An Intuitive Graphical Approach to Understanding the Split-Plot Experiment

Timothy J. Robinson
University of Wyoming

William A. Brenneman and William R. Myers
The Proctor & Gamble Company

Abstract

While split-plot designs have received considerable attention in the literature over the past decade, there seems to be a general lack of intuitive understanding of the error structure of these designs and the resulting statistical analysis. Typically, students learn the proper error terms for testing factors of a split-plot design via *expected mean squares*. This does not provide any true insight as far as why a particular error term is appropriate for a given factor effect. We provide a way to intuitively understand the error structure and resulting statistical analysis in split-plot designs through building on concepts found in simple designs, such as completely randomized and randomized complete block designs, and then provide a way for students to "see" the error structure graphically. The discussion is couched around an example from paper manufacturing.

1. Introduction

Many industrial and agricultural experiments involve two types of factors, some with levels hard or costly to change and others with levels that are relatively easy to change. Examples of hard-to-change factors include mechanical set-ups, environmental factors, and many others. When hard-to-change factors exist, it is in the practitioner's best interest to minimize the number of times the levels of these factors are changed. A common strategy is to run all combinations of the easy-to-change factors for a given setting of the hard-to-change factors. This restricted randomization of the experimental run order results in a split-plot design (SPD).

Although great technical strides have been made in terms of the design and analysis of SPDs, there seems to be a general lack of intuitive understanding of the error terms and the resulting statistical analysis. Typically, students learn the proper error terms for testing factors of a split-plot design via *expected mean squares*. While this context is certainly important, we have found in our own consulting and teaching experience that the expected mean square framework does not provide any true insight as far as why a particular error term is appropriate for a given factor effect. In this manuscript, we hope to improve the fundamental understanding of SPDs by taking a first-principles approach to describing the error structure through building on concepts already familiar to students in simple designs and provide a way for students to "see" the SPD error structure graphically. The examples provided here are all of the industrial variety and while they may be more interesting to those who teach and consult with engineers, the discussion is valid within any context of a split-plot design.

Since SPDs are essentially two or more error control experimental designs superimposed on top of one another, we follow the notation of Hinkelmann and Kempthorne (1994) by denoting a given split-plot design as SPD(D_w,D_s) where D_w and D_s refer to the designs in the whole-plot and sub-plot factors, respectively. An extensive, but not exhaustive list of references on the design and analysis of SPDs includes Letsinger, Myers, and Lentner (1996); Christensen (1996); Huang, Chen and Voelkel (1998); Rao (1998); Bingham and Sitter (2001); Webb, Lucas and Borkowski (2004); Federer and King (2007); Smith and Johnson (2007); and Kowalski, Parker and Vining (2007).

Two common SPDs are designs in which the whole-plot factor levels are assigned via a completely randomized design (CRD) and the sub-plot factors are assigned via a randomized complete block design (RCBD) [i.e. SPD(CRD,RCBD)] and designs in which both the whole-plot factor levels and sub-plot factor levels are randomly assigned within a RCBD [i.e. SPD(RCBD, RCBD)]. We begin in Section 2 with a review of the CRD and then move along to the RCBD in Section 3. In Section 4 we extend our discussion to split-plot designs. Throughout we show how the error structure dictated by the experimental design can be explored through graphical methods. A common example from paper manufacturing will be discussed in all settings in order to unify the presentation.

2. Completely Randomized Designs CRD's

The most commonly assigned design structure for experiments is the CRD. The CRD assumes the availability of a set of homogenous *experimental units* (EUs). *Experimental units* are the physical entities to which a factor level combination is applied. The experimental unit, upon exposure to a factor level combination is considered a *replicate* of the treatment combination. *Replication* or *replicated design* refers to the occurrence of two or more replicates for a given treatment combination. To illustrate terminology, we refer to a modified version of the tensile strength example from Montgomery (2001). In this example, a paper manufacturer is interested in determining the effect of three different preparation (henceforth referred to as prep) methods (Z) on the tensile strength of paper. For simplicity, we will refer to the levels of Z as 1, 2, and 3. We will assume that there are enough resources to produce nine batches of pulp (three batches for each level of Z). Since the levels of Z are randomly assigned to the batches, the batches are the experimental units. Consider the replicated 3-level design provided in Table 1. The notion of a CRD is that the order in which the prep methods are utilized to produce batches of material is randomized.

Table 1. A replicated 3-level design for paper manufacturing example

	Replicate								
	1	1	2	1	3	2	2	3	3
Prep Method	2	3	2	1	2	1	3	3	1
Tensile Strength	38.75	31	37.25	34.5	39.5	35.25	37.5	33.25	37.25

A possible model for the 3-level CRD is

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1,2,3; j = 1,2,3, \tag{1}$$

where y_{ij} is the ijth observation, μ is the overall mean, τ_i, is the ith treatment effect and ε_{ij} is the experimental error component. Experimental error describes the variation among identically and independently treated experimental units. For the CRD, it is typically assumed that that the ε_{ij} are

i.i.d. $N(0, \sigma^2)$. The experimental error variance, σ^2, describes the variance of observations on experimental units, for which the differences among the observations can only be attributed to experimental error. The magnitude of σ^2 is a function of a variety of sources, including 1. natural variation among EUs; 2. observation/measurement error; 3. inability to reproduce the treatment combinations exactly from one replicate to another; 4. interaction of treatments and replicates; and 5. other unaccounted for sources of variation.

In the CRD, the experimental error variance will be determined by the differences associated with the replicates nested within treatment. Specifically, one would look at the three prep method i replicates (y_{i1}, y_{i2}, y_{i3}) and see how their tensile strength values differ from their group mean $(\bar{y}_{i.})$. These estimated differences are then pooled together to get one estimate of the experimental error variance,

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{3}\sum_{j=1}^{3}\left(y_{ij} - \bar{y}_{i.}\right)^2}{df\ error}. \tag{2}$$

The error degrees of freedom (*df error* in (2) above) for the CRD arise from the fact that there are generally n replicates for each treatment level and one degree of freedom is used for estimation of the treatment mean. Thus, for each of the t treatment levels there are n-1 degrees of freedom

and pooling we have

$$t*(n-1)$$ (3)

error degrees of freedom. For the CRD in Table 1 with three replicates and three treatment levels we have 3*(3-1) = 6 *df* for the experimental error.

