ORIGIN := 0

Unbalanced Nested ANOVA - Sokal & Rohlf Example

prepared by
Wm Stein

This sheet offers prototyped example of "Full Sib" nested ANOVA appearing in Sokal & Rohlf *Biometry* 3rd Edition (SR) for comparison of their Least Squares direct approach in Box 10.6. with Reduced Maximum Liklihood estimation results of lme() {nlme} function in R, described by J.C. Pinheiro & D.M. Bates *Mixed-Effects Models in S and S-Plus* (PB). Variance Components are calculated comparing quantities in SR with equivalent formulas in M. Lynch & B. Walsh 1998 (LW) *Genetics and analysis of quantitative traits*, Chapter 18.

Reading Data:

X := READPRN("/Data/SRBox10.6.txt")

^ Data reformatted with col 0=Dam, 1=Sire, 2=PH response, 3=Sire means.

A := READPRN("/Data/SRBox10.6DamMeans.txt")

^ Independent factor Dam with col 0=Dam means, 1=n_i.

B := READPRN("/Data/SRBox10.6LitMeans.txt")

^ Nested factor Sire (Litter) with col 0=Sire means, 1=n_{ij}, 2=Dam means.

X = ▮

$GM := \mathrm{mean}\left(X^{\langle 2\rangle}\right)$ GM = ▮

^ grand mean

Sums of Squares:

SS_A independent factor:

$i := 0..\mathrm{length}\left(A^{\langle 0\rangle}\right) - 1$ $a := \mathrm{length}\left(A^{\langle 0\rangle}\right)$

$n := A^{\langle 1\rangle}$ $AM := A^{\langle 0\rangle}$ **< n_i & $means_i$**

$SS_A := \sum_i n_i \cdot \left(AM_i - GM\right)^2$ SS_A = ▮ B = ▮

SS_B nested factor:

$j := 0..\mathrm{length}\left(B^{\langle 0\rangle}\right) - 1$ $b := \mathrm{length}\left(B^{\langle 0\rangle}\right)$

$n := B^{\langle 1\rangle}$ $BM := B^{\langle 0\rangle}$ $AM := B^{\langle 2\rangle}$ **< n_{ij} , B $means_{ij}$, & A $means_i$**

$SS_B := \sum_j n_j \cdot \left(BM_j - AM_j\right)^2$ SS_B = ▮

SS_E within nested error:

$k := 0..\mathrm{length}\left(X^{\langle 0\rangle}\right) - 1$ $r := \mathrm{length}\left(X^{\langle 0\rangle}\right)$ A = ▮

$O := X^{\langle 2\rangle}$ $AM := X^{\langle 3\rangle}$ **< $Response_k$, B $means_k$**

$SS_E := \sum_k \left(O_k - AM_k\right)^2$ SS_E = ▮

SS_T Total SS:

$SS_T := \sum_k \left(O_k - GM\right)^2$ SS_T = ▮

$SS_A + SS_B + SS_E$ = ▮

ANOVA Table:

Sums of Squares:		degrees of freedom:		Mean Squares:	
SS_A	$SS_A = ▮$	$df_A := a - 1$	$df_A = ▮$	$MS_A := \frac{SS_A}{df_A}$	$MS_A = ▮$
$SS_{B(A)}$	$SS_B = ▮$	$df_B := b - a$	$df_B = ▮$	$MS_B := \frac{SS_B}{df_B}$	$MS_B = ▮$
SS_E	$SS_E = ▮$	$df_E := r - b$	$df_E = ▮$	$MS_E := \frac{SS_E}{df_E}$	$MS_E = ▮$
SS_T	$SS_T = ▮$	$df_T := r - 1$	$df_T = ▮$	^ These values verified SR p. 296.	

Tests of Significance:

Using Full and Reduced Linear Models approach KNNL.

For effect in independent variable A:

Null Hypothesis and Alternative:

H_0: Regression coefficient for Treatment A is zero - no independent effect in A

H_1: Regression coefficient not zero - treatment effect is evident in A

Test Statistic:

$F_a := \frac{MS_A}{MS_E}$ $F_a = ▮$ $F_a := \frac{MS_A}{MS_B}$ $F_a = ▮$ **< verified SR p. 296**

^ fixed effects denominator   **^ random effects denominator**

(without Satterthwaith's correction)

Decision Rule:

$\alpha := 0.05$ **< set as desired**

If Fs > F(1-α,df_A, df_E) then Reject H_0, otherwise accept H_0

$qF(1 - \alpha, df_A, df_E) = ▮$

Probability:

$P := \min(pF(F_a, df_A, df_E), 1 - pF(F_a, df_A, df_E))$ $P = ▮$

For effect in nested variable B(A):

Null Hypothesis and Alternative:

H_0: Regression coefficient for Nested B is zero - no effect for nested B(A)

