Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.2, No.9, 2012

IN SEARCH OF OPTIMAL ESTIMATION

K. A. Bhat and A. Ahmed

ABSTRACT

In this paper we introduce a method that can provides the best estimated value of missing observations in experimental design. We have used a different way for finding the missing observations. We use the concept of mathematical programming, we convert the statistical problem into mathematical programming problem. Numerical illustration is used to understand the technique.

Key Words: Missing observations, Mathematical Programming, LINGO software, Cycling process.

Introduction

The area of missing data and imputation in statistical models is certainly very important and is one of the most active areas of statistical research today. In 1933 Yates developed a result for missing observations using the general techniques of minimizing the error sum of squares with respect to the missing cell(s). Most of the text books on design of experiment uses that result [1,3]. After Yates many methods are developed to estimate the missing values, the use of the EM algorithms, MCMC stochastic substitution method and other more recent methods are ubiquitous in the discipline [2,5,6].

We have introduced a method in which we try to reduce all types of errors and variability in the model and then automatically the estimated value comes out is the best. We have used a different type of technique in which we use the concept of mathematical programming [4] to minimize the sum of squares of error under the conditions that the variability of sources of variations remain under given limits.

Estimation of Missing observation

Let us consider the model

$$Y = X\ \theta + \varepsilon$$

where $\theta = \mu, \tau_1 \ldots, b_1 \ldots, \ldots$ be the 'n' sources of variations.

Let there be 'k' missing observations in the model and σ_i^2 is the known variance of i^{th} source of variation without considering the missing observations.

Error sum of square (S.S.E) = Total sum of squares – Sum of squares due to sources of variations (S_i)

We introduced the mathematical programming as:

Minimizing f(x) = S.S.E.

Subject to

Variance (S_i) $\leq \sigma_i^2$

$x_i \geq 0$

We estimate the missing observations one by one and simultaneously use them in the model till the all missing observations can be estimated.

Once all missing observations are estimated use cycling, if the values of missing observations improves continue the procedure, otherwise stop the procedure.

Estimation of 'k' missing observations in model having two sources of vaiations.

The model is

$$Y = X\ \theta + \varepsilon$$

Let the sources of variations be S_1 and S_2 (say treatment and blocks) having 't' treatments and 'r' blocks in the design.

S.S.E = Total sum of squares – Treatment sum of square – Block sum of squares

S.S.E. = $\mathbf{x_i^2\{1- 1/t -1/r +1/rt\} - 2x_i\{ (Y_{.j})/t + (Y_{i.})/r - Y_{..}/rt \}}$ + constant with respect of x_i

$$\text{Variance (S1)} = 1/r\{ \sum(\mathbf{Yi.})^2 + x_i^2\} - \{1/r\ (\sum \mathbf{Yi.} + x_i)\}^2$$

$$\text{Variance (S2)} = 1/t\{ \sum(\mathbf{Y.}j)^2 + x_i^2\} - \{1/t(\sum \mathbf{Y.j} + x_i)\}^2$$

$$\text{Variance (Total)} = 1/rt\{ \sum(\mathbf{Yij})^2 + x_i^2\} - \{1/rt(\sum \mathbf{Yij} + x_i)\}^2$$

Now to estimate first missing observation the mathematical programming problem is

Minimize $f_1(x)$ = $\mathbf{x_1^2\{1- 1/t -1/r +1/rt\} - 2x_1\{ (Y_{.j})/t + (Y_{i.})/r - Y_{..}/rt \}}$

Subject to

$$1/r\{ \sum(Yi.)^2 + x_1^2\} - \{1/r\ (\sum Yi. + x_1)\}^2 \leq \sigma_1^2$$

$$1/t\{ \sum(Y.j)^2 + x_1^2\} - \{1/t(\sum Y.j\ + x_1)\}^2 \leq \sigma_2^2$$

$$1/rt\{ \sum(Yij)^2 + x_1^2\} - \{1/rt(\sum Yij\ + x_1)\}^2 \leq \sigma_T^2$$

$$x_1 \geq 0$$

where σ_1^2, σ_2^2, σ_T^2 are the first source variability, second source variability and the total variability.