Figure 1a provides a graphical representation of the experimental error upon noting the dispersion among the three prep method replicates, y_{i1}, y_{i2}, y_{i3}, within the ith prep method, averaged across the prep methods. One can gain intuition regarding the treatment effect by visualizing the dispersion among the $\bar{y}_{i.}$ (between groups variance) to the average dispersion among the replicates within each of the prep methods. Technically speaking, one compares the dispersion among the $\bar{y}_{i.}$, multiplied by the square root of the number of replicates within each treatment (and represented by the square root of the mean sum of squares for treatments), with the dispersion among the replicates within each prep (and represented by the square root of the mean sum of squares for error). In this example it is apparent that the between groups variance is only slightly greater than the experimental error variance and this is further verified by the non-significant prep method p-value (0.1038) found in Table 2.

Figure 1. Experimental error representation CRD (a). Experimental error representation for SPD[CRD,RCBD] (b).
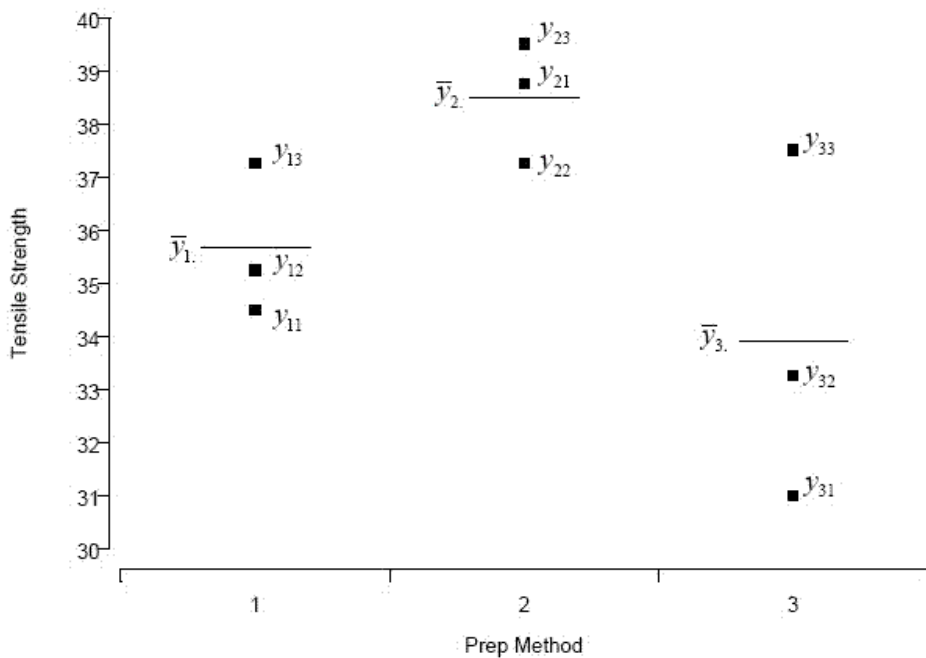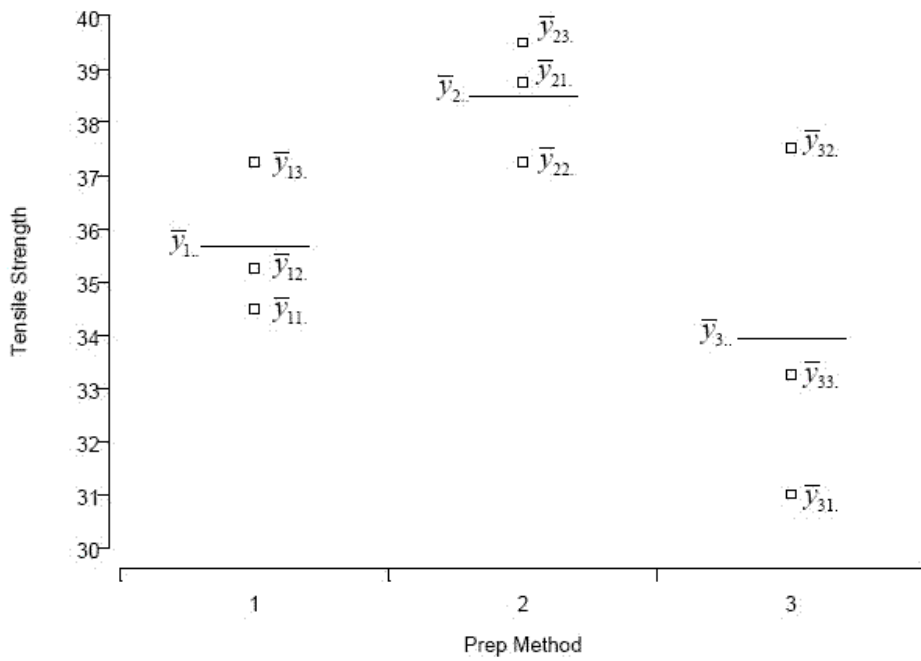
Figure 1a:

Figure 1b:

Table 2. Analysis of Variance for the CRD in Table 1.

Source	DF	SS	MS	F	Prob>F
Prep Method	2	32.097	16.048	3.384	0.1038
Error	6	28.458	4.743		
C.Total	8	60.555			

3. Randomized Complete Block Designs (RCBDs)

Suppose for the paper manufacturing example, only three batches can be produced in a given day and environmental conditions from day to day are thought to influence tensile strength. Instead of treating the design as a CRD, it is probably more efficient (lower experimental error variance) to utilize a RCBD where the day is the block. Table 3 provides the set-up of a RCBD for the modified paper manufacturing example. Note that randomization of treatment levels occurs independently within each day.

Table 3. RCBD for the 3-level Paper Manufacturing Example

		Day							
		1			2			3	
Prep Method	1	3	2	2	3	1	1	2	3
Tensile Strength	34.5	31	38.75	37.25	33.25	35.25	37.25	39.5	37.5

An appropriate model for the RCBD in Table 3 is

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, \quad i = 1,2,3; j = 1,2,3, \tag{4}$$

where y_{ij}, μ, τ_i are as defined in (1), β_j denotes the jth random day effect, and ε_{ij} denotes the experimental error. Note for the RCBD that every

prep method occurs in every day and replication of treatments occurs across days. If the day effect is ignored in the analysis then the

experimental error would include $\beta_j + \varepsilon_{ij}$. By including the day effect in the analysis, the day effect is in essence extracted from the experimental error.

