Mybo Webinar-4

on

Factorial Designs and Simplex Lattice in Pharmaceutical Research: Concepts and calculations

Prof. J. VIJAYA RATNA
(Retired)
AU College of Pharmaceutical Sciences
Andhra University, Visakhapatnam
Andhra Pradesh, India

1

Contents

2

- Introduction
- Designs of experiments
- EVOP- example
- Simplex lattice: two component, three component- examples, calculations
- Factorial Designs: one factor at different levels, two factors at two levels and three factors at two levels (with one replication and with four replications)

Prof. J. Vijaya Ratna

2

Need for optimisation: Quality by Design

3

In a Quality-by-Design system:

- The product is designed to meet patient requirements
- The process is designed to consistently meet product critical quality attributes
- The impact of formulation components and process parameters on product quality is understood
- Critical sources of process variability are identified and controlled
- The process is continually monitored and updated to assure consistent quality over time

3

QbD Approach

4

- Quality built into product & process by design, based on scientific understanding
- Knowledge rich submission – showing product knowledge & process understanding
- Specifications based on product performance requirements
- Flexible process within design space, allowing continuous improvement
- Focus on robustness – understanding and controlling variation

4

QbD Approach

5

1. Choose Experimental Design.
2. Conduct Randomized Experiments
3. Carry out Analysis of Data. Find the significant attributes.
4. Create Multidimensional Surface model.
5. Identify the levels of the independent variables for the given dependent variable.

Prof. J. Vijaya Ratna

5

Why statistical optimization ?

6

- Efficient experimentation
- Cost effective
- Effect of all factors can be studied in less time
- Interactions of variables on the results can be studied
- The best solution can be obtained at through the challenging dependent and independent variables.
- To overcome the problem of one variable at a time (OVAT) process.

Prof. J. Vijaya Ratna

6

7

Statistical Designs:

- There are two varieties of statistical optimization designs.
- Experimentation continues as optimization study is done.
- Evolutionary Operations Procedure or EVOP and the simplex methods are examples of this.
- Experimentation is completed before the optimization takes place.
- Classic mathematical and search methods are examples for this.

7

8

Response Surface Methodology

- Response Surface Methodology (RSM) is composed of a particular set of mathematical and statistical techniques used to investigate the practical relationship between one or more measured responses (i.e. dependent variables) and a number of independent variables) with the ultimate goal of obtaining an optimal problem solution.

8

Response Surface Methodology :

9

- The basic components: experimental design, regression analysis, and optimization analysis, and optimization algorithms.
- use of RSM allows the determination of an optimal set of experimental conditions that maximize or minimize a primary response.
- also useful for examining changes in the response surface over a given range of independent variable levels.

Prof. J. Vijaya Ratna

9

10

There are three phases in RSM.

- Exploratory phase
- Bounding phase
- Optimization phase

Prof. J. Vijaya Ratna

10

Three phases

11

- **Exploratory Phase :** the experimenter's objectives are to screen a large set (≥9) of independent variables to determine whether they significantly affect the measured response. Significance of variables is determined.
- Direction toward the region most likely to include the optimal response is determined. Direction of influence of variables is determined.
- **Bounding Phase :** The experimental region containing the optimum response is located. Pin pointing is done.
- **Optimization phase :** Levels of the independent variables that give the best response are determined. Optimization is done.

11

Example for RSM

12

- Operation of a bioreactor to produce an enzyme by growing a micro organism is the example.
- **Objective :** To identify the most important variables and their levels for the best yield of the enzyme.
- **Variables :** Several variables are there such as (1) amount of protein source (2) amount of carbohydrate source (3) amount of nitrogen source (4) pH (5) mixing rate (6) aeration rate (7) temperature.
- Choosing the variables and levels from experience we perform several experiments in Factorial Design.

12

Example for RSM

13

- By analysis of Variance we find their significance, main effects and interaction effects. We select most relevant independent variables and go for a DoE.
- Again we get Analysis of Variance, multiple Regression Equations and graphs.
- We develop what are known as contour plots or response surfaces.
- From these graphs we select certain optimized variables and levels
- Using them we perform further experiments and compare the results of the prepared process/product with predicted result by DoE.
- Error must be nearer to zero. This process/ product then becomes our optimized one.
- Thus we optimize the production of the enzyme.

13

Sequential Optimization

14

- Experimental Strategies for Response Surface Exploration and Optimization
- Sequential experimentation helps us to build on the knowledge gained at the completion of each part of the overall design strategy.

14

Comprehensive Experimentation

15

- If the entire experiment was designed and performed at the onset, comprehensive experimentation, the investigator would have to know
- (a) which variables were the most important
- (b) the ranges of the variables to be studied and
- (c) whether mathematical transformation of the data or responses was necessary.
- At the beginning of the investigation, the researcher cannot answer these questions
- As the experimentation proceeds, researcher becomes capable of answering these questions.

15

Evolutionary Operations

16

- Evolutionary Operation (EVOP) is a sequential optimization technique.
- First developed by Box and Droper.
- Designed to be performed by process operators during full scale manufacturing without endangering the final product.

16

Procedure for EVOP

17

- Operators perform a simple experiment -typically a factorial design with a center point- inside the range of operating conditions that will give acceptable product.
- Changes in the processing variables are usually quite small.
- Many trials are necessary to differentiate the effect of a variable from the experimental error.
- This does not matter, because a typical production operation is performed repetitively.
- When one set of data has been collected at all design point, a "cycle" is said to have been completed.

17

Cycles

18

- A single cycle is not sufficient to detect any significant change in the response, so the experiment is repeated at the same design points (i.e., Cycle-2).
- Additional cycles are performed until one or more processing variables, interactions or the mean response proves to be significantly different from the experimental error.
- An estimate of the experimental error is obtained from the cycle data.
- After a significant effect has been detected, one "phase" is said to have been completed, and the processing variables are adjusted in a direction that improves the operation.

18

EVOP

19

- The objective, is to move in the direction of the optimum response.
- To facilitate EVOP procedures by production personnel, a simple form was developed for analysis of the data during the manufacturing operation.
- Such forms have been adapted to the following example.

19

EVOP example : Suspension Manufacturing

20

- A large in-line rotor stator colloid will with recirculating capabilities was adapted to a 5000-L tank for manufacture of a suspension.
- Because of active ingredient degradation, a 5% drug flush was added to the suspension to obtain a 24month expiration dating.
- It was determined after production start up that by reducing the heat generated during suspension homogenization, less active ingredients degraded and, therefore, a more stable product could be obtained.

20

Rotor- stator gap and total homogenization time

21

- The improved stability would subsequently allow reduction of the drug flush.
- This can save money by reducing drug input and also by extending the expiration date.
- Because product demand was great, the manufacturing operation could not be stopped or disturbed.
- Evolutionary operating techniques were used to determine whether small changes in the rotor stator gap distances and total homogenization times could be made to minimize the final product temperature,
- While always functioning within the ambit of company specifications.

21

2^2 factorial design

22

- A 2^2 factorial design, centered around the current operating conditions, was chosen for the study.
- The center point, taken at the currently used gap distance and homogenization time, allowed a check on possible changes in the mean suspension temperature that would indicate curvature in the response surface and the possibility that the current process was straddling a minimum (or a maximum).

22

First phase- First cycle

23

- The operating variables were coded at the upper (+), lower(-), and midpoint (O) values, and the production lots were manufactured randomly.
- The main effects and interaction term were calculated.
- Unless a separate estimate of the experimental error is available, little information can be obtained from the first cycle.

23

Second EVOP cycle

24

- Was performed at the same operating levels
- Because none of the effects were greater than the error limits, which represent approximately a 95% confidence interval, any real effects could not be distinguished from experimental error

24

Third cycle

25

- A third cycle was performed under the same operating conditions. In this cycle, the gap setting was shown to have a significant effect at the 95% confidence level.
- This ended the first EVOP phase and the results were presented to a special expert committee.

25

Decision at the end of Phase I

26

- It was ascertained that the final suspensions were well within company specifications during phase I of the EVOP evaluation.
- Decision was made to reset the operating conditions at the high gap setting (0.016in) and, because it had no significant effect, a short homogenization time (55 min).
- This resulted in a more stable product (it showed longer shelf life) as well as a decrease in the overall production time.

26

Phase II

27

- Phase II of the project was then initiated around these new operating conditions (i.e. new design center point) and EVOP was continued to further minimize the homogenized suspension temperature.

27

Conclusion of testing

28

- This process was continued further into phase III and phase IV; until no further decrease in temperature could be achieved without the danger of spoiling the product integrity.

28

ANOVA and polynomial equation/response surface

29

Prof. J. Vijaya Ratna

Two parts in Analysis.

- **Analysis of Variance** to find the significance of main effects and interaction effects
- **Constructing a polynomial equation**/developing a response surface to find the coefficients for different effects and thus determine the relationship between the independent and dependent variables- and thus to be able to do optimisation – find X values for a given Y value.

