ORIGIN := 0

Split-plot Designs

Split-plot designs are *not* distinguished from factorial (i.e. fully crossed multi-factor) designs by the structure of the data as observed in a data table. Both appear identical, but the difference resides in how the data are collected. In many situations, it is difficult to apply full randomization to all crossed factors because some experimental or observational conditions are harder to apply than others. Thus, in classic agricultural studies, fields are divided into plots related to a hard-to-apply factor (such as irrigation) and within plots, subplots are defined by levels of more easily applied treatments (such as levels of fertilizer). Similar data structures occur in observational studies because observations occur naturally in a hierarchical arrangement such as teachers within schools within districts, etc. All of these kinds of studies have a complex covariance structure between sets of observations that violate the assumptions of crossed-factorial design (constant variance and no covariance between observations). Instead, in advance of analyzing the data, observations can be considered to have different variances for observations at different levels of the hierarchy, and observations within levels of factors are correlated. In such situations, the individual cases (lines within a dataset) are said to be "pseudoreplicated".

"Split-plot design" terminology is usually applied to experimental data involving hierarchical spatial arrangements. However, the overall approach may encompass many other experimental and observational situations variously termed "hierarchical modeling", "repeated measures designs", "nested designs" and others. In addition to terminology, differences also reside in whether the researcher views some or all individual factors to be fixed versus random, and whether some factors comprise numerical data (as in regression). Differences also arise in what the researcher wishes to extract from the analysis including treatment differences for multiple fixed factors with or without interaction, regression coefficients for numerical relationships, and estimates of variance components for random factors.

Analyzed here are classic examples of split-plot designs of differing complexity depending on the number of hierarchical factors involved, and whether the studies are "balanced" or not. Balanced data, consisting of equal numbers of smallest scale replicates per set of treatment factors, can be handled by standard ANOVA methods because sums of squares for treatments or treatments within other factors sum directly to total SS (i.e., SS's are"orthogonal"). For unbalanced data, correction factors may be sometimes be applied to standard least-squares (LS) regressions. However, newer methods and software such as R's lme() {nlme} and lmer() {lme4} functions involving maximum liklihood (ML) and restricted maximul liklihood (REML) iterative maximizations have largely supplanted LS.

Example - Balanced Case:

Example from W. N. Venables & B. D. Ripley 2002. *Modern Applied Statistics with S, p. 282.* Terminology from M.H. Kutner, C. J. Nachtsheim, J. Neter & W. Li 2005 *Applied Linear Statistical Models 5th Ed.* (KNNL).

KNNL p. 1162 asks us to note that this "split-plot" design is identical to a "two-factor experiment with repeated measures on one factor" design p. 1140, with changes in terminology.

Model:

$$Y_{ijk} = \mu + \rho_{i(j)} + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk}$$

where:

μ = a constant level (intercept)

$\rho_{i(j)}$ = random variable ~ $N(0,\sigma_\rho^2)$

α_j = fixed constant

β_k = fixed constant

$(\alpha\beta)_{jk}$ = fixed constant

ε_{ijk} = random variable ~ $N(0,\sigma^2)$

Restrictions:

$\Sigma\alpha_j = 0$

$\Sigma\beta_k = 0$

$\Sigma(\alpha\beta)_{jk} = 0$ for all j and for all k

Model Dimensions:

$s := 6$ **< number of random blocks observed in $\rho_{i(j)}$** $i := 1 .. s$

$a := 3$ **< number of treatments for fixed factor α_j** $j := 1 .. a$

$b := 4$ **< number of treatments for fixed factor β_k** $k := 1 .. b$

ANOVA Sums of Squares:

Source of Variation: **from KNNL p. 1142, not evaluated here:**

Factor A: $SSA = bs\Sigma_j(Y^{bar}.j. - Y^{bar}...)^2$

^ mean for each class of ouside factor A minus grand mean.

Factor B: $SSB = as\Sigma_k(Y^{bar}..k - Y^{bar}...)^2$

^ mean for each class of inside factor B minus grand mean.

AB Interactions: $SSAB = s\Sigma_j\Sigma_k(Y^{bar}.jk - Y^{bar}.j. - Y^{bar}..k + Y^{bar}...)^2$

^ jk factor block means

Subjects/Blocks within A: $SSS(A) = b\Sigma_i\Sigma_j(X^{bar}ij. - Y^{bar}.j.)^2$

^ ij block means - factor A class means

Error: $SSB.S(A) = \Sigma_i\Sigma_j\Sigma_k(Yijk - Y^{bar}.jk - Y^{bar}ij. + Y^{bar}.j.)^2$

^ observation minus factor AB block means minus ij block means + class A means

Total: $SSTO = \Sigma_i\Sigma_j\Sigma_k(Yijk - Y^{bar}...)^2$

^ observation minus grand mean

Degrees of Freedom:

Factor A: $a - 1 = 2$

Factor B: $b - 1 = 3$

AB Interactions: $(a - 1) \cdot (b - 1) = 6$

Subjects/Blocks within A: $a \cdot (s - 1) = 15$

Error: $a \cdot (s - 1) \cdot (b - 1) = 45$

Total: $a \cdot b \cdot s - 1 = 71$

```
> oats
     B          V      N   Y
1    I    Victory 0.0cwt 111
2    I    Victory 0.2cwt 130
3    I    Victory 0.4cwt 157
4    I    Victory 0.6cwt 174
5    I Golden.rain 0.0cwt 117
6    I Golden.rain 0.2cwt 114
7    I Golden.rain 0.4cwt 161
8    I Golden.rain 0.6cwt 141
9    I  Marvellous 0.0cwt 105
10   I  Marvellous 0.2cwt 140
11   I  Marvellous 0.4cwt 118
12   I  Marvellous 0.6cwt 156
13  II    Victory 0.0cwt  61
14  II    Victory 0.2cwt  91
15  II    Victory 0.4cwt  97
16  II    Victory 0.6cwt 100
17  II Golden.rain 0.0cwt  70
18  II Golden.rain 0.2cwt 108
19  II Golden.rain 0.4cwt 126
20  II Golden.rain 0.6cwt 149
21  II  Marvellous 0.0cwt  96
22  II  Marvellous 0.2cwt 124
23  II  Marvellous 0.4cwt 121
24  II  Marvellous 0.6cwt 144
25 III    Victory 0.0cwt  68
26 III    Victory 0.2cwt  64
27 III    Victory 0.4cwt 112
28 III    Victory 0.6cwt  86
29 III Golden.rain 0.0cwt  60
30 III Golden.rain 0.2cwt 102
31 III Golden.rain 0.4cwt  89
32 III Golden.rain 0.6cwt  96
33 III  Marvellous 0.0cwt  89
34 III  Marvellous 0.2cwt 129
35 III  Marvellous 0.4cwt 132
36 III  Marvellous 0.6cwt 124
37  IV    Victory 0.0cwt  74
38  IV    Victory 0.2cwt  89
39  IV    Victory 0.4cwt  81
40  IV    Victory 0.6cwt 122
41  IV Golden.rain 0.0cwt  64
42  IV Golden.rain 0.2cwt 103
43  IV Golden.rain 0.4cwt 132
44  IV Golden.rain 0.6cwt 133
45  IV  Marvellous 0.0cwt  70
46  IV  Marvellous 0.2cwt  89
47  IV  Marvellous 0.4cwt 104
48  IV  Marvellous 0.6cwt 117
49   V    Victory 0.0cwt  62
50   V    Victory 0.2cwt  90
51   V    Victory 0.4cwt 100
52   V    Victory 0.6cwt 116
53   V Golden.rain 0.0cwt  80
54   V Golden.rain 0.2cwt  82
55   V Golden.rain 0.4cwt  94
56   V Golden.rain 0.6cwt 126
57   V  Marvellous 0.0cwt  63
58   V  Marvellous 0.2cwt  70
59   V  Marvellous 0.4cwt 109
60   V  Marvellous 0.6cwt  99
61  VI    Victory 0.0cwt  53
62  VI    Victory 0.2cwt  74
63  VI    Victory 0.4cwt 118
64  VI    Victory 0.6cwt 113
65  VI Golden.rain 0.0cwt  89
66  VI Golden.rain 0.2cwt  82
67  VI Golden.rain 0.4cwt  86
68  VI Golden.rain 0.6cwt 104
69  VI  Marvellous 0.0cwt  97
70  VI  Marvellous 0.2cwt  99
71  VI  Marvellous 0.4cwt 119
72  VI  Marvellous 0.6cwt 121
```

F-Tests for Fixed Factors:

Hypotheses:

H_0:

all $(\alpha\beta)_{jk}$ = 0 for A with B interactions
all α_j = 0 for fixed factor A
all β_k = 0 for fixed factor B within A

H_1:

all $(\alpha\beta)_{jk}$ <> 0 for A with B interactions
all α_j <> 0 for fixed factor A
all β_k <> 0 for fixed factor B within A

Test Statistics:

$F_{\alpha\beta}$ = MSAB/MSB.S(A)
F_{α} = MSA/MSS(A)
$F\beta$ = MSB/MSB.S(A)

Critical Values for the Tests:

$\alpha := 0.05$ **< Type I error rate must be explicitly set.**

$C_{\alpha\beta}$ = qF(1-α,(a-1)(b-1),a(s-1)(b-1))
$C\alpha$ = qF(1-α, (a-1),a(s-1))
$C\beta$ = qF(1-α, (b-1),a(s-1)(b-1))

Decision Rules:

If: $F_{\alpha\beta} > C_{\alpha\beta}$ **Then Reject H_0**
$F_{\alpha} > C_{\alpha}$
$F_{\beta} > C_{\beta}$

Note: In R, these tests are obtained by the functions:
anova() for serial SS and tests, and
Anova() {car} for margianal SS and tests.

