Advances in Educational Technology and Psychology (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/aetp.2022.060703
ISSN 2371-9400 Vol. 6 Num. 7

Parameter estimation and application of non full rank linear regression model

Fan Jun, Du Wei, Jiang Yiming, Zhang Chunyan*

Anhui University School of Mathematical Sciences, Anhui Hefei 230000, China

Keywords: The non full rank model, Parameter estimation,Conditional inverse

Abstract: In the article, we first introduced the conception of non full rank model. Using the knowledge of probability theory and mathematical statistics, similar to the regular inverse of matrix, the general solution of normal equations was obtained by using the conditional inverse of matrix. In addition, by introducing the concept of estimable quantity, we found that the BLUE of the estimable quantity was unique. In addition, for the non full rank linear model, we estimated the estimable parameters and analysed the variance, and then solved the confidence interval of the parameters. Ultimately, we analysed the parameter estimation on the basis of the linear model.

Admittedly, linear models are widely used in biology, medicine, economy, management and other fields. And in the past few years, considerable attention has been paid to full-rank linear models. One common approach to estimating the parameters of a full-rank linear model is the method of least-squares. However, not all linear models are full-rank[6]. For example, an experimental comparison of the three kinds of pain relievers for effectiveness in alleviating arthritis. Therefore, it is desirable to explore the non-full-rank models.

In this paper, the conclusions we have obtained are based on the theories of full-rank models, such as the methods for estimating parameters, the approach to analysing variance, and the way of estimating confidence interval. However, what makes us special is that we replace the regular inverse with condition inverse of the matrix to solve the normal equation, and obtain the uniform form of all solutions of the equation. What's more, different from directly estimating the parameters in full-rank model, the estimable function consists of parameters is estimated in this paper.

1. The non full rank model

The less than full rank linear model is

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$y = X\beta + \varepsilon$

Where the design matrix X has non-full rank. And we assume that ε is multivariate normal and $E(\varepsilon) = 0, Var(\varepsilon) = \sigma^2 I$.

2. Parameter Estimation

2.1 Conditional inverse

Definition 1： If D is a $n \times p$ matrix, there exists a $p \times n$ matrix D^c which is known as the conditional inverse matrix of D such that it satisfies the property : $DD^cD = D$. If $n = p$ and D is nonsingular, then the $D^c = D^{-1}$.

The D^c can be figure out by following steps:

Step1: Find a reversible sub-matrix of D and it has the same rank as D ,then mark it as M ;

Step2: Construct a new matrix D^c of the same size as D^T in the following form:

$$D^c = \begin{bmatrix} M^{-1} & 0 \\ 0 & 0 \end{bmatrix}_{p \times n}$$

And D^c is the conditional inverse matrix of D that we want.

Proof: We assume M is the principal sub-matrix of D, and write

$$D = \begin{bmatrix} M & D_{12} \\ D_{21} & D_{22} \end{bmatrix}_{n \times p}$$

Thus

$$\begin{aligned} DD^cD &= \begin{bmatrix} M & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \begin{bmatrix} M^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} M & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ D_{21}M^{-1} & 0 \end{bmatrix} \begin{bmatrix} M & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \\ &= \begin{bmatrix} M & D_{12} \\ D_{21} & D_{21}M^{-1}D_{12} \end{bmatrix} \end{aligned}$$

Due to $r(M) = r(D)$, all other columns of D can be written as linear combinations of the first $r(D)$ columns. So, there is a matrix T such that

$$\begin{bmatrix} D_{12} \\ D_{22} \end{bmatrix} = \begin{bmatrix} M \\ D_{21} \end{bmatrix} T \quad \text{i.e.} \quad \begin{cases} D_{12} = MT \\ D_{22} = D_{21}T \end{cases}$$

And then we can gain that $D_{22} = D_{21}T = D_{21}M^{-1}D_{12}$.

Hence, we can prove that $DD^cD = \begin{bmatrix} M & D_{12} \\ D_{21} & D_{22} \end{bmatrix} = D$.

2.2 Solving the normal equations

Since the normal equation $X^T Xb = X^T y$ is consistent in the general linear model $y = X\beta + \varepsilon$, it always has a solution. However, for the linear model with low rank, the normal equation has countless solutions, which impedes the work of parameter estimation. Next, we're going to solve the normal equation.

