TOPICS IN THE DESIGN OF EXPERIMENTS

Part 1: Optimal Design Theory

Steve Coad

Queen Mary, University of London

Model - Design - Analysis

Model

Model for the observation y is

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

where x denotes a set of experimental conditions,
$\vartheta = (\vartheta_1, \ldots, \vartheta_p)^{\mathrm{T}}$ denotes a vector of unknown parameters
and ε denotes an observational error, a random variable.

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

Analysis is based on the observations y.
Hence, it depends on

- the design (x),
- the properties of the errors (ε),
- the structure of the model function (η).

Linear Models

Write

$$\eta(x,\vartheta) = f_1(x)\vartheta_1 + \ldots + f_p(x)\vartheta_p.$$

Then

$$y = f(x)^{\mathrm{T}}\vartheta + \varepsilon,$$

where

$$f(x)^{\mathrm{T}} = \big(f_1(x),\ldots,f_p(x)\big), \quad \vartheta = \begin{pmatrix} \vartheta_1 \\ \vdots \\ \vartheta_p \end{pmatrix},$$

or, in matrix notation,

$$Y = X\vartheta + \epsilon,$$

where

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} f_1(x_1)\ldots f_p(x_1) \\ \vdots \\ f_1(x_n)\ldots f_p(x_n) \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Linear Models

Assume that $\epsilon \sim \mathcal{N}_n(0, V)$.

Then the maximum likelihood estimator of ϑ is normally distributed:

$$\hat{\vartheta} \sim \mathcal{N}_p\{\vartheta, (X^{\mathrm{T}} V^{-1} X)^{-1}\}.$$

The model assumption about the errors and the form of the matrix X (design) are involved in the analysis.

Example 1

Growth Rate

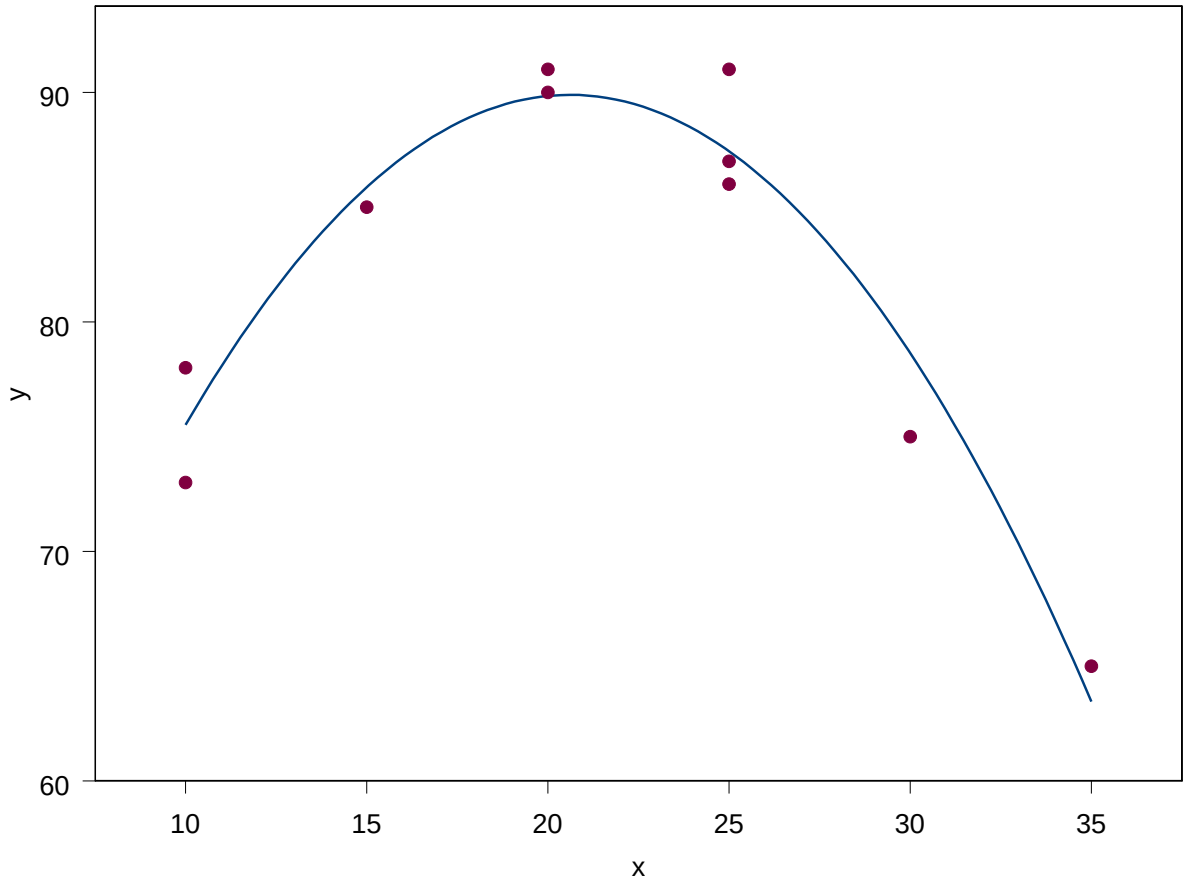
Ten experimental rats were fed with various doses of a dietary supplement and a growth rate was observed. The data are given in the table below.

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

Example 1

Growth Rate

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



Example 1

Growth Rate

A few design questions:

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

Example 1

Growth Rate - Quadratic Regression

Here,

$$\eta(x, \vartheta) = \vartheta_1 + \vartheta_2 x + \vartheta_3 x^2 = f(x)^{\mathrm{T}} \vartheta,$$

where

$$f(x)^{\mathrm{T}} = (1, x, x^2), \qquad \vartheta = \begin{pmatrix} \vartheta_1 \\ \vartheta_2 \\ \vartheta_3 \end{pmatrix}.$$

Then the matrix X is

$$X = \begin{pmatrix} f(x_1)^{\mathrm{T}} \\ f(x_2)^{\mathrm{T}} \\ \vdots \\ f(x_n)^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix}.$$

Example 1

Growth Rate - Quadratic Regression

Assuming that $\epsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$, the covariance matrix of the maximum likelihood estimator of ϑ is

$$\begin{aligned}
\mathrm{Var}(\widehat{\vartheta}) =& \sigma^2 (X^{\mathrm{T}} X)^{-1} \\
=& \sigma^2 \left\{ \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \\ x_1^2 & \dots & x_n^2 \end{pmatrix} \times \begin{pmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix} \right\}^{-1} \\
=& \sigma^2 \begin{pmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 \end{pmatrix}^{-1} .
\end{aligned}$$

We would like to make it “somehow small”.

Note that the matrix depends on the design variable x.

