

AMS 582 Fall 2021  
Design and Analysis of Experiments  
The One-Sixty-Fourth Fraction  
of the  $2^{10}$  Factorial Design

Kai Li\*

Stony Brook University — November 27, 2021

## 1 Introduction

In an experimental setting, researchers are interested in exploring and modeling the relationship between dependent variables and independent variables. A multiple linear regression model can often be expressed as  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ , where  $\boldsymbol{\beta}$  is a regression coefficient vector and  $\boldsymbol{\epsilon}$  is a random error vector [3, 4]. It is possible to create an experiment with a certain number of experimental trials to estimate  $\boldsymbol{\beta}$ , the explanatory relationship between  $\mathbf{X}$  and  $\mathbf{Y}$ . Moreover, researchers are likely to develop reduced models due to a small portion of potential important variables among a large amount of exploratory variables. Statistical inferences such as hypothesis testing, building confidence intervals for model parameters and making predictions are critical to interpret modeling results.

The objective of this paper is to design, perform and analyze an experiment for the given scenario. To achieve our goal, we follow the guidelines for designing and analyzing experiments suggested by Kutner et al. and Montgomery [2, 3]. The remainder of the paper is organized as follows. Section 1.1 discusses the recognition of and statement of the problem. Choice of experimental design with detailed reasoning is provided in Section 1.2. Section 2 outlines a comprehensive procedure for performing the experiment. Section 3 provides the statistical analysis results. Section 4 details the conclusions and additional comments on the analysis procedures performed throughout the paper. Lastly, Appendix A shows the implementation of all the experimental design and analysis processes used in this paper.

### 1.1 Current Problem

We are given a dependent variable  $Y$  and 10 independent variables, also called factors,  $A$  through  $J$  with unselected restrictive continuous values from  $-1$  to  $1$ . We are responsible for choosing an experimental design with appropriate factor values under this limited scenario. The  $\mathbf{Y}$  vector values are given from an unknown relationship  $y = f(a, \dots, j)$  after determining the design and associated factor values. We are also told that researchers believe that the true model uses five or fewer independent variables, but this is unknown apriori.

---

\*Department of Applied Mathematics and Statistics, Stony Brook University, email: kai.li@stonybrook.edu

## 1.2 Rationale of Selecting the Best Design

To select the most suitable design for the given conditions, we first decompose the current problem into the following components or structures of the design of an experiment referred by Kutner et al. [3]:

### **The set of explanatory factors included in the study.**

The 10 explanatory variables are given as factors  $A$  through  $J$ . We know little about which explanatory factors are active and important, except the belief that the set of active independent variables contains five or fewer elements. Hence, we believe that the study can be categorized as a factor screening study, also called a comparative exploratory experimental study. Kutner et al. [2] and Montgomery [3] describe factor screening studies as studies that are intended in discovering the set of active factors from a large set of factors.

### **The set of treatments in the study.**

The set of treatments to be included can be decided by the 10 factors and their respective levels, where each treatment corresponds to a combination of all the factors. At this stage of the investigation, we want to examine a manageable number of treatment combinations. In exploratory or screening studies, limiting each factor to two levels reduces the number of treatment combinations and it works very well under general settings [2, 3].

### **The set of experimental units included in the study.**

The experimental units are not specified. We can consider the experimental units as general objects or entities to which the treatments are applied.

### **The rules and procedures by which the treatments are randomly assigned to the experimental units (or vice versa).**

Randomization is critically important in all phases of an experiment, as it tends to average out all systematic extraneous effects between the treatments so that the pure treatment effects can be observed [2, 3]. Randomization is also the key to validating the assumption of error term independence [3]. Without randomization, subjective assignments can lead to selection bias [2]. In this paper, randomization is achieved by permuting treatments to experimental units and randomizing the order of individual runs using R.

### **The outcome measurements that are made on the experimental units.**

Kutner et al. [2] note that the measurement process results should be unbiased and precise to minimize the probability of measurement bias in the study. Here, since the measurements in vector  $Y$  are given based on input factor values and undetermined functional relationship of the statistical design, we assume that the response variable values are unbiased and precise.

Additionally, the choice of a proper statistical design is closely related to the three basic principles of experimental design introduced by Montgomery [3]. They are randomization, replication, and blocking. To elaborate on replication and blocking:

### **Replication**

Normally, resources are restricted. Montgomery [3] recommends that unless some of the original independent variables can be omitted, running a single replicate of the design is sufficient. In our case, since we have no idea which factors are unimportant to drop, we will adopt the strategy of running only one replicate for the experiment. However, there are drawbacks associated with a single replicate design, discussed in Section 4.1.

## Blocking

Blocking is mainly used for improving the accuracy with which comparisons among the interested factors are made, and equivalently to further reduce the nuisance factors' variability [3]. In our case, since the experimental units are set in a general sense under a factor screening study, blocking is not possible.

