a

Chapter 6 Extension

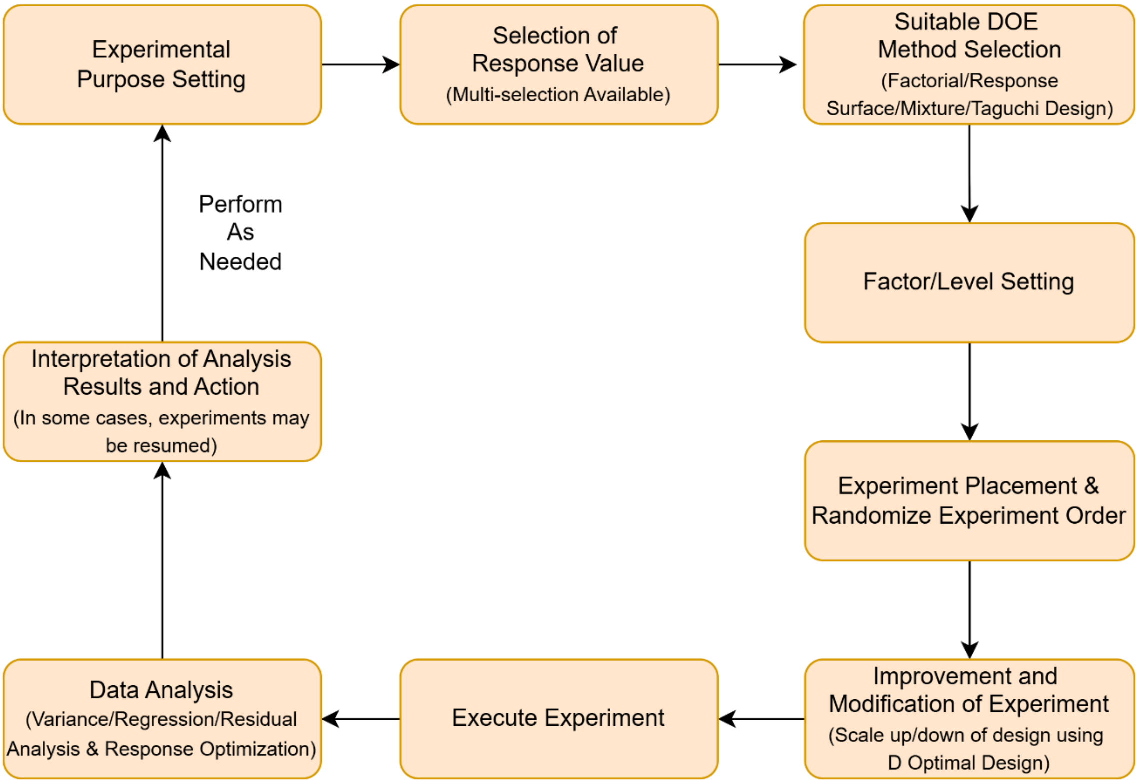
6.1 DOE

6.1.1 Getting Started with DOE

6.1.1.1 Design of Experiment (DOE)

Design of Experiment (DOE) refers to the design methods for experiments. Experimental design is the process of planning an experiment including how to collect data, selecting appropriate statistical methods for analysis, and controlling bias and variability.

DOE is as the following procedures.

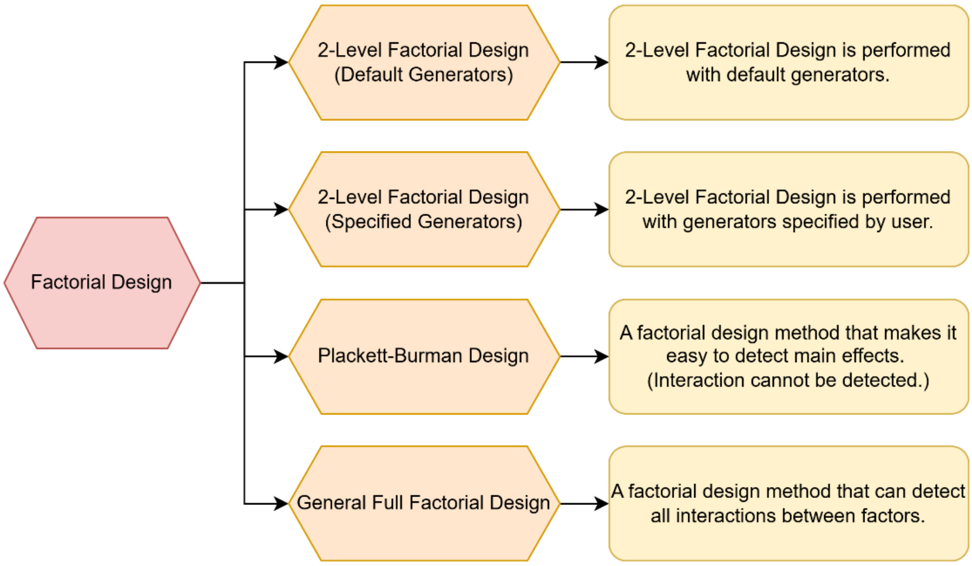


First, choose the response variable and factors, then select suitable DOE method with defined factors and levels. Experiments are conducted following the experiment plan table, and analyze data by the methods such as analysis of variance, regression analysis, residual analysis, and response optimizer methods.

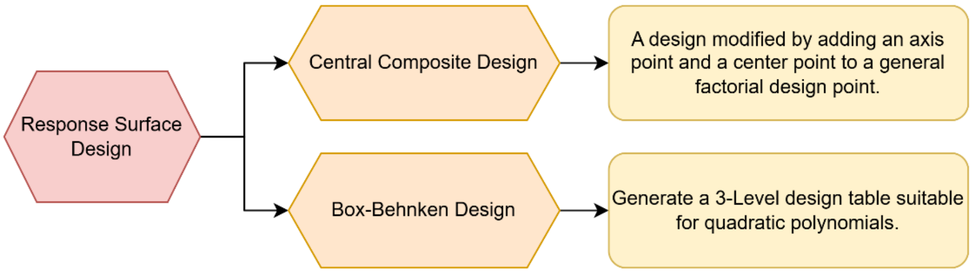
6.1.1.2 ECMiner™ DOE

ECMiner™ DOE includes Factorial Design, Response Surface Design, Mixture Design, and Taguchi Design methods.

