Tailoring DOE Constraints to the Problem

ENBIS Annual Conference
June 28th, 2022

Volker Kraft, Chris Gotwalt - JMP

Tailoring DOE Constraints to the Problem

Abstract

There are often constraints among the factors in experiments that are important not to violate but are difficult to describe in mathematical form. In this presentation, we illustrate a simple workflow of creating a simulated dataset of candidate factor values. From there, we identify a physically realisable set of potential factor combinations that is supplied to the new Candidate Set Design capability in JMP 16.

This then identifies the optimal subset of these filtered factor settings to run in the experiment. We also illustrate the Candidate Set Designer's use on historical process data, achieving designs that maximize information content while respecting the internal correlation structure of the process variables.

Our approach is simple and easy to teach. It makes setting up experiments with constraints much more accessible to practitioners with any amount of DOE experience.

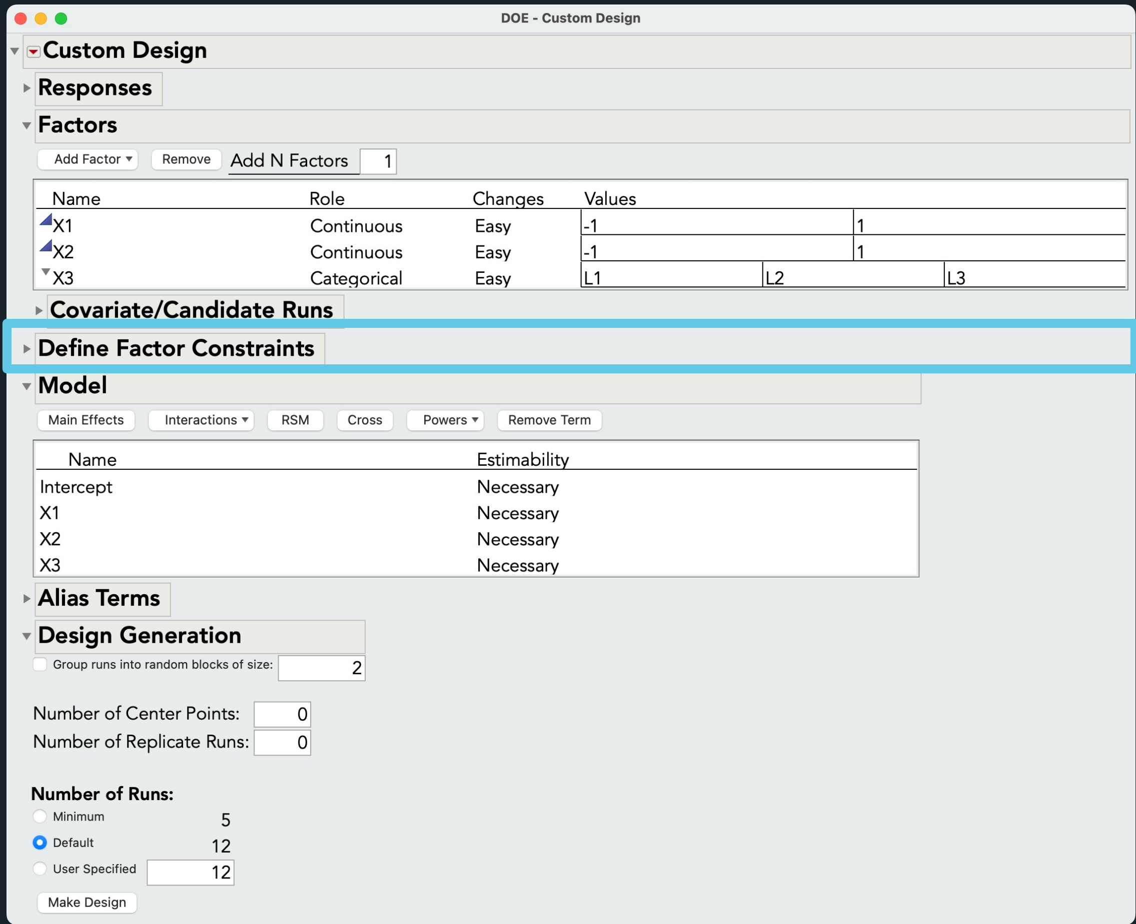
Starting point

Custom Design > Factor Constraints

- Why an optimal design?
  – We want to tailor the design to the problem at hand!
- JMP Custom Designer workflow:
  – Define response(s) and factors
  – Define the model (incl. main effects, interactions and polynomial terms)
  – Define the budget = number of runs
- If factors can be changed independently, we are all set.
  If not, we have to
  – Define the constraints

