MA 575 Linear Models:

Cedric E. Ginestet, Boston University

ANOVA, ANCOVA and Mixed Effects Models

Week 10, Lecture 2

1 Fixed Effects Models

1.1 Blocking, Randomization and Contrasts

A **grouping** or blocking of observations can be achieved by using categorical or **dummy** variables. Examples of categorical variables include gender, country of origin, job title and experimental treatment. This should be contrasted with **ordinal** variables, such as age class, highest degree attained or a score on a 5-point scale with values comprised between *strongly agree* and *strongly disagree*.

The matrix $\mathbf{X}$ is usually referred to as the **design matrix**, because it specifies the experimental design. For instance, if considering a one-way analysis of variance (ANOVA) over three different groups, where we have 2 subjects in each group. We may select one of the following two design matrices,

$$\mathbf{X}_1 := \begin{vmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{vmatrix}, \qquad \text{and} \qquad \mathbf{X}_2 := \begin{vmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{vmatrix};$$

where in the former case, $\mathbf{X}_1$ is called a **cell means design**, whereas in the latter case, $\mathbf{X}_2$ is referred to as a **reference group design**, where the mean value in the two remaining groups are expressed as *offsets* from the value attained in the first group.

A factor with m levels can be coded using $m - 1$ variables, when the model includes an *intercept*. Otherwise, when no intercept is included as in the cell means design mentioned above, we require m dummy variables. The purpose of **blocking** is to increase *between-group* variability, while controlling *within-group* variability. For instance, we may wish to control for sex in a particular experiment testing for the effect of a drug. Then, we will ensure that each experimental block contains the same number of males and females. **Random allocation** of each independent unit to treatment condition permits to isolate the between-group variation. In **repeated-measures** design, the ordering of the variables may matter, and in that case, we may use a **Latin squares design** in order to ensure that all the orderings of the treatment have been administered.

1.2 ANOVA: Analysis of Variance

The ANOVA model is well-suited for grouped data sets, such as

$$(y_{ij}) \qquad i = 1, \ldots, n_j, \quad j = 1, \ldots, m.$$

Observe that we are here allowing for different group sizes, since the n_j's depend on $j = 1, \ldots, m$. The **one-way ANOVA** model over a factor with m levels can be expressed as follows,

$$y_i = \sum_{j=1}^{m} x_{ij}\alpha_j + e_i, \qquad i = 1, \ldots, N,$$

or more succinctly using matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{e}. \tag{1}$$

If $m = 3$, the **design matrix** has a **block** structure. Adopting a **cell means design**, we have

$$\mathbf{X} = \begin{vmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1}_{n_3} \end{vmatrix},$$

where each $\mathbf{1}_{n_j}$ is a vector of one's of size n_j. Alternatively, we may make the group structure explicit through a choice of nested indexes,

$$y_{ij} = \alpha_j + e_{ij}, \qquad i = 1, \ldots, n_j; \quad j = 1, \ldots, m;$$

where the N in the non-grouped formula above can be recovered by

$$N := \sum_{j=1}^{m} n_j.$$

Here, each α_j is said to be a **fixed effect** in the sense that it is assumed to control some population distribution from which the data has been drawn. As we have seen so far, using OLS theory, such parameters can be estimated by minimizing a Euclidean cost function.

1.3 OLS Estimators

The ANOVA is a special case of the general multiple regression that we have considered so far. Thus, we can simply minimize its RSS with respect to the vector $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_m]^T$ in order to obtain as before,

$$\widehat{\boldsymbol{\alpha}} := (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}.$$

Note, however, that the simple structure of the design matrix in the ANOVA model allows us to immediately compute the values of the OLS estimators, such that since the inverse of the matrix $\mathbf{X}^T\mathbf{X}$ of order $m \times m$ is given by

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{vmatrix} n_1 & & & \\ & n_2 & & \\ & & \ddots & \\ & & & n_m \end{vmatrix}^{-1} = \begin{vmatrix} \frac{1}{n_1} & & & \\ & \frac{1}{n_2} & & \\ & & \ddots & \\ & & & \frac{1}{n_m} \end{vmatrix},$$

it then follows that we immediately obtain

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \begin{vmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_m \end{vmatrix}.$$

Therefore, the vector of OLS estimators, $\widehat{\boldsymbol{\alpha}}$ is simply the vector of group means.

