Split Plot Designs:

The Good, Not so Good, and Confusing … Using JMP

Presented by Donald K. Lewis, Ph.D.
Principal, Lewis Consulting LLC

For the Willamette Valley JMP Users' Group
March 6, 2017

Presenter

Don Lewis, Ph.D., *Principal, Lewis Consulting*

email:	dlewis@consultlewis.com
phone:	(503) 244-4223
address:	P.O. Box 2282 Lake Oswego, OR 97035
web:	www.consultlewis.com

Presentation Overview

- What is a *Split Plot (S-P) Design?* Why does it matter?
- Why the S-P design is a common scenario in actual industrial experiments.
- Why DOE users are oblivious to the issue.
- What the consequences are of failure to account for the S-P structure in the design and analysis of the data.
- When the S-P should be chosen over standard designs.
- Demonstrate the use of *JMP* in the design and analysis of these experiments.
- Kick around the importance of this in real DOE applications (where statisticians are NOT around!).

What is a Split Plot Design?

The “Experimental Unit”

Every experiment has an *experimental unit*, that unit which receives a complete repetition of the experimental inputs or “treatments” during the conduct of the experiment.

- It is crucial to identify that unit so that the sample size (# of units) is sufficient to minimize decision-making errors due to unit-to-unit variance in the response (*alpha* and *beta errors)*.
- But, a question: *Might it be possible that an experiment might have more than one unique “experimental unit”?*
- The answer: not just *possibly* but *frequently!*

Since that is the case, *what is the implication of having multiple experimental units on the design of the experiment and the analysis of the data?*

The "Rubber Band" Shooting Experiment

Assume that an experimenter desires to optimize a rubber band shooting process. He/she must consider how many experimental units (bands) to shoot, n, in order to detect a change (∂) in the mean *flight distance*. Let's say that the factor of interest is the *stretch amount* of the band.

But, any single rubber band can be stretched (and shot) more than one time (although not until it weakens). Therefore, *stretch distance* can be varied within the same rubber band.

An obvious question: *Would the experiment design not improve if the change in the stretch amount occurred within the same band, thus avoiding band-to-band variation in flight distance?*

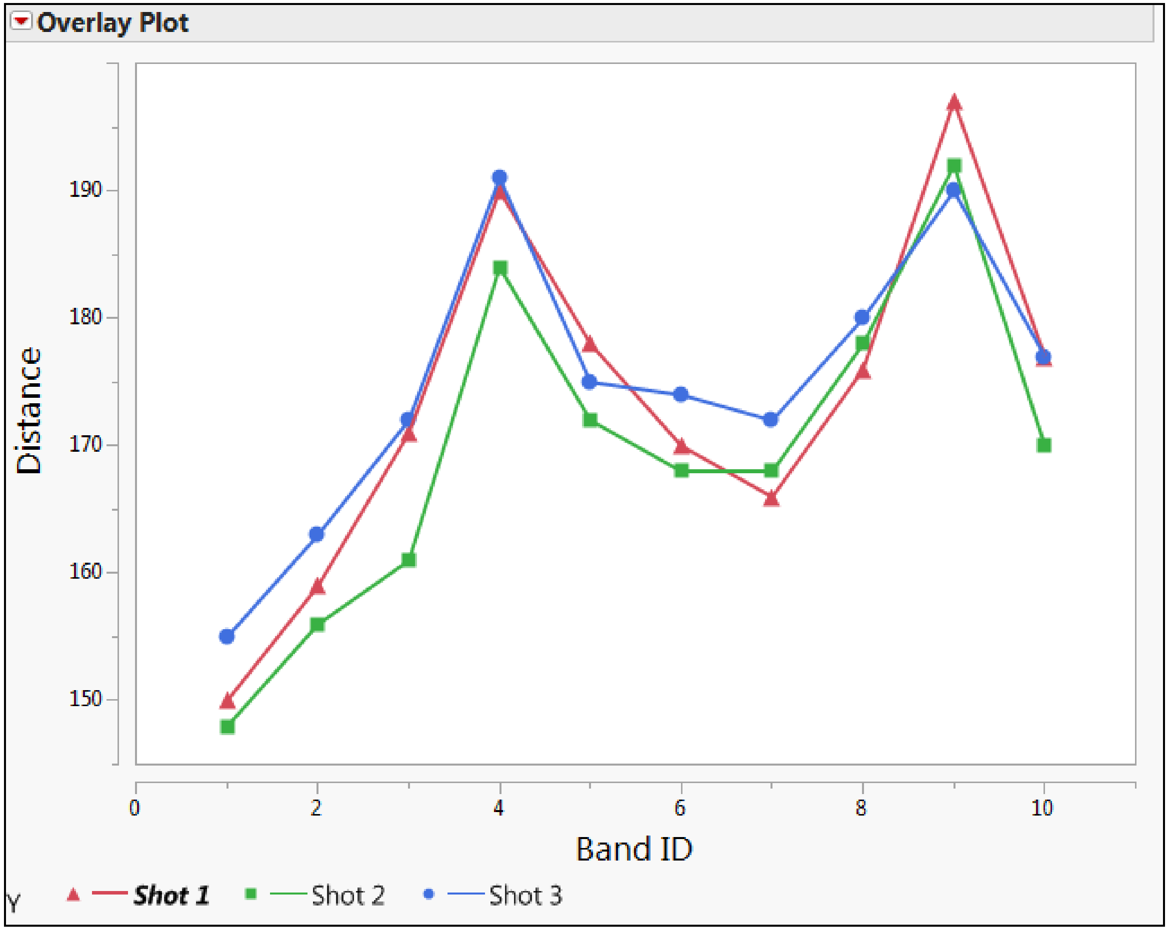
Rubber Band Shooting Data

The data to the right represent the flight distance (in inches) of 3 repeated shots of 10 rubber bands selected from a common bag of bands. They were stretched 5.5 inches and shot horizontally at the waist height of the shooter.

