

Split-Plot Designs

Bret Larget

Departments of Botany and of Statistics
University of Wisconsin—Madison

March 27, 2007

The Big Picture

- In a *split-plot design*, plots are assigned at random to a first treatment.
- Within each plot individuals are assigned at random to a second treatment.
- The key feature of a *split-plot design* is that there are two different levels for comparing treatments of interest.
- Differences in treatment effects of the first variable must be examined relative to plot level variation.
- Differences in treatment effects of the second variable must be examined relative to plot level variation.

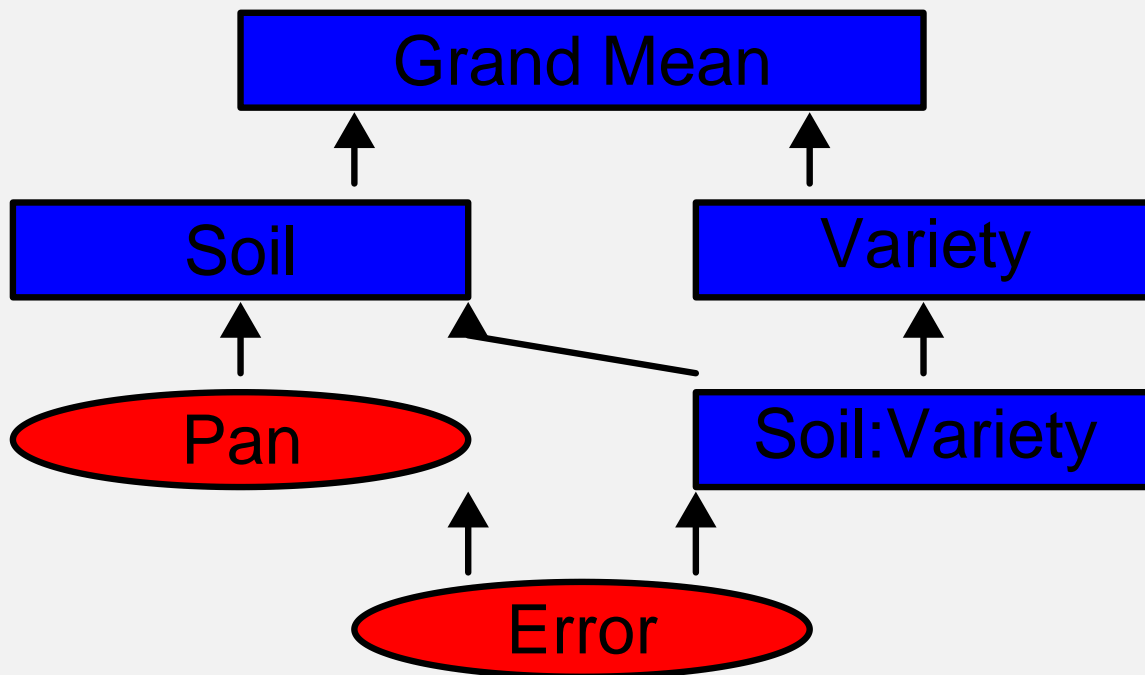
Barley Yield Experiment

- In an experiment, there are two treatments of interest, soil type (3 types) and barley variety (4 types).
- Each soil type is placed into two pans (plots).
- Within each pan, each variety is planted in a single subplot.
- The response variable is the yield of the barley.

