Saving runs in fractional factorial designs

Pere Grima, Lourdes Rodero, Xavier Tort-Martorell
Department of Statistics and Operational Research
Universitat Politècnica de Catalunya - BarcelonaTech, Spain

ABSTRACT

When it is known a priori that some contrasts are negligible in a factorial design, their expressions can be used to deduce the missing results. In this article we propose a method for using this procedure when, as in the case of fractional designs, it is not known which contrasts will be null. The method is based on first establishing an interval of possible values corresponding to each of the missing results, then identifying which contrasts are always null independently of the value of said results.

1. Introduction

In industrial contexts, conducting experiments is usually costly and the resources for carrying out experimental designs are often scarce. The seed idea for this paper came about when two of the authors were giving training and advice on DEO to a company from the aeronautic sector. In it, as part of the optimization of a welding process of a complicated component of a turbine it was decided to conduct some experiments. Runs were expensive, the material was an expensive titanium alloy and slow, something that interfered the normal production process (some experiences of our work with this company were published in Lluís Marco-Almagro et al., 2014). The first step in a sequential experimentation process was to conduct a two level fractional factorial that unfortunately could not be completed. The last two runs could not be executed. Among other ways of analyzing the data, we asked the technicians the range of values

in which it seemed reasonable to find the response of the two runs not made, and we used those intervals as a check on the validity of the conclusions of our analysis.

When a few years later, we saw the iterative method proposed by Zhou, M. and Goh, TN (2016) to save runs in two level factorial designs by detecting the point when carrying out new runs will not add relevant information, it made us think about the welding experiment and about how important it can be in certain circumstances to be able to decide –before executing all the planned runs– whether or not carrying out the missing ones will add any relevant information.

Zhou and Goh method is based on updating the effect values as the runs are carried out. In order to calculate the effects, the missing values are assigned the average of the responses that have already been obtained. When the values of the effects stabilize, the experiment has finished. In this article we present a more elaborated and precise method based on using the experimenter's technical knowledge of the possible outcome of the runs not yet realized and combing this with the responses that are already available. More specifically, the previous knowledge of the experimenter is used to establish – for each run not yet performed – an interval in which it is known with certainty what the response would be if it were carried out. So, for example, if the response is the performance of a chemical reaction, its value will surely be between 0 and 100%. In addition, it is possible that the experimenter can affirm without fear of error that under certain conditions the performance will be greater than 30% and less than 90%. This allows us to estimate the effects for all the possible results of runs not performed and to detect those effects that are not significant, regardless of what those results are. The proposed method also uses results from the work of Xampeny et al. (2018) showing which runs to skip in a two level factorial experiment so that the estimation of their results has the least possible influence on the analysis of the effects.

To estimate the response corresponding to the skipped runs we use the method proposed by Draper and Stoneman (1964). The method consists of using the estimates of contrasts that are considered negligible a priori, for example interactions of 3 or more factors. The missing values are estimated by equaling to zero the algebraic expressions of those contrasts. If you have more than one missing value, it is necessary to set up systems of equations and it is a necessary condition – although not sufficient – to have as many null contrasts as there are missing values. Box (1990) provides a very clear and didactic explanation of how to estimate the missing values by using the interactions considered null, doing so by following the original proposal of Draper and Stoneman. Goh (1996) also elaborates on the possibility of omitting the executions of as many runs as there are contrasts that can be considered null – not only those corresponding to 3 or more factors interactions, but also other contrasts based on the experimenter's previous knowledge. In this context where contrasts can be considered a priori null, Xampeny et al.

(2017) have studied which are the runs that, if omitted, can be estimated with the best properties.

Our proposal focuses on designs with 8 and 16 runs. Designs with 32 or more runs are not usual, neither in practice nor in the technical literature; furthermore, when experimental designs allow for the execution of so many runs, it is not usually in a context of scarce resources. In any case, our proposal can easily be generalized to two-level factorial design with any number of runs.

The article is organized as follows. The following section describes the possibilities and limitations of estimating missing values through expressions of contrasts that are considered negligible. Next section presents the proposed methodology and provides and illustration of its operation in 8-run designs by using an example from the book by Box, Hunter and Hunter. In the following section we present a different example from Box, Hunter and Hunter to discuss the method for saving up to two runs in 16-run designs. Finally, we outline the procedure for saving runs and provide some remarks and conclusions.

2. Estimation of missing values based on negligible contrasts

As we have already mentioned, the procedure consists of deducing the missing values by equating to zero the expressions of the contrasts considered negligible from scratch. For example, the book by Box, Hunter and Hunter (2005) presents the calculation of effects through a 2^3 design in which the factors are Temperature (T), Concentration (C) and Catalyst Type (K). The response is the performance of the chemical reaction performed under these conditions. Table 1 provides the design matrix, enlarged with the columns for all contrasts and the values of the responses (y).

Table 1: *Design example of a* 2^3 *from Box, Hunter and Hunter (2005)*

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

Figure 1 (left) presents the effects in a Normal Probability Plot (NPP) and shows that the significant ones are T, C and the interaction TK.