1.4 Hypothesis Testing

For an ANOVA model, we wish to know whether there are significant differences in the level of the observed variable in each group. The null hypothesis, in this case, is therefore whether or not the group-specific intercepts are identical,

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_m.$$

We can partition the different sources of variances in an ANOVA model, by noting that the total sum of squares, SYY or TSS can be decomposed as follows,

$$\begin{array}{ccccccc} \sum_{j=1}^{m}\sum_{i=1}^{n_j}(y_{ij}-\bar{\bar{y}})^2 & = & \sum_{j=1}^{m} n_j(\bar{y}_j-\bar{\bar{y}})^2 & + & \sum_{j=1}^{m}\sum_{i=1}^{n_j}(y_{ij}-\bar{y}_j)^2 \\ \text{Total SS} & = & \text{Between-group SS} & + & \text{Within-group SS} \\ N-1 & = & m-1 & + & N-m \end{array}$$

The partitioning of the variance in the ANOVA model can be verified by observing that the cross-terms eliminate,

$$\begin{aligned} 2\sum_{j=1}^{m}\sum_{i=1}^{n_j}(y_{ij}-\bar{y}_j)(\bar{y}_j-\bar{\bar{y}}) &= 2\sum_{j=1}^{m}\left((\bar{y}_j-\bar{\bar{y}})\sum_{i=1}^{n_j}(y_{ij}-\bar{y}_j)\right) \\ &= 2\sum_{j=1}^{m}(\bar{y}_j-\bar{\bar{y}})(n_j\bar{y}_j-n_j\bar{y}_j) = 0. \end{aligned}$$

In fact, this decomposition can be easily seen to be the finite sample analogue of the standard decomposition of the variance as the variance of the conditional expectation and the expectation of the conditional variance, such that for any two random variables Y and X,

$$\mathbb{V}\text{ar}[Y] = \mathbb{V}\text{ar}[\mathbb{E}[Y|X]] + \mathbb{E}[\mathbb{V}\text{ar}[Y|X]].$$

In our case, Y and X would correspond to the outcome variable and the group allocation, respectively.

This can be tested using an F-test, where we compare the full model with m intercepts with an **alternative** model comprising a single intercept, which would have the following form,

$$y_{ij} = \alpha + e_{ij}, \qquad i = 1, \ldots, n_j; \quad j = 1, \ldots, m.$$

The latter has only a single parameter and therefore $N-1$ degrees of freedom, which contrasts with the $N-m$ degrees of freedom of the full ANOVA model. Defining $\text{RSS}_1(\alpha) := \text{TSS}$ as the residual sum of squares for the first model with a single intercept, and $\text{RSS}_2(\alpha_1, \ldots, \alpha_m)$ as the within-group variance in the ANOVA model, we obtain

$$F = \frac{(\text{RSS}_1(\widehat{\alpha}) - \text{RSS}_2(\widehat{\boldsymbol{\alpha}}))/(m-1)}{\text{RSS}_2(\widehat{\boldsymbol{\alpha}})/(N-m)} = \frac{\text{BG-SS}\,/(m-1)}{\text{WG-SS}\,/(N-m)},$$

where BG-SS and WG-SS stands for between-group and within-group sums of squares, respectively. Thus, the F-statistic for this hypothesis would be distributed as $F \sim F(m-1, N-m)$ (see lecture 3.2).

1.5 ANCOVA: Analysis of Covariance

The term ANCOVA is not used consistently in the literature, and may refer to either fixed effects or mixed effects models. The main characteristic of this general family of models is that they comprise both continuous and discrete predictors.

