Contributions to the Planning and Analysis of Factorial Designs

Rafel Xampeny Solaní

PhD Thesis directed by

Dr. Pere Grima Cintas

Department of Statistics and Operations Research
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Thesis presented in fulfillment of the requirements of the Degree of Doctor by the Universitat Politècnica de Catalunya – BarcelonaTech

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

The thesis is presented as a compendium of five articles:

1. ***SELECTING SIGNIFICANT EFFECTS IN FACTORIAL DESIGNS: LENTH'S METHOD VERSUS USING NEGLIGIBLE INTERACTIONS***
   Communications in Statistics - Simulation and Computation
   Received: 14 November 2016, Accepted: 22 March 2017, Accepted author version posted online: 30 March 2017, Published online: 28 June 2017
   DOI: 10.1080/03610918.2017.1311917
   http://www.tandfonline.com/doi/abs/10.1080/03610918.2017.1311917

2. ***ESTIMATING MISSING VALUES FROM NEGLIGIBLE INTERACTIONS IN FACTORIAL DESIGNS***
   Quality and Reliability Engineering International, 2017; 0: 1–13
   Received: 21 December 2016, Revised: 18 April 2017, Accepted: 21 April 2017
   DOI: 10.1002/qre.2172
   https://doi.org/10.1002/qre.2172

3. ***WHICH RUNS TO SKIP IN TWO LEVEL FACTORIAL DESIGNS WHEN NOT ALL CAN BE PERFORMED***
   Quality Engineering
   Accepted author version posted online: 18 January 2018
   DOI: 10.1080/08982112.2018.1428751
   https://doi.org/10.1080/08982112.2018.1428751

4. ***CONSEQUENCES OF USING ESTIMATED RESPONSE VALUES FROM NEGLIGIBLE INTERACTIONS IN FACTORIAL DESIGNS***
   Submitted under consideration of an indexed JCR Journal: 18 December 2017

5. ***SELECTING SIGNIFICANT EFFECTS IN FACTORIAL DESIGNS: LENTH’S METHOD VERSUS BOX-MEYER APPROACH***
   Submitted under consideration of an indexed JCR Journal: 5 March 2018

To my father and mother, Rafel and Elena

To my wife and daughter, Comfort and Anna
To my brother, Ramon

ACKNOWLEDGMENTS

I would like to thank all those people who have collaborated in one way or another to the realization of this thesis. First and foremost, to my thesis director Dr. Pere Grima. A statistician with an extraordinary talent. He has supported me throughout the process, understanding my personal situation. He is a person with an outstanding quality in all senses, both professional and personal. Without him it wouldn’t have been possible.

I must also sincerely thank Dr. Xavier Tort-Martorell for his essential contribution and in fact I could say that thanks to him I am in this wonderful world of Statistics.

I would like to express my special gratitude to all professors and colleagues of the department, since I have received from all of them and without exception their knowledge in this matter.

Finally, I would like to thank my whole family for their patience with me along this path. Especially my father and my mother.

I do not have enough space on this page to express my gratitude to all of them.

ABSTRACT

The thesis is about contributions to the planning and analysis of factorial designs.

Article 1: Selecting significant effects in factorial designs: Lenth's method versus using negligible interactions

Among the many analytical techniques that have been published to analyze the significance of the effects in the absence of replications, two have emerged as the most widely used in text books as well as statistical software packages: The Lenth's method and the estimation of the variance of the effects from the values of those considered negligible. This article shows that neither is better than the other in all cases, and by analyzing the results obtained in a wide variety of situations it provides guidelines on when it is preferable to use one or the other technique.

Article 2: Estimating missing values from negligible interactions in factorial designs

When a factorial design has some missing value but there are contrasts (at least as many as there are missing values) corresponding to interactions that can be considered negligible, equating the expressions of those contrasts to zero allows the missing values to be deduced. This procedure allows estimating values that correspond to runs that have gone wrong or could not be performed as planned, as well as to reduce the experimental plan while saving some runs whose results can be estimated a posteriori. The problem is that the estimated values have a variance that is higher than that of the experimental runs, and the variance depends on which are the missing values and the contrasts used to make the estimate. This article performs a thorough analysis of the variances obtained by estimating up to two missing values in 8-run designs and up to five in 16-run designs, with tables included which provide the missing values and contrasts that should be used to obtain estimates that have minimal variability.

Article 3: Which runs to skip in two level factorial designs when not all can be performed

When a two level factorial design allows estimating contrasts that can be considered negligible from scratch, it is possible to omit some runs and later estimate their values by equating to zero the expressions of some of that contrasts. This article presents the combinations of runs to be omitted in 8 and 16 runs two level factorial designs so that the responses can be estimated in such a way as to produce the least possible impact on the desired properties of the estimated contrasts: low and equal variance and the smallest possible correlation among them.

Article 4: Consequences of using estimated response values from negligible interactions in factorial designs

This article analyzes the increase in the probability of committing type I and type II errors in assessing the significance of the effects when some properly selected runs have not been carried out and their responses have been estimated from the interactions considered null from scratch. This is done by simulating the responses from known models that represent a wide variety of practical situations that the experimenter will encounter; the responses considered to be missing are then estimated and the significance of the effects is assessed. Through comparison with the parameters of the model, the errors are then identified. To assess the significance of the effects when there are missing values, the Box-Meyer method has been used. The conclusions are that 1 missing value in 8 run designs and up to 3 missing values in 16 run designs experiments can be estimated without hardly any notable increase in the probability of error when assessing the significance of the effects.

Article 5: Selecting significant effects in factorial designs: Lenth's method versus the Box-Meyer approach

The Lenth method is conceptually simple and probably the most common approach to analyzing the significance of the effects in non-replicated factorial designs. Here, we compare it with a Bayesian approach proposed by Box and Meyer and which does not appear in the usual software packages. The comparison is made by simulating the results of 4, 8 and 16 run designs in a set of scenarios that mirror practical situations and analyzing the results provided by both methods. Although the results depend on the number of runs and the scenario considered, the use of the Box and Meyer method generally produces better results.

Table of Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1

INTRODUCTION

The thesis is about contributions to the planning and analysis of factorial designs. It has a very practical approach to guide practitioners to get the most out of constraints such as time and resources they have to commonly deal with. In this line, the thesis aims on the one hand, to handle the problem of assessing factor's statistical significance when using un-replicated factorial designs. On the other hand, how to plan an experimental design without completing it, but getting almost the same information as if it was to be completed; deeply analyzing the pros and cons of doing it following this approach. Using the last technique, missing responses in factorial designs are properly imputed when repeating the experiments is unfeasible.

1.1 Overview

In front of the scarcity of resources and time that practitioners will find in the industry, one of the most common factorial design they will probably have to bargain with is the un-replicated factorial design. Block I is devoted to analyze how we should properly discriminate active from non-active effects in this specific situation.

Block II investigate, with the same constrains as in the first, more complicated scenarios that can arise:

- To have missing responses within a factorial design and to appropriately impute them.
- To plan factorial designs knowing in advance that not all runs will be able to be performed.

1.1.1 Block I: Analyze the significance of the effects in un-replicated Factorial Designs

In the industry, the most used designs are the factorial designs at two levels, that is, 2^k designs. On the one hand we have the variable we are trying to optimize, that is our response variable, and on the other hand the factors that affect the response. The influence of the factors over the response is quantified by calculating the effects, which are orthogonal contrasts of the response vector.

When we know that the system of experimentation has high variability, several experiments are made under each experimental condition and they are called replicates. When we have replicates, we can estimate the experimental error from them, and thus, an estimator of the variance of the effects that will allow us to discriminate which effects are actives and which are not, so that important factors will be recognized.

However, given that the resources for experimentation are usually limited, replicates are typically lacking. In these cases, it is necessary to analyze the significance of the effects using other methods, which can be graphical or analytical.

We focus on analytical methods because they can be automated and can be used without the intervention of human judgment, which graphic methods need, and thus implemented in statistical software packages.

Among the different analytical methods, Lenth's method stands out for its simplicity and efficiency. However, it is known that Lenth's original method produces a low percentage of type I error with the counterpart of an increase in the type II error, probably the most important in industrial applications.

With the aim of solving this drawback of the Lenth’s method, and based on a computer simulation that guarantees what an experimenter can find in reality, as regards of the number and magnitude of active effects in factorial designs of 4, 8 and 16 runs (the most used in industrial statistics), we analyze two other techniques that we think can complement it in its weakness and improve it:

A. Estimation of the variance of the effects by using interactions that can be considered negligible from the origin.
B. Box and Meyer [1986, 1993] method.

Article 1 (***Selecting significant effects in factorial designs: Lenth’s method versus using negligible interactions***, **http://www.tandfonline.com/doi/abs/10.1080/03610918.2017.1311917**), compares Lenth’s method with method A. This also is not a sophisticated method and uses the fact that high interaction terms (from 3-way going) can reasonably assumed to be negligible and thus we can estimate the variance of the effects from them.

Article 5 (***Selecting significant effects in factorial designs: Lenth's method versus the Box-Meyer approach***) compares Lenth’s method with respect to Box and Meyer alternative, method B. It uses Bayesian Analysis to discriminate active from inert effects. The prior probability of an effect being active has to previously be fixed and this could be done by the knowledge of the expert in the field. As a result, it gives the model with highest posterior probability distinguishing active from non-active effects. It is mathematically more complicated but nowadays this wouldn’t be a restriction given the fact that computational time is not a constraint anymore in this case.

After analyzing the percentages of type I and type II errors obtained through this simulation we are able to conclude when it is better to use each one and when we should avoid using them.

Statistical software packages use by default only one method. We have proved that the best option is to have all them available and automatically, or by the knowledge of the experimenter, choose the best according to the faced scenario. Not forgetting that "practical significance", given in many cases by normal probability plots, technical knowledge and common sense, will solve the discrepancy.

1.1.2 Block II: Planning a Factorial Design in order to save runs or deal with missing response values

Experiments are not a one-shot operation, experiments are never run in one go. According to Box *et al.* [2005]:

- "The best time to run an experiment is after the experiment. You will discover things from the previous experiment that you wish you had considered the first time around.
- For the above reason, do not spend more than 20% to 25% of your time and budget on your first group of experiments. Keep some time aside to add more experiments and learn more about the system."

At some point of this sequential strategy we might be in the situation of encountering missing responses from an already performed factorial design. When a factorial design has some missing value but there are contrasts (at least as many as there are missing values) corresponding to interactions that can be considered negligible, equating the expressions of those contrasts to zero allows the missing values to be deduced. One of the problems is that the variance of the responses increases, with respect to the one of the complete design, as the number of missing values does, as expected.

Article 2 (***Estimating missing values from negligible interactions in factorial designs*, https://doi.org/10.1002/qre.2172**), performs a thorough analysis of the response variances obtained by estimating up to two missing values in 8-run designs and up to five in 16-run designs. There are tables included, which provide the missing values and contrasts that should be used to obtain estimates that have minimal variability.

To plan factorial designs knowing in advance that not all runs will be able to be performed is the second scenario we analyze. We need to pay a huge caution though, especially when it comes to imputing missing values. If this is a designed experiment, each trial is chosen to have high influence on the results. After all, you are trying to get maximum information from minimal number of data points. A part from the increase in the response variance, the effects are affected as well. Moreover, they lose the orthogonality property. In order to have the less impact on the model, you need to find appropriate combinations of responses that guarantee minimum effect's variance and minimum correlation between effects, and this will be appropriately analyzed in **Article 3** (***Which runs to skip in two level factorial designs when not all can be performed*, https://doi.org/10.1080/08982112.2018.1428751**). After performing this deep analysis, this article presents the combinations of runs to be omitted in 8 and 16 runs two level factorial designs.

Finally, it is needed to quantify the mentioned impact on the model when we estimate response values from negligible interactions. As mentioned before, the effects lose the orthogonally property, thus they are not independent anymore. Among the three methods utilized in Block I to asses effect's significance in un-replicated factorial designs, since we have used interactions that can be considered negligible to estimate the missing responses, method A cannot be used. Lenth's method requires independency among the contrasts. Box and Meyer method is used because it accepts some degree of correlation among them. **Article 4** (***Consequences of using estimated response values from negligible interactions in factorial designs***), carries out this analysis in a similar way as in Block I, exposing the consequences.

1.2 Thesis Outline

Articles are chronologically displayed as they were delivered to the journals. The first three papers (Chapters 2 to 4) have already been published and the last two (Chapters 5 and 6) are currently under revision by JCR indexed journals. Chapter 7 is devoted to summarize the contributions and global results of the thesis and future lines of research.

Block I: Analyze the significance of the effects in un-replicated Factorial Designs

- CHAPTER 2:
  SELECTING SIGNIFICANT EFFECTS IN FACTORIAL DESIGNS: LENTH'S METHOD VERSUS USING NEGLIGIBLE INTERACTIONS
  Communications in Statistics - Simulation and Computation
  Received: 14 November 2016, Accepted: 22 March 2017, Accepted author version posted online: 30 March 2017, Published online: 28 June 2017
  DOI: 10.1080/03610918.2017.1311917
- CHAPTER 6:
  SELECTING SIGNIFICANT EFFECTS IN FACTORIAL DESIGNS: LENTH'S METHOD VERSUS BOX-MEYER APPROACH
  Submitted under consideration of an indexed JCR Journal: 5 March 2018

Block II: Planning a Factorial Design in order to save runs or deal with missing response values

- CHAPTER 3:
  ESTIMATING MISSING VALUES FROM NEGLIGIBLE INTERACTIONS IN FACTORIAL DESIGNS
  Quality and Reliability Engineering International, 2017; 0: 1–13
  Received: 21 December 2016, Revised: 18 April 2017, Accepted: 21 April 2017
  DOI: 10.1002/qre.2172
- CHAPTER 4:
  WHICH RUNS TO SKIP IN TWO LEVEL FACTORIAL DESIGNS WHEN NOT ALL CAN BE PERFORMED
  Quality Engineering
  Accepted author version posted online: 18 January 2018
  DOI: 10.1080/08982112.2018.1428751
- CHAPTER 5:
  CONSEQUENCES OF USING ESTIMATED RESPONSE VALUES FROM NEGLIGIBLE INTERACTIONS IN FACTORIAL DESIGNS
  Submitted under consideration of an indexed JCR Journal: 18 December 2017

Chapter 2

ARTICLE 1

Selecting significant effects in factorial designs: Lenth's method versus using negligible interactions

2

ARTICLE 1

Selecting significant effects in factorial designs: Lenth's method versus using negligible interactions

Rafel Xampeny, Pere Grima, Xavier Tort-Martorell
Department of Statistics and Operational Research
Universitat Politècnica de Catalunya – BarcelonaTech, Spain

Communications in Statistics - Simulation and Computation
Received: 14 November 2016, Accepted: 22 March 2017, Accepted author version posted online: 30 March 2017, Published online: 28 June 2017
DOI: 10.1080/03610918.2017.1311917
http://www.tandfonline.com/doi/abs/10.1080/03610918.2017.1311917

ABSTRACT

Among the many analytical techniques that have been published to analyze the significance of The effects in the absence of replications, two have emerged as the most widely used in text books as well as statistical software packages: The Lenth's method and the estimation of the variance of the effects from the values of those considered negligible. This article shows that neither is better than the other in all cases, and by analyzing the results obtained in a wide variety of situations it provides guidelines on when it is preferable to use one or the other technique.

KEY WORDS: Lenth's method, Effects significance, Factorial design, Statistical software, Negligible interactions

2.1 Introduction

When a 2^k experimental design is carried out, the influence of the factors (X_1, X_2, … , X_k) over the response (Y) is quantified through the so-called effects. Once the effects are calculated it is necessary to analyze which are significantly different from zero. If there are replicates it is possible to estimate the variance of the response and from it the variance of the effects (σ^2_{ef}) that can be used to assess significance.

In absence of replicates, there is no estimate for the variance, and thus significance has to be assessed by other methods. Several graphical and analytical methods have been developed to solve the problem. Among the graphical are: the Pareto chart of the effects based on the sparsity principle and the idea that significant effects will present values that will stand out from those that are not; and representing the effects in a Normal Probability Plot (NPP) or Half

Normal Plot based on the idea that non-active effects follow a $N(0,\sigma_{ef}^2)$, and thus will be aligned on a straight line. Graphical methods have the inconvenient that they require human judgment and cannot be easily automated.

A lot of analytical methods have been proposed, Hamada and Balakrishnan [1998] give a very complete and deep study of many of them. Here we present a very short review.

Lenth [1989] proposed a simple but effective procedure for estimating the standard deviation of the effects that will be explained in section 2. Other authors tried to improve this method adapting it to specific situations, for example, Dong [1993] proposed a procedure useful when the number of significant effects is low (less than 20%) and Juan and Peña [1992] when the number of significant effects is large (greater than 20%). The so-called "step by step" strategy proposed by Venter and Steel [1998], also tries to improve Lenth's method for the case of many significant effects. A more general approach is proposed by Ye *et al.* [2001] with their step-down version of the Lenth's method. Box and Meyer [1986, 1993] proposed a Bayesian approach. Recently, Espinosa *et al*. [2016], proposed a Bayesian sequential method based on posterior predictive checks to screen for active effects.

In spite of this wide variety of analytical methods, Lenth's original method and to estimate the variance of the effects by using interactions that can be considered negligible from the outset are the most widespread. They are explained in commonly used textbooks as Montgomery [2013] and Box *et al.* [2005], and implemented in the most common statistical software packages for industrial applications (Fontdecaba *et al.* [2014]). Accepting to reduce the problem to choose among these two methods, it seems reasonable to suppose that it will be more adequate to apply one or the other according to the characteristics of the situation; for instance, depending on the number of interactions that can be considered negligible. However, several statistical software packages use always by default the same method regardless of the characteristics of the case being analyzed (Fontdecaba *et al.* [2014]).

The aim of the paper is to characterize in which situations it is better to use each method, based on the results of a simulation under some general scenarios representing a wide variety of situations that may occur in real life experimentation. The number of Type I and Type II errors made by each method is used to compare them.

The paper is organized as follows: in the next section we describe the two methods to be compared: Lenth's method and variance estimation using negligible interactions. In Section 2.3 we describe the simulation scenarios. In Section 2.4 we show and analyze the obtained results. Finally we provide some conclusions and final remarks in Section 2.5.

2.2 Description of the compared methods

Lenth's method consists of estimating the standard deviation of the effects based on the fact that if $X \sim N(0, \sigma^2)$, the median of $|X|$ equals 0.645σ and therefore $1.5 \cdot \mathrm{median}|X| = 1.01\sigma \cong \sigma$. Assuming that κ_i $(i = 1, \dots, n)$ are the values of the effects of interest and that their estimators c_i are distributed according to a $N(\kappa_i, \sigma_{ef}^2)$, Lenth defines $s_0 = 1.5 \cdot \mathrm{median}|c_i|$ and calculates a new median excluding the estimated effects whith $|c_i| > 2.5s_0$. By doing so he expects to exclude the effects with $\kappa > 0$ and use the others to calculate the so-called Pseudo Standard Error:

$$PSE = 1.5 \cdot \underset{|c_i|<2.5s_0}{\mathrm{median}}|c_i| \tag{2.1}$$

From this PSE, a margin of error *(ME)* can be calculated. For a 95% confidence level it will be $ME = t_{0.975,\nu} \cdot PSE$. If $|c_i| > ME$ the effect c_i is considered active.

Lenth proposes that $\nu = n/3$ where n is the number of effects considered. This is the value that has been used in some software packages (e.g. Minitab) although it has been demonstrated to produce type I error probabilities under 5% with the additional and unavoidable inconvenience to produce bigger type II errors. Ye and Hamada [2000] and Fontdecaba *et al.* [2015] proposed t values that give better results (Table 2.1).

Table 2.1: Proposed values for $t_{0.975}$ that should be applied with the PSE

Design type	Proposed values for $t_{0.975}$		
	Lenth (1989)	Ye and Hamada (2000)	Fontdecaba *et al.* (2015)
8-run	3.76	2.297	2
16-run	2.57	2.156	2

The other option is to assume from the outset that some effects are negligible and to use their values c_j $(j = 1, \dots, m)$ to estimate their variance:

$$s_{ef}^2 = \frac{\sum_{j=1}^{m} c_j^2}{m} \tag{2.2}$$

In general, interactions of three or more factors are considered negligible. In some cases technical expertise of the phenomenon being studied allow to consider negligible some particular two factor interaction and add them to the ones used to estimate s_{ef}^2.

The problem is that the two analytical methods do not always give the same result. (Montgomery [2013], p. 279), taken from Bell *et al.* [2006] presents a 2^4 design which was conducted to test new ideas to increase direct mail sales of credit cards. The response is the number of orders obtained and the factors are related to the offered conditions (Table 2.2).

Table 2.2: 2^4 *design with the results obtained from Montgomery (2013) p. 279*

A Annual fee	B Account-opening fee	C Initial interest rate	D Long-term interest rate	Y Orders
–	–	–	–	184
+	–	–	–	252
–	+	–	–	162
+	+	–	–	172
–	–	+	–	187
+	–	+	–	254
–	+	+	–	174
+	+	+	–	183
–	–	–	+	138
+	–	–	+	168
–	+	–	+	127
+	+	–	+	140
–	–	+	+	172
+	–	+	+	219
–	+	+	+	153
+	+	+	+	152

Computing Lenth's *PSE* from the effects we get $PSE = 11.4375$ and $ME = 2.57 \cdot 11.4375 = 29.40$ so with a 95% confidence the active effects are $A = 30.37$, $B = -38.88$ and $D = -37.37$. Statistical software packages use always analytical methods to suggest which effects should be considered active, even when they show the results graphically in a normal or other plot. For example, Minitab (Minitab [2010]) uses Lenth's method and marks as significant *A, B* and *D,* both in the NPP chart (Figure 2.1) and also in the Pareto chart were it draws a line at 29.4 value.

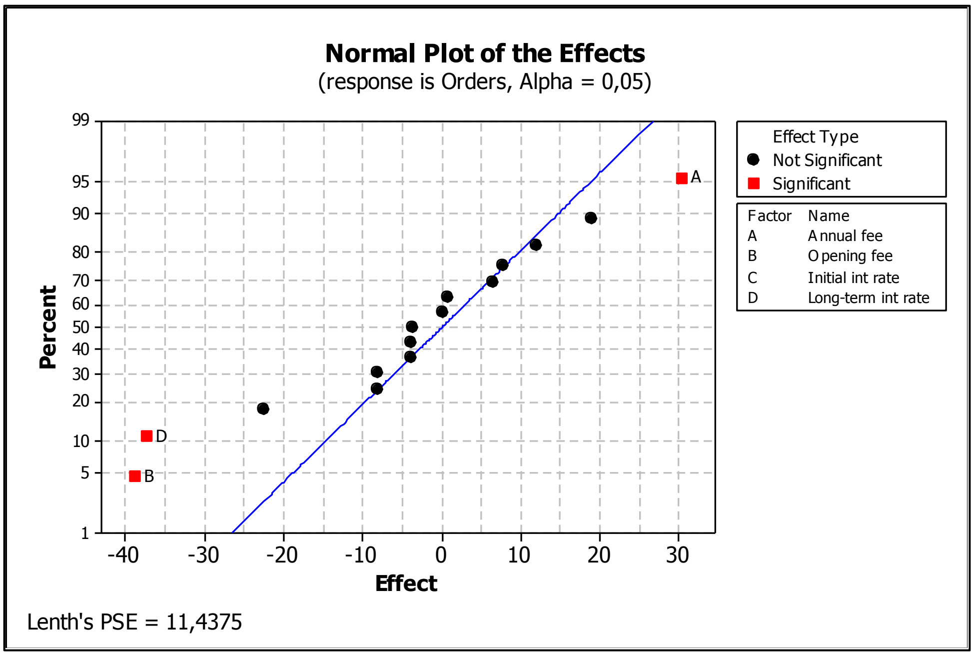

Figure 2.1: Normal Probability Plot presented by Minitab from the effects obtained in the Example of Montgomery (2013) p. 279

However, if we estimate the variance of the effects considering that the interactions of three or more factors are zero, we have $s_{ef} = 5.24$ and using also a 95% confidence level the effects that should be considered significant are the ones with a value higher than $t_{0.975,5} \cdot 5.24 = 13.46$. Therefore, $C = 18.88$ and $AB = -22.63$ will also be considered significant. This is the result given, for instance, by Statistica (Statistica [2015]) which by default analyse the significance of the effects by this method. (Figure 2.2).

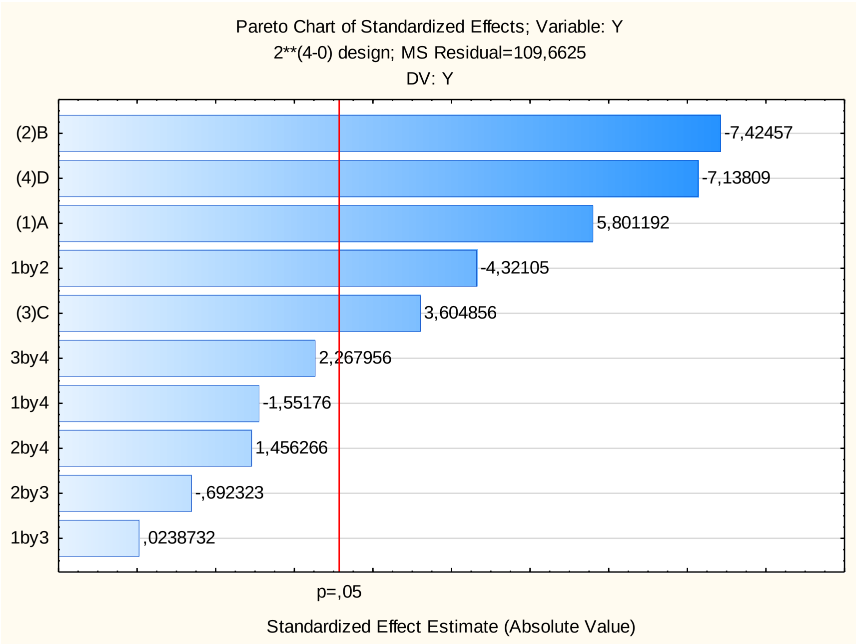

Figure 2.2: Pareto chart of effects presented by Statistica for the effects obtained in the Example of Montgomery (2013) p. 279

2.3 Tested scenarios and simulation

In order to identify the conditions under each method gives better results we have considered a set of combinations of the number and the magnitude of active effects that are a good reflection of the variety of situations that can occur in practice. In each situation we consider that k_i $(i = 1, \dots, n)$ are the effects of interest and their estimators c_i are distributed according to $N(k_i, \sigma_{ef})$. Then, without loss of generality, we can fix $\sigma_{ef} = 1$ so that the subset of j inert effects follows a $N(0, 1)$ distribution and the $(n - j)$ subset of active effects is distributed as a $N(a\Delta,\ 1)$ where a can be different for each effect and Δ are called Spacing and varies in all cases from 0.5 to 8 with increments of 0.5 as it is done in the references cited below.

For eigth-run designs we use the same testing scenarios that were used by Fontdecaba *et al.* [2015]:

S8-1: $k_1 = \cdots = k_6 = 0, k_7 = \Delta$
S8-2: $k_1 = \cdots = k_5 = 0, k_6 = k_7 = \Delta$
S8-3: $k_1 = \cdots = k_4 = 0, k_5 = k_6 = k_7 = \Delta$
S8-4: $k_1 = \cdots = k_4 = 0, k_5 = \Delta, k_6 = 2\Delta, k_7 = 3\Delta$

And for 16-run designs the same that were used by Venter and Steel [1998], and later also by Ye *et al.* [2001] and Fontdecaba *et al.* [2014]:

S16-1: $k_1 = \cdots = k_{14} = 0, k_{15} = \Delta$
S16-2: $k_1 = \cdots = k_{12} = 0, k_{13} = k_{14} = k_{15} = \Delta$
S16-3: $k_1 = \cdots = k_{10} = 0, k_{11} = \cdots = k_{15} = \Delta$
S16-4: $k_1 = \cdots = k_8 = 0, k_9 = \cdots = k_{15} = \Delta$
S16-5: $k_1 = \cdots = k_{12} = 0, k_{13} = \Delta, k_{14} = 2\Delta,\ k_{15} = 3\Delta$
S16-6: $k_1 = \cdots = k_{10} = 0, k_{11} = \Delta,\ k_{12} = 2\Delta,\ k_{13} = 3\Delta,\ k_{14} = 4\Delta,\ k_{15} = 5\Delta$

We have conducted 10,000 simulations for each scenario and each value of Spacing, using the R statistical software package [15].

These situations have been analysed by Lenth's method as well as by estimating σ_{ef}^2 from the effects considered negligible, which have been selected at random from the ones with $k = 0$. The number of negligible effects go from one to three in eight-run designs, and from one to six in 16-run ones, far beyond the normal situation of having 1 (rarely 2) in 8 runs and up to 5 in 16 runs designs.

As an example, one result from the S8-4 scenario with $\Delta = 3$, is indicated in Table 2.3. There are four inert effects: $k_{1,2,3,4} = 0$, and three active effects: $k_5 = 1\Delta = 3,\ k_6 = 2\Delta = 6$ and $k_7 = 3\Delta = 9$. With these c_i values we get a Lenth's PSE = 1.84 and using the t-value proposed by Lenth (t = 3.76) the effects that show $|c_i| > 6.92$ should be considered active. In this case only c_7. Since $k_5 = 3$ and $k_6 = 6$ are different from zero, the Lenth method has produced two type II and zero type I errors. If we use c_1 and c_2 –randomly taken from the effects with $k = 0$– to estimate the variance of the effects with two degrees of freedom, we obtain $s_{ef}^2 = 1.12$ which gives a critical value of $t_{0.975,2} \cdot s_{ef} = 4.56$. Therefore, c_6 and c_7 are considered active and c_5 not. Thus, the method has produced just one type II error.

Table 2.3: Results with the values of the effects obtained by simulation for a design with 8 runs, S8-4 scenario, Δ = 3 estimating the variance with 2 df

i	k_i	c_i	Effects significance analyzed by: Lenth's method	Variance estimation from c_1 and c_2
1	0	-1.50	Not significant	– *
2	0	-0.02	Not significant	– *
3	0	1.22	Not significant	Not significant
4	0	0.40	Not significant	Not significant
5	3	2.69	Not significant**	Not significant**
6	6	6.05	Not significant**	SIGNIFICANT
7	9	8.93	SIGNIFICANT	SIGNIFICANT

*Not analyzed.
** Type II Error

After generating 10,000 sets of effects for each scenario and Δ value and identifying active ones by the two methods, we calculated the percentage of type I and type II errors.

The total number of possible type I errors is not the same in both cases, because in the variance estimation method the effects used to estimate the variance are not analysed. For instance, with the S8-4 scenario and Δ = 3 the Lenth method can commit up to 40,000 type I errors (4 in each situation) while the variance estimation method can only commit 20,000 type I if the variance is estimated with 2 df, because those who are supposed zero and are not analysed. With both methods there are 30,000 opportunities of type II error. The results obtained in this specific case are in Table 2.4.

Table 2.4: Type I and Type II errors found in the 10,000 simulations with scenario C8-4 with Δ = 3 and estimating the variance with 2 df

	Type I error		Type II error	
	Absolute value	Percentage	Absolute value	Percentage
Lenth's method	92	$\frac{92}{40000}100 = 0.23$	14346	$\frac{14346}{30000}100 = 47.82$
Variance estimated with 2 df	966	$\frac{966}{20000}100 = 4.83$	7866	$\frac{7866}{30000}100 = 26.22$

2.4 Obtained results

In this section we present the results of all simulations. For the Lenth method the results are calculated using Lenth's original t values (still in use in several places like Minitab (Minitab, 2010) and also the t value proposed by Fontdecaba *et al.* (2015).

2.4.1 Eight-run designs

The percentage of type I errors produced by estimating the variance by the effects considered negligible is, as expected, around 5%. When the Lenth's method is applied the percentage varies depending on the scenario and spacing. Figure 2.3 shows the percentage of type I errors for each scenario and Spacing value.

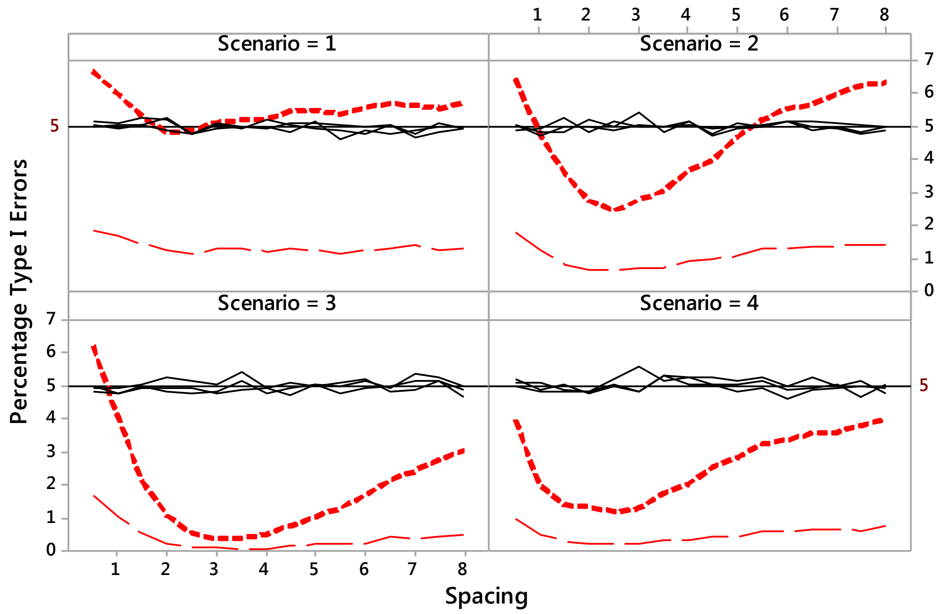

Figure 2.3: Eight-run designs. Percentage of type I errors using Lenth's method (long dashed: t=3.76; short dashed; t= 2) and estimating the effects' variance with 1, 2 and 3 df (solid lines)

The most important differences, and probably also the most relevant in this context of the design of experiments occur in type II errors (Figure 2.4). The percentage of these errors always decreases while increasing the Spacing value.

