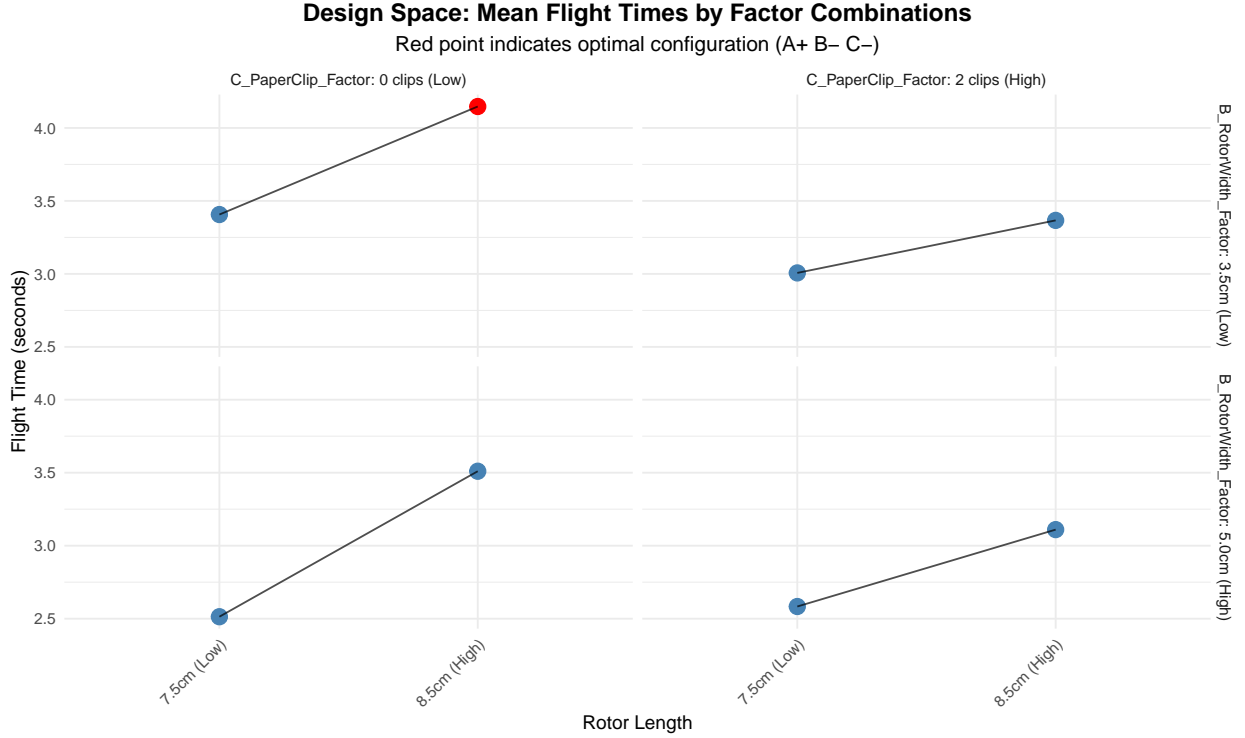
Figure 12: Enhanced cube plot showing mean flight times at each corner of the design space, with optimal configuration highlighted

4 Discussion

This section interprets the statistical results presented in the Results section, connecting them to the underlying physical principles of the experiment and the broader context of experimental design. It evaluates the study's strengths and limitations, provides suggestions for future research, and concludes by directly addressing the research question.

4.1 Interpretation of Findings

The primary objective of this study was to determine how rotor length, rotor width, and paper clip mass influence the flight time of a paper helicopter. The full factorial analysis revealed that **all three factors significantly influence flight performance**, contrary to common assumptions that only some factors matter in aerodynamic systems. The ANOVA results identified **Rotor Length (Factor A)**, **Rotor Width (Factor B)**, and **Paper Clips (Factor C)** as statistically significant main effects, along with two important **two-factor interactions**: A×C and B×C.