What does it suggest about variability?

Band	Shot 1	Shot 2	Shot 3
1	150	148	155
2	159	156	163
3	171	161	172
4	190	184	191
5	178	172	175
6	170	168	174
7	166	168	172
8	176	178	180
9	197	192	190
10	177	170	177

Analysis of Rubber Band Shooting Data



Clearly, there is far less variability in distance <u>between</u> shots of the same rubber band than across shots of <u>different</u> rubber bands. *What are the values of the variances?*

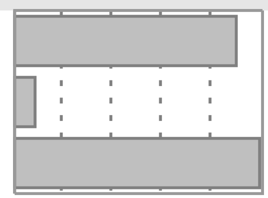
Components of Variance Analysis

Analysis of Variance

Source	DF	SS	Mean Square	F Ratio	Prob > F
Band	9	4157.333	461.926	33.5539	<.0001 *
Within	20	275.3333	13.7667		
Total	29	4432.667	152.851		

Variance Components

Component	Var Component	% of Total	20406080	Sqrt(Var Comp)
Band	149.38642	91.6		12.222
Within	13.76667	8.4		3.710
Total	163.15309	100.0		12.773



In fact, the estimated band-to-band variance is 149.4, while the shot-to-shot variance is 13.8. The required # of shots diminishes by a factor of $V_{BAND+SHOT} / V_{SHOT} = 163/13.8 = 12$! (Note: this is the basis of the advantage of the *paired design.)*

Experimental Units in Multi-factor DOE

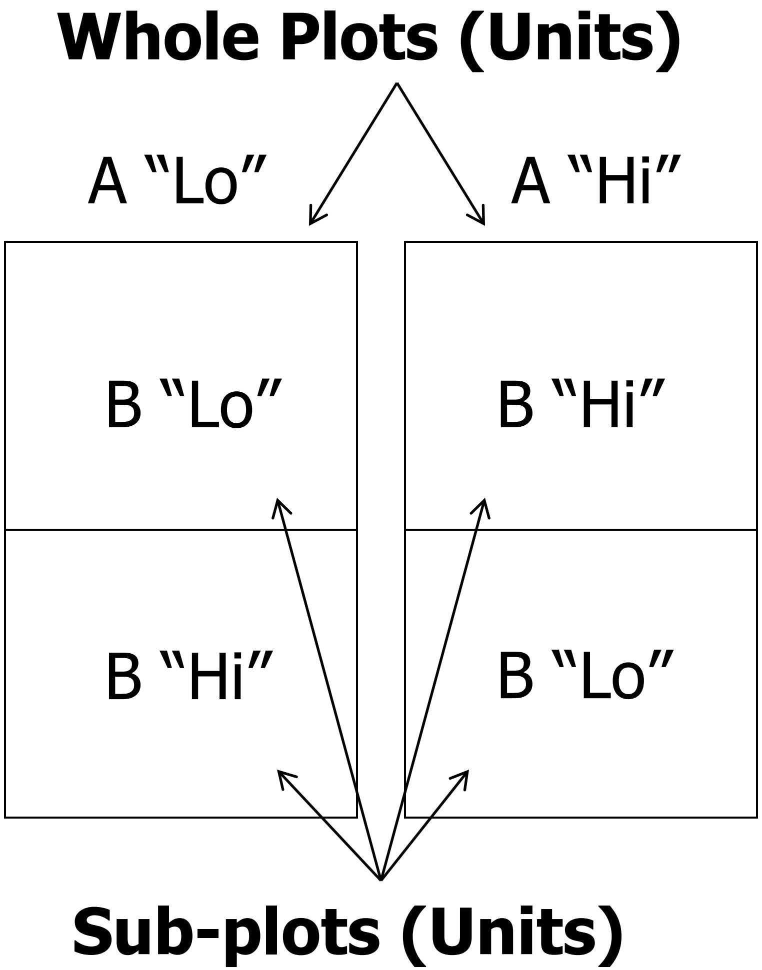
When multiple factors are varied in the same experiment, it is common that some can be varied within a particular unit, but the remainder must be varied across different units. Planning the best way to design / conduct the experiment becomes a challenge (as does the data analysis)!

Example: The *stretch distance* can be varied within the same rubber band, while *band elasticity* must vary band to band.

When this occurs, either by design or by happenstance, we have what is called a *Split Plot Design*.

Note: If "Band Elasticity" is a hard-to-change factor, this design structure is particularly appealing .

So What is a "Split-Plot" Design?



A *Split Plot Design* is a design where there is more than one type of experimental unit.

For some of the factors, changes in factor levels occur across *whole units (plots)*, while the remaining factors are changed across *sub-units* within the whole plots (more commonly called *sub-plots*.)

Examples of Split Plot Design Factors

Process	Whole Unit Factors	Split Unit Factors
Band shooting	Rubber bands (Band size)	Shots (Stretch Distance)
Cake baking	Oven runs (Temperature)	Oven positions (Cake recipes)
Plasma etching	Chamber runs (Vacuum press.)	Chamber positions (Substrate type)
Agriculture	Land plots (Aerial spray method)	Mini-plots (Plant spacing)
Engine design	Engines (Design type)	Engine runs (fuel type)

In each cell is the experimental unit for the factor in parentheses in that cell. For example, in an experiment to study the baking of a cake, *Temperature* of the oven must be varied across *Oven runs*, while ingredients involved in the *Cake recipe* can be varied across *Oven positions* within the same oven run.