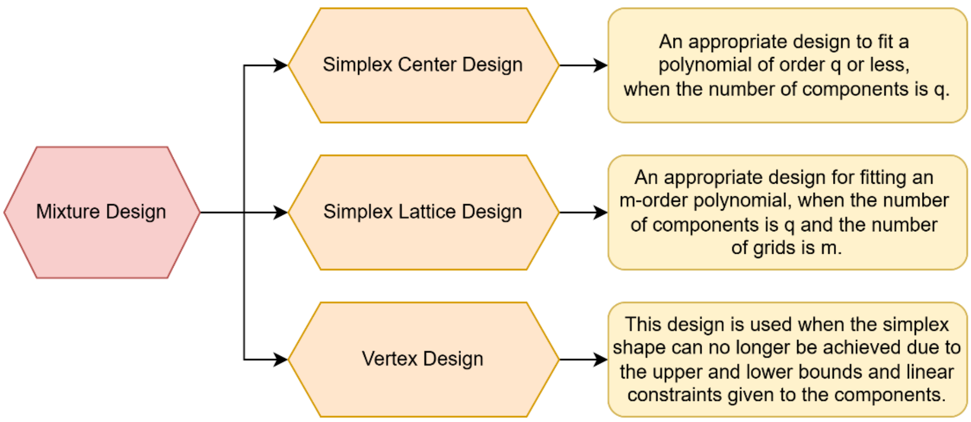
- **Factorial Design**



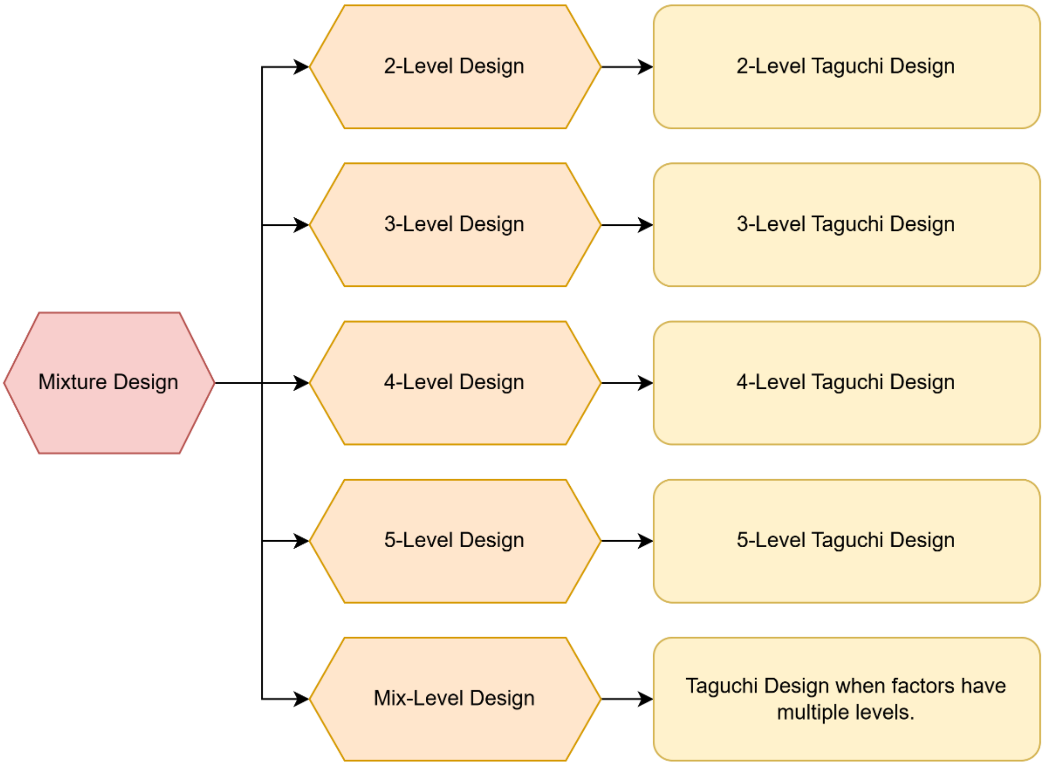
- **Response Surface Design**



- **Mixture Design**



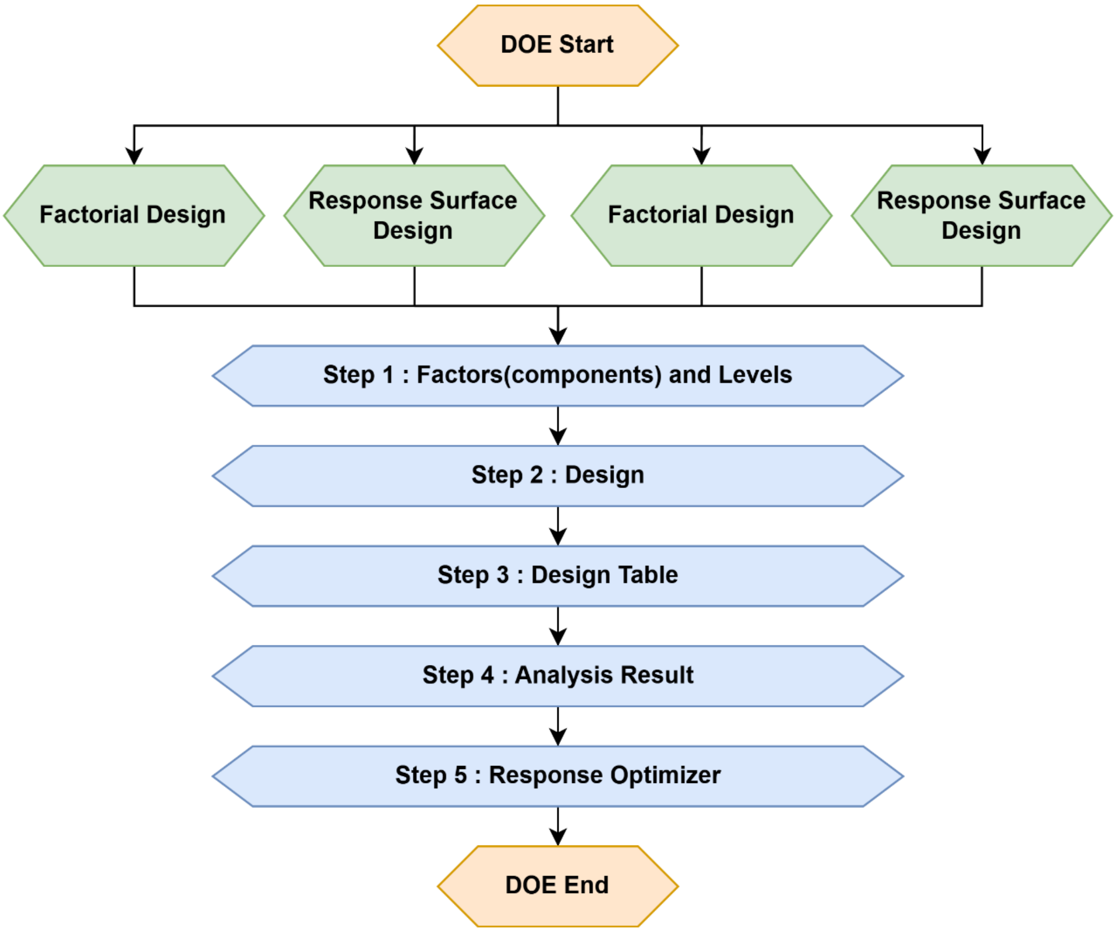
- **Taguchi Design**



6.1.2 Structure of ECMiner™ DOE

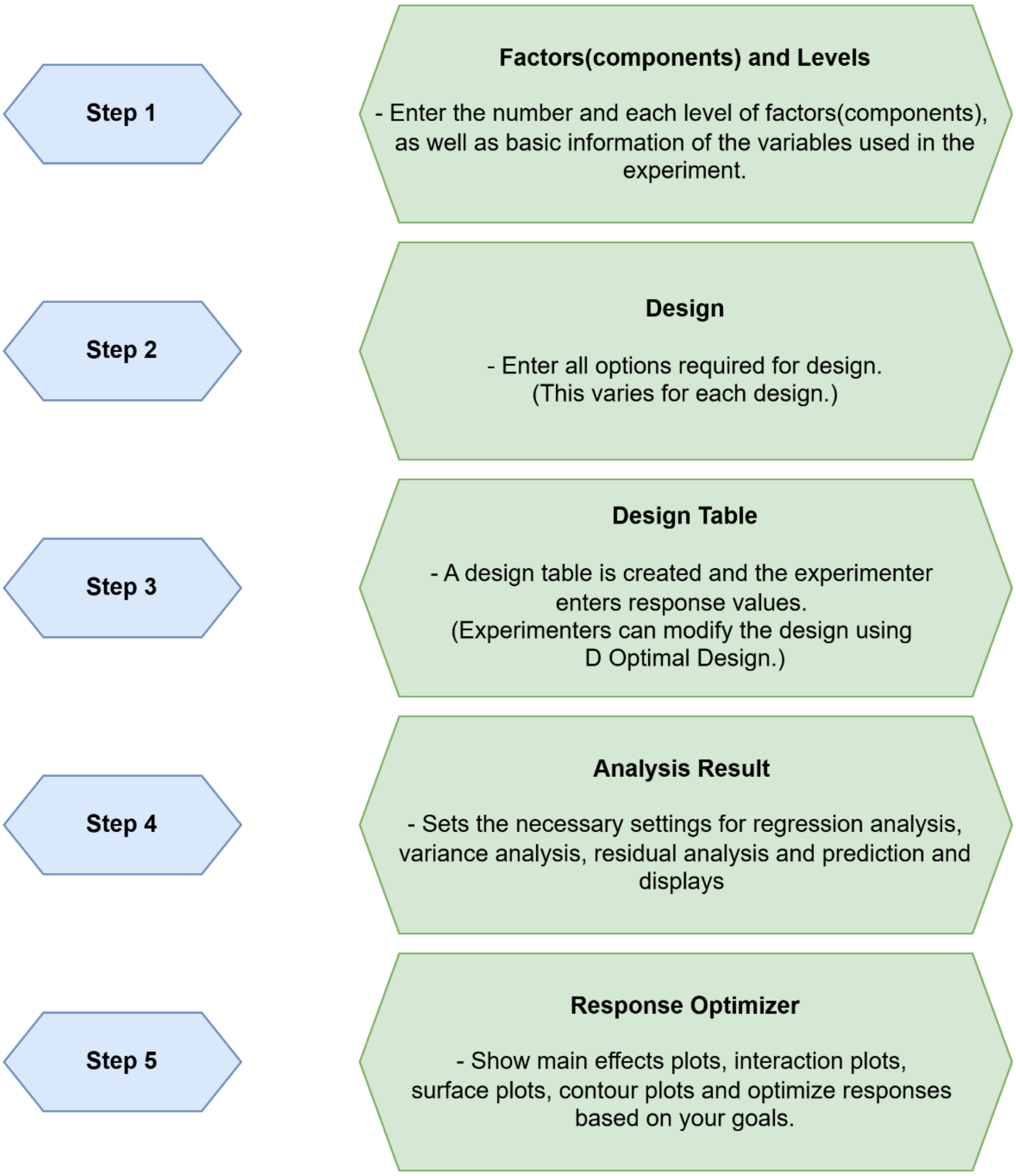
6.1.2.1 Structure of DOE

ECMiner™ DOE include Factorial Design, Response Surface Design, Mixture Design, and Taguchi Design.



6.1.2.2 DOE configuration

ECMiner™ DOE is divided into 5 steps by design methods.



6.1.3 DOE Methods

6.1.3.1 Factorial Design

There are two main types of factorial design. The one is 2-Level Factorial Design and the other is the General Full Factorial Design. 2-Level Factorial Design refers to an experiment with 2-level factors. However, even if each factor has two level values, if the experiment is performed at all grid points, 2^n (n is the number of factors) experiments must be performed. Therefore, 2-Level Factorial Design provides a way to reduce the number of such experiments to suit your purposes. There are three types of 2-Level Factorial Design provided by ECMiner™ as follows.

- **2-Level Factorial Design (Default Generators)**
- **2-Level Factorial Design (Specified Generators)**
- **Plackett-Burman Design**

ECMiner™ provides the General Full Factorial Design, which is convenient when the number of levels for each factor is different. In this experiment, because the experiment is performed at all grid points, the significance level of all interactions can be determined.

6.1.3.1.1. 2-Level Factorial Design (Default Generators)

2-level fractional design (default generators) are used for full factorial and fractional factorial design.

Introduction to the experiment

This experiment is conducted to improve the yield of a specific reaction process.

factor A is the reaction time, factor B is the reaction temperature, and factor C is the amount of ingredient. The specific experimental conditions are as follows.

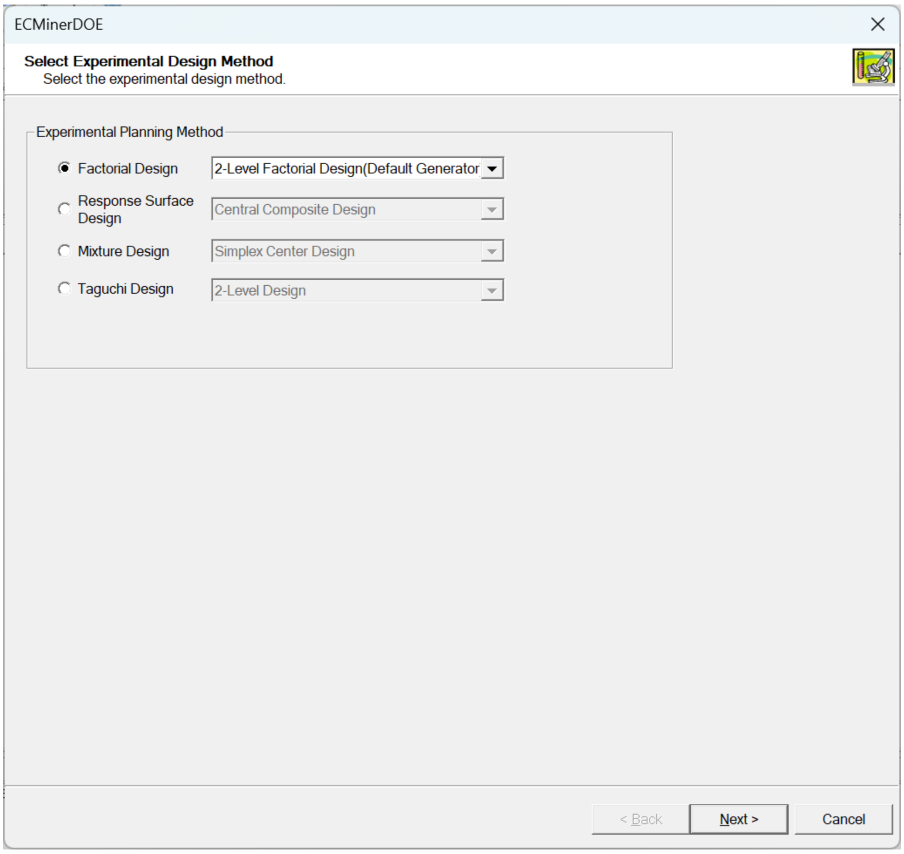
Reaction Time(A): $A_0 = 4, A_1 = 5(hour)$

Reaction Temperature(B): $B_0 = 250, B_1 = 300(℃)$

Amount of Ingredient(C): $C_0 = 3.0, C_1 = 3.5(g)$

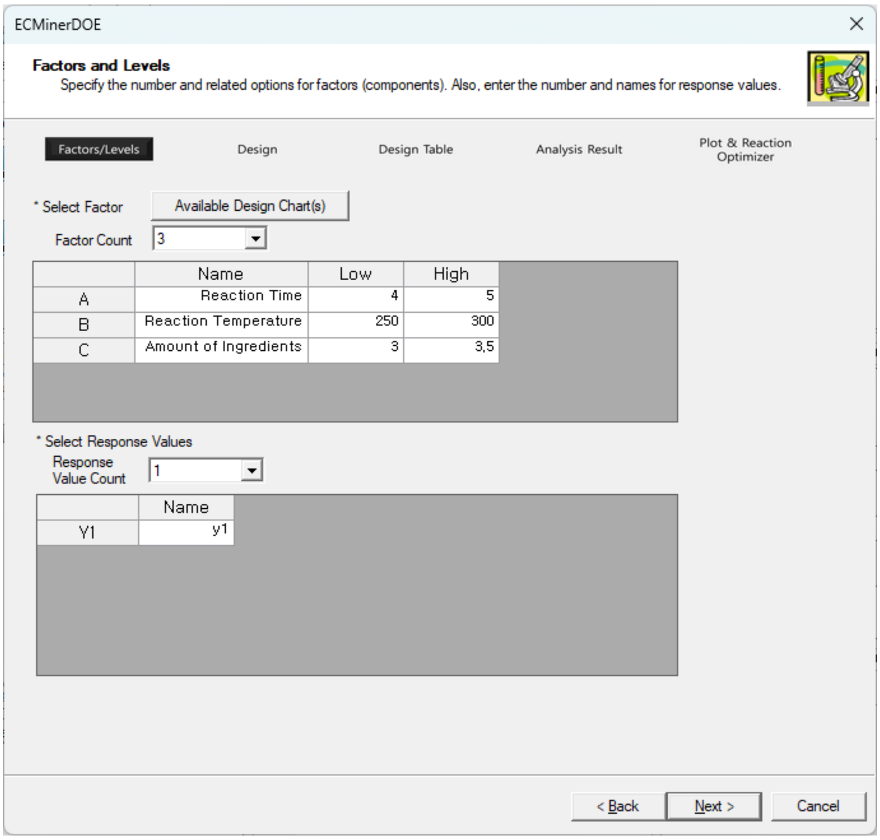
At this time, we cannot complete the experiment in one day, so we would like to experiment by confounding the interaction ABC with the day.

Select 2-Level Factorial Design (**Default Generators**) as follows.

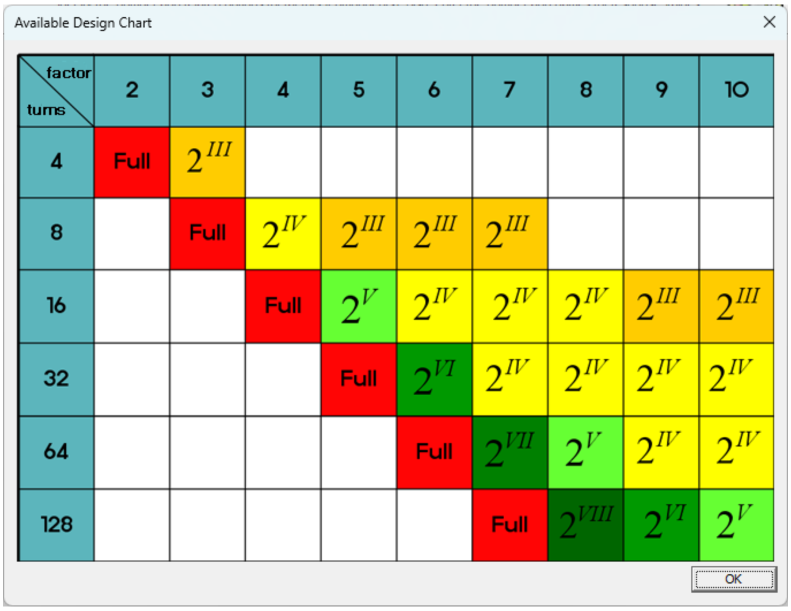


- **Step 1: Factors and Levels**

  Define factor name and enter values for the low and high levels. Set the number of responses and the name of each response value.



By clicking the **Available Design Chart(s)** button, you can see the following screen.