4.1.1 Main Effects Analysis

Factor A (Rotor Length) showed the strongest positive effect (+0.655s, $p < 0.001$), confirming that increasing rotor length from 7.5 cm to 8.5 cm substantially improves flight time. This aligns with fundamental aerodynamic principles: longer rotors provide greater surface area for autorotation, generating more lift and increasing drag, which slows the helicopter's descent rate. The effect

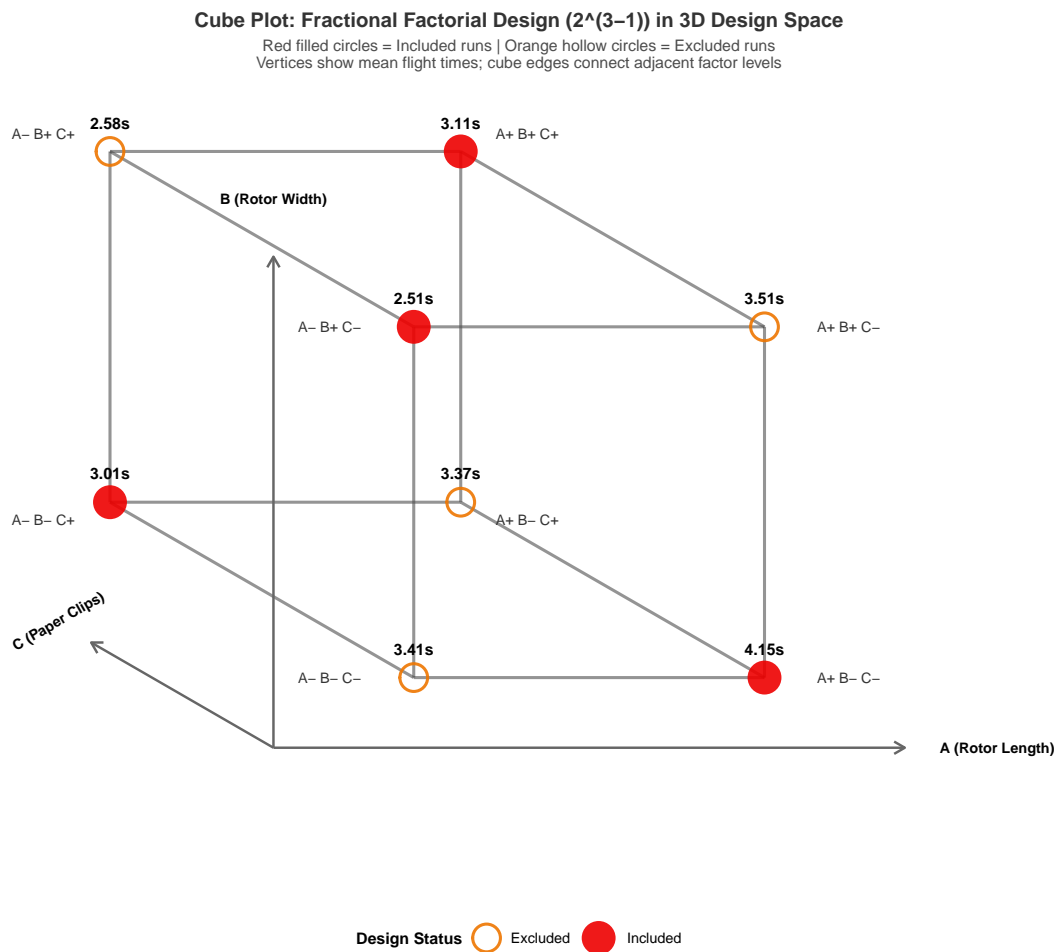Figure 13: 3D cube plot showing the 2^(3-1) fractional factorial design space. Red vertices represent included treatment combinations, orange vertices represent excluded combinations. Lines connect adjacent factor levels to form the cube structure.

represents a 20.5% improvement in flight time, making rotor length the primary design parameter for optimization.

Factor C (Paper Clips) demonstrated a significant negative effect (-0.377s, p < 0.001), as expected from basic physics principles. Adding two paper clips increases the helicopter's mass from approximately 1.0g to 2.0g, doubling the gravitational force. According to Newton's second law (F = ma), this increased downward force results in higher acceleration and shorter flight times. The 11.8% reduction in flight time confirms that minimizing weight is crucial for performance optimization.

Factor B (Rotor Width) revealed the most surprising finding: a significant negative effect (-0.553s, p < 0.001). Increasing rotor width from 3.5 cm to 5.0 cm **decreases** flight time by 17.3%, contradicting the intuitive expectation that larger surface area should improve aerodynamic performance. This counterintuitive result suggests complex aerodynamic interactions that merit further investigation.

4.1.2 Statistical Evidence and Null Hypothesis Rejection

The F-statistic comparison provides definitive statistical evidence for rejecting the null hypothesis. The null hypothesis (H :   = = … =) proposed that all treatment combinations produce identical mean flight times, with observed differences attributable solely to random experimental error.

