Boehringer Ingelheim

StatistiCs M ANCE

Evaluation of designs of experiment

to support process development and process optimization directly in the labs and facilities

Life forward

Overview

- Background
- Investigated designs
- Details of evaluation

  - supported order of regression models

  - power, accuracy of estimation/prediction, optimality criteria

  - effects in case of model selection (with AICc)

  - applicability of lack-of-fit test
- Results
- Outlook/ next steps

Background

During process development experiments are often planned and analyzed directly in the lab, with little ~~(or late) i~~nput from statisticians.

+ Good background knowledge.
  Quick and direct support.

– Risk of useless experiments,
  if design is not adequate.

- Frontload of the statistical support: Present and discuss the Pros&Cons of different designs before the actual use case. Inform about possible errors and pitfalls.

- Motivate to re-use DoEs (with clearly understood risks and benefits) instead of optimizing with software. (Understanding a good design is better than misusage of the “best“ design.)

- Identify complex situations, where an individual design of experiments with statistical support is still the best option.

Background

During process development experiments are often planned and analyzed directly in the lab, with little ~~(or late)~~ input from statisticians.

\+ Good background knowledge.

Quick and direct support.

− Risk of useless experiments,

if design is not adequate.

Current Project:

Summarizing characteristics and strengths of different designs of experiment,

supported by simulation results,

in order to develop a tool for chosing appropriate DoE strategies.

Investigated designs*

- Full factorial designs
- Central composite designs
- D-, I-, A- optimal designs
- Definitive screening designs

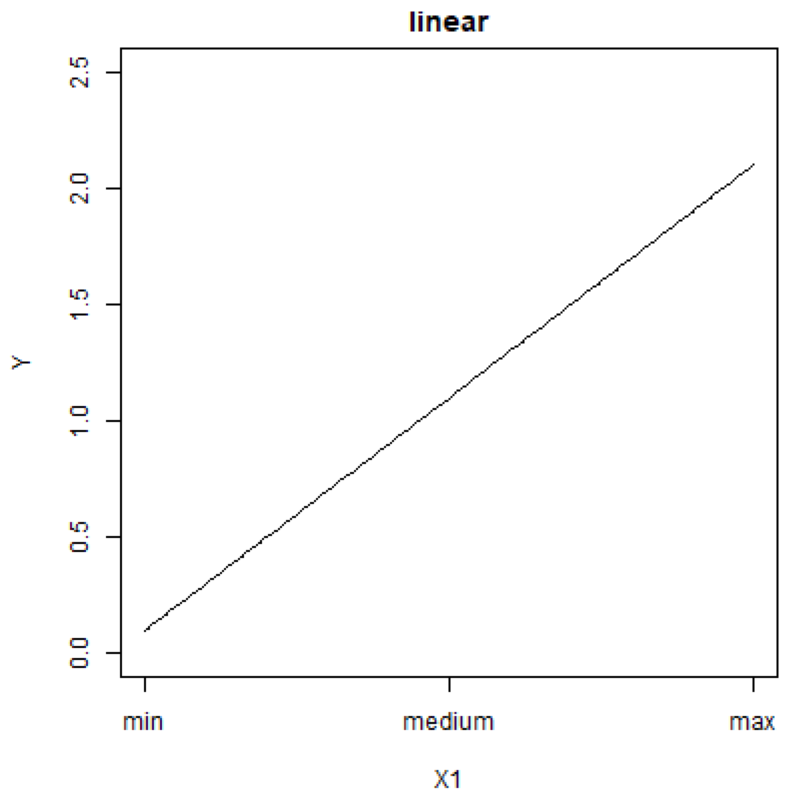
.. that are analyzed using regression models.

*Selection was motivated by what was used/misused in the labs.

Evaluation: Supported order of regression models

1.Visualization of linear, quadratic, cubic, interaction effect, as basis for discussion about expected effects.

2. Showing which designs support effects up to which order.



Evaluation: Power, optimality criteria, …

1. The power to find the expected effects is usually the most important criteria, and well understood by non-statisticians.

   → Power calculations for effect sizes between 1 and 4 were included.