29

The Simplex Lattice

30

Prof. J. Vijaya Ratna

30

The Simplex Lattice

31

- Response surfaces and optimal regions for formulation characteristics are frequently obtained from the application of simplex lattice designs.
- A simplex is a geometric figure that has one more point than the number of factors.
- So, for two factors or independent variables the simplex is represented by a triangle.
- Once the shape of a simplex has been determined, the method can employ a simplex of fixed size or of variable size that are determined by comparing the magnitudes of the responses after each successive calculation

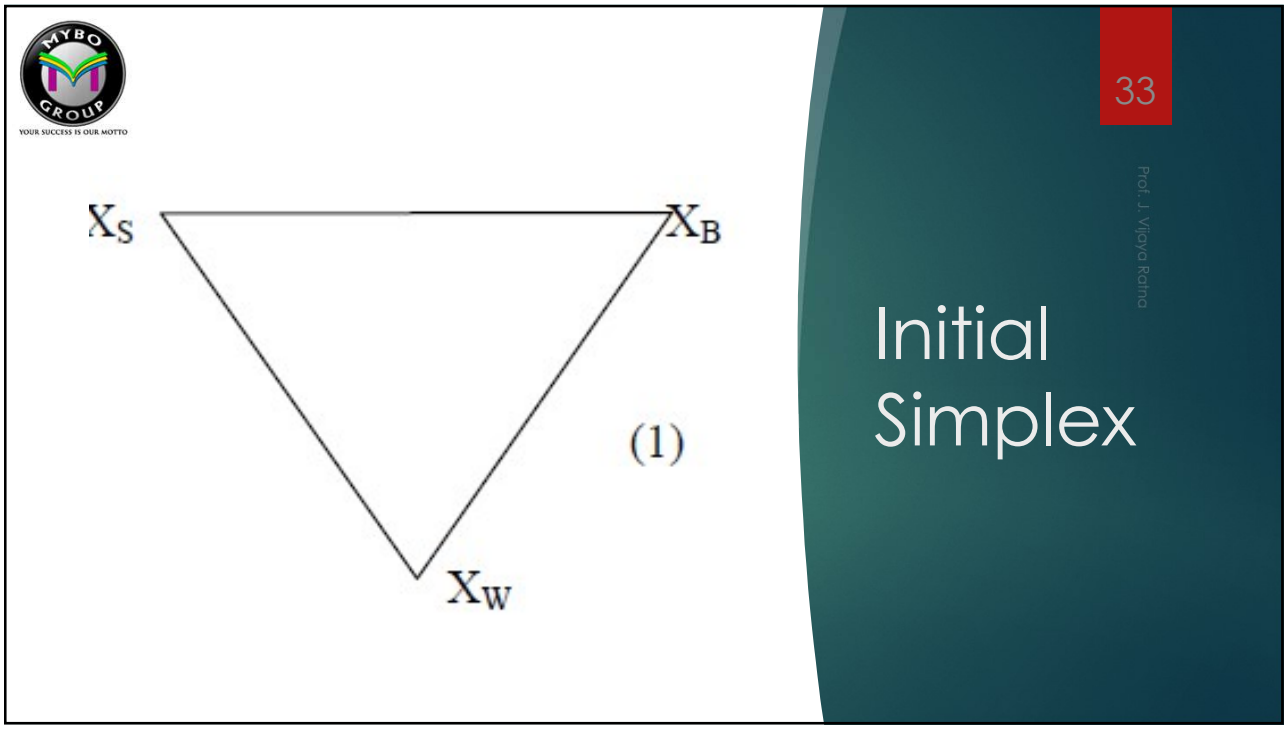
31

Initial Simplex

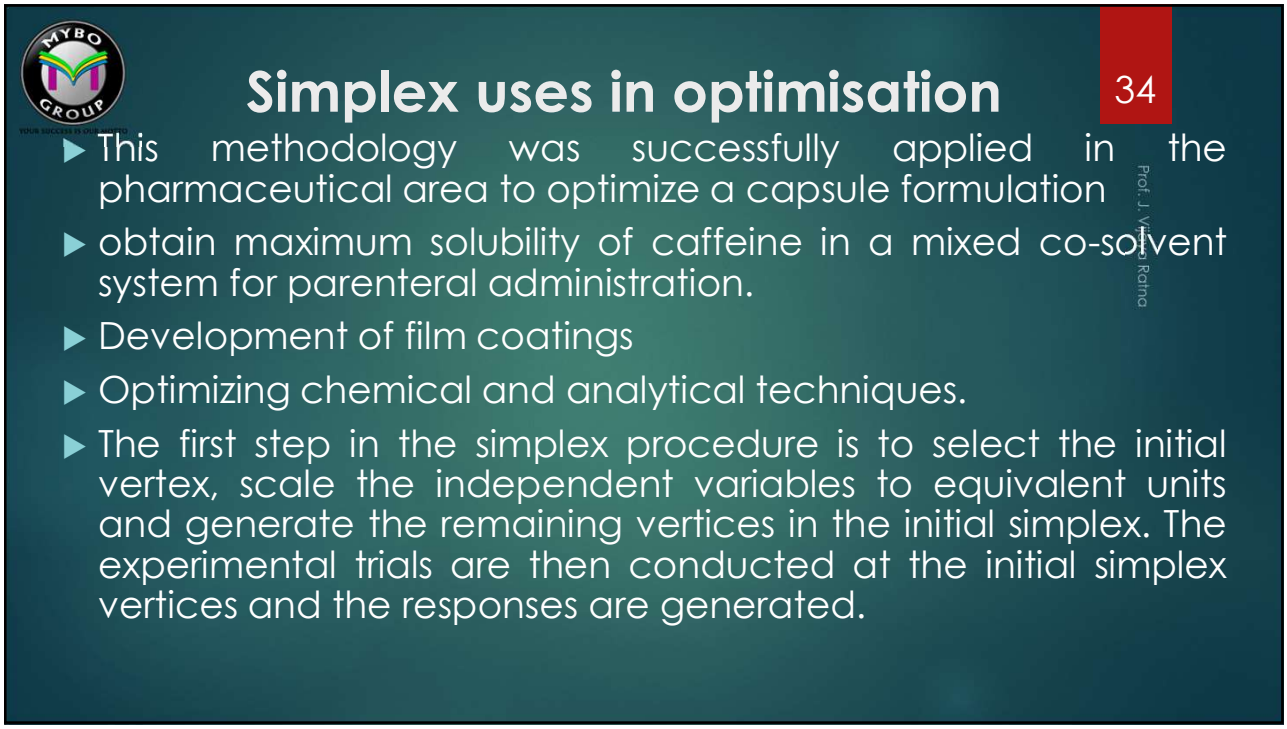
32

- The initial simplex is represented by the lowest triangle; the strategy is to move toward a better response by moving away from the worst response.
- The simplex is represented by n+1 vertices of an n-dimensional figure, where n is the number of factors being examined.
- For example, to investigate two independent variables, the simplex design is a triangle. The results of previous experiments are used to define the variables levels for the next experiment in the search for an optimum. The simplex figure is then directed along the response surface following a set of rules that sequentially moves value.

32


33
Initial Simplex
Prof. J. Vijaya Ratna
X_S
X_B
(1)
X_W

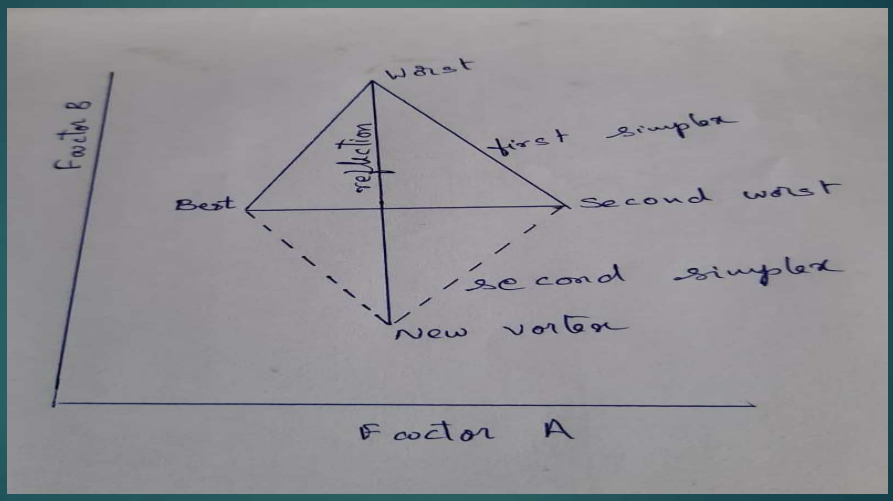
33


Simplex uses in optimisation
34
Prof. J. Vijaya Ratna
This methodology was successfully applied in the pharmaceutical area to optimize a capsule formulation
obtain maximum solubility of caffeine in a mixed co-solvent system for parenteral administration.
Development of film coatings
Optimizing chemical and analytical techniques.
The first step in the simplex procedure is to select the initial vertex, scale the independent variables to equivalent units and generate the remaining vertices in the initial simplex. The experimental trials are then conducted at the initial simplex vertices and the responses are generated.

34

Reflection

35



35

Reflection

36

- If the worst response is found at vertex X_W, the best response at vertex X_B, and the second best response at X_S, by "reflecting" away from the vertex with worst response (X_W) it is hoped that a value closer to the optimum will be located.
- Reflection away from X_W takes place through the centroid (X_O) of the remaining vertices (X_S and X_B) and the response at the new reflected vertex (X_R) is then evaluated.
- X_o is the centroid of all the remaining vertices except for x_w, x_w is the vertex in the simplex having the worst response.

36

Expansion

37

- If the response during reflection at vertex X_R is better than the response at vertex X_B, then expansion of the simplex is attempted.
- If the response at vertex X_E is better than that at X_R, then X_R is replaced by X_E and the new simplex ($X_S X_B X_E$) is evaluated. If on the other hand the response at X_E is poorer than that at X_R, the expansion process is considered unsuccessful. The vertex X_R is then retained reflection begins again.

