Análisis Secuencial Bayesiano de Diseños Experimentales Discriminantes Definitivos

Víctor Aguirre Torres

Departamento de Estadística , ITAM.

Seminario Aleatorio, 4 de Abril 2014

Credits

- Joint work with Román de la Vara (CIMAT).

Credits

- Joint work with Román de la Vara (CIMAT).
- Partial financial support by Asociación Mexicana de la Cultura, A. C.

Definitive Screening Experiments (DSE)

- Usually screening experiments consider 2 levels for each factor, 2^{m-k} plans

Definitive Screening Experiments (DSE)

- Usually screening experiments consider 2 levels for each factor, 2^{m-k} plans
- Jones and Nachtsheim (2011) JQT (JN11), introduce screening experimental plans with three levels,

Definitive Screening Experiments (DSE)

- Usually screening experiments consider 2 levels for each factor, 2^{m-k} plans
- Jones and Nachtsheim (2011) JQT (JN11), introduce screening experimental plans with three levels,
- Main effects, quadratic effects, and interaction terms for $m \geq 6$

Definitive Screening Experiments (DSE)

- Usually screening experiments consider 2 levels for each factor, 2^{m-k} plans
- Jones and Nachtsheim (2011) JQT (JN11), introduce screening experimental plans with three levels,
- Main effects, quadratic effects, and interaction terms for $m \geq 6$
- If m number of factors, require $2m + 1$ experimental runs

Definitive Screening Experiments (DSE)

- Usually screening experiments consider 2 levels for each factor, 2^{m-k} plans
- Jones and Nachtsheim (2011) JQT (JN11), introduce screening experimental plans with three levels,
- Main effects, quadratic effects, and interaction terms for $m \geq 6$
- If m number of factors, require $2m+1$ experimental runs
- There are potentially $m + \frac{m(m-1)}{2} + m + 2$ parameters!

Definitive Screening Experiments (DSE)

- Usually screening experiments consider 2 levels for each factor, 2^{m-k} plans
- Jones and Nachtsheim (2011) JQT (JN11), introduce screening experimental plans with three levels,
- Main effects, quadratic effects, and interaction terms for $m \geq 6$
- If m number of factors, require $2m+1$ experimental runs
- There are potentially $m + \frac{m(m-1)}{2} + m + 2$ parameters!
- $m = 6$ implies 13 runs and 29 parameters. Effect sparsity and effect heredity assumptions are required for analysis

Experimental Plan for m=6

Run	$x_1(A)$	$x_2(B)$	$x_3(C)$	$x_4(D)$	$x_5(E)$	$x_6(F)$
1	0	1	-1	-1	-1	-1
2	0	-1	1	1	1	1
3	1	0	-1	1	1	-1
4	-1	0	1	-1	-1	1
5	-1	-1	0	1	-1	-1
6	1	1	0	-1	1	1
7	-1	1	1	0	1	-1
8	1	-1	-1	0	-1	1
9	1	-1	1	-1	0	-1
10	-1	1	-1	1	0	1
11	1	1	1	1	-1	0
12	-1	-1	-1	-1	1	0
13	0	0	0	0	0	0

Model

$$y_i = \beta_0 + \sum_{j=1}^{m} \beta_j x_{ij} + \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} \beta_{jk} x_{ij} x_{ik} + \sum_{j=1}^{m} \beta_{jj} x_{ij}^2 + \varepsilon_i$$

$$i = 1, ..., 2m+1$$

ε_i iid $N(0, \sigma^2)$

$\frac{(m+2)(m+1)}{2} + 1$ unknown parameters: β,σ^2

JN11 suggested for Analysis

- Forward step regression programmed in JMP

JN11 suggested for Analysis

- Forward step regression programmed in JMP
- Model search based on a bias corrected version of Akaike's Information Criteria

Forward Stepwise Regression in JMP

- Follows strategy by Hamada and Wu (1992)

Forward Stepwise Regression in JMP

- Follows strategy by Hamada and Wu (1992)
- Iterative procedure. JN11 used the "Combine" option with a p-value to enter of 0.1

Forward Stepwise Regression in JMP

- Follows strategy by Hamada and Wu (1992)
- Iterative procedure. JN11 used the "Combine" option with a p-value to enter of 0.1
- Our strategy may be viewed as a adaptation of Hamada and Wu's approach to Definitive Screening Experiments and using Bayesian tools

