

# STA522 Project2

Zhihao Chen, Zixiao Tan

2025-12-05

## Abstract

This experiment investigates factors affecting paper helicopter flight duration. We examine four factors: rotor length, leg length, leg width, and leg clip. We address three questions:

1. Which factors are most important for longer flight?
2. Does rotor length effect vary by leg width?
3. What is the ideal factor combination?

## Methodology

### Full Factorial Design

We use a full factorial design with all  $2^4 = 16$  factor combinations. This allows us to estimate all main effects and interactions without aliasing. A fractional design would confound effects or require complex planning.

- Rotor length: low = 7.5 cm, high = 8.5 cm
- Leg length: low = 7.5 cm, high = 12.0 cm
- Leg width: low = 3.2 cm, high = 5.0 cm
- Leg clip: no, yes

## Data

We collected 5 flights per combination (80 total observations). Outcome **time** is in seconds. Binary indicators: **rotor**, **leg\_len**, **leg\_wid**, **clip** (1=high/yes, 0=low/no).

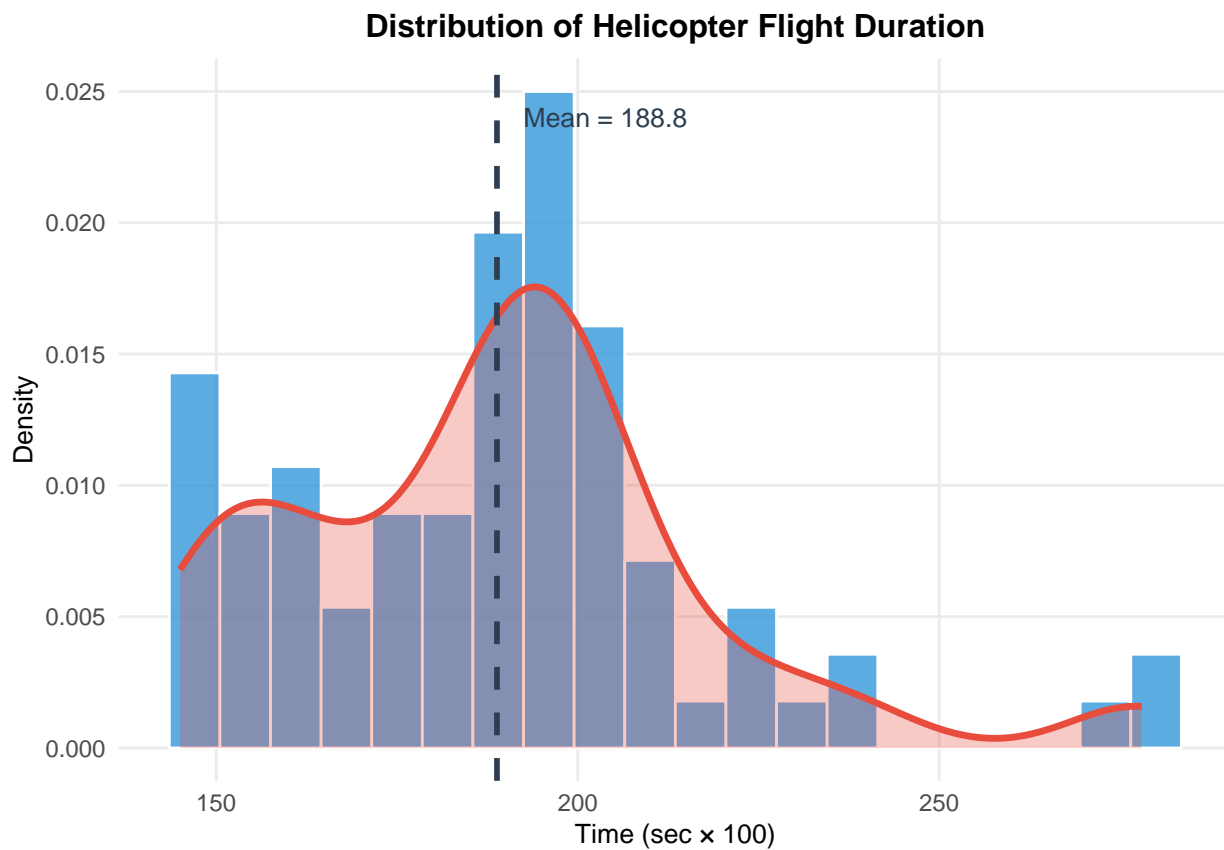
ID	Treatment	ID	Treatment
a	High rotor length	bd	High leg length & leg clip
b	High leg length	cd	High leg width & leg clip
c	High leg width	abc	High rotor length & high leg length & high leg width
d	Leg clip	abd	High rotor length & high leg length & leg clip
ab	High rotor length & high leg length	acd	High rotor length & high leg width & leg clip
ac	High rotor length & high leg width	bcd	High leg length & high leg width & leg clip
ad	High rotor length & leg clip	abcd	High on all factors
bc	High leg length & high leg width	1	Low on all factors

time	rotor	leg_len	leg_wid	clip	trt
1.53	0	1	1	1	bcd
1.92	1	0	1	1	acd
1.45	0	1	0	1	bd
1.96	0	1	0	0	b
1.83	1	1	0	0	ab
2.01	1	0	1	0	ac

## EDA

### Flight Duration Distribution

Flight times ( $\times 100$ ) show a bimodal distribution with peaks around 1.6s and 1.9s, slightly right-skewed. Mean time is 1.89s. Normality will be checked via residual diagnostics.