After getting estimated value of 'x_1' use this value in model and proceed same to estimate the second missing observation, the mathematical programming problem for estimation of second missing variable is

Minimize $f_2(x) = x_2^2\{1- 1/t -1/r +1/rt\} – 2x_2\{ (Y_{.j})/t + (Y_{i.})/r - Y_{..}/rt \}$

Subject to

$$1/r\{ \sum(Yi.)^2 + x_2^2\} - \{1/r\ (\sum Yi. + x_2)\}^2 \leq \sigma_1^2$$

$$1/t\{ \sum(Y.j)^2 + x_2^2\} - \{1/t(\sum Y.j\ + x_2)\}^2 \leq \sigma_2^2$$

$$1/rt\{ \sum(Yij)^2 + x_2^2\} - \{1/rt(\sum Yij\ + x_2)\}^2 \leq \sigma_T^2$$

$$X_2 \geq 0$$

Here we get estimation of X_2, now start the cycling to improve the estimated values. If we get improved values then continue the procedure otherwise stop.

Example

The table below is the yields of 6 varieties in a 4 replicate experiment for which two observations are missing.

S_1/S_2	1	2	3	4	5	6	Total
1	*18.5*	*15.7*	*16.2*	*14.1*	*13.0*	*13.6*	**91.1**
2	*11.7*	X_1	*12.9*	*14.4*	*16.9*	*12.5*	**68.4**
3	*15.4*	*16.6*	*15.5*	*20.3*	X_2	*21.5*	**89.3**
4	*16.5*	*18.6*	*12.7*	*15.7*	*16.5*	*18.0*	**98.0**
Total	**62.1**	**50.9**	**57.3**	**64.5**	**46.4**	**65.6**	***346.8***

$\sigma_1^2(S_1) = 1.47$ and $\sigma_1^2(S_2) = 3.36$

Variance $(S_1) = \frac{1}{4}(868.01 + x_1^2) - \{\frac{1}{4}(50.9 + x_1)\}^2$

$= 55.1 - 25.45\, x_1 + 0.1875\, x_1^2$

Variance $(S_2) = 1/6\,(952.52 + x_1^2) - \{1/6\,(68.4 + x_1)\}^2$

$= 28.79 - 3.8\, x_1 + 0.1389\, x_1^2$

Variance (Total) $= 1/23\,(5606.02 + x_1^2) - \{1/23\,(346.8 + x_1)\}^2$

$= 16.39 - 1.31\, x_1 + 0.0416\, x_1^2$

E.S.S. $= x_1^2\{1 - 1/t - 1/r + 1/rt\} - 2x_1\{(Y_{.j})/t + (Y_{i.})/r - Y_{..}/rt\}$

$= 0.625\, x_1^2 - 19.35\, x_1$

Now the resulting mathematical programming is:

(P1) Minimize $f(x) = 0.625\, x_1^2 - 19.35\, x_1$

Subject to

$25.45\, x_1 - 0.1875\, x_1^2 \geq 53.63$

$3.8\, x_1 - 0.1389\, x_1^2 \geq 25.43$

$1.31\, x_1 - 0.0416\, x_1^2 \geq 10.06$

$$x_1 \geq 0$$

using the LINGO software and solve the mathematical programming we get

$x_1 = 15.48$

use value of x_1 in original table and continue estimate the second missing observation.

