Lecture 1
Experimental optimisation

Eddie Schrevens

Experimental Optimisation

Definition of the region of interest or the factor space. Determination of the important factors and their ranges, taking into account the nature of the factor.

Development of predictive mathematical models, describing the relation between the independent variables and the dependent variables and/or interrelations between variables and/or group structures in the data.

Construction of an optimal design (or sampling strategy) that spans the factor space in relation to the assumed models.

Optimisation phase: obtaining the optimal operating conditions by investigation of the mathematical model.

Experimental region, factor space, region of interest

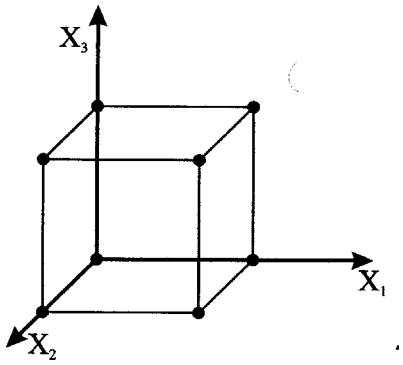
The region of interest is the hyper-volume spanned by the points in the treatment matrix. Thus a treatment matrix has a geometrical equivalent: fi for 3 experimental factors X_i with 2 levels per factor (0, 1), resulting in 8 treatments.

Treatment matrix

Treatment	X_1	X_2	X_3
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1
5	1	1	0
6	1	0	1
7	0	1	1
8	1	1	1

Geometrical representation of the experimental region

Cuboidal experimental region

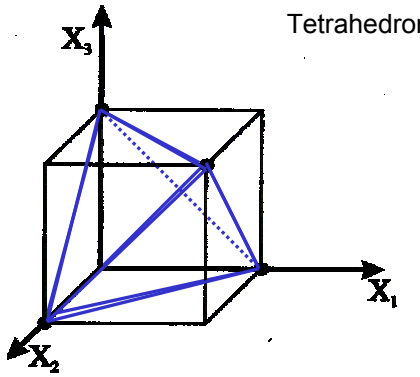


Experimental region, factor space, region of interest

Lets delete 4 treatments → treatment matrix

Treatment	X_1	X_2	X_3
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1
5	1	1	0
6	1	0	1
7	0	1	1
8	1	1	1

Geometrical representation



Tetrahedron

Defining the experimental region

- Experimental factors, fi X_1, X_2, X_3, …
- Levels for each experimental factor, fi 0, 1

↓

Range for each factor

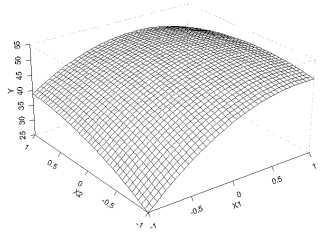
- Treatments = combinations of certain levels for each factor
- Treatment matrix=vertical concatenation of treatments

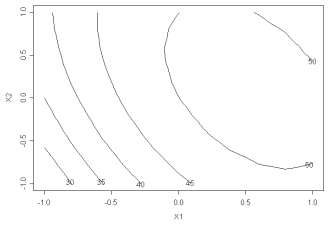
Experimental region = convex hull of the treatment matrix

Predictive models

General(ised) (non)Linear Model

Relation between independ variable and exp factors

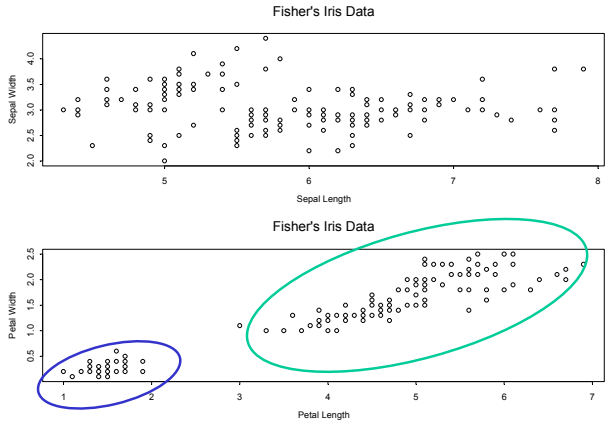




Predictive models

Classification (clustering) model

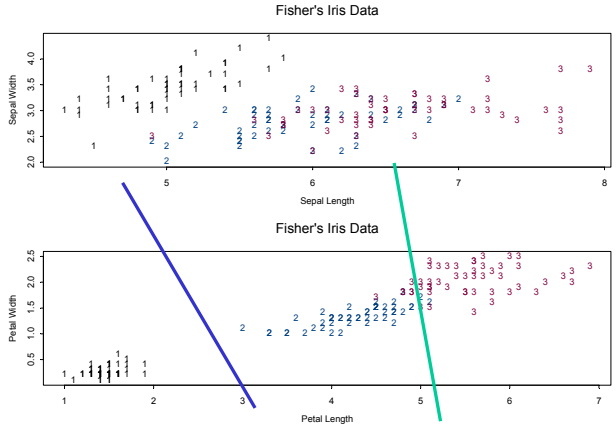
Discovering group structures



Predictive models

Discrimination model

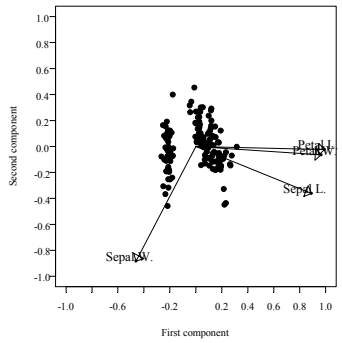
Modelling existing group structures



Predictive models

Ordination model

Modelling interrelations between variables



PCA Biplot on IRIS data

Predictive Models

Modelling the relation between indep variable and exp factors

- **It is assumed that a response y depends on a set of variables $x_1, x_2, \ldots, x_k$. It is assumed that the x's can be controlled by the experimenter with negligible error in comparison with the variance of y. In general the x's are called independent variables or experimental factors. The response y is called the dependent variable and is considered to be numeric.**

The dependence of y on the x's can be formalised as

$$y = f(x_1, x_2, \ldots, x_k)$$

General polynomial models

- When the mathematical form of the function f is unknown, in most cases this function can be approximated satisfactorily, within the experimental region, by a polynomial function in the x's. It should be emphasised that the polynomial approximation can only be used within the experimental region in other words within the region spanned by the factor combinations of the x's, because polynomials are notoriously untrustworthy when extrapolated.

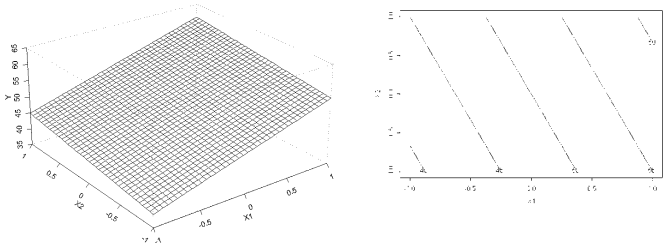
First order polynomial models

- A linear approximation function, also called a first order model can be written as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

- This approximation is useful when the function f is studied in a narrow range of the x_1, x_2, …, x_k. In other words the range of the x's is so small that little curvature is to be expected in f. The fitting of such a polynomial can be considered as a particular case of multiple linear regression. In order to estimate the regression coefficients β_i in this model each independent variable x_i should take at least 2 levels. This model is able to estimate the main effects. To fit this model at least k+1 data points are necessary.

First order polynomial models

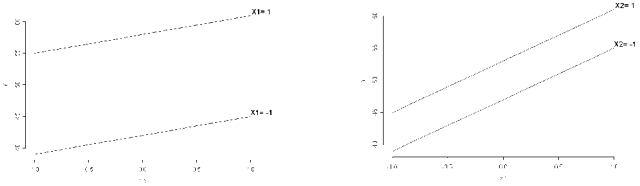


Different graphical representations of the first order model:

$y = 50 + 8X_1 + 3X_2$

Perspective plot and Contour plot

First order polynomial models



Different graphical representations of the first order model:

$y = 50 + 8X_1 + 3X_2$

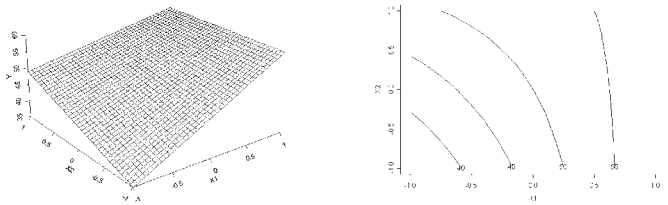
Line plot $y^*X_2=X_1$ and $y^*X_1=X_2$

First order polynomial models + interaction

- **A slightly different model, consists of the linear model expanded with 2-factor interaction terms. This is formalised as follows. To fit this model 2 levels per factor and replications are necessary. At least 1 + k + k(k-1)/2 data points are required.**

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i<j} \sum \beta_{ij} x_i x_j$$

