Split plot design and its relatives

- A comment on error
- Split plot design as an RCBD
- Another way to think about the main plot error
- Split plot design as a CRD
- Split plot design as a Latin Square
- Split plot approach to repeated measurements (time as a subplot effect)
- Repeated measurements using the REPEATED statement
- *APPENDIX: The standard split plot analysis for Example 3*

A Comment on Error

How you randomize your experimental units determines your experimental design, and your experimental design determines the error terms for all tests – there is no way around this, end of story. So, once you decide on an experimental design, the error terms associated with that design must be used consistently in F-tests, contrasts, means separations, everything.

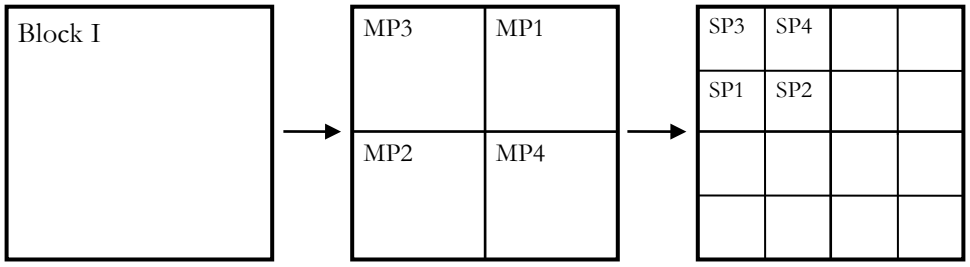
Split Plot Design as an RCBD

A split plot design results from a two-stage randomization process of a factorial treatment structure. Because of this two-stage process, one loses sensitivity in detecting differences among main plot treatments (the first level of randomization) but gains sensitivity in detecting differences among subplot treatments (the second level), as well as the significance of the MainPlot:Subplot interaction.

Example 1 *Split plot design ST&D pg. 406 [Lab9ex1.R]*

This experiment is an example of a split plot design organized as an RCBD. A significant interaction is found between main plots and subplots; thus an analysis of simple effects is required. In this study, four different varieties of oats (main plots) are randomized within four blocks, and four different seed treatments (subplots) are randomized within each of the sixteen main plots:

Blocks: Four
Main plots (varieties): Four
Subplots (seed treatments): Four (includes one control, SP1)



Note that in this case the greater precision is afforded to the seed treatments (subplots) and the lesser to the varieties (main plots). If the primary intention is to investigate differences among seed treatments, one possible reason to choose this design would be to extend the scope of the experiment across varieties.

```
#The ANOVA
oats_mod<-aov(Yield ~ VarietyA + Block + Error(Variety:Block) + TrtmtB +
      VarietyA:TrtmtB, oats_dat)
summary(oats_mod)
```

As we have seen before in some previous examples (nested experiments, mixed models), in a split plot design we need to specifically declare an error term in order to carry out the appropriate F-tests.

Output

```
Error: SeedLotA:Block
           Df Sum Sq Mean Sq F value  Pr(>F)
VarietyA    3 2848.0   949.3   13.82 0.00102 **
Block       3 2842.9   947.6   13.79 0.00103 **
Residuals   9  618.3    68.7
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Error: Within
                 Df Sum Sq Mean Sq F value  Pr(>F)
TrtmtB            3  170.5   56.85   2.799 0.05386 .
VarietyA:TrtmtB   9  586.5   65.16   3.208 0.00595 **
Residuals        36  731.2   20.31
```

The F values indicate significant differences among varieties ($p = 0.001$), no significant differences (barely) among seed treatments ($p = 0.0539$), and a significant interaction ($p = 0.0059$). If there were no significant interaction, we could analyze the main effects of the main plot and subplot (see lecture notes). BUT, since the interaction *is* significant, we cannot make comparisons among the main effects. Rather, we can use Dunnett's Test to compare the *simple effects* of each seed treatment against the control **within each variety**. For variety 1:

```
#Means comparisons
#library(multcomp)

#Comparisons among subplot levels within a common main plot level
lot1_dat<-subset(oats_dat, VarietyA == 1)
lot1_mod<-lm(Yield ~ Block + TrtmtB, lot1_dat)
anova(lot1_mod)
lot1_dunnett=glht(lot1_mod,linfct=mcp(TrtmtB="Dunnett"))
confint(lot1_dunnett)
```

You could then copy and paste this code and change a few numbers to carry out similar analyses for the other seed lots. Or you could try looping things:

```
#Using loops in R to cycle through levels of the main plot A
A_levels<-c(1:4)
for (i in A_levels) {
  with(subset(oats_dat, VarietyA == A_levels[i]), {
       print(A_levels[i])
       print(anova(lm(Yield ~ Block + TrtmtB)))
       print(confint(glht(lm(Yield ~ Block + TrtmtB),
                                    linfct=mcp(TrtmtB="Dunnett"))))
  })
}
```

Either way, you'll get a lot of output that you'll need to summarize.