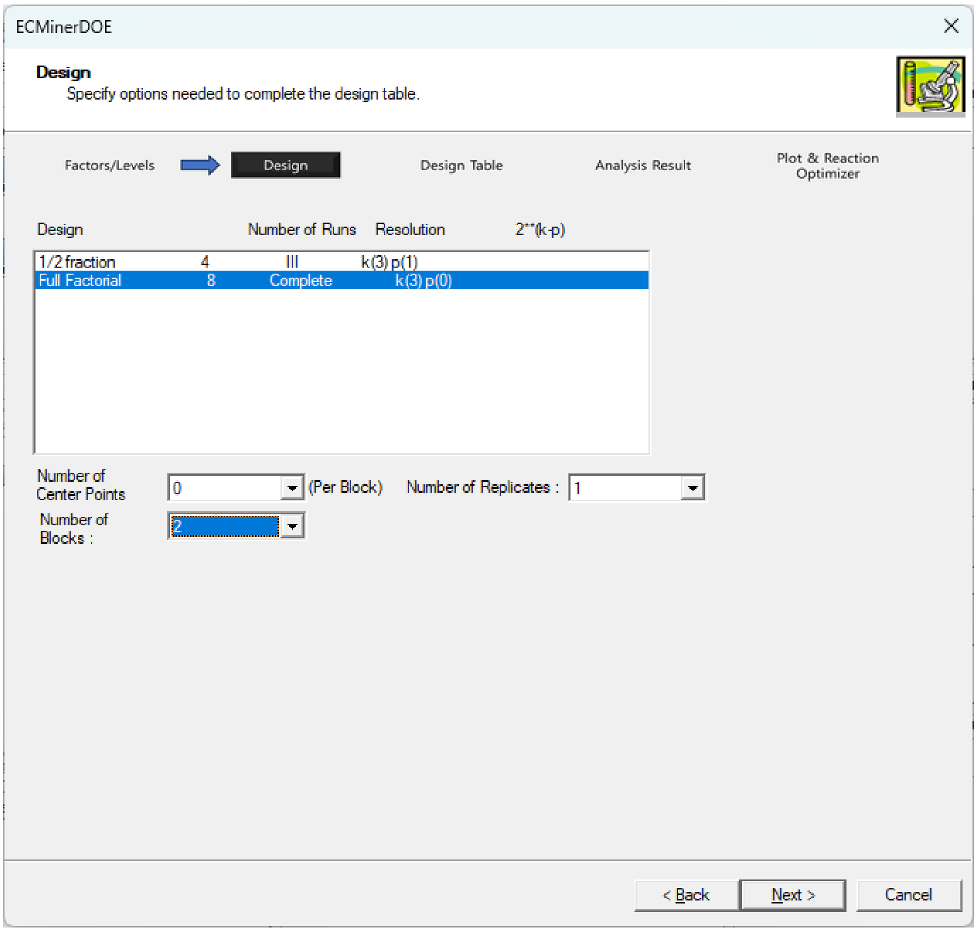
- **Step 2: Design**

  Select fractional factorial design or full factorial design. Set the number of center points, the number of setup repetitions, and number of blocks.

Number of Center Points: Assume a linear relationship between factors and the response in a 2-level factorial design. Add center points to test for non-linear (curved) relationships. Set the center points to the midpoint of each factor in a 2-level factorial design.

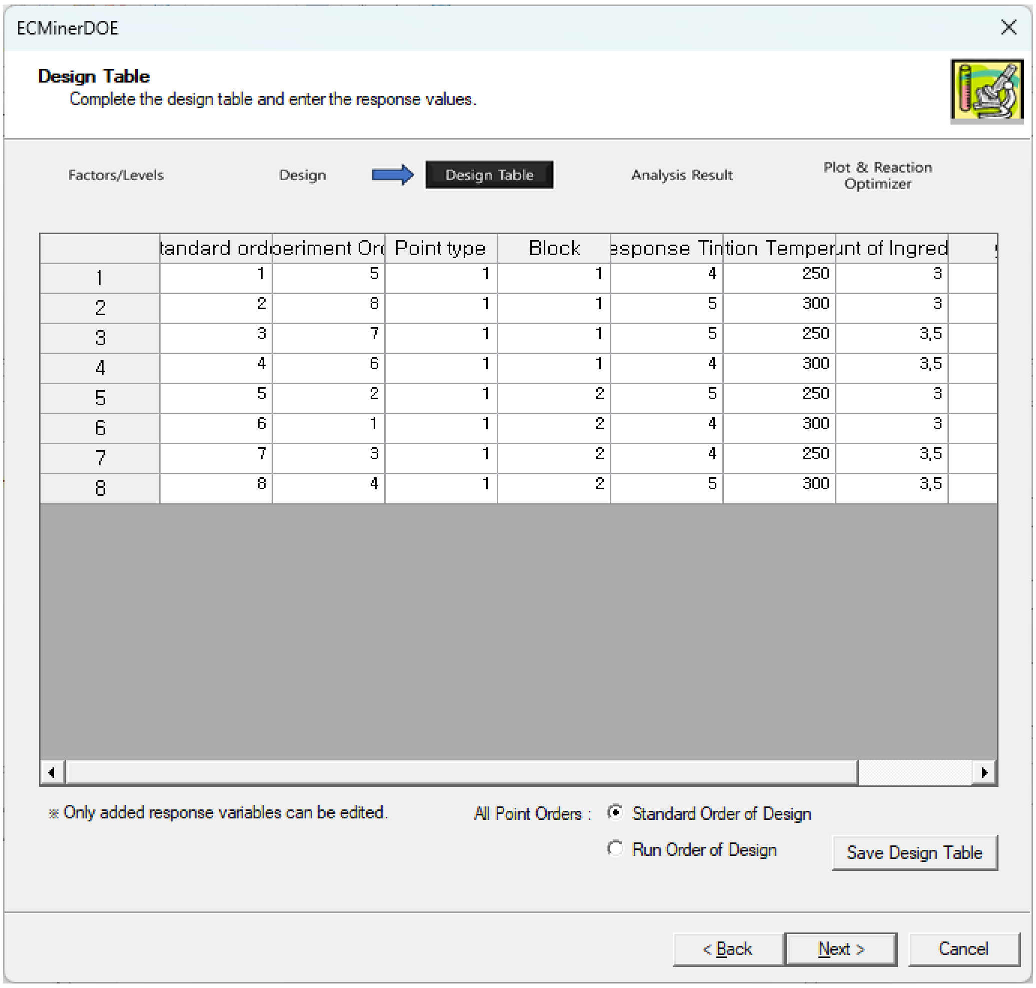
For example:

- **Factor A**: Low = 50, High = 100 → Center = 75.
- **Factor B**: Low = 10, High = 20 → Center = 15.
- **Number of Replicates**: Calculate the amount of pure error by repeating the same experiment.
- **Number of blocks**: Set blocks in factorial design. Use blocking to reduce the impact of nuisance factors (uncontrolled variables) on the experimental results.