Before we do that, let's pay attention to the general consistent system $Dx = g$. Obviously, $D^c g$ is a solution to the system, because $DD^c g = DD^c Dx = Dx = g$. Then we can deduce that $x = D^c g + (I - D^c D)z$ solves the consistent system $Dx = g$,where z is a random $p \times 1$ vector:

Proof:

$$\begin{aligned} Dx &= D\left[D^c g + (I - D^c D)z\right] \\ &= DD^c g + (D - DD^c D)z \\ &= g + (D - D)z = g \end{aligned}$$

Furthermore, if x^* is any solution to the system, then we have $x^* = D^c g + (I - D^c D)x^*$ for any D^c.

According to the above conclusion, solutions to the normal equation are also found. For some z and any $(X^T X)^c$, the solutions can be expressed as $b = (X^T X)^c X^T y + [I - (X^T X) X^T X]z$.

2.3 Estimation

Instead of directly estimating parameter β in the full rank linear models, the estimable function $t^T \beta$ is first found in the linear models with low rank and then estimated $t^T \beta$.

Definition 2: In the general linear model $y = X\beta + \varepsilon$, if there is a vector λ makes the expected value of $\lambda^T y$ is $E\left[\lambda^T y\right] = t^T \beta$, then the function $t^T \beta$ is estimable.

In other words, a quantity is estimable if there is a linear unbiased estimator for it[6]. And there are two theorems for determining whether $t^T \beta$ is estimable:

Th1: A function $t^T \beta$ is estimable if and only if linear system $X^T X z = t$ has a solution;

Th2: A function $t^T \beta$ is estimable if and only if there are some $(X^T X)^c$ that fulfils the property: $t^T (X^T X)^c X^T X = t^T$.

Now that we know how to judge the estimable function, we will introduce a well-known theorem that finds out the estimate of $t^T \beta$ and proves that the estimate of $t^T \beta$ is unique:

Th3: (A Gauss-Markov Theorem[7]) The BLUE (an abbreviation for the best linear unbiased estimate)for estimable function $t^T \beta$ is $z^T X^T y$, where z is a solution to the linear system $X^T X z = t$. Moreover, for any solution to the system, this estimate is exactly the same and can be expressed as $t^T b$,where b is any solution to the normal equation $X^T X b = X^T y$.

According to the **Th3** we can conclude that the BLUE is unique and equal to $t^T b$ .This conclusion is of great significance to the development of the theory of non-full rank model.

There is also a general conclusion for linear combinations of estimable quantities:

Let $z = a_1 t_1^T \beta + a_2 t_2^T \beta + \cdots + a_k t_k^T \beta$, where $t_1^T \beta, t_2^T \beta, \cdots$ and $t_k^T \beta$ are estimable functions. Then z is estimable, and the BLUE for z is $a_1 t_1^T b + a_2 t_2^T b + \cdots + a_k t_k^T b$.

3. Estimating σ^2 in the non full rank model

For the residuals, we define their sum of squares is SS_{Res} in the full rank model:

$$SS_{\mathrm{Res}} = (y - Xb)^T (y - Xb) = y^T [I - X(X^T X)^{-1} X^T] y.$$

Meanwhile, we assume the number of parameters is p, and the sample size is n. According to the above conditions, we estimate σ^2 by

$$s^2 = \frac{SS_{\mathrm{Res}}}{n - p}.$$

Similarly, for the normal equations, if we suppose the arbitrary solution to them is b, we can also define

$$SS_{\mathrm{Res}} = (y - Xb)^T (y - Xb),$$

in the less than full rank model.

b can vary, but Xb will not, because $X\beta$ is estimable.

Therefore, for the choice of b, SS_{Res} is constant.

Th4: $SS_{\mathrm{Res}} = y^T [I - X(X^T X)^c X^T] y$.

How do we find an estimator for σ^2?

Let's consider SS_{Res} again. Take $H = X(X^TX)^cX^T$ and remember that $HX = X$.

$E[SS_{\text{Res}}] = tr(I-H)\sigma^2$.

Since $I-H$ is idempotent and symmetric, we have $E[SS_{\text{Res}}] = r(I-H)\sigma^2 = (n-r(X))\sigma^2$.

Th5: In the general linear model $y = X\beta+\varepsilon$, suppose the rank of X is r and the mean and variance of ε are 0 and $\sigma^2 I$, respectively. Then $\frac{SS_{\text{Res}}}{n-r}$ is an unbiased estimator of σ^2.

4. Interval estimation in the non full rank model

For the full rank model, by multiplying a normal variable by the reciprocal of the square root of a χ^2 variable, we gained a t distributed quantity. And then we gained confidence intervals.

The χ^2 variable was

$$\frac{SS_{\text{Res}}}{\sigma^2},$$

and its degrees of freedom was $n-p$.

The σ^2 term was unknown, but there was another σ^2 term in the numerator, and they can offset. So we could calculate with something.

