DESIGN OF EXPERIMENTS

How To Analyze A Split-Plot Experiment

by **Kevin J. Potcner and Scott M. Kowalski**

Many quality improvement projects require some form of experimentation on a process. A chemical engineer may wish to determine the settings for certain process variables to optimize a critical quality characteristic of the resulting product. A materials engineer may run a plastic injection molding process using different grades of raw material to determine which produces the least variability in breaking strength.

The deliberate changing of input process variables with the intention of studying their effect on output variables is referred to as a designed experiment. Typically, statisticians identify a designed experiment by describing two primary components:

1. One component, referred to as the treatment structure, details the different factors (input variables) the experiment will incorporate and the different settings (levels) for those factors. For example, a 2^5 full factorial treatment structure means five factors will be used in the experiment, each studied at two levels, and all 2 x 2 x 2 x 2 x 2 = 32 treatment combinations are to be run.
2. The other component is referred to as the experimental or design structure of the experiment. This component illustrates how the experimental runs are to be carried out—for example, defining the experimental and observational units, selecting the experimental units and assigning them to the treatment combinations, choosing the randomization scheme and

In 50 Words Or Less

- **Experiments with simple design structures, such as complete randomization, are often not realistic in the real world.**
- **Typically an experiment will have some form of randomization restriction, and the split-plot method is a solution.**
- **The analysis of a split-plot experiment involves two error variances.**

deciding how the treatment combinations will be changed throughout the experiment.

In a previous article in *Quality Progress*, we illustrated the features of the split-plot design, how common the features are in industrial experimentation and how the practitioner can recognize this situation.[1] We will now illustrate the proper analysis of this particular type of design structure.

In many real experimental situations, a restriction is typically placed on the randomization of the runs.

Example of a Split-Plot Design

Consider an experiment involving the water resistant property of wood. Two types of wood pretreatment (one and two) and four types of stain (one, two, three and four) have been selected as variables of interest. A graphical representation of this type of treatment design is shown in Figure 1.

Conducting this experiment in a completely randomized fashion would require eight wood panels for each full replicate of the design. Each wood panel would be randomly assigned a particular pretreatment and stain combination. But it turns out to be very difficult to apply the pretreatment to a small wood panel.

The easiest way would be to apply each of the pretreatment types (one and two) to an entire board, then cut each board into four smaller pieces and apply the four stain types to the smaller pieces. This is shown in Figure 2.

So how exactly will the experiment be conducted? For example, how many boards will be used for each treatment combination? How many replicates of each treatment combination will be run? In what order will the experimental runs be conducted? How many measurements will be made on each small piece? These decisions should be based on both statistical and practical considerations.

Suppose the experimenter has decided to run three replicates of the pretreatment factor. This results in six boards and 3 x 2 x 4 = 24 total observations. To produce an experimental run for this process, you must first pretreat a board. After one of the randomly selected pretreatments has been applied, the board is