Types of Factors in Experiments

There are two types of factors in industrial experiments: those that are *Hard-to-Change* and those that are *Easy-to-Change*.

<u>Hard-to-Change</u>: factors whose changes in levels are challenging

- Oven temperature (annealing)
- Chamber pressure (diffusion)
- Slurry type (polishing)
- Bath concentration (plating)

<u>Easy-to-Change</u>: factors that can be changed without consequence

- Anneal time (annealing)
- Wafer type (diffusion)
- Plate RPM (polishing)
- Voltage (plating)

Combining H-to-C & E-to-C Factors in Factorial's

When combining the two types of factors in any factorial arrangement a natural way to do so is to vary the E-to-C factors within levels of the H-to-C factors. Of course, when this occurs, the design has a S-P structure (which is beneficial for decision-making on the E-to-C factors.) However, the assumption behind full and fractional-factorial designs is a *completely randomized* design.

What does that mean? That the experimental units have been completely randomized in terms of assignment to the treatment settings and that each treatment combination (and any replicates) are subject to the same experimental error. An upshot of this assumption is that many factorial designs are not optimal, both in terms of confounding and replication.

The "Completely Randomized" vs. the S-P Design

An "off-the-shelf" factorial type experiment is assumed to be a *completely randomized design*. That means that the experimental units (of only one type) are randomly assigned to the treatment combinations (t.c.'s) of the factors (and any replicates).

- There is no restriction to the order in which the t.c.'s are conducted ("randomization").
- Each response is captured from the same type of experimental unit; each response is subject to the same variance.
- The statistical analysis / modeling of the data is straight-forward. JMP's "Fit Model" default routine is followed.
- Note: The only restriction to the randomization that might occur is when the experiment is *blocked*, but JMP by default randomizes within the blocks and the data analysis assumes that the responses within the blocks are subject to the same within-block variance.

Challenge #1 of S-P Experiments: Design

The first challenge presented by the S-P alternative to the CRD is when they should be used. The answer usually depends upon:

(1) The cost and convenience of changing the factors during the conduct of the experiment, and

(2) The increased precision / power (reduced # of experimental units) provided by the S-P design.

- We have already considered (2). Clearly, if factors can be varied within the whole unit, they should be. Then, for the factors varied across the whole units, enough repetitions of the whole plots should be made to achieve the desired power.
- What is probably more common is that the S-P design is inadvertent. A CRD design is chosen (and created in software like *JMP*), but the actual conducted experiment is a S-P design.

How Does an Inadvertent S-P Design Occur?
The Experimenter Does the Following…

1. Desires to conduct factorial design (say, a 2^{K-P} design)

2. Goes to *JMP's* "Screening Design" procedure to generate it.

3. ***But that is a CRD design (or randomized block design)!***

4. Scrutinizes the design in the *JMP* table and sorts it by the H-to-C factor(s).

5. Conducts the experiment in that order.

What is wrong with that?

Common Problems with the Inadvertent S-P Experiment

- Large differences in standard errors between whole plot effects and split-plot effects.
- Way too few whole plot units are in the design (often, $n_{LO} = n_{HI} = 1$!)
- #2 is not recognized by the experimenter and the *JMP* table is filled with many rows of whole plot units.
- #1 leads to mixing of the whole plot and split plot variances.
- Inactive effects are wrongly labeled as statistically significant or vice versa.

Example: Standard 2^{4-1} Res. IV Design
A, B: Hard to Change / C, D: Easy to Change

Design Matrix

A	B	C	D
-1	-1	-1	-1
1	-1	-1	1
-1	1	-1	1
1	1	-1	-1
-1	-1	1	1
1	-1	1	-1
-1	1	1	-1
1	1	1	1

Note that at each of the four t.c.'s involving A and B, there are only two t.c.'s involving C and D. If there is no serious limit to the number of "split units" within each whole unit that can be run, *why not conduct as many t.c.'s involving the factors, C and D, as possible?* That design would involve 4 whole plots at A and B and 4 sub-plots within each of the whole plots.

Example: Conduct of 2-Level Design
A, B: Hard to Change / C, D: Easy to Change

<table>
<tr><th>A</th><th>B</th><th>C</th><th>D</th></tr>
<tr><td rowspan="4">-1</td><td rowspan="4">-1</td><td>-1</td><td>-1</td></tr>
<tr><td>-1</td><td>1</td></tr>
<tr><td>1</td><td>-1</td></tr>
<tr><td>1</td><td>1</td></tr>
<tr><td rowspan="4">-1</td><td rowspan="4">1</td><td>-1</td><td>-1</td></tr>
<tr><td>-1</td><td>1</td></tr>
<tr><td>1</td><td>-1</td></tr>
<tr><td>1</td><td>1</td></tr>
<tr><td rowspan="4">1</td><td rowspan="4">-1</td><td>-1</td><td>-1</td></tr>
<tr><td>-1</td><td>1</td></tr>
<tr><td>1</td><td>-1</td></tr>
<tr><td>1</td><td>1</td></tr>
<tr><td rowspan="4">1</td><td rowspan="4">1</td><td>-1</td><td>-1</td></tr>
<tr><td>-1</td><td>1</td></tr>
<tr><td>1</td><td>-1</td></tr>
<tr><td>1</td><td>1</td></tr>
</table>

Assume that it would be convenient to run each of the four C-D t.c.'s at each of the four A-B t.c.'s. Then the design would appear (in matrix form) to the left. It is of course now a full factorial.

Since *JMP* needs a value for each factor in each row of the table, the user would likely enter the design as a CRD 2^4 full factorial.

What is wrong with that?

Example: A 2^K Factorial with a S-P Structure

2^5 Full Factorial; A, B, C, D are H-to-C; E is E-to-C

Objective: Make paper more susceptible to ink ("wettability") through plasma processing.

Response: Contact angle between water droplet and paper surface.

Factors:

- A: Pressure
- B: Power
- C: Flow Rate
- D: Gas Type
- E: Paper Type

Design: Split plot; A, B, C, D varied over reactor runs; E varied within reactor runs

Data taken from Bisgaard, S. *Journal of Quality Technology*, Vol. 32, No. 1, 2000.

2^5 Factorial with Split Plot Arrangement

				E			
A	B	C	D	-	+	Average	Delta
-	-	-	-	48.6	57.0	52.8	8.4
+	-	-	-	41.2	38.2	39.7	-3.0
-	+	-	-	55.8	62.9	59.4	7.1
+	+	-	-	53.5	51.3	52.4	-2.2
-	-	+	-	37.6	43.5	40.6	5.9
+	-	+	-	47.2	44.8	46.0	-2.4
-	+	+	-	47.2	54.6	50.9	7.4
+	+	+	-	48.7	44.4	46.6	-4.3
-	-	-	+	5.0	18.1	11.6	13.1
+	-	-	+	56.8	56.2	56.5	-0.6
-	+	-	+	25.6	33.0	29.3	7.4
+	+	-	+	41.8	37.8	39.8	-4.0
-	-	+	+	13.3	23.7	18.5	10.4
+	-	+	+	47.5	43.2	45.4	-4.3
-	+	+	+	11.3	23.9	17.6	12.6
+	+	+	+	49.5	48.2	48.9	-1.3

Use of *JMP's* "Custom Design" Menu for S-P's

JMP is capable of providing a "one stop shopping" design and analysis of a S-P experiment. Using "DOE > Custom Design" the key steps involve:

1. The "Changes" field in the initial introduction of the Factors. By using "Hard" *JMP* is directed to create whole plots for those factor combinations. By using "Easy" *JMP* varies those factors in sub-plots.
2. The experimenter must inform *JMP* which effects beyond the main effects are desired from the experiment. That input will determine the extent of the confounding in the design.
3. The experimenter must also inform *JMP* how many whole plots (total) will be included in the experiment.

JMP's Custom Design: Factor "Changes"

