

## Chapter 13 : Experimental design

**Linear Models** –  $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$  the basic completely randomized design, or one-way ANOVA

Source	d.f.	Expected Mean Squares
Treatments	$t - 1$	$\sigma^2 + n\sigma_\tau^2$
Error	$t(n - 1)$	$\sigma^2$

### Error structure (random effects)

**Nested error terms**  $Y_{ijkl} = \mu + \tau_i + \gamma_{ij} + \delta_{ijk} + \varepsilon_{ijkl}$

Source	d.f.	Expected Mean Squares
Treatments	$t - 1$	$\sigma^2 + n\sigma_\gamma^2 + nd\sigma_\delta^2 + ndg\sigma_\tau^2$
Experimental error	$t(n - 1)$	$\sigma^2 + n\sigma_\gamma^2 + nd\sigma_\delta^2$
Sampling error	$tn(s - 1)$	$\sigma^2 + n\sigma_\gamma^2$
SubSampling Error	$tns(b - 1)$	$\sigma^2$

In this case the experimental error is the error needed for testing treatments. The other errors are not used, but the replicates appear in the coefficient of the treatment error term, so there is a gain in power.

**Experimental units and experimental error** – the unit to which the treatment is assigned is the experimental unit. The variation in these units is used to calculate an error called the experimental error. This is the error used for testing treatments.

**Sampling units and sampling error** – these units within the experimental unit improve the estimates of the means and add power to the analysis. These units also provide an additional error term nested within the experimental error.

**Subsampling units** – these units within the sampling units provide additional improvement in the estimates of the means and add power to the analysis. These units also provide an additional error term nested within the sampling error.

The multiple experimental and sampling units are called replicates. Note that replicates at every level appear in the coefficient of the treatment variance component and increase power.

**Example 1** – In agricultural experiments a field is often divided into a number of plots.

Treatments may be assigned completely at random to these plots creating a Completely Randomized Design. Imagine that there are 30 plots in a soybean field, all planted in the same way with the same variety of soybeans. The objective is to examine weed control in the plots. Five herbicides are to be tested ( $t=5$ ), each assigned at random to 6 of the plots ( $p=6$ ). The herbicides are applied and after two weeks the investigator goes to each plot and places a 0.5 m<sup>2</sup> frame at 3 randomly selected spots in each plot ( $n=3$ ). The percent coverage of weeds is estimated in each of the frames. The source table for the analysis of the percent coverage in each frame is as follows.

The sources are herbicide, plot(herbicide), and frame(plot\*herbicide). The model in SAS GLM might look like “MODEL Y = HERB PLOT(HERB)” where the term with parenthesis is a nested term and is read “plot within herbicide treatment”. In order to get the treatment (HERB) tested with the experimental error term (PLOT(HERB)) in GLM we would add the request “TEST H=HERB E=PLOT(HERB)”. The last term

(FRAME(PLOT\*HERB)) is not put in the program statements and SAS will pick this up as an error term, often referred to as the residual error term.

PROC MIXED is set up a little differently. In PROC MIXED, only fixed effects go in the model (while in GLM all sources go in the model). The random effects must go in a separate RANDOM statement. The setup in PROC MIXED would be similar to the following: MODEL Y = HERB; RANDOM PLOT(HERB);” As with the GLM, the last term (FRAME(PLOT\*HERB)) is not put in the program statements and will be picked this up as another error term.

$$Y_{ijk} = \mu + \tau_i + \delta_{ij} + \varepsilon_{ijk}$$

Source	d.f.	Expected Mean Squares
Treatments	$\tilde{t}-1$	$\sigma^2 + n\sigma_\delta^2 + np\sigma_\tau^2$
Experimental error	$t(p-1)$	$\sigma^2 + n\sigma_\delta^2$
Sampling error	$tp(n-1)$	$\sigma^2$
Total	$tpn-1$	

**Example 2** – The university lakes are to be tested to determine if levels of bacteria are different in the various lakes ( $t=4$ ). Ten sites ( $s=10$ ) are to be selected a random from each lake. At each site three replicate samples are taken in 3 one pint bottles ( $b=2$ ). In the laboratory each pint bottle is used to prepare two agar plates. The plates are incubated for 48 hours and the number of bacterial colonies is counted.