Data

	Soil 1		Soil 2		Soil 3	
	Pan 1	Pan 2	Pan 1	Pan 2	Pan 1	Pan 2
Var 1	9.2	11.8	12.9	19.1	16.8	21.7
Var 2	11.3	14.6	15.5	22.6	19.5	23.7
Var 3	6.9	9.3	12.9	16.6	16.8	21.0
Var 4	6.5	9.9	11.2	16.6	15.8	17.8

Split-Plot Design with Interaction



Model

•

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + c_{ik} + e_{ijk}$$

where

- μ is a population mean.
- α_i is the main effect of soil type (A) i , $i = 1, \dots, 3$,
 $\sum_{i=1}^3 \alpha_i = 0$.
- β_j is a main effect of barley variety (B) j , $j = 1, \dots, 4$,
 $\sum_{j=1}^4 \beta_j = 0$.
- $(\alpha\beta)_{ij}$ is the interaction effect of A and B,
 $\sum_i (\alpha\beta)_{ij} = 0$ for each j and $\sum_j (\alpha\beta)_{ij} = 0$ for each i
- $c_{ik} \sim \text{iid } N(0, \sigma_a^2)$ is the plot (pan) error distribution, $k = 1, 2$.
- $e_{ijk} \sim \text{iid } N(0, \sigma_e^2)$ is the subplot (individual) error distribution,
 $k = 1, 2$.

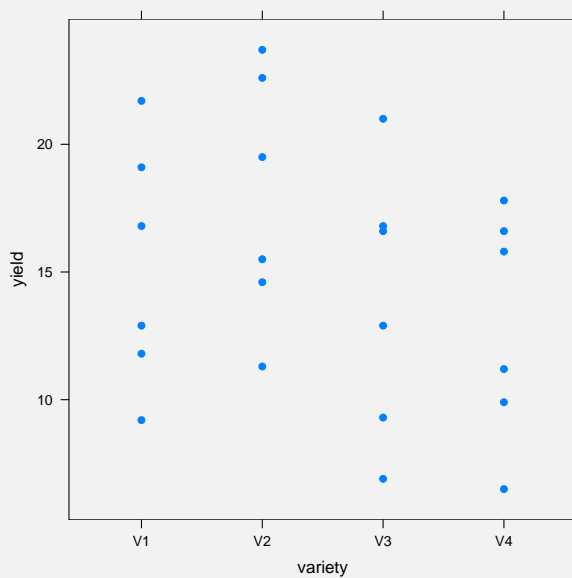
Data

```
> barley = read.table("barley.txt", header = T)
> attach(barley)
> str(barley)

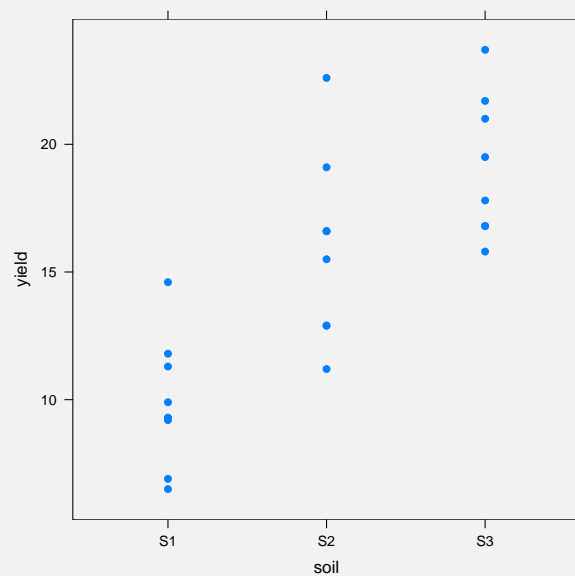
'data.frame': 24 obs. of 4 variables:
 $ variety: Factor w/ 4 levels "V1","V2","V3",...: 1 1 2 2 3 3 4 4 1 1 ...
 $ soil   : Factor w/ 3 levels "S1","S2","S3": 1 1 1 1 1 1 1 1 2 2 ...
 $ pan    : Factor w/ 2 levels "P1","P2": 1 2 1 2 1 2 1 2 1 2 ...
 $ yield  : num  9.2 11.8 11.3 14.6 6.9 9.3 6.5 9.9 12.9 19.1 ...
```

Plot of Data

```
> library(lattice)
> fig1 = xyplot(yield ~ variety, pch = 16)
> print(fig1)
```



```
> library(lattice)
> fig2 = xyplot(yield ~ soil, pch = 16)
> print(fig2)
```



