Topic 6: Randomized Complete Block Designs (RCBD's)

[ST&D sections 9.1 – 9.7 (except 9.6) and section 15.8]

6.1. Variability in the completely randomized design (CRD)

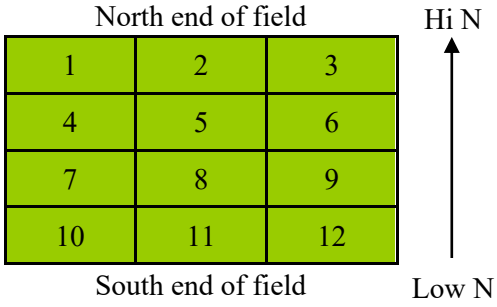
In the CRD, it is assumed that all experimental units are uniform. This is not always true in practice, and it is necessary to develop methods to deal with such variability. When comparing two methods of fertilization, if one region of the field has much greater natural fertility than the others, a treatment effect might be incorrectly ascribed to the treatment applied to that part of the field, leading to a Type I error. For this reason, when conducting a CRD, it is always advocated to include as much of the native variability of the experiment as possible *within* each experimental unit (e.u.), making each e.u. as representative of the whole experiment, and the whole experiment as uniform, as possible. In actual field studies, plots are designed to be long and narrow to achieve this objective. But if the e.u.'s are more variable, experimental error (MSE) is larger, F (MST/MSE) is smaller, and the experiment is less sensitive. And if the experiment is replicated in a variety of situations to increase its scope, the variability increases even further. This additional variability needs to be removed from the analysis so that the actual effects of treatment can be detected. This is the purpose of blocking.

6.2. Randomized complete block design (RCBD)

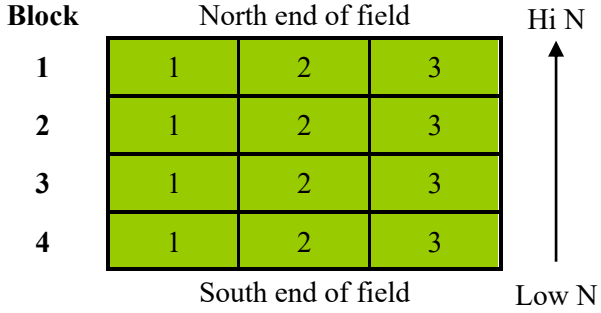
6.2.1. Definition

The RCBD assumes that a population of experimental units can be divided into a number of relatively homogeneous subpopulations or ***blocks***. The treatments are then randomly assigned to experimental units such that each treatment occurs equally often (usually once) in each block (i.e. each block contains all treatments). Blocks usually represent levels of naturally-occurring differences or sources of variation that are unrelated to the treatments, and *the characterization of these differences is not of interest to the researcher*. In the analysis, the variation among blocks can be partitioned out of the experimental error (MSE), thereby reducing this quantity and increasing the power of the test.

6.2.2. Example: Consider a field trial comparing three cultivars (A, B, and C) of sugar beet with four replications (in this case, the field is divided into 12 plots; each plot is a replication / e.u.). Suppose the native level of soil nitrogen at the field site varies from high at the north end to low at the south end (see diagram). In such a situation, yield is expected to vary from one end of the field to the other another, *regardless of cultivar differences*. This violates the assumption that the error terms are randomly distributed since the residuals will tend to be positive at the north end of the field and negative at the south end.

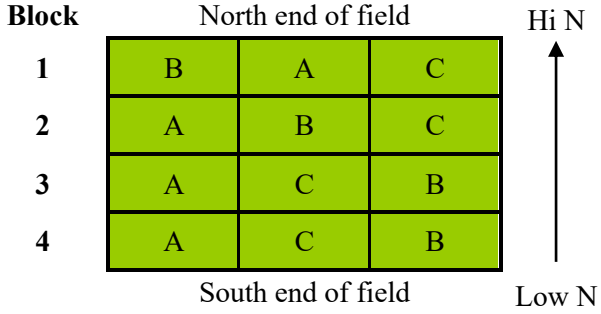


One strategy to minimize the impact of this variability in native soil fertility on the analysis of treatment effects is to divide the field into four east-west blocks of three plots each.