First order polynomial models + interaction

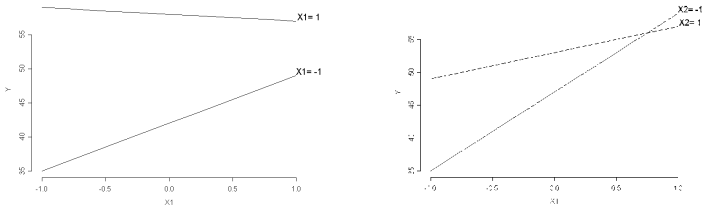


Different graphical representations of the first order model with interaction term:

$y = 50 + 8X_1 + 3X_2 - 4X_1X_2$

Perspective plot and Contour plot

First order polynomial models + interaction



Different graphical representations of the first order model with interaction term:

$y = 50 + 8X_1 + 3X_2 - 4X_1X_2$

Line plot $y^*X_2=X_1$ and $y^*X_1=X_2$

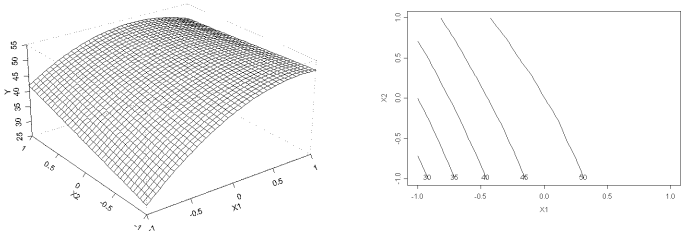
Second order polynomial models

- **If the fitting of f involves more curvature a quadratic or second order model becomes necessary.**

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i<j} \sum \beta_{ij} x_i x_j$$

- **This full quadratic estimation contains linear-, squared- and cross-product terms. To estimate the regression parameters β_i, β_{ii} and β_{ij} each variable x_i should take at least 3 levels. This model requires at least 1 + 2k + k(k-1)/2 data points.**

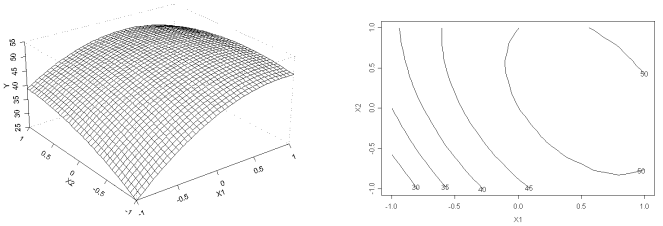
Second order polynomial models



Perspective and contour plot of the second order model:

$y = 50 + 8X_1 + 3X_2 - 7X_1^2 - 4X_1X_2$

Second order polynomial models



Perspective and contour plot of the second order model:

$y = 50 + 8X_1 + 3X_2 - 7X_1^2 - X_2^2 - 4X_1X_2$

General remark on the use of predictive models

$$y = f(x_1, x_2, \ldots, x_k)$$

- Parameter estimation
  - Physical, biological, genetical, … meaning?
    - Balanced ANOVA models
- Predicted response estimation
  - Response surface methodologies in optimisation

The method of least squares

The polynomial models can be formulated in matrix notation:

$$Y = X\beta + \varepsilon$$

Where Y is an nx1 vector of observations on the dependent variable, X equals an nxp matrix of known factor levels for each individual component, including cross-product and quadratic terms, β is a px1 vector of unknown parameters, n is the number of experimental units and ε is an nx1 vector of random errors.

In the case were the (X'X) matrix is not singular, the least squares estimation of the parameters b of β is given by:

$$b = (X'X)^{-1}X'y$$

The variance-covariance matrix of b is expressed in the following equation:

$$var(b) = \sigma^2 (X'X)^{-1}$$

with σ^2 the error variance. The elements of the matrix $(X'X)^{-1}$ are proportional to the variance and the covariances of the elements of b.

The variance of the prediction in a specific point x is given by:

$$var(xb) = \sigma^2 x'(X'X)^{-1}x$$

The method of least squares

$$Y=X\beta + \varepsilon$$

- **Parameter estimation**

$$b = (X'X)^{-1}X'y$$

variance-covariance matrix of b

$$var(b) = \sigma^2 (X'X)^{-1}$$

$(X'X)^{-1} = I$

- **Predicted response estimation**

$$Y=Xb \text{ for a specific } x$$

variance of the prediction in a specific point x

$$var(xb) = \sigma^2 x'(X'X)^{-1}x$$

Design

- The next and most important question arises: 'What data are needed to fit these polynomial models with maximum precision?' Or, 'What factor level combinations should be used as treatments?' This pertains to the definition of the experimental region and the selection of an optimal experimental design in function of the assumed model.

Model in k exp factors	Minimal number of data points
First order	k+ 1
First order + interaction	1 + k + k(k-1)/2
Second order	1 + 2k + k(k-1)/2

Definitions in experimental design (1)

- An experimental factor or independent variable is a fixed variable, which is supposed to determine at least partially the system under study (and thus affects the dependent variable(s)) and which can take at least two different values or factor levels. A factor is denoted x_i for i=1, 2, …, k and k the number of experimental factors in the study.
- An experimental treatment is a specific combination of a level for each factor. A treatment corresponds with a row vector consisting of a level for each factor ordered in an arbitrary way. Vertical concatenation of all these row vectors gives the treatment matrix.
- The number of observations of the independent variable for each treatment are called replications.

Definitions in experimental design (2)

- An experimental unit is the smallest subdivision of an experiment to which a treatment is applied in a single trial. An experimental unit corresponds to a row vector representing a treatment.
- The design matrix of an experiment consisting of n experimental units and k factors is a nxk matrix, which rows are the experimental units.
- The region of interest, the factor space of the experiment or the experimental region is the hyper-volume spanned by the points in the treatment matrix. Thus a treatment matrix has a geometrical equivalent. In classical experimental design theory both hyper-cubical and hyper-spherical experimental regions are most common.
- The size of an experiment equals the number of treatments multiplied by the number of replications per treatment or the number of experimental units n.

Basic principles of experimental design

Replications

Measurements of the dependent variables for repetitions of the same treatment

Randomisation

Protection against unknown bias introducing factors

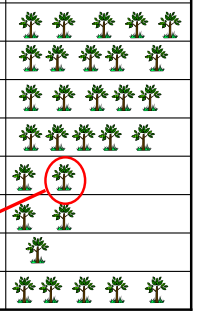
Blocking

Protection against known bias introducing factors

Basic principles of experimental design

Replications: measurements of the dependent variables for repetitions of the same treatment

Treatment	X_1	X_2	X_3	Reps: yield/tree
1	0	0	0	
2	1	0	0	
3	0	1	0	
4	0	0	1	
5	1	1	0	
6	1	0	1	
7	0			
8	1	1	1	



Basic principles of experimental design

Replication

Replication or measurement of the dependent variables for repetitions of the same treatment has two goals:

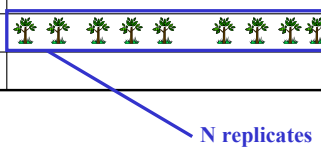
- Firstly it allows to obtain an estimate of the experimental error. This estimate of error becomes a basic unit of measurement to determine whether observed differences in the dependent variable are really statistically different. Thus the size of a detectable effect (precision) is determined by the number of replications.
- Secondly replication permits to obtain more precise estimates of possible differences between treatments.

Basic principles of experimental design

Replication

Estimating experimental error

Treatment	X_1	X_2	X_3	Reps: measurements of fruit yield per tree
…				
T_i	12	-10	123	
…				



Fruit yield / tree has inherent biological variability

Random variable

Follows a normal probability distribution

Estimation of the Population Mean, μ (T_t).

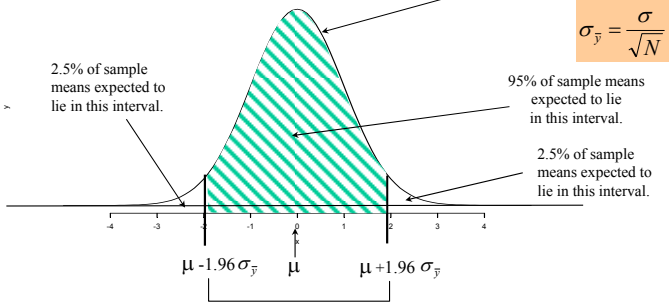
Point Estimate of μ is the sample mean: $\bar{y}$

Interval Estimate of μ is the confidence interval:

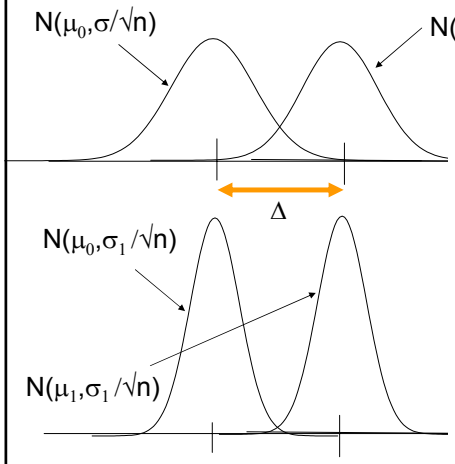
The Central Limit Theorems tell us the shape of the sampling distribution

$N(\mu, \sigma_{\bar{y}})$

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{N}}$$



Increasing precision in comparing two treatments



$\sigma_1 < \sigma$

Decreasing σ decreases the spread in the sampling dist of $\bar{y}$

Same thing happens if you increase n.