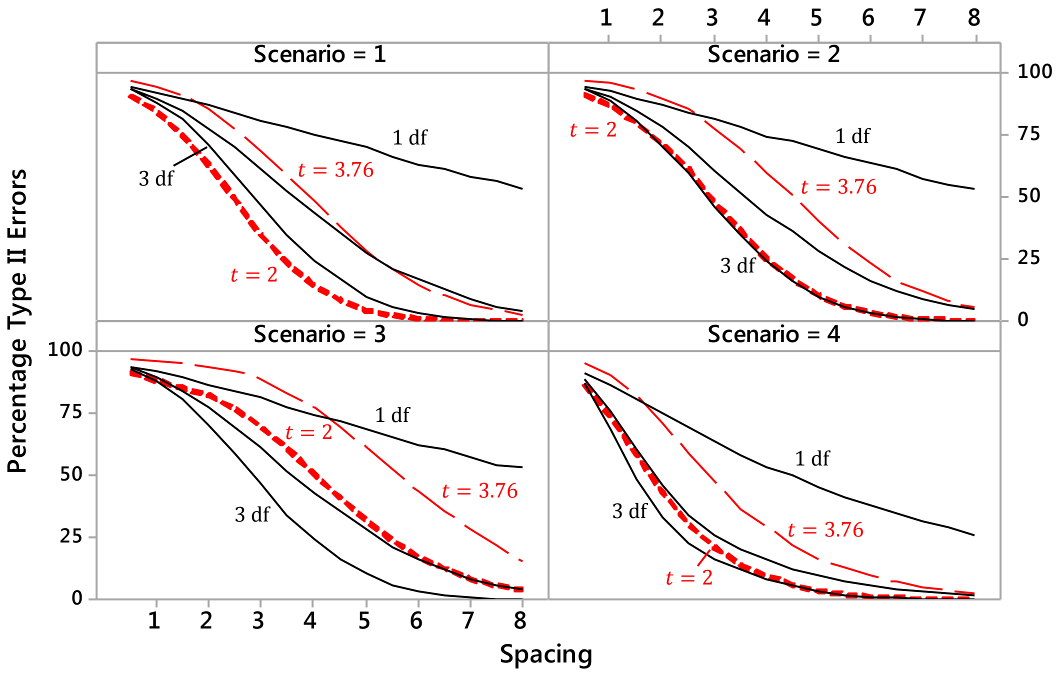

Figure 2.4: Percentage of type II errors using Lenth's method (long dashed: t=3.76; short dashed; t= 2) and estimating the effects' variance with 1, 2 and 3 df (solid lines)

Table 2.5 shows the number of degrees of freedom used to estimate the variance of the effects that are needed to get a better result than Lenth's method. Results from Lenth method are calculated with $t = 3.76$ and $t = 2$. For instance, for scenario 4, if we use two degrees of freedom or more, type II errors will be smaller than using the method of Lenth with $t = 3.76$, while with $t = 2$ three degrees of freedom or more are needed to get smaller type II errors.

Table 2.5: Eight runs. Degrees of freedom needed for the negligible interaction method to give lower probability of type II errors than the Lenth's method

		Scenario			
		S8-1	S8-2	S8-3	S8-4
Lenth's method $t =$	3.76	2	2	2	2
	2	>3	>3	2	3

Therefore, in an eight-run design, to analyze the significance of the effects by estimating their variance with only one degree of freedom has, in all scenarios considered, bigger probabilities of type II error than with the Lenth method.

2.4.2 Sixteen-run designs

Figure 2.5 shows type I errors in 16-run designs' scenarios. As in the eight run designs case, the negligible interaction method produces errors at around 5% whereas Lenth's method varies depending on the scenario and the spacing.

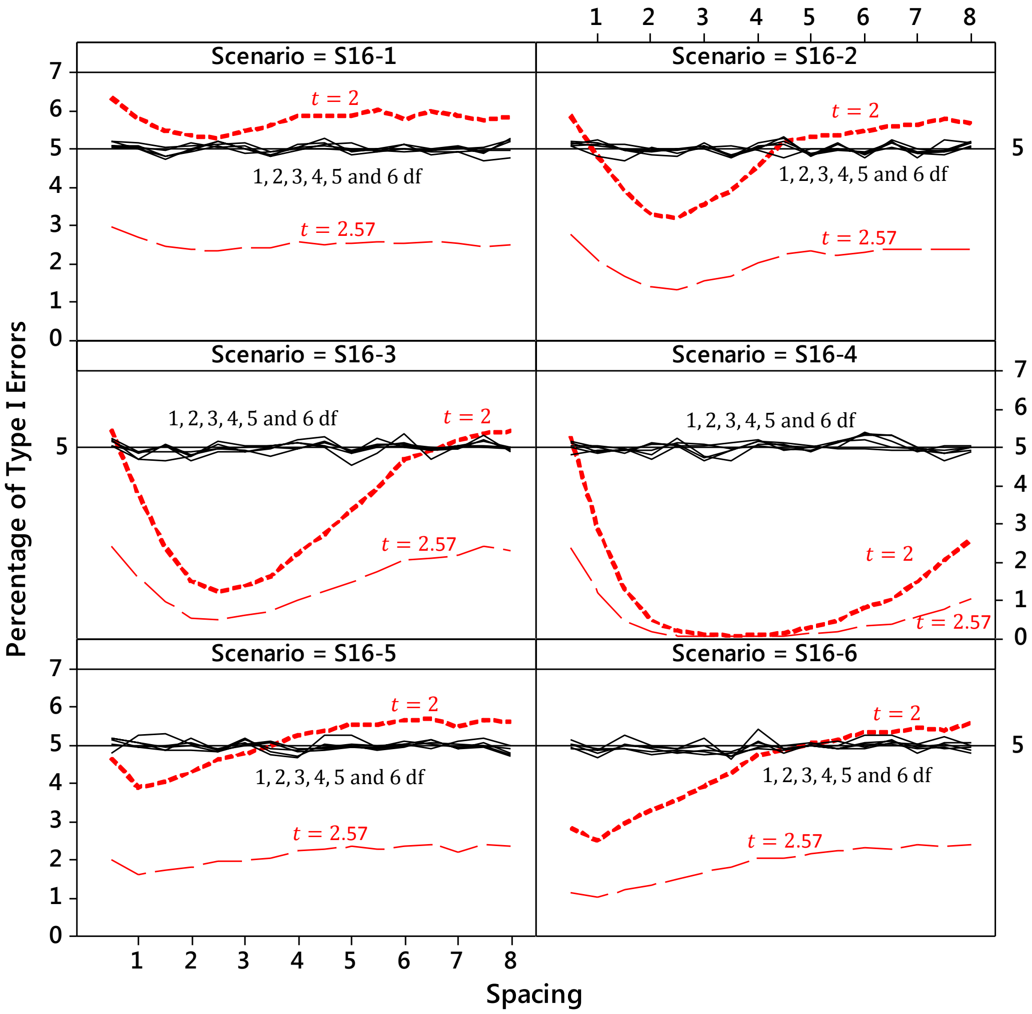

Figure 2.5: Sixteen-run design. Percentage of type I errors using Lenth's method (long dashed: t=2.57; short dashed; t= 2) and estimating the effects variance with 1 up to 6 df (solid lines)

As before, in the negligible interaction method the proportion of type II errors decreases when the number of degrees of freedom increases and in Lenth's method this proportion is always smaller with $t = 2$ (Figure 2.6). Table 2.6 indicates the number of degrees of freedom the negligible interactions method needs to outperform Lenth's method for different scenarios.

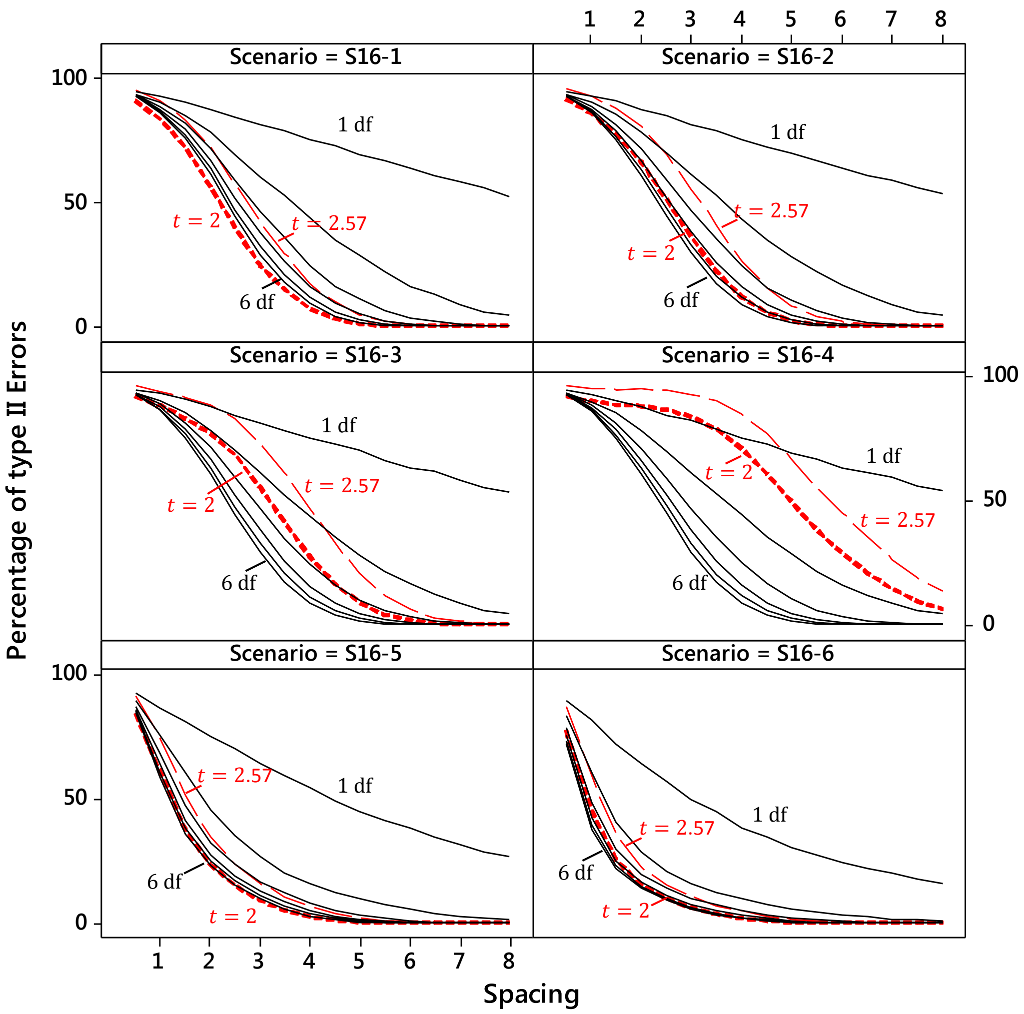

Figure 2.6: Sixteen-run design. Percentage of type II errors using Lenth's method (long dashed: t=3.76; short dashed; t= 2) and estimating the effects variance with 1 up to 6 df (solid lines, from top to bottom)

Table 2.6: Sixteen runs. Degrees of freedom needed for the negligible interaction method to give better results than the Lenth's method

		Scenario					
		S16-1	S16-2	S16-3	S16-4	S16-5	S16-6
Lenth's method $t =$	2.57	4	3	3	2	4	3
	2	>6	5	3	2	5	5

2.5 Conclusions and final remarks

We have analyzed the two most widely used analytical methods to judge the significance of effects in un-replicated factorial designs, and we have seen that they do not always produce the same result and that in such cases the best method is not always the same. However, we have seen that in some situations one is clearly better than the other. In the bullet points below we summarize the conclusions from the study and give some recommendations to practitioners, as well as to software makers:

- There are better alternatives for the t values than those proposed by Lenth. This is not new (see, for instance, Ye and Hamada [2000] or Fontdecaba *et al.* [2015]) and the study makes it evident once again. An improvement point for several widely used statistical software packages.
- To estimate the variance of the effects with a single degree of freedom is a bad practice and nearly always worse than to apply the method of Lenth. Some software packages analyse by default the significance of the effects considering negligible the interactions of three or more factors, and they do this also for 2^3 designs in which, obviously, there is only one three factor interaction.
- In eight-run designs, except when we face scenario 3, at least three degrees of freedom are needed for the negligible interaction method to be better than Lenth's method. Practically we will never know a priori that three of the seven contrasts are negligible, and therefore, as a general rule, in eight runs designs it is better to apply the method of Lenth.
- In 16-run designs, the negligible interactions method provides better results (or almost the same in scenario 1) when 5 or more degrees of freedom can be used for variance estimation. Naturally, this happens in complete 2^4 when, interactions of three or more factors are considered negligible.
- A final recommendation to practitioners is to not forget that in many cases normal probability plots, technical knowledge and common sense will solve the discrepancy.

Applying this recommendations to Montgomery's [2013] p. 279 example, the advice is to use the negligible interactions method, and using it the effects of C and AB would be considered active. A very reasonable assessment by looking at the normal probability plot.

2.6 References

Bell, G.H., Ledolter, J. & Swersey, A.J. (2006). Experimental design on the front lines of marketing: Testing new ideas to increase direct mail sales, *International Journal of Research in Marketing,* 23:309-319.

Box, G.E.P., Hunter, J.S. & Hunter, W.G. (2005). *Statistics for experimenters: design, innovation, and discovery*. New Jersey: Wiley.

Box, G.E.P. & Meyer, R.D. (1986). An Analysis for Unreplicated Fractional Factorials. *Technometrics,* 28:11-18.

Box, G.E.P. and Meyer, R.D. (1993). Finding the Active Factor in Fractionated Screening Experiments. *Journal of Quality Technology,* 25:94-105.

Dong, F. (1993). On the identification of active contrasts in unreplicated fractional factorials. *Statistica Sinica,* 3:209-217.

Espinosa, V., Dasgupta, T. & Rubin, D.B. (2016). A Bayesian Perspective on the Analysis of Unreplicated Factorial Experiments Using Potential Outcomes. *Technometrics,* 58: 62-73.

Fontdecaba, S., Grima, P. & Tort-Martorell, X. (2014). Analyzing DOE with satistical software sackages: controversies and proposals. *The American Statistician,* 68:205-211.

Fontdecaba, S., Grima, P. & Tort-Martorell, X. (2015). Proposal of a single critical value for the lenth method. *Quality Technology and Quantitative Management,* 12:41-51.

Hamada, M. & Balakrishnan, N. (1998). Analyzing unreplicated factorial experiments: a review with some new proposal. *Statistica Sinica,* 8:1-41.

Juan, J. & Peña, D. (1992). A simple method to identify significant effects in unreplicated two-level factorial designs. *Communications in Statistics. Theory and Methods,* 21: 1383-1403.

Lenth, R.V. (1989). Quick and easy analysis of unreplicated factorials. *Technometrics,* 31: 469-473.

Minitab 17 Statistical Software (2010). [Computer software]. State College, PA: Minitab, Inc.

Montgomery, D.C. (2013). *Design and analysis of experiments*. Singapore: Willey.

R Development Core Team. (2011). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing.

Statistica 13.0 (2013). [Computer software]. Dell Software.

Venter, J.H. & Steel, S.J. (1998). Identifying active contrasts by stepwise testing. *Technometrics,* 40:304-313.

Ye, K.Q. & Hamada, M. (2000). Critical values of the Lenth method for unreplicated factorial designs. *Journal of Quality Technology,* 32:57-66.

Ye, K.Q., Hamada, M. & Wu, C.F.J. (2001). A step-down length method for analyzing unreplicated factorial designs. *Journal of Quality Technology,* 33:140-152.

Chapter 3

ARTICLE 2

Estimating missing values from negligible interactions in factorial designs

3

ARTICLE 2

Estimating missing values from negligible interactions in factorial designs

Rafel Xampeny, Pere Grima, Xavier Tort-Martorell
Department of Statistics and Operations Research
Universitat Politècnica de Catalunya – BarcelonaTech, Spain

Quality and Reliability Engineering International,
2017; 0: 1–13
Received: 21 December 2016, Revised: 18 April 2017, Accepted: 21 April 2017
DOI: 10.1002/qre.2172
https://doi.org/10.1002/qre.2172

ABSTRACT

When a factorial design has some missing value but there are contrasts (at least as many as there are missing values) corresponding to interactions that can be considered negligible, equating the expressions of those contrasts to zero allows the missing values to be deduced. This procedure allows estimating values that correspond to runs that have gone wrong or could not be performed as planned, as well as to reduce the experimental plan while saving some runs whose results can be estimated a posteriori. The problem is that the estimated values have a variance that is higher than that of the experimental runs, and the variance depends on which are the missing values and the contrasts used to make the estimate. This article performs a thorough analysis of the variances obtained by estimating up to two missing values in 8-run designs and up to five in 16-run designs, with tables included which provide the missing values and contrasts that should be used to obtain estimates that have minimal variability.

KEYWORDS: Factorial design, missing values, negligible interactions, estimating missing values, saving runs.

3.1 Introduction

Although experimental plans are usually designed assuming that, at the end, we will have the response values in all experimental conditions, in practice it is not strange to have missing responses. This may occur because an unforeseen problem has hampered performing some runs, or because we doubt the validity of some results and prefer to ignore them. If it is possible to carry out these missing runs once again, this is undoubtedly the best choice.

However, if this is not possible, the experimenter has no other choice but to try and get the most out of the available results.

If among the effects to estimate there are some that can be considered negligible, as is generally the case with three or more factor interactions, one way to tackle the analysis is to estimate the values of the missing responses from the expressions of the negligible interactions equated to zero. Naturally, this procedure can also be applied to estimate the response corresponding to runs not conducted in order to make the experimental plan more economical.

The idea is not new. Draper and Stoneman [1964] already proposed estimating missing responses through expressions of interactions that are considered negligible. It is especially interesting their proposal to select the sets of interactions that should be considered negligible. They do it by estimating the missing values from different sets of interactions that can reasonably be considered negligible, then they calculate the effects and analyze how reasonable these effects are, based on their representation in a Normal Probability Plot (NPP). Many years later, Box [1990] provides a very didactic explanation of Draper and Stoneman's proposal for estimating missing responses, and he stresses the importance of being attentive to all the details that would allow learning from the experiment. Goh [1996] uses the same strategy but with the aim of reducing the number of runs by as many as the number interactions considered negligible before conducting the experiment. Almimi *et al.* [2008] extend the Draper and Stoneman method to Split-plot designs.

More recently, Zhou and Goh [2016] insist on the idea of reducing the number of runs and propose a sequential strategy. The method consists of taking the mean value of the responses already obtained and assigning it to the response of runs that are still not performed. Each time a new run is performed, the effects are recalculated by replacing the value that was assigned to its response with the value that was actually obtained. When the values of the effects are stabilized, the experiment finishes and it is generally expected that runs will be saved.

In this article we return to the Draper and Stoneman proposal for using interactions that can be considered negligible in order to estimate the missing values. The problem that arises when applying this procedure is that the variance of the estimated values is always greater than the values obtained directly from the experiment. This greater variability depends on what the estimated values are and also on which sets of interactions are used. There are also situations with 4 or 5 missing values in which it is impossible to estimate their values even if there are 5 negligible interactions. Knowing which situations are the most advantageous for estimating the missing values and which should be avoided is undoubtedly useful when the experimental plans face a shortage of resources and we opt to save some experiment and estimate its value a posteriori.

The paper is organized as follows: Sections two and three provide the variances obtained when we have up to two missing values in eight-run designs and up to five in 16-run designs, depending on the missing value combinations and the interactions used. Section four presents designs in which these recommendations may be more useful, as they present contrasts that correspond only to interactions of 3 or more factors and, finally, section five provides some conclusions. The appendix details the combinations of missing values and interactions that should be used in each case for estimating the missing values with minimal variance.

3.2 Eight-run experiments

3.2.1 One missing response

Let us consider the case of a 2^3 design. Table 3.1 shows its table of Contrast Coefficients.

Table 3.1: Table of Contrast Coefficients for a 2^3 design.

	Contrast	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
	A	-1	1	-1	1	-1	1	-1	1
	B	-1	-1	1	1	-1	-1	1	1
	C	-1	-1	-1	-1	1	1	1	1
1	AB	1	-1	-1	1	1	-1	-1	1
2	AC	1	-1	1	-1	-1	1	-1	1
3	BC	1	1	-1	-1	-1	-1	1	1
4	ABC	-1	1	1	-1	1	-1	-1	1

If the ABC interaction is negligible, we have:

$$-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8 = 0 \tag{3.1}$$

And from this expression we can calculate any response value depending on the rest. If σ_y^2 is the variance of the response, it is obvious that the variance of an estimated value will be $7\sigma_y^2$, regardless of that value.

If a two-factor interaction is also considered null, it is possible to estimate the missing value with a smaller variance by averaging the estimates obtained with each of the interactions. To determine the variance of the estimate in this case, we need to keep in mind that any pair of contrasts shares the same sign in half of the positions (whose product gives the + sign) and has a different sign in the other half (whose product gives the - sign). Thus, if the missing value has the same sign in two interactions being considered, the two expressions that give their estimate (one with each interaction) will have 4 terms with a different sign and 3 with the same.

For example, let y_2 be the missing value and BC and ABC be the interactions that can be considered null. We will use the expression $\hat{y}_{i,p}$ to refer to the estimate of y_i based on the interaction whose order number in Table 3.1 is p. The expressions will be:

$$\hat{y}_{2,3} = -y_1 + y_3 + y_4 + y_5 + y_6 - y_7 - y_8$$
$$\hat{y}_{2,4} = +y_1 - y_3 + y_4 - y_5 + y_6 + y_7 - y_8$$

In a similar way, $\hat{y}_{i,pq}$ will denote that the estimation was made using the average of the values obtained from interactions numbers p and q. Thus, in our example:

$$\hat{y}_{2,34} = \frac{\hat{y}_{2,3} + \hat{y}_{2,4}}{2} = \frac{2y_4 + 2y_6 - 2y_8}{2}$$

A similar situation occurs when the missing value has a different sign in the two interactions considered. The sign is changed for the terms of one expression while not for the other, and therefore there also remain 4 signs that are the same and 3 that are different. For example, in the case of estimating y_2 with AC and ABC, we have:

$$\hat{y}_{2,2} = +y_1 + y_3 - y_4 - y_5 + y_6 - y_7 + y_8$$
$$\hat{y}_{2,4} = +y_1 - y_3 + y_4 - y_5 + y_6 + y_7 - y_8$$

And in this case,

$$\hat{y}_{2,24} = \frac{2y_1 - 2y_5 + 2y_6}{2}$$

Therefore, we always have $V(\hat{y}_{i,pq}) = 3\sigma_y^2$, no matter the missing value or the interactions considered negligible.

3.2.2 Two missing values

Given a pair of null interactions, we can estimate the values for a pair of missing values, but not for just any pair. If, in the system of equations, the coefficients of the missing values are linearly dependent, the system will be inconsistent. For each possible pair of null interactions, Table 3.2 uses the same background color (white or gray) to show the response values that are linearly dependent. Therefore, given a pair of null interactions, only two values can be estimated if they are located in zones of different color.

Table 3.2: Pairs of negligible interactions to be used for estimating two missing responses in an eight-run design

		y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
1	AB	1	-1	-1	1	1	-1	-1	1
	ABC	-1	1	1	-1	1	-1	-1	1
2	AC	1	-1	1	-1	-1	1	-1	1
	ABC	-1	1	1	-1	1	-1	-1	1
3	BC	1	1	-1	-1	-1	-1	1	1
	ABC	-1	1	1	-1	1	-1	-1	1

For example, if the interactions that can be considered null are AB and ABC, and the missing values are y_1 and y_5, we will have:

$$\begin{aligned} y_1 + y_5 &= y_2 + y_3 - y_4 + y_6 + y_7 - y_8 \\ -y_1 + y_5 &= -y_2 - y_3 + y_4 + y_6 + y_7 - y_8 \end{aligned} \tag{3.2}$$

Therefore,

$$\begin{aligned} 2y_1 &= 2y_2 + 2y_3 + 2y_4 \\ 2y_5 &= 2y_6 + 2y_7 + 2y_8 \end{aligned} \tag{3.3}$$

And $V(y_1) = V(y_5) = 3\sigma_y^2$.

This result is independent of the missing values and the interactions used. As we have already mentioned, any pair of interactions shares the same sign in half of the positions and has a different sign in the other half. Of the two missing values, one will always have the same sign in the two interactions and another will have a different sign; then, on the right-hand side of the equations (2), we will always have half the responses with the same sign and the other half with a different sign, which leads to always having expressions similar to the one we have in (3). Therefore, when it is possible to estimate a pair of missing values, their variances will always be equal to $3\sigma_y^2$.

3.3 Sixteen-run experiments

3.3.1 One missing response

Let us now consider a 2^4 design with factors $A, B, C,$ and D. Table 3.3 shows the contrasts corresponding to three and four factor interactions.

Table 3.3: Contrasts –in standard order– corresponding to the 3 and 4 factor interactions in a 2^4design.

Interaction		y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}
1	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
2	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
3	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
4	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
5	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1

By equating the expression of any of the interactions to zero, we can find out the value of the missing response and its variance will be equal to $15\sigma_y^2$; but this value can also be reduced by using the average of the 5 values obtained from each of the 5 interactions as an estimator. We will continue to use the notation $\hat{y}_{i,p}$ to refer to the estimate of $\hat{y}_i$ from the interaction p, where p is now its order number in Table 3.3. Also, similarly to the previous case, we write $\hat{y}_{i,pq\ldots}$ to refer to the estimate of $\hat{y}_i$calculated with the average of the estimates obtained from the interactions $p, q, \ldots$

Therefore:

$$\hat{y}_{i,12345} = \frac{\hat{y}_{i,1} + \hat{y}_{i,2} + \hat{y}_{i,3} + \hat{y}_{i,4} + \hat{y}_{i,5}}{5}$$

Since the different estimates of y_i are not independent, we have:

$$V(\hat{y}_{i,12345}) = \frac{1}{25}\left(\sum_{p=1}^{5} V(\hat{y}_{i,p}) + 2\sum_{p\neq q} \mathrm{Cov}(\hat{y}_{i,p};\ \hat{y}_{i,q})\right) \qquad (3.4)$$

Expressions $\hat{y}_{i,p}$ and $\hat{y}_{i,q}$ can be written as follows:

$$\hat{y}_{i,p} = x_{i,pq} + z_{i,pq}$$
$$\hat{y}_{i,q} = x_{i,pq} - z_{i,pq}$$

where $x_{i,pq}$ is the sum of the responses that have the same sign in $\hat{y}_{i,p}$ and in $\hat{y}_{i,q}$, while $z_{i,pq}$ is the sum of those with a different sign. For example, for $i = 5$, $p = 1$and $q = 2$, we have:

$$y_{5,1} = y_6 + y_7 - y_8 + y_9 - y_{10} - y_{11} + y_{12} + (y_1 - y_2 - y_3 + y_4 - y_{13} + y_{14} + y_{15} - y_{16})$$
$$y_{5,2} = y_6 + y_7 - y_8 + y_9 - y_{10} - y_{11} + y_{12} - (y_1 - y_2 - y_3 + y_4 - y_{13} + y_{14} + y_{15} - y_{16})$$

Hence,

$$x_{5,12} = y_6 + y_7 - y_8 + y_9 - y_{10} - y_{11} + y_{12}$$
$$z_{5,12} = y_1 - y_2 - y_3 + y_4 - y_{13} + y_{14} + y_{15} - y_{16}$$

Eliminating the x and z subscripts to lighten the notation (they are always $x_{i,pq}$ and $z_{i,pq}$), and given that the product of two contrasts leads to a new contrast, we have:

$$\begin{aligned} Cov(\hat{y}_{i,p}, \hat{y}_{i,q}) &= Cov(x + z, x - z) = \\ &= E[(x + z)(x - z)] - E(x + z)E(x - z) = \\ &= E(x^2) - E(z^2) - [E(x)]^2 + [E(z)]^2 = \\ &= V(x) - V(z) = \\ &= n_i\sigma_y^2 - (15 - n_i)\sigma_y^2 \end{aligned}$$

where n_i is the number of terms that have the same sign in $\hat{y}_{i,p}$ and $\hat{y}_{i,q}$. Using a rationale that is similar to that for estimating a missing value from two null interactions in eight-run designs, we can say always that $n_i = 7$.

Substituting the obtained values in expression (4):

$$V(\hat{y}_{i,12345}) = \frac{1}{25}\left(5 \cdot 15\sigma_y^2 + 2\binom{5}{2}\ (7-8)\sigma_y^2\right) = \frac{11}{5}\sigma_y^2$$

In general, if some of the considered interactions cannot be taken as negligible, we have:

$$V(\hat{y}_{i,1\ldots k}) = \frac{1}{k^2}\left(k \cdot 15\sigma_y^2 + 2\ \binom{k}{2}\left(7\sigma_y^2 - 8\sigma_y^2\right)\right) \quad (3.5)$$

where $k \geq 2$. No matter what the missing value is, it will always be possible to estimate it from any interaction considered negligible. Table 3.4 summarizes the values of the variance of the estimate based on the number of interactions considered negligible.

Table 3.4: Variance of the missing response estimated depending on the number of negligible interactions used.

Missing value	Number of null interactions (no matter which ones)	Variance of the estimate
Any	1	$15\sigma_y^2$
	2	$7\sigma_y^2$
	3	$4.33\sigma_y^2$
	4	$3\sigma_y^2$
	5	$2.2\sigma_y^2$

3.3.2 Two missing responses

In equating two negligible interactions to zero, we get a system of two equations where the unknowns are the missing responses. As in the case of estimating two missing values in an eight-run design, it is not possible to use just any pair of interactions because –depending on which ones are the missing responses– some pairs will lead to linearly dependent systems. Table 3.5 shows the 10 pairs that can be formed with the 5 given interactions. The pairs are divided into two groups (white background and grey background) in such a way that, given two missing responses, only the pairs of interactions which have one response in each group are useful.

Table 3.5: Pairs of negligible interactions to be used for estimating two missing responses.

		y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}
1	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
2	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
3	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
4	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1
5	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
6	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
7	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1
8	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
9	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1
10	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1

Furthermore, the variance of the estimate in this case can be reduced by averaging the values estimated from several pairs of interactions. The analytical study of each case is long and tedious. However, an exhaustive analysis can be performed by programming and analysing all possible cases, which we did by using the R statistical software package [2011]. From this analysis, we can conclude that the pairs of missing responses can be divided into two groups: those that allow the use of up to 4 interaction pairs (subset A) and those that allow the use of up to 6 (subset B). When the missing responses are in subset A, the minimum variance of the estimated value that can be attained is $4\sigma_y^2$; and when they are in subset B, it is e $2.33\sigma_y^2$. Therefore, if it is possible to choose the missing responses, it is better to pick them from subset B. Table 3.6 contains the obtained variances according to the subset to which the pair of missing values belong and the number of pairs of interactions used. The set of pairs of missing values that can be estimated with minimal variance can be found in Table 3.12 of the Appendix, along with the interactions that need to be used in each case.

Table 3.6: Obtained variances when we have two missing values.

Pairs of interactions considered	Pairs of missing values that are possible to use(1)	Null interactions used. Number of times that each of them appears	Variance for both missing values	Interest
1	Subset A or B	2: 1,1	$7\sigma_y^2$	When only 2 interactions can be considered null
2	Subset A or B	3: 1,2,1	$5\sigma_y^2$	When only 3 interactions can be considered null
	Subset B	4: 1, 1, 1, 1	$3\sigma_y^2$	When only 4 interactions can be considered null and the missing values are in subset B
3	Subset A or B	4: 1, 3, 1, 1	$4.33\sigma_y^2$	When only 4 interactions can be considered null and the missing values are in subset A
	Subset B	4: 1, 2, 2, 1 5: 1, 2, 1, 1, 1	$3.44\sigma_y^2$ $2.56\sigma_y^2$	
4	Subset A	5: 1, 4, 1, 1, 1	$4\sigma_y^2$	When the missing values are from subset A.
	Subset B	4: 2, 2, 2, 2 5: 1, 2, 2, 2, 1	$3\sigma_y^2$ $2.5\sigma_y^2$	Same result as using two pairs of interactions
5	Subset B	5: 1, 2, 2, 2, 3	$2.52\sigma_y^2$	
6	Subset B	5: 2, 2, 2, 3, 3	$2.33\sigma_y^2$	When the missing values are from subset B. Best combinations of missing values (minimum variance) considering the five negligible interactions. See details in Table 3.12.

(*) Subset A: 40 Pairs of missing values that can be estimated with up to 4 pairs of interactions
Subset B: 80 Pairs of missing values that can be estimated with up to 6 pairs of interactions

3.3.3 Three missing responses

A trio of null interactions is needed for estimating three missing responses. With five negligible interactions, we can form 10 trios. However, as in the previous cases, the trios of interactions that can be used depend on the missing values. Table 3.7 shows the responses divided into 4 groups for each interaction trio. For the equation system to have a solution, the 3 missing responses should belong to different groups.