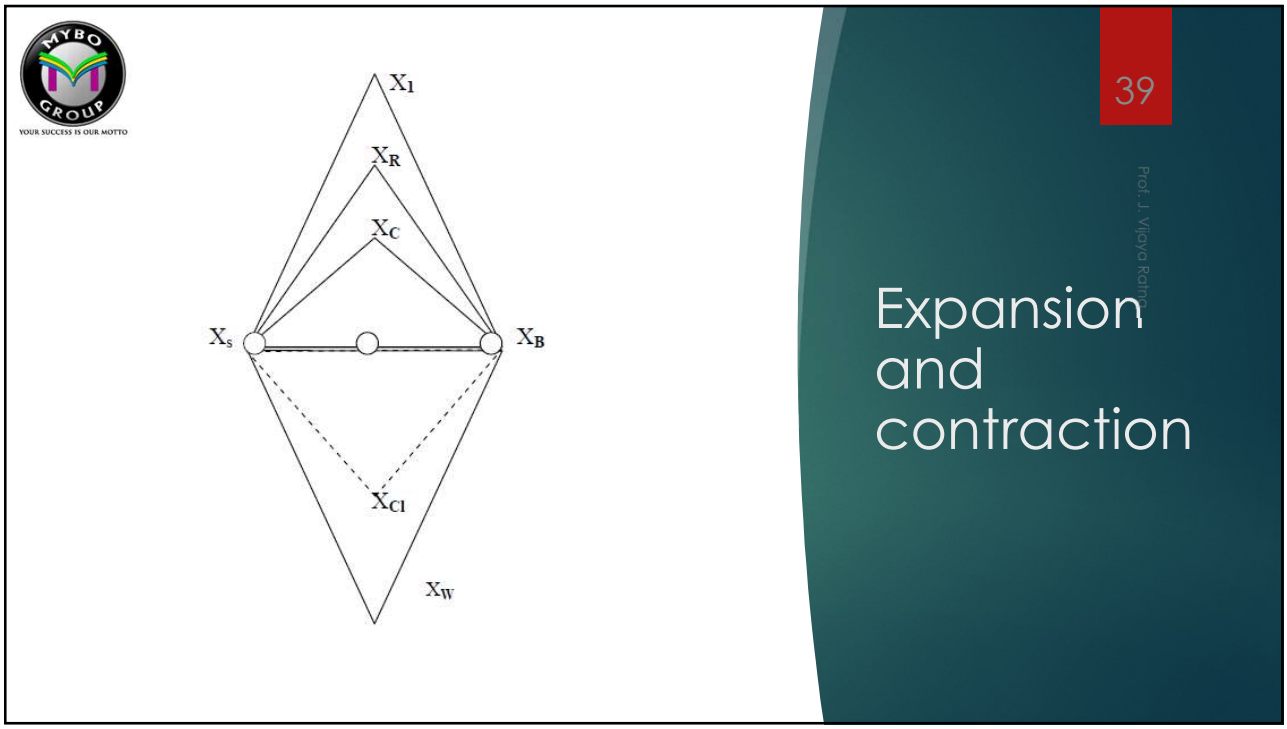
37

Contraction

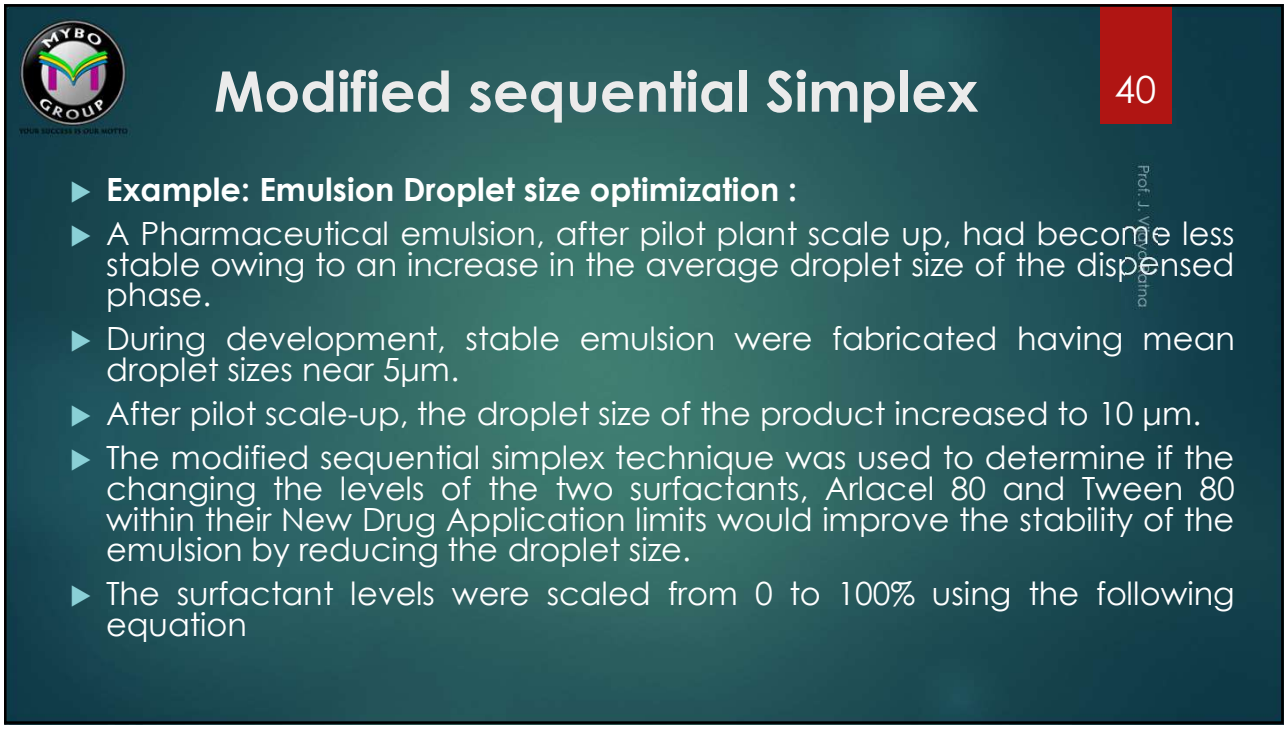
38

- If during reflection process, the response at X_R is worse than all the other responses in the simplex except for that at vertex X_W, we replace vertex X_W with vertex X_R and "contract" the new simplex $X_S X_B X_R$.
- Thus the new X_W will be X_R.
- The reflection and expansion are performed until an optimum response is located.

38


39
Expansion and contraction
Prof. J. Vijaya Ratna
X_1
X_R
X_C
X_S
X_B
X_{C1}
X_W

39


Modified sequential Simplex
40
Prof. J. Vijaya Ratna
Example: Emulsion Droplet size optimization :
A Pharmaceutical emulsion, after pilot plant scale up, had become less stable owing to an increase in the average droplet size of the dispensed phase.
During development, stable emulsion were fabricated having mean droplet sizes near 5µm.
After pilot scale-up, the droplet size of the product increased to 10 µm.
The modified sequential simplex technique was used to determine if the changing the levels of the two surfactants, Arlacel 80 and Tween 80 within their New Drug Application limits would improve the stability of the emulsion by reducing the droplet size.
The surfactant levels were scaled from 0 to 100% using the following equation

40

Example

41

$$F_S = \frac{F_i - F_{Lo}}{F_{Hi} - F_{Lo}} \times 100$$

- Where F_S was the scaled surfactant level in percent, F_i was the actual surfactant level and F_{HI} and F_{LO} were the high and low limits for the surfactant level found in the NDA.

$$\text{Centroid} = [\text{Tween level} - \text{Arlaced level}] = X_O$$

$$= \frac{1}{2}\ 35\% + 10\% - 10\% + 60\% = 22.5\% - 35\%$$

- Vertex 4 = [Tween level – Arlaced level] = X_R

= (1+1) [22.5% - 35.0%] – [10% - 10%] = [35% - 60%]

Results of Modified simple Simplex:

41

Example

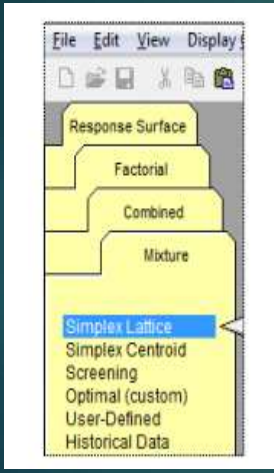
42

- The sequential simplex was a technique continued until a minimum droplet size of 4 µm was obtained at a Tween 80 level of 55.5% and an Arlaced level of 64.3% resulting in an emulsion exhibiting excellent stability properties.
- This is the optimized condition.

42

SIMPLEX LATTICE

43



- It is mixture design for 2 to 3 components where all the components have the same range and there are no constraints on the design space. The points chosen are pure components and then enough points between those to estimate the polynomial is chosen. By default, this design is augmented to include the center point and axial check blends.
- Simplex design provide the advantage of changing the quantity of different ingredients in the formulation in a systematic manner, yet keeping their total amount constant.

43

Formulation and Development of Buoyant Matrices of Dipyridamole

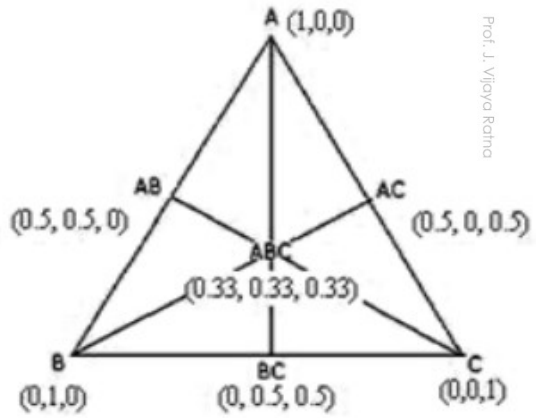
44

- Experimental design A Simplex Lattice Design (SLD) was adopted to optimize the formulation variables.
- In this design 3 factors were evaluated by changing their concentrations simultaneously and keeping total concentration constant.
- The amounts of matrix forming agent (HPMC K4M): X1
- Gas-generating agent (sodium bicarbonate): X2
- Floating enhancer (ethyl cellulose): X3 were selected as independent variables

44

45

- The floating lag time (FLT): Y1
- Percentage of drug release at 6th hour: Y2 were taken as response variables.
- Response Surface Plot and Contour Plot were developed and used in the process of optimization.



Prof. J. Vijaya Ratna

45

46

Factorial Experiments

Prof. J. Vijaya Ratna

46

Factorial Experiments

47

- When the effects of different factors, each at different levels, on the results of the experiments are to be elucidated, FACTORIAL EXPERIMENTS are carried out.
- Designs of choice for simultaneous determination of the main effects and interaction effects of several factors.

47

Definitions

48

- Factor
- A factor is an assigned variable such as concentration ,temperature, lubricating agent , drug treatment or diet. A factor can be quantitative or qualitative.

48

Quantitative factor

49

Prof. J. Vijaya Ratna

- A quantitative factor has a numerical value assigned to it.
- Ex: concentration: 1%,2%,3%
- A qualitative factor has a name assigned to it rather than a value.
- Ex: treatments, diets, batches of materials, laboratories, analysts or tablet diluents

49

Factors

50

Prof. J. Vijaya Ratna

- Single factor designs fit the category of one way ANOVA design.

50

Levels

51

- The levels of a factor are the values or designations assigned to the factor.
- Ex: For the factor "temperature", levels may be 30° or 40°.
- For the factor "concentration", levels may be 0.1 molar or 0.2 molar.
- For the factor "drug treatment" levels may be drug and placebo.

51

Runs or trials

52

- The runs or trials that comprise factorial experiments consist of all combinations of all levels of all factors.

52

Example

53

- Ex: Investigation of the effects of drug concentration and lubricant concentration on dissolution time of a tablet.
- If both factors were at two levels (two concentrations for each factor), four runs (dissolution determinations for four formulations) would be required.

53

2^2 factorial

54

Drug	Lubricant	Symbol	Formulation
Low	Low	(1)	Low drug and low lubricant concentration.
Low	High	(A)	Low drug and high lubricant concentration
High	Low	(B)	High drug and low lubricant concentration
High	High	(Ab)	High drug and high lubricant conc.

54

EFFECTS

55

- The effect of a factor is the change in response caused by varying the levels of the factor.
- The main effect is the effect of a factor averaged over all levels of the other factors.
- In the example the main effect due to drug would be the difference between the average response when drug is at the high level (runs b and ab) and the average response when drug is at the low level (runs (1) and a).

55

Interaction

56

- Interaction may be defined as a lack of "additivity of factor effects".
- Ex: In a two factor experiment if factor A has an effect equal to 5 and if factor B has an effect of 10, additivity would be evident if an effect of 15 were observed when both A and B are at their high levels.

56

Interaction

57

- If the effect is greater than 15 when both factors are at their high levels, the result is synergistic with respect to the two factors.
- If the effect is less than 15 when both factors are at their high levels, an antagonistic effect is said to exist.
- The drug effect measured when the lubricant is at the low level (a-(1)) is DIFFERENT from the drug effect measured when the lubricant is at the high level (ab-b).

57

Interaction effect

58

- Factorial designs are the designs of choice for simultaneous determination of the effects of several factors and their interactions.

58

Interaction effect

59

- When a particular experiment involves more than one independent variable, the traditional method is to keep one variable (A) constant and change the other (B).
- After establishing the effect of B, in the next cycle B is kept constant and A is changed the effects of changing of A and B on the dependent variable Y are then separately reported. The problem in this type of experimentation is, we cannot detect interaction effect

59

Advantages

60

- In the absence of interaction, factorial designs have maximum efficiency in estimating main effects.
- If interactions exist, factorial designs are necessary to reveal and identify the interactions.
- Since factors' effects are measured over varying levels of other factors, conclusions apply to a wide range of conditions.

60

Advantages

61

- Maximum use is made of the data since all main effects and interactions are calculated from all of the data.
- Factorial designs are orthogonal, all estimated effects and interactions are independent of effects of other factors.

61

Applications

62

- The results of a factorial experiment may be used
- To help interpret the mechanism of an experimental system.
- To determine whether the main effects and the interaction effects are significant or not.
- To recommend to implement a practical procedure or set of conditions in an industrial manufacturing situation.

62

Factorial Design

63

- Many experiments are designed to find the following:
- What is the effect of a single independent variable on a single dependent variable?
- But experiments may be designed like this also:
- What are the effects of multiple independent variables on a single dependent variable?
- **Factorial Designs include multiple independent variables.**

63

Experimental Design Names

64

The name of an experimental design depends on:

- The number of independent variables
- The number of levels of each independent variable

64

Objectives of experiments

65

- Find whether the effects of the independent variables on the dependent variable is significant or not.
- Find whether there is interaction or not
- Separate the effects of the different independent variables
- Optimisation-means- find the exact level of the independent variable for a given value of the dependent variable-prepare the product with the predicted value, evaluate and find the error.

65

The Process

66

- Factorial designs and other optimisation designs give the trials that are to be carried out
- Analysis of variance is used to find the significance of the main effects and the interaction effects
- Depending on the design, simple graphs involving linear regression or multiple regression equations or contour plots and response surfaces are used to do optimisation.

66

Experimental design names

67

- The effect of a single variable is known as a main effect
- The effect of two variables considered together is known as an interaction

67

Experiment having only one factor

68

- One factor only at three levels

  Example:
- We alter the binder concentration at three levels, 10%, 20% and 30% levels, with the objective of producing a tablet with good hardness but disintegrates within 15 minutes.
- Then the table of formulations would be like this.

68

ONE FACTOR

69

Type of tablets	Hardness	Disintegration time
Tablets with 10% binder	---	---
Tablets with 20% binder	---	---
Tablets with 30% binder	---	---

69

One factor at three levels

70

- Here, there is only one factor being tested, concentration of binder, there is no question of interaction effect, hence by one-way ANOVA, we will determine whether the difference in hardness and disintegration time being observed is significant or not, whether it is really due to the change in the concentration of the binder or not.
- Then, we take concentration of binder on the X axis and draw lines for hardness and disintegration time on the two sides of Y axis.
- For disintegration time at 15 minutes, we can determine the hardness and the corresponding binder concentration on X axis. This would be the optimised formula. Here, Y is taken as a simple function of X.

70

One factor

71

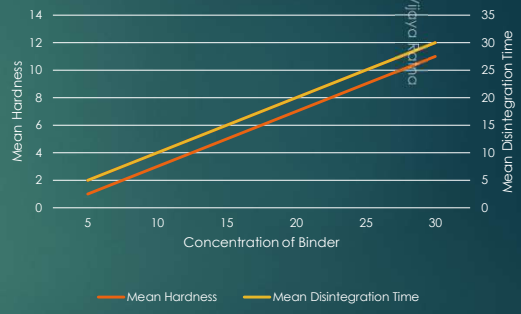
- If we take one factor at only two levels, we would get linear relationship for any dependent variable and we would miss any curvature that is present in the relationship. This is not the right way of experimenting.
- Taking one factor at three levels is only slightly better. Taking the single factor at as many levels as we can, within the space of the experimental region (say minimum required binder concentration to maximum advisable binder concentration, say 5% to 30%, is divided into six levels, 5%, 10%,------30%) would give more reliable results.
- Now, if the results are analysed taking Y as a function of X, the highest binder concentration that allows tablet disintegration within 15 minutes may be taken as the optimised concentration.

71

72

Tablets	Mean Hardness (Kg/cm²)	Mean Disintegration Time (mins)
2	1	5
4	3	10
6	5	15
8	7	20
10	9	25
12	11	30

Tablets	Mean Disintegration Time (mins)
5	10
10	20
15	30
20	40
25	50
30	60



Effect of Binder concentration on hardness and disintegration

72

ANALYSIS OF VARIANCE

73

- Analysis of variance is a statistical procedure designed to analyze the difference between the means of two or more samples.
- Rationale: Total variance of all the data in an experiment can be separated and attributed to two sources, variance between groups and variance within groups.
- Variance within groups is also called error variance.
- If the experimental treatment has been effective, the between groups differences will be expected to be greater than can be accounted for by chance.

73

PURPOSE OF ANOVA

74

- To establish whether variation between groups is likely to be a function of chance or not.
- Example: Comparison of three batches of tablets with their dissolution times

Tablets	Dissolution times				
Batch I	77	81	71	76	80
Batch II	72	58	74	66	70
Batch III	76	85	82	80	77

74

Sum of Squares

75

- 1. Sum of squares within groups is the sum of squares of the deviations of each individual score in each group from the mean of the group.

$$SS_W = \Sigma(X_1 - \overline{X}_1)^2 + \Sigma(X_2 - \overline{X}_2)^2 + \Sigma(X_3 - \overline{X}_3)^2$$

through all the groups.

SS_W is sum of squares within the groups

X_1 is mean of first group

X_1 is an individual score in the first group

K is number of groups.

75

Partitioning the sum of squares

76

- The basic ingredient of ANOVA is the sum of squares.
- Sum of squares is a measure of variability.
- 2. Total sum of squares is the sum of squares of the deviations of each of the observations from the grand mean.

$$SS_T = \Sigma\left(X - \overline{X}_T\right)^2$$

Where SS_T is total sum of Squares

X is individual score

$\overline{X}_T$ is mean of all the scores.

76

Sum of squares

77

- 3. Sum of squares between groups is the variation that may be due to the experimental treatment. It is the sum of squares of the deviation of each separate group mean from the grand mean.

$$SS_b = n_1(X_1 - \overline{X}_t)^2 + n_2(X_2 - \overline{X}_t)^2 + n_3(X_3 - \overline{X}_t)^2$$

through all the groups

SS_b → Sum of squares between

X_1 → mean of first group

n_1 → the number in the first group

X_t → the mean of all the scores

$$MS_W = \frac{SS_W}{d.f._W} = \frac{SS_W}{N-K} \qquad \mathrm{MS_b} = \frac{\mathrm{SS_b}}{\mathrm{d.f_b}} = \frac{SS_b}{k-1}$$

N-1 = (N-K) + (K-1)

Total = within + Between

$F = MS_b / MS_W$

Prof. J. Vijaya Ratna

77

Computational formulas for sums of squares

78

$$SS_T = \Sigma(X - \overline{X}_T)^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

- SS_T → Total sum of Squares
- X → Each Individual score
- $\overline{X}_T$ → Grand mean of all the N cases
- N → Total Number of observations
  $= n_1 + n_2 + n_3$
- Σx^2 → Sum of the squares of each raw score
- $(\Sigma x)^2$ → Square of the sum of the raw scores

Prof. J. Vijaya Ratna

78

Computational formulas for sums of squares

79

Prof. J. Vijaya Ratna

2. Sum of squares between groups

$$SS_b = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \text{-----} \frac{(\Sigma X_k)^2}{n_k} - \frac{(\Sigma X)^2}{N}$$

- SS_b Sum of squares between the groups
- ΣX_1 Sum of the first group
- N_1 Number in the first group
- ΣX Sum of all the scores
- N Total number of scores.

