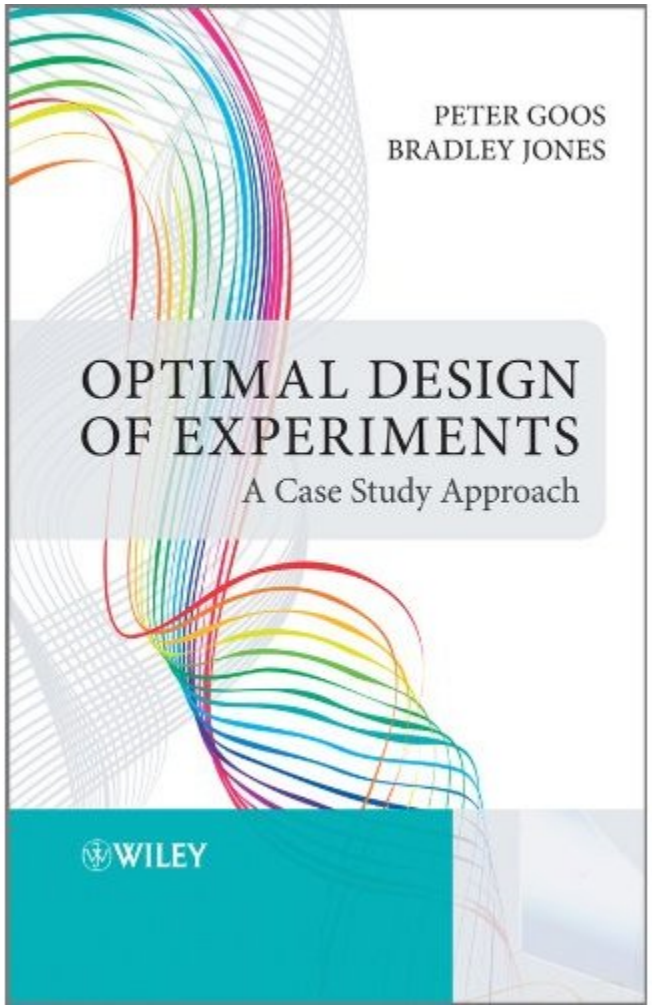
Modern Design of Experiments for Rapid and Active Learning

Bradley Jones
Distinguished Research Fellow
JMP Division/SAS

Outline

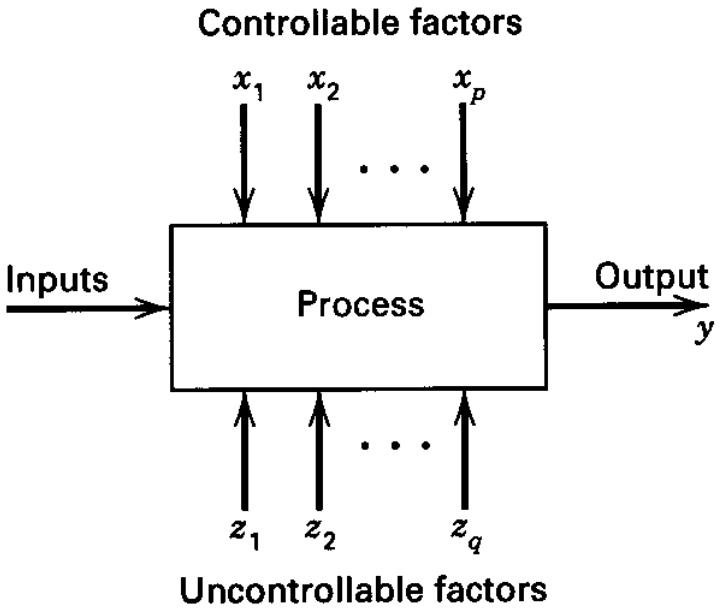
1. Introduction
2. Textbook Designs – restrictions and limitations
3. Examples of Computer Generated Designs
    a. When the region of interest is not regular
    b. When design runs need to be run in groups – blocking
    c. When characteristics of the experimental material are known in advance

Most examples are based on this book.