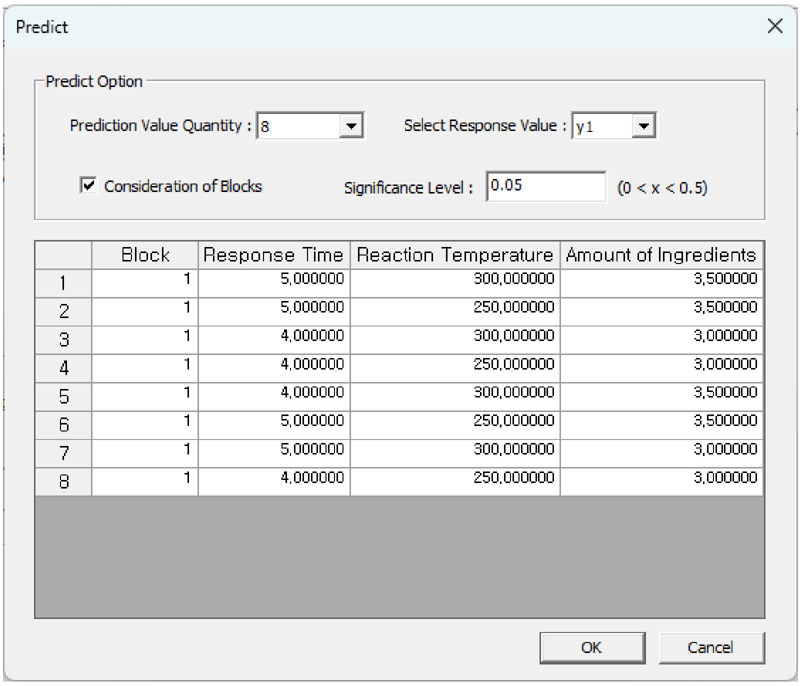
- **Step 3: Design Table**

Design table is generated from design settings. Set the appropriate Response values.

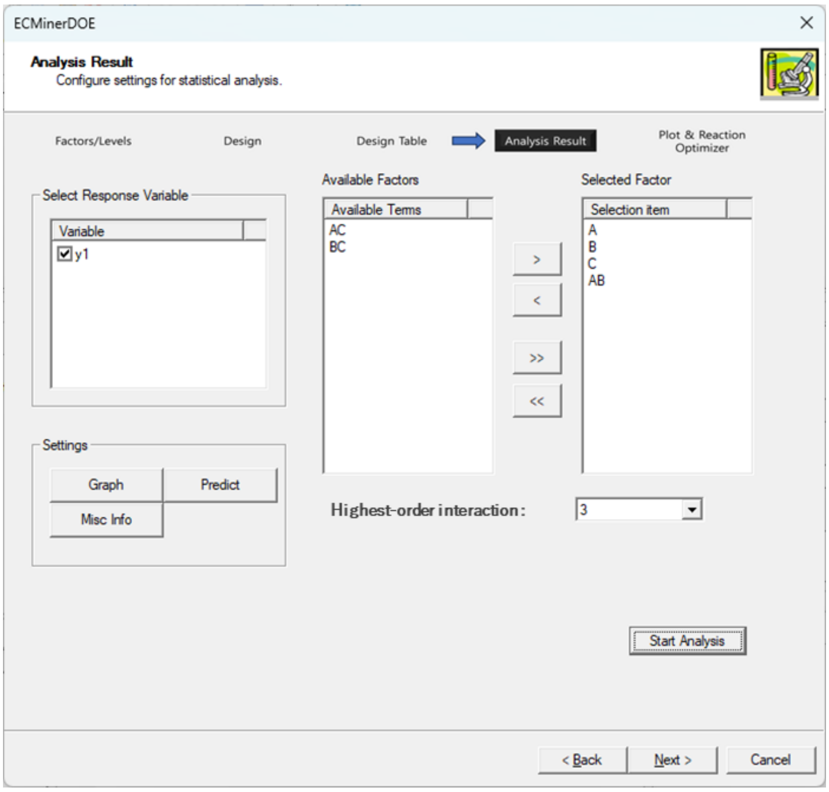


- **Step 4: Analysis Result**

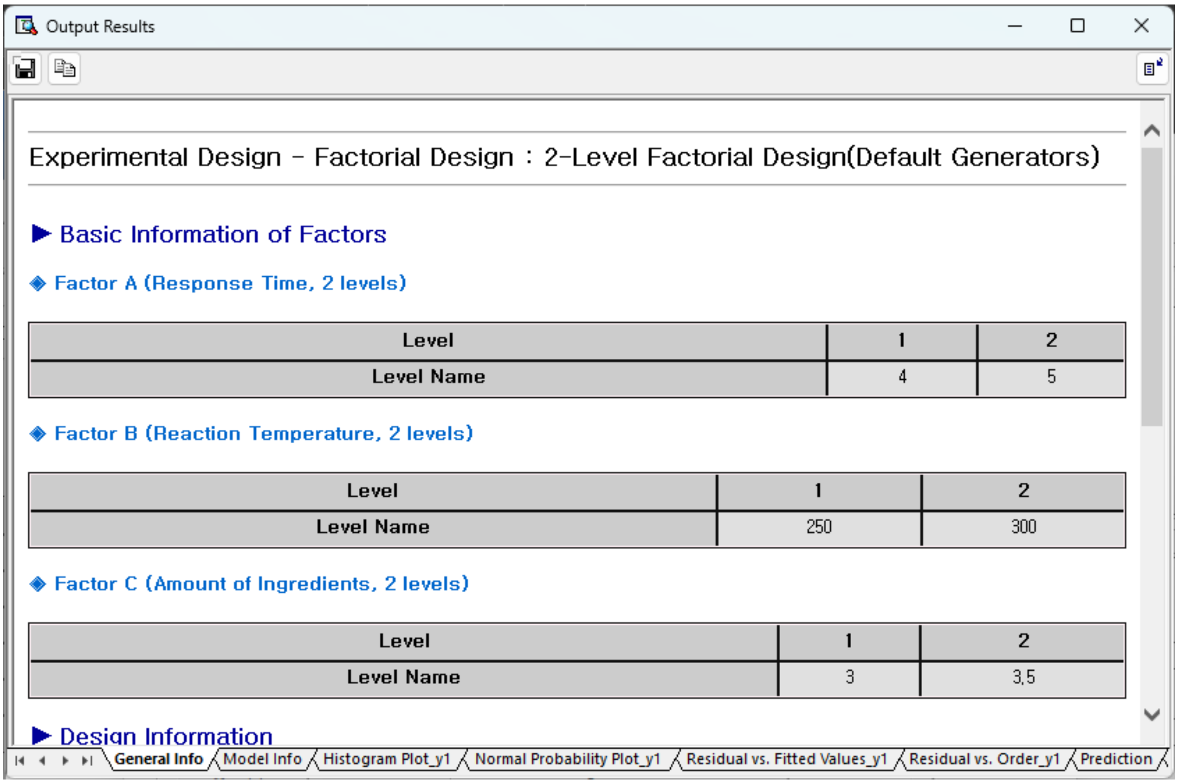
  This step provides the settings necessary for regression analysis, residual analysis, and analysis of variance. First, click the **Graph**, **Predict**, and **Misc Info** buttons to complete the settings. Set the desired number of prediction value quantity and the appropriate values in the Predict option as below.



