THE MISSING-PLOT TECHNIQUE

H.LUO

The origin of this method is [Allan & Wishart] back into 1930, who listed a few reasons why *missing-plots* may occur in some agriculture situation:

> "Cases that have occurred are: (a) a weight was missed from one plot of a Latin Square; (b) the corner plot of a similar square gave a very poor yield, the reason, being that it adjoined the farm road and had evidently suffered badly from trampling. There was thus a good case for rejection of the particular plot; (c) in figures for a randomised block experiment recently submitted to this department, one figure, a protein determination, was missing and could not be retrieved."

There are two reasons why a missing-plot should be paid a great deal of attention, one is pointed by Prof.Maceachern that a missing entry might destroy the orthogonal structure of the design matrix and thus bring difficulties in calculations in old days; such a concern also applied to the lost of robustness when there is a missing plot.[1] The other is that "Clearly this method sacrifices a good deal of information that the plots rejected would have been capable of furnishing..."[2]. To remedy the second flaw, many researchers in 1930-1940s paid a lot of efforts, among them Frank Yates was certainly a star. His later work [Yates1940] has a far-reaching influence which discussed the situation where there are data yet the control is not neatly done. That is to say, the design is incomplete, not all treatment combinations can be replicated in each block. By re-weighting, we could still estimate the contrasts. But the missing-plot is another situation, where no data of a certain plot is available in any block.

Although [Allan & Wishart] has developed estimation of one missing-plot for randomized block design and Latin squares, the formal two-stage remedy of this situation is firstly stated by [Yates1933] in the setting of analysis of variance. A later introductory article [R.Anderson] extends to all other kinds of designs. A more modern statement of this can be found in [Kendall3] pp.124 Sec37.50-37.57. I summarized points from Kendall's as follows.

Yate's philosophy is simple, given an observation $\mathbf{y} = \begin{pmatrix} \mathbf{z} \\ \mathbf{u} \end{pmatrix}$ where $\mathbf{z}$ are observed data and $\mathbf{u}$ are missing data. We can regard it as an estimation problem that we estimate $\beta, \mathbf{x}$ given $\mathbf{z}$ and the design matrix $\mathbf{X}$ using the model $\mathbf{y} = \beta\mathbf{X} + \epsilon$.

The first step is to estimate $\mathbf{u}, \beta$ by LSE simultaneously. Which is to say, the sum of squares of residuals $S = (\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta)$ should be minimized by solving the normal equations $\begin{cases} \frac{\partial S}{\partial \beta} = 0 \\ \frac{\partial S}{\partial \mathbf{u}} = 0 \end{cases}$, denote the LS results by $\begin{cases} \widehat{\mathbf{u}} \\ \widehat{\beta(\mathbf{u})} \end{cases}$ and the least sums of squares $S_0(\mathbf{u})$.

The second step is to minimize, using variational equation w.r.t $\mathbf{u}$, the $S_0(\mathbf{u})$.

In general[3] we can proceed using the method of successive minimizations which must yield the same result, first minimize S w.r.t $\mathbf{u}$. Partition the design matrix $\mathbf{X} = \begin{pmatrix} \mathbf{X}_z \\ \mathbf{X}_u \end{pmatrix}$, plug it into the

[1] [Kendall3] pp.123 Sec37.50

[2] [Allan & Wishart]

[3] This paragraph is rephrased from [Kendall3] pp.124 Sec37.52

S we have that $S = \left[\left(\begin{array}{c}\mathbf{z}\\ \mathbf{u}\end{array}\right) - \left(\begin{array}{c}\mathbf{X}_z\\ \mathbf{X}_u\end{array}\right)\beta\right]' \left[\left(\begin{array}{c}\mathbf{z}\\ \mathbf{u}\end{array}\right) - \left(\begin{array}{c}\mathbf{X}_z\\ \mathbf{X}_u\end{array}\right)\beta\right] = (\mathbf{z}-\mathbf{X}_z\beta)'(\mathbf{z}-\mathbf{X}_z\beta) + (\mathbf{u}-\mathbf{X}_u\beta)'(\mathbf{u}-\mathbf{X}_u\beta)$. Now the term $(\mathbf{z}-\mathbf{X}_z\beta)'(\mathbf{z}-\mathbf{X}_z\beta)$ is not including $\mathbf{u}$ any more, the term $(\mathbf{u}-\mathbf{X}_u\beta)'(\mathbf{u}-\mathbf{X}_u\beta)$ is minimized when $\mathbf{u} = \mathbf{X}_u\beta$. $S = (\mathbf{z}-\mathbf{X}_z\beta)'(\mathbf{z}-\mathbf{X}_z\beta)$ can be minimized using LS methods as $\widehat{\beta(\mathbf{u})} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}_z'\mathbf{z}+\mathbf{X}_u'\mathbf{u})$ and thus we have a functional equation $\mathbf{u} = \mathbf{X}_u\widehat{\beta(\mathbf{u})} = \mathbf{X}_u(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}_z'\mathbf{z}+\mathbf{X}_u'\mathbf{u})$.