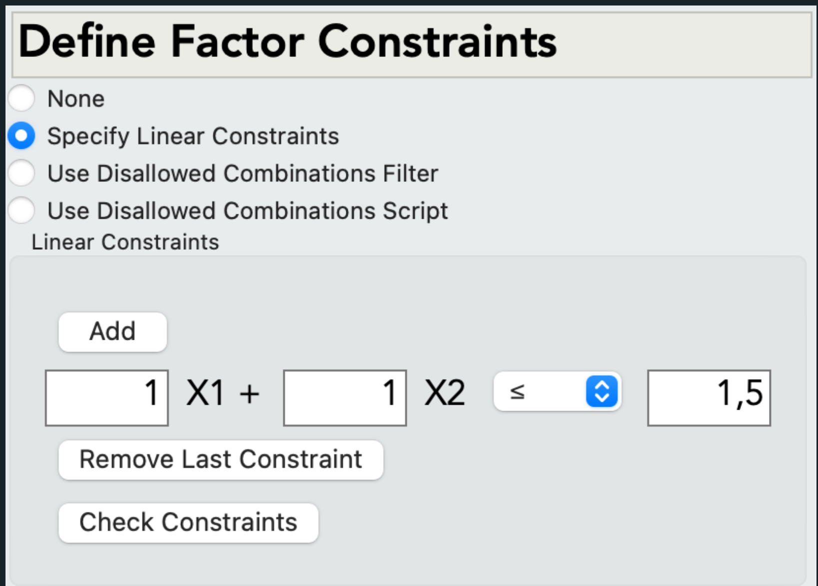
Starting point

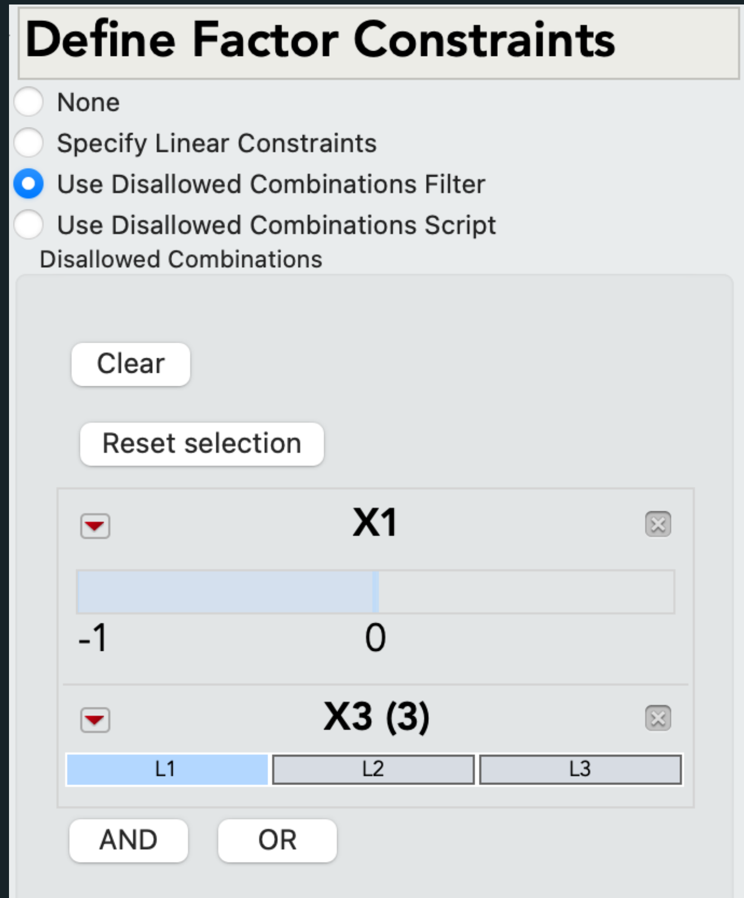
Custom Design > Factor Constraints



Starting point

Custom Design > Factor Constraints





More tailored constraints

Custom Design > Covariate/Candidate Runs

Covariate/Candidate Runs

Select Covariate Factors — Load a set of candidate runs for covariates from the current data table.

- In a nutshell, think about covariates as:
  - Input variable(s) we want to account for, but don't have complete control.
  - Uncontrolled, but observable ahead of time.

Runners Covariate...

3/0 Cols
100/0

	Miles	Weight	Strike Point
1	10,4	143	Midfoot
2	12,4	140	Midfoot
3	10,4	147	Heel
4	7,6	150	Forefoot
5	8,4	145	Heel
6	11,2	145	Midfoot
7	14,0	139	Midfoot
8	13,2	145	Midfoot
9	9,2	146	Midfoot
10	12,0	142	Heel
11	7,6	145	Midfoot
12	9,2	152	Midfoot
13	13,2	143	Midfoot

These covariates were measured on runners who are candidates for an experiment on running shoe wear. (JMP Sample Data)

jmp STATISTICAL DISCOVERY

What’s next?

Custom Design > Covariate/Candidate Runs

- Look at examples defining constraints by candidate runs
  - first one as live demo
- Mention other use cases of covariate factors
- Share some technical insight into optimal design using covariates
  - JMP and JMP Pro, version 16+

Use cases of candidate set designs

Custom Design > Covariate/Candidate Runs

- The range of a continuous factor depends on the levels of a categorical factor
- Physical constraints on factors that cannot vary completely independently (i.e. temperature and pressure)
  - Historical data design
  - Filtered design
- Mixture problems
  - Especially sub-mixture problems
- Nonlinear constraints via the filtering method
  - Formula column based filtered designs

$$X1^{2} + 4 \cdot X3 + X4 \cdot X5 \geq 100$$

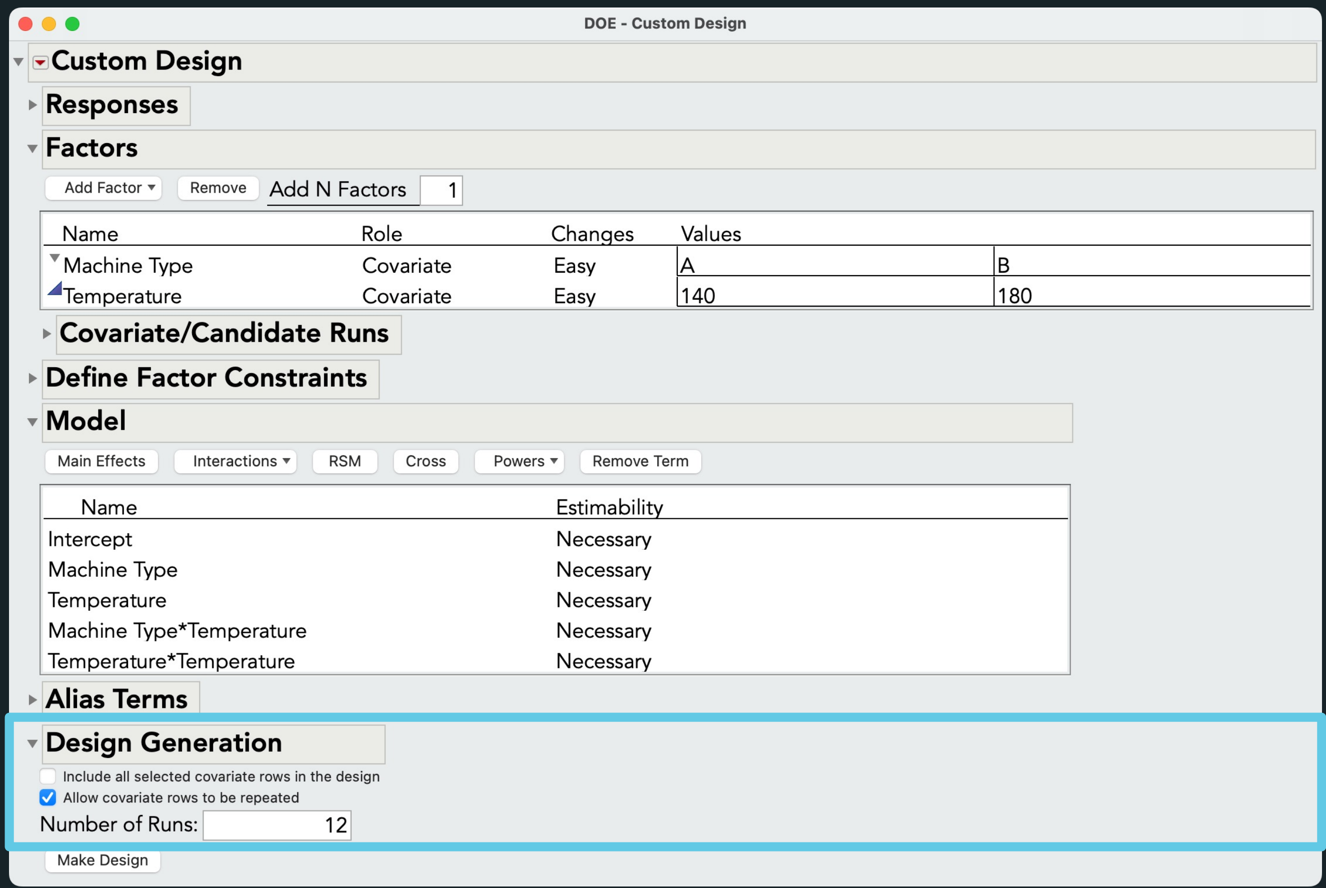
#1: Suppliers with different operator regions

The problem

- 12-run DOE with two factors:
  - Machine Supplier (A and B)
  - Temperature
- Supplier A:
  - Temperature controls from 150C to 170C controllable to 5C
- Supplier B:
  - Temperature controls from 140C to 180C controllable to 10C

#1: Suppliers with different operator regions

Custom Design



#1: Suppliers with different operator regions

Chosen candidate runs selected

Machine and Temperature Candidate Set.jmp

Machine and Temperature Candidate Set.jmp
DOE leads to missing factors

Columns (2/0)
Machine Type
Temperature

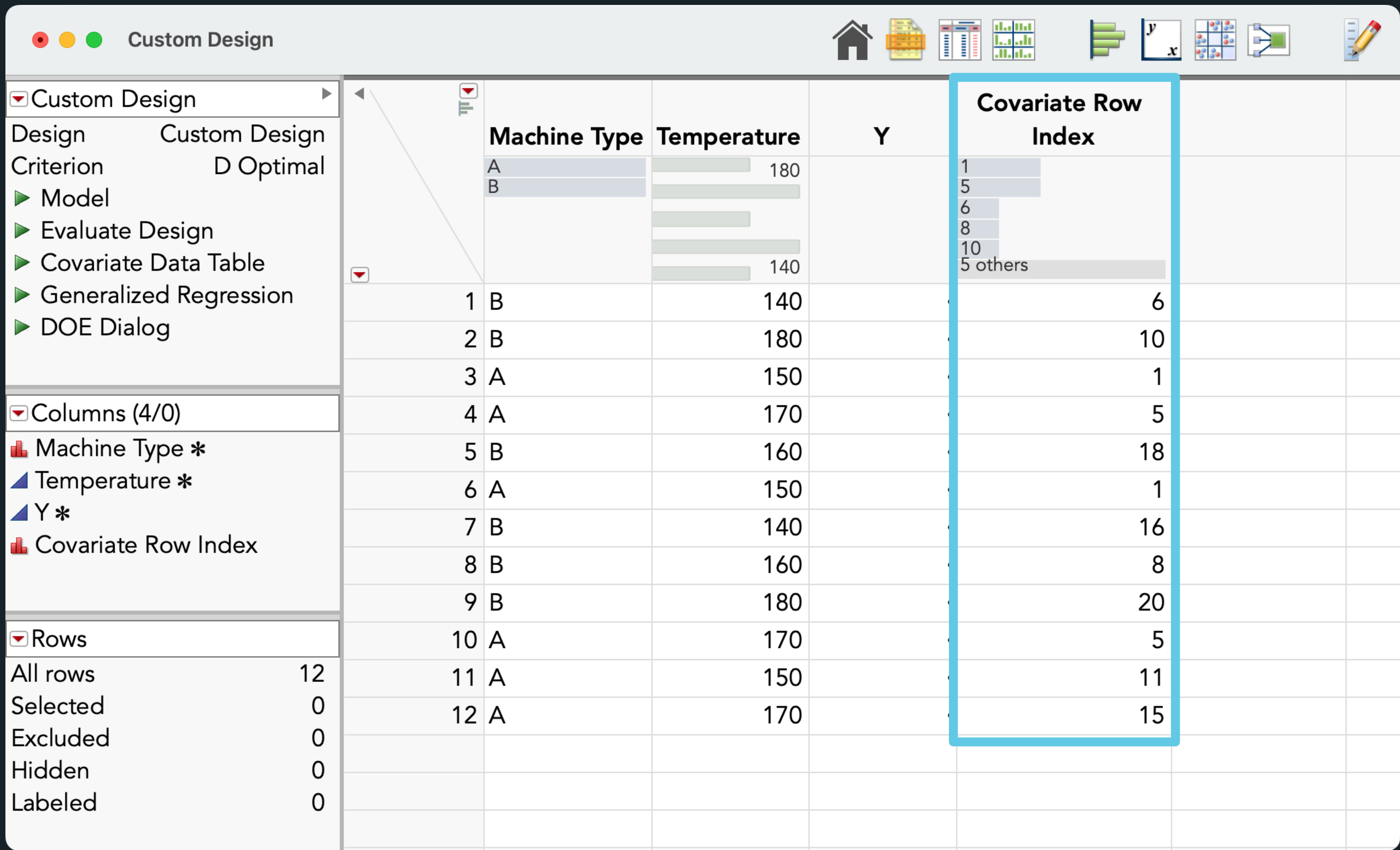
Rows	
All rows	20
Selected	10
Excluded	0
Hidden	0
Labeled	0

