PLOT SPLITTING

The Plot to Split The Plot

Some designed experiments are undermined by unwitting plot splitting

by Lynne B. Hare

Once there was an aspiring research baker named Jake. He worked in R&D at a large commercial baked goods company, having recently received an advanced degree in the technology. He was ready to set the world, if not the bread, on fire.

Jake's boss gave him his first assignment: Make the bread faster, better and cheaper. "Piece of cake," he thought. "I'll show her how I can use my spiffy new technology, coupled with my newly acquired group dynamics skills, to do just that."

Jake coordinated with his newfound colleagues, former professors, and new companions in related technologies, and he accumulated all the corporate bread lore he could—the good and the bad, the fabled, the real, the fancy and the facts. He learned about the corporate culture and power structure, who was up and who was down, who could make things happen, who said frog, who jumped, and how high.

To accomplish his assigned task, of course, he had to experiment, but the experimental turf, to his chagrin, was not his own, nor his boss's, nor hers, nor even that of the research upper crust.

Jake had to go, toque in hand, to the production people, and he knew the production people were not interested in research. They were interested in—and rewarded for—making more and more bread to bring home the bacon. They were indeed good corporate citizens, understanding at a deliberate distance the value of research, as long as the time commitment was held to a minimum. Doing that was the butter on their bread, so to speak.

Five factors

What factors might be involved in faster, better and cheaper bread production, given a sacred corporate formula? Jake's conclusion, based on interviews and aggregate views, was:

1. Oven temperature.
2. Oven throughput rate.
3. Oven air flow.
4. Air moisture.
5. Recipe active yeast content.

Recognize, please, that the oven is not your mom's conventional kitchen oven. To the contrary, it is a very long, very dark tunnel similar to that found on a railroad, but somewhat smaller and without the trains. Bread bakes as the dough runs this gauntlet.

Given the long batch baking time, reticence among production people to sacrifice output to research time, and five factors to study, Jake knew he had to experiment efficiently.

As it turned out, while attending an earlier short course on experimental strategy, he had learned about the benefits of first using screening designs for many factors, following up using more complicated plans focusing on the factors that showed promising effects from screening. Screening designs employed in the first phase of experimentation help identify the most important factors for further study in subsequent phases.

Jake looked up a five-factor screening design and went to his local, friendly statistician for help. Together, they arrived at a design characterized by two levels of each factor. It can be used to estimate the five factor effects, along with the 10 possible two-factor interactions. It was not capable of finding the sweet spot—that is, the optimal combination for faster, better and cheaper bread.

It could, however, help narrow the field of major contributing factors.

iStock.com/Betelgejze/Converted

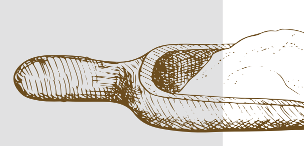


TABLE 1

Bread baking screening design

Batch	Oven temperature	Throughput	Air flow	Air moisture	Yeast	Speed	Quality	Cost
1	1	1	1	1	1			
2	-1	-1	1	-1	-1			
3	-1	-1	-1	1	-1			
4	-1	-1	1	1	1			
5	-1	1	-1	-1	-1			
6	1	1	-1	1	-1			
7	-1	1	-1	1	1			
8	1	-1	-1	-1	-1			
9	1	1	1	-1	-1			
10	1	-1	1	-1	1			
11	1	-1	-1	1	1			
12	-1	1	1	-1	1			
13	1	-1	1	1	-1			
14	-1	-1	-1	-1	1			
15	1	1	-1	-1	1			
16	-1	1	1	1	-1			

Note: -1 represents the low level of the factor, and 1 represents the high level of the factor to be determined after further discussion. Batches are listed in randomized order deliberately to minimize the influence of extraneous factors such as personnel changes and exterior temperature and humidity changes.

Formally, it is called a half replicate of a two-to-the-fifth factorial (2^{5-1}) experimental design, but Jake didn't want to tell anyone that because of the fear associated with the name. Besides, they'd call him a nerd.

Instead, Jake prepared a worksheet to show the specific settings of each of the 16 bread batches he wanted to produce.

Jake showed his worksheet to his boss and R&D colleagues, explaining the pluses and minuses, the randomization and the intention to take the experimental process in phases, with Table 1 outlining the first phase. They liked it, and they admired his diligence.

With their positive comments as confidence builders, he went to the production people. That visit didn't work out as well as he had hoped.

Feedback from production

The production people examined the plan and said they would ditch the randomization and instead run all the low temperature batches first, raise the oven temperature until it reached the new target—this might take a few hours—and they'd run the rest of the batches. Jake disagreed. "You've got to follow the design," he said. "Otherwise, the modeling won't work, and the outcome won't be repeatable."

"Son," they said, "do you know how long it takes to change the temperature of these ovens? You've got us changing up and down nine times. That would shut us down for a week! You're a good guy. We like you, but we can't run your design. Sorry."

Crestfallen, Jake went back to the statistician and told her their design wouldn't work. The production people said there were too many oven temperature changes. "Of course," the statistician said. "I should have thought of that! We'll develop a new design. Production will still have to change the temperature, but not quite as often."

"Why change the temperature more than once?" asked Jake. "Because any factor must be evaluated against its repeats, and the repeats must be random. Otherwise, you can't assure the boss that if you did it again, you'd get it again. And bosses always ask that question. It's their job."

"Look," the statistician said. "If you were only going to evaluate the effect of temperature, you'd run for a while at the low level, then the high level and then repeat, right?" Jake nodded.

"Well, we can do the same thing here for temperature, but we don't have to do it for the other factors," the statistician said. "It's called splitting the plot. The term comes from agricultural experimentation where field plots were assigned different levels of a main treatment and also sets of other treatments within plots. The design type is a 'split plot,' but we don't have to say that out loud."

Here it is in Online Table 1, which can be found on this column's webpage at qualityprogress.com. It was necessary to add four temperature changes, but that shouldn't break the bank.

Well, the production people grumbled a little, but Jake got his first "split-plot design ever" run. The production people were relieved because they only had to change the temperature three times. And Jake's boss was happy because her charge was moving nicely ahead with his new assignment. Jake hit the ground running!

And his new colleagues started calling him Jake the Baker. QP

Lynne B. Hare is a statistical consultant. He holds a doctorate in statistics from Rutgers University in New Brunswick, NJ. He is past chair of the ASQ Statistics Division and a fellow of ASQ and the American Statistical Association.

ONLINE TABLE 1
BREAD BAKING SPLIT PLOT DESIGN

Batch	Whole plots	Oven temperature	Throughput	Air flow	Air moisture	Yeast	Speed	Quality	Cost
1	1	-1	1	-1	1	-1			
2	1	-1	-1	1	1	1			
3	1	-1	1	-1	-1	1			
4	1	-1	1	1	-1	-1			
5	1	-1	-1	-1	-1	-1			
6	2	1	1	1	-1	-1			
7	2	1	1	-1	1	1			
8	2	1	-1	1	1	1			
9	2	1	-1	-1	-1	1			
10	2	1	-1	-1	1	-1			
11	3	-1	-1	1	-1	1			
12	3	-1	-1	-1	1	1			
13	3	-1	-1	1	1	-1			
14	3	-1	1	1	1	1			
15	3	-1	1	-1	-1	-1			
16	4	1	1	1	-1	1			
17	4	1	-1	1	-1	-1			
18	4	1	-1	-1	1	1			
19	4	1	1	-1	-1	-1			
20	4	1	1	1	1	-1			

Note: *The key is the same Table 1's. Oven temperature is difficult to change, so it is not fully randomized, but it is replicated. The remaining factors are randomized.*

EXPERIMENTAL DESIGN

How To Recognize A Split-Plot Experiment

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

The application of statistically designed experiments is becoming increasingly important in organizations engaged in Six Sigma and other quality initiatives. In conducting these experiments and analyzing the resulting data, experimenters become aware of the treatment structure of the design: the number of factors to be studied and the various factor level combinations.

For example, most practitioners know a 2^3 full factorial design consists of three factors, each at two levels, where all eight treatment combinations are studied. However, practitioners often neglect the details of how the experimental runs are performed and thus fail to see how this component, along with the treatment structure, determines which statistical approach to use.

Most would choose to run the eight treatment combinations in a completely randomized order, known as a 2^3 full factorial completely randomized design. Unfortunately, limitations involving time, material, cost and experimental equipment can make it inefficient and, at times, impossible to run a completely randomized design. In particular, it may be difficult to change the level for one of the factors. In this case, practitioners typically fix the level of the difficult-to-change factor and run all the combinations of the other factors—the split-plot design.

In 50 Words Or Less

- **Not incorporating the experimental approach into an analysis can result in incorrect conclusions.**
- **One type of statistical experimental design, known as the split-plot, is often more common in experimental situations than the completely randomized design.**
- **Several examples will help practitioners recognize the split-plot design.**

Recognizing a Split-Plot Design

Split-plot experiments began in the agricultural industry. Because one factor in the experiment is usually a fertilizer or irrigation method, it can only be applied to large sections of land called whole plots. The factor associated with this is therefore called a whole plot factor.

Within the whole plot, another factor, such as seed variety, is applied to smaller sections of the land, which are obtained by splitting the larger section of the land into subplots. This factor is therefore referred to as the subplot factor.

These same experimental situations are also common in industrial settings. Split-plot designs have three main characteristics:

1. The levels of all the factors are not randomly determined and reset for each experimental run. Did you hold a factor at a particular setting and then run all the combinations of the other factors?
2. The size of the experimental unit is not the same for all experimental factors. Did you apply one factor to a larger unit or group of units involving combinations of the other factors?
3. There is a restriction on the random assignment of the treatment combinations to the experimental units. Is there something that prohibits assigning the treatments to the units completely randomly?

The following industrial examples will help you recognize when it would be best to use a split-plot experiment.

Example A

Let's say you want to examine the image quality of a printing process by varying three factors:

- A = blanket type.
- B = cylinder gap.
- C = press speed.

Figure 1 illustrates a simple image of this part of a printing press.

You plan to study two different blanket types (1 and 2), three different cylinder gaps (low, medium and high) and two press speeds (low and high), and will run all 12 treatment combinations (see Figure 2) in the experiment. A completely randomized design would require you to run the 12 treatment combinations in a random order.

To change the cylinder gap and press speed, you simply make an adjustment on a control panel while the printing press is still running. Factors such as these are called easy-to-change factors. To change the blanket type, however, you must stop the press and manually replace the blanket. A change such as this is called a hard-to-change factor.

Now imagine the first three runs in your experiment are (A = 1, B = -1 and C = -1), (A = 2, B = 1 and C = -1) and (A = 1, B = 0 and C = 1). This means you would have to install the blanket three times (1 to 2, then back to 1), and using a completely randomized design would require you to frequently stop the press, thereby extending the time required to run the experiment.

A more time efficient approach, and one that fits

FIGURE 1 The Printing Press

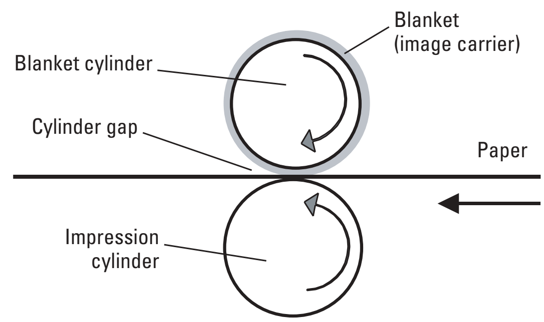
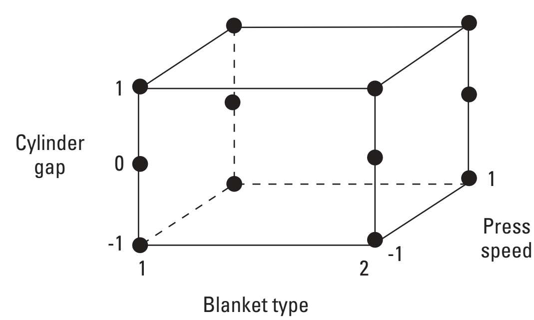


FIGURE 2 Factors That Affect Image Quality



into the split-plot framework, would be to randomly choose one of the blanket types (1 or 2), install it on the printing press, and run the six treatment combinations in cylinder gap and press speed in a random order. Then you would change the blanket type and run the six treatments in another random order, repeating the process until you reached the desired number of replicates for the blanket type factor.

The way in which the factor levels for blanket type are changed in the second approach involves a different randomization scheme from that of the factor levels for the other two factors. This different randomization structure is one feature of a split-plot design and is common when some of the factors are difficult to change. If the experiment were conducted in this manner, it would be incorrect to analyze the data as if you had run the experiment as a completely randomized design.

FIGURE 3 Factors That Affect Wood's Water Resistance

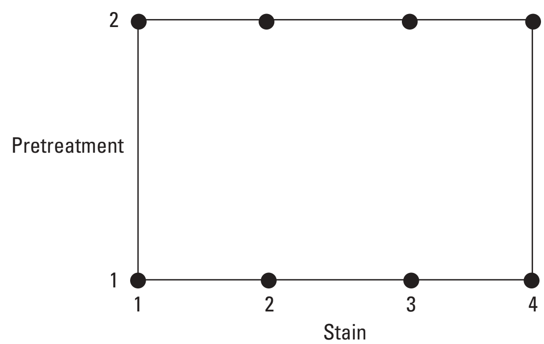
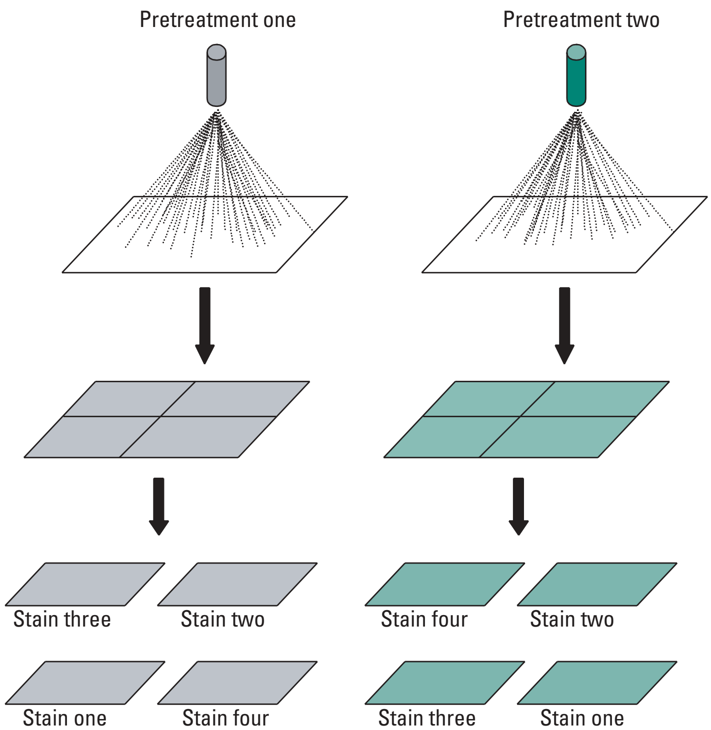


FIGURE 4 Treatment Application



Example B

Now let's look at an experiment involving the water resistance property of wood in which you select two types of wood pretreatment (1 and 2) and four types of stain (1, 2, 3 and 4) as variables (see Figure 3).

To conduct this experiment in a randomized fashion, you would need eight wood panels for each full replicate of the design. You would then randomly assign a particular pretreatment and stain combination to each wood panel.

That's when you discover how difficult it is to apply the pretreatment to a small wood panel. The easiest way to do it would be to apply each of the pretreatment types (1 and 2) to an entire board, then cut each board into four pieces and apply the four stain types to the smaller pieces (see Figure 4).

The experimental units for the two factors in this experiment are not the same. For the pretreatment factor, the experimental unit is the entire board, but for the stain factor, the experimental unit is one of the small panels cut from the large board. Varying sizes of experimental units is another feature of split-plot designs.

Example C

Let's say you want to examine the effect the four following

FIGURE 5 Factors That Affect the Strength of Plastic

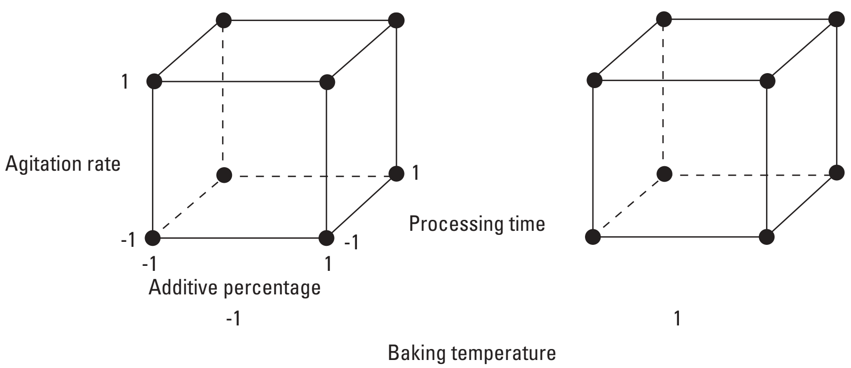
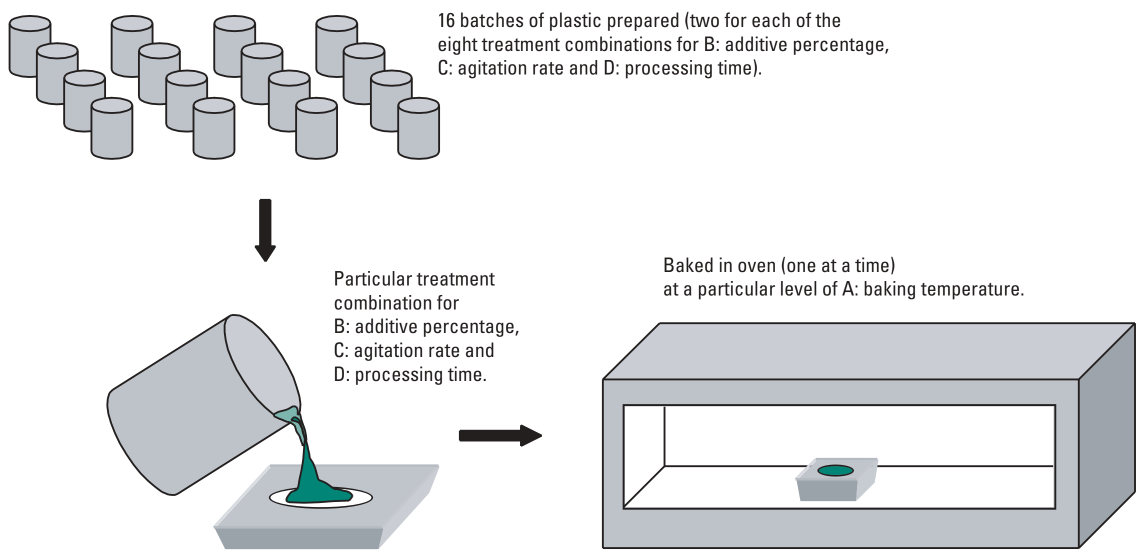


FIGURE 6 Individual Baking Process



various factors have on the strength of plastic:

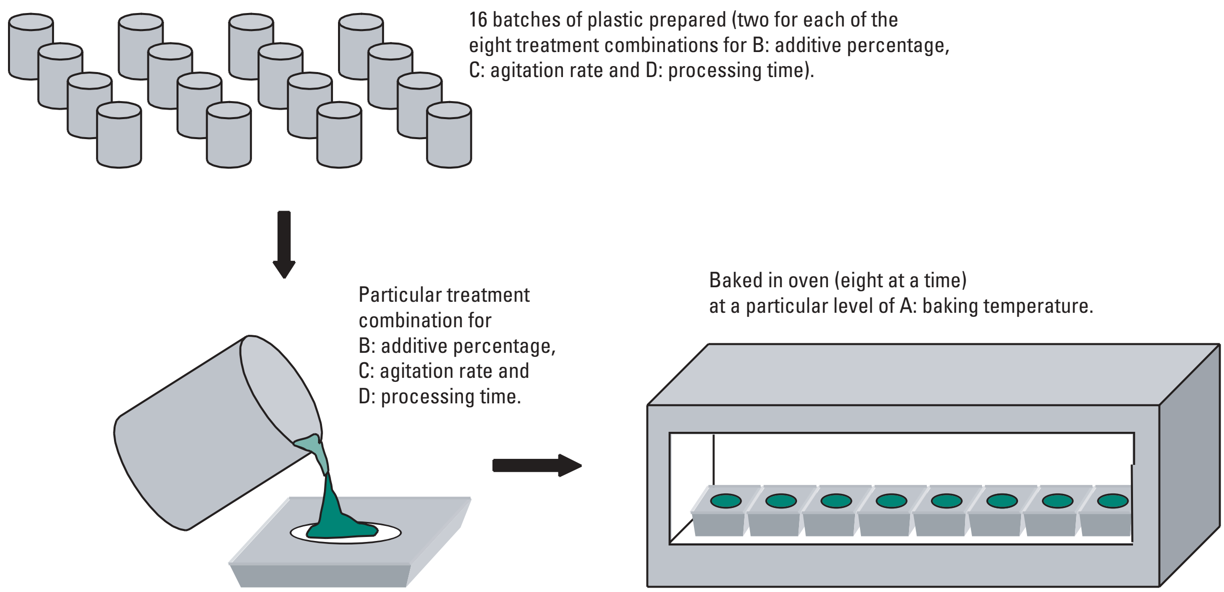
- A = baking temperature.
- B = additive percentage.
- C = agitation rate.
- D = processing time.

You plan to study each factor at two levels: low = -1 and high = 1. See Figure 5 for a graphical representation of this treatment design.

To conduct this experiment as a completely randomized design, you would run all 16 treatment combinations in a random order. After obtaining the required 16 batches of plastic, two for each of the eight different combinations of factors B, C and D, you would pour the plastic into molds and bake each individually at one of two temperatures (see Figure 6).

If you conducted the experiment in this way, you would have to frequently change the baking oven's temperature, which may take some time to stabilize. A completely randomized design also implies each run of the oven is a true experimental run. That means 16 separate runs of the oven are

FIGURE 7 Group Baking Process



needed, therefore adding considerable time to the experiment.

A more efficient approach would be to bake all eight molds for one temperature setting at the same time. You would follow this by a single run of the oven at the other temperature level, repeating the process until you have run the desired number of replicates for the temperature factor (see Figure 7).

Now you no longer have a completely randomized design but a split-plot design. Why? There are three reasons:

1. For each of the three factors—additive percentage, agitation rate and processing time—one experimental unit equals one batch of plastic, while for the temperature factor, one experimental unit equals all eight batches.
2. Temperature can be thought of as a hard-to-change factor, and the three easy-to-change factors are varied within a level of the hard-to-change factor.
3. The temperature factor uses a different randomization scheme from the other factors. The molds are assigned to the temperature factor in groups of eight as opposed to individually.

Split-Plot Design Affects Analysis

Many practitioners fail to see there is more to knowing the correct analysis than just being able to identify the treatment structure. The analysis of designed experiments directly follows from the way the runs were carried out.

For example, when a designed experiment uses blocks such as days or batches, the analysis of the experiment includes a term for these blocks. When a designed experiment is performed by fixing a factor and then running the combinations of the other factors, using different sized experimental units or using a different randomization for the factors (a split-plot design), the analysis should incorporate these features.

In example C, the complete 2^3 factorial treatment design was replicated twice using the split-plot approach. This resulted in the 32 response values shown in Table 1. The responses were first analyzed incorrectly as if they came from a completely randomized design. The responses were then correctly analyzed as a split-plot experiment. (Our intention is not to teach the analysis, but interested readers can look at Table 2 (p. 66)

for a summary of the two different analyses.)

The results of the incorrect analysis, a completely randomized design, indicate that, at the 0.05 significance level, the main effects for A (baking temperature) and D (processing time) are significant, as are the AC (baking temperature/agitation rate) and AD (baking temperature/processing time) interactions.

The results of the correct, split-plot analysis indicate the main effects for B (additive percentage) and D (processing time) are significant at the 0.05 level and A (baking temperature) is not. In addition to the AC (baking temperature/agitation rate) and AD (baking temperature/processing time) interactions, the CD (agitation rate/processing time) interaction is also significant.

Experimental Error

Two interesting results appear when the two analysis approaches are compared:

1. The effect of the baking temperature was thought to be significant when analyzed as a completely randomized design but was actually insignificant when analyzed correctly. The whole plot error of 56.29 is much larger than the error of 14.21 from the completely randomized design analysis. This would cause you to incorrectly assume baking temperature is an important effect.
2. Effects at the subplot level that were not significant when analyzed as a completely randomized design are seen as significant when analyzed correctly. The subplot error of 9.78 is smaller than the one that arises when a completely randomized design is incorrectly assumed. As a result, important effects at the subplot level that were missed in the incorrect analysis are now seen.

Why did this happen? In the completely randomized design, all factor effects use the mean square error as the estimate of experimental error. In a split-plot experiment, however, there are two different experimental error structures: one for the whole plot factor and one for the subplot factors. This is a result of the two separate randomizations that occur when the experiment is run.

Experimental error is caused when the actual experimental conditions are replicated. This could include the preparation and mixing of the plastic batches or the setup and temperature stabilization of the oven. For the baking temperature factor, there are only four experimental units—each set of eight molds placed together in the oven.