Output Summary

VarietyA	TrtmtB p-value	S / NS
1	9.6e-06	***
2	0.7312	NS
3	0.6093	NS
4	0.2109	NS

Rankings and pairwise comparisons of treatments (2, 3, 4) vs. control (1) within each variety

	Variety A			
	1	**2**	**3**	**4**
TrtmtB Dunnett Results	2 *	2	4	2
	3 *	4	3	4
	4	3	2	3

Notice that the seed treatment (Factor B) is found to have a significant effect in Variety 1, in contradiction to the original NS ANOVA result for the subplot (p = 0.0539). This is why it is important to look at simple effects when there is a significant interaction.

Another way to think about the Main Plot error

Chew on this: The correct error term produces the same F- and p-values for the main plot effect (A) that you get *if you simply average the subplots (B)*. Let's look again at the RCBD example from above; but this time we'll average over the subplots [using ddply()], essentially removing the subplot treatments from the experiment:

```
#Revisiting the same analysis for A by first averaging the subplots...
library(plyr)
oats_Bmeans_dat <- ddply(oats_dat, c("VarietyA", "Block"), summarise,
                         Yield = mean(Yield, na.rm=TRUE))
#...then carrying out the following ANOVA
oats_Bmeans_mod<-aov(Yield ~ VarietyA + Block, oats_Bmeans_dat)
summary(oats_Bmeans_mod)
```

Lo and behold, the output we get matches the previous results exactly:

```
            Df Sum Sq Mean Sq F value  Pr(>F)
VarietyA     3  712.0  237.34   13.82 0.00102 **
Block        3  710.7  236.91   13.79 0.00103 **
Residuals    9  154.6   17.17
```

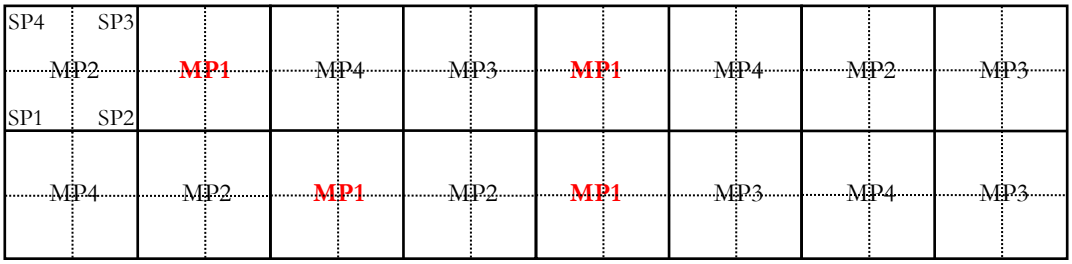
In addition to confirming that Block:VarietyA is the appropriate error term to use for VarietyA, this result illustrates that in a split plot design the main plot effect is totally insensitive to the variation among subplots (i.e. when comparing main plot effects, subplots act as subsamples).

Split Plot Design as a CRD

Recall that in a CRD, "Replication" does not appear in the lm() statement because variation among replications within a given treatment level is the source of error for the experiment. In a split plot design organized as a CRD, however, the Replication:A interaction is needed as the appropriate error term for the main plot; to use it in this way, we must include it in the linear model. ***BE AWARE***: When you include "Replication:A" in the linear model without including "Replication" by itself, R automatically produces a SS labeled "Replication:A" that includes SS for "Replication" *AND* "Replication:A" combined. In general:

```
lm ( y ~ A + A:B)      * Produces SS (A:B) = SS (A:B) + SS (B);
lm ( y ~ B + A:B)      * Produces SS (A:B) = SS (A:B) + SS (A);
lm ( y ~ A:B)          * Produces SS (A:B) = SS (A:B) + SS (A) + SS (B);
```

Here is how the experiment above might look had the main plots been randomized as a CRD.



To accommodate this new design, the R code would be changed by first substituting Reps for Blocks in the input and then modifying the lm() statement appropriately:

Example 2 *[Lab9ex2.R]*

```
#The ANOVA
oats_mod<-aov(Yield ~ VarietyA + Error(VarietyA:Rep) + TrtmtB +
      VarietyA:TrtmtB, oats_dat)
summary(oats_mod)
```

It is worth mentioning that the inclusion of the VarietyA:Rep term in the model is the only thing that distinguishes this experiment from a CRD factorial.

