STATISTICS
TEXAS A&M UNIVERSITY

Handout 06

Multi Factor Designs and Blocking

Full Factorial Design

2^k Factorial Design

Copyright (c) Derya Akleman 2017

Full Factorial

- Most experiments involve two or more factors. These factors have two or more levels and can be quantitative or qualitative.

- A design is said to be a full factorial design if all possible combinations of factor levels are included in the experiment.

- A large number of runs are required for a full factorial as the number of factors increase.

  For example, for a full factorial design with *f* factors such that factor *i* has n_i levels, the number of runs is given by $n_1 * n_2 * n_3 \ldots * n_f$.

Copyright (c) Derya Akleman 2017

STATISTICS
TEXAS A&M UNIVERSITY

Full Factorial

Advantages:

- Factorials reveal interactions
- Factorials are more efficient than experiments that use one factor at a time
- Combinations of factor levels provide replication for individual factors when factors are removed from the design

Disadvantage: It requires too many runs

Example: Examine the effects amount of Catalyst, the level of Pressure and the Stirring Rate on the Filtration Rate of a chemical product. If each factor has 5 levels then you must perform 5*5*5=125 runs

Exercise: How many runs would there be for a full factorial experiment with 4 factors with 3 levels each?

Copyright (c) Derya Akleman 2017

Full Factorial

2 x 2 Factorial Design

		Drug Therapy	
		Placebo	***Prozac***
Psycho-therapy	***None***	Control	Prozac
	CBT	CBT	Combined Therapy

2 x 3 Factorial Design

		Drug Therapy	
		Placebo	***Prozac***
Psycho-therapy	***None***	Control	Prozac
	CBT	CBT	CBT + Prozac
	EFT	EFT	EFT + Prozac

Going 3D: 2 x 2 x 2 Factorial Design

Female

		Drug Therapy	
		Placebo	***Prozac***
Psycho-therapy	***None***	♀ Control	♀ Prozac
	CBT	♀ CBT	♀ Combo

Male

		Drug Therapy	
		Placebo	***Prozac***
Psycho-therapy	***None***	♂ Control	♂ Prozac
	CBT	♂ CBT	♂ Combo

Full Factorial model

2x2 model

$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$, where

$\mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$,

i=1,…,a, j=1,…,b, k=1,..,n_{ij},   a=2, b=2

2x3 model

$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$, where

$\mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$

i=1,…,a, j=1,…,b, k=1,..,n_{ij},   a=2, b=3

2x2x2 model

$y_{ijkl} = \mu_{ijk} + \varepsilon_{ijkl}$, where

$\mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}$

i=1,…,a, j=1,…,b, k=1,..,c, l=1,..,n_{ijk},   a=2, b=2,c=2

Copyright (c) Derya Akleman 2017

Full Factorial

Two types of effects can emerge in multi-factorial designs:

Main Effects: When one independent variable (factor) has an effect on its own. That is, the mean for some pair of levels of the independent variable differ significantly from one another.

Interaction Effects: When the effect of one independent variable is different for different levels of another independent variable .

These are NOT mutually exclusive

Copyright (c) Derya Akleman 2017

Battery Life Experiment

Temperature (Categorical with levels 15,70, 125)
Type of Plate material (Categorical with levels 1,2,3)

Battery Life

Goal: maximize the Battery life

<u>Generating the full factorial design in JMP:</u>

Select DOE-Full Factorial Design

Under Responses, change Y to Battery Life

Under factors,

select Categorical-3 level then change X1 to Temperature and type the levels 15, 70, 125 in the same order

select Categorical-3 level then change X2 to Type and type the levels 1, 2, 3 in the same order

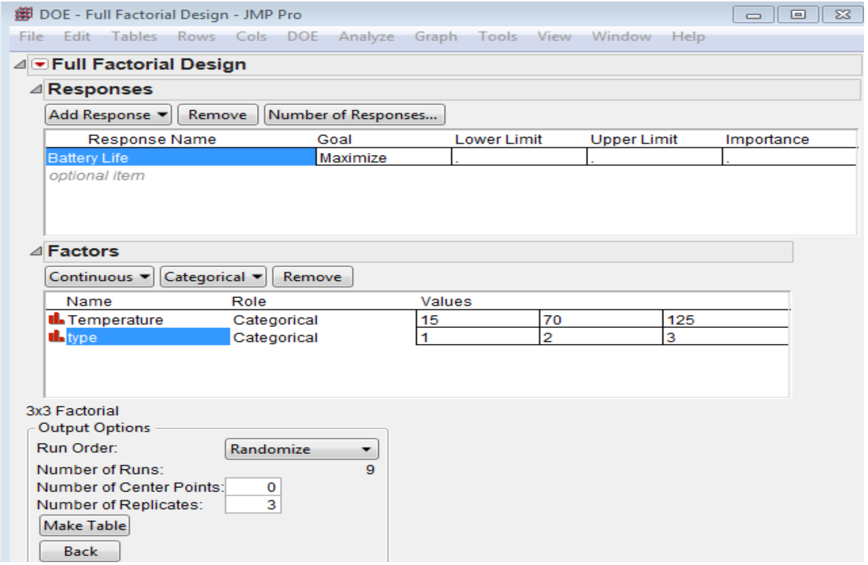
To save as a separate JMP table for future use: *Save factors or Save responses* clicking the red triangle next to Full Factorial Design.

If you do not need to save, click Continue and type 3 for the number of replicates.

Copyright (c) Derya Akleman 2017

STATISTICS
TEXAS A&M UNIVERSITY

Battery.JMP



Response Battery Life

Whole Model

Actual by Predicted Plot

Summary of Fit

RSquare	0.76521
RSquare Adj	0.695642
Root Mean Square Error	25.98486
Mean of Response	105.5278
Observations (or Sum Wgts)	36

Analysis of Variance

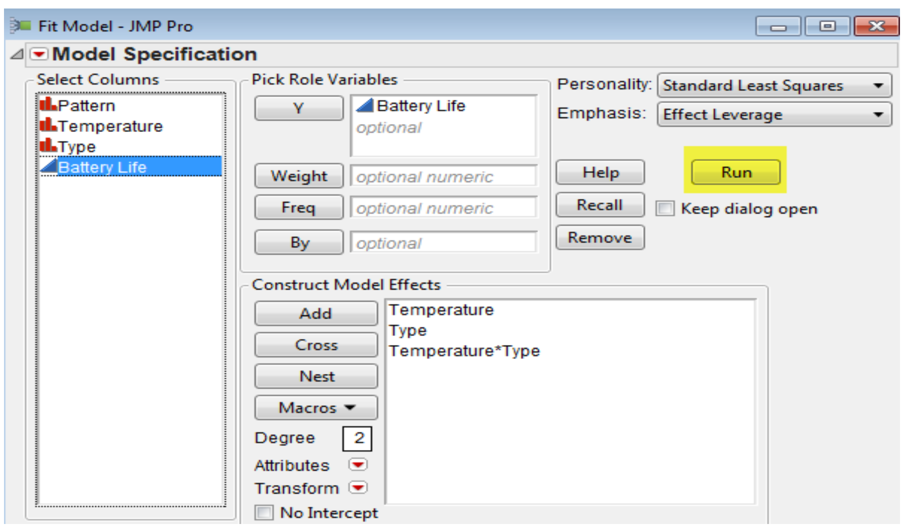
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	8	59416.222	7427.03	10.9995
Error	27	18230.750	675.21	Prob > F
C. Total	35	77646.972		<.0001*

Parameter Estimates

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Temperature	2	2	39118.722	28.9677	<.0001*
Type	2	2	10683.722	7.9114	0.0020*
Temperature*Type	4	4	9613.778	3.5595	0.0186*

Analyze-Fit model



Battery.JMP

<u>Look at the following and comment on the correct approach.</u>

LS Means Plot for Temperature without regard to Type

LS Means Plot for Type without regard to Temperature

LS Means Plot for Interaction (Type*Temperature)

To maximize the Battery Life, what should you set Temperature and Type at?

Copyright (c) Derya Akleman 2017

Battery.JMP

Look at the following and comment on the correct approach.

LS Means Plot for Temperature without regard to Type

LS Means Plot for Type without regard to Temperature

LS Means Plot for Interaction (Type*Temperature)

To maximize the Battery Life, what should you set Temperature and Type at?