FIGURE 1 Factors That Affect Wood's Water Resistance

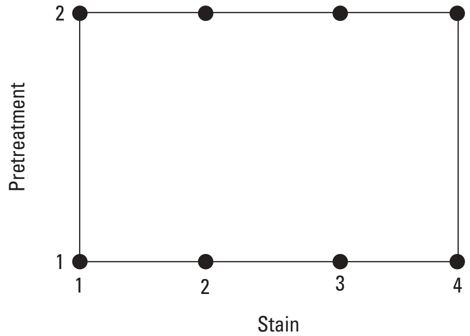
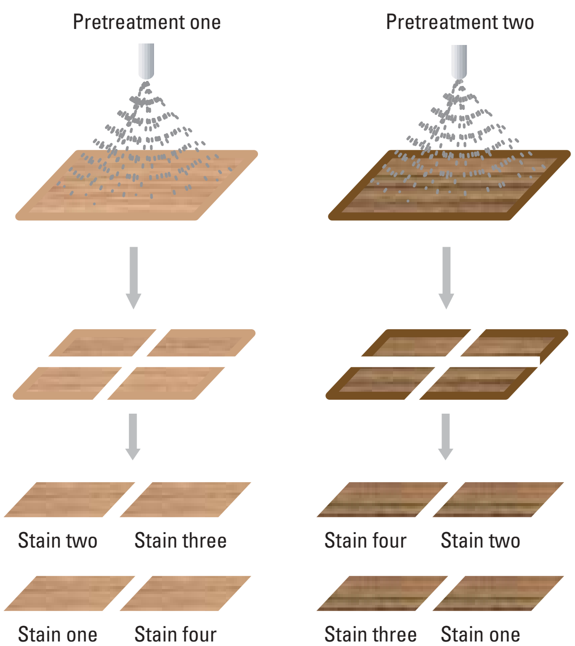
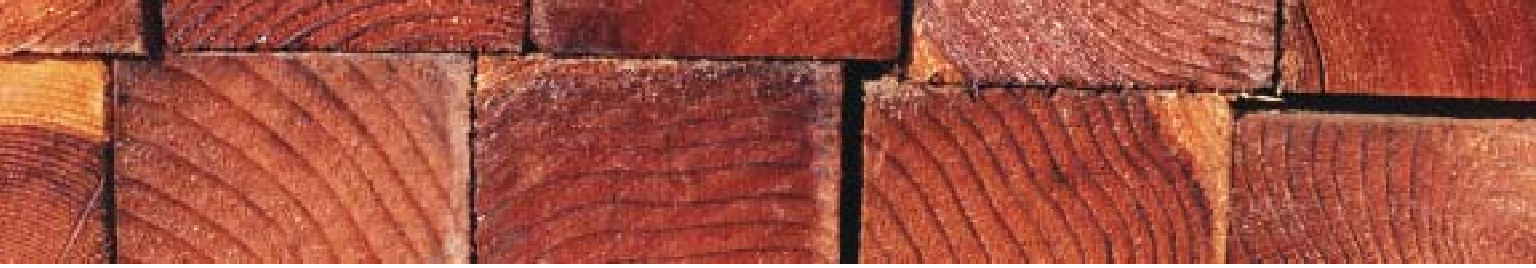


FIGURE 2 Treatment Application





cut into four pieces and then stained using one of the four stains selected at random.

The reader should recognize this is a split-plot design for four reasons:

1. For the pretreatment factor, an experimental unit is the entire board or a set of four pieces of the board after they are cut. For the stain factor, an experimental unit is an individual piece of the board. Having unequal sized experimental units for the different factors is one key element of a split-plot design.
2. Each factor uses a different randomization scheme. In contrast, a complete randomized design would use one randomization scheme for all 24 experimental runs.
3. Note for a single run at one level of pretreatment, four separate runs are conducted for the stains. As a result, pretreatment could be thought of as a hard-to-change factor, while stain could be considered an easy to change factor.
4. The number of experimental replicates is not the same for each factor. Pretreatment has only three experimental replicates for each of the two factor levels, while stain has six experimental replicates for the stain factor levels.

Because of these features, we would say the experimenter has run a 2 x 4 full factorial treatment structure within a split-plot design structure. Each of the six whole-plots (entire boards) has four subplots (smaller pieces of board), resulting in three replicates at the whole-plot level and six replicates at the subplot level.

TABLE 1 Data for Wood Example

Pretreat	Stain	WP error	Resistance
2	2	4	53.5
2	4	4	32.5
2	1	4	46.6
2	3	4	35.4
2	4	5	44.6
2	1	5	52.2
2	3	5	45.9
2	2	5	48.3
1	3	1	40.8
1	1	1	43.0
1	2	1	51.8
1	4	1	45.5
1	2	2	60.9
1	4	2	55.3
1	3	2	51.1
1	1	2	57.4
2	1	6	32.1
2	4	6	30.1
2	2	6	34.4
2	3	6	32.2
1	1	3	52.8
1	3	3	51.7
1	4	3	55.3
1	2	3	59.2

WP = whole plot

TABLE 2 Whole-Plot Analysis Using The Averages of Resistance In Each Whole Plot

Analysis of variance for the average resistance

Source	DF	SS	MS	F	P
Pretreat	1	195.51	195.51	4.03	0.115
Error	4	193.84	48.46		
Total	5	389.35			

DF = degrees of freedom
SS = sums of squares
MS = mean square
F = F-statistic
P = p-value

How To Analyze the Experiment

The simplest experiment from a statistical analysis perspective is what's called a completely randomized design structure. This, however, would require all 8 x 3 = 24 experimental runs to be conducted in a completely random order. For the experiment to be run in this way, each of the 24 runs would need to be a "true" experimental run. This would include a complete preparation and setup of the experimental materials and equipment.

As you can imagine, this experimental approach is not always efficient, practical or at times even possible to run. In many real experimental situations, a restriction is typically placed on the randomization of the runs. Such restriction, however, affects the statistical analysis.

TABLE 3 Incorrect Completely Randomized Design Analysis For Water Resistance of Wood

Source	DF	SS	MS	F	P
Pretreat	1	782.04	782.04	13.49	0.002
Stain	3	266.00	88.67	1.53	0.245
Pretreat x Stain	3	62.79	20.93	0.36	0.782
Error	16	927.88	57.99		
Total	23	2038.72			

DF = degrees of freedom
SS = sums of squares
MS = mean square
F = F-statistic
P = p-value

TABLE 4 Correct Split-Plot Analysis For Water Resistance of Wood

Analysis of variance for resistance, using adjusted SS for tests

Source	DF	SS	MS	F	P
Pretreat	1	782.04	782.04	4.03	0.115
WP (pretreat)	4	775.36	193.84	15.25	*
Stain	3	266.01	88.67	6.98	0.006
Pretreat x stain	3	62.79	20.93	1.65	0.231
Error	12	152.52	12.71		
Total	23	2038.72			

WP = whole-plot errors
DF = degrees of freedom
SS = sums of squares
MS = mean square
F = F-statistic
P = p-value

An example illustrates the correct analysis of split-plot experiments. Consider the previously described experiment involving the water resistant property of wood. Two types of wood pretreatment (one and two) and four types of stain (one, two, three and four) have been selected as variables of interest.

A graphical representation of the experiment is shown in Figure 2 on p. 68 (for each pretreatment the stains have been randomly assigned to the four panels). Table 1 (p. 69) gives the design as it was carried out: First a randomly selected pretreatment is applied, then the wood is cut into four panels and the stains are applied in random order.

The null hypothesis for all factors is H_0: There is no effect due to the factor. A test statistic is necessary to test this hypothesis. In this paper, the test statistics are all F-statistics, which are the ratio of the mean square (MS) for the factor of interest to the correct mean square error

$$F = \frac{MS_{Factor}}{MS_{CorrectError}}$$

Once the F-statistic has been calculated, a p-value can be computed and used to test the null hypothesis (we typically reject H_0 if the p-value < 0.05). The p-value is the probability the test statistic will take on a value at least as extreme as the observed value of the statistic, assuming the null hypothesis is true.

It is sometimes easier to think of the analysis of a split-plot experiment as two separate experiments corresponding to the two levels of the split-plot experiment: the whole-plot (WP) level and the sub-plot level.

Whole-Plot Level Only

Again, suppose the experiment is carried out using three replicates of the pretreatment factor. This involves six boards (three for pretreatment number one and three for pretreatment number two). For now, let's focus on only these six boards (before they are cut and the stains are applied) and break down the degrees of freedom (df).

Because these six boards are randomly assigned a pretreatment level, this part of the experiment is essentially a completely randomized design with one 2-level factor (pretreatment) and three replicates. Therefore, there is 6 – 1 = 5 total df for this whole-plot level of the experiment.

Because the only factor has two levels, pretreatment has 1 df. This leaves 4 df for the error term at the whole-plot level. Notice how thinking of the experiment in this manner clearly shows the pretreatment variable has its own error term, "whole-plot error." The split-plot design simply exploits the fact that each of the six pretreated boards can be cut into four pieces and another factor (stain) can also be studied.

Once all the data are collected, we could write the model as:

Average response =
pretreatment factor + WP error

in which average response is the mean of the four different stain responses in each whole plot, and WP error is the error term for the whole-plot factor (pretreatment).

The whole-plot experimental error is estimated by examining the variability that occurs between the three whole plots within each of the two pretreatment settings. Using these six averages will yield the correct F-test for pretreatment (pretreatment is not significant with p = 0.115, as shown in Table 2, p. 69). However, the sums of squares will not be the same as the correct overall split-plot analysis (they will be off by a factor of 4 = the number of subplots in each whole-plot).