Numerical Summaries

```
> sapply(split(yield, soil), mean)

      S1      S2      S3
9.9375 15.9250 19.1375

> sapply(split(yield, soil), sd)

      S1      S2      S3
2.647337 3.710121 2.798948

> sapply(split(yield, variety), mean)

      V1      V2      V3      V4
15.25000 17.86667 13.91667 12.96667

> sapply(split(yield, variety), sd)

      V1      V2      V3      V4
4.750895 4.868128 5.239625 4.448221

> sapply(split(yield, soil:variety), mean)

S1:V1 S1:V2 S1:V3 S1:V4 S2:V1 S2:V2 S2:V3 S2:V4 S3:V1 S3:V2 S3:V3 S3:V4
10.50 12.95  8.10  8.20 16.00 19.05 14.75 13.90 19.25 21.60 18.90 16.80

> sapply(split(yield, soil:variety), sd)

S1:V1 S1:V2 S1:V3 S1:V4 S2:V1 S2:V2 S2:V3 S2:V4
1.838478 2.333452 1.697056 2.404163 4.384062 5.020458 2.616295 3.818377
S3:V1 S3:V2 S3:V3 S3:V4
3.464823 2.969848 2.969848 1.414214
```

Mixed Effects Analysis

```
Linear mixed-effects model fit by REML
Formula: bf
      AIC      BIC logLik MLdeviance REMLdeviance
 92.44 107.8 -33.22      65.4      66.44
Random effects:
Groups   Name          Variance Std.Dev.
soil:pan (Intercept) 8.88775  2.9812
Residual                0.65512  0.8094
number of obs: 24, groups: soil:pan, 6

Fixed effects:
              Estimate Std. Error t value
(Intercept) 15.00000    1.22825  12.213
soil1        -5.06250    1.73700  -2.915
soil2         0.92500    1.73700   0.533
variety1      0.25000    0.28616   0.874
variety2      2.86667    0.28616  10.018
variety3     -1.08333    0.28616  -3.786
soil1:variety1 0.31250    0.40470   0.772
soil2:variety1 -0.17500    0.40470  -0.432
soil1:variety2 0.14583    0.40470   0.360
soil2:variety2 0.25833    0.40470   0.638
soil1:variety3 -0.75417    0.40470  -1.864
soil2:variety3 -0.09167    0.40470  -0.227

Correlation of Fixed Effects:
              (Intr) soil1  soil2  varty1  varty2  varty3  sl1:v1  sl2:v1  sl
soil1          0.000
soil2          0.000 -0.500
variety1        0.000  0.000  0.000
variety2        0.000  0.000  0.000 -0.333
variety3        0.000  0.000  0.000  0.000  0.000
soil1:variety1  0.000  0.000  0.000  0.000  0.000  0.000
soil2:variety1  0.000  0.000  0.000  0.000  0.000  0.000  0.000
soil1:variety2  0.000  0.000  0.000  0.000  0.000  0.000  0.000  0.000
soil2:variety2  0.000  0.000  0.000  0.000  0.000  0.000  0.000  0.000
soil1:variety3  0.000  0.000  0.000  0.000  0.000  0.000  0.000  0.000
soil2:variety3  0.000  0.000  0.000  0.000  0.000  0.000  0.000  0.000
soil2:vrty1    0.000  0.000  0.000  0.000  0.000  0.000  0.000 -0.500
```

```
> library(lme4)
```

```
> options(contrasts = c("contr.sum", "contr.poly"))
```

```
> bf = formula(yield ~ soil * variety + (1 | soil:pan))
```

