RSM Split-Plot Designs & Diagnostics Solve Real-World Problems

Shari Kraber
Stat-Ease, Inc.
2021 E. Hennepin Ave.
Minneapolis, MN 55413
Shari@statease.com

Pat Whitcomb
Stat-Ease, Inc.
2021 E. Hennepin Ave.
Minneapolis, MN 55413

Martin Bezener
Stat-Ease, Inc.
2021 E. Hennepin Ave.
Minneapolis, MN 55413

FTC 2017

FTC 2017 1

Agenda Transition

- **Review split plot concept**
- Sizing response surface designs
- Case Study: Building a split plot *(Corn Milling)*
- Analyzing a split plot *(Corn Milling)*
- Diagnostics *(Corn Milling)*

FTC 2017 2

George Box[1]

Must We Randomize Our Experiment *(1 of 2)*

"You do not need to randomize if you believe your system is, to a sufficiently good approximation, in a state of control and can be relied on to stay in that state while you make experimental changes (that presumably you have never made before).

"Sufficiently good approximation to a state of control means that over the period of experimentation, differences due to the slight degree of non-stationarity, will be small compared with differences due to the treatments.

"In making such a judgment bear in mind that belief is not the same as wishful thinking."

FTC 2017 3

George Box

Must We Randomize Our Experiment *(2 of 2)*

George's advice:

- In those cases where randomization only slightly complicates the experiment, always randomize.
- In those cases where randomization would make the experiment impossible or extremely difficult to do, but you can make an honest judgment about being in a state of control, of the kind outlined on the previous slide, run the experiment anyway without randomization.
- If you believe the process is so unstable that without randomization the results would be useless and misleading, and randomization would make the experiment impossible or extremely difficult to do, then do not run the experiment. Work instead on stabilizing the process or on getting the information some other way.
- A compromise design that sometimes helps to overcome some of these difficulties is the split-plot arrangement.

FTC 2017 4

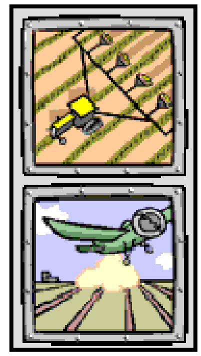
Split-Plot Designs

Problem

Often in designed experiments some of the factors are more difficult or expensive to vary than others. In some cases conducting completely randomized design isn't practical.

Solution

Restrict the randomization so it is practical to conduct the design. Sometimes the restriction results in a "split-plot" design.



FTC 2017 5

Split-Plot Designs

The "split-plot" design originated in the field of agriculture. Experimenters applied one treatment to a large area of land, called a "whole plot" and other treatments to smaller areas of land within the whole plot called "subplots".

whole plot				whole plot			
subplot	subplot	subplot	subplot	subplot	subplot	subplot	subplot
subplot	subplot	subplot	subplot	subplot	subplot	subplot	subplot

FTC 2017 6

Split-plot Designs

Split-Plot Designs:

- Split plots have two types of factors: “Hard-to-change” (HTC) and “Easy-to-change” (ETC).
- The randomization of HTC factor is restricted.
- Split plots naturally arise in many DOE studies.
- Building and analyzing a split-plot design was tricky until the wide spread availability of good DOE packages.

FTC 2017 7

Pros
Advantages of a Split-Plot Design

Pros include:

- **Practical**: Randomizing hard-to-change (HTC) factors in groups, rather than randomizing every run, is much less labor and time intensive.
- **Malleable**: Factors that naturally have large experimental units can be easily combined with factors having smaller experimental units.
- **More powerful**: Tests for the subplot effects from the easy-to-change (ETC) factors generally have higher power due to partitioning the variance sources.
- **Adaptable**: New treatments can be introduced to experiments that are already in progress.

