

# **Reducing simulation time with Design of Experiments**

Jacek Kalowski

Manhaz/NCBJ

Let's make some...



# How to measure pizza?

- Taste?
- No
- Crisp?
- Maybe

On a scale 1-10

Called „response”

# How many samples?

Depends on number of factors considered

$n$  – number of factors

For two level experiment it is  $2^n$

It can be reduced (but not for free)...

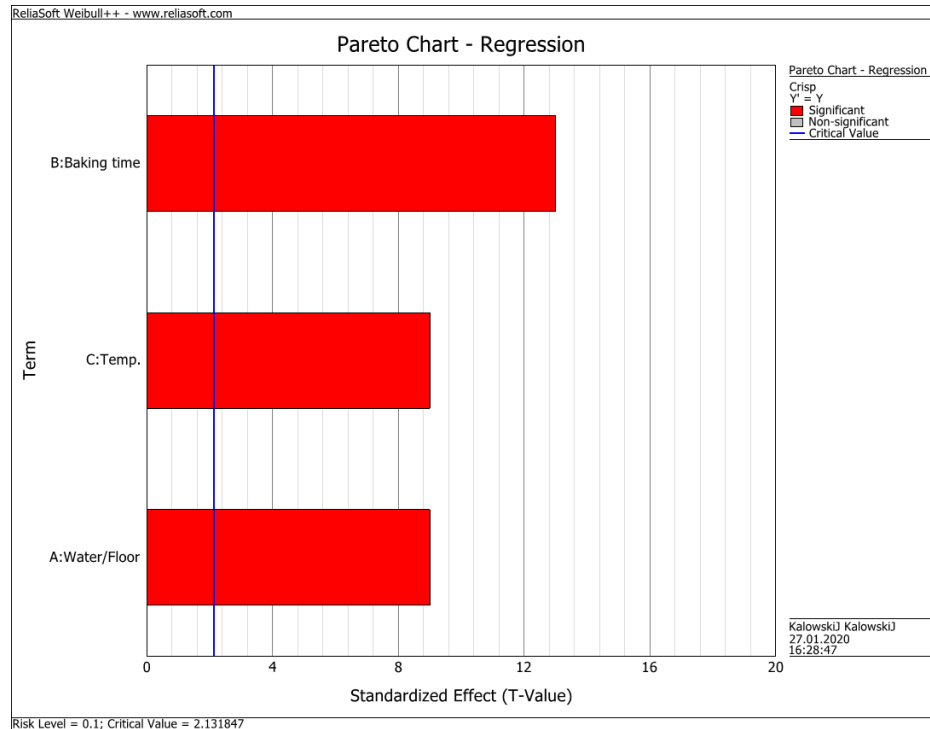
A:Water/Floor	B:Baking time (min)	C:Temp. (deg.C)
0.6	4	250
0.75	6	400

# Example

Let's consider factors (and response) values:

A:Water/Floor	B:Baking time (min)	C:Temp. (deg.C)	Crisp
0.75	6	250	6
0.75	6	400	8
0.6	6	400	10
0.75	4	250	2
0.75	4	400	5
0.6	4	400	7
0.6	4	250	5
0.6	6	250	8

# Results



- All factors are significant
- Model is reduced to 3 main factors
- Interactions are not considered

# Boring results

**ANOVA Table**

ANOVA Table					
Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Model	3	41.375	13.791667	110.333333	0.000266
Main Effects	3	41.375	13.791667	110.333333	0.000266
Residual	4	0.5	0.125		
Lack of Fit	4	0.5	0.125		
Total	7	41.875			

