Pareto plot threshold for multiple management factors in design of experiments

T. VAN HECKE

Ghent University, Faculty of Engineering and Architecture
Voskenslaan 270, Ghent, Belgium

July 22, 2016

Abstract

Economical, technical and strategic reasons make unreplicated experimental designs necessary and popular in industrial and management settings. This paper investigates the legitimate use of well known methods that appoint effects as significant when they are larger than a defined threshold. We found that established estimation methods, such as Lenth's method and Dong's method, are too tolerant and a larger threshold should be used before an effect can be determined as influencing. This comparing analysis was done by using the degree of freedom that is available when the resolution of the design is one less than the maximum value. Using the Pareto Principle we suggest a more accurate threshold estimator for significantly influencing factors and/or interactions based on analysis of variance.

Keywords: design of experiments; significant effect; sum of squares; factorial design; experimental error

1 Introduction

DOE (Design of Experiments) [1] provides a powerful means to achieve breakthrough improvements in product quality and process efficiency [8, 6]. From the viewpoint of manufacturing fields, DOE can reduce the number of required experiments when taking into account the numerous factors or treatments possibly affecting experimental results. DOE strives to carry out the fewest number of experiments while maintaining the most important information. When there are no replications the experimental error cannot be measured without additional assumptions. Consequently the traditional methods based on analysis of variance (ANOVA) are useless without these additional assumptions. The traditional solution to this problem is the construction of normal plots of the estimated effects [4] in case of the specific type of designs where all of the effects are orthogonal to each other (and hence can be estimated separately) . Those that depart from linearity by a substantial distance are determined as

Table 1: k main factors ($2 \leq k \leq 4$) and their interactions lead to a design with m columns or possibly active effects.

		main	interactions		
k	m	effects	1-way	2-way	3-way
2	3	A B	AB		
3	7	A B C	AB AC BC	ABC	
4	15	A B C D	AB AC AD BC BD CD	ABC ABD ACD BCD	ABCD

significant. In order to eliminate the subjectivity of this graphical methods for detecting significant effects, other methods were proposed, such as Lenth's method [7] and Dong's method [5]. Here a sharp distinction between the significant and non-significant effect is provided, based on the pseudo standard error, as explained in section 3.1 and section 3.2. The statistical software program Minitab [12] uses Lenth's threshold to distinguish the significant factors and/or their interactions.

Our contribution consists in evaluating the quality of Lenth's and Dong's threshold. In section 3.3 we elaborate an alternative based on the Pareto Principle, i.e. a large proportion of process variation is associated with a small proportion of the process variables [3]. This assumption of effect sparsity makes it possible to use the highest order interaction to estimate the experimental error, as it has the smallest probability to be active. The observations at the different levels of that inactive highest order interaction may be regarded as replicates at each combination of levels of the other, possibly active factors. This alternative threshold is numerically evaluated in section 4 for a management example. The numerical results show that a larger threshold than the estimates of Lenth and Dong should be used to separate the inactive from the active factors and/or their interactions.

2 Full factorial and fractional factorial screening designs

A popular design when screening a large amount of factors, uses only two levels (low and high) for each factor, assuming that there is no curvature in the range of the factor of interest. In a 2^k-design where k is the number of factors, $m = 2^k - 1$ possibly active effects of factors (=main effect) and their interactions of different orders are under consideration (see Table 1). Here 2^k refers to the number of possible combinations of factor settings. When analyzing factorial experiments with replicated measurements at each treatment combination, an estimate of internal variance or noise may be calculated from the replicates at each treatment combination. The most important contribution of DOE is modeling the dependency of the response variable on the active factors and/or interactions with a limited number of experiments. Full factorial designs include all factors and their interactions, while fractional factorial designs [2]

use confounding patterns that express the relation between the effects. The shorter the defining relations or confounding patterns, the lower the resolution of the design. This makes that we have to carry out fewer fractional factorial experiments compared to a full factorial experiment. Famous fractional factorial designs are Plackett and Burman designs [9], Taguchi designs [10, 11] and others. Taguchi designs use orthogonal arrays as design schedules and also, the optimal conditions are determined by analyzing the S/N ratio (Signal-to-Noise ratio) as a performance measure.