Table 3.7: Trios of negligible interactions to be used for estimating three missing responses.

		y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}
	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
1	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1
	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
2	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1
	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
3	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1
	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
4	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
5	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1
	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
6	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1
	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
7	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
8	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1
	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
9	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
10	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1

The 560 possible trio combinations of missing responses can be divided into 3 subsets:

- Subset A: Formed by the 160 trios of missing responses that can be estimated with up to 3 interaction trios; for instance, y_1, y_2 and y_3, which can only be estimated with interaction trios 1, 4 and 7 (Table 3.7). The missing responses are estimated with variances: $7\sigma_y^2$ one of them, and $4.33\sigma_y^2$ the other two.

- Subset B: Formed by 240 trios of missing responses that can be estimated with up to 4 interaction trios. For instance, y_1, y_2 and y_7, can be estimated with interaction trios 1, 2, 7 and 9. The missing responses are estimated with variances: $3\sigma_y^2$ one of them, and $5\sigma_y^2$ the other two.

- Subset C: Formed by the remaining 160 trios of missing responses. They can be estimated with up to 7 interaction trios, such as y_1, y_4 and y_6, which can be estimated by the interaction trios 1, 2, 3, 4, 7, 9 y 10. In this case, the three responses have the same variance: $2.43\sigma_y^2$.

Table 3.8: Obtained variances when we have three missing values.

Trios of interactions considered	Trios of missing values that are possible to use(*)	Null interactions used. Number of times that each of them appears	Variance ($\times\ \sigma_y^2$) for missing values	Interest
1	Subset A, B or C	3: 1, 1, 1	7, 7, 7	When only 3 interactions can be considered null
2	Subset A, B or C	4: 1, 2, 2, 1	7, 5, 5	Missing values are in subset A or B and only four interactions can be considered null
	Subset B	5: 1, 1, 2, 1, 1	5, 5, 3	
	Subset C	4: 1, 2, 2, 1	7, 3, 3	
		5: 1, 1, 2, 1, 1,	5, 3, 3	
3	Subset A, B or C	5: 1, 2, 3, 2, 1	7, 4.33, 4.33	Missing values are in subset A. Worst combinations of missing values (maximum variance) considering the five negligible interactions.
	Subset B	5: 2, 3, 2, 1, 1	5.22, 5.22, 3.44	
	Subset C	5: 2, 3, 2, 1, 1	5.22, 3.44, 2.56	
		5: 3, 3, 1, 1, 1	7, 2.56, 2.56	
		4: 2, 2, 2, 2	3.44, 3.44, 3.44	
		5: 2, 2, 3, 1, 1	2.56, 3.44, 3.44	
		5: 2, 2, 2, 2, 1	3.44, 2.56, 2.56	
4	Subset B	5: 2, 4, 2, 2, 2	5, 5, 3	Missing values are in subset B
	Subset C	5: 2, 4, 2, 2, 2	5, 2.5, 2.5	
		5: 3, 4, 2, 2, 1	4, 3, 2.5	
		4: 3, 3, 3, 3	3, 3, 3	Missing values are in subset C and only four interactions can be considered null
		5: 2, 2, 2, 3, 3	3, 3, 2.5	
		5: 1, 2, 3, 3, 3	3, 2.5, 2.5	
5	Subset C	5: 2, 2, 3, 3, 5	3.16, 2.84, 2.84	
		5: 1, 3, 3, 4, 4	3.16, 2.52, 2.52	
		5: 2, 2, 3, 4, 4	2.84, 2.52, 2.52	
6	Subset C	5: 2, 3, 5, 4, 4	2.56, 2.56, 2.56	
7	Subset C	5: 3, 3, 5, 5, 5	2.43, 2.43, 2.43	Missing values are in subset C. Best combinations of missing values (minimum variance) considering the five negligible interactions. See details in Table 3.13.

(*) Subset A: 160 Trios of missing values that can be estimated with up to 3 trios of interactions
Subset B: 240 Trios of missing values that can be estimated with up to 4 trios of interactions
Subset C: 160 Trios of missing values that can be estimated with up to 7 trios of interactions

In the Appendix (Table 3.13) can be found the list of the 160 trios of missing responses that can be estimated with minimum variance using 7 interaction trios.

3.3.4 Four missing responses

When there are four missing responses, we need groups of four interactions to estimate their values; and we can create 5 such groups (quartets). Each of those quartets in Table 3.9 has been divided into 8 zones in such a way that, for the system of equations to have a solution, it is a necessary condition –but not sufficient– that there not be two missing values in the same zone. It is not sufficient because, inside the quartet, the responses may be linearly dependent. This happens in quartet 1, for example, where the responses to be estimated are y_1, y_2, y_3, and y_4, they are located in different zones. But $y_4 = y_1 \cdot y_2 \cdot y_3$ and, therefore, it cannot be used.

Table 3.9: Quartets of negligible interactions to be used for estimating four missing responses.

		y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}
1	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1
2	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1
3	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1
4	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1
5	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1

Among the 1820 possible sets of four missing responses, there are 100 of them that are impossible to estimate because any interaction quartet gives an incompatible system of equations (Table 3.14 in the Appendix presents the list of these 100 combinations). The responses of the remaining 1720 sets can be estimated by using from 1 up to 5 systems of equations. For each subset, the best estimates are obtained using the maximum number of systems of equations (Table 3.10).

When we can choose the combination of missing responses, the best option is among the 40 combinations that can be estimated with 2 systems of equations. They lead to an estimate of them with variance equal to $2.5\sigma_y^2$, which is the minimum that can be attained. This counterintuitive result is due to the fact that, when using more than two systems of equations, they are not independent among themselves. Table 3.15 in the Appendix contains the list of those 40 sets of missing responses, along with the interactions that should be used.

Table 3.10: Obtained variances when we have four missing values.

Quartets of interactions considered	Quartets of missing values that are possible to use (*)	Null interactions used. Number of times that each of them appears	Variance ($\times \sigma_y^2$) for missing values	Interest
-	-	-	-	There are 100 quartets of missing values impossible to be estimated. See details in Appendix Table 3.14.
1	Subset A	4: 1, 1, 1, 1	3, 3, 3, 3	Best option if only 4 interactions can be considered null
			7, 7, 7, 7	
			15, 7, 7, 7	
2	Subset A	5: 1, 1, 2, 2, 2	2.5, 2.5, 2.5, 2.5	The 40 combinations of missing values that can be estimated with minimum variance. See details in Appendix Table 3.15.
			7, 3, 3, 3	
			7, 7, 3, 3	
			7, 7, 7, 3	
			7, 7, 5, 5	
			13, 5, 5, 5	
			13, 7, 7, 5	Worst combinations of missing values after considering the five negligible interactions.
3	Subset B	5: 3, 3, 2, 2, 2	7, 2.56, 2.56, 2.56	
			3.89, 3.89, 3, 3	
			6.11, 6.11, 3.44, 3.44	
			7, 6.11, 3.44, 3.44	
4	Subset B	5: 3, 3, 3, 3, 4	4.5; 2.5; 2.5; 2.5	
			3, 3, 3, 3	
			6, 6, 3, 3	
5	Subset C	5: 4, 4, 4, 4, 4	2.68, 2.68, 2.68, 2.68	

(*) Subset A: 1720 quartets of missing values that can be estimated with up to 2 quartets of interactions
Subset B: 560 quartets of missing values that can be estimated with up to 4 quartets of interactions
Subset C: 80 quartets of missing values that can be estimated with up to 5 quartets of interactions

3.3.5 Five missing responses

Among the 4368 possible quintets of missing responses, it is not possible to estimate the values in 1360 cases by using a five-equation system obtained from five negligible interactions. For instance, if the missing responses are y_1, y_2, y_3, y_4 and y_5, the equations will come from columns y_1 to y_5 of Table 3.3 and, since $y_4 = y_1 \cdot y_2 \cdot y_3$, the system of equations is inconsistent.

When it is possible to solve the system of five equations, the variances of the estimated values depend on which ones are the missing responses, and they might not be the same for all of them. Table 3.11 provides the variances of the estimated values for all the possibilities that can occur in this case. In cases where the experimenter can choose the responses to be estimated, Table 3.16 in the Appendix gives the 16 combinations of missing responses that can be estimated with minimum variance ($2.56\sigma_y^2$ for all of them).

Table 3.11: Obtained variances when we have five missing values.

Quintets of interactions considered	Quintets of missing values that are possible to use	Null interactions used. Number of times that each of them appears	Variance ($\times\, \sigma_y^2$) for missing values	Interest
-	-	-	-	There are 1360 quintets of missing values that are impossible to be estimated
1	All quintets that can be estimated (3008 quintets)	5: 1, 1, 1, 1, 1	31, 15, 15, 7, 7	Worst combinations of missing values
			15, 15, 7, 7, 7	
			31, 7, 7, 7, 7	
			15, 7, 7, 7, 7	
			7, 7, 7, 7, 7	
			7, 7, 3, 3, 3	
			2.56, 2.56, 2.56, 2.56, 2.56	Best combinations of missing values. See details Table 3.16.

3.4 Cases of greater practical interest. Recommendations

In this section we discuss the situation on which the experimenter wants to save costs or time by skyping some runs. We do so for 8 and 16 run designs that have contrasts to estimate three or more factor interactions, the ones usually considered negligible a priori. For this designs, we provide the experimental conditions whose results can be estimated with minimum variance and which interactions to use among the ones that are available.

3.4.1 2^3 Design

Only the interaction of three factors can be considered negligible a priori; therefore, only one missing value can be estimated. No matter what this value is, it is always estimated with the same variance.

3.4.2 2^4 Design

There are five interactions of three or more factors that can be considered negligible a priori; therefore, up to five missing values can be estimated. If the experimental plan is designed to stop performing 1 to 5 runs, the missing values will be estimated with the minimum variance by choosing them as follows:

- One missing value: No matter what the missing value, it can be estimated with minimum variance by making the estimate using each of the five negligible interactions and averaging the estimates obtained.

- Two missing values: They must be in what we have called subset B, which consists of the 80 pairs of missing values that can be estimated with up to six pairs of null interactions. These 80 pairs and the interactions that should be chosen for each of them are detailed in Table 3.12.

- Three missing values: The missing values must be in subset C, which consists of 160 trios of missing values that can be estimated with up to 7 triplets of interactions. These 160 trios and the 7 trios of interactions that should be used with each of them can be found in Table 3.13.

- Four missing values: Of the 1720 missing quartets that can be estimated with two quartets of interactions, there are 40 that can be estimated with minimal variance. The relationships between these quartets of missing values and the interactions that should be used for each of them are found in Table 3.15. It should be noted that there are also 100 combinations of 4 missing values that cannot be estimated with only 5 null interactions (see their relationships in Table 3.14).

- Five missing values: Of the 4368 possible quintets of missing values, it is not possible in 1360 cases to estimate their values from the 5 interactions considered null. There are 16 combinations of missing values that can be estimated with minimum variance using a five equation system with five unknowns. These are found in Table 3.16.

3.4.3 2^{6-2} Design

Using any of the sets of generators that provide maximum resolution and minimum aberration (e.g., E=ABC and F=BCD) there are, in their alias structure, two contrasts of aliased effects where only three-factor interactions occur. We can therefore estimate one or two missing values by equating the value of these contrasts to zero. If it is desired to estimate only one missing value no matter what it is, the variance of the estimate is always the same. In the case that it is desired to estimate two values when the above-mentioned design generators are used, the contrast containing the ABD interaction and the one containing ACD –which correspond to pair 5 in Table 3.5– can be considered null. Therefore, the missing values must be one from each of the following groups: Group 1: $y_1, y_2, y_7, y_8, y_9, y_{10}, y_{15}, y_{16}$; Group 2: $y_3, y_4, y_5, y_6, y_{11}, y_{12}, y_{13}, y_{14}$.

3.4.4 2^{7-3} Design

Using any of the sets of generators that provide maximum resolution and minimum aberration (in this case, for example, E=ABC, F=BCD and G=ACD), there is a contrast of aliased effects where only three or more interactions occur. In this case we can estimate only one missing value. No matter what that value is, the variance of the estimate will always be the same.

3.5 Conclusions

Missing values in two level factorial designs can be estimated via contrasts, in general corresponding to high order interactions, that are considered negligible from scratch. The variance of this estimates is different depending on the number of runs to be estimated, the number of contrasts available and the relation between them.
The paper provides the variances of the estimates for one and two missing values in 8 run designs and for up to five missing values in 16 run designs and all possible sets of a priori negligible interactions.

The results are especially interesting in the situations in which the experimenter wants to reduce the experimental plan saving some runs. It is clear that in this case the experimenter can choose which runs to skip. Then the recommendations and tables provided can be very useful to, given the interactions considered a priori negligible, choose the runs to estimate the missing values with minimum variance.

3.6 References

Almimi, A.A., Kulahci, M., Montgomery, D.C. Estimation of missing observations in two-level split-plot designs. *Quality and Reliability Engineering International* 2008; 24:127–152. DOI: 10.1002/qre.871

Box, G.E.P. A simple way to deal with missings observations from DOE. *Quality Engineering* 1990; 3(2):249-254.

Draper, N., Stoneman, D. Estimating missing values in unreplicated two-level factorial and fractional factorial designs. *Biometrics* 1964; 20(3):443-458.

Goh, T.N. Economical experimentation via 'lean design'. *Quality and Reliability Engineering International* 1996; 12:383–388.

Zhou, M., Goh, T.N. Iterative Designed Experiment Analysis (IDEA). *Quality and Reliability Engineering International* 2016; 32:2977-2986. DOI: 10.1002/qre.1981.

3.7 Appendix: Useful information to choose the interactions and/or missing values in 16-run designs

We consider negligible the five interactions of three or more factors in 2^4 designs, and those that are aliased with them in the case of fractional designs.

In the tables that follow, the numbers that appear on the columns "Missing responses" correspond to the rows of the design matrix in standard order.

3.7.1 Two missing values

Table 3.12: List of the 80 pairs of missing responses that can be estimated with minimum variance using six pairs of interactions (numbered as in Table 3.5)

Missing responses	Pairs of Interactions	Missing responses	Pairs of Interactions	Missing responses	Pairs of Interactions	Missing responses	Pairs of Interactions
1,4	2,3,5,6,9,10	3,5	1,2,6,7,8,9	5,10	2,4,5,7,8,10	8,12	2,3,4,5,6,7
1,6	1,3,5,7,8,10	3,6	1,2,3,7,9,10	5,11	3,4,6,7,8,9	8,13	1,4,5,6,9,10
1,7	1,2,6,7,8,9	3,8	1,3,5,7,8,10	5,14	1,2,4,6,8,10	8,14	1,3,4,5,8,9
1,8	1,2,3,7,9,10	3,9	1,3,4,5,8,9	5,15	1,3,4,5,8,9	8,15	1,2,4,6,8,10
1,10	1,2,4,6,8,10	3,10	1,4,5,6,9,10	5,16	1,4,5,6,9,10	9,12	2,3,5,6,9,10
1,11	1,3,4,5,8,9	3,12	1,2,4,6,8,10	6,7	2,3,5,6,9,10	9,14	1,3,5,7,8,10
1,12	1,4,5,6,9,10	3,13	3,4,6,7,8,9	6,9	2,4,5,7,8,10	9,15	1,2,6,7,8,9
1,13	2,3,4,5,6,7	3,15	2,3,4,5,6,7	6,10	2,3,4,5,6,7	9,16	1,2,3,7,9,10
1,14	2,4,5,7,8,10	3,16	2,4,5,7,8,10	6,12	3,4,6,7,8,9	10,11	2,3,5,6,9,10
1,15	3,4,6,7,8,9	4,5	1,2,3,7,9,10	6,13	1,2,4,6,8,10	10,13	1,3,5,7,8,10
2,3	2,3,5,6,9,10	4,6	1,2,6,7,8,9	6,15	1,4,5,6,9,10	10,15	1,2,3,7,9,10
2,5	1,3,5,7,8,10	4,7	1,3,5,7,8,10	6,16	1,3,4,5,8,9	10,16	1,2,6,7,8,9
2,7	1,2,3,7,9,10	4,9	1,4,5,6,9,10	7,9	3,4,6,7,8,9	11,13	1,2,6,7,8,9
2,8	1,2,6,7,8,9	4,10	1,3,4,5,8,9	7,11	2,3,4,5,6,7	11,14	1,2,3,7,9,10
2,9	1,2,4,6,8,10	4,11	1,2,4,6,8,10	7,12	2,4,5,7,8,10	11,16	1,3,5,7,8,10
2,11	1,4,5,6,9,10	4,14	3,4,6,7,8,9	7,13	1,3,4,5,8,9	12,13	1,2,3,7,9,10
2,12	1,3,4,5,8,9	4,15	2,4,5,7,8,10	7,14	1,4,5,6,9,10	12,14	1,2,6,7,8,9
2,13	2,4,5,7,8,10	4,16	2,3,4,5,6,7	7,16	1,2,4,6,8,10	12,15	1,3,5,7,8,10
2,14	2,3,4,5,6,7	5,8	2,3,5,6,9,10	8,10	3,4,6,7,8,9	13,16	2,3,5,6,9,10
2,16	3,4,6,7,8,9	5,9	2,3,4,5,6,7	8,11	2,4,5,7,8,10	14,15	2,3,5,6,9,10

3.7.2 Three missing values

Table 3.13. List of the 160 trios of missing responses that can be estimated with minimum variance using seven trios of interactions. (numbered as in Table 3.7)

Missing responses	Trios of Interactions	Missing responses	Trios of Interactions	Missing responses	Trios of Interactions	Missing responses	Trios of Interactions
1,4,6	1,2,3,4,7,9,10	2,5,16	1,2,3,4,5,8,9	3,10,16	1,2,3,4,6,8,10	5,14,15	1,4,5,6,7,9,10
1,4,7	1,2,3,4,7,9,10	2,7,9	1,2,5,6,7,8,9	3,12,13	1,2,5,6,7,8,9	6,7,9	1,2,3,4,5,6,7
1,4,10	1,4,5,6,7,9,10	2,7,11	2,3,5,6,8,9,10	3,12,15	2,4,5,7,8,9,10	6,7,12	1,2,3,4,5,6,7
1,4,11	1,4,5,6,7,9,10	2,7,12	1,3,5,6,7,8,10	3,13,16	1,2,3,4,5,6,7	6,7,13	1,4,5,6,7,9,10
1,4,14	1,2,3,4,5,6,7	2,7,13	1,3,5,6,7,8,10	4,5,9	2,3,5,6,8,9,10	6,7,16	1,4,5,6,7,9,10
1,4,15	1,2,3,4,5,6,7	2,7,14	2,3,5,6,8,9,10	4,5,10	1,3,5,6,7,8,10	6,9,12	1,2,3,4,5,6,7
1,6,7	1,2,3,4,7,9,10	2,7,16	1,2,5,6,7,8,9	4,5,11	1,2,5,6,7,8,9	6,9,15	1,2,3,4,6,8,10
1,6,10	2,4,5,7,8,9,10	2,8,11	1,2,3,4,6,8,10	4,5,14	1,2,5,6,7,8,9	6,9,16	1,3,5,6,7,8,10
1,6,12	1,2,3,4,5,8,9	2,8,12	3,4,6,7,8,9,10	4,5,15	1,3,5,6,7,8,10	6,10,13	2,4,5,7,8,9,10
1,6,13	2,4,5,7,8,9,10	2,8,13	1,2,3,4,6,8,10	4,5,16	2,3,5,6,8,9,10	6,10,15	2,3,5,6,8,9,10
1,6,15	1,2,3,4,5,8,9	2,8,14	3,4,6,7,8,9,10	4,6,7	1,2,3,4,7,9,10	6,10,16	3,4,6,7,8,9,10
1,7,11	3,4,6,7,8,9,10	2,9,12	1,4,5,6,7,9,10	4,6,9	1,2,3,4,6,8,10	6,12,13	1,2,5,6,7,8,9
1,7,12	1,2,3,4,6,8,10	2,9,14	2,4,5,7,8,9,10	4,6,10	3,4,6,7,8,9,10	6,12,15	1,2,3,4,5,8,9
1,7,13	3,4,6,7,8,9,10	2,9,16	1,2,5,6,7,8,9	4,6,15	1,2,3,4,6,8,10	6,13,16	1,4,5,6,7,9,10
1,7,14	1,2,3,4,6,8,10	2,11,13	1,2,3,4,6,8,10	4,6,16	3,4,6,7,8,9,10	7,9,12	1,2,3,4,5,6,7
1,8,10	1,2,5,6,7,8,9	2,11,14	2,3,5,6,8,9,10	4,7,9	1,2,3,4,5,8,9	7,9,14	1,2,3,4,5,8,9
1,8,11	1,3,5,6,7,8,10	2,11,16	1,2,3,4,5,8,9	4,7,11	2,4,5,7,8,9,10	7,9,16	1,2,5,6,7,8,9
1,8,12	2,3,5,6,8,9,10	2,12,13	1,3,5,6,7,8,10	4,7,14	1,2,3,4,5,8,9	7,11,13	3,4,6,7,8,9,10
1,8,13	2,3,5,6,8,9,10	2,12,14	3,4,6,7,8,9,10	4,7,16	2,4,5,7,8,9,10	7,11,14	2,3,5,6,8,9,10
1,8,14	1,3,5,6,7,8,10	2,13,16	1,2,3,4,5,6,7	4,9,14	1,2,3,4,5,8,9	7,11,16	2,4,5,7,8,9,10
1,8,15	1,2,5,6,7,8,9	3,5,8	1,2,3,4,7,9,10	4,9,15	1,2,3,4,6,8,10	7,12,13	1,3,5,6,7,8,10
1,10,11	1,4,5,6,7,9,10	3,5,9	3,4,6,7,8,9,10	4,9,16	2,3,5,6,8,9,10	7,12,14	1,2,3,4,6,8,10
1,10,13	2,4,5,7,8,9,10	3,5,10	1,2,3,4,6,8,10	4,10,11	1,4,5,6,7,9,10	7,13,16	1,4,5,6,7,9,10
1,10,15	1,2,5,6,7,8,9	3,5,15	3,4,6,7,8,9,10	4,10,15	1,3,5,6,7,8,10	8,10,11	1,2,3,4,5,6,7
1,11,13	3,4,6,7,8,9,10	3,5,16	1,2,3,4,6,8,10	4,10,16	3,4,6,7,8,9,10	8,10,13	1,2,3,4,5,8,9
1,11,14	1,3,5,6,7,8,10	3,6,9	1,3,5,6,7,8,10	4,11,14	1,2,5,6,7,8,9	8,10,15	1,2,5,6,7,8,9
1,12,13	2,3,5,6,8,9,10	3,6,10	2,3,5,6,8,9,10	4,11,16	2,4,5,7,8,9,10	8,11,13	1,2,3,4,6,8,10
1,12,14	1,2,3,4,6,8,10	3,6,12	1,2,5,6,7,8,9	4,14,15	1,2,3,4,5,6,7	8,11,14	1,3,5,6,7,8,10
1,12,15	1,2,3,4,5,8,9	3,6,13	1,2,5,6,7,8,9	5,8,10	1,2,3,4,5,6,7	8,12,13	2,3,5,6,8,9,10
1,14,15	1,2,3,4,5,6,7	3,6,15	2,3,5,6,8,9,10	5,8,11	1,2,3,4,5,6,7	8,12,14	3,4,6,7,8,9,10
2,3,5	1,2,3,4,7,9,10	3,6,16	1,3,5,6,7,8,10	5,8,14	1,4,5,6,7,9,10	8,12,15	2,4,5,7,8,9,10
2,3,8	1,2,3,4,7,9,10	3,8,10	1,2,3,4,5,8,9	5,8,15	1,4,5,6,7,9,10	8,14,15	1,4,5,6,7,9,10
2,3,9	1,4,5,6,7,9,10	3,8,12	2,4,5,7,8,9,10	5,9,14	2,4,5,7,8,9,10	9,12,14	1,2,3,4,7,9,10
2,3,12	1,4,5,6,7,9,10	3,8,13	1,2,3,4,5,8,9	5,9,15	3,4,6,7,8,9,10	9,12,15	1,2,3,4,7,9,10
2,3,13	1,2,3,4,5,6,7	3,8,15	2,4,5,7,8,9,10	5,9,16	2,3,5,6,8,9,10	9,14,15	1,2,3,4,7,9,10
2,3,16	1,2,3,4,5,6,7	3,9,12	1,4,5,6,7,9,10	5,10,11	1,2,3,4,5,6,7	10,11,13	1,2,3,4,7,9,10
2,5,8	1,2,3,4,7,9,10	3,9,15	3,4,6,7,8,9,10	5,10,15	1,3,5,6,7,8,10	10,11,16	1,2,3,4,7,9,10
2,5,9	2,4,5,7,8,9,10	3,9,16	1,3,5,6,7,8,10	5,10,16	1,2,3,4,6,8,10	10,13,16	1,2,3,4,7,9,10
2,5,11	1,2,3,4,5,8,9	3,10,13	1,2,3,4,5,8,9	5,11,14	1,2,5,6,7,8,9	11,13,16	1,2,3,4,7,9,10
2,5,14	2,4,5,7,8,9,10	3,10,15	2,3,5,6,8,9,10	5,11,16	1,2,3,4,5,8,9	12,14,15	1,2,3,4,7,9,10

3.7.3 Four missing values

Table 3.14: List of the 100 combinations of 4 missing responses that cannot be estimated from 5 null interactions

1,2,3,4	1,6,9,14	2,7,10,15	4,6,12,14	6,8,9,11
1,2,5,6	1,6,11,16	2,8,9,15	4,7,10,13	6,8,10,12
1,2,7,8	1,7,9,15	2,8,10,16	4,7,12,15	6,8,13,15
1,2,9,10	1,7,10,16	3,4,5,6	4,8,9,13	6,8,14,16
1,2,11,12	1,8,9,16	3,4,7,8	4,8,10,14	7,8,9,10
1,2,13,14	2,3,6,7	3,4,9,10	4,8,11,15	7,8,11,12
1,2,15,16	2,3,10,11	3,4,11,12	4,8,12,16	7,8,13,14
1,3,5,7	2,3,14,15	3,4,13,14	5,6,7,8	7,8,15,16
1,3,6,8	2,4,5,7	3,4,15,16	5,6,9,10	9,10,11,12
1,3,9,11	2,4,6,8	3,5,11,13	5,6,11,12	9,10,13,14
1,3,10,12	2,4,9,11	3,5,12,14	5,6,13,14	9,10,15,16
1,3,13,15	2,4,10,12	3,6,11,14	5,6,15,16	9,11,13,15
1,3,14,16	2,4,13,15	3,7,9,13	5,7,9,11	9,11,14,16
1,4,5,8	2,4,14,16	3,7,10,14	5,7,10,12	9,12,13,16
1,4,9,12	2,5,10,13	3,7,11,15	5,7,13,15	10,11,14,15
1,4,13,16	2,5,12,15	3,7,12,16	5,7,14,16	10,12,13,15
1,5,9,13	2,6,9,13	3,8,9,14	5,8,9,12	10,12,14,16
1,5,10,14	2,6,10,14	3,8,11,16	5,8,13,16	11,12,13,14
1,5,11,15	2,6,11,15	4,5,12,13	6,7,10,11	11,12,15,16
1,5,12,16	2,6,12,16	4,6,11,13	6,7,14,15	13,14,15,16

Table 3.15: List of the 40 combinations of 4 missing responses that can be estimated with minimum variance. Quartets of interactions numbered as in Table 3.9.

Missing values	Quartets of interac.	Missing values	Quartets of interac.	Missing values	Quartets of interac.	Missing values	Quartets of interac.
1,4,6,7	1,5	2,3,5,8	1,5	3,5,9,15	4,5	4,6,9,15	1,4
1,4,10,11	2,5	2,3,9,12	2,5	3,5,10,16	1,4	4,6,10,16	4,5
1,4,14,15	1,2	2,3,13,16	1,2	3,6,9,16	2,4	4,7,9,14	1,3
1,6,10,13	3,5	2,5,9,14	3,5	3,6,10,15	3,4	4,7,11,16	3,5
1,6,12,15	1,3	2,5,11,16	1,3	3,6,12,13	2,3	5,8,10,11	1,2
1,7,11,13	4,5	2,7,9,16	2,3	3,8,10,13	1,3	5,8,14,15	2,5
1,7,12,14	1,4	2,7,11,14	3,4	3,8,12,15	3,5	6,7,9,12	1,2
1,8,10,15	2,3	2,7,12,13	2,4	4,5,9,16	3,4	6,7,13,16	2,5
1,8,11,14	2,4	2,8,11,13	1,4	4,5,10,15	2,4	9,12,14,15	1,5
1,8,12,13	3,4	2,8,12,14	4,5	4,5,11,14	2,3	10,11,13,16	1,5

3.7.4 Five missing values

Table 3.16: List of the 16 combinations of 5 missing responses that can be estimated with minimum variance

1,4,6,10,15	1,8,12,14,15	2,7,9,12,14	3,6,10,13,16
1,4,7,11,14	2,3,5,9,16	2,7,11,13,16	4,5,9,14,15
1,6,7,12,13	2,3,8,12,13	3,5,8,10,15	4,5,10,11,16
1,8,10,11,13	2,5,8,11,14	3,6,9,12,15	4,6,7,9,16

Chapter 4

ARTICLE 3

Which runs to skip in two level factorial designs when not all can be performed

4

ARTICLE 3

Which runs to skip in two level factorial designs when not all can be performed

Rafel Xampeny, Pere Grima, Xavier Tort-Martorell
Department of Statistics and Operational Research
Universitat Politècnica de Catalunya – BarcelonaTech, Spain

Quality Engineering
Accepted author version posted online: 18 January 2018
DOI: 10.1080/08982112.2018.1428751
https://doi.org/10.1080/08982112.2018.1428751

ABSTRACT

When a two level factorial design allows estimating contrasts that can be considered negligible from scratch, it is possible to omit some runs and later estimate their values by equating to zero the expressions of some of that contrasts. This article presents the combinations of runs to be omitted in 8 and 16 runs two level factorial designs so that the responses can be estimated in such a way as to produce the least possible impact on the desired properties of the estimated contrasts: low and equal variance and the smallest possible correlation among them.
KEYWORDS: Factorial designs, missing values, negligible interactions, saving runs, effects' variance, effects' correlation.

4.1 Introduction

Having limited resources available can make it unfeasible to conduct all the desired runs in a two level experimental design. It is then reasonable to choose which runs to omit in such a way that would allow obtaining the maximum information and, especially considering practitioners needs, in the simplest possible way.

Several authors have studied the problem of analyzing factorial designs with missing values and proposed different solutions. In a review paper, Jarret [1978] highlights the solution used by Wilkinson [1958a, 1958b], who analyzes the available results without estimating the missing values. Draper and Stoneman [1964] proposed using the contrasts that can be considered null for estimating the results of experiments that could not be performed due to unforeseen circumstances. Their method can also be applied when one has already anticipated that some runs cannot be performed due to a lack of resources. John [1979] underscores the problems that are inevitably involved in the analysis of factorial designs with missing values. In a short article written for didactic purposes, Box [1990] advocates deducing the missing values by following the procedure proposed by Draper and Stoneman and, he suggests not estimating more than one missing value in 8 run designs and no more than two in 16 run designs. Srivastava *et al.* [1991] study the robustness of several types of designs against missing data. More recently, Godolphin [2006] proposes a procedure to assess the impact of missing values in certain types of blocked factorial designs. Almini *et al.* [2007] extend Draper's and Stoneman's method to two-level split-plot designs.

Already in the line of our proposal, that is with the aim of saving runs and conducting the most economical experimental plans possible, Goh [1996] proposes the so-called "Lean Design", which is also based on the idea of estimating missing values from negligible contrasts. Along the same lines, Zhou and Goh [2016] recently proposed a sequential strategy aimed at reducing – when possible – the number of runs performed. Another possible approach is to design a D-optimal experiment with the required number of runs. D-optimal designs are chosen to maximize the determinant of the information matrix $X'X$, where X is the contrast matrix. This requires the experimenter to specify the model he or she wants to estimate. One good property of these designs is that the average variance of the effects is minimized. Johnson et al. [2011] and Robinson and Anderson-Cook [2011] provide a detailed view of the characteristics and possibilities of D-optimal designs, and the book by Goos and Jones [2011] presents their advantages and practical possibilities with a variety of real-world examples. In addition, there are a number of statistical packages that support D-optimal designs. All this seems to indicate that this type of designs are very well suited to solving the problem of reducing the number of runs required by two level factorial designs. However, they present some inconveniences that hinder their being used by practitioners without solid training in experimental design. The main hindrance is that, in general, the estimated effects have different variances that complicate the analysis of significance. Also, the provided designs often change the factor levels, even if only slightly. On top of that, there is the fact that practitioners may lose control of the situation, specifically by getting the impression that they are conducting an optimum experiment without really understanding that it is in fact optimum in a very specific sense and only in the case that the specified model is correct.

In this paper we identify the runs that can be omitted in 8- and 16-run two-level factorial designs in order to make the estimation of their values and analysis of the experimental results as easy as possible. Our proposal is based on the Draper and Stoneman method of using contrasts that can be considered negligible a priori (such as three or more factor interactions) in order to estimate the response of the runs to be omitted. This allows calculating the effects and assessing their significance as if all runs have been conducted, which provides an easy way out for practitioners with limited training in DOE, such as Six Sigma Black Belts. In Section 4.2, we provide the variance of the estimated effects and the correlation coefficients that arise among them, depending on the number and position of the runs omitted. In Section 4.3 we discuss how to choose which runs to leave out and also provide a fast and simple way for practitioners to do so (a detailed and complete discussion of all possibilities is provided in the Appendix). Section 4.4 offers 2 examples and, finally, in Section 4.5 we summarize our findings and provide some conclusions.