2. Optimality criteria are often used as rationale for a chosen design. However, often a different design may also be not far from optimality, while on the other hand optimal designs in one criterion might be low performer in other criteria. See for example:

   MWANGI, Wangui Patrick; ANAPAPA, Ayubu; OTIENO, Argwings. **Selection of Second Order Models' Design Using D-, A-, E-, T-Optimality Criteria.** *Asian Journal of Probability and Statistics*, 2019, 5. Jg., Nr. 2, S. 1-15.

Evaluation: Model Selection / AICc

Often model selection by best AIC or AICc is applied:

$$AIC = 2k - \ln(\hat{L})$$

k, number of parameters

$\hat{L}$, maximum Likelihood

$$AICc = AIC + \frac{2k^2 + 2k}{n - k - 1}$$

k, number of parameters

n, sample size

Remember, that AICc can be stricter or less strict than p-value selection (which is usually used in power calculation):
Example of possible pitfall due to model selection:

Design: 2 variables (-1,0,1)
D-optimal design for quadratic model.
Sample size N=9.
True Model: $Y \sim A + B + A^2$
Power for an effect size of 2 in the quadratic effect: >90%.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.0751     0.8138  -1.321  0.24372
A             3.1130     0.5559   5.600  0.00251 **
B             1.8099     0.6304   2.871  0.03494 *
I(A^2)        2.8236     0.9856   2.865  0.03520 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.362 on 5 degrees of freedom
Multiple R-squared:  0.8998,    Adjusted R-squared:  0.8397
F-statistic: 14.97 on 3 and 5 DF,  p-value: 0.006233
```

Model	AICc
$Y \sim A + B + A{:}B + A^2 + B^2$	147.7
$Y \sim A + B + A^2 + B^2$	75.7
$Y \sim A + B + A{:}B + A^2$	79.8
$Y \sim A + B + A{:}B + B^2$	88.5
$Y \sim A + B + A^2$	55.8
$Y \sim B + B^2$	63.5
$Y \sim A + B$	52.6
$Y \sim A + A^2$	52.6
$Y \sim B$	56.3
$Y \sim A$	48.4
$Y \sim 1$	52.5

Boehringer Ingelheim

Evaluation: Model Selection / AICc

Often model selection by best AIC or AICc is applied:

$$AIC = 2k - \ln(\hat{L})$$

k, number of parameters

$\hat{L}$, maximum Likelihood

$$AICc = AIC + \frac{2k^2 + 2k}{n - k - 1}$$

k, number of parameters

n, sample size

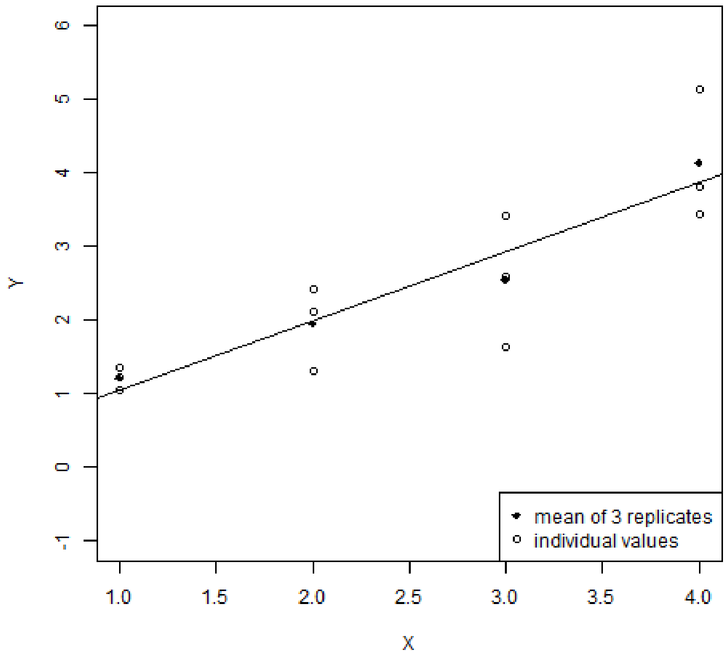