Output

```
Error: SeedLotA:Rep
           Df Sum Sq Mean Sq F value Pr(>F)
VarietyA    3   2848   949.3   3.291 0.0581 .
Residuals  12   3461   288.4
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Error: Within
                  Df Sum Sq Mean Sq F value  Pr(>F)
TrtmtB             3  170.5   56.85   2.799 0.05386 .
VarietyA:TrtmtB    9  586.5   65.16   3.208 0.00595 **
Residuals         36  731.2   20.31
```

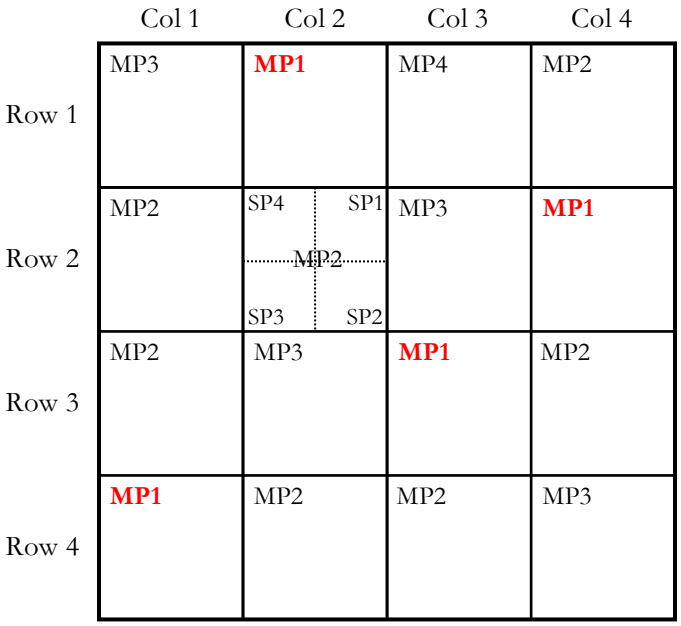
Analyzing this set of data as though it were a CRD leads to a NS main effect of variety (p = 0.0581), in contrast to its significant effect under the RCBD. Does this make sense? Now, because there is a significant interaction, our analysis must continue with the simple effects. The code is similar to the RCBD case:

```
#Means comparisons
#library(multcomp)

#Comparisons among subplot levels within a common main plot level
lot1_dat<-subset(oats_dat, VarietyA == 1)
lot1_mod<-lm(Yield ~ TrtmtB, lot1_dat)
anova(lot1_mod)
lot1_dunnett=glht(lot1_mod,linfct=mcp(TrtmtB="Dunnett"))
confint(lot1_dunnett)

#Etc....or...
#Using loops in R to cycle through levels of the main plot A
A_levels<-c(1:4)
for (i in A_levels) {
  with(subset(oats_dat, VarietyA == A_levels[i]), {
       print(A_levels[i])
       print(anova(lm(Yield ~ TrtmtB)))
       print(confint(glht(lm(Yield ~ TrtmtB),linfct=mcp(TrtmtB="Dunnett"))))
  })
}
```

And the results can be summarized in a manner similar to the one shown above for the RCBD case.

Split Plot Design as a Latin Square

How would this split plot look as a Latin Square?



In a Latin Square, “Row” and "Col" are both blocking variables and therefore both need to appear in the MODEL statement. The error term for the main plot (VarietyA) is now "Row:Col:VarietyA," and R will automatically include in this term all possible two-way interactions among these effects (e.g. Row:Col, Row:VarietyA, and Col:VarietyA):

```
#The ANOVA
oatsLS_mod<-aov(Yield ~  Row + Col + SeedLotA +
                         Error(Row:Col:SeedLotA) +
                         TrtmtB +
                         SeedLotA:TrtmtB,
                         oats_dat)
```

Repeated Measures (time as a subplot effect)

By repeatedly measuring the same experimental unit at different points in time, one can gain insight into the effect of time on the observations. In this way, "time" is similar to a subplot effect in a split plot design; there are two major differences, however:

1. Whereas true subplots can be assigned randomly, "time" cannot.
2. Because the observations are made on the same experimental units, the observations are not independent from one another and the degrees of freedom must be adjusted appropriately.

