Split-plot Designs

ORIGIN ≡ 1

Split-plot designs involve situations where it is difficult to apply full randomization to all crossed factors because some experimental or observational conditions are harder to apply than others. In agricultural studies, such as the classic "Oats" study described here, experimental fields are often divided into replicate blocks. Within each block, plots are set up related to a hard-to-apply factor (in this case plant variety set in place by an automated planter) and within each plots, subplots are defined by levels of more easily applied treatments (here levels of N fertilizer). Similar data structures occur in observational studies because data sometimes occur naturally in a hierarchical arrangement such as teachers within schools within districts, etc.

Because the factors of the study (Blocks, Plots, Subplots) exist in a hierarchical arrangement, the data collected 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, one conducts a mixed-model (Type III) analysis. Example from W. N. Venables & B. D. Ripley 2002. *Modern Applied Statistics with S*, p. 282.

Example - Balanced Case: Oats

Data Structure:

BLOCK	PLOT	TREATMENT SUBPLOTS			
		0.0cwt	*0.2cwt*	*0.4cwt*	*0.6cwt*
I	*Victory*	111	130	157	174
	Golden.rain	117	114	161	141
	Marvellous	105	140	118	156
II	*Victory*	61	91	97	100
	Golden.rain	70	108	126	149
	Marvellous	96	124	121	144
III	*Victory*	68	64	112	86
	Golden.rain	60	102	89	96
	Marvellous	89	129	132	124
IV	*Victory*	74	89	81	122
	Golden.rain	64	103	132	133
	Marvellous	70	89	104	117
V	*Victory*	62	90	100	116
	Golden.rain	80	82	94	126
	Marvellous	63	70	109	99
VI	*Victory*	53	74	118	113
	Golden.rain	89	82	86	104
	Marvellous	97	99	119	121

Although perhaps appearing as a cross-factor study in R's data long form, the data was collected in the hierarchical format above.

OATS := READPRN("c:/DATA/Models/Oats.txt")

Y := OATS **< Only vector Y from the data set is imported here**

> 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
```

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 for Plots within Blocks ~ $N(0,\sigma_\rho^2)$

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 V α_j $j := 1..a$

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

Grand Mean:

$GM := mean(Y)$ $GM = 103.972222$

Block Means:

From R:

```
#BLOCKS:
B1=Y[B=="I"]
B2=Y[B=="II"]
B3=Y[B=="III"]
B4=Y[B=="IV"]
B5=Y[B=="V"]
B6=Y[B=="VI"]
BLOCKS=cbind(B1,B2,B3,B4,B5,B6)
```

```
> BLOCKS
       B1  B2  B3  B4  B5  B6
 [1,] 111  61  68  74  62  53
 [2,] 130  91  64  89  90  74
 [3,] 157  97 112  81 100 118
 [4,] 174 100  86 122 116 113
 [5,] 117  70  60  64  80  89
 [6,] 114 108 102 103  82  82
 [7,] 161 126  89 132  94  86
 [8,] 141 149  96 133 126 104
 [9,] 105  96  89  70  63  97
[10,] 140 124 129  89  70  99
[11,] 118 121 132 104 109 119
[12,] 156 144 124 117  99 121
```

```
Bb=tapply(oats$Y,oats$B,mean) #blocking factor means
> Bb
       I       II      III       IV        V       VI 
135.3333 107.2500  95.9167  98.1667  90.9167  96.2500 
```

Block Means:

$$B := \begin{pmatrix} 135.3333 \\ 107.2500 \\ 95.9167 \\ 98.1667 \\ 90.9167 \\ 96.2500 \end{pmatrix}$$

Plot Means:

```
#PLOT FACTOR V:
V1=Y[V=="Victory"]
V2=Y[V=="Golden.rain"]
V3=Y[V=="Marvellous"]
PLOTS=cbind(V1,V2,V3)
```

```
Vb=tapply(Y,oats$V,mean) #PLOT factor level means
> Vb
Golden.rain  Marvellous     Victory
    104.500     109.792      97.625
```

Plot Means:

$$V := \begin{pmatrix} 97.625 \\ 104.5 \\ 109.791667 \end{pmatrix}$$