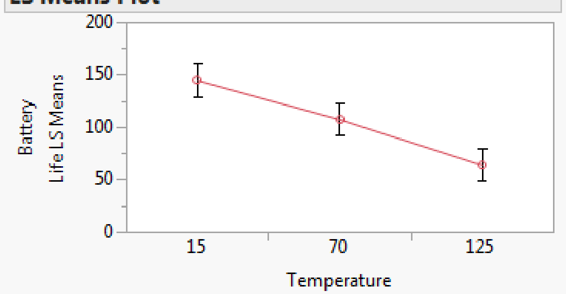
Temperature

Leverage Plot

Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
15	144.83333	7.5011830	144.833
70	107.58333	7.5011830	107.583
125	64.16667	7.5011830	64.167

LS Means Plot



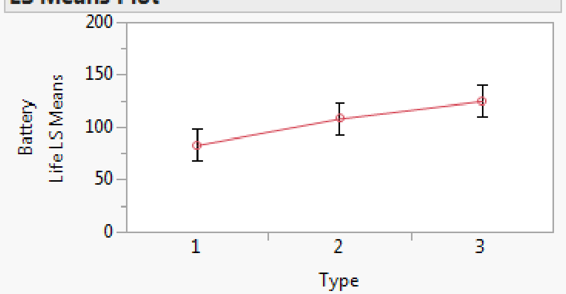
Type

Leverage Plot

Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	83.16667	7.5011830	83.167
2	108.33333	7.5011830	108.333
3	125.08333	7.5011830	125.083

LS Means Plot



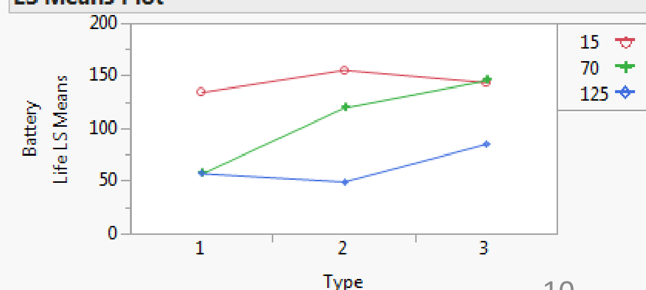
Temperature*Type

Leverage Plot

Least Squares Means Table

Level	Least Sq Mean	Std Error
15,1	134.75000	12.992430
15,2	155.75000	12.992430
15,3	144.00000	12.992430
70,1	57.25000	12.992430
70,2	119.75000	12.992430
70,3	145.75000	12.992430
125,1	57.50000	12.992430
125,2	49.50000	12.992430
125,3	85.50000	12.992430

LS Means Plot



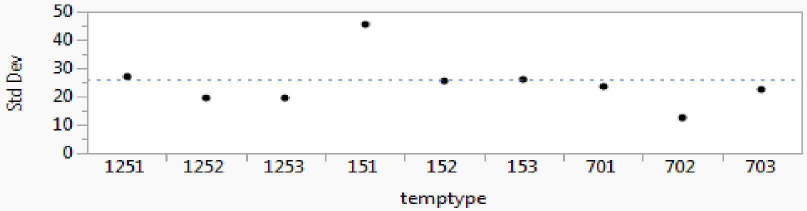
Copyright (c) Derya Akleman 2017

Battery.JMP

Are the variances from each population different?

Oneway Analysis of Battery Life By temptype

Tests that the Variances are Equal

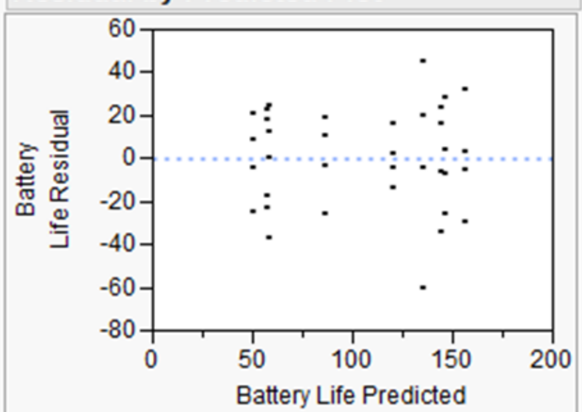


Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
1251	4	26.85144	18.75000	18.50000
1252	4	19.26136	14.50000	14.50000
1253	4	19.27866	14.50000	14.50000
151	4	45.35324	32.75000	32.75000
152	4	25.61738	17.75000	17.75000
153	4	25.97435	20.00000	20.00000
701	4	23.59908	20.25000	20.25000
702	4	12.65899	9.25000	9.25000
703	4	22.54440	16.25000	16.25000

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	0.9471	8	27	0.4958
Brown-Forsythe	0.7996	8	27	0.6081
Levene	0.9019	8	27	0.5289
Bartlett	0.6544	8	.	0.7321

Warning: Small sample sizes. Use Caution.