After a careful decomposition of the problem, we have to analyze which statistical experiment to use. First, screening experimental designs are helpful in factor screening studies. Factorial designs are one of the screening experimental designs that are effective for characterizing the overall effect of several factors [2, 3]. Every possible combination of the levels of the factors is investigated in a factorial design in each replicate of the experiment, including interaction terms to avoid misleading conclusions.

Because we have two levels, say low and high, for each factor, the  $2^k$  factorial design seems to be good. In particular, the  $2^k$  design is especially helpful in the early stages of factor screening experiments [3]. The  $2^k$  design with only one replicate is also called an unreplicated factorial design [2]. In the  $2^k$  design, the coding scheme of  $\pm 1$  for all factors is called orthogonal coding, sometimes known as effects coding. There are several numerical advantages related to the design matrix  $\mathbf{X}$  using the orthogonal coding scheme, which is discussed in detail in Section 4.2. Since the current problem involves factor values from  $-1$  to  $1$ , orthogonal coding is an appropriate scheme for inputting factor values.

However, it may be impractical to run a single complete replicate screening 10 factors. Theoretically, we would need to run  $2^{10} = 1024$  times in a  $2^{10}$  full factorial experiment given a limited amount of resources. A full factorial design can be very wasteful in this case because of the sparsity of effects principle, which states that response variables are mostly determined by a small number of main effects and low-order interactions [2]. There is no internal estimate of the pure error for an unreplicated factorial experiment [2]. Hence, we analyze such an experiment by combining the mean squares of negligible high-order interactions to estimate the pure error, based on the sparsity of effects principle. To improve resource efficiency, subsets of the treatment combinations are considered. It is essential to note that for factorial designs, a carefully selected subset or fraction of the treatments can be used with little or no information lost from the main effects and key low-order interactions [2, 3]. The experimental design of a selected subset of treatment combination from a full factorial design is called a fractional factorial design. Fractional factorial designs allow researchers to examine a large number of factors with relatively few runs by confounding main and low-order interaction effects with ignorable higher-order interactions [2].

We searched the summary tables of useful fractional factorial designs presented by Box et al. [1] and Montgomery [3]. There are four useful fractional factorial designs for 10-factor factorial experiments. Namely, they are  $2_{\text{V}}^{10-3}$ ,  $2_{\text{IV}}^{10-4}$ ,  $2_{\text{IV}}^{10-5}$ , and  $2_{\text{III}}^{10-6}$  with the number of runs 128, 64, 32, and 16, respectively. In general, a  $2^{k-f}$  design represents a fractional factorial design with  $k$  factors and fraction  $f$ . Besides the number of runs and the fraction, the resolutions of the designs, written as the subscripts, are also different among the four designs. In this paper, we shall choose the experiment with the lowest runs due to the running cost. That is, we select the  $2_{\text{III}}^{10-6}$  fractional factorial design, the one-sixty-fourth fraction of the  $2^{10}$  factorial design. Of course, there are disadvantages of this design compared to the rest of the three designs. We comment and discuss the potential drawbacks and limitations of the  $2_{\text{III}}^{10-6}$  fractional factorial experimental design in Section 4.3. Table 8 in Appendix A.1 provides the alias relationship for the selected  $2_{\text{III}}^{10-6}$  fractional factorial design. Note that more than one aliases relationships exist depending on the design generators. The one we include in Table 8 and we shall use is the minimum aberration design. In practice, minimum aberration designs are preferred in particular when researchers want to perform a factor screening experiment with multiple factors using fewer runs [5].

## 2 Methods

### 2.1 Construction of the Design

The  $2_{III}^{10-6}$  fractional factorial design with the dependent variable values for the current problem is given in Table 1. A detailed alias relationship for the design is included in Table 8.

Table 1: A  $2_{III}^{10-6}$  Design for the Problem

Run	Basic Design				$E$	$F$	$G$	$H$	$I$	$J$	$Y$
	$A$	$B$	$C$	$D$	$(ABC)$	$(BCD)$	$(ACD)$	$(ABD)$	$(ABCD)$	$(AB)$	
1	−	+	+	+	−	+	−	−	−	−	68719.99
2	+	+	+	+	+	+	+	+	+	+	34907.43
3	−	+	−	−	+	+	−	+	−	−	31080.38
4	+	−	+	+	−	−	+	−	−	−	35371.26
5	+	−	−	+	+	+	−	−	+	−	32897.43
6	+	−	−	−	+	−	+	+	−	−	36039.24
7	+	−	+	−	−	+	−	+	+	−	31676.99
8	−	−	+	−	+	+	+	−	−	+	61663.00
9	+	+	−	−	−	+	+	−	+	+	27566.36
10	+	+	−	+	−	−	−	+	−	+	37931.20
11	−	+	+	−	−	−	+	+	+	−	30249.03
12	−	−	−	−	−	−	−	−	+	+	32145.47
13	−	+	−	+	+	−	+	−	+	−	-8863.36
14	−	−	+	+	+	−	−	+	+	+	34265.42
15	+	+	+	−	+	−	−	−	−	+	38565.08
16	−	−	−	+	−	+	+	+	−	+	40734.15