The day by treatment interaction and the experimental error are confounded in the RCBD. The intuition behind this statement can be seen through our example where we note that during a single day, the three prep methods are randomized and the resulting tensile strengths are recorded. With just a single day, no replication exists and there would not be any way to test for the prep method effect since the experimental error is confounded with any observed difference in prep methods. However, when three days worth of experiments are performed in this manner, there will be replicates of the 3-level experiment, one replicate for each day. In order to make sure we account for the fact that different days may result in different tensile strength results, the correct experimental error term would be one that gets at the change in the observed differences between the 3-level prep variable from day to day. But this is exactly the definition of an interaction between prep method and days. Consequently, the day by prep method interaction and the experimental error are confounded. For this reason one must assume that any differences in prep method across the days is not the result of an actual interaction effect but instead the result of experimental error. Thus, the degrees of freedom for the expected error term for the RCBD are degrees of freedom for the day by prep method interaction [(3-1) × (3-1)]. In general, the degrees of freedom are

$$[(b - 1) \text{ x } (t - 1)] \tag{5}$$

where b is the number of blocks and t is the number of treatments.

Figure 2a provides a graphical representation of the experimental error for the RCBD in Table 3 under the assumption of no prep by day interaction. The comparison of prep method 2 to prep method 1 is denoted by Δ_1 for day 1, Δ_2 in day 2, and Δ_3 in day 3. A different set of deltas would exist upon comparing prep methods 1 vs. 3 and 2 vs. 3. The experimental error is obtained as follows. Look at the variation in the deltas for comparing prep methods 2 and 1. Also look at the variation in the deltas for comparing prep methods 3 and 1, and the variation in the deltas for comparing prep methods 3 and 2. These three sets of variations in the deltas are pooled for an estimate of the experimental error variance. Notice that the variation in the deltas corresponds to a lack of parallelism in the lines in Figure 2a. No variation corresponds to parallel lines and thus, negligible experimental error. Large variation results in lines that are not parallel and thus larger experimenteral error. In a two-factor analysis of variance, a lack of parallelism is an indication of the existence of an interaction between the factors. In the RCBD, experimental error and block by treatment interactions are confounded. Thus, variation in the deltas (used to estimate experimental error) and lack of parallelism (an indication of an interaction) provide the same information about experimental error, assuming no interactions actually exist. The experimental error is quantified by the square root of the mean squared error.

Figure 2a also provides the overall means for each of the prep methods and one can get a general idea of the treatment effect by the magnitude of the differences in the treatment means relative to the experimental error (deviation from parallel lines). In general, if the profiles are relatively parallel and widely separated then there is a significant treatment effect. The lack of parallelism is quantified by the mean squared error in Table 4 (2.267) while the separation among the prep methods is quantified by the mean square for prep method (16.048) in Table 4. Similar to what was observed in the CRD, when making the actual assessment of a treatment effect, the variation in the overall means for each prep method must be inflated by a factor of the square root of the number of replicates (here replicates are blocks) and is represented by the mean sum of squares for treatments. Here, the dispersion among the prep method means is substantially larger than the experimental error variance, thus suggesting a significant prep method effect, a fact evidenced by the small p-value (0.0485) for prep method in Table 4. Note the reduction in the error sum of squares from the CRD (SSE = 28.458 in Table 2) to the RCBD (SSE = 9.069 in Table 4). In summary, the experimental error in a CRD is represented by (the average of) the dispersions among the replicates within each treatment and the experimental error in a RCBD by the variation in treatment differences from block to block.

Table 4. Analysis of Variance Table for RCBD Analysis of Data in Table 3.

Source	DF	SS	MS	F	Prob>F
Prep Method	2	32.097	16.048	7.078	0.0485
Day	2	19.389	9.694		
Error	4	9.069	2.267		
C.Total	8	60.555			

