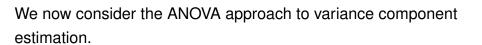
ANOVA Variance Component Estimation



The ANOVA approach is based on the method of moments.

Suppose

$$y = X\beta + Zu + e$$
, where

u and e are independent random vectors satisfying

$$E(u) = 0$$
 and $E(e) = 0$.

Furthermore, suppose Z can be partitioned as

$$oldsymbol{Z} = [oldsymbol{Z}_1, \dots, oldsymbol{Z}_m]_{n imes q_m}$$

and u can be partitioned correspondingly as

$$u = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}$$

so that

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

Suppose

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

Then

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

where $\mathbf{Z}_{m+1} \equiv \mathbf{I}$ and $\sigma_{m+1}^2 = \sigma_e^2$.

By Lemma 4.1,

$$E(\mathbf{y}'A\mathbf{y}) = tr(\mathbf{A}\operatorname{Var}(\mathbf{y})) + [E(\mathbf{y})]'\mathbf{A}E(\mathbf{y})$$

$$= \sum_{j=1}^{m+1} \sigma_j^2 tr(\mathbf{A}\mathbf{Z}_j\mathbf{Z}_j') + \beta' X' \mathbf{A}X\beta$$

$$= \sum_{j=1}^{m+1} \sigma_j^2 tr(\mathbf{Z}_j'\mathbf{A}\mathbf{Z}_j) + \beta' X' \mathbf{A}X\beta.$$

Now suppose we choose a set of matrices $A_1, \ldots, A_{m+1} \ni$

$$X'A_iX = 0 \quad \forall i = 1, \ldots, m+1.$$

Then

$$E(\mathbf{y}'\mathbf{A}_{i}\mathbf{y}) = \sum_{i=1}^{m+1} \sigma_{j}^{2} tr(\mathbf{Z}'_{j}\mathbf{A}_{i}\mathbf{Z}_{j}) \quad \forall \ i = 1, \dots, m+1.$$

If we use the method of moments of moments, we replace $E(y'A_iy)$ with its observed value $y'A_iy$ to obtain the equations

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

We can write these equations in matrix form as

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

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

The matrices A_1, \ldots, A_{m+1} are usually chosen so that

$$y'A_1y,\ldots,y'A_{m+1}y$$

correspond to sums of squares from an ANOVA table.

Many strategies for choosing A_1, \ldots, A_{m+1} have been proposed.

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

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

Let

$$s = [y'A_1y, y'A_2y, \dots, y'A_{m+1}y]'.$$

If M is nonsingular, then the vector of ANOVA variance component estimates is

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

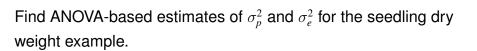
Recall M and s were chosen so that

$$m{M}m{\sigma}^2 = E(m{s}), \quad ext{where} \quad m{\sigma}^2 = egin{bmatrix} \sigma_1^2 \ dots \ \sigma_{m+1}^2 \end{bmatrix}.$$

Thus,

$$E(\hat{\sigma}^2) = E(M^{-1}s) = M^{-1}E(s)$$
$$= M^{-1}M\sigma^2 = \sigma^2.$$

 \therefore the ANOVA estimator of σ^2 is unbiased.



Recall

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

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

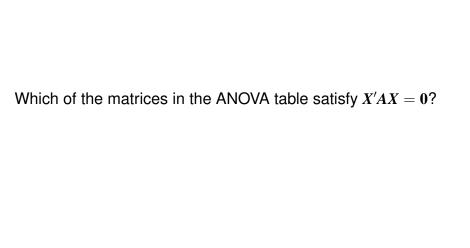
We need to identify matrices A_1 and A_2 such that

$$X'A_1X = 0$$
 and $X'A_2X = 0$.

Of course, we also require the quadratic forms $y'A_1y$ and $y'A_2y$ to contain information about σ_p^2 and σ_e^2 .

To find appropriate A_1 and A_2 , start by writing down an ANOVA table with columns labeled *Source*. *Matrix*. and *DF*.

Source	Matrix	DF
Intercept	P_1	1
Genotype	$P_X - P_1$	2 - 1 = 1
Pot(Genotype)	$P_Z - P_X$	4 - 2 = 2
Error	$I - P_Z$	10 - 4 = 6
Total	I	10



Let

$$A_1 = P_Z - P_X$$
 and $A_2 = I - P_Z$.

$$:: \mathcal{C}(X) \subset \mathcal{C}(Z), P_Z X = X.$$

$$\therefore X'(P_Z - P_X)X = X'(P_Z X - P_X X)$$
$$= X'(X - X) = 0.$$

Also,
$$X'(I - P_Z)X = X'(X - P_ZX) = X'(X - X) = 0$$
.

$$\therefore X'A_iX = \mathbf{0} \quad \forall \ i = 1, 2.$$

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

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

$$\iff$$

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

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 $y'A_1y = y'(P_Z - P_X)y$ and $y'A_2y = y'(I - P_Z)y$ using summation notation.