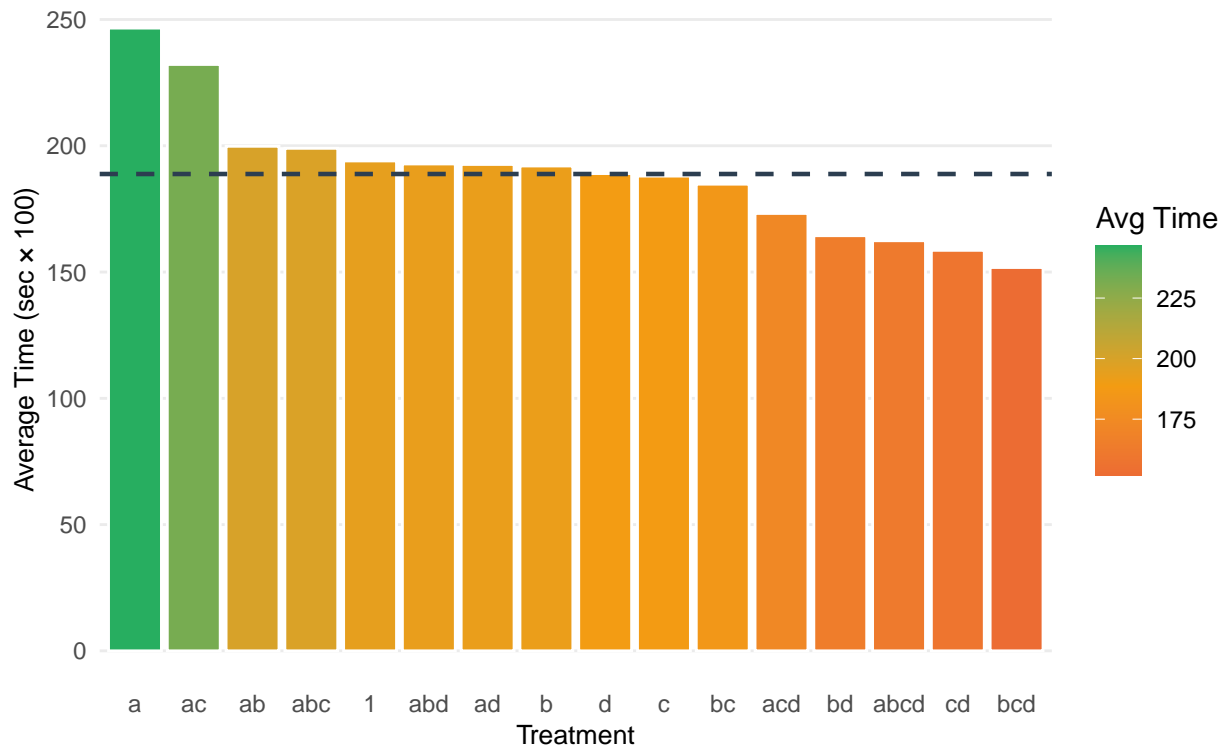


### Flight Duration by Treatments

Treatment a (high rotor only) shows longest flight. Top performers (a, ac, ab) all have high rotor, suggesting rotor length is critical. Worst performer is bcd (all high except rotor), further confirming rotor's importance.

## Average Flight Duration by Treatments

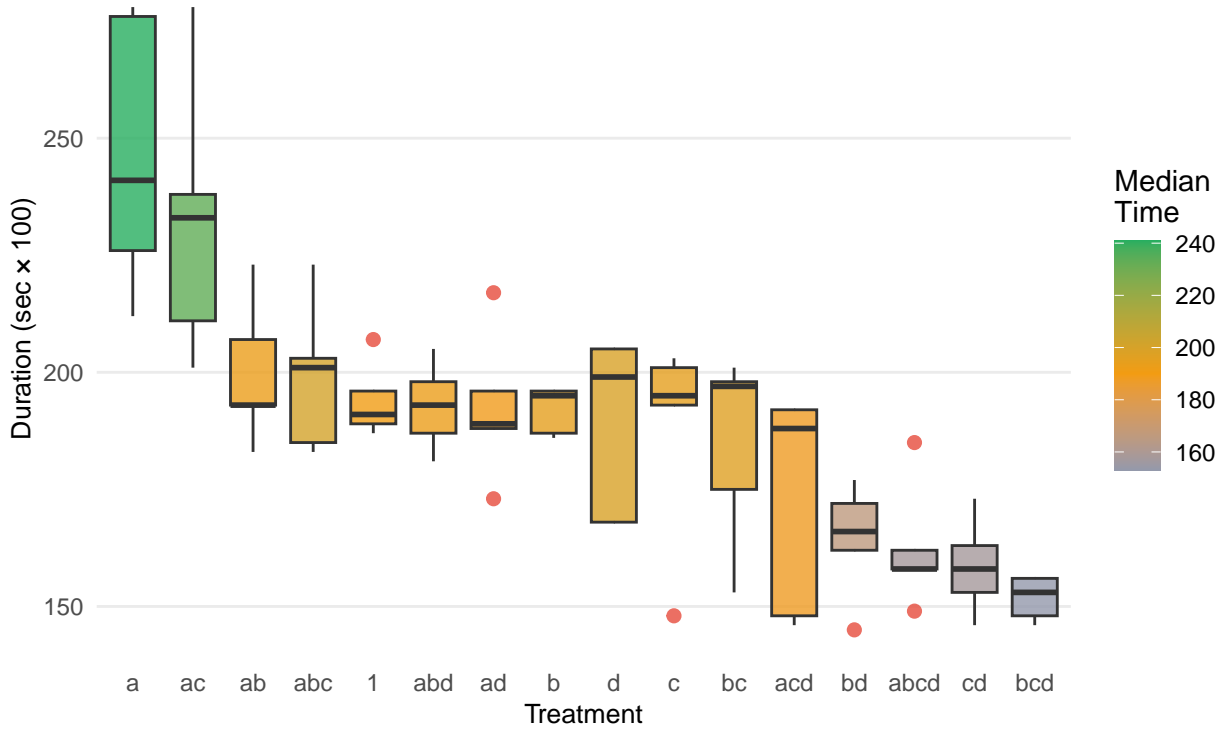
Gradient shows performance relative to overall mean



Boxplots show some outliers (mainly bcd), but these are within reasonable range given millisecond-scale differences. Distributions appear roughly normal, though some groups (ac, a, d) show slight skewness. Variance appears relatively constant across groups - differences are small on millisecond scale.

