SEMINAR AT ROFEL ON E.D. PLEASE READ IT!!!!!!!!!!!!

SCREENING & OPTIMIZATION
USING EXPERIMENTAL DESIGNS

What is Optimization

Optimization of a formulation or process is finding the

best possible composition or operating conditions

Thus, for a formulation, the type and components may be selected, according to the previous experience, by expert knowledge or by systemic screening. Then the relative and or total proportions of the excipients are varied to obtain the best endpoint

For a process, the process is chosen, and a study is

carried out to determine the best operating conditions

to obtain the desired formulation properties

Screening, factor studies, and optimization

The type of study carried out will depend on the stage of the project. In particular, experimental design may be carried out in stages, and the experiments of a factor study may be augmented by further experiments to a design giving the detailed information needed for true optimization

Historical Review

Approaches for Optimization

1. Statistically designed experiment approach

4. Non-systemic (Trial and Error) Approach

**Process Models for DOE**

Experimental designs for optimization

A. Completely randomized designs

B. Randomized block designs

 Latin squares

 Graeco-Latin squares

 Hyper-Graeco-Latin squares

C. Full factorial designs

D. Fractional factorial designs

E. Plackett-Burman designs

F. Response surface (second-order) designs
 Central composite designs
 Box-Behnken designs

Factorial Design in Method Development

* Chromatographic analysis was done by gas chromatography–mass spectrometry in the selected ion monitoring mode
* Using the peak area ratio (morphine-to-internal standard) as the response, they investigated the independent variables that could influence the acid hydrolysis, including temperature (range 70–130 °C), acid volume (range 500–1,000 µL) and time (range 15–90 min)
* A 23 full factorial design for the screening and a response surface methodology, including a central composite design for optimization, were applied
* The factors which influenced the response to a greater extent were temperature and its interaction both with time and acid volume
* By application of a multiple regression analysis to the experimental data, a second-order polynomial equation was obtained
* The optimal predicted conditions for morphine 3-glucuronide acid hydrolysis were 115 °C, 38 min and 500 µL for temperature, time and acid volume, respectively
* Using design of experiments, instead of the one factor at a time approach, we achieved the optimum combination of all factor values, and this allowed the best results to be obtained, simultaneously optimizing resources
* In addition, time and money can be saved, since other approaches are in general more time-consuming and laborious, and do not take into account the interactions between factors

 **Central Composite Designs (CCD)**

Box-Wilson Central Composite Designs

* A Box-Wilson Central Composite Design, commonly called `a central composite design,' contains an imbedded factorial or fractional factorial design with center points that is augmented with a group of `star points' that allow estimation of curvature
* If the distance from the center of the design space to a factorial point is ±1 unit for each factor, the distance from the center of the design space to a star point is ±α with | α | > 1
* The precise value of α depends on certain properties desired for the design and on the number of factors involved

*Diagram of central composite design generation for two factors*

* A central composite design always contains twice as many star points as there are factors in the design
* The star points represent new extreme values (low and high) for each factor in the design
* Table summarizes the properties of the three varieties of central composite designs
* The diagrams in Figure illustrate the three types of central composite designs for two factors
* Note that the CCC explores the largest process space and the CCI explores the smallest process space
* Both the CCC and CCI are [rotatable](http://www.itl.nist.gov/div898/handbook/pri/section7/pri7.htm) designs, but the CCF is not
* In the CCC design, the design points describe a circle *circumscribed* about the factorial square
* For three factors, the CCC design points describe a sphere around the factorial cube

**Determining α in Central Composite Designs**

* To maintain rotatability, the value of depends on the number of experimental runs in the factorial portion of the central composite design:

 α=[number of factorial runs]1/4

* If the factorial is a full factorial, then

 α=[2k]1/4

* However, the factorial portion can also be a fractional factorial design of resolution V

Typical value of α as a function of the number of factors

* The value of α also depends on whether or not the design is orthogonally blocked
* That is, the question is whether or not the design is divided into blocks such that the block effects do not affect the estimates of the coefficients in the 2nd order model
* Under some circumstances, the value of α allows simultaneous rotatability and orthogonality
* One such example for *k* = 2 is shown below:

 **Box-Behnken designs**

* The Box-Behnken design is an independent quadratic design in that it does not contain an embedded factorial or fractional factorial design
* In this design the treatment combinations are at the midpoints of edges of the process space and at the center. These designs are rotatable (or near rotatable) and require 3 levels of each factor
* The designs have limited capability for orthogonal blocking compared to the central composite designs.