$$Y_{ijkl} = \mu + \tau_i + \gamma_{ij} + \delta_{ijk} + \varepsilon_{ijkl}$$

Source	d.f.	Expected Mean Squares
Treatments	$\tilde{t}-1$	$\sigma^2 + n\sigma_\gamma^2 + nb\sigma_\delta^2 + nbs\sigma_\tau^2$
Experimental error	$t(s-1)$	$\sigma^2 + n\sigma_\gamma^2 + nb\sigma_\delta^2$
Sampling error	$ts(b-1)$	$\sigma^2 + n\sigma_\gamma^2$
Subsampling error	$tsb(n-1)$	$\sigma^2$
Total	$tsbn-1$	

**Blocks** – these are sources of variation that are handled like treatments, but they are not sources that are of interest to the investigator. If not addressed, this variation will add extra variation to the error term and reduce power. To prevent this, these sources are placed in the model. These sources of variation are usually considered “random” effects. I will not use the term “replicates” to refer to blocks; this term will be reserved for multiple observations for a treatment or within an experimental unit. The analysis is called a **Randomized Block Design**.

**Blocks are a form of replication**, and are often referred to as “replicates”. If an experiment cannot be conducted all at one time, we may do one “replicate” now, including all of the treatment levels at least once. The experiment may then be repeated at different times as often as needed to achieve the desired number of replicates. Unfortunately, since the experiment was not done all at once, there may be some differences in the experimental material or ambient conditions between the different replicate periods. To account for this additional variation, the periods are put in the model as blocks.

**Blocks are a means of removing or accounting for a source of variation that cannot be avoided.** Perhaps we can do all of our replicates at the same time, but not in the same place. As a result, there is an inability to treat all replicates equally. For example, if all of our replicates are Petri dishes that cannot be placed in the same incubator, or set up as plots in a field or placed in the same greenhouse, we may have to split-up our replicates (i.e. use several incubators, fields

or greenhouses). In this case the differences between incubators, fields or greenhouses may become a source of variation.

When experiments are repeated in different years or at different sites the variation due to these sites or years are addressed by blocking. This type of replication, or variation, is often included in an experiment on purpose so an experimental result will be show to work under broader conditions than are represented by a single year or a single site.

**Blocks interactions represent an experimental error term.** The blocks themselves are factored out of the model as an effect with  $b-1$  degrees of freedom. The treatments, of course, represent another  $t-1$  d.f. After these sources are put in the model, an additional  $(t-1)(b-1)$  degrees of freedom remain. These represent an experimental error term, and if included in the model are represented by a “block by treatment interaction” (e.g. block \* treatment).

**Other forms of replication may be included in the experiment in addition to the blocks.** If treatments occur only once in each block, then the block by treatment interaction is the only error term. However, additional experimental units (plots in a field, potted plants or Petri dishes) may be employed in each block, providing additional replication. Sampling units and Subsampling units may also be taken in each block by treatment unit, also providing additional replication. Examples below include a randomized block design without additional replicates and with additional replicates.

**Example 3** – Sometimes the objective of blocking is purely to provide additional replication. If an experimenter needs to test 3 dietary treatments on fish, but only has 3 tanks available, the experiment can be set up in the 3 tanks and repeated as many times as necessary to provide replication.

In the experiment below an investigator is testing 4 “sunless tanning” products for skin tone. He pays volunteer subjects to apply each of the 4 treatments to a small area on the subject’s lower leg. Each subject gets all 4 treatments, but each subject provides only a single observation. Replication is provided by recruiting numerous subjects for the study. The day after application of the treatment each subject returns to have the color tone evaluated. This experiment is outlined below where subjects are the blocks.

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

Source	d.f.	Expected Mean Squares
Block	$\tilde{b} - 1$	$\sigma^2 + t\sigma_\beta^2$
Treatment	$\tilde{t} - 1$	$\sigma^2 + b\sigma_\tau^2$
Block * Treatment	$(\tilde{t} - 1)(\tilde{b} - 1)$	$\sigma^2$

**Example 4** – In agricultural experiments a field is often divided into a number of plots.

Occasionally it is noted that the field is not homogeneous; fertility, soil moisture or other factors important to the experiment vary across the field. If the investigator is aware of these patterns, he can divide the field in subunits, perhaps east to west, that are more homogeneous. If there are 30 plots placed in a heterogeneous field, they could be placed in 6 rows such that the section of the field in each row could be more homogeneous. The variability due to heterogeneity in the field can then be taken up by the blocks. If the objective is to examine weed control in the plots, the five herbicides are to be tested ( $t=5$ ), each assigned at random to one of 6 blocks ( $b=6$ ) such that all five treatments occur once in each block. The herbicides are applied and after two weeks the investigator goes to each plot and places a 0.5 m<sup>2</sup> frame at 3 randomly selected spots in each plot ( $n=3$ ). The percent coverage of weeds is estimated in each of the frames. The source table for the analysis of the percent coverage in each frame is as follows.

