

# SOME PRINCIPLES OF DESIGNING EXPERIMENTS

Barbara Bogacka

Queen Mary, University of London

School of Mathematical Sciences

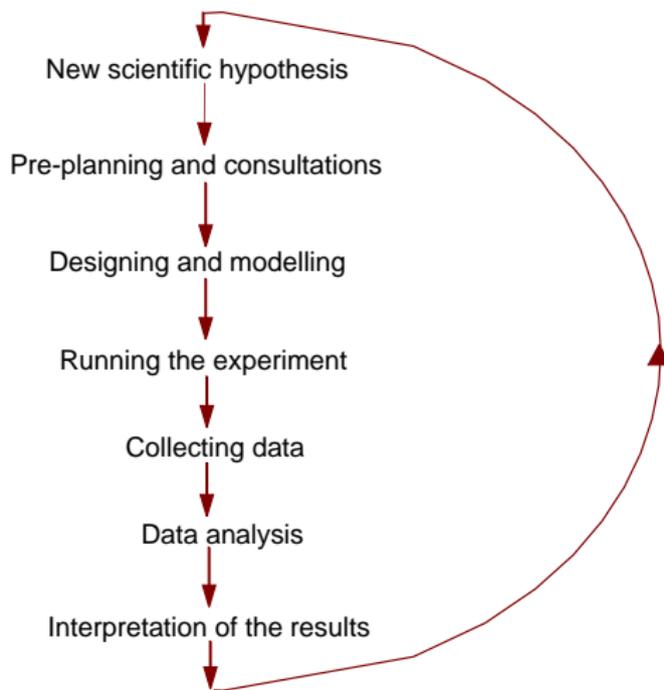
# Stages of a Statistically Designed Experiment

## General Work-flow



# Stages of a Statistically Designed Experiment

## General Work-flow



# Stages of a Statistically Designed Experiment

## New Scientific Hypothesis

- ▶ What is the purpose of your experiment:
  - ▶ Test a hypothesis that a new treatment is not better than the old one?
  - ▶ Establish a relationship among variables?
  - ▶ Find best treatment among a set of possible treatments?
- ▶ What is the population you want to make the inference about?
- ▶ What is already known about the phenomenon you are going to experiment on (your experience, subject literature, discussion with colleagues)?

# Stages of a Statistically Designed Experiment

## Pre-planning and Consultation

Discuss your **new scientific hypothesis** with a statistician:

- ▶ What is the purpose of the experiment, population to infer about, what is known about it up to date?
- ▶ What are your available resources (experimental material, time, people)?
- ▶ Is there any structure in the experimental material which may matter for your observations?
- ▶ What are the factors you are interested in?
- ▶ Are there any nuisance factors which may also influence your observations?

# Stages of a Statistically Designed Experiment

## Pre-planning and Consultation

- ▶ What will you observe?
- ▶ How the data are going to be collected (machine, people)?
- ▶ How accurate the observations are likely to be?
- ▶ Are there any technical or other constraints or limitations?

# Stages of a Statistically Designed Experiment

## Designing and Modelling

**Pre-planning and consultation** will help to find a good design for your experiment.

To design **efficiently**, that is to obtain maximum information with minimum experimental effort, you need to:

- ▶ design for the purpose of your experiment,
- ▶ model your observations,
- ▶ include a good **randomization** scheme.

# Stages of a Statistically Designed Experiment

## Designing and Modelling

Two large classes of experiments are:

- ▶ comparative experiments where **treatments** are allocated to **experimental units**,
- ▶ non-comparative experiments where we choose a **sampling scheme** within a **design region**.

# Stages of a Statistically Designed Experiment

## Running the Experiment

A good **design** and an appropriate **model** will help you to run the experiment.

Make sure that:

- ▶ the experiment is done according to the design,
- ▶ the randomization is properly performed,
- ▶ the external conditions surrounding the experiment are kept fixed as much as possible,
- ▶ **replications** are done in the same conditions.

# Stages of a Statistically Designed Experiment

## Collecting Data

Make sure that:

- ▶ You have planned ahead of time the way to collect and to record your data.
- ▶ In your spreadsheet (on paper or in an electronic version)
  - ▶ include space for such information as hour and date of the data collection and name of the person who did it (if relevant, or even if you think not),
  - ▶ allow space for comments,
  - ▶ be as accurate as possible - do not round the numbers, do not record the averages only,
  - ▶ record the actual settings, not those intended.

# Stages of a Statistically Designed Experiment

## Collecting Data

- ▶ If an observation is very unusual check as soon as possible what is a possible reason for it (error in recording, special circumstances).
- ▶ Do not change the units of measurements during the course of data collection.
- ▶ Always keep the original spreadsheet even if the data were copied to another file.
- ▶ If recorded electronically, keep a printed version as well.

# Stages of a Statistically Designed Experiment

## Data Analysis

- ▶ Data analysis makes use of a statistical model of the observations and it should be also planned at the stage of designing the experiment.
- ▶ It aims at giving an answer to your initial question (scientific hypothesis).
- ▶ It may suggest further studies.

# Stages of a Statistically Designed Experiment

## Data Analysis

- ▶ The analysis is usually done using some statistical software such as Genstat, SPSS, SAS, Minitab, Statistica or more specialist software, such as PopDesign.
- ▶ The analysis often includes such elements as ANOVA tables, lists of means, p-values, estimates of the model parameters, their confidence intervals, fitted model, residual diagnostics.
- ▶ Special computer programs are written for data analysis in non-standard models.

# Stages of a Statistically Designed Experiment

## Interpretation of the Results

- ▶ Always look critically at any computer output (there may be mistakes in the data file or in the procedure).
- ▶ Discuss the output with a statistician.
- ▶ Use your common sense as well as the statistical output in the interpretation.
- ▶ The results may give a definite answer or may suggest what was wrong or may reveal a new scientific hypothesis.

HAVE YOU EVER PLANNED AN EXPERIMENT?

# Comparative Experiments: CRD

## Example 1

An experimenter wants to compare the effects of three diets on milk production.