FTC 2017 8

Agenda Transition

- Review split plot concept
- **Sizing response surface designs**
- Case Study: Building a split plot *(Corn Milling)*
- Analyzing a split plot *(Corn Milling)*
- Diagnostics *(Corn Milling)*

FTC 2017 9

Sizing a Response Surface Design

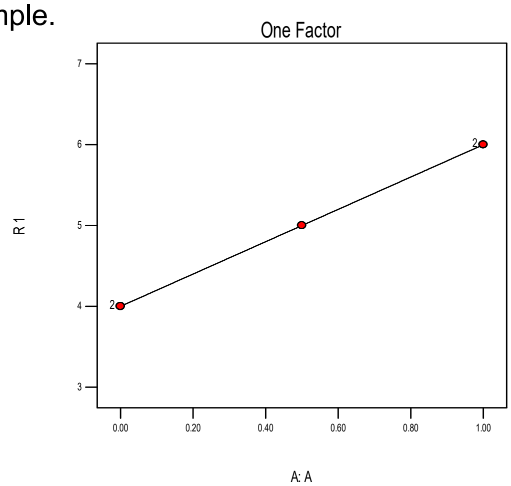
Factorial DOE	Response Surface Methods
During screening and characterization (factorials) emphasis is on identifying factor effects. What are the important design factors? For this purpose power is an ideal metric to evaluate design suitability.	When the goal is optimization (usually the case for RSM) emphasis is on the fitted surface. How well does the surface represent true behavior? For this purpose precision (FDS) is a good metric to evaluate design suitability.

Apply strategy of experimentation and DOE process all the way!

FTC 2017 10

Sizing for Precision

To illustrate sizing for precision we will start with a simple one-factor example.



FTC 2017 11

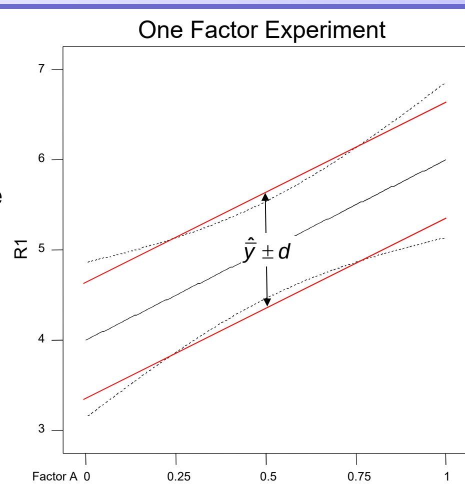
One Factor Experiment
Linear Model

The solid center line is the fitted model; $\hat{y}$ is the expected value or mean prediction.

The curved dotted lines are the computer generated confidence limits, or the actual precision.

d Is the half-width of the desired confidence interval, or the desired precision. It is used to create the outer straight lines.

Note: The actual precision of the fitted value depends on where we are predicting.



FTC 2017 12

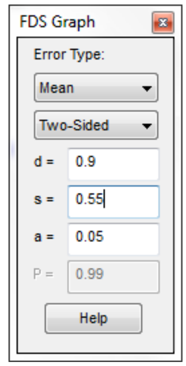
One Factor Experiment
Linear Model - Example

- Want a linear surface to represent the true response value within ± 0.90 with 95% confidence.
- The overall standard deviation for this response is 0.55.