### 2.2 Preliminary Examination

It is essential to inspect the magnitude and direction of factor effects to observe which factors are possibly active in experiments involving  $2^k$  designs [2]. Before undertaking any formal analysis procedure to investigate the factor level effects, it is normally useful to informally examine the factor effects from estimated factor level means plots, such as a main effects plot [2]. If the line is horizontal to the  $x$ -axis in the main effects plot, there is no main effect. In contrast, if the line is not horizontal, the main effect arises with the steeper the slope, the bigger the main effects magnitude.

One disadvantage of using the main effects plot is that we cannot formally conclude the statistical significance of the differences between factor level means. Additional details are discussed in Section 4.4. Therefore, we use stepwise selection, also called stepwise regression, to increase the certainty of the justification of significant factors. Stepwise regression is an algorithm to find the subset of variables that gives the best model fit, using a selected criterion. We also follow the recommendation from Montgomery et al. [4] to perform the stepwise regression algorithm and then backward elimination to compare the selection result. We use the Bayesian information criterion (BIC) to assess the factors' performance in stepwise regression.

Furthermore, the analysis of variance can also typically be used to verify the interpretation from the plots. On the other hand, we should be cautious about the interpretation of the main effects from ANOVA because ANOVA alone does not convey the complete conclusion for factors having a significant difference [2]. Therefore, we consider the magnitude and direction of effects from the informal examination along with the ANOVA result to make the best decision.

## 2.3 General Analysis Procedure

The use of resolution III designs for initial screening experiments is generally followed by additional experiments on the active factors identified previously [2]. More specifically, suppose we have  $l$  factors justified as important in Section 2.2, we then study a follow-up  $2^l$  complete factorial experiment to investigate the interactions among these  $l$  factors. The interaction effects matrix, a matrix of all two-factor interaction plots, provides an informal picture to understand how the relationship between the factors and the response changes by the factor interactions. For each plot in the matrix, if the two lines are parallel, no interaction occurs. Oppositely, intersecting lines indicate an interaction. The more nonparallel the lines are, the more significant the strength of the interaction. Formally, we still require results from ANOVA to consolidate the conclusion of the statistical significance in interaction effects.

Once we have determined if interactions should be added to the model, we can proceed to build the final model for the problem. Model validation, experiment underlying assumptions and model adequacy checking should be performed to ensure model appropriateness.

## 2.4 ANOVA Diagnostics and Remedial Measures

Similar to regression models, departures from ANOVA model are likely to occur. Diagnosis of departures procedures is also similar to regression models such as residual plots and normal probability plots. Each of the possible departures, guided by Kutner et al. [2], should be studied and diagnosed so that the appropriateness of the models is examined. For instance, the residual plot should be structureless if the ANOVA model is adequate; the normal probability plot will resemble a nearly straight line if the residuals underlying error distribution is normal; residual plot against fitted values and normal probability plot can also be used to detect outliers [3].

Note that some of the departures do not significantly impact the analysis of variance while some can have serious effects. For example, ANOVA is robust to normality, and thus moderate departures from normality are not often an issue; what is more, the  $F$ -test is only slightly affected by the nonconstancy of variance in the balanced fixed effects model. Finally, lack of dependence of the error terms has a drastic negative effect on statistical inference for all ANOVA models [2, 3]. Even if inadequacies and violations of the model underlying assumptions occur, multiple remedial measures can be done. For instance, a standard remedial measure of heterogeneity of variances is to use weighted least squares [2]. Transformation of the response can be done to fix the nonconstancy of variance and nonnormality of the error term distribution [2, 3].

## 2.5 Statistical Inference

After a verification that the final model is appropriate for the current problem, we can make statistical inferences such as hypothesis testing and interval estimation. In particular, researchers are highly interested in the confidence intervals on individual regression coefficients [2, 4]. Regression model  $F$ -test and confidence/prediction intervals on the mean response are also important.

# 3 Results

In this section, we shall provide the numerical and graphical results based on the analysis procedures discussed in Section 2.

