

ORIGIN = 0

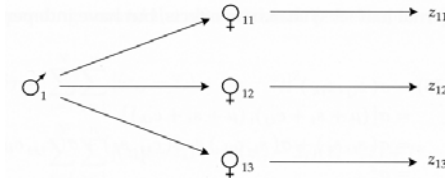
Prototype of Half-Sib Analysis from Lynch & Walsh

W. Stein

Half-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, especially *additive genetic variance*. Example below is derived from Chapter 18 Example 2 from (LW) Michael Lynch & Bruce Walsh 1998 *Genetics and Analysis of Quantitative Traits*, Sinauer Associates. The authors suggest that the ANOVA approaches described in Chapter 18 work as well as any technique for balanced designs involving equal numbers of dams within sires (i.e., replicates). For unbalanced designs, they present corrections based in large measure (but with one noted modification), on Searle et al. 1992 *Variance Components*. The authors also imply that more recent maximum likelihood (ML & REML) techniques have largely supplanted ANOVA as the technique of choice for studies with unbalanced data. See their Chapters 26 & 27 for an overview.

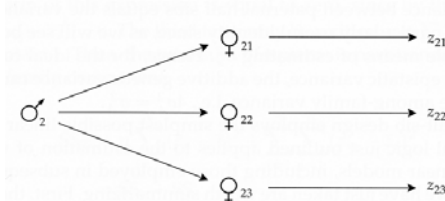
Data Design:

Single offspring are collected from different females mated to a single male, thus comprising "half-sib" families.



ANOVA Model:

$$z_{ij} = \mu + s_i + e_{ij} \quad \text{< One-Way Treatment Effects ANOVA model}$$



Assumptions:

$$e_{ij} \sim N(0, \sigma^2) \quad s_i \sim N(0, \sigma_s^2)$$

all families have equal variance = σ^2

Example: LW Example 2 p. 568

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

Reported Dimensions & ANOVA table results:

$N := 20$ < wild grown self-incompatible plants = number of families $i := 0..N - 1$

$$n_r := \begin{pmatrix} 10 & 5 \\ 9 & 8 \\ 8 & 4 \\ 7 & 1 \\ 6 & 1 \\ 4 & 1 \end{pmatrix} \quad n_r \langle 0 \rangle = \begin{pmatrix} 10 \\ 9 \\ 8 \\ 7 \\ 6 \\ 4 \end{pmatrix} \quad \text{< } n_i = \text{number of progeny = replicates}$$

$$n_r \langle 1 \rangle = \begin{pmatrix} 5 \\ 8 \\ 4 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{< number of families with the associated } n_i$$

$T := 171$ < Total number of observations = replicates summed over all families

$$n_r \langle 0 \rangle^T \cdot n_r \langle 1 \rangle = (171)$$

$MS_S := 0.1156$ among < From original
 $MS_e := 0.0370$ within < ANOVA table

n_i in vector form ^

(10)
 10
 10
 10
 10
 9
 9
 9
 9
 9
 9
 9
 9
 8
 8
 8
 8
 7
 6
 (4)

Table 18.1 giving general formulas for unbalanced designs. Should also work for balanced designs because the latter are a logical subset of the former.

Table 18.1 Summary of a one-way ANOVA involving N independent families, the i th of which contains n_i individuals.

Factor	df	SS	MS	$E(\text{MS})$
Among-families	$N - 1$	$SS_s = \sum_{i=1}^N n_i (\bar{z}_i - \bar{z})^2$	$SS_s / (N - 1)$	$\sigma_c^2 + n_0 \sigma_s^2$
Within-families	$T - N$	$SS_e = \sum_{i=1}^N \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_i)^2$	$SS_e / (T - N)$	σ_e^2
Total	$T - 1$	$SS_T = \sum_{i=1}^N \sum_{j=1}^{n_i} (z_{ij} - \bar{z})^2$	$SS_T / (T - 1)$	σ_z^2

Note: The total sample size is $T = \sum_{i=1}^N n_i$, and $n_0 = [T - (\sum n_i^2 / T)] / (N - 1)$, which reduces to n with equal family sizes. Degrees of freedom are denoted by df, observed sums of squares by SS, and expected mean squares by $E(\text{MS})$.

