ALTERNATIVE APPROACH FOR ESTIMATING E(S^2) VALUES FOR SUPERSATURATED DESIGNS

J. I. Mbegbu[1] and C. O. Todo[2]

[1]Department of Mathematics, University of Benin, Benin City, Nigeria

[2]Department of Statistics, School of Computing and IT, Delta State Polytechnic, Otefe-Oghara, Nigeria

ABSTRACT

The $E(s^2)$ criterion measures the average correlation among the columns of the design matrix for supersaturated designs (SSDs). In this light, Bulutoglu and Cheng [3] constructed some SSD's, and then obtained their corresponding values $E(s^2)$ by computational method. In this paper however, we estimated the values of $E(s^2)$ by a regression method using existing SSDs. The results obtained were in line with the results of Bulutoglu and Cheng and the lower bound derived by Nguyen [6] and Tang and Wu [7] were satisfied.

Keywords: $E(s^2)$ criterion, supersaturated designs, regression, lower bound.

Correspondence: Prof. J. I. Mbegbu, Department of Mathematics University of Benin, Benin City, Nigeria Julian.mbegbu@yahoo.com.

1. INTRODUCTION

Let n be the number of runs in a two-level experiment in m factors, then the design is called a saturated design (SD) if $m = n - 1$[1]. An SD is called a supersaturated design (SSD) if $m > (n-1)$. An optimal design is one which is good in terms of some meaningful statistical criterion or a family of criteria with respect to a given problem [2].

The design and analysis of supersaturated designs rely on the effect of sparcity principle, that is, when the number of relatively important factors in a factorial experiment is small [3]. Again, according to Xu and Wu [4], the popular criterion in the literature as a measure of goodness or for comparing SSD is the $E(s^2)$ criterion. $E(s^2)$ criterion, which was proposed by Booth and Cox [5] for the construction of a two-level supersaturated design, measures the average correlation among the columns of the design matrix of an SSD. In the design with n run and m factors, each of the factors (columns) has 2 levels and we require that $\frac{n}{2}$ of the entries in each column be +1 and the others -1. In the literature, most of the values of $E(s^2)$ are obtained directly from constructed SSD by the use of computational method.

However, in this research work, we shall view this problem through the method of regression. This is achieved using existing SSD's constructed by Bulutoglu and Cheng [3] and then estimating the value of $E(s^2)$ by method of regression rather than computational method.

2. MATERIALS AND METHODS

Explicit computational method to obtain values of $E(s^2)$ for supersaturated designs, SSD's: Let x denote the incidence matrix of the SSD (n, m), then $X = \lfloor a_{ij} \rfloor_{nxm}$.

$$[S_{ij}]_{m \times m} = X^T X \tag{2.1}$$

where $i, j = 1, 2, \cdots, m$.

Evidently [3]

$$E(s^2) = \frac{\sum_{i<j} s_{ij}^2}{\binom{m}{2}} \tag{2.2}$$

where s_{ij} is the value of the entry at the ith row and jth column of $X^T X$.

In order to compare SSDs that have the same number of runs and the same number of factors but different values of $E(s^2)$, we minimize $E(s^2)$. For a two-level supersaturated design,

$$s_{ij} = \begin{cases} 1, \text{ if factor j occurs in the high level in run i} \\ \text{-1 factor j occurs in the low level in run i} \end{cases}$$

Theorem 2.1: Bulutoglu and Cheng [3]: Suppose $n-1$ is an odd prime power, q is an even divisor of $n-2$ with $q \neq n-2$, x is a primitive element of $GF(n-1)$ *and* T is a subset of z_q of size $\frac{q}{2}$. Let e be the smallest positive integer such that $T+e=T$. Then, the $e(n-1)$ sets $\{S_{r,a} : r=0,\cdots,e-1, a \in GF(n-1)\}$,

where $S_{r,a} = \{x^{iq+I} + a : 0 \leq j \leq \frac{(n-2)}{q} - 1, \ i \in T+r\}$ are distinct and constitute a balanced incomplete block design $(n-1, e(n-1), (n-2)/q-1)$.

Furthermore, if $(n-2)/q$ is odd and U is a subset of size $\frac{e}{2}$ of $\{0,\cdots,e-1\}$ such $U^* = U + \frac{q}{2}$, where U^* is the complement of U in $\{0,\ldots,e-1\}$ and the addition is reduced modulo q, then the $e(n-1)/2$ sets $\{S_{r,a} : r \in U, a \in GF(n-1)\}$ constitute a balanced incomplete block design with distinct blocks.

By computational method (2.2), Bulutoglu and Cheng [3] obtained the corresponding values of $E(s^2)$ for some existing supersaturated designs (SSD's) shown in Table1.

Table 1: Values of $E(s^2)$ for some existing supersaturated designs (SSD's)

Supersaturated Designs	Computational Value of $E(s^2)$ [3]
SSD (10, 14)	5.0549
SSD (10, 15)	5.5238
SSD (14, 17)	4.9412
SSD (14, 18)	5.6732
SSD (14, 19)	6.0585

Regression method to estimate the values of $E(s^2)$ for supersaturated designs (SSD's)

We shall use the existing SSD constructed by Bulutoglu and Cheng [3] to estimate the values of $E(s^2)$ for supersaturated designs.