$\sigma_2^2(S_1) = 3.07$ and $\sigma_2^2(S_2) = 6.48$

Variance $(S_1) = \frac{1}{4}(726.86 + x_2^2) - \{\frac{1}{4}(46.4 + x_2)\}^2$

$= 47.155 - 5.8x_2 + 0.1875\, x_2^2$

Variance $(S_2) = 1/6\,(1627.31 + x_2^2) - \{1/6\,(89.3 + x_2)\}^2$

$= 49.71 - 4.96x_2 + 0.1389x_2^2$

Variance (Total) $= 1/24\,(5845.65 + x_2^2) - \{1/24\,(362.28 + x_2)\}^2$

$= 15.71 - 1.26\, x_2 + 0.04\, x_2^2$

E.S.S. $= x_2^2\{1 - 1/t - 1/r + 1/rt\} - 2x_2\{(Y_{.j})/t + (Y_{i.})/r - Y_{..}/rt\}$

$= 0.625x_2^2 - 22.76x_2$

Now the resulting mathematical programming is:

(P2) Minimize $f_2(x) = 0.625x_2^2 - 22.76x_2$

Subject to

$$5.8x_2 - 0.1875\, x_2^2 \geq 44.087$$

$$4.96x_2 - 0.1389x_2^2 \geq 43.23$$

$$1.26x_2 - 0.04x_2^2 \geq 9.66$$

$$x_2 \geq 0$$

solution is **$x_2 = 17.48$.**

we have estimated both missing observations, now to check whether there may be some possibilities of improvement we use cycling to estimate x_1 keeping x_2 as it is and continue.

(P1.1) Minimize $f(x) = 0.625\, x_1^2 - 17.89\, x_1$

Subject to

$$25.45\, x_1 - 0.1875\, x_1^2 \geq 53.63$$

$$3.8\, x_1 - 0.1389\, x_1^2 \geq 25.43$$

$$1.26\, x_1 - 0.04\, x_1^2 \geq 9.76$$

$$x_1 \geq 0$$

the solution of this problem is **$x_1 = 14.31$.**

(P2.1) Minimize $f(x) = 0.625x_2^2 - 22.88x_2$

Subject to

$$5.8x_2 - 0.1875\, x_2^2 \geq 44.087$$

$$4.96x_2 - 0.1389x_2^2 \geq 43.23$$

$$1.25x_2 - 0.04x_2^2 \geq 9.59$$

$$x_2 \geq 0$$

solution is **$x_2 = 17.48$**

The solution of (P2) and (P2.1) is same so stop the procedure.

The best estimated values of the missing observations are:

$x_1 = 14.31$ **$x_2 = 17.48$**

Reference:

[1] Cochran, W. G. and Cox,. “Experimental Designs”. Wiley Classical Library, (1992).

[2] Dempster. A.P., Laird, N.M., Rubin, D.B. “Maximum Likelihood from incomplete data via the EM Algorithm”. Journal of the Royal Statistical Society. Series B (Methodological) 39 (1): 1-38. (1977).

[3] Hicks, C. R. “Fundamental Concept in the design of Experiment”. Oxford University Press, USA, (1999)

[4] Harvey M. Wagner. “Principles of Operations Research”. Prentice-Hall India (1999).

[5] Jarrett, R. G., “ The analysis of Designed Experiments with Missing Observations”. App. Statist (1978), 27, No.1, 38-46.

[6] Murray, L. W., “Estimation of Missing Cells in Randomized Block and Latin square Design”. The American statistician, November 1986, vol. 40, N0. 4.

This academic article was published by The International Institute for Science, Technology and Education (IISTE). The IISTE is a pioneer in the Open Access Publishing service based in the U.S. and Europe. The aim of the institute is Accelerating Global Knowledge Sharing.

More information about the publisher can be found in the IISTE's homepage: http://www.iiste.org

The IISTE is currently hosting more than 30 peer-reviewed academic journals and collaborating with academic institutions around the world. **Prospective authors of IISTE journals can find the submission instruction on the following page:** http://www.iiste.org/Journals/

The IISTE editorial team promises to the review and publish all the qualified submissions in a fast manner. All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Printed version of the journals is also available upon request of readers and authors.

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar