When analyzing randomized block designs there are two different variations in design that must be recognized. These are, 1) randomized block designs without replication within blocks and 2) randomized block designs with replication within blocks. The first is the most common and includes randomized incomplete block designs (not all treatments occur in every block) and the simple randomized complete block design (every treatment occurs exactly once in every block). These designs generally have greater precision than do designs with replication within blocks, because block size is smaller thus reducing the variance within blocks. On the other hand, when treatments are replicated within blocks the analysis includes the block*treatment interaction which allows the researchers to evaluate the consistency of treatment responses across blocks (different experimental conditions). In some cases this is important knowledge especially in field research.

Regardless of which design variation is used, the block effect is generally considered a *random effect* while the treatment effect may be either fixed or random. A fixed effect is something that you or another researcher could repeat with some knowledge about the outcome based on the completed experiment. For example if you use 4 levels of an antibiotic, the effect is considered fixed because once the experiment is completed you could, with some confidence, predict the effect on the response variable in a repetition of study. Fixed effects are fit by least squares techniques. A random effect is generally not repeatable, or at least something you wouldn't try to replicate. Values of a random effect are assumed to be chosen from a normal distribution of possible effects and replication of the study would result in a different sample of those effects from the population. That is for a random block effect, a repetition of the study would not include the same blocks. Ideally the sampling of the population would be random. When the sampling is not random, and it’s often not, then the researcher must identify the population that he/she believes is represented by the random factor. Blocks would generally be considered random unless a replication of the experiment would use exactly the same blocks, but treatments are generally fixed effects. Therefore, the analysis of randomized block designs are best analyzed as a "mixed" model, which allows the random effects to be fit using maximum likelihood techniques. As its name suggests, the MIXED procedure allows you to specify such a model, while many other anova procedures do not recognize random effects while fitting the model. Because those procedures do not recognize random effects, standard errors and/or tests of hypotheses are frequently incorrect.

Randomized blocks without replication within blocks

If your design *does not* include replication within blocks, the only change as compared to analysis of a CRD, is to include the block effect in your analysis. The residual variance is the appropriate measure of experimental variance and MIXED correctly uses the residual variance to conduct tests of hypotheses.

However, there is one case, although not common, that deserves further consideration. This is when blocking is so ineffective that the estimate of the block variance component is zero. In least squares anova the estimate of the block variance component would have been negative. In this case, one might be tempted to drop the block effect from the analysis and rerun the analysis as a CRD. Unlike least squares anova you don't need to consider dropping these effects in the MIXED procedure. The restricted maximum likelihood restricts the estimate of variance components to be non-negative values. The remaining random variances are estimated given the zero estimate and therefore deleting the source from the model will not change the remaining estimates of variance. The issue is the degrees of freedom for tests of significance. That is, what should happen to the degrees of freedom associated with the zero variance random effect? When the block variance component is zero, the residual variance is equivalent to pooling the block effect with the residual. The question is should the degrees of freedom for tests of significance using the residual include the block degrees of freedom? There is no agreement on this issue. I would argue that one should not get the benefit of those degrees of freedom. The degrees of freedom assigned to blocks is the price that one must pay to use a blocked design and we should not reward researchers that, due to poor planning and ineffective use of blocking. Thus, allowing them to pretend that they never intended to block their experiment and allowing them to pool the block degrees of freedom with the residual degrees of freedom seems inappropriate.

An additional issue of concern when using randomized block designs is what sources of variation should be included in the estimates of the standard errors of the treatment means. The SEM is not only a measure of the precision of the estimated mean, but is also an estimate of how much treatment means would be expected to vary over repetitions of the experiment. However, many anova programs produce what are called *narrow sense estimates* of the SEMs. A narrow sense estimate represents variation over repetitions of the experiment only if one uses exactly the same blocks (thus the same EUs) and simply rerandomizes the assignment of treatments to the experimental units. On the other hand, if blocks are defined as random, MIXED produces *broad sense estimates* of the SEM. A broad sense estimate corresponds to repetitions of the experiment with another sample of blocks (thus different EUs) from the population. This matches the concept of random blocks, since at least in theory, blocks should represent a random sample from a population of potential blocks. I would suggest that you use the MIXED procedure with blocks defined as random for blocked experiments, and then appropriately define the population that is represented by your research. In many cases the population will be quite narrow since most experiments simply use available experimental material. If blocks are in fact fixed effects, then give them a meaningful name that does not imply that there are random replications as is commonly assumed with the word 'blocks'.

PROC MIXED ratio covtest;

The RATIO and COVTEST options are used to better evaluate the relative magnitude of the variance components (VC) and the precision of their estimates by printing confidence limits.

CLASS <blk trt>**;**

The independent variables identifying both block and treatment must appear in the CLASS statement.

MODEL <dep var> **=** <trt> **/**
ddfm=kr outp=<residuals dataset name>;

The MODEL statement contains only fixed sources of variation.

RANDOM <blk>**;**

The RANDOM statement contains only random sources of variation.

ESTIMATE . . .
CONTRAST . . .

ESTIMATE and CONTRAST statements, produce appropriate test of significance and estimates of linear combinations of the fixed effects and their standard errors of the estimates.

LSMEANS <trt> **/ pdiff;**

The LSMEANS statement produces means and standard errors of the means. F ratios and t values are appropriate even if there is no block variance. SEMs are broad sense estimates and therefore contain both block and residual variance, while the SEDs will contain only the residual variance.

Randomized blocks with replicated EUs within blocks

If your design *includes* experimental unit replication within blocks, then the analysis must include both block and block*treatment interactions as random sources of variation. I would recommend that only the MIXED model procedure be used in this case, since the opportunity for producing inappropriate analyses using fixed model least squares programs is great. I find that even those who understand the problems of using fixed model programs on mixed model data frequently fail to recognize when a problem has occurred. In many cases, fixed model programs can be modified to obtain reasonable mixed model analyses, but the appropriate modifications are often overlooked and incorrect tests are commonly reported. MIXED, on the other hand, uses procedures that appropriately fit both random and fixed effects, and uses the expected mean squares for the fixed effects to construct appropriate tests of significance. Therefore, as long as the user correctly identifies the random and fixed sources of variation, there is relatively little chance of incorrectly testing hypotheses as compared to using fixed effect programs.

PROC MIXED ratio covtest;
CLASS <blk trt>**;**

The RATIO option prints the ratio of each VC to the residual and the COVTEST option prints the standard error of the variance and computes the Z ratio.

MODEL <dep var> = <trt> /
ddf=kr outp=<residuals dataset name>**;**

Again note that only the fixed sources of variation appear in the model statement.

RANDOM <blk blk*trt>**;**

*The RANDOM statement includes both random sources of variation, the block and the block*treatment interaction. The MIXED procedure computes the expected mean squares, E(MS), for the fixed effects and constructs appropriate F ratios and t values for tests of the fixed effects from combinations of the random variance components.*

ESTIMATE . . .
CONTRAST . . .

Tests of hypotheses using mean comparison procedures, ESTIMATEs and CONTRASTs are also based on the expected mean squares.

LSMEANS <trt> **/ pdiff;**

*SEMs are the broad sense estimates, in this case containing block, block*treatment and residual variance, while the SEDs will contain variance due to block*treatment and residual.*

Randomized block designs (w/o reps within blocks) with heterogeneous treatment variances

As stated previously, the primary reason for using the MIXED procedure is to more correctly model the random variance as compared to other linear models procedures. Block designs present an even greater challenge to the user who is trying to adequately model the random variance. In this example, we will extend the techniques for fitting heterogeneous variances, that were first presented for the CRD in Lab #3, to randomized block designs.

The data used for this example is based on a constant block variance, but heterogeneous experimental (residual) variances among treatments. There are six complete blocks of four treatments in a 2x2 factorial arrangement. I am presenting a factorial treatment structure to illustrate how the treatment structure may sometimes be used to solve the heterogeneous variance problem. I could have selected a factorial for the CRD and unstructured treatments for the block design to illustrate the two solutions. What is important is that one example be a factorial treatment structure, where the solution to the variance heterogeneity may be associated with only one of the factors.

PROC MIXED ratio covtest
scoring=<# of iterations>**;**
CLASS <classification and variance group variables>**;**

It is sometime difficult to achieve convergence for RCB, heterogeneous variance models. This generally occurs when some variances are zero or near zero, while others are large. SCORING will frequently improve convergence. I commonly use SCORING=10, but more is sometimes required.

MODEL <dependent var>**=**<fixed sources of variation>
/ ddfm=kr
outp=<residuals dataset name>**;**

When both RANDOM and REPEATED statements are used in the MIXED procedure the
The RANDOM statement may be used more than once in the same MIXED procedure. This RANDOM statement defines the blocking variable.

RANDOM <blk>**;**
RANDOM <blk>**/ group=**<trt or highest order factorial source>**;**

This RANDOM statement plus the previous RANDOM is used when you want a separate residual variance for <u>each</u> treatment or treatment combination.

REPEATED / GROUP=<variance group>**;**

The REPEATED statement plus the first random statement (omit the second) is used to request partitioned variances when some treatments or treatment combinations variances are to be pooled.

LSMEANS <fixed sources of variation>
/ pdiff;

SAS program and listing for example randomized block designs

```
TITLE1 Lab #7:  ANALYZING RANDOMIZED BLOCK DESIGNS;

OPTIONS LS=66 PS=54 PAGENO=1;

TITLE2 RANDOMIZED COMPLETE BLOCK DESIGN (RCBD);
TITLE3 A Block is Five Adjacent Trees in a Row;
TITLE4 Data are Number of Dead Citrus Snout Beetles;
DATA rcbd;
INPUT trt blk dead;
LOGDEAD=LOG10(dead);
LINES;

  DATA LINES ENTERED HERE

RUN;
```

```
NOTE: The data set WORK.RCBD has 25 observations and 4 variables.
NOTE: DATA statement used:
      real time           0.39 seconds
      cpu time            0.05 seconds
```

COMMENTS

This is the simplest of the block designs. Every block contains every treatment once and only once.

```
PROC PRINT;
QUIT;
```

Lab #7: ANALYZING RANDOMIZED BLOCK DESIGNS 1
RANDOMIZED COMPLETE BLOCK DESIGN (RCBD)
A Block is Five Adjacent Trees in a Row
Data are Number of Dead Citrus Snout Beetles

Obs	trt	blk	dead	LOGDEAD
1	1	1	24	1.38021
2	1	2	27	1.43136
3	1	3	38	1.57978
4	1	4	44	1.64345
5	1	5	28	1.44716
6	2	1	67	1.82607
7	2	2	54	1.73239
8	2	3	163	2.21219
9	2	4	129	2.11059
10	2	5	134	2.12710
11	3	1	30	1.47712
12	3	2	42	1.62325
13	3	3	87	1.93952
14	3	4	72	1.85733
15	3	5	107	2.02938
16	4	1	117	2.06819
17	4	2	77	1.88649
18	4	3	70	1.84510
19	4	4	161	2.20683
20	4	5	212	2.32634
21	5	1	145	2.16137
22	5	2	99	1.99564
23	5	3	202	2.30535
24	5	4	182	2.26007
25	5	5	251	2.39967

These are the same data that were used in the CRD example. The experiment was in fact blocked and the print of the data at the left now contains the block id code so that it can be analyzed as a RCBD. The block (location) was five adjacent trees in a row. The analysis is on log10(DEAD), since the examination of residuals indicated the need for a log transformation. To keep these examples simple we will assume that the assumptions have been examined and that the data are in the appropriate scale for analysis for all of the data sets analyzed in this chapter.

```
PROC MIXED RATIO COVTEST;
CLASS blk trt;
MODEL logdead = trt / DDFM=KR;
RANDOM blk;
LSMEANS trt / PDIFF;
QUIT;
```

```
NOTE: Convergence criteria met.
NOTE: PROCEDURE MIXED used:
      real time           0.53 seconds
      cpu time            0.07 seconds
```

Block (blk) and treatment (trt) are both in the CLASS statement. The trt source of variation is in the MODEL statement, while the block source of variation is in the RANDOM statement. The RANDOM statement requests that variance components be estimated for the random sources. The rules for creating Expected Mean Squares requires identification of random and fixed sources of variation and the number of levels of each factor. MIXED then generates the expected mean squares for the fixed effects and uses the estimates of the variance components to compute appropriate F ratios and t values for tests of hypotheses about the fixed effects.

```
Lab #7:  ANALYZING RANDOMIZED BLOCK DESIGNS          2
  RANDOMIZED COMPLETE BLOCK DESIGN (RCBD)
  A Block is Five Adjacent Trees in a Row
Data are Number of Dead Citrus Snout Beetles

             Model Information

Data Set                     WORK.RCBD
Dependent Variable           LOGDEAD


          Class Level Information

  Class    Levels    Values

  blk           5    1 2 3 4 5
  trt           5    1 2 3 4 5


             Dimensions

      Covariance Parameters             2
      Observations Used                25


          Iteration History

Iteration    Evaluations    -2 Res Log Like       Criterion

        0              1        -2.32192229
        1              1        -9.85557990      0.00000000


        Convergence criteria met.
```

Just a reminder, throughout this chapter, I have omitted portions of the mixed procedure output, when it is not relevant to the example being presented.

The class level information now contains both blk and trt information since they were both required in the CLASS statement.

```
Lab #7:  ANALYZING RANDOMIZED BLOCK DESIGNS              3
     RANDOMIZED COMPLETE BLOCK DESIGN (RCBD)
      A Block is Five Adjacent Trees in a Row
   Data are Number of Dead Citrus Snout Beetles

                The Mixed Procedure

          Covariance Parameter Estimates

                                 Standard        Z
Cov Parm       Ratio   Estimate      Error    Value       Pr Z

blk           1.1284    0.01848    0.01543     1.20     0.1155
Residual      1.0000    0.01638   0.005791     2.83     0.0023


                  Fit Statistics

        -2 Res Log Likelihood            -9.9
        AIC (smaller is better)          -5.9
        AICC (smaller is better)         -5.1
        BIC (smaller is better)          -6.6


          Type 3 Tests of Fixed Effects

                      Num     Den
        Effect         DF      DF    F Value    Pr > F

        trt             4      16      24.29    <.0001
```

You should compare this analysis with the analysis of the same data analyzed as a CRD in Lab #3. Look for and explain differences in the degrees of freedom, estimates of variance, F ratios, SEMs, etc. Ask yourself: 1) Was giving up 4 error df to block a good trade-off? 2) Did the test of significance of the fixed effect result in greater sensitivity? 3) In general, what did we gain or lose by blocking?

The estimate column contains estimates of the components of variance for the random effects. The SE of the VC relative to the size of the VC (Z=VC/SE) is also a reflection of the precision. In ordinary least squares (OLS) analysis, as presented in the lecture and text, df are part of the calculations for SS and MS. However, OLS does not result in df for variance components (VC). In this case the BLK MS would be based on 5 blocks (4 df), but the df associated with the BLK VC is not 4. There are 25 residuals with 16 df for the residual MS and since the MS=VC the df would be 16 for both. The precision of an estimate of variance is a function of the number of df. The greater the df the more precise the estimates of variance. DF are also needed to determine probabilities for tests of significance. In some MIXED results when it is necessary to estimate df, MIXED uses $2Z^2$ as an estimate of df for VCs. For the BLK VC, df = $2*1.20^2 \approx 3$ df and for the residual VC, df = $2*2.83^2 = 16$ df the same as for a OLS analysis.

For the RCBD this is the correct test of significance for the fixed source of variation according to the expected mean squares, where blocks are random and treatments are fixed. Compare the F ratio here with the one computed in the CRD Lab.

Lab #7: ANALYZING RANDOMIZED BLOCK DESIGNS 4
RANDOMIZED COMPLETE BLOCK DESIGN (RCBD)
A Block is Five Adjacent Trees in a Row
Data are Number of Dead Citrus Snout Beetles

The Mixed Procedure

Least Squares Means

Effect	trt	Estimate	Standard Error	DF	t Value	Pr > |t|
trt	1	1.4964	0.08350	9.41	17.92	<.0001
trt	2	2.0017	0.08350	9.41	23.97	<.0001
trt	3	1.7853	0.08350	9.41	21.38	<.0001
trt	4	2.0666	0.08350	9.41	24.75	<.0001
trt	5	2.2244	0.08350	9.41	26.64	<.0001

Differences of Least Squares Means

Effect	trt	_trt	Estimate	Standard Error	DF	t Value	Pr > |t|
trt	1	2	-0.5053	0.08094	16	-6.24	<.0001
trt	1	3	-0.2889	0.08094	16	-3.57	0.0026
trt	1	4	-0.5702	0.08094	16	-7.04	<.0001
trt	1	5	-0.7280	0.08094	16	-8.99	<.0001
trt	2	3	0.2163	0.08094	16	2.67	0.0167
trt	2	4	-0.06492	0.08094	16	-0.80	0.4343
trt	2	5	-0.2227	0.08094	16	-2.75	0.0142
trt	3	4	-0.2813	0.08094	16	-3.47	0.0031
trt	3	5	-0.4391	0.08094	16	-5.42	<.0001
trt	4	5	-0.1578	0.08094	16	-1.95	0.0689

For the least squares means, the standard errors (SEM) are estimates of the expected variation among treatment means over repetitions of the experiment. Most fixed model anova programs compute narrow sense SEMs. Narrow sense estimates correspond to repetitions of this experiment using the same blocks and thus the same trees, and simply randomly reassigning the treatments to trees within the blocks. However, if one were to repeat the experiment, it seems likely that one would choose different blocks (locations) and thus different trees. To estimate the variation for the broad sense repetition of the experiment, block variance must be included in the estimates of SEMs. The MIXED procedure estimates of the SEMs are broad sense estimates, when BLK is in the random statement and narrow sense when BLK is included in the model statement.

For the differences of least squares means, the standard errors (SED) are the standard errors of the difference between treatment means. This SED does not include the block variance, since the experiment was designed so that treatments are comparable within blocks. That is, repetitions of the study would not be expected to affect the SEDs. These tests of differences among treatment means are also derived from the expected mean squares.

```
TITLE2 RANDOMIZED INCOMPLETE BLOCK DESIGN (RIBD);
TITLE3 Litters are Blocks and Drugs are Treatments;
TITLE4 Data are lymphocyte counts in pigs;
DATA ribd;
INPUT drug$ litter lym_cnt;
LINES;

  DATA LINES ENTERED HERE

RUN;

NOTE: The data set WORK.RIBD has 17 observations and 3 variables.
NOTE: DATA statement used:
      real time           0.01 seconds
      cpu time            0.01 seconds


PROC PRINT;
QUIT;

NOTE: There were 17 observations read from the data set WORK.RIBD.
NOTE: PROCEDURE PRINT used:
      real time           0.01 seconds
      cpu time            0.01 seconds
```

In incomplete block designs, not all blocks contain all treatments. This may be useful when it does not make sense to allow the number of treatments to determine block size. In this case the litters form natural blocks, but the number of useable EUs in a litter varies from two to five. Incomplete block designs may also occur when data are lost from a RCBD. The ideal incomplete block design is the "balanced incomplete block" (BIB) design. In BIB designs every treatment occurs with every other treatment, within blocks, exactly the same number of times, over the complete experiment. An advantage of BIB designs is that the analysis is easily done by hand. If you use MIXED to analyze your data, a BIB design is not necessary. However, the BIB principle, of equally replicated pairs of treatments, should be used when constructing incomplete block designs. That is, one should find a design such that the number of times treatment pairs occur together within blocks is nearly equal.

```
PROC PRINT;
QUIT;
```

```
Lab #7:  ANALYZING RANDOMIZED BLOCK DESIGNS          4
   RANDOMIZED INCOMPLETE BLOCK DESIGN (RIBD)
  Litters are Blocks and Drugs are Treatments
       Data are lymphocyte counts in pigs
```

Obs	drug	litter	lym_cnt
1	A	1	7.1
2	B	1	6.7
3	C	1	7.1
4	P	1	6.7
5	C	1	6.9
6	A	2	6.1
7	B	2	5.1
8	C	2	5.8
9	P	2	5.4
10	A	3	6.9
11	B	3	5.9
12	C	3	6.2
13	A	4	5.6
14	B	4	5.1
15	P	4	5.2
16	C	5	6.2
17	P	5	5.3

An experiment was conducted to examine the effects of drugs (A, B, C and P=placebo) on lymphocyte counts in pigs. The test was to be conducted only on gilts (females) and there were 5 litters available: 1 litter of five, 1 litter of four, 2 litters of three, and 1 litter of two. The treatments were assigned to litters so as to balance as closely as possible the direct comparisons (within block) of the treatment pairs (AB, AC, AP, BC, BP, CP). Each of the six possible pairs occurred within a block either 3 or 4 times (i.e., AC occurs twice in litter 1, and once each in litters 2 and 3). Near equal replication and balancing of the frequency of within block treatment pairs will result in about the same SEDs, and thus sensitivity, for tests of significance among the treatment means. Of course the researcher may intentionally use unequal replication and unbalanced pairing of treatments to increase the sensitivity for tests among selected treatment pairs at the expense of less sensitivity among other pairs.

```
PROC MIXED RATIO COVTEST;
CLASS litter drug;
MODEL lym_cnt = drug / DDFM=KR;
RANDOM litter;
LSMEANS drug / PDIFF;
QUIT;

NOTE: Convergence criteria met.
NOTE: PROCEDURE MIXED used:
      real time           0.06 seconds
      cpu time            0.01 seconds
```