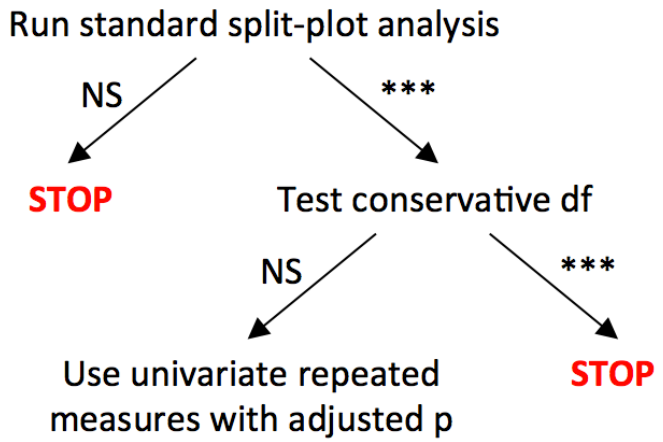
Before going through a specific example, let's first take a moment and outline the analysis protocol you should follow if you have an experiment in which you are carrying out repeated measures:

1. First, treat the repeated measures as a subplot effect of time in a split plot design, run the analysis as you would any split plot, and look at the resultant ANOVA.

2. Then, *by hand*, reduce the subplot degrees of freedom to 1 (this will also affect the mainplot*subplot interaction df and the error df) and recalculate p-values for the subplot and the mainplot*subplot interaction(s). Note that this procedure will not affect your F-values; it only affects the df of the *critical* F-value and thus the corresponding p-values.

3. Next, compare the results of the normal split plot ANOVA [from Step 1] and the conservative df ANOVA [from Step 2]:

- If both are significant, the effects are significant. [STOP]
- If both are NS, the effects are NS. [STOP]
- If the full df ANOVA is significant but the conservative df ANOVA is not, perform a repeated measures analysis using R. [Go to Step 4]



4. When you perform a repeated measures analysis, you will obtain two tables in the output:

- **Unadjusted (split-plot) ANOVA table**: You will refer to this table for tests on the mainplot effects. The F- and p-values in this table will match those in the full df ANOVA, as long as you specified the correct mainplot error. That is, these values are the p-values generated using

the full df; they match the p-values in your original split plot ANOVA and should not be used for the subplot effects because they result from assuming that all the measurements in your experiment are perfectly independent from one another (not the case when taking repeated measures on the same experimental units). The results for the main plot are valid.

- **Sphericity-adjusted ANOVA table**: You refer to this table for tests on the subplot effects. There are two columns of p-values in this table:

  1) "p[GG]" These p-values are generated using adjusted df based on the actual correlations found in the data. In reality, your measurements are not perfectly independent (full df) or perfectly correlated (conservative df) – they are somewhere in between. The G-G procedure tries to find this middle ground. USE THESE P-VALUES.

  2) "p[HF]" Another method of adjusting df, but not as conservative as G-G.

A Final Comment: All this being said, if you have a strongly significant Orthogonal Components Sphericity Test ($p < 0.001$), you should use extreme caution when interpreting any of these results, even the G-G values. Failing this test is akin to failing Levene's for an ANOVA. Either some remediation needs to be done to the data so that you no longer fail it (i.e. transformation), or you need to find some other way to analyze your data (e.g. MANOVA).

Repeated Measures Using the ezANOVA() function

So, your full-df ANOVA was significant but your conservative-df ANOVA was not. R to the rescue: The ezANOVA() function in the "ez" package provides a slick way of adjusting for correlations among repeated observations of the same experimental units.

Example 3 *[Lab9ex3.R]*

In this example, the main plot itself has an underlying 2x2 factorial treatment structure; the subplot is Time. Specifically, sixteen dogs (experimental units) were randomly assigned to four groups (main plots). Within each group, dogs received either morphine or trimethaphan (levels M or T, variable DRUG) and had either depleted or intact histamine levels (levels D or I, variable HIST). The response variable was the blood concentration of histamine at 0, 1, 3, and 5 minutes after injection of the drug.

```
#Inform R about which variables are factors
hist_dat$Time<-as.factor(hist_dat$Time)
hist_dat$Dog<-as.factor(hist_dat$Dog)

#library(ez)
ezANOVA(
  data = hist_dat,
  dv = H,
  wid = Dog,
  within = Time,
  between = Drug:Hist,
  return_aov = TRUE
)
```

Output and Commentary

	Effect	DFn	DFd	F	p	p<.05	ges
2	Drug	1	12	2205.7921794	5.673929e-15	*	0.97182999
3	Hist	1	12	1239.7188966	1.758533e-13	*	0.95095465
5	Time	3	36	38.4860031	2.522014e-11	*	0.72262619
4	Drug:Hist	1	12	74.4477314	1.719674e-06	*	0.53797133
6	Drug:Time	3	36	3.5882896	2.286217e-02	*	0.19543200
7	Hist:Time	3	36	0.3006148	8.247187e-01		0.01994375
8	Drug:Hist:Time	3	36	4.8650415	6.087231e-03	*	0.24774164

