BIOSTATISTICS 667
Estimability and Related Issues in Linear Models
By Bahjat Qaqish

Note: This material is not officially part of BIOS 667. However, knowledge of it is assumed.

Take the model $\eta = X\beta$. Given a vector a, what do we mean when we say that $a^\top \beta$ is estimable? How do we find out whether $a^\top \beta$ is estimable or not? Theorems and formulae that give the answers can be found in most books on linear models. Here, we will take an approach based on a fundamental understanding. All questions about estimability can be answered by remembering a simple rule:

η is estimable

This is not just a rule, it is the very *definition of estimability.* We shall call it the golden rule of estimability (GRE), and it is essentially all we have to remember about estimability; everything else follows as a logical conclusion. It is to be interpreted as follows: Only linear combinations of η are estimable and if something is estimable it must be a linear combination of η. If something is not a linear combination of η then it is not estimable.

Let us take as an example one-way analysis of variance (ANOVA) with three groups (or treatments)

$$\eta_i = \mu + \alpha_i, \qquad (i = 1, 2, 3)$$

The vector β is $\beta = (\mu, \alpha_1, \alpha_2, \alpha_3)^\top$. Now, is μ estimable? To answer this question using the GRE, we look for a linear combination of η that is equal to μ. That is, we try to find coefficients $c = (c_1, c_2, c_3)^\top$ such that

$$\mu = c^\top \eta = c_1\eta_1 + c_2\eta_2 + c_3\eta_3.$$

Now we substitute the model $\eta_i = \mu + \alpha_i$ into the above to get

$$\mu = c_1(\mu + \alpha_1) + c_2(\mu + \alpha_2) + c_3(\mu + \alpha_3).$$

Then we equate the coefficients of μ and the α_i's on the two sides of the equation to get

$$\begin{array}{rrcl} \mu: & 1 & = & c_1 + c_2 + c_3 \\ \alpha_1: & 0 & = & c_1 \\ \alpha_2: & 0 & = & c_2 \\ \alpha_3: & 0 & = & c_3 \end{array}$$

Obviously, such a vector c does not exist. This says that μ is *not* a linear combination of η. By the GRE this implies that μ is *not* estimable.

In the above model, is α_1 estimable? We follow the same procedure. We try to find coefficients $c = (c_1, c_2, c_3)^\top$ such that

$$\alpha_1 = c^\top \eta = c_1\eta_1 + c_2\eta_2 + c_3\eta_3.$$

Now we substitute the model $\eta_i = \mu + \alpha_i$ into the above to get

$$\alpha_1 = c_1(\mu + \alpha_1) + c_2(\mu + \alpha_2) + c_3(\mu + \alpha_3)$$

Now we equate the coefficients of μ and the α_i's on the two sides of the equation to get

$$\begin{array}{rcl} \mu: & 0 & = & c_1 + c_2 + c_3 \\ \alpha_1: & 1 & = & c_1 \\ \alpha_2: & 0 & = & c_2 \\ \alpha_3: & 0 & = & c_3 \end{array}$$

Again, such a vector c does not exist. This says that α_1 is *not* a linear combination of η, which means that α_1 is *not* estimable.

In the above model, is $\alpha_3 - \alpha_1$ estimable? We can check that $\eta_3 - \eta_1 = \alpha_3 - \alpha_1$, so the answer is yes; the coefficients are $c = (-1, 0, 1)^\top$.

Let us study this again. By the GRE, $a^\top \beta$ is estimable only if we can write $a^\top \beta = c^\top \eta$ for some vector c. If that is the case, then

$$a^\top \beta = c^\top \eta = (c^\top X)\beta.$$

As β is free to range over R^p, this implies that $a^\top = c^\top X$ or $a = X^\top c$. In other words, the GRE says that $a^\top \beta$ is estimable only if a (the column vector) is a linear combination of the rows of X (viewed as column vectors), or $a^\top$ (the row vector) is a linear combination of the rows of X (viewed as row vectors). The equation

$$a = X^\top c$$

looks like a linear regression model (with no error) in which a is the "response", $X^\top$ is the "design matrix" and c is the "regression coefficients". If we regress a on $X^\top$ we should get a "perfect fit". The fitted values from the regression of a on $X^\top$ are $X^\top (XX^\top)^- Xa$. So a consequence of the GRE is that $a^\top \beta$ is estimable only if

$$a = X^\top (XX^\top)^- Xa$$

or, transposing,

$$a^\top = a^\top X^\top (XX^\top)^- X.$$

These two forms are often given in books as a theorem, possibly with a matrix A (or $A^\top$) in place of the vector a.