The MIXED syntax is exactly the same for RIBD as for the RCBD.

```
        Lab #7:  ANALYZING RANDOMIZED BLOCK DESIGNS          6
         RANDOMIZED INCOMPLETE BLOCK DESIGN (RIBD)
        Litters are Blocks and Drugs are Treatments
             Data are lymphocyte counts in pigs

                     Model Information

   Data Set                     WORK.RIBD
   Dependent Variable           lym_cnt


                  Class Level Information

     Class     Levels    Values

     litter         5    1 2 3 4 5
     drug           4    A B C P


                        Dimensions

            Covariance Parameters            2
            Observations Used               17


                     Iteration History

  Iteration    Evaluations    -2 Res Log Like       Criterion

          0              1        32.36760557
          1              2        19.76629824      0.90806069
          .
          .
          .
          8              1        14.73018252      0.00000000

                 Convergence criteria met.
```

Incomplete block structures are frequently more difficult to fit than are complete block structures, which is why it required 8 iterations to converge.

Lab #7: ANALYZING RANDOMIZED BLOCK DESIGNS 7
RANDOMIZED INCOMPLETE BLOCK DESIGN (RIBD)
Litters are Blocks and Drugs are Treatments
Data are lymphocyte counts in pigs

Covariance Parameter Estimates

Cov Parm	Ratio	Estimate	Standard Error	Z Value	Pr Z
litter	8.0654	0.3486	0.2549	1.37	0.0857
Residual	1.0000	0.04322	0.02029	2.13	0.0166

The large variance component for litter relative to the residual (Ratio) is evidence of the importance of using the natural blocking (by litter). Any attempt to try to mix gilts from different litters to make complete blocks would likely have resulted in a smaller block variance and a less efficient design.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
drug	3	9.15	11.13	0.0021

This is the F test of the fixed effects.

Least Squares Means

Effect	drug	Estimate	Standard Error	DF	t Value	Pr > |t|
drug	A	6.3823	0.2858	5.07	22.33	<.0001
drug	B	5.6573	0.2858	5.07	19.79	<.0001
drug	C	6.1476	0.2818	4.82	21.82	<.0001
drug	P	5.7150	0.2852	5.04	20.04	<.0001

These SEMs are the broad sense estimates. Even though they are still based on pooled variances (variance components in this case) only the drugs A and B have exactly the same standard errors, but the differences in the SEMs are quite small. These differences in the SEMs are due to the differences in replication (r=4 for A, B, and P; and r=5 for C) and differences in distribution of the treatments across the five blocks.

Differences of Least Squares Means

Effect	drug	_drug	Estimate	Standard Error	DF	t Value	Pr > |t|
drug	A	B	0.7250	0.1470	9.07	4.93	0.0008
drug	A	C	0.2348	0.1479	9.18	1.59	0.1463
drug	A	P	0.6673	0.1559	9.19	4.28	0.0020
drug	B	C	-0.4902	0.1479	9.18	-3.31	0.0088
drug	B	P	-0.05772	0.1559	9.19	-0.37	0.7196
drug	C	P	0.4325	0.1463	9.14	2.96	0.0158

The equality (or inequality) of the standard errors is primarily a reflection of the differences in replication and in the degree of balance for the direct comparison of the treatment pairs. In this case there is about a 6% difference between the largest and smallest SED, which seems to be reasonable balance in the SEs. If one examines these differences in SEDs prior to the start of study, then one can assign the treatments such that tests of greater interest are associated with smaller SEDs.

```
TITLE2 RCBD, EACH CULTIVAR (CV) REPLICATED TWICE PER BLOCK;
TITLE3 Blocks are Three Greenhouse Benches;
TITLE4 Data are heights of plants;
DATA rcbd2x;
INPUT blk rep cv$ height;
LINES;

   DATA LINES ENTERED HERE

RUN;
```

```
NOTE: The data set WORK.RCBD2X has 24 observations and 4
      variables.
NOTE: DATA statement used:
      real time           0.01 seconds
      cpu time            0.01 seconds
```

```
PROC PRINT;
QUIT;
```

```
NOTE: There were 24 observations read from the data set
      WORK.RCBD2X.
NOTE: PROCEDURE PRINT used:
      real time           0.01 seconds
      cpu time            0.01 seconds
```

Randomized block designs may also have more EUs per block than treatments. In this case the experiment contains two levels of replication. The first is replicated EUs within blocks, and the second is the block replication. The present example is a greenhouse experiment examining differences in growth among four different cultivars of a house plant. They have six plants of each cultivar and the assigned greenhouse area has three tables (locations) which would form natural blocks. The experiment is designed with two pots of each cultivar randomly assigned to a location on each table (4 cultivars, 2 pots each = 8 pots per table).