He will use 11 Holstein dairy cows at similar points in their lactation cycles. Cows will be fed the diets for three weeks. During the third week, the average daily milk production will be recorded for each cow, in pounds per day.

# Comparative Experiments: CRD

## Scientific Hypothesis

The diets do not differ in a sense of their effect on milk production.

# Comparative Experiments: CRD

## Pre-planning and Consultation

- ▶ What else, apart from the diet may have an effect on the milk production?
- ▶ What population does the sample of cows represent?
- ▶ Are there any technical restrictions which may influence the response?
- ▶ How the cows will be fed?

# Comparative Experiments: CRD

## Designing and Modelling

- ▶ The cows will be fed individually.
- ▶ They will be also observed individually.
- ▶ All stay on the same farm.

MODEL:

$$y_{ij} = \mu + \mu_i + \varepsilon_{ij}, \quad i = 1, 2, 3, \quad j = 1, \dots, n_i,$$

where  $\mu$  denotes the general mean,  $\mu_i$  is the effect of diet  $i$ ,  $\varepsilon_{ij}$  denotes random error of observation  $ij$ . It is often assumed that  $\varepsilon_{ij}$  are independently and identically distributed and

$$\varepsilon_{ij} \sim N(0, \sigma^2).$$

# Comparative Experiments: CRD

## Designing and Modelling

- ▶ Ideally, the cows would be randomly chosen from a large herd.
- ▶ However, there are only 11 cows on the farm at similar point of lactation, i.e.,  $N = 11$ .
- ▶ The cows have to be randomly assigned (permuted) to the diets.
- ▶ Number of diets is 3; denote the diets by letter A, B, C.

DESIGN:

cow	5	2	3	10	1	4	8	9	6	11	7
diet	A	A	A	A	B	B	B	B	C	C	C

This is one of many possible **Completely Randomized Designs**.

Is it a good design?

# Comparative Experiments: CRD

## Running the Experiment

What should be taken care of while running this experiment?

- ▶ The cows stay in the same conditions over the time of the experiment.
- ▶ They are all fed in the same time and in the same way (apart from the diet differences).
- ▶ ?

# Comparative Experiments: CRD

## Collecting Data

A spreadsheet for collecting data may look like this one:

Cow	diet	day 1	comments	measured by	...	day 7	comments	measured by
1	B				...			
2	A				...			
3	A				...			
4	B				...			
5	A				...			
6	C				...			
7	C				...			
8	B				...			
9	B				...			
10	A				...			
11	C				...			

# Comparative Experiments: CRD

## Data Analysis

The model:

$$y_{ij} = \mu + \mu_i + \varepsilon_{ij}, \quad i = 1, 2, 3, \quad j = 1, \dots, n_i$$

and the data, the model fit and the residuals:

Diet	Yield Data	Model Fit	Residual
A	60.7	61.2	-0.5
A	59.7	61.2	-1.5
A	61.9	61.2	0.7
A	62.5	61.2	1.3
B	55.6	54.3	1.3
B	52.9	54.3	-1.4
B	52.7	54.3	-1.6
B	56.0	54.3	1.7
C	62.8	58.2	4.6
C	55.8	58.2	-2.4
C	56.0	58.2	-2.2

# Comparative Experiments: CRD

## Data Analysis

Now, we may state the following general hypothesis:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = 0$$

versus the alternative hypothesis

$$H_1 : H_0 \text{ is not true}$$

We may also be interested in more specific hypotheses, such as

$$H_0 : \mu_i = \mu_j \quad (\mu_i - \mu_j = 0); \quad i, j = 1, 2, 3.$$

versus

$$H_1 : \mu_i \neq \mu_j \quad (\mu_i - \mu_j \neq 0)$$

# Comparative Experiments: CRD

## Data Analysis

ANOVA table

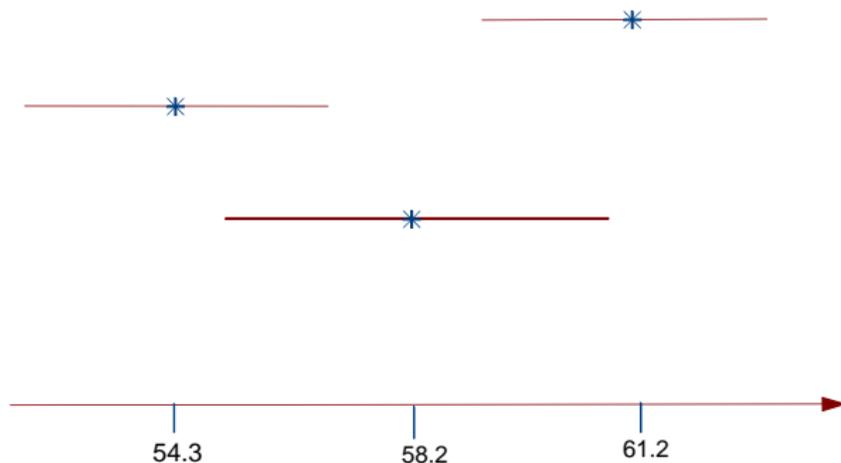
Source	DF	SS	MS	F	p
Diets	2	95.66	47.83	8.40	0.011
Error	8	45.54	5.69		
Total	10	141.20			

Level	n	Mean	StDev
A	4	61.200	1.249
B	4	54.300	1.742
C	3	58.200	3.985
Pooled StDev	=	2.386	

# Comparative Experiments: CRD

## Data Analysis

Individual 95% CIs For Mean Based on Pooled StDev



# Comparative Experiments: CRD

## Interpretation of the Results

- ▶ There is a strong evidence to reject the general null hypothesis ( $p = 0.011$ ).
- ▶ Hence, we may say that the different diets lead to different milk yield produced by Holstein dairy cows.
- ▶ The confidence intervals for the mean effects suggest that diets A and B may differ significantly; however, diet C may not be different from either of them.
- ▶ The result for diet C are affected by one unusual observation.
- ▶ It would have been better to have had four observations for diet C rather than three.