We can proceed in a similar for the non full rank model.

The procedure for finding a confidence interval for the non full rank model has two small differences with the full rank case.

Firstly, we can only find the confidence interval of estimator!

Secondly, we use the conditional inverse $(X^TX)^c$, instead of the inverse $(X^TX)^{-1}$.

The remaining steps are the same.

We have

$$\begin{aligned} Var\ t^Tb &= Var\ t^T(X^TX)^cX^Ty \\ &= t^T(X^TX)^cX^T\sigma^2 IX(X^TX)^ct \\ &= \sigma^2 t^T(X^TX)^ct. \end{aligned}$$

Therefore, when the degree of freedom is $n-r$, we gained a t distribution

$$\frac{(t^Tb-t^T\beta)/\sigma\sqrt{t^T(X^TX)^ct}}{\sqrt{s^2/\sigma^2}}.$$

For the (estimable) quantity $t^T\beta$, when the degree of freedom is $n-r$, we use a t distribution to acquire its confidence interval:

$$t^Tb \pm t_{\alpha/2}s\sqrt{t^T(X^TX)^ct}.$$

For those estimable individual parameters, their confidence intervals can also be found by this expression.

5. Results

Through the analysis of this article, we get these following results:

First and foremost, for some z and any conditional inverse $(X^TX)^c$, we obtain the general solution of the normal equations, which can be expressed as $b=(X^TX)^cX^Ty+[I-(X^TX)X^TX]z$.

Next, let $z = a_1t_1^T\beta + a_2t_2^T\beta+\cdots+a_kt_k^T\beta$, where $t_1^T\beta, t_2^T\beta,\cdots$ and $t_k^T\beta$ are estimable functions. Then z is estimable, and the BLUE for z is $a_1t_1^Tb+a_2t_2^Tb+\cdots+a_kt_k^Tb$.

Then, in the general linear model $y = X\beta + \varepsilon$, suppose the rank of X is r and the mean and variance of ε are 0 and $\sigma^2 I$, respectively. Then $\frac{SS_{\text{Res}}}{n-r}$ is an unbiased estimator of σ^2.

Last but not least, for the (estimable) quantity $t^T\beta$, when the degree of freedom is $n-r$, we use a t distribution to acquire its confidence interval: $t^T b \pm t_{\alpha/2} s\sqrt{t^T (X^T X)^c t}$. Furthermore, for those estimable individual parameters, their confidence intervals can also be found by this expression.

6. Discussion

The non-full rank linear regression model also plays an important role in medical, financial, agricultural and other field, so it is of practical significance to its research.

In this paper, we make point and interval estimation of the parameters of the non-full rank linear model. However, we can derive these results based on the assumption that the errors ε have mean 0 and variance $\sigma^2 I$. In practice, these assumptions do not always hold, which requires further study. Furthermore, another important research direction in the future is to explore the estimation and application of parameters in non-full rank and nonlinear models.

It is worth mentioning that the biased estimation of some parameters in the linear model is better than the unbiased estimation (e.g. multi-collinear phenomenon) in certain conditions. There are quite a few papers studying the biased estimation of linear model parameters, and some common biased estimation methods are Stein estimation, Ridge estimation, Liu estimation, and so on. Future research can be expected to apply the biased estimation methods to non-full rank linear model.

References

[1] YJ Wu, E Levina, Z Ji(2017). Generalized linear models with low rank effects for network data, arXiv.org, Ithaca.

[2] A Ozdemir, MA lwen, S Aviyente(2015). Locally linear low-rank tensor approximation, 2015 GLOBALSIP, pp:839-843.

[3] T Piotrowski, R Cavalcante, I Yamada(2009). Stochastic MV-PURE Estimator—Robust Reduced-Rank Estimator for Stochastic Linear Model. IEEE transactions on signal processing, vol.57, no.4, pp:1293-1303.

[4] HAL Thi, VT Ho(2021). Alternating DCA for reduced-rank multitask linear regression with covariance matrix estimation.

[5] ZHANG Wei-Guo, WANG Li-Yan, CHEN Wei(2004). Parameter Estimation of Generalizing Linear Model Under Restriction Condition, Journal of Hebei Normal University (Natural Science Edition), vol.28, no.3, pp:241-244.

[6] N. R. Mohan Madhyastha, S. Ravi, A. S. Praveena(2020). A First Course in Linear Models and Design of Experiments[M], Springer, Singapore.

[7] Latif, AHMM, Bretz, F, Brunner, E. Robustness considerations in selecting efficient two-color microarray designs, Bioinformatics, vol.25, no.18, pp: 2355-2361.