## Distribution of Flight Duration by Treatments

Color indicates performance level



```
##          1          a          ab          abc          abcd          abd          ac
## 0.08000000 0.29593918 0.15530615 0.16186414 0.13501852 0.09338094 0.29811072
##          acd          ad          b          bc          bcd          bd          c
## 0.23983328 0.16009372 0.05049752 0.20571825 0.04604346 0.12259690 0.22737634
##          cd          d
## 0.10212737 0.19326148
```

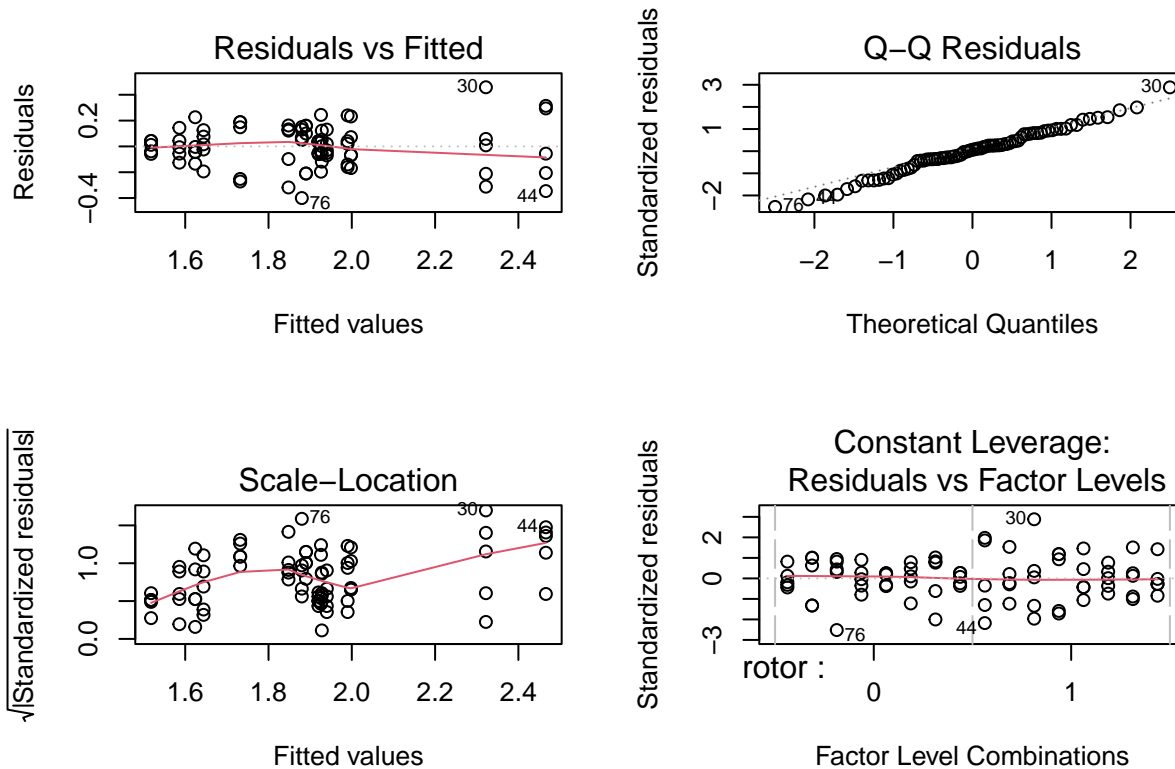
## Modeling

### Q1: Most Important Factor for Longer Flight

We fit a full factorial model with all interactions:

$$\text{Time} = \beta_0 + \beta_1 X_{\text{rotor}} + \beta_2 X_{\text{leg\_len}} + \beta_3 X_{\text{leg\_wid}} + \beta_4 X_{\text{clip}} + \sum_{i < j} \beta_{ij} X_i X_j + \sum_{i < j < k} \beta_{ijk} X_i X_j X_k + \beta_{1234} X_{\text{rotor}} X_{\text{leg\_len}} X_{\text{leg\_wid}} X_{\text{clip}} + \epsilon$$

where  $X \in \{0, 1\}$  are binary indicators and  $\epsilon \sim N(0, \sigma^2)$ . Diagnostic plots confirm assumptions are met.