	Machine Type	Temperature
1	A	150
2	A	155
3	A	160
4	A	165
5	A	170
6	B	140
7	B	150
8	B	160
9	B	170
10	B	180
11	A	150
12	A	155
13	A	160
14	A	165
15	A	170
16	B	140
17	B	150
18	B	160
19	B	170
20	B	180

jmp STATISTICAL DISCOVERY

#1: Suppliers with different operator regions

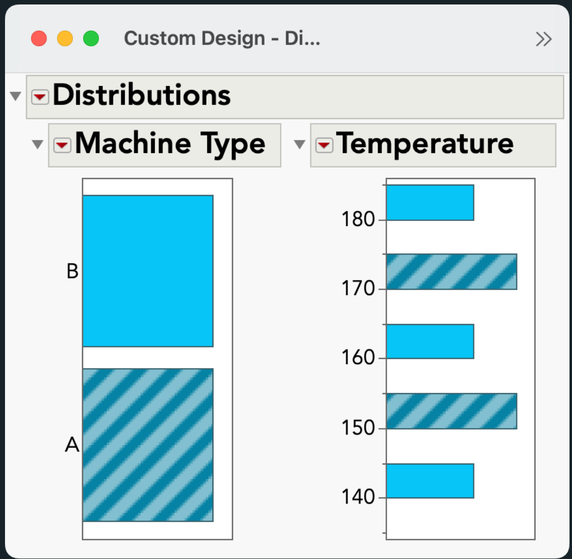
Custom Design table



	Machine Type	Temperature	Y	Covariate Row Index
1	B	140		6
2	B	180		10
3	A	150		1
4	A	170		5
5	B	160		18
6	A	150		1
7	B	140		16
8	B	160		8
9	B	180		20
10	A	170		5
11	A	150		11
12	A	170		15

#1: Suppliers with different operator regions

Forcing a missing value into the design (Machine A at 160°C)



Covariate/Candidate Runs

Select Covariate Factors — Load a set of candidate runs for covariates from the current data table.

Machine Type	Temperature
A	150
A	155
A	160
A	165
A	170
B	140
B	150
B	160
B	170
B	180
A	150
A	155
A	160
A	165
A	170
B	140
B	150
B	160
B	170
B	180

Define Factor Constraints

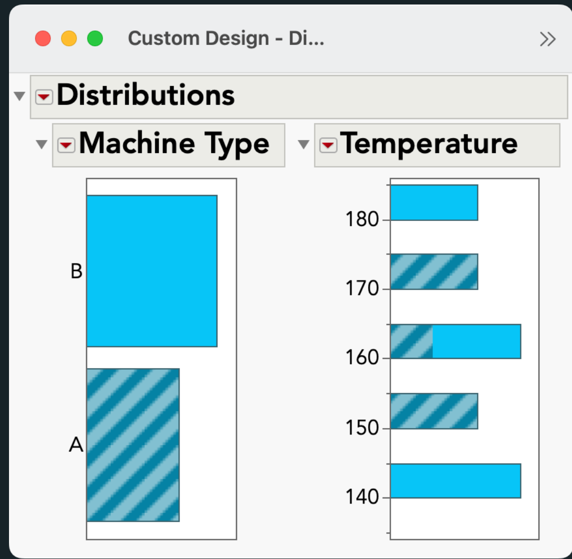
Model

Alias Terms

Design Generation

- [x] Include all selected covariate rows in the design
- [x] Allow covariate rows to be repeated

Number of Runs: 12



#2: Using historical process data

Powder Metallurgy Historical Data.jmp

Powder Metallurgy Historical Data.jmp
Source: Philip J. Rams… Used with Permission
Notes: A company manufactures steel drive sh

Columns (18/0)
- Shrinkage
- Formation Method *
- Compaction Method
- Compaction Pressure
- Sintering Time
- Sintering Temp
- Sintering Method
- Powder Source
- Powder Type
- Ratio
- Unit
- Dwell
- Particle Size
- P1
- P2
- P3
- P4
- P5