The **analysis of covariance** model combines the features of the ANOVA with the ones of multiple regression. It can handle a data structure of the following form,

$$(y_{ij}, x_{ij}) \qquad i = 1, \ldots, n_j, \quad j = 1, \ldots, m.$$

It incorporates both group-specific intercepts and slope coefficients for continuous variables. Using a **block structure**, an ANCOVA model over m groups and with a single covariate can be written as follows,

$$y_{ij} = \alpha_j + x_{ij}\beta + e_{ij}, \qquad i = 1, \ldots, n_j; \quad j = 1, \ldots, m.$$

The design matrix for an ANCOVA model over $m = 3$ groups and a single continuous covariate has a **block design matrix** of the form,

$$\mathbf{X} = \begin{vmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \mathbf{0} & \mathbf{x}_{n_1} \\ \mathbf{0} & \mathbf{1}_{n_2} & \mathbf{0} & \mathbf{x}_{n_2} \\ \mathbf{0} & \mathbf{0} & \mathbf{1}_{n_3} & \mathbf{x}_{n_3} \end{vmatrix}.$$

In ANCOVA, we are primarily concerned with the categorical variable. In general, we introduce a continuous covariate in order to **control** for group differences on that particular covariate. The main hypothesis of interest will therefore be related to testing for the main effect of the experimental design while *controlling* for the extraneous continuous covariates. Thus, we could compare the model above with one including a single intercept, α, thereby testing for the null hypothesis,

$$H_0 : \alpha_1 = \ldots = \alpha_m.$$

As before, we need to compare the ANCOVA model above with a simpler model, based on a single intercept,

$$y_{ij} = \alpha + x_{ij}\beta + e_{ij},$$

which provides us with a reference $\text{RSS}_1(\alpha, \beta)$ for model comparison. This gives the following F-statistic,

$$F = \frac{(\text{RSS}_1(\widehat{\alpha}, \widehat{\beta}) - \text{RSS}_2(\widehat{\boldsymbol{\alpha}}, \widehat{\beta}))/(m-1)}{\text{RSS}_2(\widehat{\boldsymbol{\alpha}}, \widehat{\beta})/(N-m-1)}.$$

Thus, the F-statistic for this null hypothesis would be distributed as $F \sim F(m-1, N-m-1)$, due to the fact that we now also have to estimate a slope coefficient for the continuous covariate (see lecture 3.2).

1.6 Data Types for Fixed Effects Models

We can summarize the three types of data sets that we have dealt with, thus far, in the following table:

Regression	(y_i, x_i)	$i = 1, \ldots, N.$
ANOVA	(y_{ij})	$i = 1, \ldots, n_j;\ j = 1, \ldots, m.$
ANCOVA	(y_{ij}, x_{ij})	$i = 1, \ldots, n_j;\ j = 1, \ldots, m.$

These three models can be straightforwardly estimated using OLS or MLE, and the estimators are available in **closed-form**, as we have seen in previous lectures.

2 Mixed Effects Model

2.1 Motivating Example

A classical motivating example for the use of mixed effects model is a standard situation, where we have sets of data points (y_i, x_i), with $i = 1, \ldots, n$. This can be fitted using the standard OLS machinery, such that

$$y_i = \alpha + \beta x_i + e_i, \qquad i = 1, \ldots, n.$$

However, it may turn out that the data at hand, in fact, exhibits a group structure, such that we have

$$(y_{ij}, x_{ij}) \qquad i = 1, \ldots, n_j, \quad j = 1, \ldots, m.$$

Hence, we may instead assume that the single intercept in our original model has been drawn from a **population of intercepts**. This produces a so-called **two-level hierarchical model**,

$$\begin{aligned} y_{ij} &= \alpha_j + \beta x_{ij} + e_{ij}, \\ \alpha_j &= \alpha + b_j. \end{aligned}$$

for $i = 1, \ldots, n$, and $j = 1, \ldots, m$. At first, this may resemble the ANCOVA model that we have presented earlier. Similarly, this model explicitly specifies a set of parallel lines with identical slopes, but distinct intercepts. However, if this model is fitted using a mixed model structure, then the intercepts, b_j's, will be treated as random variables. Altogether, there are three main motivations for using mixed effects models:

i. First and foremost, the main advantage of using mixed effects models over standard fixed effects models is that they avoid **over-fitting** the data. Indeed, we are here only estimating the variance of the random effects as opposed to the random effects themselves.