Because these blocks run perpendicular to the nitrogen gradient, the soil within each of these blocks will be relatively uniform. This is the basic idea of the randomized complete block design. Remember that in the *completely randomized design (CRD)*, each e.u. in the experiment has an equal chance of being assigned any treatment level (i.e. a single randomization is performed for the entire experiment). This is not the case in an RCBD. In the *randomized complete block design (RCBD)*, each e.u. *in a given block* has the same chance of being chosen for each treatment (i.e. a separate randomization is performed for each block). Within each block, a fixed number (often 1) of e.u.'s will be assigned to each treatment level. The term "complete" refers to the fact that all treatment levels are represented in each block (and, by symmetry, that all blocks are represented in each treatment level).

After the four separate randomizations, one for each block, the field could look like this:



6.2.3. The linear model

In the case of a single replication per block-treatment combination (like the example above), the underlying linear model that explains each observation is:

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

Here, as before, τ_i represents the effect of Treatment i (i = 1,...,t), such that the average of each treatment level is $\overline{T}_i = \mu + \tau_i$. Now, in a similar way, β_j represents the effect of Block j (j = 1,...,r), such that the average of each block is $\overline{B}_j = \mu + \beta_j$. As always, ε_{ij} are the residuals, the deviations of each observation from their expected values. The model in dot notation:

$$Y_{ij} = \overline{Y}_{..} + (\overline{Y}_{i.} - \overline{Y}_{..}) + (\overline{Y}_{.j} - \overline{Y}_{..}) + (Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})$$

And the sum of squares:

$$\sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \overline{Y}_{..})^2 = r\sum_{i=1}^{t}(\overline{Y}_{i.} - \overline{Y}_{..})^2 + t\sum_{j=1}^{r}(\overline{Y}_{.j} - \overline{Y}_{..})^2 + \sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})^2$$

TSS = SST + SSB + SSE

Since the variance of means of n observations is σ^2/n, the coefficients ***r*** and ***t*** (within SST and SSB, respectively) ensure that all mean squares are estimates of the same σ^2 when there are no block or treatment effects. This is another example of *partitioning* of variance, made possible because the sums of squares of blocks and treatments are *orthogonal* to one another. This orthogonality is a direct result of the completeness of the block design.

6.2.4. ANOVA

ANOVA table for the RCBD (one replication per block-treatment combination):

Source	df	SS	MS	F
Total	rt - 1	TSS		
Treatments	t - 1	SST	SST/(t-1)	MST/MSE
Blocks	r - 1	SSB	SSB/(r-1)	
Error	(r-1)(t-1)	TSS-SST-SSB	SSE/(r-1)(t-1)	

ANOVA table for the CRD:

Source	df	SS	MS	F
Total	rt – 1	TSS		
Treatments	t - 1	SST	SST/(t-1)	MST/MSE
Error	t(r - 1)	TSS-SST	SSE/r(t-1)	

Due to the additional factor in the linear model, the ANOVA table for the RCBD has an additional row (Block) relative to that for the CRD. Notice that one consequence of this is that there are fewer degrees of freedom for error in the RCBD design than in the CRD design [(r-1)(t-1) vs. t(r-1), or (r - 1) fewer degrees of freedom]. In the RCBD, these (r – 1) degrees of freedom have been partitioned from the error and assigned to the blocks.

Situation 1: No differences among blocks (i.e. no block effects)

If the RCBD design were applied to an experiment in which the blocks were really no different from one another (i.e. there were no significant block effect), the MSE for the CRD would be smaller than the MSE for the RCBD simply due to the differences in error degrees of freedom. For example, if t = 3 and r = 4, MSE_{CRD} = SSE/9 and MSE_{RCBD} = SSE/6. Therefore, the F statistic for the CRD would be larger, meaning the CRD would be the more powerful (sensitive) design.