Calculated Estimates:

$$n_0 := \frac{T - \frac{\sum (n_i)^2}{T}}{N - 1} \qquad n_0 = 8.5367805 \qquad < \text{"weighted family size"} >$$

Estimated variance components:

$$\begin{aligned} \text{Var}_s &:= \frac{MS_s - MS_e}{n_0} & \text{Var}_s &= 0.0092072 & < \text{estimate of } \sigma_s^2 > \\ \text{Var}_e &:= MS_e & \text{Var}_e &= 0.037 & < \text{estimate of } \sigma_e^2 > \\ \text{Var}_z &:= \text{Var}_s + \text{Var}_e & \text{Var}_z &= 0.0462072 & < \text{estimate of } \sigma_z^2 > \end{aligned}$$

Eq 18.12a-c:

Intraclass correlation & heritability:

$$\begin{aligned} t &:= \frac{\text{Var}_s}{\text{Var}_z} & t &= 0.1992593 & < \text{"intraclass correlation"} > \\ h_{sq} &:= 4 \cdot t & h_{sq} &= 0.7970372 & < \text{"heritability"} > \end{aligned}$$

Variance and Standard errors:

Eq 18.21:

$$\begin{aligned} \text{Var}_t &:= \frac{2 \cdot (1 - t)^2 \cdot [1 + (n_0 - 1) \cdot t]^2}{N \cdot n_0 \cdot (n_0 - 1)} & \text{Var}_t &= 0.0062373 & < \text{variance of intraclass correlation} > \\ SE_h &:= 4 \cdot \sqrt{\text{Var}_t} & SE_h &= 0.3159075 & < \text{standard error of heritability} > \end{aligned}$$

Additive genetic variance:

^ all confirmed p. 568

$$\sigma_A := 4 \cdot \left(\frac{MS_s - MS_e}{n_0} \right) \qquad \sigma_A = 0.0368289 \qquad < \text{"additive genetic variance"} >$$

**^ my interpretation based on definition in Eq 18.2 ff
However, this needs independent verification by a qualified geneticist!
LW are not as direct in definition as I might have liked.**

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:

$$F := \frac{MS_s}{MS_e}$$

$$F = 3.1243$$

Critical Value of the Test:

$$\alpha := 0.01 \quad < \text{Probability of Type I error must be explicitly set}$$

$$N - 1 = 19$$

$$T - N = 151$$

$$CV := qF[1 - \alpha, (N - 1), (T - N)]$$

$$CV = 2.0279$$

Decision Rule:

^ all confirmed p. 568

IF $F > C$, THEN REJECT H_0 OTHERWISE ACCEPT H_0 **Probability:**

$$P := 1 - pF(F, N - 1, T - N)$$

$$P = 5.0635 \times 10^{-5}$$

Sampling Variance & Standard Errors for Balanced Data:

$$n_b := 5 \quad < \text{set for appropriate replicate (within family) size}$$

$$\text{Var}_{MSE} := 2 \cdot \frac{(MS_e)^2}{T - N + 2}$$

$$\text{Var}_{MSE} = 1.7895 \times 10^{-5}$$

$$SE_{MSE} := \sqrt{\text{Var}_{MSE}}$$

$$SE_{MSE} = 0.0042$$

$$\text{Var}_{\sigma_A} := \frac{2}{n_b} \cdot \left(\frac{MS_s^2}{N + 1} + \frac{MS_e^2}{T - N + 2} \right)$$

$$\text{Var}_{\sigma_A} = 0.0003$$

$$SE_{\sigma_A} := \sqrt{\text{Var}_{\sigma_A}}$$

$$SE_{\sigma_A} = 0.0161$$

$$\text{Cov}_{MSE\sigma_A} := \frac{2 \cdot MS_e^2}{n_b \cdot (T - N + 2)}$$

$$\text{Cov}_{MSE\sigma_A} = 3.5791 \times 10^{-6}$$

< Eq. 18.20a-c

Confidence Intervals for Balanced Data:

$$\alpha := 0.05 \quad < \text{set as desired}$$

$$n_b := 5 \quad < \text{set for appropriate replicate (within family) size}$$

For Within Family Variance:

$$SS_e := MS_e \cdot (T - N)$$

$$\chi_L := qchisq\left(\frac{\alpha}{2}, T - N\right)$$

$$\chi_L = 118.8714$$

$$\chi_U := qchisq\left(1 - \frac{\alpha}{2}, T - N\right)$$

$$\chi_U = 186.9136$$

$$L_e := \frac{SS_e}{\chi_L} \quad U_e := \frac{SS_e}{\chi_U}$$

$$CI_e := (L_e \quad MS_e \quad U_e)$$

$$CI_e = (0.047 \quad 0.037 \quad 0.0299)$$

For Among Family Variance (Additive Genetic Variance):

$$F := \frac{MS_s}{MS_e}$$

Eq 18.18 also from the test above.

using 9999 for infinity:

$$F_A := qF\left(\frac{\alpha}{2}, N - 1, T - N\right)$$

$$F_A = 0.4568$$

$$F_B := qF\left(\frac{\alpha}{2}, T - N, N - 1\right)$$

$$F_B = 0.5492$$

$$F_{infA} := qF\left(\frac{\alpha}{2}, N - 1, 9999\right)$$

$$F_{infA} = 0.4686$$

$$F_{infB} := qF\left(\frac{\alpha}{2}, 9999, N - 1\right)$$

$$F_{infB} = 0.5779$$

$$L_s := \frac{MS_e}{n_b} \cdot \left[\frac{F}{F_{\text{infA}}} - 1 - \left(\frac{F_A}{F_{\text{infA}}} - 1 \right) \cdot \left(\frac{F_A}{F} \right) \right] \quad L_s = 0.042 \quad < \text{Eq 18.23a-b}$$

$$U_s := \frac{MS_e}{n_b} \cdot \left[F \cdot F_{\text{infB}} - 1 + \left[\left(1 - \frac{F_{\text{infB}}}{F_B} \right) \cdot \left(\frac{1}{F_B} \right) \right] \right] \quad U_s = 0.0053$$

$$CI_{\sigma_A} := (L_s \quad \sigma_A \quad U_s) \quad CI_{\sigma_A} = (0.042 \quad 0.0368 \quad 0.0053) \quad < \text{Note: data above were unbalanced, so these numbers don't apply. They also don't make sense.}$$

For heritability:

$$F_L := qF\left(\frac{\alpha}{2}, N-1, T-N\right) \quad F_L = 0.4568 \quad F_U := \frac{1}{qF\left(\frac{\alpha}{2}, T-N, N-1\right)} \quad F_U = 1.8207$$

$$h_L := 4 \cdot \left[\frac{\left(\frac{F}{F_U}\right) - 1}{\left(\frac{F}{F_U}\right) + n_b - 1} \right] \quad h_U := 4 \cdot \left[\frac{\left(\frac{F}{F_L}\right) - 1}{\left(\frac{F}{F_L}\right) + n_b - 1} \right] \quad CI_h := (h_L \quad h_{sq} \quad h_U) \quad CI_h = (0.501 \quad 0.797 \quad 2.1549)$$

Sampling Variance & Standard Errors for Unbalanced Data:

$$\text{Var}_{\text{MSE}} := 2 \cdot \frac{(MS_e)^2}{T - N + 2}$$

$$\text{Var}_{\text{MSE}} = 1.7895 \times 10^{-5}$$

$$\text{Var}_{\sigma_A} := \frac{2}{n_0 \cdot (N+1)} \cdot \left[\frac{(T-1) \cdot \text{Var}_e^2}{n_0 \cdot (T-N)} + (2 \cdot \text{Var}_e \cdot \text{Var}_s) + \left[\frac{\sum_i (n_i)^2 + \left[\frac{\sum_i (n_i)^2}{T} \right]^2 - \frac{2 \cdot \sum_i (n_i)^3}{T}}{n_0 \cdot (N-1)} \right] \cdot \text{Var}_s^2 \right]$$

$$\text{Var}_{\sigma_A} = 1.7921 \times 10^{-5}$$

$$\text{Cov}_{\text{MSE}\sigma_A} := \frac{2 \cdot MS_e^2}{n_0 \cdot (T - N + 2)}$$

$$\text{Cov}_{\text{MSE}\sigma_A} = 2.0963 \times 10^{-6}$$

Confidence Intervals for Unbalanced Data:

For Within Family Variance:

$$SS_e := MS_e \cdot (T - N)$$

Balanced data approach still valid for unbalanced data.

$$\chi_L := qchisq\left(\frac{\alpha}{2}, T - N\right) \quad \chi_L = 118.8714 \quad \chi_U := qchisq\left(1 - \frac{\alpha}{2}, T - N\right) \quad \chi_U = 186.9136$$

$$L_e := \frac{SS_e}{\chi_L} \quad U_e := \frac{SS_e}{\chi_U} \quad CI_e := (L_e \quad \text{Var}_e \quad U_e) \quad CI_e = (0.047 \quad 0.037 \quad 0.02989)$$

For Among Family Variance (Additive Genetic Variance):

^ backwards? but I may be mis-reading LW p. 562 Eq 18.22

LW p. 567 refer us to "complicated" formulas in Searle et al (1992, p. 76-78) or preferably, it seems, to use of ML/REML techniques...

Prototype in R:**COMMANDS:**

#2008 HALF-SIB ANALYSIS IN LYNCH & WALSH CH18

#DATA

N=20

T=171

n=c(10,10,10,10,10,9,9,9,9,9,9,9,8,8,8,8,7,6,4)

length(n)

MSS=0.1156

MSE=0.0370

#CALCULATED ESTIMATES

 $n0=(T-(t(n) \% \% n)/T)/(N-1)$

n0

Vars=(MSS-MSE)/n0

Vars

Vare=MSE

Vare

Varz=Vars+Vare

Varz

t=Vars/Varz

t

hsq=4*t

hsq

 $Vart=((2*(1-t)^2)*(1+(n0-1)*t)^2)/(N*n0*(n0-1))$

Vart

SEh=4*sqrt(Vart)

SEh

AGV=4*((MSS-MSE)/n0)

AGV

#F-TEST OF BETWEEN-FAMILY VARIANCE=0

alpha=0.01

F=MSS/MSE

F

CV=qf(1-alpha,N-1,T-N)

CV

P=1-pf(F,N-1,T-N)

RESULTS=c(F,CV,P)

RESULTS

#CONFIDENCE INTERVALS

alpha=0.05

#CI FOR Vare=SSE:

SSE=MSE*(T-N)

XL=qchisq(alpha/2,T-N)

XU=qchisq(1-alpha/2,T-N)

Le=SSE/XL

Ue=SSE/XU

RESULTS=c(Le,Vare,Ue)

RESULTS

RETURNS:

> n0

[,1]

[1,] 8.53678

> Vars=(MSS-MSE)/n0

> Vars

[,1]

[1,] 0.009207218

> Vare=MSE

> Vare

[1] 0.037

> Varz=Vars+Vare

> Varz

[,1]

[1,] 0.04620722

> t

[,1]

[1,] 0.1992593

> hsq

[,1]

[1,] 0.7970372

> Vart

[,1]

[1,] 0.006237346

> SEh

[,1]

[1,] 0.3159075

> AGV=4*((MSS-MSE)/n0)

> AGV

[,1]

[1,] 0.03682887

> #F-TEST OF BETWEEN-FAMILY VARIANCE=0

> RESULTS

[1] 3.124324e+00 2.027862e+00 5.063512e-05

> #CONFIDENCE INTERVALS

> alpha=0.05

> #CI FOR Vare=SSE:

> RESULTS

[1] 0.04700036 0.03700000 0.02989082

Other calculations (for balanced data) not attempted...