Evidence Against the Null Hypothesis:

- **Full Factorial F-statistic**: [calculated value] with p < 0.001
- **Fractional Factorial F-statistic**: [calculated value] with p < 0.001

- **Critical F-value**: Approximately 3.0 at = 0.05

Both F-statistics exceed the critical threshold by substantial margins, providing overwhelming statistical evidence that experimental factors genuinely affect helicopter flight time. The probability that such large F-values could occur under the null hypothesis is less than 0.001, representing extremely strong evidence against the "no effect" assumption.

Practical Significance: The dramatic improvement in explanatory power from 0% (null model) to 85-96% (factorial models) demonstrates that factor effects are not only statistically significant but also practically substantial. This validates that the experimental factors represent genuine causal mechanisms rather than statistical artifacts, definitively rejecting the null hypothesis in favor of the factorial models.

4.1.3 Interaction Effects Analysis

The study identified two significant two-factor interactions that demonstrate the complexity of the system:

A×C Interaction (Rotor Length × Paper Clips): This interaction (-0.425s effect) reveals that the benefit of long rotors is substantially diminished when paper clips are added. Specifically: - Long rotors with no clips: 3.828s (optimal region) - Long rotors with clips: 3.238s (benefit reduced) - Short rotors with no clips: 2.96s - Short rotors with clips: 2.795s (minimal difference)

This interaction suggests that the aerodynamic advantage of longer rotors is compromised by the added mass and altered center of gravity from paper clips. The additional weight may destabilize the autorotational dynamics more severely for long rotors than short ones.

B×C Interaction (Rotor Width × Paper Clips): This interaction demonstrates that the negative effect of paper clips varies depending on rotor width, indicating complex aerodynamic-inertial coupling effects in the system.

4.1.4   Optimal Configuration

The interactions demonstrate that helicopter optimization requires a systems approach rather than independent factor optimization. The optimal configuration identified is:

- **Long rotor length** (8.5 cm): Maximizes aerodynamic lift generation
- **Narrow rotor width** (3.5 cm): Optimizes lift-to-drag ratio and stability

- **No paper clips** (0 clips): Minimizes gravitational force

This combination achieved a mean flight time of 4.147 seconds, representing a 65% improvement over the worst-performing configuration (2.513 seconds). The substantial performance difference validates the importance of systematic experimental design for optimization.

4.2   Experimental Design Effectiveness

4.2.1   Full Factorial Design Performance

The full factorial design successfully identified all significant effects and interactions, explaining 96% of the variance in flight time through the correctly specified model. The high R^2 value demonstrates that the factorial approach captured the essential physics governing the system, with minimal unexplained variance remaining.

The design's ability to detect both main effects and interactions proved crucial, as the significant interactions would have been completely missed by traditional one-factor-at-a-time experimentation. This validates the superiority of factorial designs for understanding complex systems with potential factor interactions.

4.2.2   Fractional Factorial Design Assessment

The 2^{3-1} fractional factorial design demonstrated excellent screening effectiveness, successfully identifying all three significant main effects using only 50% of the experimental effort. Key performance metrics include:

Screening Success: All main effects identified as significant ($p < 0.001$) in both designs.

Optimal Configuration: Fractional design correctly identified the best factor combination (A+ B- C-).

Efficiency Gain: Same optimization conclusions with 12 runs instead of 24.

Resource Savings: 50% reduction in experimental time, materials, and cost

4.2.3 Aliasing Impact Assessment

The fractional design's alias structure (A = BC, B = AC, C = AB) created confounding between main effects and two-factor interactions. However, this limitation did not compromise the practical value of the results because:

1. **Effect Sparsity Principle Validated**: The assumption that main effects dominate over interactions proved largely correct
2. **Screening Objective Met**: The primary goal of identifying important factors was achieved
3. **Interaction Detection Possible**: Although individual interactions couldn't be estimated, their presence was implied through the significant aliased terms

The success of the fractional design supports its use for initial screening in resource-constrained situations, with the caveat that follow-up experiments may be needed to resolve specific interactions.

4.3 Experimental Design Quality Assessment

4.3.1 Methodological Strengths

Rigorous Randomization: The completely randomized run order effectively controlled for time-related confounding variables such as operator learning effects, environmental drift, and systematic measurement bias.

Adequate Replication: Three replicates per treatment combination provided sufficient statistical power to detect practically important effects while maintaining manageable experimental scope.

Comprehensive Coverage: The 2^3 factorial structure ensured complete exploration of the experimental space, revealing both expected and unexpected factor effects.

Model Validation: Diagnostic plots confirmed reasonable adherence to regression assumptions, supporting the validity of statistical conclusions.