To think of this another way, consider the general form of a confidence interval for the difference between two means (H_0: $\overline{Y}_A - \overline{Y}_B = 0$):

$$\left(\overline{Y}_A - \overline{Y}_B\right) \pm CritStatistic_{\alpha, df_{MSE}} \sqrt{MSE \frac{2}{r}}$$

If there are no block effects, the half-length of this confidence interval will be smaller for the CRD than for the RCBD for *two* reasons:

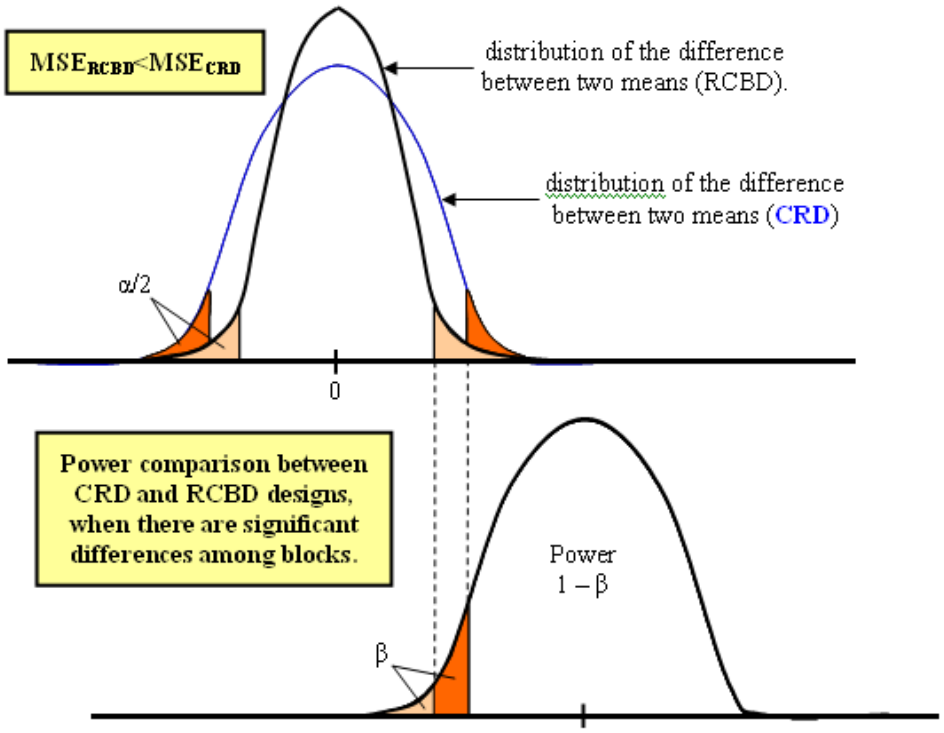
1. The CRD will have a smaller critical value in the above formula due to its larger error degrees of freedom.
2. $MSE_{CRD} < MSE_{RCBD}$ due to difference in error degrees of freedom.

The larger critical value and the larger MSE in the RCBD moves the threshold of rejection further from the mean than in the CRD. This change in the rejection threshold affects the Type II error (β) and the power of the test (1- β). Under this scenario, the probability of accepting a false null hypothesis (β) will be smaller in the CRD than in the RCBD. In other words, the CRD would in this situation be more powerful.

Situation 2: Significant difference among blocks

Now suppose that there really are substantial differences among blocks as well as among treatments (H_0 is false). In a CRD, this variation due to differences among blocks would remain in the error (i.e. would not be partitioned from the error). This larger MSE would make the F statistic (MST/MSE) for the CRD smaller (less significant) than the F statistic for the RCBD.

Under this scenario, the RCBD would still have a larger **critical** (i.e. tabular) F value because of the lost degrees of freedom; but this may be more than compensated by the smaller MSE. If the effect of the reduced MSE (increased F statistic) outweighs the effect of the larger critical value (rejection threshold further from 0), the net result will be a smaller β and thus a larger power in the RCBD relative to the CRD.



Obviously, one should only use the RCBD when the variation explained by the blocks more than offsets the penalty associated with having fewer error degrees of freedom. So how can one determine when an RCBD is appropriate? This question is answered using the concept of *efficiency*, introduced in Section 1.4.4.6 and elaborated upon in section 6.3.

6.2.5. Example (from Little and Hills)

This experiment was conducted to investigate the effect of estrogen on weight gain in sheep.

The four treatments in the experiment are a factorial combinations of two separate factors: Gender of sheep (male and female) and amount of estrogen (S0 and S3). Although this

experiment could be analyzed as a factorial, in this example we are treating the four treatments and four levels of a single factor (gender-estrogen combination).

Sheep from four different ranches were involved in the experiment. Anticipating that differences in herd management may affect the results, the researchers blocked by ranch. The completeness of an RCBD demanded, therefore, that each ranch volunteer four sheep to the experiment, two males and two females, providing one replication of each treatment level from each ranch.