# Comparative Experiments

## Relevant Definitions

**A Treatment** is the entire description of what can be applied to an experimental unit (Ex.1: diet, at three levels).

**An Experimental Unit** is the smallest unit to which a treatment is applied (Ex.1: cow).

**An Observational Unit** is the smallest unit on which a response will be measured (Ex.1: cow).

**Replication of a treatment** A number of experimental units to which same treatment is applied (Ex.1:  $n_1 = 4, n_2 = 4, n_3 = 3$ ).

IS ANYTHING HERE SIMILAR TO YOUR EXPERIMENTS?

# Comparative Experiments: CRD with Two Factors

## Treatment Structure

The experimenter wants to see whether adding vitamins to any of the diets, A, B and C, would improve the milk yield. She is interested in comparing diets as before, but now, she is also interested in comparing the effects of adding, or not adding, the vitamins.

- ▶ Now we have two **factors**: diet, on three levels (A, B, C) and vitamins, on two levels (Yes, No);
- ▶ A treatment is a combination of the levels of the factors;
- ▶ The two factors may act independently or they may **interact**.

# Comparative Experiments: CRD with Two Factors

## Treatment Structure

Comparative Experiments may have various **treatment structures**, for example, all combinations of two or more factors, all combinations of two factors and a control or they may just be unstructured.

# Comparative Experiments: CRD with Two Factors

## Treatment Structure

A simple CRD with treatments as a combination of two factors might be:

cow	5	11	10	4	1	8	6	2	7	9	12	3
diet	A	A	A	A	B	B	B	B	C	C	C	C
vit.	Y	Y	N	N	Y	Y	N	N	Y	Y	N	N

Note: here we have 12 experimental units, 6 treatments, 2 replications of each treatment.

# Comparative Experiments: CRD with Two Factors

## Treatment Structure

Model:

$$y_{lkj} = \mu + \underbrace{\alpha_l + \beta_k + \gamma_{lk}}_{\mu_i} + \varepsilon_{lkj}, \quad \varepsilon_{lkj} \underset{iid}{\sim} N(0, \sigma^2),$$

where the  $i$ -th treatment effect  $\mu_i$  is now a sum of the effects of diet, vitamins and their interaction.

Null Hypotheses:

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0 \quad H_0 : \beta_1 = \beta_2 = 0 \quad H_0 : \gamma_{lk} = 0 \text{ for all } lk.$$

# Comparative Experiments: CRD with Two Factors

## Treatment Structure

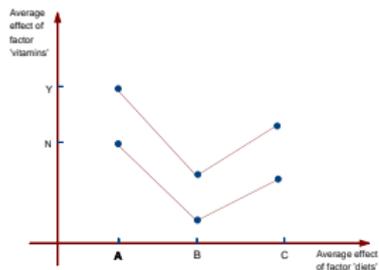
Skeleton ANOVA

Source	DF
Diets	2
Vitamins	1
Interaction	2
Error	6
Total	11

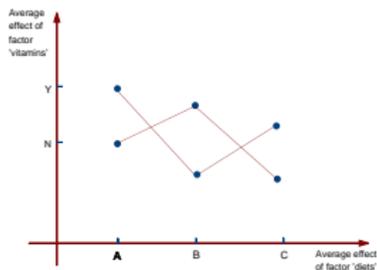
# Comparative Experiments: CRD with Two Factors

## Treatment Structure

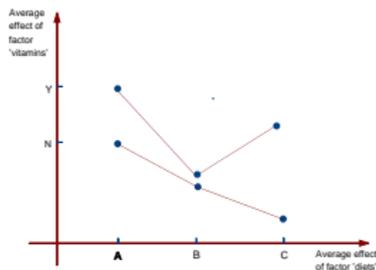
Interaction: mean response to levels of one factor changes in a different way depending on levels of the other factor.



no interaction



interaction



interaction

DO YOU HAVE MORE THAN ONE FACTOR IN YOUR  
EXPERIMENT?

# Comparative Experiments: Block Designs

## Units Structure

If experimental units have a clear structure which may differently affect the response, they may be divided into groups called **blocks**, where a block will have the units which are most alike in some sense, reflecting the structure.

Types of blocks:

- ▶ natural discrete divisions,
- ▶ continuous grading,
- ▶ for trial management.

If possible,

- ▶ blocks should have equal size,
- ▶ blocks should be big enough to allow each treatment to occur at least once in each block.

# Comparative Experiments: Block Designs

## Units Structure

For example, assume that the dietary experiment is to be done in two different farms. We may expect, that despite the same breed of cows and the same treatments used in both farms, there may be some specific effects on the milk yield inherent to each farm.

If we include this effect in the model, we will get more precise estimates of the treatment effects.

# Comparative Experiments: Block Designs

## Units Structure

We could just have a CRD on each farm, independently randomized. For the unstructured treatments it might be

farm 1	cow	5	2	3	6	1	4
	diet	A	A	B	B	C	C
	vit.	Y	N	Y	N	Y	N
farm 2	cow	1	2	4	3	5	6
	diet	A	A	B	B	C	C
	vit.	Y	N	Y	N	Y	N

Here we have 12 experimental units and 2 replications for each treatment level.

This a small experiment - if possible, it would be better to have more replications.

# Comparative Experiments: Block Designs

## Units Structure

Model:

$$y_{jlki} = \mu + \theta_j + \alpha_l + \beta_k + \gamma_{lk} + \varepsilon_{jlki}, \quad \varepsilon_{jlki} \stackrel{iid}{\sim} N(0, \sigma^2),$$

where  $\theta_j$  denotes  $j$ -th block effect,  $\alpha_l, \beta_k, \gamma_{lk}$  denote the effects of diet, vitamins and the interaction.

Skeleton ANOVA

Source	DF
Farms	1
Diets	2
Vits	1
Interaction	2
Residual	5
Total	11

IS BLOCKING RELEVANT TO YOUR EXPERIMENTAL  
SITUATION?

# Comparative Experiments

## Observational Units Within Experimental Units

A different units structure would be if the cows were put in pens, four in each pen, and then eat at lib, but the milk yield measured for each individual cow.

This is a situation when the experimental units contain observational units.

