ANOVA Variance Component Estimation

We now consider the ANOVA approach to variance component estimation.

The ANOVA approach is based on the method of moments.

Suppose

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}, \text{ where}$$

$\boldsymbol{u}$ and $\boldsymbol{e}$ are independent random vectors satisfying

$$E(\boldsymbol{u}) = \boldsymbol{0} \quad \text{and} \quad E(\boldsymbol{e}) = \boldsymbol{0}.$$

Furthermore, suppose $\underset{n\times q}{\boldsymbol{Z}}$ can be partitioned as

$$\boldsymbol{Z} = [\underset{n\times q_1}{\boldsymbol{Z}_1}, \ldots, \underset{n\times q_m}{\boldsymbol{Z}_m}]$$

and $\boldsymbol{u}$ can be partitioned correspondingly as

$$\boldsymbol{u} = \begin{bmatrix} \boldsymbol{u}_1 \\ \vdots \\ \boldsymbol{u}_m \end{bmatrix}$$

so that

$$\boldsymbol{Z}\boldsymbol{u} = \sum_{j=1}^{m} \boldsymbol{Z}_j \boldsymbol{u}_j.$$

Suppose

$$\begin{aligned}\text{Var}(\boldsymbol{u}) &= \text{Var}\left(\begin{bmatrix}\boldsymbol{u}_1\\ \vdots \\ \boldsymbol{u}_m\end{bmatrix}\right) = diag(\sigma_1^2\mathbf{1}'_{q_1}, \ldots, \sigma_m^2\mathbf{1}'_{q_m}) \\ &= \begin{bmatrix}\sigma_1^2 \boldsymbol{I}_{q_1\times q_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \sigma_m^2 \boldsymbol{I}_{q_m\times q_m}\end{bmatrix}.\end{aligned}$$

Then

$$\text{Var}(\boldsymbol{y}) = \sum_{j=1}^{m+1} \sigma_j^2 \boldsymbol{Z}_j \boldsymbol{Z}_j',$$

where $\boldsymbol{Z}_{m+1} \equiv \underset{n \times n}{\boldsymbol{I}}$ and $\sigma_{m+1}^2 = \sigma_e^2$.

By Lemma 4.1,

$$\begin{aligned} E(\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}) &= tr(\boldsymbol{A}\text{Var}(\boldsymbol{y})) + [E(\boldsymbol{y})]'\boldsymbol{A}E(\boldsymbol{y}) \\ &= \sum_{j=1}^{m+1} \sigma_j^2 tr(\boldsymbol{A}\boldsymbol{Z}_j\boldsymbol{Z}_j') + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta} \\ &= \sum_{j=1}^{m+1} \sigma_j^2 tr(\boldsymbol{Z}_j'\boldsymbol{A}\boldsymbol{Z}_j) + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta}. \end{aligned}$$

Now suppose we choose a set of matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_{m+1} \ni$

$$\boldsymbol{X}'\boldsymbol{A}_i\boldsymbol{X} = \boldsymbol{0} \quad \forall\, i = 1, \ldots, m+1.$$

Then

$$E(\boldsymbol{y}'\boldsymbol{A}_i\boldsymbol{y}) = \sum_{j=1}^{m+1} \sigma_j^2 tr(\boldsymbol{Z}_j'\boldsymbol{A}_i\boldsymbol{Z}_j) \quad \forall\, i = 1, \ldots, m+1.$$

If we use the method of moments of moments, we replace $E(\boldsymbol{y}'\boldsymbol{A}_i\boldsymbol{y})$ with its observed value $\boldsymbol{y}'\boldsymbol{A}_i\boldsymbol{y}$ to obtain the equations

$$\boldsymbol{y}'\boldsymbol{A}_i\boldsymbol{y} = \sum_{j=1}^{m+1} \hat{\sigma}_j^2 tr(\boldsymbol{Z}_j'\boldsymbol{A}_i\boldsymbol{Z}_j) \quad \forall\, i = 1, \ldots, m+1.$$