S = 0.353553

R-sq = 98.81%

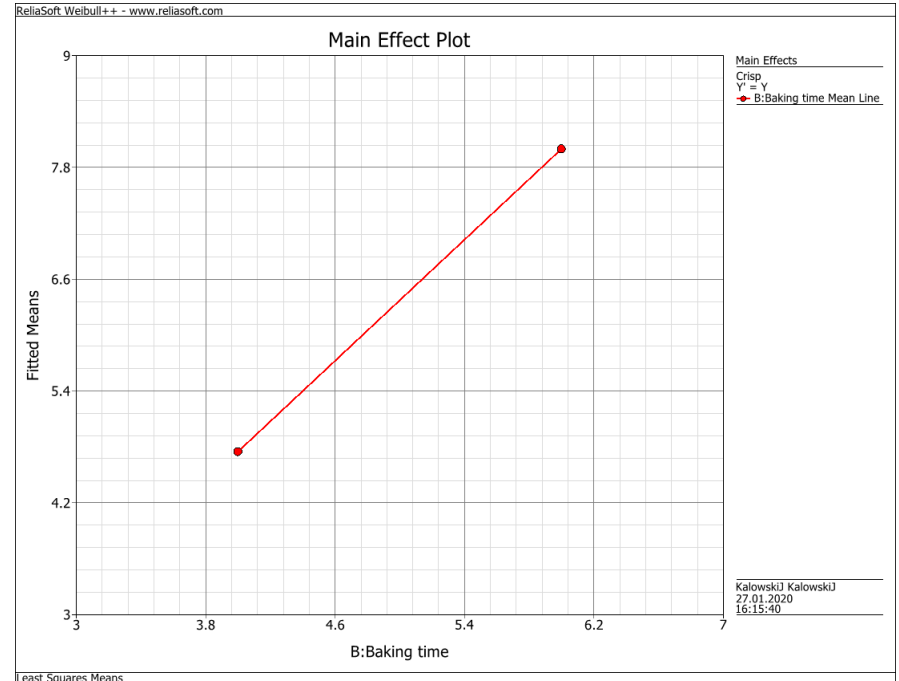
R-sq(adj) = 97.91%

**Regression Table**

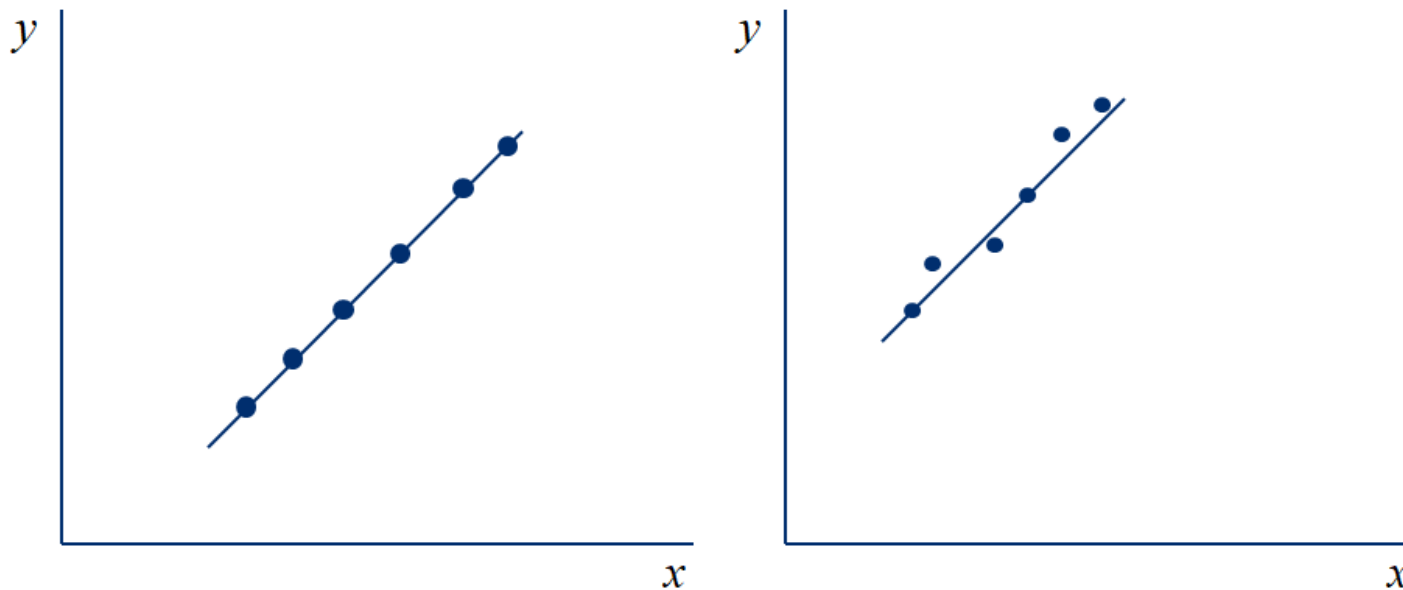
Regression Information							
Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
Intercept		6.375	0.125	6.108519	6.641481	51	8.846230E-7
A:Water/Floor	-2.25	-1.125	0.125	-1.391481	-0.858519	-9	0.000844
B:Baking time	3.25	1.625	0.125	1.358519	1.891481	13	0.000202
C:Temp.	2.25	1.125	0.125	0.858519	1.391481	9	0.000844

# Test hypothesis slope=0

- Showing the mean effect on response when Baking time is low vs when it is high
- To see if it is significant one has to look at variability of results



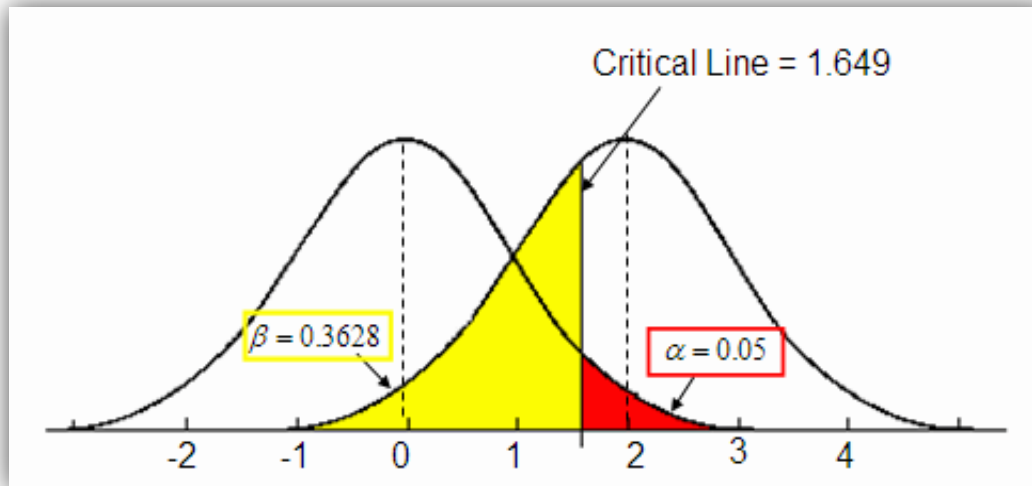
# Is it significant?



- Response variability explained by the model (left)
- Not all variability explained by the model (right)
- Is the model significant (compared to variability)? Are we trying to fit the model to the noise?

# Avoiding errors

	H0 is true	H1 is true
Fail to Reject H0	Correct Decision	Type II error, $\beta$
Reject H0	Type I error, $\alpha$	Correct Decision, Power = $1 - \beta$



- Alfa is selected at the beginning (safe error)
- Beta should be always considered – how likely it is I will miss significant effect because I have a sample that is too small

# Power of the experiment

Power Study		
	Degrees of Freedom	Power
<b>A:Water/Floor</b>	1	0.322736
<b>B:Baking time</b>	1	0.322736
<b>C:Temp.</b>	1	0.322736
<b>Residual</b>	4	-

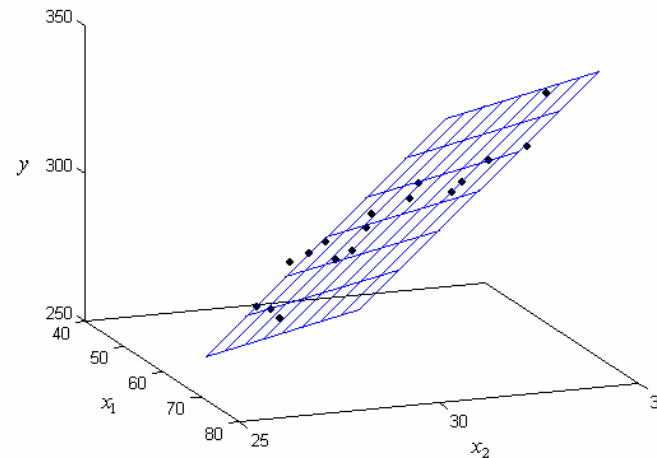
- In DoE software you can calculate Power of the experiment (1-beta). In this case beta is very high at almost 68%
- In our case all effects are significant - but if they are not, how can one be sure that significant effect was not missed?

# Regression model

Observation Number	X1	X2	Yield
1	41.9	20.46	251.3
2	43.4	28.06	251.3
3	43.9	21.99	248.3
4	44.5	21.7	267.5
5	47.3	24.11	273
6	47.5	27.11	276.5
7	47.9	23.3	270.3
8	50.2	23.29	274.9
9	52.8	20.87	285
10	53.2	26.72	290
11	56.7	23.56	297
12	57	28.66	302.5
13	63.5	24.12	304.5
14	65.3	29.05	309.3
15	71.1	20.39	321.7
16	77	20.88	330.7
17	77.8	26.19	349

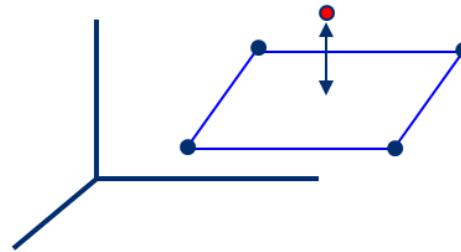
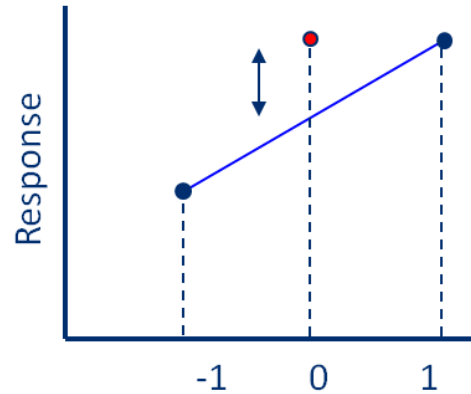
Using least squares, the model fitted to the data set is:

$$y = 135.54 + 2.3898x_1 + 0.8521x_2 + \varepsilon$$



- In most cases regression model should contain only factors found to be significant (reduced order model)
- Regression models are useful for predictions and they can use categorical factors too (1/0 i.e. present/not present)

# Non linear?



- If nonlinearity is found or expected in the data one can run tests for curvature, if additional center point measurements are done

# Options

- One Factor Designs: **For comparison**
- Factorial Designs: **For screening vital factors**
  - Full Factorial
    - Two Level Design
    - General Level Design
  - Fractional Factorial (**sample size reduction**)
    - Two Level Design
    - Plackett-Burman Design
    - Taguchi's Orthogonal Arrays
- Response Surface Designs: **For system optimization**
- Taguchi Robust Designs: **For system robustness**
- Reliability DOE: **For life test and data analysis**

# Opportunities for simulations

- Reduced order model (finding area of interest)
- Looking for minima/maxima
- Faster optimization
- Robustness study

Note: the general idea behind this presentation is that simulation time can be potentially reduced if simulation settings for each run are selected based on DoE techniques with minimal sample. This allows to calculate results faster and identify most useful additional simulation runs i.e learn in steps.