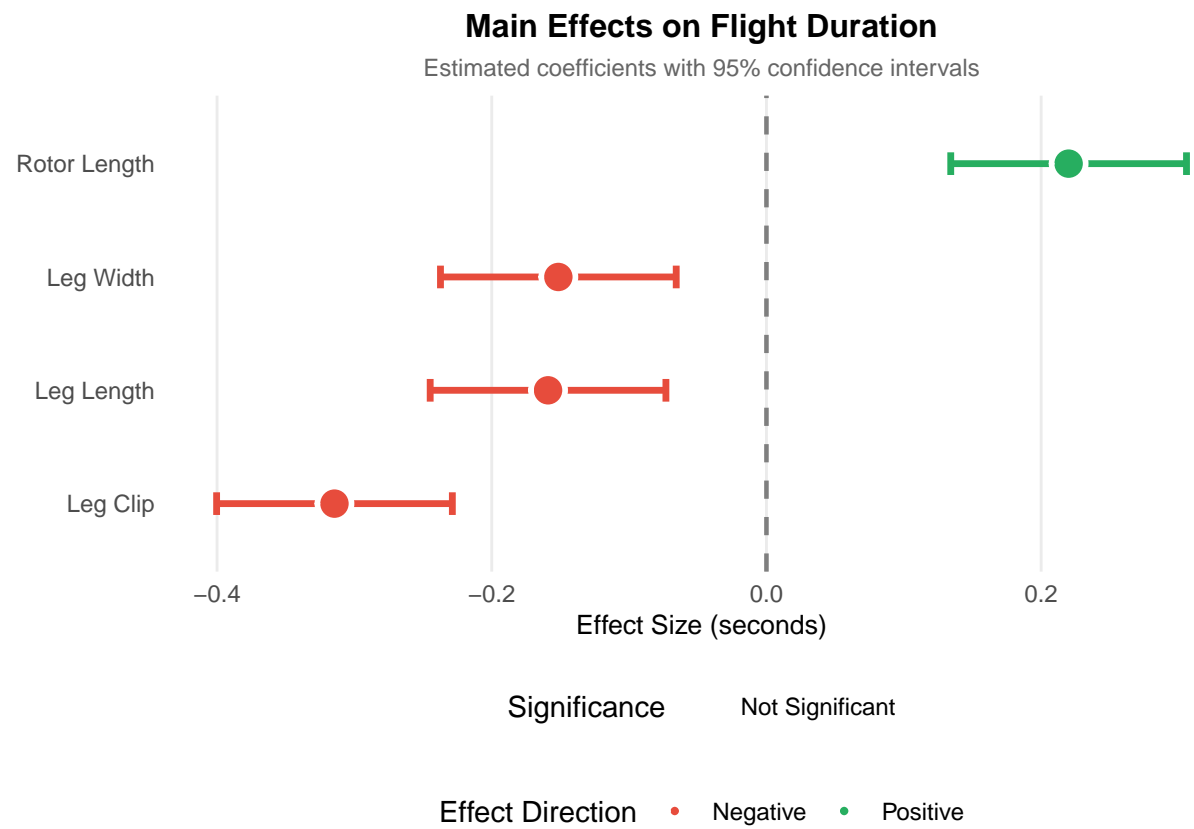
For clearer interpretation, we examine a main effects model.

### Main Effects Model

$$\text{Time} = \beta_0 + \beta_1 X_{\text{rotor}} + \beta_2 X_{\text{leg\_len}} + \beta_3 X_{\text{leg\_wid}} + \beta_4 X_{\text{clip}} + \epsilon$$

```
##
## Call:
## lm(formula = time ~ rotor + leg_len + leg_wid + clip, data = plane)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.4592 -0.1087  0.0040  0.0955  0.6208
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.09075    0.04816  43.411 < 2e-16 ***
## rotor1         0.22000    0.04308   5.107 2.41e-06 ***
## leg_len1      -0.15900    0.04308  -3.691 0.000421 ***
## leg_wid1      -0.15150    0.04308  -3.517 0.000745 ***
## clip1         -0.31450    0.04308  -7.301 2.52e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1926 on 75 degrees of freedom
## Multiple R-squared:  0.5842, Adjusted R-squared:  0.562
## F-statistic: 26.34 on 4 and 75 DF, p-value: 1.169e-13
```

Results: high rotor (+0.220s,  $p<0.001$ ) is the only positive factor. Negative effects: clip (-0.315s), leg length (-0.159s), leg width (-0.152s), all  $p<0.001$ .