Rows

All rows	6.253
Selected	0
Excluded	0
Hidden	0
Labeled	0

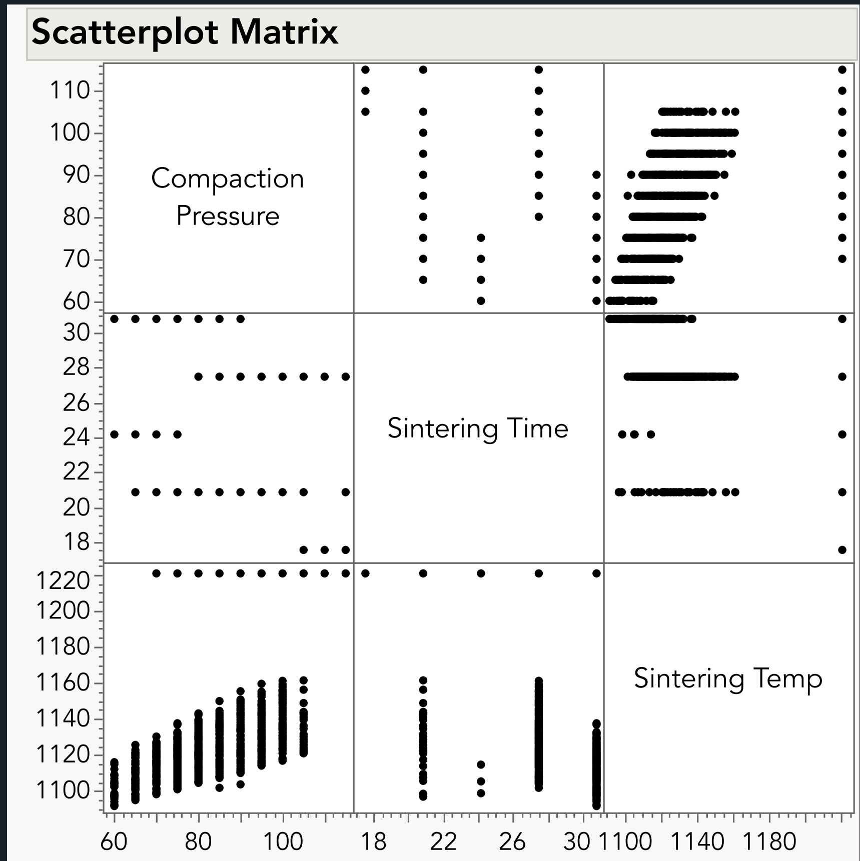
9/0 Cols

	Shrinkage	Formation Method	Compaction Method	Compaction Pressure	Sintering Time	Sintering Temp	Sintering Method	Powder Source	Powder Type
	0,76 / -2,2	Die Compaction, Isostatic Press, Extrusion	Cold, Warm	115 / 60	30,8 / 17,5	1221 / 1092	Atmosphere, Vacuum	D, N	3, 1, 2, 5, 4
1	0,19	Die Compaction	Cold	90	27,48	1133	Vacuum	N	1
2	-0,37	Extrusion	Cold	90	27,48	1143	Atmosphere	D	3
3	-0,16	Extrusion	Cold	100	27,48	1121	Atmosphere	D	3
4	0,41	Extrusion	Cold	95	27,48	1122	Atmosphere	D	3
5	-0,65	Extrusion	Cold	85	27,48	1118	Atmosphere	D	4
6	0,41	Extrusion	Cold	95	27,48	1122	Atmosphere	D	4
7	-0,23	Extrusion	Cold	100	27,48	1137	Atmosphere	D	3
8	0,12	Extrusion	Cold	100	27,48	1126	Atmosphere	D	3
9	0,41	Extrusion	Cold	105	20,85	1149	Atmosphere	D	4
10	-0,72	Extrusion	Cold	85	27,48	1126	Atmosphere	D	3
11	-1,08	Extrusion	Cold	85	27,48	1117	Atmosphere	D	5
12	-0,65	Extrusion	Cold	80	27,48	1124	Atmosphere	D	3
13	-0,79	Extrusion	Cold	85	27,48	1140	Atmosphere	D	4
14	-0,16	Extrusion	Cold	100	27,48	1129	Atmosphere	D	5
15	-0,65	Extrusion	Cold	85	27,48	1118	Atmosphere	D	2
16	0,12	Extrusion	Cold	100	27,48	1126	Atmosphere	D	2
17	-0,72	Extrusion	Cold	85	27,48	1126	Atmosphere	D	3
18	-0,44	Extrusion	Cold	85	27,48	1135	Atmosphere	D	5
19	-0,65	Extrusion	Cold	80	27,48	1124	Atmosphere	D	3
20	-1,15	Extrusion	Cold	80	27,48	1120	Atmosphere	D	5

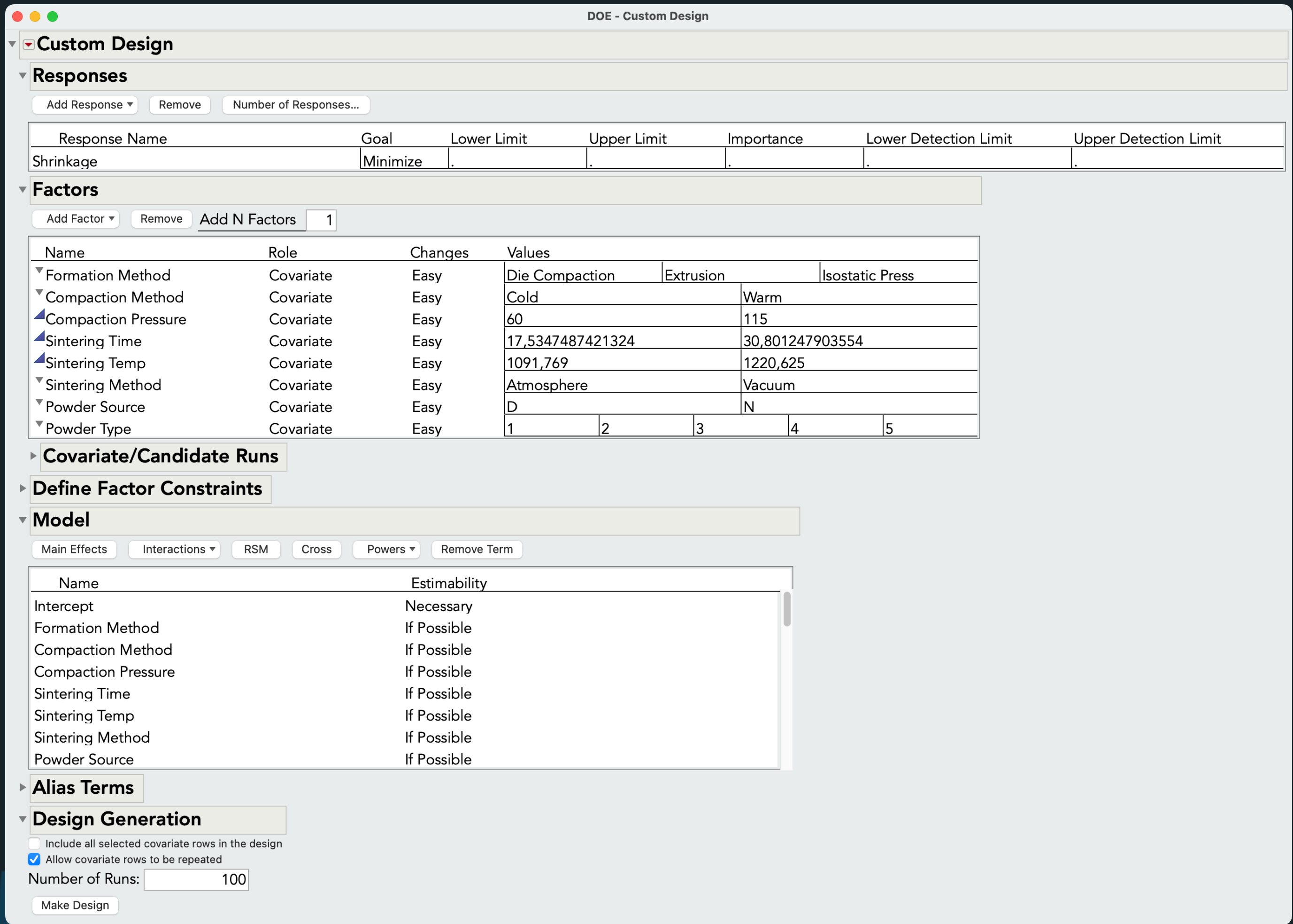
jmp STATISTICAL DISCOVERY

#2: Using historical process data

Implicit constraints suggest DOE



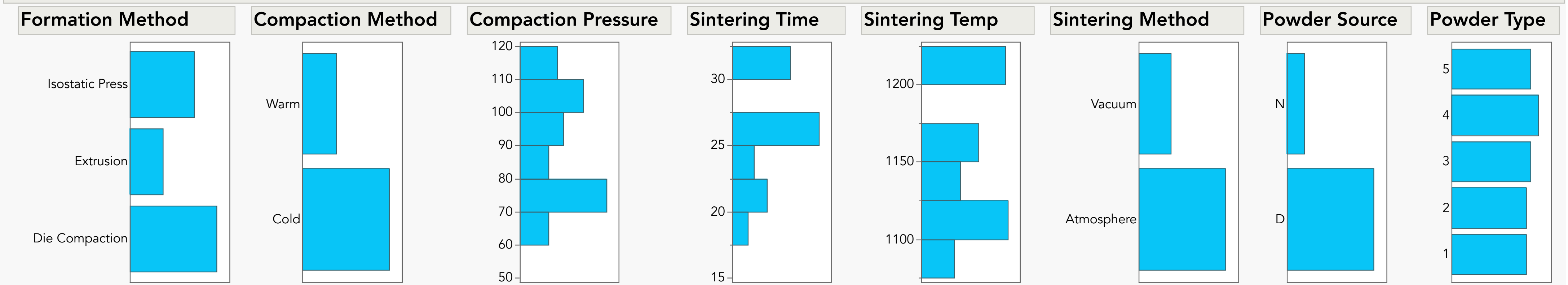
#2: Using historical process data



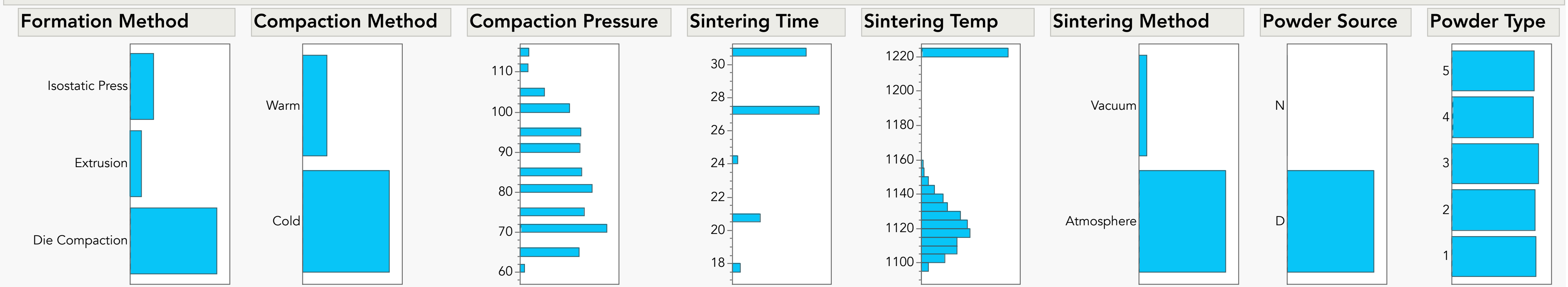
- Set response
- Load (controllable) covariate factors
- Choose RSM model, estimate all terms except intercept "If possible" (Baysian I-optimal candidate set design)
- Allow optimally chosen replication
- Choose # of runs to 100
- Make design

#2: Using historical process data

Distributions (Candidate set design)



Distributions (all data)



#2: Using historical process data

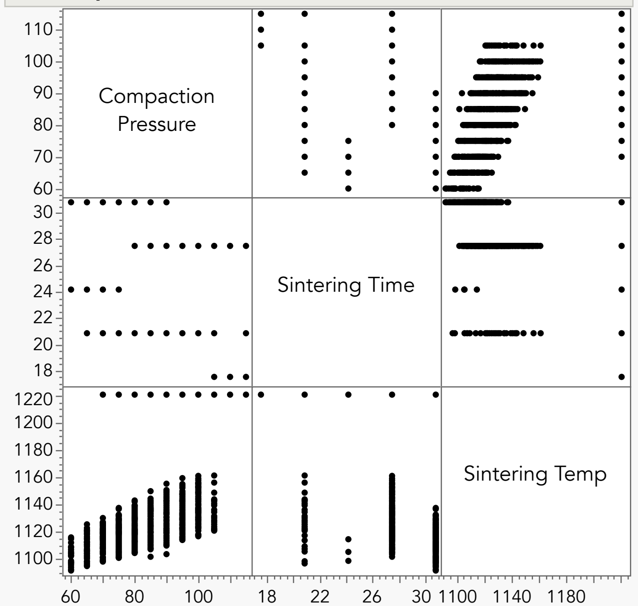
Multivariate (all data)

Correlations

	Compaction Pressure	Sintering Time	Sintering Temp
Compaction Pressure	1,0000	-0,6489	0,4945
Sintering Time	-0,6489	1,0000	-0,7343
Sintering Temp	0,4945	-0,7343	1,0000

The correlations are estimated by Row-wise method.

Scatterplot Matrix



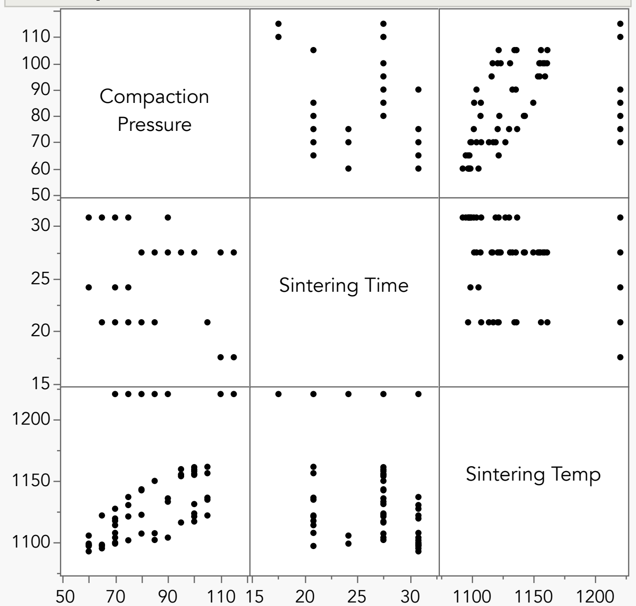
Multivariate (Candidate set design)

Correlations

	Compaction Pressure	Sintering Time	Sintering Temp
Compaction Pressure	1,0000	-0,4367	0,4464
Sintering Time	-0,4367	1,0000	-0,4100
Sintering Temp	0,4464	-0,4100	1,0000

The correlations are estimated by Row-wise method.

Scatterplot Matrix



JMP STATISTICAL DISCOVERY

#3: Machine suppliers with temperature and pressure

Space Filling Design

Responses

Add Response ▾ | Remove | Number of Responses...

Response Name	Goal	Lower Limit	Upper Limit	Importance	Lower Detection Limit	Upper Detection Limit
Y	Maximize	.	.	.	.	.