Same thing happens if Δ is increased.

Basic principles of experimental design

Replication

To gain some intuitive insight in the effect of replication, the relation between the number of replicates on the standard error of the mean of a treatment is demonstrated.

Recall that the standard error of the mean is given by the standard deviation divided by the square root of the number of replicates or the sample size.

Suppose that 45 measurements of activity of the poly phenol oxidase (PPO) enzyme of 45 plant extract of the same treatment were carried out. The enzymatic activity is expressed in change of spectrofotometric absorption per minute.

Basic principles of experimental design

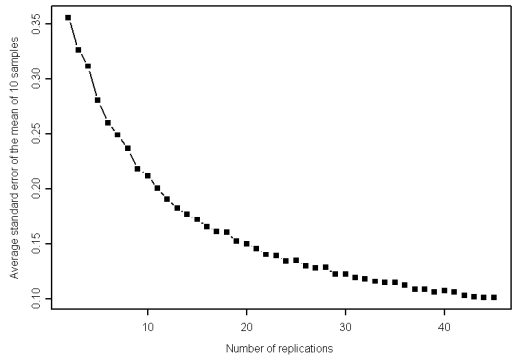
Replication

The main question arises: 'How many replicates does the experimenter need to achieve a 'satisfactory' estimate of the mean enzymatic activity for that particular treatment?' To simulate the effect of replication from the population of 45 measurements 10 samples are taken respectively of 2, 3, 4, …, 44 measurements. For each sample the standard error of the mean is calculated. For a series of samples with the same sample size the average of the standard errors is computed and plotted against the sample size.

	Mean	Standard error of the mean	Number of replications
PPO	2.55	0.10	45

Basic principles of experimental design

Replication



Basic principles of experimental design

Replication

The standard error of the mean shows a sharp decrease when the number of replicates is increasing, but the decrease tends to stabilize around the true population value at large numbers of replications.

Recall that, under normality assumptions, the 95% confidence interval around the true population average μ is approximately:

$$\overline{X} - 2stderr \le \mu \le \overline{X} + 2stderr$$

It is obvious that if the number of replicates increases the accuracy of the sample estimate of μ increases tremendously until an asymptotic value is reached.

For 45 replications	For 10 replications
$2.5 - 0.2 \le \mu \le 2.55 + 0.2$	$2.5 - 0.42 \le \mu \le 2.55 + 0.42$

Basic principles of experimental design

Randomisation

Protection against unknown bias introducing factors

By randomisation is meant that both the allocation of the treatments to the experimental units as well as the order in which measurements are taken should be determined at random. The different treatments should be randomised over the experimental units when carrying out the experiment. This means that each treatment should have the same probability of being exposed to bias generating factors not included in the experiment. Thus randomisation of the treatments excludes the introduction of systematic bias into an experiment, by averaging out the possible effects of extraneous factors.

In the experimental practice many bias introducing factors are possible; for example differences in light flux density in growing rooms, etc

The use of randomisation is the keystone of the application of statistical theory to the design of experiments.

Basic principles of experimental design

Randomisation

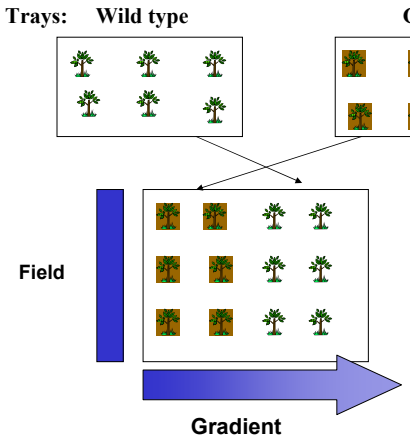
Protection against unknown bias introducing factors

In the experimental practice many bias introducing factors are possible fi

- Locational factors
- Subjective selection
- Measurement order in lab
- Systematic bias

Example 1

Locational factors

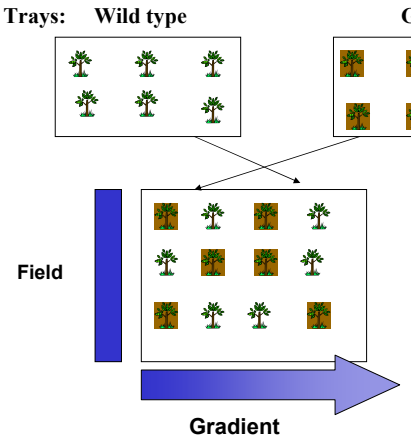


Confounding of two experimental factors
- Plant aspect
- Soil moisture

Due to non-randomisation

Example 1

Locational factors: randomised



Random allocation of experimental units in the field

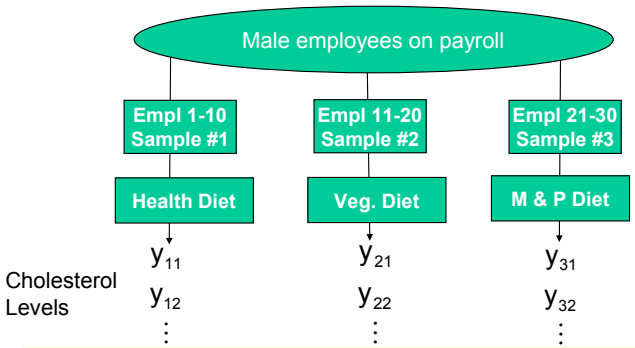
Each plant has the same probability of being exposed to bias generating factors like soil moisture

Treatment mean not affected by moisture content

Within treatment variance is increasing

Example 2

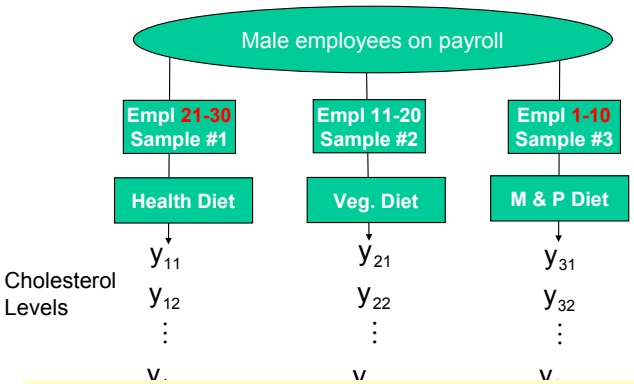
Subjective selection



What if payroll is ordered from top- to low salaries?
Confounding of salary and diet

Example 2

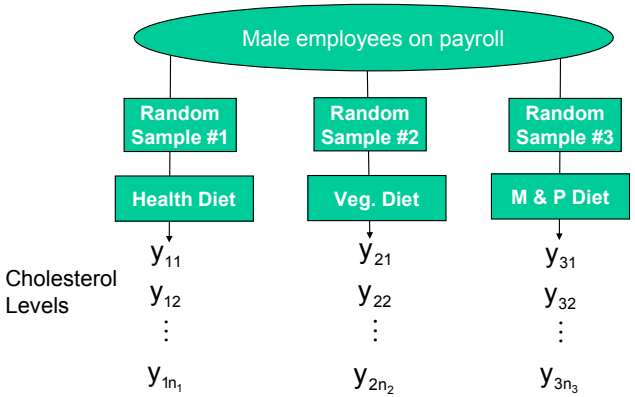
Subjective selection: switch order



Confounding of salary and diet

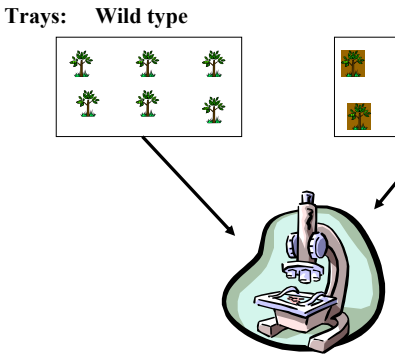
Example 2

Randomised experimental study



Example 3

Order of measurement: non- randomised

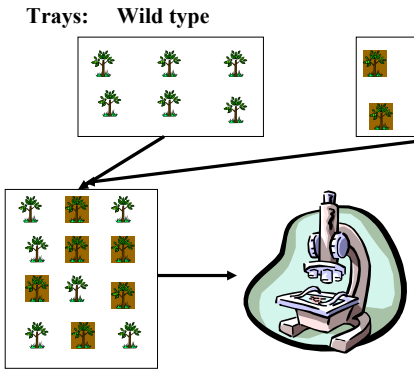


- Confounding of two experimental factors
  - Plant aspect
  - Equipment failure
- Invalid experiment
- Due to non-randomisation

Measurement equipment becomes unstable after some time

Example 3

Order of measurement: randomised



Randomisation step

- Treatment means not affected by equipment failure
- Treatment means can be compared without confounding
- Valid experiment
- Variability on the means is affected by equipment failure

Example from industry

Sampling bottles on a conveyor belt

Virtual experimentation environment

Basic principles of experimental design

Randomisation

Protection against unknown bias introducing factors

The use of randomisation is the keystone for the application of statistical theory to the design of experiments

Non-randomisation leads to invalid experiments due to confounding between the experimental factors under study and some unknown bias generating factors, resulting in invalid results and interpretation

Randomisation of the experimental units in an experimental design leads to a completely randomised designs.