What Is a Designed Experiment?

a structured set of tests of a system or process



Fundamental to designed experiments are…

1. Response(s) - outputs
2. Factor(s) - inputs
3. Mathematical Model

Why design experiments?

1. To establish causal relationships between inputs & outputs
2. To use this understanding to match desired outputs
3. To do the above efficiently & inexpensively

How do you start doing designed experiments?

The most frequently used industrial experiment is a factor screening experiment.

Screening experiments are a useful starting point.

What is a screening experiment?

… screening experiments – [are] experiments in which many factors are considered and the objective is to identify those factors (if any) that have large effects.

Montgomery, D.C. (2005) Design and Analysis of Experiments, Wiley, New York, page 283.

A textbook screening design

Run	Methanol x_1	Ethanol x_2	Propanol x_3	Butanol x_4	pH x_5	Time x_6	Yield (mg)
1	0	0	0	10	6	1	10.94
2	0	10	0	0	9	1	15.79
3	0	10	0	10	9	2	25.96
4	10	10	10	0	6	1	35.92
5	0	0	10	0	6	2	22.92
6	0	10	10	10	6	1	23.54
7	10	10	0	0	6	2	47.44
8	10	0	0	0	9	1	19.80
9	10	0	10	10	9	1	29.48
10	0	0	10	0	9	2	17.13
11	10	10	10	10	9	2	43.75
12	10	0	0	10	6	2	40.86

Bie, Xiaomei, et. al. “Screening the main factors affecting extraction of the antimicrobial substance from Bacillus sp. fmbJ using the Plackett–Burman method” World Journal of Microbiology & Biotechnology (2005) 21: 925–928 Springer 2005

Implicit Restrictive Assumptions

1. Factor effects are linear

   Why? – Every factor has only two settings

2. All factor combinations are feasible

   Why? – All pairs of low and high factor settings appear

More textbook design limitations

1. Designs are prespecified – the number of required runs is fixed
2. Applicable models are limited
3. Lack of flexibility in numbers of runs in a group
4. Lack of flexibility in the types of factors
5. Lack of flexibility in dealing with hard-to-change factors
6. No ability to use knowledge about the experimental material
7. Lack of flexibility for replicating runs to estimate error variance

Consequence of Textbook Design Limitation

Experimenters simplify the problem description to accommodate the use of a prespecified design.

Preferable alternative to textbook design

- use computer algorithms to construct designs that match all the problem specific requirements

“Design to fit the problem – Don’t change the problem to fit the design”

Software Demonstration

All Main Effects Model

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	27.79	0.71	39.0	<.0001*
Methanol(0,10)	8.41	0.71	11.8	<.0001*
Ethanol(0,10)	4.27	0.71	6.0	0.0019*
Propanol(0,10)	1.00	0.71	1.4	0.2214
Butanol(0,10)	1.29	0.71	1.8	0.1292
pH(6,9)	-2.48	0.71	-3.5	0.0178*
Time(1,2)	5.22	0.71	7.3	0.0007*

DOE when not all factor combinations are feasible

Accommodating a Constrained Factor Space

Goals

1. Show the practical need for design when there are restrictions on the factor settings.
2. Provide an example of an experiment with factor constraints that also accommodates curvature in the factor-response relationship.

Reasons to Impose Factor Constraints

1. Certain combinations of factor settings may not work and you know this in advance.
2. Other combinations of factor settings may be dangerous.
3. Other combinations could damage the processing system.

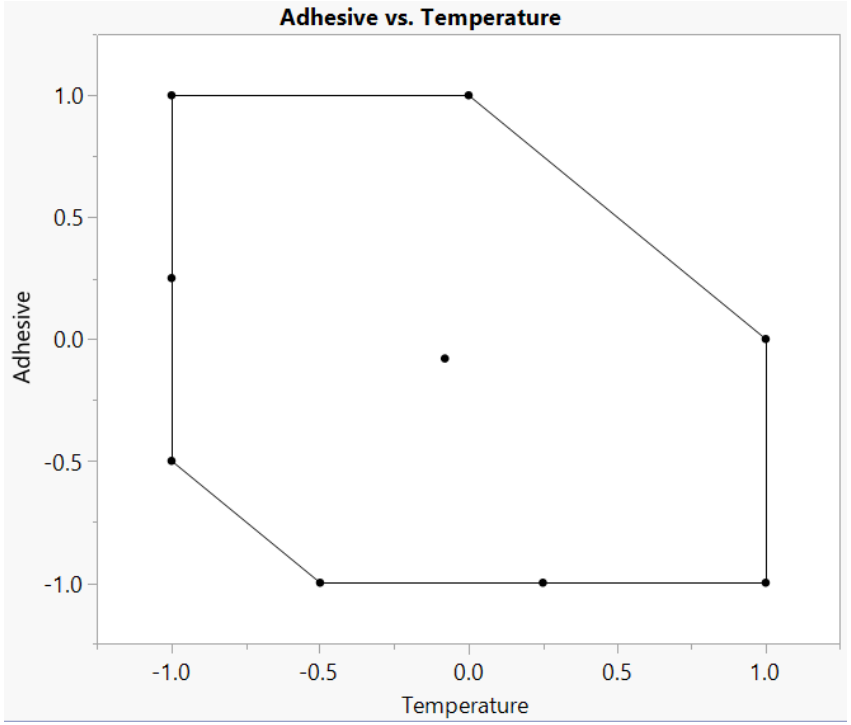
Design employing constrained factors

Scenario – Bonding electronic chip to a pacemaker

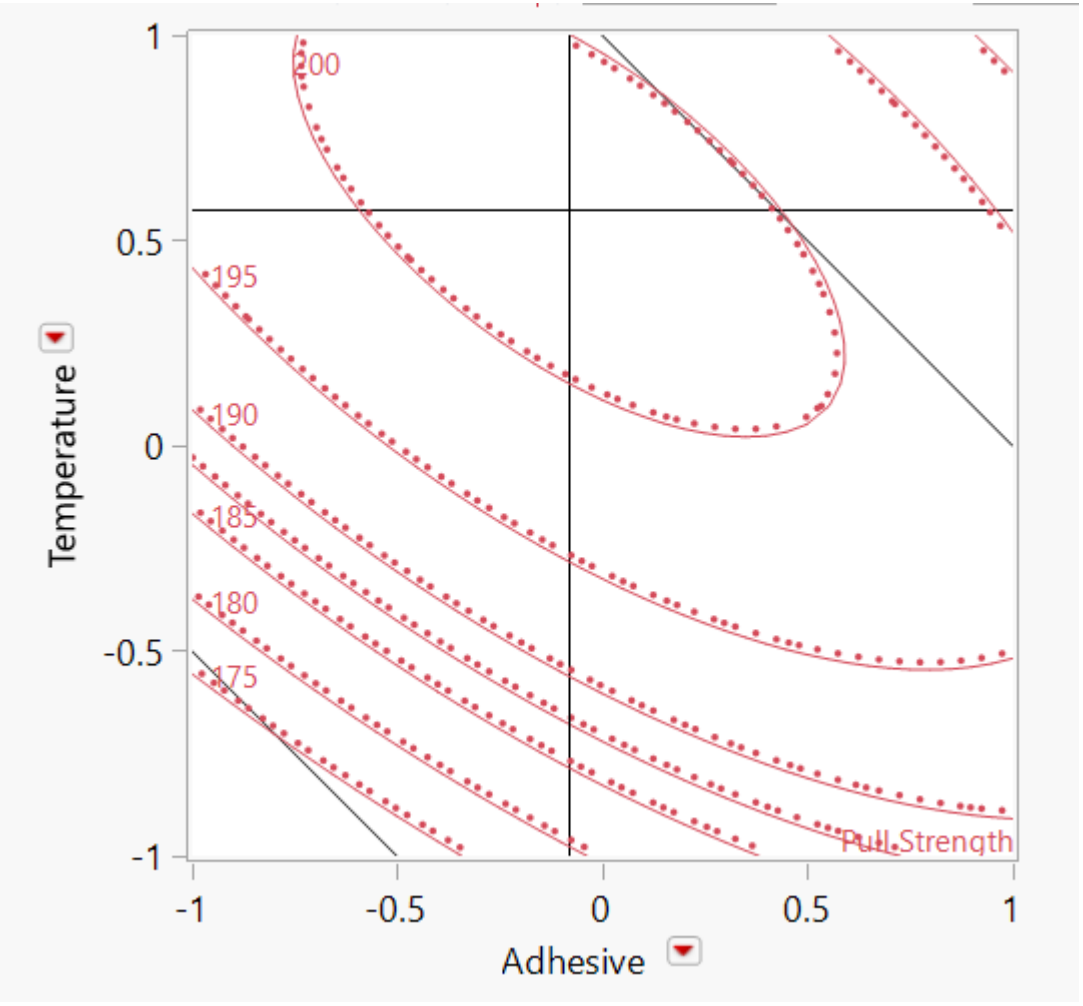
1. Two factors – adhesive and temperature
2. Response is peel strength of the bond
3. Combinations involving high settings of both factors or low settings of both factors are known to result in low yield.

Design with Response Data

Adhesive	Temperature	Pull Strength
-0.5	-1	173.9
1	0	195.9
-0.08	-0.08	201.2
-1	1	199.3
1	-1	190.3
0	1	198.5
-1	0.25	190.5
0.25	-1	179.5
-1	-0.5	176.7
1	0	198.1
0	1	200.5
-0.08	-0.08	197.1



Software Demonstration



A design with homogeneous groups of runs - blocks

1. Many processes have sources of variability that are not controllable factors.
2. Example: day-to-day variability
3. This leads to groups of observations.
4. The groups are called blocks.
5. The grouping variable, day, in this case is called a blocking factor.
6. We expect that runs within a day are more similar than runs between days

Vitamin Stability Experiment

ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis

Volume 16, Issue 2, October 1997, Pages 275-280

Evaluation of the methods for the determination of the stability constant of cyclodextrin–chlorambucil inclusion complexes

Yannis L. Loukas

Vitamin Stability Experiment

Vitamins degrade when exposed to light

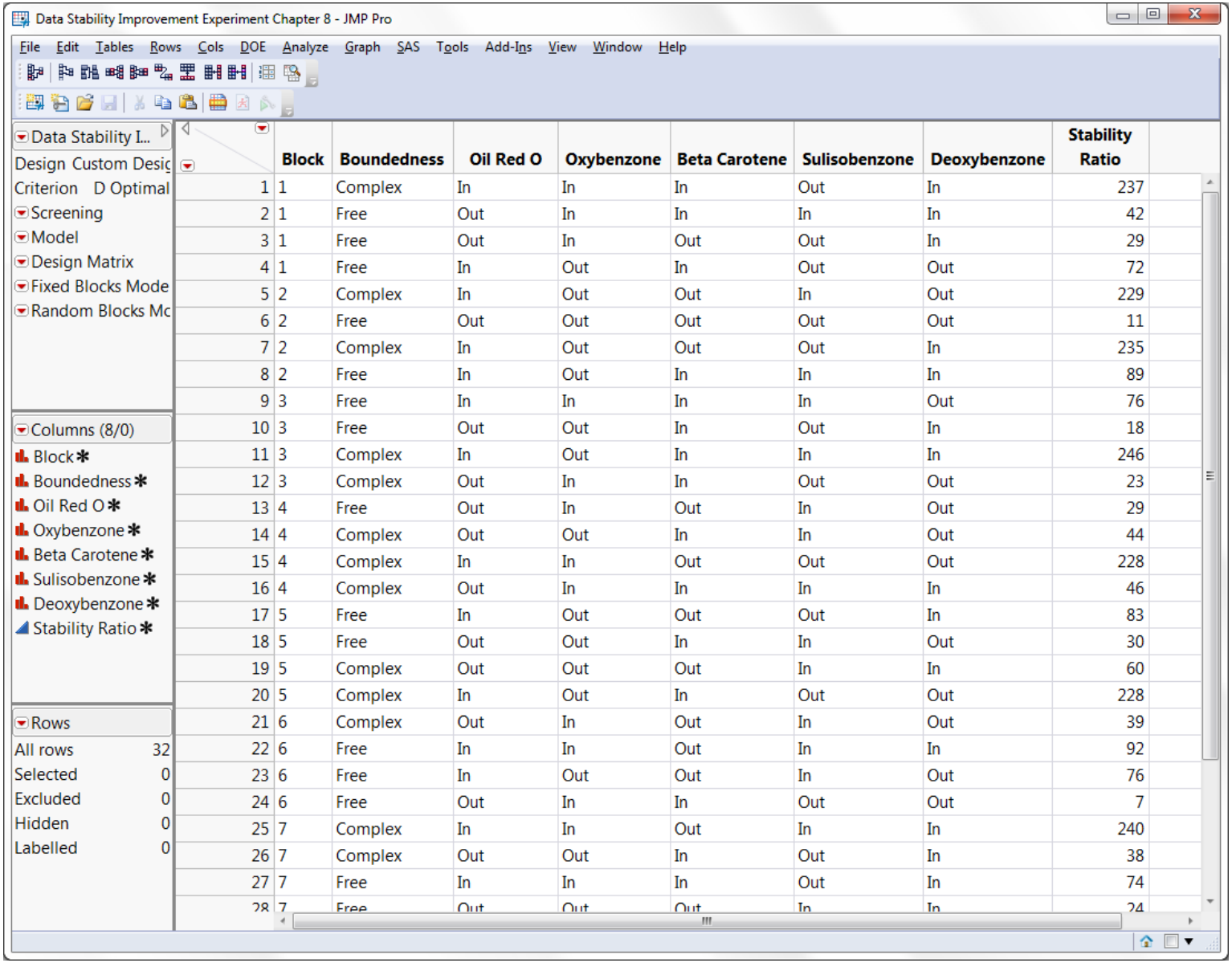
Can be stabilized when embedded in a fatty molecule