Incorrect Completely Randomized Design Analysis

If the 24 pieces involving the four stains are incorrectly viewed as their own completely randomized experiment, then there would be 24 – 1 = 23 total df. This would involve 2 - 1= 1 df for pretreatment, 4 – 1 = 3 df for stain and (2 - 1)(4 - 1) = 3 df for the pretreatment by stain interaction. Therefore, there would be 23 – 7 = 16 df for error. The incorrect completely randomized model is:

Response = pretreatment + stain + pretreatment x stain interaction + error.

Notice, however, this analysis is incorrect because it does not remove the sums of squares and 4 df for whole-plot error discussed above (this is viewing the experiment only at the subplot level). The error term in this model is the sum of the whole-plot error and the subplot error.

When the whole-plot error is not removed from the completely randomized analysis, the error term used for testing the subplot factors is inflated. Therefore, the F-test for all terms in the model would use the wrong error term. This can result in F-tests that are insignificant for some subplot factors while overstating significance for the whole-plot factor.

Table 3 shows the analysis. Notice the pretreatment factor is incorrectly identified as significant (p = 0.002), while the stain factor is insignificant (p = 0.245). We have seen in the earlier whole-plot analysis that pretreatment is not significant, and we will see later in the correct split-plot analysis that stain is significant.

Correct Split-Plot Analysis

The split-plot model is:

Response = pretreatment + WP error + stain + pretreatment x stain interaction + SP error

in which SP error is the error for the subplot factor (stain) and the whole-plot by subplot interaction (pretreatment x stain). To get the correct analysis of variance table with all sources of variation including the two error terms involves removing the sums of squares and df for the whole-plot error from the reported error term in the incorrect completely randomized analysis. This can be done manually, but then all F-tests and p-values will have to be generated manually as well. Fortunately, many software packages can be tricked to do this for you automatically by using a nested model:

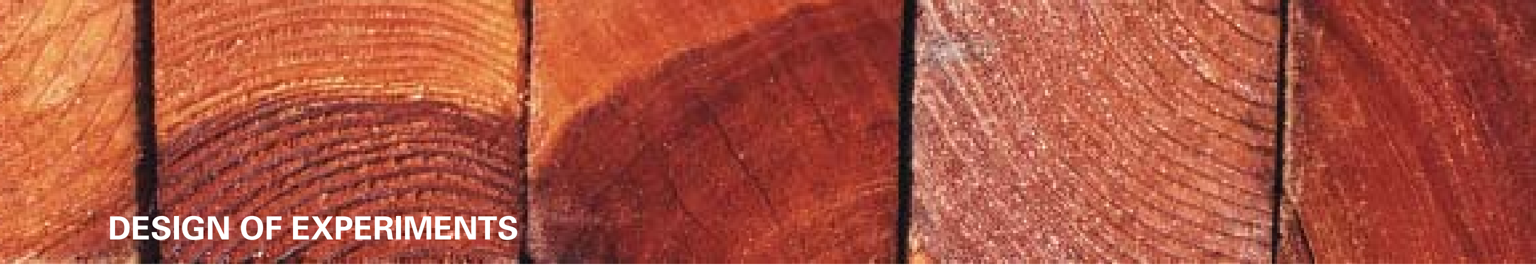
Response = pretreatment + WP (pretreatment) + stain + pretreatment x stain interaction + SP error,

The limitations and challenges of experimenting in the real world result in these simple experiments being the exception rather than the norm.

in which WP is a variable that goes from one to six indicating each whole-plot and must be declared as a random factor.

Specifying the model in this way allows the creation of two separate estimates of experimental error, an ingredient of the split-plot design. The nested term WP (pretreatment) comes from the fact the whole plots are nested within pretreatment.[2]

This term will be the correct error term for the pretreatment factor, and most software packages will correctly use this term for the F-test of pretreatment. The df will also be correctly calculated as 2(3 - 1) = 4 in which 2 represents the number of



levels for the pretreatment factor and 3 represents the number of replicates at the pretreatment level of the experiment.