Custom Design

Responses

Factors

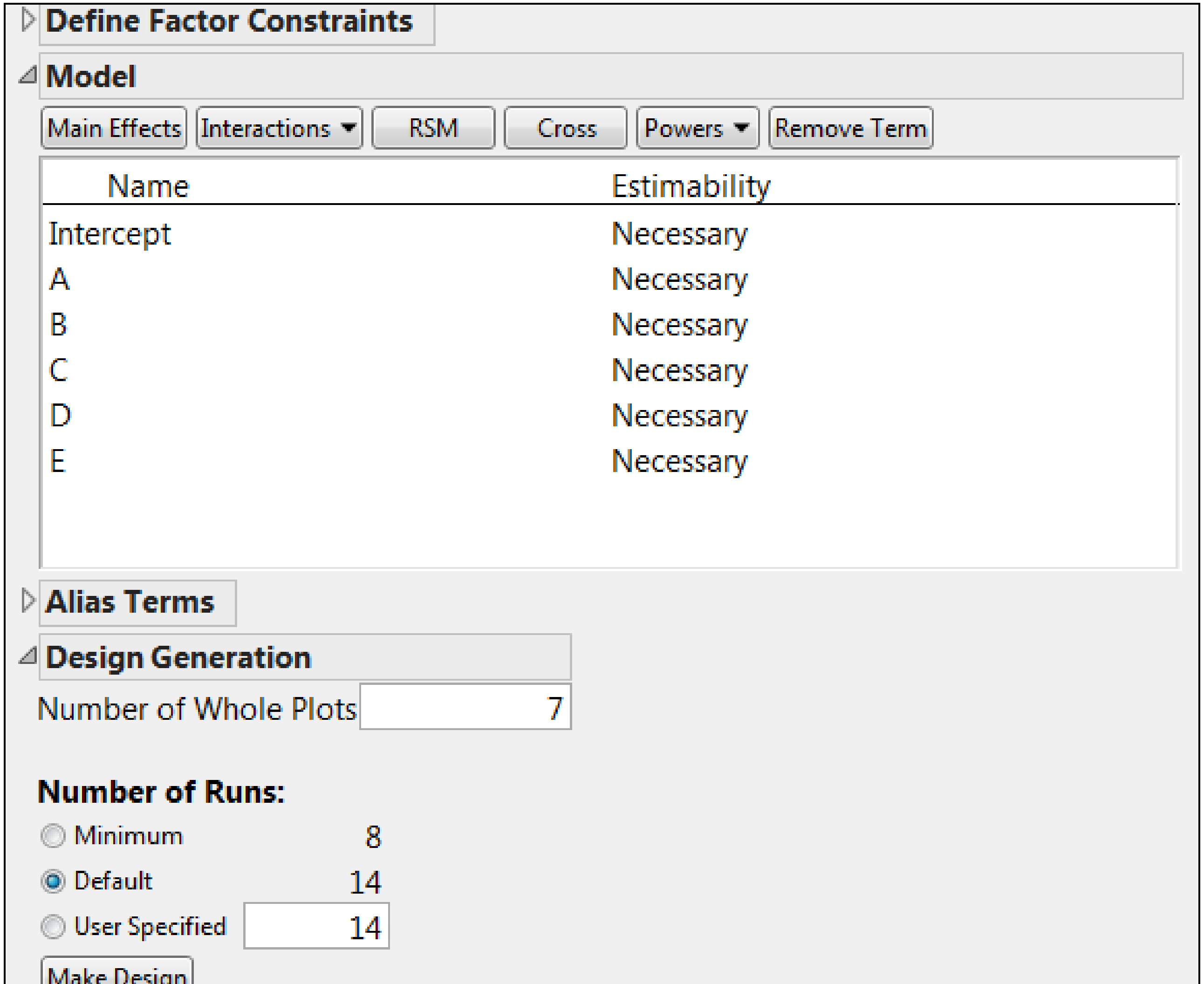
Add Factor | Remove | Add N Factors 1

Name	Role	Changes	Values	
A	Continuous	Hard	-1	1
B	Continuous	Hard	-1	1
C	Continuous	Hard	-1	1
D	Continuous	Hard	-1	1
E	Continuous	Easy	-1	1

Define Factor Constraints

With this designation the factors A, B, C, and D will define the whole-plot t.c.'s. E will define the sub-plot levels within the whole plots.

JMP's Custom Design: Model Terms



The default model is a main effects only model. The *JMP* user must add in any additional terms desired to be estimated from the S-P experiment.

The number of whole plots must also be specified.

Example: Plasma Processing Full Factorial

Factors

Add N Factors 1

Nam	Role	Change	Value
A	Discrete Numeri	Hard	-1 1
B	Discrete Numeri	Hard	-1 1
C	Discrete Numeri	Hard	-1 1
D	Discrete Numeri	Hard	-1 1
E	Discrete Numeri	Easy	-1 1

Design Generation

Number of Whole Plots 16

Number of Runs:

○ Minimum	16
◉ Default	32
○ User Specified	32

Model

Name	Estimabilit
Intercep	Necessary
A	Necessary
B	Necessary
C	Necessary
D	Necessary
E	Necessary
A*B	Necessary
A*C	Necessary
A*D	Necessary
A*E	Necessary
B*C	Necessary
B*D	Necessary
B*E	Necessary
C*D	Necessary
C*E	Necessary
D*E	Necessary

JMP's Custom Design Design Generation

Custom Design

Design

Run	Whole Plots	A	B	C	D	E
1	1	1	-1	-1	1	1
2	1	1	-1	-1	1	-1
3	2	1	-1	1	-1	1
4	2	1	-1	1	-1	-1
5	3	1	1	1	-1	1
6	3	1	1	1	-1	-1
7	4	-1	1	1	1	-1
8	4	-1	1	1	1	1
9	5	-1	-1	1	-1	-1
10	5	-1	-1	1	-1	1
11	6	1	1	-1	-1	1
12	6	1	1	-1	-1	-1

Listed to the left are the first 12 of the 32 runs in the design.

JMP describes the S-P structure in this table in the "Whole Plot" column.

Check to make sure that is the case!

Challenge #2 of S-P Experiments: Data Analysis

The analysis of the data from a S-P design can be a challenge, even with DOE software like *JMP*. If the DOE user is not aware of the true S-P structure of the experiment, the analysis will almost surely be wrong.

Why? Remember that the whole unit (plot) factors are subject to one response variance (actually two: between unit and within unit variance), while the sub-unit factors are subject only to the sub-unit response variance.

As we saw from the rubber band data, *each variance is likely to be quite different in size from the other!*

Analysis Method #1:

1. For each whole plot t.c., list all factor levels and sub-plot responses in one row. Average the responses.

2. Calculate all effects for these whole plot factors and determine the significant / active effects.
   - Use *JMP's* "Fit Model" routine, as usual
   - Plot coefficients on normal plot
   - Calculate p-values, as usual

3. At each whole plot t.c., compute the various sub-plot effects involving the responses at the sub-plot t.c.'s.

4. Using *JMP's* "Fit Model" determine the significant effects.
   - Same analysis as in #2 above.

2^5 Factorial with Split Plot Arrangement

A	B	C	D	E: -	E: +	Average	Delta
-	-	-	-	48.6	57.0	52.8	8.4
+	-	-	-	41.2	38.2	39.7	-3.0
-	+	-	-	55.8	62.9	59.4	7.1
+	+	-	-	53.5	51.3	52.4	-2.2
-	-	+	-	37.6	43.5	40.6	5.9
+	-	+	-	47.2	44.8	46.0	-2.4
-	+	+	-	47.2	54.6	50.9	7.4
+	+	+	-	48.7	44.4	46.6	-4.3
-	-	-	+	5.0	18.1	11.6	13.1
+	-	-	+	56.8	56.2	56.5	-0.6
-	+	-	+	25.6	33.0	29.3	7.4
+	+	-	+	41.8	37.8	39.8	-4.0
-	-	+	+	13.3	23.7	18.5	10.4
+	-	+	+	47.5	43.2	45.4	-4.3
-	+	+	+	11.3	23.9	17.6	12.6
+	+	+	+	49.5	48.2	48.9	-1.3