Corrected Akaike's Information Criteria AICc

- This procedure is based on a bias correction for Akaike's Information Criteria given in Hurvich and Tsai (1989)

Corrected Akaike's Information Criteria AICc

- This procedure is based on a bias correction for Akaike's Information Criteria given in Hurvich and Tsai (1989)
- Let $n = 2m + 1$ and $k(< n)$ be the number of parameters in β in a given model M

$$AIC_c = n \ln(\widehat{\sigma}^2) + n \frac{1 + m/n}{1 - (m+2)/n}$$

Corrected Akaike's Information Criteria AICc

- This procedure is based on a bias correction for Akaike's Information Criteria given in Hurvich and Tsai (1989)
- Let $n = 2m + 1$ and $k(< n)$ be the number of parameters in β in a given model M

$$AIC_c = n\ln(\widehat{\sigma}^2) + n\frac{1 + m/n}{1 - (m+2)/n}$$

- JN mention that with $n = 13$, 10 is the maximum number of predictors that can be included with AICc

Corrected Akaike's Information Criteria AICc

- This procedure is based on a bias correction for Akaike's Information Criteria given in Hurvich and Tsai (1989)
- Let $n = 2m + 1$ and $k (< n)$ be the number of parameters in β in a given model M

$$AIC_c = n \ln(\widehat{\sigma}^2) + n \frac{1 + m/n}{1 - (m+2)/n}$$

- JN mention that with $n = 13$, 10 is the maximum number of predictors that can be included with AICc
- JN11 fitted all possible models having 10 or fewer predictors, $16,628,808$ models

Conjugate Analysis of the Linear Model

- Box and Tiao (1992)

Conjugate Analysis of the Linear Model

- Box and Tiao (1992)
- Let $n = 2m + 1$ and $k(< n)$ be the number of parameters in β in a given model M

Conjugate Analysis of the Linear Model

- Box and Tiao (1992)
- Let $n = 2m + 1$ and $k(< n)$ be the number of parameters in β in a given model M
- Prior density of (β, σ^2) is $\pi(\beta, \sigma^2) \propto \sigma^{-2}$ (improper and non-informative)

Conjugate Analysis of the Linear Model

- Box and Tiao (1992)
- Let $n = 2m + 1$ and $k(< n)$ be the number of parameters in β in a given model M
- Prior density of (β, σ^2) is $\pi(\beta, \sigma^2) \propto \sigma^{-2}$ (improper and non-informative)
- The joint posterior distribution of β is multivariate t with $\nu = n - k$ degrees of freedom

Conjugate Analysis of the Linear Model

- Box and Tiao (1992)
- Let $n = 2m + 1$ and $k(< n)$ be the number of parameters in β in a given model M
- Prior density of (β,σ^2) is $\pi(\beta,\sigma^2) \propto \sigma^{-2}$ (improper and non-informative)
- The joint posterior distribution of β is multivariate t with $\nu = n - k$ degrees of freedom
- Then the posterior distribution of each β_j is

$$\beta_j|\mathbf{y}, M \sim t(\widehat{\beta}_j, s^2 c_{jj}, \nu)$$

Bayesian Evaluation of Significance

- Posterior probability intervals $P(a < \beta_j < b|\mathbf{y}, M) = 1 - \alpha$

Bayesian Evaluation of Significance

- Posterior probability intervals $P(a < \beta_j < b|\mathbf{y}, M) = 1 - \alpha$
- Odds that the parameter is positive or negative using $P(\beta_j > 0|\mathbf{y}, M)$ or $P(\beta_j < 0|\mathbf{y}, M)$

Bayesian Evaluation of Significance

- Posterior probability intervals $P(a < \beta_j < b|\mathbf{y}, M) = 1 - \alpha$
- Odds that the parameter is positive or negative using $P(\beta_j > 0|\mathbf{y}, M)$ or $P(\beta_j < 0|\mathbf{y}, M)$
- Note: Care has to be taken so that $k < n$, up to $k = n - 1$

Some Words About Odds

1. A an event: odds in favor $\frac{P(A)}{1-P(A)}$.; odds against $\frac{1-P(A)}{P(A)}$

Some Words About Odds

1. A an event: odds in favor $\frac{P(A)}{1-P(A)}$.; odds against $\frac{1-P(A)}{P(A)}$
2. Interpretation: odds in favor greater than one, then it is the number of times that A is more likely to happen than its complement.