# Comparative Experiments

## Observational Units Within Experimental Units

### Design

cow	15	4	24	10	21	8	18	9	6	11	7	23
diet	A				B				C			
cow	3	2	20	1	16	19	12	22	5	14	17	13
diet	A				B				C			

Here we have 24 observational units, 6 experimental units and 2 replications for each treatment level.

# Comparative Experiments

## Observational Units Within Experimental Units

Skeleton ANOVA

Source		DF
Pens	diet	2
	residual	3
	total	5
Cows		18
Total		23

We test the treatment effects within the Pens stratum.

WHAT ARE YOUR EXPERIMENTAL UNITS  
(OBSERVATIONAL UNITS)?

## PART II

# Non-comparative Experiments: Regression Models

## Purpose of an experiment - stats point of view

- ▶ Estimation of the parameters and their functions (i.g. contrasts) and further testing statistical hypothesis (we know the family of models).
- ▶ Model building (no information about the family of models).
- ▶ Model discrimination (there are two or more plausible families of models).
- ▶ A combination of estimation and model discrimination.
- ▶ Other.

# Non-comparative Experiments: Regression Models

## Example 2. Simple Linear Regression

Model:

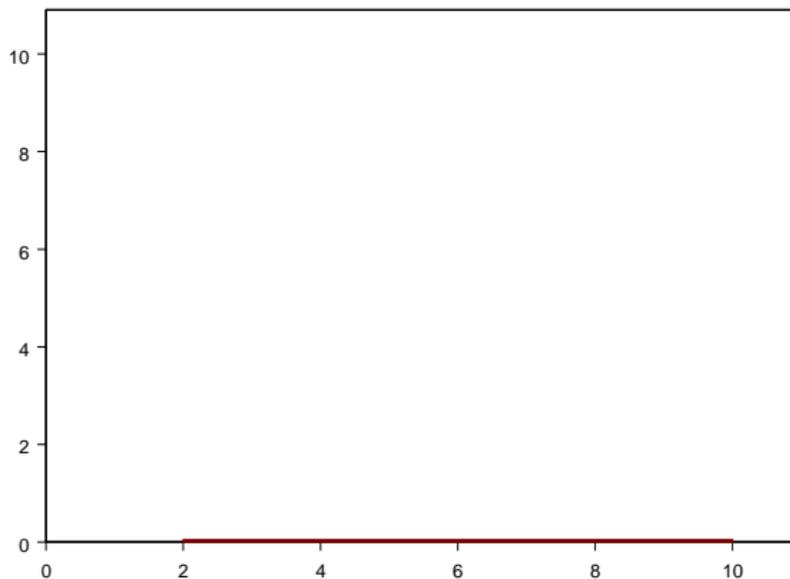
$$y_i = \theta_0 + \theta_1 x_i + \varepsilon_i, \quad \varepsilon_i \underset{iid}{\sim} N(0, \sigma^2),$$

where  $y_i$  is the response to  $x_i$ , the  $i$ -th value of the explanatory variable.  $(\theta_0, \theta_1)$  are unknown parameters (intersection and slope).

How should we design an experiment, if we knew that the response is linear, as in this equation, and we were interested in estimation of the unknown parameters?

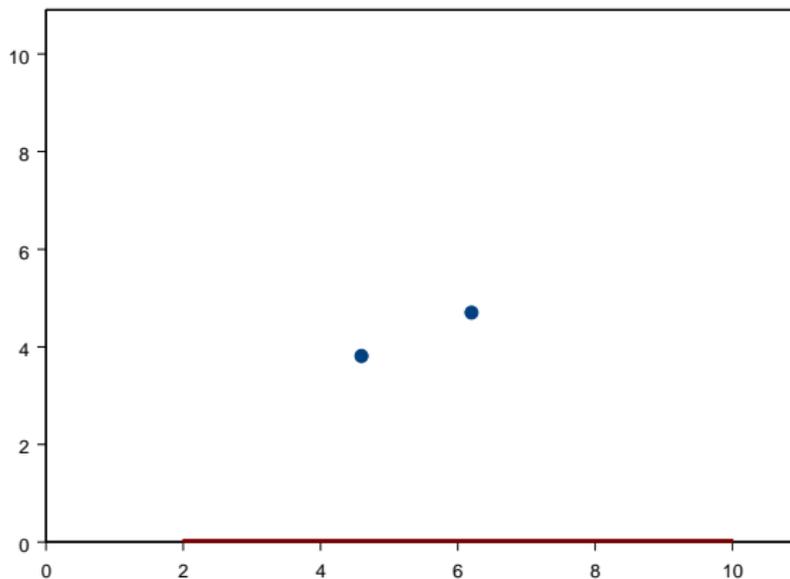
# Non-comparative Experiments: Regression Models

## Example 2. Simple Linear Regression



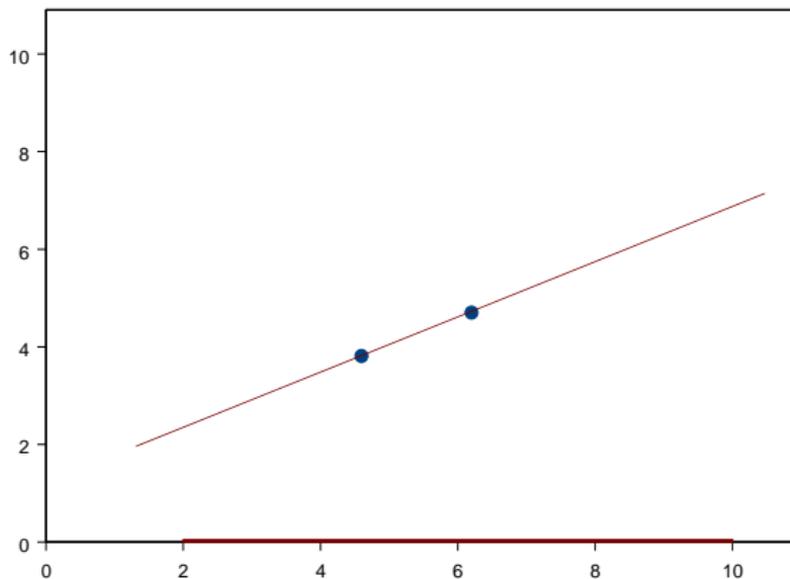
# Non-comparative Experiments: Regression Models