## 3.1 Preliminary Examination

Preliminary analysis plots and the ANOVA tables illustrated in Section 2.2 are shown here. First, we have the main effects plot in Figure 1, implemented in Appendix A.4:

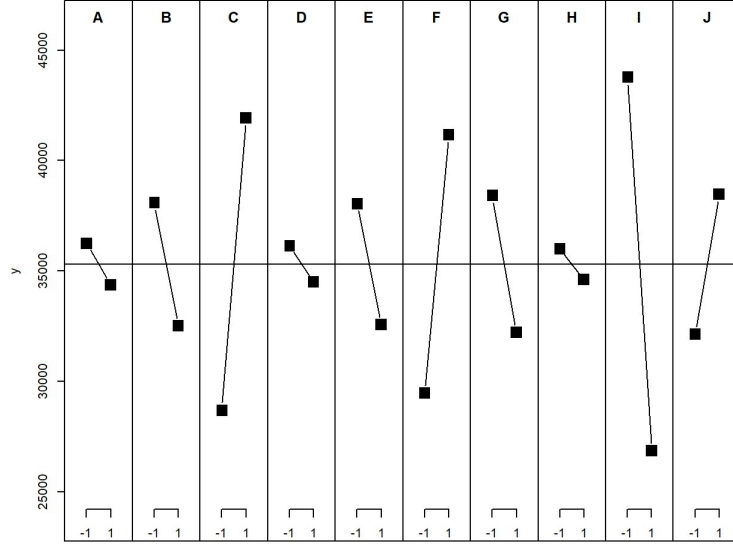


Figure 1: Main Effects Plot

The factors that are most obviously different, visually speaking, are  $C$ ,  $F$ , and  $I$ . Again, we need additional tests to solid the argument. In particular, both stepwise regression and backward selection procedures indicate that  $C$ ,  $F$ , and  $I$  are statistically significant in the ANOVA model containing all first-order terms. Implementation details are provided in Appendix A.3. Moreover, we have the following ANOVA results for first-order terms in Table 2, implemented in Appendix A.2. Note that the ANOVA  $p$ -values for all factors are not significant at 5 percent significance level. On the other hand, the  $p$ -values for factors  $C$ ,  $F$ , and  $I$  are close enough to 0.05, compared to the rest whose  $p$ -values deviate heavily. Hence, we justify that factors  $C$ ,  $F$ , and  $I$  are active.

Table 2: Analysis of Variance for all First-Order Factors

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	$F_0$	P-Value
A	14,135,880	1	14,135,880	0.074	0.797
B	124,527,952	1	124,527,952	0.648	0.457
C	700,758,290	1	700,758,290	3.646	0.114
D	10,598,320	1	10,598,320	0.055	0.837
E	120,120,664	1	120,120,664	0.625	0.465
F	546,886,110	1	546,886,110	2.846	0.152
G	153,852,120	1	153,852,120	0.801	0.412
H	7,813,926	1	7,813,926	0.041	0.848
I	1,143,446,322	1	1,143,446,322	5.950	0.059
J	160,067,701	1	160,067,701	0.833	0.403
Error	960,963,779	5	192,192,756		
Total	3,943,171,064	15			

To conclude the preliminary examination, we have the following model specification in Table 3 illustrating which variables should be included in or excluded from the analysis done previously.

Table 3: Model Specification

Variables	A	B	C	D	E	F	G	H	I	J
Model	Out	Out	In	Out	Out	In	Out	Out	In	Out

### 3.2 General Analysis Procedure

After determining the three active factors, we proceed to do a follow-up  $2^3$  factorial experiment to investigate the interactions, as suggested in Section 2.3. We first visually examine the interaction effects matrix for the three factors. The plots are shown in Figure 2, implemented in Appendix A.4. The factor names are printed on the diagonal of the plot matrix. Note that the two lines for each plot (interaction) are unparallel. This may indicate a statistically insignificant conclusion for the interaction effects to be added to the model.