Enter: d = 0.90

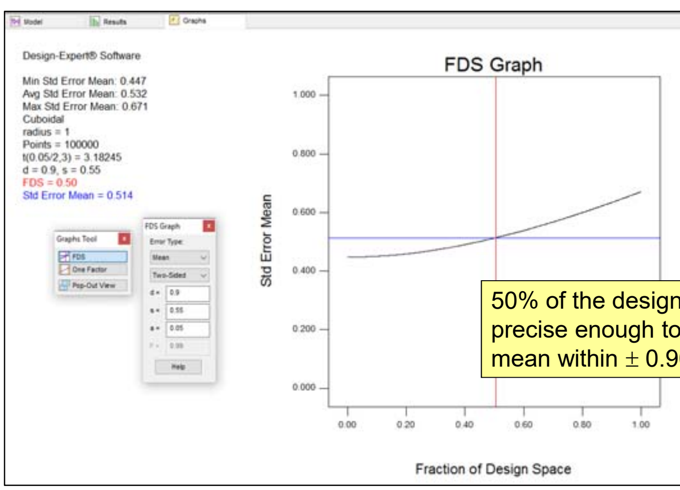
s = 0.55

a (α) = 1 – 0.95 = 0.05



FTC 2017 13

One Factor Experiment
Linear Model

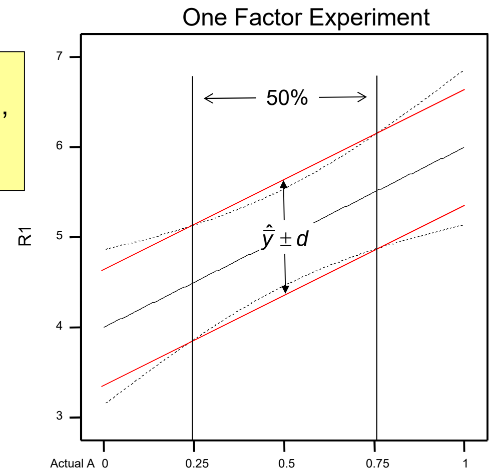


50% of the design space is precise enough to predict the mean within ± 0.90.

FTC 2017 14

One Factor Experiment
Linear Model

50% of the design space has the desired precision, i.e. is inside the solid straight (red) lines.



FTC 2017 15

FDS
Fraction of Design Space

Fraction of Design Space:

- Calculates the volume of the design space having a prediction variance (PV) less than or equal to a specified value.
- The ratio of this volume to the total volume of the design volume is the fraction of design space.
- Produces a single plot showing the cumulative fraction of the design space on the x-axis (from zero to one) versus the PV on the y-axis.

FTC 2017 16

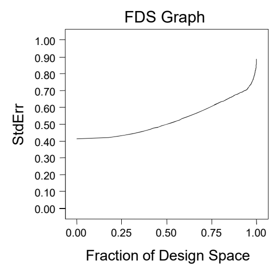
FDS Plot
Fraction of Design Space

Statistical detail

1. Pick random points (50,000) in the design space.
2. Calculate the standard error of the expected value

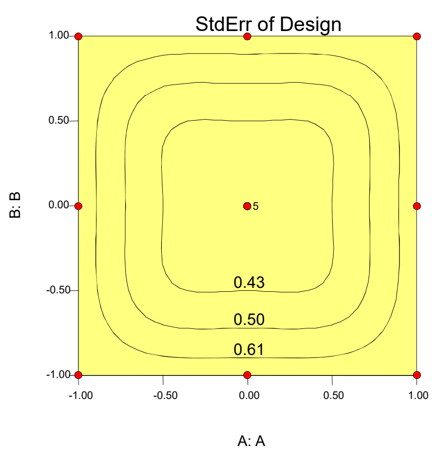
$$\text{StdErr}(x_0)$$

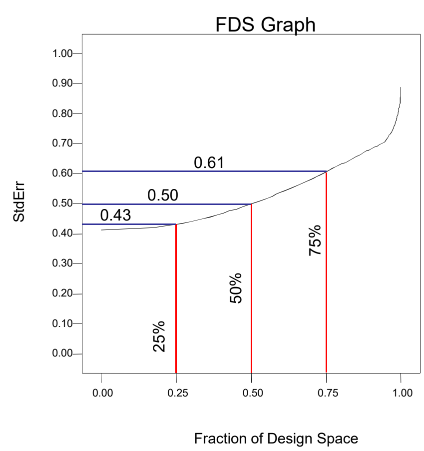
3. Plot the standard error as a fraction of the design space.



FTC 2017 17

FDS – StdErr
Two-Factor Face Centered CCD





FTC 2017 18

Sizing for Precision
FDS is Good Enough?

How good is good enough? Rules of thumb:
- For exploration want FDS ≥ 80%
- For verification want FDS of 95-100%

What can be done to improve precision?
- Manage expectations; i.e. increase d
- Decrease noise; i.e. decrease s
- Increase risk of Type I error; i.e. increase alpha
- Increase the number of runs in the design

Choose design appropriate to the problem:
- Size the design for the precision required.