Analysis of Plasma Experiment: Whole Plot Effects

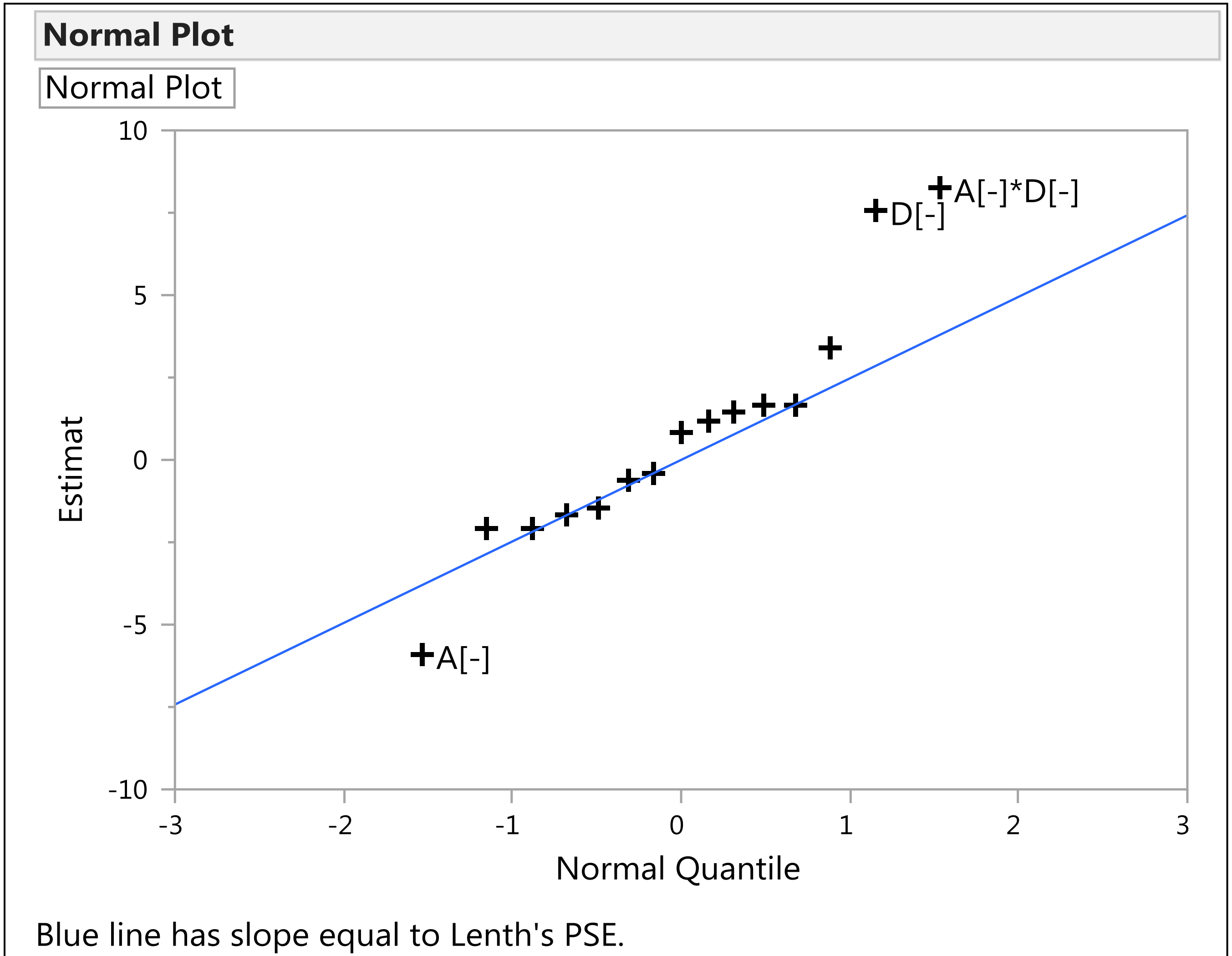
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercep	40.98125	1.908998	21.47	<.0001 *
A[-]	-5.9125	1.908998	-3.10	0.0269 *
B[-]	-2.1125	1.908998	-1.11	0.3188
C[-]	1.69375	1.908998	0.89	0.4156
D[-]	7.55	1.908998	3.95	0.0108 *
A[-]*B[-	-2.10625	1.908998	-1.10	0.3201
A[-]*C[-	1.4875	1.908998	0.78	0.4711
A[-]*D[-	8.28125	1.908998	4.34	0.0074 *
B[-]*C[-]	-0.425	1.908998	-0.22	0.8326
B[-]*D[-	-1.65625	1.908998	-0.87	0.4253
C[-]*D[-	0.8375	1.908998	0.44	0.6792

P-values are based upon an estimate of variance from higher-order effects: ABC, ABD, ACD, etc.

A, D, and AD are significant.

Analysis of Plasma Experiment: Whole-plot Factors



A, D, and AD are significant.

Note: The model needs to include all of the higher-order effects for this plot to be most useful.