However, this is achieved by using regression method rather than computational method. We define the regression equation as:

$$y = a + bx \tag{2.1.1}$$

where $a = \bar{y} - b\bar{x}$ (2.1.1a)

and $b = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ (2.1.1b)

The regression equation was implemented on a computer using the FORECAST syntax, FORECAST (x, known_ y's, known_x's). The equation shall be used in the following way:

The dependent variable y is replace by $E(s^2)$, while the independent variable x is replace by a combination of $n \text{ and } m$. Many combinations of $n \text{ and } m$ were considered, such as $(n+m),\ n, m, (n-m),\ (m-n),\ nm,\ \frac{n}{m},\ \frac{m}{n},\ \frac{m}{n^2},\ \frac{n^2}{m},\ nm^2$. The accuracy of the combination was ascertained by using it to predict already known $E(s^2)$ values. The combination that gives the most accurate prediction of $E(s^2)$ is nm^2 and hence, it was chosen to be the value of the independent variable x. Now, in the regression equation (2.1.1), we shall use:

$$E(s^2) = a + b(nm)^2 \tag{2.1.2}$$

$$a = \bar{E}(s^2) - b\left(\overline{nm^2}\right) \tag{2.1.3}$$

$$b = \frac{\sum\left(nm^2 - (\overline{nm^2})\right)\left(E(s^2) - \bar{E}(s^2)\right)}{\sum\left(nm^2 - (\overline{nm^2})\right)^2} \tag{2.1.4}$$

3. RESULT AND DISCUSSION

Using regression method, we obtained the estimates of $E(s^2)$ for existing supersaturated designs (Table 2). These results were obtained by incorporating the n runs and m factors in existing SSDs and their corresponding computational values of $E(s^2)$ in equations (2.1.2), (2.1.3) and (2.1.4) respectively.

Table 2: Estimates of $E(s^2)$ for existing supersaturated designs

Existing-Supersaturated Designs	Computational Value of $E(s^2)$	Value of $E(s^2)$ Estimated by Method of Regression
SSD (10, 14)	5.0549	5.2696
SSD (10, 15)	5.5238	4.9731
SSD (14, 17)	4.9412	5.7158
SSD (14, 18)	5.6732	5.6067
SSD (14, 19)	6.0585	5.4058

Validity of the estimated value of $E(s^2)$**:** Comparing the values of $E(s^2)$ estimated by regression method with an $E(s^2)$ lower bound [6, 7]

$$E(s^2) \geq \frac{\left(m(n^2+n-1)-n^3\right)}{\left(n(m-1)\right)}$$

we obtain Table 3

Table 3: Comparison of the values of $E(s^2)$ estimated by regression method with an $E(s^2)$ lower bound

Existing-Supersaturated Designs	Value of $E(s^2)$ Estimated by Method of Regression	Established lower bound for $E(s^2)$: $\frac{\left(m(n^2+n-1)-n^3\right)}{\left(n(m-1)\right)}$
SSD (10, 14)	5.2696	4.0462
SSD (10, 15)	4.9731	4.5358
SSD (14, 17)	5.7158	3.6116
SSD (14, 18)	5.6067	4.2773
SSD (14, 19)	5.4058	4.8690

In Table 2, a list of estimated values of $E(s^2)$ by method of regression was shown. Table 3 shows that the values of $E(s^2)$ estimated by the method of regression satisfy the bound $E(s^2) \geq \frac{\left(m(n^2+n-1)-n^3\right)}{\left(n(m-1)\right)}$. In other words, the validity of the estimated values of $E(s^2)$ is guaranteed.

4. CONCLUSION

The estimated values of $E(s^2)$ for existing supersaturated designs obtained by regression approach is valid since these values satisfy an $E(s^2)$ lower bound [6]. Hence, we have established an alternative computation of $E(s^2)$-values by regression approach.

REFERENCES

[1] Nguyen, N. and Cheng, S. (2008). New $E(s^2)$-optimal supersaturated designs constructed from incomplete block designs. Technometrics, 50(1): 26-31.

[2] Mbegbu, J. I. (2007). Optimal balanced incomplete block designs with cost constraints. PhD Thesis, Rivers State University of Science and Technology, Port-Harcourt, Nigeria.

[3] Bulutoglu, D. A. and Cheng, C. (2004). Construction of $E(s^2)$-optimal supersaturated designs. The Annals of Statistics, 32 (4): 1662-1678.

[4] Xu, H. and Wu, C. F. J. (2005). Construction of optimal multi-level supersaturated designs. The Annals of Statistics, 33(6): 2811-2836.

[5] Booth, K. H. and Cox, D. R. (1962). Some systematic supersaturated designs. Technometrics, 4: 489-495.

[6] Nguyen, N. K. (1996). An algorithmatic approach for constructing supersaturated designs. Technometrics, 38: 67-73.

[7] Tang, B. X. and Wu, C. F. J. (1997). A method for constructing supersaturated designs and its $E(s^2)$-optimality. Canadian Journal of Statistics, 25: 191-201.