Design of Experiment
What does DOE mean?

DOE stands for Design Of Experiments

DOE techniques are used to generate a series of designs which satisfy different requisites according to the objective of the analysis, which can be however always summarized as:

“have the best with the smallest effort”

The main concern is to determine the relationship between factors (inputs) affecting a process and the output of that process with the lowest number of experiments as possible.
In other words….

• Have you flown a paper airplane before? (Hopefully not in this class)

• Do you always use the same type of paper?
• Do you always use the same design?

• Do you want it to fly straight or do tricks?
What does DOE mean in this case?

- Design of experiment is used here to test paper airplane flight distance.
- We want the planes to fly as far as they can.
- We need to think about how we are going to design and perform the experiment.
- What things do we need to think about? (Think about the steps of the Scientific Method)

- What question are we trying to answer?
  - We want to design an experiment to test how the addition of paper clips will affect the flight distance of the paper airplane.
  
  - How does adding paper clips to a paper airplane affect its flight?
Procedure

• How are we going to perform the experiment?
  - What do we need to do?
  - What needs to be kept constant?
  - What is our control?
  - Which are our independent variables?
  - What are we going to observe? How?

<table>
<thead>
<tr>
<th># of paper clips</th>
<th>Flight Distance (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>
Making a graph of your data

How adding paper clips affects paper airplane flight distance.

1. Number of paper clips for x-axis.
2. Determine interval and label for x-axis.
3. Determine interval and label for y-axis.
4. Plot your data.
5. Connect the data points.
The choice on experiments

**Aim of the Activity**: have a good sample from laboratory tests for statistic study

**Cost per Experiment**: 1000 $

6 Random entries  
Cost of the Campaign = 6,000 $

8 Full Factorial entries  
Cost of the Campaign = 8,000 $

64 Full Factorial entries  
Cost of the Campaign = 64,000 $

😊 Cost  
😊 Quality

😍 Cost  
😊 Quality

😢 Cost  
😢 Quality
Curse of dimensionality

- is the problem caused by the exponential increase in volume associated with adding extra dimensions to a space.

For example, 100 evenly-spaced sample points suffice to sample a unit interval with no more than 0.01 distance between points; an equivalent sampling of a 10-dimensional unit hypercube with a lattice with a spacing of 0.01 between adjacent points would require $10^{20}$ sample points.
Random & SOBOL

- The DOE Random & Sobol Sequences cover sufficiently the domain of the functions.

- The mathematical theory is the Random Number Generation.
  - Sequence Random (function with “many” variables)
  - Sequence Sobol (function with “few” variables < 10)

- Random sequences of experiments allow the sampling of a configuration space with continuous and discrete variables without pre-defined interactions

- The use of random sequences avoids the risk of “correlated sampling” even in the case of limited sampling
DOE: space fillers

Random and Sobol

Random sequence

Sobol sequence

Pseudo-random
High number of variables
Suitable for GA

Better distributed designs
Suitable for a low number of variables (<10)
Suitable for RSM, GA
Factorial DOE

Full factorial

Number of generated designs: $m^n$
$m = \text{variable level}$
(number of “possible” states of a variable)
$n = \text{variables number}$

Full Factorial 2 levels

$2^n$ designs allow to correctly capture the first order interaction (e.g. $x*y$)

Full Factorial 3 levels

$3^n$ designs allow to correctly capture the second order interaction (e.g. $x^2*y$)
## DOE: space fillers

### Reduced factorial

Number of generated designs = \(2^p\)

\(p < n\) (number of variables)

<table>
<thead>
<tr>
<th>n. design</th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
<th>x4 (=x1*x2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Removing the second order information for \(x1*x2\)
DOE: space fillers

Cubic face centered

$2^n + 2n + 1$ designs

allow to correctly capture the second order interaction (e.g. $x^2y$)

3 variables
15 designs
DOE: space fillers

Box Behnken

Very similar to the face centered algorithm, it uses the mid-side nodes and the center of the (hyper-) cube

3 variables
13 designs
DOE: space fillers

Latin Square

The designs number \( (m^2, \text{where } m \text{ is the required level}) \) does not depend on the number of variables