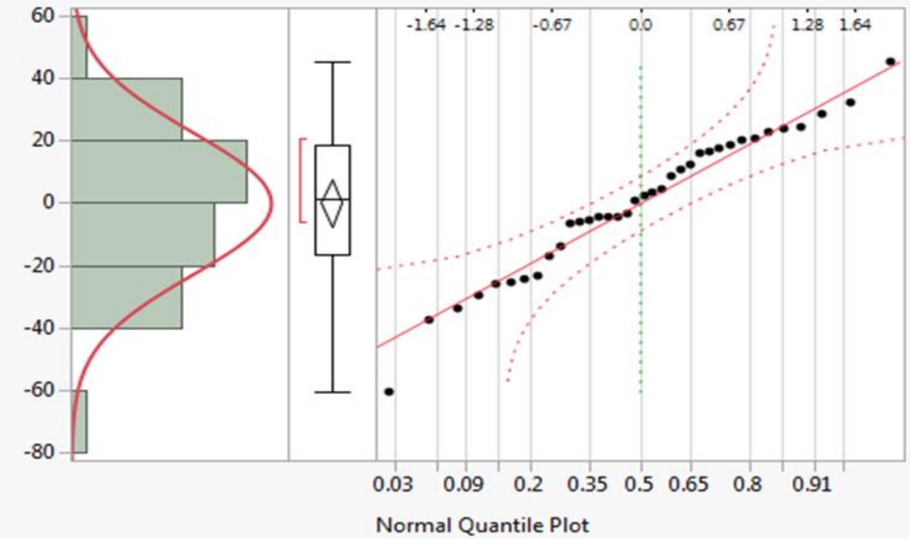
Residual by Predicted Plot



Are the residuals normally distributed?

Distributions

Residual Battery Life



Normal(7.1e-15,22.8228)

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	7.105e-15	-7.722112	7.7221125
Dispersion	σ	22.822764	18.511124	29.770882

-2log(Likelihood) = 326.36218445802

Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.976057	0.6117

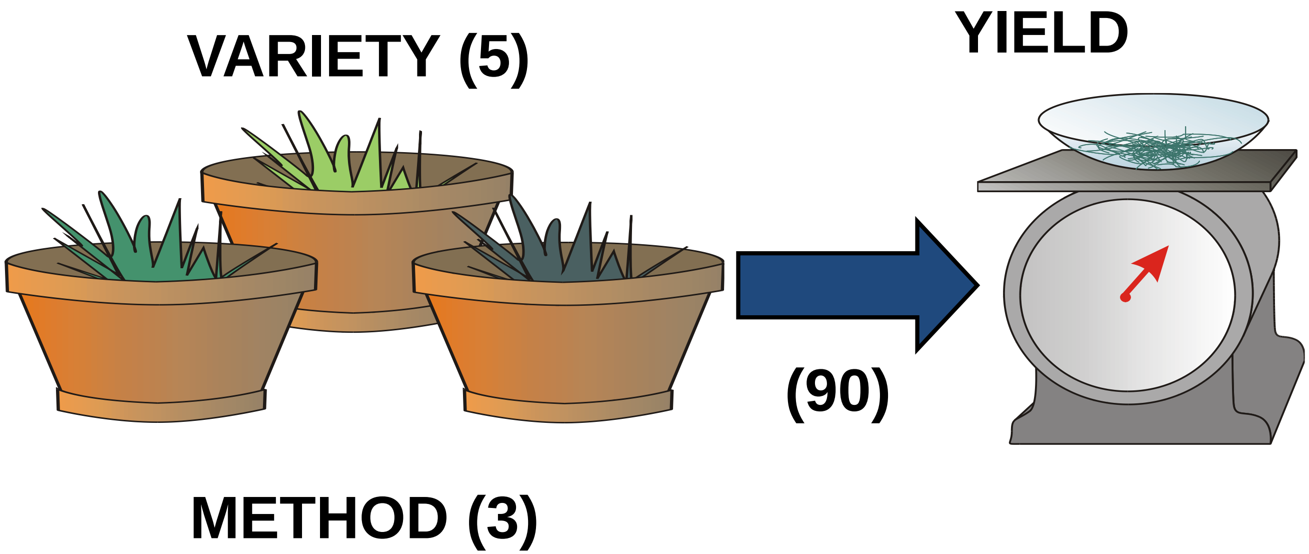
Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Copyright (c) Derya Akleman 2017

Grass Example

- Three seed growth methods are applied to seeds from each of randomly chosen five varieties of turf grass.
- Six pots are planted with seeds from each method by variety combination.
- The 90 pots are randomly placed in a uniform growth chamber and dry matter yields are measured from clippings at the end of four weeks.

method	variety	yield
A	1	22.1
A	1	24.1
A	1	19.1
A	1	22.1
A	1	25.1
A	1	18.1
A	2	29.1
A	2	17.1
A	2	21.6
A	2	28.6
A	2	17.1
A	2	26.6
A	3	25.3
A	3	25.8
A	3	22.8
A	3	28.3
A	3	21.3
A	3	18.3
A	4	19.8
A	4	28.3
...	...	...



Copyright (c) Derya Akleman 2017

Define the Mixed Model

$$y_{ijk} = \mu + \alpha_i + b_j + (\alpha b)_{ij} + \varepsilon_{ijk}$$

method effect, fixed

variety effect, random

Method* variety effect, random

$$b_j \sim N(0, \sigma_b^2), (\alpha b)_{ij} \sim N(0, \sigma_{\alpha b}^2), \varepsilon_{ijk} \sim N(0, \sigma^2)$$

therefore,

$$E(y_{ijk}) = \mu + \alpha_i$$
$$Var(y_{ijk}) = \sigma_b^2 + \sigma_{\alpha b}^2 + \sigma^2$$

Copyright (c) Derya Akleman 2017

Plotting the Data and Fitting the Two-Way Mixed Model

Which of the following is a correct statement?

(i) **`method`** is found to be significant only for the five varieties of grass seed included in the study.

(ii) **`method`** is found to be not significant only for the five varieties of grass seed included in the study.

(iii) **`method`** is found to be significant for all varieties of grass seed.

(iv) **`method`** is found to be not significant for all varieties of grass seed.

GrassExample.sas
PROC MEANS,
PROC SGPLOT,
PROC MIXED

Copyright (c) Derya Akleman 2017

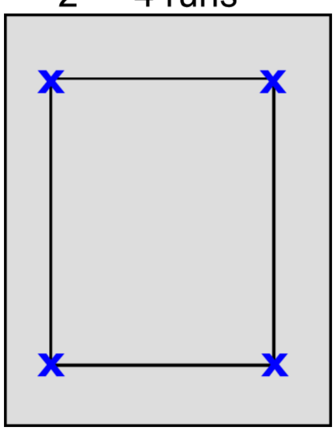
2^k Factorial Designs

The 2^k factorial design

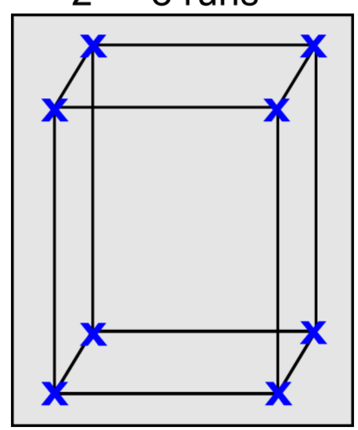
- is a special case of the full factorial design, introduced to reduce the number of design points
- is widely used in industrial experimentation
- forms basic building blocks for other designs
- assumes that the response is linear in terms of the factors in the design.

Two level Full Factorial Design

2^2 = 4 runs



2^3 = 8 runs



Copyright (c) Derya Akleman 2017

Coded Levels in 2^k Factorial Designs

A 2^k design is one such that each factor has exactly two levels. These two levels are usually called low and high, and usually denoted by -1 and +1.

Coded value=$\frac{value-(high\ value+low\ value)/2}{(high\ value-low\ value)/2}$

With k factors, there will be:

k main effects, $\binom{k}{2}$ two-factor interactions, $\binom{k}{3}$ three-factor interactions,……………

……., 1 *k*-factor interactions

<u>Example:</u> Design with 5 factors (or 2^5 design), will have 5 main effects, 10 two-factor interactions, 10 three-factor interactions, 5 four-factor interactions and 1 five-factor interaction so the total number of effects is 31. If this experiment is run with no replications and only pairwise interactions then the total number of effects is 15.