Lab #7: ANALYZING RANDOMIZED BLOCK DESIGNS 7
RCBD, EACH CULTIVAR (CV) REPLICATED TWICE PER BLOCK
Blocks are Three Greenhouse Benches
Data are heights of plants

OBS	BLK	REP	CV	HEIGHT
1	1	1	A	19.3
2	1	2	A	17.2
3	1	1	B	20.1
4	1	2	B	19.4
5	1	1	C	17.4
6	1	2	C	16.6
7	1	1	D	16.6
8	1	2	D	15.7
9	2	1	A	16.7
10	2	2	A	15.5
11	2	1	B	21.2
12	2	2	B	20.8
13	2	1	C	14.4
14	2	2	C	13.6
15	2	1	D	13.5
16	2	2	D	12.9
17	3	1	A	17.7
18	3	2	A	19.8
19	3	1	B	21.0
20	3	2	B	21.9
21	3	1	C	15.8
22	3	2	C	17.4
23	3	1	D	12.8
24	3	2	D	14.7

```
PROC MIXED RATIO COVTEST;
CLASS  blk cv;
MODEL height = cv / DDFM=KR;
RANDOM blk blk*cv;
LSMEANS cv / PDIFF;
QUIT;
```

The key difference in this case is the addition of the BLK*CV interaction to the RANDOM statement. The BLK*CV variance component is a measure of the inconsistency among cultivars to perform the same in different blocks (different environments).

```
        Lab #7:  ANALYZING RANDOMIZED BLOCK DESIGNS         10
    RCBD, EACH CULTIVAR (CV) REPLICATED TWICE PER BLOCK
            Blocks are Three Greenhouse Benches
                 Data are heights of plants

                    The Mixed Procedure

                     Model Information

   Data Set                     WORK.RCBD2X
   Dependent Variable           height


                  Class Level Information

     Class    Levels    Values

     blk           3    1 2 3
     cv            4    A B C D


                        Dimensions

            Covariance Parameters             3
            Observations Used                24


                     Iteration History

  Iteration    Evaluations    -2 Res Log Like       Criterion

          0              1        78.60076245
          1              1        72.16951599      0.00000000


                  Convergence criteria met.
```

The class level information indicates that the data set contains three tables (blk) and four treatments.

```
    Lab #7:  ANALYZING RANDOMIZED BLOCK DESIGNS        11
  RCBD, EACH CULTIVAR (CV) REPLICATED TWICE PER BLOCK
        Blocks are Three Greenhouse Benches
            Data are heights of plants

              Covariance Parameter Estimates

                                  Standard       Z
Cov Parm       Ratio   Estimate      Error   Value      Pr Z
blk           0.6409     0.5432     0.9226    0.59    0.2780
blk*cv        1.1813     1.0012     0.8407    1.19    0.1168
Residual      1.0000     0.8475     0.3460    2.45    0.0072


              Type 3 Tests of Fixed Effects

                    Num     Den
        Effect       DF      DF    F Value    Pr > F
        cv            3       6      15.82    0.0030


                   Least Squares Means

                         Standard
Effect   cv   Estimate      Error     DF   t Value   Pr > |t|
cv       A     17.7000     0.8100   6.51     21.85     <.0001
cv       B     20.7333     0.8100   6.51     25.60     <.0001
cv       C     15.8667     0.8100   6.51     19.59     <.0001
cv       D     14.3667     0.8100   6.51     17.74     <.0001


            Differences of Least Squares Means

                            Standard
 Effect   cv   _cv  Estimate      Error     DF  t Value  Pr > |t|
 cv       A    B     -3.0333     0.9747      6    -3.11    0.0208
 cv       A    C      1.8333     0.9747      6     1.88    0.1090
 cv       A    D      3.3333     0.9747      6     3.42    0.0141
 cv       B    C      4.8667     0.9747      6     4.99    0.0025
 cv       B    D      6.3667     0.9747      6     6.53    0.0006
 cv       C    D      1.5000     0.9747      6     1.54    0.1747
```

We now have estimates of variance components for all three random sources of variation. Relative to the residual, the block source of variation is smaller than the residual variance while the block*cultivar is a little greater than the residual.

The test of the fixed effects indicates that there are significant cultivar differences.

Again these are the broad sense estimates of the standard errors of the means.

These SEDs include both residual and blk*cv variances, since observed differences between the cultivar means would be influenced by both of these variances, but not by the block variance.

```
TITLE2 RCBD, FITTING HETEROGENEOUS TREATMENT VARIANCES;
TITLE3 2x2 Factorial Treatment Structure;
DATA rcbhv;
INPUT blk a b y;
LINES;

   DATA LINES ENTERED HERE

RUN;
```

This experiment is a randomized complete block design with 6 blocks of 4 treatments. The treatment structure is a 2x2 factorial.

```
NOTE: The data set WORK.RCBHV has 24 observations and 4 variables.
NOTE: DATA statement used:
      real time           0.01 seconds
      cpu time            0.01 seconds


PROC PRINT;
QUIT;

NOTE: There were 24 observations read from the data set
      WORK.RCBHV.
NOTE: PROCEDURE PRINT used:
      real time           0.01 seconds
      cpu time            0.01 seconds
```

Lab #7: ANALYZING RANDOMIZED BLOCK DESIGNS 10

RCBD, FITTING HETEROGENEOUS TREATMENT VARIANCES

2x2 Factorial Treatment Structure

OBS	BLK	A	B	Y
1	1	1	1	13.6
2	1	1	2	16.8
3	1	2	1	23.8
4	1	2	2	23.5
5	2	1	1	13.8
6	2	1	2	17.0
7	2	2	1	20.5
8	2	2	2	15.9
9	3	1	1	11.4
10	3	1	2	12.5
11	3	2	1	27.6
12	3	2	2	26.1
13	4	1	1	24.4
14	4	1	2	28.3
15	4	2	1	35.0
16	4	2	2	31.8
17	5	1	1	26.2
18	5	1	2	28.5
19	5	2	1	40.8
20	5	2	2	38.2
21	6	1	1	24.3
22	6	1	2	25.3
23	6	2	1	36.9
24	6	2	2	43.3

```
TITLE4 Standard RCB analysis;
PROC MIXED RATIO COVTEST;
CLASS blk a b;
MODEL y=a b a*b / DDFM=KR;
RANDOM blk;
QUIT;
```