We can write these equations in matrix form as

$$\begin{bmatrix} tr(\boldsymbol{Z}_1'\boldsymbol{A}_1\boldsymbol{Z}_1) & \dots & tr(\boldsymbol{Z}_{m+1}'\boldsymbol{A}_1\boldsymbol{Z}_{m+1}) \\ tr(\boldsymbol{Z}_1'\boldsymbol{A}_2\boldsymbol{Z}_1) & \dots & tr(\boldsymbol{Z}_{m+1}'\boldsymbol{A}_2\boldsymbol{Z}_{m+1}) \\ \vdots & \vdots & \vdots \\ tr(\boldsymbol{Z}_1'\boldsymbol{A}_{m+1}\boldsymbol{Z}_1) & \dots & tr(\boldsymbol{Z}_{m+1}'\boldsymbol{A}_{m+1}\boldsymbol{Z}_{m+1}) \end{bmatrix} \begin{bmatrix} \hat{\sigma}_1^2 \\ \hat{\sigma}_2^2 \\ \vdots \\ \hat{\sigma}_{m+1}^2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}'\boldsymbol{A}_1\boldsymbol{y} \\ \boldsymbol{y}'\boldsymbol{A}_2\boldsymbol{y} \\ \vdots \\ \boldsymbol{y}'\boldsymbol{A}_{m+1}\boldsymbol{y} \end{bmatrix}.$$

Solving for $\hat{\sigma}_1^2, \sigma_2^2, \dots, \hat{\sigma}_{m+1}^2$ gives the ANOVA estimates.

The matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_{m+1}$ are usually chosen so that

$$\boldsymbol{y}'\boldsymbol{A}_1\boldsymbol{y}, \ldots, \boldsymbol{y}'\boldsymbol{A}_{m+1}\boldsymbol{y}$$

correspond to sums of squares from an ANOVA table.

Many strategies for choosing $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_{m+1}$ have been proposed.

The book Variance Components by Searle, Casella, and McCulloch contains an extensive discussion of several strategies.

Let $\boldsymbol{M}$ denote the matrix whose $(i,j)^{th}$ element is $tr(\boldsymbol{Z}_j'\boldsymbol{A}_i\boldsymbol{Z}_j)$.

Let

$$\boldsymbol{s} = [\boldsymbol{y}'\boldsymbol{A}_1\boldsymbol{y}, \boldsymbol{y}'\boldsymbol{A}_2\boldsymbol{y}, \ldots, \boldsymbol{y}'\boldsymbol{A}_{m+1}\boldsymbol{y}]'.$$

If $\boldsymbol{M}$ is nonsingular, then the vector of ANOVA variance component estimates is

$$\hat{\boldsymbol{\sigma}}^2 \equiv \begin{bmatrix} \hat{\sigma}_1^2 \\ \vdots \\ \hat{\sigma}_{m+1}^2 \end{bmatrix} = \boldsymbol{M}^{-1}\boldsymbol{s}.$$

Recall $\boldsymbol{M}$ and $\boldsymbol{s}$ were chosen so that

$$\boldsymbol{M}\boldsymbol{\sigma}^2 = E(\boldsymbol{s}), \quad \text{where} \quad \boldsymbol{\sigma}^2 = \begin{bmatrix} \sigma_1^2 \\ \vdots \\ \sigma_{m+1}^2 \end{bmatrix}.$$

Thus,

$$\begin{aligned} E(\hat{\boldsymbol{\sigma}}^2) &= E(\boldsymbol{M}^{-1}\boldsymbol{s}) = \boldsymbol{M}^{-1}E(\boldsymbol{s}) \\ &= \boldsymbol{M}^{-1}\boldsymbol{M}\boldsymbol{\sigma}^2 = \boldsymbol{\sigma}^2. \end{aligned}$$

$\therefore$ the ANOVA estimator of $\boldsymbol{\sigma}^2$ is unbiased.

Find ANOVA-based estimates of σ_p^2 and σ_e^2 for the seedling dry weight example.