4.2 Variance and correlation of effects calculated from estimated responses

Our study focuses on two level factorial designs that require conducting 8 or 16 runs. Designs with 32 or more runs are not common, neither in practice nor in the technical literature. What is more, when so many runs are planned, it will rarely be in a context of limited resources that require omitting some of them.

Among the designs with 8 and 16 runs, we study those that allow estimating contrasts that may be considered null a priori, i. e., contrasts that estimate strings of interactions of three or more factors. These are 2^3 and 2^4 full factorials and 2^{6-2} and 2^{7-3} fractional designs that – with appropriate generators– allow estimating contrasts (respectively, two and one) which include only interactions of three or more factors. In other 8 and 16 run designs, unless the experimenter has a strong technical knowledge allowing him to disregard some two factor interactions, there are no a priori negligible contrast. Thus, neither this nor other methodologies based in this fact can be applied.

4.2.1 2^3 Design

In this design, it can be considered that the interaction of three factors (ABC) is negligible and therefore the result of one missing run can be estimated.
If σ_y^2 is the experimental variance and the eight experimental results are available, the variance of the effects equals $\sigma_y^2/2$ (see, for example, Box *et al.* [2005]). However, if we have a missing value, say y_1, we can estimate it by:

$$y_1 = y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8$$

For example, the main effect of factor A will be:

$$\begin{aligned} A = \frac{1}{4}(- \ y_2 - \ y_3 + \ y_4 - \ y_5 + \ y_6 + \ y_7 - \ y_8 \\ + \ y_2 - \ y_3 + \ y_4 - \ y_5 + \ y_6 - \ y_7 + \ y_8) = \\ = \frac{1}{4}(-2y_3 + 2y_4 - 2y_5 + 2y_6) \end{aligned}$$

Therefore:

$$V(A) = \sigma_y^2$$

It can be verified that all the effects have the same variance except, of course, the one considered negligible that is zero with null variance.

Another problem is that the effects are not independent. For example, we have:

$$B = \frac{1}{4}(-2y_2 + 2y_4 - 2y_5 + 2y_7)$$

Therefore:

$$V(A+B) = V\left[\frac{1}{4}(-2y_2 - 2y_3 + 4y_4 - 4y_5 + 2y_6 + 2y_7)\right] =$$

$$= 3\sigma_y^2$$

Since $V(A+B) = V(A) + V(B) + 2\mathrm{Cov}(A,B)$, we get that $\mathrm{Cov}(A,B) = 0.5\sigma_y^2$. It is easily verifiable that all the effects share the same covariance in absolute values. Therefore, in this case it does not matter which is the missing value, as its influence on the effects is always the same.

To better assess the degree of correlation that occurs among the effects, we can calculate the correlation coefficient, which in this case is $\rho = \frac{0.5\sigma_y^2}{\sigma_y \sigma_y} = 0.5$.

4.2.2 2^4 Design

By considering the interactions of three or more factors negligible, up to 5 missing values can be estimated in this type of design; therefore up to 5 runs can be skipped.

When only one value is missing, it can be estimated from any of the negligible interactions in the same way as in 2^3 designs. Table 4.1 presents the contrasts that correspond to interactions of 3 and 4 factors in 2^4 factorial designs.

Table 4.1: Contrasts – in standard order – corresponding to the 3 and 4 factor interactions in a 2^4 design.

Interaction		y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}
1	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
2	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
3	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
4	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
5	$ABCD$	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1

If, for example, the value of y_1 is missing, we can estimate it from the expression of the interaction $ABCD$ equated to zero:

$$y_1 = y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8 + y_9 - y_{10} - y_{11} + y_{12} - y_{13} + y_{14} + y_{15} - y_{16}$$

So that, for example, the effect A will be estimated as:

$$A = \frac{1}{8}(+y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8 - y_9 + y_{10} - y_{11} + y_{12} - y_{13} + y_{14} - y_{15} + y_{16} -$$
$$-y_2 - y_3 + y_4 - y_5 + y_6 + y_7 - y_8 - y_9 + y_{10} + y_{11} - y_{12} + y_{13} - y_{14} - y_{15} + y_{16}) =$$
$$= \frac{1}{8}(-2y_3 + 2y_4 - 2y_5 + 2y_6 - 2y_9 + 2y_{10} - 2y_{15} + 2y_{16})$$

From which it follows that $V(A) = 0.5\sigma_y^2$. Just as in the 2^3 design, the variance is the same for all effects of interest (the main effects and the interactions of two factors) and its value does not depend on which is the missing value. However, in this case the variance of the effects can be reduced by estimating the missing value from each of the five interactions considered null and using the average of them as the estimate. For example, if the missing value is y_1, we have:

Null interaction	Estimate of y_1
ABC	$+y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8 - y_9 + y_{10} + y_{11} - y_{12} + y_{13} - y_{14} - y_{15} + y_{16}$
ABD	$+y_2 + y_3 - y_4 - y_5 + y_6 + y_7 - y_8 + y_9 - y_{10} - y_{11} + y_{12} + y_{13} - y_{14} - y_{15} + y_{16}$
ACD	$+y_2 - y_3 + y_4 + y_5 - y_6 + y_7 - y_8 + y_9 - y_{10} + y_{11} - y_{12} - y_{13} + y_{14} - y_{15} + y_{16}$
BCD	$-y_2 + y_3 + y_4 + y_5 + y_6 - y_7 - y_8 + y_9 + y_{10} - y_{11} - y_{12} - y_{13} - y_{14} + y_{15} + y_{16}$
ABCD	$+y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8 + y_9 - y_{10} - y_{11} + y_{12} - y_{13} + y_{14} + y_{15} - y_{16}$
Mean	$\frac{1}{5}(3y_2 + 3y_3 - y_4 + 3y_5 - y_6 - y_7 - y_8 + 3y_9 - y_{10} - y_{11} - y_{12} - y_{13} - y_{14} - y_{15} + 3y_{16})$

Then that the main effect of the factor A will be estimated as:

$$A = \frac{1}{8}[+y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8 - y_9 + y_{10} - y_{11} + y_{12} - y_{13} + y_{14} - y_{15} + y_{16} -$$
$$- \frac{1}{5}(3y_2 + 3y_3 - y_4 + 3y_5 - y_6 - y_7 - y_8 + 3y_9 - y_{10} - y_{11} - y_{12} - y_{13} - y_{14} - y_{15} + 3y_{16})] =$$
$$= \frac{1}{8}\left(\frac{2}{5}y_2 - \frac{8}{5}y_3 + \frac{6}{5}y_4 - \frac{8}{5}y_5 + \frac{6}{5}y_6 - \frac{4}{5}y_7 + \frac{6}{5}y_8 - \frac{8}{5}y_9 + \frac{6}{5}y_{10} - \frac{4}{5}y_{11} + \frac{6}{5}y_{12} - \frac{4}{5}y_{13} + \right.$$
$$\left. + \frac{6}{5}y_{14} - \frac{4}{5}y_{15} + \frac{2}{5}y_{16}\right)$$

From which it follows that $V(A) = 0.3\sigma_y^2$. It can also be verified that this value is the same for all the effects of interest and that it is independent of the run whose result is estimated.

Below we discuss the situations in which we want to skip more than one run and, thus, we have to estimate more than one missing value. In these cases the variances of the effects depend on which runs have been skipped and on the interactions used to estimate them. Depending on the runs omitted the variances of the estimated effects will be equal or different and bigger or smaller. To judge the different situations that arise and decide which runs to skip and which interactions to use for estimating them, we use a scatterplot of two indicators: the average of the variances and the maximum variance of the estimated effects.

Although when there is only one missing value the situation is clear and the plot is unnecessary, we use it for illustration. Figure 4.1 (left) shows the maximum versus the mean value of the variances of the effects according to the number of null interactions – and therefore the number of equations – that have been used. The dashed line represents points that have a mean value that is equal to its maximum value. In this simple case, since all the effects have the same variance, all points fall on this line.

To assess the degree of dependence between effects, we create a similar figure in which we plot the correlation coefficients (in absolute values) between the effects on a plot of maximum versus mean values. By looking at the two plots together, it is easy to choose the best option. In this case, both plots point in the same direction: a greater effects variance is also accompanied by a higher correlation coefficient. Therefore, as expected, the greater the number of negligible interactions used to estimate the missing values, the better the characteristics of the estimated effects.

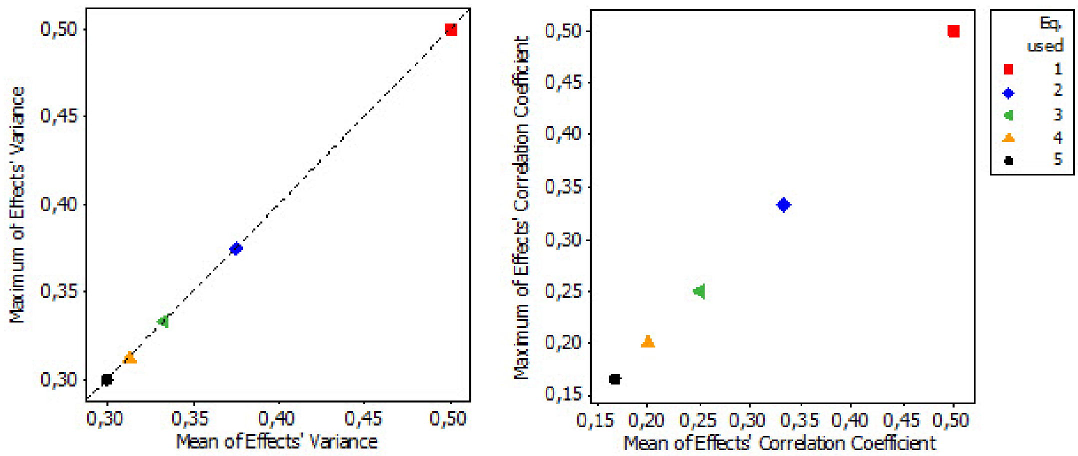

Figure 4.1: One missing value: Maximum versus mean of the effects variance (left) and the correlation coefficients in absolute values (right).

If we want to skip two runs, the two missing values can be estimated by equating two negligible interactions to zero to get a system of two equations with two unknowns. Since there are 5 a priori null interactions, in theory ten systems can be used. But, it is not possible to use just any pair of interactions because – depending on which ones are the missing responses – some pairs will lead to linearly dependent systems. In this case, of the 120 possible pairs of missing values, 80 can be estimated with 6 systems of equations while the remaining 40 pairs can be estimated with only 4 systems.

Table 4.2 (Xampeny *et al.* [2017]) provides information on which pairs of missing values can be estimated with which interactions. In it, the 10 pairs that can be formed with the 5 given interactions are divided into two groups (white background and grey background) in such a way that, given two missing responses, only the pairs of interactions which have one response in each group are useful.

Table 4.2: Pairs of negligible interactions to be used for estimating two missing responses.

		y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}
1	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
2	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
3	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
4	ABC	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1
5	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
6	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
7	ABD	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1
8	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
9	ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1
10	BCD	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
	ABCD	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1

For example, if the missing values are y_1 and y_2, their values cannot be estimated with the first pair of interactions (ABC and ABD) because the equations obtained are linearly dependent. In this case, they can be estimated only with pairs 3, 6, 8 and 10 (according to the number scheme shown in Table 4.3). However, if the missing values are y_1 and y_4, these can be estimated with 6 systems of equations, namely those corresponding to the pairs of interactions 2, 3, 5, 6, 9 and 10.

To evaluate the possible choices of which pairs of runs to skip, we produce the same two plots (Figure 4.2) as before. Since deducing the effects variances on a case-by-case basis is tedious, we have developed some routines in R (R Core Team [2016]) for evaluating all possible cases.

In this case there is no set of missing values that is clearly better than the others. Using six systems of equations, we obtain the minimum mean value of the variance of the effects of interest ($0.3583\sigma_y^2$). Another interesting possibility is to use only two systems of equations involving only 4 of the 5 available null interactions. This produces a similar mean value of the variance ($0.3750\sigma_y^2$), with the advantage of it being the same for all the effects of interest, which facilitates the analysis of its statistical significance.

With regard to the correlation coefficients, these are only slightly higher when using the two systems of equations (Framed points in Figure 4.2. In fact the framed points are diamonds that in this case are covered by triangles).

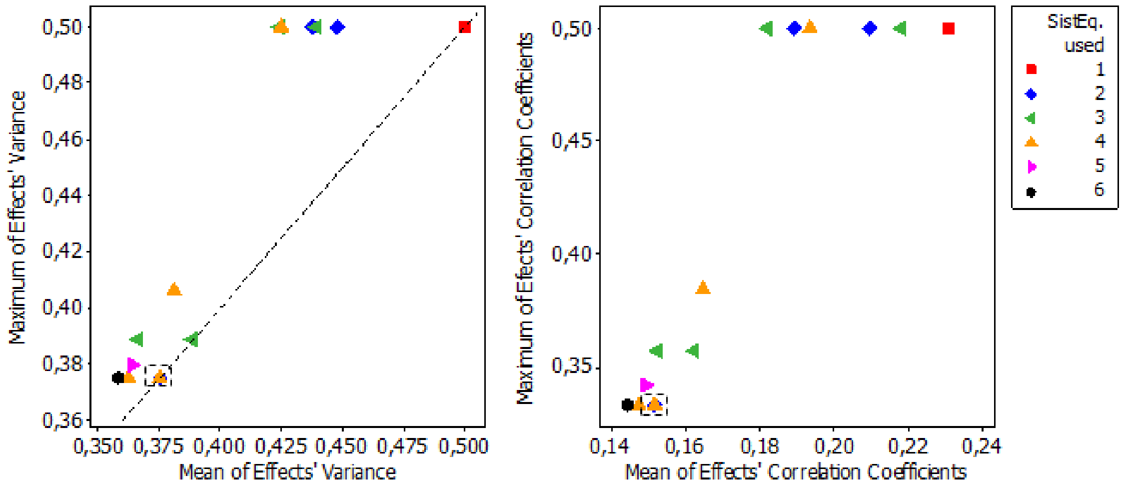

Figure 4.2: Two missing values: Maximum versus mean of the variance effects of interest (left) and correlation coefficients (right). The most interesting points are framed.

If we want to skip three runs, we will have three missing values to estimate. 160 of the 560 possible trios can be estimated using 7 systems of 3 equations with 3 unknowns, and these are the ones that allow obtaining the minimum mean value ($0.4214\sigma_y^2$) for the variances of the effects of interest.

However, similarly to when two values are missing, there are other trios of missing values that allow estimating the effects with a mean variance that is only slightly higher ($0.4375\sigma_y^2$) and with the added advantage of being the same for all the effects of interest. These are the trios that can be estimated with four systems of three equations using only four null interactions. (Framed point in Figure 4.3).

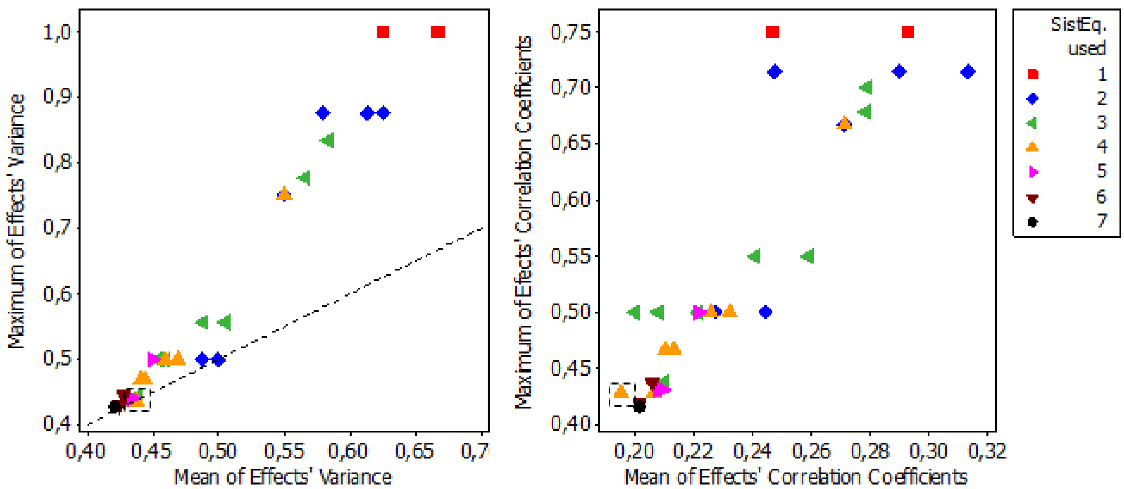

Figure 4.3: Three missing values: Maximum versus mean of the variance effects of interest (left) and correlation coefficients (right). The most interesting points are framed.

The case of skipping four runs is curious, because it does not follow the general rule that using more systems of equations deliver better results. Thus, of the 1820 possible sets of four missing values, there are only 80 that can be estimated with 5 systems of equations (the maximum) and the estimated effects have an average variance of $0.4920\sigma_y^2$. Yet, there also exist 40 combinations that can be estimated with only two systems of equations, and a slightly smaller mean variance ($0.4875\sigma_y^2$).

Nevertheless, we consider that the best option is to use one of the 40 quartets that can be estimated with a single system of 4 equations. This cases give the minimum mean value for the effects correlation and a very close to the minimum mean variance of the effects ($0.5\sigma_y^2$), the same for all of them (Figure 4.4).

It should be noted that there are also 100 quartets whose values are impossible to estimate (see Xampeny *et al.* [2017]) for the list of these quartets). Care should be taken not to omit quartets of runs from this set.

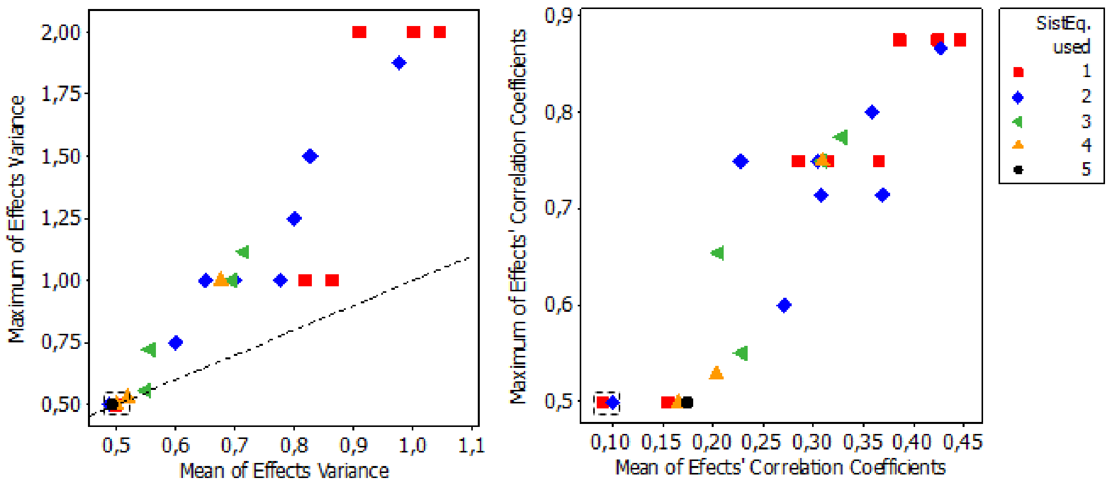

Figure 4.4: Four missing values: Maximum versus mean of the variance effects of interest (left) and correlation coefficients (right). The most interesting points are framed.

In the case of wanting to skip 5 runs and thus having 5 missing values to estimate, there is no other option but to set up a system of 5 equations with 5 unknowns. Of the 4368 possible combinations of missing values, there are 1360 that are impossible to calculate because the system of equations is inconsistent. There exists an exceptional group of 16 combinations of missing values that allow estimating the effects with a minimum variance ($0.5\sigma_y^2$) that in addition is also the same for all the effects of interest and have relatively low correlation among them (Figure 4.5).

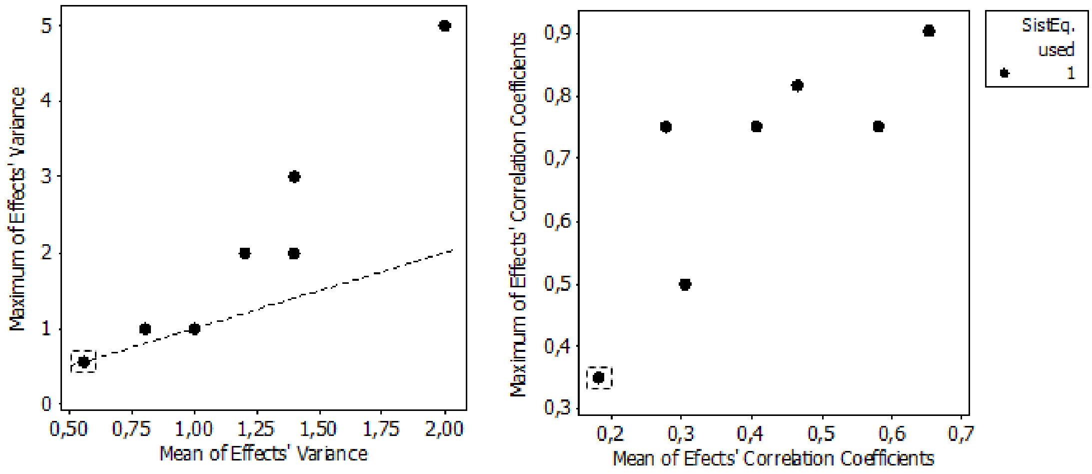

Figure 4.5: Five missing values: Maximum versus mean of the variance effects of interest (left) and correlation coefficients (right). The points of interest are framed.

4.2.3 Fractional factorial designs

So far, we have considered full factorial designs. The idea of saving runs and estimating the response by equating negligible interactions to zero is in general not applicable to fractional factorial designs. In them, the number of runs has already been cut off to the point that the same contrast is used to estimate several effects in what is called confounding and higher order interactions are usually confounded with the effects of interest: the main effects or two factor interactions. Thus, it is not possible to use them for estimating missing values.

Among the 16 run designs there are two exceptions: 2_{IV}^{6-2} and 2_{IV}^{7-3}. They have contrasts that are formed exclusively by interactions that can be considered negligible.

For example, a 2_{IV}^{6-2} design with generators E=ABC and F=BCD has the alias structure provided in Table 4.3, which shows that the last two contrasts estimate strings of three factor interactions, which can thus be used to estimate the missing values.

Table 4.3: Alias structure of a 2^{6-2} design with generators E=ABC and F=BCD. In bold, the effects corresponding to a 2^4 design

I + ABCE + ADEF + BCDF

A + BCE + DEF + ABCDF
B + ACE + CDF + ABDEF
C + ABE + BDF + ACDEF
D + AEF + BCF + ABCDE
E + **ABC** + ADF + BCDEF
F + ADE + **BCD** + ABCEF
AB + CE + ACDF + BDEF
AC + BE + ABDF + CDEF
AD + EF + ABCF + BCDE
AE + **BC** + DF + ABCDEF
AF + DE + **ABCD** + BCEF
BD + CF + ABEF + ACDE
BF + **CD** + ABDE + ACEF
ABD + ACF + BEF + CDE
ABF + **ACD** + BDE + CEF

If we want to save one run, its value can be estimated as the average of the values obtained from the expression of the two contrasts that are considered null. In this case, the effects of interest are estimated with a variance of $0.375\sigma_y^2$, and it is the same for all of them (Figure 4.1). If we want to save two runs, only 64 pairs of the 120 possible pairs of runs can be omitted. Table 4.3 can be used to identify the suitable pairs.

The case of the 2_{IV}^{7-3} design is simpler. There is only one contrast that involves only higher-order interactions. Therefore, only one run can be saved and the variance of the effects will be $0.5\sigma_y^2$, as seen above.

4.3 Selection of runs to skip

As already stated above, it is convenient to skip runs in such a way that estimating their responses produce effects with similar and small variances and that are among them as independent as possible. This facilitates the analysis of their statistical significance through standard procedures, such as their representation on a Normal Probability Plot (NPP). Fortunately, as we have shown when the effects are estimated with minimum variance, they also tend to have the lowest correlations among them; therefore, the problem can be reduced to finding the best behavior of the variance.

Figure 4.6 illustrates the importance of properly choosing the runs to skip by showing how the variance and correlation of the effects vary between the best (recommended) and worst options in the case of estimating up to 5 values in a 2^4 design.

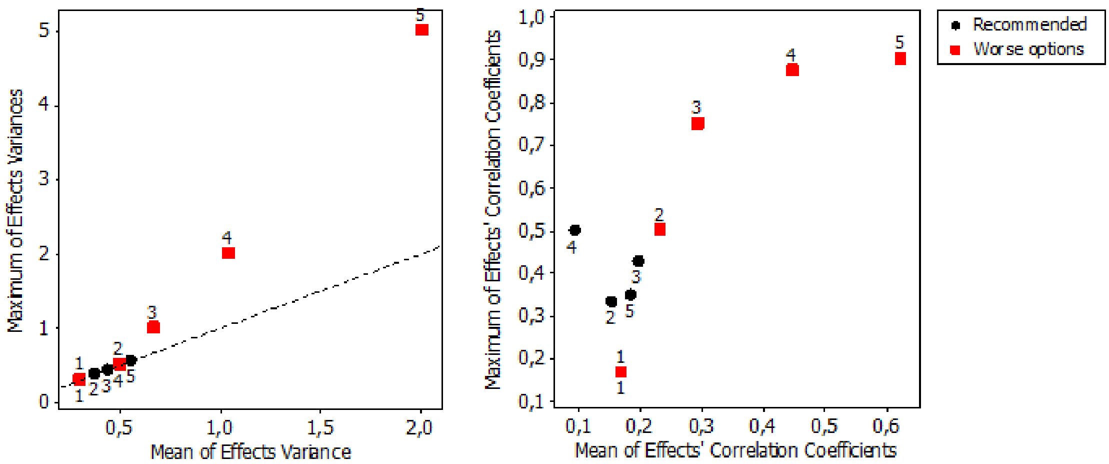

Figure 4.6: Behavior of variances (left) and correlation coefficients (right) for a good (circles) and bad (boxes) selection of runs to omit. The number next to each item indicates the number of missing values that are estimated.

Table 4.4 presents the number of missing value combinations that provide the desired properties of minimum value when the variance is constant (and there is therefore low correlation among the effects). In it, we can also see the number of combinations providing the worst properties (maximum variance). It is clear in Table 4.4 that if no criterion is used to select the experimental conditions in which the results will be estimated, there is a high risk of ending up with a poor selection, i.e., a selection with bad properties.

Table 4.4: Number of missing value combinations with the best and worst characteristics

	Runs to skip				
	1	2	3	4	5
Number of combinations of missing values	16	120	560	1820	4368
Number of recommended combinations	16	80	160	40	16
Number of worse combinations	-	40	160	480	480
Combinations impossible to estimate	-	-	-	100	1360

Table 4.5 presents a guide for quickly and simply identifying which runs to skip in any 8 and 16 runs designs with high order interactions that can be considered zero from scratch. The guide is addressed to practitioners. The proposed runs have been chosen from the exhaustive lists presented in the Appendix, such that they not only provide the desired properties but they also prioritize using four factor interactions whenever possible. In cases with several possible options, one was chosen at random.

Table 4.5: Practical guide for selecting the runs to skip and the interactions to use

Design	Number of runs to skip	Runs to skip*	How to estimate the missing results	Interactions to use
2^3	1	Any	From equaling to zero the null interaction.	ABC (The only one that can be considered negligible)
2^4	1	Any	Mean of the 5 values obtained from equaling to zero the 5 null interactions.	The five that can be considered negligible
	2	y_6, y_{12}	For each missing value: Mean of the two results obtained solving two systems of equations	First system: ABC and BCD Second system: ABD and ABCD
	3	y_4, y_6, y_{10}	For each missing value: Mean of the four results obtained from solving four systems of equations	First system: ABD, ACD and ABCD Second system: ABC, ACD and ABCD Third system: ABC, ABD and ABCD Fourth system: ABC, ABD and ACD
	4	y_2, y_7, y_9, y_{16}	Results obtained from one system of four equations	ABC, ABD, BCD and ABCD
	5	$y_2, y_3, y_5, y_9, y_{16}$	Results obtained from one system of five equations	The five that can be considered negligible
2^{6-2}	1	Any	Mean of the two results obtained from each null contrast	The two contrasts that contain interactions of three or more factors.
	2	y_2, y_3	Results obtained from one system of two equations	The two contrasts that contain interactions of three or more factors.
2^{7-3}	1	Any	From equaling to zero the null contrast	The only contrast that contains interactions of three or more factors.

*In the standard order of the design matrix

4.4 Application examples

To illustrate how to use the proposed method and the results it provides, we will use examples from Box, Hunter and Hunter [2005].

First, let us consider the 2^3 design in the pilot plant example (p. 177) used to introduce factorial designs. The three factors are: Temperature (T), Concentration (C) and type of catalyst (K). The response (Y) is the performance obtained from a chemical reaction. Table 4.6 shows the contrast matrix and the response.

Table 4.6: Response and contrast matrix for the Pilot Plant Example.

Run	T	C	K	TC	TK	CK	TCK	y
1	−	−	−	+	+	+	−	60
2	+	−	−	−	−	+	+	72
3	−	+	−	−	+	−	+	54
4	+	+	−	+	−	−	−	68
5	−	−	+	+	−	−	+	52
6	+	−	+	−	+	−	−	83
7	−	+	+	−	−	+	−	45
8	+	+	+	+	+	+	+	80

According to Table 4.5 if we want to skip one run it does not matter which one we omit. Let us suppose that run number 1 has not been carried out. Considering that the value of the TCK interaction is negligible, we can equate to zero its contrast and estimate $\hat{y}_1 = 62$. Then we can calculate the 7 effects of interest and represent them in in NPP. We obtain the graph in Figure 4.7 (right), which is compared with what is obtained when all the results are available (left).

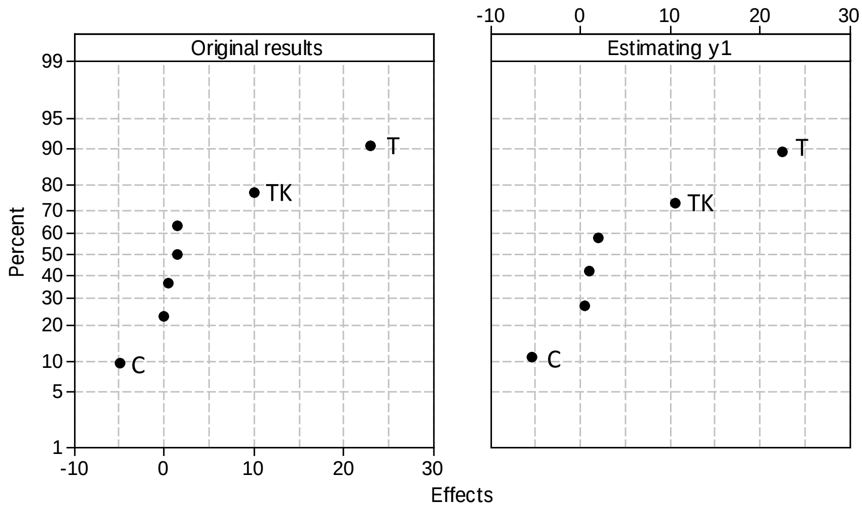

Figure 4.7: Pilot plant example. Analysis of the significance of effects with the original results and the estimated value of y_1.

When estimating a response by equating an interaction to zero, this interaction is *known* and therefore is not included when analyzing the significance of the effects. Having one point less can at times make it more difficult to analyze statistical significance.

This problem does not occur in designs with 16 experiments since –even after eliminating the interactions equated to zero– the number of points remaining is sufficient for discriminating significant from non-significant effects.

We will see this occur in the 2^4 design presented on page 200 of the same book. The example presents a process development experiment in which the factors are: amount of catalyst (A), temperature (B), pressure (C), and concentration of one reactant (D). The response (Y) is the percent conversion. Table 4.7 shows the contrast matrix and the response.

Table 4.7: Response and contrast matrix for the Process Development Experiment.

Run	A	B	C	D	ABC	ABD	ACD	BCD	ABCD	Y
1	−	−	−	−	−	−	−	−	+	70
2	+	−	−	−	+	+	+	−	−	60
3	−	+	−	−	+	+	−	+	−	89
4	+	+	−	−	−	−	+	+	+	81
5	−	−	+	−	+	−	+	+	−	69
6	+	−	+	−	−	+	−	+	+	62
7	−	+	+	−	−	+	+	−	+	88
8	+	+	+	−	+	−	−	−	−	81
9	−	−	−	+	−	+	+	+	−	60
10	+	−	−	+	+	−	−	+	+	49
11	−	+	−	+	+	−	+	−	+	88
12	+	+	−	+	−	+	−	−	−	82
13	−	−	+	+	+	+	−	−	+	60
14	+	−	+	+	−	−	+	−	−	52
15	−	+	+	+	−	−	−	+	−	86
16	+	+	+	+	+	+	+	+	+	79

As before, if we want to save resources by not performing one experiment, it does not matter which one we omit. Again, let us suppose that the one omitted is the first one (y_1). We can estimate its value from the expressions of the negligible interactions with the following results:

Equation	Interaction used in each equation	Estimated value for y_1:
1	ABC	64
2	ABD	74
3	ACD	68
4	BCD	64
5	ABCD	72
	Mean	68.4

Following the recommendations of Table 4.5, if we want to omit two experiments, we will choose runs 6 and 12 and will estimate y_6 and y_{12} from two systems of equations using ABC and BCD interactions for the first system and ABC and ABCD for the second (see Table 4.9 for more possibilities). The responses will be estimated by the mean of the results obtained in each system.

