# Inferences for Saturated Simplex Designs

A. A. Al-Shiha
*King Saud University, Department of Statistics and Operations Research*
*College of Science, P.O.Box 2455, Riyadh 11451,*
*Riyadh , Saudi Arabia*
*E-mail: aalshiha@ksu.edu.sa*

**Abstract**
A robust procedure is described for examining the significance of the effects in saturated simplex designs. The procedure described in this paper is motivated by the procedure used for analyzing unreplicated factorial experiments proposed by Aboukalam and Al-Shiha(2001).

**1. Introduction:**

Experiments used in many fields of applied sciences are characterized by the tendency of including a large number of potential factors that might affect the response. Accurate methods accommodating a large number of factors with as few as possible experimental units are usually desired. If the assumption of no interactions exist between the factors, non-replicated orthogonal designs such as simplex designs can be efficiently used to detect significant factors.

Let $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ be the general first-order linear model in $p$ factors where $\mathbf{y}$ is an $n\times1$ vector of responses, $\beta^t=(\beta_0, \beta_1, \ldots, \beta_p)$ is an $n\times1$ vector of unknown parameters, $\varepsilon^t=(\varepsilon_1,\ldots,\varepsilon_n)$ is an $n\times1$ vector of random errors, and $\mathbf{X}$ is an $n\times n$ matrix of settings of the input factors. We assume that $\varepsilon\sim N_n(\mathbf{0},\sigma^2\mathbf{I}_n)$, where $\mathbf{0}$ is an $n\times1$ zero vector and $\mathbf{I}_n$ is the identity matrix of dimension $n$. Consequently, $\mathbf{y}\sim N_n(\mathbf{X}\beta ,\sigma^2\mathbf{I}_n)$. The number of observations, $n$, in a saturated design is equal to the number of parameters in the model, i.e. $n=p+1$. Therefore, the error variance has zero degrees-of-freedom and it cannot be independently estimated. As a consequence, the usual $F$-type tests of significance of the regression coefficients can not be performed.

For the saturated orthogonal design, we have $\mathbf{X}^t\mathbf{X}=diag(b_0, b_1, \ldots, b_p)$ and the best linear unbiased estimate (*BLUE*) of $\beta$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y}$. Based on the assumption given earlier, it can be shown that $\hat{\boldsymbol{\beta}} \sim N_{p+1}(\boldsymbol{\beta},\sigma^2 diag(b_0^{-1},b_1^{-1},\ldots,b_p^{-1}))$, and that $\hat{\beta}_0$ is nothing but the sample mean of the observed responses, $\bar{y}$, whereas the remaining elements $\hat{\beta}_1,\ldots$, and $\hat{\beta}_p$ are estimates of the effects of factors. Now, define the vector of factors' coefficients to be $\theta'=(\theta_1, \theta_2,\ldots, \theta_p)$, where $\theta_i = \beta_i$ ($i$=1, 2, ..., $p$). Consequently, the *BLUE* of $\theta$ is $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_p) = (\hat{\beta}_1,\ldots,\hat{\beta}_p)$. Moreover, $\hat{\boldsymbol{\theta}} \sim N_p(\boldsymbol{\theta},\sigma^2\mathbf{D})$ where $D = diag(b_1^{-1},b_2^{-1},\ldots,b_p^{-1})$. When the design is chosen (or transformed) in such a way that the lengths of columns of the factors are all equal, i.e., $b_1= \ldots =b_p=b$, then $\hat{\boldsymbol{\theta}} \sim N_p(\boldsymbol{\theta},\tau^2\mathbf{I}_p)$ where $\tau^2=\sigma^2/b$. The statistical problem is to identify significant factors that affects the response. Therefore, our objective is to infer, if any, of the means $\theta_1, \theta_2, \ldots, \theta_p$ significantly differs from zero. Under the null hypothesis $H_O$: $\theta_1= \theta_2 = \ldots = \theta_p = 0$, the set of observations $\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_p$ is a random sample from the normal distribution $N(0,\tau^2)$. Using an appropriate estimate of the scale $\tau$, say $S$, the individual hypothesis $H_O$: $\theta_j=0$ ($j$=1, 2, ..., $p$) is rejected if the absolute standardized estimate $|\hat{\theta}_j|/S$ is large, as described by Aboukalam and Al-Shiha (2001). Since $\hat{\boldsymbol{\theta}} \sim N_p(\boldsymbol{\theta},\tau^2\mathbf{I}_p)$, we suggest using the procedure of Aboukalam and Al-Shiha (2001) for examining the significance of the parameters in a saturated orthogonal design for a first-order model.