Suitable for statistical analysis

Only the zero order interactions can be captured

Example: Latin Square with 3 variables \((x_1, x_2, x_3)\) and 3 levels

- \( x_1 \) (1,2,3)
- \( x_2 \) (A,B,C)
- \( x_3 \) (a,b,c)
DOE: space fillers

Latin Square

\[ x \times y \times z \]

\[ x \{1, 2, 3, 4, 5\} \]
\[ y \{1, 2, 3, 4, 5\} \]
\[ z \{1, 2, 3, 4, 5\} \]
DOE: Reliability and Robustness

Montecarlo and Latin Hypercube

These techniques generate a series of randomly distributed designs according to a given probability density function (Normal, Cauchy, Weibull,…)

Define linear correlations between variables
DOE: Reliability and Robustness

Cauchy

Normal
DOE: Reliability and Robustness

Montecarlo versus Latin Hypercube
DOE: Reliability and Robustness

Montecarlo versus Latin Hypercube
DOE: hands on

Try to show the differences between Montecarlo and latin Hypercube

For a two variables problem

Generate a Montecarlo DOE
Generate a Latin Hypercube DOE

Plot hystograms, curve fitting, correlation matrix, Q-Q plots, compare results
Statistical analysis

- Statistical tools can be used to **analyze distributions**, coming from experiments or from DOE, to obtain information from the system (e.g., what is the most responsible cause of failures)

- Statistical tools can be used to **find correlations**, in particular which input variables have most influences in the system outputs; these results can be obtained from an available database, from a DOE, or after an optimisation phase
Statistical analysis

Several statistical tools are available:

- Density and cumulative Distribution
- Box-Whiskers
- Quantile Plot
- Statistics summary
- ANOVA
- Broken Constraints
- Main and interaction effects
- Student
- Correlation Matrix
- Scatter Matrix
- SOM

Tools for distribution analysis

Tools for correlation analysis
Probability Density Plot

- The Probability Density chart summarizes the distribution of a data set (min, max, mean, variance,…)

- This plot is obtained by splitting the range of the data into equal sized classes

- The number of points that fall into each class are counted

- It reveals:
  - the kind of distribution
  - where the data is located
  - how spread out the data are
Statistical analysis for distributions (Density and Cumulative Distribution)

Probability density function (for any input/output variable);

Cumulative Distribution function (for any input/output variable);

- Select theoretical distribution from Properties>Distribution
- Distribution parameters are available in Properties>General Information
Statistical analysis for distributions (Box Whiskers)

Q1 (first quartile): cut off lowest 25% of data
Q3 (third quartile): cut off highest 25% of data

U1 (upper fence) = Q3+1.5*(Q3-Q1) : it is the limit over which points are considered as outliers
L1 (lower fence) = Q1-1.5*(Q3-Q1) : it is the limit below which points are considered as outliers

MEDIAN: 50% of the distribution data are expected to be lower (or greater) than the Median value.

DENSEST HALF RANGE: smallest range that contains half of the distribution samples
CONFIDENCE INTERVAL: 95% of confidence that the mean is inside this range

Reports statistical data (symbols and summary table) for any input/output variable
Statistical analysis for distributions *(quantile-quantile plot)*

Quantile-quantile plot for any input/output variable: It is used to determine if the data points (abscissa) can be represented by an analytic distribution (ordinate, you can change the type in properties)

- best fitting distribution would have points on the green 45° line
- if some points are outside the region bounded by red lines (Lillenfor’s test), distributions are different
Statistical analysis for distributions (Statistic summary)

Automatic creation of all the previous chart on the selected input/output variables
ANOVA: import database

- import database (or use existing one in Design table)
- each column is a different distribution

e.g., each column is a different production line, and shows the distribution of defects per day
Question: which line is most defective?
Box Whiskers can give a preliminary indication of which is most defective:

LizT seems to have highest mean, but we need a tool to validate distribution analysis: this is ANOVA
Statistical analysis for distributions (ANOVA)

• Select ANOVA tool
• Select table
• Select variables (distributions) to analyze

• in this case LizT is compared with other two production lines
Statistical analysis for distributions (ANOVA)

- **ANOVA analysis summary** is given to compare influence of different variables

- **Variance check**: only if variances of variables is similar, ANOVA can be performed

- **Hartley's test** ($F_{\text{max}}$ test) is performed to verify that different groups have a similar variance, an assumption needed for ANOVA.