Basic principles of experimental design

Randomisation

Protection against unknown bias introducing factors

Completely Randomized Design (one-factor design)

- Experimental units are *relatively homogeneous*.
- Experiment will use very *few replicates*.
- Treatments are assigned to experimental units at *random*.
- Each treatment replicated the same number of times.
- No accommodation made for *disturbing variables* (other sources of variation).

Basic principles of experimental design

Blocking

Protection against known bias introducing factors

When reasons exist to believe that a certain factor, initially not included in the experiment could influence the response, this factor should be introduced as a blocking factor.

The blocking factor should be designed as independent (orthogonal) as possible with the other experimental factors. If this factor is not interacting with other factors, the blocking factor has a pure additive effect and is not interfering with the interpretation of the main factors.

The blocking factor measures the effect of the known bias introducing process, even in situations where this factor is not specifically defined.

Blocking

Practical Situations

In many situations, the researcher:

- Does not have sufficient homogeneous experimental material or conditions in one group (location, batch, etc) to effectively use the completely randomized design (i.e. *resource constraints*)
- The study objectives require examining treatments over a broad range of experimental units in order that results can be extended to more situations (i.e. *breadth of study objectives*).
- The experimental material must be grouped for administrative or implementation purposes (i.e. *implementation constraints*).

If the researcher knows something about the characteristics of the experimental material or conditions, it is often possible to *group* experimental units into sets of *relatively homogeneous* material and then *compare treatment level means within these groups (=blocks)*.

Blocking example 1

A scientist was interested in the use of three chemicals and water on their effectiveness in extracting sulfur from Florida soils. The chemicals of interest are:

- •Calcium Chloride $CaCl_2$
- •Ammonium Acetate NH_4OAc
- •Mono Calcium Phosphate $Ca(H_2PO_4)_3$
- •Water H_2O

Five soils were chosen for this experiment:

•Troup	Jackson Co.	Paleudults soil
•Lakeland	Walton Co.	Quartzipsamments soil
•Leon	Duval Co.	Haplaquads soil
•Chipley	Jackson Co.	Quartzipsamments soil
•Norfolk	Alachua Co.	Paleudults soil

Blocking example 1

Blocking for control of extraneous variation

The main interest in the experiment is the comparison of the four extraction methods. Which one is the best?

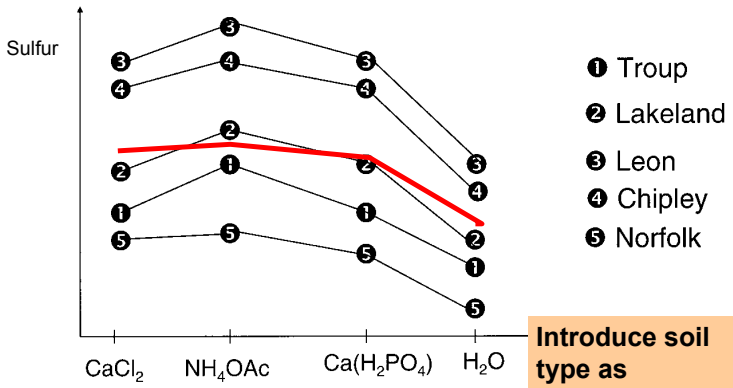
The variation imposed on the extraction procedure by the five different soil types represents a source of *extraneous variation*. Unless *controlled* for in the experiment, this variation has the potential to "swamp" or overwhelm the differences among the extraction procedure, resulting in the high probability of concluding there are no treatment effects when in fact there actually are treatment effects present.

Fair comparisons only occur among extractions within a soil type.

We wish to use the combined experience across soil types to make a stronger statement about the extraction procedures.

Blocking example 1

Graphical View



Introduce soil type as blocking factor

Randomized Block Design

Any experimental design in which the randomization of treatments is restricted to groups of experimental units within a predefined block of units assumed to be internally homogeneous is called a **randomized block design**. Blocks of units are created to control known sources of variation in expected (mean) response among experimental units.

Rules for blocking:

- Carefully examine the situation at hand and identify those factors which are know to affect the proposed response.
- Choose one or two of these factors as the basis for creating blocks.

Blocking factors are sometimes referred to as *disturbing factors*.

Blocking factors are additional experimental factors.

Other examples of blocking factors

Disturbing Variable	Experimental Unit
Nutrient gradient Water moisture gradient Slope differences Soil composition	Field Plot
Orientation to sun Flow of air Distribution of heat	Location in Greenhouse
Age Local density	Tree
Gender Age Socio-demographics	Person

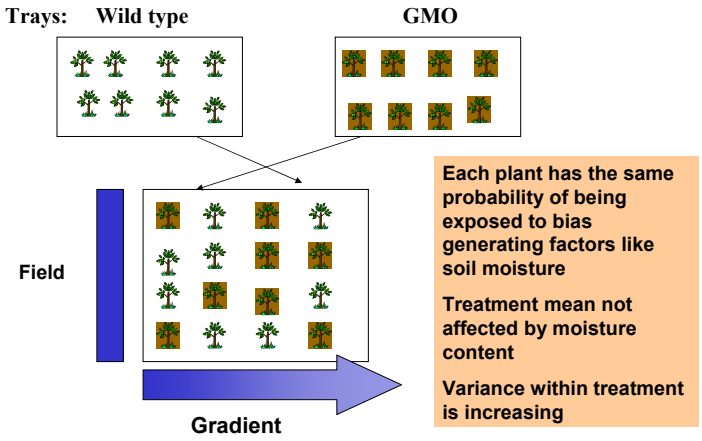
Blocking importance

How blocks are formed is critical to the effectiveness of the analysis.

Introducing blocks should maximizes *within block homogeneity* while simultaneously maximizing *among block heterogeneity*.

Blocking example 2

Completely randomised design



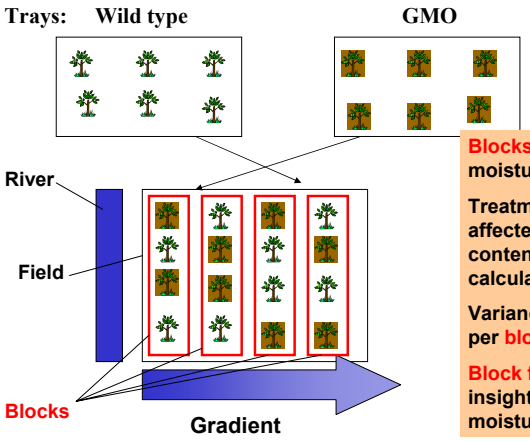
Each plant has the same probability of being exposed to bias generating factors like soil moisture

Treatment mean not affected by moisture content

Variance within treatment is increasing

Blocking example 2

Randomised block design



Blocks perpendicular to moisture gradient

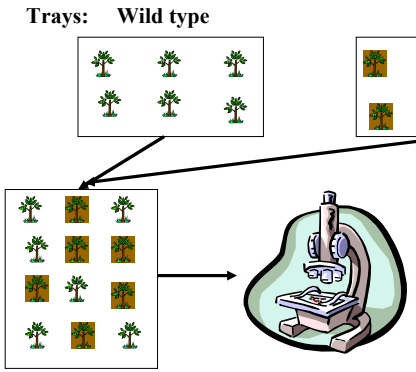
Treatment means not affected by moisture content, because calculated within block

Variance within treatment per block has decreased

Block factor can give insight in the effect of the moisture gradient

Blocking example 3

Order of measurement: completely randomised



Randomisation step

- Treatment means not affected by equipment failure
- Treatment means can be compared without confounding
- Valid experiment
- Variability on the means is affected by disturbing factors

Blocking example 3

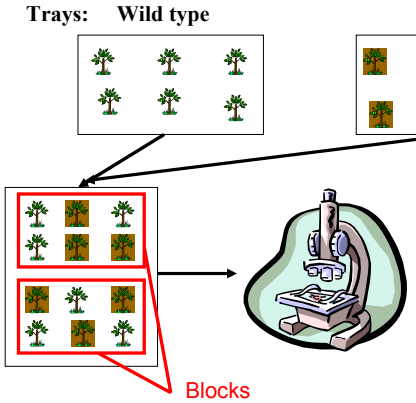
Order of measurement: completely randomised

Suppose

- Amount of measurements is to large to finish on one day
- Operators are switched at regular times
- Part of the measures are taken on different equipment or in different labs
- Part of the basic material was subjected to different pre-processing conditions
- Temperature rise in the lab during a midsummer day

Blocking example 3

Order of measurement: completely randomised



Introducing a blocking factor takes care of the disturbing process

Makes experimentation more precise, due to smaller error variance

Advantages and Disadvantages of blocking

Advantage of a Blocked Design

- To control a *single* extraneous source of variation and remove its effect from the estimate of experimental error (reducing variance).
- Allows more *flexibility in experimental layout*.
- Allows more *flexibility in experimental implementation* and administration.

Disadvantage of a Blocked Design

- Generally unsuited when there is a large number of treatments because of possible *loss of within block homogeneity*.
- Serious problem with the analysis if a *block factor by treatment interaction effect* actually exists and no replication within blocks has been included. (solution: use replication within blocks when possible).

Complete or incomplete blocked designs

Can all treatments be accommodated in each block?

Complete Block Design: Every treatment occurs in each block.
Incomplete Block Design: Every treatment does not occur in each block.

Complete

A	B	C	D
B	D	C	A
D	C	A	B

Incomplete

A	B	C
B	D	A
D	C	A

Randomization in Blocked Designs

For all 'one' blocking classification designs:

- Randomization of treatments to experimental units takes place within each block.
- A separate randomization is required for each block.
- The design is said to have "one restriction on randomization'.

A **completely randomized design** requires only one randomization.

Two disturbing factors lead to two restrictions on randomisation

Latin square design

Randomization in practice

All major statistical packages provide tools to generate random numbers and/or generate randomisation and blocking schemes for a great many standard experimental designs

SAS, S-Plus, design-expert, …

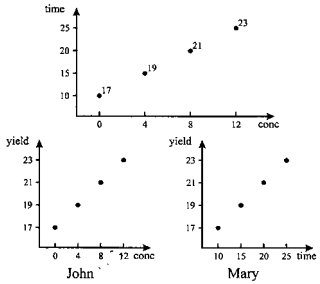
S-plus, randomise a design matrix : *mydesign <- fac.design(rep(3,2))*
randomize (mydesign)

S-plus, generate n random numbers: *round(runif(n, min=0, max=100))*

Properties of experimental designs

Example: yield of a new reaction under development
Probably influenced by:
concentration of additive 1 (0% - 12%)
reaction time (10 min - 20 min).

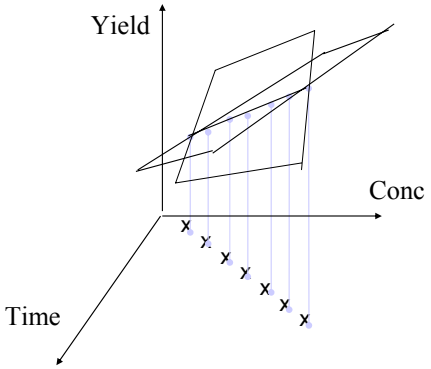
Researcher nr. 1 (John): concentration is important.
Researcher nr. 2 (Mary): time is important.
Four carefully measured runs will be done:
Experiment 1:



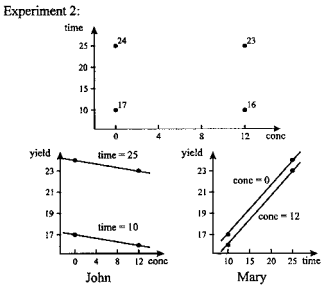
Problem: an increase in yield cannot be attributed to any of the predictor variables in particular. We can only say that there's a joint effect.

Effects of time and concentration are completely confounded

Extreme Multicollinearity



Properties of experimental designs



Conclusion: time has a positive effect on yield; increasing time (while keeping all other things equal) increases yield. Concentration has a small negative effect.

From the second experiment we learned much more than from the first one.

➡ **Orthogonal design**

Properties of experimental designs

Treatment	Time	Conc	Yield
1	10	0	17
2	10	12	16
3	25	0	24
4	25	12	23
means	17.5	6	

➡ Centering

Treatment	Time	Conc	Yield
1	-7.5	-6	17
2	-7.5	6	16
3	7.5	-6	24
4	7.5	6	23

Calculation of var-cov matrix of the treatment matrix T

$$T'T/n\text{-}1 = \begin{pmatrix} 75 & 0 \\ 0 & 48 \end{pmatrix}$$

1. Center the var Time and Conc: extract column means
2. Calculate T'T/n-1, with n number of observations

⬇ **Orthogonal design**

Properties of experimental designs

Standard coding as an orthogonal treatment matrix

Treatment vectors are centred and rescaled to constant variance

Treatment	Time	Conc	Yield
1	-1	-1	17
2	-1	1	16
3	1	-1	24
4	1	1	23

$$T'T/n\text{-}1 = \begin{pmatrix} 1.33 & 0 \\ 0 & 1.33 \end{pmatrix}$$

Coding

$$x_i^c = 2\left(\frac{x_i - \bar{x}}{\Delta_i}\right)$$

x_i^c the coded factor level
x_i the original factor level
$\bar{x}$ the average of and x_i
Δ_i the range of

Calculations of the variance-covariance matrix in S-plus

```
#calculation of VCV of centred X
> X<-matrix(c(-7.5,-6,-7.5,6,7.5,-6,7.5,6), nrow=4, ncol=2, byrow=T)
> X
     [,1] [,2]
[1,] -7.5   -6
[2,] -7.5    6
[3,]  7.5   -6
[4,]  7.5    6
> VCVmatrix<- t(X)%*%X/(nrow(X)-1)
> VCVmatrix
     [,1] [,2]
[1,]   75    0
[2,]    0   48
>
> var(X)
     [,1] [,2]
[1,]   75    0
[2,]    0   48
```

S-plus calculations of the variance-covariance matrix on the standard coded design matrix

```
#calculation of VCV of standard coded X
> Xs<-matrix(c(1,1,1,-1,-1,1,-1,-1), nrow=4, ncol=2, byrow=T)
> Xs
     [,1] [,2]
[1,]    1    1
[2,]    1   -1
[3,]   -1    1
[4,]   -1   -1
> VCVsmatrix<- t(Xs)%*%Xs/(nrow(Xs)-1)
> VCVsmatrix
         [,1]     [,2]
[1,] 1.333333 0.000000
[2,] 0.000000 1.333333
> var(Xs)
         [,1]     [,2]
[1,] 1.333333 0.000000
[2,] 0.000000 1.333333
```

Properties of experimental designs

Orthogonality

Reconsider the first order polynomial

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

In matrix notation

$Y = X\beta + \varepsilon$

The variance-covariance matrix of b (estimate of β) is expressed in the following equation:

$\text{var}(b) = \sigma^2 (X'X)^{-1}$

with σ^2 the error variance.

Properties of experimental designs

Orthogonality

A design for a first order model is orthogonal when its design matrix D is an orthogonal matrix. The product of an orthogonal design matrix with his transpose is a diagonal matrix, thus D'D. Recall that the variance-covariance matrix of the model parameters in a least squares fit of a linear model is:

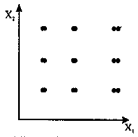
$\text{var}(b) = \sigma^2 (D'D)^{-1}$

with σ^2 the error variance.

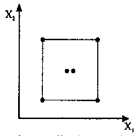
Thus the covariances of the model parameters are zero, in other words the parameter estimates can be assessed independently. Thus no confounding between experimental factors arises and effects can be interpreted directly.

Recipes for orthogonality

combining each level of each factor an equal number of times, always leads to orthogonality :

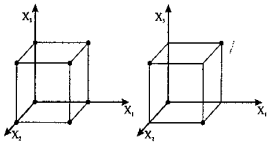


adding points in the center does not destroy orthogonality :



Factorial type designs

orthogonality has much to do with *balance* :



Balance

Balancing refers to the specific assignment of treatments to experimental units such that *comparisons of treatment effects are done with equal precision*. This is usually accomplished by equally replicating each treatment.

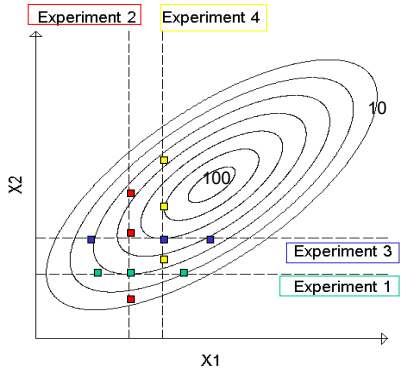
Balanced Block Design: The variance of the difference between two treatment means is the same regardless of which two treatments are compared. This usually implies that the overall replication (disregarding which blocks they are in) for the comparison of two treatments is the same for all pairs of treatments.

Partially Balanced Design: The variance of the difference between two treatments depends on which two treatments are being considered. This usually implies different replication for different treatments.

Unbalanced Designs: Usually what you end up with - not a design.