This is the same randomized block syntax as the first analysis presented in this chapter except that the model contains the sources of variation for the factorial analysis.

```
          Lab #7:  ANALYZING RANDOMIZED BLOCK DESIGNS          14
        RCBD, FITTING HETEROGENEOUS TREATMENT VARIANCES
               2x2 Factorial Treatment Structure
                     Standard RCB analysis

                       Model Information

      Data Set                      WORK.RCBHV
      Dependent Variable            y


                    Class Level Information

        Class    Levels    Values

        blk           6    1 2 3 4 5 6
        a             2    1 2
        b             2    1 2


                          Dimensions

               Covariance Parameters             2
               Observations Used                24
```

```
        Lab #7:  ANALYZING RANDOMIZED BLOCK DESIGNS          15
      RCBD, FITTING HETEROGENEOUS TREATMENT VARIANCES
             2x2 Factorial Treatment Structure
                   Standard RCB analysis

                    The Mixed Procedure

               Covariance Parameter Estimates

                                        Standard        Z
Cov Parm        Ratio    Estimate        Error    Value        Pr Z

blk            4.9566     53.2377      35.3823     1.50      0.0662
Residual       1.0000     10.7408       3.9220     2.74      0.0031


                      Fit Statistics

            -2 Res Log Likelihood          126.6
            AIC (smaller is better)        130.6
            AICC (smaller is better)       131.3
            BIC (smaller is better)        130.2


               Type 3 Tests of Fixed Effects

                     Num     Den
        Effect        DF      DF    F Value    Pr > F

        a              1      15      57.08    <.0001
        b              1      15       0.31    0.5875
        a*b            1      15       1.63    0.2211
```

If we find that variances are homogeneous, then the single pooled residual variance of 10.7408 is our best estimate of the experimental variance for all treatments.

At this point, the primary interest is to obtain the model fitting information for comparison with other candidate models.

Note that for this model, the pooled residual variance has Den DF=15. The validity of these tests is dependent on the appropriateness of the assumption of equality of experimental variances among treatments and these tests should be ignored until we complete our examination of assumptions.

```
TITLE4 Fitting separate variances for each treatment;
PROC MIXED RATIO COVTEST SCORING=5;
CLASS blk a b;
MODEL y=a b a*b / DDFM=KR;
RANDOM blk;
RANDOM blk / GROUP=a*b;
QUIT;
```

The new statement here is the second RANDOM statement. This statement requests MIXED to fit five variance components in addition to the BLK variance component from the first RANDOM statement. These are the four variances for each treatment plus an overall residual. The double RANDOM statements sometimes requires many iterations (>30) to converge. The SCORING option on the PROC statement frequently reduces the required number of iterations.

```
          Lab #7:  ANALYZING RANDOMIZED BLOCK DESIGNS          16
        RCBD, FITTING HETEROGENEOUS TREATMENT VARIANCES
               2x2 Factorial Treatment Structure
         Fitting separate variances for each treatment

                       Model Information

     Data Set                    WORK.RCBHV
     Dependent Variable          y
     Group Effect                a*b

                    Class Level Information

       Class    Levels    Values
       blk           6    1 2 3 4 5 6
       a             2    1 2
       b             2    1 2

                          Dimensions
              Covariance Parameters             6
              Observations Used                24

                       Iteration History

  Iteration    Evaluations    -2 Res Log Like       Criterion
          0              1       147.09549494
          1              2       115.77655062      0.00264838
          .
          .
          .
         35              1       114.65019332      0.00000001

                   Convergence criteria met.
```

Lab #7: ANALYZING RANDOMIZED BLOCK DESIGNS 18
RCBD, FITTING HETEROGENEOUS TREATMENT VARIANCES
2x2 Factorial Treatment Structure
Fitting separate variances for each treatment

Covariance Parameter Estimates

Cov Parm	Group	Ratio	Estimate	Standard Error	Z Value	Pr Z
blk		20501511	44.6070	16.2882	2.74	0.0031
blk	a*b 1 1	0	0	.	.	.
blk	a*b 1 2	659529	1.4350	0.9076	1.58	0.0569
blk	a*b 2 1	5317455	11.5697	4.2246	2.74	0.0031
blk	a*b 2 2	15804372	34.3870	12.5564	2.74	0.0031
Residual		1.0000	2.176E-6	0	.	.

Fit Statistics

-2 Res Log Likelihood	114.7
AIC (smaller is better)	124.7
AICC (smaller is better)	128.9
BIC (smaller is better)	123.6

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
a	1	15.9	48.98	<.0001
b	1	15.9	0.26	0.6146
a*b	1	15.9	1.40	0.2541

Note that one of the treatment residual variances is zero. When one of the residual variances is zero, it often requires many iterations to find a solution.

We should now assess the model fitting information to compare this heterogeneous variance model which requires estimating 6 parameters with the standard RCB which requires estimating 2 parameters.

