STAT 8200 — Design of Experiments for Research Workers
Lab 9 – Due: Thursday, Oct. 31, 2013

Example:

Suppose that in the redwing flaxseed example, the experiment was conducted in the same way as described in lecture, with the exception that in each of the six plots per block, two samples of flaxseed were taken and the percent oil content was measured on each sample. That is, the design and data are as follows:

T6 36.4,37.4	T4 36.8,36.1
T5 36.3,37.4	T1 34.4,36.2
T3 34.4,34.5	T2 33.3,33.0

Block 1

T6 37.3,37.3	T4 36.6,35.6
T3 34.0,34.5	T5 34.9,33.8
T1 35.9,35.9	T2 31.9,30.8

Block 2

T1 36.0,37.8	T4 37.0,38.5
T5 35.9,36.0	T2 34.9,34.3
T6 37.7,37.2	T3 34.5,34.1

Block 3

T4 36.4,36.1	T5 37.1,35.6
T6 36.7,36.4	T2 37.1,36.7
T3 33.1,32.2	T1 34.1,36.0

Block 4

In this example, we have 2 sub-samples (aka pseudo-replicates) in each of the 24 plots. There are two ways to analyze such data. The methods are equivalent for inference on the block and treatment effects and means. However, in the first method we do not bother to quantify variability due to sub-sampling (which is different from error variability) and in the second method we do quantify this sub-sampling variability.

Copy the file redwing3.sas from the public directory to your USB drive, run it, and examine the output.

Method 1: Since the experimental unit is the plot, we have 2 measurements per experimental unit. We can collapse the two measurements on each plot to one by taking their mean and then analyze one measurement per experimental unit (the mean of the two pseudo-replicate values). The analysis then goes through exactly as in a usual RCBD (for example, like in redwing1.sas, if we treat blocks as fixed, or as in redwing2.sas, if we treat blocks as random).

See redwing3.sas and its output. The means are computed by block and treat in PROC MEANS and placed in the output data set eudata (experimental unit data). In this

data set eumean is the mean of the two observations per plot. The data set eudata is printed on p.1 of the output. Then eumean is used as the response variable in the usual RCBD analysis. Here, we will assume blocks are random, which seems more appropriate here, and do the analysis using PROC MIXED. The results from this analysis appear on pp.2–3. Notice that the dependent variable is EUMEAN. The F statistics for the main effect of treat and for the contrast comparing control with inoculation are 4.86 and 7.63, respectively. Both are statistically significant. Note that the main effect of blocks should not be tested.

Method 2: Alternatively, we can analyze the original data (which are taken at the pseudo-replicate level, not the experimental unit level). This analysis is done in the second call to PROC MIXED, which also treats blocks as random. Notice that the data set being analyzed is now redwing2 (the original data) and the dependent variable is OIL. In this design we do not have true replication and the sum of squares labeled block*treat is really the error sum of squares. The "residual" sum of squares actually quantifies subsampling variability, not error variability among the true experimental units. Therefore, the appropriate F test for treat is

$$F = \frac{MS_{Treat}}{MS_{block*treat}} \sim F(\text{d.f.}_{Treat}, \text{d.f.}_{block*treat})$$

PROC MIXED detects this automatically and computes the F test for treat appropriately as

$$F = \frac{MS_{Treat}}{MS_{block*treat}} = \frac{13.89}{2.86} = 4.86$$

rather than the usual MS_{Treat}/MS_E.

Note that if PROC GLM had been used to do the analysis (e.g., if we had regarded blocks as fixed), it would not have known automatically to use the appropriate denominator for the F test on treat. Instead, by default it always uses MS_E as the denominator for all F tests. This can be over-ridden by telling SAS to create the right test statistics with the TEST statement*. The syntax is

TEST H=*source to be tested* E=*mean squares for denom of test statistic*;

For our contrast, we also must use the E= option on the CONTRAST statement to obtain the correct denominator for our contrast F test. Notice that the F statistics agree with what we obtained in method 1. Finally, with this method we can estimate the pseudo-replicate variance. Our estimate is what SAS lists as Mean Square for Error, or 0.479.

* See the comment in redwing3.sas in the middle of the call to PROC GLM for an alternative syntax to accomplish the same thing.

STAT 8200 — Lab 9

Name:_____

Exercise:

An experiment was conducted in which three treatments and a control treatment were applied to soybean seeds. The four treatments were randomly allocated to four plots within each of five blocks corresponding to five regions within a large soybean field. The response variable of interest was the percentage of soybean plants which failed to emerge after planting. In each plot the response variable was measured on each of three samples of 100 plantings. The data are as follows:

Treatment	Block				
	1	2	3	4	5
Control	8	10	12	13	11
	9	7	14	11	12
	8	9	11	12	8
Arasan	2	6	7	11	5
	1	3	8	15	3
	4	6	5	14	6
Spergon	4	10	9	8	10
	5	7	5	8	14
	4	10	8	9	8
Fermate	9	7	5	5	3
	8	3	7	10	5
	10	12	5	5	7

Produce the correct ANOVA Table for analyzing these data based on

a. All 60 data.

b. One observation per experimental unit.

Include Source, SS, d.f., MS, and F in your ANOVA Tables. A data set including all 60 measurements is contained in the file "soybean.dat".