Even though each of these eight molds comes from a different treatment combination of the other three factors, they were all processed in a single run of the oven. They do not provide an estimate of experimental error for the whole plot factor. The experimental error for the whole plot factor comes

TABLE 1 The 32 Response Values

	Temperature	Additive	Rate	Time	Strength
	A	B	C	D	Y
1	1	1	-1	-1	51.9
2	1	1	-1	1	66.8
3	1	1	1	-1	66.2
4	1	1	1	1	70.8
5	1	-1	1	-1	61.3
6	1	-1	1	1	68.5
7	1	-1	-1	1	59.5
8	1	-1	-1	-1	58.5
9	-1	1	-1	-1	57.4
10	-1	1	-1	1	57.5
11	-1	-1	1	-1	56.5
12	-1	1	1	1	63.9
13	-1	-1	1	1	56.4
14	-1	1	1	-1	58.1
15	-1	-1	-1	1	53.2
16	-1	-1	-1	-1	59.5
	Temperature	Additive	Rate	Time	Strength
	A	B	C	D	Y
17	-1	-1	-1	-1	66.6
18	-1	-1	-1	1	63.9
19	-1	1	1	-1	62.6
20	-1	1	1	1	63.2
21	-1	-1	1	-1	56.1
22	-1	1	-1	1	63.3
23	-1	-1	1	1	62.7
24	-1	1	-1	-1	65.0
25	1	-1	-1	-1	59.5
26	1	-1	-1	1	64.2
27	1	-1	1	1	68.0
28	1	-1	1	-1	58.6
29	1	1	-1	-1	65.6
30	1	1	1	1	73.3
31	1	1	-1	1	61.5
32	1	1	1	-1	64.0

TABLE 2 Summary of Incorrect and Correct Analyses for Example

Incorrect completely random			Correct split-plot	
Term	Significance	Variability	Significance	Variability
Temperature ▲	Significant	14.21	Not significant	56.29
Additive ▲	Not significant	14.21	Significant	9.78
Rate	Not significant	14.21	Not significant	9.78
Time	Significant	14.21	Significant	9.78
Temperature/additive	Not significant	14.21	Not significant	9.78
Temperature/rate	Significant	14.21	Significant	9.78
Temperature/time	Significant	14.21	Significant	9.78
Additive/rate	Not significant	14.21	Not significant	9.78
Additive/time	Not significant	14.21	Not significant	9.78
Rate/time ▲	Not significant	14.21	Significant	9.78
		One error term for all		Two error terms

▲ Shows terms that have a different interpretation between the two analyses.

from the variation experienced when the temperature is changed. This whole plot error is typically larger than the error from a completely randomized design.

In conducting a split-plot experiment, you need to be sure there is true replication in the whole plot factor. If each level of baking temperature was run only once and not replicated as it was here, there would be no estimate of whole plot experimental error and, therefore, no statistical test for this factor.

The challenges faced by practitioners result in completely randomized experiments being the exception, not the norm. Unfortunately, split-plot and other noncompletely randomized experimental designs have not received proper attention in most Black Belt statistical training courses because the mathematical concepts are usually more complicated or more general than those in the completely randomized design.

Fortunately, the availability of statistical software has slowly started to ease the analysis and interpretation of more complicated experimental structures, such as split-plot experiments.

Knowledge of the split-plot design gives practitioners another option with which to conduct experiments more efficiently.

BIBLIOGRAPHY

Box, G.E.P., "Split-Plot Experiments," *Quality Engineering*, Vol. 8, No. 3, pp. 515-520.

Box, G.E.P., and Stephen Jones, "Split-Plots for Robust Product and Process Experimentation," *Quality Engineering*, Vol. 13, No. 1, pp. 127-134.

Montgomery, D.C., *Design and Analysis of Experiments*, fifth edition, John Wiley & Sons, 2001.

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.*

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.*

Split-Plot Tables 1 & 2 Quality Progress November 2003

```
options nodate nonumber ls=80 nocenter;
data plastic;
input temp   oven   additive      rate   time   y @@;
datalines;
1      1      1      -1     -1     51.9 1       1      1      -1     1      66.8
1      1      1      1      -1     66.2 1       1      1      1      1      70.8
1      1      -1     1      -1     61.3 1       1      -1     1      1      68.5
1      1      -1     -1     1      59.5 1       1      -1     -1     -1     58.5
-1     2      1      -1     -1     57.4 -1      2      1      -1     1      57.5
-1     2      -1     1      -1     56.5 -1      2      1      1      1      63.9
-1     2      -1     1      1      56.4 -1      2      1      1      -1     58.1
-1     2      -1     -1     1      53.2 -1      2      -1     -1     -1     59.5
-1     1      -1     -1     -1     66.6 -1      1      -1     -1     1      63.9
-1     1      1      1      -1     62.6 -1      1      1      1      1      63.2
-1     1      -1     1      -1     56.1 -1      1      1      -1     1      63.3
-1     1      -1     1      1      62.7 -1      1      1      -1     -1     65.0
1      2      -1     -1     -1     59.5 1       2      -1     -1     1      64.2
1      2      -1     1      1      68.0 1       2      -1     1      -1     58.6
1      2      1      -1     -1     65.6 1       2      1      1      1      73.3
1      2      1      -1     1      61.5 1       2      1      1      -1     64.0
;
proc glm;
class temp additive rate time;
model y =
temp additive rate time temp*additive temp*rate temp*time additive*rate additive*time rate*time;
run;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	464.7606250	46.4760625	3.27	0.0106
Error	21	298.2490625	14.2023363		
Corrected Total	31	763.0096875			

Source	DF	Type III SS	Mean Square	F Value	Pr > F	
temp	1	85.47781250	85.47781250	6.02	0.0230	←
additive	1	45.36281250	45.36281250	3.19	0.0884	
rate	1	41.17781250	41.17781250	2.90	0.1034	
time	1	75.95281250	75.95281250	5.35	0.0310	←
temp*additive	1	1.08781250	1.08781250	0.08	0.7847	
temp*rate	1	78.43781250	78.43781250	5.52	0.0286	←
temp*time	1	62.44031250	62.44031250	4.40	0.0483	←
additive*rate	1	27.93781250	27.93781250	1.97	0.1754	
additive*time	1	2.94031250	2.94031250	0.21	0.6538	
rate*time	1	43.94531250	43.94531250	3.09	0.0931	