## Example 2. Simple Linear Regression



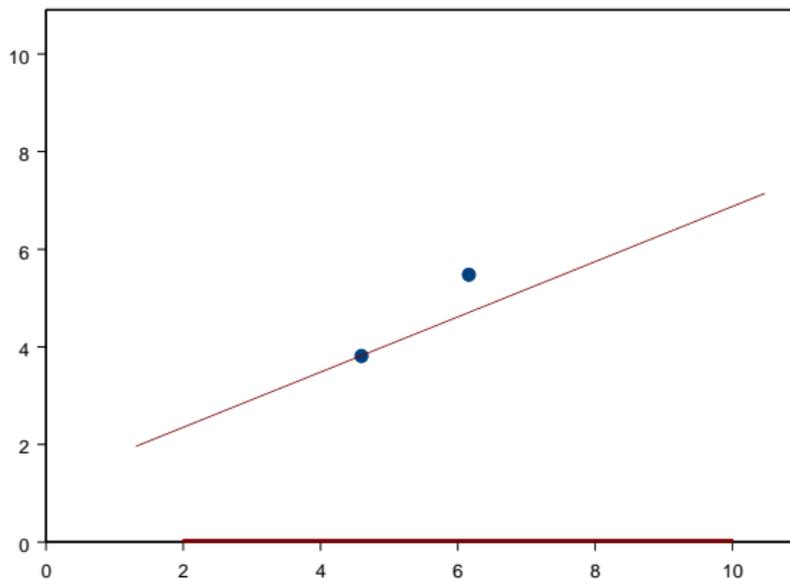
# Non-comparative Experiments: Regression Models

## Example 2. Simple Linear Regression



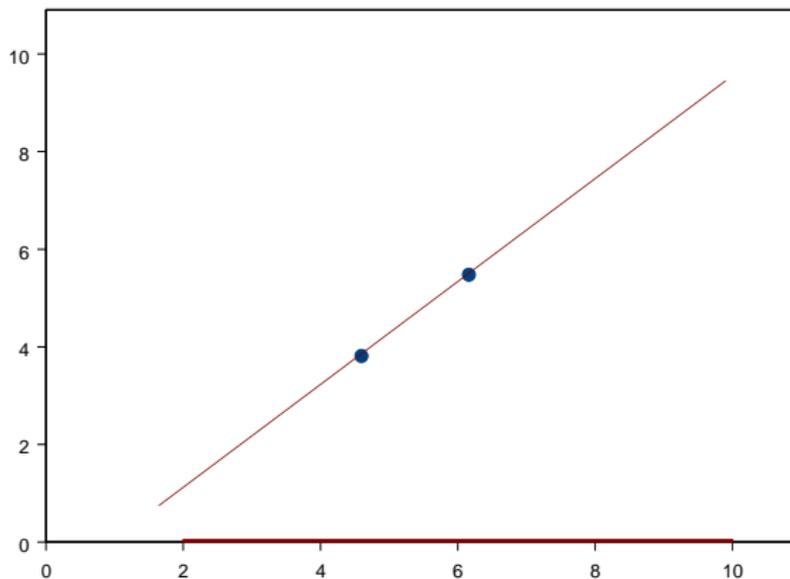
# Non-comparative Experiments: Regression Models

## Example 2. Simple Linear Regression



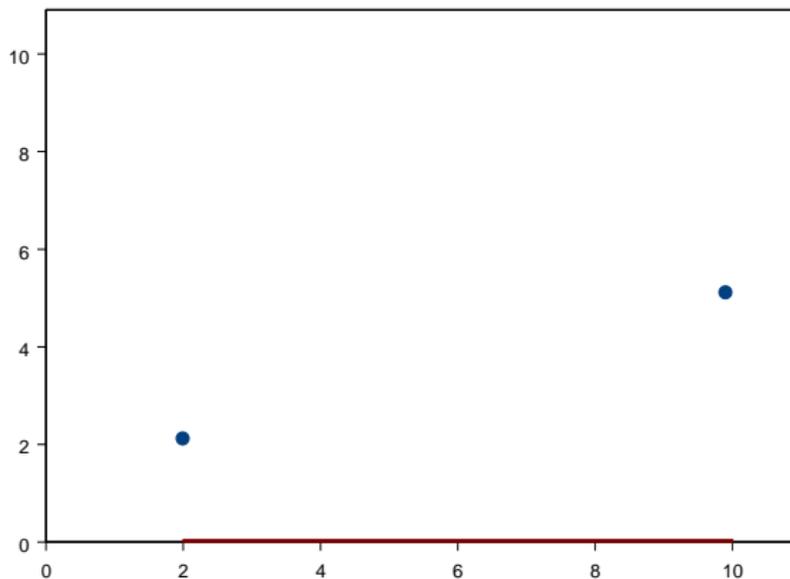
# Non-comparative Experiments: Regression Models

## Example 2. Simple Linear Regression



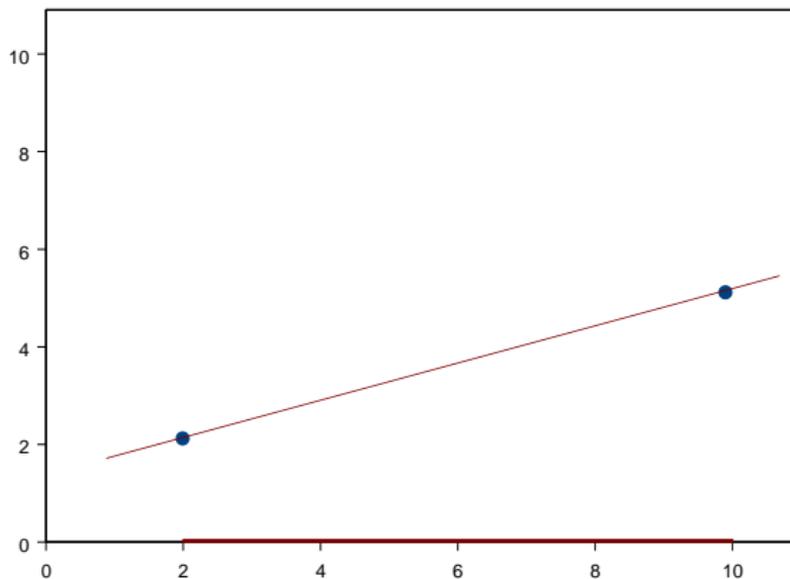
# Non-comparative Experiments: Regression Models