Figure 2. Experimental error representation for RCBD (a). Whole-plot experimental error and whole-plot treatment effect representation for SPD[RCBD,RCBD] (b).
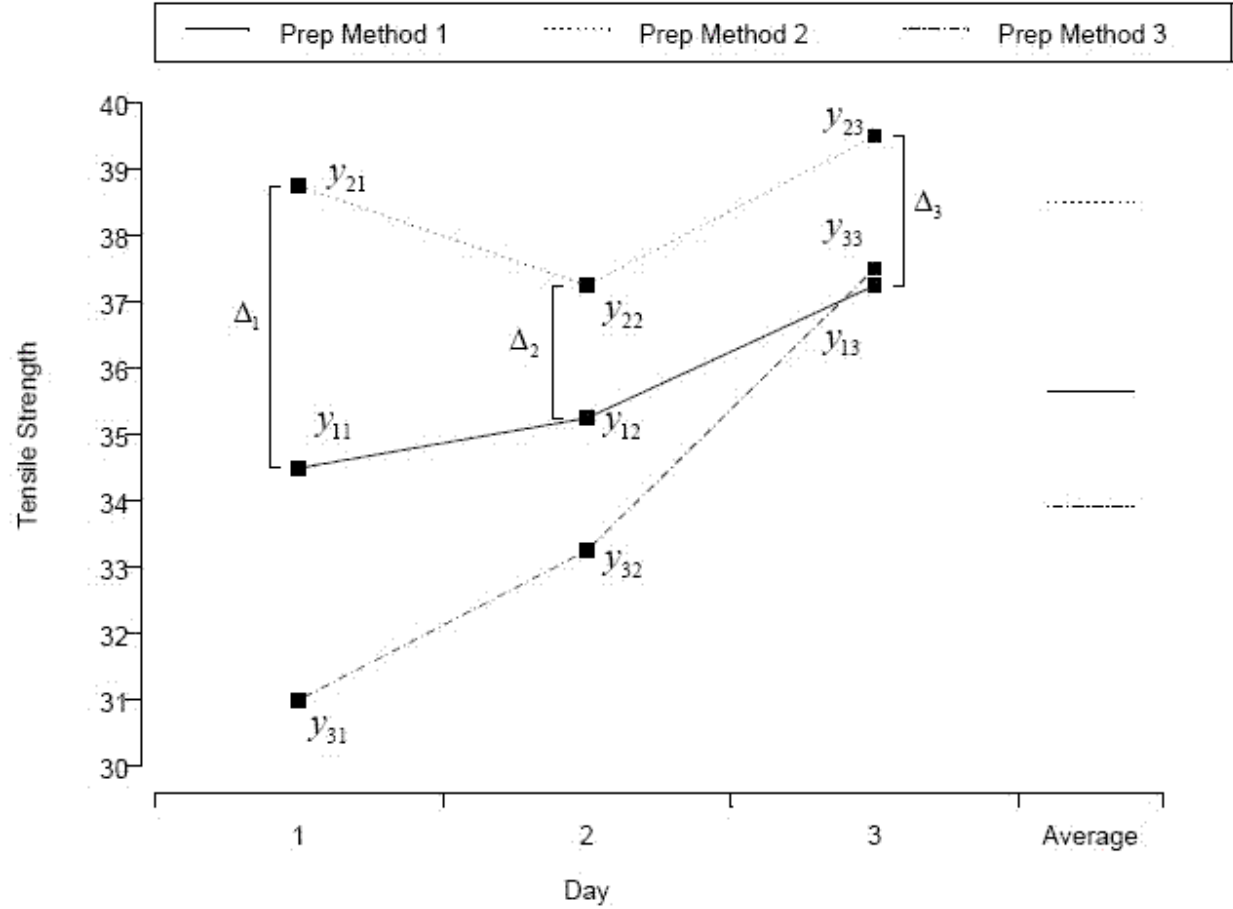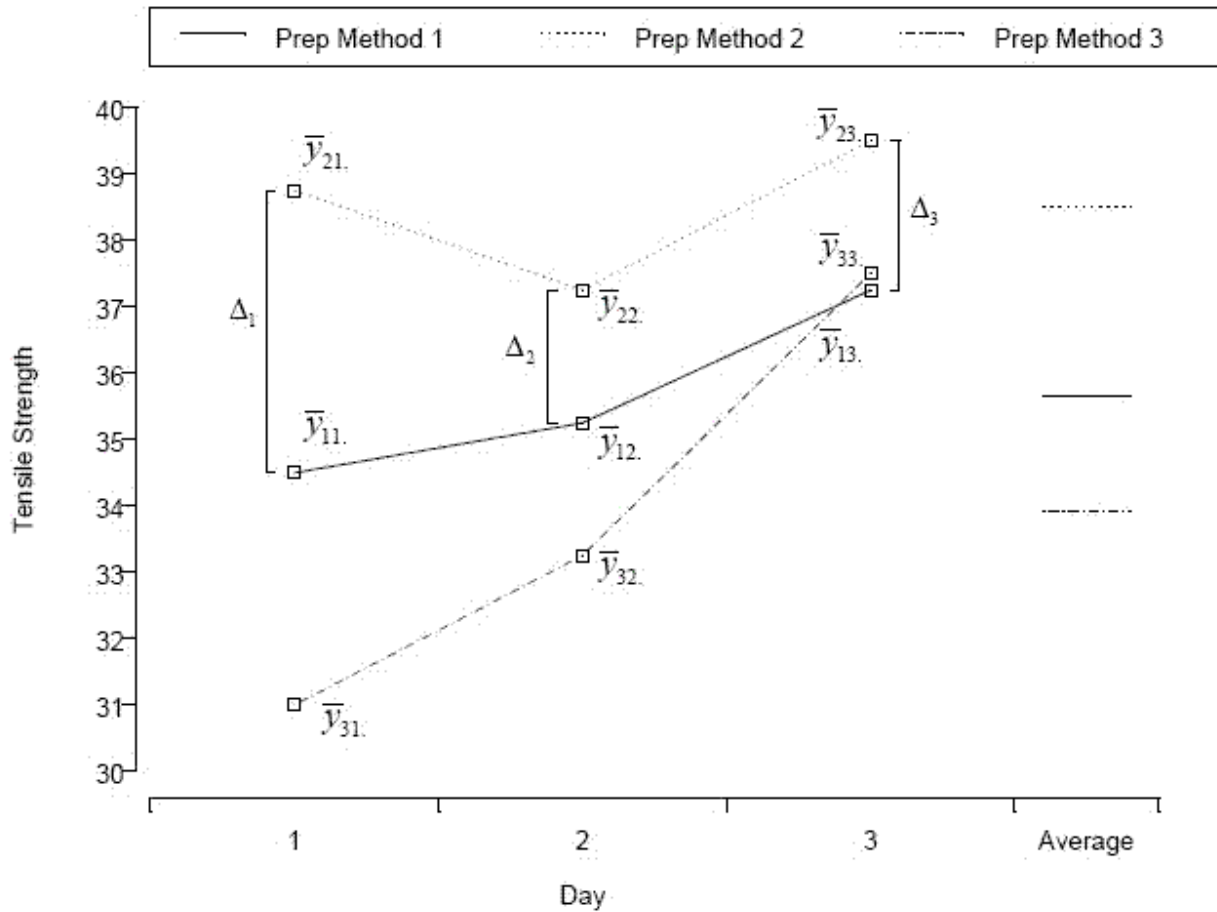
Figure 2a:

Figure 2b:

In the next section we demonstrate how the intuition of the experimental error in the CRD and RCBD can be extended to the split-plot design setting.

4. Split Plot Designs

Suppose there is interest in investigating the effect of a second factor, cooking temperature on tensile strength. In this experiment, once a batch is constructed with a particular prep method, the batch is split into sub-units for cooking. Here, the batches are the *whole-plot units* with prep method as the *whole-plot factor* and the sub-units are cooking portions with cooking temperature (henceforth referred to as temp) as the *sub-plot factor*. Prep method can be considered as the hard to change factor whereas temp is an easy to change factor since its levels are easily randomized once the batch is constructed with a given prep method. For the SPD there are two separate randomizations and thus two separate experimental errors, one for the whole-plot factor levels and another for the sub-plot factor levels. In this section, we will discuss a scenario in which the whole-plot factor levels are fully randomized and a second scenario in which the whole-plot factor levels are randomized within a block, resulting in a RCBD for the whole-plot factor. Note that the sub-plot randomization is always restricted in the sense that randomization takes place separately within each whole-plot, making each whole-plot a block for the sub-plot factor levels.

4.1 SPD's With Completely Randomized Whole-plot Levels, SPD[CRD,RCBD]

Let's assume that all nine batches of pulp can be made on the same day and that no blocking is necessary. In this case, the three prep methods would be randomized to the nine batches, much like a single variable experiment at three levels would be randomized with three replicates, i.e., a CRD. Table 5 presents a possible randomization structure at the whole-plot level.