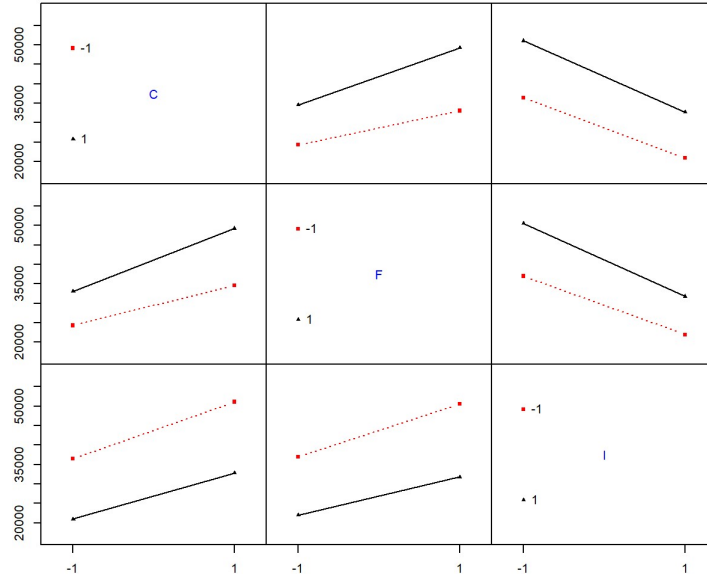


Figure 2: Interaction Effects Matrix

To formally test the significance of the interaction terms, we build a three-factor ANOVA model for the three factors. The pooling of three-factor interaction is used here because we assume three-factor interactions are small concerning the main and two-factor interaction terms for unreplicated two-level studies. The complete ANOVA results with a decomposition of interaction sum of squares are shown in Table 4 and 5. The implementation code is included in Appendix A.2. The  $p$ -value of the interaction shows the insignificance to be added to the model. Hence, our final model contains the three first-order terms without interactions.

Table 4: Analysis of Variance for Significance of Regression and Interaction

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	$F_0$	$P$ -Value
Regression	2,391,090,722	3	797,030,241	4.796	0.029
Interaction	56,438,583	3	18,812,861	0.113	0.950
Error	1,495,641,759	9	166,182,418		
Total	3,943,171,064	15			

Table 5: Decomposition of Interaction Sum of Squares

Source of Variation	Sum of Squares	Degrees of Freedom
$C \times F$	34,488,777	1
$C \times I$	7,813,926	1
$F \times I$	14,135,880	1

The regression version of the three-factor ANOVA model can be stated as follows:

$$Y_i = \beta_0 + \beta_1 C_i + \beta_2 F_i + \beta_3 I_i + \epsilon_i. \quad (1)$$

The implementation in Appendix A.2 outputs the following regression equation:

$$\hat{Y}_i = 35309.318 + 6617.960 C_i + 5846.399 F_i - 8453.721 I_i. \quad (2)$$

The two widely used methods for assessing the overall adequacy of the model are multiple R-squared and adjusted R-squared. They are 0.6064 and 0.508, respectively. Tables 6 and 7 provide additional details on the significance of regression and its decomposition into sums of squares.

Table 6: Analysis of Variance for Significance of Regression

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	$F_0$	P-Value
Regression	2,391,090,722	3	797,030,241	6.162	0.00887
Error	1,552,080,342	12	129,340,028		
Total	3,943,171,064	15			

Table 7: Decomposition of Regression Sum of Squares

Source of Variation	Sum of Squares	Degrees of Freedom
$C$	700,758,290	1
$F$	546,886,110	1
$I$	1,143,446,322	1

### 3.3 ANOVA Diagnostics and Remedial Measures

Model adequacy checking using the classical graphical methods enables researchers to decide if remedial measures are needed to improve the model effectiveness. Figure 3 is the diagnostic residual versus fitted value plot. The diagnostic normal probability plot is provided in Figure 4. Appendix A.5 shows the code to generate the plots.

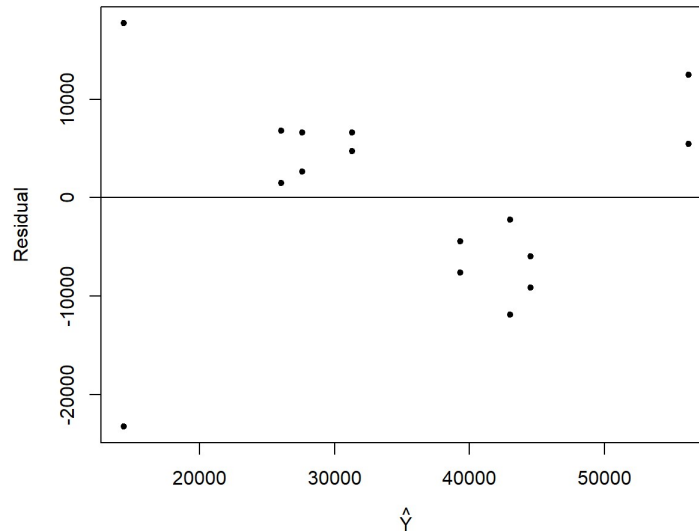


Figure 3: Residual against the Fitted Values

As discussed previously in Section 2.4, we shall investigate every possible departure from the ANOVA model suggested by Kutner et al. [2] to finish ANOVA diagnostics.



