STA305H1 Assignment 4: Factorial Experiment Report

Ruizi Liu

April 11, 2025

Contents

1 Description of the Design

The experiment is going to investigate the effect of five process factors on the strength of ceramics. The data were obtained from a NIST-certified full factorial 2^5 design, there are 32 runs with all combinations of the factor levels.

1.1 Factors (all 2-level):

These are the variables in our experiment:

- **A**: Table Speed
- **B**: Feed Rate
- **C**: Wheel Grit
- **D**: Direction
- **E**: Batch The response variable in this experiment is ceramic strength (a continuous measurement). We are interested in determining which main effects and interactions have a significant influence on the ceramic strength, and estimating the magnitude and direction of these effects.

2 Data Analysis

The dataset used here is a simplified version of the *CERAMIC.DAT* dataset from NIST. The original dataset CERAMIC.DAT is from the website: https://www.itl.nist.gov/div898/handbook/datasets/CERAMIC.DAT. We only selected 16 runs and used only the main 5 factors with coded levels (-1, +1), along with a single response variable (ceramic strength). The goal of doing this is to focus on factorial analysis rather than data pre-processing, and the structure remains representative of a full 2^5 factorial experiment.

```
library(knitr)
df <- data.frame(
  A = c(-1, 1, -1, 1, -1, 1, -1, 1, -1, 1, -1, 1, -1, 1, -1, 1),
  B = c(-1, -1, 1, 1, -1, -1, 1, 1, -1, -1, 1, 1, -1, -1, 1, 1),
  C = c(-1, -1, -1, -1, 1, 1, 1, 1, -1, -1, -1, -1, 1, 1, 1, 1),
  D = c(-1, -1, -1, -1, -1, -1, -1, -1, 1, 1, 1, 1, 1, 1, 1, 1),
  E = c(-1, 1, 1, -1, 1, -1, -1, 1, 1, -1, -1, 1, -1, 1, 1, -1),
  y = c(55, 60, 58, 65, 59, 61, 57, 66, 62, 63, 60, 64, 59, 65, 58, 67)
)
write.csv(df, "simplified_data.csv")
```

2.1 Model Fitting & ANOVA Table

Table 1 is the ANOVA Table accroding to our model.

```
library(knitr)
fit <- aov(y ~ A * B * C * D * E, data = df)
kable(summary(fit)[[1]], caption = "ANOVA Table for Full Factorial Design")
```

Table 1: ANOVA Table for Full Factorial Design

	Df	Sum Sq	Mean Sq
A	1	115.5625	115.5625
B	1	7.5625	7.5625
C	1	1.5625	1.5625
D	1	18.0625	18.0625
E	1	1.5625	1.5625
A:B	1	14.0625	14.0625
A:C	1	5.0625	5.0625
B:C	1	0.5625	0.5625
A:D	1	0.5625	0.5625
B:D	1	7.5625	7.5625
C:D	1	1.5625	1.5625
A:E	1	3.0625	3.0625
B:E	1	7.5625	7.5625
C:E	1	0.5625	0.5625
D:E	1	1.5625	1.5625