In the main screen of Step 4, select the maximum degree of the terms to be included. You can select the maximum degree as many as the number of selected factors (3), but since the block is currently created using the block generator called I = ABC, the term ABC cannot be selected. Click the **Start Analysis** button to view the analysis results.

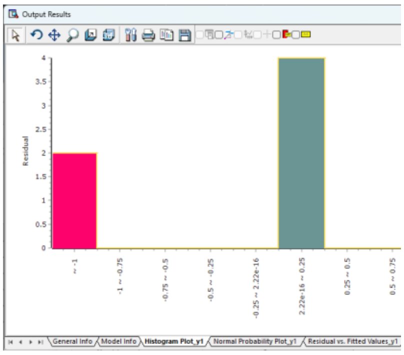
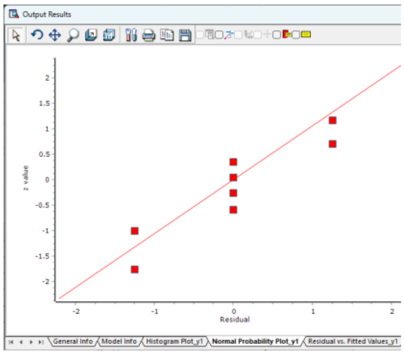
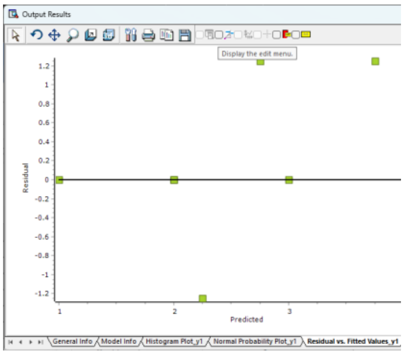
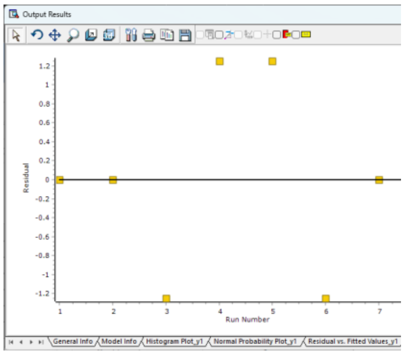


General Info: Basic information, design information, and alias for each factor.



Model Info: View results of regression analysis, analysis of variance, and abnormal observations (extreme leverage, standardized residuals).

Residual Plots: Display Residual Histogram Plot, Normal Probability Plot, Residual vs. Fitted Values, Residual vs. Order.

Prediction: Obtain the predicted value based on the input you specified for prediction.

Output Results

Experimental Design – Factorial Design : 2-Level Factorial Design(Default Generators)

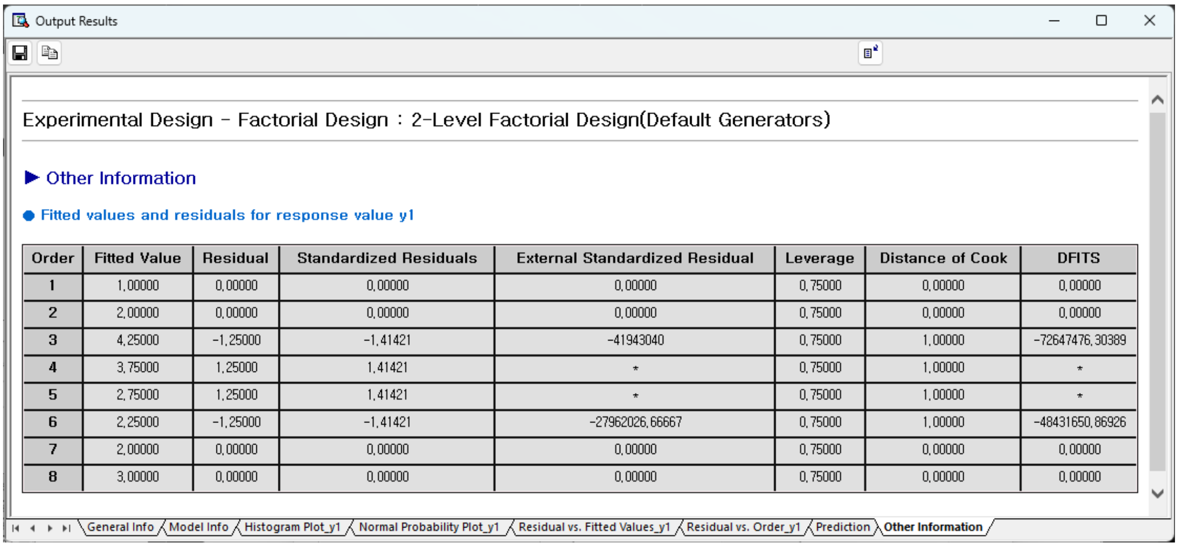
▶ Prediction

● Predictions for response value y1

No.	Block	Response Time	Reaction Temperature	Amount of Ingredients	Fitted Value	SE Fitting Value	95.000000% CI	95.000000% PI
1	1	5	300	3.50000	3.25000	1.53093	(-3.337065, 9.837065)	(-6.811908, 13.311908)
2	1	5	250	3.50000	4.25000	1.53093	(-2.337065, 10.837065)	(-5.811908, 14.311908)
3	1	4	300	3	2.50000	1.53093	(-4.087065, 9.087065)	(-7.561908, 12.561908)
4	1	4	250	3	1	1.53093	(-5.587065, 7.587065)	(-9.061908, 11.061908)
5	1	4	300	3.50000	3.75000	1.53093	(-2.837065, 10.337065)	(-6.311908, 13.811908)
6	1	5	250	3.50000	4.25000	1.53093	(-2.337065, 10.837065)	(-5.811908, 14.311908)
7	1	5	300	3	2	1.53093	(-4.587065, 8.587065)	(-8.061908, 12.061908)
8	1	4	250	3	1	1.53093	(-5.587065, 7.587065)	(-9.061908, 11.061908)

General Info / Model Info / Histogram Plot_y1 / Normal Probability Plot_y1 / Residual vs. Fitted Values_y1 / Residual vs. Order_y1 / Prediction / Other Information

Other Information: residual and several other statistics.