Let us now suppose that one of the runs has not been carried out due to unforeseen problems or intentionally because the available resources allowed for only 7 runs to be carried out. Considering the value of the TCK interaction to be negligible, we can write:

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

And from this expression we can estimate the value of any result based on the others. For example, if y_5 is missing we have:

$$\hat{y}_5 = y_1 - y_2 - y_3 + y_4 + y_6 + y_7 - y_8 = 50$$

By calculating the effects with this new value, we obtain the results in Figure 1 (right), in which the interaction whose value has been forced to equal zero is not represented because it does not represent the variability of the effects.
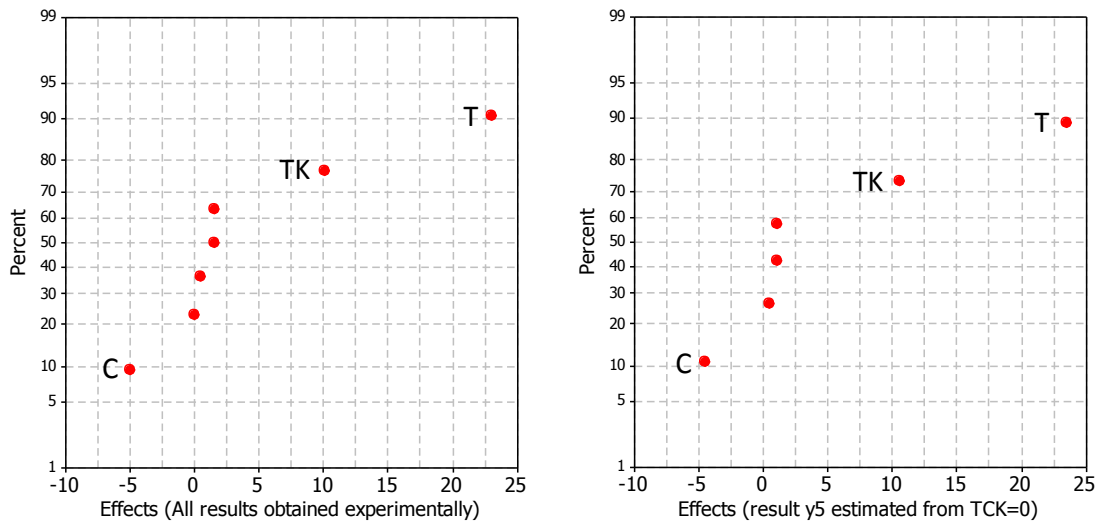
Figure 1: NPP of the effects obtained with all the results (left) and by estimating one of them (right)

This practice also has some drawbacks. First, there is the fact that estimating the response for a run from others responses provokes the appearance of correlation between the effects, they stop being independent. In addition, the variance of the estimated response is greater than that obtained directly from the experimentation. In a 2^3 design, if we call σ_y^2 the variance of the response, it follows immediately from the way the missing response is estimated that its variance will be $7\sigma_y^2$. In turn, this increase inflates the variance of the effects.

In two level factorial designs, when all runs are conducted all the effects have the same variance: $V(\text{effect}) = \frac{4\sigma_y^2}{N}$ where N is the total number of runs (see, for example, Box et al. 2005). For instance, the expression for the T main effect is:

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

And from here we get $V(T) = \frac{\sigma_y^2}{2}$.

Things are different when one run is missing and its response is estimated. For example if y_5 is missing, the T main effect expression becomes:

$$T = \frac{1}{4}[-y_1 + y_2 - y_3 + y_4 - (y_1 - y_2 - y_3 + y_4 + y_6 + y_7 - y_8) + y_6 - y_7 + y_8]$$

Where y_5 is estimated, according to the procedure explained above, by the expression between parentheses. Operating we get:

$$T = \frac{1}{4}(-2y_1 + 2y_2 - 2y_7 + 2y_8)$$

And thus that $V(T) = \sigma_y^2$. Double the value than when all responses are obtained experimentally.

The correlation presented by the effects as well as their increase in variance make it difficult to identify the active effects. It must also be borne in mind that while the existence of one negligible contrast is always sufficient to estimate a single missing value, when the number of missing values is bigger than one the number of negligible contrasts necessary to estimate them may be bigger than the number of missing values.

Xampeny *et al.* (2017) have studied these drawbacks, showing that the greater the ratio of available contrasts in relation to results to be estimated, the lower the variance of the latter. They also show that when the variance of the effects is low, they also tend to have a low correlation between them. So, the problem can be reduced to minimize the increment in the variance of the effects. As said, this is achieved by having the greatest possible number of contrasts that can be considered null.

This article is directed at facilitating the estimation of one missing value in designs with 8 runs and up to two in designs with 16 runs even when there are no contrasts that can be considered negligible from scratch. Of course, the procedure is easy to generalize to any number of missing values, but the increase in the variance of the effects may compromise the validity of the

conclusions obtained. Already in 1990 Box said: "Usually, I would start to feel uncomfortable with the analysis when there was more than one missing observation in an eight-run experiment, or more than two observations missing from a 16-run experiment".

3.   Eight-run design. Saving one run

Box, Hunter and Hunter (2005) present a 2^{7-4} design to study how a set of 7 factors affect the time a cyclist takes to climb a hill. The factors are related to the condition of the bicycle and also of the cyclist. They are found in Table 2 together with the response.