FTC 2017 19

Agenda Transition

- Review split plot concept
- Sizing response surface designs
- **Case Study: Building a split plot *(Corn Milling)***
- Analyzing a split plot *(Corn Milling)*
- Diagnostics *(Corn Milling)*

FTC 2017 20

Building Split-Plot Design
By example: Corn Milling

Engineers at a corn mill want to model the effect of 4 factors on the amount of grits produced from one minute of grinding:

- **moisture** content of the feed (HTC),
- **roll gap**,
- **screen size**,
- **roller speed**.

To prepare corn for the experiment, 30 kg batches are tempered to the desired moisture content. These batches can then be split into 3 runs (sub-batches) of 10 kg each for the other (ETC) processing factor settings. These sets of 3 runs per batch become the “groups” for the design – grouping on the HTC moisture content.

FTC 2017 21

Corn Milling
Design Customization

The scientists want to study each of the 4 factors at exactly three levels each. Since there are groups (sub-batches) of 3 runs each, a design with a multiple of 3 runs would be nice (although not necessary).

Design choice: RSM Split-Plot Optimal (Custom)

Factor type: discrete (to get exactly 3 levels for each factor)

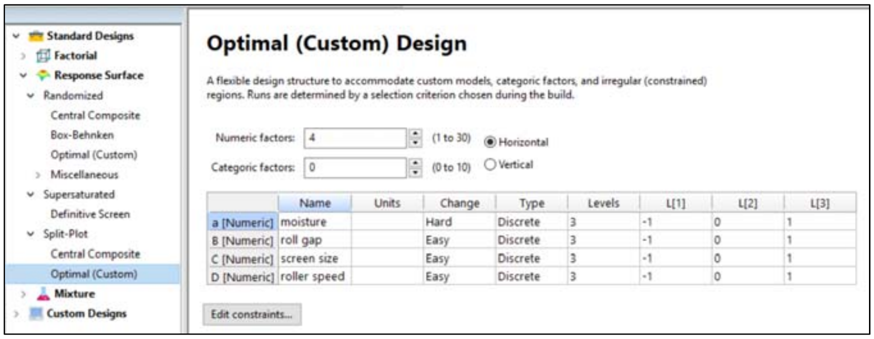
Customizing: edit number of groups and number of model points to get exactly 10 groups and 30 total runs (3 runs per batch)

Let’s build it >>>

FTC 2017 22

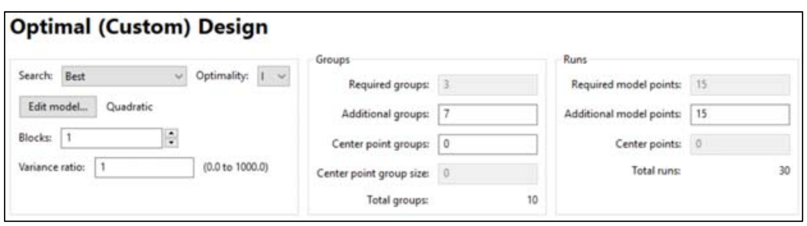
Corn Milling
Optimal Three-Level Split Plot *(page 1 of 2)*

A "**Response Surface**", "**Split-Plot**", "**Optimal (custom)**" design for four factors. Make the first factor **HTC** and make all the factor types "**Discrete**" with "**3**" levels:



Corn Milling
Optimal Three-Level Split Plot *(page 2 of 2)*

Need **10** groups and **30** runs to match Corn Milling example:



There is one response "**Yield**".