Five different fatty molecule possibilities

Binding with sugar might help as well to stabilize the vitamins

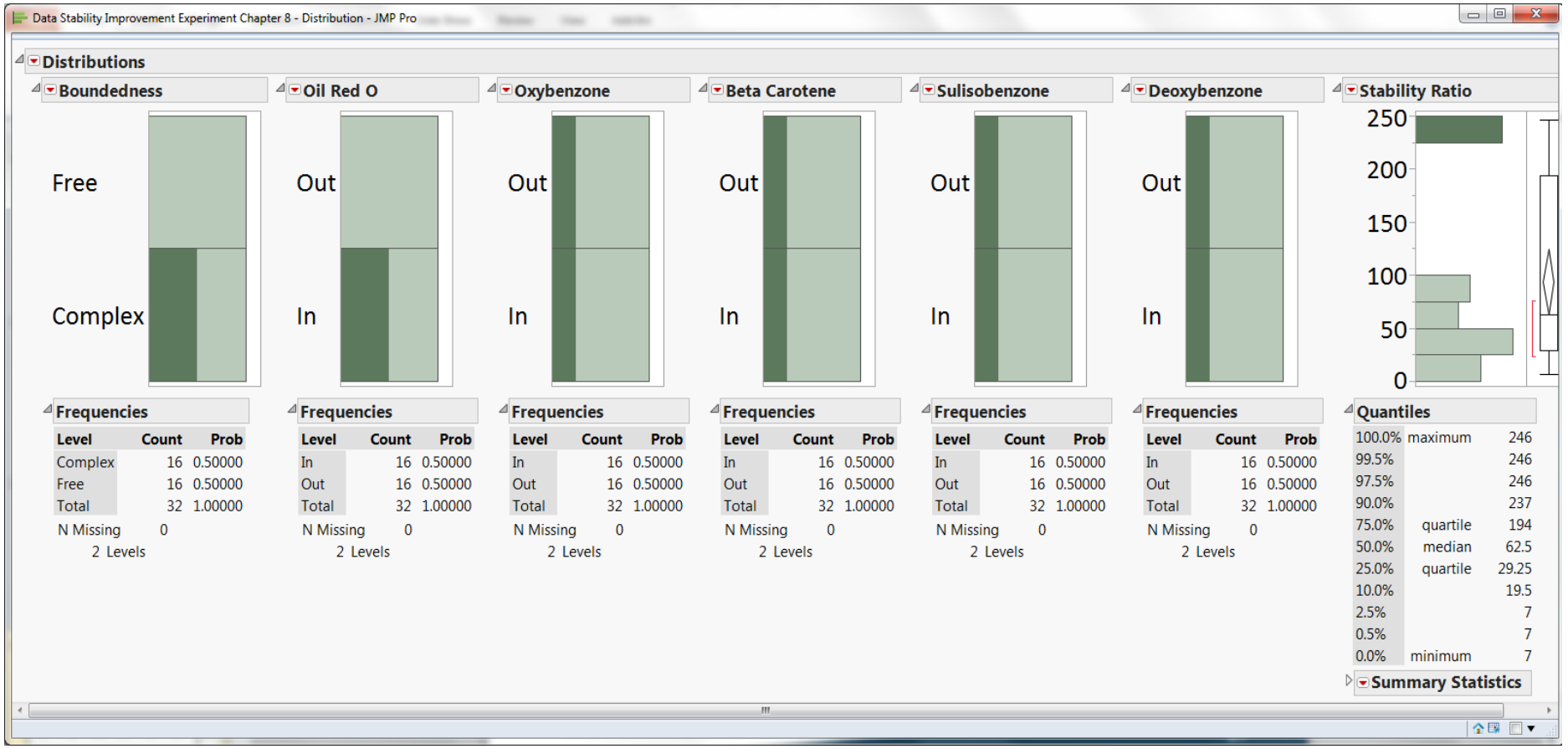
Day-to-day variation due to calibration of measurement apparatus