79

Computational formulas for sums of squares

80

Prof. J. Vijaya Ratna

3. Within groups sum of squares

$$SS_w = \Sigma X_1^2 - \frac{(\Sigma X_1)^2}{n_1} + \Sigma X_2^2 - \frac{(\Sigma X_2)^2}{n_2} + \text{-----} + \Sigma X_K^2 - \frac{(\Sigma X_k)^2}{n_k}$$

- SS_W = Sum of squares within the groups
- εX_1^2 = Sum of the squares of the scores in the first group.
- n_1 = the number in the first group
- SS_W = SS_T - SS_b

80

SCHEMATIC OF ONE – WAY ANOVA

81

Source of Variation	Degrees of freedom	Sum of Squares	Mean Square	F
Treatments (Between)	K-1	SS_{Tr} or SS_b	$MS_b = SS_b/K-1$	$MS_b/MS_{W=}$ F
Error (Within)	K (n-1) or N-K	SS_E or SS_W	$MS_W = SS_W / N-K$	
Total	$K_{n-1\ or}$ N-1	SS_T		

81

Test Procedure

82

- 1. Null hypothesis : $\mu_1 = \mu_2 = \mu_3$
  Alternative hypothesis : μ's are not all equal.
- 2. Level of significances : $\alpha = 0.05$
- 3. Criterion: Reject the null hypothesis if F>3.89, the value of F0.05 for K-1=3-1=2 and N-K=15-3=12 degrees of freedom where F is to be determined by an analysis of variance, otherwise accept it.

82

Calculations

83

- 4. Calculations:

$$SS_T = \sum X_i^2 - \frac{(\sum X_i)^2}{N}$$

$$SSb = \frac{(\sum X_i)^2}{n1} + \frac{(\sum X_2)^2}{n2} + \text{--------} - \frac{(\sum X_i)^2}{n}$$

$$SS_W = SS_T - SS_b$$

- 5. ANOVA table
- 6. Decision: If F value obtained exceeds 3.89 reject H_0, otherwise accept H_0

83

One way ANOVA

84

84

Partitioning

85

Variance is partitioned between two sources,

- Variance between groups and
- Variance within groups.

Prof. J. Vijaya Ratna

85

Chance Variance

86

- Chance Variance
- A variance which occurs due to chance reasons, which is very small in scale and which occurs on both positive and negative sides.
- Also called as error variance.

Prof. J. Vijaya Ratna

86

Real Variance

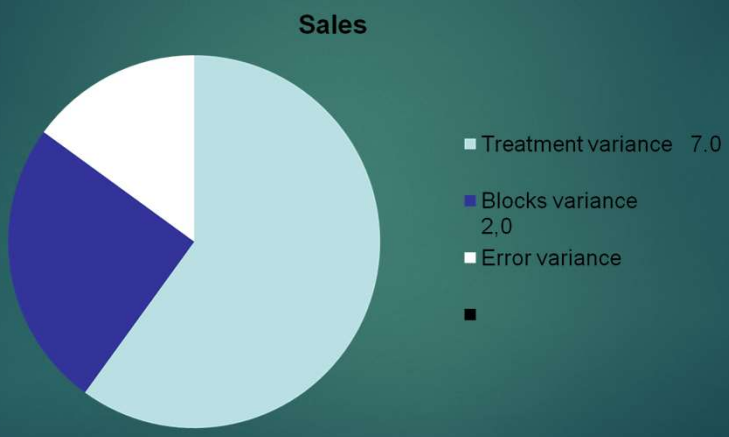
87

- Real variance is due to a real reason and often is the effect of a treatment.
- When a treatment is effective there is a clear, significant difference in the result.
- So ANOVA consists of separating the variance into two categories, that which is due to a real reason and that which is due to a chance reason.

87

Two way ANOVA

88



88

TWO WAY ANALYSIS OF VARIANCE

89

- $SS_T = SS\ (T_r) + SSB + SSE$
- $SSE = SST - [SS\ (T_r) + SSB]$
- SSB → Sum of squares between Blocks
- $SS\ (T_r)$ → Sum of squares between Treatments
- SSE → Error sum of Squares

89

SCHEMATIC OF TWO WAY ANOVA

90

Source of variation	Degrees of freedom	Sum of squares	Mean Square	F
Treatments	K-1	$SS\ (T_r)$	$MS\ (T_r) = SS(T_r)/\ K-1$	$\frac{MS(T_r)}{MSE}$
Blocks	n-1	SSB	$MSB = SSB/n-1$	$\frac{MSB}{MSE}$
Error	(K-1) (n-1)	SSE	$MSE = \frac{SSE}{(K-1)(n-1)}$	
Total	(N-1)	SST		

90

Three way ANOVA

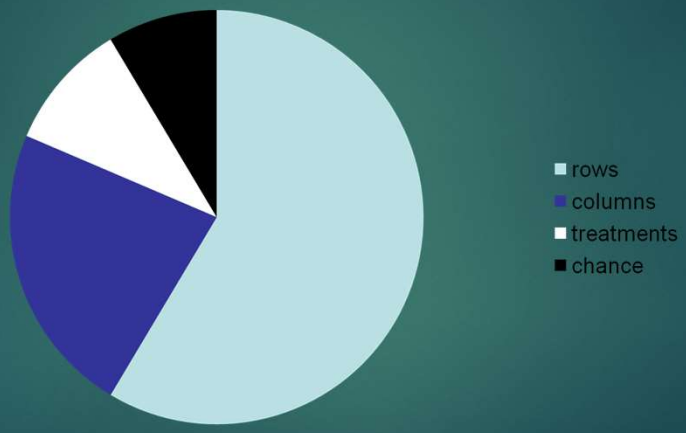
91

- In Three way ANOVA, the total variability in the data is subdivided into four categories- rows, columns, treatments, and chance.
- Three F ratios are calculated.
- Three decisions are taken.
- Significance of column effect, row effect and treatment effect are determined.

91

Three way ANOVA

92



92

THREE WAY ANALYSIS OF VARIANCE

93

$$SS_T = \Sigma(X_1)^2 - \frac{(\Sigma Xi)^2}{N}$$

$$SS_{rows} = \frac{(\Sigma X_{1r})^2}{n_{1r}} + \frac{(\Sigma X_{2r})^2}{n_{2r}} + \ldots\ldots\ldots\frac{(\Sigma X_i)^2}{N}$$

$$SS_{columns} = \frac{(\Sigma X_{1c})^2}{n_{1c}} + \frac{(\Sigma X_{2c})^2}{n_{2c}} + \ldots\ldots\ldots\frac{(\Sigma X_i)^2}{N}$$

$$SS_{Treatments} = \frac{(\Sigma X_{1tr})^2}{n_{1tr}} + \frac{(\Sigma X_{2tr})^2}{n_{2tr}} + \ldots\ldots\ldots\frac{(\Sigma X_i)^2}{N}$$

$$SS_E = SS_T - [SS_R + SS_C + SS_{tr}]$$

93

SCHEMATIC OF THREE WAY ANOVA

94

Source of Variation	Degrees of freedom	Sum of squares	Mean square	F
Rows	r-1	SSR	MSR = $\frac{SSR}{r-1}$	$\frac{MSR}{MSE}$
Columns	r-1	SSC	MSC = $\frac{SSC}{r-1}$	$\frac{MSC}{MSE}$
Treatments	r-1	$SS(T_r)$	$MS(T_r) = \frac{SS(T_r)}{r-1}$	$\frac{MS(T_r)}{MSE}$
Error	(r-1) (r-2)	SSE	MSE= $\frac{SSE}{(r-1)(r-2)}$	
Total	r^2-1	SST		

94

Linear regression

95

- When we are testing to find the significance of one factor, we do one way analysis of variance to find the significance of the factor.
- To carry out optimisation by interpolation, we plot y on X, or write an equation for Y on X and solve it for a given Y. This is done by linear regression or by transforming the variables if the relationship is not linear.

95

Statistical design of experiments

96

- When the number of independent variables is more than one we have to go for "statistical design of experiments and multiple linear regression analysis".

96

Optimization

97

- Optimization techniques help us to
- (1) select experimental points so that the entire area of interest is covered or considered;
- (2) separate the effects of the variables, i.e. we can tell which variable caused a particular result.

Prof. J. Vijaya Ratna

97

Factorial designs in optimisation

98

- Factorial Design is the most widely used method of optimization.
- We set up the design,
- We carry out the experiments
- Generate the data
- We write multiple regression equations that express the relationships between the variables. These equations are the basis of optimization and they define the response surface for the system under investigation

Prof. J. Vijaya Ratna

98

Example

99

- Example : Imagine an old balance with a zero error of 1kg. A fat man comes and stands on it and the balance says 122kg. We announce his weight as 121kg.
- A child comes and stands on the balance and it shows 23kg. We announce that the child's weight is 22kg.

99

Example

100

- In this experiment we could have missed some interaction effect.
- The balance's zero error may be changing with weight.