Factors

Name	Role	Values	
Temperature	Continuous	1080	1160
Pressure	Continuous	60	120
Machine	Categorical	A	B

Define Factor Constraints

3 Factors

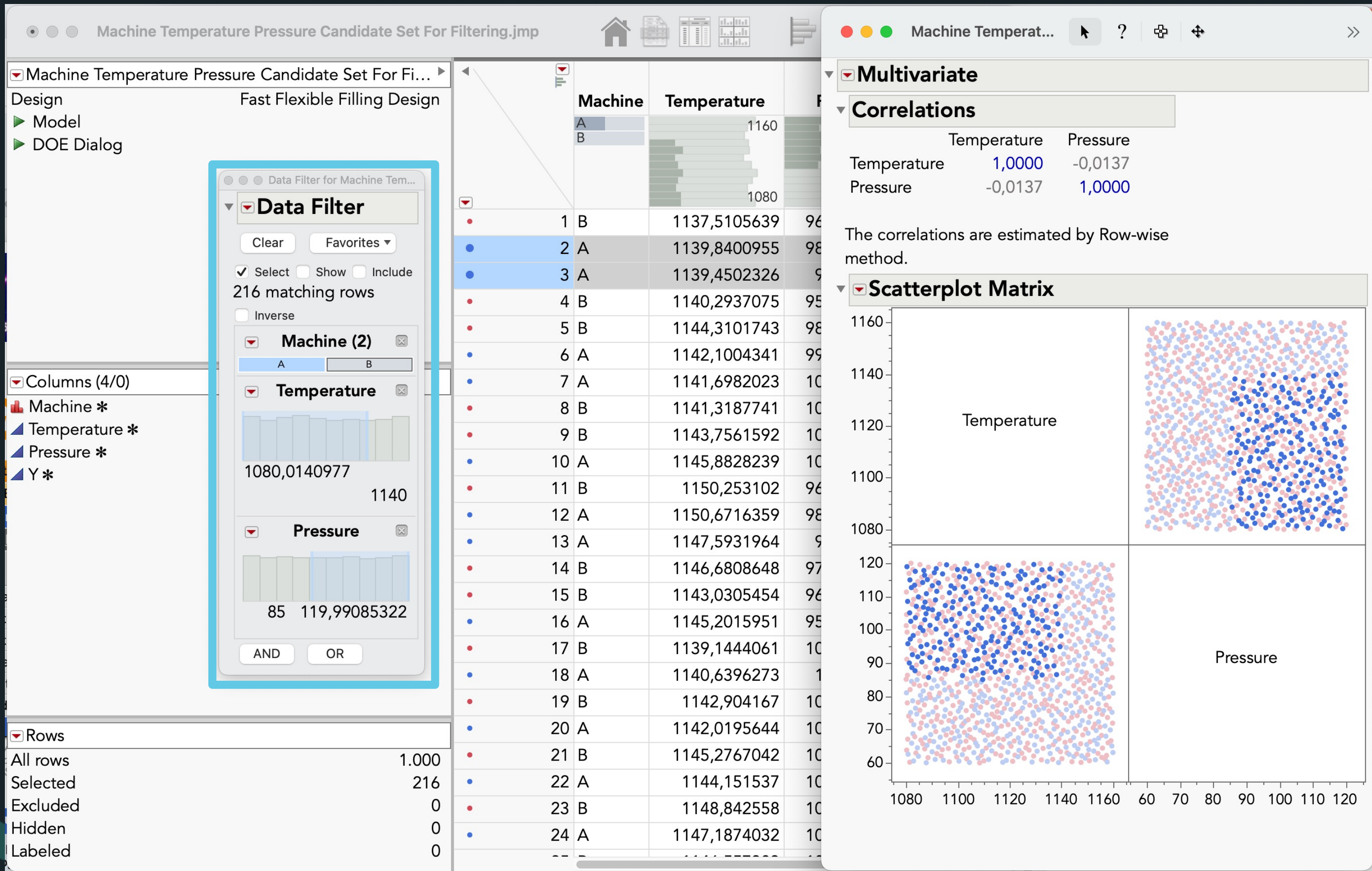
Space Filling Design Methods

Number of Runs: 1000

Sphere Packing	Optimal Packing of Spheres Inside of a Cube.
Latin Hypercube	Latin HyperCube with Optimal Spacing
Uniform	Uniform Design
Minimum Potential	Minimum energy designs in a spherical region.
Maximum Entropy	Maximum entropy designs for a Gaussian process
Gaussian Process IMSE Optimal	Integrated mean square error optimal designs for a Gaussian process
Fast Flexible Filling	Space filling design through clustering.