Recall

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{Z} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Because there is only one variance component besides the error variance, we have $m = 1$, $\boldsymbol{Z}_1 = \boldsymbol{Z}$, and $\boldsymbol{Z}_2 = \boldsymbol{I}$.

We need to identify matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ such that

$$\boldsymbol{X}'\boldsymbol{A}_1\boldsymbol{X} = \boldsymbol{0} \text{ and } \boldsymbol{X}'\boldsymbol{A}_2\boldsymbol{X} = \boldsymbol{0}.$$

Of course, we also require the quadratic forms $\boldsymbol{y}'\boldsymbol{A}_1\boldsymbol{y}$ and $\boldsymbol{y}'\boldsymbol{A}_2\boldsymbol{y}$ to contain information about σ_p^2 and σ_e^2.

To find appropriate $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, start by writing down an ANOVA table with columns labeled *Source*, *Matrix*, and *DF*.

Which of the matrices in the ANOVA table satisfy $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} = \boldsymbol{0}$?

To apply the ANOVA variance component estimation method, we need to set up the equations

$$\begin{bmatrix} tr(\boldsymbol{Z}_1'\boldsymbol{A}_1\boldsymbol{Z}_1) & tr(\boldsymbol{Z}_2'\boldsymbol{A}_1\boldsymbol{Z}_2) \\ tr(\boldsymbol{Z}_1'\boldsymbol{A}_2\boldsymbol{Z}_1) & tr(\boldsymbol{Z}_2'\boldsymbol{A}_2\boldsymbol{Z}_2) \end{bmatrix} \begin{bmatrix} \hat{\sigma}_p^2 \\ \hat{\sigma}_e^2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}'\boldsymbol{A}_1\boldsymbol{y} \\ \boldsymbol{y}'\boldsymbol{A}_2\boldsymbol{y} \end{bmatrix}$$

$$\Longleftrightarrow$$

$$\begin{bmatrix} tr(\boldsymbol{Z}'(\boldsymbol{P_Z} - \boldsymbol{P_X})\boldsymbol{Z}) & tr(\boldsymbol{P_Z} - \boldsymbol{P_X}) \\ tr(\boldsymbol{Z}'(\boldsymbol{I} - \boldsymbol{P_Z})\boldsymbol{Z}) & tr(\boldsymbol{I} - \boldsymbol{P_Z}) \end{bmatrix} \begin{bmatrix} \hat{\sigma}_p^2 \\ \hat{\sigma}_e^2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}'(\boldsymbol{P_Z} - \boldsymbol{P_X})\boldsymbol{y} \\ \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_Z})\boldsymbol{y} \end{bmatrix}.$$

Let n_{ij} = number of seedlings in the j^{th} pot for genotype i.

Thus, we have $n_{11} = 3, n_{12} = 2, n_{21} = 3, n_{22} = 2$.

Write $\boldsymbol{y}'\boldsymbol{A}_1\boldsymbol{y} = \boldsymbol{y}'(\boldsymbol{P_Z} - \boldsymbol{P_X})\boldsymbol{y}$ and $\boldsymbol{y}'\boldsymbol{A}_2\boldsymbol{y} = \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_Z})\boldsymbol{y}$ using summation notation.

Find $tr(\boldsymbol{Z}'(\boldsymbol{P_Z} - \boldsymbol{P_X})\boldsymbol{Z})$, $tr(\boldsymbol{Z}'(\boldsymbol{I} - \boldsymbol{P_Z})\boldsymbol{Z})$, $tr(\boldsymbol{P_Z} - \boldsymbol{P_X})$, and $tr(\boldsymbol{I} - \boldsymbol{P_Z})$.

Find expressions for the ANOVA estimators of $\hat{\sigma}_p^2$ and $\hat{\sigma}_e^2$.

The ANOVA estimates of the variance components are sometimes equal to the REML estimates.

This equality occurs for certain types of balanced designs and when the ANOVA estimates are positive.

For the seedling dry weight example, the REML and ANOVA estimates agree when the latter are positive.

However, if last pot contained a 3 seedlings instead of 2 (for example), the REML and ANOVA estimates would differ.