Table 5: Randomization of the Whole-Plot Factor Prep Method Replicated Three Times.

	Replication								
	1	1	2	1	3	2	3	2	3
Prep Method	2	1	1	3	1	2	2	3	3

Next, for a given prep method, the four levels of temp are randomly applied to the batch sub-units (the sub-plot units). The second level of randomization and order in which the experimental runs would be performed is provided in Table 6.

Table 6: Randomization of the SPD[CRD,RCBD] in Prep Method and Temp

	Replication								
	1	1	2	1	3	2	3	2	3
Prep Method	2	1	1	3	1	2	2	3	3
Temp	275	200	275	200	275	275	200	200	225
	250	225	250	225	250	200	250	250	250
	225	275	225	250	200	225	225	275	200
	200	250	200	275	225	250	275	225	275

In this situation the whole-plot factor, prep method, at three levels with three replicates is randomized and then the sub-plot factor, temp, is randomized at the sub-plot level. Since the levels of prep method are randomly assigned at the batch level, the batch effect must be assessed by comparison to a batch experimental error term which reflects the natural spread across batches. Similarly, since the levels of temp are randomly assigned at the sub-unit level, the temp effect must be assessed by comparison to a sub-unit experimental error term reflecting natural dispersion across sub-units. Contrast this to a CRD setting involving prep method and temp in which one would need 12 batches (one for each combination of prep method and temp) for a single replicate and 36 batches for three replicates. A CRD would only have one experimental error term which would reflect batch dispersion. The SPD offers cost efficiency for the hard to change factor prep method as three replicates of the SPD would only require nine changes of prep method versus the 36 required for the CRD.

An appropriate model for the SPD[CRD,RCBD] described above is

$$y_{ijk} = \mu + \tau_i + \delta_{j(i)} + \gamma_k + (\tau\gamma)_{ik} + \varepsilon_{ijk}, \quad i = 1,2,3; j = 1,2,3; k = 1,2,3,4, \tag{6}$$

where y_{ijk} is the response on the jth day for prep method i at temp k. The parameter μ is the overall mean, τ_i is the fixed effect due to the ith whole-plot treatment (prep method), $\gamma_{j(i)}$ is the whole-plot error, γ_k is the fixed effect due to the kth sub-plot treatment (temp), $(\tau\gamma)_{ik}$ is the whole-plot by sub-plot interaction and ε_{ijk} is the sub-plot error. It is typically assumed that the $\delta_{j(i)}$ are *i.i.d.* $N(0,\sigma^2_\delta)$ with σ^2_δ denoting the experimental error variance of the whole-plot units. The ε_{ijk} are assumed *i.i.d.* $N(0,\sigma^2_\varepsilon)$ with σ^2_ε denoting the experimental error variance of the sub-plot units. Finally, it is assumed that the $\delta_{j(i)}$ and ε_{ijk} are independent of one another.

In providing intuition for the two experimental error components of the SPD[CRD,RCBD], we first begin with CRD at the whole-plot level. For the whole-plot experiment, we replicated the 3-level prep methods three times and completely randomized the run order. In other words, we took the 9 runs of the prep method (1,1,1,2,2,2,3,3,3) and fully randomized them. Recall from our discussion in Section 2, the experimental error variance for a CRD is determined by the differences associated with the replicates nested within each prep method. For the SPD, the contribution of the ith whole-plot level for the jth replicate is summarized by taking the mean response across the sub-plot levels, $\bar{y}_{ij\cdot}$. In our example, $\bar{y}_{ij\cdot}$ is

the average response on the jth day for the ith prep method averaged across the observed cooking temps $\left(\text{i.e. } \bar{y}_{ij.} = \sum_{k=1}^{4} y_{ijk}/4 \right)$.

Since the whole-plot design is a CRD, the whole plot experimental error variance will be determined by the differences associated with the whole plot replicates nested within the whole plot treatment. Specifically, one would look at the three prep method i replicates $(\bar{y}_{i1.}, \bar{y}_{i2.}, \bar{y}_{i3.})$ and see how these average tensile strength values differ from their group mean $(\bar{y}_{i..})$. These estimated squared differences are then pooled together to get one estimate of the whole plot experimental error variance,

$$\hat{\sigma}_{\varepsilon}^2 = \frac{\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\bar{y}_{ij.} - \bar{y}_{i..}\right)^2}{df\ error}, \tag{7}$$

where $\hat{\sigma}_{\varepsilon}^2$ denotes the estimate of whole-plot error variance. Note the direct parallel between the expression in (7) with that given in (2).

The degrees of freedom associated with the whole-plot error are calculated just as they were in (3) from Section 2. Due to the presence of both whole-plots and sub-plots now, we will modify the notation and use

$$t_w(n_w - 1) \tag{8}$$

to denote the degrees of freedom error. Here, t_w denotes the number of whole-plot treatment combinations and n_w denotes the number of whole-plot replicates nested within each whole-plot treatment combination.

The whole plot experimental error variance is easily visualized in Figure 1b by noting the dispersion among the three prep method replicates, $\bar{y}_{i1.}, \bar{y}_{i2.}, \bar{y}_{i3.}$, within the ith prep method. Figure 1b is equivalent to Figure 1a but with the y_{ij} (from Figure 1a) $= \bar{y}_{ij.}$ (from Figure 1b). Therefore, the interpretation of the whole-plot treatment effects is analogous to the discussion of the treatment effect of the CRD in Section 2. More specifically, one can get a general idea of the whole-plot treatment effects by comparing the dispersion in the treatment means (here, the dispersion among $\bar{y}_{1..}, \bar{y}_{2..},$ and $\bar{y}_{3..}$) with respect to the whole plot experimental error variance. Note that the F-statistic for prep method in Table 7 is identical to that in the CRD analysis found in Table 2. The equivalent results are due to the fact that when the data are balanced, taking the mean across the levels of the sub-plot factor within a whole-plot and then performing a CRD analysis of the means is analogous to the split-plot analysis for the whole-plot factor. Note also that the whole-plot treatment and whole-plot error sums of squares are four times that of the CRD, due to the four sub-plots within each whole-plot, therefore the F-ratio for the whole-plot treatment is unaffected.