See *Biostatistics* Worksheet 400 for details.

Probabilities:

$P_{\alpha\beta}$ = 1-pF($F_{\alpha\beta}$,(a-1)(b-1),a(s-1)(b-1))
P_{α} = 1-pF(F_{α},(a-1),a(s-1))
P_{β} = 1-pF(F_{α},(b-1),a(s-1)(b-1))

t-Tests for Fixed Regression Terms:

These tests are directed toward individual regression slope parameters underlying ANOVA. They are marginal tests to determine whether particular slopes are zero. See *2010 Biostatistics 38* for details.

Hypotheses:

H_0: - a single regression coefficient β is 0 meaning that this regressor is unimportant.
H_1: - non-zero effect for this regressor.

Test Statistics:

$t_j = b_j/sb_j$ **< estimate regression coefficient divided by standard error for that coefficient**

Sampling Distributions:

If Assumptions hold and H_0 is true, then t ~t(n-k)

where: n = number of cases in dataset
k = number of independent variables in model

Prototype in R:

```
#SPLIT-PLOT ANOVA
setwd("c:/DATA/Models/")
#Oats in Venables & Ripley 2002 p. 282
library(MASS)
oats
attach(oats)
S=B  #subjects/blocks
A=V #outside factor
B=N #inside factor
LM1=aov(Y~A*B+Error(S/A))
summary(LM1)

#USING LINEAR MIXED MODELS:
library(nlme)
library(lme4)
LME1a=lme(Y~A*B,random=~1|S/A)
LME1b=lmer(Y~A*B+(1|S/A))
anova(LME1a)
anova(LME1b)
summary(LME1a)
summary(LME1b)
```

< Note: Variables S,A,B are treated as factors by R because values in the dataset are alphanumeric. The authors use an *ordered factor* variable for N, but results here are the same.

```
> summary(LM1)

Error: S
          Df  Sum Sq Mean Sq F value Pr(>F)
Residuals  5 15875.3  3175.1

Error: S:A
          Df Sum Sq Mean Sq F value Pr(>F)
A          2 1786.4   893.2  1.4853 0.2724
Residuals 10 6013.3   601.3

Error: Within
          Df  Sum Sq Mean Sq F value    Pr(>F)
B          3 20020.5  6673.5 37.6856 2.458e-12 ***
A:B        6   321.7    53.6  0.3028    0.9322
Residuals 45  7968.7   177.1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

^ The ANOVA reports strata with error partitioned according to stratum. Sums of Squares and degrees of freedom for errors at each level presumably sum to total SSE and df_E.

```
> anova(LME1a)
            numDF denDF  F-value p-value
(Intercept)     1    45 245.14299  <.0001
A               2    10   1.48534  0.2724
B               3    45  37.68561  <.0001
A:B             6    45   0.30282  0.9322
```

```
> anova(LME1b)
Analysis of Variance Table
    Df  Sum Sq Mean Sq F value
A    2   526.1   263.0  1.4853
B    3 20020.5  6673.5 37.6856
A:B  6   321.8    53.6  0.3028
```

^ Similar and complemetary reports by lme() and lmer() functions in R:

```
> summary(LME1a)
Linear mixed-effects model fit by REML
 Data: NULL
       AIC      BIC    logLik
  559.0285 590.4437 -264.5143

Random effects:
 Formula: ~1 | S
        (Intercept)
StdDev:    14.64496

 Formula: ~1 | A %in% S
        (Intercept) Residual
StdDev:    10.29862 13.30727

Fixed effects: Y ~ A * B
                      Value Std.Error DF   t-value p-value
(Intercept)        80.00000  9.106958 45  8.784492  0.0000
AMarvellous         6.66667  9.715026 10  0.686222  0.5082
AVictory           -8.50000  9.715026 10 -0.874933  0.4021
B0.2cwt            18.50000  7.682958 45  2.407927  0.0202
B0.4cwt            34.66667  7.682958 45  4.512151  0.0000
B0.6cwt            44.83333  7.682958 45  5.835426  0.0000
AMarvellous:B0.2cwt  3.33333 10.865343 45  0.306786  0.7604
AVictory:B0.2cwt   -0.33333 10.865343 45 -0.030679  0.9757
AMarvellous:B0.4cwt -4.16667 10.865343 45 -0.383482  0.7032
AVictory:B0.4cwt    4.66667 10.865343 45  0.429500  0.6696
AMarvellous:B0.6cwt -4.66667 10.865343 45 -0.429500  0.6696
AVictory:B0.6cwt    2.16667 10.865343 45  0.199411  0.8428
```

```
> summary(LME1b)
Linear mixed model fit by REML
Formula: Y ~ A * B + (1 | S/A)
 AIC   BIC logLik deviance REMLdev
 559 593.2 -264.5    595.9     529
Random effects:
 Groups   Name        Variance Std.Dev.
 A:S      (Intercept) 106.06   10.299
 S        (Intercept) 214.48   14.645
 Residual             177.08   13.307
Number of obs: 72, groups: A:S, 18; S, 6