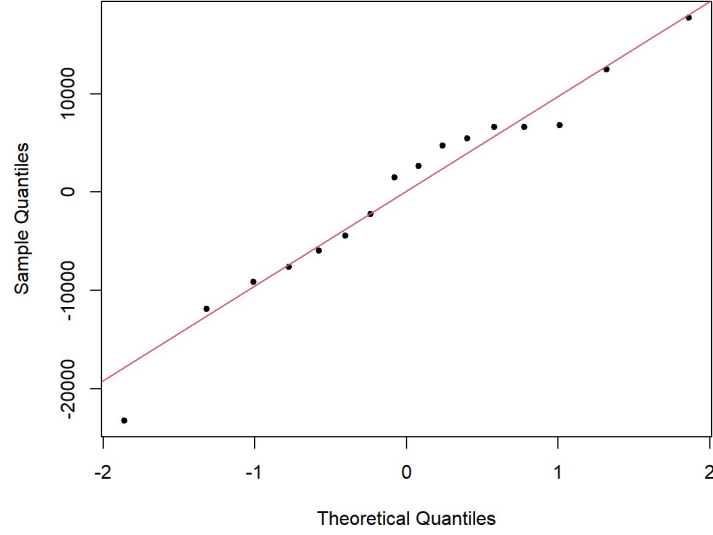


Figure 4: Normal Q-Q Plot

### Nonconstancy of Error Variance

The residual plot against the fitted values in Figure 3 shows no unusual pattern. Especially, since the sample size is comparatively small, it is normal that the residual scatters around zero for each factor level are slightly different. Thus, we claim the error variance is constant.

### Nonindependence of Error Terms

Because randomization is enforced in the experiment, the assumption of error term independence is not violated. What is more, the nonindependence of error terms normally requires verification for experiments with time-related effects [2, 3].

### Outliers

Residual plots versus fitted values facilitate the detection of outliers. In particular, Figure 3 reveals a structureless pattern. Moreover, due to the relatively small sample size for the resource constraints, a few observations that are seemingly different than other observations may be normal. Finally, Kutner et al. [2] recommend discarding possibly outlying observations only when the observations are identified as incorrectly measured or entered. In our case, we have no evidence to delete any values and thus, no remedial measures are needed to take place.

### Omission of Important Explanatory Variables

This assumption automatically holds because of the background given in the current problem. If an investigation is required, residual analysis can also be applied to study if the model is adequate [2]. The residual plot shows no apparent assumption-deviation pattern.

### Nonnormality of Error Terms

The normal Q-Q plot in Figure 4 depicts a nicely straight line on the central values of the plot. Since the number of observations is relatively small, it is not unusual to have one or two extreme residuals deviating slightly from the straight line. Thus, we conclude the normality assumption is not violated.

## 3.4 Statistical Inference

Here, we present several important statistical inference results, as discussed in Section 2.5. First, the  $p$ -value for the test of regression, from Table 6, is 0.00887. This implies that we reject the

null hypothesis that none of the regressors contribute significantly to the model at 5 percent level of significance. Furthermore, recall the point estimates for regression coefficients in Equation 2 are  $\hat{\beta}_0 = 35309.318$ ,  $\hat{\beta}_1 = 6617.960$ ,  $\hat{\beta}_2 = 5846.399$  and  $\hat{\beta}_3 = -8453.721$ , respectively. We obtain the following 95 percent confidence intervals for  $\beta_j$ s:

$$\begin{aligned} 29114.53 &\leq \beta_0 \leq 41504.11 \\ 423.173 &\leq \beta_1 \leq 12812.75 \\ -348.388 &\leq \beta_2 \leq 12041.19 \\ -14648.51 &\leq \beta_3 \leq -2258.93 \end{aligned}$$

We can predict the mean of the response by the settings of the control variables. For example, suppose we are interested in the response for  $C = 1$ ,  $F = -1$ , and  $I = 1$ . We have  $\hat{Y}_h = 27627.158$ . The 95 percent prediction interval for  $Y_{h(\text{new})}$ :  $-76.774 \leq Y_{h(\text{new})} \leq 55331.09$ . The 95 percent confidence interval for  $\mu_Y$ :  $15237.58 \leq \mu_Y \leq 40016.73$ . The calculation implementation is shown in Section A.6.

## 4 Discussion

The main advantage of applied statistical methods is to increase objectivity in a variety of decision-making scenarios [3]. In this paper, we have been improving our confidence in designing an experiment and modeling the given data. However, the conclusions do not prove factors  $C$ ,  $F$ , and  $I$  have particular effects. They only serve as guidelines to increase the reliability and validity of results. The following subsections provide additional insights into different perspectives of our research.

### 4.1 Unreplicated Two-Level Studies

Montgomery [3] proposes risk of conducting an unreplicated experiment is that the fitting model may be pure noise. That is, the estimated factor effects may be close to zero. Therefore, the experimenter may falsely conclude the insignificance of the factors. If the variability of the response is smaller, the probability of an erroneous conclusion will be smaller [3].