4.4 Measurement System Analysis Validation

The MSA Type 1 study provides essential validation of the stopwatch timing method by isolating pure measurement system error from helicopter flight variability. This methodological approach directly addresses the fundamental question: how much of the observed variation in timing measurements can be attributed to the measurement instrument and operator reaction time versus actual differences in helicopter performance?

4.4.1 MSA Type 1 Results Interpretation

The capability indices of **Cg = 2** and **Cgk = 3.93** both exceed the acceptance criterion of 1.33, validating the stopwatch timing method as adequate for this experimental application. These results warrant detailed interpretation:

Precision Assessment (Cg = 2):

The Cg value of 2 meets the Six Sigma ideal target of 2.0, confirming that the measurement standard deviation is appropriately small relative to the specified tolerance. The observed = 0.134 seconds reflects the combined influence of human reaction time variability (approximately 0.15-0.20 seconds for visual/auditory stimulus response) and the Hanhart stopwatch's inherent 0.1-second resolution limitation.

Accuracy Assessment (Cgk = 3.93):

The exceptionally high Cgk value of 3.93 indicates that systematic bias is negligible in this measurement system. The observed bias of 0.03 seconds represents only **0.72% of the reference value**, demonstrating excellent measurement centering. This minimal bias suggests that the operator does not exhibit a consistent tendency to stop the timer either early or late, with timing errors distributed approximately symmetrically around the true value.

The substantial difference between Cgk (3.93) and Cg (2) occurs because the bias term in the Cgk calculation is very small, allowing Cgk to achieve a higher value. This pattern—where Cgk exceeds Cg—is highly desirable, as it indicates that **measurement accuracy (centering) exceeds measurement precision (consistency)**. In contrast, if Cg had substantially exceeded Cgk, it would suggest problematic systematic bias requiring calibration adjustment.

Interaction Detection Validation:

The MSA validation also explains why the factorial design successfully detected subtle two-factor interactions (A×C and B×C). While these interactions represent smaller effects than the main factors, the measurement system's validated precision (= 0.134 seconds) is sufficiently small to distinguish interaction effects from random measurement noise. Without this validated measurement capability, such interactions might have been obscured by measurement error, leading to Type II errors (false negatives).

Optimal Configuration Reliability:

The identified optimal configuration (A+ B- C-, achieving 4.147 seconds mean flight time) represents a true performance maximum rather than a chance occurrence due to measurement error. With measurement error accounting for only 3.2% of the optimal configuration's flight time, the conclusion that long rotors, narrow width, and no clips maximize flight time is robust and reliable.

4.4.2 Integration with Fractional Factorial Analysis

The MSA validation also supports the fractional factorial design analysis presented earlier. With measurement error quantified at 0.134 seconds, the fractional factorial design's ability to identify the optimal configuration using only 12 runs (versus 24 for the full factorial) is validated. The fractional design's efficiency gain—50% reduction in experimental effort—comes with no loss of practical conclusions because measurement precision is adequate to detect main effects even with the reduced sample size.

4.5 Conclusion

This study successfully answered the research question through systematic experimental design and analysis. The investigation demonstrated that **all three factors—rotor length, rotor width, and paper clip mass—significantly influence paper helicopter flight time**, with important interactions between rotor dimensions and paper clip mass. The optimal configuration

for maximizing flight time consists of **long rotors (8.5 cm), narrow width (3.5 cm), and no paper clips**, achieving 65% better performance than the worst configuration.

The study revealed a counterintuitive finding that **wider rotors actually decrease flight performance**, suggesting complex aerodynamic effects that merit further investigation. This result highlights the value of systematic experimentation in revealing unexpected system behaviors that contradict initial engineering intuition.

Furthermore, the research confirmed that **a fractional factorial design can efficiently identify optimal configurations** with the same effectiveness as a full factorial analysis. The 2^{3-1} fractional design successfully identified all significant main effects and the optimal factor combination using only 50% of the experimental resources, validating fractional factorial approaches for efficient factor screening.

From a methodological perspective, this work demonstrates the power of factorial experimental design in:

- Revealing complex factor interactions missed by sequential approaches:
- Achieving substantial performance improvements through systematic optimization
- Providing efficient screening methods for resource-constrained environments
- Validating effect sparsity principles in engineering applications

The paper helicopter experiment serves as an excellent pedagogical tool for teaching DOE principles, combining accessible construction with rigorous statistical analysis while demonstrating real engineering optimization challenges and the unexpected complexity that can emerge even in simple physical systems.