Table 6.1 RCBD. Effect of estrogen on weight gain in sheep (lbs).

	Ranch (i.e. block)				Treatment	
Treatment	**I**	**II**	**III**	**IV**	**Total**	**Mean**
F-S0	47	52	62	51	212	53
M-S0	50	54	67	57	228	57
F-S3	57	53	69	57	236	59
M-S3	54	65	74	59	252	63
Block Total	208	224	272	224	928	
Block Mean	52	56	68	56		58

Table 6.2 RCBD ANOVA

Source	df	SS	MS	F
Total	15	854		
Blocks	3	576	192.00	24.69**
Treatment	3	208	69.33	8.91**
Error	9	70	7.78	

Table 6.3 CRD ANOVA

Source	df	SS	MS	F
Totals	15	854		
Treatment	3	208	69.33	1.29 NS
Error	12	646	53.83	

Since each treatment is present at the same level of replication within each block, differences among blocks are not the result of treatment effects. Differences among blocks are entirely independent of treatment effects and are due only to differences associated with the four ranches. Therefore, this component (SSB) can be perfectly partitioned from the total SS. Ultimately, this reduces the experimental error. To see this, compare the two tables above (Tables 6.2 and 6.3), paying close attention to the degrees of freedom and the SS in each analysis.

6.3. Relative efficiency [ST&D 221, and Topic 1 section 1.4.4.6]

We saw earlier that if the variation among blocks is large then we can expect the RCBD to be more sensitive to treatment effects than the CRD; conversely, if this variation is small, the CRD may be more sensitive (i.e. more powerful). The concept of *relative efficiency* formalizes the comparison between two experimental methods by quantifying this balance between loss of degrees of freedom and reduction in experimental error.

Recall that the F statistic = MST/MSE. The experimental design primarily affects the MSE since the degrees of freedom for treatments is always (t – 1) and the variation due to treatments is independent of (i.e. orthogonal to) the variation due to blocks and the experimental error. The information per replication in a given design is:

$$I = \frac{1}{\sigma_\varepsilon^2}$$

Therefore, the relative efficiency of one design another is

$$RE_{1:2} = \frac{I_1}{I_2} = \frac{\frac{1}{\sigma_{\varepsilon 1}^2}}{\frac{1}{\sigma_{\varepsilon 2}^2}} = \frac{\sigma_{\varepsilon 2}^2}{\sigma_{\varepsilon 1}^2}$$

In reality, we never know the true experimental error (σ_ε^2); we only have an *estimate* of it (MSE). To pay for this lack of knowledge, a correction factor is introduced into the expressions for information (I) and relative efficiency (RE) (Cochram and Cox, 1957). The following formulas include this correction factor and give an estimate of the relative amount of information provided by two designs:

$$I = \frac{1}{\sigma_\varepsilon^2} \approx \left(\frac{df_{MSE} + 1}{df_{MSE} + 3} \right) \frac{1}{MSE}$$

$$RE_{1:2} = \frac{I_1}{I_2} \approx \frac{\left(\frac{df_{MSE1} + 1}{df_{MSE1} + 3} \right) \frac{1}{MSE_1}}{\left(\frac{df_{MSE2} + 1}{df_{MSE2} + 3} \right) \frac{1}{MSE_2}} = \frac{(df_{MSE1} + 1)(df_{MSE2} + 3) MSE_2}{(df_{MSE2} + 1)(df_{MSE1} + 3) MSE_1}$$

where MSE_i is the mean square error from experimental design i. If this ratio is greater than 1, it means that Design 1 provides more information per replication and is therefore more efficient than Design 2. If $RE_{1:2} = 2$, for example, each replication in Design 1 provides twice as much information as each replication in Design 2. Design 1 is twice as efficient.

The main problem with the approach is how to estimate MSE for the alternative design. Suppose an experiment is conducted as an RCBD. The MSE for this design is simply given by the analysis (MSE_{RCBD}). But now we wish to ask the question: What *would have been* the value of the MSE *if* the experiment had been conducted as a CRD? In fact, it was not conducted as a CRD. The treatments were not randomized according to a CRD. Because of this, one cannot just re-analyze the data as though it were a CRD and use the MSE from the analysis as a valid estimate of MSE_{CRD}.