ii. Moreover, the fixed effects are likely to be better *identified* (see lecture 11.1). That is, we will have more **degrees of freedom** to estimate the fixed effects by treating some of the parameters as random.

iii. Ultimately, the final decision of treating some parameters as fixed and others as random lies in the hands of the experimenter. This is perhaps why, we sometimes use the phrase **nuisance** parameters to refer to the quantities that are not of primary interest.

Finally, observe that mixed effects models are *implicitly* assuming the existence of a **group structure**.

2.2 VARCOMP: Variance Components

The variance components model is the simplest example of a mixed effects model, which contains both fixed and random effects. It is a generalization of the ANOVA model, where each intercept is random.

$$y_{ij} = \alpha_j + e_{ij}, \qquad i = 1, \ldots, n_j; \quad j = 1, \ldots, m,$$

with again $\alpha_j := \alpha + b_j$. We additionally make the assumption that these random effects are normally distributed with mean α and variance σ_b^2, such that

$$b_j \overset{\text{iid}}{\sim} N(0, \sigma_b^2), \qquad \text{and} \qquad e_{ij} \overset{\text{iid}}{\sim} N(0, \sigma_e^2),$$

where in general, $\sigma_b^2 \neq \sigma_e^2$. Equivalently, we could have re-written the first condition as $\alpha_j \overset{\text{iid}}{\sim} N(\alpha, \sigma_b^2)$. Therefore, the entire variance components model can be expressed in the following fashion,

$$y_{ij} = \alpha + b_j + e_{ij}, \qquad i = 1, \ldots, n_j; \quad j = 1, \ldots, m.$$

Here, the **random effects** b_j and the **error terms** e_{ij} are assumed to be (statistically) independent, which signifies that for every i and every j, we have

$$\mathbb{V}\text{ar}[b_j, e_{ij}] = 0.$$

Variance components are especially useful, when the number of groups exceeds the number of within-group elements. By contrast, when the number of groups is small, it is preferable to use ANOVA. This can be summarized with the following rule of thumb (see Demidenko, 2004, p.6):

1. *ANOVA* is preferable, when $m < \min n_j$;
2. *VARCOMP* is preferable, when $m > \max n_j$.

2.3 General Mixed Effects Models

When considering mixed effects models, it is common to denote **fixed effects** by Greek letters, and **random effects** by Roman ones. The main mixed effects model equation can be written as follows,

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{b}_j + \mathbf{e}_j, \qquad j = 1, \ldots, m. \tag{2}$$

where each term has dimensions:

- $\mathbf{y}_j$ is an $n_j \times 1$ vector of observations.
- $\mathbf{X}_j$ is an $n_j \times p^*$ design matrix for the fixed effects.
- $\mathbf{Z}_j$ is an $n_j \times k$ design matrix for the random effects.
- $\boldsymbol{\beta}$ is a $p^* \times 1$ vector of fixed effects.
- $\mathbf{b}_j$ is a $k \times 1$ vector of random effects.
- $\mathbf{e}_j$ is an $n_j \times 1$ vector of error terms.