## Example 2. Simple Linear Regression



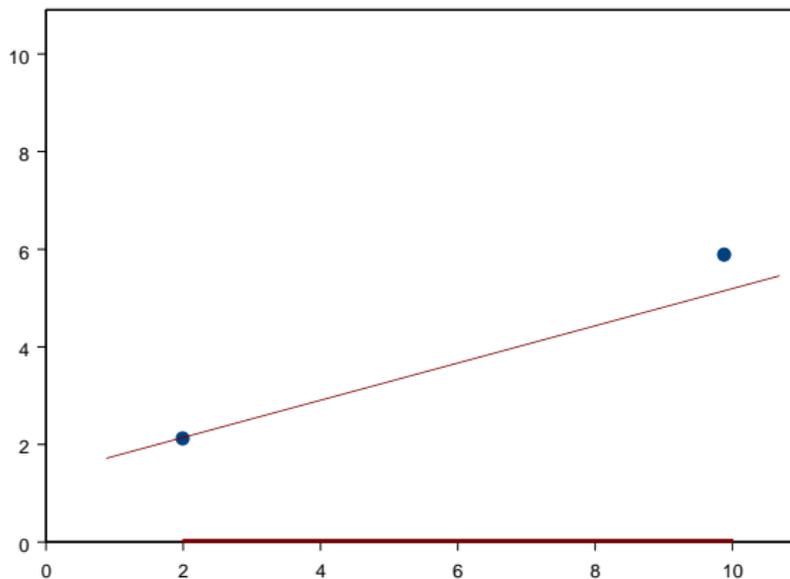
# Non-comparative Experiments: Regression Models

## Example 2. Simple Linear Regression



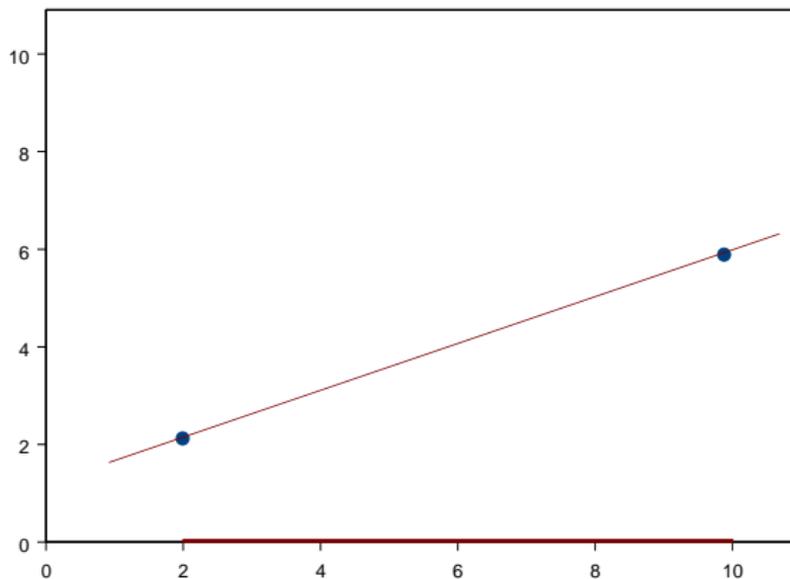
# Non-comparative Experiments: Regression Models

## Example 2. Simple Linear Regression



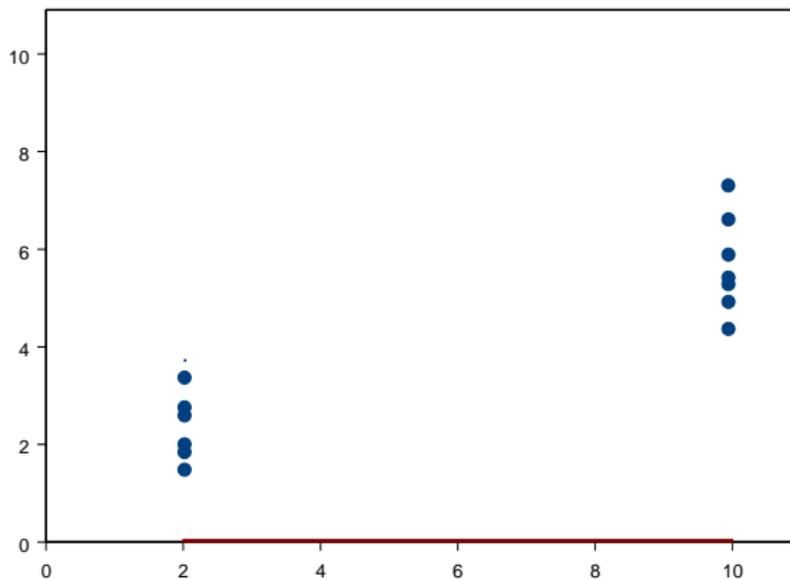
# Non-comparative Experiments: Regression Models