In the one-way ANOVA model with I groups, let α denote the vector $(\alpha_1, \cdots, \alpha_I)^\top$. Suppose we consider an $I \times 1$ vector d with the property that $1^\top d = d_1 + \cdots + d_I = 0$. We can check that $d^\top \eta = d_1(\mu + \alpha_1) + \cdots + d_I(\mu + \alpha_I) = \mu(1^\top d) + d^\top \alpha = d^\top \alpha$. This means that $d^\top \alpha$ is a linear combination of η which by the GRE means that it is estimable. This form of estimable functions is called a *contrast*. Note that $d^\top \alpha$ can be written as $c^\top \beta$ by defining c as $c := (0, d^\top)^\top$

Most regression coefficients are contrasts! More precisely, if X has full rank and (say) the first column is the intercept, then each β_j, $j > 1$, is a contrast. That is, for $j > 1$ we have $\beta_j = c_j^\top \eta$ where c_j is a vector of coefficients with $c_j^\top 1_n = 0$. It is not too hard to see why. Note that $c_j^\top$ is the jth row of $(X^\top X)^{-1} X^\top$. Since $(X^\top X)^{-1} X^\top X = I_{p \times p}$, and 1_n is the first column of X, we see that $c_j^\top 1_n = 1$ for $j = 1$ and $c_j^\top 1_n = 0$ for $j = 2, \cdots, p$.

Less Than Full Rank Models

Let us revisit the one-way ANOVA model

$$\eta_i = \mu + \alpha_i, \qquad (i = 1, 2, 3)$$

The design matrix (assuming one obervation at each treatment level is sufficient) is 3×4, and has less than full column rank. The parameter vector is $\beta = (\mu, \alpha_1, \alpha_2, \alpha_3)^\top$. We saw above that none of the 4 individual parameters $(\mu, \alpha_1, \alpha_2, \alpha_3)$ is estimable. There are many values of β that lead to the same η. For example, both

$$\beta = (1, 2, 3, 4)^\top$$

and

$$\beta = (4, -1, 0, 1)^\top$$

lead to exaclty the same η

$$\eta = (3, 4, 5)^\top.$$

This is not really a problem. Even though β is not unique, η is unique and well-defined. But, it is desirable to have "standard" methods for finding a vector β from a given η and X. There are three methods to achieve this. The methods may look different, but deep down they are very similar. Their connections will be pointed out. All methods will be illustrated with the ANOVA model given above.

1. *Estimability restrictions:*

   Warning Warning Warning: "Estimability restrictions" *are not actual restrictions on the parameter space.* They are merely a tool for generating estimates such that all estimable functions have unique estimates. Treating or interpreting estimability restrictions as actual restrictions on the parameter space is a serious error. For further clarification see Nelder(1998, Food Quality and Preference, 157–159), Nelder(1994, Statistics and Computing, 221–234), Nelder(2008, International Statistical Review, 134–139).

   Since α_1 is not estimable, we can apply the restriction $\alpha_1 = 0$. This leaves the parameters $(\mu, \alpha_2, \alpha_3)$ which all become estimable;

   $$\mu = \eta_1, \quad \alpha_2 = \eta_2 - \eta_1, \quad \alpha_3 = \eta_3 - \eta_1.$$

   The choice we made $\alpha_1 = 0$ is not unique. We could have chosen $\alpha_2 = 0$ or $\mu = 0$.

   Only a non-estimable function can be restricted. However, this should be done sequentially, not wholesale. It should be done "one at a time", with checking after each one. Once $\alpha_1 = 0$ has been applied, all the remaining parameters become estimable and no further restriction is possible.

   An example of the sequential nature of these restrictions: In the two-way ANOVA model

   $$\eta_{ij} = \mu + \alpha_i + \beta_j, \qquad (i = 1, \cdots, I; j = 1, \cdots, J),$$

   none of the $1 + I + J$ parameters is (individually) estimable. However, once we restrict $\alpha_1 = 0$, all the remaining α_i's become estimable (as evidenced by $\alpha_i = \eta_{ij} - \eta_{1j}$). But, μ and the β_j's are still non-estimable. If we now add the restriction $\mu = 0$, all the remaining parameters become estimable.