With fractional factorial designs a compromise between experimental effort and results is chosen. The number of trials n varies between $k+1$ and 2^{k-1} (half of a full factorial design, so one replication for each setting of factors), with k the number of factors. The minimal value of $n = k + 1$ is used by Taguchi's designs and de Plackett-Burman (PB) designs. These saturated designs are very useful for economically detecting large main effects, assuming all interactions are negligible when compared with the few important main effects.

3 Thresholds in a Pareto plot

A Pareto plot (see Figure 1) is a graphical overview of the process factors and/or interactions of influence, in ranking order of the most influencing down to the least influencing. A threshold separates the influencing and non-influencing ones.
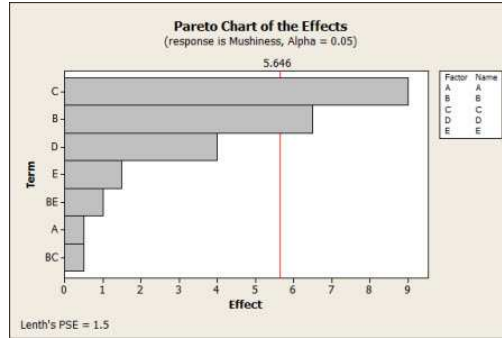

Figure 1: a Pareto plot as produced by Minitab

3.1 Lenth's Pareto threshold

Lenth's pseudo standard error (PSE) [7] is used for calculating a threshold or critical value $E_c^{[L]}$ for the effects when there are no replicates. It is based on the fact that the standard deviation of a sample from a Gaussian $N(0, \sigma)$ distribution may be estimated as $M = 1.5 \times$median(absolute values of the E_i-values). Here E_i ($1 \leq i \leq m$) is the ith effect or the difference of the mean output

value at the higher level of the factor/interaction and the mean output value at the lower level of the factor/interaction. A refinement is made by deleting effects that exceed 2.5 times this estimate and by recomputing the median. The critical value for the effects associated with a confidence of $(1-\alpha)\,100\%$ and expressed in (1), equals the PSE multiplied by the critical value for a t-distribution with $m/3$ degrees of freedom, where m is the number of effects being assessed.

$$E_c^{[L]} = \text{PSE} \cdot t_{1-\alpha/2}(m/3) \qquad \text{with } P(u > t_{1-\alpha/2}(m/3)) = 1 - \alpha/2, \ u : t(m/3) \tag{1}$$

3.2   Dong's Pareto threshold

Just like Lenth, Dong [5] used an adaptive standard error based on the trimmed mean of squared contrasts rather than the trimmed median of the unsigned contrasts:

$$E_c^{[D]} = \sqrt{\frac{1}{m_{inactive}} \sum_{|E_i| < 2.5\,M} E_i^2} \cdot t_{(1+0.98^{1/m})/2}(m_{inactive}) \tag{2}$$

where E_i is the ith calculated effect and $m_{inactive}$ is the number of effects that exceed 2.5 times M with M defined in section 3.1.

3.3   The Pareto threshold based on ANOVA

Let X be a factor or interaction included in a design, where only two levels ($+$ and $-$) are examined. The effect E_X of X defined by (3) as the difference between the mean outcome for X at higher and lower level respectively, is directly related to the sum of squares SSX, as shown in (4). Here a balanced design is assumed, so exactly half of the sample values are measures for X at level $+$.

$$E_X = \bar{y}_+ - \bar{y}_- \tag{3}$$

$$
\begin{aligned}
SSX &= \frac{n}{2}\left((\bar{y}_+ - \bar{y})^2 + (\bar{y}_- - \bar{y})^2\right) \\
&= \frac{n}{2}\left((\bar{y}_+ - \frac{\bar{y}_+ + \bar{y}_-}{2})^2 + (\bar{y}_- - \frac{\bar{y}_+ + \bar{y}_-}{2})^2\right) \\
&= \frac{n}{4}(\bar{y}_+ - \bar{y}_-)^2 \\
&= \frac{n}{4}E_X^2 \tag{4}
\end{aligned}
$$

SSX is a traditional tool to construct the F-value (5) with a $F(1, df_E)$-distribution,

$$F_X = \frac{SSX/1}{SSEr/df_{Er}} \tag{5}$$

4

Table 2: Factors assessed in a study on revenue per year

Factor	Low(-)	High(+)
A: quality control	no	yes
B: packing price	10	12
C: marketing budget	50	80

used to decide whether the nulhypothesis

H_0 : the factor/interaction X has no significant influence on the outcome y

is accepted or rejected with a confidence of $100\,(1-\alpha)\%$.