Consequently, X is the so-called **design matrix**. The **design** is the set of x values

$$\xi = \{x_1 \; x_2 \; \ldots \; x_n\} = \left\{ \begin{array}{cccc} x_1 & x_2 & \ldots & x_s \\ \frac{r_1}{n} & \frac{r_2}{n} & \ldots & \frac{r_s}{n} \end{array} \right\},$$

where the r_i are the **replications** of the **support points** x_i of the design. Here,

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

Example 2

A linear model for an experiment with 5 treatments and 3 blocks of size 4

Let the allocation of treatments to blocks be as follows:

block 1	1	2	3	5
block 2	1	4	3	2
block 3	5	1	4	2

Then the linear model may be written as

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij},$$

where
μ denotes the overall mean,
τ_i denotes the effect of the ith treatment, $i = 1, 2, \ldots, 5$, and
β_j denotes the effect of the jth block, $j = 1, 2, 3$.

Example 2

A linear model for an experiment with 5 treatments and 3 blocks of size 4

block 1	1	2	3	5
block 2	1	4	3	2
block 3	5	1	4	2

The matrix X and the vector ϑ are of the form

$$Y = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{31} \\ y_{32} \\ y_{42} \\ y_{43} \\ y_{51} \\ y_{53} \end{pmatrix}, \quad X = \left(\begin{array}{c|cccc|ccc} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right), \quad \vartheta = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \tau_5 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}.$$

Example 2

A linear model for an experiment with 5 treatments and 3 blocks of size 4

block 1	1	2	3	5
block 2	1	4	3	2
block 3	5	1	4	2

The so-called **incidence matrix** N indicates the design as well:

$$N = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

Another way of presenting the design is

$$\xi = \left\{ \begin{array}{cccccccccccc} \binom{1}{1} & \binom{1}{2} & \binom{1}{3} & \binom{2}{1} & \binom{2}{2} & \binom{2}{3} & \binom{3}{1} & \binom{3}{2} & \binom{4}{2} & \binom{4}{3} & \binom{5}{1} & \binom{5}{3} \\ \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} \end{array} \right\}.$$

Non-linear models

Suppose that

$$y_i = \eta(x_i, \vartheta) + \varepsilon_i.$$

Then a Taylor series expansion about the prior ϑ^o yields

$$\eta(x, \vartheta) = \eta(x, \vartheta^o) + f(x, \vartheta^o)^{\mathrm{T}}(\vartheta - \vartheta^o) + \frac{1}{2}(\vartheta - \vartheta^o)^{\mathrm{T}} f..(x, \vartheta^o)(\vartheta - \vartheta^o) + \ldots,$$

where

$$f(x, \vartheta^o)^{\mathrm{T}} = \left(\frac{\partial \eta(x, \vartheta)}{\partial \vartheta_1}, \frac{\partial \eta(x, \vartheta)}{\partial \vartheta_2}, \ldots, \frac{\partial \eta(x, \vartheta)}{\partial \vartheta_p} \right) \Big|_{\vartheta = \vartheta^o}$$

and $f..(x, \vartheta^o)$ is the matrix of second derivatives with respect to the parameters evaluated at ϑ^o.

Thus, a linear approximation to the model is

$$\eta(x, \vartheta) - \eta(x, \vartheta^o) = f(x, \vartheta^o)^{\mathrm{T}}(\vartheta - \vartheta^o)$$

or

$$\eta(x, \vartheta) = \text{const} + f(x, \vartheta^o)^{\mathrm{T}} \vartheta.$$

Non-linear models

Differences between linear models and linearised non-linear models

1. We have $\hat{\vartheta} \sim \mathcal{N}_p\{\vartheta, (X^{\mathrm{T}} V^{-1} X)^{-1}\}$ **asymptotically.**
2. Matrix $X = \begin{pmatrix} f(x_1, \vartheta^o)^{\mathrm{T}} \\ f(x_2, \vartheta^o)^{\mathrm{T}} \\ \vdots \\ f(x_n, \vartheta^o)^{\mathrm{T}} \end{pmatrix}$ depends on ϑ^o and on x.
3. The **non-linearity** of the model is ignored.

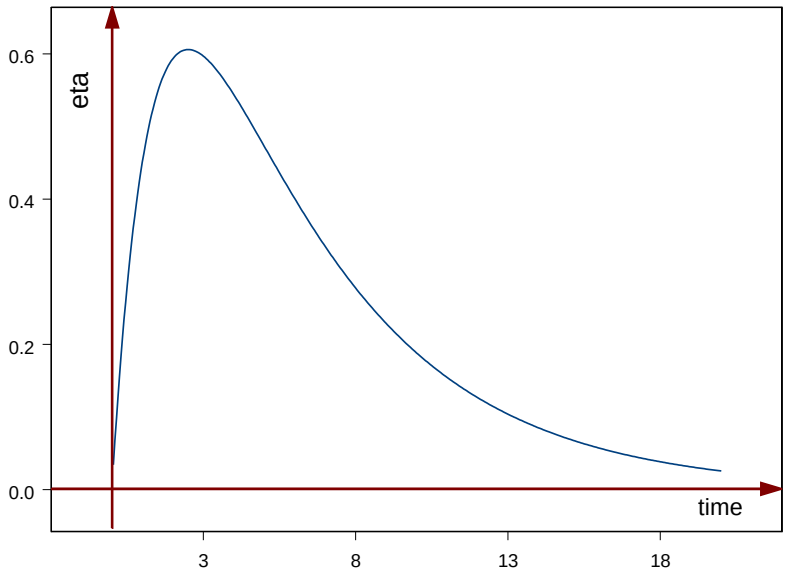
Example 3

One-compartment pharmacokinetic model

The concentration of a drug in the blood is often expressed in the form

$$y = \underbrace{\frac{k_a}{k_a - k_e}\left(e^{-k_e t} - e^{-k_a t}\right)}_{\eta(x,\vartheta)} + \varepsilon,$$

where k_a and k_e denote the rates of absorption and elimination of the drug. Here, $x = t$ and $\vartheta = (k_a, k_e)^{\mathrm{T}}$.

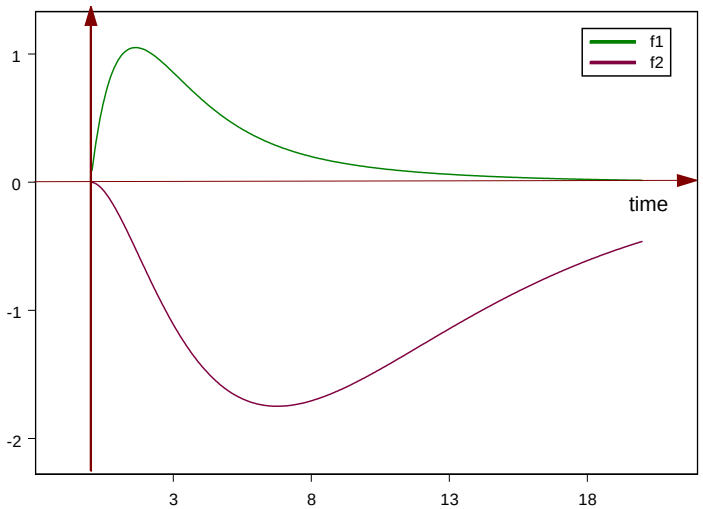


Example 3

One-compartment pharmacokinetic model

In our example, we have

$$f(t,\vartheta^o) = \left(\begin{array}{c} \frac{\partial \eta}{\partial k_a} \\ \\ \frac{\partial \eta}{\partial k_e} \end{array} \right) \Big|_{\vartheta=\vartheta^o} = \left(\begin{array}{l} -\frac{1}{(k_a^o - k_e^o)} \left\{ \frac{k_e^o}{k_a^o - k_e^o} e^{-k_e^o t} - (t k_a^o + \frac{k_e^o}{k_a^o - k_e^o}) e^{-k_a^o t} \right\} \\ -\frac{k_a^o}{(k_a^o - k_e^o)} \left\{ \frac{1}{k_a^o - k_e^o} e^{-k_a^o t} - (t + \frac{1}{k_a^o - k_e^o}) e^{-k_e^o t} \right\} \end{array} \right).$$



Partial derivatives of η with respect to the parameters k_a and k_e at $k_a^o = 0.7$ and $k_e^o = 0.2$.

Example 3

One-compartment pharmacokinetic model

Hence, after linearisation and adjusting for a constant, the model can be written as

$$y = k_a f_1(t, \vartheta^o) + k_e f_2(t, \vartheta^o) + \varepsilon,$$

where

$$f_1(t, \vartheta^o) = \frac{\partial \eta}{\partial k_a}\Big|_{\vartheta=\vartheta^o} \text{ and } f_2(t, \vartheta^o) = \frac{\partial \eta}{\partial k_e}\Big|_{\vartheta=\vartheta^o}.$$

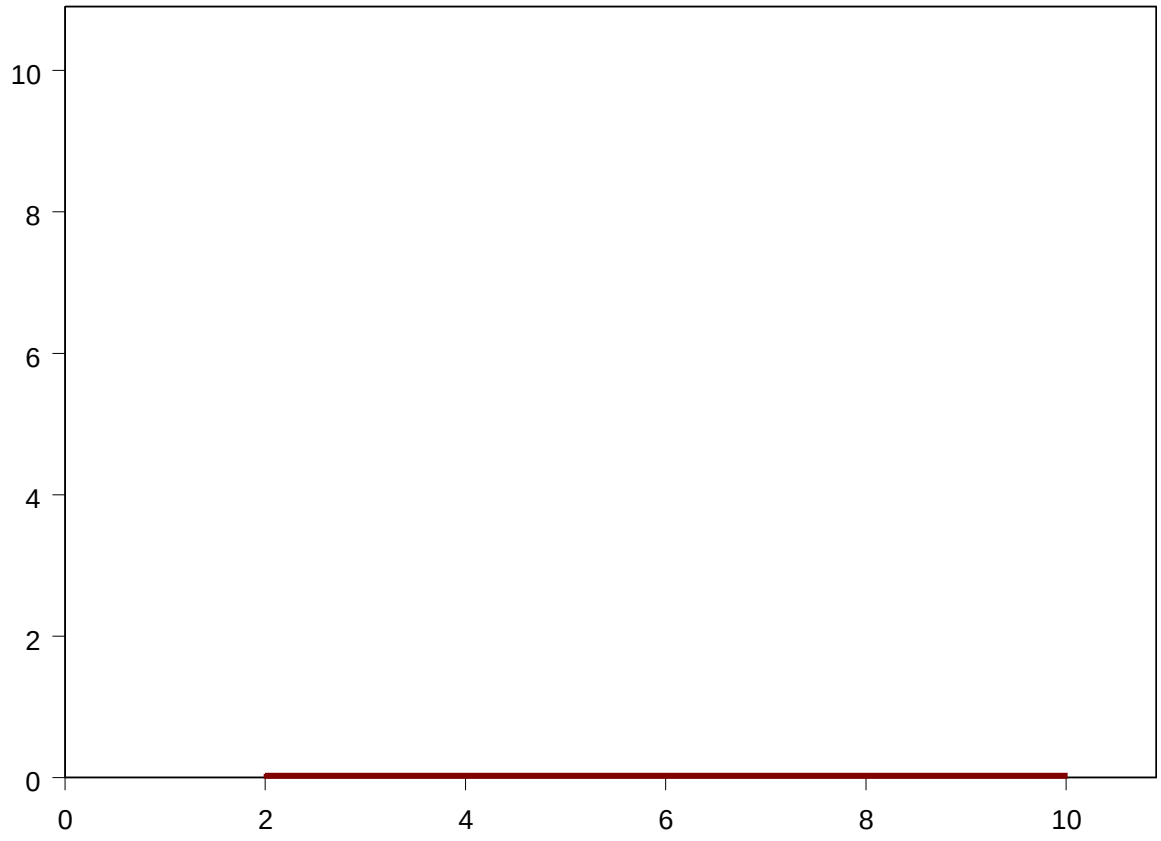
Compare it with the simple linear regression model

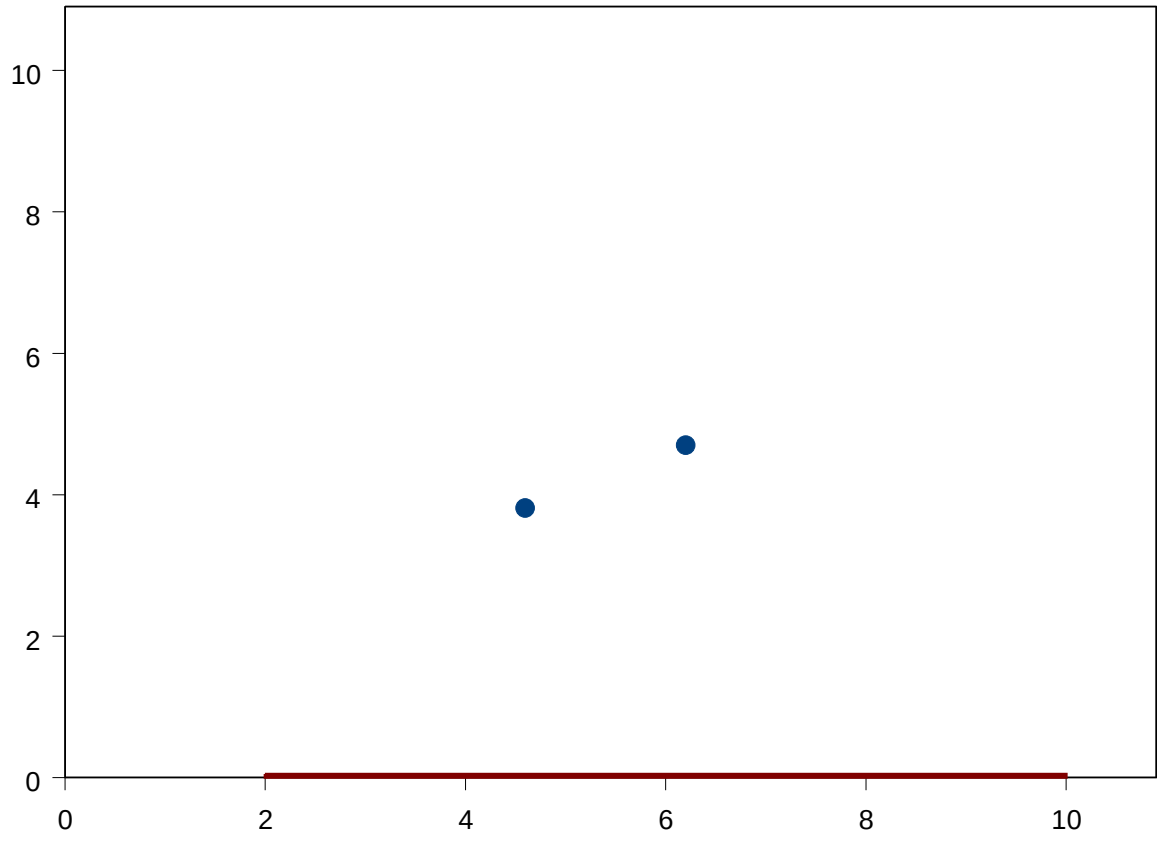
$$y = \vartheta_1 + \vartheta_2 t + \varepsilon$$

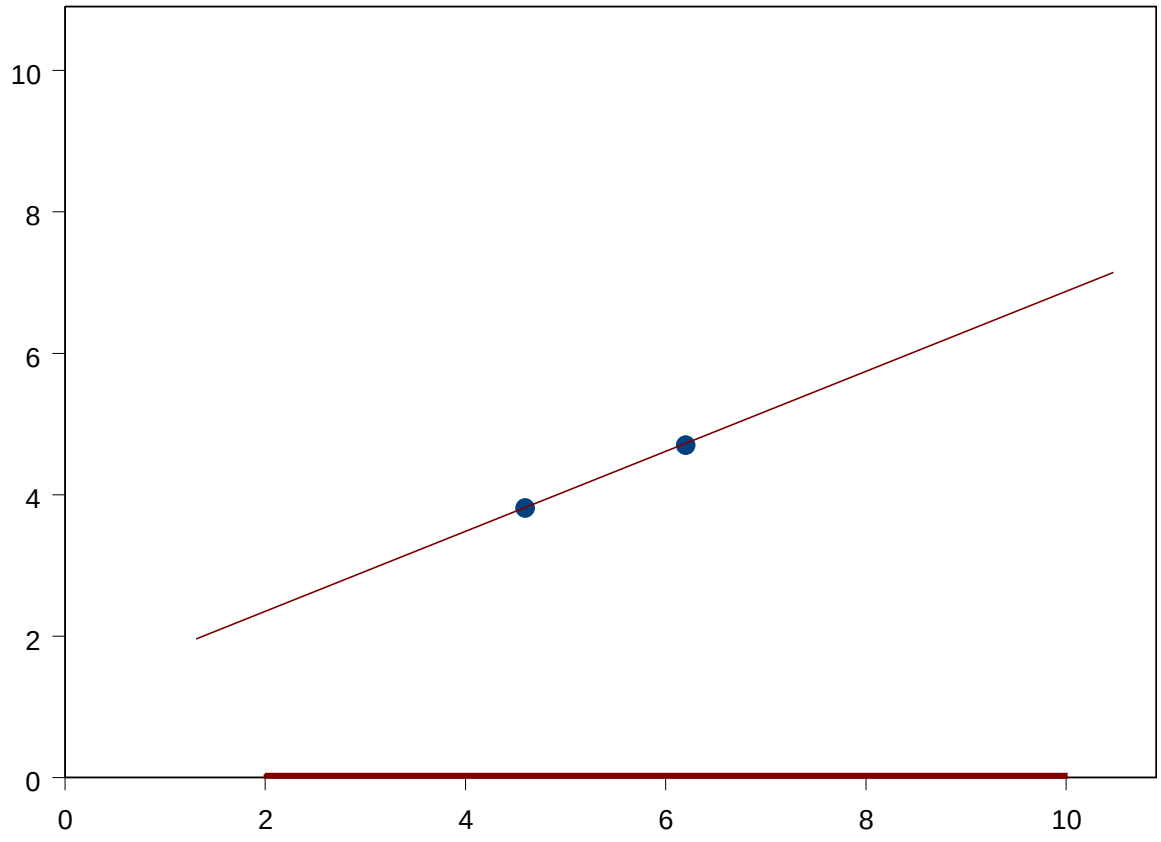
in which

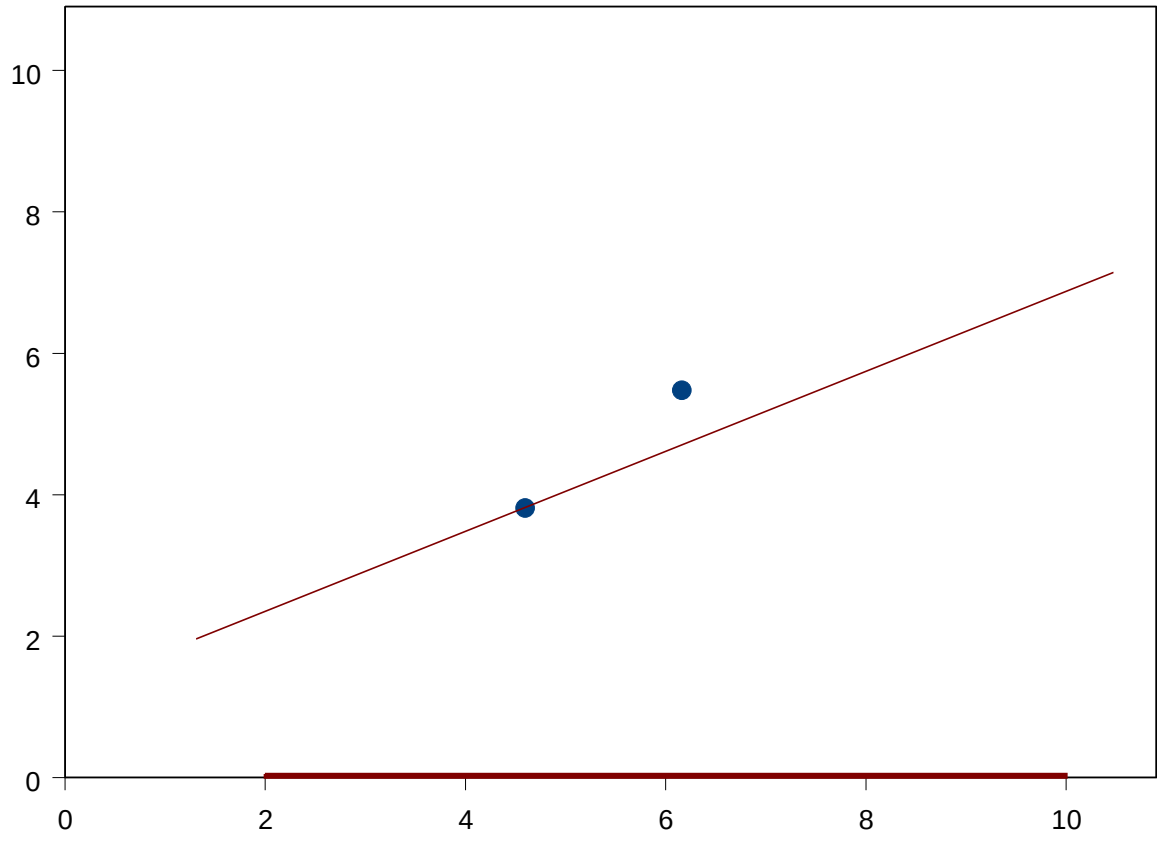
$$f_1 = 1, \quad f_2 = t.$$

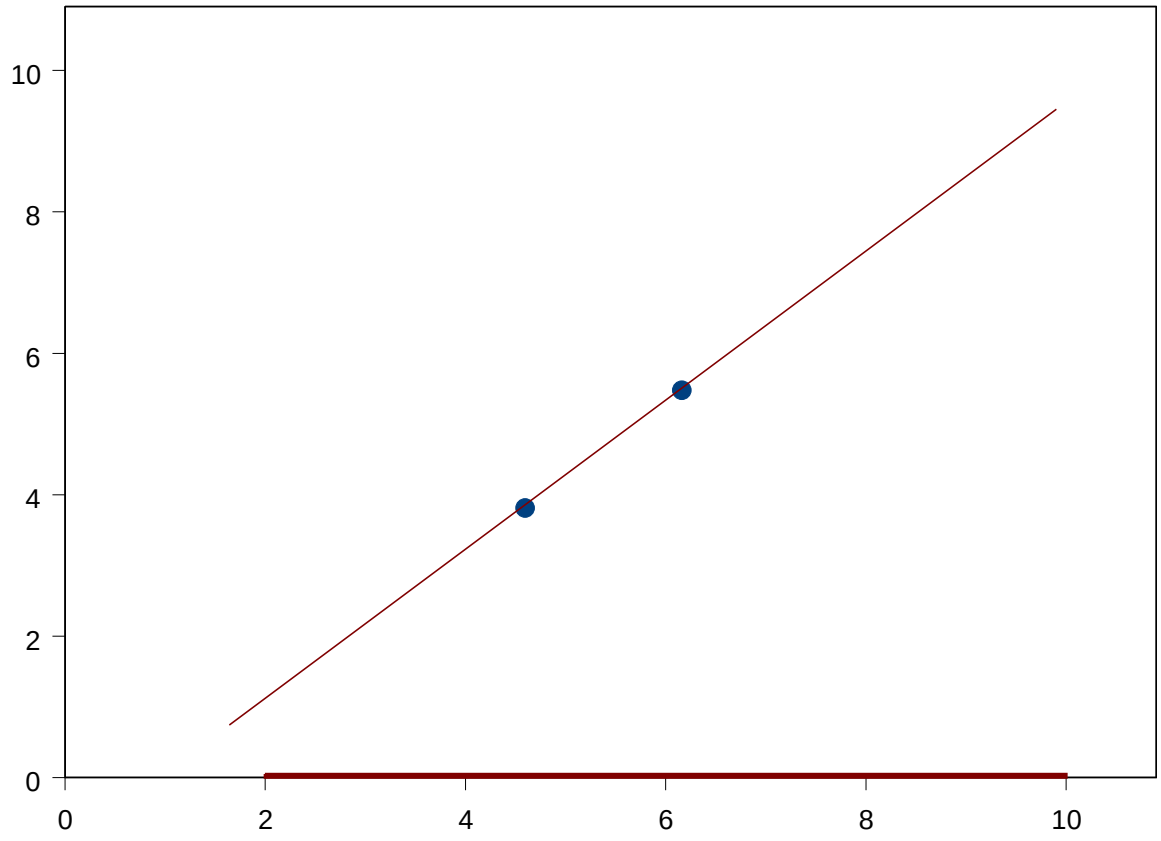
What would be a good design for such a model?

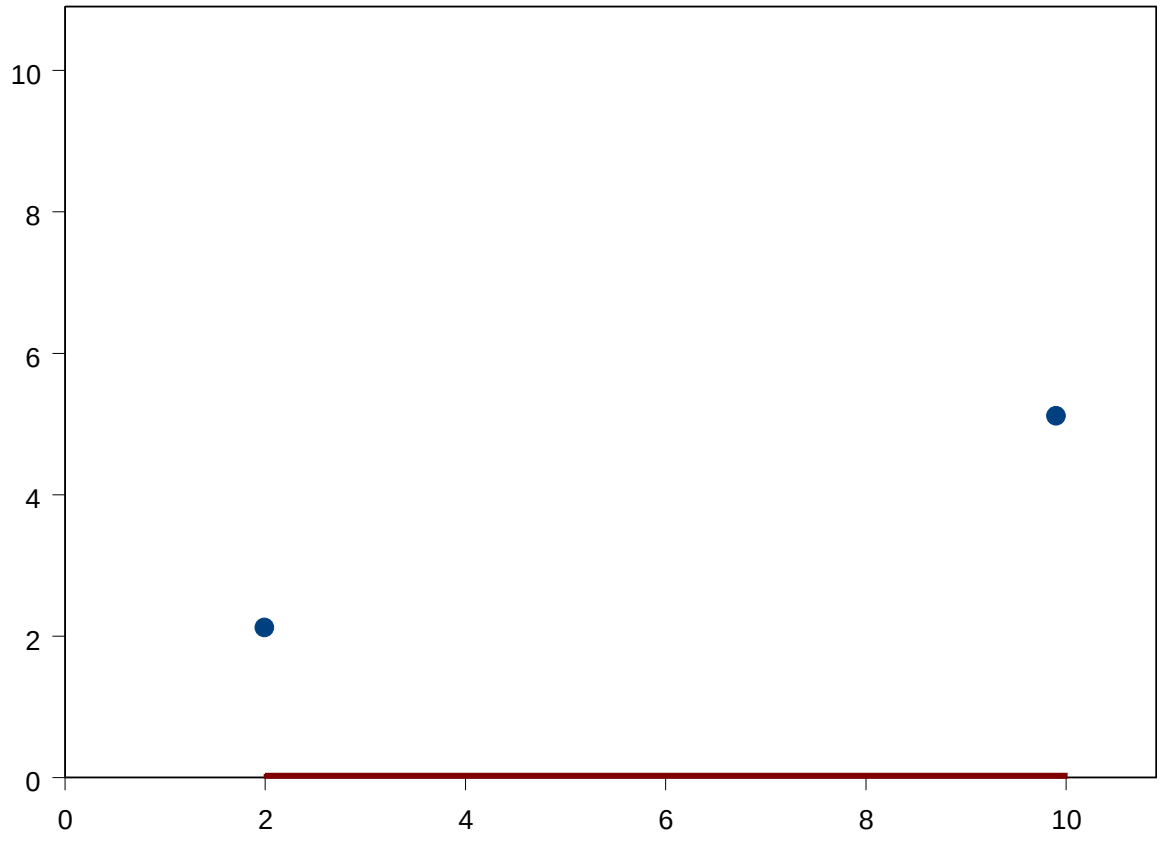

10
8
6
4
2
0
0
2
4
6
8
10

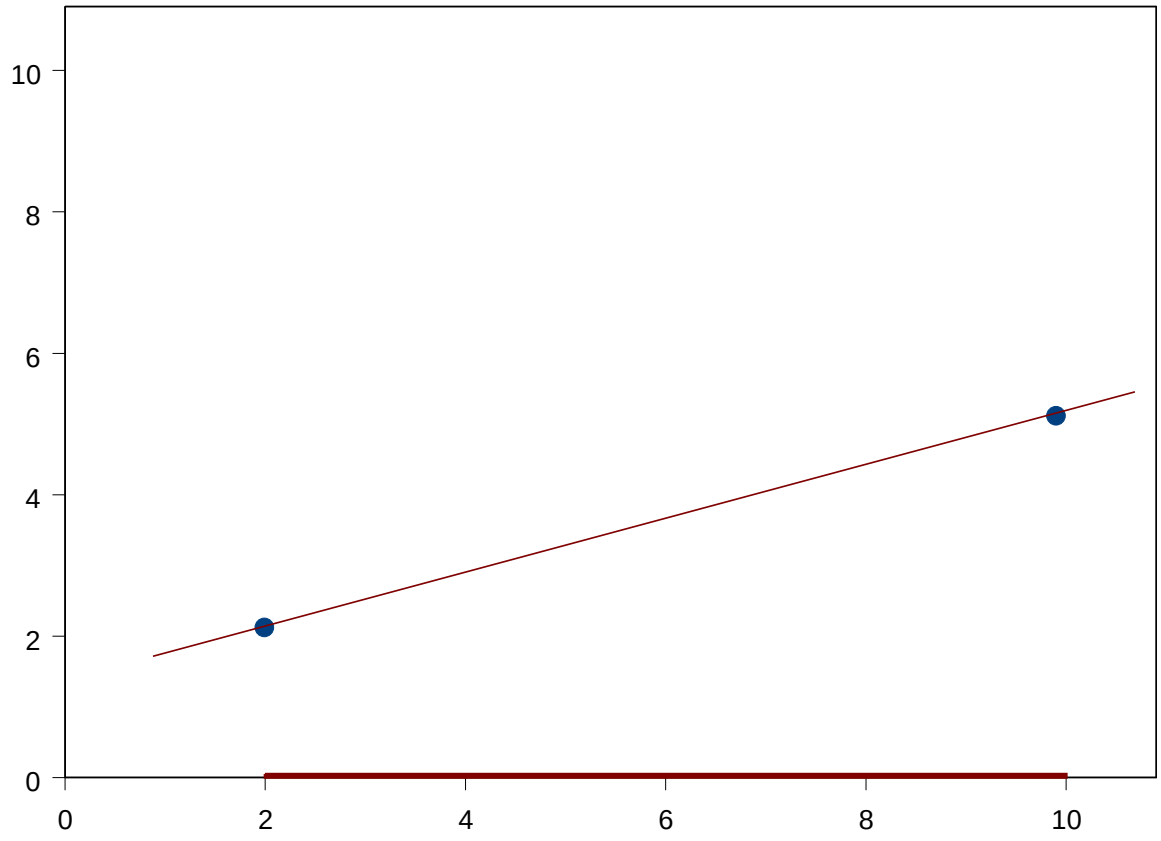

10
8
6
4
2
0
0
2
4
6
8
10

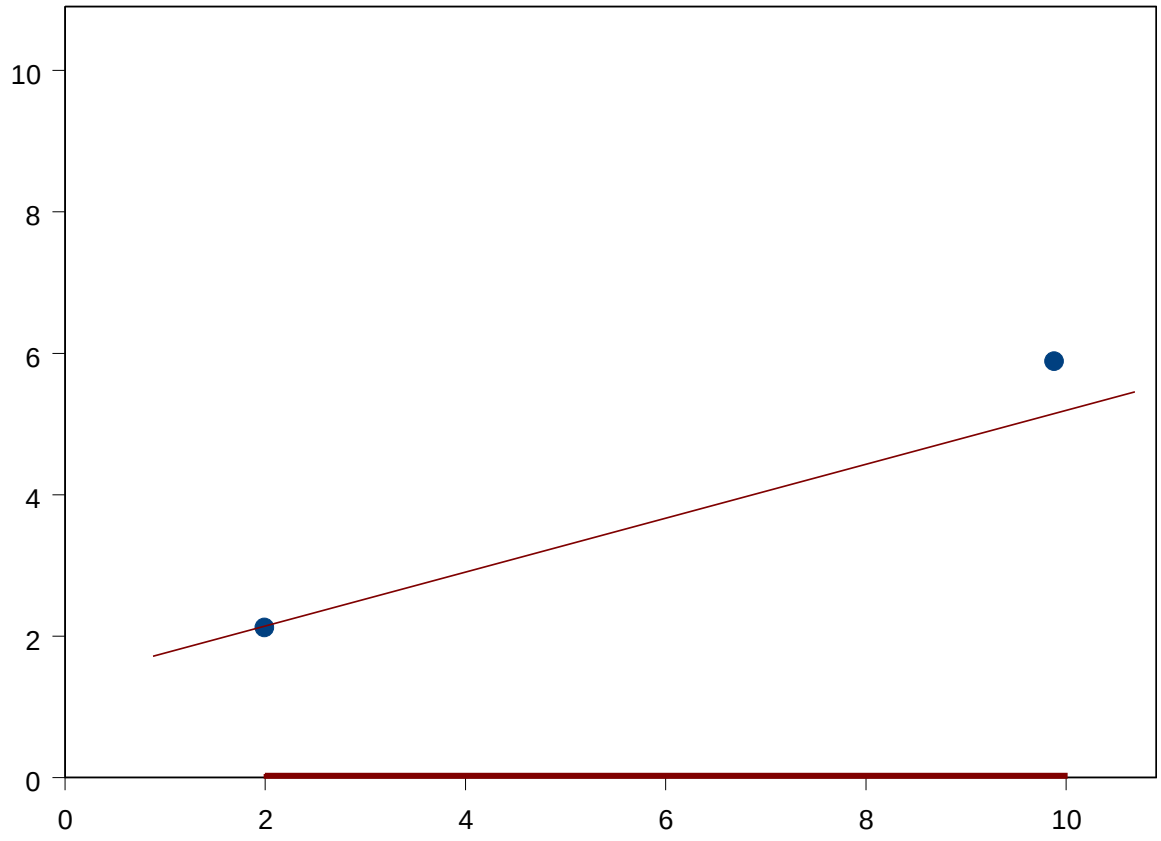

10
8
6
4
2
0
0
2
4
6
8
10

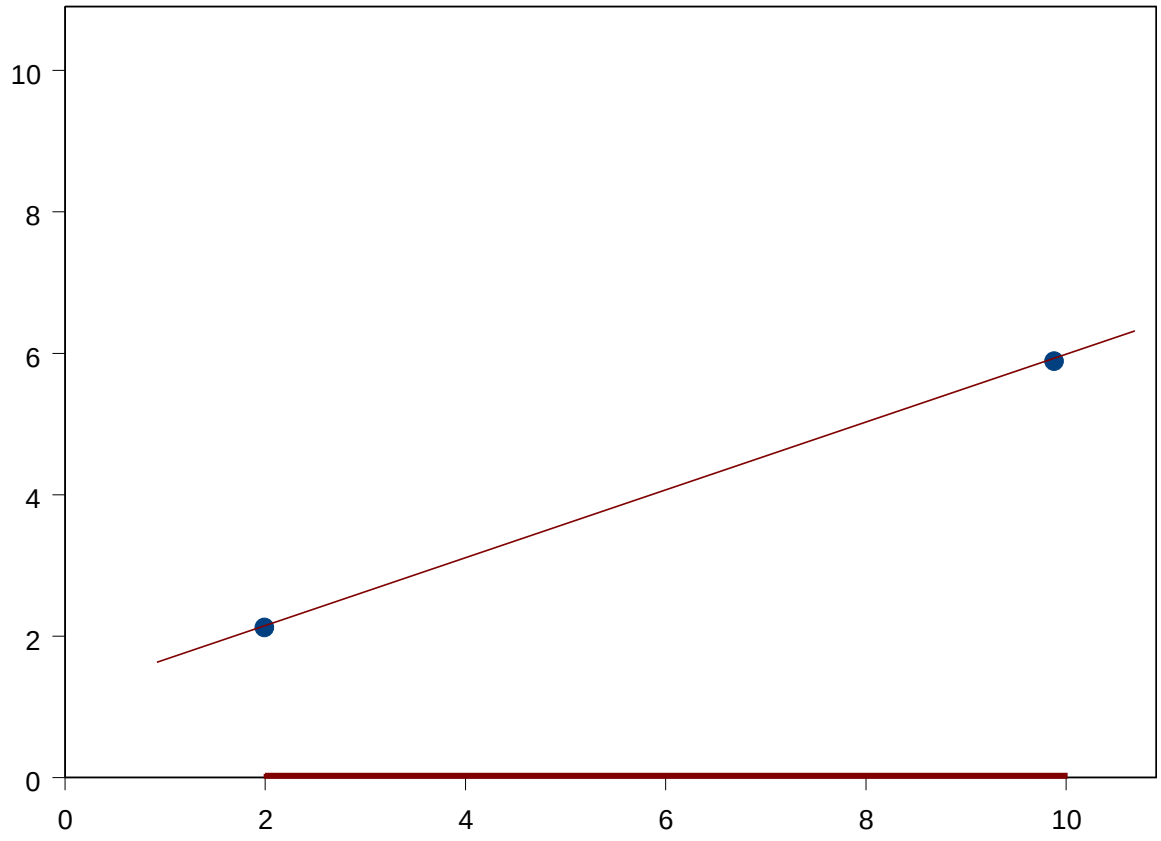

10
8
6
4
2
0
0
2
4
6
8
10

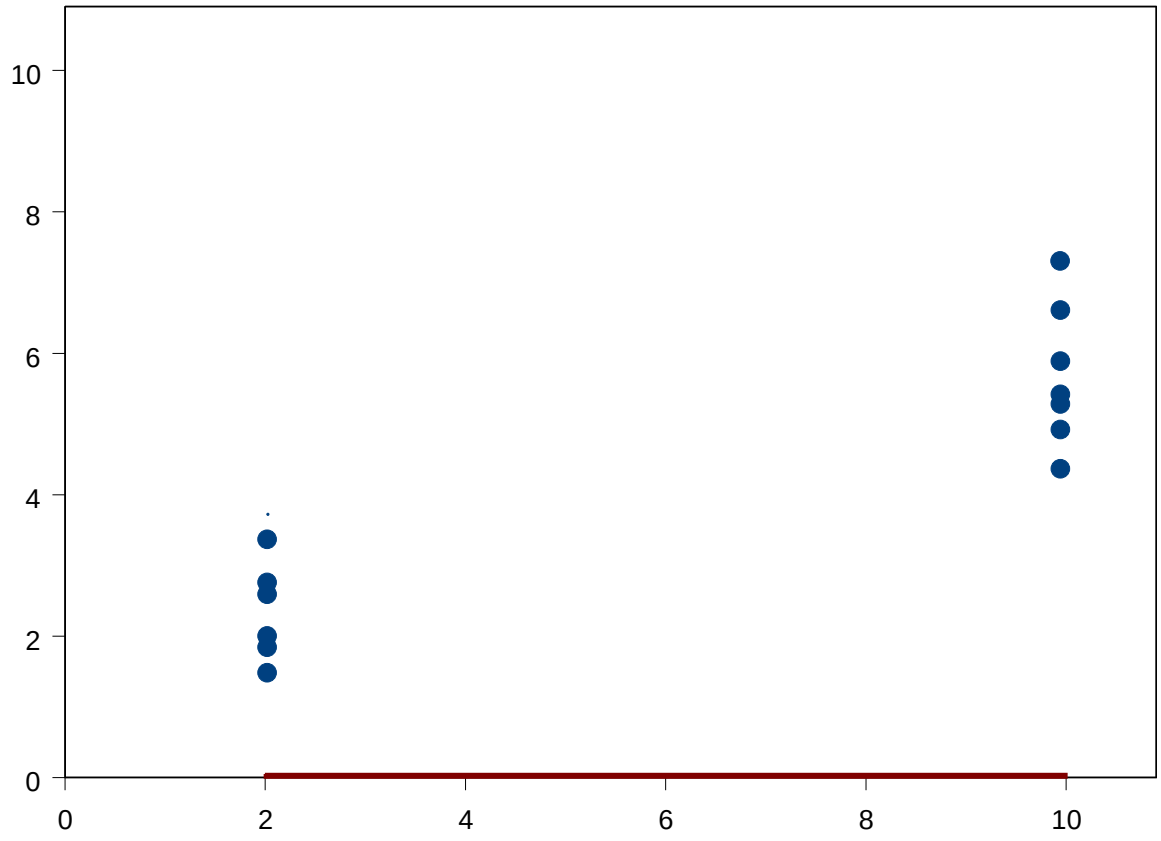

10
8
6
4
2
0
0
2
4
6
8
10

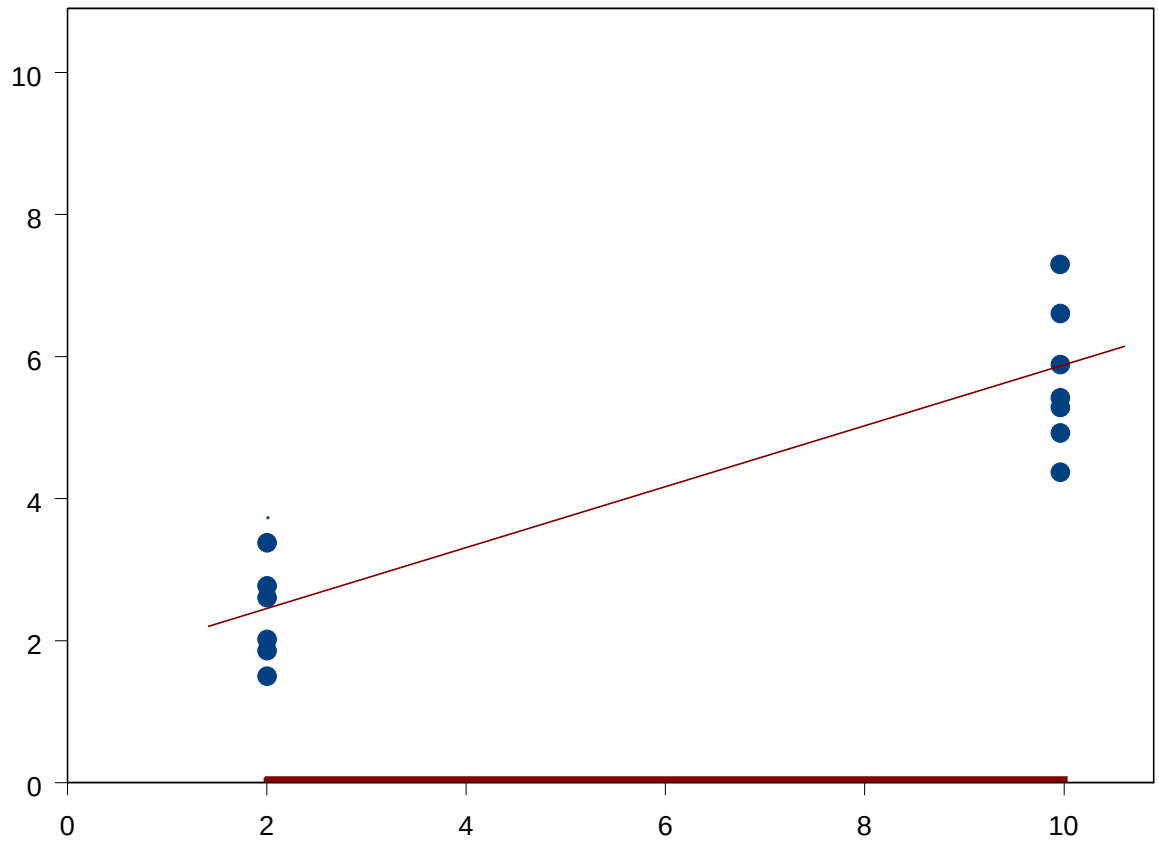

10
8
6
4
2
0
0
2
4
6
8
10


10
8
6
4
2
0
0
2
4
6
8
10


10
8
6
4
2
0
0
2
4
6
8
10


10
8
6
4
2
0
0
2
4
6
8
10


10
8
6
4
2
0
0
2
4
6
8
10



Well-designed experimental data for fitting a simple linear regression model.

Big Question

What is the best choice of the design points when the model is non-linear?

Purpose of an experiment - statistical point of view

- estimation of the parameters and their functions (e.g. contrasts) and further testing of statistical hypotheses (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.

What is a Design of an Experiment?

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

- Classical (combinatorial) design
  - way of allocating treatments to experimental units,
  - usually applied to linear models,
  - treatment structure and unit structure have to be defined and matched.

What is a Design of an Experiment?

- Continuous (approximate) design
  Here,
$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_s \\ w_1 & w_2 & \dots & w_s \end{array} \right\},$$
  where
$$\sum_{i=1}^{s} w_i = 1, \quad w_i > 0.$$
  - 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.

Optimum Design of Experiments

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

How can we measure the precision of estimation?

- via variance and bias of the estimator

Optimum Design of Experiments

- **When there are several competing models, then**
  - Purpose:
    - **discrimination** between the models,
    - **estimation of the parameters and discrimination**.
  - Design: to optimise 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 optimise for the specific objective.

Optimum Design of Experiments

Problems

- In linear models, a combinatorial optimum design for particular treatment and unit structures may not exist.
- In non-linear models, it is possible to find an approximate (continuous) optimum design, but it depends on
  - prior values of the unknown parameters,
  - curvature of the assumed model,
  - usually there are no closed-form solutions.
- In any case, the optimum design depends on the assumptions regarding the variability and correlation of the observed response.

Criteria for design optimality

For parameter estimation

Most of the criteria for parameter estimation were introduced for linear models and are functions of the Fisher information matrix

$$M = \left\{ \mathrm{E}\left(\frac{\partial l}{\partial \vartheta_i} \frac{\partial l}{\partial \vartheta_j} \right) \right\}_{i,j=1,2,\ldots,p},$$

where $l = \log L(\vartheta; y)$ and $L(\vartheta; y)$ is the likelihood function for ϑ given y.

This comes from the (asymptotic) properties of maximum likelihood estimators:

$$\widehat{\vartheta} \underset{approx.}{\sim} \mathcal{N}_p\left(\vartheta, M^{-1}\right).$$

Criteria for design optimality

For parameter estimation

For a normal linear model for the observations

$$Y \sim \mathcal{N}_n(X\vartheta, V),$$

the likelihood function is

$$L(\vartheta; y) = \frac{1}{(2\pi)^{n/2}\sqrt{|V|}} \exp\left\{-\frac{1}{2}(y - X\vartheta)^{\mathrm{T}} V^{-1}(y - X\vartheta)\right\}.$$

Hence,

$$l = \text{const} - \frac{1}{2}(y - X\vartheta)^{\mathrm{T}} V^{-1}(y - X\vartheta),$$

$$\frac{\partial l}{\partial \vartheta} = X^{\mathrm{T}} V^{-1}(y - X\vartheta)$$

and

$$M = \mathrm{E}\left\{X^{\mathrm{T}} V^{-1}(Y - X\vartheta)(Y - X\vartheta)^{\mathrm{T}} V^{-1} X\right\} = X^{\mathrm{T}} V^{-1} X.$$

Criteria for design optimality

For parameter estimation

For a non-linear model for the observations, we have

$$X = \begin{pmatrix} f(x_1, \vartheta^o)^{\mathrm{T}} \\ \vdots \\ f(x_n, \vartheta^o)^{\mathrm{T}} \end{pmatrix}, \quad f(x_i, \vartheta^o)^{\mathrm{T}} = \begin{pmatrix} \frac{\partial \eta(x_i, \vartheta^o)}{\partial \vartheta_1}, & \ldots, & \frac{\partial \eta(x_i, \vartheta^o)}{\partial \vartheta_p} \end{pmatrix}.$$

The quantity

$$\frac{\partial \eta(x_i, \vartheta)}{\partial \vartheta_j} \text{ for } j = 1, 2, \ldots, p \text{ and } i = 1, 2, \ldots, n$$

is called the **parameter sensitivity**.

Criteria for design optimality

For parameter estimation

Common criteria (functions of the Fisher information matrix) introduced by Wald (1943), Elving (1952), Kiefer (1959), Kiefer (1975) are

- **A** - minimises the average variance of contrasts of treatment effects,
- **D** - minimises the volume of the confidence ellipsoid for the unknown model parameters,
- **E** - minimises the widest confidence interval among confidence intervals for all parameters,
- **G** - minimises the variance of prediction of the model function,
- **c** - minimises the variance of a function of the parameters (i.e. area under the curve),
- Φ_p - a general class of criteria, which includes A, D and E,
- **Universal** - a very general convex function, which includes an even wider class of criteria.