In addition, we will also assume that both the error terms and the random effects are **uncorrelated** across groups and between each other in the sense that when $j \neq j'$, we have

$$\mathbb{E}[\mathbf{e}_j\mathbf{e}_{j'}^T] = \mathbf{0}, \qquad \mathbb{E}[\mathbf{b}_j\mathbf{b}_{j'}^T] = \mathbf{0}, \qquad \text{and} \qquad \mathbb{E}[\mathbf{e}_j\mathbf{b}_{j'}^T] = \mathbf{0};$$

where the first zero matrix is of order $n_j \times n_{j'}$, the second zero matrix is of order $k \times k$, and the third one of order $n_j \times k$. Finally, when $j = j'$, we have

$$\mathbb{E}[\mathbf{e}_j\mathbf{e}_j^T] = \sigma^2\mathbf{I}_{n_j}, \qquad \mathbb{E}[\mathbf{b}_j\mathbf{b}_j^T] = \sigma^2\mathbf{D}, \qquad \text{and} \qquad \mathbb{E}[\mathbf{e}_j\mathbf{b}_j^T] = \mathbf{0}.$$

Here, we have assumed that both the variance of both the error terms and the random effects are controlled by the same parameter, σ^2, but we will see that this assumption can be relaxed.

2.4 Distributional Properties

For convenience, we here assume that the variance of the error terms and the one of the random effects are controlled by the same parameter σ^2. This gives the following two distributions,

$$\mathbf{e}_j \overset{\text{ind}}{\sim} \text{MVN}_{n_j}(\mathbf{0}, \sigma^2\mathbf{I}_{n_j}), \qquad \mathbf{b}_j \overset{\text{iid}}{\sim} \text{MVN}_k(\mathbf{0}, \sigma^2\mathbf{D}),$$

for every $j = 1, \ldots, m$. Firstly, we can reformulate equation (2), in order to combine all of the groups into a single equation. This gives

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e}. \tag{3}$$

Here, $\boldsymbol{\beta}$ remains identical –that is, a vector of dimension $p^* \times 1$, while the other matrices are **stacked** into block matrices,

$$\mathbf{y} := \begin{vmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_m \end{vmatrix}, \qquad \mathbf{X} := \begin{vmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{vmatrix}, \qquad \mathbf{Z} := \begin{vmatrix} \mathbf{Z}_1 & \ldots & \mathbf{0} \\ \ddots & \ddots & \ddots \\ \mathbf{0} & \ldots & \mathbf{Z}_m \end{vmatrix}, \qquad \mathbf{b} := \begin{vmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_m \end{vmatrix}, \qquad \text{and} \qquad \mathbf{e} := \begin{vmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_m \end{vmatrix};$$

where $\mathbf{y}$, $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{b}$ and $\mathbf{e}$ are of dimensions $(N \times 1)$, $(N \times p^*)$, $(N \times mk)$, $(mk \times 1)$, and $(N \times 1)$. Also, recall that

$$N := \sum_{j=1}^{m} n_j.$$

Next, we can further compress the definition of the mixed effects model in equation (3) by combining the random effects and the error terms, in order to obtain

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}.$$

where

$$\boldsymbol{\eta} := \begin{vmatrix} \boldsymbol{\eta}_1 \\ \vdots \\ \boldsymbol{\eta}_m \end{vmatrix} = \begin{vmatrix} \mathbf{Z}_1\mathbf{b}_1 + \mathbf{e}_1 \\ \vdots \\ \mathbf{Z}_m\mathbf{b}_m + \mathbf{e}_m \end{vmatrix}.$$

Clearly, it is easy to verify that this term has a null expectation. For clarity of presentation, we drop the dependency of the expectation on $\mathbf{X}$, such that $\mathbb{E}[\cdot] := \mathbb{E}[\cdot|\mathbf{X}]$,

$$\mathbb{E}[\boldsymbol{\eta}] = \begin{vmatrix} \mathbf{Z}_1\mathbb{E}[\mathbf{b}_1] + \mathbb{E}[\mathbf{e}_1] \\ \vdots \\ \mathbf{Z}_m\mathbb{E}[\mathbf{b}_m] + \mathbb{E}[\mathbf{e}_m] \end{vmatrix} = \mathbf{0}.$$