Table 7: Analysis of variance for SPD[CRD,RCBD]

Source	DF	SS	MS	F	Prob>F
Prep Method	2	128.39	64.19	3.38	0.1038
Reps (Prep Method) *Whole-plot Error*	6	113.83	18.97	.	.
Temp	3	434.08	144.69	36.43	< 0.0001
Prep Method x Temp	6	75.17	12.53	3.15	0.0271
Sub-plot Error	18	71.50	3.97	.	.

4.2 Whole-plot Experimental Error Variance for the SPD[RCBD,RCBD]

When there is a blocking factor, the whole-plot factor levels are randomized within the blocks. Consider again the scenario from Section 3 where it is only possible to make three batches of pulp in a given day and environmental conditions from day to day are thought to influence tensile strength. Here, we have two blocking factors: one at the whole-plot level (prep methods randomly assigned within a day) and the other at the sub-

plot level (sub-plot levels randomly assigned within a whole-plot level). As in Section 4.1, the sub-plot factor is temp and the randomization of the levels of temp takes place within each whole-plot, making the whole-plots blocks for the sub-plot factor. The randomization and run order for both the whole-plots and sub-plots is provided in Table 8. Contrast the randomization for the SPD to a RCBD with two factors. In the RCBD, each day would require 12 batches, one for each of the 12 combinations of prep method and temp. A single randomization would take place, namely the order in which the 12 batches are run within a day. This design would not be feasible here since it was stated that only three batches can be run on a given day. The SPD overcomes the necessity for so many batches to be run in a given day by incorporating two levels of randomization.

First the order of the three prep methods (whole-plot levels) would be randomized for a given day, and then, separately, the levels of temp (sub-plot levels) are randomized to the cooking portions within each batch. Thus, for a given day the SPD would require only three batches of material to be produced.

Table 8: Randomization of the SPD[RCBD,RCBD] in Prep Method and Temp

		Day 1			Day 2			Day 3	
Prep Method	2	3	1	1	3	2	3	1	2
	275 (42)	200 (29)	275 (36)	200 (28)	275 (40)	275 (40)	200 (32)	200 (31)	225 (40)
Temp	250 (38)	225 (26)	250 (37)	225 (32)	250 (32)	200 (31)	250 (39)	250 (41)	250 (39)
	225 (41)	275 (36)	225 (35)	250 (40)	200 (31)	225 (36)	225 (34)	275 (40)	200 (35)
	200 (34)	250 (33)	200 (30)	275 (41)	225 (30)	250 (42)	275 (45)	225 (37)	275 (44)

An appropriate model for the SPD described above is

$$y_{ijk} = \mu + \tau_i + \beta_j + \delta_{ij} + \gamma_k + (\tau\gamma)_{ik} + \varepsilon_{ijk}, \quad i = 1,2,3; j = 1,2,3; k = 1,2,3,4, \tag{9}$$

where μ, τ_i, β_j, γ_k, and $(\tau\gamma)_{ik}$ are as defined in (4) and (6), δ_{ij} denotes the whole-plot error, and ε_{ijk} is the sub-plot error. The same distributional assumptions made with the SPD[CRD,RCBD] for the error terms are made here. Similar to the discussion in Section 4.1, in considering the design at the whole-plot level, it is helpful to view the responses as the $\bar{y}_{ij\cdot}$'s, where $\bar{y}_{ij\cdot} = \sum_{k=1}^{4} y_{ijk}/4$ is the average strength of the four cooking portions for prep method i on day j. Since the whole-plot treatments (prep methods) are randomized according to a RCBD, the block (day) by treatment interaction and the whole-plot experimental error are confounded (see the discussion in Section 3). For this reason one must assume that any differences in prep method across the days are not the result of an actual interaction effect but instead the result of whole plot experimental error. Thus, the degrees of freedom for the whole plot error term are the degrees of freedom for the day by prep method interaction [(3-1) × (3-1)]. In general, the whole-plot error df are given by

$$(t_w - 1)*(n_w - 1) \tag{10}$$

where t_w is the number of whole-plot levels and n_w is the number of whole-plot blocks. Note that the degrees of freedom for the denominator of the prep method F-statistic in the SPD[RCBD,RCBD] has four degrees of freedom instead of six in the SPD[CRD, RCBD] case.

Figure 2b provides a graphical representation of the experimental error for the whole-plot factor and is identical to Figure 2a but with $y_{ij} = \bar{y}_{ij\cdot}$. Here the magnitude of the whole-plot experimental error variance is reflected in the degree of differences among the deltas (lack of parallelism in Figure 2b) for all possible treatment comparisons. The whole-plot experimental error variance is estimated by averaging the variation in the deltas when comparing prep methods 2 and 1, the variation in the deltas when comparing prep methods 3 and 1, and the variation in the deltas

when comparing prep methods 3 and 2. Figure 2b also provides the overall mean for each of the prep methods and one can get a general idea of the whole-plot treatment effects by the magnitude of the differences in the means with respect to the experimental error (i.e. , deviation from parallel lines). As with the RCBD, when making the actual assessment of a treatment effect, the variation in the overall means for each prep method must be inflated by a factor of the square root of the number of replicates (here replicates are blocks) and is represented by the mean sum of squares for prep method (128.39) in Table 9. If the profiles in the block by whole-plot treatment interaction plot, (Figure 2b for our example), are relatively parallel and widely separated then there is a significant whole-plot effect. Thus, the visualization of a significant whole plot effect via Figure 2b is identical to the visualization of a treatment effect in the RCBD using Figure 2a. This is further evidenced by the fact that the p-value for prep method in Table 9 is precisely the same as that given in Table 4. Unlike the standard RCBD where only one type of experimental unit exists, the presence of whole-plot and sub-plot units in the SPD implies that one needs to be careful in the interpretation of the whole-plot effect in the case of a possible interaction between the whole-plot and sub-plot effects. More will be discussed regarding whole plot treatment interactions with subplot interactions in Section 4.3.

Table 9: Analysis of variance for the SPD[RCBD,RCBD] case

Source	DF	SS	MS	F	Prob>F
Day	2	77.56	38.78	.	.
Prep Method	2	128.39	64.19	7.08	0.0485
Day x Prep Method _Whole-plot Error_	4	36.28	9.07	.	.
Temp	3	434.08	144.69	36.43	< 0.0001
Prep Method x Temp	6	75.17	12.53	3.15	0.0271
Sub-plot Error	18	71.50	3.97	.	.