Fixed effects:
                    Estimate Std. Error t value
(Intercept)          80.0000     9.1064   8.785
AMarvellous           6.6667     9.7150   0.686
AVictory             -8.5000     9.7150  -0.875
B0.2cwt              18.5000     7.6830   2.408
B0.4cwt              34.6667     7.6830   4.512
B0.6cwt              44.8333     7.6830   5.835
AMarvellous:B0.2cwt   3.3333    10.8653   0.307
AVictory:B0.2cwt     -0.3333    10.8653  -0.031
AMarvellous:B0.4cwt  -4.1667    10.8653  -0.383
AVictory:B0.4cwt      4.6667    10.8653   0.430
AMarvellous:B0.6cwt  -4.6667    10.8653  -0.430
AVictory:B0.6cwt      2.1667    10.8653   0.19
```

Example - Balanced Case:

Example from Crawley 2007 *The R Book*.

Split-plot example "splityield.txt" in his set of datafiles.

Prototype in R:

```
#SPLIT-PLOT ANOVA
Y=read.table("splityield.txt",header=T)
Y
attach(Y)
Y=yield
A=factor(block)
B=factor(irrigation)
C=factor(density)
D=factor(fertilizer)

#CROSSED DESIGN:

LM1=lm(Y~B*C*D)  #A (blocks) ignored
anova(LM1)
```

```
> anova(LM1)
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
B          1 8277.6  8277.6 59.5746 2.813e-10 ***
C          2 1758.4   879.2  6.3276 0.0033972 **
D          2 1977.4   988.7  7.1160 0.0018070 **
B:C        2 2747.0  1373.5  9.8853 0.0002197 ***
B:D        2  953.4   476.7  3.4310 0.0395615 *
C:D        4  304.9    76.2  0.5486 0.7008151
B:C:D      4  234.7    58.7  0.4223 0.7918283
Residuals 54 7503.0   138.9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
LM2=aov(Y~B*C*D) #same result as above using aov()
summary(LM2)
```

```
> summary(LM2)

            Df Sum Sq Mean Sq F value    Pr(>F)
B            1 8277.6  8277.6 59.5746 2.813e-10 ***
C            2 1758.4   879.2  6.3276 0.0033972 **
D            2 1977.4   988.7  7.1160 0.0018070 **
B:C          2 2747.0  1373.5  9.8853 0.0002197 ***
B:D          2  953.4   476.7  3.4310 0.0395615 *
C:D          4  304.9    76.2  0.5486 0.7008151
B:C:D        4  234.7    58.7  0.4223 0.7918283
Residuals   54 7503.0   138.9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

> Y

	yield	block	irrigation	density	fertilizer
1	90	A	control	low	N
2	95	A	control	low	P
3	107	A	control	low	NP
4	92	A	control	medium	N
5	89	A	control	medium	P
6	92	A	control	medium	NP
7	81	A	control	high	N
8	92	A	control	high	P
9	93	A	control	high	NP
10	80	A	irrigated	low	N
11	87	A	irrigated	low	P
12	100	A	irrigated	low	NP
13	121	A	irrigated	medium	N
14	110	A	irrigated	medium	P
15	119	A	irrigated	medium	NP
16	78	A	irrigated	high	N
17	98	A	irrigated	high	P
18	122	A	irrigated	high	NP
19	83	B	control	low	N
20	80	B	control	low	P
21	95	B	control	low	NP
22	98	B	control	medium	N
23	98	B	control	medium	P
24	106	B	control	medium	NP
25	74	B	control	high	N
26	81	B	control	high	P
27	74	B	control	high	NP
28	102	B	irrigated	low	N
29	109	B	irrigated	low	P
30	105	B	irrigated	low	NP
31	99	B	irrigated	medium	N
32	94	B	irrigated	medium	P
33	123	B	irrigated	medium	NP
34	136	B	irrigated	high	N
35	133	B	irrigated	high	P
36	132	B	irrigated	high	NP
37	85	C	control	low	N
38	88	C	control	low	P
39	88	C	control	low	NP
40	112	C	control	medium	N
41	104	C	control	medium	P
42	91	C	control	medium	NP
43	82	C	control	high	N
44	78	C	control	high	P
45	94	C	control	high	NP
46	60	C	irrigated	low	N
47	104	C	irrigated	low	P
48	114	C	irrigated	low	NP
49	90	C	irrigated	medium	N
50	118	C	irrigated	medium	P
51	113	C	irrigated	medium	NP
52	119	C	irrigated	high	N
53	122	C	irrigated	high	P
54	136	C	irrigated	high	NP
55	86	D	control	low	N
56	78	D	control	low	P
57	89	D	control	low	NP
58	79	D	control	medium	N
59	86	D	control	medium	P
60	87	D	control	medium	NP
61	85	D	control	high	N
62	89	D	control	high	P
63	83	D	control	high	NP
64	73	D	irrigated	low	N
65	114	D	irrigated	low	P
66	114	D	irrigated	low	NP
67	109	D	irrigated	medium	N
68	131	D	irrigated	medium	P
69	126	D	irrigated	medium	NP
70	116	D	irrigated	high	N
71	136	D	irrigated	high	P
72	133	D	irrigated	high	NP

