eqr014

Lenth's Method for the Analysis of Unreplicated Experiments

Russell V. Lenth

Department of Statistics and Actuarial Science

The University of Iowa

Iowa City, IA   USA 52242

Voice 319-335-0814   FAX 319-335-3017

russell-lenth@uiowa.edu*

June 30, 2006

Keywords   Lenth's method, Unreplicated factorial experiments, Screening experiments, Saturated model.

*File names: eqr014.tex and eqr014.pdf; prepared using pdfLaTeX.

Abstract

Lenth's method is an objective method for deciding which effects are active in the analysis of unreplicated experiments, when the model is saturated and hence there are no degrees of freedom for estimating the error variance. The article reviews its formulation and computation and examples, and discusses the choice of critical values and presentation and interpretation of the results.

1 Introduction

Consider a 2^{7-4} experimental design (*see* eqr007) involving 7 factors, each at 2 levels, and only 8 experimental runs. This is an example of a saturated experiment, in that we must estimate 8 parameters (7 main effects plus the mean), leaving no degrees of freedom for error.

The analysis of such experiments has historically been rather subjective. A popular method has been that of Daniel [1] (*see* eqr009), whereby the absolute values of the 7 main-effect estimates are plotted against half-normal scores (i.e., equally-spaced quantiles of the half-normal distribution). If all the true main effects are zero, these points should fall near a straight line; however, if there is a small number of "active" (nonzero) effects while the remaining ones are inactive, the inactive ones should follow a straight line and the active ones should look like outliers.

The half-normal-plot method requires one to make a judgment about what is an outlier and what is not. In an effort to remove this element of subjectivity, Lenth [2] proposed a method for estimating a standard-error-like quantity, called the pseudo standard error or PSE. The effects are then judged relative to the PSE to decide whether there is enough evidence for them to be deemed active.

Both the Daniels method and the Lenth method rely on the concept of "effect sparsity"—that only a few effects are active and the rest are inactive. This concept is discussed more rigorously in Box and Meyer [3], along with another objective method for analyzing such experiments.

2 The method

Lenth's method assumes that we have m independent contrast or effect estimates, and that they all have the same variance τ^2. In an orthogonal two-level experiment on N observations, for example, each contrast has the form $c = \bar{y}_+ - \bar{y}_-$, where $\bar{y}_+$ is the average of the $N/2$ observations at the "high" level of a factor and $\bar{y}_-$ is the average of the $N/2$ observations at the "low" level. Each such contrast has variance $\tau^2 = 4\sigma^2/N$, where σ^2 is the error variance. Let $c_1, c_2, \ldots, c_m$ denote the m contrast estimates. Usually, owing to the fact that the model is saturated, we have $m = N - 1$.

The computation involves two stages. First, let

$$s_0 = 1.5 \cdot \text{median}\ \{|c_j|\}$$

Then, let

$$\text{PSE} = 1.5 \cdot \text{median}\ \{|c_j| : |c_j| \leq 2.5 s_0\}$$

Note that the second step is just like the first, except that we exclude those effects that exceed $2.5s_0$ in absolute value. PSE is termed the pseudo standard error, and it is an estimate of τ.

$<$ Table 1 near here $>$

Once the PSE is obtained, one can multiply it by a factor from Table 1 to obtain a margin of error (ME) for the contrasts. Contrasts that exceed the ME in absolute value are deemed active at the 5% significance level. Alternatively, one may compute a simultaneous margin of error (SME) using a different factor from Table 1. The distinction is that there is at most a 5% chance that one individual inactive contrast will exceed the ME, while there is at most a 5% chance that *any* inactive contrast will exceed the SME.

The recommended practice is to display the contrasts in a bar chart or a Pareto chart, with decision lines added for the ME and the SME (see the example below). A contrast that extends beyond the SME line is clearly and active effect. One that exceeds the ME but not the SME should be viewed with some caution, as it may be an artifact of testing several contrasts.

Some software implementations of Lenth's method produce a half-normal plot of the contrasts with a reference line having a slope based on the PSE; however, this is a poor alternative to the bar chart and is not recommended because it does not show the decision thresholds.

The values in Table 1 are excerpted from tables in Ye and Hamada [4] that cover several additional values of m and significance levels other than .05.

3 Example

Box et al. [5] present data from an 8-run experiment involving 7 2-level factors. The response variable is the time it takes to climb a particular hill on a bicycle, and the factors are such things as seat height, dynamo off or on, etc. The data are displayed in Table 2. The factor levels shown in the headings correspond to the respective labels "−" and "+" in the table. The final row gives the contrast estimates c_j. (To improve clarity, the factor levels are altered from those in the reference.)