4.3 Sub-plot Experimental Error Variance

 As mentioned earlier, although there are different randomization schemes possible for the whole-plot factor levels, any randomization scheme for the sub-plot factors will be restricted since sub-plot factor levels are always randomized within whole-plots. To conceptualize the sub-plot experimental design, it is helpful to focus upon a single level of the whole-plot factor. In our example, imagine formulating three batches (i.e. three whole-plot replicates) of pulp using a single prep method and then splitting each of these batches into four equal cooking portions (i.e. 12 total cooking portions). For each batch separately, the levels of temp are randomly assigned to the four cooking portions. Table 10 presents an example of this randomization structure. Note that the sub-plot design is simply an RCBD where the blocks are the replicates of the specific whole-plot level (prep method). Thus, to understand the sub-plot error term, all one needs to do is to identify the variable(s) in the data set which uniquely define(s) the whole-plot replicates (batches). In the SPD[CRD,RCBD] case, batches of pulp uniquely define the whole-plot replicate variable while in the SPD[RCBD,RCBD] case, the day variable uniquely defines the replicates. For both types of SPDs, the sub-plot experimental error variance within a given prep method is estimated via the whole-plot replicate variable by sub-plot variable (temp) interaction. The error degrees of freedom would be given by

$$(n_w - 1)*(t_s - 1) \tag{11}$$

where n_w is the number of whole-plot replicates within a given whole-plot level and t_s is the number of sub-plot treatment levels.

Table 10: Sub-Plot Structure for one Prep Method

Batch Number Nested within Prep Method											
(Blocking variable at the Sub-Plot Level)											
1				2				3			
Temp 200	225	250	275	200	225	250	275	200	225	250	275

The sub-plot error structure for prep method 1 is visualized in Figure 3a where the whole-plot replicate variable by temp interaction is plotted for prep method 1. The individual points in Figure 3a are the tensile strengths for prep method 1 across each of the j whole plot replicates and k cooking temperatures (i.e. the $y_{1_{jk}}$'s). Recall for RCBDs the block by treatment interaction is confounded with experimental error and any difference in treatment effects observed across blocks is assume to be experimental error variance. Note that Δ_{11}, Δ_{12} and Δ_{13} represent the observed tensile strength differences at a temp of 225 and a temp of 200 across the three whole plot replicates for prep method 1. A different set of deltas would be observed for each of the other temp level pairwise comparisons. If the Δ's differ from one replicate to the next, i.e. , lack of parallelism, this suggests a replicate (block) by temp interaction. Since the design structure is an RCBD, the differences in the Δ's represent the sub-plot error variance. This is identical to the discussions and illustrations for the RCBD in Sections 3 and 4.2 regarding Figure 2a and Figure 2b.

Since there are a total of three prep methods, the overall estimate of sub-plot error variance would be one in which the sub-plot error variances are pooled across all of the levels of the whole-plot variable (prep method). One would have to look at all three plots (Figures 3a, Figure 3b and Figure 3c) to get a sense for the overall sub-plot error variance. The magnitude of the sub-plot error variance would be reflected by the overall lack of parallelism across Figures 3a, 3b and 3c. The overall degrees of freedom for the sub-plot experimental error would then be

$$t_w(n_w - 1)*(t_s - 1) \tag{12}$$

where the expression in (12) is simply that of (11) multiplied by the number of whole-plot levels t_w. Note that the expression for the sub-plot error degrees of freedom in (12) does not depend on the type of design at the whole-plot level since the whole-plot replicates (whether true replicates or replicates across blocks) form the blocks for the sub-plot design. This fact is illustrated in Table 7 [SPD(CRD,RCBD)] and Table 9 [SPD(RCBD,RCBD)] where we use the interaction effect of replication (day) by temp [(3-1)*(4-1)] nested within the three prep methods to estimate the sub-plot error for a total of (3-1)*(4-1)*3 = 18 degrees of freedom.

To visualize the sub-plot effect Figures 3a, 3b and 3c provide the group means for each of the levels of temp. Let us first focus on Figure 3a where one can get a general idea of the sub-plot treatment effect by the magnitude of the differences in the overall sub-plot (temp) means relative to the lack of parallelism. In observing the differences among the four temp means versus the mild lack of parallelism, one would anticipate a possible temp effect for prep method 1. A similar evaluation would be done for prep method 2 and prep method 3 by looking at Figures 3b and 3c. Overall, if the profiles are relatively parallel and widely separated for each of the whole-plot levels (prep method) then that would indicate a potentially significant sub-plot effect.

At this point, it is important to remember that any observed sub-plot effect should not be interpreted until one has evaluated whether or not there is a significant interaction between the whole-plot and sub-plot effects. Observing Figures 3a, 3b and 3c one can also assess a potential whole-plot by sub-plot interaction. For example, in prep method 1 (Figure 3a), the means for 275 and 250 are much closer to each other than they are in prep method 3 (Figure 3c). This indicates a possible whole-plot by sub-plot interaction. Note for this example, the whole-plot by sub-plot interaction is indeed significant (p-value = 0.0271 in Tables 7 and 9. The sub-plot error variance is used to assess the whole-plot by sub-plot interaction.