#WITH ERROR STRATUM A SPECIFIED:
LM3=aov(Y~B*C*D+Error(A))
summary(LM3)

> summary(LM3)

```
Error: A
          Df  Sum Sq Mean Sq F value Pr(>F)
Residuals  3 194.444  64.815

Error: Within
          Df Sum Sq Mean Sq F value    Pr(>F)
B          1 8277.6  8277.6 57.7618 6.157e-10 ***
C          2 1758.4   879.2  6.1350 0.0040972 **
D          2 1977.4   988.7  6.8994 0.0022288 **
B:C        2 2747.0  1373.5  9.5845 0.0002926 ***
B:D        2  953.4   476.7  3.3266 0.0438563 *
C:D        4  304.9    76.2  0.5319 0.7128342
B:C:D      4  234.7    58.7  0.4095 0.8009729
Residuals 51 7308.6   143.3
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

< Variance including sum of squares and degrees of freedom associated with variable A (block) is now partitioned out of Residual SS.

#FULL FACTORIAL MODEL
#WITH ENTIRE HIERARCHICAL STRUCTURE CORRECTLY SPECIFIED:
LM4=aov(Y~B*C*D+Error(A/B/C))
summary(LM4)

> summary(LM4)

```
Error: A
          Df  Sum Sq Mean Sq F value Pr(>F)
Residuals  3 194.444  64.815

Error: A:B
          Df Sum Sq Mean Sq F value  Pr(>F)
B          1 8277.6  8277.6  17.590 0.02473 *
Residuals  3 1411.8   470.6
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: A:B:C
          Df  Sum Sq Mean Sq F value  Pr(>F)
C          2 1758.36  879.18  3.7842 0.05318 .
B:C        2 2747.03 1373.51  5.9119 0.01633 *
Residuals 12 2787.94  232.33
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: Within
          Df  Sum Sq Mean Sq F value    Pr(>F)
D          2 1977.44  988.72 11.4493 0.0001418 ***
B:D        2  953.44  476.72  5.5204 0.0081078 **
C:D        4  304.89   76.22  0.8826 0.4840526
B:C:D      4  234.72   58.68  0.6795 0.6106672
Residuals 36 3108.83   86.36
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

< Residual sums of squares for all levels add to total SS in LM1 above.

Significance in tests changes as pseudoreplication is removed from the study. This is the appropriate split-plot analysis for balanced data verified in Crawley 2007 p. 470. For unbalanced data, functions lme() {nlme} or lmer() {lme4} are instead recommended.

From this starting point, he shows additional model simplification in lme() by removing elements from the full factorial model B*C*D.

Note: it is unnecessary (but not harmful) to specify smallest level of the hierarchical structure in the Error(A/B/C/D) term since the smallest level is already assumed by R.

```
#USING LINEAR MIXED MODELS:
library(nlme)
library(lme4)
LM6=lme(Y~B*C*D, random=~1|A/B/C)
anova(LM6)

LM7=lmer(Y~B*C*D+(1|A/B/C))
anova(LM7)
```

Corresponding ANOVA Results:

lme() result:

```
> anova(LM6)
            numDF denDF   F-value p-value
(Intercept)     1    36 2674.6301  <.0001
B               1     3   30.9207  0.0115
C               2    12    3.7842  0.0532
D               2    36   11.4493  0.0001
B:C             2    12    5.9119  0.0163
B:D             2    36    5.5204  0.0081
C:D             4    36    0.8826  0.4841
B:C:D           4    36    0.6795  0.6107
```

lmer() result:

```
> anova(LM7)
Analysis of Variance Table
      Df  Sum Sq Mean Sq F value
B      1 2670.19 2670.19 30.9206
C      2  653.58  326.79  3.7842
D      2 1977.44  988.72 11.4493
B:C    2 1021.07  510.53  5.9119
B:D    2  953.44  476.72  5.5204
C:D    4  304.89   76.22  0.8826
B:C:D  4  234.72   58.68  0.6795
```

^ For this balanced case, the two procedures produce mostly identical results to each other and to the split-plot aov() analysis above (only the F-value for B differs much). What the anova() wrapper reports for the mixed-model functions is different but complementary.

$$1 - \mathrm{pF}(30.9207, 1, 3) = 0.011474$$

^ for B

< probabilities may be calculated using the corresponding pf() function in R, specifying F value and associated numerator and denominator degrees of freedom.