The data

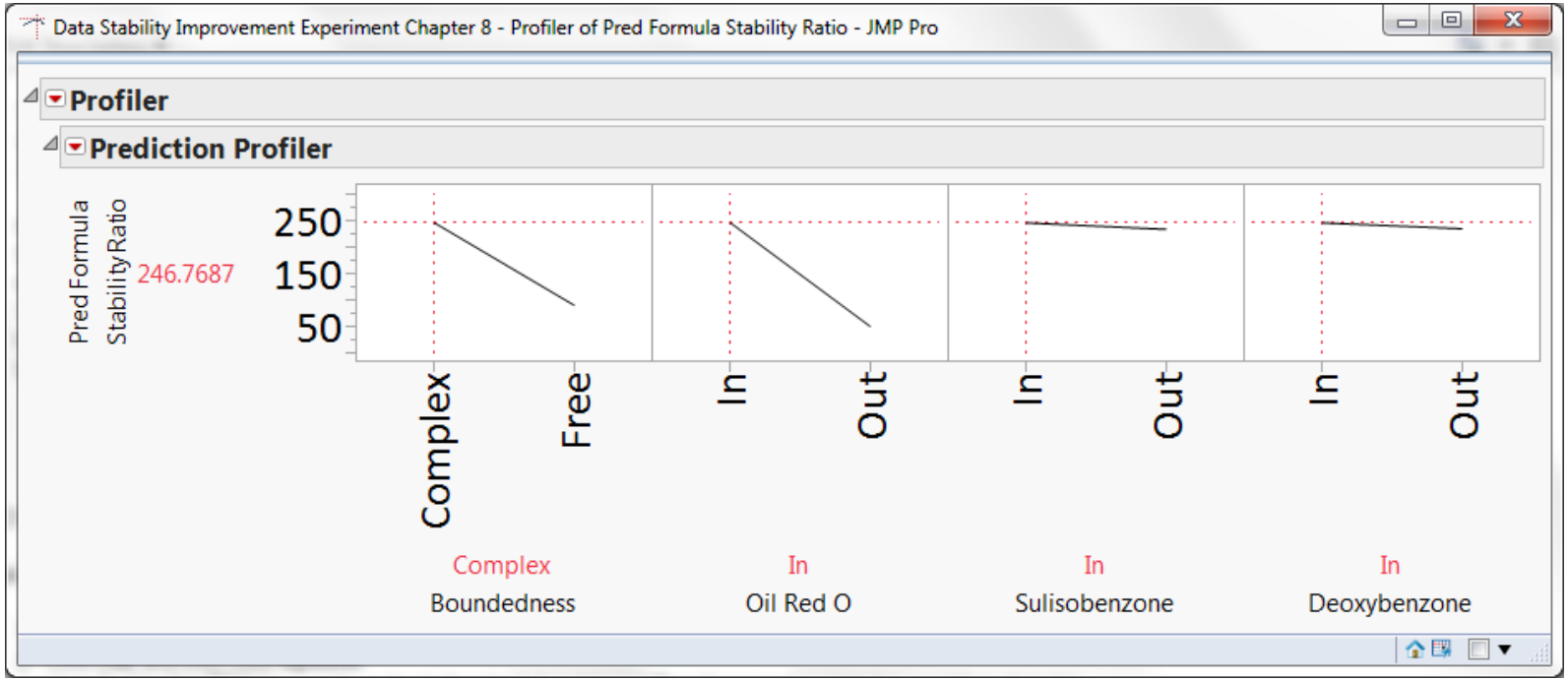


	Block	Boundedness	Oil Red O	Oxybenzone	Beta Carotene	Sulisobenzone	Deoxybenzone	Stability Ratio
1	1	Complex	In	In	In	Out	In	237
2	1	Free	Out	In	In	In	In	42
3	1	Free	Out	In	Out	Out	In	29
4	1	Free	In	Out	In	Out	Out	72
5	2	Complex	In	Out	Out	In	Out	229
6	2	Free	Out	Out	Out	Out	Out	11
7	2	Complex	In	Out	Out	Out	In	235
8	2	Free	In	Out	In	In	In	89
9	3	Free	In	In	In	In	Out	76
10	3	Free	Out	Out	In	Out	In	18
11	3	Complex	In	Out	In	In	In	246
12	3	Complex	Out	In	In	Out	Out	23
13	4	Free	Out	In	Out	In	Out	29
14	4	Complex	Out	Out	In	In	Out	44
15	4	Complex	In	In	Out	Out	Out	228
16	4	Complex	Out	In	In	In	In	46
17	5	Free	In	Out	Out	Out	In	83
18	5	Free	Out	Out	In	In	Out	30
19	5	Complex	Out	Out	Out	In	In	60
20	5	Complex	In	Out	In	Out	Out	228
21	6	Complex	Out	In	Out	In	Out	39
22	6	Free	In	In	Out	In	In	92
23	6	Free	In	Out	Out	In	Out	76
24	6	Free	Out	In	In	Out	Out	7