Figure 3. Replication(Block) by Temp interaction for Prep Method 1 (3a). Replication(Block) by Temp interaction for Prep Method 2 (3b). Replication(Block) by Temp interaction for Prep Method 3 (3c).
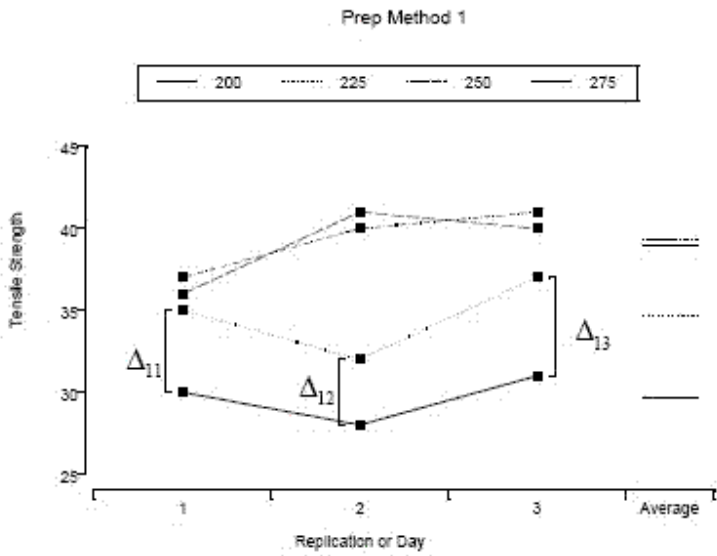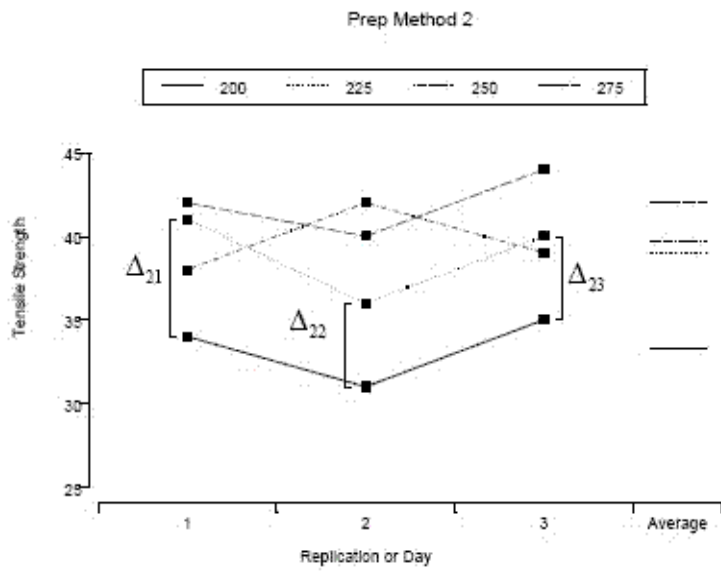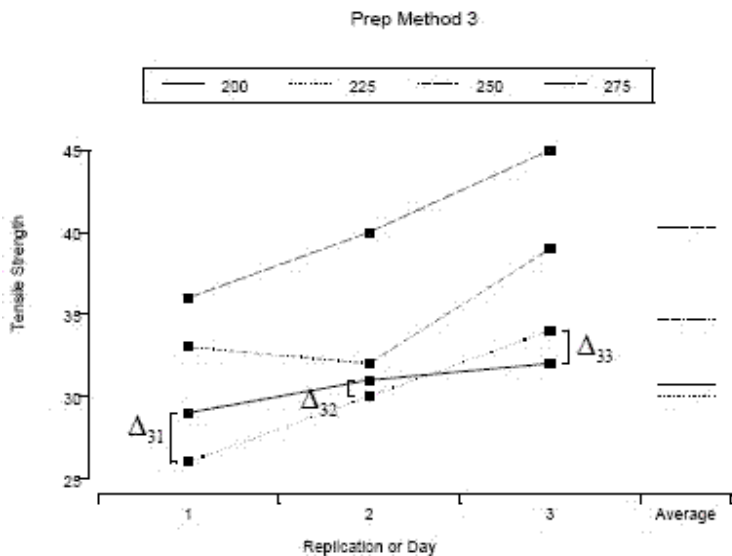
Figure 3a:

Figure 3b:

Figure 3c

5. Conclusions

Providing the intuition behind the analysis of SPDs is not an easy task. In this paper we show that the whole-plot and sub-plot error structure can be broken down into easy to understand CRD or RCBD designs. The whole-plot error is estimated by the effect of the replication variable nested within the whole-plot factor for a CRD at the whole-plot level while the whole-plot error is estimated by the block by whole-plot factor interaction effect for a RCBD at the whole-plot level. We also showed that at the sub-plot level, the error is estimated by pooling the replicate (block) by sub-plot factor interaction effects over the whole-plot levels. All of these concepts were illustrated in an intuitive graphical approach, thus allowing students to "see" the error structure and gain intuition of the statistical analysis by associating each source of variation in the SPD ANOVA table with a corresponding plot.

References

Bingham, D. and Sitter, R. S. (2001). "Design Issues for Fractional Factorial Experiments," *Journal of Quality Technology*, 33, 2-15.

Box, G.E.P. and Jones, S. (1992). "Split-Plot Designs for Robust Product Experimentation," *Journal of Applied Statistics*, 19, 3-26.

Christensen, R. (1996). *Analysis of Variance, Design and Regression,* New York: Chapman & Hall.

Federer, W.T. and King, F. (2007). *Variations on Split Plot and Split Block Experiment Designs*, New Jersey: Wiley.

Hinkelmann, K.H. and Kempthorne, O. (1994). *Design and Analysis of Experiments, Vol. 1*, New York: John Wiley & Sons.

Huang, P., Chen, D., and Voelkel, J. O. (1998). "Minimum-Aberration Two-Level Split-Plot Designs," *Technometrics*, 40, 314-326.

Kowalski, S.M., Parker, P.A. and Vining, G.G. (2007). "Tutorial: Industrial Split-plot Experiments," *Quality Engineering*, 19, 1-16.

Letsinger, J. D., Myers, R. H., and Lentner, M. (1996). "Response Surface Methods for Bi-Randomization Structure," *Journal of Quality Technology*, 28, 381-397.

Montgomery, D.C. (2001). *Design and Analysis of Experiments, 5th edition*, New York: John Wiley & Sons.

Rao, P. V. (1998). *Statistical Research Methods in Life Sciences*, New York: Duxbury Press.

Smith, C. and Johnson, D. (2007). "Comparing analyses of unbalanced split-plot Experiments," *Journal of Statistical Computation and Simulation*, 77, 119-129.

Webb, D., Lucas, J. M. and Borkowski, J. J. (2004). "Factorial Experiments when Factor Levels Are Not Necessarily Reset," *Journal of Quality*

Technology, 36, 1-11.

Yates, F. (1935). "Complex Experiments," *Journal of the Royal Statistical Society,* Supplement 2, 181-247.

Yates, F. (1937). "The Design and Analysis of Factorial Experiments," *Commonwealth Bureau of Soil Science*, Tech. Comm., No. 35.

Timothy J. Robinson
Associate Professor of Statistics
University of Wyoming
Laramie,WY 82071
tjrobin@uwyo.edu

William A. Brenneman
Principle Statistician
Department of Statistics
The Procter & Gamble Company
Brenneman.wa.@pg.com

William R. Myers
Section Head
Department of Statistics
The Procter & Gamble
Myers.wr@pg.com