Comparing Linear Models as Regression:

```
#COMPARING FIXED EFFECTS
#MARGINAL t-TESTS:
summary(LM1)
summary(LM6)
summary(LM7)
```

Note: hierarchical specification of variance is not possible with lm().

```
> summary(LM1)

Call:
lm(formula = Y ~ B * C * D)

Residuals:
   Min      1Q Median     3Q    Max
-34.25  -5.75   0.50   5.75  23.75

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)                80.500      5.894  13.659  < 2e-16 ***
Birrigated                 31.750      8.335   3.809 0.000359 ***
Clow                        5.500      8.335   0.660 0.512143
Cmedium                    14.750      8.335   1.770 0.082432 .
DNP                         5.500      8.335   0.660 0.512143
DP                          4.500      8.335   0.540 0.591491
Birrigated:Clow           -39.000     11.787  -3.309 0.001673 **
Birrigated:Cmedium        -22.250     11.787  -1.888 0.064459 .
Birrigated:DNP             13.000     11.787   1.103 0.274976
Birrigated:DP               5.500     11.787   0.467 0.642665
Clow:DNP                    3.250     11.787   0.276 0.783818
Cmedium:DNP                -6.750     11.787  -0.573 0.569264
Clow:DP                    -5.250     11.787  -0.445 0.657820
Cmedium:DP                 -5.500     11.787  -0.467 0.642665
Birrigated:Clow:DNP         7.750     16.670   0.465 0.643867
Birrigated:Cmedium:DNP      3.750     16.670   0.225 0.822863
Birrigated:Clow:DP         20.000     16.670   1.200 0.235470
Birrigated:Cmedium:DP       4.000     16.670   0.240 0.811276
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.79 on 54 degrees of freedom
Multiple R-squared: 0.6842,     Adjusted R-squared: 0.5847
F-statistic: 6.881 on 17 and 54 DF,  p-value: 2.298e-08
```

> summary(LM6)

```
Linear mixed-effects model fit by REML
 Data: NULL
       AIC      BIC    logLik
  481.6212 525.3789 -218.8106

Random effects:
 Formula: ~1 | A
        (Intercept)
StdDev: 0.0006600339

 Formula: ~1 | B %in% A
        (Intercept)
StdDev:    1.982461

 Formula: ~1 | C %in% B %in% A
        (Intercept) Residual
StdDev:    6.975554 9.292805

Fixed effects: Y ~ B * C * D
                        Value Std.Error DF   t-value p-value
(Intercept)             80.50  5.893741 36 13.658558  0.0000
Birrigated              31.75  8.335008  3  3.809234  0.0318
Clow                     5.50  8.216282 12  0.669403  0.5159
Cmedium                 14.75  8.216282 12  1.795216  0.0978
DNP                      5.50  6.571005 36  0.837010  0.4081
DP                       4.50  6.571005 36  0.684827  0.4978
Birrigated:Clow        -39.00 11.619577 12 -3.356404  0.0057
Birrigated:Cmedium     -22.25 11.619577 12 -1.914872  0.0796
Birrigated:DNP          13.00  9.292805 36  1.398932  0.1704
Birrigated:DP            5.50  9.292805 36  0.591856  0.5576
Clow:DNP                 3.25  9.292805 36  0.349733  0.7286
Cmedium:DNP             -6.75  9.292805 36 -0.726368  0.4723
Clow:DP                 -5.25  9.292805 36 -0.564953  0.5756
Cmedium:DP              -5.50  9.292805 36 -0.591856  0.5576
Birrigated:Clow:DNP      7.75 13.142011 36  0.589712  0.5591
Birrigated:Cmedium:DNP   3.75 13.142011 36  0.285344  0.7770
Birrigated:Clow:DP      20.00 13.142011 36  1.521837  0.1368
Birrigated:Cmedium:DP    4.00 13.142011 36  0.304367  0.7626

Number of Observations: 72
Number of Groups:
              A        B %in% A C %in% B %in% A
              4               8              24
```

> summary(LM7)

```
Linear mixed model fit by REML
Formula: Y ~ B * C * D + (1 | A/B/C)
   AIC   BIC logLik deviance REMLdev
 481.6 531.7 -218.8    529.5   437.6
Random effects:
 Groups   Name        Variance   Std.Dev.
 C:(B:A)  (Intercept) 4.8657e+01 6.9755e+00
 B:A      (Intercept) 3.9306e+00 1.9826e+00
 A        (Intercept) 7.0693e-13 8.4079e-07
 Residual             8.6356e+01 9.2928e+00
Number of obs: 72, groups: C:(B:A), 24; B:A, 8; A, 4

Fixed effects:
                       Estimate Std. Error t value
(Intercept)              80.500      5.894  13.659
Birrigated               31.750      8.335   3.809
Clow                      5.500      8.216   0.669
Cmedium                  14.750      8.216   1.795
DNP                       5.500      6.571   0.837
DP                        4.500      6.571   0.685
Birrigated:Clow         -39.000     11.620  -3.356
Birrigated:Cmedium      -22.250     11.620  -1.915
Birrigated:DNP           13.000      9.293   1.399
Birrigated:DP             5.500      9.293   0.592
Clow:DNP                  3.250      9.293   0.350
Cmedium:DNP              -6.750      9.293  -0.726
Clow:DP                  -5.250      9.293  -0.565
Cmedium:DP               -5.500      9.293  -0.592
Birrigated:Clow:DNP       7.750     13.142   0.590
Birrigated:Cmedium:DNP    3.750     13.142   0.285
Birrigated:Clow:DP       20.000     13.142   1.522
Birrigated:Cmedium:DP     4.000     13.142   0.304
```