MSE_{CRD} can be *estimated*, however, by the following formula (ST&D 222):

$$\hat{MSE}_{CRD} \cong \frac{df_B MSB_{RCBD} + (df_T + df_e) MSE_{RCBD}}{df_B + df_T + df_e}$$

where MSB and MSE are the block and error mean squares in the original design (RCBD), and df_B, df_T, and df_e are the block, treatment, and error degrees of freedom in the original design. To obtain this formula, the total SS of the two designs are assumed equal. This equation is then expanded such that the SS are rewritten in terms of the underlying variance components of the expected MS. Simplification of the terms generates the above estimate (for a complete derivation, see Sokal & Rohlf 1995, Biometry 838-839).

From the sheep experiment, $MSE_{RCBD} = 7.78$ and $MSB_{RCBD} = 192.0$. Therefore:

$$\hat{MSE}_{CRD} \cong \frac{df_B MSB_{RCBD} + (df_T + df_e) MSE_{RCBD}}{df_B + df_T + df_e} = \frac{3(192) + (3+9)7.78}{3+3+9} = 44.62$$

And…

$$RE_{RCBD:CRD} \cong \frac{(df_{MSE1} + 1)(df_{MSE2} + 3)\hat{MSE}_{CRD}}{(df_{MSE2} + 1)(df_{MSE1} + 3)MSE_{RCBD}} = \frac{(9+1)(12+3)44.62}{(12+1)(9+3)7.78} = 5.51$$

Interpretation: It takes 5.51 replications in the CRD to produce the same amount of information as one replication in the RCBD. Or, the RCBD is 5.51 time more efficient than the CRD in this case. It was a very good idea to block by ranch.

6.4. Assumptions of the model

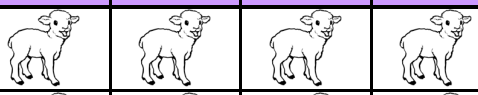
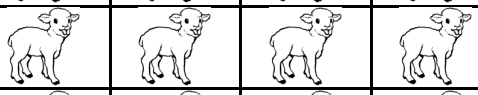
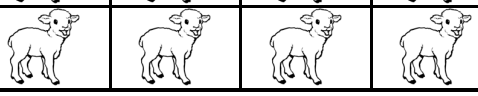
Again, the model for the RCBD with a single replication per block-treatment combination is:

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

As in the CRD, it is assumed that the residuals (ε_{ij}) are independent, homogeneous, and normally distributed. Also as in the CRD, it is assumed that the variance within each treatment levels is homogeneous across all treatment levels. But now, in an RCBD without replication (i.e. with a

single replication per block-treatment combination), there is a third assumption of the model: Additivity of main effects.

Recall that experimental error is defined as the variation among experimental units *that are treated alike*. With that in mind, consider the following schematic of our sheep experiment:

	Ranch			
Trtmt	**1**	**2**	**3**	**4**
M Est$_0$				
M Est$_3$				
F Est$_0$				
F Est$_3$				

In this experiment, while there are four reps of each level of treatment and four reps of each block, there is **no true replication** vis-à-vis calculation of experimental error. For example, there is only one male sheep at Ranch 1 that received no estrogen. Normally, our estimate of the experimental error would come from looking at the variation among two or more sheep treated alike (e.g. two or more sheep of the same gender, at the same ranch, receiving the same estrogen treatment). So if we have no ability to calculate the experimental error, what is the ε_{ij} in our linear model?

There is an *expected* value for each of the 16 cells in the above diagram, given by:

$$\text{Expected } Y_{ij} = \mu + \tau_i + \beta_j$$

In this design, we use the deviation of the observed values from their expected value as estimates of the experimental error. Technically, though, these deviations are the combined effects of experimental error *and* any nonzero block*treatment interaction for that cell. With only one replication per cell, we are unable to separate these two effects. So when we use these deviations (observed – expected) as an estimate of the experimental error, we are assuming that there are **no** significant block*treatment interactions (i.e. no significant non-additive effects).

Said another way, in this model:

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

The residuals are the results of experimental error and any non-additive treatment*block interactions:

$$\varepsilon_{ij} = \tau_i * \beta_j + \text{error}_{ij}$$

Thus, when we use ε_{ij} as estimates of the true experimental error, we are assuming that $\tau_i * \beta_j = 0$.

