ORIGIN ≡ 0   **Prototype of Full-Sib Analysis from Lynch & Walsh**   **W. Stein**

Full-Sib Analysis involves data derived from "families" of one parent (often the male = "sire" to reduce maternal effects) mated with several partners (often the female = "dam"). The typical objective is to estimate variance components. Example below is derived from Chapter 18 Example 3 from (LW) Michael Lynch & Bruce Walsh 1998 *Genetics and Analysis of Quantitative Traits*, Sinauer Associates. The authors show calculations for balanced & unbalanced designs following from their prior discussion in Chapter 18 of half-sib analysis. Definition and a test for significance of additive genetic variance is provided. At the end of their Example, they show an example calculation of how to estimate variance components including: (1) additive genetic variance, (2) dominance genetic variance, (3) maternal-effects variance & (4) variance for "unobserved causes (presumably special environmental effects)"

Data Design:

Multiple offspring (full-sibs) are collected as replicates from different females mated to a single male, thus comprising "half-sib" families.
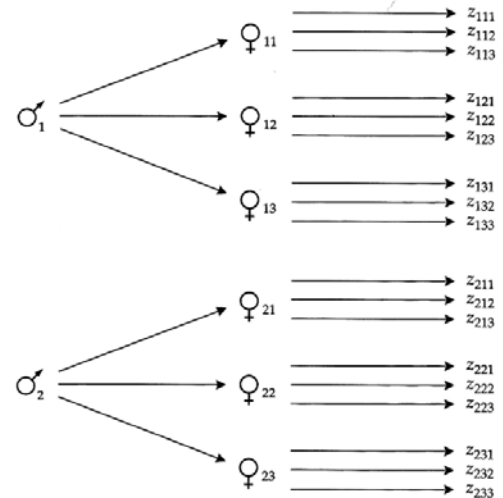
ANOVA Model:

$z_{ijk} = \mu + s_i + d_{j(i)} + e_{ijk}$ **< Nested ANOVA model**

^ modified slightly from LW to show nesting similar to terminlolgy in KNNL

Assumptions:

$e_{ij} \sim N(0, \sigma_e^2)$ $s_i \sim N(0, \sigma_S^2)$ $d_{j(i)} \sim N(0, \sigma_d^2)$

all families have equal variance = σ^2

Factor variances uncorrelated

Example: **LW Example 3 p. 578**

The example proceeds from results of an ANOVA table, presumably published by the orignal authors (Berenbaum et al. 1986), not from the original data.

Reported Dimensions & ANOVA table results:

	df	Mean Squares			
				$N := 29 + 1$	$N = 30$	**< number sires = flour beetle families**
Sires	$df_s := 29$	$MSS := 5.949$	$k_1 := 9.1$			
Dams within Sires	$df_d := 56$	$MSD := 3.925$	$k_2 := 9.1$	$M := \dfrac{56}{N+1}$	$M = 1.8065$	**< *average* number of dams within sires**
Sibs within Dams	$df_e := 695$	$MSE := 1.314$	$k_3 := 25.7$	$T := 695 + N \cdot M$	$T = 749.1935$	

Calculated Estimates:

Estimated variance components:

Eq 18.32a-c:

$$Var_s := \frac{MSS - MSE - \left(\dfrac{k_2}{k_1}\right) \cdot (MSD - MSE)}{k_3}$$

$Var_s = 0.078755$ **< estimate of σ_S^2**

$$Var_d := \frac{MSD - MSE}{k_1}$$

$Var_d = 0.286923$ **< estimate of σ_d^2**

$Var_e := MSE$

$Var_e = 1.314$ **< estimate of σ_e^2**

$Var_z := Var_s + Var_d + Var_e$

^ verified LW p. 578

Table 13.3 giving general formulas for unbalanced designs. Note terms k_1, k_2, k_3 and *average* M.

Table 18.3   Summary of a nested analysis of variance involving N sires, M_i dams within the ith sire, and n_{ij} offspring within the ijth full-sib family.