Remember, that AICc can be stricter or less strict than p-value selection (which is usually used in power calculation):
Example of possible pitfall due to model selection:

Design: 2 variables (-1,0,1)
D-optimal design for quadratic model.
Sample size N=9.
True Model: $Y \sim A + B + A^2$
Power for an effect size of 2 in the quadratic effect: >90%.

Simulation shows that in this design model selection by best AICc misses the correct model in 77% of the runs!

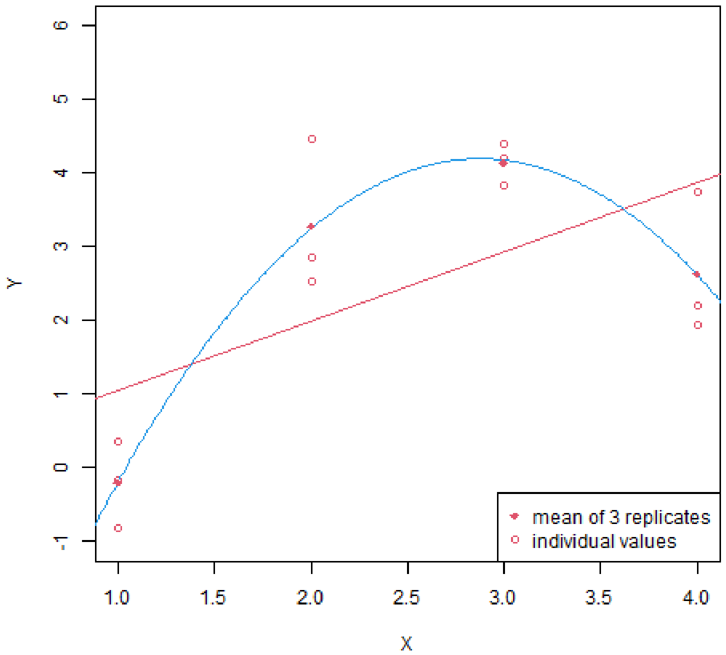
Evaluation: Lack of fit test

Lack of fit tests are included in the regression analysis of most GUI software (e.g. Design Expert, Minitab).



Design Expert /Minitab:
F-test for Variance ratio:
Observed variation around prediction versus variation of replicates around their mean.

F: 0.7598
p: 0.4988



Warning: False positive test results in unbalanced designs or under heteroscedasticity.

False negative test results in case of low number of replicates and lack of fit points.

→ False positive and false negative rate evaluated with simulation runs

Current presentation of results & Outlook

Information and simulation results stored in an Excel table:

Manually select similar designs to compare them (e.g. 3 variables, 21 runs):

1				LoF Test significant				Power		
2				3rd order effects (QL-I) with effect size:			/lodel,	Reduced Quadr. Model (all 3 lin, 1 2FI, 1 Quad, Δ=1)		
3	Design	# variables	Sample Size	Δ=2	Δ=1	Δ=0.5	2FI	Lin.	Quad.	2FI
16	Central composite (radial), with 7 center poi	3	21	98%	44%	13%	25%	41%	95%	26%
17	Central composite (k=1), with 7 center point	3	21	69%	23%	9%	25%	32%	57%	26%
18	D-opt, 3-desr., 3 Var, Quadr, 15 Mod, 0 LOF,	3	21	24%	9%	7%	37-38%	44-46%	48%	39%
19	D-opt, 5-desr., 3 Var, Quadr, 13 Mod, 4 LOF,	3	21	16%	9%	6%	26-28%	36-39%	54%	29%
20	D-opt, 5-desc., 3var, Cubic, 20 Mod + 1 CP	3	21	x	x	x	31-32%	42-43%	40%	33%
22	I-opt, 3-desr., 3 Var, Quadr, 15 Mod, 3 rep, 3	3	21	36%	12%	7%	23-24%	34-37%	56%	26%
26	Def.Screening (max 10)	3	21	50%	14%	6%	40%	51%	32%	47%

<u>Future enhancements:</u>

Add search options and filters, and additional designs.

Rshiny App that allows to upload an own design and compare the results to the designs in the table?

Provide training and Q&A.

Disclaimer