Some Words About Odds

1. A an event: odds in favor $\frac{P(A)}{1-P(A)}$.; odds against $\frac{1-P(A)}{P(A)}$
2. Interpretation: odds in favor greater than one, then it is the number of times that A is more likely to happen than its complement.
3. Examples: 2-1 implies $P(A) = 2/3$; 20-1 implies $P(A) = 0.9523$

Some Words About Odds

1. A an event: odds in favor $\frac{P(A)}{1-P(A)}$.; odds against $\frac{1-P(A)}{P(A)}$
2. Interpretation: odds in favor greater than one, then it is the number of times that A is more likely to happen than its complement.
3. Examples: 2-1 implies $P(A) = 2/3$; 20-1 implies $P(A) = 0.9523$
4. Roulette: to win betting "Premiere douzaine " in American roulette odds against are 2-1

Some Words About Odds

1. A an event: odds in favor $\frac{P(A)}{1-P(A)}$.; odds against $\frac{1-P(A)}{P(A)}$
2. Interpretation: odds in favor greater than one, then it is the number of times that A is more likely to happen than its complement.
3. Examples: 2-1 implies $P(A) = 2/3$; 20-1 implies $P(A) = 0.9523$
4. Roulette: to win betting "Premiere douzaine " in American roulette odds against are 2-1
5. If you have odds in favor 10-1 or 20-1 of winning in a game, would you bet on it?

Bayesian Strategy

- **Step 1**. Fit a model with all main effects. Choose those that show some potential of significance. Be generous.

Bayesian Strategy

- **Step 1**. Fit a model with all main effects. Choose those that show some potential of significance. Be generous.
- **Step 2**. Using the effects chosen in Step 1, fit a model with main, quadratic and interaction effects. Choose those that show potential of significance. Be strict.

Bayesian Strategy

- **Step 1**. Fit a model with all main effects. Choose those that show some potential of significance. Be generous.
- **Step 2**. Using the effects chosen in Step 1, fit a model with main, quadratic and interaction effects. Choose those that show potential of significance. Be strict.
- **Step 3**. Fit a model with **all main effects** again, and quadratic and interaction effects that showed significance in Step 2. Choose those that show potential of significance. Be strict.

Bayesian Strategy

- **Step 4**. Using the effects chosen in Step 3, fit a model with main, quadratic and interaction effects that have shown significance. If any new main effect should be included add it with any quadratic or interaction effects that have not been tested before. Choose those that show potential of significance. Be strict.

Bayesian Strategy

- **Step 4**. Using the effects chosen in Step 3, fit a model with main, quadratic and interaction effects that have shown significance. If any new main effect should be included add it with any quadratic or interaction effects that have not been tested before. Choose those that show potential of significance. Be strict.
- **Final Step**. Fit a model with effects that have shown evidence of being significance.

Simulated Example

$$m = 6$$

$$y_j = 20 + 4A + 3B - 2C - D + 5BC + 6AA + \varepsilon_i$$

$$\varepsilon_i \sim N(0, 1)$$

Stepwise Procedure in JMP

Stepwise Results

Stepwise Fit for Y

Current Estimates

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Intercept	20.0576471	1	0	0.000	1
☐	☑	X1	3.408	2	192.6436	407.638	2.9e-6
☐	☑	X2	2.748	2	200.7311	424.752	2.62e-6
☐	☑	X3	-1.309	2	142.3509	301.218	6.15e-6
☐	☑	X4	-0.851	2	9.407917	19.907	0.00416
☐	☐	X5	0	1	0.26896	1.179	0.33862
☐	☐	X6	0	1	0.43264	2.311	0.20309
☐	☐	X1*X2	0	1	0.152025	0.591	0.48501
☐	☐	X1*X3	0	1	0.131951	0.503	0.51737
☐	☐	X1*X4	0	1	0.156025	0.609	0.47889
☐	☐	X1*X5	0	2	0.435212	0.875	0.50199
☐	☐	X1*X6	0	2	0.779243	2.906	0.19864
☐	☑	X2*X3	5.595	1	125.2161	529.920	2.88e-6
☐	☐	X2*X4	0	1	0.461888	2.568	0.18434
☐	☐	X2*X5	0	2	0.70216	2.197	0.25839
☐	☐	X2*X6	0	2	0.43281	0.867	0.50442
☐	☐	X3*X4	0	1	0.01918	0.066	0.80991
☐	☐	X3*X5	0	2	0.417108	0.819	0.52037
☐	☐	X3*X6	0	2	0.473473	1.003	0.46388
☐	☐	X4*X5	0	2	0.279325	0.464	0.66724
☐	☐	X4*X6	0	2	0.598892	1.542	0.34625
☐	☐	X5*X6	0	3	0.857625	1.766	0.38153
☐	☑	X1*X1	-7.2714706	1	76.49896	323.747	0.00001
☐	☐	X2*X2	0	1	0.01918	0.066	0.80991
☐	☐	X3*X3	0	1	0.152025	0.591	0.48501
☐	☑	X4*X4	1.22352941	1	2.165907	9.166	0.02916
☐	☐	X5*X5	0	2	0.520167	1.180	0.41876
☐	☐	X6*X6	0	2	0.683847	2.061	0.27334