Table 2: 2^{7-4} design to study how 7 factors affect a response (Box, Hunter and Hunter, 2005)

Run Number	Seat Up/Down A	Dynamo Off/On B	Handlebars Up/Down C	Gear Low/Medium D (=AB)	Raincoat On/off E (=AC)	Breakfast Yes/No F (=BC)	Tires Hard/Soft G (=ABC)	Climb Hill (sec) y
1	-	-	-	+	+	+	-	69
2	+	-	-	-	-	+	+	52
3	-	+	-	-	+	-	+	60
4	+	+	-	+	-	-	-	83
5	-	-	+	+	-	-	+	71
6	+	-	+	-	+	-	-	50
7	-	+	+	-	-	+	-	59
8	+	+	+	+	+	+	+	88

The contrasts are calculated as before with the only difference that since it is a fractional design each contrast represents the sum of several effects. The string of effects associated to each contrast provides what is called the alias structure. We ignore it here, as this is not relevant to the topic at hand. For example, the contrast associated to the main effect of factor A represents the sum of 16 effects and it is calculated as:

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

The values obtained for each contrast are:

$$l_A = 3.5 ; \quad \boldsymbol{l_B = 12.0}; \quad l_C = 1.0; \quad \boldsymbol{l_{AB} = 22.5}; \quad l_{AC} = 0.5; \quad l_{BC} = 1.0; \quad l_{ABC} = 2.5$$

. Based on the variability obtained in previous experiments, it can be deduced that only the contrasts l_B and l_{AB} (in bold) are statistically significant.

Let us suppose that by randomizing the run order, run number 5 appears in last place and the cyclist does not have time to perform it. We can establish a range of values within which it is

reasonable to consider that its response will lay. The responses of previous runs together with the knowledge and experience of the experimenter can help to establish this interval. An automatic criterion can also be applied based on the values already obtained. For example, if M is the maximum value of the results obtained and m is the minimum value, the interval could be: $[m - k(M - m), \ M + k(M - m)]$, with k being a value of the order of 0.2. We have analyzed several examples from the literature, and have found that this value is rarely exceeded. In any case the experimenter can choose to be more conservative and use a bigger k value. Naturally, if the extremes are outside the range of possible values (for example, they go beyond 100% in the performance of a chemical reaction) the minimum/maximum possible values will be used.

By applying the automatic criterion, in our case the interval would be established between 42.4 and 95.6 seconds. But let us be more conservative and consider the value to lie somewhere between 40 and 100. In that interval, we place 101 equidistant values (in our case: 40, 40.6, …, 100), and for each of them we calculate the values of the contrasts. For example, with the first of them, $y_5 = 40$, the values of the contrasts are:

$$l_A = 11.25 \ ; \quad l_B = 19.75; \quad l_C = -6.75; \quad l_{AB} = 14.75; \quad l_{AC} = 8.25; \quad l_{BC} = 8.75; \quad l_{ABC} = -5.25$$

Next we identify which are statistically significant. We will use Lenth's method, which is a general method and probably the best known and most used. This 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 $1.5 \cdot \text{median}|X| = 1.01\sigma \cong \sigma$. Considering that $\kappa_i \ (i = 1, \dots, n)$ 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 $s_0 = 1.5 \cdot \text{median}|c_i|$ and calculates a new median by excluding the estimated effects with $|c_i| > 2.5s_0$. By doing so he expects to exclude the effects with $\kappa > 0$ and use the others to calculate the so-called Pseudo Standard Error:

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

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

In the original article (Lenth, 1989), it is proposed that $v = n/3$ where n is the number of effects considered, obtaining the values of 3.76 and 2.57 for t in designs with, respectively, 8 and 16 runs. However, several authors have shown that these values produce type I errors (considering a contrast significant when in reality it is not) with an error probability that is clearly lower than the intended 5%; furthermore, there is the negative consequence of a higher probability of type II error (considering a contrast not significant when it actually is). For example, Loughin (1998) proposes $t = 2.300$ and 2.152 for 8- and 16-run designs; Ye and Hamada (2000) propose very

similar values (2.297 and 2.156); while Fontdecaba et al. (2015) propose using $t = 2$ in designs with both 8 and 16 runs. Here we use their proposal because it is the one that generally presents the best results and also because being the one that proposes a lower value of t is the most demanding when declaring an effect as non-significant. So it is the most conservative with respect to our methodology. Additionally, we have verified that the results are very similar to those obtained using the values of Loughin and of Ye and Hamada.

From the contrast values obtained with $y_5 = 40$, we get $PSE = 13.125$, and since all the contrast values are within the interval $0 \pm 2 \cdot PSE$, none will be significant in this case. Figure 2 shows which values of y_5 are significant for each of the contrasts after scanning the 101 values in which we have divided the range of possible values.
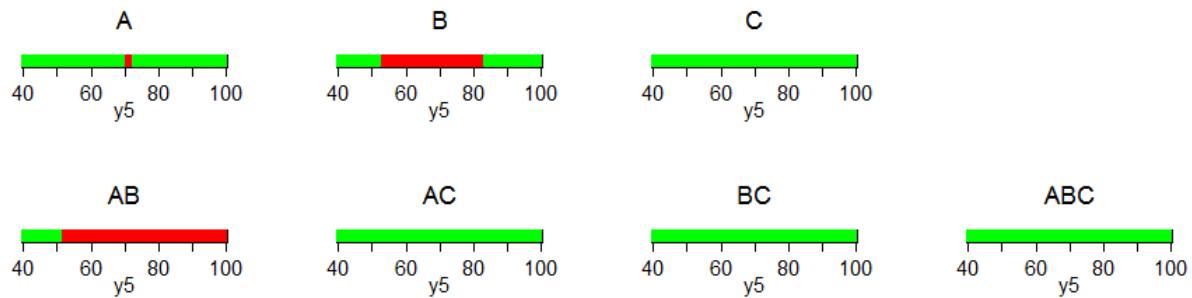