Here $SSEr = \sum_{i=1}^{n/2}(y_{i,+} - \bar{y}_+)^2 + \sum_{i=1}^{n/2}(y_{i,-} - \bar{y}_-)^2$ represents the residual sum of squares with $df_{Er} = 1$ degree of freedom.

The factor or interaction X is statistically considered as being significantly influencing if $F_X > F_\alpha(1, df_{Er})$, where $P(v > F_\alpha(1, df_{Er})) = \alpha$ with v : $F(1, df_{Er})$. This is equivalent to (6) when (4) and $F_\alpha(1, \nu) = (t_{1-\alpha/2}(\nu))^2$ are taken into account.

$$E_X > 2\sqrt{\frac{SSE}{n\,df_{Er}}}\ t_{1-\alpha/2}(df_{Er}) = E_c^{[F]} \tag{6}$$

The degrees of freedom are noticed between brackets for the F-distribution as well as for the t-distribution. $E_c^{[F]}$ is the improved pareto plot threshold for X to be significantly influencing.

4   Numerical example

As part of building an underlying business model, the effects of changing three process input factors on revenue per year were studied in a 2^3 experiment. The factors included in the study, along with their experimental levels, are explained in Table 2.

A design with 3 factors, 6 possibly active effects, resolution 3 and 8 runs as described in Table 3 at the right, is used to determine $E_c^{[F]}$ where the interaction column ABC is omitted compared to the full factorial design. Figure 2 and Figure 3 give a scatter plot of Lenth's threshold and Dong's threshold respectively (horizontally) and our threshold $E_c^{[F]}$ (vertically) for the effects when 100 trials are used. In the experimental setup for one factor an effect of magnitude $35\,\sigma$ is imposed. A successrate of more than 85% is reached for both, defined as the ratio of the number of trials where the correct active effect is detected over the total number of trials.

Figure 2 and Figure 3 suggest that Lenth's threshold is often too tolerant and should be refined. This means that also in cases where no alternative threshold

Table 3: Taguchi's method L_4 or geometric 3-factor PB design (left) and a fractional factorial resolution 3 design with 3 factors (right)

	factor					factor		interaction		
run	A	B	C	run	A	B	C	AB	AC	BC
1	+	+	+	1	+	+	+	+	+	+
2	+	-	-	2	+	+	-	+	-	-
3	-	+	-	3	+	-	+	-	+	-
4	-	-	+	4	+	-	-	-	-	+
				5	-	+	+	-	-	+
				6	-	+	-	-	+	-
				7	-	-	+	+	-	-
				8	-	-	-	+	+	+

estimation method is available due to lack of replicates (such as the Taguchi L_4 method or the geometric PB design as described in Table 3 at the left), Lenth's method assigns too much effects as being significant.
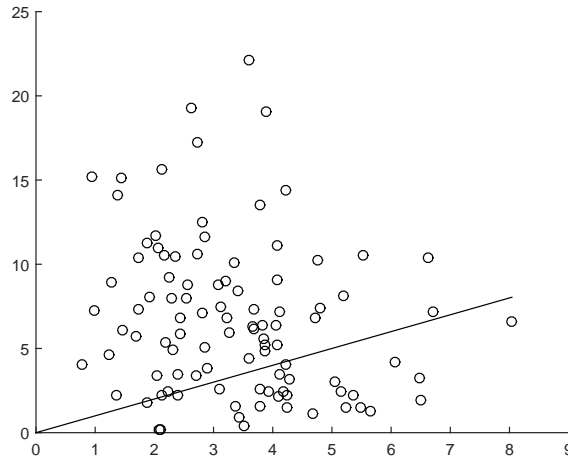

Figure 2: Scatter plot of the points $(E_c^{[L]}, E_c^{[F]})$ for 100 trials with effect=$35\,\sigma$.