### 4.2 Advantages of Orthogonal Coding

Recall that  $\mathbf{X}$  is the regressor matrix, also called the design matrix, storing all factor values. When a balanced factorial experiment is carried out at two levels for each factor and a  $\pm 1$  orthogonal coding is employed, the  $\mathbf{X}^T \mathbf{X}$  matrix is greatly simplified. Consider the  $2^k$  factorial design we have discussed with one replication in Section 1.2. The  $\mathbf{X}^T \mathbf{X}$  matrix is simplified because:

- Any two columns of the  $\mathbf{X}$  matrix are orthogonal; that is,  $\mathbf{X}_q^T \mathbf{X}_{q'} = 0$ .
- The sum of squares of the elements in each column,  $\mathbf{X}_q^T \mathbf{X}_q$ , is always  $2^k$ . It leads for balanced two-level factorial designs to a diagonal  $\mathbf{X}^T \mathbf{X}$  matrix.

Thus, the least squares and maximum likelihood estimators of  $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$  can be written in a simple form of  $\frac{1}{2^k} \mathbf{X}^T \mathbf{Y}$  [2]. Furthermore, if we write the regression model as  $\mathbf{Y} = \mathbf{X}\beta + \epsilon = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \epsilon$ , the sum of squares of  $\beta_1$  and the sum of squares of  $\beta_2$  measure the contribution of the regressors in  $\mathbf{X}_1$  and  $\mathbf{X}_2$ , respectively, to the model unconditionally. In other words, if the independent variables are orthogonal, we can uniquely determine each variable's effects on the model. Because of this, researchers are often willing to have orthogonal coding regressors in their data collection experiments [4].

### 4.3 Resolution

There is a trade-off between the resolution level and the number of observations in a study. For higher resolutions, the assumptions of including unimportant interactions in the model become less restrictive [3]. In this paper, we used the one-sixty-fourth fraction of the  $2^{10}$  factorial design with resolution III in 16 runs. It is of interest for experimenters to choose the highest possible resolution of design [2, 3]. Recall the possible alternatives we have are  $2_V^{10-3}$ ,  $2_{IV}^{10-4}$  and  $2_{IV}^{10-5}$  with 128, 64 and 32 runs, respectively. Therefore, for future research, we can choose a  $2_V^{10-3}$  design if we believe the loss of sacrificing higher-order interactions information outweighs the cost of runs, given sufficient resources.

### 4.4 Plots of Estimated Factor Level Means

Kutner et al. [2] suggest the following four types of plots to informally inspect the factor level effects: a line plot, a bar graph, a main effects plot, and a normal probability plot. We use the main effects plot to visualize. Remark that the four plots themselves do not provide any information on the standard errors. Thus, we cannot make a solid conclusion that the differences are statistically significant between the factor level means. Bar-interval graph and interval plot are enhanced bar graph and main effects plot, respectively, such that the confidence limits for each factor level mean are included [2]. Also, note that a normal probability plot is a suitable informal examination when the sample sizes are equal and the number of factors is sufficiently large, such as greater than or equal to 10 [2]. Therefore, we can also adopt the normal probability plot in this paper to visualize the factor level means.

## A Technical Implementation in R

### A.1 Factorial Design Generation

Computer statistical software is broadly used to assist experimenters in choosing and building experimental designs [3]. The following code builds a  $2_{III}^{10-6}$  statistical design in random order to follow the randomization principle discussed in Section 1.2. The complete alias relationship for the design used in this paper is included in Table 8.

```
dae.r

library(FrF2)
dsg <-
  FrF2(nruns = 16, nfactors = 10, factor.names = LETTERS[1:10],
       generators = c("ABC", "BCD", "ACD", "ABD", "ABCD", "AB"),
       seed = 123)
summary(dsg)
```

### A.2 Analysis of Variance and Linear Model

Suppose we have input the dataset in Table 1 named data in the code. We can obtain ANOVA and linear model results using functions `aov()` and `lm()`, respectively. Table 2 is the ANOVA output from `anova(aov1)`; Tables 4 and 5 are from `anova(aov2)`; Tables 6 and 7 are from `anova(aov3)`. `summary()` provides summaries of model fitting functions such as model function coefficients,  $F$ -statistic and its  $p$ -value. Equation 2 can be obtained from the summary of `lm3`.

Table 8: Alias Relationship for  $2^{10-6}$  Fractional Factorial Design $2^{10-6}$ ; 1/64 fraction of 10 factors in 16 runs

Resolution III

Design Generators

$$E = ABC \quad F = BCD \quad G = ACD \quad H = ABD \quad I = ABCD \quad J = AB$$