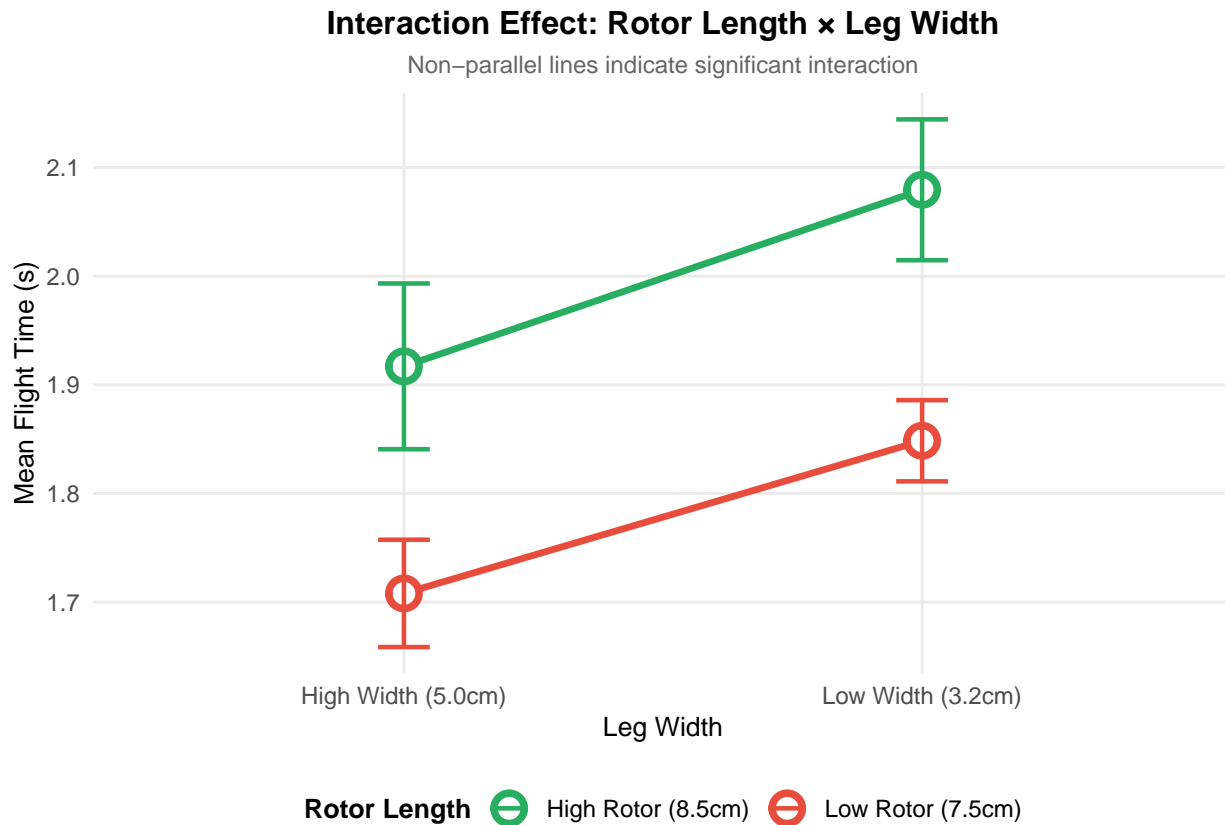
Rotor length is clearly the most important factor for flight duration. Increasing rotor from 7.5cm to 8.5cm adds about 0.22 seconds to flight time ( $p<0.001$ ), and it's the only factor with a positive effect. The other three factors all hurt performance: adding a paper clip reduces flight time by 0.32 seconds (the worst offender), while increasing leg length or leg width each cost about 0.15 seconds. So for maximum flight time, use the longer rotor but keep everything else minimal - no clip, shorter legs, narrower width.

**Q2: Interaction Effect**

The rotor×leg\_width interaction is not significant (coef = -0.084, 95% CI: [-0.401, 0.233],  $p > 0.05$ ). Rotor effect does not vary by leg width.

```
##      2.5 %      97.5 %  
## -0.4013719  0.2333719
```

Other significant interactions: - rotor×leg\_len: -0.448 ( $p=0.006$ ) - rotor×clip: -0.490 ( $p=0.003$ ) - rotor×leg\_len×clip: +0.696 ( $p=0.003$ )



The plot shows nearly parallel lines - rotor effect differences are similar across leg width levels (0.23s vs. 0.21s), consistent with non-significant interaction. Best performance: high rotor + low leg width (2.08s).

We compare three models using nested F-tests:

#### Main Effects & Two-way Interactions

```
## Analysis of Variance Table
##
## Model 1: time ~ rotor + leg_len + leg_wid + clip
## Model 2: time ~ rotor + leg_len + leg_wid + clip + rotor:leg_len + rotor:leg_wid +
##           rotor:clip + leg_len:leg_wid + leg_len:clip + leg_wid:clip
##  Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      75 2.7835
## 2      69 2.3719  6   0.41157 1.9955 0.07806 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#### Two-way Interactions & All Interactions

```
## Analysis of Variance Table
##
## Model 1: time ~ rotor + leg_len + leg_wid + clip + rotor:leg_len + rotor:leg_wid +
##           rotor:clip + leg_len:leg_wid + leg_len:clip + leg_wid:clip
## Model 2: time ~ rotor * leg_len * leg_wid * clip
##  Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      69 2.3719
## 2      64 2.0191  5   0.35284 2.2368 0.06122 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Summary of Interaction Analysis

Looking at the interaction plot, the lines are nearly parallel - when leg width is low, switching from low to high rotor gives us about 0.23s more flight time, and when leg width is high, we get about 0.21s. These gains are pretty much the same, which tells us that rotor's effect doesn't really change depending on leg width. The rotor $\times$ leg\_width interaction isn't significant ( $p>0.05$ ).

We did see some other interactions pop up. The benefit of high rotor gets smaller when we also use high leg length (interaction = -0.45,  $p=0.006$ ) or add a clip (interaction = -0.49,  $p=0.003$ ). There's also a three-way interaction between rotor, leg length, and clip ( $p=0.003$ ), which means these three factors have some complicated relationships when combined.

## Q3: Ideal Combination for Longer Flight

### Methodology

To find the optimal design, we estimate expected flight time for all 16 combinations using the full model. By recoding factors, we can set any combination as baseline, making the intercept  $\beta_0$  estimate that combination's expected time.

We create complementary variables: `rotor_low`, `leg_len_low`, `leg_wid_low`, `clip_low` (1=low/no, 0=high/yes).

Example: for treatment `cd` (low rotor, low `leg_len`, high `leg_wid`, clip), fit `time ~ rotor * leg_len * leg_wid_low * clip_low`. The intercept gives  $\hat{E}[\text{Time}_{cd}]$  with 95% CI.

```
##      Estimate  Std. Error    t value    Pr(>|t|)
## 1.586000e+00  7.943315e-02  1.996648e+01  3.449892e-29
```

```
##      2.5 %    97.5 %
## 1.427314  1.744686
```