The sources are herbicide, block, block\*herbicide, and frame(plot\*herbicide). The model in SAS GLM might look like “MODEL Y = BLOCK HERB BLOCK\*HERB” where the term with parenthesis is an interaction and is read “block by herbicide interaction”. In order to get the treatment (HERB) tested with the experimental error term (BLOCK\*HERB) in GLM we would add the request “TEST H=HERB E= BLOCK\*HERB”. The last term (FRAME(PLOT\*HERB)) is not put in the program statements and SAS will pick this up as an error term, often referred to as the residual error term.

PROC MIXED is set up a little differently. In PROC MIXED, only fixed effects go in the model and the random effects must go in a separate RANDOM statement. The setup in PROC MIXED would be similar to the following: MODEL Y = HERB; RANDOM BLOCK BLOCK\*HERB” As with the GLM, the last term (FRAME(PLOT\*HERB)) is not put in the program statements and will be picked this up as another error term.

$$Y_{ijk} = \mu + \tau_i + \beta_j + \tau\beta_{ij} + \varepsilon_{ijk}$$

Source	d.f.	Expected Mean Squares
Block	$\bar{b} - 1$	$\sigma^2 + n\sigma_{\tau\beta}^2 + nt\sigma_{\beta}^2$
Treatment	$\bar{t} - 1$	$\sigma^2 + n\sigma_{\tau\beta}^2 + nb\sigma_{\tau}^2$
Block * Treatment	$(\bar{t} - 1)(\bar{b} - 1)$	$\sigma^2 + n\sigma_{\tau\beta}^2$
Rep(block*treatment)		$\sigma^2$

### Treatment arrangements

**Single factor** – the experiments that we have generally seen have had a single treatment. We will call these treatment arrangements a “single factor” treatment arrangement, and they can occur with either Completely Randomized Designs (CRD) or Randomized Block Designs (RBD). These experiments are relatively simple and are represented by all of the examples above.

**Factorial** – this treatment arrangement is known as two-way Analysis of Variance or factorial treatment arrangements. The design can be either CRD or RBD. The arrangement consists of two (or more) treatments that are cross classified such that each level of the first treatment occurs with each level of the second treatment and vice versa. These are one of the most common and power types of experimental design. An example of a 3 by 5 cross classified treatment arrangement for treatments A and B is given below.

**CRD factorial:**  $Y_{ijk} = \mu + \tau_i + \tau_{2i} + \tau_1\tau_{2ij} + \varepsilon_{ijk}$

**RBD factorial with no additional reps:**  $Y_{ijk} = \mu + \beta_i + \tau_{1j} + \tau_{2k} + \tau_1\tau_{2jk} + \tau_1\tau_2\beta_{ijk}^*$

**RBD factorial with additional reps:**  $Y_{ijkl} = \mu + \beta_i + \tau_{1j} + \tau_{2k} + \tau_1\tau_{2jk} + \tau_1\tau_2\beta_{ijk}^* + \varepsilon_{ijkl}$

\* In randomized block designs, all block by treatment interactions are usually pooled into a single block interaction term. In SAS, with possible interactions  $\beta\tau_1$ ,  $\beta\tau_2$  and  $\beta\tau_1\tau_2$ , if the lower order interactions are not included ( $\beta\tau_1$  and  $\beta\tau_2$ ), they will automatically be pooled with the higher order term ( $\beta\tau_1\tau_2$ ).