Defining relation

$$\begin{aligned}
0 &= ABCE = BCDF = ADEF = ACDG = BDEG = ABFG = CEFG = ABDH \\
&= CDEH = ACFH = BEFH = BCGH = AEGH = DFGH = ABCDEFGH = ABCDI \\
&= DEI = AFI = BCEFI = BGI = ACEGI = CDFGI = ABDEFGI = CHI \\
&= ABEHI = BDFHI = ACDEFHI = ADGHI = BCDEGHI = ABCFGHI \\
&= EFGHI = ABJ = CEJ = ACDFJ = BDEFJ = BCDGJ = ADEGJ = FGJ \\
&= ABCEFGJ = DHJ = ABCDEHJ = BCFHJ = AEFHJ = ACGHJ = BEGHJ \\
&= ABDFGHJ = CDEFGHJ = CDIJ = ABDEIJ = BFIJ = ACEFIJ = AGIJ \\
&= BCEGIJ = ABCDFGIJ = DEFGIJ = ABCHIJ = EHII = ADFHIJ \\
&= BCDEFHIJ = BDGHIJ = ACDEGHIJ = CFGHIJ = ABEFGHIJ
\end{aligned}$$

Aliases

$$\begin{aligned}
A &= FI = BJ & I &= DE = AF = BG = CH \\
B &= GI = AJ & J &= AB = CE = FG = DH \\
C &= HI = EJ & AC &= BE = DG = FH \\
D &= EI = HJ & AD &= EF = CG = BH \\
E &= DI = CJ & AE &= BC = DF = GH \\
F &= AI = GJ & AG &= CD = BF = EH = IJ \\
G &= BI = FJ & AH &= BD = CF = EG \\
H &= CI = DJ
\end{aligned}$$

2 blocks of 8:  $AG = CD = BF = EH = IJ$ 

dae.r

```

aov1 = aov(y ~ ., data = data)
anova(aov1)

aov2 = aov(y ~ C*F*I, data = data)
anova(aov2)

aov3 = aov(y ~ C + F + I, data = data)
anova(aov3)

lm3 = lm(y ~ C + F + I, data = data)
summary(lm3)

```

### A.3 Stepwise Regression

Stepwise selection and backward elimination can be done by the function `step()`. Note that, by default, R uses the Akaike information criterion (AIC). That is,  $k = 2$  is used by default in the following code, where  $k$  is the multiple of the number of degrees of freedom used for the penalty. To use the BIC, we simply have  $k = \log(\text{nrow}(\text{data}))$  instead.

```
dae.r

step(aov1, k = log(nrow(data)))

step(aov1, direction = "backward", k = log(nrow(data)))
```

### A.4 Main Effects Plot and Interaction Effects Matrix

FrF2 package has two useful functions, `MEPlot()` and `IAPlot()`, to graphically display the main effects plot and the interaction plots, shown in Figure 1 and Figure 2, respectively.

```
dae.r

MEPlot(aov1, main = NULL)

IAPlot(aov2, main = NULL)
```

### A.5 Residual Analysis

Residual analysis, discussed in Section 2.4, is one the most important procedure to evaluate model adequacy in applied statistics. We present a residual plot versus the fitted values shown in Figure 3 and a normal probability plot of the residuals in Figure 4. `resid()` calculates the residuals for all observations from the linear model. The `plot()` and `abline()` functions plot the fitted values of the linear model with the residuals to visualize the random pattern. The normal Q-Q plot is created using the traditional commands `qqnorm()` and `qqline()`.

```
dae.r

res = resid(lm3)

plot(fitted(lm3), res, xlab = expression(hat(Y)),
     ylab = "Residual", pch = 20)
abline(0, 0)

qqnorm(res, main = NULL, pch = 20)
qqline(res, col = 2)
```

## A.6 Confidence and Prediction Intervals

Statistical inference results can be obtained in R. For confidence intervals of regression coefficients, `confint()` directly provides the  $(1 - \alpha)\%$  confidence intervals for all coefficients (95 percent confidence interval by default). To obtain confidence and prediction intervals of the mean of a new response, we first predict the response given the factor values. Then the interval estimations are given using `predict()` with the type of interval estimation specified. The following code retrieves the predictions of all possible combinations of the factor values with interval estimations returned.

```
dae.r

confint(lm3)

pred_data <- expand.grid(C = c(-1, 1), F = c(-1, 1), I = c(-1, 1))
predict(lm3, newdata = pred_data, interval = "confidence")
predict(lm3, newdata = pred_data, interval = "prediction")
```

## References

- [1] G. E. P. Box, J. S. Hunter, and W. G. Hunter. *Statistics for Experimenters: Design, Innovation, and Discovery*. Wiley, 2nd edition, May 2005.
- [2] M. H. Kutner, C. Nachtsheim, J. Neter, and W. Li. *Applied Linear Statistical Models*. McGraw-Hill, 5th edition, November 2004.
- [3] D. C. Montgomery. *Design and Analysis of Experiments*. Wiley, 10th edition, July 2019.
- [4] D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to Linear Regression Analysis*. Wiley, 5th edition, April 2012.
- [5] P. C. Wang. A simple method for obtaining minimum aberration designs. *Communications in Statistics - Theory and Methods*, 39(18):3363–3370, September 2010.