Figure 2: Range of values for the missing result (in red) and in which the contrast is significant

The contrasts that contain the effects C, AC, BC and ABC are not significant, regardless of the value that y_5 takes; therefore, their expressions can be used to estimate the missing value. By equating to zero their expressions, we have:

$$\hat{y}_{5(C)} = +y_1 + y_2 + y_3 + y_4 - y_6 - y_7 - y_8$$

$$\hat{y}_{5(AC)} = +y_1 - y_2 + y_3 - y_4 + y_6 - y_7 + y_8$$

$$\hat{y}_{5(BC)} = +y_1 + y_2 - y_3 - y_4 - y_6 + y_7 + y_8$$

$$\hat{y}_{5(ABC)} = +y_1 - y_2 - y_3 + y_4 + y_6 + y_7 - y_8$$

As shown in Xampeny et al. (2017), the best estimator will be the average of the values obtained from each of the null contrasts. In this case:

$$\hat{y}_5 = \frac{1}{4} 4 \cdot y_1 = 69$$

Therefore, the $\text{Var}(\hat{y}_5) = \text{Var}(y_1) = \sigma_y^2$. Curiously, in this occasion, the effects will have the same variance as if all runs would have been conducted.

Table 3 shows which ones are the null interactions (contrasts) and the estimate obtained for the missing value depending on what it is. Figure 3 shows the values of the contrasts depending on what the missing response is (the number 0, with a darker gray bar represents the value when there is no missing response).

Table 3: Interactions that are null and the estimated value based on the missing response

Missing result	Null interactions for all possible values of missing result	Omitted run estimation (current value)
y_1	C, AC, BC, ABC	71 (69)
y_2	C, AC, BC, ABC	50 (52)
y_3	C, AC, BC	62 (60)
y_4	C, AC, BC	86. 3 (83)
y_5	C, AC, BC, ABC	69 (71)
y_6	C, AC, BC, ABC	52 (50)
y_7	C, AC, BC	57 (59)
y_8	C, AC, BC	84. 7 (88)
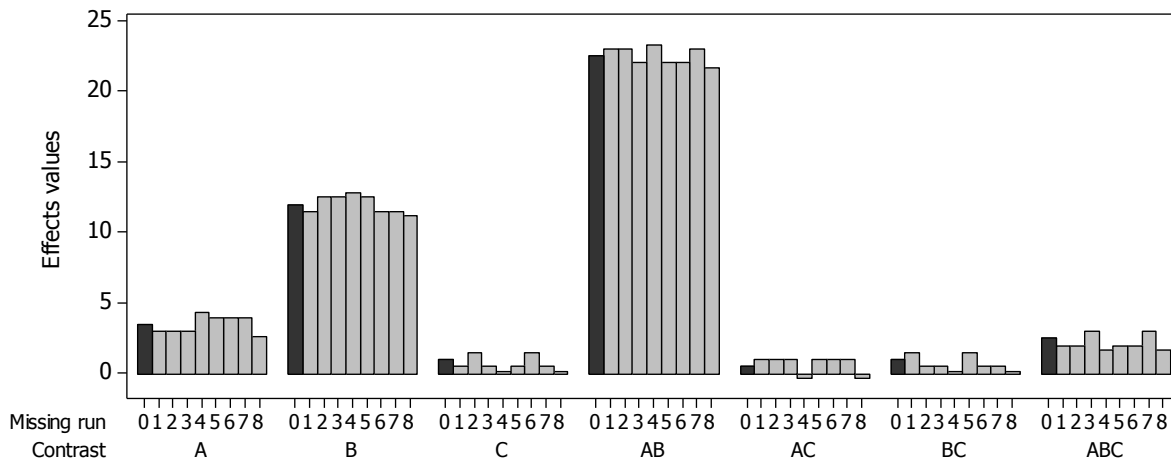
Figure 3: The contrast values obtained according to the missing run (in standard order). The value 0 represents situations in which no run is missing.

It can be observed that regardless of which run is not carried out, the conclusions will be the same as if all runs have been conducted.

4. Sixteen-run design. Saving up to two runs

In 8 run designs, there are only 7 contrasts available; while 16 run designs have 15. By having more contrasts, it is expected that there will also be a greater number of them that are negligible, such that the value of a missing result can be estimated with greater precision than in designs with 8 runs. In addition, with a bit of luck you can also estimate two missing results without the conclusions being affected. If there is no luck, one more run will have to be done; and if you have very bad luck you will have to do them all.

We continue with the text of Box, Hunter and Hunter (2005), in which a 2^{5-1} design is presented for studying the performance of a chemical reaction based on 5 factors (A: Feed rate, B: Catalyst, C: Agitation rate, D: Temperature, E: Concentration). Table 4 contains the design matrix along with the values of the response and the effects obtained.