This assumption of no interaction in a two-way ANOVA is referred to as the assumption of **additivity** of the main effects. If this assumption is violated, it's an indication that your blocks are not behaving as you expected; in other words, there is something of interest lurking within your blocking variable that you need to better understand.

Example: A significant interaction term will result if the effect of the two factors A and B on the response variable Y is multiplicative rather than additive. This is one form of non-additivity.

	Factor A			
Factor B	$\tau_1 = +1$	$\tau_2 = +2$	$\tau_3 = +3$	
$\beta_1 = +1$	2	3	4	Additive effects
	1	2	3	Multiplicative effects
	0	0.30	0.48	Log of multiplicative effects
$\beta_2 = +5$	6	7	8	Additive effects
	5	10	15	Multiplicative effects
	0.70	1.00	1.18	Log of multiplicative effects

In the above table, additive and multiplicative treatment effects are shown in a hypothetical two-way ANOVA. Let us assume that the population mean is $\mu = 0$. Then the mean of the e.u.'s subjected to level 1 of factor A and level one of factor B should be 2 by the conventional additive model. Similarly, the expected subgroup mean subjected to level 3 of factor A and level 2 of factor B is 8, since the respective contributions to the mean are 3 and 5. If the process is multiplicative rather than additive, however, as occurs in a variety of physicochemical and biological phenomena, the expected values are quite different. For treatment A_3B_2, the expected value is 15, the product of 3 and 5.

If multiplicative data of this sort are analyzed by a conventional ANOVA, the interaction SS will be large due to the nonadditivity of the treatment effects. If this SS is embedded in the SSE, as in the case of an RCBD with one e.u. per block-treatment combination, the estimate of the experimental error will be artificially large, thereby making all F tests artificially insensitive.

In the case of multiplicative effects, there is a simple remedy. Transforming the variable by taking the log of each mean will restore additivity. The third line in each cell gives the logarithm of the expected value, assuming multiplicative relations. After the transformation, the increments are strictly additive again ($\tau_1=0$, $\tau_2=0.30$, $\tau_3=0.48$, $\beta_1=0$, $\beta_1=0.70$). This is a good illustration of how transformations of scale can be used to meet the assumptions of analysis of variance.

6.4.1 Tukey's 1-df test for nonadditivity [ST&D 395]

John Tukey devised a very clever method of testing for significant non-additive effects (i.e. interactions) in datasets that lack the degrees of freedom necessary to include such effects (i.e. interactions) directly in the model. Here's the logic behind the test:

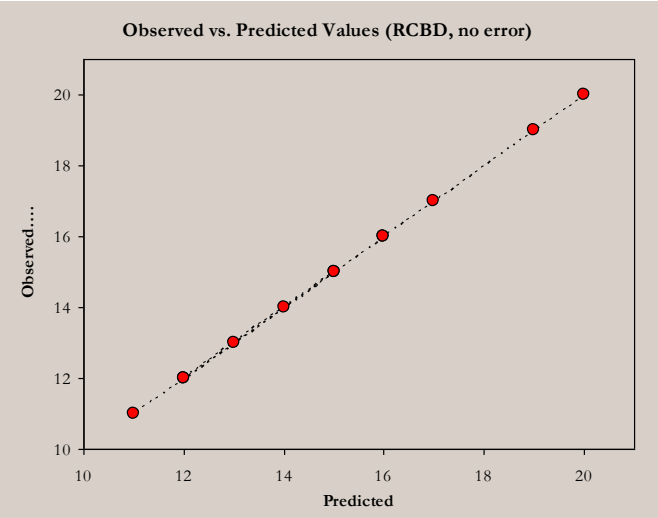
To begin, recall that under our linear model, each observation is characterized as:

$$y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij}$$

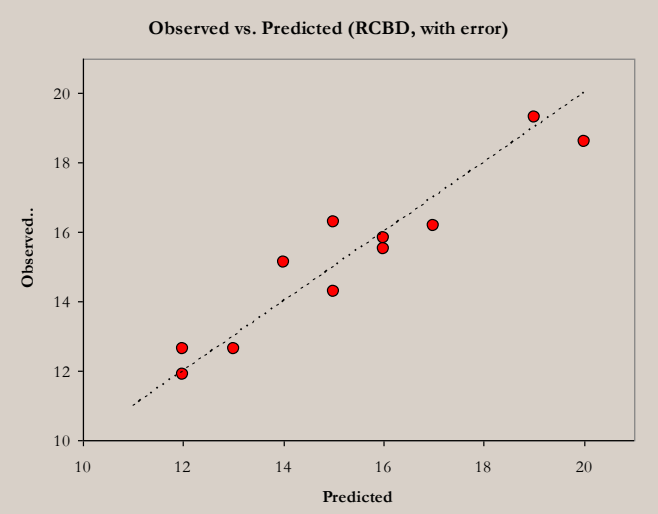
Therefore, the predicted value of each individual is given by:

$$pred_{ij} = \mu + \beta_i + \tau_j$$

In looking at these two equations, the first thing to notice is the fact that, if we had no error in our experiment (i.e. if $\varepsilon_{ij} = 0$), the observed data would exactly match its predicted values and a correlation plot of the two would yield a perfect line with slope = 1:



Now let's introduce some error. If the errors in the experiment are in fact random and independent (criteria of the ANOVA and something achieved by proper randomization from the outset), then ε_{ij} will be a *random variable* that causes *no systematic deviation* from this linear relationship, as indicated in the next plot:



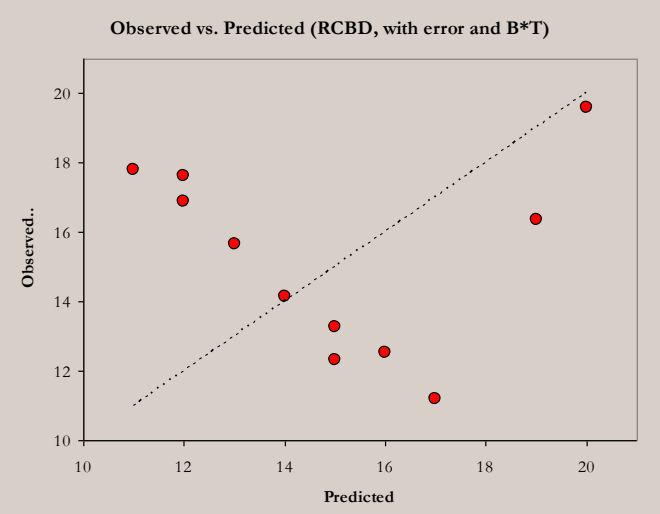
As this plot shows, while random error may decrease the overall strength of correlation, it will not systematically compromise its underlying linear nature.

But what happens when you have an interaction (e.g. Block * Treatment) but lack the degrees of freedom necessary to include it in the linear model (e.g. when you have only 1 replication per block*treatment combination)? In this case, the df and the variation assigned to the interaction are relegated to the error term simply because we need a nonzero df_{error} to carry out our F tests. Under such circumstances, you can think of the error term as now containing two separate components:

$$\varepsilon_{ij} = \varepsilon_{RANDOMij} + \text{B*T Interaction Effects}$$

While the first component is random and will not affect the underlying linear correlation seen above, the second component is non-random and will cause systematic deviations from linearity. Indeed, if this interaction component is too large, the observed vs. predicated correlation will become detectably non-linear, thereby violating the ANOVA assumption of random and independent error, not to mention making your F tests much less sensitive.

The plot on the following page illustrates the deviation from linearity that results when significant multiplicative effects (one kind of nonadditive effect) cannot be accommodated by the model. The quadratic (i.e. non-linear) trend is unmistakable.



SO, if the observed and predicted values obey a linear relationship, then the non-random Interaction Effects buried in the error term are sufficiently small to uphold our assumption of random, independent error.

Seen in this light, our test for unaccounted-for nonadditivity [significant nonadditive (i.e. interaction) effects] becomes a simple test for linear regression, which is what Tukey's 1-df test is. It is a regression of the observed values with the *squares* of the predicted values. Why the squares? Because, as mentioned before when talking about contrasts, *to establish the existence of a linear relationship (as opposed to a correlation of a higher power), one must test for (and successfully reject H_o for) a quadratic trend*.

Please note: This test is necessary **ONLY** when there is **one observation** per block-treatment combination. If there are two or more replications per block-treatment combination, the block*treatment interaction can be tested directly in an exploratory model.