Corn Milling
Design Layout

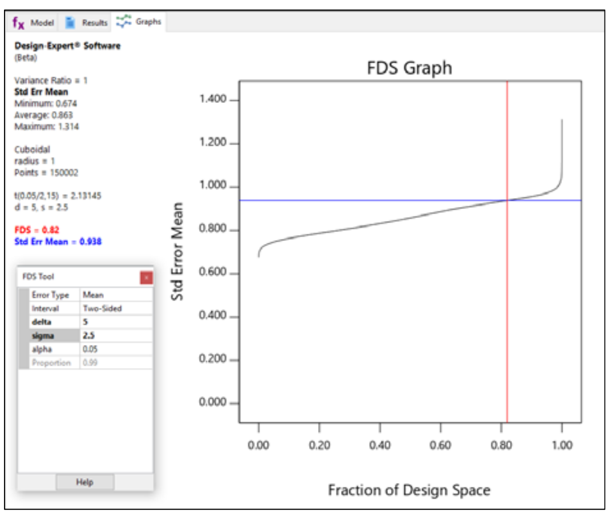
Group	Run	Factor 1 a:moisture	Factor 2 B:roll gap	Factor 3 C:screen size	Factor 4 D:roller speed
1	1	1	1	1	1
1	2	1	-1	-1	-1
1	3	1	-1	1	-1
2	4	1	1	-1	0
2	5	1	1	-1	-1
2	6	1	-1	1	0
3	7	-1	0	-1	-1
3	8	-1	0	1	0
3	9	-1	-1	0	1
4	10	0	-1	-1	0
4	11	0	0	1	1
4	12	0	-1	0	-1
			⋮		
9	25	1	1	1	-1
9	26	1	1	0	1
9	27	1	-1	-1	1
10	28	-1	-1	-1	1
10	29	-1	1	-1	-1
10	30	-1	-1	-1	-1

FTC 2017 25

Corn Milling
Check design for sample size

Precision (d) = 5.0
Std dev (s) = 2.5

FDS = 82%



FTC 2017 26

Corn Milling
with response data

Group	Run	Factor 1 a:moisture	Factor 2 B:roll gap	Factor 3 C:screen size	Factor 4 D:roller speed	Response 1 Yield
1	1	1	1	1	1	505
1	2	1	-1	-1	-1	493
1	3	1	-1	1	-1	491
2	4	1	1	-1	0	498
2	5	1	1	-1	-1	504
2	6	1	-1	1	0	500
3	7	-1	0	-1	-1	494
3	8	-1	0	1	0	498
3	9	-1	-1	0	1	498
4	10	0	-1	-1	0	496
4	11	0	0	1	1	503
4	12	0	-1	0	-1	496
			⋮			
9	25	1	1	1	-1	505
9	26	1	1	0	1	500
9	27	1	-1	-1	1	490
10	28	-1	-1	-1	1	494
10	29	-1	1	-1	-1	497
10	30	-1	-1	-1	-1	495

FTC 2017 27

Agenda Transition

- Review split plot concept
- Sizing response surface designs
- Case Study: Building a split plot *(Corn Milling)*
- **Analyzing a split plot *(Corn Milling)***
- Diagnostics *(Corn Milling)*

FTC 2017 28

Analyzing a Split-Plot Design
By example: Corn Milling

Split plot analysis:

- REML variance component estimation
- GLS parameter estimation
- Model reduction
- ANOVA – Kenward-Roger approximate F-tests
- Diagnostics for model parameter estimates (fixed)
- Diagnostics for variance components (random)

FTC 2017 29

Split Plots
Fixed and Random Effects

In a split plot design there are fixed and random effects:

$$\hat{y} = \beta_0 + \beta_1 a + \ldots + \beta_4 D + \beta_{12} aB + \ldots + \beta_{34} CD + \beta_{11} a^2 + \ldots + \beta_{44} D^2 + \sigma_\gamma^2 + \sigma_\varepsilon^2$$

- **Fixed** effects are the factor effects:
  The model coefficients (the **β**s). These effects are **fixed** by the factor levels used in the design.
- **Random** effects are the variance estimates:
  There are "group" (σ_γ^2) and "residual" (σ_ε^2) variances. These variances are sampled (i.e. they are random) while the design is run. The variances are not fixed by the design.