```
proc glm;
class temp oven additive rate time;
model y = temp oven(temp) additive rate time temp*additive temp*rate temp*time additive*rate
additive*time rate*time;
random oven(temp) / test;
run;
```

Tests of Hypotheses for Mixed Model Analysis of Variance

Dependent Variable: y

Source	DF	Type III SS	Mean Square	F Value	Pr > F
temp	1	85.477812	85.477812	1.52	0.3427
Error	2	112.390625	56.195313		

Error: MS(oven(temp))

Source	DF	Type III SS	Mean Square	F Value	Pr > F	
oven(temp)	2	112.390625	56.195313	5.74	0.0112	
additive	1	45.362812	45.362812	4.64	0.0443	←
rate	1	41.177812	41.177812	4.21	0.0542	
time	1	75.952812	75.952812	7.76	0.0118	←
temp*additive	1	1.087812	1.087812	0.11	0.7424	
temp*rate	1	78.437812	78.437812	8.02	0.0107	←
temp*time	1	62.440312	62.440312	6.38	0.0206	←
additive*rate	1	27.937812	27.937812	2.86	0.1074	
additive*time	1	2.940312	2.940312	0.30	0.5899	
rate*time	1	43.945312	43.945312	4.49	0.0474	←
Error: MS(Error)	19	185.858438	9.782023			

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

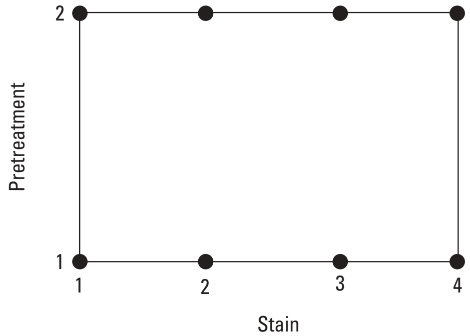
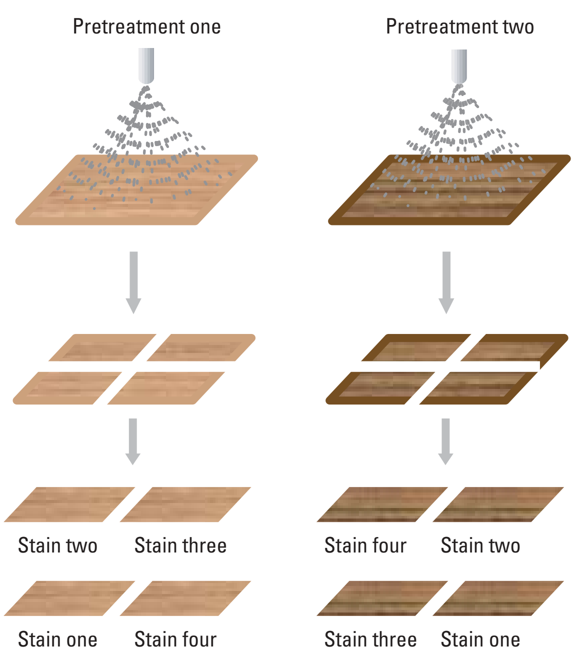
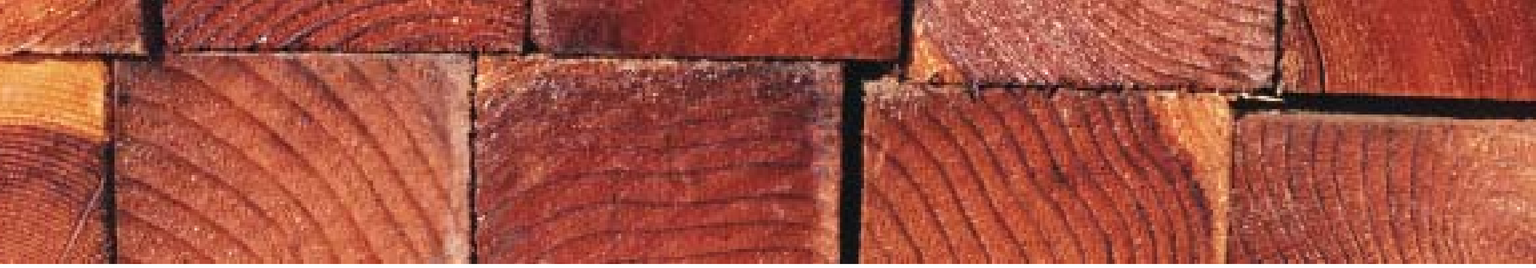


FIGURE 2 Treatment Application





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.

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 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

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:

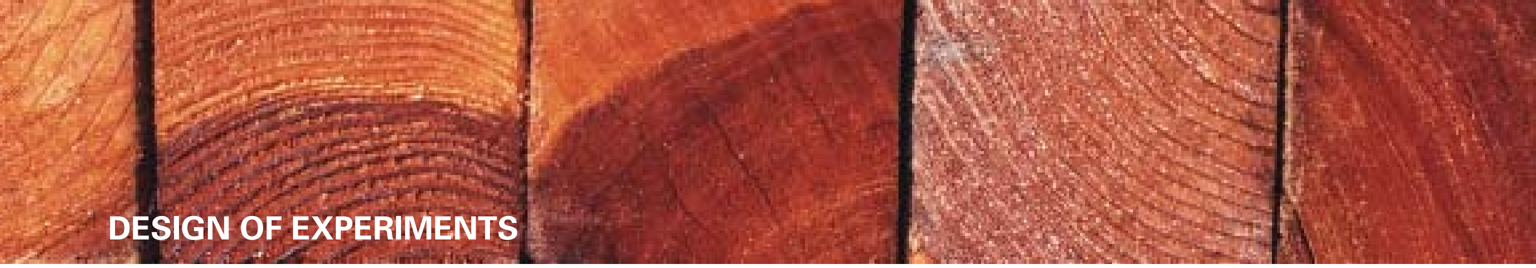
Response = pretreatment + WP (pretreatment) + stain + pretreatment x stain interaction + SP error,

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.

Split-Plot Tables 1,2,3,4 Quality Progress December 2004