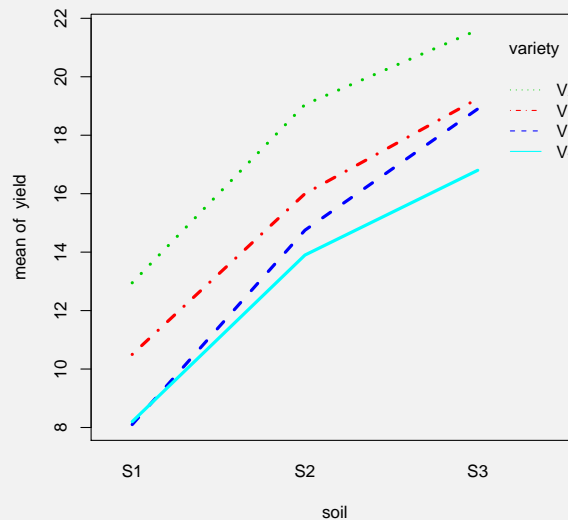
```
> barley.lmer = lmer(bf)
```

```
> summary(barley.lmer)
```

Interaction Plots

```
> interaction.plot(soil, variety, yield, col = 2:5)
```

- Points are of group means.
- Lines indicate change, but are meaningless between x values.
- Lines that are nearly parallel indicate little potential *interaction*.
- Lines that are far from parallel indicate possible interaction.



Likelihood Ratio Tests

- Two *nested models* can be fit with a *likelihood ratio test*.
- When L_1 is the likelihood of the smaller model and L_2 is the likelihood of the larger model,
- the statistic $-2 \times (\log(L_1) - \log(L_2))$ has approximately a chi-square distribution with degrees of freedom equal to the difference in the number of parameters in the model.
- The `anova` function is useful to carry out likelihood ratio tests for models fit with `lmer`.

Nested Models

```

> barley1 = lmer(yield ~ 1 + (1 | soil:pan))
> barleyA = lmer(yield ~ soil + (1 | soil:pan))
> barleyB = lmer(yield ~ variety + (1 | soil:pan))
> barleyAB = lmer(yield ~ soil + variety + (1 | soil:pan))
> barleyABi = lmer(yield ~ soil * variety + (1 | soil:pan))

```

Testing for a Soil Effect 1

```

> anova(barley1, barleyA)

Data:
Models:
barley1: yield ~ 1 + (1 | soil:pan)
barleyA: yield ~ soil + (1 | soil:pan)

```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
barley1	2	127.392	129.748	-61.696				
barleyA	4	123.665	128.377	-57.832	7.7267		2	0.02100 *

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Testing for a Soil Effect 2

```
> anova(barleyB, barleyAB, barleyABi)

Data:
Models:
barleyB: yield ~ variety + (1 | soil:pan)
barleyAB: yield ~ soil + variety + (1 | soil:pan)
barleyABi: yield ~ soil * variety + (1 | soil:pan)
          Df      AIC      BIC  logLik  Chisq Chi Df Pr(>Chisq)
barleyB    5  92.860  98.750 -41.430
barleyAB   7  88.786  97.032 -37.393  8.0739    2  0.01765 *
barleyABi 13  91.396 106.711 -32.698  9.3899    6  0.15281
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparison to F test

```
> barley.lm = lm(yield ~ variety + soil/pan)
> anova(barley.lm)
```

Analysis of Variance Table

```
Response: yield
          Df Sum Sq Mean Sq F value    Pr(>F)
variety    3  81.53   27.18  42.331 1.473e-07 ***
soil       2 348.83  174.41 271.673 1.655e-12 ***
soil:pan   3 109.09   36.36  56.642 2.055e-08 ***
Residuals 15   9.63    0.64
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> 1 - pf(174.41/36.36, 2, 3)
```

```
[1] 0.1162684
```


Observations

- Using the likelihood ratio test, there is strong evidence that soil has an effect on yield both in the absence ($p = 0.02$) and the presence ($p = 0.18$) of variety.
- There is little evidence that the interaction between soil and yield is important ($p = 0.15$).
- Using an F test, there is much less evidence for the effect of soil in the presence of variety ($p = 0.12$).
- Notice that the larger apparent effect due to soil based on the magnitude of the difference in sample means is offset by the much larger standard errors in estimating the size of the effect.