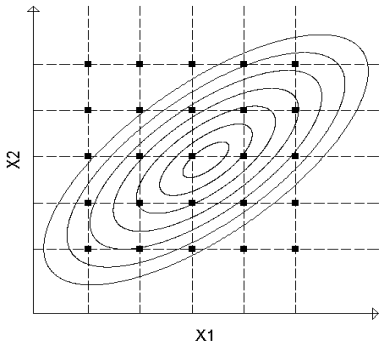
Choosing an optimisation strategy

'One-factor-at-a-time' approach



Choosing an optimisation strategy

Multi-factor approach



Choosing an optimisation strategy

- **The sum of all the experimental units of all the one-factor experiments is most of the time much larger than the number factorial combinations. Taking into account that the one-factor-at-a-time does not guarantee to find the optimum, while the factorial does, it is obvious that factorial type experiments are by far more efficient in optimisation strategies.**
- **Secondly the efficiency of finding significant effects increases with the number of factor combinations included in the experiment.**
- **With factorial type designs investigation of interactions is possible**

Choosing an optimisation strategy

Example: One-factor-at-a-time' experiment

	Vitamins		
Treatment	N	P	T
1	-1	-1	-1
2	+1	-1	-1
3	-1	+1	-1
4	-1	-1	+1

The effect of N can be assessed as the difference between the response at low level and the response at high level, thus treatment 2 minus treatment 1, both P and T being constant. Obviously in this approach the conclusions about the effect of N can only be drawn at the low levels of P and T. In the same way the effect of P is estimated as responses of treatment 3 minus treatment 1 and the effect of T as 4 minus 1.

If the observations on each treatment (replicates) are independent with equal variance σ^2, the variances of these quantities are $2\sigma^2$.

Essentially this experiment consists of 3 sub-experiments with 2 treatments each, as a consequence three one-way-ANOVA models are necessary for the analysis.

Choosing an optimisation strategy

Example: Multi-factor or factorial experiment

	Vitamins		
Treatment	N	P	T
1	+1	+1	+1
2	+1	+1	-1
3	+1	-1	+1
4	+1	-1	-1
5	-1	+1	+1
6	-1	+1	-1
7	-1	-1	+1
8	-1	-1	-1

In this factorial experiment the effect of a vitamin is estimated as the difference between the sum of all high levels and the sum of all low levels divided by 4, thus for example for N:

Effect of N = [(Treatment 1 + 2 + 3 + 4) – (Treatment 5 + 6 + 7 + 8)] / 4

Here the conclusions about the effect of N are based on all level combinations of the factor P and T. Moreover all possible interaction effects between 2 and 3 vitamins can be evaluated through one three-way-ANOVA-model.

The variance of this quantity equals $\sigma^2/2$, thus 4 times smaller than in the 'one-factor-at-a-time' approach. This again demonstrates the superiority of the factorial experiment.

Factorial Experiment

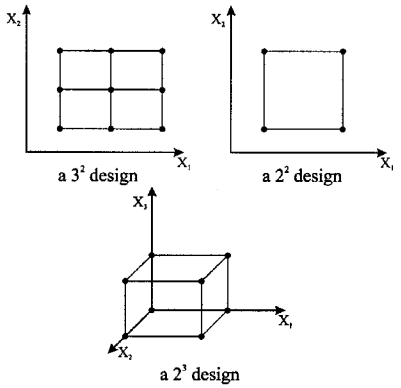
Factorial Experiment - an experiment in which the response *y* is observed at all factor level combinations.

An *experiment is not a design.* (e.g. one can perform *a factorial experiment in a completely randomized design* or in a *randomized complete block design*, or in a *Latin square design*.)

Design relates to how the experimental units are arranged, grouped, selected and how treatments are allocated to units.

Experiment relates to how the treatments are formed. In a factorial experiment, treatments are formed as combinations of factor levels.

Factorial Experiment



Number of treatments

a*b*c*d …

with

a number of levels of factor A

b number of levels of factor B

c number of levels of factor C

d number of levels of factor D

Size of the experiment

R*(a*b*c*d)

with

R number of replications

Statistical models for the 'factorial type' approach

As demonstrated previously experimenting with all important factors simultaneously included in a 'factorial' type experiment is a far better approach.

Once this multi-factor experimental region is defined, a model form has to be chosen, that will be used to relate the experimental factors with the measured plant response within the experimental region.

In a 'Response Surface' approach the preferred models are polynomials of first and second degree, namely, linear and quadratic models. The emphasis lays on predicted response estimation. Dependent and independent variables are numeric.

In an ANOVA approach a sums of squares partitioning approach is preferred. The accent lays on parameter estimation. Dependent variable is numeric, independent variables are categorical or fixed numeric. (next chapter)

Determination of the important factors and their ranges

I. <u>Sub-optimal operating conditions are known.</u> Starting from the actual operating conditions for a certain process, an optimal 'factorial type' design is constructed to explore the neighbourhood of this 'sub-optimum. In this case the experimental region is defined as a design centred at the 'sub-optimum'. Thus for each factor a range is chosen with as centre the sub-optimum. Some particular combinations of the levels of each experimental factor describe the vertices of a hyper-volume, centred at the sub-optimum. The choice of the actual range per factor depends on a priori knowledge about the effect of the factor on the response variable, on the nature of the factors and on the assumed model.

A first order design and model leads to consecutive small experiments: 'method of steepest ascent'.

A second order design and model needs a small number of large experiments.

Determination of the important factors and their ranges

II. <u>No or not enough a priori knowledge about the process.</u> When a priori knowledge about the importance of experimental factors or about the approximate location of the optimum is missing, an overall optimisation in multiple steps is necessary.

In the first steps screening experiments are used to separate important components from less important ones over a relatively large experimental region, thus involving relatively large ranges for each factor. Mostly first order models.

Once the key factors are determined, the approach sub I can be followed to determine the region of interest.

Statistical models for the 'factorial type' approach

First order models with or without interaction terms, are appropriate in three situations:

1. Screening experiments to select the important factors out of set of possible factors of influence.

2. Experiments in so narrow ranges that the expected effect on the response variable can be assumed to be linear. This approach is especially suitable in the 'Method of Steepest Ascent'.

3. When the real model is known to be first order linear.

A first order model fits a hyper-plane over the experimental region, as a consequence the maximal response will always occur on the border of the region if the plane has slope. This model only provides information about the main effects of the factors.

If the fitted surface has no slope the experimental factors have no main effects within the experimental region; in other words a zone of equal response is found.

Designs for first order models

- These models need at least experiments with 2 levels per factor and k+1 treatments.
- For screening experiments with large amounts of components the Plackett-Burman- and fractional factorial designs are most appropriate.

Designs for first order models

Plackett-Burman designs

A Plackett-Burman design is an orthogonal screening design, consisting of a fraction of a 2^k full factorial with as main property a tremendous reduction of the amount of treatments. As a consequence no interactions between factors are estimable, but still all main effects can be interpreted trough a first order model.

These designs allow screening from 6 to 23 factors simultaneously.

Designs for first order models

12 treatments Plackett-Burman designs

- low level (-1)
+ high level (+1)

	Experimental factors										
Treatment	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1	+	+	-	+	+	+	-	-	-	+	-
2	+	-	+	+	+	-	-	-	+	-	+
3	-	+	+	+	-	-	-	+	-	+	+
4	+	+	+	-	-	-	+	-	+	+	-
5	+	+	-	-	-	+	-	+	+	-	+
6	+	-	-	-	+	-	+	+	-	+	+
7	-	-	-	+	-	+	+	-	+	+	+
8	-	-	+	-	+	+	-	+	+	+	-
9	-	+	-	+	+	-	+	+	+	-	-
10	+	-	+	+	-	+	+	+	-	-	-
11	-	+	+	-	+	+	+	-	-	-	+
12	-	-	-	-	-	-	-	-	-	-	-

Designs for first order models

Plackett-Burman screening designs

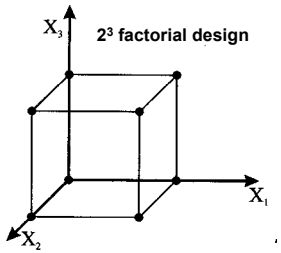
- To apply these designs, the experimenter assigns a column to each chosen experimental factor under study (columns in the table) and then allocates treatments randomly according to the chosen columns.
- As a rule of thumb the number of treatments should be at least 5 more than the number of experimental factors to allow the estimation of experimental error and valid test statistics about the model.
- The data are analysed with a first order regression model.

Number of exp factors	Possible design
4 < 7	12 treatments (frac fact is better)
7 – 15	20 treatments
13 – 23	28 treatments

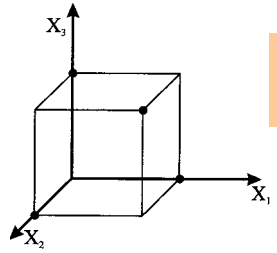
Designs for first order models

Fractional factorial designs

- Also these designs are fractions of 2^k full factorial designs, but only a limited amount of fractions is allowed because these designs emphasise 'balance' in estimating factor effects. In other words the estimation of each specific main effect should consist of geometrically balanced differences of measured responses on high and low factor levels.
- Of course the reduction in the number of treatments results in the impossibility of estimating high order interactions, depending on the fraction chosen .



