THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

CHE384, From Data to Decisions: Measurement, Uncertainty, Analysis, and Modeling

Lecture 68
Factorial Design of Experiments

Chris A. Mack
Adjunct Associate Professor

http://www.lithoguru.com/scientist/statistics/

Design of Experiments Process

- First, define the objectives of the experiment
- Second, define the inputs to the process
  - Controlled inputs: those inputs under the direct control of the experimenter
  - Uncontrolled, measured inputs: those inputs not controlled by the experimenter, but which are measured (nuisance variables)
  - Uncontrolled, unmeasured inputs: inputs, often unknown to the experimenter, that are neither controlled nor measured but can affect the output response
- Inputs can be continuous (temperature, current) or discrete (which tool, operator, or factory)
- For each controlled input (called a factor), define a range and number of levels

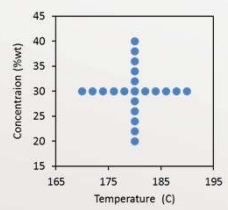
Circular Experimental Design

- An experimental design sets which predictor variables to vary, over what range, and with what distribution of values (sampling)
  - Given the nature of the response function (our model), we can easily decide how to sample it
- For exploratory or screening experiments, we are trying to understand which variables matter, and what the basic nature of the response is
  - How do we decide what our design should be?
  - The best time to design an experiment is after the experiment is finished!

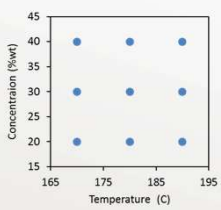
Exploratory Designs

- If we don't know the final form of our model, explore with a few data points first
  - Sequential approach: a simple screening experiment first, followed by a more thorough investigation
  - Use up to 25% of the total amount of data we plan to collect for screening
  - We want to know: What variables are most important, and over what ranges?

Example Design Choice



Changing one variable at a time ignores the possibility of interactions



This design is more efficient and effective at covering the design space

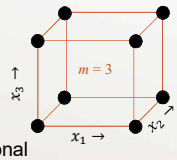
Full Factorial Design

- For factor A, use a levels, for factor B use b levels, for factor C use c levels, etc.
- It is common to normalize each factor to run from -1 to 1 (called coded variables)
- Perform n complete replicates of the experiment
- A full factorial design uses all combinations of all factors
  - Total number of measurements = $nabc\ldots$

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Two-Level Factorial Design (Screening)

- For each variable (factor), run every combination at two levels (min and max values, labeled -1 and +1)
  - For m factors, there will be 2^m experimental runs for a full factorial design (times the number of replications)
  - With this we can detect linear variations, including interactions
  - This design is completely orthogonal
  - We can add more data later, filling in the design space after seeing the initial results
  - Replications are required to estimate errors



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Hierarchy Principle

- The Hierarchy Principle in model development: only include a higher-order term if the constituent lower-order terms are also included
  - If $x_1^2 x_2$ is in the model, then one must also include x_1, x_2, and x_1^2
  - Only hierarchical models are invariant under linear transformation (e.g., a shift in zero position)
- Note: higher order and interaction terms can often mask the main effects (lower order terms)

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Fractional Factorial Design

- Consider a two-level, 7-factor full factorial design (for a total of 128 runs)

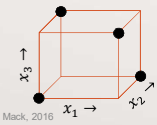
average	Main effects	2-factor interact	3-factor interact	4-factor interact	5-factor interact	6-factor interact	7-factor interact
1	7	21	35	35	21	7	1

- Many of the higher order interactions will be negligible (sparsity-of-effects principle), and are thus redundant
  - We can reduce our number of runs by eliminating the higher-order model interactions

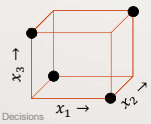
THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Fractional Factorial Design

- How to choose a sub-set of the full factorial design?
  - We want a balanced (all combinations have the same number of observations) and orthogonal (effects of any factor sum to zero across the effects of the other factors) design



Two half-factorial designs for three variables



Here main effects are confounded (aliased) with 2-factor interactions

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Two-Level Half-Factorial Design

- Half-factorial design for four factors (8 runs):

X_1	X_2	X_3	X_4
-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

- Start with a full two-level factorial design for three factors ($x_1 - x_3$)
- Let the entries in the fourth column be given by $x_4 = x_1 x_2 x_3$
- The alternative design sets $x_4 = -x_1 x_2 x_3$

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Fractional Factorial Aliasing

X_1	X_2	X_3	X_4	X_1X_2	X_3X_4
-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	+1	-1	-1	-1
-1	+1	+1	-1	-1	-1
+1	+1	+1	+1	+1	+1

- For four factors there is one intercept, four main effects, and 5 two-factor interactions (11 terms)
  - Not all of these terms can be distinguished in 8 runs
  - We say that some terms are "aliased" or confounded
- For example, in the 4-factor half-factorial, x_1x_2 is aliased with (indistinguishable from) x_3x_4, as is x_1x_3 with x_2x_4

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Projections

- In a half-factorial design, if one of the factors proves to have no effect on the response, the k factor half-factorial design collapses to a k-1 factor full factorial design

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Adding the Center Point

- In a full factorial design, it is common to add the center point, often with repeats
  - The center point is often the plan of record (POR), and significant data may already exist about its response
  - The center points do not affect the orthogonality of the design and do not change any model parameter except the intercept (overall mean) β_0
- A repeated center point can be used to check linear model validity
  - Are second-order terms significant?

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Lecture 68: What have we learned?

- What is full factorial design?
- What kind of design is often used for screening variables?
- What is the Hierarchy Principle?
- What is fractional factorial design and what are its advantages and disadvantages?
- Explain aliasing in fractional factorial design