System of 2 equations	Interactions used in each system	Estimated values y_6	y_{12}
1	ABC, BCD	62	76
2	ABD, ABCD	61	79
	Mean	61.5	77.5

If we omit three experiments, according also to the recommendations of Table 4.5, we will choose not to perform runs 4, 6 and 10 and estimate them from four systems of three equations, each one using the indicated interactions (see Table 4.5 for more possibilities). The results obtained are:

System of 3 equations	Interactions used in each system	Estimated values		
		y_4	y_6	y_{10}
1	ABD, ACD, ABCD	83	61	50
2	ABC, ACD, ABCD	83	58	53
3	ABC, ABD, ABCD	80	61	53
4	ABC, ABD, ACD	80	58	50
	Mean	81.5	59.5	51.5

If we are willing to have 4 missing values, Table 4.5 tells us to skip runs 2, 7, 9 and 16 and to estimate their responses using a single system of four equations (see Table 4.5 for more possibilities). In the example, the estimated values are:

System equations	Interactions used	Estimated values			
		y_2	y_7	y_9	y_{16}
1	ABC, ABD, BCD, ABCD	58.5	84.5	58.5	81.5

Finally, if we want to skip 5 runs, Table 4.5 recommends skipping runs 2, 3, 5, 9 and 16 (more options in Table 4.5). In this case the estimates are:

System equations	Interactions used	Estimated values				
		y_2	y_3	y_5	y_9	y_{16}
1	ABC, ABD, ACD, BCD, ABCD	58.33	89.33	72.33	58.33	81.33

Since the selection of runs omitted as well as their estimation has been done in accordance with the recommendations of Table 4.5, all the effects of interest have the same variance; thus, their significance can be assessed by a NPP. Figure 4.8 shows the NPP representation of the original results from the Process Development Experiment (upper left panel) together with the representation of the effects when 1, 2, 3, 4 and 5 runs have been omitted and the responses have been estimated by the proposed method.

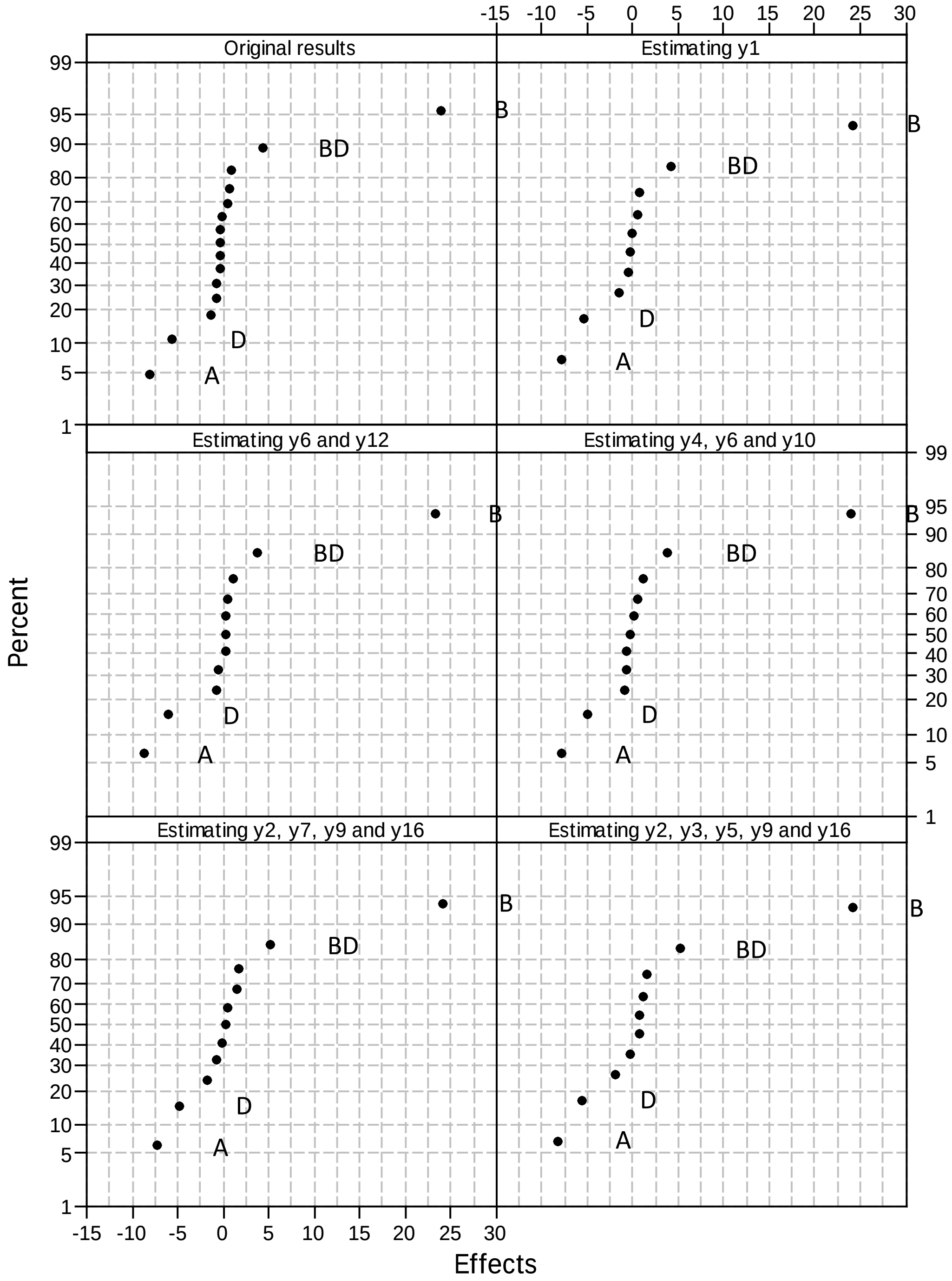

Figure 4.8: Analysis of the effects obtained in example of Box, Hunter and Hunter (2005, pg. 200) with the original results and 1 to 5 estimated results.

When the 5 negligible interactions have been used, the graph shows only the 10 points that correspond to the 10 effects of interest (4 main effects and 6 two-factor interactions). When only four negligible interactions are used to estimate the missing values, the value of the 5th interaction is also calculated and represented, because this helps to distinguish the non-significant effects. In this particular case the plots are very similar independently of the number of runs estimated. In general, estimating 1 or 2 runs does not affect the plot but as the as the number of estimated runs increases the difficulty of estimating their significance increases.

4.5 Summary and conclusions

In two level factorial designs it is possible to save as many runs as there are contrasts that can be considered negligible a priori. Omitting experiments may be an interesting option for saving resources, but it has undesirable consequences: it increases the variability of the estimated effects and provokes the appearance of correlations among them. These undesirable consequences can be minimized by adequately choosing which runs to omit and using an appropriate method to estimate the skipped runs.

The problem can be tackled in different ways, one of which is using D-optimal designs. In comparison, the method we propose is simple and easy to understand. Furthermore, it produces estimates of the effects that not only have similar and small variances, but are also as independent as possible from each other. This produces the additional advantage that – once the missing values have been estimated from the expressions of the contrasts – the analysis procedure is the same as if there were no missing runs.

We believe that this approach may be useful for practitioners and experimenters who lack a deep theoretical knowledge of optimal designs and linear models. The appendix provides tables showing which runs to skip and how to estimate them depending on the type of design and the number of missing values.

4.6 References

Almimi, A.A., Kulahci, M., Montgomery, D.C. 2008. Estimation of missing observations in two-level split-plot designs. *Quality and Reliability Engineering International* 24:127–152.

Box, G.E.P. 1990. A simple way to deal with missings observations from DOE. *Quality Engineering* 3(2):249-254.

Box, G.E.P., Hunter, J.S., Hunter, W.G. 2005. *Statistics for experimenters: design, innovation, and discovery*. Hoboken: John Wiley & Sons.

Draper, N., Stoneman, D. 1964. Estimating missing values in unreplicated two-level factorial and fractional factorial designs. *Biometrics* 20(3):443-458.

Godolphin, J.D. 2006. Reducing the Impact of Missing Values in Factorial Experiments Arranged in Blocks. *Quality and Reliability Engineering International* 22:669-682.

Goh, T.N. 1996. Economical experimentation via 'lean design'. *Quality and Reliability Engineering International* 12:383–388.

Goos, P. and Jones, B. 2011. *Optimal Design of Experiments. A Case Study Approach*. Chichester (UK): John Wiley & Sons.

Jarrett, R.G. 1978. The Analysios of Designed Experiments with Missing Observations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 27, No. 1 (1978) , pp. 38-46.

John, P.W.M. 1979. Missing Points in 2^n and 2^{n-k} Factorial Designs. Technometrics. Vol. 21, No. 2, pp: 225-228.

Johnson, R. T., Montgomery, D. C. and Jones, B.A. 2011. An Expositori Paper on Optimal Design. *Quality Engineering. 23:287-301*

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Srivastava, R., Gupta, V.K., Dey, A. 1991. Robustness of some designs against missing data. *Journal of Applied Statistics*. Vol. 18, No. 3.

Robinson, T.J. and Anderson-Cook. 2011. A Closer Look ar D-Optimality for Screening Designs. *Quality Engineering. 23:1-14.*

Wilkinson, G.N. (1958a). Estimation of missing values for the analysis of incomplete data. *Biometrics*, 14, 257-286.

--------- (1958b). The analysis of variance and derivation of standard errors for incomplete data. *Biometrics*, 14, 360-384.

Xampeny, R., Grima, P. and Tort-Martorell, X. 2017. Estimating missing values from negligible interactions in factorial designs. *Quality and Reliability Engineering International*. 33(6):1235-1247.

Zhou, M., Goh, T.N. 2016. Iterative Designed Experiment Analysis (IDEA). *Quality and Reliability Engineering International*. 32:2977-2986.

4.7 Appendix: Complete list of possible runs to save in 2^4 Designs

Table 4.8 provides a guide for selecting the missing values in 2^4, 2^{6-2} and 2^{7-3} designs. Combined with Tables 4.9 to 4.12, this selection allows identifying all possible combinations of runs to be omitted that fulfill the desired properties.

Table 4.8: Designs with 16 experiments. Variance of the effects, depending on how the missing values are estimated.

Design	Results to estimate	Selection of missing values	Variance effects ($\times \sigma_y^2$) [Number of cases]	Correlation Coeff. Mean - Maximum	Variance characteristics
2^4	1	Any	0.3 [10]	0.167 – 0.167	All equal.
	2	Pairs that can be estimated with up to 6 systems of 2 equations (Table 4.9)	0.375 [6], 0.3333 [4]	0.144 – 0.333	Minimum mean value.
		Pairs that can be estimated with 2 systems of equations using 4 null interactions (Table 4.9)	0.375 [10]	0.152 – 0.333	Mean value only slightly higher than the previous case. All equal.
	3	Trios that can be estimated with 7 systems of 3 equations (Table 4.10)	0.4286 [9], 0.3571 [1]	0.202 – 0.417	Minimum mean value.
		Trios that can be estimated with 4 systems of equations using only four interactions (Table 4.10)	0.4375 [10]	0.195 – 0.429	Mean value only slightly higher than the previous case. All equal.
	4	Subset of the quartets that can be estimated with 2 systems of equations (Table 4.11)	0.5 [9], 0.375 [1]	0.100 – 0.5	Minimum mean value.
		Quartets that can be estimated with 5 systems of equations	0.5 [6], 0.48 (4)	0.175 – 0.5	Mean value only slightly higher than the previous case and with fewer differences among them.
		Subset of the quartets that can be estimated using only four null interactions (Table 4.11)	0.5 [10]	0.091 -0.5	Mean value only slightly higher than the previous case. All equal.
	5	Subset of quintets that can be estimated with a system of 5 equations (Table 4.12)	0.5555 (10)	0.183 – 0.350	Lowest mean value. All equal.
2^{6-2}	1	Any	0.375 [10]	0.333 – 0.333	All equal.
	2	Any of the 64 pairs of missing values that can be estimated with two null interactions.	0,5 [10]	0.231 – 0.5	All equal.
2^{7-3}	1	Any	0.5 [10]	0.5 – 0.5	All equal.

Table 4.9 has the list of 80 missing values whose estimates allow obtaining the effects with the minimum mean variance by means of the six systems of equations indicated in the first row. Using any of the system pairs in the second row (4 of the possible 6 are indicated), a slightly higher variance is obtained, but it is the same for all the effects. The values of the responses correspond to the standard order of the design matrix, and the values of the interactions are those indicated in Table 4.2.

Table 4.9: Saving two runs. List of the recommended 80 runs to save and interactions to use for estimating them.

Missing respon.	Pairs of Interactions	Missing respon.	Pairs of Interactions	Missing respon.	Pairs of Interactions	Missing respon.	Pairs of Interactions
1,4	2,3,5,6,9,10 2,6; 2,10; 3,5; 3,9	3,5	1,2,6,7,8,9 1,8; 1,9; 2,6; 2,7	5,10	2,4,5,7,8,10 2,7; 2,10; 4,5; 4,8	8,12	2,3,4,5,6,7 2,6; 2,7; 3,5; 3,7
1,6	1,3,5,7,8,10 1,8; 1,10; 3,5; 3,7	3,6	1,2,3,7,9,10 1,9; 1,10; 2,7; 2,10	5,11	3,4,6,7,8,9 3,7; 3,9; 4,6; 4,8	8,13	1,4,5,6,9,10 1,9; 1,10; 4,5; 4,6
1,7	1,2,6,7,8,9 1,8; 1,9; 2,6; 2,7	3,8	1,3,5,7,8,10 1,8; 1,10; 3,5; 3,7	5,14	1,2,4,6,8,10 1,8; 1,10; 2,6; 2,10	8,14	1,3,4,5,8,9 1,8; 1,9; 3,5; 3,9
1,8	1,2,3,7,9,10 1,9; 1,10; 2,7; 2,10	3,9	1,3,4,5,8,9 1,8; 1,9; 3,5; 3,9	5,15	1,3,4,5,8,9 1,8; 1,9; 3,5; 3,9	8,15	1,2,4,6,8,10 1,8; 1,10; 2,6; 2,10
1,10	1,2,4,6,8,10 1,8; 1,10; 2,6; 2,10	3,10	1,4,5,6,9,10 1,9; 1,10; 4,5; 4,6	5,16	1,4,5,6,9,10 1,9; 1,10; 4,5; 4,6	9,12	2,3,5,6,9,10 2,6; 2,10; 3,5; 3,9
1,11	1,3,4,5,8,9 1,8; 1,9; 3,5; 3,9	3,12	1,2,4,6,8,10 1,8; 1,10; 2,6; 2,10	6,7	2,3,5,6,9,10 2,6; 2,10; 3,5; 3,9	9,14	1,3,5,7,8,10 1,8; 1,10; 3,5; 3,7
1,12	1,4,5,6,9,10 1,9; 1,10; 4,5; 4,6	3,13	3,4,6,7,8,9 3,7; 3,9; 4,6; 4,8	6,9	2,4,5,7,8,10 2,7; 2,10; 4,5; 4,8	9,15	1,2,6,7,8,9 1,8; 1,9; 2,6; 2,7
1,13	2,3,4,5,6,7 2,6; 2,7; 3,5; 3,7	3,15	2,3,4,5,6,7 2,6; 2,7; 3,5; 3,7	6,10	2,3,4,5,6,7 2,6; 2,7; 3,5; 3,7	9,16	1,2,3,7,9,10 1,9; 1,10; 2,7; 2,10
1,14	2,4,5,7,8,10 2,7; 2,10; 4,5; 4,8	3,16	2,4,5,7,8,10 2,7; 2,10; 4,5; 4,8	6,12	3,4,6,7,8,9 3,7; 3,9; 4,6; 4,8	10,11	2,3,5,6,9,10 2,6; 2,10; 3,5; 3,9
1,15	3,4,6,7,8,9 3,7; 3,9; 4,6; 4,8	4,5	1,2,3,7,9,10 1,9; 1,10; 2,7; 2,10	6,13	1,2,4,6,8,10 1,8; 1,10; 2,6; 2,10	10,13	1,3,5,7,8,10 1,8; 1,10; 3,5; 3,7
2,3	2,3,5,6,9,10 2,6; 2,10; 3,5; 3,9	4,6	1,2,6,7,8,9 1,8; 1,9; 2,6; 2,7	6,15	1,4,5,6,9,10 1,9; 1,10; 4,5; 4,6	10,15	1,2,3,7,9,10 1,9; 1,10; 2,7; 2,10
2,5	1,3,5,7,8,10 1,8; 1,10; 3,5; 3,7	4,7	1,3,5,7,8,10 1,8; 1,10; 3,5; 3,7	6,16	1,3,4,5,8,9 1,8; 1,9; 3,5; 3,9	10,16	1,2,6,7,8,9 1,8; 1,9; 2,6; 2,7
2,7	1,2,3,7,9,10 1,9; 1,10; 2,7; 2,10	4,9	1,4,5,6,9,10 1,9; 1,10; 4,5; 4,6	7,9	3,4,6,7,8,9 3,7; 3,9; 4,6; 4,8	11,13	1,2,6,7,8,9 1,8; 1,9; 2,6; 2,7
2,8	1,2,6,7,8,9 1,8; 1,9; 2,6; 2,7	4,10	1,3,4,5,8,9 1,8; 1,9; 3,5; 3,9	7,11	2,3,4,5,6,7 2,6; 2,7; 3,5; 3,7	11,14	1,2,3,7,9,10 1,9; 1,10; 2,7; 2,10
2,9	1,2,4,6,8,10 1,8; 1,10; 2,6; 2,10	4,11	1,2,4,6,8,10 1,8; 1,10; 2,6; 2,10	7,12	2,4,5,7,8,10 2,7; 2,10; 4,5; 4,8	11,16	1,3,5,7,8,10 1,8; 1,10; 3,5; 3,7
2,11	1,4,5,6,9,10 1,9; 1,10; 4,5; 4,6	4,14	3,4,6,7,8,9 3,7; 3,9; 4,6; 4,8	7,13	1,3,4,5,8,9 1,8; 1,9; 3,5; 3,9	12,13	1,2,3,7,9,10 1,9; 1,10; 2,7; 2,10
2,12	1,3,4,5,8,9 1,8; 1,9; 3,5; 3,9	4,15	2,4,5,7,8,10 2,7; 2,10; 4,5; 4,8	7,14	1,4,5,6,9,10 1,9; 1,10; 4,5; 4,6	12,14	1,2,6,7,8,9 1,8; 1,9; 2,6; 2,7
2,13	2,4,5,7,8,10 2,7; 2,10; 4,5; 4,8	4,16	2,3,4,5,6,7 2,6; 2,7; 3,5; 3,7	7,16	1,2,4,6,8,10 1,8; 1,10; 2,6; 2,10	12,15	1,3,5,7,8,10 1,8; 1,10; 3,5; 3,7
2,14	2,3,4,5,6,7 2,6; 2,7; 3,5; 3,7	5,8	2,3,5,6,9,10 2,6; 2,10; 3,5; 3,9	8,10	3,4,6,7,8,9 3,7; 3,9; 4,6; 4,8	13,16	2,3,5,6,9,10 2,6; 2,10; 3,5; 3,9
2,16	3,4,6,7,8,9 3,7; 3,9; 4,6; 4,8	5,9	2,3,4,5,6,7 2,6; 2,7; 3,5; 3,7	8,11	2,4,5,7,8,10 2,7; 2,10; 4,5; 4,8	14,15	2,3,5,6,9,10 2,6; 2,10; 3,5; 3,9

Table 4.10 has the list of the recommended 160 trios of missing values (in the standard order of the design matrix) when we want to save 3 runs. The minimum mean value of the effect variance is obtained with the 7 systems of 3 equations that are indicated. With the first 4 – or with the last 3 plus the underlined one that is among the first – a somewhat higher mean variance is obtained, but it is the same for all the interactions of interest. The three systems of equations are codified as follows:

1: ACD, BCD, ABCD	2: ABD, BCD, ABCD	3: ABD, ACD, ABCD	4: ABD, ACD, BCD	5: ABC, BCD, ABCD
6: ABC, ACD, ABCD	7: ABC, ACD, BCD	8: ABC, ABD, ABCD	9: ABC, ABD, BCD	10: ABC, ABD, ACD

Table 4.10: Saving three runs. List of the 160 recommended runs to save and the interactions to use for estimating them.

Missing respon.	Trios of interact.	Missing respon.	Trios of interact.	Missing respon.	Trios of interact.	Missing respon.	Trios of interact.
1,4,6	1,2,3,4,7,9,10	2,5,16	1,2,3,4,5,8,9	3,10,16	1,2,3,4,6,8,10	5,14,15	1,5,6,7,4,9,10
1,4,7	1,2,3,4,7,9,10	2,7,9	1,5,6,7,2,8,9	3,12,13	1,5,6,7,2,8,9	6,7,9	1,2,3,4,5,6,7
1,4,10	1,5,6,7,4,9,10	2,7,11	2,5,8,9,3,6,10	3,12,15	2,5,8,9,4,7,10	6,7,12	1,2,3,4,5,6,7
1,4,11	1,5,6,7,4,9,10	2,7,12	1,5,6,7,3,8,10	3,13,16	1,2,3,4,5,6,7	6,7,13	1,5,6,7,4,9,10
1,4,14	1,2,3,4,5,6,7	2,7,13	1,5,6,7,3,8,10	4,5,9	2,5,8,9,3,6,10	6,7,16	1,5,6,7,4,9,10
1,4,15	1,2,3,4,5,6,7	2,7,14	2,5,8,9,3,6,10	4,5,10	1,5,6,7,3,8,10	6,9,12	1,2,3,4,5,6,7
1,6,7	1,2,3,4,7,9,10	2,7,16	1,5,6,7,2,8,9	4,5,11	1,5,6,7,2,8,9	6,9,15	1,2,3,4,6,8,10
1,6,10	2,5,8,9,4,7,10	2,8,11	1,2,3,4,6,8,10	4,5,14	1,5,6,7,2,8,9	6,9,16	1,5,6,7,3,8,10
1,6,12	1,2,3,4,5,8,9	2,8,12	3,6,8,10,4,7,9	4,5,15	1,5,6,7,3,8,10	6,10,13	2,5,8,9,4,7,10
1,6,13	2,5,8,9,4,7,10	2,8,13	1,2,3,4,6,8,10	4,5,16	2,5,8,9,3,6,10	6,10,15	2,5,8,9,3,6,10
1,6,15	1,2,3,4,5,8,9	2,8,14	3,6,8,10,4,7,9	4,6,7	1,2,3,4,7,9,10	6,10,16	3,6,8,10,4,7,9
1,7,11	3,6,8,10,4,7,9	2,9,12	1,5,6,7,4,9,10	4,6,9	1,2,3,4,6,8,10	6,12,13	1,5,6,7,2,8,9
1,7,12	1,2,3,4,6,8,10	2,9,14	2,5,8,9,4,7,10	4,6,10	3,6,8,10,4,7,9	6,12,15	1,2,3,4,5,8,9
1,7,13	3,6,8,10,4,7,9	2,9,16	1,5,6,7,2,8,9	4,6,15	1,2,3,4,6,8,10	6,13,16	1,5,6,7,4,9,10
1,7,14	1,2,3,4,6,8,10	2,11,13	1,2,3,4,6,8,10	4,6,16	3,6,8,10,4,7,9	7,9,12	1,2,3,4,5,6,7
1,8,10	1,5,6,7,2,8,9	2,11,14	2,5,8,9,3,6,10	4,7,9	1,2,3,4,5,8,9	7,9,14	1,2,3,4,5,8,9
1,8,11	1,5,6,7,3,8,10	2,11,16	1,2,3,4,5,8,9	4,7,11	2,5,8,9,4,7,10	7,9,16	1,5,6,7,2,8,9
1,8,12	2,5,8,9,3,6,10	2,12,13	1,5,6,7,3,8,10	4,7,14	1,2,3,4,5,8,9	7,11,13	3,6,8,10,4,7,9
1,8,13	2,5,8,9,3,6,10	2,12,14	3,6,8,10,4,7,9	4,7,16	2,5,8,9,4,7,10	7,11,14	2,5,8,9,3,6,10
1,8,14	1,5,6,7,3,8,10	2,13,16	1,2,3,4,5,6,7	4,9,14	1,2,3,4,5,8,9	7,11,16	2,5,8,9,4,7,10
1,8,15	1,5,6,7,2,8,9	3,5,8	1,2,3,4,7,9,10	4,9,15	1,2,3,4,6,8,10	7,12,13	1,5,6,7,3,8,10
1,10,11	1,5,6,7,4,9,10	3,5,9	3,6,8,10,4,7,9	4,9,16	2,5,8,9,3,6,10	7,12,14	1,2,3,4,6,8,10
1,10,13	2,5,8,9,4,7,10	3,5,10	1,2,3,4,6,8,10	4,10,11	1,5,6,7,4,9,10	7,13,16	1,5,6,7,4,9,10
1,10,15	1,5,6,7,2,8,9	3,5,15	3,6,8,10,4,7,9	4,10,15	1,5,6,7,3,8,10	8,10,11	1,2,3,4,5,6,7
1,11,13	3,6,8,10,4,7,9	3,5,16	1,2,3,4,6,8,10	4,10,16	3,6,8,10,4,7,9	8,10,13	1,2,3,4,5,8,9
1,11,14	1,5,6,7,3,8,10	3,6,9	1,5,6,7,3,8,10	4,11,14	1,5,6,7,2,8,9	8,10,15	1,5,6,7,2,8,9
1,12,13	2,5,8,9,3,6,10	3,6,10	2,5,8,9,3,6,10	4,11,16	2,5,8,9,4,7,10	8,11,13	1,2,3,4,6,8,10
1,12,14	1,2,3,4,6,8,10	3,6,12	1,5,6,7,2,8,9	4,14,15	1,2,3,4,5,6,7	8,11,14	1,5,6,7,3,8,10
1,12,15	1,2,3,4,5,8,9	3,6,13	1,5,6,7,2,8,9	5,8,10	1,2,3,4,5,6,7	8,12,13	2,5,8,9,3,6,10
1,14,15	1,2,3,4,5,6,7	3,6,15	2,5,8,9,3,6,10	5,8,11	1,2,3,4,5,6,7	8,12,14	3,6,8,10,4,7,9
2,3,5	1,2,3,4,7,9,10	3,6,16	1,5,6,7,3,8,10	5,8,14	1,5,6,7,4,9,10	8,12,15	2,5,8,9,4,7,10
2,3,8	1,2,3,4,7,9,10	3,8,10	1,2,3,4,5,8,9	5,8,15	1,5,6,7,4,9,10	8,14,15	1,5,6,7,4,9,10
2,3,9	1,5,6,7,4,9,10	3,8,12	2,5,8,9,4,7,10	5,9,14	2,5,8,9,4,7,10	9,12,14	1,2,3,4,7,9,10
2,3,12	1,5,6,7,4,9,10	3,8,13	1,2,3,4,5,8,9	5,9,15	3,6,8,10,4,7,9	9,12,15	1,2,3,4,7,9,10
2,3,13	1,2,3,4,5,6,7	3,8,15	2,5,8,9,4,7,10	5,9,16	2,5,8,9,3,6,10	9,14,15	1,2,3,4,7,9,10
2,3,16	1,2,3,4,5,6,7	3,9,12	1,5,6,7,4,9,10	5,10,11	1,2,3,4,5,6,7	10,11,13	1,2,3,4,7,9,10
2,5,8	1,2,3,4,7,9,10	3,9,15	3,6,8,10,4,7,9	5,10,15	1,5,6,7,3,8,10	10,11,16	1,2,3,4,7,9,10
2,5,9	2,5,8,9,4,7,10	3,9,16	1,5,6,7,3,8,10	5,10,16	1,2,3,4,6,8,10	10,13,16	1,2,3,4,7,9,10
2,5,11	1,2,3,4,5,8,9	3,10,13	1,2,3,4,5,8,9	5,11,14	1,5,6,7,2,8,9	11,13,16	1,2,3,4,7,9,10
2,5,14	2,5,8,9,4,7,10	3,10,15	2,5,8,9,3,6,10	5,11,16	1,2,3,4,5,8,9	12,14,15	1,2,3,4,7,9,10

Table 4.11 shows the 40 quartets of missing values that can be estimated with a single system of 4 equations for estimating all effects with the same variance and minimum correlation. Each quartet appears twice since it can be estimated with the same properties through two systems of different equations (e. g., 1,4,6,7 can be estimated with the quartet of interactions 1 and also 5).

Table 4.11: Saving four runs. Missing values that can be estimated using only one system of four equations that allow estimating all the effects with the same variance and minimum correlation

1 ABD, ACD, BCD, ABCD	2 ABC, ACD, BCD, ABCD	3 ABC, ABD, BCD, ABCD	4 ABC, ABD, ACD, ABCD	5 ABC, ABD, ACD, BCD
1,4,6,7	1,4,10,11	1,6,10,13	1,7,11,13	1,4,6,7
1,4,14,15	1,4,14,15	1,6,12,15	1,7,12,14	1,4,10,11
1,6,12,15	1,8,10,15	1,8,10,15	1,8,11,14	1,6,10,13
1,7,12,14	1,8,11,14	1,8,12,13	1,8,12,13	1,7,11,13
2,3,5,8	2,3,9,12	2,5,9,14	2,7,11,14	2,3,5,8
2,3,13,16	2,3,13,16	2,5,11,16	2,7,12,13	2,3,9,12
2,5,11,16	2,7,9,16	2,7,9,16	2,8,11,13	2,5,9,14
2,8,11,13	2,7,12,13	2,7,11,14	2,8,12,14	2,8,12,14
3,5,10,16	3,6,9,16	3,6,10,15	3,5,9,15	3,5,9,15
3,8,10,13	3,6,12,13	3,6,12,13	3,5,10,16	3,8,12,15
4,6,9,15	4,5,10,15	3,8,10,13	3,6,9,16	4,6,10,16
4,7,9,14	4,5,11,14	3,8,12,15	3,6,10,15	4,7,11,16
5,8,10,11	5,8,10,11	4,5,9,16	4,5,9,16	5,8,14,15
6,7,9,12	5,8,14,15	4,5,11,14	4,5,10,15	6,7,13,16
9,12,14,15	6,7,9,12	4,7,9,14	4,6,9,15	9,12,14,15
10,11,13,16	6,7,13,16	4,7,11,16	4,6,10,16	10,11,13,16

Finally, Table 4.12 contains 16 combinations of 5 missing values that can be estimated with a system of 5 equations using 5 unknowns, which allow estimating the effects with minimum mean variance and with the same value for all the effects of interest.

Table 4.12: Saving five runs. List of recommended 16 combinations of 5 missing responses

1,4,6,10,15	1,8,12,14,15	2,7,9,12,14	3,6,10,13,16
1,4,7,11,14	2,3,5,9,16	2,7,11,13,16	4,5,9,14,15
1,6,7,12,13	2,3,8,12,13	3,5,8,10,15	4,5,10,11,16
1,8,10,11,13	2,5,8,11,14	3,6,9,12,15	4,6,7,9,16

Chapter 5

ARTICLE 4

Consequences of using estimated response values from negligible interactions in factorial designs

5

ARTICLE 4

Consequences of using estimated response values from negligible interactions in factorial designs

Rafel Xampeny, Pere Grima, Xavier Tort-Martorell
Department of Statistics and Operations Research
Universitat Politècnica de Catalunya – BarcelonaTech, Spain

Quality and Reliability Engineering International
Submitted under consideration of an indexed JCR Journal: 18 December 2017

ABSTRACT

This article analyzes the increase in the probability of committing type I and type II errors in assessing the significance of the effects when some properly selected runs have not been carried out and their responses have been estimated from the interactions considered null from scratch. This is done by simulating the responses from known models that represent a wide variety of practical situations that the experimenter will encounter; the responses considered to be missing are then estimated and the significance of the effects is assessed. Through comparison with the parameters of the model, the errors are then identified. To assess the significance of the effects when there are missing values, the Box-Meyer method has been used. The conclusions are that 1 missing value in 8 run designs and up to 3 missing values in 16 run designs experiments can be estimated without hardly any notable increase in the probability of error when assessing the significance of the effects.

KEYWORDS: Factorial design, missing values, negligible interactions, Lenth method, significant effects.

5.1 Introduction

In a factorial design, it is possible to estimate as many missing response values as there are interactions that can be considered negligible (Draper and Stoneman [1964], Box [1990]). Take, for example, a 2^3 design with a table of contrasts such as Table 5.1.

Table 5.1: Contrasts and responses for a 2^3design

A	B	C	AB	AC	BC	ABC	Y
-1	-1	-1	1	1	1	-1	y_1
1	-1	-1	-1	-1	1	1	y_2
-1	1	-1	-1	1	-1	1	y_3
1	1	-1	1	-1	-1	-1	y_4
-1	-1	1	1	-1	-1	1	y_5
1	-1	1	-1	1	-1	-1	y_6
-1	1	1	-1	-1	1	-1	y_7
1	1	1	1	1	1	1	y_8

If the ABC interaction is negligible we have:

$$-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8 = 0$$

And from this expression we can deduce any response value depending on the remainder.

This procedure can be very useful when it is not possible to perform all the runs required by the chosen design, but it also has undesired consequences. It is straightforward to see that if σ_y^2 is the variance of the responses obtained from the experimentation, the variance of the estimated response will be $7\sigma_y^2$. We will discuss later how this fact affects the analysis of the significance of the effects.

Another problem with this procedure is that the estimation of missing values is not always possible. For example, if in a 2^3 design there were two missing values and the interactions BC and ABC could be considered negligible, we would have 28 possible pairs of missing values and only the values of 16 of them could be estimated. Table 5.2 shows the contrasts associated with interactions BC and ABC. Their expressions can provide a system of two equations with two unknowns to deduce, for example, the values of y_1 and y_2; however, this cannot be done to deduct y_1 and y_3 since the system of equations is inconsistent.