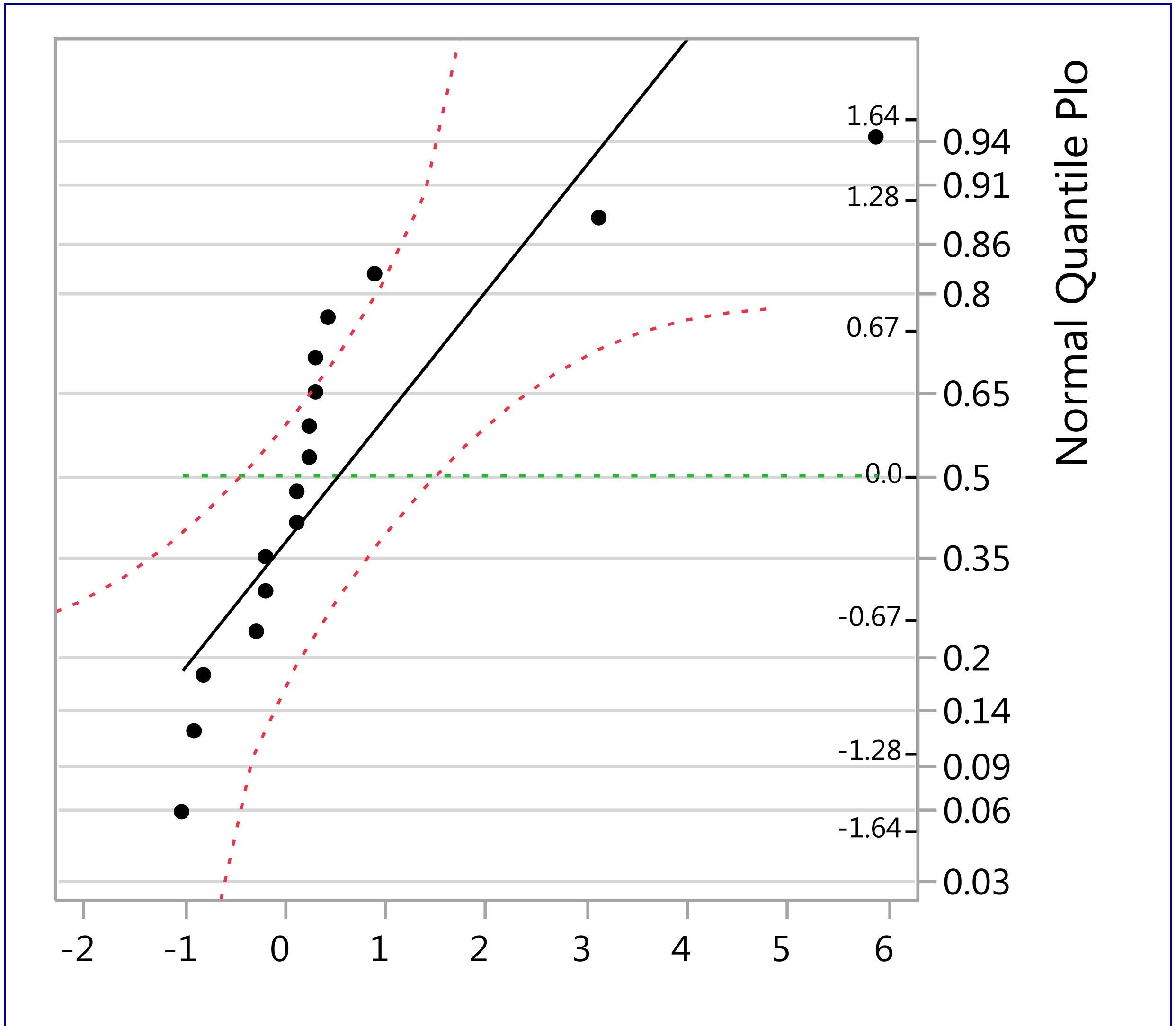
Analysis of Plasma Experiment: Sub-plot Factors

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercep	3.1375	0.508144	6.17	<.0001 *
A[-]	5.9	0.508144	11.61	<.0001 *
B[-]	0.3	0.508144	0.59	0.5669
C[-]	0.1375	0.508144	0.27	0.7917
D[-]	-1.025	0.508144	-2.02	0.0688

It is important to realize that the "Intercept" coefficient represents the E main effect. *Why?* Because the data being analyzed are the response differences from E-hi to E-lo. The remaining main effects actually estimate the interactions of the whole plot factors with E. The p-values are based upon an estimate of variance from interactions involving A, B, C, and D. These actually represent the higher-order effects: ABE, ACE, ADE, etc. Only E and AE are significant.

Analysis of Plasma Experiment: Sub-plot Effects



Note: To make this plot, save all of the coefficients to a *JMP* table and use *JMP's* "Analyze > Distribution" module to create a normal quantile plot.

The two extreme points represent the E and AE effects.

ANOVA Tables from the Whole Plot Model: *What Does It Demonstrate?*

Summary of Fit

RSquare	0.81755
RSquare Adj	0.771938
Root Mean Square Error	6.911532
Mean of Response	40.98125
Observations (or Sum Wgts	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	2568.6281	856.209	17.9239
Error	12	573.2312	47.769	**Prob > F**
C. Tota	15	3141.8594		<.0001 *

ANOVA Tables from the Sub-plot Model:
What Does It Demonstrate?