Factor	df	Sums of Squares	MS	$E(\text{MS})$
Sires	$N-1$	$\displaystyle\sum_{i=1}^{N}\sum_{j=1}^{M_i} n_{ij}(\bar{z}_i - \bar{z})^2$	SS_s/df_s	$\sigma_e^2 + k_2\sigma_d^2 + k_3\sigma_s^2$
Dams (sires)	$N(\overline{M}-1)$	$\displaystyle\sum_{i=1}^{N}\sum_{j=1}^{M_i} n_{ij}(\bar{z}_{ij} - \bar{z}_i)^2$	SS_d/df_d	$\sigma_e^2 + k_1\sigma_d^2$
Sibs (dams)	$T - N\overline{M}$	$\displaystyle\sum_{i=1}^{N}\sum_{j=1}^{M_i}\sum_{k=1}^{n_{ij}} (z_{ijk} - \bar{z}_{ij})^2$	SS_e/df_e	σ_e^2
Total	$T-1$	$\displaystyle\sum_{i=1}^{N}\sum_{j=1}^{M_i}\sum_{k=1}^{n_{ij}} (z_{ijk} - \bar{z})^2$		

$$k_1 = \frac{1}{N(\overline{M}-1)}\left(T - \sum_{i=1}^{N}\frac{\sum_j^{M_i} n_{ij}^2}{n_i}\right)$$

$$k_2 = \frac{1}{N-1}\left(\sum_{i=1}^{N}\frac{\sum_j^{M_i} n_{ij}^2}{n_i} - \frac{\sum_i^{N}\sum_j^{M_i} n_{ij}^2}{T}\right)$$

$$k_3 = \frac{1}{N-1}\left(T - \frac{\sum_i^{N} n_i^2}{N}\right)$$

Note: T is the total number of individuals in the experiment, $\overline{M}$ is the mean number of dams/sire, and n_i is the total number of offspring of the ith sire. MS denotes an observed mean square, $E(\text{MS})$ denotes its expected value, and df denotes degrees of freedom.

Intraclass correlations:

$$t_{\text{PHS}} := \frac{\text{Var}_s}{\text{Var}_z}$$

$t_{\text{PHS}} = 0.046887$ **< partial half sibs**

$$t_{\text{FS}} := \frac{\text{Var}_s + \text{Var}_d}{\text{Var}_z}$$

$t_{\text{FS}} = 0.2177$ **<full sibs**

Eq 18.33a-b:

"If all resemblance between relatives were due to additive genetic variance, we would expect $t_{\text{FS}} \sim 2t_{\text{PHS}}$" LW p. 578.

^ verified LW p. 578

Variance & standard errors of intraclass correlations:

$$\text{Var}_{\text{MSS}} := \frac{2\cdot\text{MSS}^2}{\text{df}_s + 2} \qquad \text{Var}_{\text{MSE}} := \frac{2\cdot\text{MSE}^2}{\text{df}_e + 2} \qquad \text{Var}_{\text{MSD}} := \frac{2\cdot\text{MSD}^2}{\text{df}_d + 2} \qquad \phi := \frac{\left(\dfrac{k_2}{k_3}\right) - 1}{k_1}$$

$$\text{Var}_{\text{tPHS}} := \frac{\text{Var}_{\text{MSS}} + \left(\dfrac{k_2}{k_1^2}\right)\cdot\text{Var}_{\text{MSD}} + \left[1 - \left(\dfrac{k_2}{k_1}\right)\right]^2\cdot\text{Var}_{\text{MSE}}}{\left(k_3\cdot\text{Var}_z\right)^2}$$

$\text{Var}_{\text{tPHS}} = 0.0013$ 0.0098

$$\text{Var}_{\text{tFS}} := \frac{\text{Var}_{\text{MSS}} + k_3^2\cdot\left[\phi^2\cdot\text{Var}_{\text{MSD}} + (1+\phi)^2\cdot\text{Var}_{\text{MSE}}\right]}{\left(k_3\cdot\text{Var}_z\right)^2}$$

$\text{Var}_{\text{tFS}} = 0.0037$ 0.0127

these do NOT match LW p. 578 ^   **LW ^**

Either *they* or *I* have mis-typed these equations...

$SE_{tPHS} := \sqrt{Var_{tPHS}}$ $SE_{tPHS} = 0.0354$ **Standard Errors:**

$SE_{tFS} := \sqrt{Var_{tFS}}$ $SE_{tFS} = 0.0607$ **< these do NOT match LW p. 578 ^**

Note: LW p. 579 mention "more precise" estimates of variance for intraclass correlations based on their Eq 18.37a-b. However, these equations were presented by LW specifically under the context of *balanced design* leaving open definitions of n & M in their equations for this *unbalanced problem*. I don't think their discussion in Problem 3 makes much sense here.