Next, the variance of $\boldsymbol{\eta}$ can be obtained for any index $j = 1, \ldots, m$,

$$\begin{aligned} \mathbb{V}\text{ar}[\boldsymbol{\eta}_j] &= \mathbb{E}[\boldsymbol{\eta}_j\boldsymbol{\eta}_j^T] - \mathbb{E}[\boldsymbol{\eta}_j]\mathbb{E}[\boldsymbol{\eta}_j]^T \\ &= \mathbb{E}[(\mathbf{Z}_j\mathbf{b}_j + \mathbf{e}_j)(\mathbf{Z}_j\mathbf{b}_j + \mathbf{e}_j)^T] \\ &= \mathbb{E}[\mathbf{Z}_j\mathbf{b}_j\mathbf{b}_j^T\mathbf{Z}_j^T + \mathbf{Z}_j\mathbf{b}_j\mathbf{e}_j^T + \mathbf{e}_j\mathbf{b}_j^T\mathbf{Z}_j^T + \mathbf{e}_j\mathbf{e}_j^T] \\ &= \mathbf{Z}_j\mathbb{E}[\mathbf{b}_j\mathbf{b}_j^T]\mathbf{Z}_j^T + \mathbf{Z}_j\mathbb{E}[\mathbf{b}_j\mathbf{e}_j^T] + \mathbb{E}[\mathbf{e}_j\mathbf{b}_j^T]\mathbf{Z}_j^T + \mathbb{E}[\mathbf{e}_j\mathbf{e}_j^T] \\ &= \sigma^2(\mathbf{I}_{n_j} + \mathbf{Z}_j\mathbf{D}\mathbf{Z}_j^T), \end{aligned}$$

since both $\mathbb{E}[\mathbf{b}_j\mathbf{e}_j^T] = \mathbf{0}$ and $\mathbb{E}[\mathbf{e}_j\mathbf{b}_j^T] = \mathbf{0}$, by assumption. This derivation can also be verified straightforwardly using the properties of the variance operator,

$$\mathbb{V}\text{ar}[\boldsymbol{\eta}_j] = \mathbb{V}\text{ar}[\mathbf{Z}_j\mathbf{b}_j + \mathbf{e}_j] = \mathbf{Z}_j\,\mathbb{V}\text{ar}[\mathbf{b}_j]\mathbf{Z}_j^T + \mathbb{V}\text{ar}[\mathbf{e}_j] + 2\mathbf{Z}_j\,\mathbb{C}\text{ov}[\mathbf{b}_j, \mathbf{e}_j].$$

Moreover, it is easy to see that $\mathbb{V}\text{ar}[\boldsymbol{\eta}_j\boldsymbol{\eta}_{j'}^T] = 0$, whenever $j \neq j'$. Putting it altogether, we obtain an $N \times N$ matrix that has a **block-diagonal** structure,

$$\mathbb{V}\text{ar}[\boldsymbol{\eta}] = \sigma^2 \begin{vmatrix} \mathbf{I}_{n_1} + \mathbf{Z}_1\mathbf{D}\mathbf{Z}_1^T & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_2} + \mathbf{Z}_2\mathbf{D}\mathbf{Z}_2^T & \ldots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{I}_{n_m} + \mathbf{Z}_m\mathbf{D}\mathbf{Z}_m^T \end{vmatrix}.$$

Finally, we can then express the full **likelihood** of our model using the variance of the $\boldsymbol{\eta}$, such that

$$\mathbf{y}_j \overset{\text{ind}}{\sim} \text{MVN}_{n_j}\left(\mathbf{X}_j\boldsymbol{\beta},\ \sigma^2(\mathbf{I}_{n_j} + \mathbf{Z}_j\mathbf{D}\mathbf{Z}_j^T)\right), \qquad j = 1, \ldots, m.$$

References

Demidenko, E. (2004). *Mixed Models: Theory and Applications.* Wiley, London.