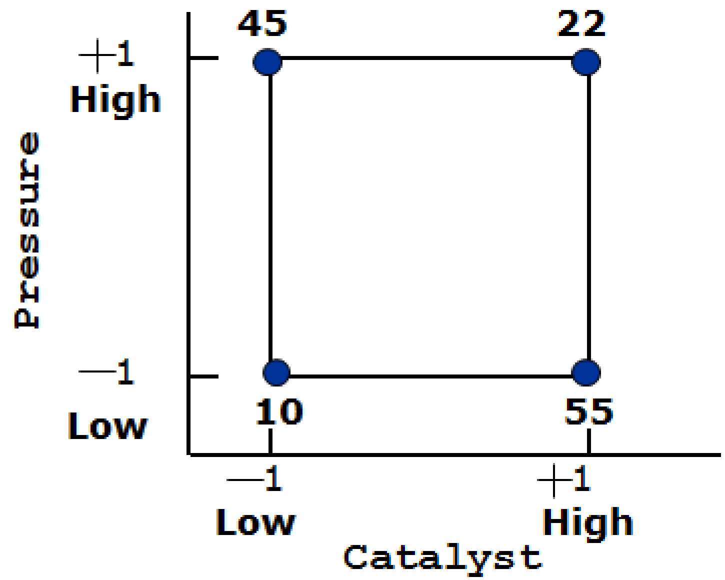
If the strong assumption of "no interaction" is made but not true, is the estimate of the error inflated?

Why?

How are the test results effected?

Copyright (c) Derya Akleman 2017

Two Factor Factorial Experiment



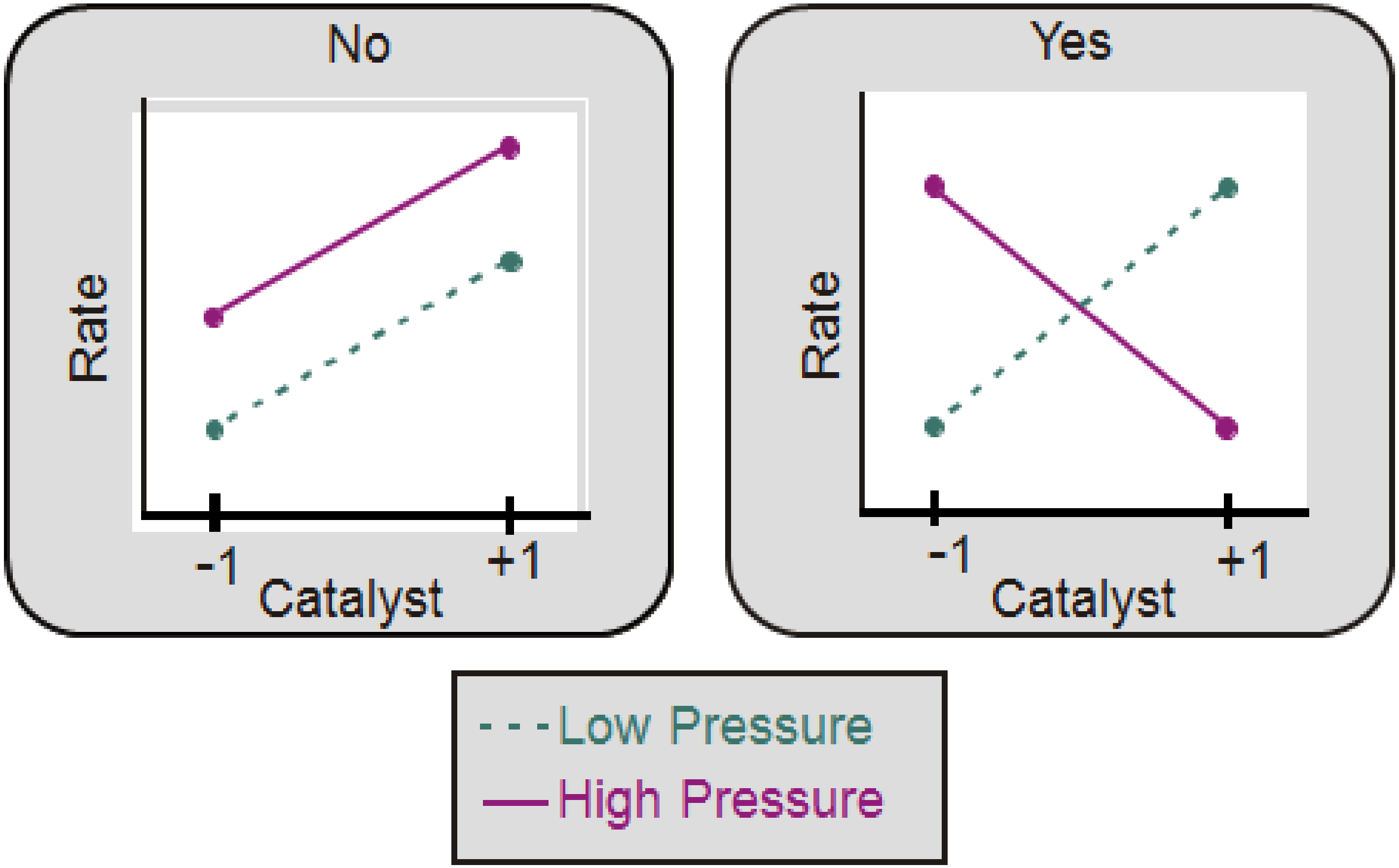
The main effect Catalyst=(55+22)/2 - (45+10)/2 = 11