Table 5.2: Contrasts associated with the BC and ABC interactions in a 2^3 design

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
BC	1	1	-1	-1	-1	-1	1	1
ABC	-1	1	1	-1	1	-1	-1	1

In addition, when there is more than one missing response, the variances of the estimated values depend on which those responses are and also on the interactions used for their estimation. In Xampeny *et al.* [2017], it is shown that if in a 2^4 design the five three or more factors interactions can be considered negligible, there will be 4368 possible quintets of missing responses, of which it is impossible to estimate the values of 1360 of them due to their systems of equations being inconsistent. For the combinations that can be estimated, there are notable differences in the variances of the estimated values, depending on the missing responses. For example, the combination of missing values y_1, y_2, y_3, y_8 and y_{12} is one of the 480 that lead to estimates with maximum variances, namely: $31\sigma_y^2$, $15\sigma_y^2$, $15\sigma_y^2$, $7\sigma_y^2$ and $7\sigma_y^2$, respectively; while the combination y_1, y_4, y_6, y_{10} and y_{15} is one of the 16 that present lower values in the variance of the estimates, precisely $2.56\sigma_y^2$ for all of them. Naturally, the bigger the variance of the estimated values, the bigger the variance of the effects.

An additional problem is that since some values of the response are deduced from others, the effects are correlated among them. For example, if in a 2^3 design we have experimentally obtained the eight values of the response, the main effect of, let us say factor A, will be:

$$A = \frac{1}{4}(-y_1 + y_2 - y_3 + y_4 + y_5 + y_6 - y_7 + y_8)$$

From which we get that the variance of A, $V(A)$ is equal to $\sigma_y^2/2$. However, if we have a missing value, for instance y_1, we have:

$$y_1 = +y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8$$

Then the main effect of A will be:

$$\begin{aligned} A = \frac{1}{4}(- & \ y_2 - y_3 + y_4 - y_5 + y_6 + y_7 - y_8 \\ + & \ y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8) = \\ = \frac{1}{4}(& -2y_3 + 2y_4 - 2y_5 + 2y_6) \end{aligned}$$

From which it follows that in this case $V(A) = \sigma_y^2$, which is double that obtained when all the response values have been obtained experimentally.

Additionally, as said above, when some response values have been estimated, the effects are not independent. Following with the previous example we have:

$$B = \frac{1}{4}(-2y_2 + 2y_4 - 2y_5 + 2y_7)$$

And therefore:

$$\begin{aligned} V(A+B) &= V\left[\frac{1}{4}(-2y_2 - 2y_3 + 4y_4 - 4y_5 + 2y_6 + 2y_7)\right] = \\ &= 3\sigma_y^2 \end{aligned}$$

As $V(A+B) = V(A) + V(B) + 2\text{Cov}(A,B)$ it immediately follows that $\text{Cov}(A,B) = 0.5\sigma_y^2$.

Xampeny *et al.* [2018] provide recommendations on which runs to omit and how to estimate them when not all of them can be done for all two level 8 and 16 runs factorial designs containing contrasts formed only by interactions of three or more factors. When these recommendations, detailed in table 5.3, are followed, the effects are estimated with the following properties: 1) same variance for all of them, 2) minimum increase in variance compared to what would occur without missing values, and 3) minimum value of the correlation between effects.

This approach for saving runs has also disadvantages, and the objective of this article is to quantify them. This is done by simulation: a series of scenarios are proposed (varying the numbers and values of the significant effects) and, in each of them, we compare the number of errors made in the analysis of the importance of the effects when all the runs are available with the number of errors when some runs have been estimated following the recommendations in Table 5.3.

Below are detailed which scenarios these are, the methods followed for the assessing the significance of the effects, the results obtained and, finally, the conclusions that can be drawn from this work.

Table 5.3: Recommended runs to skip for obtaining effects with minimum variance and the same for all of them[4].

Design	Results to estimate	Missing runs recommended	Example of missing runs and interactions used	How to estimate the missing responses
2^3	1	Unimportant	Either ABC	From equating to zero the null interaction.
2^4	1	Unimportant	Either ABC, ABD, ACD, BCD, ABCD	Mean of the 5 values obtained from equating to zero the 5 null interactions.
	2	Pairs that can be estimated with 2 systems of 2 equations using 4 null interactions	y_6, y_{12} First System: ABC, ACD Second System: ABD, BCD	For each missing value: Mean of the two results obtained with two systems of equations
	3	Trios that can be estimated with 4 systems of 3 equations using only 4 interactions	y_1, y_4, y_5 First System: ACD, BCD, ABCD Second System: ABD, BCD, ABCD Third system: ABD, ACD, ABCD Fourth system: ABD, ACD, BCD	For each missing value: Mean of the four results obtained solving four systems of equations
	4	Subset of the quartets that can be estimated using only 4 null interactions	y_1, y_4, y_6, y_7 ABD, ACD, BCD, ABCD	Results obtained from a single system of four equations
	5	Subset of the quintets that can be estimated with a system of 5 equations	$y_1, y_4, y_6, y_{10}, y_{15}$ ABC, ABD, ACD, BCD, ABCD	Results obtained from a single system of five equations
2^{6-2}	1	Unimportant	Anyone The two negligible contrasts	Mean of the two results obtained from each null contrast
	2	Any of the 64 pairs of missing values that can be estimated with two null interactions.	y_1, y_3* The two negligible contrasts	Results obtained from a system of two equations
2^{7-3}	1	Unimportant	Anyone The negligible contrast	From equating the null contrast to zero

* With generators E = ABC and F = BCD

5.2 Simulation scenarios

To study the probabilities of error in the analysis of the significance of the effects, we have proposed a series of scenarios that aim to represent the most common situations that the experimenter can encounter. These scenarios consider that part of the effects are null: that is, that their values belong to a distribution of $N(\mu = 0;\ \sigma_{ef})$. The rest have an average equal to Δ or a multiple of this value. With no loss of generality, $\sigma_{ef} = 1$ is taken and, following the criteria of Ye *et al*. [2001], the values of Δ are called Spacing and they vary from 0.5 to 8 in increments of 0.5.

For 8 run designs, we consider the 4 scenarios that were already used by Fontdecaba *et al*. [2015] to analyze the behavior of Lenth's [1989] method.

$S8_1$: $\mu_1 = \Delta,\ \mu_2 = \cdots = \mu_7 = 0$
$S8_2$: $\mu_1 = \mu_2 = \Delta,\ \mu_3 = \cdots = \mu_7 = 0$
$S8_3$: $\mu_1 = \mu_2 = \mu_3 = \Delta,\ \mu_4 = \cdots = \mu_7 = 0$
$S8_4$: $\mu_1 = \Delta,\ \mu_2 = 2\Delta,\ \mu_3 = 3\Delta,\ \mu_4 = \cdots = \mu_7 = 0$

And for 16 run designs we consider those that were used for the first time by Venter and Steel [1998], then later also by Ye *et al*. [2001] and by Fontdecaba *et al*. [2015]:

$S16_1$: $\mu_1 = \Delta, \mu_2 = \cdots = \mu_{15} = 0,$
$S16_2$: $\mu_1 = \mu_2 = \mu_3 = \Delta,\ \mu_4 = \cdots = \mu_{15} = 0$
$S16_3$: $\mu_1 = \cdots = \mu_5 = \Delta,\ \mu_6 = \cdots = \mu_{15} = 0$
$S16_4$: $\mu_1 = \cdots = \mu_7 = \Delta,\ \mu_8 = \cdots = \mu_{15} = 0$
$S16_5$: $\mu_1 = \Delta, \mu_2 = 2\Delta,\ \mu_3 = 3\Delta,\ \mu_4 = \cdots = \mu_{15} = 0$
$S16_6$: $\mu_1 = \Delta,\ \mu_2 = 2\Delta,\ \mu_3 = 3\Delta,\ \mu_4 = 4\Delta,\ \mu_5 = 5\Delta,\ \mu_6 = \cdots = \mu_{15} = 0,$

From the model provided by each scenario, the factors' effects are obtained by simulation. They are analyzed below to identify those that are considered significant. By comparing the results of this analysis with the coefficients of the model, the errors committed are identified.

For the missing values, we proceed as follows: From the values generated for the effects and an arbitrary value for the mean we calculate the response values. Then, the response values that are considered missing are replaced by their estimates – which are calculated through the established procedure in each case. Finally, we calculate the effects again and analyze their significance.

For example, if the values of the randomly generated effects in scenario $S8_3$ with a Spacing value of $\Delta = 5$ are:

A	B	C	AB	AC	BC	ABC
5.25	-4.32	6.07	-0.50	-0.68	-0.27	1.39

then, by assessing their significance by means of their representation in a Normal Probability Plot (NPP) (Figure 5.1, left), the effects that are truly different from zero (A, B and C) appear as significant and, therefore, in this case no error would be made.

From the values of the effects and with an average equal to 100 (arbitrary value), the following responses are obtained (in the standard order of the design matrix):

i:	1	2	3	4	5	6	7	8
y_i:	95.76	102.22	93.60	96.28	102.81	107.85	97.33	104.15

As in a 2^3design, it does not matter which run we do not perform, we randomly choose one of the response values and consider it missing, for example y_4. Next, by equating to zero the expression of the interaction ABC, we estimate its value and in this case we obtain $\hat{y}_4 = 101.84$. With this estimated response we calculate the effects, and we get:

A	B	C	AB	AC	BC	ABC
6.64	-2.93	4.68	0.89	-0.71	-1.66	0.00

By ignoring the existence of a certain correlation among the effects and excluding the ABC interaction whose equal to zero value has been forced and, therefore, does not represent the variability of the null effects, we have represented these values in NPP (Figure 1, right), and only the effects A and C appear to be significant. Therefore, a type II error is committed, since in reality B is different from zero.

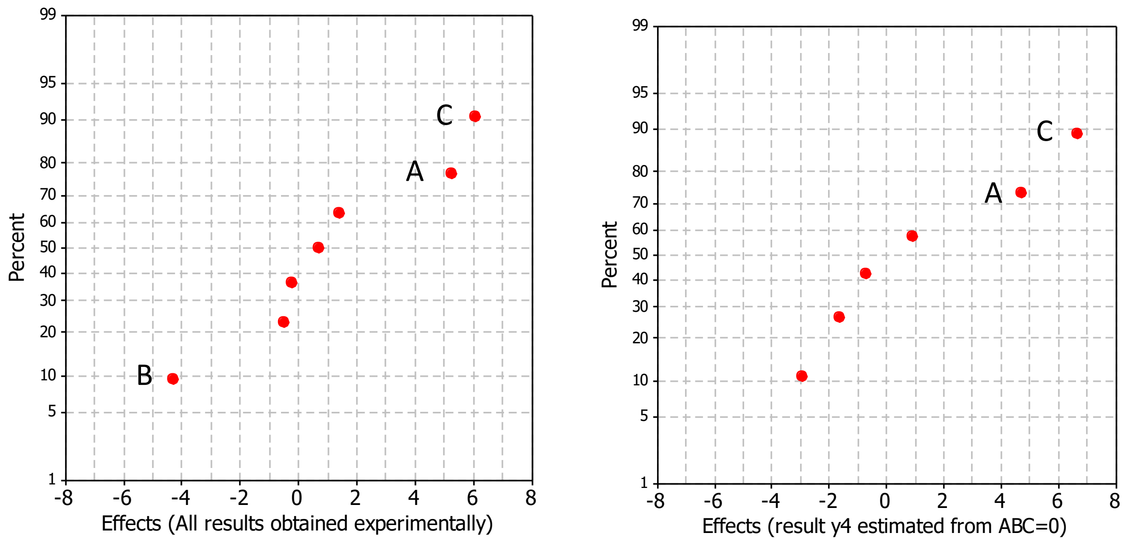

Figure 5.1: Analysis of the significance of the effects in a 2^3design with all the responses obtained experimentally (left) and estimating one of them by equating the ABC interaction to zero.

For each design and for each number of missing runs, we select which ones to skip by following the recommendations in Table 5.3. In each case, and for each scenario and spacing value, 10,000 situations are simulated; and for each one of those, the percentage of type I and type II errors that have been committed are determined. Table 5.4 summarizes the conditions under which the simulations are carried out.

Table 5.4: Summary of the simulations carried out

	Design			
	2^3	2^4	2^{6-2}	2^{7-3}
Num. of missing runs	1	From 1 to 5	1 and 2	1
Runs to skip / Interactions used	Those indicated in Table 5.3			
Scenarios:	From 1 to 4	From 1 to 6	From 1 to 6	From 1 to 6
Simulations:	10,000			

Since in the above example (scenario $S8_3$) there are four null effects, up to 4 type I errors can be made. Therefore, the 10,000 simulations provide opportunities for 40,000 type I errors. On the other hand, having 3 non-null effects there are 30,000 type II error opportunities. After applying the Lenth method with the value of t proposed by Ye and Hamada [2000], and once all the experiments have been carried out, the results indicated in Table 5.5 are obtained. These values will be compared with those obtained when there are estimated response values.

Table 5.5: Error types produced in the 10,000 simulations of the values of the effects in configuration $S8_3$ using the Lenth's Method with Δ = 3.

Type I error		Type II error	
Absolute value	Percentage	Absolute value	Percentage
641	$\frac{641}{40000}100 = 1{,}60$	20928	$\frac{20928}{30000}100 = 69{,}76$

The problem lies in how to assess the significance of the effects automatically in such a way that it can be implemented in the simulation programs. This issue is dealt with in the following section.

5.3 Assessing the significance of the effects

Among the disadvantages involved in using estimated responses, first place is given to the great difficulty in assessing the significance of the effects. When all the runs have been carried out, this task can be done, as many software packages do, either using the variability of effects based on the values of those that can be considered null or by using the method of Lenth. It also can be done manually representing the effects in a Normal Probability Diagram (NPP), a task that requires the analyst's judgment. An analysis of how some well-known statistical software packages address the issue of assessing the significance of the effects can be found in Fontdecaba *et al.* [2014].

Neither of the three methods is appropriate in our case. The judgement by representing the effects in NPP cannot be automated. Nor can we estimate the variance of the effects using those considered null, since they have been used to deduct the missing values. And with respect to Lenth's method is based on if $X \sim N(0,\sigma)$, then the median of $|X|$ is equal to 0.6475 and thus $1.5 \cdot \text{median}|X| = 1{,}01\sigma \cong \sigma$. This circumstance is exploited in order to define $s_0 = 1.5 \cdot \text{median}|c_i|$, where c_i are the values of the effects. Naturally, s_0 is not a good estimator of σ_{ef}, since the values of the active effects also intervene in its calculation. To eliminate them, a new median is calculated by excluding the values $|c_i| > 2.5s_0$. In this way we get what is called the Pseudo Standard Error (PSE), from which is defined an interval of $0 \pm t \cdot PSE$ that contains the effects that are considered inert and where t depends on the confidence level and number of effects being considered. The procedure is very attractive both for its simplicity and for being well-known and commonly used. The above only holds if effects are independent, which never occurs when there are missing values. In addition, if the effects whose values have been forced to zero are excluded, the probabilities of error increase rapidly when considering less than 7 effects. On the other hand, including effects whose values have been forced to zero decreases the PSE, which also leads to major errors.

Hamada and Balakrishnan [1998] discuss and compare a great variety of procedures for assessing the significance of the effects in factorial designs without replicas. From among all of them, we have chosen the Bayesian approach of Box-Meyer, since it is a recognized method that is not restricted to a specific number of effects and does not require independence. In addition, there is an R package that allows it to be applied automatically.

The method of Box and Meyer (1986, 1993) considers the set of all possible models: $M_0, M_1, \cdots, M_m$ that can be contemplated. The value of m is equal to $2^a - 1$, with a being the number of effects that are going to be analyzed. So for example, in a 2^3design with factors A, B and C, we will have $m = 127$, with M_0 being a model that does not include any significant effect until M_{127}, which includes the 7 effects considered: A, B, C, AB, AC, BC and ABC. This requires using the Bayes theorem to determine the probability of each model M_i, given the response vector $\boldsymbol{y}$. In other words:

$$p(M_i|\boldsymbol{y}) = \frac{p(M_i)f(\boldsymbol{y}|M_i)}{\sum_{h=0}^{m} p(M_h)f(\boldsymbol{y}|M_h)}$$

The calculation of $p(M_i)$ is simple. If the total number of effects considered is N, the probability that an effect is active is π, and f_i is the number of active effects in the model M_i, then we have $p(M_i) = \pi^{f_i}(1-\pi)^{N-f_i}$. The value of π must be previously fixed. Box and Meyer propose the value of 0.25 and that is the one we have used.

For calculating $f(\boldsymbol{y}|M_i)$, it is necessary to assign an a priori distribution for the values of the effects. Box and Meyer propose using $N(0,\gamma^2\sigma^2)$, where the mean is 0 due to the direction of each effect being unknown a priori and the magnitude of the effect relative to the experimental noise is captured through the parameter γ. By also following the suggestion of these authors for each case, we have taken the value of γ that minimizes the probability that all effects are null. The expression of $f(\boldsymbol{y}|M_i)$ and the details of deducing it can be seen in the Appendix of the second article of Box and Meyer [1993].

Barrios [2013] has developed the BsMD package for R [2016] that allows determining the probabilities $p(M_i|\boldsymbol{y})$. By introducing the design matrix, the response vector and the values π

and γ, a list of models is obtained in order of their assigned probability. The effects that the model contains are those most likely to be taken as significant.

We have established the reference of what would happen in the case of no missing runs, by using both methods: Box and Meyer and Lenth. There is controversy about which values of t should be used. For a confidence level of 95%, Lenth proposed the values of 3.76 and 2.57 for designs with 8 and 16 experiments, respectively. These values have been discussed by authors such as Loughin [1998], Ye and Hamada [2000], and Fontdecaba *et al*.[2015], all of whom show that a type I error closer to 5% is obtained and that there is a notable decrease in type II errors when using lower values of t. In our study, we used the values proposed by Ye and Hamada: 2.297 and 2.156 for 8 and 16 experiments respectively.

5.4 Results in a 2^3 designs

As an a priori estimate of the proportion of active effects, we have used Box and Meyer's [1993] recommended value of $\pi = 0.25$. When we have a missing response value, forcing an effect to be null leads to think that the proportion of active effects will be greater. However, we have also tested with a value of $\pi = 0.30$, and the results do not improve. Therefore, we have maintained the same value regardless of whether we have all the responses or there is a missing value.

Choosing the value of γ is more complicated. In their first article, Box and Meyer use a different metric that they call k, which is related to γ in the form of $k^2 = n\gamma^2 + 1$, where n is the number of experiments that the design requires. After analyzing a set of cases in this first article, they observe that the values of k vary between 2.7 ($\gamma = 0.89$) and 18 ($\gamma = 6.35$); so they propose using the value of $k = 10$ ($\gamma = 3.52$), because it is a round number that represents approximately the average of the observed values. In their second article they propose choosing the value of γ that minimizes the probability of obtaining a model with all the effects null; and this is the criterion we have used.

To determine those values of γ, we simulated 1000 cases for each Scenario-Spacing combination, which identified for each case the value of γ that minimizes the probability that all effects are null. The value chosen for each Scenario-Spacing combination is the average of the 1000 values obtained. The calculations were made with the help of the `BsSProb` function included in the `BsMD` package of R that calculates the probability associated with each of the models that can be proposed. In each case, probabilities have been evaluated for 20 values of γ that are equidistant within the range of $\gamma = 0.5$ to $\gamma = 10$, which is wider than the one proposed by Box and Meyer in their first article. The values obtained are those we have used in our study, and they are shown in Table 5.6.

Table 5.6: Values of γ used in each Spacing-Scenario combination of values for 2^3 designs.

Spacing	Scenario			
	1	2	3	4
0.5	0.74	0.79	0.76	0.77
1	0.75	0.75	0.77	0.96
1.5	0.81	0.87	0.73	1.33
2	0.89	0.88	0.79	1.67
2.5	1.08	1.09	0.84	2.17
3	1.22	1.24	1.01	2.64
3.5	1.45	1.38	1.15	3.11
4	1.64	1.79	1.28	3.57
4.5	1.79	2.00	1.51	4.24
5	2.02	2.18	1.83	4.56
5.5	2.22	2.44	2.07	5.03
6	2.42	2.72	2.36	5.57
6.5	2.58	2.92	2.82	5.85
7	2.80	3.20	3.02	6.25
7.5	2.96	3.36	3.25	6.64
8	3.21	3.67	3.52	7.04

Instead of previously calculating average values of γ, we could have calculated its value in each case. However, we have verified that the best one obtained is not relevant and doing it in this way greatly extends the computing time, especially when working with 16 experiments in which for each value of γ it is necessary to calculate the probability of the 2^{15} models that can be built.

Figure 5.2 shows the obtained results and also includes – for reference – those of the Lenth method when there are no missing values. The differences are barely noticeable for type I errors and are not relevant for type II errors, especially when using the results of the Lenth method as a reference.

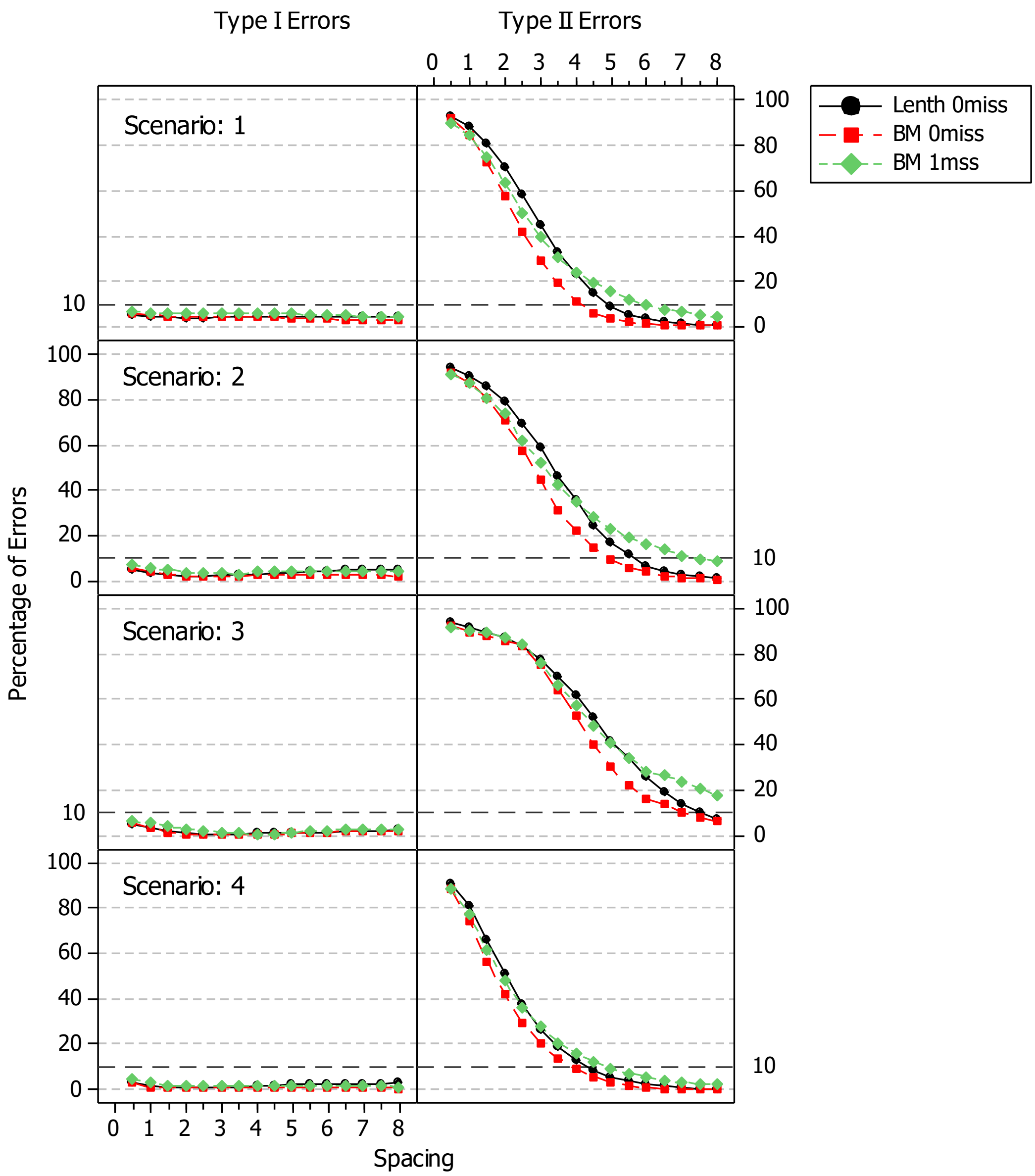

Figure 5.2: 2^3 Designs. Percentage of effects for which a type I and type II error is committed in the analysis of their statistical significance. Without missing values (Lenth and Box-Meyer method) and with one missing value (Box-Meyer).

5.5 Results in 2^4 designs

The same procedure has been applied as for 2^3 designs. The value of $\pi = 0.25$ has also been taken and the values of γ are the average of those obtained by performing 1000 simulations in each Scenario-Spacing combination. To find the value that minimizes the probability that all effects are null in this case, the range of γ values is 0. 5 to 8 (also slightly wider than the one proposed by Box and Meyer). The values obtained for γ are those listed in Table 5.7.

Table 5.7: Values of γ used in each Spacing-Scenario combination of values for 2^4 designs

Spacing	Scenario					
	1	2	3	4	5	6
0.5	0.53	0.54	0.54	0.53	0.54	0.58
1	0.54	0.54	0.53	0.53	0.62	0.86
1.5	0.55	0.55	0.54	0.52	0.86	1.35
2	0.57	0.58	0.55	0.51	1.12	1.80
2.5	0.61	0.65	0.58	0.52	1.37	2.26
3	0.66	0.71	0.63	0.54	1.69	2.66
3.5	0.72	0.87	0.74	0.57	1.98	3.11
4	0.78	0.97	0.91	0.58	2.23	3.62
4.5	0.87	1.14	1.04	0.66	2.59	4.08
5	0.95	1.34	1.26	0.74	2.84	4.58
5.5	1.06	1.45	1.46	0.90	3.12	4.92
6	1.18	1.56	1.67	1.12	3.44	5.44
6.5	1.26	1.69	1.88	1.39	3.67	5.75
7	1.38	1.82	1.98	1.62	4.00	6.11
7.5	1.48	1.97	2.16	1.74	4.32	6.40
8	1.59	2.12	2.28	1.98	4.57	6.74

The results obtained (Figure 5.3) show that the percentage of type I errors increases, in general, when the number of missing values increases. However, it remains at values below 10%, except in the worst case of 5 missing values (Scenario 1), where it rises to around 15%. Regarding the proportion of type II errors, the increase is either not relevant or it even drops, except with 4 missing values, in which case it clearly increases. In scenario 4, a singular behavior occurs in which it remains above 80% even for high Spacing values, especially for 4 missing values.

5.6 Results in other designs

In 2^{6-2} designs with the right generators, for example $E = ABC$ and $F = BCD$, there are two contrasts in which only interactions of 3 or more factors intervene. Therefore, values of 1 or 2 missing values can be estimated. The results obtained with this design are summarized in Figure 5.4. It can be seen that type I errors are maintained at similar values in all scenarios and the same is true for type II errors in scenarios 5 and 6. In scenarios 1-3, there is an increase in the proportion of type II errors when there are missing values, although only for some spacing values. In scenario 4, with spacing values between 3.5 and 5.5, the Box-Meyer method performs poorly both with and without missing values.

For 2^{7-3} designs with a missing value (Figure 5.5) the result is similar to the one we just discussed.

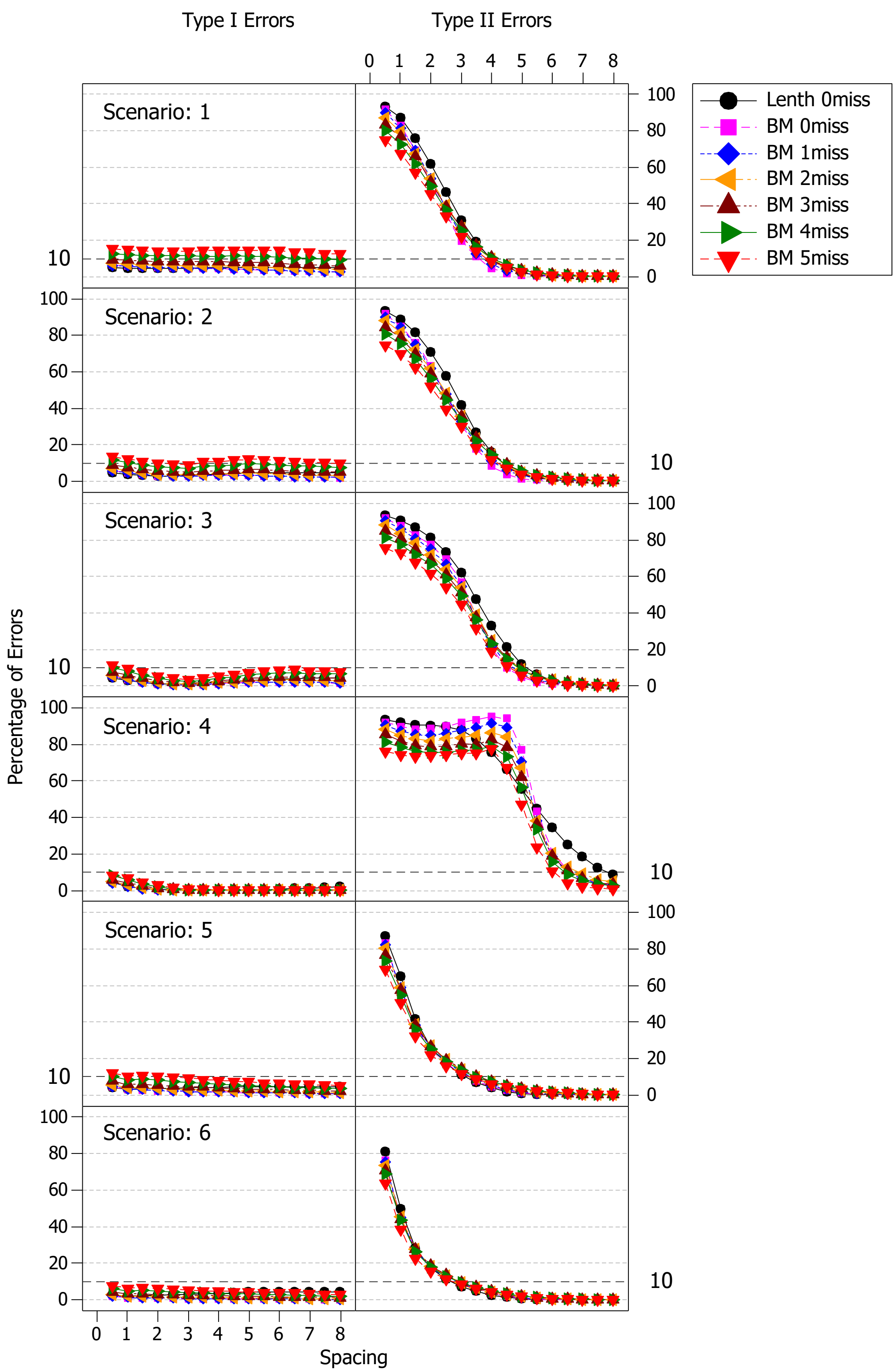

Figure 5.3: 2^4 *Designs. Percentage of effects for which a type I or type II error is committed in the analysis of its statistical significance with and without missing values. Box-Meyer method.*

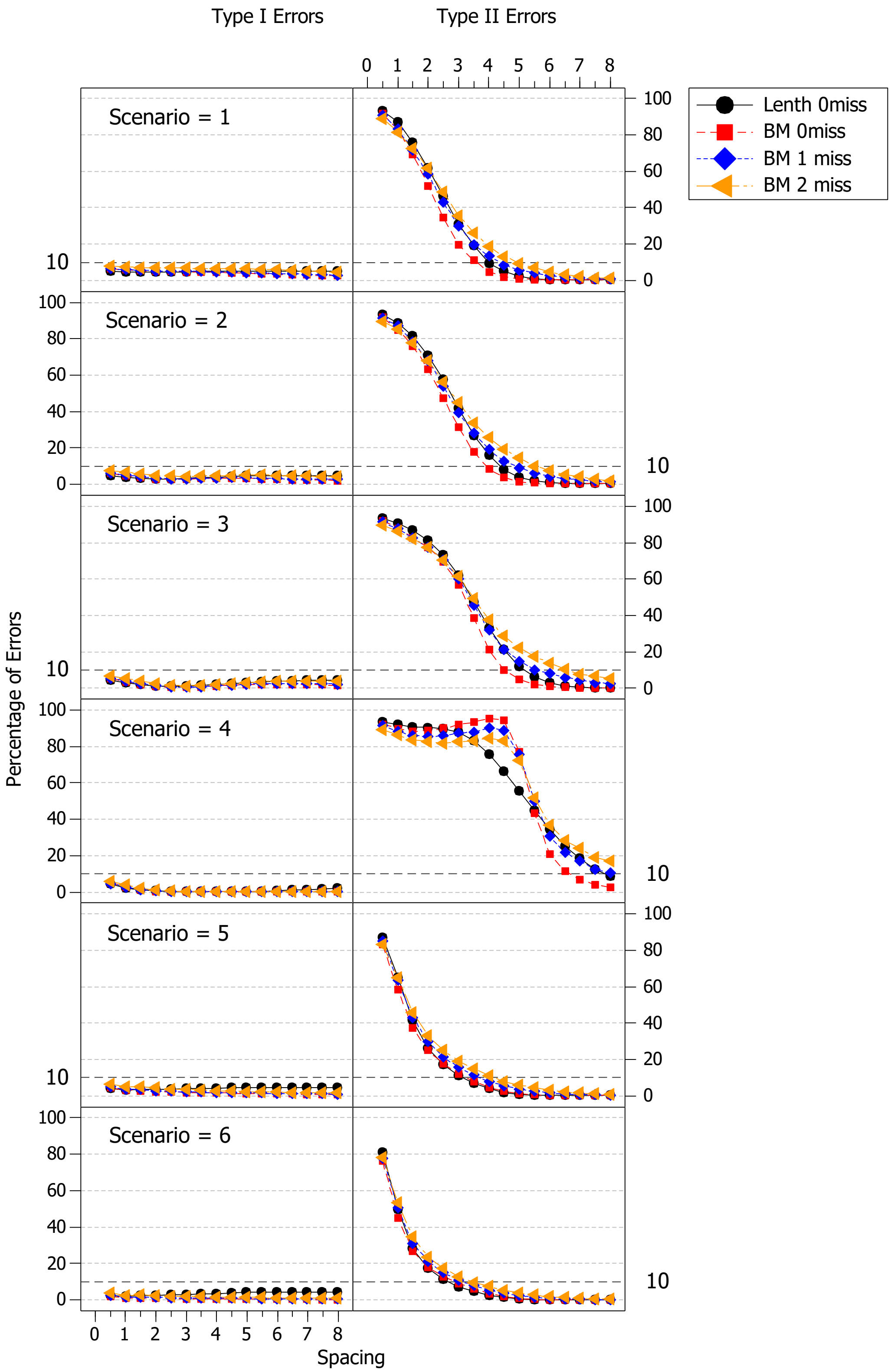

Figure 5.4: 2^{6-2} *Designs. Percentage of effects for which a type I or type II error is committed in the analysis of its statistical significance with and without missing values. Methods of Lenth and of Box-Meyer.*