Table 4: Design matrix and estimated effects (significant ones in bold) from the 2^{5-1} chemical experiment

Run	A	B	C	D	E (=ABCD)	y (% reacted)	Effects
1	−	−	−	−	+	56	$I(+ABCDE) = 62.25$
2	+	−	−	−	−	53	$A(+BCDE) = -2.0$
3	−	+	−	−	−	63	$\boldsymbol{B(+ACDE) = 20.5}$
4	+	+	−	−	+	65	$C(+ABDE) = 0.0$
5	−	−	+	−	−	53	$\boldsymbol{D(+ABCE) = 12.15}$
6	+	−	+	−	+	55	$AB(+CDE) = 1.5$
7	−	+	+	−	+	67	$AC(+BDE) = 0.5$
8	+	+	+	−	−	61	$AD(+BCE) = -0.75$
9	−	−	−	+	−	69	$BC(+ADE) = 1.5$
10	+	−	−	+	+	45	$\boldsymbol{BD(+ACE) = 10.75}$
11	−	+	−	+	+	78	$CD(+ABE) = 0.25$
12	+	+	−	+	−	93	$\boldsymbol{ABC(+DE) = -9.50}$
13	−	−	+	+	+	49	$ABD(+CE) = 2.25$
14	+	−	+	+	−	60	$ACD(+BE) = 1.25$
15	−	+	+	+	−	95	$BCD(+AE) = 1.25$
16	+	+	+	+	+	82	$\boldsymbol{ABCD(+E) = -6.25}$

Let us suppose that runs 5 and 10 could not be performed but we know that the performance will be equal to or greater than 40% and, of course, equal to or less than 100%. By scanning the values of y_5 and y_{10}, in the same way as for a single missing value, and calculating the effects and analyzing their statistical significance, we obtain the graph in Figure 4, where the red zones correspond to combinations of the values y_5 and y_{10}, for which the effect indicated is significant.
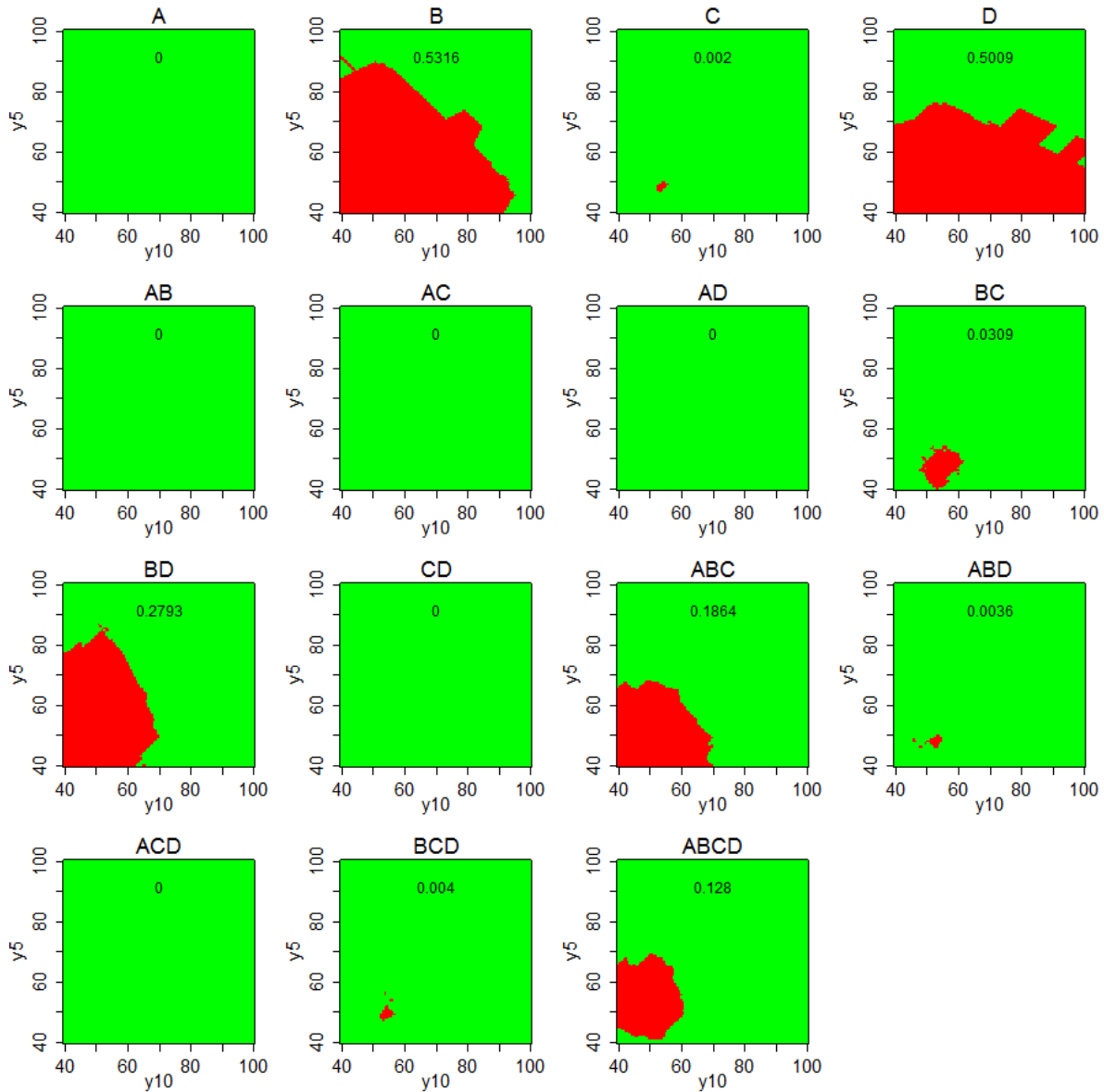
Figure 4: *Significant effects (red zone) depending on the values of the missing results.*