**A Box-Behnken Design for Three Factors**

* The geometry of this design suggests a sphere within the process space such that the surface of the sphere protrudes through each face with the surface of the sphere tangential to the midpoint of each edge of the space

 **Comparisons of response surface designs**

* Table in next slide contrasts the structures of four common quadratic designs one might use when investigating three factors
* The table combines CCC and CCI designs because they are structurally identical
* For three factors, the Box-Behnken design offers some advantage in requiring a fewer number of runs
* For 4 or more factors, this advantage disappears.

**Structural Comparisons of CCC (CCI), CCF, and Box-Behnken Designs for Three Factors**

* Table illustrates the factor settings required for a central composite circumscribed (CCC) design and for a central composite inscribed (CCI) design (standard order), assuming three factors, each with low and high settings of 10 and 20, respectively
* Because the CCC design generates new extremes for all factors, the investigator must inspect any worksheet generated for such a design to make certain that the factor settings called for are reasonable
* In table on next slide, treatments 1 to 8 in each case are the factorial points in the design; treatments 9 to 14 are the star points; and 15 to 20 are the system-recommended center points
* Notice in the CCC design how the low and high values of each factor have been extended to create the star points
* In the CCI design, the specified low and high values become the star points, and the system computes appropriate settings for the factorial part of the design inside those boundaries

**Factor Settings for CCC and CCI Designs for Three Factors**

* Table illustrates the factor settings for the corresponding central composite face-centered (CCF) and Box-Behnken designs
* Note that each of these designs provides three levels for each factor and that the Box-Behnken design requires fewer runs in the three-factor case

**Factor Settings for CCF and Box-Behnken Designs for Three Factors**

**Number of Runs Required by Central Composite and Box-Behnken Designs**

**Summary of Properties of Classical Response Surface Designs**

 **Contour Plot**

* A contour plot is a graphical technique for representing a 3-dimensional surface by plotting constant *z* slices, called contours, on a 2-dimensional format. That is, given a value for *z*, lines are drawn for connecting the (x,y) coordinates where that *z* value occurs.
* The contour plot is an alternative to a 3-D surface plot
* The contour plot is formed by:
	+ Vertical axis: Independent variable 2
	+ Horizontal axis: Independent variable 1
	+ Lines: Iso-response values
* The independent variables are usually restricted to a regular grid
* The actual techniques for determining the correct iso-response values are rather complex and are almost always computer generated
* An additional variable may be required to specify the Z values for drawing the iso-lines
* Some software packages require explicit values
* Other software packages will determine them automatically
* If the data (or function) do not form a regular grid, you typically need to perform a 2-D interpolation to form a regular grid
* The contour plot is used to answer the question, How does Z change as a function of X and Y?
* For univariate data, a [run sequence plot](http://www.itl.nist.gov/div898/handbook/eda/section3/runseqpl.htm) and a [histogram](http://www.itl.nist.gov/div898/handbook/eda/section3/histogra.htm) are considered necessary first steps in understanding the data
* For 2-dimensional data, a [scatter plot](http://www.itl.nist.gov/div898/handbook/eda/section3/scatterp.htm) is a necessary first step in understanding the data
* In a similar manner, 3-dimensional data should be plotted
* Small data sets, such as result from designed experiments, can typically be represented by [block plots](http://www.itl.nist.gov/div898/handbook/eda/section3/blockplo.htm), [decks mean plots](http://www.itl.nist.gov/div898/handbook/eda/section3/dexmeanp.htm), and the like (here, "DEX" stands for "Design of Experiments“)
* For large data sets, a contour plot or a 3-D surface plot should be considered a necessary first step in understanding the data
* The [dex contour plot](http://www.itl.nist.gov/div898/handbook/eda/section3/eda33a1.htm) is a specialized contour plot used in the design of experiments
* In particular, it is useful for [full](http://www.itl.nist.gov/div898/handbook/pri/section3/pri333.htm) and [fractional](http://www.itl.nist.gov/div898/handbook/pri/section3/pri334.htm) designs
* Contour plots are available in most general purpose statistical software programs
* They are also available in many general purpose graphics and mathematics programs
* These programs vary widely in the capabilities for the contour plots they generate
* Many provide just a basic contour plot over a rectangular grid while others permit color filled or shaded contours
* [Dataplot](http://www.itl.nist.gov/div898/handbook/eda/section4/eda44.htm) supports a fairly basic contour plot
* Most statistical software programs that support design of experiments will provide a dex contour plot capability