25	7	Complex	In	In	Out	In	In	240
26	7	Complex	Out	Out	In	Out	In	38
27	7	Free	In	In	In	Out	In	74
28	7	Free	Out	Out	Out	In	In	24

Graphical descriptive analysis



Software Demonstration



DOE With Covariate Factors Known in Advance

Designing around prior information about experimental units

Goals

1. Explain the practical utility of taking information about the experimental units into account in design
2. Provide an example of an experiment with a covariate factor.

Reasons to Use Covariate Factor Information

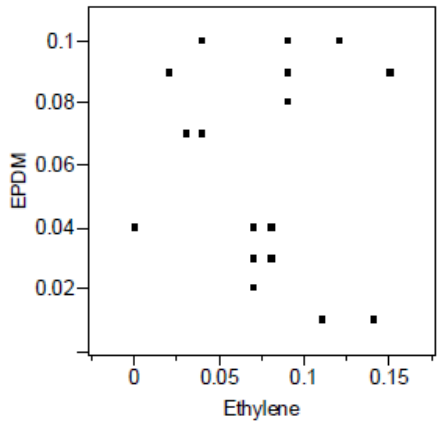
1. You may have more experimental material than necessary and need to choose which material to use.
2. Certain characteristics of the experimental material may strongly influence the response and you want to make sure that these characteristics are as little correlated with the effects of your control factors as possible.