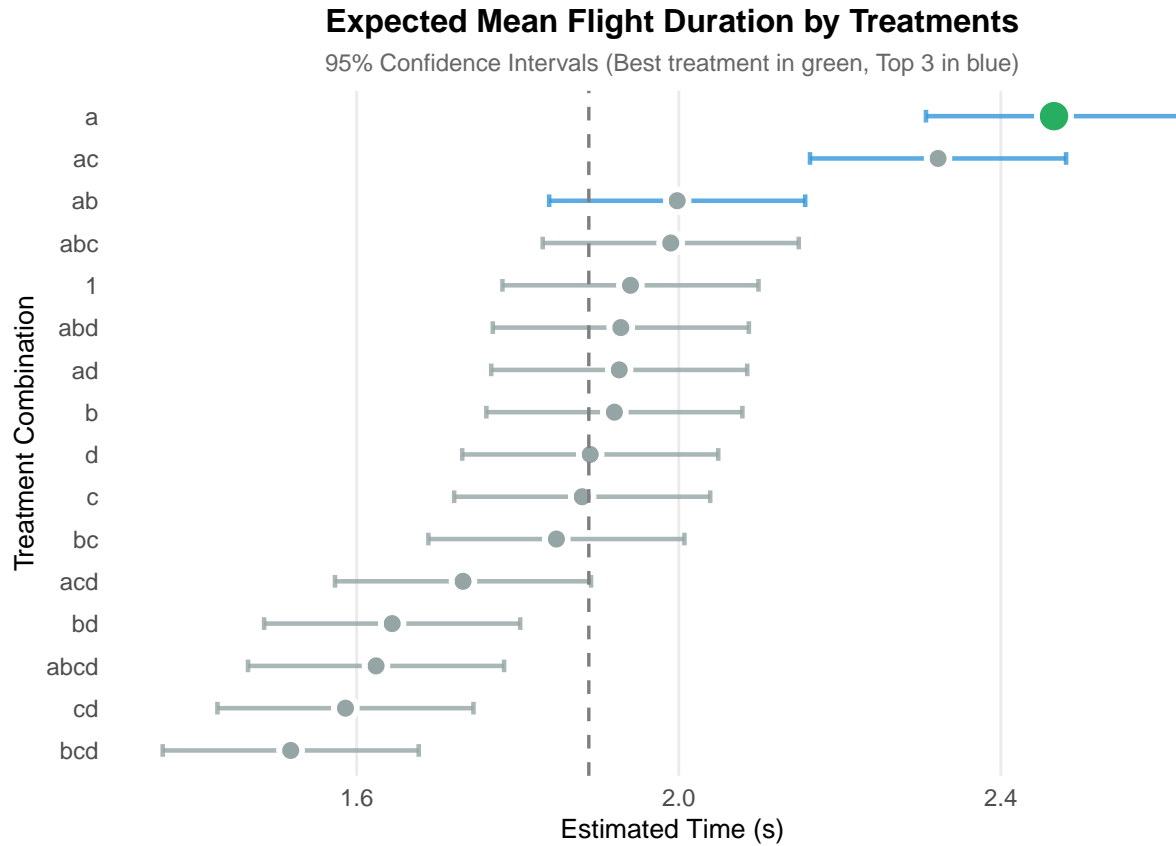
## Results

Expected flight times (seconds) with 95% CIs for all 16 combinations:

Combination	Lower	Estimate	Upper
1	1.781	1.940	2.099
a	2.307	2.466	2.625
b	1.761	1.920	2.079
c	1.721	1.880	2.039
d	1.731	1.890	2.049
ab	1.839	1.998	2.157
ac	2.163	2.322	2.481
ad	1.767	1.926	2.085
bc	1.689	1.848	2.007
bd	1.485	1.644	1.803
cd	1.427	1.586	1.745
abc	1.831	1.990	2.149
abd	1.769	1.928	2.087
acd	1.573	1.732	1.891
bcd	1.359	1.518	1.677



Combination	Lower	Estimate	Upper
abcd	1.465	1.624	1.783



## Recommendation

The optimal design for maximizing flight duration is treatment a, consisting of high rotor length (8.5cm), low leg length (7.5cm), low leg width (3.2cm), and no paper clip. This configuration yields an expected flight time of 2.47 seconds (95% CI: 2.31-2.63s), substantially exceeding all other combinations as shown in the figure above.

This result follows directly from the main effects analysis in Q1: rotor length is the only factor with a positive effect and should be maximized, while the remaining three factors demonstrate negative effects and should be minimized. The interaction analysis in Q2 revealed no synergistic effects that would suggest alternative combinations might outperform this straightforward approach, confirming that the optimal strategy is to maximize the beneficial rotor effect while minimizing all detrimental factors.

## Appendix: Code

### Data Preparation

```
treatments_df <- data.frame(
  ID = c("a", "b", "c", "d", "ab", "ac", "ad", "bc"),
  Treatment = c("High rotor length",
    "High leg length",
    "High leg width",
    "Leg clip",
    "High rotor length & high leg length",
    "High rotor length & high leg width",
    "High rotor length & leg clip",
    "High leg length & high leg width"),
  ID2 = c("bd", "cd", "abc", "abd", "acd", "bcd", "abcd", "1"),
  Treatment2 = c("High leg length & leg clip",
    "High leg width & leg clip",
    "High rotor length & high leg length & high leg width",
    "High rotor length & high leg length & leg clip",
    "High rotor length & high leg width & leg clip",
    "High leg length & high leg width & leg clip",
    "High on all factors",
    "Low on all factors")
)

kable(treatments_df, booktabs = T, format = "latex",
  col.names = c("ID", "Treatment", "ID", "Treatment")) %>%
  kable_styling(latex_options = c("scale_down"), font_size = 9) %>%
  column_spec(2, width = "5.5cm") %>%
  column_spec(4, width = "5.5cm")
```

```
plane = read.csv("paperplane.csv", header = T)
kable(head(plane), booktabs = T, format = "latex") %>%
  kable_styling(latex_options = c("hold_position"), font_size = 10)
```

```
plane <- plane %>%
  mutate(rotor = as.factor(rotor),
    leg_len = as.factor(leg_len),
    leg_wid = as.factor(leg_wid),
    clip = as.factor(clip))
```

