Easier Experimenting When You Can't Randomize!

Feb 2016 MN ASQ Meeting

Shari L Kraber, PE
Stat-Ease, Inc., Minneapolis, MN
shari@statease.com

ASQ
The Global Voice of Quality™

Learning Objectives

- Why randomize?
- How to handle hard-to-change factors via split-plot design of experiments.
- The price that must be paid for restricting randomization to make a test plan more convenient.

What you should be able to do <u>after</u> this talk:
Recognize when a split plot might the DOE of choice.

2

ASQ

The gist of this talk

- R. A. Fisher warned that *"Designing an experiment is like gambling with the devil: only a random strategy can defeat all his betting systems."*[1]
- George Box countered that *"Often, it is not only most convenient but also most effective to run industrial experiments without randomizing all the factors."*[2]
- By conveniently grouping hard-to-change factors, split plots provide a middle ground between complete randomization versus none at all. ☺
- But a price must be paid in statistical power. ☹

1. Source: Appendix of Box, Hunter and Hunter, *Statistics for Experimenters*, 2nd Ed., 2005.
2. "Quality Quandaries", *Quality Engineering* 8(3), 515-520,1996.

Agenda

- **Devilish things that happen if you do not randomize**
- How split plots facilitate hard-to-change (HTC) factors
- George Box solves a heated HTC quality quandary
- Helicopter testing made easy via split plot
- Conclusion

Devilish things happen if you do not randomize

Consider a simple two-level designed experiment (DOE) on three factors. Here is the layout in standard order. Why not run it this way, designating a hard-to-change (HTC) factor—say oven temperature—as "C"?

Std	A	B	C
1	–	–	–
2	+	–	–
3	–	+	–
4	+	+	–
5	–	–	+
6	+	–	+
7	–	+	+
8	+	+	+

5

The Lurking Devils 

Most processes are subjected on a daily basis or even longer periods to time-related variables:

- Heat rising
- Humidity going up
- Operators tiring out
- Machines wearing down
- What else? How many of you see this happening in your area?

"<u>Stuff" happens</u>!

This creates something dreaded by statisticians: "Bias" (a 4-letter word!)—what comes out of the analysis as being the main effect of factor C may well be one of these lurking factors. <u>Only by randomizing* can this source of bias be defeated</u>.

**(Or via a "trend-free" run order, which is beyond this talk's scope.)*

6

Agenda

- Devilish things that happen if you do not randomize
- <u>How split plots facilitate hard-to-change (HTC) factors</u>
- George Box solves a heated HTC quality quandary
- Helicopter testing made easy via split plot
- Conclusion

7

How split plots facilitate hard-to-change factors

Split plots provide an intermediate structure for design of experiments (DOE) between complete randomization versus not at all. In these designs:

- Hard-to-change (HTC) factor combinations remain fixed within groups.
  - *If at all practical, randomize the order of the groups!*
- Easy-to-change (ETC) factors remain completely randomized within these groups.

8

A split-plot field layout

Fisher and all at Rothamsted Agricultural Research Station in UK pioneered the application of split plots such as this one on:

A. Six varieties of sugar beets (number 1 through 6) sown either
B. Early (E) or late (L).*

Doing it completely randomized (CR) as shown on top could not be done due to the time of sowing being hard to change. Thus the split plot (SP) below is the design of choice.

CR:	E5	L1	L4	E2	E6	E3	L3	E1	L6	L5	E4	L2
SP:	E4	E1	E6	E5	E3	E2	L2	L3	L6	L5	L1	L4

This confounds time of sowing with blocks of land thus providing no statistical power for assessing this factor (early versus late). The only way around this is to replicate the whole plots, that is, repeat this entire experiment on a number of fields.

*Source: D. R. Cox, *Split Unit Principle*, 1958, Wiley & Sons, New York.

ASQ

Power loss

The devil is in the statistical details: As noted in the sugar beet spilt plot the restriction in randomization causes a loss of power for seeing HTC effects (time of sowing) relative to the ETCs (types of beets).

You must assess the cost of convenience in terms of power loss.

Let's get a feel for this via two case studies.

But first we should review the concept of statistical power.

ASQ

What is Power? (No Factor Effect; H_0: $\Delta = 0$)

Power = (1−β)*100%

Power is the probability of revealing an active effect of size delta (Δ) relative to the noise (σ) as measured by signal to noise ratio (Δ/σ).

It should be high (at least 80%!) for the effect size of interest.

Effect?		ANOVA says:	
		Retain H_0	***Reject H_0***
Truth:	***No***	OK☺	Type I Error (alpha) *False Alarm*
	Yes	Type II Error (beta) *Failure to detect*	OK☺

ASQ

Agenda

- ➢ Devilish things that happen if you do not randomize
- ➢ How split plots facilitate hard-to-change (HTC) factors
- ➢ George Box solves a heated HTC quality quandary
- ➢ Helicopter testing made easy via split plot
- ➢ Conclusion

ASQ

George Box solves heated quality quandary

George Box in a Quality Quandaries column on split-plot experiments* detailed a clever experiment that discovered a highly corrosion-resistant coating for steel bars. The engineers wanted to test two key factors:

A. Heat at three levels--360, 370, and 380 Deg C
B. Four coatings baked onto steel bars.

However it took a long time to change temperature in the oven whereas coatings could be readily applied. Box solved this problem by grouping the temperatures in the test plan shown on the following slide.

**Quality Engineering*, 8(3), 515-520, 1996

13

Split plot solves quandary of HTC heats

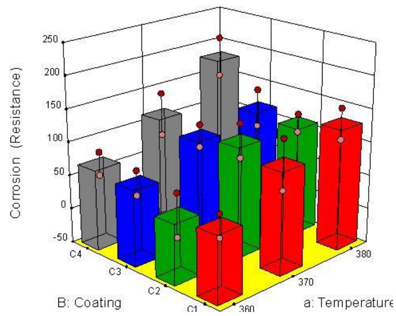
Group	**Heats** (Deg C)	**Coatings by Position** (Corrosion resistance in parentheses)			
1	360	C2 (73)	C3 (83)	C1 (67)	C4 (89)
2	370	C1 (65)	C3 (87)	C4 (86)	C2 (91)
3	380	C3 (147)	C1 (155)	C2 (127)	C4 (212)
4	380	C4 (153)	C3 (90)	C2 (100)	C1 (108)
5	370	C4 (150)	C1 (140)	C3 (121)	C2 (142)
6	360	C1 (33)	C4 (54)	C2 (8)	C3 (46)

Observe that the ETC factor coatings are randomly positioned, and how Box made it even easier, in addition to grouping by heats, by increasing the furnace temperature run-by-run and then decreasing it gradually. This is not ideal—better to randomize groups with re-sets-but a practical compromise.

14

Outcome of Corrosion Split Plot

The HTC factor (temperature) created so much noise in this process that in a randomized design it would have overwhelmed the effect of coating. The split plot structure overcomes this variability by grouping the heats—in essence filtering out the temperature differences.



Power increase in subplots

In this case the heats were grouped into what statisticians refer to as “whole plots”, within which the engineers applied coatings randomly in “subplots”. The two-factor interaction aB between temperature and coating achieved significance at the $p<0.05$ threshold. The main effect of coating (B) also came out significant.

If this experiment had been run completely randomized, p-values for the coating effect and the coating-temperature interaction come out to approximately 0.4 and 0.85 p-values; respectively, that is, nowhere near to being statistically significant.

In other words, the split plot structure provides more power to the subplots.

This comes at a cost: Less power for the whole plot factor(s).