Criteria	Model Parameters Two	Six
AICC	131.3	125.9
-2 Res Log Likelihood	126.6	114.7

LRT = 126.6-114.7 = 11.9, df=6-2=4, P<.01

The conclusion is that fitting block variance plus four treatment variances is a better fit than the standard analysis with only block and residual variances.

```
TITLE5 Fitting separate variances for levels of factor A;
PROC MIXED SCORING=5 COVTEST;
CLASS blk a b;
MODEL y=a b a*b / DDFM=KR;
RANDOM blk;
REPEATED / GROUP=a;
LSMEANS a b a*b / PDIFF;
QUIT;
```

The above approach is used only when you want to fit a residual variance for each of the treatment combinations in a simple RBD. If you decide to pool some treatment variances, then replace the second RANDOM with the REPEATED statement as presented in Lab #3. In this example, the GROUP=a requests that MIXED compute separate variances for the two levels of factor A, while pooling across the levels of B. This was suggested as an alternative model by examination of the variance components from the analysis of the previous model.

```
         Lab #7:  ANALYZING RANDOMIZED BLOCK DESIGNS          19
       RCBD, FITTING HETEROGENEOUS TREATMENT VARIANCES
              2x2 Factorial Treatment Structure
        Fitting separate variances for each treatment
      Fitting separate variances for levels of factor A

                      Model Information
    Data Set                      WORK.RCBHV
    Dependent Variable            y
    Group Effect                  a

                   Class Level Information
      Class    Levels    Values
      blk           6    1 2 3 4 5 6
      a             2    1 2
      b             2    1 2

                         Dimensions
             Covariance Parameters            3
             Observations Used               24

                      Iteration History
  Iteration    Evaluations    -2 Res Log Like       Criterion
          0              1       147.09549494
          1              2       131.62005019      0.05765364
          .
          .
          .
          9              1       117.94242002      0.00000000


                  Convergence criteria met.
```

Lab #7: ANALYZING RANDOMIZED BLOCK DESIGNS 20
RCBD, FITTING HETEROGENEOUS TREATMENT VARIANCES
2x2 Factorial Treatment Structure
Fitting separate variances for each treatment
Fitting separate variances for levels of factor A

Covariance Parameter Estimates

Cov Parm	Group	Estimate	Standard Error	Z Value	Pr Z
blk		45.1508	28.7801	1.57	0.0583
Residual	a 1	0.7271	0.4660	1.56	0.0593
Residual	a 2	26.9044	12.3198	2.18	0.0145

Fit Statistics

-2 Res Log Likelihood	117.9
AIC (smaller is better)	123.9
AICC (smaller is better)	125.4
BIC (smaller is better)	123.3

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
a	1	10.1	44.37	<.0001
b	1	10.1	0.24	0.6355
a*b	1	10.1	1.27	0.2863

This is the goodness of fit info for heterogeneous variances for the levels of factor A and homogeneous variances for the levels of factor B. Here I used a RANDOM statement to specify the block VC and the REPEATED to request the two VCs for the levels of factor A.

These estimates of random variances are listed under the column labeled Estimate.

Now compare the model fitting information for all three models.

Criteria	# of Model Parameters Two	Three	Six
AICC	131.3	125.4	125.9
-2 LL	126.6	117.9	114.7

LRT = 126.6-117.9 = 8.7, df=3-2=1, P<.002
LRT = 117.9-114.7 = 3.2, df=6-3=3, p>.2

The conclusion is that fitting the three variance parameters is better than the two, and that fitting six is not better than three variance parameters.

The partitioning of the variances has resulted in a decrease in DDF from 15 to 10.1 as compared to a single pooled variance, but 10.1 as compared to 8.55 when a variance was computed for each treatment.

Lab #7: ANALYZING RANDOMIZED BLOCK DESIGNS 21
RCBD, FITTING HETEROGENEOUS TREATMENT VARIANCES
2x2 Factorial Treatment Structure
Fitting separate variances for each treatment
Fitting separate variances for levels of factor A

The Mixed Procedure

Least Squares Means

Effect	a	b	Estimate	Standard Error	DF	t Value	Pr > |t|
a	1		20.1750	2.7542	5	7.33	0.0007
a	2		30.2833	3.1252	7.94	9.69	<.0001
b		1	24.8583	2.9456	6.48	8.44	<.0001
b		2	25.6000	2.9456	6.48	8.69	<.0001
a*b	1	1	18.9500	2.7652	5.08	6.85	0.0009
a*b	1	2	21.4000	2.7652	5.08	7.74	0.0005
a*b	2	1	30.7667	3.4654	10.6	8.88	<.0001
a*b	2	2	29.8000	3.4654	10.6	8.60	<.0001

Differences of Least Squares Means

Effect	a	b	_a	_b	Estimate	Standard Error	DF	t Value	Pr > |t|
a	1		2		-10.1083	1.5174	10.1	-6.66	<.0001
b		1		2	-0.7417	1.5174	10.1	-0.49	0.6355
a*b	1	1	1	2	-2.4500	0.4923	4.87	-4.98	0.0045
a*b	1	1	2	1	-11.8167	2.1460	10.1	-5.51	0.0003
a*b	1	1	2	2	-10.8500	2.1460	10.1	-5.06	0.0005
a*b	1	2	2	1	-9.3667	2.1460	10.1	-4.36	0.0014
a*b	1	2	2	2	-8.4000	2.1460	10.1	-3.91	0.0029
a*b	2	1	2	2	0.9667	2.9947	9.54	0.32	0.7538

Again, the appropriate VC are used to estimate the standard errors for both means and differences.

One still needs to check normality. In this case, the distribution of residuals should be normal within levels of factor A.