Case Study Using Covariate Information

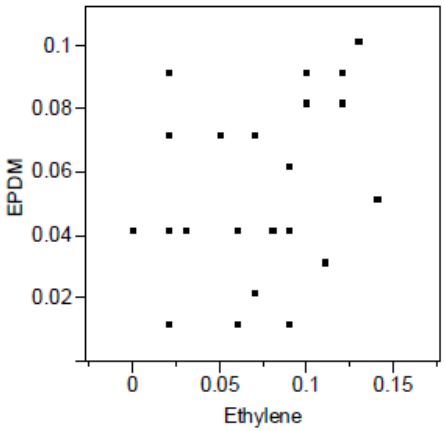
Scenario

1. You have 40 batches of polypropylene to use in your study.
2. You have a budget of 18 runs.
3. You have information about two continuous and one categorical characteristic for each batch.

Graphical View of Covariate Information

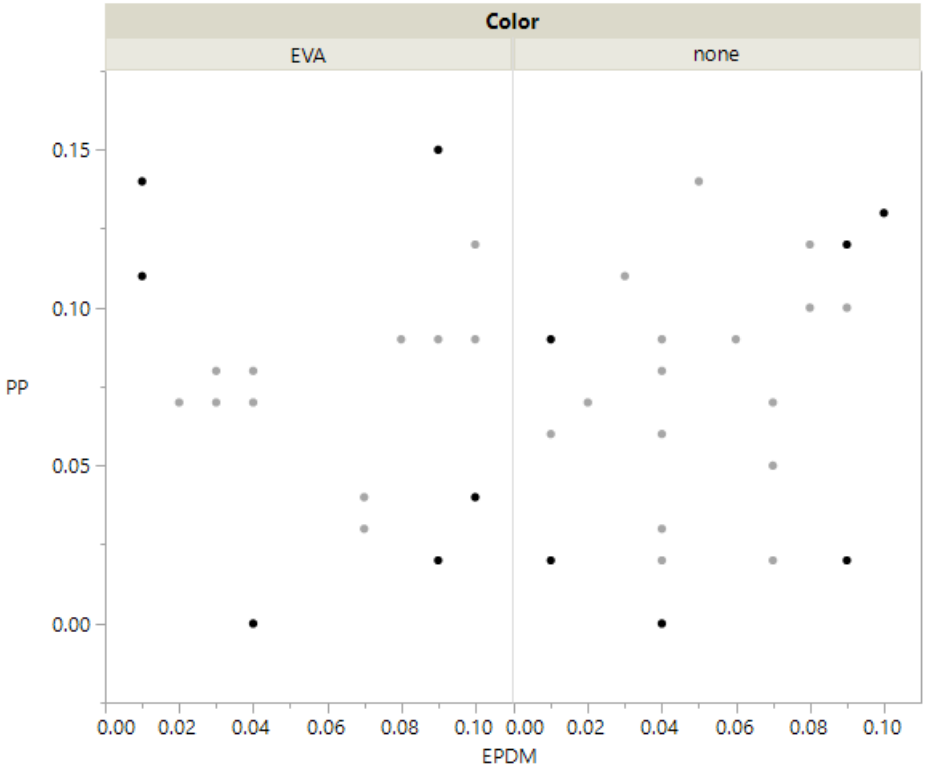


(a) Samples containing EVA



(b) Samples not containing EVA

Software Demonstration



Summary of DOE with Covariate Factors

1. Including covariate information, when it is available, is a powerful way to account for known sources of variability.

2. This approach can also be used to adjust the time order of runs when the response is subject to linear drift over time.

Summary

1. This talk provided a taste of a new approach to industrial design of experiments.

2. Each of the examples emphasizes the approach of designing experiments to solve the problem as stated rather than to change the requirements in order to use a textbook design.