Aalto University
School of Chemical
Technology

Design of experiments

Mikko Mäkelä

Aalto University, School of Chemical Engineering
Department of Bioproducts and Biosystems
Espoo, Finland

Session 1

Introduction

- Why experimental design

Factorial design

- Design matrix
- Model equation = coefficients
- Residual
- Response contour

Session 2

Factorial design

- Research problem
- Design matrix
- Model equation = coefficients
- Degrees of freedom
- Predicted response
- Residual
- ANOVA
- R^2
- Response contour

Research problem

A chemist is interested on the effect of temperature (A), catalyst concentration (B) and time (C) on the molecular weight of polymer produced

- She performed a 2^3 factorial design

Parameter	Low	High
A (°C)	100	120
B (%)	4	8
C (min)	20	30

Myers et al., Response Surface Methodology (3rd ed.); 2009: 131.

Design matrix

N:o	A	B	C	Resp.
1	100	4	20	2400
2	120	4	20	2410
3	100	8	20	2315
4	120	8	20	2510
5	100	4	30	2615
6	120	4	30	2625
7	100	8	30	2400
8	120	8	30	2750

Build a design matrix with interactions and determine the coefficients

Model

Empirical model

$y_c = f(A, B, C)$

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

In matrix notation

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} \rightarrow \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ 1 & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

Measure (y) **Choose** (X) **Unknown!** (b, e)

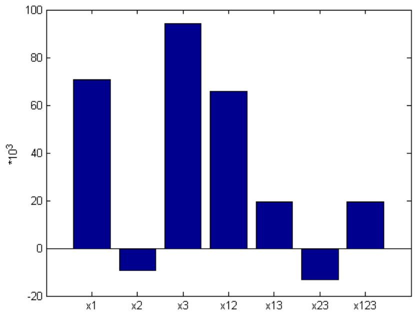
Coefficients

Model coefficients

- Least squares calculation

Coefficient significance?

- 8 model terms



Aalto University
School of Chemical Technology

25.3.2019
7

Degrees of freedom

Degrees of freedom (df) lost by imposing linear constraints on a sample

- E.g. sample variance:

$$s^2 = \frac{\sum (y-\bar{y})^2}{n-1} \quad \text{where} \quad \sum (y-\bar{y}) = 0$$

→ *n - 1* residuals can be used to completely determine the others

In regression models, dfs are lost due to the constraints imposed by the coefficients

- Residual df *n - p* with *n* observations and *p = k + 1* model terms

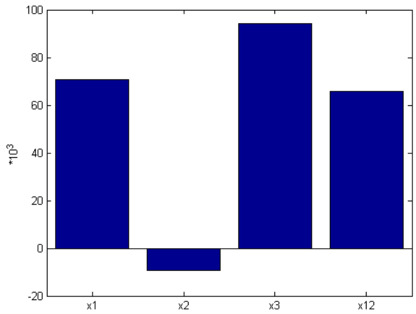
Aalto University
School of Chemical Technology

25.3.2019
8

Coefficients

Remove unimportant model terms

and calculate new coefficients



Aalto University
School of Chemical
Technology

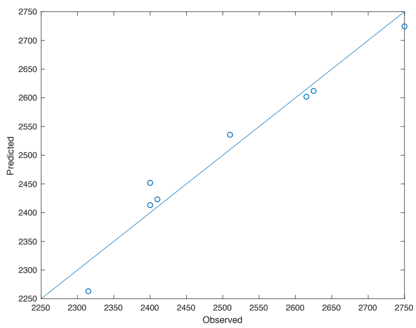
25.3.2019
9

Predicted response

Calculate predicted response,

- Based on **X** and **b**

Graphical tools are useful



Aalto University
School of Chemical
Technology

25.3.2019
10

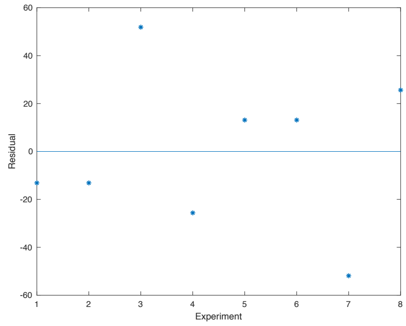
Residuals

Calculate model residuals

A common way is to scale the residuals

- E.g. standardized residual

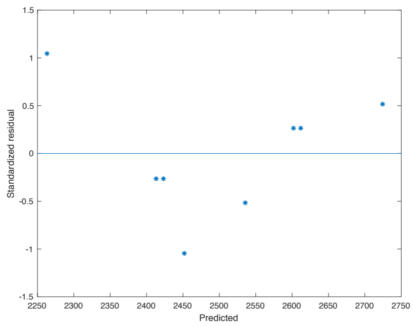
→ Need an error approximation



Residuals

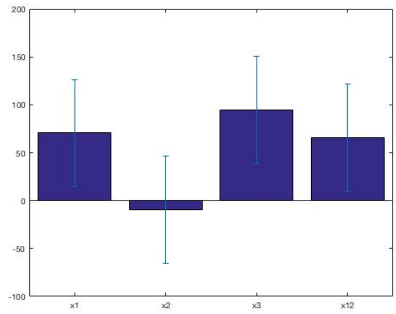
Standardized residuals

- >|3| generally a potential outlier, why?