Heritibility:

$h := 4 \cdot t_{PHS}$ $h = 0.1875$

^ confirmed LW p. 579

Test of σ_s^2:   **Works with both balanced or unbalanced data LW p. 567.**

Hypotheses:

H_0: $\sigma_s^2 = 0$

H_1: $\sigma_s^2 <> 0$

Test Statistic:

$$F := \frac{k_1 \cdot MSS + (k_2 - k_1) \cdot MSE}{k_2 \cdot MSD}$$ $F = 1.5157$

Critical Value of the Test:

$\alpha := 0.05$ **< Probability of Type I error must be explicitly set**

Satterthwaite numerator degree of freedom (r) - LW Eq 18.35a-c:

$c_s := \dfrac{k_1}{k_2}$ $c_e := \dfrac{(k_2 - k_1)}{k_2}$ $Q := c_s \cdot MSS + c_e \cdot MSE$ **< Q is a linear combination**

$$r := \frac{Q^2}{\dfrac{(c_s \cdot MSS)^2}{N - 1} + \dfrac{(c_e \cdot MSE)^2}{T - N \cdot M}}$$ $r = 29$

$CV := qF(1 - \alpha, r, df_d)$ $df_d = 56$ **< use of df_d here makes this an** $CV = 1.6699$
                                                        ANOVA Type II test

Decision Rule: **^ F, r, & df_d**
                                                                                 confirmed LW p. 578

IF F > C, THEN REJECT H_0 OTHERWISE ACCEPT H_0

Probability:

$P := 1 - pF(F, r, df_d)$ $P = 0.0908$

Note: LW p. 576 explicitly state: "Recalling that the variance associated with sires is an estimate of $\sigma_A^2/4$, the F ratio defined by Eq 18.35c provides a test for significant <u>additive genetic variance</u>." This is as definite as I can find for the relationship: $4\sigma_s^2 = \sigma_A^2$. The definition clearly distinguishes both *population* (i.e., theoretical) *variances:* $\sigma_s^2, \sigma_d^2, \sigma_e^2$ & σ_z^2 from their *estimates:* Var_s & Var_e and from associated *sample measures:* MSS & MSE. In the context of this study design the *estimate* of σ_s^2, based on observed data, is Var_s defined above. By implication, a test of $\sigma_s^2 = 0$ is the same as a test of $\sigma_A^2 = 4\sigma_s^2 = 0$.

Test of $\sigma_d{}^2$:   **Works with both balanced or unbalanced data LW p. 567.**

Hypotheses:

$$H_0: \quad \sigma_d{}^2 = 0$$

$$H_1: \quad \sigma_d{}^2 <> 0$$

Test Statistic:

$$F := \frac{MSD}{MSE}$$ F = 2.9871

Critical Value of the Test:

$\alpha := 0.01$ **< Probability of Type I error must be explicitly set**

$CV := qF\left(1 - \alpha, df_d, df_e\right)$ $df_d = 56$ $df_e = 695$ CV = 1.5228

Decision Rule:

^ F, df_d, & df_e confirmed p. 578

IF F > C, THEN REJECT H_0 OTHERWISE ACCEPT H_0

As above, LW's table for CV seems a little off...

Probability:

$P := 1 - pF\left(F, r, df_d\right)$ P = 0.0002

Final Estimates:

LW p. 579 last paragraph of Example 3 in Chapter 18 seems to do some heavy lifting...

Thus, we see that the heritability estimates based on paternal half sibs and on the paternal-offspring regression are quite consistent, averaging about 0.15. The fact that maternal half sibs are much more similar than paternal half sibs, and that offspring are much more similar to mothers than to fathers, suggests that maternal effects are a significant source of variation. In the absence of significant epistatic sources of variation, the covariance between maternal half sibs is $\sigma_A^2/4 + \sigma_{E_c}^2$. Thus, multiplying the maternal half-sib correlation by four inflates the heritability estimate by $4\sigma_{E_c}^2/\sigma_z^2$. This suggests that the fraction of the phenotypic variance that is due to variation in maternal effects is approximately $(0.41 - 0.15)/4 = 0.06$. Finally, we recall that twice the intraclass correlation between full sibs (0.436) actually estimates $[\sigma_A^2 + (\sigma_D^2/2) + 2\sigma_{E_c}^2]/\sigma_z^2$. Thus, with our previous results, the contribution to the phenotypic variance from dominance can be estimated as $2[0.436 - 0.15 - (2 \times 0.06)] \simeq 0.33$. In summary, assuming that epistasis is of negligible importance, these results suggest that the variance in developmental rate is approximately partitioned as: 15% additive genetic variance, 33% dominance genetic variance, 7% maternal-effects variance, and 45% unobserved causes (presumably special environmental effects).

h = 0.1875   **< our estimate above**

h := 0.15   **< LW's "average" estimate for this paragraph.**

Ultimate partion of variance seems important. Presumably their description of how to do it has something to do with LW Eq 18.29a-c & 18.31a-c, but I'll need some genetic advice/translation to de-gookify it.