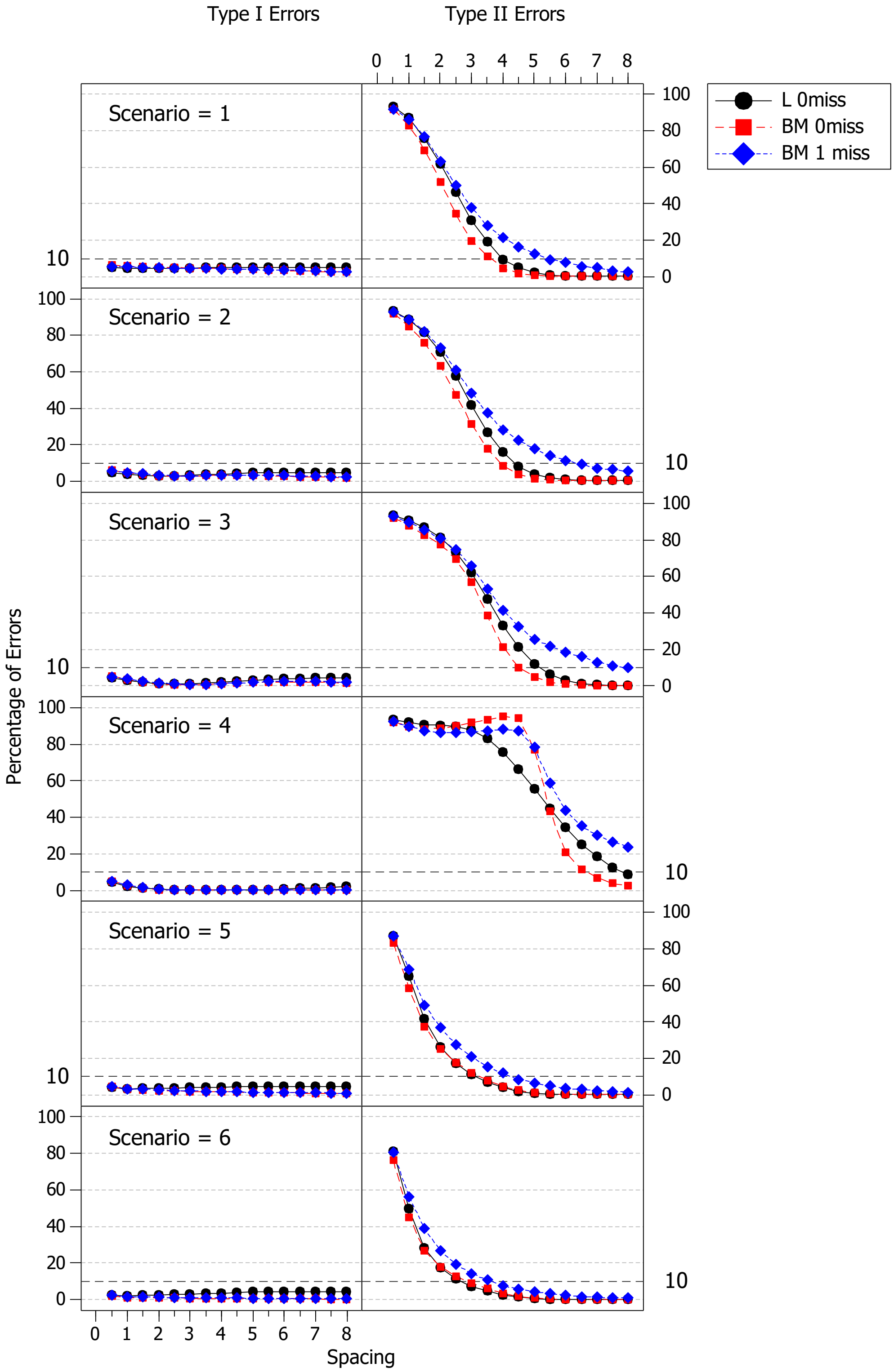

Figure 5.5: 2^{7-3} *Designs. Percentage of effects for which a type I or type II error is committed in the analysis of its statistical significance with and without missing values. Methods of Lenth and of Box-Meyer.*

5.7 Summary and conclusions

We have studied the increase in the probability of committing type I and type II errors in assessing the significance of the effects in 8 and 16 run designs when some properly selected runs have not been carried out and their responses have been estimated from the interactions considered null from scratch.

The only 8 run design with a suitable interaction is the 2^3 design. In it a missing response value can be estimated by clearing its value from the expression of the ABC interaction – which is considered null – equated to zero. The problem that arises is that the variance of the estimated value is greater than that of the values obtained directly from the experimentation; and this in turn causes a greater variance of the effects that, moreover, cease to be independent.

In 16 run designs there are more possibilities. The 2^4 allows to estimate up to 5 missing values since there are 5 interactions of 3 or more factors that can be considered null. In addition, the 2^{6-2} design allows to estimate up to two missing values since it has two contrasts that only estimate interactions of 3 or more factors and the 2^{7-3} design has one suitable contrast and thus allows the estimation of one missing value. In these cases, the variance of the missing values depends on which runs have been skyped as well as which interactions are used and how they are used to perform the estimation. Xampeny *et al*. [2018] have identified which is the best strategy in each case, and that is the one that has been followed in this work.

One consequence of having estimated values is that it complicates the task of assessing the significance of the effects. The degrees of freedom that could be used to estimate the effect variance are used to estimate the missing values and, therefore, this method cannot be used. The conditions for applying Lenth's method are not met either, and therefore using it would lead to important errors. A good possibility that we have used in this paper is the Box-Meyer method.

Another consequence is the greater probability of error when assessing the significance of the effects. By analyzing simulations – in a wide variety of situations – of the proportion of type I and type II errors that have been discussed, our conclusions are:

- Estimating one response value, no matter which one, in 2^3 designs is barely noticeable in terms of the difference in the proportion of type I errors. For type II errors, the difference is slightly bigger but hardly relevant. The analysis also serves to show the good performance of the Box-Meyer method compared to Lenth's. It is interesting to note that the proportion of errors when applying the Lenth method without missing values is approximately the same as when the Box-Meyer method is applied to a 2^3 design with one estimated value.

- In 2^4 designs, working with up to 3 missing values does not produce relevant changes in the proportion of errors, whether they be type I or type II. With 4 and 5 missing values, there is indeed an increase in the proportion of errors – whether they be type I, type II, or both.

- In 2^{6-2} designs when a single missing value is estimated, the results hardly change. When in this same design two missing values are estimated – or one is estimated in a 2^{7-3} design

– the increase in the proportion of errors is indeed noticeable, especially in some scenarios and for certain spacing values.

George Box said [1990] “do not rely on your results if you have too many missing observations. Usually, I would start to feel uncomfortable with the analysis when there was more than one missing observation in an 8-run experiment, or more than two observations missing from a 16-run experiment”. Box refers to situations in which the number of runs has not been planned or there is a result suspected of being anomalous and which one prefers to disregard. Our results are consistent with this statement, and we can add that if one can choose the missing runs, up to 3 runs can be omitted in 16-run designs.

References

Barrios E (based on Daniel Meyer's code). BsMD: Bayes Screening and Model Discrimination. R package version 2013.0718 (2013). https://CRAN.R-project.org/package=BsMD

Box GEP, Meyer RD. An Analysis for Unreplicated Fractional Factorials. *Technometrics* 1986, 28(1):11-18.

Box GEP, Meyer RD. Finding the Active Factor in Fractionated Screening Experiments. *Journal of Quality Technology* 1993, 25(2):94-105.

Box GEP. George’s Column: A simple Way to Deal with Missing Observations from Designed Experiments. *Quality Engineering* 1990; 3(2):249-254.

Draper NR, Stoneman DM. Estimating missing values in unreplicated two-level factorial and fractional designs. *Biometrics* 1964; 20(3):443-458. DOI: 10.2307/2528487.

Fontdecaba S, Grima P, Tort-Martorell X. Analyzing DOE with satistical software sackages: controversies and proposals. *The American Statistician* 2014, **68**(3):205–211.

Fontdecaba S, Grima P, Tort-Martorell X. Proposal of a single critical value for the Lenth method. *Quality Technology and Quantitative Management* 2015, 12(1):41–51.

Hamada M, Balakrishnan N. Analyzing unreplicated factorial experiments: a review with some new proposal, *Statistica Sinica* 1998, 8(1):1–41.

Lenth RV. Quick and easy analysis of unreplicated factorials. *Technometrics* 1989, 31(4):469-473

Loughin TM, Calibration of the length test for unreplicated factorial designs. *Journal of Quality Technology* 1998, **30**(2):171-175.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2016). https://www.R-project.org/

Venter JH, Steel SJ. Identifying active contrasts by stepwise testing. *Technometrics* 1998, **40**(4):304-313

Xampeny R, Grima P, Tort-Martorell X. Estimating missing values from negligible interactions in factorial designs. *Quality and Reliability Engineering International* 2017; **33**(6):1235-1247. DOI: 10.1002/qre.2172.

Xampeny R, Grima P, Tort-Martorell X. Which runs to skip in two level factorial designs when not all can be performed. *Quality Engineering* (2018) DOI: 10.1080/08982112.2018.1428751

Ye KQ, Hamada M. Critical values of the Lenth method for unreplicated factorial designs. *Journal of Quality Technology* 2000, **32**(1):57–66.

Ye KQ, Hamada M, Wu CFJ. A step-down Lenth method for analyzing unreplicated factorial designs.
Journal of Quality Technology 2001; **33**(2), 140-152.

Chapter 6

ARTICLE 5

Selecting significant effects in factorial designs: Lenth's method versus the Box-Meyer approach

6

ARTICLE 5

Selecting significant effects in factorial designs: Lenth's method versus the Box-Meyer approach

Rafel Xampeny, Pere Grima, Xavier Tort-Martorell
Department of Statistics and Operations Research
Universitat Politècnica de Catalunya – BarcelonaTech, Spain

Journal of Applied Statistics
Submitted under consideration of an indexed JCR Journal: 5 March 2018

ABSTRACT

The Lenth method is conceptually simple and probably the most common approach to analyzing the significance of the effects in non-replicated factorial designs. Here, we compare it with a Bayesian approach proposed by Box and Meyer and which does not appear in the usual software packages. The comparison is made by simulating the results of 4, 8 and 16 run designs in a set of scenarios that mirror practical situations and analyzing the results provided by both methods. Although the results depend on the number of runs and the scenario considered, the use of the Box and Meyer method generally produces better results.

KEYWORDS: Factorial design, significant effects, Lenth method, Box-Meyer method, four-run experiments.

6.1 Introduction

Through experimentation, two-level factorial designs provide a great number of possibilities for efficiently analyzing how a set of variables affect a response – particularly in industrial environments. This influence is quantified by calculating the effects, which are orthogonal contrasts of the response vector. Since the effects are affected by random variability – which is inherited from the variability of the response –it is necessary to analyze whether its value is significantly different from zero.

When there are replicas, that is to say, when the experiment has been conducted several times at each experimental condition, we can estimate the experimental error and from it we can get an estimate of the variance of the effects. This estimate can be used to perform significance tests for each effect in the usual way. However, given that the resources for experimentation are usually limited, replicas are typically lacking. In cases it is necessary to analyze the significance of the effects using other methods, which can be graphical or analytical.

Among the graphical methods is the Pareto diagram of effects – where the value of the significant effects is expected to stand out from the rest – and the representation of the effects on a Normal Probability Plot (NPP) (Daniel [1959]). When the effects are represented in NPP, it is expected that the non-significant ones (which belong to a Normal distribution with average $\mu = 0$) will fall on a line that passes through the point $(0, 0.5)$. A variant of NPP is the Half Normal Plot; and in this case the line goes through the point $(0, 0)$.

Representing the effects with NPP is very useful. However, it is not always easy to interpret, especially when there are few effects, as in designs with 8 or fewer runs (a study on the topic can be seen in De León *et al.* [2011]). Furthermore, it cannot be used for making automatic decisions in statistical software packages. The impossibility of automating its use prevents comparison of its effectiveness with other methods and thus the method is not included in this study.

There are many analytical methods for testing the significance of effects in the absence of replicas. Hamada and Balakrishnan [1998] analyze the advantages and disadvantages of a wide selection of them. The one that appears in the most typical textbooks (such as Box, Hunter and Hunter [2005] and Montgomery [2013]) as well as in the most usual statistical software packages for industrial applications (see Fontdecaba *et al.* [2014]) is the Lenth method (Lenth [1989]), which is conceptually simple and provides good results.

Box and Meyer [1986, 1993] published a method using a Bayesian approach. However, due probably to its greater complexity, it did not become widely used and is not among those usually considered when analyzing the significance of effects; nor does it appear as an option in the statistical software packages that are most commonly used by practitioners (Fontdecaba *et al.* [2014]).

In this article, we defend the Box-Meyer method, showing its effectiveness in a wide variety of scenarios that endeavor to represent practical situations. The article is organized as follows. First, the Lenth and Box-Meyer methods are described. Next, we present the situations in which the two methods are compared and the comparison criteria are described. Next, the results obtained are analyzed, showing that the Box-Meyer method performs best in most situations.

6.2 Lenth and Box-Meyer Methods

Lenth's method consists of estimating the standard deviation of the effects based on the fact that if $X \sim N(0, \sigma)$, the median of $|X|$ is equal to 0.645σ and therefore $1.5 \cdot \text{median}|X| = 1.01\sigma \cong \sigma$. Supposing that κ_i $(i = 1, \dots, n)$ are the values of the effects of interest and that their estimators c_i are distributed according to $N(\kappa_i, \sigma_{ef})$, then s_0 Is defined as $1.5 \cdot \text{median}|c_i|$and this value is used to calculate a new median by excluding the estimates of the effects with the value $|c_i| > 2.5s_0$ in order to exclude those with $\kappa > 0$. In this way you get the so-called *Pseudo Standard Error*:

$$PSE = 1.5 \cdot \underset{|c_i|<2.5s_0}{\text{median}} |c_i|$$

From the *PSE* you can calculate a margin of error, *ME*, which, for a confidence level of 95% will be $ME = t_{0.975,\nu} \times PSE$. If $|c_i| > ME$, then the effect $c_i$is considered significant.

Lenth [1989] includes a table with the values of $t_{0.975}$ for designs 2^{k-p} with values of $k-p$ that are understood to be between 3 and 8, that is, designs with between 8 and 256 runs. No examples or references to designs with $k-p=2$ (4 runs) are included; but some software packages also use it in this case (perhaps because the original article does not explicitly discourage its use). On the other hand, Lenth proposes using $\nu = n/3$, with n being the number of effects considered; and this is the value that has been used in some known software packages [6], although it has been shown that it produces type I error probabilities below 5%, which is counterbalanced by higher probabilities of type II error. Ye and Hamada [2000] and Fontdecaba et al. [2015] have proposed values of t that deliver better results (Table 6.1).

Table 6.1: *Proposed values for the value of* $t_{0.975}$ *that must be applied together with the PSE*

Estimated effects	Proposed values for $t_{0.975}$		
	Lenth	Ye and Hamada	Fontdecaba *et al.*
7	3.76	2.297	2
15	2.57	2.156	2

The Box and Meyer method [1986], [1993] considers the set of all possible models that can be proposed: $M_0, M_1, \cdots, M_m$. The value of m is equal to $2^n - 1$, with n being the number of effects that are going to be analyzed. So, for example, in a 2^3 design with factors A, B and C, we will have $m = 127$, with M_0 being a model that does not include any significant effect until M_{127}, which includes the 7 effects considered: A, B, C, AB, AC, BC and ABC. This is to determine – by means of Bayes' theorem – the probability of each model M_i given the response vector $\boldsymbol{y}$, that is to say:

$$p(M_i|\boldsymbol{y}) = \frac{p(M_i)f(\boldsymbol{y}|M_i)}{\sum_{h=0}^{m} p(M_h)f(\boldsymbol{y}|M_h)}$$

Calculating $p(M_i)$ is easy. If the total number of effects considered is N, the probability that an effect is active is π and f_i is the number of active effects in model M_i; then $p(M_i) = \pi^{f_i}(1-\pi)^{N-f_i}$. The value of π must be previously fixed. Box and Meyer use the value of 0.25 in the examples they present.

For the calculation of $f(\boldsymbol{y}|M_i)$, it is necessary to assign an a priori distribution to the effects values. Box and Meyer propose using $N(0, \gamma^2\sigma^2)$. Where the mean is 0 due to the lack of a priori knowledge regarding the direction of each effect, and the parameter γ captures the magnitude of the effect relative to the experimental noise. It is suggested to assign to γ the value that minimizes the probability that all the effects are null. The expression of $f(\boldsymbol{y}|M_i)$ and the details for deducing it can be seen in the Appendix of Box and Meyer's second article [1993].

6.3 Test scenarios

To study the probabilities of error in the effects significance analysis, we have proposed a series of scenarios that try to represent situations that the experimenter can find in practice. These scenarios consider part of the effects to be null, that is, that their values belong to a distribution $N(\mu = 0; \sigma_{ef})$. The rest have an average that is equal to Δ or a multiple of this value. With no loss of generality, $\sigma_{ef} = 1$ is taken and, following the criteria of Ye *et al*. [2001], the values of Δ are designated Spacing and they vary between 0.5 and 8 in increments of 0.5.

We perform simulations for designs with 4, 8 and 16 runs and omit designs with more runs since they are not widely used. What is more, this designs allow estimating a lot of effects, many of which – according to the effect sparsity principle – will be zero. In this circumstance, identifying those that are significant is an easy task with any procedure.

At the opposite end are the designs with 4 runs. Although they are not usually considered in articles that deal with effects significance analysis, it is not unusual to have two factors remaining for study in the last steps of a sequential experimentation process. What is certain is that with only three effects (those obtained from a design with 4 runs) it is difficult to select those that should be considered significant when no information is available on the experimental error. In this circumstances the usual methods are totally ineffective. We have seen practitioners and students surprised to see in the information provided by the software they are using that none of the two factors they are considering have an influence on the response – even though everything indicated that at least one should. Thus, we have included in our analysis the case of designs with only 4 runs. We will see that also in this case and in spite of not presenting extraordinary results the Box-Meyer method improves the Lenth method.

In the three cases, 4, 8 and 16 run designs we propose 6 scenarios.

For the 4 run designs the scenarios cover from the case when the three effects are null up to when all three effects are active. In this last case, the effects can have the same average value, Δ, or average values of Δ, 2Δ, 3Δ (Table 6.2).

Table 6.2: *Effect values in scenarios considered in 4-run designs*

Scenarios	Effects		
	1	2	3
$S4_1$	0	0	0
$S4_2$	**Δ**	0	0
$S4_3$	**Δ**	**Δ**	0
$S4_4$	**Δ**	**2Δ**	0
$S4_5$	**Δ**	**Δ**	**Δ**
$S4_6$	**Δ**	**2Δ**	**3Δ**

For 8 run designs, we first considered scenarios 1, 2, 3 and 5 that were used by Fontdecaba *et al*. [2015] to analyze the performance of Lenth's method. And then we added scenarios 4 and 6, in which there exists the possibility that 4 significant effects also exist (Table 6.3).

Table 6.3: *Effect values in scenarios considered in 8-run designs*

Scenarios	Effects						
	1	2	3	4	5	6	7
$S8_1$	**Δ**	0	0	0	0	0	0
$S8_2$	**Δ**	**Δ**	0	0	0	0	0
$S8_3$	**Δ**	**Δ**	**Δ**	0	0	0	0
$S8_4$	**Δ**	**Δ**	**Δ**	**Δ**	0	0	0*
$S8_5$	**Δ**	**2Δ**	**3Δ**	0	0	0	0
$S8_6$	**Δ**	**2Δ**	**3Δ**	**4Δ**	0	0	0*

For 16 run designs, we use the same scenarios that were used for the first time by Venter and Steel [1998], and later by Ye *et al*. [2001] and Fontdecaba *et al*. [2015]. From 1 to 7 significant effects are considered (Table 6.4):

Table 6.4: *Effect values in scenarios considered in 16-run designs*

Scenarios	Effects														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$S16_1$	**Δ**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$S16_2$	**Δ**	**Δ**	**Δ**	0	0	0	0	0	0	0	0	0	0	0	0
$S16_3$	**Δ**	**Δ**	**Δ**	**Δ**	**Δ**	0	0	0	0	0	0	0	0	0	0
$S16_4$	**Δ**	**Δ**	**Δ**	**Δ**	**Δ**	**Δ**	**Δ**	0	0	0	0	0	0	0	0
$S16_5$	**Δ**	**2Δ**	**3Δ**	0	0	0	0	0	0	0	0	0	0	0	0
$S16_6$	**Δ**	**2Δ**	**3Δ**	**4Δ**	**5Δ**	0	0	0	0	0	0	0	0	0	0

6.4 Simulation

For each scenario, and within each scenario for each Spacing value we have simulated 10,000 situations. Each of them has been analyzed using Lenth's (with $t = 3.76$ and $t = 2.297$) and Box-Meyer's methods.

To apply the Lenth method in designs with 8 or 16 runs, we perform the analysis using values of $t_{0.975}$ (which were proposed in the original article (Lenth [1989])) and also those proposed by Ye and Hamada [2000]. When using this in 4-run designs the value of $t_{0.975}$ with a single degree of freedom ($t = 12.71$) – which would be the one obtained by following Lenth's proposed general rule and which is still used by some statistical software packages – gives very bad results, practically never detects the active effects[1]. We have studied how the probabilities of type I and type II errors vary according to the value of t in the scenarios considered. It is observed (Figure 6.1) that even when dropping down to a value of $t = 2$, the active effects are barely detected. For $t = 2/3$, the type I error ratios are similar to those obtained with the Box-Meyer method, so we present the comparison with the values obtained for this value of t.

When we apply the Box-Meyer method in designs with 8 or 16 runs, we followed the authors' recommendation both for the a priori proportion of significant effects (π = 0.25), and for the estimation of the parameter γ (the value that minimizes the probability of the model having all null effects). In designs with 4 runs, it is reasonable to consider that the 3 effects may be null or may be active; thus, in this case we have taken the value of $\pi = 0.50$. Regarding the value of γ, we have analyzed the type I and type II error proportions in all scenarios and for all Spacing values (Figure 6.2). Some values give good results in some scenarios but bad in others. We have chosen $\gamma = 2$, since this value reasonably balances the goodness of the results in all scenarios.

To determine the probabilities $p(M_i|\boldsymbol{y})$, we have used the BsMD package developed by Barrios [2013] for the statistical software R [2016]. Introducing the design matrix, the response vector and the values for π and γ delivers a list of models that are ordered according to the posterior probability of being correct of each of them. The effects considered significant are those contained in the model with the greatest probability. The package also includes a function to identify the value of γ that minimizes the probability that all effects are null. We have used this value in 8 and 16 run designs.

To illustrate the procedure followed, let us take as an example the results from one of the 10000 simulations performed in scenario $S8_2$ with a Spacing value $\Delta = 3$. The values of the effects are those indicated in Table 6.5 (Effects, c_i). Applying the Lenth method delivers a $PSE = 0.5625$, if we use the value of $t = 3.76$, the effects that present $|c_i| > 2.115$ must be considered significant. In this case effect 1. Since those that are actually active are effects 1 and 2, a type II error is committed because 2 is not considered significant. If we apply the Box-Meyer method, we first determine the value of γ that minimizes $p(M_0|\boldsymbol{y})$, it is $\gamma = 2.5$. Using this value, the model with the greatest a posteriori probability is the one that includes the effects 1, 2 and 4. As only effects 1 and 2 are really active, the Box-Meyer method succeeds in identifying them as such; but it is also mistaken in considering effect number 4 to be significant and thus commits a type I error in this case. Table 6.5 summarizes the results obtained.

[1]R.V. Lenth was aware that his method could not be applied to 4 run designs and in his paper never tries to do that. Unfortunately several statistical packages apply it to all two level designs independently of the number of runs.

Table 6.5: *Results with the values of the effects obtained by simulation for a design with 8 experiments, scenario S82,* $\Delta = 3$.

Effects #	c_i	Actual Fact	Effects significance analyzed by: Lenth Method ($t = 3.76$)		Box and Meyer Method	
1	4.44	Active	SIGNIFICANT	(Correct)	SIGNIFICANT	(Correct)
2	1.75	Active	Not significant	(Type II error)	SIGNIFICANT	(Correct)
3	-0.13	Inert	Not significant	(Correct)	Not significant	(Correct)
4	1.18	Inert	Not significant	(Correct)	SIGNIFICANT	(Type I Error)
5	-0.48	Inert	Not significant	(Correct)	Not significant	(Correct)
6	0.27	Inert	Not significant	(Correct)	Not significant	(Correct)
7	-0.08	Inert	Not significant	(Correct)	Not significant	(Correct)

After performing 10 000 simulations, the errors of each type and for each method are added together and their percentage of all total possibilities is calculated. Thus, in scenario $S8_2$ you can commit up to 50 000 type I errors (in each simulation there are 5 inert effects that, erroneously, can be considered active), and you have 20 000 options for a type II error (in each simulation there are 2 active effects that may not be identified). The results obtained in the case of our example ($S8_2$, $\Delta = 3$) are indicated in Table 6.6.

Table 6.6: *Types of error produced in the 10 000 simulations of the values of the effects in scenario* $S8_2$ *with* $\Delta = 3$

	Type I error Absolute value	Type I error Percentage	Type II error Absolute value	Type II error Percentage
Lenth Method (t=3.76)	356	$\frac{356}{50000}100 = 0.712$	15544	$\frac{15544}{20000}100 = 77.72$
Box-Meyer Method	1619	$\frac{1619}{50000}100 = 3.238$	10090	$\frac{10090}{20000}100 = 50.45$

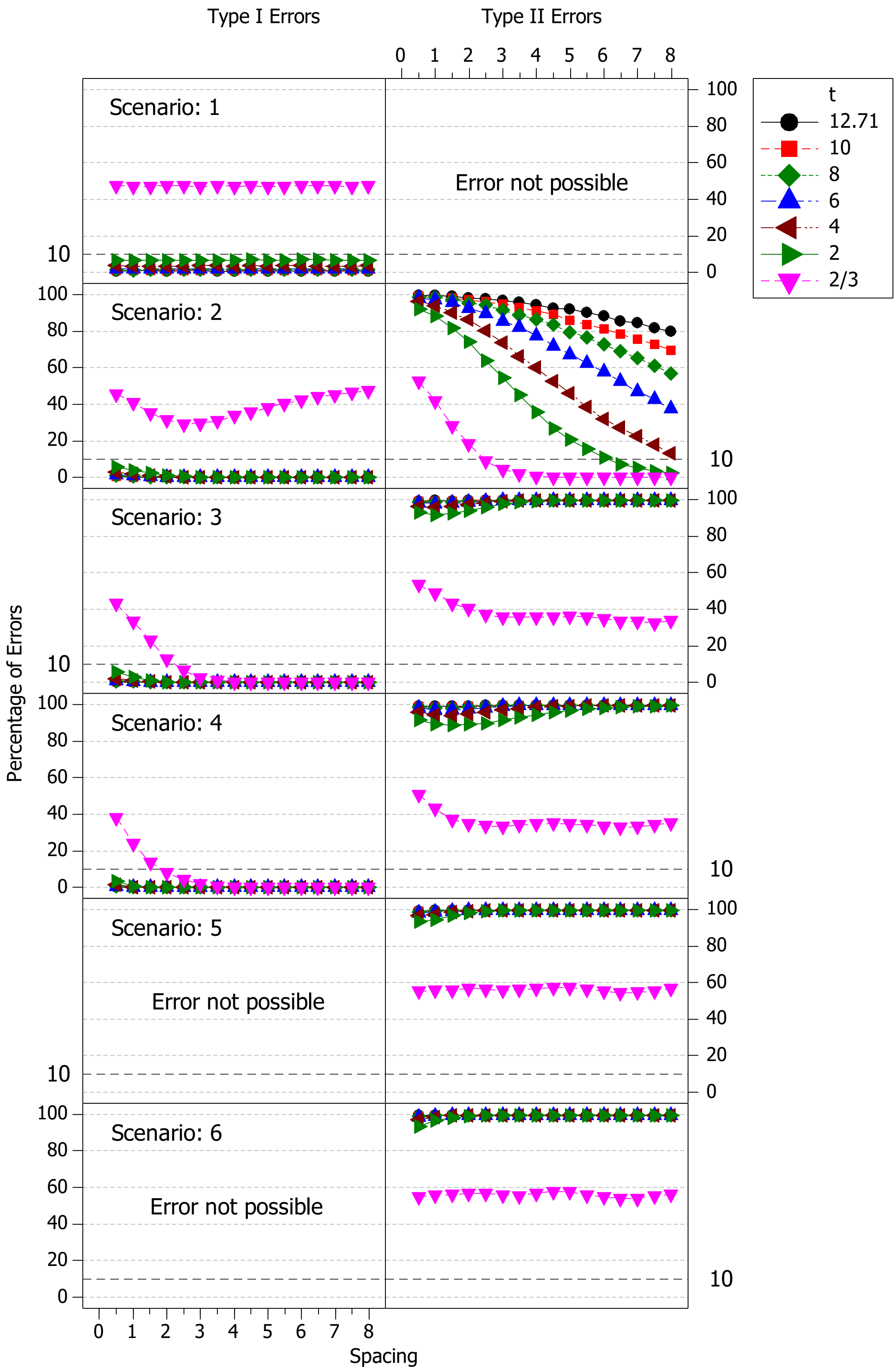

Figure 6.1: Lenth method. Proportion of errors in designs with 4 runs depending on the value of t chosen.

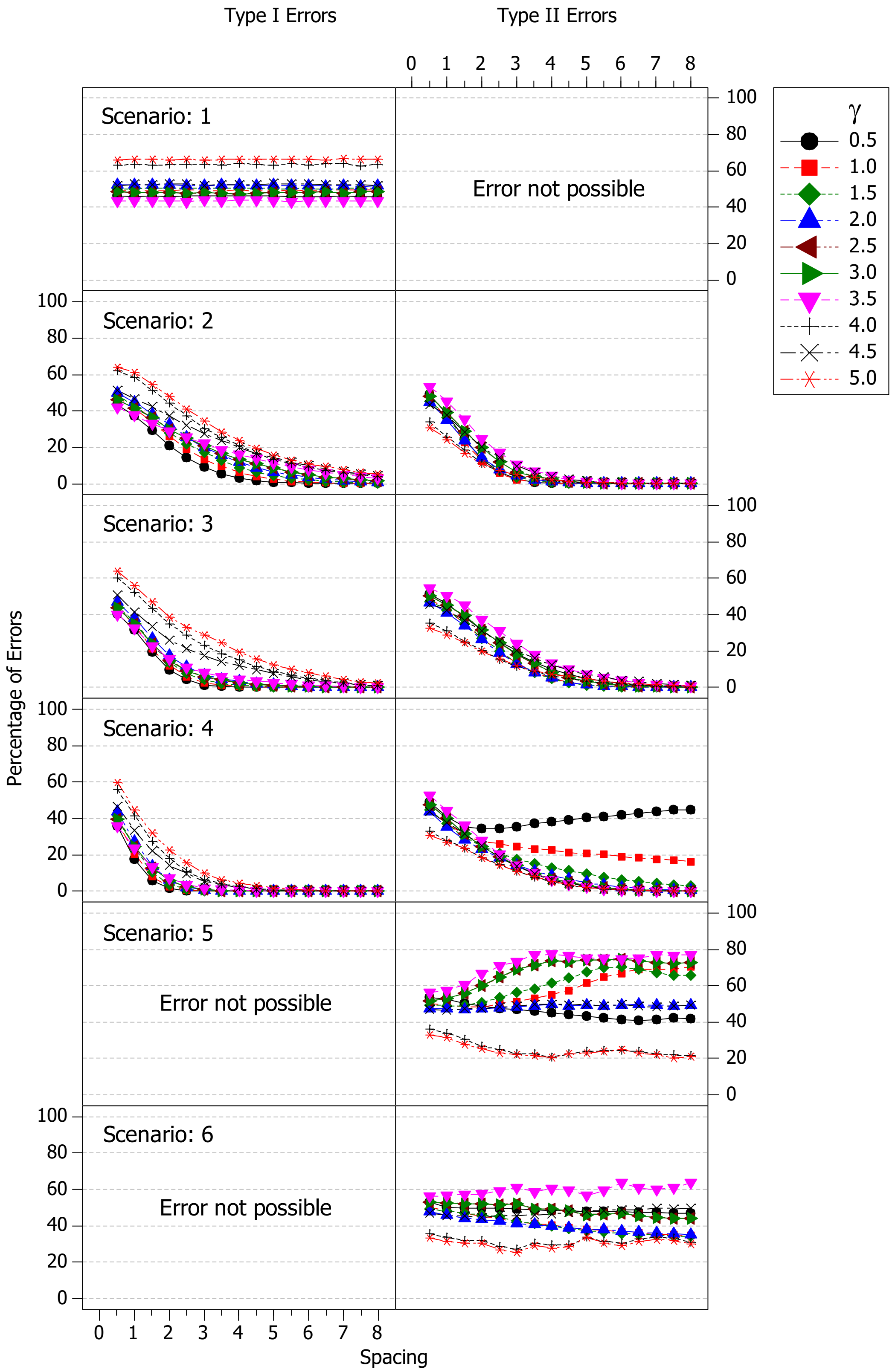

Figure 6.2: Box-Meyer method. Proportion of errors in designs with 4 runs depending on the value of γ

6.5 Results

Figure 6.3 shows the comparison of the results obtained for 4 run designs. In scenario 1 there is no probability of type II error since no effect is active. Nor can there be any type I error in scenarios 5 and 6, since all effects are active. Lenth's method always gives a lower proportion of type I error, but at the expense of systematically ignoring type II errors in all scenarios except for 2. The Box-Meyer method produces a greater proportion of type I errors, especially in scenario 1, but type II errors fall significantly in all scenarios. We cannot say that the Box-Meyer method is excellent in this case, but the results are clearly better than with Lenth's method. In any case, it seems important to us that the experimenter is aware of the shape of these error curves.