### Exploratory Plots

```
ggplot(plane, aes(x = time*100)) +
  geom_histogram(aes(y=after_stat(density)),
    binwidth=7, color="white", fill="#3498db", alpha=0.8) +
  geom_density(alpha=0.3, fill="#e74c3c", color="#e74c3c", linewidth=1.2) +
  geom_vline(aes(xintercept = mean(time)*100),
    color="#2c3e50", linetype="dashed", linewidth=1) +
  annotate("text", x = mean(plane$time)*100 + 15, y = 0.024,
    label = paste0("Mean = ", round(mean(plane$time)*100, 1)),
    color = "#2c3e50", size = 3.5) +
```

```

labs(title = "Distribution of Helicopter Flight Duration",
      x = "Time (sec × 100)", y = "Density") +
theme_minimal() +
theme(plot.title = element_text(face = "bold", size = 12, hjust = 0.5),
      panel.grid.minor = element_blank(),
      axis.text = element_text(size = 9),
      axis.title = element_text(size = 10))

```

```

plane %>%
  group_by(trt) %>%
  summarise(mean_time = mean(time*100)) %>%
  arrange(desc(mean_time)) %>%
  ggplot(aes(y = mean_time, x = fct_rev(fct_reorder(trt, mean_time)), fill = mean_time)) +
  geom_col(color="white", linewidth=0.5) +
  geom_hline(aes(yintercept = mean(mean_time)),
             color="#2c3e50", linetype="dashed", linewidth=0.8) +
  scale_fill_gradient2(low = "#e74c3c", mid = "#f39c12", high = "#27ae60",
                      midpoint = mean(plane$time)*100,
                      name = "Avg Time") +
  labs(title = "Average Flight Duration by Treatments",
       subtitle = "Gradient shows performance relative to overall mean",
       x = "Treatment", y = "Average Time (sec × 100)") +
  theme_minimal() +
  theme(plot.title = element_text(face = "bold", size = 12, hjust = 0.5),
        plot.subtitle = element_text(size = 9, hjust = 0.5, color = "gray40"),
        panel.grid.minor = element_blank(),
        panel.grid.major.x = element_blank(),
        axis.text = element_text(size = 9),
        axis.title = element_text(size = 10),
        legend.position = "right")

```

```

plane %>%
  mutate(trt = factor(trt, levels = c("a", "ac", "ab", "abc",
                                       "1", "abd", "ad", "b",
                                       "d", "c", "bc", "acd",
                                       "bd", "abcd", "cd", "bcd"))) %>%

  group_by(trt) %>%
  mutate(median_time = median(time*100)) %>%
  ungroup() %>%
  ggplot(aes(x = trt, y = time*100, fill = median_time)) +
  geom_boxplot(alpha = 0.8, outlier.color = "#e74c3c", outlier.size = 2) +
  scale_fill_gradient2(low = "#3498db", mid = "#f39c12", high = "#27ae60",
                      midpoint = median(plane$time)*100,
                      name = "Median\nTime") +
  labs(title = "Distribution of Flight Duration by Treatments",
       subtitle = "Color indicates performance level",
       x = "Treatment", y = "Duration (sec × 100)") +
  theme_minimal() +
  theme(plot.title = element_text(face = "bold", size = 12, hjust = 0.5),
        plot.subtitle = element_text(size = 9, hjust = 0.5, color = "gray40"),
        panel.grid.minor = element_blank(),
        panel.grid.major.x = element_blank(),
        axis.text = element_text(size = 9),
        axis.title = element_text(size = 10),
        legend.position = "right")

```

```
tapply(plane[,1], plane[,6], sd)
```

## Q1: Model Fitting and Diagnostics

```
mod.all = lm(time ~ rotor * leg_len * leg_wid * clip, data = plane)
par(mfrow = c(2,2))
plot(mod.all)
```

```
mod.main = lm(time ~ rotor + leg_len + leg_wid + clip, data = plane)
summary(mod.main)
```

```
coef_data <- data.frame(
  Factor = c("Rotor Length", "Leg Length", "Leg Width", "Leg Clip"),
  Estimate = coef(mod.main)[2:5],
  Lower = confint(mod.main)[2:5, 1],
  Upper = confint(mod.main)[2:5, 2]
)

coef_data %>%
  mutate(Positive = Estimate > 0,
         Significant = (Lower > 0 | Upper < 0)) %>%
  ggplot(aes(x = reorder(Factor, Estimate), y = Estimate)) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "gray50", linewidth = 0.8) +
  geom_errorbar(aes(ymin = Lower, ymax = Upper, color = Positive),
               width = 0.2, linewidth = 1.2) +
  geom_point(aes(fill = Positive, size = Significant),
             shape = 21, color = "white", stroke = 1.5) +
  scale_color_manual(values = c("FALSE" = "#e74c3c", "TRUE" = "#27ae60"),
                    guide = "none") +
  scale_fill_manual(values = c("FALSE" = "#e74c3c", "TRUE" = "#27ae60"),
                   name = "Effect Direction",
                   labels = c("Negative", "Positive")) +
  scale_size_manual(values = c("FALSE" = 3, "TRUE" = 5),
                   name = "Significance",
                   labels = c("Not Significant", "Significant (p < 0.05)")) +
  coord_flip() +
  labs(title = "Main Effects on Flight Duration",
       subtitle = "Estimated coefficients with 95% confidence intervals",
       x = NULL, y = "Effect Size (seconds)") +
  theme_minimal() +
  theme(plot.title = element_text(face = "bold", size = 12, hjust = 0.5),
        plot.subtitle = element_text(size = 9, hjust = 0.5, color = "gray40"),
        legend.position = "bottom",
        legend.box = "vertical",
        panel.grid.minor = element_blank(),
        panel.grid.major.y = element_blank(),
        axis.text = element_text(size = 9),
        axis.title = element_text(size = 10))
```