$<$ Table 2 near here $>$

The calculations are as follows. The median of the absolute contrasts $\{3.5, 12, 1.0, 22.5, .5, 1.0, 2.5\}$ is 2.5; hence, $s_0 = 1.5 \times 2.5 = 3.75$ and $2.5s_0 = 9.375$. Since 2 of the 7 contrasts exceed this bound, we compute the PSE from the remaining 5: $\text{PSE} = 1.5 \cdot \text{median}\,\{3.5, 1.0, .5, 1.0, 2.5\} = 1.5 \times 1 = 1.5$. Now, using Table 1 with $m = 7$, we have $\text{ME} = 2.297 \times 1.5 = 3.45$ and $\text{SME} = 4.867 \times 1.5 = 7.30$. Figure 1 shows a bar chart of the contrasts with reference lines added for $\pm ME$ and $\pm SME$.

< Figure 1 near here >

The contrasts for Gear and Dynamo exceed the SME, and so they are identified as active effects. Both are positive effects, meaning that the higher gear ratio and the presence of the dynamo both increase the expected time to climb the hill. The contrast for Seat barely surpasses the ME in the negative direction, providing a hint that having the seat up might reduce the time. However, this contrast does not surpass the SME, which means that if we adjust for multiple testing, it is not identified as active.

4 Discussion

Subsequent to the publication of [2], it has been found that the critical values suggested there are too conservative; hence, the tables in [4] and excerpted in Table 1 are preferable. These critical values are based on the complete null case—where all the contrasts are assumed inactive. When one or more of the contrasts are actually active, that tends to inflate PSE somewhat, making the procedure somewhat conservative. It grows more conservative as the number of truly active effects increases, and is most conservative when the magnitude of the active effect is only moderate (less than 2τ or so). (Very large active effects lead to large contrasts that have a very high chance of exceeding $2.5s_0$ and thus being discarded in the second step of computation.)

The main advantage of Lenth's method is its simplicity and ease of use, making it easy to incorporate in programs or spreadsheet computations, or even to do by hand. Its main disadvantage is that it is based on a restrictive set of assumptions: independent contrast estimates that all have the same variance.

Lenth's method should not be used in replicated experiments, where the standard error can be estimated more efficiently by traditional analysis-of-variance methods.

There are other objective methods that can be used instead of Lenth's method. Most notable among these is the Bayesian method of Box and Meyer [3], which (unlike the Lenth method) is extensible to a broad class of regression variable-selection problems (see [6]). There is also a formulation in [7] of the Box-Meyer method that analyzes which factors (rather than which effects) are active. A survey and Monte Carlo comparison of several methods is given in [8].

5 References

1. Daniel C. Use of half-normal plots in interpreting factorial two-level experiments. Technometrics 1959 1: 311-340.

2. Lenth R. Quick and easy analysis of unreplicated factorials. Technometrics 1989 31: 469–473.

3. Box G, Meyer R. An analysis for unreplicated fractional factorials. Technometrics 1986 28: 11–18.

4. Ye K, Hamada M. Critical values of the Lenth method for unreplicated factorial designs.

Journal of Quality Technology 2000 32: 57–66.

5. Box G, Hunter J, Hunter W. Statistics for Experimenters (2nd ed). Wiley: New York, 2005; p 245.

6. Chipman H, Hamada M, Wu C. A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. Technometrics 1997 39: 372–381.

7. Box G, Meyer R. Finding the active factors in fractionated screening experiments. Journal of Quality Technology 1993 25: 94–105.

8. Balakrishnan N, Hamada M. Analyzing unreplicated factorial experiments: a review with some new proposals. With discussions. Statistica Sinica 1998 8: 1–41.

Table 1: Multipliers to obtain the ME and SME at the $\alpha = .05$ level.

m	for ME	for SME	m	for ME	for SME
7	2.297	4.867	19	2.120	4.118
8	2.201	4.868	23	2.097	4.017
11	2.211	4.438	26	2.082	3.985
15	2.156	4.240	27	2.077	3.964
17	2.138	4.164	31	2.064	3.925

Table 2: Data from the bicycle experiment.

Time (sec)	Seat down/up	Dynamo no/yes	Handlebars down/up	Gear low/med	Raincoat off/on	Breakfast no/yes	Tires soft/hard
50	−	−	−	−	−	+	+
52	−	−	+	−	+	−	−
88	−	+	−	+	−	−	−
83	−	+	+	+	+	+	+
71	+	−	−	+	+	+	−
69	+	−	+	+	−	−	+
59	+	+	−	−	+	−	+
60	+	+	+	−	−	+	−
Contrast:	−3.5	12.0	−1.0	22.5	−0.5	−1.0	−2.5

Figure 1: Bar chart of the contrast estimates for the bicycle example.