^ In this balanced case, results are equivalent, although small differences are noted in very small variance components for the mixed-models.

Example - Unbalanced Case:

```
#UNBALANCED EXAMPLE
#Crawley 2007 p. 472
Y=read.table("splityieldunbalanced.txt")
Y
attach(Y)
Y=yield
A=factor(block)
B=factor(irrigation)
C=factor(density)
D=factor(fertilizer)
```

```
LMu1=lm(Y~B*C*D)
LMu4=aov(Y~B*C*D+Error(A/B/C))
LMu6=lme(Y~B*C*D, random=~1|A/B/C)
# lme() doesn't like NA in dataset
LMu7=lmer(Y~B*C*D+(1|A/B/C))

anova(LMu1)
anova(LMu6)
anova(LMu7)
summary(LMu4)

summary(LMu1)
summary(LMu6)
summary(LMu7)
```

> anova(LMu1)

```
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
B          1 7826.2  7826.2 55.6097 8.292e-10 ***
C          2 1653.2   826.6  5.8736 0.0049652 **
D          2 1869.2   934.6  6.6409 0.0026690 **
B:C        2 2771.5  1385.8  9.8467 0.0002311 ***
B:D        2  911.5   455.7  3.2383 0.0471141 *
C:D        4  337.5    84.4  0.5995 0.6646007
B:C:D      4  228.1    57.0  0.4053 0.8040196
Residuals 53 7458.9   140.7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

> summary(LMu4)

```
Error: A
          Df  Sum Sq Mean Sq F value Pr(>F)
B          1   0.075   0.075   9e-04 0.9788
Residuals  2 167.704  83.852

Error: A:B
          Df Sum Sq Mean Sq F value  Pr(>F)
B          1 7829.9  7829.9 21.9075 0.04274 *
C          1  564.4   564.4  1.5792 0.33576
Residuals  2  714.8   357.4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: A:B:C
          Df  Sum Sq Mean Sq   F value  Pr(>F)
C          2 1696.47  848.24    3.4044 0.07066 .
D          1    0.01    0.01 2.774e-05 0.99589
B:C        2 2786.75 1393.37    5.5924 0.02110 *
Residuals 11 2740.72  249.16
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: Within
          Df  Sum Sq Mean Sq F value    Pr(>F)
D          2 1959.36  979.68 11.1171 0.0001829 ***
B:D        2  993.59  496.79  5.6375 0.0075447 **
C:D        4  273.56   68.39  0.7761 0.5482571
B:C:D      4  244.49   61.12  0.6936 0.6014280
Residuals 35 3084.33   88.12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

> anova(LMu6)

```
            numDF denDF   F-value p-value
(Intercept)     1    35 2615.6915  <.0001
B               1     3   29.5564  0.0122
C               2    12    3.5946  0.0598
D               2    35   10.8796  0.0002
B:C             2    12    5.8611  0.0168
B:D             2    35    5.4231  0.0089
C:D             4    35    0.8447  0.5065
B:C:D           4    35    0.6645  0.6208
```

> anova(LMu7)

```
Analysis of Variance Table
      Df  Sum Sq Mean Sq F value
B      1 2610.67 2610.67 29.5567
C      2  635.00  317.50  3.5945
D      2 1921.93  960.96 10.8796
B:C    2 1035.40  517.70  5.8611
B:D    2  958.02  479.01  5.4231
C:D    4  298.43   74.61  0.8447
B:C:D  4  234.79   58.70  0.6645
```

^ F-values correspond closely between the mixed-models, and correspond more-or-less to estimates made with aov() using correctly specified hierarchical variance. F-values do not correspond to F-values in the crossed-design (LMu1) with pseudoreplication.