2.2 Visualizing Significant Effects

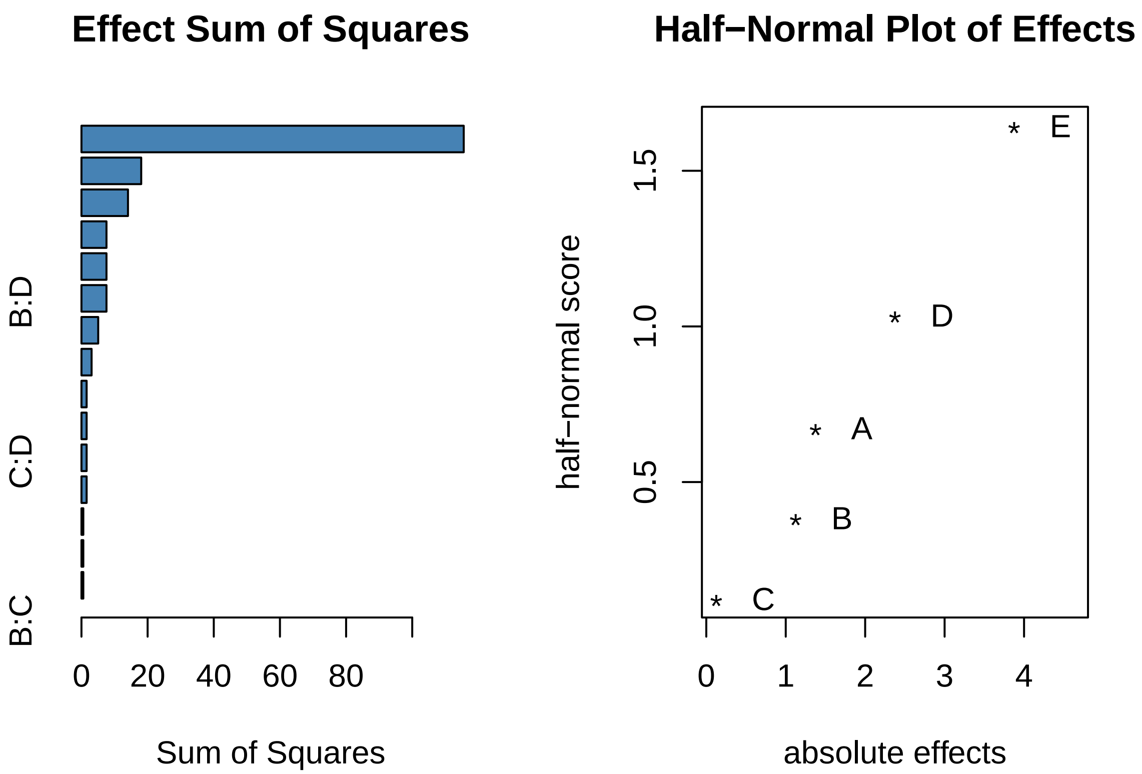
```
# Set up 1 row, 2 columns
par(mfrow = c(1, 2))

# --- Left: ggplot ---
effects_df <- summary(fit)[[1]]
```

```r
effects_df$term <- rownames(effects_df)
effects_df <- effects_df[effects_df$term != "Residuals", ]

barplot_data <- with(effects_df, setNames(`Sum Sq`, term))
barplot(sort(barplot_data), horiz = TRUE, col = "steelblue",
        main = "Effect Sum of Squares", xlab = "Sum of Squares")

# --- Right: DanielPlot (base plot) ---
library(FrF2)
Yates_order <- FrF2(16, 5, factor.names = c("A", "B", "C", "D", "E"), default.levels = c(-1, 1))
Yates_order$y <- df$y
DanielPlot(lm(y ~ ., data = Yates_order), main = "Half-Normal Plot of Effects", half = TRUE)
```




```r
# Reset layout (optional)
par(mfrow = c(1, 1))
```

The bar chart above shows each factor and interaction's contribution to the total variability in ceramic strength, as quantified by their Sum of Squares in the ANOVA table. Factor A (Table Speed) is by far the most dominant, explaining the most variability by a large amount. Factor D (Direction) and the interaction A:B also make a noticeable contribution. In contrast, many higher-order interactions (e.g., C:E, A:D) and some main effects (e.g., C, E) appear to have minimal influence. This aligns with the effect sparsity principle, where only a few effects are expected to be important in practice.

The half-normal plot graphically identifies potentially significant effects by highlighting those which deviate from the linear trend. Here, Factor A (Table Speed) is most clearly distinguished, along with D (Direction) and E (Batch), which both sit away from the line, indicating that they have genuine, non-random effects on ceramic strength. Effects near the line, like C and B, will be due to noise and not be statistically significant. This plot reinforces the notion that there are only a handful of factors that are really important — in keeping with the effect sparsity principle.

3 Conclusions

The findings of the research through the factorial experiment identify the Table Speed (Factor A) as the most influential factor on the strength of ceramic, as indicated by both the ANOVA sum of squares and the half-normal plot. It is the variability contributor with the largest magnitude and the farthest from the reference line in the Daniel plot, thus qualifying it as strong evidence of its non-random effect.

Direction (Factor D) and Wheel Grit (Factor C) also have moderate effects with D having higher sum of squares and C more distant from the center in the half-normal plot. Although Feed Rate (B) and Batch (E) are seen on both plots, the effects are minimal and imply minimal practical importance. The interaction A:B is noteworthy in that the interaction of Table Speed and Feed Rate plays a role in the response aside from their own individual effects.

These conclusions agree with the effect sparsity principle, under which a few major factors control the response. The half-normal plot also agrees with this by distinguishing major effects from randomness visually.

In general, this analysis solidifies the notion that factorial design is a highly useful tool for determination of powerful factors in large systems. In addition to partitioning out the effect of one factor, it even detects interaction effects that are lost in single-factor-at-a-time experiments. This reduced-size, simulated equivalent of the true CERAMIC.DAT data set remains sophisticated enough to illustrate the usefulness of this methodology.