FTC 2017 30

Split-plot Statistical Details
Assumptions

Split-plot analysis assumptions:

- Whole-plot errors are iid as N(0, σ_γ^2)
- Sub-plot errors are iid as N(0, σ_ε^2)
- Whole-plot and sub-plot errors are mutually independent
- Randomization and factor reset are used to execute the DOE

FTC 2017 31

Corn Milling
ANOVA for Selected Model

Start with the full quadratic, reduce via Backward selection using AICc criterion and adding back terms to maintain Hierarchy:

REML (REstricted Maximum Likelihood) analysis
Kenward-Roger p-values

Source	Term df	Error df	F-value	p-value	
Whole-plot	1	22.00	10.88	0.0033	significant
a-moisture	1	22.00	10.88	0.0033	
Subplot	6	22.00	11.71	< 0.0001	significant
B-roll gap	1	22.00	3.03	0.0955	
C-screen size	1	22.00	23.80	< 0.0001	
D-roller speed	1	22.00	0.0688	0.7956	
aB	1	22.00	15.08	0.0008	
BD	1	22.00	14.13	0.0011	
CD	1	22.00	24.95	< 0.0001	

FTC 2017 32

Split Plot Background
REML

Maximum Likelihood Estimation:

The aim of maximum likelihood estimation is to find the parameter value(s) that makes the observed data most likely.

Restricted Maximum Likelihood Estimation:

A way to estimate variances. In the split plot case, REML estimates the Group variance for whole-plot factors (σ^2_γ) and the Residual variance for subplot factors (σ^2_ε).

Once the variances are estimated, Generalized Least Squares (GLS) is used to estimate the factor effects.

Then Kenward-Roger's method is used to produce F-tests.

FTC 2017 33

Corn Milling
Variance Components (random)

The variance components as estimated by REML and model summary statistics. Focus on the Adjusted R-Squared:

Source	Variance	Standard Error	95% CI Low	95% CI High
Group	0.0000	0.0000	0.0000	0.0000
Residual	6.99	2.11	4.18	14.01
Total	6.99			

Std. Dev.	2.64		**R^2**	0.7747
Mean	496.80		**Adjusted R^2**	0.6733
C.V. %	0.5323			

FTC 2017 34

Corn Milling Split Plot
Regression coefficients (fixed)

Coefficients in terms of coded factors:

Source	Cofficient Estimate	Standard Error	VIF
Intercept	496.52	0.4894	
Whole-plot Terms:			
a-moisture	1.82	0.5517	1.04
Subplot Terms:			
B-roll gap	0.9531	0.5471	1.02
C-screen size	2.68	0.5483	1.03
D-roller speed	-0.1444	0.5505	1.03
aB	2.33	0.6010	1.03
BD	-2.26	0.6015	1.03

FTC 2017 35

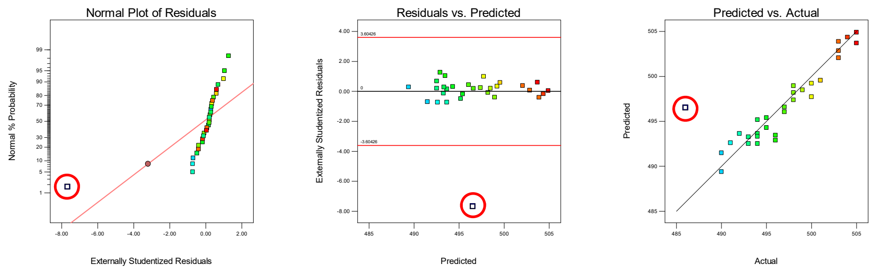
Agenda Transition

- Review split plot concept
- Sizing response surface designs
- Case Study: Building a split plot *(Corn Milling)*
- Analyzing a split plot *(Corn Milling)*
- **Diagnostics *(Corn Milling)***

FTC 2017 36

Corn Milling
An Outlier?

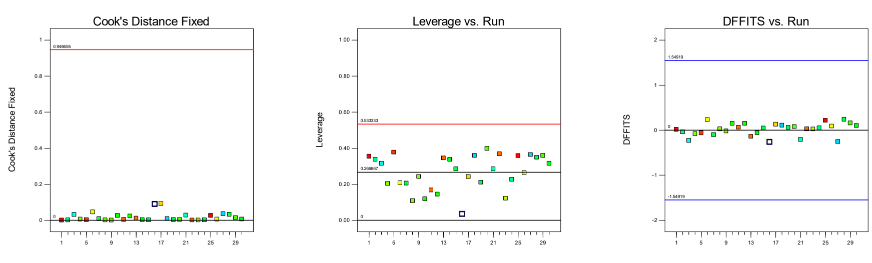
Run #16 appears discrepant with the rest of the data:



FTC 2017 37

Corn Milling
An Outlier?