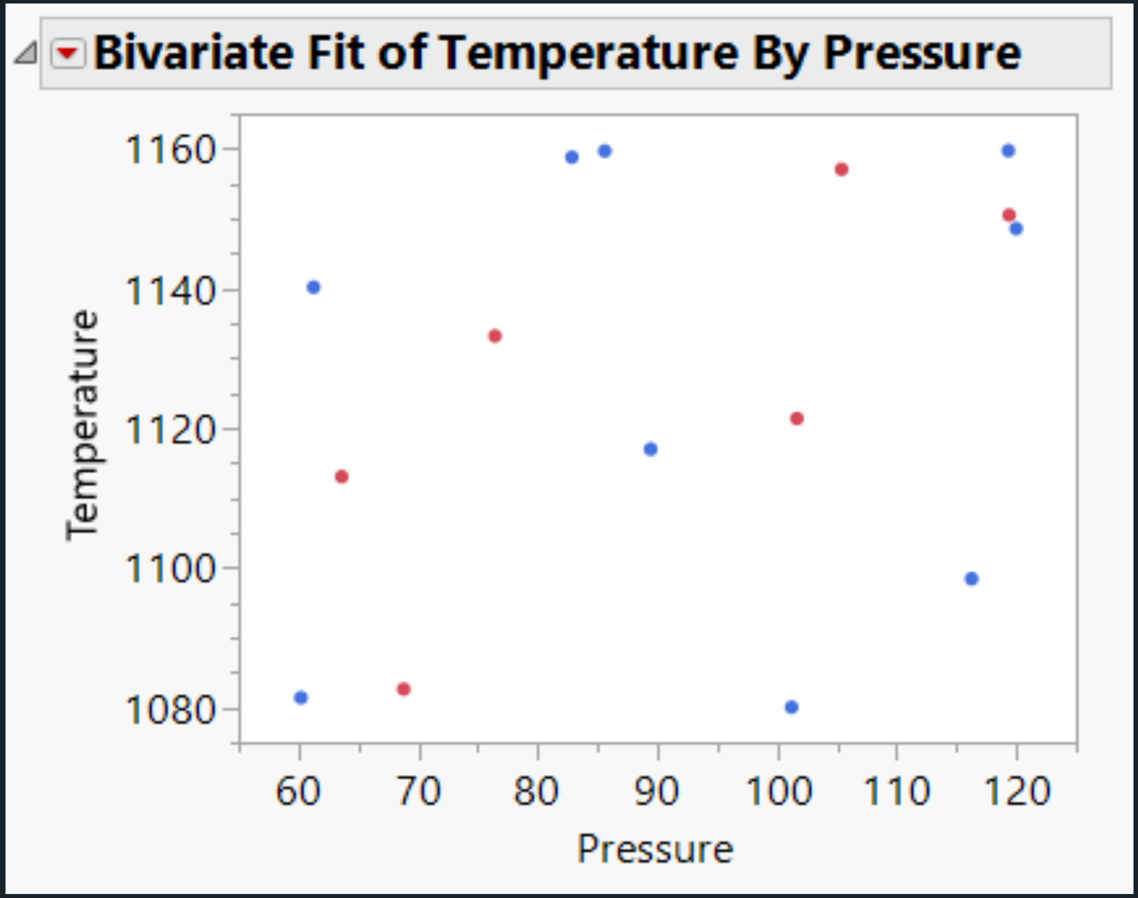
Back

#3: Machine suppliers with temperature and pressure



#3: Machine suppliers with temperature and pressure

Sample design



Covariate Factors

More use cases: Using all rows

- We have observable, uncontrollable factors for our experimental units that we want to account for
  - Assign two treatments to a group of patients.
- We want to force a particular structure for a subset of inputs (that may even be controllable)
  - 12 runs, 5 two-level factors, but factor A needs 2/3 of the levels at "L1" and 1/3 at "L2".
- DSD-type designs
  - 3-level categorical factor
- Complex block structures
- Account for other covariate information (time trend, for example)

jmp STATISTICAL DISCOVERY

Covariate Factors

More use cases: Using a subset of rows

- Pick an optimal subset (NOT a random subset)
  - Pick a subset of patients from a larger group
  - Hard-to-change covariates (such as multiple measurements per student)
- Augmentation - force original runs into the design
- Allow runs to be used more than once
- Constraints: Disallowed combinations alternative
  - Create candidate set (Full Factorial, Space-Filling, Simulation) > Filter & select runs

$$X1^{2} + 4 \cdot X3 + X4 \cdot X5 \geq 100$$

Covariate Factors

"Behind the scenes" in JMP 16+

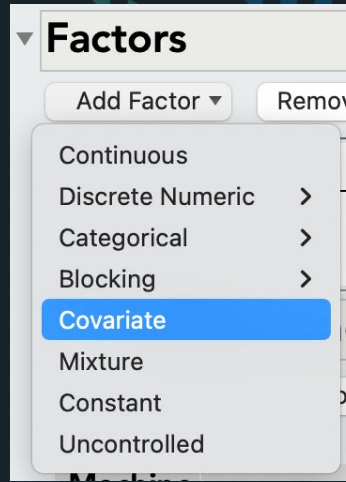
- JMP versions 10 – 15: Two-step optimization
  1. Row-exchange on the covariates (D-optimality)
     - Switching rows, starts with candidate set
  2. Coordinate-exchange on the controllable factors
     - Swapping values element-by-element, no candidate set needed
- JMP versions 16+: Hybrid approach
  – Allows for different optimality criteria to be considered for the covariates.
  – More flexibility for your design creation (best of both worlds).
  – Option much more visible to the user.

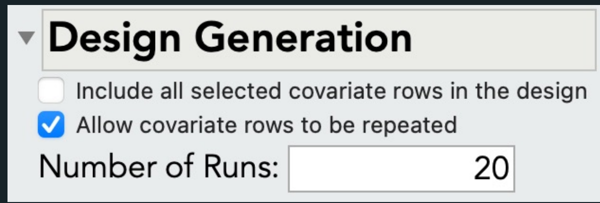
JMP STATISTICAL DISCOVERY

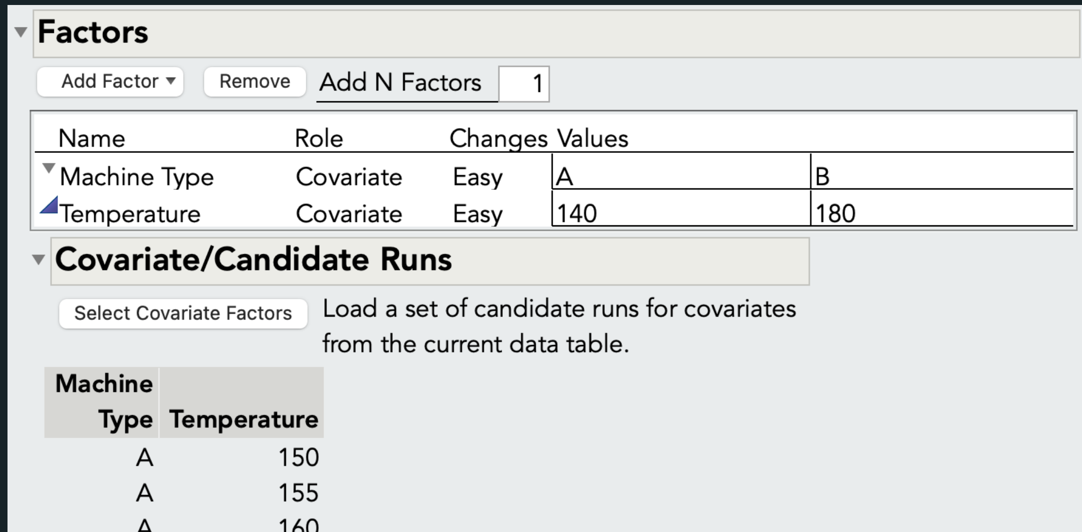
Covariate Factors

High flexibility

- Covariates can be combined with other Custom Design settings
  - Other factor types, e.g. mixtures
  - Split-plot designs, e.g. hard-to-change covariates
  - Blocking
  - Model effects, e.g. interactions
- No JMP Pro needed







JMP STATISTICAL DISCOVERY

Suggested resources

- JMP Developer Tutorial: “Handling Covariates Effectively when Designing Experiments”, Ryan Lekivetz, Principal Research Statistician Developer, Oct 2021
- JMP Blog: “What is a covariate in design of experiments?”, Ryan Lekivetz, Principal Research Statistician Developer, Feb 2021
- JMP Help: Design of Experiments Guide >
  - Custom Designs > Build a Custom Design > Factor Types
  - Examples of Custom Designs > Experiments with Covariates
- Goos, P., and Jones, B. (2011). “Optimal Design of Experiments: A Case Study Approach”, chapter 9. New York: John Wiley & Sons.

jmp.com