   Formally, estimability restrictions can be described by $M\beta = 0$, where the $p - r$ rows of M represent the $p - r$ restrictions needed. In order for M to represent a valid set of estimability restrictions it must satisfy two requirements (but is otherwise arbitrary). The two requirements are (Scheffe, 1959, page 17): a) The rows of M must be linealry independent of the rows of X, and b) The columns of the augmented matrix

$$G = \begin{bmatrix} X \\ M \end{bmatrix}$$

are linearly independent (i.e. G has full column rank). This leads to a *unique* β that satisfies the two requirements

$$\eta = X\beta \qquad \text{and} \qquad 0 = M\beta.$$

The unique solution is

$$\beta = (G^\top G)^{-1} X^\top \eta$$

Notice that this defines a unique β *once M has been chosen*. But keep in mind that estimability restrictions are not unique since there is a lot of flexibility in the choice of M.

In the one-way ANOVA model, the restriction $\alpha_1 = 0$ corresponds to

$$M = [0 \quad 1 \quad 0 \quad 0].$$

Exercise: Check that this M satisfies the two conditions given above. The unique solution is obvious, but check that it can be obtained by $\beta = (G^\top G)^{-1} X^\top \eta$

2. *Removing columns from the design matrix:* When X is not full rank, we can find a column that is linearly dependent on the other columns. Then we can remove that column and also remove the corresponding element from β. This gives a new design matrix X^* that is $n \times (p-1)$ and a new parameter vector β^* that is $(p-1) \times 1$. If X^* does not have full column rank, we apply the same procedure again. This process is repeated until we have a design matrix that has full column rank and a corresponding parameter vector.

   Note that the choice of which column to remove is arbitrary. In the one-way ANOVA model we can remove any of the four columns. The remaining matrix will have full rank. Note that removing a column is equivalent to restricting the corresponding coefficient to zero. Removing the first column is equivalent to the restriction $\mu = 0$, removing the second column is equivalent to the restriction $\alpha_1 = 0$, and so forth. So this approach can be viewed as a special form of the previous approach.

3. *Using a generalized inverse (g-inverse):* In this approach we simply use any generalized inverse $(X^\top X)^-$ to define a vector β by

$$\beta = (X^\top X)^- X^\top \eta.$$

   Example: In the one-way ANOVA model (with 3 groups)

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \; X^\top X = \begin{bmatrix} 3 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

   A g-inverse of $X^\top X$ is

$$A_1 = \frac{1}{16} \begin{bmatrix} 3 & 1 & 1 & 1 \\ 1 & 11 & -5 & -5 \\ 1 & -5 & 11 & -5 \\ 1 & -5 & -5 & 11 \end{bmatrix}$$

and it leads to

$$\beta_* = \begin{pmatrix} \mu_* \\ \alpha_{1*} \\ \alpha_{2*} \\ \alpha_{3*} \end{pmatrix} = A_1 X^\top \eta = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 \\ 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} \eta_1 + \eta_2 + \eta_3 \\ 3\eta_1 - \eta_2 - \eta_3 \\ 3\eta_2 - \eta_1 - \eta_3 \\ 3\eta_3 - \eta_1 - \eta_2 \end{pmatrix}$$

Exercise: Verify that η can be recovered as $X\beta_*$. Even though β is not unique, η is unique and can be recovered. This is consistent with the GRE. Another g-inverse of $X^\top X$ is

$$A_2 = \begin{bmatrix} 1 & -1 & -1 & 0 \\ -1 & 2 & 1 & 0 \\ -1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and it leads to

$$\beta_{**} = A_2 X^\top \eta = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \begin{pmatrix} \eta_3 \\ \eta_1 - \eta_3 \\ \eta_2 - \eta_3 \\ 0 \end{pmatrix}$$

Exercise: Verify that η can be recovered as $X\beta_{**}$.