```
options nodate nonumber ls=80 nocenter;
data wood;
input pretreat      stain  board  resistance @@;
cards;
2      2      1      53.5 2 4      1      32.5 2 1      1      46.6
2      3      1      35.4 2 4      2      44.6 2 1      2      52.2
2      3      2      45.9 2 2      2      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      3      32.1 2 4      3      30.1
2      2      3      34.4 2 3      3      32.2 1 1      3      52.8
1      3      3      51.7 1 4      3      55.3 1 2      3      59.2
;

proc sort;
by pretreat board;
run;

proc means;
var resistance;
by pretreat board;
run;
```

pretreat=1 board=1

N	Mean	Std Dev	Minimum	Maximum
4	45.2750000	4.7549097	40.8000000	51.8000000

pretreat=1 board=2

N	Mean	Std Dev	Minimum	Maximum
4	56.1750000	4.0966450	51.1000000	60.9000000

pretreat=1 board=3

N	Mean	Std Dev	Minimum	Maximum
4	54.7500000	3.3271610	51.7000000	59.2000000

pretreat=2 board=1

N	Mean	Std Dev	Minimum	Maximum
4	42.0000000	9.7846819	32.5000000	53.5000000

pretreat=2 board=2

N	Mean	Std Dev	Minimum	Maximum
4	47.7500000	3.3391616	44.6000000	52.2000000

pretreat=2 board=3

N	Mean	Std Dev	Minimum	Maximum
4	32.2000000	1.7568912	30.1000000	34.4000000

```
data justwholeplot;
input pretreat y;
cards;
1 45.275
1 56.175
1 54.75
2 42.00
2 47.75
2 32.20
;

proc glm;
class pretreat;
model y = pretreat;
run;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	195.5104167	195.5104167	4.03	0.1150
Error	4	193.8404167	48.4601042		
Corrected Total	5	389.3508333			

```
proc glm data=wood;
class pretreat stain;
model resistance = pretreat stain pretreat*stain;
run;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	1110.838333	158.691190	2.74	0.0452
Error	16	927.880000	57.992500		
Corrected Total	23	2038.718333			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
pretreat	1	782.0416667	782.0416667	13.49	0.0021
stain	3	266.0050000	88.6683333	1.53	0.2454
pretreat*stain	3	62.7916667	20.9305556	0.36	0.7820

```
proc glm data=wood;
class pretreat stain board;
model resistance = pretreat board(pretreat) stain pretreat*stain;
random board(pretreat) / test;
run;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	1886.200000	171.472727	13.49	<.0001
Error	12	152.518333	12.709861		
Corrected Total	23	2038.718333			

Tests of Hypotheses for Mixed Model Analysis of Variance

Dependent Variable: resistance

Source	DF	Type III SS	Mean Square	F Value	Pr > F
pretreat	1	782.041667	782.041667	4.03	0.1150
Error	4	775.361667	193.840417		
Error: MS(board(pretreat))					

Source	DF	Type III SS	Mean Square	F Value	Pr > F
board(pretreat)	4	775.361667	193.840417	15.25	0.0001
stain	3	266.005000	88.668333	6.98	0.0057
pretreat*stain	3	62.791667	20.930556	1.65	0.2309
Error: MS(Error)	12	152.518333	12.709861		