H_1: Regression coefficient not zero - effect evident in B(A)

Test Statistic:

$F_b := \frac{MS_B}{MS_E}$ $F_b = ▮$ **< verified SR p. 297**

Decision Rule:

$\alpha := 0.05$ **< set as desired**

If Fs > F(1-α,df_B, df_E) then Reject H_0, otherwise accept H_0

$qF(1 - \alpha, df_B, df_E) = ▮$

Probability:

$P := \min(pF(F_b, df_B, df_E), 1 - pF(F_b, df_B, df_E))$ $P = ▮$

Calculating Variance Components:

Calculations based on LW pp. 573-574. **Equivalent notation:**

$a = ▮$ **number of outer factor (Dams)** $N := a$ $N = ▮$

$b = ▮$ **number of inner factor (Sires)** $Mbar := \frac{b}{a}$ $Mbar = ▮$

$r = ▮$ **total number of individuals** $T := r$ $T = ▮$

$$i := 0 .. \text{length}\left(A^{\langle 0 \rangle}\right) - 1$$

$$BB := \begin{pmatrix} 2 & 4^2 + 4^2 \\ 2 & 5^2 + 4^2 \\ 3 & 4^2 + 4^2 + 5^2 \\ 2 & 3^2 + 4^2 \\ 3 & 5^2 + 4^2 + 3^2 \\ 3 & 4^2 + 4^2 + 5^2 \\ 3 & 4^2 + 5^2 + 5^2 \\ 2 & 3^2 + 5^2 \\ 2 & 4^2 + 4^2 \\ 2 & 5^2 + 4^2 \\ 2 & 5^2 + 5^2 \\ 2 & 5^2 + 4^2 \\ 3 & 5^2 + 3^2 + 4^2 \\ 3 & 4^2 + 4^2 + 5^2 \\ 3 & 5^2 + 5^2 + 5^2 \end{pmatrix} \qquad BB = \begin{pmatrix} 2 & 32 \\ 2 & 41 \\ 3 & 57 \\ 2 & 25 \\ 3 & 50 \\ 3 & 57 \\ 3 & 66 \\ 2 & 34 \\ 2 & 32 \\ 2 & 41 \\ 2 & 50 \\ 2 & 41 \\ 3 & 50 \\ 3 & 57 \\ 3 & 75 \end{pmatrix}$$

^ n_i 's **^$\Sigma n_{ij}{}^2$'s from counts in B** **^ number of sires within dams**

$$k_1 := \frac{1}{N \cdot (Mbar - 1)} \cdot \left[T - \sum_i \left[\frac{\left(BB^{\langle 1 \rangle}\right)_i}{\left(A^{\langle 1 \rangle}\right)_i}\right]\right] \qquad k_1 = ▮$$

SR n_0

$$k_2 := \frac{1}{N - 1} \cdot \left[\sum_i \left[\frac{\left(BB^{\langle 1 \rangle}\right)_i}{\left(A^{\langle 1 \rangle}\right)_i}\right] - \frac{\sum_i \left(BB^{\langle 1 \rangle}\right)_i}{T}\right] \qquad k_2 = ▮$$

SR n_0'

$$A^{\langle 1 \rangle} = ▮$$

$$k_3 := \frac{1}{N - 1} \cdot \left[T - \frac{\sum_i \left[\left(A^{\langle 1 \rangle}\right)_i\right]^2}{T}\right] \qquad k_3 = ▮$$

SR $(nb)_0$

^ Note: One must use correction (from SR) of LW Table 18.3 for k_3: T (not N) in denominator

^ values verified SR p. 297

From SR pp 296-297 Intermediate calculation quantities 1-4:

$$q1 := r \qquad q1 = ▮ \qquad q3 := \sum_i \left[\left(A^{\langle 1 \rangle}\right)_i\right]^2 \qquad q3 = ▮$$

$$q2 := \sum_i \left(BB^{\langle 1 \rangle}\right)_i \qquad q2 = ▮ \qquad q4 := \sum_i \left[\frac{\left(BB^{\langle 1 \rangle}\right)_i}{\left(A^{\langle 1 \rangle}\right)_i}\right] \qquad q4 = ▮$$

From LW Eq. 18.32a-c & Table 18.3:

Mean Squares	Expected MS	Expected Variance Components	
$MS_A = ▮$	$br\sigma_A{}^2 + r\sigma_B{}^2 + \sigma_E{}^2$	$\sigma s_A := \dfrac{MS_A - MS_E - \left(\frac{k_2}{k_1}\right) \cdot \left(MS_B - MS_E\right)}{k_3}$	$\sigma s_A = ▮$
$MS_B = ▮$	$r\sigma_B{}^2 + \sigma_E{}^2$	$\sigma s_B := \dfrac{\left(MS_B - MS_E\right)}{k_1}$	$\sigma s_B = ▮$
$MS_E = ▮$	$\sigma_E{}^2$	$\sigma s_E := MS_E$	$\sigma s_E = ▮$

values verified SR p. 298 ^

Prototype in R:

```
library(nlme)     # {nlme} for lme() & invervals()
library(ape)      # {ape} for varcomp()
#library(help=ape) # prototype for finding package index
#+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
#CALCULATING VARIANCE COMPONENTS
#READ STRUCTURED DATA TABLE WITH NUMERIC CODED NESTED FACTORS
K=read.table("c:/2008LinearModelsData/SRBox10.6NEWR.txt")
K
attach(K)
Y=PH        #response variable
A=factor(Dam) #OUTER FACTOR A
B=factor(Sire) #INNER FACTOR B inside A
r=dim(K)[[1]] #TOTAL NUMBER OF INDIVIDUALS
r         #quantity 1 in SR
BB=summary(B)
BBsq=BB^2
q2=sum(BBsq)
q2        #quantity 2 in SR
a=length(levels(A))
AA=summary(A)
AAsq=AA^2
q3=sum(AAsq)
q3        #quantity 3 in SR
b=length(levels(B))
L=table(A,B)
BBsum=rowSums(L^2)
M=BBsum/AA
q4=sum(M)
q4        #quantity 4 in SR
k1= (1/(a*((b/a)-1)))*(r-q4)
k1        #quantity k1 in LW
k2= (1/(a-1))*(q4-(q2/r))
k2        #quantity k2 in LW
k3= (1/(a-1))*(r-(q3/r))
k3        #quantity k3 in LW
```

```
> r        #quantity 1 in SR
[1] 160
> q2       #quantity 2 in SR
[1] 708
> q3       #quantity 3 in SR
[1] 1800
> q4       #quantity 4 in SR
[1] 65.68956
> k1       #quantity k1 in LW
[1] 4.286838
> k2       #quantity k2 in LW
[1] 4.37604
> k3       #quantity k3 in LW
[1] 10.625
```

```
#LEAST SQUARES VARIANCE COMPONENTS:
sigA = (MSA-MSE-(k2/k1)*(MSB-MSE))/k3
sigA  #variance component for outside factor A
sigB = (MSB-MSE)/k1
sigB  #variance component for nested factor B inside A
sigE = MSE
sigE  #variance component for error
LEASTSQvar=c(sigA,sigB,sigE)
#MAXIMUM LIKLIHOOD VARIANCE COMPONENTS using lme():
FMe=lme(Y~1,random=~1|A/B,data=K)
FMe
summary(FMe)
varcomp(FMe)    # {ape} Variance Components calculated
MAXLIKvar=c(varcomp(FMe)[1],varcomp(FMe)[2],varcomp(FMe)[3])
intervals(FMe)  # {nlme} PB Confidence Intervals
PBINTlower=c(intervals(FMe)$reStruct$A[1,1]^2,
            intervals(FMe)$reStruct$B[1,1]^2,
            intervals(FMe)$sigma[1]^2)
PBINTest  =c(intervals(FMe)$reStruct$A[1,2]^2,
            intervals(FMe)$reStruct$B[1,2]^2,
            intervals(FMe)$sigma[2]^2)
PBINTupper = c(intervals(FMe)$reStruct$A[1,3]^2,
            intervals(FMe)$reStruct$B[1,3]^2,
            intervals(FMe)$sigma[3]^2)
results=cbind(MS,LEASTSQvar,MAXLIKvar,PBINTlower,PBINTest,PBINTupper)
#LEAST SQUARES AND MAXIMUM LIKLIHOOD ESTIMATES OF VARIANCE COMPONENTS:
results
detach(K)
```

RETURNS:

```
> #LEAST SQUARES AND MAXIMUM LIKLIHOOD ESTIMATES OF VARIANCE COMPONENTS:
> results
```

	MS	LEASTSQvar	MAXLIKvar	PBINTlower	PBINTest	PBINTupper
A	127.15526	8.521288	8.895675	3.0426872	8.895675	26.00761
B	36.37440	2.714905	2.645517	0.3670645	2.645517	19.06683
Within	24.73604	24.736043	24.807877	19.3155079	24.807877	31.86200

KEY TO THE ABOVE TABLE:

MS: mean squares from lm() anova() table.
LEASTSQvar: Least Squares Variance Components calculated from SR & LW formulas.
MAXLIKvar: REML variance components reported by varcomp() {ape}.
PBINT: Approximate 95% Confidence Interval reported by intervals() {nlme}.
(Note all values in intervals() are reported in Standard Deviations, so are squared in this table to allow direct comparison with other values.)

- **PBINTlower: lower C.I. limit.**
- **PBINTest: Point Estimate from intervals() {nlme}.**
- **PBINTupper: upper C.I. limit.**