The general form of the two examples above is that given the model $\eta = X\beta$ and a matrix A that is a g-inverse of $X^\top X$, we can obtain a corresponding β as $\beta_* = AX^\top\eta$ and recover η as $X\beta_* = XAX^\top\eta = X(X^\top X)^- X^\top\eta = \eta$. Exercise: Show that $B := AX^\top$ is a g-inverse of X, i.e. show that $XBX = X$, hence XB is idempotent. Further, show that XB is symmetric. So, conclude that XB is a projection matrix. Show that it projects (orthogonally) onto the column space of X.

Note that the estimability restrictions discussed above can be viewed as simply a way of picking a g-inverse. Indeed $(G^\top G)^{-1}$ is a g-inverse of $X^\top X$. Exercise: verify this claim. So, picking a set of restrictions $M\beta = 0$ is equivalent to picking a g-inverse of $X^\top X$.

Verify that A_1 given above is $(G_1^\top G_1)^{-1}$ where G_1 is obtained by imposing the restriction $\mu - (\alpha_1 + \alpha_2 + \alpha_3) = 0$, i.e. $M_1 = [1, -1, -1, -1]$ while $A_2 = (G_2^\top G_2)^{-1}$ where G_2 is obtained by imposing the restriction $\alpha_3 = 0$, i.e. $M_2 = [0, 0, 0, 1]$

Note: We use "g-inverse" as an abbreviation of "generalized inverse". This is not universal practice. Additionally, different books define generalized inverses differently. Here we say that B is a generalized inverse of A if $ABA = A$. We don't require any other properties. Other names found in the literature are "conditional inverse", "pseudo inverse" and "p-inverse".

Model Equivalence

As we have seen above, what really defines a model is η, not X and not β. If there is an $n \times q$ matrix Z and a $q \times 1$ vector γ such that η can be expressed as $\eta = X\beta$ and also as $\eta = Z\gamma$, we say that $X\beta$ and $Z\gamma$ are *equivalent representations of the model.* This is often stated in short form by saying that *the models $X\beta$ and $Z\gamma$ are equivalent*, or, simply, *the design matrices X and Z are equivalent.* We also say that $Z\gamma$ is a *reparametrization* of the model $\eta = X\beta$, and vice versa.

Let us expand what model equivalence really means. It means that given any $\beta \in R^p$, there exists a $\gamma \in R^q$ such that $X\beta = Z\gamma$, *and* given any $\gamma \in R^q$, there exists a $\beta \in R^p$ such that $X\beta = Z\gamma$. Another way to say this is that *any* linear combination of the columns of X can be expressed as a linear combination of the columns of Z, *and* vice versa. This clearly means the the column span of X is identical to the column span of Z.

One way to check this is to regress each column of X on the matrix Z, and vice versa. The fit will be a "perfect fit" in each case if the column spans are identical (i.e. if the two models are equivalent).

The fitted values from regressing a column t on X are $X(X^\top X)^- X^\top t$. The fitted values, being a projection, are unique no matter which g-inverse is used. If t is a linear combination of the columns of X then $t = X(X^\top X)^- X^\top t$. Similarly, if each column of Z is a linear combination of the columns of X then

$$Z = X(X^\top X)^- X^\top Z,$$

and if each column of X is a linear combination of the columns of Z then

$$X = Z(Z^\top Z)^- Z^\top X.$$

These are the formulae often given in books.

The transformation from β to γ follows as

$$\gamma = (Z^\top Z)^- Z^\top X\beta$$

and in the opposite direction as

$$\beta = (X^\top X)^- X^\top Z\gamma$$

with the understanding that these may not be unique.

Example: In the one-way ANOVA model (with 3 groups) two possible parametrizations are

$$\eta_i = \mu + \alpha_i \quad (i = 1, 2, 3)$$

and

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \end{bmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{pmatrix}.$$

In this example, it is easy to see that $\gamma_0 = (\eta_1 + \eta_2 + \eta_3)/3 = \mu + (\alpha_1 + \alpha_2 + \alpha_3)/3$, $\gamma_1 = (\eta_3 - \eta_1)/2 = (\alpha_3 - \alpha_1)/2$ and $\gamma_2 = (\eta_1 + \eta_3 - 2\eta_2)/6 = (\alpha_1 + \alpha_3 - 2\alpha_2)/6$. Note that γ_1 and γ_2 are contrasts; γ_1 is a *linear* contrast while γ_2 is a *quadratic* contrast. This terminology is used in connection with orthohogonal polynomials.