- **ANOVA table**: determines if variables have different significance
Statistical analysis for distributions (ANOVA)

- Box-Whiskers, table of means and differences: if previous tests are valid, indicates the most significant variable

- here we conclude that LizT is statistically the most defective line
Tools for correlation analysis

- Main and interaction effects
- Student
- Correlation Matrix
- Scatter Matrix
Statistical analysis for correlations (Effect matrix)

• Effect of one input (abscissa: range is split in – and + half) to one output (ordinate: mean value)

• Same effect: output are in this case represented by bars (mean and STDEV)
Details on definition of Effect

**Definition of Effect of input A for output Obj:**

- Medium value of the function for every variable (computed for half range + or -):
  
  \[ A^-; \quad A^+ \]

  \[ \text{Effect} = (\text{Mean } +) - (\text{Mean } -) \]

- The same for the interactions between the variables:
  
  \[ AB +/ - - \quad (\text{concord}); \quad AB +/ - + \quad (\text{discord}) \]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>OBJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>65.6</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>79.3</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>51.3</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>69.6</td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>59.8</td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>77.7</td>
</tr>
<tr>
<td>7</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>74.2</td>
</tr>
<tr>
<td>8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>87.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>AB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean +</td>
<td>75</td>
<td>70.8</td>
<td>78.8</td>
<td>69.5</td>
<td></td>
</tr>
<tr>
<td>Mean -</td>
<td>66.5</td>
<td>70.6</td>
<td>62.8</td>
<td>71.8</td>
<td></td>
</tr>
<tr>
<td>Effect</td>
<td>8.5</td>
<td>0.2</td>
<td>16</td>
<td>-2</td>
<td></td>
</tr>
<tr>
<td>Mean AB+</td>
<td>76.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean AB-</td>
<td>64.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effect</td>
<td>15.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Statistical analysis for correlations (Student charts)

- t-Student chart shows the effect of each input variable accordingly to a selected output

<table>
<thead>
<tr>
<th>Factor</th>
<th>Effect Size</th>
<th>Significance</th>
<th>Critical t (p = 0.05)</th>
<th>Critical t (p = 0.01)</th>
<th>Critical t (p = 0.001)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1 + X2 + X3</td>
<td>7.60E01</td>
<td>0.009</td>
<td>1.294E00</td>
<td>1.905E00</td>
<td>2.571E00</td>
</tr>
<tr>
<td>X1 + X2</td>
<td>-3.07E01</td>
<td>0.009</td>
<td>1.294E00</td>
<td>1.905E00</td>
<td>2.571E00</td>
</tr>
<tr>
<td>X1 + X3</td>
<td>2.47E01</td>
<td>0.009</td>
<td>1.294E00</td>
<td>1.905E00</td>
<td>2.571E00</td>
</tr>
</tbody>
</table>

- Accordingly to t-Student parameter definition, overall 3D chart shows the normalised effect of each input variable accordingly to each output

Significance indicates the probability that response variable and the factor are not correlated (i.e. it is the probability that a difference in the response at factor variation is due to chance).
Details on Student test test

- **Effect size** indicates the kind of relationship between the factor and the response variable: a value less than zero indicates that the relationship is inverse.

- An high value of **Significance** parameter (max value 0.5) indicates that there is a very high probability that the factor doesn’t influence the response variable.
Details on Student test

- $n_+$ and $n_-$ are the numbers of values in the upper and lower parts of domain of the input variable
- $M_+$ and $M_-$ are the means of the values for the output variable $x$ in the upper and lower parts of domain of the input variable
- $S_G^2$ is the general variance
- $S_+^2$ and $S_-^2$ are the variances of the population for the output variable $x$ in the upper and lower parts of domain of the input variable

$$ t = \frac{|M_- - M_+|}{\sqrt{\frac{S_G^2}{n_-} + \frac{S_G^2}{n_+}}} $$

$$ S_G^2 = \frac{(n_- - 1)S_-^2 + (n_+ - 1)S_+^2}{n_+ + n_- - 2} $$

$$ S_+^2 = \frac{\sum (x_+ - M_+)^2}{n_+ - 1} $$

$$ S_-^2 = \frac{\sum (x_- - M_-)^2}{n_- - 1} $$

If $t$ follows a well known distribution called Student distribution then $M_-$ and $M_+$ are not statistically distinct i.e. probably there is no correlation between factor and response variable (significance close to 0.5)
Details on Student test

<table>
<thead>
<tr>
<th>Factors</th>
<th>Effect Size</th>
<th>Significance</th>
<th>t-Student</th>
<th>critical t (α = 0.10)</th>
<th>critical t (α = 0.05)</th>
<th>critical t (α = 0.01)</th>
</tr>
</thead>
<tbody>
<tr>
<td>obj1</td>
<td>9.6499E-1</td>
<td>0.129</td>
<td>1.1688E0</td>
<td>1.3304E0</td>
<td>1.7341E0</td>
<td>2.5524E0</td>
</tr>
</tbody>
</table>

• **critical t (α=0.1)=1.33** → means that if \( t \geq 1.33 \) the response variable and the factor are correlated with a significance = 10% (i.e. there is 10% of probability that the difference between the range + and - of the response variable is due to chance).

• **critical t (α=0.05)=1.73** → means that if \( t \geq 1.73 \) the response variable and the factor are correlated with a significance = 5% (i.e. there is 5% of probability that the difference is due to chance).

• **critical t (α=0.01)=2.55** → means that if \( t \geq 2.55 \) the response variable and the factor are correlated with a significance = 1% (i.e. there is 1% of probability that the difference is due to chance).

In the example \( t =1.688 \) and the significance is 0.129 → means that the probability that the difference between the range + and - of the response variable is due to chance is 12.9%.

The significance \( \alpha \) is always between 0, i.e. correlation between factor and response variable, and 0.5, i.e. not correlation between factor and response variable (50% of probability that that the difference in the response variable ranges is due to chance).
Overall Student chart

For each response (S and V), effect of inputs are reported in an overall chart on a common scale
Statistical analysis for correlations (Correlation chart and Scatter matrix)

Correlation chart:
- +1 max direct correlation
- -1 max inverse correlation
- 0 no correlation

Scatter matrix:
- Report correlation, scatter plot and distribution charts for a pair of variables
Details on Correlation Matrix

- The correlation is a number (between -1 and 1) describing the degree of relationship between two variables.
- The correlation is a measure of the linear association.
- If it is exactly equal to 1, the two variables are perfectly positively correlated and the values all lie on a straight line with positive slope.
- If it is equal to zero, the variables are uncorrelated, that is linearly unassociated.
- If it is equal to -1, then the two variables are perfectly negatively correlated.
Statistical Analysis: Interaction Effects

Neither the t-Student test nor the correlation matrix are able to assess if interaction effects between factors exist.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Effect Size</th>
<th>Significance</th>
<th>t-Student</th>
<th>critical t ($\alpha = 0.10$)</th>
<th>critical t ($\alpha = 0.05$)</th>
<th>critical t ($\alpha = 0.01$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>x[1]</td>
<td>6.8079E1</td>
<td>0.009</td>
<td>2.7074E0</td>
<td>1.3450E0</td>
<td>1.7613E0</td>
<td>2.6245E0</td>
</tr>
<tr>
<td>x[0]</td>
<td>-3.1086E-15</td>
<td>0.500</td>
<td>1.0016E-16</td>
<td>1.4149E0</td>
<td>1.8946E0</td>
<td>2.9980E0</td>
</tr>
</tbody>
</table>
Statistical Analysis: Interaction Effects

By using **DOE Interaction Effects** and **Effects Matrix** is possible to assess if factors interactions play a role in the response variable.
Statistical Analysis: Interaction Effects

Doe Interaction Effects

If $x[0]$ and $x[1]$ are both set at upper or lower values, the response variable is higher, further information can be obtained by the effects matrix chart.
Statistical Analysis: Interaction Effects

It is possible to see that, though \( x[0] \) has not a main effect on the response variable, increasing both \( x[0] \) and \( x[1]\) an higher value of the response variable is obtained. On the other side if both variables are set to lower values there is no such an effect.
Example 1

How to use modeFRONTIER to get the most relevant qualitative information from a data-base of experiments
Statistical Analysis: Example 1

This experiment was conducted on a *catapult* – a table-top wooden device often used to teach design of experiments and statistical process control. The catapult has several controllable factors and a response easily measurable.
Statistical Analysis: Example 1

**Variables:**

- **Response Variable** $Y = $ distance
- **Factor 1 = band height** (height of the pivot point for the rubber bands – levels were 3.25 and 4.75 inches with a centerpoint level of 4);
- **Factor 2 = start angle** (location of the arm when the operator releases—starts the forward motion of the arm – levels were 0 and 20 degrees with a centerpoint level of 10 degrees);
- **Factor 3 = rubber bands** (number of rubber bands used on the catapult– levels were 1 and 2 bands);
- **Factor 4 = arm length** (distance the arm is extended – levels were 0 and 4 inches with a centerpoint level of 2 inches);
- **Factor 5 = stop angle** (location of the arm where the forward motion of the arm is stopped and the ball starts flying – levels were 45 and 80 degrees with a centerpoint level of 62 degrees)

A reduced factorial technique was used, the number of designs evaluated is $2^{5-1} = 16 + 4$ center points = 20 designs.
### Statistical Analysis: Example 1

<table>
<thead>
<tr>
<th>Pattern</th>
<th>band height</th>
<th>start angle</th>
<th>rbands</th>
<th>arm length</th>
<th>stop angle</th>
<th>Y_distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>----+</td>
<td>3.25</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>80</td>
<td>28</td>
</tr>
<tr>
<td>00000</td>
<td>4</td>
<td>10</td>
<td>2</td>
<td>2</td>
<td>62</td>
<td>99</td>
</tr>
<tr>
<td>++++</td>
<td>4.75</td>
<td>20</td>
<td>2</td>
<td>4</td>
<td>80</td>
<td>126.5</td>
</tr>
<tr>
<td>++++</td>
<td>4.75</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>45</td>
<td>126.5</td>
</tr>
<tr>
<td>++++</td>
<td>3.25</td>
<td>20</td>
<td>2</td>
<td>4</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>++++</td>
<td>4.75</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>45</td>
<td>35</td>
</tr>
<tr>
<td>00000</td>
<td>4</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>62</td>
<td>45</td>
</tr>
<tr>
<td>++++</td>
<td>4.75</td>
<td>20</td>
<td>1</td>
<td>0</td>
<td>80</td>
<td>28.25</td>
</tr>
<tr>
<td>++++</td>
<td>4.75</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>80</td>
<td>85</td>
</tr>
<tr>
<td>++++</td>
<td>3.25</td>
<td>20</td>
<td>1</td>
<td>0</td>
<td>45</td>
<td>8</td>
</tr>
<tr>
<td>++++</td>
<td>4.75</td>
<td>20</td>
<td>1</td>
<td>4</td>
<td>45</td>
<td>36.5</td>
</tr>
<tr>
<td>++++</td>
<td>3.25</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>45</td>
<td>33</td>
</tr>
<tr>
<td>00000</td>
<td>4</td>
<td>10</td>
<td>2</td>
<td>2</td>
<td>62</td>
<td>84.5</td>
</tr>
<tr>
<td>++++</td>
<td>4.75</td>
<td>20</td>
<td>2</td>
<td>0</td>
<td>45</td>
<td>28.5</td>
</tr>
<tr>
<td>++++</td>
<td>3.25</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>45</td>
<td>33.5</td>
</tr>
<tr>
<td>++++</td>
<td>3.25</td>
<td>20</td>
<td>2</td>
<td>0</td>
<td>80</td>
<td>36</td>
</tr>
<tr>
<td>++++</td>
<td>4.75</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>80</td>
<td>84</td>
</tr>
<tr>
<td>++++</td>
<td>3.25</td>
<td>20</td>
<td>1</td>
<td>4</td>
<td>80</td>
<td>45</td>
</tr>
<tr>
<td>00000</td>
<td>4</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>62</td>
<td>37.5</td>
</tr>
</tbody>
</table>
Statistical Analysis: Example 1

The probability density function and the box whiskers chart show the large spread of the data and a pattern that should be explained by the analysis.
Statistical Analysis: Example 1

Student chart is useful to see the relevance and the effects of the different parameters.
Statistical Analysis: Example 1

Interaction effects can play a role as well. This can be assessed by the DOE interaction effects chart: it is easy to understand if some parameters have a strong interaction. Further details can be gained by the matrix interaction charts.
Statistical Analysis: Example 1

- Run order interacts with arm length, band height and number of rubber bands.

- Bands height and number of rubber bands interact.

- Arm length and number of rubber bands interact.
Statistical Analysis

- An **accurate assessment** of the DOE is useful in any case: it gives a better insight of the problem and can reduce the complexity, limiting the number of variables and the variables definition range.