JMP finds as active: A, B, C, D, BC, AA, DD. Type I error to include DD. Notice: AA reported in JMP is -7.27 (vs reported true=6)

Minimization of Hurvich and Tsai's AICc

$$y_j = 20 + 4A + 3B - 2C - D + 5BC + 6AA + \varepsilon_i$$

1. Minimizing AIC_c gives model: A, B, C, BC, AA with $AIC_c = 70.63$. Type II error not to include D.

Minimization of Hurvich and Tsai's AICc

$$y_j = 20 + 4A + 3B - 2C - D + 5BC + 6AA + \varepsilon_i$$

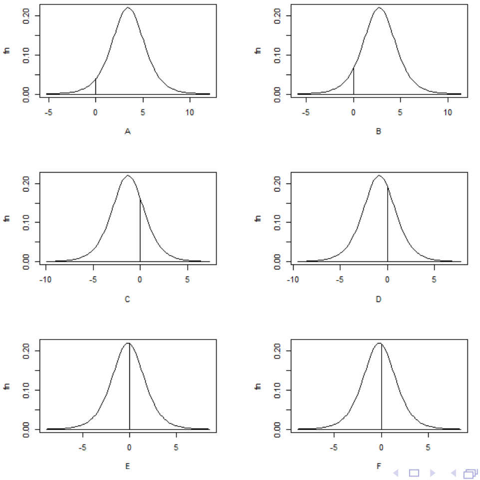
1. Minimizing AIC_c gives model: A, B, C, BC, AA with $AIC_c = 70.63$. Type II error not to include D.
2. $AIC_c = 83.72$ for model found by JMP: A, B, C, D, BC, AA, DD. Type I error to include DD.

Minimization of Hurvich and Tsai's AICc

$$y_j = 20 + 4A + 3B - 2C - D + 5BC + 6AA + \varepsilon_i$$

1. Minimizing AIC_c gives model: A, B, C, BC, AA with $AIC_c = 70.63$. Type II error not to include D.
2. $AIC_c = 83.72$ for model found by JMP: A, B, C, D, BC, AA, DD. Type I error to include DD.
3. $AIC_c = 71.25$ for true model A, B, C, D, BC, AA

Step 1. Posterior Distributions

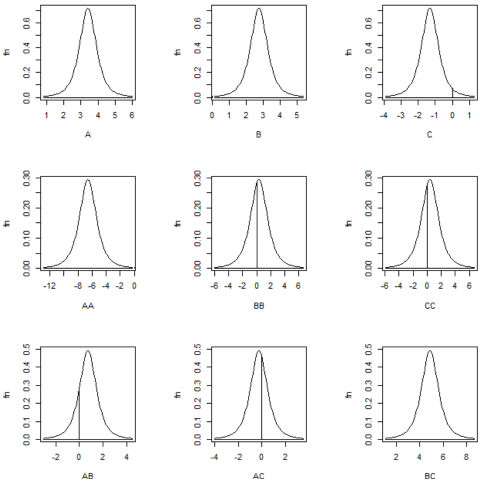


Step 1. Posterior Probability Intervals and Odds

Effects	Estim	Q025	Q975	P+	P-	Odds +	Odds -
A	3.408	-0.839	7.655	0.951	0.049	19.6	0.1
B	2.748	-1.499	6.995	0.918	0.082	11.2	0.1
C	-1.309	-5.556	2.938	0.24	0.76	0.3	3.2
D	-0.851	-5.098	3.396	0.321	0.679	0.5	2.1
E	-0.164	-4.411	4.083	0.464	0.536	0.9	1.2
F	-0.208	-4.455	4.039	0.454	0.546	0.8	1.2