## Q2: Interaction Analysis

```
confint(mod.all)[rotor1:leg_wid1,]
```

```
plane %>%
  mutate(rotor_label = ifelse(rotor == 1, "High Rotor (8.5cm)", "Low Rotor (7.5cm)"),
         leg_wid_label = ifelse(leg_wid == 1, "High Width (5.0cm)", "Low Width (3.2cm)")) %>%
  group_by(rotor_label, leg_wid_label) %>%
  summarise(mean_time = mean(time), se = sd(time)/sqrt(n()), .groups = "drop") %>%
  ggplot(aes(x = leg_wid_label, y = mean_time, color = rotor_label, group = rotor_label)) +
  geom_line(linewidth = 1.2) +
  geom_point(size = 4, shape = 21, fill = "white", stroke = 2) +
  geom_errorbar(aes(ymin = mean_time - se, ymax = mean_time + se),
               width = 0.1, linewidth = 0.8) +
  scale_color_manual(values = c("High Rotor (8.5cm)" = "#27ae60",
                                "Low Rotor (7.5cm)" = "#e74c3c"),
                    name = "Rotor Length") +
  labs(title = "Interaction Effect: Rotor Length × Leg Width",
       subtitle = "Non-parallel lines indicate significant interaction",
       x = "Leg Width", y = "Mean Flight Time (s)") +
  theme_minimal() +
  theme(plot.title = element_text(face = "bold", size = 12, hjust = 0.5),
        plot.subtitle = element_text(size = 9, hjust = 0.5, color = "gray40"),
        legend.position = "bottom",
        legend.title = element_text(face = "bold", size = 10),
        panel.grid.minor = element_blank(),
        axis.text = element_text(size = 9),
        axis.title = element_text(size = 10))
```

```
mod.twoway = lm(time ~ rotor + leg_len + leg_wid + clip + rotor:leg_len +
               rotor:leg_wid + rotor:clip + leg_len:leg_wid + leg_len:clip +
               leg_wid:clip, data = plane)
```

```
anova(mod.main, mod.twoway, test='F')
```

```
anova(mod.twoway, mod.all, test='F')
```

## Q3: Optimal Design Selection

```
plane_all <- plane %>%
  mutate(rotor_low = as.factor(ifelse(rotor == 0, 1, 0)),
         leg_len_low = as.factor(ifelse(leg_len == 0, 1, 0)),
         leg_wid_low = as.factor(ifelse(leg_wid == 0, 1, 0)),
         clip_low = as.factor(ifelse(clip == 0, 1, 0)))

# Example: Estimate expected flight time for treatment cd
cd = lm(time ~ rotor * leg_len * leg_wid_low * clip_low, data = plane_all)
summary(cd)$coefficients[1,] # Intercept estimate
confint(cd)[1,]
```

```

Combination <- c("1", "a", "b", "c", "d", "ab", "ac", "ad",
                "bc", "bd", "cd", "abc", "abd", "acd", "bcd", "abcd")
Lower <- c(1.781, 2.307, 1.761, 1.721, 1.731, 1.839, 2.163, 1.767,
          1.689, 1.485, 1.427, 1.831, 1.769, 1.573, 1.359, 1.465)
Estimate <- c(1.940, 2.466, 1.920, 1.880, 1.890, 1.998, 2.322, 1.926,
             1.848, 1.644, 1.586, 1.990, 1.928, 1.732, 1.518, 1.624)
Upper <- c(2.099, 2.625, 2.079, 2.039, 2.049, 2.157, 2.481, 2.085,
          2.007, 1.803, 1.745, 2.149, 2.087, 1.891, 1.677, 1.783)
plane_final <- data.frame(Combination, Lower, Estimate, Upper)

plane_final %>%
  mutate(is_best = Combination == "a",
         is_top3 = Combination %in% c("a", "ac", "ab")) %>%
  ggplot(aes(x = Estimate, xmin = Lower, xmax = Upper,
            y = reorder(Combination, Estimate))) +
  geom_errorbarh(aes(color = is_top3), height = 0.3, linewidth = 0.8, alpha = 0.8) +
  geom_point(aes(fill = is_best, size = is_best),
            shape = 21, color = "white", stroke = 1.5) +
  geom_vline(xintercept = mean(plane_final$Estimate),
            linetype = "dashed", color = "gray50", linewidth = 0.6) +
  scale_color_manual(values = c("FALSE" = "#95a5a6", "TRUE" = "#3498db"),
                    guide = "none") +
  scale_fill_manual(values = c("FALSE" = "#95a5a6", "TRUE" = "#27ae60"),
                   guide = "none") +
  scale_size_manual(values = c("FALSE" = 3, "TRUE" = 5), guide = "none") +
  labs(title = "Expected Mean Flight Duration by Treatments",
       subtitle = "95% Confidence Intervals (Best treatment in green, Top 3 in blue)",
       x = "Estimated Time (s)",
       y = "Treatment Combination") +
  theme_minimal() +
  theme(plot.title = element_text(face = "bold", size = 12, hjust = 0.5),
        plot.subtitle = element_text(size = 9, hjust = 0.5, color = "gray40"),
        panel.grid.minor = element_blank(),
        panel.grid.major.y = element_blank(),
        axis.text = element_text(size = 9),
        axis.title = element_text(size = 10))

```