Coefficients

Standard error / confidence interval of model coefficients



Aalto University
School of Chemical
Technology

25.3.2019
13

ANOVA

Analysis of variance (ANOVA) allows to statistically test the model

- F test for variances
- H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$
- H_1: at least one $\beta \neq 0$

Parameter	df	Sum of squares (SS)	Mean square (MS)	F-value	p-value
Total corrected	n-1	SStot			
Model	k	SSmod	MSmod	MSmod /MSres	<0.05-0.10
Residual	n-k-1	SSres	MSres		

Aalto University
School of Chemical
Technology

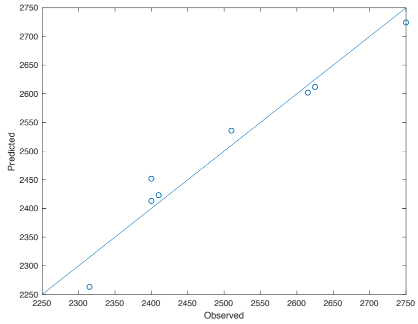
25.3.2019
14

R^2 statistic

Data variation explained by the model

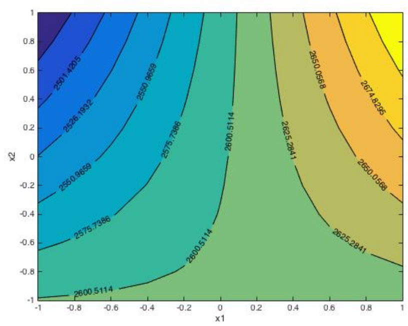
- Compares model and total sum of squares

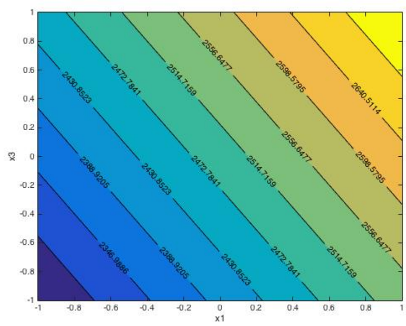
R^2 = 95%



Response contour

For three factors, 2D contours require one constant factor

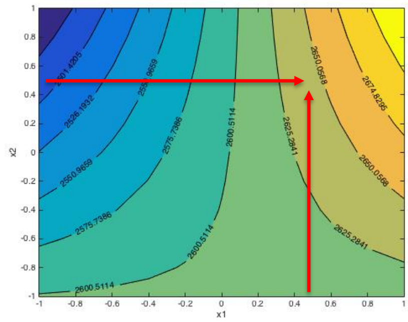




Response contour

Use of the model

- Prediction
- Finding an optimum
- Verification



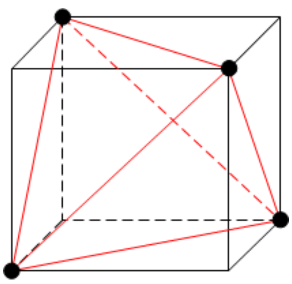
A = 115 and B = 7

- $\mathbf{x}_m$ = [1 0.5 0.5 1 0.25]
- $y_m$hat = 2645

Fractional factorials

With many factors factorial designs require a lot of experiments (2^k)

→ Fractional factorials



Denoted e.g. 2^{3-1}

- Half-fraction of a 2^3 design
- Enables only a main effect model
- Watch out for aliases

Fractional factorials

Defining relation I = ABC

A	B	C	AxBxC
-	-	-	-
+	-	-	+
-	+	-	+
+	+	-	-
-	-	+	+
+	-	+	-
-	+	+	-
+	+	+	+

A	B	C	AxBxC
-	-	-	-
+	+	-	-
+	-	+	-
-	+	+	-

A	B	C	AxBxC
+	-	-	+
-	+	-	+
-	-	+	+
+	+	+	+

Alias structure

Some terms cannot be independently estimated due to aliasing

- I = ABC → A = BC, B = AC, C = AB

A	B	C	AB	AC	BC	AxBxC
-	-	-	+	+	+	-
+	+	-	+	-	-	-
+	-	+	-	+	-	-
-	+	+	-	-	+	-

Session 2

Factorial design

- Research problem
- Design matrix
- Model equation = coefficients
- Degrees of freedom
- Predicted response
- Residual
- ANOVA
- R^2
- Response contour
- Fractional factorials

Nomenclature

Design matrix

Coefficient

Degrees of freedom

Prediction

Residual

Outlier

ANOVA

Sum of squares

Mean square

Contour

Alias

Thank you!