	A1	A2	A3	A4	A5
B1	a <sub>1</sub> b <sub>1</sub>	a <sub>2</sub> b <sub>1</sub>	a <sub>3</sub> b <sub>1</sub>	a <sub>4</sub> b <sub>1</sub>	a <sub>5</sub> b <sub>1</sub>
B2	a <sub>1</sub> b <sub>2</sub>	a <sub>2</sub> b <sub>2</sub>	a <sub>3</sub> b <sub>2</sub>	a <sub>4</sub> b <sub>2</sub>	a <sub>5</sub> b <sub>2</sub>
B3	a <sub>1</sub> b <sub>3</sub>	a <sub>2</sub> b <sub>3</sub>	a <sub>3</sub> b <sub>3</sub>	a <sub>4</sub> b <sub>3</sub>	a <sub>5</sub> b <sub>3</sub>

**Nested** – this treatment arrangement is not common, and occurs when the levels of a treatment have levels of a second treatment associated with them. The nested treatment arrangement differs from the factorial in that the levels of the second treatment are not the same for each of the first treatments. For example, there are three levels of treatment A below with 8 levels of treatment B. Note that the levels of B are different for each level of A. This treatment arrangement is not as common as the two above (single factor and factorial), so we will not cover it vary heavily.

	A1			A2			A3	
B	a <sub>1</sub> b <sub>1</sub>	a <sub>1</sub> b <sub>2</sub>	a <sub>1</sub> b <sub>3</sub>	a <sub>2</sub> b <sub>4</sub>	a <sub>2</sub> b <sub>5</sub>	a <sub>2</sub> b <sub>6</sub>	a <sub>3</sub> b <sub>7</sub>	a <sub>3</sub> b <sub>8</sub>

**Fixed and random effects** – While blocks and error terms are almost invariably random effects, treatments may be either fixed or random. The distinction is particularly important with a factorial treatment arrangement because of its effect on the tests that should be done in the various cases.

Initially, expected mean squares (EMS) are developed the same for random and fixed effects. Once developed, the EMS are modified so that fixed effects reflect the fact that they estimate the sum of squared fixed effects. This is often designated simply as a subscripted Q, where the q represents the whole fixed effects term. (e.g.

$$Q_{\tau} = n \sum_{i=1}^t \left( \frac{\tau_i^2}{t-1} \right).$$

Note how the three models below change for a factorial CRD with two random effects, with mixed effects and with both effects fixed. Recall that each term is tested with an error term that contains all variance components except the term being tested.

$$Y_{ij} = \mu + \tau_{1i} + \tau_{2j} + \tau_{1\tau_{2ij}} + \varepsilon_{ijk}$$

Source	d.f.	Expected Mean Squares (Random)
Treatment 1	$\tilde{t}_1 - 1$	$\sigma^2 + n\sigma_{\tau_1\tau_2}^2 + n\tilde{t}_2\sigma_{\tau_1}^2$
Treatment 2	$\tilde{t}_2 - 1$	$\sigma^2 + n\sigma_{\tau_1\tau_2}^2 + n\tilde{t}_1\sigma_{\tau_2}^2$
Interaction	$(\tilde{t}_1 - 1)(\tilde{t}_2 - 1)$	$\sigma^2 + n\sigma_{\tau_1\tau_2}^2$
Error	$\tilde{t}_1\tilde{t}_2(n-1)$	$\sigma^2$

Source	d.f.	Expected Mean Squares (Mixed)
Treatment 1	$\tilde{t}_1 - 1$	$\sigma^2 + n\sigma_{\tau_1\tau_2}^2 + Q_{\tau_1}$
Treatment 2	$\tilde{t}_2 - 1$	$\sigma^2 + n\sigma_{\tau_1\tau_2}^2 + n\tilde{t}_1\sigma_{\tau_2}^2$
Interaction	$(\tilde{t}_1 - 1)(\tilde{t}_2 - 1)$	$\sigma^2 + n\sigma_{\tau_1\tau_2}^2$
Error	$\tilde{t}_1\tilde{t}_2(n-1)$	$\sigma^2$

Source	d.f.	Expected Mean Squares (Fixed)
Treatment 1	$\tilde{t}_1 - 1$	$\sigma^2 + Q_{\tau_1}$
Treatment 2	$\tilde{t}_2 - 1$	$\sigma^2 + Q_{\tau_2}$
Interaction	$(\tilde{t}_1 - 1)(\tilde{t}_2 - 1)$	$\sigma^2 + Q_{\tau_1\tau_2}$
Error	$\tilde{t}_1\tilde{t}_2(n-1)$	$\sigma^2$

**Other possible patterns of terms exist.** Blocks are not limited to a single term. It is possible to have several levels of blocking. It is also possible to have a second set of treatments nested within the first set of treatments (a split – plot design). Frequently, repeated measures are taken on an experimental unit giving a “repeated measures design”. These advanced designs will not be covered in detail.