> summary(LMu1)

```
Call:
lm(formula = Y ~ B * C * D)
Residuals:
    Min      1Q  Median      3Q     Max
-34.250  -5.750   0.500   5.625  23.750
Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)              80.500      5.932  13.571  < 2e-16 ***
Birrigated               31.750      8.389   3.785 0.000393 ***
Clow                      5.500      8.389   0.656 0.514880
Cmedium                  14.750      8.389   1.758 0.084460 .
DNP                       5.500      8.389   0.656 0.514880
DP                        4.500      8.389   0.536 0.593894
Birrigated:Clow         -39.000     11.863  -3.287 0.001798 **
Birrigated:Cmedium      -22.250     11.863  -1.876 0.066229 .
Birrigated:DNP           13.000     11.863   1.096 0.278108
Birrigated:DP             5.500     11.863   0.464 0.644818
Clow:DNP                  3.250     11.863   0.274 0.785182
Cmedium:DNP              -6.750     11.863  -0.569 0.571767
Clow:DP                  -5.250     11.863  -0.443 0.659894
Cmedium:DP               -5.500     11.863  -0.464 0.644818
Birrigated:Clow:DNP       7.750     16.777   0.462 0.646014
Birrigated:Cmedium:DNP    1.833     17.123   0.107 0.915139
Birrigated:Clow:DP       20.000     16.777   1.192 0.238530
Birrigated:Cmedium:DP     4.000     16.777   0.238 0.812474
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.86 on 53 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared: 0.6765,     Adjusted R-squared: 0.5727
F-statistic: 6.519 on 17 and 53 DF,  p-value: 6.337e-08
```

< Comparison of crossed design (left) with mixed-model designs (below).

Estimates of parameter match in all cases. Standard errors and t-test with match closely between mixed-model designs, but do not match the crossed design (although roughly similar in reported values).

> summary(LMu6)

```
Linear mixed-effects model fit by REML
 Data: NULL
       AIC      BIC    logLik
  474.7823 518.1288 -215.3912

Random effects:
 Formula: ~1 | A
        (Intercept)
StdDev: 0.0006541865
 Formula: ~1 | B %in% A
        (Intercept)
StdDev:    2.009935
 Formula: ~1 | C %in% B %in% A
        (Intercept) Residual
StdDev:      6.9847 9.398264

Fixed effects: Y ~ B * C * D
                           Value Std.Error DF   t-value p-value
(Intercept)             80.50000  5.940396 35 13.551285  0.0000
Birrigated              31.75000  8.400989  3  3.779317  0.0325
Clow                     5.50000  8.279897 12  0.664259  0.5191
Cmedium                 14.75000  8.279897 12  1.781423  0.1001
DNP                      5.50000  6.645576 35  0.827618  0.4135
DP                       4.50000  6.645576 35  0.677142  0.5028
Birrigated:Clow        -39.00000 11.709543 12 -3.330617  0.0060
Birrigated:Cmedium     -22.25000 11.709543 12 -1.900160  0.0817
Birrigated:DNP          13.00000  9.398264 35  1.383234  0.1754
Birrigated:DP            5.50000  9.398264 35  0.585214  0.5622
Clow:DNP                 3.25000  9.398264 35  0.345809  0.7316
Cmedium:DNP             -6.75000  9.398264 35 -0.718218  0.4774
Clow:DP                 -5.25000  9.398264 35 -0.558614  0.5800
Cmedium:DP              -5.50000  9.398264 35 -0.585214  0.5622
Birrigated:Clow:DNP      7.75000 13.291152 35  0.583095  0.5636
Birrigated:Cmedium:DNP   3.89578 13.638668 35  0.285643  0.7768
Birrigated:Clow:DP      20.00000 13.291152 35  1.504760  0.1414
Birrigated:Cmedium:DP    4.00000 13.291152 35  0.300952  0.7652
```

> summary(LMu7)

```
Linear mixed model fit by REML
Formula: Y ~ B * C * D + (1 | A/B/C)
   AIC   BIC logLik deviance REMLdev
 474.8 524.6 -215.4    523.2   430.8
Random effects:
 Groups   Name        Variance Std.Dev.
 C:(B:A)  (Intercept) 48.7861  6.9847
 B:A      (Intercept)  4.0395  2.0098
 A        (Intercept)  0.0000  0.0000
 Residual             88.3275  9.3983
Number of obs: 71, groups: C:(B:A), 24; B:A, 8; A, 4

Fixed effects:
                       Estimate Std. Error t value
(Intercept)              80.500      5.940  13.551
Birrigated               31.750      8.401   3.779
Clow                      5.500      8.280   0.664
Cmedium                  14.750      8.280   1.781
DNP                       5.500      6.646   0.828
DP                        4.500      6.646   0.677
Birrigated:Clow         -39.000     11.710  -3.331
Birrigated:Cmedium      -22.250     11.710  -1.900
Birrigated:DNP           13.000      9.398   1.383
Birrigated:DP             5.500      9.398   0.585
Clow:DNP                  3.250      9.398   0.346
Cmedium:DNP              -6.750      9.398  -0.718
Clow:DP                  -5.250      9.398  -0.559
Cmedium:DP               -5.500      9.398  -0.585
Birrigated:Clow:DNP       7.750     13.291   0.583
Birrigated:Cmedium:DNP    3.896     13.639   0.286
Birrigated:Clow:DP       20.000     13.291   1.505
Birrigated:Cmedium:DP     4.000     13.291   0.301
```