```
> PLOTS
      V1  V2  V3
 [1,] 111 117 105
 [2,] 130 114 140
 [3,] 157 161 118
 [4,] 174 141 156
 [5,]  61  70  96
 [6,]  91 108 124
 [7,]  97 126 121
 [8,] 100 149 144
 [9,]  68  60  89
[10,]  64 102 129
[11,] 112  89 132
[12,]  86  96 124
[13,]  74  64  70
[14,]  89 103  89
[15,]  81 132 104
[16,] 122 133 117
[17,]  62  80  63
[18,]  90  82  70
[19,] 100  94 109
[20,] 116 126  99
[21,]  53  89  97
[22,]  74  82  99
[23,] 118  86 119
[24,] 113 104 121
```

Subplot Means:

```
#SUBPLOT FACTOR N:
N1=Y[N=="0.0cwt"]
N2=Y[N=="0.2cwt"]
N3=Y[N=="0.4cwt"]
N4=Y[N=="0.6cwt"]
SUBPLOTS=cbind(N1,N2,N3,N4)

> tapply(Y,oats$N,mean)
  0.0cwt   0.2cwt   0.4cwt   0.6cwt
 79.3889  98.8889 114.2222 123.3889
```

```
> SUBPLOTS
       N1  N2  N3  N4
 [1,] 111 130 157 174
 [2,] 117 114 161 141
 [3,] 105 140 118 156
 [4,]  61  91  97 100
 [5,]  70 108 126 149
 [6,]  96 124 121 144
 [7,]  68  64 112  86
 [8,]  60 102  89  96
 [9,]  89 129 132 124
[10,]  74  89  81 122
[11,]  64 103 132 133
[12,]  70  89 104 117
[13,]  62  90 100 116
[14,]  80  82  94 126
[15,]  63  70 109  99
[16,]  53  74 118 113
[17,]  89  82  86 104
[18,]  97  99 119 121
```

Subplot Means:

$$N := \begin{pmatrix} 79.388889 \\ 98.888889 \\ 114.222222 \\ 123.388889 \end{pmatrix}$$

Cell Means:

```
#CELLS C:
C11=subset(oats,V=="Victory"&N=="0.0cwt")$Y
C12=subset(oats,V=="Victory"&N=="0.2cwt")$Y
C13=subset(oats,V=="Victory"&N=="0.4cwt")$Y
C14=subset(oats,V=="Victory"&N=="0.6cwt")$Y
C21=subset(oats,V=="Golden.rain"&N=="0.0cwt")$Y
C22=subset(oats,V=="Golden.rain"&N=="0.2cwt")$Y
C23=subset(oats,V=="Golden.rain"&N=="0.4cwt")$Y
C24=subset(oats,V=="Golden.rain"&N=="0.6cwt")$Y
C31=subset(oats,V=="Marvellous"&N=="0.0cwt")$Y
C32=subset(oats,V=="Marvellous"&N=="0.2cwt")$Y
C33=subset(oats,V=="Marvellous"&N=="0.4cwt")$Y
C34=subset(oats,V=="Marvellous"&N=="0.6cwt")$Y

C1=cbind(mean(C11),mean(C12),mean(C13),mean(C14))
C2=cbind(mean(C21),mean(C22),mean(C23),mean(C24))
C3=cbind(mean(C31),mean(C32),mean(C33),mean(C34))
> rbind(C1,C2,C3)
        [,1]     [,2]    [,3]    [,4]
[1,] 71.5000  89.6667 110.833 118.500
[2,] 80.0000  98.5000 114.667 124.833
[3,] 86.6667 108.5000 117.167 126.833
```