### Schematic layouts of some CRD experiments.

1) CRD: Treatment levels are  $t=5$  where each treatment occurs in 3 cells (experimental units :  $n=3$ ) and each cell is sampled once. This is a CRD with only experimental error.

<b>A</b>	<b>A</b>	<b>D</b>
<b>B</b>	<b>C</b>	<b>B</b>
<b>D</b>	<b>A</b>	<b>E</b>
<b>E</b>	<b>D</b>	<b>C</b>
<b>B</b>	<b>E</b>	<b>C</b>

Model	$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$
Source	Expected Mean Squares
Treatments	$\sigma^2 + n\sigma_\tau^2$
Error	$\sigma^2$

2) CRD: Treatment levels are  $t=5$  where each treatment occurs in 3 cells (experimental units :  $n=3$ ) and each cell is sampled twice (sampling units :  $d=2$ ). This is a CRD with both experimental error and sampling error.

<b>A</b>	<b>A</b>	<b>A</b>	<b>A</b>	<b>D</b>	<b>D</b>
<b>B</b>	<b>B</b>	<b>C</b>	<b>C</b>	<b>B</b>	<b>B</b>
<b>D</b>	<b>D</b>	<b>A</b>	<b>A</b>	<b>E</b>	<b>E</b>
<b>E</b>	<b>E</b>	<b>D</b>	<b>D</b>	<b>C</b>	<b>C</b>
<b>B</b>	<b>B</b>	<b>E</b>	<b>E</b>	<b>C</b>	<b>C</b>

Model	$Y_{ijk} = \mu + \tau_i + \delta_{ij} + \varepsilon_{ijk}$
Source	Expected Mean Squares
Treatment	$\sigma^2 + n\sigma_\delta^2 + nd\sigma_\tau^2$
Experimental Error	$\sigma^2 + n\sigma_\delta^2$
Sampling Error	$\sigma^2$

3) CRD?: Treatment levels= $5$ ; each treatment occurs in one cell (experimental unit) and each cell is sampled three times. CRD with no experimental error, this experiment has sampling error only (Pseudo-replication).

<b>C</b>	<b>C</b>	<b>C</b>
<b>A</b>	<b>A</b>	<b>A</b>
<b>D</b>	<b>D</b>	<b>D</b>
<b>E</b>	<b>E</b>	<b>E</b>
<b>B</b>	<b>B</b>	<b>B</b>

Model	$Y_{ijk} = \mu + \tau_i + \varepsilon_{ijk}$
Source	Expected Mean Squares
Treatments	$\sigma^2 + n\sigma_\delta^2 + nd\sigma_\tau^2$
Sampling Error	$\sigma^2$

### Schematic layouts of some RBD experiments.

4) RBD: Three blocks (b=3) and five treatment levels (t=5) where each block has 5 cells (experimental units) and each treatment occurs once in each block and each is sampled once.  
RBD with experimental error only

Block 1	Block 2	Block 3
C	B	D
A	C	B
D	A	E
E	D	A
B	E	C

Model	$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$
Source	Expected Mean Squares
Treatment	$\sigma^2 + nb\sigma_\tau^2$
Block	$\sigma^2 + nt\sigma_\beta^2$
Error	$\sigma^2$

5) RBD: Three blocks (b=3) and five treatment levels (t=5) where each block has 5 cells, and each treatment occurs once in each block and each cell is sampled twice (d=2). In this case the residual error term measures only the within cell variation.

Block 1	Block 2	Block 3
C C	B B	D D
A A	C C	B B
D D	A A	E E
E E	D D	A A
B B	E E	C C

Model	$Y_{ijk} = \mu + \tau_i + \beta_j + \tau\beta_{ij} + \varepsilon_{ijk}$
Source	Expected Mean Squares
Treatment	$\sigma^2 + n\sigma_{\tau\beta}^2 + nd\sigma_\tau^2$
Block	$\sigma^2 + n\sigma_{\tau\beta}^2 + nd\sigma_\beta^2$
Block*Treatment	$\sigma^2 + n\sigma_{\tau\beta}^2$
Error	$\sigma^2$

6) RBD: Blocks=3, Treatment levels=5; each block has 10 cells and each treatment occurs twice in each block. Each cell (treatment by block and within block replicate combination) is sampled once. In this case the residual error term measures between cell variation.