The contrasts that contain the effects A, AB, AC, AD, CD and ACD are null for any combination of values for y_5 and y_{10} within the range where they can conceivably be found. The number included in each square indicates the proportion of the surface in which the effect appears to be significant. When this value is very low (for example, less than 0.01) this effect could also be considered null.

Estimating two missing responses must be done through a system of two equations, each one based on an interaction. However, not all the systems that can appear are consistent. For example, from the interactions of A and AB equaled to zero, we obtain an inconsistent system of equations. However, with A and AC, we have:

From A: $\quad \hat{y}_5 - \hat{y}_{10} = -y_1 + y_2 - y_3 + y_4 + y_6 - y_7 + y_8 - y_9 - y_{11} + y_{12} - y_{13} + y_{14} - y_{15} + y_{16} =$
$$= -8$$

From AC: $\hat{y}_5 + \hat{y}_{10} = \quad y_1 - y_2 + y_3 - y_4 + y_6 - y_7 + y_8 + y_9 + y_{11} - y_{12} - y_{13} + y_{14} - y_{15} + y_{16} =$
$$= 102$$

In matrix notation:

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \hat{y}_5 \\ \hat{y}_{10} \end{pmatrix} = \begin{pmatrix} -8 \\ 102 \end{pmatrix}$$

From which $\hat{y}_5 = 47$ and $\hat{y}_{10} = 55$ are obtained. Notice that to be able to solve the system of equations, the matrix of the coefficients of the missing values must be invertible; that is, their columns must be linearly independent.

With the 6 contrasts found to be null we can set up 15 systems of two equations, but only 9 are consistent for estimating the values of the missing responses y_5 and y_{10}. Table 5 allows us to deduce which contrasts can be used to formulate compatible systems and it also indicates, in each case, the estimates obtained for the example. For the system to be compatible, it is necessary that the two missing values have a different background color.

Table 5: *Identification of systems of equations that allow deducing the values of* y_5 *and* y_{10}.

	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}	$\hat{y}_5$	$\hat{y}_{10}$
A	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-	-
AB	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1		
A	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	47	55
AC	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1		
A	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	48	56
AD	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1		
A	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	46	54
CD	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1		
A	-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		
AB	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	49	53
AC	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1		
AB	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	50	54
AD	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1		
AB	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	48	52
CD	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1		
AB	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		
AC	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1	-	-
AD	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1		
AC	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1	-	-
CD	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1		
AC	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1	50	52
ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1		
AD	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	-	-
CD	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1		
AD	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	51	53
ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1		
CD	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	49	51
ACD	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1		
															Mean estimates		48.7	53.3

A situation could occur in which it is impossible to deduce the values of two missing responses. For example, if the two omitted runs are 8 and 12, the contrasts that appear null for any value of the missing responses are A, CD, ACD and BCD. Table 6 shows that all systems of two equations that can be considered with these interactions for deducing the values of y_8 and y_{12} are inconsistent. It can also happen that no compatible system can be set up with the null interactions found, as in our example when the missing values are y_1 and y_6, despite the fact that we have 7 null interactions (A, C, AB, AD, BC, CD, BCD) and therefore the possibility of constructing 21 systems of two equations. Finally, we have another possible situation that

would occur if the missing values were y_6 and y_7, in which case no contrast is null in all the combinations of possible values, so these values could not be estimated either.

Table 6: If the missing values are y_8 and y_{12}, all the systems that can be formulated are inconsistent.

	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}
A	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1
CD	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1
A	-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
A	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
CD	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
CD	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
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

In 16-run designs there are 120 possible combinations of 2 missing responses. In our example, in which the criterion establishes that the responses will be between 40 and 100, only 66 of those values can be estimated. It is not possible for the other 54 because either the systems that can be formulated are inconsistent or no contrast appears to be null, something that only happens with the pair of missing values y_6 and y_7, as mentioned previously.

If we take a certain risk and consider that the contrast is null when we have a very small surface corresponding to value combinations that cause the effect to be active, as in the case of the effects C (0.002), ABD (0.0036) or BCD (0.004) (see Figure 4), then more null contrasts are available and a larger number of pairs of missing values can be estimated. By placing the threshold at 0.05, we can estimate the values for of 108 out of the 120 possible pairs of missing values. Among the pairs whose values we can now estimate, we find those discussed above: y_8 and y_{12} as well as y_6 and y_7. Figure 5 shows the number of times that each effect appears to be significant in those 108 pairs of missing responses. The maximum variance of the effects turns out to be, in this case, $0.3\sigma_y^2$. It would be $0.25\sigma_y^2$ if there were no missing values. In the 108 cases (that is, in all of them), we can identify as significant the same effects that were considered to be so when using all the results obtained experimentally, although it is also true that in 24 cases some effect is considered significant that does not appear to be so when using the experimentally obtained results (Figure 5).