The other estimate of experimental error, called the subplot error, is estimated by examining the variation that occurs between the 12 pairs of experimental runs that have the same pretreatment and stain setting minus the whole-plot experimental error.

The whole-plot experimental error is used to test the significance of the whole-plot factor, pretreatment. The subplot experimental error is used to test the significance of the subplot factor, stain and pretreatment by stain interaction. Therefore, the tests use a different mean square error in the denominator of the F-ratio.

Table 4 (p. 70) shows the F-statistic for the effect of pretreatment, the whole-plot factor, is:

$$F = \frac{\text{Mean square}_{\text{pretreatment}}}{\text{Mean square}_{\text{WP (pretreatment)}}} = \frac{782.04}{193.84} = 4.03.$$

Note the p-value of 0.115 indicates this factor is not significant. The F-test for the effect of stain, the subplot factor, is:

$$F = \frac{\text{Mean square}_{\text{stain}}}{\text{Mean square}_{\text{error}}} = \frac{88.67}{12.71} = 6.98.$$

Note the p-value of 0.006 indicates this factor is significant. The F-test for the effect of the pretreatment by stain interaction is:

$$F = \frac{\text{Mean square}_{\text{pretreat x stain}}}{\text{Mean square}_{\text{error}}} = \frac{20.93}{12.71} = 1.65.$$

Note the p-value of 0.231 indicates the interaction effect is not significant. Notice for both pretreatment and stain, these are different conclusions from the analysis assuming a completely randomized design.

Many experiments in industry involve two-level factors. In the wood experiment, the four stains could actually be a 2^2 in stain type and amount. All this does is add a little more structure to the experiment and the breakdown of the degrees of freedom.

For example, the previous 3 df for stain can now be broken down into 1 df for stain type, 1 df for amount and 1 df for the stain type by amount interaction. This is also true for the previous pretreatment by stain interaction, which is now 1 df for pretreatment by stain type interaction, 1 df for pretreatment by amount interaction and 1 df for the pretreatment by stain type by amount interaction.

Another Example

Consider another example with one hard-to-change factor (Z), three easy-to-change factors (A, B, C) and all factors at two levels. The hard-to-change factor is replicated so there are four whole plots, each with eight subplots.

Table 5 gives the design as it was carried out: First a level for Z is randomly selected, then the eight combinations of A, B and C are carried out in random order. The correct and incorrect analyses are shown in Table 6. Notice the incorrect analysis indicates Z is significant, while the Z x A and A x B interactions are shown as not significant.

Extensions on the Split-Plot

An astute reader can probably now surmise the split-plot framework can be expanded to even more complicated experiments. Several extensions that can be made to the split-plot scenario are:

- It can have more than one hard-to-change factor. (Make sure the extra factor(s) is really hard to change and not just inconvenient to change.)
- The whole-plot level design may involve blocks instead of being completely randomized.
- There may be several easy-to-change factors, which may necessitate using a fractional factorial design at the subplot level (you must be very careful because the alias structure is much more complicated in split-plot designs).
- More factors could be added that are subplots for one factor while at the same time whole plots for other factors. This results in a split-split-plot design.[3]

The design and analysis of industrial experiments involves understanding not only the treatment structure but also the three principles of the design structure: randomization, replication and controlling for known sources of variation (typically through blocking).

The experimenter should be made aware of an

TABLE 5 Data for the Second Example

Z	A	B	C	WP	Response
1	-1	1	1	1	108.4
1	1	-1	1	1	131.6
1	-1	-1	-1	1	124.0
1	1	-1	-1	1	134.9
1	-1	1	-1	1	103.7
1	1	1	-1	1	112.9
1	1	1	1	1	113.4
1	-1	-1	1	1	122.3
-1	-1	-1	-1	3	119.3
-1	1	1	-1	3	120.9
-1	1	1	1	3	123.0
-1	1	-1	1	3	127.9
-1	-1	1	1	3	117.3
-1	-1	-1	1	3	120.9
-1	1	-1	-1	3	129.9
-1	-1	1	-1	3	115.4
1	-1	1	1	2	100.8
1	1	1	-1	2	114.4
1	1	-1	1	2	132.8
1	1	-1	-1	2	131.4
1	-1	-1	-1	2	118.4
1	-1	1	-1	2	104.4
1	1	1	1	2	111.7
1	-1	-1	1	2	121.1
-1	1	1	-1	4	116.7
-1	-1	1	-1	4	112.8
-1	-1	1	1	4	112.2
-1	1	-1	1	4	127.7
-1	-1	-1	-1	4	118.4
-1	1	1	1	4	120.9
-1	1	-1	-1	4	127.0
-1	-1	-1	1	4	119.4