- **Be aware**: the statistical tools need DOE tables able to represent correctly all the design space.
Example 2

Choosing the proper DOE for statistical analysis
Statistical Analysis: Example 2

\[ F_1(x, y) = -[1 + (A_1 + B_1)^2 + (A_2 + B_2)^2] \quad F_2(x, y) = -[(x + 3)^2 + (y + 1)^2] \]

\[ A_i = \sum_{j=1}^{n} (a_{i,j} \cdot \sin(\alpha_j) + b_{i,j} \cdot \cos(\alpha_j)) \]

\[ B_i = \sum_{j=1}^{n} (a_{i,j} \cdot \sin(\beta_j) + b_{i,j} \cdot \cos(\beta_j)) \]

\[ a = \begin{bmatrix} 0.5 & 1.0 \\ 1.5 & 2.0 \end{bmatrix} \quad b = \begin{bmatrix} -2.0 & -1.5 \\ -1.0 & -0.5 \end{bmatrix} \quad \alpha = \begin{bmatrix} 1.0 & 2.0 \end{bmatrix} \]

\[ \beta = (x, y) \quad \in \quad [-\pi, \pi] \]

Two different mathematical functions
## Statistical Analysis: Example 2

16 Designs computed with Full Factorial

<table>
<thead>
<tr>
<th>Id</th>
<th>x</th>
<th>y</th>
<th>dummy1</th>
<th>dummy2</th>
<th>o1</th>
<th>o2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-3.14</td>
<td>-3.14</td>
<td>-100</td>
<td>-100</td>
<td>-9.45804</td>
<td>-4.5992</td>
</tr>
<tr>
<td>1</td>
<td>-3.14</td>
<td>3.14</td>
<td>-100</td>
<td>-100</td>
<td>-9.45883</td>
<td>-17.1592</td>
</tr>
<tr>
<td>2</td>
<td>3.14</td>
<td>-3.14</td>
<td>-100</td>
<td>-100</td>
<td>-9.45444</td>
<td>-42.2792</td>
</tr>
<tr>
<td>3</td>
<td>3.14</td>
<td>3.14</td>
<td>-100</td>
<td>-100</td>
<td>-9.4553</td>
<td>-54.8392</td>
</tr>
<tr>
<td>4</td>
<td>-3.14</td>
<td>-3.14</td>
<td>-100</td>
<td>100</td>
<td>-9.45804</td>
<td>-4.5992</td>
</tr>
<tr>
<td>5</td>
<td>-3.14</td>
<td>3.14</td>
<td>-100</td>
<td>100</td>
<td>-9.45883</td>
<td>-17.1592</td>
</tr>
<tr>
<td>6</td>
<td>3.14</td>
<td>-3.14</td>
<td>-100</td>
<td>100</td>
<td>-9.45444</td>
<td>-42.2792</td>
</tr>
<tr>
<td>7</td>
<td>3.14</td>
<td>3.14</td>
<td>-100</td>
<td>100</td>
<td>-9.4553</td>
<td>-54.8392</td>
</tr>
<tr>
<td>8</td>
<td>-3.14</td>
<td>-3.14</td>
<td>100</td>
<td>-100</td>
<td>-9.45804</td>
<td>-4.5992</td>
</tr>
<tr>
<td>9</td>
<td>-3.14</td>
<td>3.14</td>
<td>100</td>
<td>-100</td>
<td>-9.45883</td>
<td>-17.1592</td>
</tr>
<tr>
<td>10</td>
<td>3.14</td>
<td>-3.14</td>
<td>100</td>
<td>-100</td>
<td>-9.45444</td>
<td>-42.2792</td>
</tr>
<tr>
<td>11</td>
<td>3.14</td>
<td>3.14</td>
<td>100</td>
<td>-100</td>
<td>-9.4553</td>
<td>-54.8392</td>
</tr>
<tr>
<td>12</td>
<td>-3.14</td>
<td>-3.14</td>
<td>100</td>
<td>100</td>
<td>-9.45804</td>
<td>-4.5992</td>
</tr>
<tr>
<td>13</td>
<td>-3.14</td>
<td>3.14</td>
<td>100</td>
<td>100</td>
<td>-9.45883</td>
<td>-17.1592</td>
</tr>
</tbody>
</table>

**Initial input variables**

**Added input variables**

**Results**
Statistical Analysis: Example 2

dummy1 and dummy2 have significance 0.5 in both functions.

Hint: "The number of design variables can be reduced."