Ex. from Box, Hunter and Hunter (2005) p.: 263

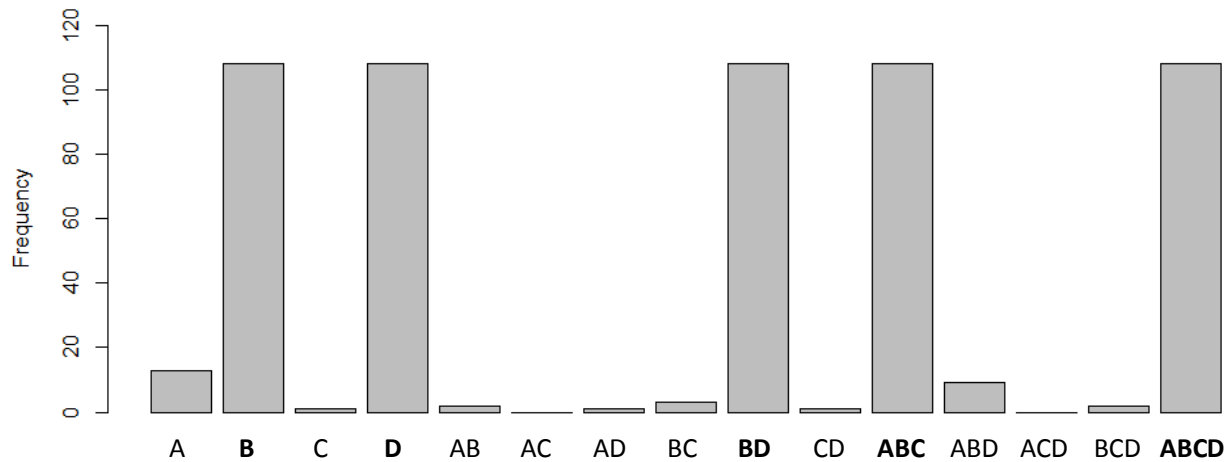Lenth Method with t = 2, Missings interval: 40-100

Figure 5: *Number of pairs of missing values in which the effects appear to be significant (in bold)*

The combination of missing results y_1 and y_6 are found among the 12 that still cannot be estimated. In this case there would be no choice but to conduct one more run, and the last one can definitely be omitted. For example, if y_1 is carried out and y_6 is left to be estimated, the results in Figure 6 are obtained, where the red intervals indicate the values of y_6 in which the effect is significant.
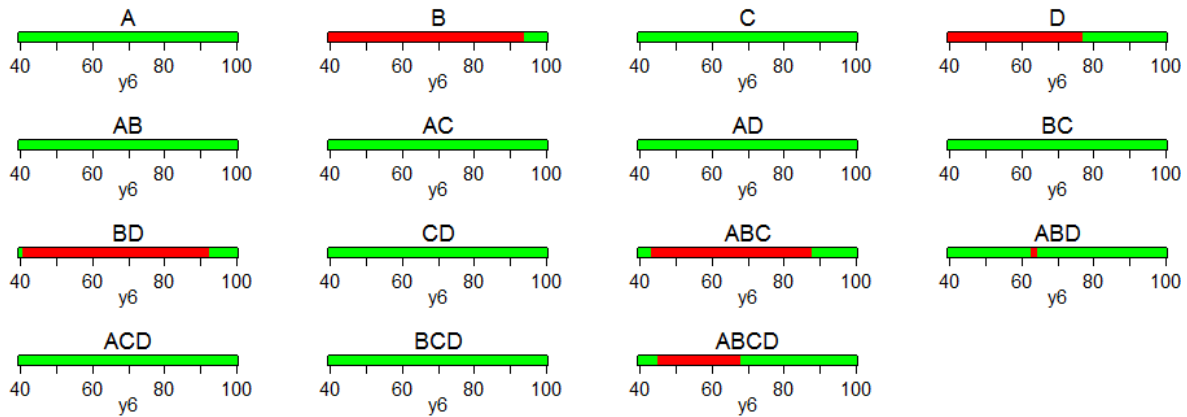
Figure 6: *Intervals of values of* y_6 *(red color) in which the effect is significant*

Table 7 contains the estimates of y_6 obtained from each of the contrasts considered null, and the average of all of them which is the best estimate and the one used. The maximum variance that is obtained for the effects is $0.278\sigma_y^2$ and the conclusions that are obtained using the estimated value of y_6 are identical to those using the original value (Figure 7).