Z = hard-to-change factors
A, B and C = easy-to-change factors
WP = whole-plot errors

TABLE 6 Summary for Second Example

Correct split-plot analysis

Source	DF	SS	MS	F	P
Z	1	59.13	59.13	2.94	0.228
WP (Z)	2	40.17	20.08	6.83	*
A	1	597.72	597.72	203.13	0.000
B	1	1226.36	1226.36	416.77	0.000
C	1	1.49	1.49	0.51	0.486
Z x A	1	14.72	14.72	5.00	0.038
Z x B	1	285.01	285.01	96.86	0.000
Z x C	1	3.71	3.71	1.26	0.275
A x B	1	13.13	13.13	4.46	0.048
A x C	1	0.81	0.81	0.28	0.605
B x C	1	1.16	1.16	0.40	0.537
Error	19	55.91	2.94		
Total	31	2299.32			

Incorrect completely randomized analysis

Source	DF	SS	MS	F	P
Z	1	59.13	59.13	12.92	0.002
A	1	597.72	597.72	130.65	0.000
B	1	1226.36	1226.36	268.05	0.000
C	1	1.49	1.49	0.33	0.575
Z x A	1	14.72	14.72	3.22	0.087
Z x B	1	285.01	285.01	62.30	0.000
Z x C	1	3.71	3.71	0.81	0.378
A x B	1	13.13	13.13	2.87	0.105
A x C	1	0.81	0.81	0.18	0.678
B x C	1	1.16	1.16	0.25	0.619
Error	21	96.08	4.58		
Total	31	2299.32			

Z = hard-to-change factors
A, B and C = easy-to-change factors
WP = whole-plot errors
DF = degrees of freedom
SS = sums of squares
MS = mean square
F = F-statistic
P = p-value

important point about the experimental replication in a split-plot design. The effect of the whole-plot factor, which will have the least number of experimental replicates, is estimated less precisely than the subplot factors, which will have more experimental replicates. Thus, if allowed a choice when planning a split-plot experiment, the experimenter should try to put the most important factors at the subplot level.

Getting Beyond Academics

Many practitioners of experimentation are beginning to incorporate the principles and methodology of designed experiments developed in the statistical literature over the last 75 years. The first experiments learned in typical statistical and quality methodology training courses are those with simple design structures, such as the completely randomized design.

In practice, however, the limitations and challenges of experimenting in the real world result in these simple experiments being the exception rather than the norm. Typically, an experiment will contain some form of a restriction on the randomization. We fear that more often than not, these features are not being incorporated into the planning and analysis of the experiment.

With the recent growth and interest in the use of the statistical sciences in today's businesses, however, we expect the sophistication and understanding of experimentation will increase, and designs such as the split plot will become more readily recognized and properly analyzed.

REFERENCES

1. S.M. Kowalski and K.J. Potcner, "How To Recognize a Split-Plot Experiment," *Quality Progress*, Vol. 36, No. 11, pp. 60-66.
2. D.C. Montgomery, *Design and Analysis of Experiments*, fifth edition, Wiley, 2001.
3. Ibid.

BIBLIOGRAPHY

G.E.P. Box, "Split-Plot Experiments," *Quality Engineering*, Vol. 8, pp. 515-520.

KEVIN J. POTCNER *is a technical trainer at Minitab Inc. in State College, PA. He earned a master's degree in statistics from the Rochester Institute of Technology in New York and is a member of ASQ.*

SCOTT M. KOWALSKI *is a technical trainer at Minitab Inc. in State College, PA. He earned a doctorate in statistics from the University of Florida, Gainesville, and is a member of ASQ.*

Please comment

If you would like to comment on this article, please post your remarks on the *Quality Progress* Discussion Board at www.asq.org, or e-mail them to editor@asq.org.