**2. Analyzing Saturated Simplex Designs:**

In the screening stage, a large number of factors are considered. Sometimes, and based on past experience, it is believed that no interactions exist between the factors of interest. The simplex designs are very helpful in such cases. Researchers can test for the significance of the different factors using few observations when they utilize a simplex design.

The saturated simplex design is an orthogonal design. If $p$ factors are involved in the first-order model, then the design consists of $n=p+1$ design points given by the vertices of a regular $p$-dimension simplex, which for $p=2$ is an equilateral triangle, for $p=3$ is a tetrahedron, etc.

Box (1952) gave the procedure for the construction of a simplex design in $k$ dimensions. According to Khuri and Cornell (1987), an alternative procedure for constructing the setting matrix **X** of an $n$-point simplex design corresponding to $p=n-1$ factors is described as follows. Let $\mathbf{X} = \sqrt{n}\ \mathbf{P}$ = $\sqrt{n}\left(\frac{1}{\sqrt{n}}\mathbf{j}_n,\ \ \mathbf{p}_1,\ \ \Lambda\ ,\ \ \mathbf{p}_p\right)_{n\times n}$ , where **P** is an $n\times n$ orthogonal matrix with equal elements in its first column and $\mathbf{j}_n$ is an $n$ vector of ones. It is clear that $\mathbf{X}^t\mathbf{X}=n\mathbf{I}_n$. Therefore, the *BLUE* of b is $\hat{\mathbf{a}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y} = \frac{1}{\sqrt{\mathrm{n}}}\mathbf{P}^t\mathbf{y}$. Equivalently, $\hat{\mathbf{a}}^t = \left(\bar{y},\ \ \frac{1}{\sqrt{n}}\mathbf{p}_1^t\mathbf{y},\ \ \Lambda\ ,\ \ \frac{1}{\sqrt{n}}\mathbf{p}_p^t\mathbf{y}\right)$. Consequently, $\hat{\mathbf{e}} = \left(\frac{1}{\sqrt{n}}\mathbf{p}_1^t\mathbf{y}, \frac{1}{\sqrt{n}}\mathbf{p}_2^t\mathbf{y}, \Lambda\ ,\ \frac{1}{\sqrt{n}}\mathbf{p}_p^t\mathbf{y}\right)$. The first element of $\hat{\mathrm{b}}$, $\hat{\beta}_0$, is simply the mean of the observed responses, whereas the remaining elements $\hat{\beta}_1, \Lambda$ , and $\hat{\beta}_p$ are estimates of the main effects of the factors. It can be shown that $\hat{\mathbf{a}} \sim N_{p+1}\left(\hat{\mathbf{a}}, \tau^2\,\mathbf{I}_{p+1}\right)$ and $\hat{\mathbf{e}} \sim N_p(\hat{\mathbf{e}}, \tau^2\,\mathbf{I}_p)$, where $\tau^2=\sigma^2/n$. In this case, $b=n$. Consequently, we can directly apply the method proposed by Aboukalam and Al-Shiha (2001) to test for significance of the parameters $\beta_1, \ldots, \beta_p$ in a simplex design.

## 3. Comments and Concluding Remarks:

It is possible that any test in the literature designed for detecting significant effects in unreplicated $2^p$ factorials can be used as a tool for detecting significant factors in an unreplicated simplex design. One of the most popular tests is that of Lenth (1989). Aboukalam and Al-Shiha (2001) have proposed a simple and robust test that outperforms that of Lenth. Therefore, we have adopted it in our paper.

In screening stages, a large number of factors are usually included in the model. In order to include as small number of factors as necessary in the subsequently refinement stages, we recommend that small values of the significance level, $\alpha$, such as $\alpha=0.01$, are used.

## REFERENCES


Aboukalam, M. A. F. and Al-Shiha, A. A. (2001). A robust analysis for unreplicated factorial experiments. Computational Statistics and Data Analysis 36, 1, 31-46.

Box, G. E. P. (1952). Multi-factor designs of first order. Biometrika 39, 49-57.

Khuri, A. I., and Cornell, J. A. (1987). Response Surfaces Designs and Analyses. New York: Marcel Dekker, Inc.

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


## RESUME


*Une robuste procedure est décrite pour examiner la signification des effects dans des desseins simplexes saturés. La procedure décrite dans cet article est motivée far la procedure utilisée pour analyser des desseins factorials saturées, proposée par Aboukalam and Al-Shiha(2001).*