Potentially Significant A, B, and C

Step 2. Posterior Distributions

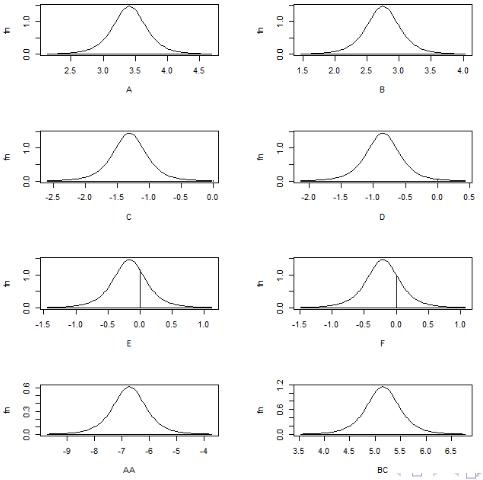


Step 2. Posterior Probability Intervals and Odds

Effects	Estim	Q025	Q975	P+	P-	Odds +	Odds -
A	3.408	1.77	5.046	0.996	0.004	284.3	0.0
B	2.748	1.11	4.386	0.994	0.006	154.8	0.0
C	-1.309	-2.947	0.329	0.042	0.958	0.0	22.7
AA	-6.579	-10.557	-2.602	0.007	0.993	0.0	148.7
BB	0.323	-3.655	4.301	0.594	0.406	1.5	0.7
CC	0.4	-3.577	4.378	0.615	0.385	1.6	0.6
AB	0.722	-1.652	3.095	0.798	0.202	3.9	0.3
AC	-0.249	-2.622	2.124	0.38	0.62	0.6	1.6
BC	4.874	2.5	7.247	0.996	0.004	273.7	0.0

Potentially Significant A, B, C, AA, BC

Step 3. Posterior Distributions

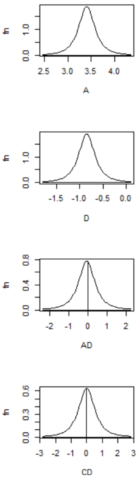


Step 3. Posterior Probability Intervals and Odds

Effects	Estim	Q025	Q975	P+	P-	Odds +	Odds -
A	3.408	2.694	4.122	1	0	10670.0	0.0
B	2.748	2.034	3.462	1	0	4600.6	0.0
C	-1.309	-2.023	-0.595	0.004	0.996	0.0	283.4
D	-0.851	-1.565	-0.137	0.015	0.985	0.0	66.4
E	-0.164	-0.878	0.55	0.279	0.721	0.4	2.6
F	-0.208	-0.922	0.506	0.232	0.768	0.3	3.3
AA	-6.725	-8.404	-5.045	0	1	0.0	5368.4
BC	5.152	4.25	6.054	1	0	21648.4	0.0