Run #16 does <u>not</u> influence the <u>fixed</u> effects, that is, *the model coefficients themselves,* very much:



The "Fixed effects" Cook's Distance measures the change in betas (coefficients) as each run is removed. This is similar to the non-split-plot case.

FTC 2017 38

Corn Milling
An Outlier?

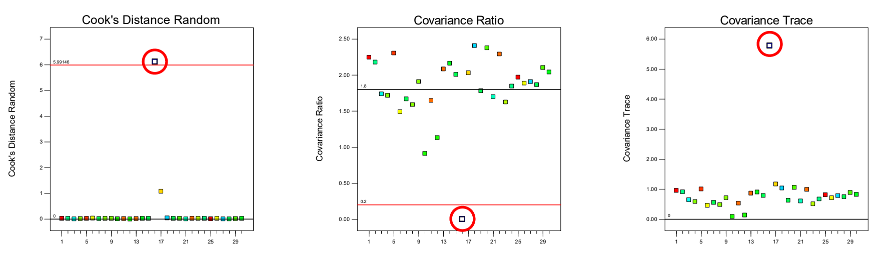
Run #16 does <u>not</u> influence the model much:

With run #16		without run #16	
Yield	=	Yield	=
+496.52		+498.29	
+1.82	* a	+1.83	* a
		-1.70	* a^2
+0.95	* B	+0.95	* B
+2.68	* C	+2.68	* C
-0.14	* D	-0.16	* D
+2.33	* aB	+2.34	* aB
-2.26	* BD	-2.35	* BD
+3.01	* CD	+3.10	* CD

FTC 2017 39

Corn Milling
An Outlier?

As indicated by these three diagnostic plots, Run #16 <u>does</u> influence the <u>random</u> effects, that is, the *standard errors* of the coefficients:



The "Random effects" Cook's Distance measures the change in the group/error variance components as each run is removed. A run that is influential on the fixed effects is not necessarily influential on the random effects and vice versa.

FTC 2017 40

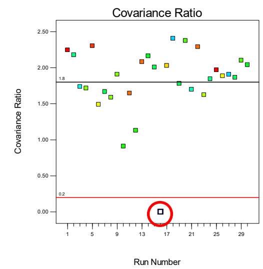
Covariance Ratio and Trace

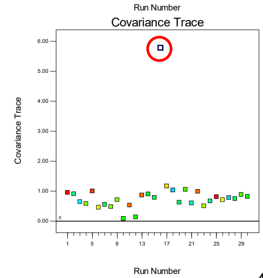
Covariance Ratio: watch for low values

1. Measures the change in precision of the estimates of the model coefficients (not the actual variance component estimates).
2. Takes into account the determinant of the variance-covariance matrix of the coefficients (similar to D-optimality).

Covariance Trace: watch for high values

1. Measures the change in precision of the estimates of the model coefficients.
2. Takes into account the trace of the variance-covariance matrix of the coefficients (similar to A-optimality).





FTC 2017 41

Corn Milling
An Outlier?

Standard errors with run #16 (left) much higher than without (right).

Source	Coef Est	Std Err
Intercept	496.52	0.49
a-moisture	1.82	**0.55**
B-roll gap	0.95	**0.55**
C-screen size	2.68	**0.55**
D-roller speed	-0.14	**0.55**
aB	2.33	**0.60**
BD	-2.26	**0.60**
CD	3.01	**0.60**

Source	Coef Est	Std Err
Intercept	498.29	0.53
a-moisture	1.83	0.25
a^2	-1.70	0.59
B-roll gap	0.95	0.24
C-screen size	2.68	0.24
D-roller speed	-0.16	0.25
aB	2.34	0.27
BD	-2.35	0.27
CD	3.10	0.27

FTC 2017 42

Corn Milling
An Outlier?