100

Interaction effect

101

- If (ab-b) = (a-1)
- then there is no interaction. The difference between these two is the interaction effect

101

Example

102

- So what should be done is this.
- Take the weight of the empty balance (1)
- Take the weight of the balance + child (a)
- Take the weight of the balance + fat man (b)
- Take the weight of the balance + child + fat man (ab)
- The weight of the child $= \frac{((ab)-(b) + ((a)-(1)}{2}$
- The weight of the fat man $= \frac{(b-1) + (ab-a)}{2}$

102

Example

103

- If (ab-b) = (a-1)
- then there is no interaction. The difference between these two is the interaction effect

103

Optimization Procedure

104

- In this procedure we construct an equation that describes the experimental results as a function of the factor levels.
- A polynomial equation can be constructed, in the case of a factorial design, where the coefficients in the equation are related to the effects and interactions of the factors.
- If we describe a factorial design as 2^n, n stands for the number of factors and 2 stands for the number of levels.

104

More than one factor

105

- The polynomial for a 2^n design will be of the form.

$$Y=B_0+B_1X_1+B_2X_2+B_3X_3+B_{12}X_1X_2+B_{13}X_1X_3+B_{23}X_2X_3+B_{123}X_1X_2X_3+$$

- Where Y is the measured response X_i is the level of the i^{th} factor, B_i, B_{ij} etc. represent coefficients computed from the responses of the formulations in the design.

Prof. J. Vijaya Ratna

105

Two factors each at two levels

106

- We want to prepare a tablet which has a high dissolution rate. So dissolution rate is response Y.
- We know that two factors are having effect on this response Y, they are (a) drug concentration and (b) lubricant concentration.

Prof. J. Vijaya Ratna

106

We decide to test each factor at two levels. The design of the experiment is like this

107

Drug	lubricant	Symbol	Formulation
X_1	X_2		
Low	Low	(1)	Low drug and low lubricant concentration
Low	High	(a)	Low drug and high lubricant concentration
High	Low	(b)	High drug and low lubricant concentration
High	High	(ab)	High drug and high lubricant concentration

107

2^2

108

- (1) We prepare the four formulations. We determine the dissolution rates for them. Then we find the effects like this:
- Main effect of drug = $\frac{[b+ab-(1)-a]}{2}$
- Main effect of lubricant = $\frac{[ab+a-(b)-(1)]}{2}$
- Interaction effect will be the difference between (ab-b) and (a-1)
- and the difference between (ab-a) and (b-1)

108

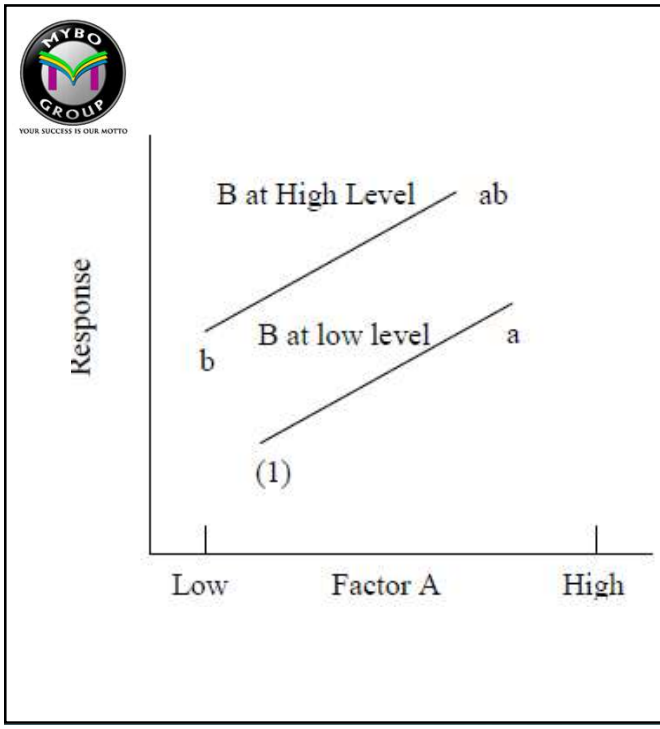
Interaction

109

Prof. J. Vijaya Ratna

- Second step is to find the interaction, graphically and numerically.

109



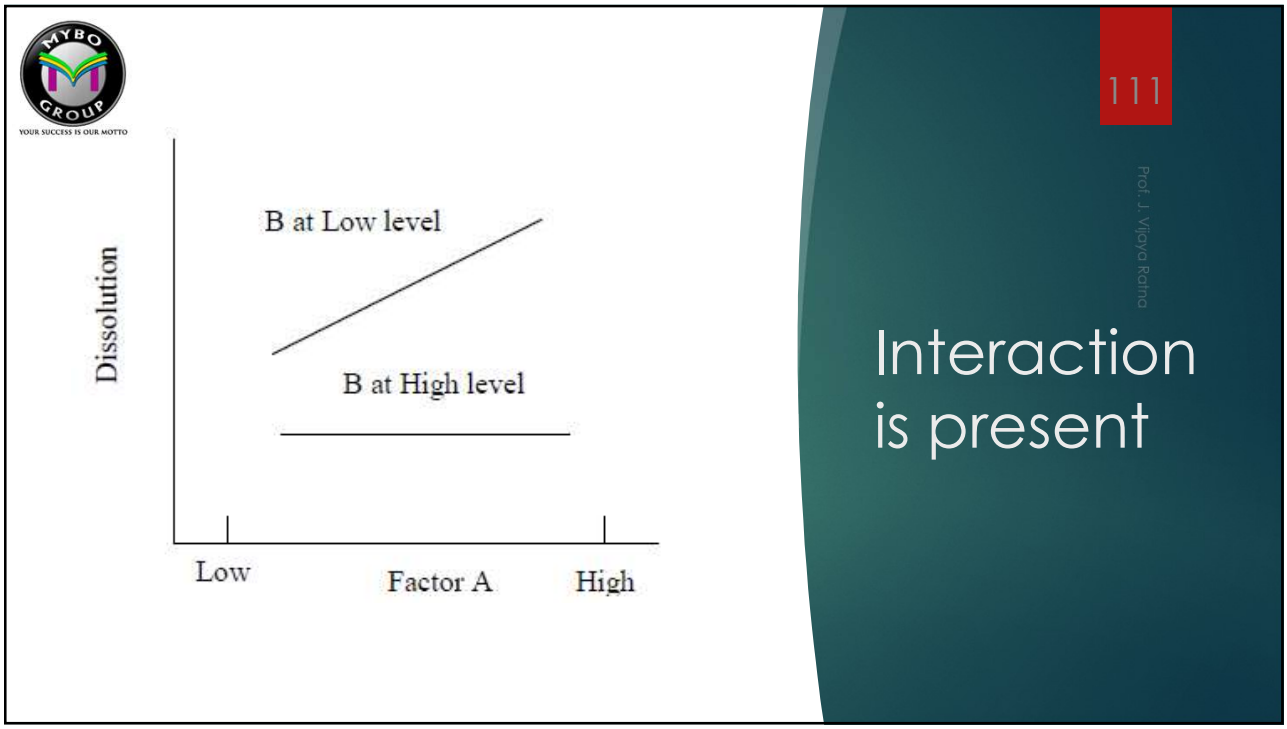
110

Prof. J. Vijaya Ratna

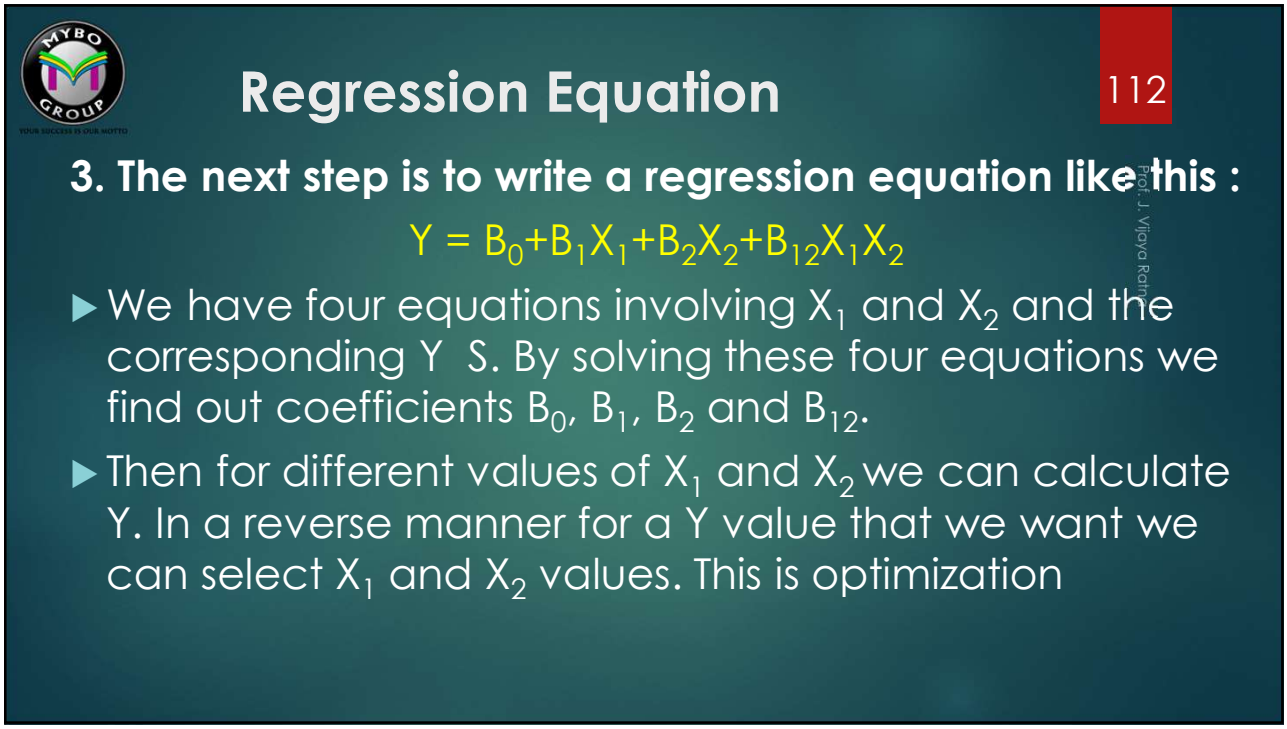
No interaction

If there is no interaction they will look like this.