The main effect Pressure=(45+22)/2 - (55+10)/2= 1

The interaction (Catalyst*Pressure) effect=(10+22)-(45+55)/2=-34

Copyright (c) Derya Akleman 2017

Interactions



Full Factorial

Catalyst		Pressure: Low	Pressure: High
	High	$(+1-1)_1$ $(+1-1)_2$	$(+1+1)_1$ $(+1+1)_2$
	Low	$(-1-1)_1$ $(-1-1)_2$	$(-1+1)_1$ $(-1+1)_2$

$(-1-1)_1$
$(-1+1)_1$
$(-1+1)_2$
$(+1+1)_1$
$(+1+1)_2$
$(+1-1)_1$
$(-1-1)_2$
$(+1-1)_2$

4 treatments
2 runs/treatment
(1 replicate)
order of runs are random
(CRD)

In CRD, the order of the design points is selected at random.

The model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

Rate=Base level+Catalyst+Pressure+Catalyst*Pressure+unaccounted for variation

Copyright (c) Derya Akleman 2017

2^k Factorial

With no replication:

- These are designs with one observation at each corner of the *cube*.
- One unusual value of the response could cause an experimenter to draw incorrect conclusions.
- The experimenter runs the risk of modeling only noise and not the effects of the factors.
- You cannot test for interaction
- These designs are widely used.

Replication:

- Replication allows for an estimation of pure error and for a test of lack of fit.
- Center points can be used for replication, if the factors are quantitative.

Copyright (c) Derya Akleman 2017

Yield Experiment

A researcher wants to study the effects of varying four continuous factors, Temperature, Pressure, Concentration and Time on the yield of the chemical process.

Temperature (continuous) 225 & 250	**Pressure (continuous)** 60 & 80
Concentration (continuous) 14 & 18	**Time (continuous)** 2.5 & 3

number of factors?

number of treatments?

number of levels for each factor?

At temperature 225, Coded value = $\frac{225-(225+250)/2}{(250-225)/2}$ =- 1

At temperature 250, Coded value = $\frac{250-(225+250)/2}{(250-225)/2}$ = 1

Copyright (c) Derya Akleman 2017

2^4 Factorial Experiment with no Replication

Select DOE-Screening Design

Change Y to Yield

Under Factors,

change 1 in the box next to Continuous into 4

select Add to enter all factors with their levels

select continue

select 16 Full Factorial >6 - Full Resolution

select continue

Select Make a Table

Now use Yield.JMP data to analyze.

Copyright (c) Derya Akleman 2017

Example

You are interested in minimizing the fuel efficiency (Fuel, measured in pounds of fuel per hour) for a company's fleet of aircraft. Consider an experiment with 16 planes. The company determined these factor of interest with the upper and lower values as shown:

Factor	Description	Low level	High level
Weight	Airplane weight (in thousands of pts)	80	120
Center	Center of gravity minus the center of lift (in feet)	-15	15
Altitude	Cruising altitude (in thousand of feet)	15	35
Speed	True air speed (in kilometers per hour)	300	450

Choose the appropriate platform and generate a two-level full factorial design with 16 runs. Be sure to define the factors and levels and define the response.

Why is 16 is the minimum number of runs?

Use Plane.JMP

Analyze this data and determine if any effects are significant at 5%.

If significant effects are found, determine the significant ones.

Check the model assumptions

Copyright (c) Derya Akleman 2017