## Example 2. Simple Linear Regression



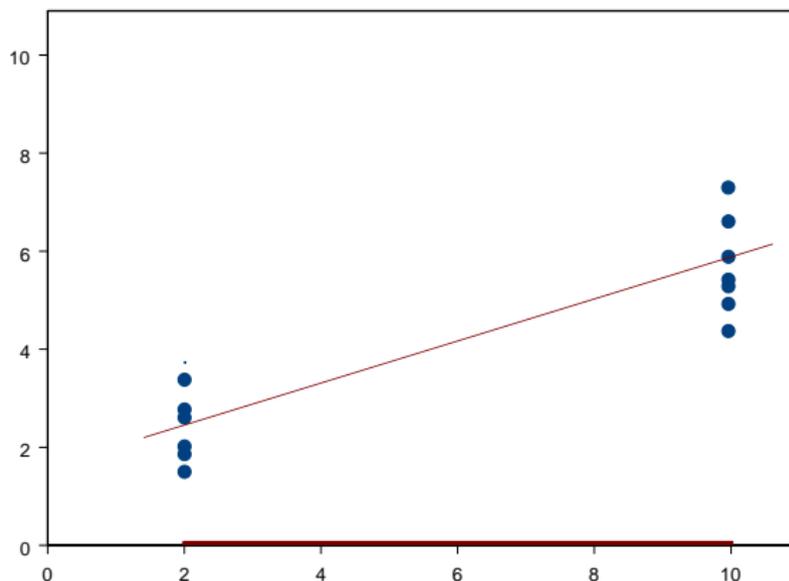
# Non-comparative Experiments: Regression Models

## Example 2. Simple Linear Regression



# Non-comparative Experiments: Regression Models

## Example 2. Simple Linear Regression



A well designed experimental data for fitting a SLR model.

# Optimum Design of Experiments

## Comparative and Non-comparative Experiments

- ▶ A **criterion** of design optimality has to be specified.
- ▶ The criterion will depend on the purpose of the experiment and on your knowledge of the model
  - ▶ When a general form of the model is known, then
    - ▶ Purpose: estimation of unknown parameters or their functions, or hypothesis testing.
    - ▶ Design for **precision of estimation**.

How can we measure the precision of estimation?

- ▶ via variance and bias of the estimator.

# Optimum Design of Experiments

## Comparative and Non-comparative Experiments

- ▶ When there are several competing models, then
  - ▶ Purpose:
    - ▶ discrimination between the models,
    - ▶ estimation of the parameters and discrimination.
  - ▶ Design: to optimize for most powerful discrimination and for precise estimation.
- ▶ When there is no information about the model at all, then
  - ▶ Purpose: to identify the model or some specific values of interest,
  - ▶ Design: to optimize for the specific objective.

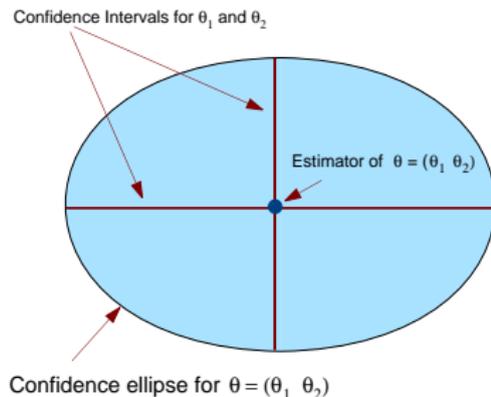
# Optimum Design of Experiments

## Problems

- ▶ A combinatorial (comparative) optimum design for a particular treatment and unit structures may not exist.
- ▶ In regression problems it is usually possible to find an approximate optimum design but it may depend on:
  - ▶ prior values of the unknown parameters,
  - ▶ curvature of the assumed model.
- ▶ In any case, the optimum design depends on the assumptions regarding the variability and correlation of the observed response.

# Optimum Design for Regression Models

## D-optimality



$100(1 - \alpha)\%$  confidence region for parameters  $\theta_1$  and  $\theta_2$ .

D-optimum design minimizes the area of the region.

# Optimum Design for Regression Models

## Example 3. Growth Rate

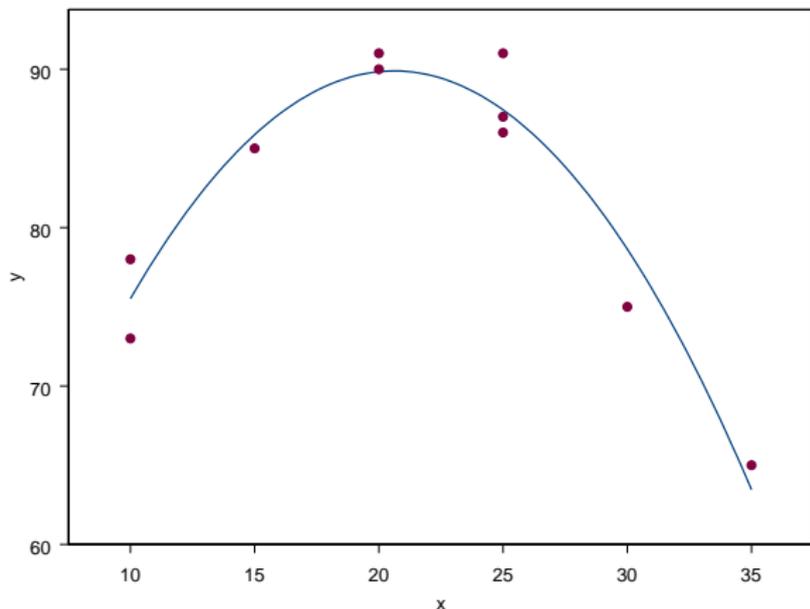
Ten experimental rats were fed with various doses of a dietary supplement and a growth rate was observed. The experimenter was interested in finding a response function. The data are given in the table below.

amount of supplement [g]	growth rate [coded]
10	73
10	78
15	85
20	90
20	91
25	87
25	86
25	91
30	75
35	65

# Optimum Design for Regression Models

## Example 3. Growth Rate

The data and the fitted quadratic polynomial are shown in the figure below.



# Optimum Design for Regression Models

## Example 3. Growth Rate

A few design questions:

1. Why these particular doses were used?
2. What was known about a plausible response before the experiment was performed?
3. Could the doses be selected in a better way?
4. How to decide what doses to select and apply?

# Optimum Design for Regression Models

## Example 3. Growth Rate

We will denote a design in the following way:

$$\xi = \{x_1 \ x_2 \ \dots \ x_N\} = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_s \\ \frac{r_1}{N} & \frac{r_2}{N} & \dots & \frac{r_s}{N} \end{array} \right\},$$

where  $r_i$  are replications (weights) of the **support points**  $x_i$  of the design such that

$$\sum_{i=1}^s r_i = N, \quad r_i > 0.$$

# Optimum Design for Regression Models

## Example 3. Growth Rate

For a full quadratic model a D-optimum design has three support points, all with equal replications. Two of the points are on the borders of the design region and one in the middle.