```
> cbind(C11,C12,C13,C14)
     C11 C12 C13 C14
[1,] 111 130 157 174
[2,]  61  91  97 100
[3,]  68  64 112  86
[4,]  74  89  81 122
[5,]  62  90 100 116
[6,]  53  74 118 113
> cbind(C21,C22,C23,C24)
     C21 C22 C23 C24
[1,] 117 114 161 141
[2,]  70 108 126 149
[3,]  60 102  89  96
[4,]  64 103 132 133
[5,]  80  82  94 126
[6,]  89  82  86 104
> cbind(C31,C32,C33,C34)
     C31 C32 C33 C34
[1,] 105 140 118 156
[2,]  96 124 121 144
[3,]  89 129 132 124
[4,]  70  89 104 117
[5,]  63  70 109  99
[6,]  97  99 119 121
```

Cell Means:

$$C := \begin{pmatrix} 71.5 & 89.6667 & 110.8333 & 118.5 \\ 80 & 98.5 & 114.6667 & 124.8333 \\ 86.6667 & 108.5 & 117.16667 & 126.8333 \end{pmatrix}$$

ANOVA Sums of Squares: Degrees of Freedom:

BLOCK Factor B:

$$SSB := b \cdot a \cdot \sum_i \left(B_i - GM\right)^2 \qquad SSB = 15875.2312 \qquad s - 1 = 5$$

PLOT Factor V:

$$SSV := b \cdot s \cdot \sum_j \left(V_j - GM\right)^2 \qquad SSV = 1786.3612 \qquad a - 1 = 2$$

SUBPLOT Factor N:

$$SSN := a \cdot s \cdot \sum_k \left(N_k - GM\right)^2 \qquad SSN = 20020.4999 \qquad b - 1 = 3$$

VN Interactions:

$$SSVN := s \cdot \sum_j \sum_k \left(C_{j,k} - N_k - V_j + GM\right)^2 \qquad SSVN = 321.7487 \qquad (a-1)\cdot(b-1) = 6$$

Total:

$$m := 1 .. \text{length}(Y)$$

$$SSTO := \sum_m \left(Y_m - GM\right)^2 \qquad SSTO = 51985.944 \qquad a \cdot b \cdot s - 1 = 71$$

Error:

$$SSE := SSTO - SSB - SSV - SSN - SSVN \qquad SSE = 13982.1035 \qquad a \cdot (s-1) \cdot (b-1) = 45$$

ANOVA Table in R:

```
LM1=aov(Y~V*N+Error(B/V),data=oats)
> summary(LM1)

Error: B
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  5  15875    3175

Error: B:V
          Df Sum Sq Mean Sq F value Pr(>F)
V          2   1786     893    1.49   0.27
Residuals 10   6013     601

Error: Within
          Df Sum Sq Mean Sq F value  Pr(>F)
N          3  20020    6674    37.7 2.5e-12 ***
V:N        6    322      54     0.3    0.93
Residuals 45   7969     177
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note that Error Sum of Squares is partitioned in the R output, but sums to SSE above.

$$6013 + 7969 = 13982$$

F-Test: **Three tests shown together here:**

Hypotheses:

H_0:

all $(\alpha\beta)_{jk} = 0$ for V with N interactions
all $\alpha_j = 0$ for fixed factor V
all $\beta_k = 0$ for fixed factor N within V

H_1:

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

Test Statistics:

Mean Squares from ANOVA Table:

F for VN Interaction: $F_{NV} := \frac{54}{177}$ $F_{NV} = 0.305$

F for Factor V: $F_V := \frac{893}{601}$ $F_V = 1.486$

F for Factor N: $F_N := \frac{6674}{177}$ $F_N = 37.706$

Critical Values for the Tests:

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

for VN Interaction: $C_{NV} := qF[1-\alpha,(a-1)\cdot(b-1),a\cdot(s-1)\cdot(b-1)]$ $C_{NV} = 2.308$

for V: $C_V := qF[1-\alpha,(a-1),a\cdot(s-1)]$ $C_V = 3.682$

for N: $C_N := qF[1-\alpha,(b-1),a\cdot(s-1)\cdot(b-1)]$ $C_N = 2.812$

Decision Rules:

If: **$F_{NV} > C_{NV}$** **Then Reject H_0, otherwise accept H_0**
$F_V > C_V$
$F_N > C_N$

Probabilities:

for VN Interaction: $P_{NV} := 1 - pF[F_{NV},(a-1)\cdot(b-1),a\cdot(s-1)\cdot(b-1)]$ $P_{NV} = 0.931$

for V: $P_V := 1 - pF[F_V,(a-1),a\cdot(s-1)]$ $P_V = 0.258$

for N: $P_N := 1 - pF[F_N,(b-1),a\cdot(s-1)\cdot(b-1)]$ $P_N = 2.436 \times 10^{-12}$

^ slight difference for P_V noted here versus R, probably due to rounding...