Potentially Significant A, B, C, D, AA, BC

Step 4. Posterior Distributions

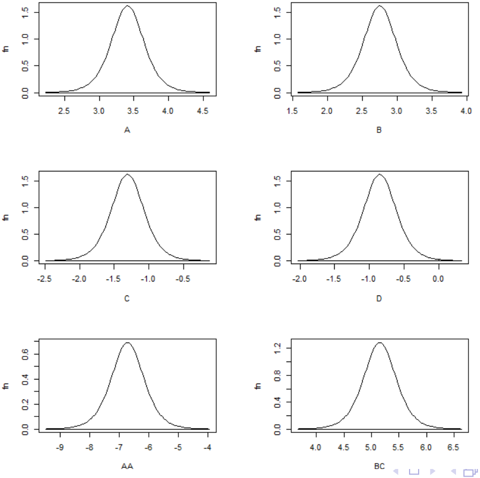


Step 4. Posterior Probability Intervals and Odds

Effects	Estim	Q025	Q975	P+	P-	Odds +	Odds -
A	3.408	2.602	4.214	0.998	0.002	664.2	0.0
B	2.748	1.942	3.554	0.998	0.002	432.5	0.0
C	-1.309	-2.115	-0.503	0.01	0.99	0.0	99.7
D	-0.851	-1.657	-0.045	0.023	0.977	0.0	43.3
AA	-7.12	-13.982	-0.258	0.023	0.977	0.0	41.8
DD	1.264	-3.578	6.106	0.811	0.189	4.3	0.2
AD	-0.072	-2.029	1.885	0.444	0.556	0.8	1.3
BC	5.665	1.439	9.891	0.986	0.014	68.5	0.0
BD	-0.249	-1.417	0.919	0.228	0.772	0.3	3.4
CD	0.002	-2.402	2.406	0.501	0.499	1.0	1.0

Potentially Significant A, B, C, D, AA, BC

Final Step. Posterior Distributions



Final Step. Posterior Probability Intervals and Odds

$$y_j = 20 + 4A + 3B - 2C - D + 5BC + 6AA + \varepsilon_i$$

Effects	Estim	Q005	Q995	P+	P-	Odds +	Odds -
A	3.408	2.53	4.28	1	0	288050.3	0.0
B	2.748	1.87	3.62	1	0	82345.0	0.0
C	-1.309	-2.19	-0.43	0.001	0.999	0.0	1371.8
D	-0.851	-1.73	0.03	0.006	0.994	0.0	175.6
AA	-6.725	-8.78	-4.67	0	1	0.0	103581.7
BC	5.152	4.05	6.26	1	0	827895.2	0.0

Final Remarks

- The approach is not a blind search, it makes use of a strategy, figures, posterior distributions and posterior odds

Final Remarks

- The approach is not a blind search, it makes use of a strategy, figures, posterior distributions and posterior odds
- Requires fitting a very limited set of models

Final Remarks

- The approach is not a blind search, it makes use of a strategy, figures, posterior distributions and posterior odds
- Requires fitting a very limited set of models
- If possible, it should be used in conjunction with the other methods

Final Remarks

- The approach is not a blind search, it makes use of a strategy, figures, posterior distributions and posterior odds
- Requires fitting a very limited set of models
- If possible, it should be used in conjunction with the other methods
- It could be adapted to other screening designs

Final Remarks

- The approach is not a blind search, it makes use of a strategy, figures, posterior distributions and posterior odds
- Requires fitting a very limited set of models
- If possible, it should be used in conjunction with the other methods
- It could be adapted to other screening designs
- It is programmed in R

Things that may be done

- To obtain posterior probability intervals for a mean response: $P(a < \mathbf{x}_0^T \boldsymbol{\beta} < b|\mathbf{y}, M) = 1 - \alpha$

Things that may be done

- To obtain posterior probability intervals for a mean response: $P(a < \mathbf{x}_0^T \boldsymbol{\beta} < b | \mathbf{y}, M) = 1 - \alpha$
- To get posterior prediction intervals $P(a < y_0' < b | \mathbf{y}, M, \mathbf{x}_0) = 1 - \alpha$

Things that may be done

- To obtain posterior probability intervals for a mean response: $P(a < \mathbf{x}_0^T \boldsymbol{\beta} < b|\mathbf{y}, M) = 1 - \alpha$
- To get posterior prediction intervals $P(a < y_0' < b|\mathbf{y}, M, \mathbf{x}_0) = 1 - \alpha$
- Compare the procedures by simulation

Things that may be done

- To obtain posterior probability intervals for a mean response: $P(a < \mathbf{x}_0^T \boldsymbol{\beta} < b|\mathbf{y}, M) = 1 - \alpha$
- To get posterior prediction intervals $P(a < y_0' < b|\mathbf{y}, M, \mathbf{x}_0) = 1 - \alpha$
- Compare the procedures by simulation
- It may be applied to Definitive Screening Experiments when the response is non-normal (GLM). MCMC

Thank you very much!

aguirre@itam.mx

References

- Hamada, M. S. and Wu, C. F. J. (1992). “Analyzing Designed Experiments with Complex Aliasing”. *Journal of Quality Technology* 24, pp. 30–37.
- Hurvich, C. M. and Tsai, C.-L. (1989). “Regression and Time Series Model Selection in Small Samples”. *Biometrika* 76, pp. 297–330.
- Jones, B. and Nachtsheim, C. J. (2011). "A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects" *Journal of Quality Technology* 43, pp. 1-14.