 **DEX Contour Plot**

* The dex contour plot is a specialized contour plot used in the analysis of [full](http://www.itl.nist.gov/div898/handbook/pri/section3/pri333.htm) and [fractional](http://www.itl.nist.gov/div898/handbook/pri/section3/pri334.htm) experimental designs
* These designs often have a low level, coded as "-1" or "-", and a high level, coded as "+1" or "+" for each factor
* In addition, there can optionally be one or more center points. Center points are at the mid-point between the low and high level for each factor and are coded as "0".
* The dex contour plot is generated for two factors. Typically, this would be the two most important factors as determined by previous analyses (e.g., through the use of the [dex mean plots](http://www.itl.nist.gov/div898/handbook/eda/section3/dexmeanp.htm) and a [Yates analysis](http://www.itl.nist.gov/div898/handbook/eda/section3/eda35i.htm)).
* If more than two factors are important, you may want to generate a series of dex contour plots, each of which is drawn for two of these factors.
* You can also generate a matrix of all pairwise dex contour plots for a number of important factors (similar to the [scatter plot matrix](http://www.itl.nist.gov/div898/handbook/eda/section3/scatplma.htm) for scatter plots).
* The typical application of the dex contour plot is in determining settings that will maximize (or minimize) the response variable. It can also be helpful in determining settings that result in the response variable hitting a pre-determined target value.
* The dex contour plot plays a useful role in determining the settings for the next iteration of the experiment. That is, the initial experiment is typically a fractional factorial design with a fairly large number of factors.
* After the most important factors are determined, the dex contour plot can be used to help define settings for a full factorial or response surface design based on a smaller number of factors.

The following are the primary steps in the construction of the dex contour plot.

* The *x* and *y* axes of the plot represent the values of the first and second factor (independent) variables
* The four vertex points are drawn. The vertex points are (-1,-1), (-1,1), (1,1), (1,-1)
* At each vertex point, the average of all the response values at that vertex point is printed
* Similarly, if there are center points, a point is drawn at (0,0) and the average of the response values at the center points is printed
* The **linear** dex contour plot assumes the model:
* where µ is the overall mean of the response variable. The values of β1, β2, β12 and µ are estimated from the vertex points using a [Yates analysis](http://www.itl.nist.gov/div898/handbook/eda/section3/eda35i.htm) (the Yates analysis utilizes the special structure of the 2-level full and fractional factorial designs to simplify the computation of these parameter estimates)
* Note that for the dex contour plot, a full Yates analysis does not need to performed, simply the calculations for generating the parameter estimates
* In order to generate a single contour line, we need a value for *Y*, say *Y*0
* Next, we solve for *U*2 in terms of *U*1 and, after doing the algebra, we have the equation:
* We generate a sequence of points for *U*1 in the range -2 to 2 and compute the corresponding values of *U*2
* These points constitute a single contour line corresponding to *Y* = *Y*0
* The user specifies the target values for which contour lines will be generated
* The above algorithm assumes a linear model for the design
* Dex contour plots can also be generated for the case in which we assume a quadratic model for the design
* The algebra for solving for *U*2 in terms of *U*1 becomes more complicated, but the fundamental idea is the same
* Quadratic models are needed for the case when the average for the center points does not fall in the range defined by the vertex point (i.e., there is curvature)

 ***Interaction Significance:***

* Note the appearance of the contour plot
* If the contour curves are linear, then that implies that the interaction term is not significant; if the contour curves have considerable curvature, then that implies that the interaction term is large and important
* In our case, the contour curves do not have considerable curvature, and so we conclude that the X1\*X2 term is not significant

 ***Best Settings:***

* To determine the best factor settings for the already-run experiment, we first must define what "best" means
* For the Eddy current data set used to generate this dex contour plot, "best" means to **maximize** (rather than minimize or hit a target) the response
* Hence from the contour plot we determine the best settings for the two dominant factors by simply scanning the four vertices and choosing the vertex with the **largest** value (= average response).
* In this case, it is (X1 = +1, X2 = +1)
* As for factor X3, the contour plot provides no best setting information, and so we would resort to other tools: the main effects plot, the interaction effects matrix, or the ordered data to determine optimal X3 settings

Question Bank

1. What is optimization? Discuss various approaches for optimization
2. Explain the terms: dependent variable, independent variable, confounding, experimental domain
3. What is optimization? Discuss various phases of optimization
4. Describe factorial designs with advantages
5. Give the factorial design layout for 3 factors each at 3 levels
6. Describe fractional factorial designs with advantages
7. Explain development of full and reduced model
8. Define CCD for RSM. Describe Box-Behnken design
9. Describe Box-Wilson method for RSM
10. What are contour plots? Discuss their significance