Agenda

- ➢ Devilish things that happen if you do not randomize
- ➢ How split plots facilitate hard-to-change (HTC) factors
- ➢ George Box solves a heated HTC quality quandary
- ➢ **Helicopter testing made easy via split plot**
- ➢ Conclusion

Helicopter testing made easy via split plot

Inspired by news of a supreme paper—Conqueror CX22—made into an airplane that broke the Guinness World Record™ for greatest distance flown*, engineers at Stat-Ease designed a two-level factorial experiment on paper helicopters. They tested:

A. Paper: 24# Navigator (standard) vs 26.6# CX22 (supreme)
B. Wing Length: Short vs Long
C. Body Length: Short vs Long
D. Body Width: Narrow vs Wide
E. Clip: Off vs On
F. Drop: Bottom vs Top

Manufacturing factors A-D were hard to change, thus a split plot experiment made it far easier to get this done.

*Sean Hutchinson in "The Perfect Paper Airplane," *Mental Floss*, January 14, 2014.

Split plot experiment on helicopters *(partial list)*

Group	Run	a:Paper	b:Wing	c:Body Length	d:Body Width	E:Clip	F:Drop
1	1	Nav Ultra	Long	Short	Narrow	Off	Top
1	2	Nav Ultra	Long	Short	Narrow	On	Bottom
2	3	CX22	Short	Short	Narrow	Off	Top
2	4	CX22	Short	Short	Narrow	On	Bottom
3	5	Nav Ultra	Long	Long	Narrow	Off	Bottom
3	6	Nav Ultra	Long	Long	Narrow	On	Top
…	…	…	…	…	…	…	…
16	31	CX22	Short	Long	Wide	Off	Top
16	32	CX22	Short	Long	Wide	On	Bottom

Staff made 16 'copters per HTC factors a-d and engineers flew them varying ETC factor E-F. Not only did this save half the time, it also conserved on the extremely expensive premium CX22 paper.

Trade-Off of Power for Convenience

Of the two responses, flight time and deviation from target, the latter featured the lowest signal-to-noise ratio—the driver for statistical power. (Note: we dropped each copter three times and averaged the results.)

Response	Signal	Noise	Signal/Noise
Time avg	0.5 sec	0.15 sec	3.33
Target avg	5.0 cm	2.00 cm	**2.50**

Here is the trade-off in power to see a meaningful effect on copter accuracy (distance off target) for a completely randomized versus a split-plot design:

Design	HTC (a-d)	ETC (E, F)
CR	97.5%	97.5%
SP	82.1%	99.9%

Not bad! Well worth taking the loss in power on HTC!

Agenda

- ➢ Devilish things that happen if you do not randomize
- ➢ How split plots facilitate hard-to-change (HTC) factors
- ➢ George Box solves a heated HTC quality quandary
- ➢ Paper helicopter testing made easy via split plot
- ➢ Conclusion

21

Summary: Must we randomize our experiment?

Box addressed this frequently-asked question as follows*:

- ➢ Always randomize in those cases where it creates little inconvenience.
- ➢ When an experiment becomes impossible being subjected to randomization and you can safely assume your process is stable, that is, any chance-variations will be small compared to factor effects, then run it as you can in non-random order; but, if due to process variation, the results would be “useless and misleading” without randomization, abandon it and first work on stabilizing the process;
- ➢ Consider a split-plot design.

*1989 report “Must We Randomize Our Experiment”, #47, Center for Quality and Productivity Improvement, University of Wisconsin – Madison

22

Learning Objectives: Recap

- Why randomize?
- How to handle hard-to-change factors via split-plot design of experiments.
- The price that must be paid for restricting randomization to make a test plan more convenient.

What you should be able to do after this talk:
Recognize when a split plot might the DOE of choice.

ASQ

23

Easier Experimenting When You Can't Randomize!

Best of luck for your experimenting!

Thanks for listening!

-- Shari

Shari Kraber, PE
Stat-Ease, Inc.
shari@statease.com

ASQ

24