Table 4 shows the success rate for 100 samples for different values of the maximum effect E and learns that the more separated the distribution of the largest effect from a zero effect, the higher the success rate of Lenth's threshold.
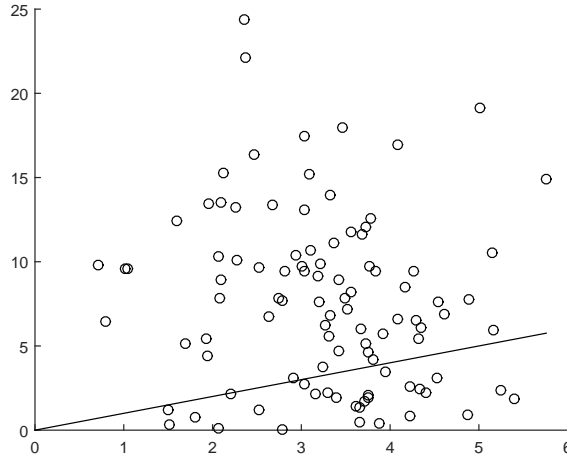
Figure 3: Scatter plot of the points $(E_c^{[D]}, E_c^{[F]})$ for 100 trials with effect$=35\,\sigma$.

Table 4: Success rate of Lenth's and Dong's threshold for different values of the maximum effect size

$E =$	$10\,\sigma$	$20\,\sigma$	$30\,\sigma$	$40\,\sigma$	$50\,\sigma$
Lenth's success rate=	60%	82%	85%	87%	88%
Dong's success rate=	64%	83%	86%	89%	90%

5 Conclusion

The thresholds of the pareto plot designed by Lenth and Dong were partly determined ad hoc and appear to be too tolerant: they notice too easily factors and/or interactions as significantly influencing. When no replications are supplied in the experiment, we suggest an improved threshold estimate by omitting the highest order interaction. This gives the opportunity to evaluate the experimental error, which is a crucial element in the estimation of the pareto plot threshold.

References

[1] ANTONY, J. (2003). *Design of Experiments for Engineers and Scientists*, Elsevier Science and Technology Books, Oxford.

[2] BOX, G., HUNTER, W. AND HUNTER, J. (1978). *Statistics for Experimenters*, Wiley-Interscience, New York.

[3] Box G. AND Meyer, R. (1986). An analysis for unreplicated fractional factorials. *Technometrics* **28,** 11–18.

[4] Daniel, C. (1959). Use of Half-normal plots in interpreting factorial two-level experiments. *Technometrics* **1,** 311–341.

[5] Dong, F. (1993). Asymptotic properties of quantiles for truncated and contaminated data. *Comm. Statist. Theory Methods* **22,** 3255–3261.

[6] Goujot, D., Meyer, X. and Courtois, F. (2012). Identification of a rice drying model with an improved sequential optimal design of experiments. *Journal of Process Control* **22 (1),** 95–107.

[7] Lenth, R. (1989). Quick and Easy Analysis of Unreplicated Factorials, *Technometrics* **31 (4)**, 469–473.

[8] Munoz-Tamayo, R., Martinon, P., Bougaran, G., Mairet, F. and Bernard, O. (2014). Getting the most out of it: Optimal experiments for parameter estimation of microalgae growth models. *Journal of Process Control* **24 (6)**, 991–1001.

[9] Plackett, R. and Burman, J. (1946). The Design of Optimal Multifactorial Experiments. *Biometrika* **33**, 305–325.

[10] Roy, R. (2001). *Design of experiments using the Taguchi Approach*, Wiley-Interscience Publication, John Wiley & Sons.

[11] Taguchi, G. and Yokoyama, K. (1993). *Taguchi Methods Design of Experiments*, Quality Engineering Series 4, American Supplier Institute Press.

[12] http://www.minitab.com/ accessed on 1/9/2015