2^3 factorial design



½ fraction of a 2^3 factorial design

16 treatments fractional factorial screening designs to fit a first order model in 6, 7 and 8 factors

Treatment	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	-1	-1	-1	+1	+1	+1	-1	+1
2	+1	-1	-1	-1	-1	+1	+1	+1
3	-1	+1	-1	-1	+1	-1	+1	+1
4	+1	+1	-1	+1	-1	-1	-1	+1
5	-1	-1	+1	+1	-1	-1	+1	+1
6	+1	-1	+1	-1	+1	-1	-1	+1
7	-1	+1	+1	-1	-1	+1	-1	+1
8	+1	+1	+1	+1	+1	+1	+1	+1
9	+1	+1	+1	-1	-1	-1	+1	-1
10	-1	+1	+1	+1	+1	-1	-1	-1
11	+1	-1	+1	+1	-1	+1	-1	-1
12	-1	-1	+1	-1	+1	+1	+1	-1
13	+1	+1	-1	-1	+1	+1	-1	-1
14	-1	+1	-1	+1	+1	+1	+1	-1
15	+1	-1	-1	+1	+1	-1	+1	-1
16	-1	-1	-1	-1	-1	-1	-1	-1

Designs for first order models

Fractional factorial designs

- **To apply the design on the previous slide, select the first 6, 7 or 8 columns. No free choice of columns due to balance.**
- **Design tables are available in literature and software.**

Number of factors	Possible design
<4	Full factorial
4	½ fraction of 2^4: 8 treatments
5	1/4 fraction of 2^5: 8 treatments
6	1/4 fraction of 2^6: 16 treatments
7	1/8 fraction of 2^7: 16 treatments
8	1/16 fraction of 2^8: 16 treatments

First order screening designs Adding centre points

- **The centre of a coded design corresponds with the treatment coding (0, 0, …, 0).**
- **Addition of replicated centre points provides the possibility to test lack of fit, in other words to test if the real response surface is first order.**
- **This lack of fit is calculated as the difference between the mean responses of all design points (equally replicated) and the mean response in the centre point.**
- **Of course to test the curvature for each individual factor a design for a second order model is necessary.**

Designs for first order models with interaction terms

- **These models need at least experiments with 2 levels per factor and $1 + k + k(k-1)/2$ treatments.**
- **For the estimation of first order models with interaction reflected Plackett-Burman- and fractional factorial designs are most appropriate.**

Designs for first order models with interaction terms

Reflected Plackett-Burman designs

- A reflected Plackett-Burman design arises from a specific Plackett-Burman design augmented with the same design but all '+' changed to'-' and vice versa.
- Thus the 12 run design is augmented to a 24 treatment design, the 20 run to 40 runs and the 28 to 56 treatments.
- The addition of these treatments makes the independent estimation of main effects and 2-factor interactions possible.

Number of factors	Possible design
<7	24 treatments, fractional factorial is better
7-15	40 treatment
13-23	56 treatment

Designs for first order models with interaction terms

Fractional factorial designs

- When a larger fraction of a 2^k full factorial is taken, not only main effects but also two-factor interactions can be estimated
- Additional fractional factorial design plans can be found in literature and software

Number of factors	Possible design
<5	Full factorial
5	1/2 fraction of 2^5: 16 treatments
6	1/2 fraction of 2^6: 32 treatments
7	1/2 fraction of 2^7: 64 treatments
8	1/4 fraction of 2^8: 64 treatments

Statistical models for the 'factorial type' approach

Second order models provide information about linear, interaction and curvature effects with respect to all or most of the independent numeric variables.

These models are widely applicable to describe experimental data in which system curvature is abundantly present. Thus these models are most appropriate to optimise a response over the experimental region.

Designs for these models need at least experiments with 3 levels per factor and $1 + 2k + k(k-1)/2$ treatments.

An obvious choice for such a design would be a 3^k factorial, namely a factorial experiment with each of k factors at three levels in all possible combinations.

Designs for second order order models

- For small numbers of factors (2 or 3, resulting in respectively 9 and 27 treatments) this approach is still feasible.
- For 4 factors, 81 treatments are involved to fit only fifteen model parameters (one intercept, four first order coefficients, four full quadratic and six cross-products) in a full quadratic second order model.
- Starting with 5 factors the number of treatments becomes prohibitively large ($3^5 = 243$).

Designs for second order models

Several authors have suggested specific second order designs that compromise between relative precision in estimating the model parameters and the amount of experimental effort, that is the number of treatments necessary.

The Box-Behnken and Central Composite designs are mostly used. Typically, these designs are appropriate for second order models in two to eight factors.

If more than eight factors are involved also these designs become unpractically large. In this case a preliminary screening experiment (first order model) is considered to select (reduce) the number of experimental factors.

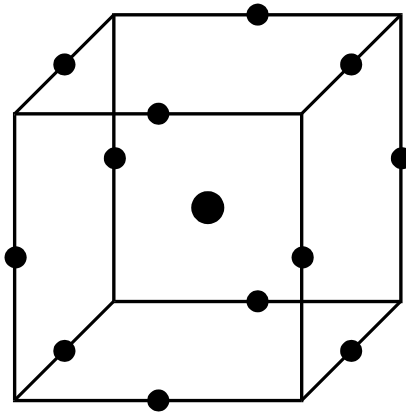
Designs for second order order models

Box-Behnken designs

- Box-Behnken designs are subsets of 3^k factorial designs.
- Except for the centre points, all points are centroids of the edges (or faces) of a hypercube with dimensions equal to the number of factors. Thus all these treatments lay on a single hyper-sphere and thus are equally distant from the centre. These designs have a hyper-spherical geometry.
- This geometrical property is associated with rotatability, or the variance of the predicted response depends only on the distance to the centre of the design and not on the direction.
- Box-Behnken designs exist for 3 to 7 experimental factors

Designs for second order order models

3-factor Box-Behnken designs



12 centroids on the perimeter

Centre point 3 times replicated

15 treatments

3-factor Box-Behnken designs

Treatment	x_1	x_2	x_3
1	+1	+1	0
2	+1	-1	0
3	-1	+1	0
4	-1	-1	0
5	+1	0	+1
6	+1	0	-1
7	-1	0	+1
8	-1	0	-1
9	0	+1	+1
10	0	+1	-1
11	0	-1	+1
12	0	-1	-1
13	0	0	0
14	0	0	0
15	0	0	0

Designs for second order order models

Box-Behnken designs

Box-Behnken designs are practical for 3 to 7 experimental factors

Number of factors	Number of centroids	Replication of the centre point	Total number of treatments
3	12	3	15
4	24	3	27
5	40	6	46
6	48	6	54
7	56	6	62

Designs for second order order models

Central Composite designs

- The Central Composite designs are by far the mostly used designs for second order models.
- These designs consist of a 2^k full factorial or a fractional factorial, augmented with 2k star points and n_c centre points.
- If a fractional factorial is used, the main and two-factor interaction effects should be estimable independently (see designs for first order models with two-factor interaction terms).
- The star points are located on the main axes of the coded design on the same distance from the centre as the factorial points. In other words, both the factorial and the star points lay on a hyper-sphere around the centre of the design.

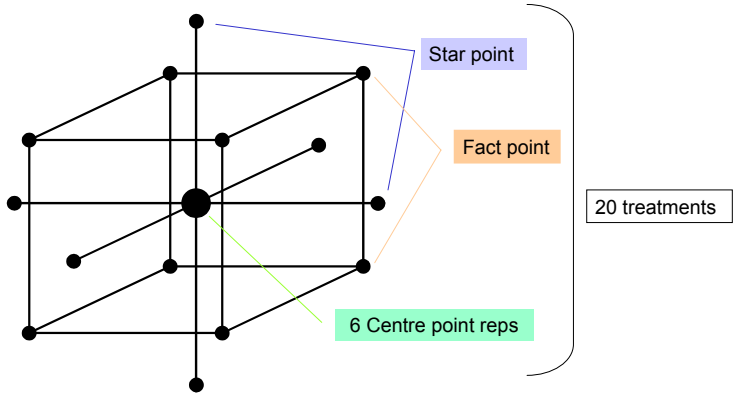
Designs for second order order models

Central Composite designs

- The centre point is replicated to estimate pure error.
- This design involves five levels for each factor coded as ($-\alpha$, -1, 0, 1, α), α being the distance from each star point to the centre.
- To provide the design the desirable property of rotatability, α has to be chosen in such a way that star and factorial points belong to the same hyper-sphere.

Designs for second order order models

3-factor Central Composite designs



Designs for second order order models

Central Composite designs

The sequential possibilities of these designs are very useful in many problems.

Firstly the factorial part of the design, including centre points, can be run to estimate a first order model. If the lack of fit test, based on the replicated centre points, indicates that additional curvature is necessary, in a second step the design can be augmented with the star points to provide degrees of freedom for a second order model.