My observation here is that: If we treat the operator $\mathbf{C}$

$$\mathbf{C}:\mathbf{u}\mapsto \mathbf{X}_u(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}_z'\mathbf{z}+\mathbf{X}_u'\mathbf{u}) = \left[\mathbf{X}_u(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_z'\mathbf{z}\right] + \left[\mathbf{X}_u(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}_u'\mathbf{u}\right] = \mathbf{C}_z + \left[\mathbf{X}_u(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}_u'\mathbf{u}\right]$$

where $\mathbf{C}_z$ is a constant vector. $\mathbf{X}_u(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}_u'\mathbf{u}$ yields a smaller matrix norm, so this operator $\mathbf{C}$ is a *contraction* w.r.t. matrix norm of any kind, by the Brouwer fixed pointed theorem over the matrix space, we know that the equation $\mathbf{u} = \mathbf{C}\mathbf{u}$ must have a solution. The solution to this equation should be denoted as $\widehat{\mathbf{u}}$. Then the following paragraph is just a usual algorithm of solving such an equation.

To estimate the missing plot $\mathbf{u}$[4], we firstly did perform $\mathbf{u} = 0$ as an initial value, and then $\widehat{\beta(\mathbf{u})} = \widehat{\beta(\mathbf{0})} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_z'\mathbf{z}$. After that, $\widehat{\mathbf{u}} = \mathbf{Proj^{-1}}_{\mathbf{X}_u^{\perp}}\left(\mathbf{X}_u\widehat{\beta(\mathbf{u})}\right) = \left[\mathbf{I}-\mathbf{X}_u(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_u'\right]^{-1}\mathbf{X}_u\widehat{\beta(\mathbf{0})}$. Update the observation vector as $\mathbf{y} = \left(\begin{array}{c}\mathbf{z}\\ \widehat{\mathbf{u}}\end{array}\right)$. There is result that ensures such an algorithm converges.

So the whole philosophy of Yates' method based on such a philosophy that we should insert such a value to missing entry such that the "completion" of this set of data have the smallest sum of squares of residuals $S = (\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)$. This philosophy is rather consistent with R.Fisher's thinking, so it is no curiosity that Yates invented this method under the influence of Fisher.[5]

The earliest readable introduction of this missing-plot technique is as early as [R.Anderson] in 1940s, while this technique is absorbed by missing data researchers in [Roderick & Rubin] as late as 1980s.

References

[R.Anderson] Anderson, RICHARD L. "Missing-plot techniques." Biometrics Bulletin 2.3 (1946): 41-47.

[Yates1933] Yates, Frank. "The analysis of replicated experiments when the field results are incomplete." Empire Journal of Experimental Agriculture 1.2 (1933): 129-142.

[Cochran & Cox] Cochran, William G., and Gertrude M. Cox. "Experimental designs ." (1957).

[Kendall3] Kendall, M. G., A. Stuart, and J. K. Ord. "Vol. 3: Design and analysis, and time series." (1983).

[Roderick & Rubin] Little, Roderick JA, and Donald B. Rubin. "On jointly estimating parameters and missing data by maximizing the complete-data likelihood." The American Statistician 37.3 (1983): 218-220.

[T.W.Anderson] TW (Theodore Wilbur) Anderson. An introduction to multivariate statistical analysis. John Wiley & Sons, 1958.

[Allan & Wishart] Allan, F. E., and J. Wishart. "A method of estimating the yield of a missing plot in field experimental work." The Journal of Agricultural Science 20.03 (1930): 399-406.

[Yates1940] Yates, Frank. "THE RECOVERY OF INTERBLOCK INFORMATION IN BALANCED INCOMPLETE BLOCK DESIGNS." Annals of Eugenics 10.1 (1940): 317-325.

---

[4] This paragraph is rephrased from [Kendall3] pp.125 Sec37.53-54

[5] [Cochran & Cox] pp.81:"For this reason Yates(3.9), following a suggestion by Fisher, considered inserting values for the missing observations so as to obtain a set of complete data." The discussion there shows the underlying philosophy yet without an example.