This first ANOVA table should be referenced for the main plot effects, ignoring the within-dog effects. The F-values here are correct since under the ezANOVA() procedure the correct main plot error term is used. The results here match the results you would obtain by running the data as a simple split plot with Time as the subplot effect [see Appendix]. The significant Drug*Hist interaction requires you to analyze the simple effects of these two factors.

```
$`Mauchly's Test for Sphericity`
          Effect         W         p p<.05
5           Time 0.6442253 0.4533546
6      Drug:Time 0.6442253 0.4533546
7      Hist:Time 0.6442253 0.4533546
8 Drug:Hist:Time 0.6442253 0.4533546
```

This next table presents the results of Mauchly's sphericity test on the orthogonal components of the covariance matrix. Somewhat simplified, the Sphericity Test tests the assumption that the variances and correlations are homogeneous across the various dependent variables. You can think of it as an assumption test for repeated measures analysis the same way Levene's is an assumption test for ANOVA generally. Since the Orthogonal Components Sphericity Test is NS ($p > 0.05$), one can use the GG-adjusted p-values in the next table without reservation. [Note: The Greenhouse-Geisser Epsilon is used to multiply the numerator and denominator degrees of freedom before determining significance levels for the F-tests.]

```
$`Sphericity Corrections`
          Effect       GGe        p[GG] p[GG]<.05      HFe        p[HF] p[HF]<.05
5           Time 0.8047628 1.661685e-09         * 1.021859 2.522014e-11         *
6      Drug:Time 0.8047628 3.317977e-02         * 1.021859 2.286217e-02         *
7      Hist:Time 0.8047628 7.820166e-01           1.021859 8.247187e-01
8 Drug:Hist:Time 0.8047628 1.101469e-02         * 1.021859 6.087231e-03         *
```

Finally, this table reports the subplot effects (i.e. within-dog effects of Time). Looking at the p[GG] column, the results indicate that Time, Time:Drug, and Time:Drug:Hist are all significant. The significant Time:Drug interaction, for example, indicates that the effect of time on the blood concentration of histamine is different for the two drugs being studied.

Appendix: The standard split plot analysis for Example 3

If you were given the histamine data to analyze, the first thing you would do is to treat it like a standard split plot. In this case, the main plot effect has a factorial treatment structure and the R code would look something like this:

```
#The split-plot ANOVA
hist_mod<-aov(H ~ Drug + Hist + Drug:Hist + Error(Drug:Hist:Dog) +
      Drug*Hist*Time, hist_dat)
summary(hist_mod)
```

Output

```
Error: Drug:Hist:Dog
           Df Sum Sq Mean Sq F value   Pr(>F)
Drug        1 24.738  24.738 2205.79 5.67e-15 ***
Hist        1 13.904  13.904 1239.72 1.76e-13 ***
Drug:Hist   1  0.835   0.835   74.45 1.72e-06 ***
Residuals  12  0.135   0.011
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: Within
                 Df Sum Sq Mean Sq F value   Pr(>F)
Time              3 1.8682  0.6227  38.486 2.52e-11 ***
Drug:Time         3 0.1742  0.0581   3.588  0.02286 *
Hist:Time         3 0.0146  0.0049   0.301  0.82472
Drug:Hist:Time    3 0.2362  0.0787   4.865  0.00609 **
Residuals        36 0.5825  0.0162
```

Notice first of all that the F- and p-values exactly match those found in the first ANVA table produced by the ezANOVA() function. This is a nice verification that our split plot programming and error term assignments are correct.

At this point, recompute the p-values *by hand* for the subplot effect (and subplot interactions) using conservative degrees of freedom. The adjusted table:

Source	DF	Adj DF
Error	36	12

Source	DF	Adj DF	F Value	(df num, df den)	Pr > F	
Time	3	1	38.49	(1, 12)	<.0001	***
Drug*Time	3	1	3.59	(1, 12)	0.0825	NS
Hist*Time	3	1	0.30	(1, 12)	0.5939	NS
Drug*Hist*Time	3	1	4.87	(1, 12)	0.0476	*

Now compare the results of the full df analysis with those of the conservative df analysis. Due to agreement between the two approaches, you can declare the Time:Hist:Drug interaction to be significant;

you can also declare the Hist:Time interaction to be NS. The Drug:Time interaction, however, was found to be significant ($p = 0.0229$) with full df but NS ($p = 0.0825$) with conservative df. Because of this discrepancy, we needed to carry out the repeated measures analysis documented in the lab.