For 8 run designs, the results are summarized in Figure 6.4. Regarding type I errors, the differences are small and in all cases reasonable values are presented. Regarding type II errors, the greater probability of error in the 4 scenarios emerges when using the value of $t = 3.76$, as already shown in Fontdecaba *et al.* [2015]. The Box-Meyer method has lower values of type II error in all scenarios and for all Spacing values.

In 16 run designs the results are presented in Figure 6.5. In this case the number of type I errors are also reasonable in all cases. Regarding the proportion of type II errors, the worst performance of the Lenth method occurs with $t = 2.57$, especially in scenarios 1, 2 and 3; and the highest proportion of type II errors with the Box-Meyer method occurs in scenario 4, especially with Spacing values above 3. In this case, the problem lies in having 46.7% of active effects, a value that is far from the $\pi = 0.25$ that is generally assumed. If $\pi = 0.50$ is considered, the results are practically identical to those from the Lenth method.

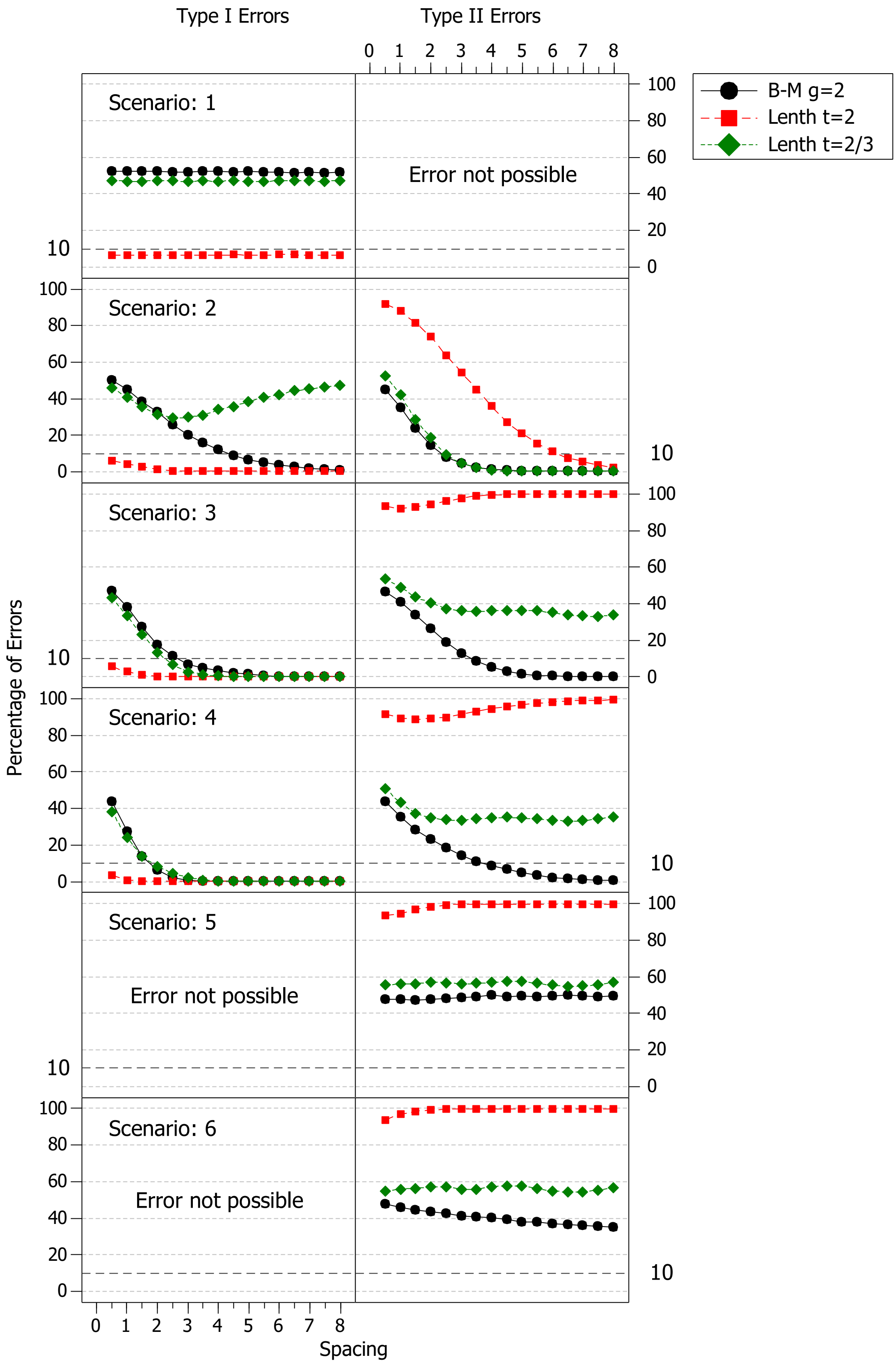

Figure 6.3: Designs with 4 runs. Comparison of the Lenth and Box-Meyer methods

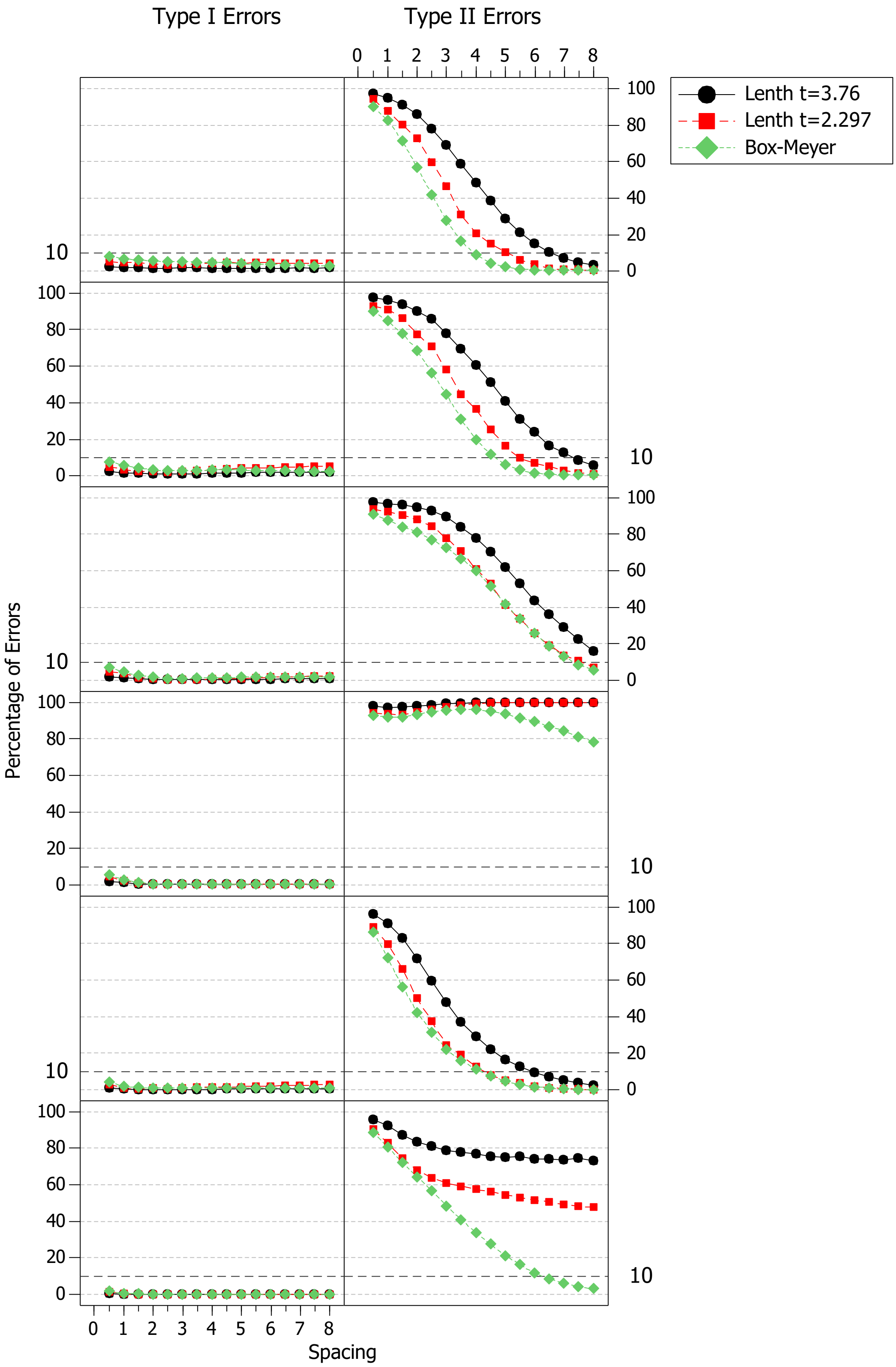

Figure 6.4: Designs with 8 runs. Comparison of the Lenth and Box-Meyer methods. The values obtained in the example of Table 6 have been circled

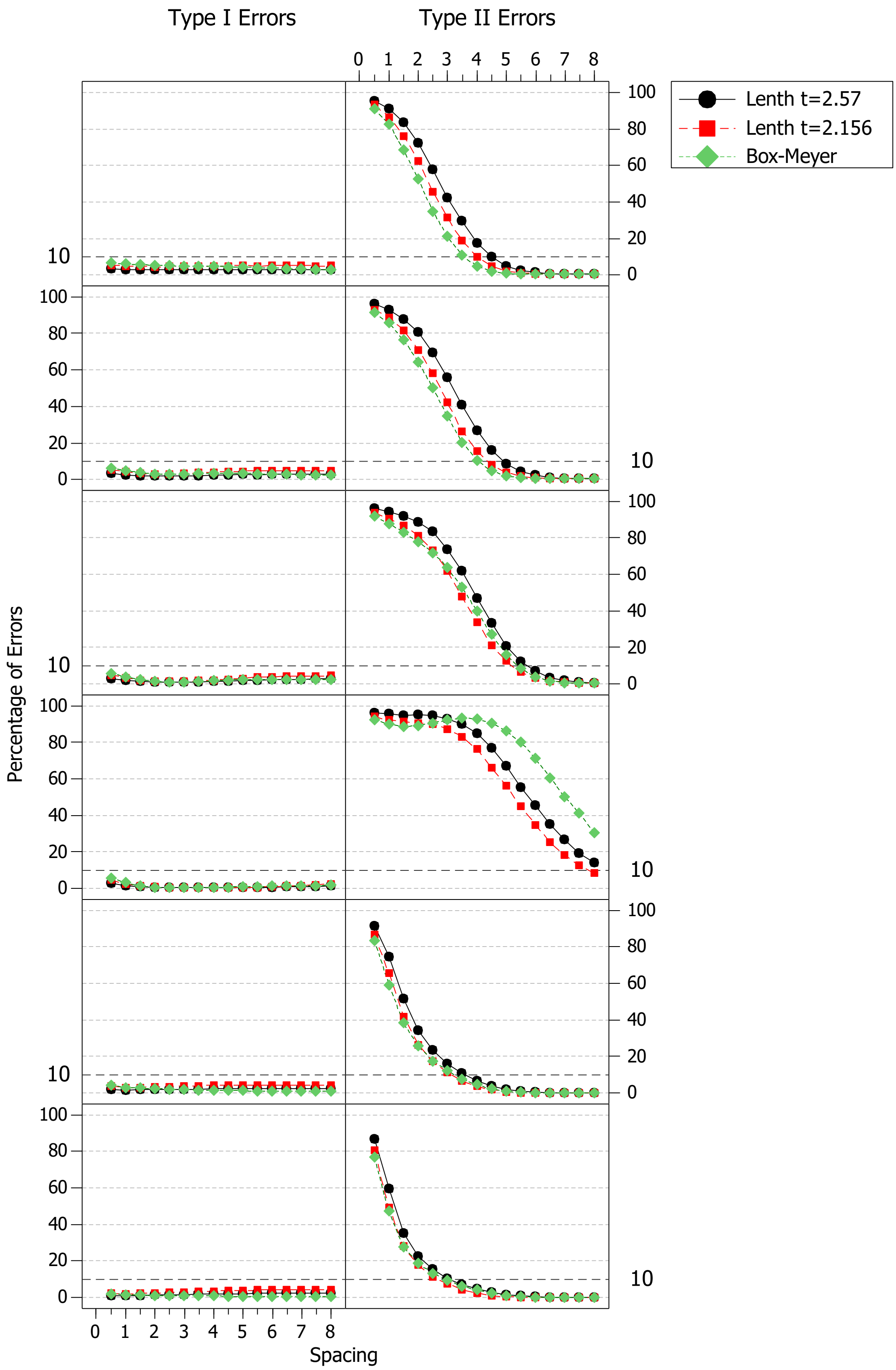

Figure 6.5: Designs with 16 runs. Comparison of the Lenth and Box-Meyer methods

6.6 Summary and recommendations

Our conclusions and recommendations after the thorough comparison between Lenth and Box-Meyer methods are:

- 4 run designs: This is a neglected situation in the literature on factorial designs. In this case, both the representation of effects in NPP and Lenth's method are – by their very nature – ineffective at identifying which effects should be considered significant. With only 3 effects, it is not possible to discriminate “those that separate from the line” when the NPP is used. Also, the Lenth method is not reliable and – especially if the recommended value of t is used – practically in no case does it detect the active effects. Naturally, miracles cannot be expected with only 3 effects, and the Box-Meyer method does not deliver excellent results either, but they are – in all scenarios considered – better than those delivered by the Lenth method, even when using the value of t that favors it more.

  It is important to be aware that if a design with 4 runs is carried out without prior information about the variability of the response, it is not possible to analyze the significance of the effects with reasonable error probabilities. If the experiment is carried out at the end of a process of sequential experimentation, the best option is to estimate the experimental error from the values of the non-significant effects obtained in the previous experiments, and estimate the variance of effects from it.

- 8 run designs: Of the two most usual designs (8 and 16 runs), these are the most difficult to analyze. The smaller number of effects makes it difficult to discriminate between those that are significant and those that are not. In this case the Box-Meyer method performs better than the Lenth method (better than when using the original value of $t = 3.76$, of course, but also when using the $t = 2.297$ value proposed by Ye and Hamada), in all scenarios and for all Spacing values.

- 16 run designs: In this case the differences are barely noticeable, except in scenarios 1 and 2, in which the Box-Meyer method is slightly better (lower proportion of type II errors); but it is slightly worse in scenario 3 and notably worse in scenario 4. In scenario 4 the proportion of active effects is close to 50%, a value that is far from the 25% assumed a priori. In both Scenario 3 and Scenario 4, if a proportion of significant effects is considered at around 50%, the results are similar to those obtained with Lenth's method.

This study clearly shows that the Box-Meyer method gives– in general – better results than the widely adopted Lenth one. Therefore, we strongly advocate for the incorporation of the Box-Meyer method to statistical packages. Having it available as an alternative or even complementary to another method will help the experimenter make better informed decisions.

A last point, worth mentioning, is that the simulation carried out confirms what other authors have already shown (see, for example Ye and Hamada [2000], Fontdecaba *et al.* [2015]), namely: the value of t that appears in the original article on the Lenth method and that is still used in the most widely distributed packages of statistical software (Fontdecaba et al. [7]) produces, on the one hand, a probability of type I errors smaller than the intended 5%;

causing, as a counterpart, a high probability of type II error, that is, it does not consider effects to be active when they actually are. In all the designs and in all the scenarios considered, the value of t proposed by Ye and Hamada produces a type I error probability that is closer to 5% and a lower probability of type II error.

6.7 References

E. Barrios (based on Daniel Meyer's code). BsMD: Bayes Screening and Model Discrimination. R package version 2013.0718, 2013; software available at https://CRAN.R-project.org/package=BsMD

G.E.P. Box and R.D. Meyer, *An analysis for unreplicated fractional factorials*. *Technometrics* 28 (1986), pp. 11-18.

G.E.P. Box and R.D. Meyer, *Finding the Active Factor in Fractionated Screening Experiments*. Journal of Quality Technology 25 (1993), pp. 94-105.

G.E.P. Box, J.S. Hunter and W.G. Hunter, *Statistics for experimenters: Design, innovation, and discovery*. Wiley, New Jersey, 2005.

C. Daniel, *Use of half-normal plots in interpreting factorial two-level experiments*, *Technometrics* 1 (1959), pp. 311-341.

G. De León, P. Grima and X. Tort-Martorell, Comparison of normal probability plots and dot plots in judging the significance of effects in two level factorial designs, J. Appl. Stat. 38 (2011), pp. 161-174.

S. Fontdecaba, P. Grima and X. Tort-Martorell, *Analyzing DOE with statistical software packages: Controversies and proposals.* Amer. Statist. 68 (2014), pp. 205-211.

S. Fontdecaba, P. Grima and X. Tort-Martorell, *Proposal of a single critical value for the Lenth method*. Quality Technology and Quantitative Management 12 (2015), pp. 41–51.

M. Hamada and N. Balakrishnan, *Analyzing unreplicated factorial experiments: A review with some new proposal.* Statistica Sinica 8 (1998), pp. 1–41.

R.V. Lenth, *Quick and easy analysis of unreplicated factorials*, Technometrics 31 (1989), pp. 469-473.

D.C. Montgomery. *Design and analysis of experiments*. Willey, Singapore, 2013.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2016; software available at https://www.R-project.org/

J.H. Venter and S.J. Steel, *Identifying active contrasts by stepwise testing*. Technometrics 40 (1998), pp. 304-313.

K.Q. Ye and M. Hamada, *Critical values of the Lenth method for unreplicated factorial designs.* Journal of Quality Technology 32 (2000), pp. 57–66.

K.Q. Ye, M .Hamada and C.F.J. Wu, *A step-down Lenth method for analyzing unreplicated factorial designs*. Journal of Quality Technology 33 (2001), pp. 140-152.

Chapter 7

SUMMARY OF CONTRIBUTIONS

7
SUMMARY OF CONTRIBUTIONS

In this last chapter, we describe the main contributions extracted from the thesis in Section 7.1 to finally propose some guidelines (Section 7.2) of what we think it would be appealing to perform as a future research.

7.1 Contributions

We divide this section in two Blocks, the first Block corresponding to assess factor's statistical significance when using un-replicated factorial designs and the second Block on how to plan an experimental design in order to skip some runs when there are time and expense's constraints and on the other hand when missing responses appear and they have to be estimated from high-order interactions that can be considered negligible.

7.1.1 Block I: Analyze the significance of the effects in un-replicated Factorial Designs

Lenth's method versus using negligible interactions

1. To estimate the variance of the effects with a single degree of freedom is a bad practice and nearly always worse than to apply the method of Lenth. Some software packages analyse by default the significance of the effects considering negligible the interactions of three or more factors, and they do this also for 2^3 designs in which, obviously, there is only one three factor interaction.
2. As a general rule, in eight runs designs it is better to apply the method of Lenth.
3. In 16-run designs, the negligible interactions method provides better results when 5 or more degrees of freedom can be used for variance estimation. Naturally, this happens in complete 2^4 when, interactions of three or more factors are considered negligible.

Lenth's method versus the Box-Meyer approach

4. This study clearly shows that the Box-Meyer method gives– in general – better results than the widely adopted Lenth one. Therefore, we strongly advocate for the incorporation of the Box-Meyer method to statistical packages. Having it available as an alternative or even complementary to another method will help the experimenter make better informed decisions.

5. 4 run designs: This is a neglected situation in the literature on factorial designs. In this case, both the representation of effects in NPP and Lenth's method are – by their very nature – ineffective at identifying which effects should be considered significant. With only 3 effects, it is not possible to discriminate "those that separate from the line" when the NPP is used. Naturally, miracles cannot be expected with only 3 effects, and the Box-Meyer method does not deliver excellent results either, but they are – in all scenarios considered – better than those delivered by the Lenth method, even when using the value of t that favors it more.
6. It is important to be aware that if a design with 4 runs is carried out without prior information about the variability of the response, it is not possible to analyze the significance of the effects with reasonable error probabilities. If the experiment is carried out at the end of a process of sequential experimentation, the best option is to estimate the experimental error from the values of the non-significant effects obtained in the previous experiments, and estimate the variance of effects from it.
7. 8 run designs: Of the two most usual designs (8 and 16 runs), these are the most difficult to analyze. The smaller number of effects makes it difficult to discriminate between those that are significant and those that are not. In this case the Box-Meyer method performs better than the Lenth method (better than when using the original value of $t = 3.76$, of course, but also when using the $t = 2.297$ value proposed by Ye and Hamada).
8. 16 run designs: In this case the differences are barely noticeable and the results are similar to those obtained with Lenth's method.

A last point, worth mentioning, is that the simulation carried out confirms what other authors have already shown (see, for example Ye and Hamada [2000], Fontdecaba *et al.* [2015]), namely: the value of t that appears in the original article on the Lenth method and that is still used in the most widely distributed packages of statistical software (Fontdecaba *et al.* [7]) produces, on one hand, a probability of type I errors smaller than the intended 5%; causing, as a counterpart, a high probability of type II error, that is, it does not consider effects to be active when they actually are. In all the designs and in all the scenarios considered, the value of t proposed by Ye and Hamada produces a type I error probability that is closer to 5% and a lower probability of type II error.

7.1.2 Block II: Planning a Factorial Design in order to save runs or deal with missing response values

Estimating missing values from negligible interactions in factorial designs

1. Missing values in two level factorial designs can be estimated via contrasts, in general corresponding to high order interactions that are considered negligible from scratch. The variance of this estimates is different depending on the number of runs to be estimated, the number of contrasts available and the relation between them.
2. We provide the variances of the estimates for one and two missing values in 8 run designs and for up to five missing values in 16 run designs and all possible sets of a priori negligible interactions.

3. The results are especially interesting in the situations in which the experimenter wants to reduce the experimental plan saving some runs. It is clear that in this case the experimenter can choose which runs to skip. Then the recommendations and tables provided can be very useful, given the interactions considered a priori negligible, to choose the runs to estimate the missing values with minimum variance.

Which runs to skip in two level factorial designs when not all can be performed

4. In two level factorial designs it is possible to save as many runs as there are contrasts that can be considered negligible a priori. Omitting experiments may be an interesting option for saving resources, but it has undesirable consequences: it increases the variability of the estimated effects and provokes the appearance of correlations among them. These undesirable consequences can be minimized by adequately choosing which runs to omit and using an appropriate method to estimate the skipped runs.
5. The problem can be tackled in different ways, one of which is using D-optimal designs. In comparison, the method we propose is simple and easy to understand. Furthermore, it produces estimates of the effects that not only have similar and small variances, but are also as independent as possible from each other. This produces the additional advantage that – once the missing values have been estimated from the expressions of the contrasts – the analysis procedure is the same as if there were no missing runs.
6. We believe that this approach may be useful for practitioners and experimenters who lack a deep theoretical knowledge of optimal designs and linear models. The appendix provides tables showing which runs to skip and how to estimate them depending on the type of design and the number of missing values.

Consequences of using estimated response values from negligible interactions in factorial designs

By analyzing simulations – in a wide variety of situations – of the proportion of type I and type II errors that have been discussed, our conclusions are:

7. Estimating one response value, no matter which one, in 2^3 designs is barely noticeable in terms of the difference in the proportion of type I errors. For type II errors, the difference is slightly bigger but hardly relevant. The analysis also serves to show the good performance of the Box-Meyer method compared to Lenth's. It is interesting to note that the proportion of errors when applying the Lenth method without missing values is approximately the same as when the Box-Meyer method is applied to a 2^3 design with one estimated value.
8. In 2^4 designs, working with up to 3 missing values does not produce relevant changes in the proportion of errors, whether they be type I or type II. With 4 and 5 missing values, there is indeed an increase in the proportion of errors – whether they be type I, type II, or both.
9. In 2^{6-2} designs, when a single missing value is estimated, the results hardly change. When in this same design two missing values are estimated – or one is estimated in a 2^{7-3} design – the increase in the proportion of errors is indeed noticeable, especially in some scenarios and for certain spacing values.

10. George Box [1990] said "do not rely on your results if you have too many missing observations. Usually, I would start to feel uncomfortable with the analysis when there was more than one missing observation in an 8-run experiment, or more than two observations missing from a 16-run experiment". Box refers to situations in which the number of runs has not been planned or there is a result suspected of being anomalous and which one prefers to disregard. Our results are consistent with this statement, and we can add that if one can choose the missing runs, up to 3 runs can be omitted in 16-run designs.

7.2 Future Research

We will provide some guidelines that we think are appealing and interesting in order to complement our work.

7.2.1 Block I: Analyze the significance of the effects in un-replicated Factorial Designs

1. On the one hand Hamada and Balakrishnan [1998] proved that Box and Meyer [1993] method is quite competitive for small number of active effects and on the other hand Lenth's method [1989] performs quite well for large number of active effects. In order to take advantage of both methods' properties, a hybrid method could be created, so that its synergy will outperform its individual use.
2. Due to the fact that computation time is currently not a restriction, we think that the improvement in assessing the significance of the effects in non-replicated factorial designs could be better addressed using Bayesian analysis. Thus new research should be addressed towards this direction.

7.2.2 Block II: Planning a Factorial Design in order to save runs or deal with missing response values

3. Another method that allows tackling the problem of finding missing responses or planning experimental designs saving some runs is through D-optimal designs. New research should be addressed in this filed.

4. There could be alternative modeling techniques that handle this situations such as Partial Least Square Regression, Neural Networks and Partition Models. These could be other options for fitting models to handle these situations differently.

Chapter 8

LIST OF REFERENCES OF ALL CHAPTERS

8

LIST OF REFERENCES OF ALL CHAPTERS

Almimi, A.A., Kulahci, M., Montgomery, D.C. Estimation of missing observations in two-level split-plot designs. *Quality and Reliability Engineering International*; 24:127–152. DOI: 10.1002/qre.871, (2008)

Barrios E. (based on Daniel Meyer's code). BsMD: Bayes Screening and Model Discrimination. R package version 2013.0718 https://CRAN.R-project.org/package=BsMD, (2013)

Bell, G.H., Ledolter, J. & Swersey, A.J. Experimental design on the front lines of marketing: Testing new ideas to increase direct mail sales, *International Journal of Research in Marketing,* 23:309-319. (2006)

Box, G.E.P. A simple way to deal with missings observations from DOE. *Quality Engineering*; 3(2):249-254. (1990)

Box, G.E.P., Hunter, J.S. & Hunter, W.G. *Statistics for experimenters: design, innovation, and discovery*. New Jersey: Wiley. (2005).

Box, G.E.P. & Meyer, R.D. An Analysis for Unreplicated Fractional Factorials. *Technometrics,* 28:11-18. (1986).

Box, G.E.P. and Meyer, R.D. Finding the Active Factor in Fractionated Screening Experiments. *Journal of Quality Technology,* 25:94-105. (1993).

Daniel, C. *Use of half-normal plots in interpreting factorial two-level experiments*, *Technometrics* 1, pp. 311-341. (1959)

Dunn, K. 5. Design and Analysis of Experiments — Process Improvement using Data. Retrieved March 20, 2018, from https://learnche.org/pid/design-analysis-experiments/index (2018)

G. De León, P. Grima and X. Tort-Martorell, Comparison of normal probability plots and dot plots in judging the significance of effects in two level factorial designs, *Journal of Applied Statistics* 38, pp. 161-174. (2011)

Dong, F. On the identification of active contrasts in unreplicated fractional factorials. *Statistica Sinica,* 3:209-217. (1993)

Draper, N., Stoneman, D. Estimating missing values in unreplicated two-level factorial and fractional factorial designs. *Biometrics*; 20(3):443-458. (1964)

Espinosa, V., Dasgupta, T. & Rubin, D.B. A Bayesian Perspective on the Analysis of Unreplicated Factorial Experiments Using Potential Outcomes. *Technometrics,* 58: 62-73. (2016)

Fontdecaba, S., Grima, P. & Tort-Martorell, X. Analyzing DOE with satistical software sackages: controversies and proposals. *The American Statistician,* 68:205-211. (2014)

Fontdecaba, S., Grima, P. & Tort-Martorell, X. Proposal of a single critical value for the lenth method. *Quality Technology and Quantitative Management,* 12:41-51. (2015)

Godolphin, J.D. Reducing the Impact of Missing Values in Factorial Experiments Arranged in Blocks. *Quality and Reliability Engineering International* 22:669-682. (2006)

Goh, T.N. Economical experimentation via 'lean design'. *Quality and Reliability Engineering International*; 12:383–388. (1996)

Goos, P. and Jones, B. *Optimal Design of Experiments. A Case Study Approach*. Chichester (UK): John Wiley & Sons. (2011)

Hamada, M. & Balakrishnan, N. Analyzing unreplicated factorial experiments: a review with some new proposal. *Statistica Sinica,* 8:1-41. (1998)

Jarrett, R.G. The Analysis of Designed Experiments with Missing Observations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 27, No. 1 (1978), pp. 38-46. (1978)

John, P.W.M. Missing Points in 2^n and 2^{n-k} Factorial Designs. Technometrics. Vol. 21, No. 2, pp: 225-228. (1979)

Johnson, R. T., Montgomery, D. C. and Jones, B.A. An Expositori Paper on Optimal Design. *Quality Engineering. 23:287-301.* (2011)

Juan, J. & Peña, D. A simple method to identify significant effects in unreplicated two-level factorial designs. *Communications in Statistics. Theory and Methods,* 21: 1383-1403. (1992)

Lenth, R.V. Quick and easy analysis of unreplicated factorials. *Technometrics,* 31: 469-473. (1989)

Loughin TM, Calibration of the length test for unreplicated factorial designs. *Journal of Quality Technology*, 30(2):171-175. (1998)

Minitab 17 Statistical Software. [Computer software]. State College, PA: Minitab, Inc. (2010)

Minitab. Quality Trainer Content Outline - Minitab. Retrieved March 20, 2018, from http://www.minitab.com/en-us/products/quality-trainer/content-outline/ (2018)

Montgomery, D.C. *Design and analysis of experiments*. Singapore: Willey. (2013)

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; software available at https://www.R-project.org/ (2016)

Robinson, T.J. and Anderson-Cook. A Closer Look ar D-Optimality for Screening Designs. *Quality Engineering. 23:1-14.* (2011)

Srivastava, R., Gupta, V.K., Dey, A. Robustness of some designs against missing data. *Journal of Applied Statistics*. Vol. 18, No. 3. (1991)

Statistica 13.0. [Computer software]. Dell Software. (2013)

Venter, J.H. & Steel, S.J. Identifying active contrasts by stepwise testing. *Technometrics,* 40:304-313. (1998)

Wilkinson, G.N. Estimation of missing values for the analysis of incomplete data. *Biometrics*, 14, 257-286. (1958a)

--------- The analysis of variance and derivation of standard errors for incomplete data. *Biometrics*, 14, 360-384. (1958b)

Xampeny, R., Grima, P. and Tort-Martorell, X. Estimating missing values from negligible interactions in factorial designs. *Quality and Reliability Engineering International*. 33(6):1235-1247. (2017)

Xampeny R, Grima P, Tort-Martorell X. Which runs to skip in two level factorial designs when not all can be performed. *Quality Engineering.* DOI: 10.1080/08982112.2018.1428751 (2018)

Ye, K.Q. & Hamada, M. Critical values of the Lenth method for unreplicated factorial designs. *Journal of Quality Technology,* 32:57-66. (2000)

Ye, K.Q., Hamada, M. & Wu, C.F.J. A step-down length method for analyzing unreplicated factorial designs. *Journal of Quality Technology,* 33:140-152. (2001)

Zhou, M., Goh, T.N. Iterative Designed Experiment Analysis (IDEA). *Quality and Reliability Engineering International*; 32:2977-2986. DOI: 10.1002/qre.1981. (2016)