Block 1	Block 2	Block 3
D C	A E	C A
A E	B A	D B
A E	E C	A E
B C	C D	D B
D B	B D	E C

Model	$Y_{ijk} = \mu + \tau_i + \beta_j + \tau\beta_{ij} + \varepsilon_{ijk}$
Source	Expected Mean Squares
Treatment	$\sigma^2 + n\sigma_{\tau\beta}^2 + nd\sigma_\tau^2$
Block	$\sigma^2 + n\sigma_{\tau\beta}^2 + nd\sigma_\beta^2$
Block*Treatment	$\sigma^2 + n\sigma_{\tau\beta}^2$
Error	$\sigma^2$

**Layouts of some other experiments.**

7) CRD Split-plot : Main plot treatment levels are  $t_1=5$ , sub-plot treatment levels  $t_2=2$  where each main plot treatment occurs in  $d=3$  cells (experimental units) and each cell is subdivided into two parts ( $d=2$ ) for one application of each level of the sub-plot treatment.

<b>E<sub>2</sub>   E<sub>1</sub></b>	<b>B<sub>2</sub>   B<sub>1</sub></b>	<b>D<sub>1</sub>   D<sub>2</sub></b>
<b>A<sub>2</sub>   A<sub>1</sub></b>	<b>C<sub>1</sub>   C<sub>2</sub></b>	<b>B<sub>1</sub>   B<sub>2</sub></b>
<b>D<sub>1</sub>   D<sub>2</sub></b>	<b>B<sub>1</sub>   B<sub>2</sub></b>	<b>C<sub>2</sub>   C<sub>1</sub></b>
<b>E<sub>2</sub>   E<sub>1</sub></b>	<b>D<sub>2</sub>   D<sub>1</sub></b>	<b>A<sub>1</sub>   A<sub>2</sub></b>
<b>A<sub>1</sub>   A<sub>2</sub></b>	<b>E<sub>2</sub>   E<sub>1</sub></b>	<b>C<sub>1</sub>   C<sub>2</sub></b>

Model	$Y_{ijk} = \mu + \tau_{1i} + \delta_{ij} + \tau_{2k} + \tau_1\tau_{2ik} + \varepsilon_{ijk}$
Source	Expected Mean Squares
Treatment	$\sigma^2 + nd\sigma_{\delta}^2 + n\sigma_{\tau_1\tau_2}^2 + nd t_2\sigma_{\tau_1}^2$
Error a	$\sigma^2 + t_2\sigma_{\delta}^2$
SubPlot Treatment	$\sigma^2 + n\sigma_{\tau_1\tau_2}^2 + nt_1\sigma_{\tau_2}^2$
Interaction	$\sigma^2 + n\sigma_{\tau_1\tau_2}^2$
Error b	$\sigma^2$

8) RBD: Blocks=3, Treatment Levels=5; each block has 10 cells and each treatment occurs twice in each block. Each cell (treatment by block combination) is sampled twice.

Block				Block				Block			
<b>D</b>	<b>D</b>	<b>C</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>E</b>	<b>E</b>	<b>C</b>	<b>C</b>	<b>A</b>	<b>A</b>
<b>A</b>	<b>A</b>	<b>E</b>	<b>E</b>	<b>B</b>	<b>B</b>	<b>A</b>	<b>A</b>	<b>D</b>	<b>D</b>	<b>B</b>	<b>B</b>
<b>A</b>	<b>A</b>	<b>E</b>	<b>E</b>	<b>E</b>	<b>E</b>	<b>C</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>E</b>	<b>E</b>
<b>B</b>	<b>B</b>	<b>C</b>	<b>C</b>	<b>C</b>	<b>C</b>	<b>D</b>	<b>D</b>	<b>D</b>	<b>D</b>	<b>B</b>	<b>B</b>
<b>D</b>	<b>D</b>	<b>B</b>	<b>B</b>	<b>B</b>	<b>B</b>	<b>D</b>	<b>D</b>	<b>E</b>	<b>E</b>	<b>C</b>	<b>C</b>