110


111
Prof. J. Vijaya Ratna
B at Low level
B at High level
Dissolution
Low
Factor A
High
Interaction is present

111


Regression Equation
112
Prof. J. Vijaya Ratna
3. The next step is to write a regression equation like this :
$Y = B_0+B_1X_1+B_2X_2+B_{12}X_1X_2$
We have four equations involving X_1 and X_2 and the corresponding Y S. By solving these four equations we find out coefficients B_0, B_1, B_2 and B_{12}.
Then for different values of X_1 and X_2 we can calculate Y. In a reverse manner for a Y value that we want we can select X_1 and X_2 values. This is optimization

112

Three factors, each at two levels

113

- In doing analysis of a 2^3 factorial design experiment, we carry out analysis of variance to find the significance of each and every factor and to find the interaction effects.
- We use contour plots and multiple regression equations for finding the values of the independent variables at the given value of dependent variable.

113

Example of worked out problem

114

- Experiment was done with three factors, each at two levels.
- There is no replication, means there is no replication of the design. Each effect value is a mean value of three or more batches under the same condition.
- Objective of the experiment was to determine the effects of three factors- lubricant, drug and binder on the thickness of the tablets.

114

2^3

115

- Each factor was taken at only two levels. Taking them at a higher level would involve more trials and more budget.
- Because there are only two levels, non-linear responses cannot be obtained.
- Researcher felt this would be sufficient to design the final formulation.

115

Factors: lubricant Drug binder Response: (thickness)

116

(1)	-	-	-	475
A	+	-	-	487
B	-	+	-	421
Ab	+	+	-	426
C	-	-	+	525
Ac	+	-	+	546
Bc	-	+	+	472
abc	+	+	+	522

116

2^3

117

Factor	Low level (mg)	High level (mg)
A (lubricant)	0.5	1.5
B (drug)	60.0	120.0
C (binder)	30.0	50.0

117

2^3

118

- Calculation of main effects, interaction effects and ANOVA can be done by hand in simple designs like this.
- If design is more complex, computers may be used.
- If number of factors is n, then n-way analysis of variance is to be done.

118

Calculations in 2^3

119

- For two level factorials, effects can be calculated by applying + and – signs arithmetically for each of the eight responses.
- We construct a table by placing a + or – sign in columns A,B and C, depending on whether or not the appropriate factor is at the high level or low level in the specific run.
- If the letter appears in the specific factor combination, a + appears in the column corresponding to that letter.

119

2^3

120

- For the combination ab, a + appears in columns A and B and a – appears in column C.
- For column A, runs a, ab, ac, abc show a +, because in these runs, A is at high level. For runs, 1, b, c, and bc, a – appears in column A, because these runs have A at its low level.
- Columns labelled as AB, AC, BC and ABC indicate interaction terms. The signs in these columns are obtained by multiplying the signs of the individual components.

120

2^3

121

- To get the sign of AB, we multiply the signs of A and B in that run; in (1), it is +, in a, it is –in c, it is + and so on.

121

Main effects

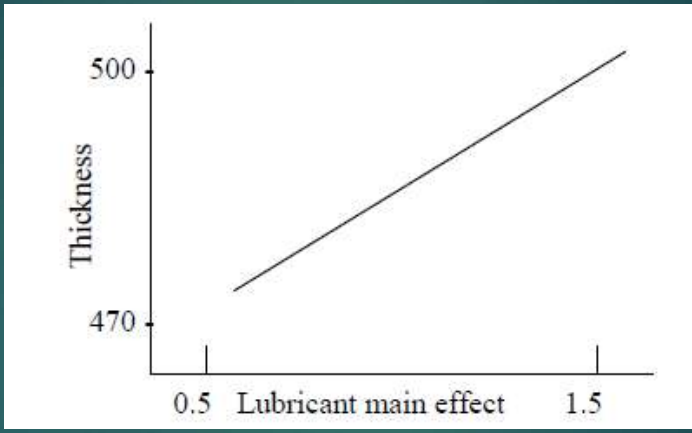
122

- Main effect of A is the average of all results at the high level of A minus the average of all results at the low level of A. Main effect of
- A= {a+ab+ac+abc} /4 - {(1)+b+c+bc}/4
- Substituting the results given in the table, we get 0.022 cm.
- Interpretation: The net effect of increasing the lubricant from its low level to its high level (averaged over all other factors), is to increase the tablet thickness by 0.022 cm.

122

Main effect

123



123

Interaction effect

124

- The interaction AC is one half of the difference between the effect of A when C is at the high level and the effect of A when C is at the low level. Interaction
- AC= ¼ { (abc+ac-bc-c) – [ab+a-b-(1)]
- At high binder level, when lubricant is increased from low to high the thickness of tablet increases by 0.0355 cm.

124

2^3

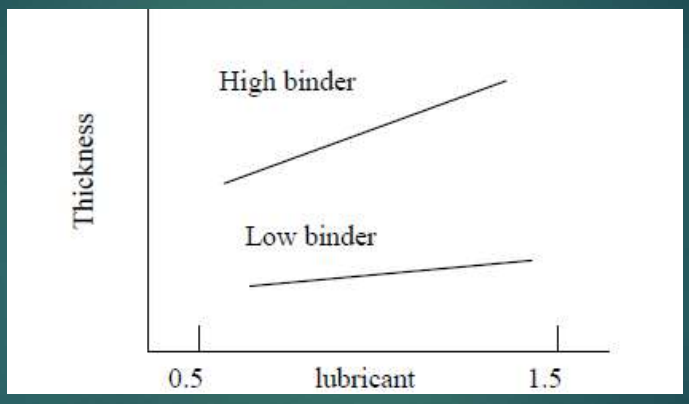
125

- At low level of binder, increasing lubricant concentration results in an increased thickness of 0.0085 cm.
- Indication of possible lubricant and binder interaction; lubricant has a greater effect at higher binder concentration.

125

Interaction

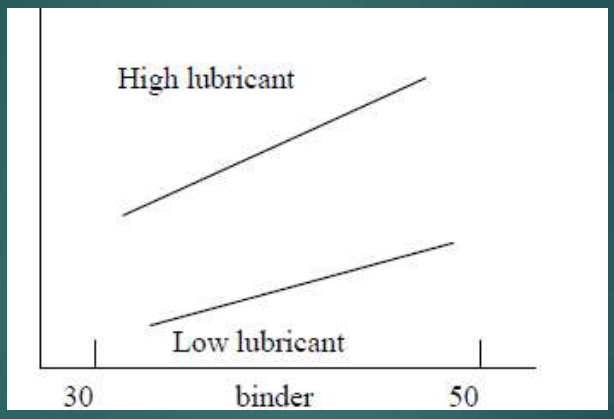
126



126

Interaction

127



127

Interaction

128

- The effect of lubricant is greater at high binder concentrations
- The effect of binder concentration is greater at the high lubricant concentration.

128

Yate's method to do ANOVA

129

- Computers are used to do analysis in Factorial designs. But doing calculations with calculator helps us understand the design.
- Method for 2^n factorial: Data are first arranged in standard order. The data are first added in pairs, followed by taking differences in pairs.

129

Yate's method

130

- 475+487=962
- 421+426=847
- 525+546=1071
- 472+522=994
- 487-475=12
- 426-421=5
- 546-525=21
- 522-472=50

Run	Thickness ($\times 10^3$)
(1)	475
a	487
b	421
ab	426
c	525
ac	546
bc	472
abc	522

130

Yate's method

131

- This addition and subtraction process is repeated sequentially on the n columns. N is the number of factors.
- The process is repeated in column 2 , working on the results in column 1. Column 3 comes from working on values in column 2.
- The process is repeated n times and by the n^{th} column the grand total value of all runs comes at the top of the column.

131

Yate's method

132

- Column 3 is divided by 2^{n-1} .
- This is 4 for 3 factors.
- This will give the effect of the run in the 4^{th} column.
- The mean square of a run is obtained by dividing the square of the value in n^{th} column by 2^n .
- The mean square of factor A is $(88)^2/ 8=968$.

132

Yate's method- finding effects

133

Run	Thickness ($x 10^3$)	(1)	(2)	(3)	Effect ($x 10^3$) (3)/ 4	Mean square ($x 10^6$) $(3)^2$ / 8
(1)	475	962	1809	3874	----	----
a	487	847	2065	88	22.0	968
b	421	1071	17	-192	- 48.0	4608
ab	426	994	71	22	5.5	60.5
c	525	12	-115	256	64.0	8192
ac	546	5	-77	54	13.5	364.5
bc	472	21	-7	38	9.5	180.5
abc	522	50	29	36	9.0	162

133

Analysis of variance

134

Factor	Source	d.f	Mean square ($x 10^6$)	F	Probability	Significance
A	lubricant	1	968	7.2	P< 0.1	NS
B	drug	1	4608	34.3	P< 0.01	S
C	binder	1	8192	61.0	P< 0.01	S
AB	Lubricant x drug	1	60.5			
AC	Lubricant x binder	1	364.5	2.7		
BC	Drug x binder	1	180.5			
ABC	Lubricant x drug x binder	1	162			

134

Analysis of variance

135

- F is calculated based on a error mean square based on AB, BC and ABC interactions, 3 d.f.
- For Factor A, $p < 0.1$
- For Factors B and C, $p < 0.01$
- The effects of drug and binder are significant.