In this exercise including run # 16 primarily increases the residual variance while having little effect on the model coefficients:

with run # 16

Variance Components

Source	**Variance**
Group	*0.000*
Residual	*6.99*
Total	6.99

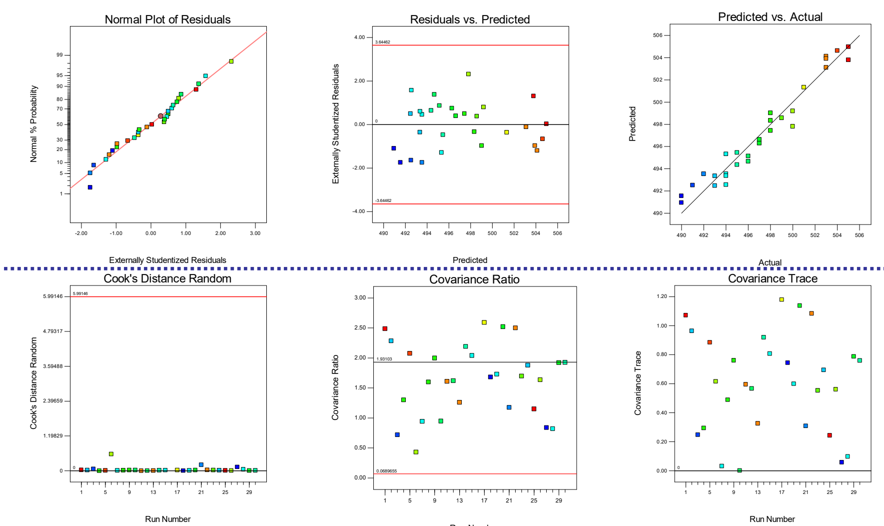
without run # 16

Variance Components

Source	**Variance**
Group	*0.000*
Residual	*1.39*
Total	1.39

FTC 2017 43

Corn Milling *(run #16 ignored)*
Top: *Fixed Effects* Bottom: *Random Effects*



FTC 2017 44

All diagnostics look good!

Corn Milling

Keep or Ignore Run #16? *(page 1 of 2)*

If we had been present while the experiments were run, then we could have:

- Investigated to see if anything unusual happened during run #16.
- Perhaps re-run run #16.

Ideally runs are ignored only because we discover an operational reason as to why the data should be excluded.

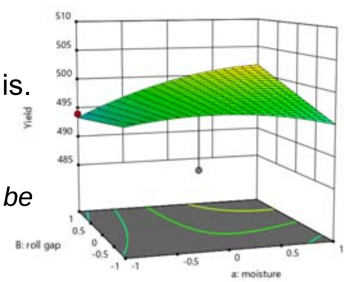
FTC 2017 45

Corn Milling

Keep or Ignore Run #16? *(page 2 of 2)*

In this case all we have is the data:

- Luckily the fixed model is robust to run #16. So it doesn't really matter if it is ignored or not.
- Run #16 has a huge effect on the variance estimates.
  - Consider what you expected the variances to be.
  - Consider how extreme the result is.
  - Look at where the point is in the DOE space. *(Since run #16 is a center point, the yield would have to be vastly different at that point.)*



FTC 2017 46

RSM Split-Plot Designs & Diagnostics

Summary

- RSM split-plot designs can be used to save time and effort.
- The restriction on randomization requires advanced techniques like REML to do a correct analysis.
- Diagnostic plots can uncover unique data problems and determine whether those problems effect the parameter estimates or the variances.

All DOE design and analysis should be driven by subject matter knowledge!

FTC 2017 47

RSM Split-Plot Designs & Diagnostics Solve Real-World Problems

Shari Kraber
Stat-Ease, Inc.
2021 E. Hennepin Ave.
Minneapolis, MN 55413
Shari@statease.com

Pat Whitcomb
Stat-Ease, Inc.
2021 E. Hennepin Ave.
Minneapolis, MN 55413

Martin Bezener
Stat-Ease, Inc.
2021 E. Hennepin Ave.
Minneapolis, MN 55413

Thank you for attending!

FTC 2017 48