Model	$Y_{ijk} = \mu + \tau_i + \beta_j + \tau\beta_{ij} + \delta_{ij} + \varepsilon_{ijl}$
Source	Expected Mean Squares
Treatment	$\sigma^2 + n\sigma_{\tau}^2 + ns\sigma_{\tau\beta}^2 + nsb\sigma_{\tau}^2$
Block	$\sigma^2 + n\sigma_{\gamma}^2 + ns\sigma_{\tau\beta}^2 + nst\sigma_{\beta}^2$
Treatment*Block	$\sigma^2 + n\sigma_{\gamma}^2 + ns\sigma_{\tau\beta}^2$
Rep(Treatment*Block)	$\sigma^2 + n\sigma_{\gamma}^2$
Error	$\sigma^2$

9) Latin Square Design (LSD): Three rows ( $r=3$ ) and three columns ( $c=3$ ) for three treatment levels ( $t=3$ ) where each treatment occurs in each row and column. Experimental units are row by column combination cells. There could also be several separate squares (blocks) and several sampling units within cells for Latin Square Designs.

<b>C</b>	<b>B</b>	<b>A</b>
<b>A</b>	<b>C</b>	<b>B</b>
<b>B</b>	<b>A</b>	<b>C</b>

Model	$Y_{ijk} = \mu + \rho_i + \beta_j + \tau_k + \varepsilon_{ijk}$
Source	Expected Mean Squares
Row	$\sigma^2 + nr\sigma_{\rho}^2$
Column	$\sigma^2 + nr\sigma_{\beta}^2$
Treatment	$\sigma^2 + nr\sigma_{\tau}^2$
Error	$\sigma^2$

### Orthogonal Polynomial multipliers (equally spaced X)

levels = 3		
X	l	q
1	-1	1
2	0	-2
3	1	1

levels = 4			
X	l	q	c
1	-3	1	-1
2	-1	-1	3
3	1	-1	-3
4	3	1	1

levels = 5				
X	l	q	c	q
1	-2	2	-1	1
2	-1	-1	2	-4
3	0	-2	0	6
4	1	-1	-2	-4
5	2	2	1	1

levels = 6					
1	-5	5	-5	1	-1
2	-3	-1	7	-3	5
3	-1	-4	4	2	-10
4	1	-4	-4	2	10
5	3	-1	-7	-3	-5
6	5	5	5	1	1

levels = 7						
1	-3	5	-1	3	-1	1
2	-2	0	1	-7	4	-6
3	-1	-3	1	1	-5	15
4	0	-4	0	6	0	-20
5	1	-3	-1	1	5	15
6	2	0	-1	-7	-4	-6
7	3	5	1	3	1	1

levels = 8							
1	-7	7	-7	7	-7	1	-1
2	-5	1	5	-13	23	-5	7
3	-3	-3	7	-3	-17	9	-21
4	-1	-5	3	9	-15	-5	35
5	1	-5	-3	9	15	-5	-35
6	3	-3	-7	-3	17	9	21
7	5	1	-5	-13	-23	-5	-7
8	7	7	7	7	7	1	1

levels = 9								
1	-4	28	-14	14	-4	4	-1	1
2	-3	7	7	-21	11	-17	6	-8
3	-2	-8	13	-11	-4	22	-14	28
4	-1	-17	9	9	-9	1	14	-56
5	0	-20	0	18	0	-20	0	70
6	1	-17	-9	9	9	1	-14	-56
7	2	-8	-13	-11	4	22	14	28
8	3	7	-7	-21	-11	-17	-6	-8
9	4	28	14	14	4	4	1	1

For levels of X that are not equally spaced there is a SAS IML instruction that will produce the orthogonal polynomial multipliers. The following statements will do this if you have SAS IML available.

```
OPTIONS PS=60 LS=78;
```

```
PROC IML;  
  RESET PRINT;  
  X={1 , 2 , 3 , 4 , 8};  
  O=ORPOL(X,3); RUN;  
  QUIT;
```

where the X vector gives the levels of the quantitative variable.

The orpol function needs one parameter specifying the name of the quantitative variable vector and a second parameter specifying the number of orthogonal polynomials levels desired.