

## SAS for Randomized Complete Block Designs

We use fictitious data from a fictitious agronomy study. This is the same data we used in the lecture notes. There are 12 plots (experimental units), the blocks are the physical location of the plots (West, Middle, East), and the treatments are various fertilizers (numbered 1 through 4). The response variable is the yield of some crop.

### SAS Code

```
options ls=76 nodate nopage;

data agronomy;
input block $ treatment yield;
datalines;
West 1 40.5
West 2 39.4
West 3 38.3
West 4 38.1
Mid 1 45.4
Mid 2 44.1
Mid 3 43.0
Mid 4 42.0
East 1 48.3
East 2 47.0
East 3 46.2
East 4 46.1
;

title 'Ignoring Blocks';
proc glm data=agronomy;
class treatment;
model yield = treatment / ss3;
lsmeans treatment / stderr pdiff;
run;

title 'Randomized Complete Block Design';
proc glm data=agronomy;
class treatment block;
model yield = treatment block / ss3;
lsmeans treatment block / stderr pdiff;
run;
```

This is the output from the first GLM statement, in which we ignored blocks. The most important thing to notice is that the ANOVA F test is not significant ( $p=0.8442$ ), so we would declare that there is not a significant difference among the treatments. All of the LSMEANS are fairly close to each other (42.07 to 44.73). Relative to these differences, the MSE (15.5) is large. The small differences between the treatment means could be the result of the large error variation, so we would declare that there is no difference between the treatments.

Ignoring Blocks

Dependent Variable: yield

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	12.6466667	4.2155556	0.27	0.8442
Error	8	124.0933333	15.5116667		
Corrected Total	11	136.7400000			

  

R-Square	Coeff Var	Root MSE	yield Mean
0.092487	9.116864	3.938485	43.20000

  

Source	DF	Type III SS	Mean Square	F Value	Pr > F
treatment	3	12.6466667	4.2155556	0.27	0.8442

  

Least Squares Means

treatment	yield LSMEAN	Standard Error	Pr >  t	LSMEAN Number
1	44.7333333	2.2738856	<.0001	1
2	43.5000000	2.2738856	<.0001	2
3	42.5000000	2.2738856	<.0001	3
4	42.0666667	2.2738856	<.0001	4

  

Least Squares Means for effect treatment  
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: yield

i/j	1	2	3	4
1		0.7113	0.5070	0.4310
2	0.7113		0.7638	0.6676
3	0.5070	0.7638		0.8961
4	0.4310	0.6676	0.8961	

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.

The following pages contain the output from the second GLM statement, in which we included blocks.

The first thing we should look at is the ANOVA F test ( $F = 298.05$ ,  $p < .0001$ ). This test is testing whether there are any treatment effects or any block effects. If this test is not significant, then there are no treatment effects but there are also no block effects. This is contrary to what we believed when we created the blocks. In other words, if we do not reject the ANOVA F test, then our blocking did not help eliminate excess variation from the error variance. In a successful RCB design, we should always reject the ANOVA F test.

The GLM Procedure					
Dependent Variable: yield					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	136.1916667	27.2383333	298.05	<.0001
Error	6	0.5483333	0.0913889		
Corrected Total	11	136.7400000			
R-Square	Coeff Var	Root MSE	yield Mean		
0.995990	0.699782	0.302306	43.20000		

For our example, we do reject the ANOVA F test, so the next thing we should look at are the tests for the individual components of the model. Make sure the test for blocks is significant (because we want to have significant differences between the blocks). In our example, this test has  $F = 675.93$  and  $p < .0001$ , so the blocks were successful at removing excess variability among the experimental units.

Look at the F test for treatments. In more complicated treatment structures, there may be many tests (like there was in two-way ANOVA analysis). In our example, there are simply four treatments. The differences between them are significant ( $F = 46.13$ ,  $p = 0.0002$ ), so we determine which means are different by looking at the estimates of the means and the pairwise differences between means.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
treatment	3	12.6466667	4.2155556	46.13	0.0002
block	2	123.5450000	61.7725000	675.93	<.0001

The estimates for the least squares means are exactly the same as they were in the previous model, but now the MSE is small (0.091) relative to the differences in means. This allows for more precise inference, and we now detect significant differences among the means. Each of the p-values associated with the LS means are testing whether or not the mean is equal to zero. These tests are usually not of interest, since we anticipate that these means will not be zero. This is the case with our example; we reject all of these individual tests.

Least Squares Means				
treatment	yield LSMEAN	Standard Error	Pr >  t	LSMEAN Number
1	44.7333333	0.1745364	<.0001	1
2	43.5000000	0.1745364	<.0001	2
3	42.5000000	0.1745364	<.0001	3
4	42.0666667	0.1745364	<.0001	4

Next we look at the differences between means. This is a major component of the analysis of the RCB design, but we only look here if the earlier F tests are significant. The p-values below have not been adjusted for multiple comparisons (we did not include an 'adjust' option on the lsmeans statement), so we will need to make this adjustment manually. Using Bonferroni's adjustment, we will reject if the p-value is less than  $0.05/6$ , or 0.0083. From the table below, we see that treatments 3 and 4 have similar means, but all other pairs of means are statistically different.

Least Squares Means for effect treatment				
Pr >  t  for H0: LSMean(i)=LSMean(j)				
Dependent Variable: yield				
i/j	1	2	3	4
1		0.0025	0.0001	<.0001
2	0.0025		0.0067	0.0011
3	0.0001	0.0067		0.1297
4	<.0001	0.0011	0.1297	

The last part of the SAS output are the means and pairwise differences for the blocks. This is not usually a primary focus of the analysis, but it provides information about the usefulness of the blocks. We see that no two blocks have the same mean, so the blocks are capturing the excess variability among the experimental units.

block	yield LSMEAN	Standard Error	Pr >  t	LSMEAN Number
East	46.9000000	0.1511530	<.0001	1
Mid	43.6250000	0.1511530	<.0001	2
West	39.0750000	0.1511530	<.0001	3

Least Squares Means for effect block  
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: yield

i/j	1	2	3
1		<.0001	<.0001
2	<.0001		<.0001
3	<.0001	<.0001	

## Summary

Fertilizer 1 produces the largest crop yield (44.73), which is significantly different from the yield produced by Fertilizer 2 (43.50). The next largest yield is generated by Fertilizer 3 (42.50), which is statistically different from both Fertilizers 1 and 2, but indistinguishable from Fertilizer 4 (42.07).