$$y'A_{1}y = y'(P_{Z} - P_{X})y = \|(P_{Z} - P_{X})y\|^{2}$$

$$= \|P_{Z}y - P_{X}y\|^{2}$$

$$= \|\begin{bmatrix} \bar{y}_{11} & \mathbf{1} \\ \bar{y}_{12} & \mathbf{1} \\ \bar{y}_{21} & \mathbf{1} \\ \bar{y}_{22} & \mathbf{1} \end{bmatrix} - \begin{bmatrix} \bar{y}_{1} & \mathbf{1} \\ \bar{y}_{1} & \mathbf{1} \\ \bar{y}_{2} & \mathbf{1} \end{bmatrix}^{2}$$

$$= 3(\bar{y}_{11} - \bar{y}_{1})^{2} + 2(\bar{y}_{12} - \bar{y}_{1})^{2}$$

$$+ 3(\bar{y}_{21} - \bar{y}_{2})^{2} + 2(\bar{y}_{22} - \bar{y}_{2})^{2}$$

$$= \sum_{i=1}^{2} \sum_{j=1}^{2} n_{ij}(\bar{y}_{ij} - \bar{y}_{i})^{2}.$$

$$y'A_2y = y'(I - P_Z)y$$

$$= ||(I - P_Z)y||^2$$

$$= ||y - P_Zy||^2$$

$$= \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij}.)^2.$$

Find $tr(\mathbf{Z}'(\mathbf{P}_{\mathbf{Z}} - \mathbf{P}_{\mathbf{X}})\mathbf{Z})$, $tr(\mathbf{Z}'(\mathbf{I} - \mathbf{P}_{\mathbf{Z}})\mathbf{Z})$, $tr(\mathbf{P}_{\mathbf{Z}} - \mathbf{P}_{\mathbf{X}})$, and $tr(\mathbf{I} - \mathbf{P}_{\mathbf{Z}})$.

Note that
$$P_X Z = \frac{1}{5} \begin{bmatrix} 3 & 2 & 0 & 0 \\ 3 & 2 & 0 & 0 \\ 3 & 2 & 0 & 0 \\ 3 & 2 & 0 & 0 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 3 & 2 \end{bmatrix}$$
. Thus, $Z' P_X Z = \frac{1}{5} \begin{bmatrix} 9 & 6 & 0 & 0 \\ 6 & 4 & 0 & 0 \\ 0 & 0 & 9 & 6 \\ 0 & 0 & 6 & 4 \end{bmatrix}$.

Thus,
$$\mathbf{Z}'\mathbf{P}_{\mathbf{X}}\mathbf{Z} = \frac{1}{5} \begin{bmatrix} 9 & 6 & 0 & 0 \\ 6 & 4 & 0 & 0 \\ 0 & 0 & 9 & 6 \\ 0 & 0 & 6 & 4 \end{bmatrix}$$

$$\mathbf{Z}'\mathbf{P}_{\mathbf{Z}}\mathbf{Z} = \mathbf{Z}'\mathbf{Z} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

Thus,

$$tr(\mathbf{Z}'\mathbf{A}_{1}\mathbf{Z}) = tr(\mathbf{Z}'(\mathbf{P}_{Z} - \mathbf{P}_{X})\mathbf{Z})$$

$$= tr(\mathbf{Z}'\mathbf{P}_{Z}\mathbf{Z} - \mathbf{Z}'\mathbf{P}_{X}\mathbf{Z})$$

$$= tr(\mathbf{Z}'\mathbf{Z}) - tr(\mathbf{Z}'\mathbf{P}_{X}\mathbf{Z})$$

$$= 10 - \frac{26}{5} = 4.8.$$

$$tr(\mathbf{Z}'\mathbf{A}_{2}\mathbf{Z}) = tr(\mathbf{Z}'(\mathbf{I} - \mathbf{P}_{\mathbf{Z}})\mathbf{Z})$$

$$= tr(\mathbf{Z}'(\mathbf{Z} - \mathbf{P}_{\mathbf{Z}}\mathbf{Z}))$$

$$= tr(\mathbf{Z}'(\mathbf{Z} - \mathbf{Z}))$$

$$= 0.$$

$$tr(\mathbf{A}_1) = tr(\mathbf{P}_{\mathbf{Z}} - \mathbf{P}_{\mathbf{X}}) = tr(\mathbf{P}_{\mathbf{Z}}) - tr(\mathbf{P}_{\mathbf{X}})$$
$$= rank(\mathbf{Z}) - rank(\mathbf{X}) = 4 - 2 = 2.$$
$$tr(\mathbf{A}_2) = tr(\mathbf{I} - \mathbf{P}_{\mathbf{Z}}) = 10 - 4 = 6.$$

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

The ANOVA/Method of Moments equations are

$$\begin{bmatrix} tr(\mathbf{Z}'(\mathbf{P}_{\mathbf{Z}} - \mathbf{P}_{\mathbf{X}})\mathbf{Z}) & tr(\mathbf{P}_{\mathbf{Z}} - \mathbf{P}_{\mathbf{X}}) \\ tr(\mathbf{Z}'(\mathbf{I} - \mathbf{P}_{\mathbf{Z}})\mathbf{Z}) & tr(\mathbf{I} - \mathbf{P}_{\mathbf{Z}}) \end{bmatrix} \begin{bmatrix} \hat{\sigma}_{p}^{2} \\ \hat{\sigma}_{e}^{2} \end{bmatrix} = \begin{bmatrix} \mathbf{y}'(\mathbf{P}_{\mathbf{Z}} - \mathbf{P}_{\mathbf{X}})\mathbf{y} \\ \mathbf{y}'(\mathbf{I} - \mathbf{P}_{\mathbf{Z}})\mathbf{y} \end{bmatrix}$$

$$\iff \begin{bmatrix} 4.8 & 2 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} \hat{\sigma}_{p}^{2} \\ \hat{\sigma}_{e}^{2} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{2} \sum_{j=1}^{2} n_{ij} (\bar{y}_{ij} - \bar{y}_{i}..)^{2} \\ \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij}.)^{2} \end{bmatrix}$$

$$\iff \hat{\sigma}_{e}^{2} = \frac{\sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij}.)^{2} - 2\hat{\sigma}_{e}^{2}}{\hat{\sigma}_{e}^{2}}.$$

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.