Table 7: Estimated values for y_6 from each of the interactions considered null

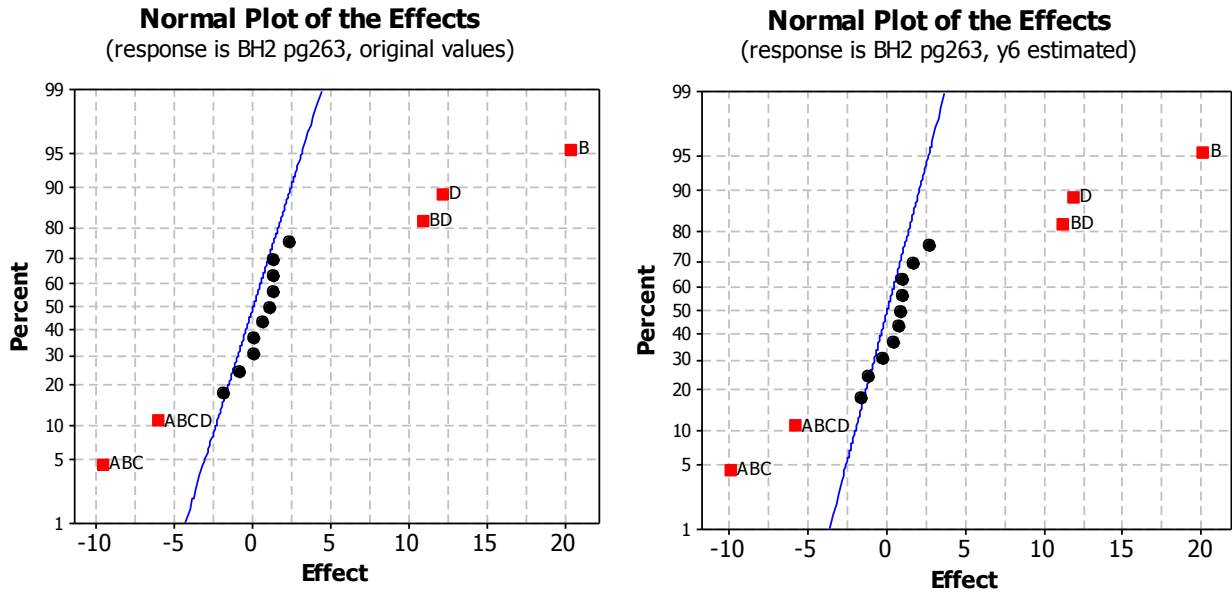| Contrast:	A	C	AB	AC	AD	BC	CD	ACD	BCD	Mean
$\hat{y}_6$:	71	55	67	51	49	67	57	65	45	58.56

Figure 7: Analysis of the contrasts in NPP using all the results obtained from the experimentation (left) and with the estimated value for y_6 (right)

5. Summary and conclusions

Next, we define the steps to follow for running the experiment and analyzing the results sequentially, which on many occasions will allow omitting one run in 8-run designs and up to two in 16-run designs. The steps to follow are:

1. Randomize the execution order of the runs.

2. Perform the first 7 in 8-run designs or the first 14 in 16-run designs.

3. Establish an interval in which the results of runs not performed are believed to be. An automatic criterion can also be established based on the results that have already been obtained (Section 3).

4. Perform a scan of the missing values within the intervals being considered. For each value or combination of values, identify what the significant effects are. This scan can be done with

the help of a small program constructed using the statistical software R (R Core Team, 2016), and it can graphically represent the results obtained just as they are in this article.

5. Identify the contrasts that are not significant, independently of the value of the missing result (8-run designs) or results (16-run designs).

6. Use the contrasts that always appear to be non-significant for estimating the values of the missing results. In the event that there is no contrast that always appears to be non-significant, we can set a threshold for the percentage of values that lead to the contrast being significant and below which it can be considered null. If establishing a reasonable threshold, for example 5%, no null contrasts appear there will be no choice but to perform a new run.

7. In 16-run designs, if there are less than two non-significant contrasts or there are two or more but the system of equations to estimate them is inconsistent, then there will be no choice but to perform a new run. With the 15 runs performed, the procedure must be repeated with the hope that there is at least one contrast that can be considered negligible.

8. Finally, in view of the results obtained in the analyzed examples – and as a rule of thumb – if the variance of the effects exceeds $0.67\sigma_y^2$ in 8-run designs 8 or $0.33\sigma_y^2$ in 16-run designs (33% increases in both cases), it is prudent to perform a new run to lower this variance and increase the probability that the conclusions are exactly the same as if all the runs had been performed.

The proposed method expands on the method of Draper and Stoneman, as it allows estimating the response in runs that have not been performed even in cases where no contrast can be considered a priori negligible. As in the method proposed by Zhou and Goh (2016), this method is applied with the aim of reducing the number of runs and it also uses the previous knowledge about the experimental system. However, the method proposed here has a solid rationale base, makes a more intense and better use of the experimenter previous knowledge and also provides a clear guideline of when a new run is needed and when the available information makes it unnecessary.

This method can also be used to improve estimates when it comes to designs in which it is known a priori that some contrasts will be null, typically in complete designs. Indeed, there will most likely be more null contrasts than those that are known a priori, and our procedure allows experimenters to detect them. Having a greater number of contrasts that can be considered null allows estimating the missing values with less variance and, therefore, having greater certainty that the conclusions will not be affected by having performed a smaller number of runs.

References

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: Wiley.

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

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

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

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

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

Marco-Almagro, Ll., Grima, P., Tort-Martorell, X. 2014. Light Consulting after Design of Experiments Training: Lessons from a Case Study. *Quality and Reliability Engineering International* 30:437–446.

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.

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.

Xampeny, R., Grima, P. and Tort-Martorell, X. 2018. Which runs to skip in two-level factorial designs when not all can be performed. *Quality Engineering*. Published Online: 4 Apr 2018 (to be actualized).

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

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