Full factorial (or reduced factorial) gives a complete information on variables if t-Student test is used.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Effect Size</th>
<th>Significance</th>
<th>t-Student</th>
<th>critical t (α = 0.10)</th>
<th>critical t (α = 0.05)</th>
<th>critical t (α = 0.01)</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>3.565E-3</td>
<td>0.000</td>
<td>1.6194E1</td>
<td>1.4149E0</td>
<td>1.8946E0</td>
<td>2.9590E0</td>
</tr>
<tr>
<td>y</td>
<td>-4.230E-4</td>
<td>0.000</td>
<td>8.636E-1</td>
<td>1.4149E0</td>
<td>1.8946E0</td>
<td>2.9590E0</td>
</tr>
<tr>
<td>dummy1</td>
<td>0.00000E0</td>
<td>0.500</td>
<td>1.3450E0</td>
<td>1.7613E0</td>
<td>2.6245E0</td>
<td></td>
</tr>
<tr>
<td>dummy2</td>
<td>0.00000E0</td>
<td>0.500</td>
<td>0.00000E0</td>
<td>1.7613E0</td>
<td>2.6245E0</td>
<td></td>
</tr>
</tbody>
</table>
### Statistical Analysis: Example 2

16 Designs computed with Random DOE

<table>
<thead>
<tr>
<th>Id</th>
<th>x</th>
<th>y</th>
<th>dummy1</th>
<th>dummy2</th>
<th>o1</th>
<th>o2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-1.83551</td>
<td>-1.05044</td>
<td>46.17462</td>
<td>-17.9918</td>
<td>-42.0142</td>
<td>-1.35857</td>
</tr>
<tr>
<td>1</td>
<td>2.912641</td>
<td>2.762534</td>
<td>93.55936</td>
<td>-98.7799</td>
<td>-9.76235</td>
<td>-49.116</td>
</tr>
<tr>
<td>2</td>
<td>-0.64596</td>
<td>-0.95748</td>
<td>89.43894</td>
<td>87.41874</td>
<td>-61.4466</td>
<td>-5.5433</td>
</tr>
<tr>
<td>3</td>
<td>-2.41208</td>
<td>1.699224</td>
<td>-41.1941</td>
<td>1.290129</td>
<td>-4.06747</td>
<td>-7.63147</td>
</tr>
<tr>
<td>4</td>
<td>-0.76467</td>
<td>-2.2626</td>
<td>31.9732</td>
<td>-68.6569</td>
<td>-39.9631</td>
<td>-6.59087</td>
</tr>
<tr>
<td>5</td>
<td>-3.1086</td>
<td>0.145397</td>
<td>38.9939</td>
<td>61.05611</td>
<td>-5.06738</td>
<td>-1.32373</td>
</tr>
<tr>
<td>6</td>
<td>-0.11462</td>
<td>0.279802</td>
<td>48.79488</td>
<td>-71.5972</td>
<td>-32.432</td>
<td>-9.9633</td>
</tr>
<tr>
<td>7</td>
<td>0.774715</td>
<td>-1.97997</td>
<td>15.43154</td>
<td>-59.0159</td>
<td>-22.5719</td>
<td>-15.2088</td>
</tr>
<tr>
<td>8</td>
<td>-2.02205</td>
<td>0.253423</td>
<td>-97.8798</td>
<td>-67.7968</td>
<td>-17.0172</td>
<td>-2.52746</td>
</tr>
<tr>
<td>9</td>
<td>-0.66229</td>
<td>-1.77334</td>
<td>94.77948</td>
<td>-50.9151</td>
<td>-52.725</td>
<td>-6.06293</td>
</tr>
<tr>
<td>10</td>
<td>2.449131</td>
<td>-2.89945</td>
<td>-13.5914</td>
<td>53.3753</td>
<td>-6.8771</td>
<td>-33.3009</td>
</tr>
<tr>
<td>11</td>
<td>-2.38758</td>
<td>0.957482</td>
<td>18.47185</td>
<td>31.0331</td>
<td>-3.9244</td>
<td>-4.20679</td>
</tr>
<tr>
<td>12</td>
<td>-0.78728</td>
<td>-0.23019</td>
<td>96.87969</td>
<td>-58.6559</td>
<td>-49.5447</td>
<td>-5.48876</td>
</tr>
<tr>
<td>13</td>
<td>0.026065</td>
<td>3.133719</td>
<td>-33.2733</td>
<td>-11.3511</td>
<td>-13.0261</td>
<td>-26.2447</td>
</tr>
</tbody>
</table>

- **Initial input variables**
- **Added input variables**
- **Results**
Random DOE does not provide reasonable coverage of the experiments space unless the number of samples is large enough to cover uniformly the variables range.

The variable significances are not correct.