HAVE YOU EVER THOUGH OF OPTIMIZING YOUR  
EXPERIMENT IN A STATISTICAL SENSE?

# Optimum Design for Regression Models

## Example 4. Enzyme Kinetics

In a typical enzyme kinetics reaction enzymes bind substrates and turn them into products. The binding step is reversible while the catalytic step irreversible:



$S$ ,  $E$  and  $P$  denote substrate, enzyme and product, respectively.

# Optimum Design for Regression Models

## Example 4. Enzyme Kinetics

The reaction rate is represented by the Michaelis-Menten model

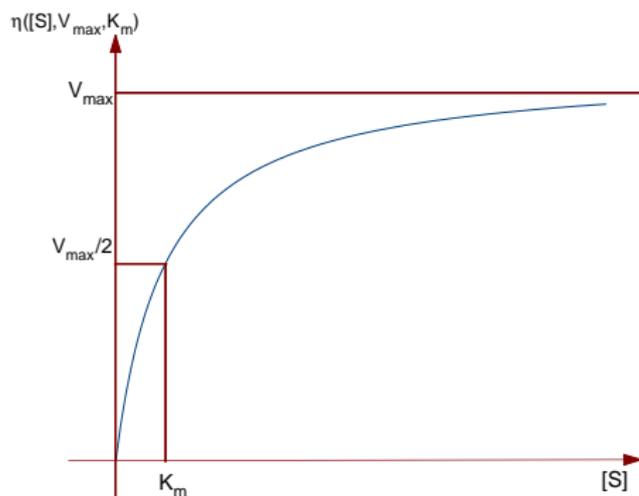
$$v = \frac{V_{max}[S]}{K_m + [S]},$$

where  $[S]$  is the concentration of the substrate and  $V_{max}$  and  $K_m$  are the model parameters:

- ▶  $V_{max}$  denotes the maximum velocity of the enzyme and
- ▶  $K_m$  is Michaelis-Menten constant; it is the value of  $[S]$  at which half of the maximum velocity  $V_{max}$  is reached.

# Optimum Design for Regression Models

## Example 4. Enzyme Kinetics



Michaelis-Menten Model. The response function:  
 $\eta([S]; V_{max}, K_m)$  for the point priors  $V_{max}^o = 1, K_m^o = 1$ .

# Optimum Design for Regression Models

## Example 4. Enzyme Kinetics

This model is non-linear with respect to the parameter  $K_m$ , but linear with respect to  $V_{max}$ .

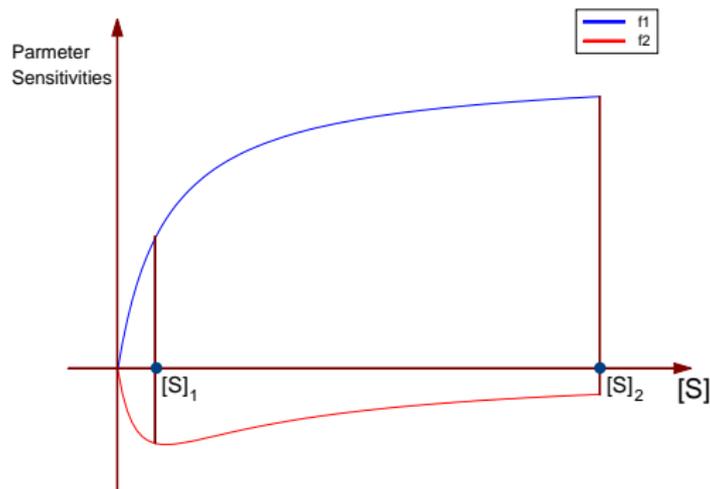
Typically, a non-linear model is linearized using a Taylor series expansion at  $\theta^0$ . Then it can be written as a linear regression model, but the coefficients  $\theta$  depend on the model parameters. They are the model derivatives with respect to the parameters.

The linearized model can then be written as

$$y_i = f_1(x_i; \theta^0)\theta_1 + f_2(x_i; \theta^0)\theta_2 + \varepsilon_i.$$

# Optimum Design for Regression Models

## Example 4. Enzyme Kinetics



The derivative  $f_1$  does not have a proper maximum; its largest value is at the border of the design region.

The best design consists of the two support points, each one equally replicated, but the first point is inside the region (not at the border as in the simple linear regression).

# Conclusions

## What is a Design of Experiment

A plan showing where/when to take observations during an experiment.

- ▶ **Classical (combinatorial) design:**
  - ▶ way of allocating treatments to experimental units;
  - ▶ treatment structure and unit structure have to be defined and matched;
  - ▶ usually applied for linear models of observations.

# Conclusions

## What is a Design of Experiment

### ► Continuous (approximate) design

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_s \\ w_1 & w_2 & \dots & w_s \end{array} \right\}$$

where

$$w_i > 0, \quad \sum_{i=1}^s w_i = 1.$$

- probability measure on a discrete set of points;
- usually applied to regression models, linear or non-linear;
- treatments and units may not be so clear as they are in the combinatorial design.

# Conclusions

## Model-Design-Analysis

**Model** of the observation  $y$

$$y = \eta(x, \vartheta) + \varepsilon,$$

$x$  denotes a set of experimental conditions,

$\vartheta = (\vartheta_1, \dots, \vartheta_p)^T$  denotes a vector of unknown parameters,

$\varepsilon$  denotes an observational error, a random variable.

**Design** tells what experimental conditions  $x$  (what levels) should we use in the study.

**Analysis** is based on the observations  $y$ .

Hence, it depends on

- ▶ the design ( $x$ ),
- ▶ the properties of the errors ( $\varepsilon$ ),
- ▶ the structure of the model function ( $\eta$ ).

THANK YOU

ANY QUESTIONS?

Reading:

A. C. Atkinson, A.N. Donev and R.D.Tobias (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press.

R.A. Bailey (2008). *Design of Comparative Experiments*. Cambridge University Press.