135

Analysis of variance

136

- We calculate: Total sum of squares,
  - Lubricant sum of squares
  - Drug sum of squares,
  - Binder sum of squares, and
  - Interactions sum of squares for ab, ac, bc and abc.
- We calculate F values and referring to the tables we can tell whether each observed difference is significant or not.
- We write a polynomial equation of the type
- $Y = B_0+B_1X_1+B_2X_2+B_3X_3+B_{12}X_1X_2+B_{13}X_1X_3+B_{23}X_2X_3+B_{123}X_1X_2X_3+\ldots$
- We write 8 equations based on our results and solve them to get B_0, B_1, B_2, B_3, B_{12}, B_{13}, B_{123} and then we can develop graphs.

136

Factorial in optimization

137

- We use these graphs and these equations to select X_1, X_2 and X_3 values for expected values of Y.
- This is optimization.

137

138

- For a 2^3 design, we set up the design and we do the experiments. We determine the result Y, it may be the hardness or disintegration time of the tablet. So we have 8 different Y values for 8 different formulations.
- We write 8 different simultaneous equations. For example

  (1) $0X_1+0X_2+0X_3=Y_1$ (some Y value determined)

  (4) $2X_1+1X_2+0X_3=Y_4$ (some determined Y value)

138

139

- We solve these 8 equations (with the help of a computer) to get the coefficients

  B_0, B_1, B_2, B_3, B_{12}, B_{23}, B_{123} etc.

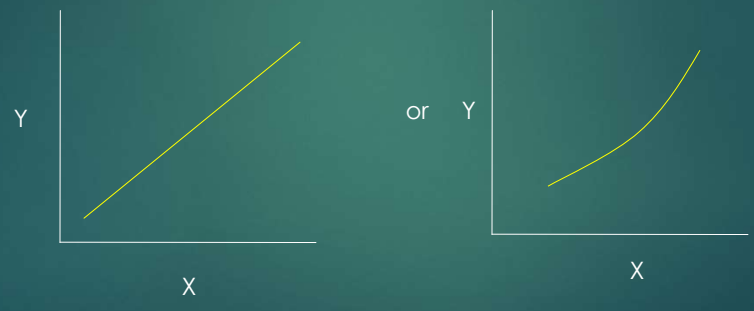
- Then we write the polynomial equation of the form,

$$Y = B_0+B_1X_1+B_2X_2+B_3X_3+ \ldots\ldots +B_{12}X_1X_2+B_{13}X_1X_3+ \ldots\ldots +B_{123}X_1X_2X_3$$

139

140

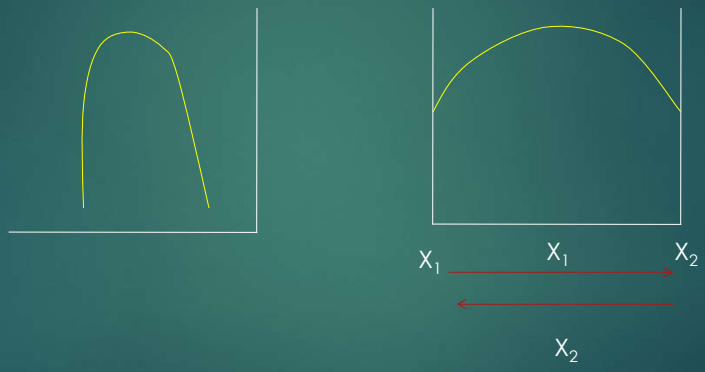
- Using this equation we can calculate the response Y for unknown X_1, X_2, X_3 values.
- Using this equation we can also develop a response surface. If the number of a. X is 1 the graph will be like



140

(b) X is 2 the graph will be like

141



141

142

- Utilizing the equation and the graph we can predict the response for a different set of X_i^s and we can select that set of X_i^s which can give us the "optimum" or "as perfect as possible under the given conditions" response Y.
- This is optimization using factorial design.

142

Factorial design calculation

143

- Take the case of 2^3 :three factors with two levels;

Factor 1	A	low	high
Factor 2	B	low	high
Factor 3	C	low	high

- Result is judged.
- Total 8 treatments are there. Let us repeat the entire design 4 times. In each replication treatments are allotted randomly. We call each treatment as one block. So the design is as following.
- A 2^3 factorial design, laid out in four blocks, to determine the effect of different kinds of factors such as A, B and C on Y, the dependent variable.

143

2^3 factorial design in four replications

144

- A 2^3 factorial design, laid out in four blocks, to determine the effect of different kinds of factors such as A, B and C.
- 1 means all three factors are at low level
- A means A is at high level and b and c are at low levels
- B means only b is at high level
- C means only c is at high level
- ABC means all three factors are at high level

144

2^3 factorial design in four replications

145

- Do calculations as for ANOVA.
- Total SS; RSS, CSS and ESS in the usual way. Find f ratios and determine significance

145

Yate's method in 2^3 factorial design in four replications

146

- Then to calculate main effects and interaction effects take in a table, totals of each row; add r1and r2, r3 and r4, r5 and r6, and r7 and r8. You will get four values. Then do r2-r1, r4-r3, r6-r5 and r8-r7 .

146

Yate's method in 2^3 factorial design in four replications

147

- Continue this process till you get in the first row the grand total.
- This column gives the main effects.
- These should be divided by a divisor to get the value of mean sum of squares.

147

Yate's method in 2^3 factorial design in four replications

148

- The divisor is no. of replications multiplied by 2^3, in this case $4*2^3=32$.
- So each SS is to be squared and divided by 32.
- This MSS is to be divided by ESS to get F ratio which is to be compared with critical value to get result of significance or not.

148

Randomization

149

- The experimental design consists of doing 8 runs, in four replications.
- Each replication is a block containing all runs.
- Which run takes which place in the block is decided by going for randomisation.

149

Randomization

150

- Although we write our test matrix in this standard order, we should actually perform the 8 experiments in RANDOM order.
- Randomization will make any factor we overlooked likely to contribute to random uncertainty rather than systematic error

150

151

Block 1

AB 284	I 98	AC 372	A 257
B 111	ABC 446	C 302	BC 361

Block 2

AC 385	ABC 429	C 317	I 105
A 283	B 95	BC 328	AB 311

Block 3

C 305	AB 320	I 85	B 13
BC 335	A 267	AC 399	ABC 464

Block 4

BC 351	AC 422	A 298	ABC 452
B 99	I 126	AB 277	C 313

151

152

Calculation

	Block 1	Block 2	Block 3	Block 4
I	98	105	85	126
A	257	283	267	298
B	111	95	133	97
AB	284	311	320	277
C	302	307	305	313
AC	372	385	399	422
BC	361	328	335	351
ABC	446	429	464	452

152

Treatment Totals

153

Treatment	Total	Square	Square/4
I	414	1,71,396	42,849
A	1105	12,21,025	3,05,256.25
B	463	1,90,096	47,524
AB	1192	14,20,864	3,55,216
C	1227	15,05,529	3,76,382.25
AC	1578	24,90,084	6,22,521
BC	1375	18,90,625	4,72,656.25
ABC	1791	32,07,681	8,01,920.25

153

Block Totals

154

Block	1	2	3	4
Total	2231	2253	2308	2336
Square	4,97,7361	50,76,009	53,26,864	5,45,56,896
Square/8	6,22,170	6,34,501	6,65,858	6,82,112

Grand total of all values = 9128

154

155

- Grand total = 9,128
- Correction factor = 26,03,762
- Sum of squares of blocks = 8,75,879.3
- Sum of squares of treatments = 1,21,21,940

155

ANOVA Table

156

Source of variation	SS	df	MSS	F	Table F	Result
Total	433,618	31	13,987.68			
RSS	426,723	7	60,960.43	212.8	2.49	S
CSS	879.25	3	293.08	1.23	3.07	NS
ESS	6015.75	21	1286.464			

156

Yate's method

157

Treatment				
I	414	1519	3147	9128
A	1105	1628	5981	2204
B	436	2815	1447	460
AB	1192	3166	757	140
C	1237	691	109	2834
AC	1578	756	351	-690
BC	1375	341	65	242
ABC	1791	416	75	10

157

Main Effects and Interaction Effects

158

Treatment	df	Yate's total	$(\text{Yate's Total})^2/32$	F	Table Value of F	Result
I	1	9128				
A	1	2204	151,800	530	4.32	S
B	1	460	6612.5	23.08	4.32	S
AB	1	140	612.5	2.1	4.32	NS
C	1	2834	250,986	877.5	4.32	S
AC	1	-690	14,878	52.02	4.32	S
BC	1	242	1830.1	6.39	4.32	S
ABC	1	10	3.125	0.01		NS

S = Significant
NS = Not Significant

158

Conclusions

159

- In pharmaceutical and in biotechnological research, process and product development are very critical.
- When we employ appropriate designs in their development, the products would be robust and would always meet the requirements of the regulators and the patients.

159

References

160

- Mandlik Satish K, Adhikari Saugat, Deshpande Ameya A, Application of Simplex Lattice Design in Formulation and Development of Buoyant Matrices of Dipyridamole. Journal of Applied Pharmaceutical Science Vol. 2 (12), pp. 107-111, December, 2012.
- Schwartz, J., Schnaare, R., & O'Connor, R. (2002). *Optimization Techniques in Pharmaceutical Formulation and Processing. Drugs and the Pharmaceutical Sciences. Marcel Dekker Inc.*
- Websters New Collegiate Dictionary, G. & C. Merriam, Springfield, MA, 1974.
- L. Cooper and D. Stienberg, Introduction to Methods of Optimization, W.B. Saunders, Philadelphia, 1970.
- S. Bolton and C. Bon. Pharmaceutical Statistics Practical and Clinical Applications. Drugs and the Pharmaceutical Sciences. Fourth Edition, Marcel Dekker Inc. P 506-538

160


Thank You
For any queries
feel free to drop a mail at
vijaya.ratna@gmail.com
mybogroup@gmail.com
161
Prof. J. Vijaya Ratna

161