Designs for second order order models

Central Composite designs

Number of factors	Full or fractional factorial	Value for α	Number of factorial treatments	Number of star points	Number of replicated centre points	Total number of treatments
3	full	1.68	8	6	6	20
4	full	2	16	8	6	30
5	full	2.38	32	10	8	50
5 (1/2)	½ relication	2	16	10	8	34
6	full	2.83	64	12	10	86
6 (1/2)	½ relication	2.38	32	12	10	54
7	full	3.63	128	14	10	152
7 (1/2)	½ relication	2.83	64	14	10	88

3-factor Central Composite design

Treatment	X_1	X_2	X_3	
1	1.00	1.00	-1.00	2^3 full factorial
2	-1.00	1.00	-1.00	
3	1.00	-1.00	-1.00	
4	-1.00	-1.00	-1.00	
5	1.00	1.00	1.00	
6	-1.00	1.00	1.00	
7	1.00	-1.00	1.00	
8	-1.00	-1.00	1.00	
9	1.68 (α)	0.00	0.00	star points
10	-1.68 ($-\alpha$)	0.00	0.00	
11	0.00	1.68 (α)	0.00	
12	0.00	-1.68 ($-\alpha$)	0.00	
13	0.00	0.00	1.68 (α)	
14	0.00	0.00	-1.68 ($-\alpha$)	
15	0.00	0.00	0.00	centre points
16	0.00	0.00	0.00	
17	0.00	0.00	0.00	
18	0.00	0.00	0.00	
19	0.00	0.00	0.00	
20	0.00	0.00	0.00	

Practical considerations

What if the wrong model is assumed?

It is good practice to introduce a treatment in an experiment that can be used as a checkpoint to control the lack of fit. A simple approach is to fit the model based on all treatments except the check point and then compare the predicted response in the check point with the real measured value. For instance, this is an adequate procedure to test for possible curvature while fitting a first order model.

The statistical approach to lack of fit consists of the partitioning of the error sums of squares into pure experimental error and lack of fit.

Investigation of the fitted model (1)

In the optimisation step the fitted response surface is investigated in search for the optimum. Different situations occur in function of the location of the optimum:

1. The optimum is located on the border of the experimental region. In this situation a principal direction of improvement can be defined, directed to the new optimum. A new experiment has to be carried out centred at the new sub-optimum, followed by re-evaluation. This is actually a procedure in search of the region of maximal response by consecutively fitting of first order models. This procedure where an experimenter proceeds sequentially along the path of maximum increase in response is called the 'steepest ascent procedure'. Ones the approximate location of the optimum is found a second order model is used to determine the optimum.

Investigation of the fitted model (2)

2. Within the experimental region no optimum is found. This means that the response is not affected by the compositional changes within the range of the experiment. In other words the experimental region can be considered as a region of sub-optimal response. If desirable a new experiment can be set up centred at the sub-optimum but increasing the ranges of all factors. When fitting first order models carry out the lack of fit test to make sure that no curvature is present in the real model.

Investigation of the fitted model (3)

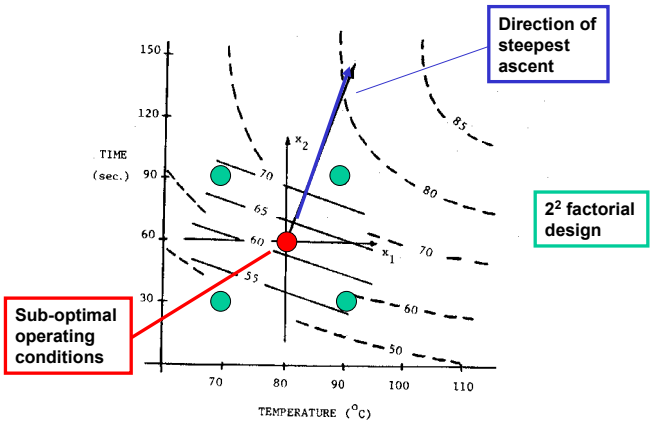
3. The optimum is found within the experimental region. The objective is reached. Double-check in case of local optima can be necessary. The fitted second order model can be investigated and the optimum can be determined.

Method of steepest ascent

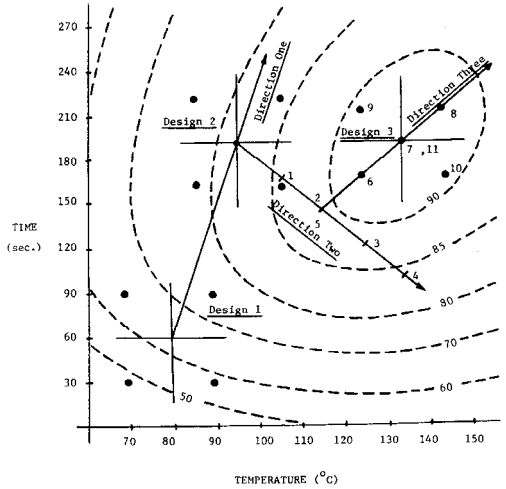
The steepest ascent procedure consists of performing a sequence of experiments along the path of maximum increase in response.

- **The procedure begins with a locale approximation of the true response surface with a hyper-plane (first order model) fitted over a small first order design.**
- **This hyper-plane provides information to determine a direction toward which an increasing value of the response can be expected.**
- **In this direction a new first order experiment is set up and a first order model fitted.**
- **Again the direction of maximum increase is determined. And so on.**
- **The union of these directions form the path of steepest ascent.**

Method of steepest ascent



Method of steepest ascent



Determination of the optimum for a second order model

Contour and perspective plots

A contour plot is a two-dimensional smoothed graph showing contours of constant response in an axes system defined by two experimental factors x_i and x_j, while the other experimental factors are kept constant. In most practical situations the fitted model should be plotted to allow preliminary evaluation of the model and determination of the optimum.

Also perspective plots are a useful aid in interpretation of fitted models.

Determination of the optimum for a second order model

Mathematical procedures

If a second order model is fitted, the direct way to find the optimum is by differentiating the response function with respect to the x_i's. Setting al partial derivatives

$$\frac{\partial y}{\partial x_i} = 0$$

will determine the stationary point if it exists. It has to be emphasised that this point can be either a maximum, a minimum or a saddle point of the fitted function

Determination of the optimum for a second order model

Mathematical procedures

To determine the nature of the fitted surface and the stationary point, a canonical analysis has to be carried out. This analysis consists of two steps.

Firstly the origin of the axes system is translated to the stationary point (centring the design matrix).

In the second step the axes system is rotated in such a way that the new axes correspond to the principal axes of the response surface or contour system.

In the new coordinate system W_i the canonical form of the response function is given by:

$$y = y_0 + \lambda_1 W_1^2 + \lambda_2 W_2^2 + \ldots + \lambda_k W_3^2$$

With y_0 the estimated response at the stationary point.

Determination of the optimum for a second order model

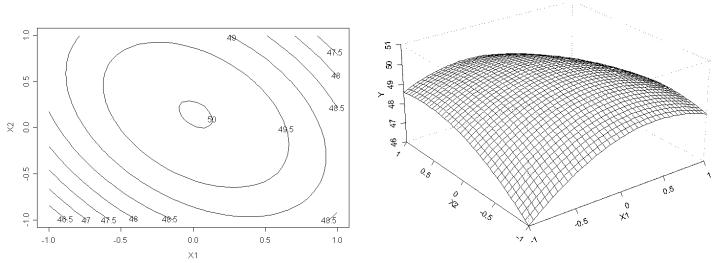
Mathematical procedures

If all λ_i are negative, a move in any direction from the stationary point results in a decrease in y. Therefore the stationary point represents the point of maximum response.

If all λ_i are positive the stationary point is a minimum.

When λ_i differ in sign the stationary point will be a saddle point.

Contour and perspective plots

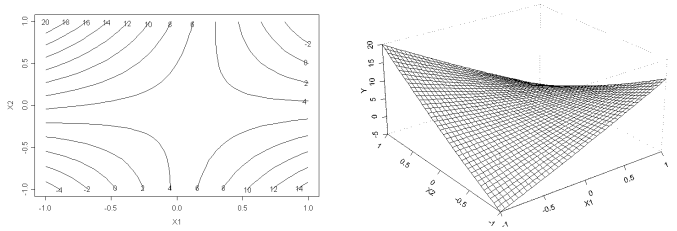


Contour and perspective plot of the second order model:

$y = 50 + 0.2X_1 + 0.3X_2 - 1.5X_1^2 - 1X_2^2 - 1X_1X_2$.

The stationary point is a maximum.

Contour and perspective plots



Contour and perspective plot of the second order model:

$y = 5 - X_1 + 1.5X_2 + 0.5X_1^2 + X_2^2 - 11X_1X_2$.

The stationary point is a saddle point.

Experimental Optimisation
Response Surface Methodology

1. Definition of the region of interest
2. Determination of the important factors and their ranges
3. Construction of an optimal design:
   - Screening designs for first order models without/with interaction
     - Plackett-Burman — Reflected P-B
     - Fractional factorial — Larger fraction Fr Fact
   - Designs for second order models
     - Box Behnken
     - Central composite
3. Fitting of a polynomial model, describing the relation between a numeric independent variable and the numeric dependent variables
4. Optimisation phase: obtaining the optimal operating conditions by investigation of the mathematical model.

Experimental Optimisation
Response Surface Methodology

An example

SCHNEIDER AND STOCKETT (1963) PERFORMED AN EXPERIMENT AIMED AT REDUCING THE UNPLEASANT ODOR OF A CHEMICAL PRODUCT WITH SEVERAL FACTORS.

PETER W. M. JOHN, STATISTICAL DESIGN AND ANALYSIS OF EXPERIMENTS MACMILLAN 1971.

SAS program

SAS output in html

Experimental Optimisation
Response Surface Methodology