Summary of Fit

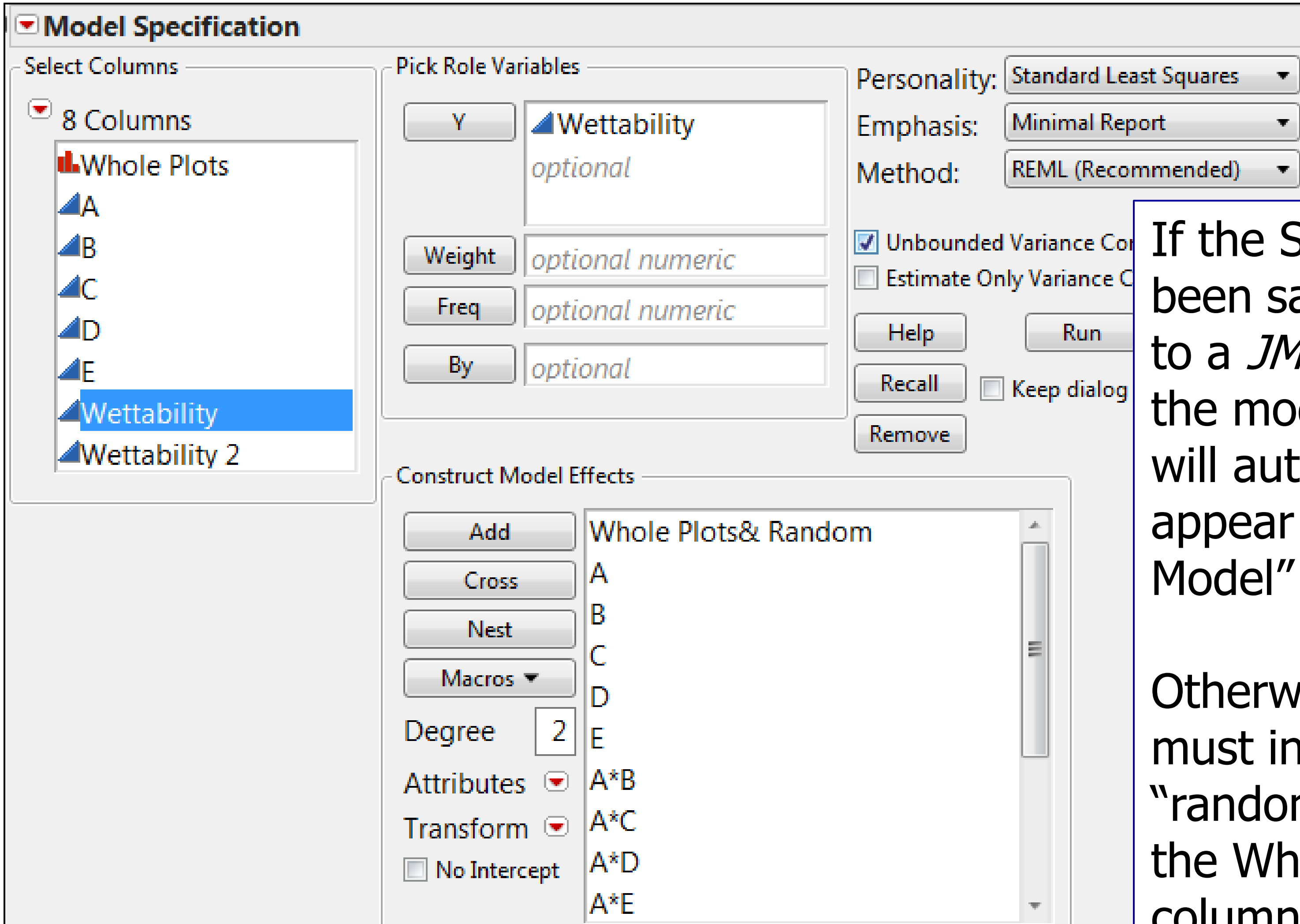
RSquare	0.896937
RSquare Adj	0.889576
Root Mean Square Error	2.138048
Mean of Response	3.1375
Observations (or Sum Wgts	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	556.96000	556.960	121.8398
Error	14	63.99750	4.571	**Prob > F**
C. Tota	15	620.95750		<.0001 *

The RMS error for the differences is 2.1 vs. 6.9 for the averages. The required sample size is thus ***11 times smaller!***

Direct S-P Model in JMP: Model Specification



If the S-P design has been saved directly to a *JMP* table, then the model to the left will automatically appear in the "Fit Model" routine.

Otherwise, the model must include a "random" term for the Whole Plots column.

Example: Split Plot Model of Plasma Data

Parameter Estimates

Term	Estimate	Std Error	DFDen	t Ratio	Prob>|t|
Intercep	40.98125	1.908998	5	21.47	<.0001 *
A	5.9125	1.908998	5	3.10	0.0269 *
B	2.1125	1.908998	5	1.11	0.3188
C	-1.69375	1.908998	5	-0.89	0.4156
D	-7.55	1.908998	5	-3.95	0.0108 *
E	1.56875	0.254072	11	6.17	<.0001 *
A*B	-2.10625	1.908998	5	-1.10	0.3201
A*C	1.4875	1.908998	5	0.78	0.4711
A*D	8.28125	1.908998	5	4.34	0.0074 *
A*E	-2.95	0.254072	11	-11.61	<.0001 *
B*C	-0.425	1.908998	5	-0.22	0.8326
B*D	-1.65625	1.908998	5	-0.87	0.4253
B*E	-0.15	0.254072	11	-0.59	0.5669
C*D	0.8375	1.908998	5	0.44	0.6792
C*E	-0.06875	0.254072	11	-0.27	0.7917
D*E	0.5125	0.254072	11	2.02	0.0688

Note: The p-values are exactly the same as those from the Method #1 analyses. However, the coefficients and Std Error's for the sub-plot effects are ½ the size in this table. That is due to the fact that this analysis essentially divides the differences by two.

The A, D, E, AD, and AE effects remain the only significant ones. DE is on the edge.

Summary: *Split Plot Designs*

In this presentation we have discussed:

1. The *Split Plot Design* involves more than one experimental unit in the same experiment. The *completely randomized design* is the default design in the minds of most experimenters.
2. Often the only reasonable strategy for conducting an industrial experiment, given cost and convenience constraints, is the *Split Plot structure.*
3. The benefit of the S-P design is the increased power associated with decisions about the split unit factors.
4. A drawback of the S-P design may be the insufficient number of whole plots in the experiment (and poor power as a consequence) for decisions about the whole plot factors.
5. Analysis of the data from a S-P design can be tricky. Use *JMP's* "Custom Design" to design the experiment and analyze the data using "Fit Model" with a "Whole Plots" term in the model.