Output Results

Experimental Design - Factorial Design : 2-Level Factorial Design(Default Generators)

▶ Other Information

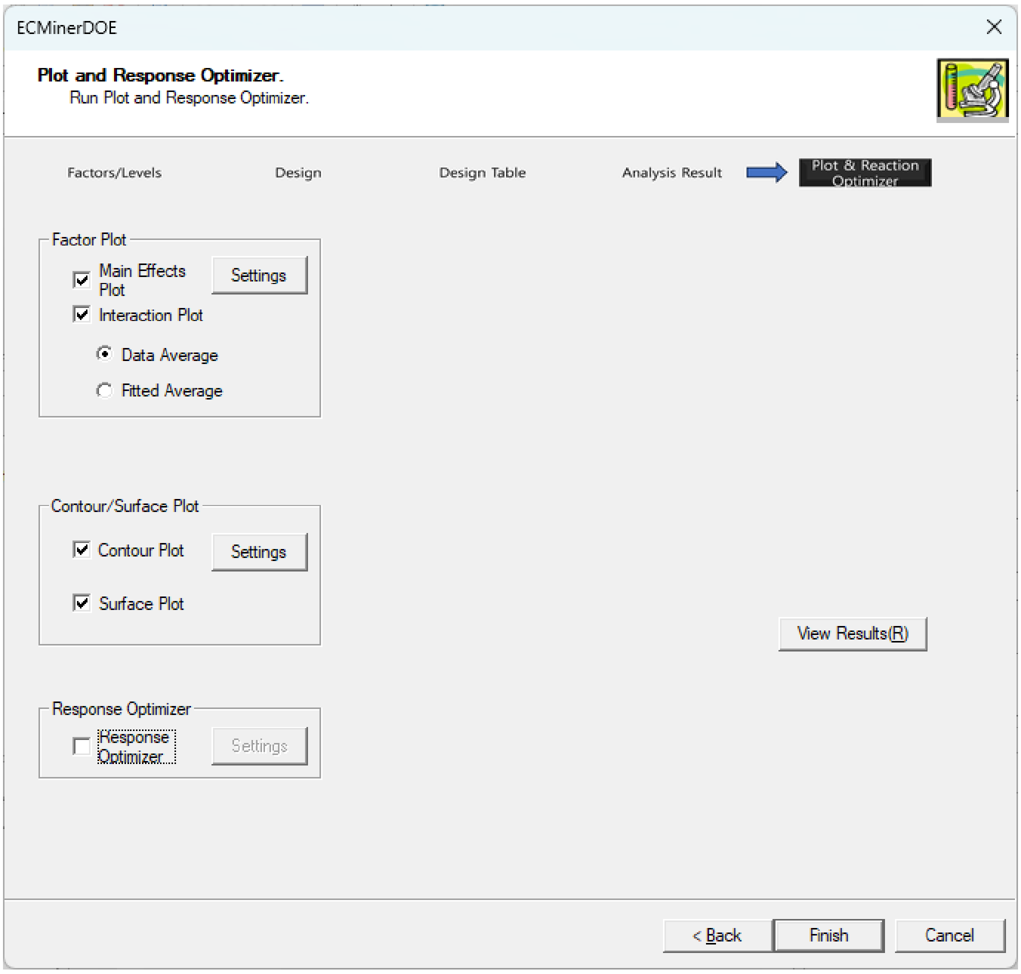
● Fitted values and residuals for response value y1

Order	Fitted Value	Residual	Standardized Residuals	External Standardized Residual	Leverage	Distance of Cook	DFITS
1	1,00000	0,00000	0,00000	0,00000	0,75000	0,00000	0,00000
2	2,00000	0,00000	0,00000	0,00000	0,75000	0,00000	0,00000
3	4,25000	-1,25000	-1,41421	-41943040	0,75000	1,00000	-72647476,30389
4	3,75000	1,25000	1,41421	*	0,75000	1,00000	*
5	2,75000	1,25000	1,41421	*	0,75000	1,00000	*
6	2,25000	-1,25000	-1,41421	-27962026,66667	0,75000	1,00000	-48431650,86926
7	2,00000	0,00000	0,00000	0,00000	0,75000	0,00000	0,00000
8	3,00000	0,00000	0,00000	0,00000	0,75000	0,00000	0,00000

General Info | Model Info | Histogram Plot_y1 | Normal Probability Plot_y1 | Residual vs. Fitted Values_y1 | Residual vs. Order_y1 | Prediction | Other Information

For detailed explanation, see 6.1.4. See Settings and Analysis.

- **Step 5: Plot and Response Optimizer**



In Step 5, there are various types of charts. In response optimizer, an optimization algorithm is performed to obtain the response value desired by the user. The following are the plots by 2-Level Factorial Design (Default Generators).

Main Effect Plot

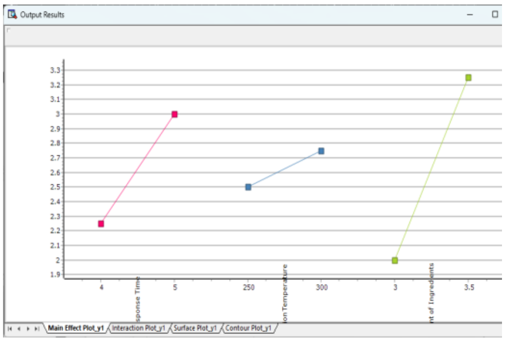
Interaction Plot

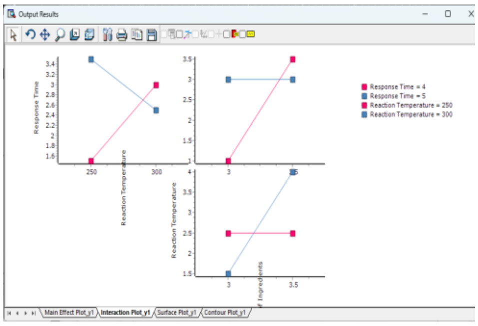
Surface Plot

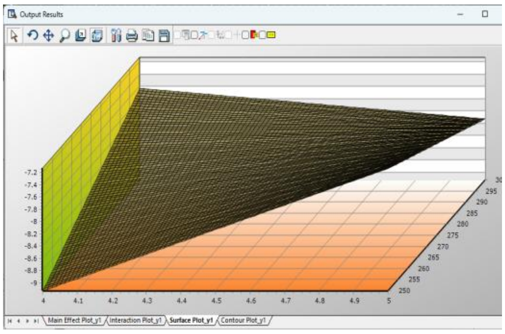
Contour Plot

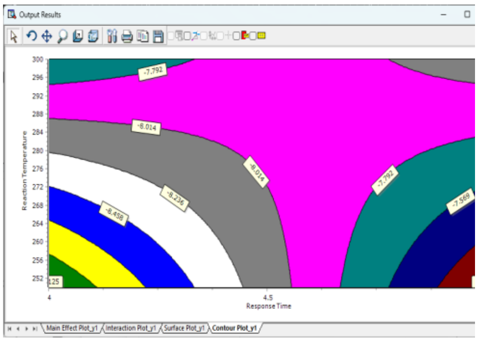
For the main effect plot and interaction plot, first select whether to use the data average or the fitted average. And through the **Setting** button, you can select which factors to plot for.

For a **surface plot** and **contour plot**, you must enter a fixed value for one or more factors when there are three or more factors. Here's an example screen you'll get after making these settings.

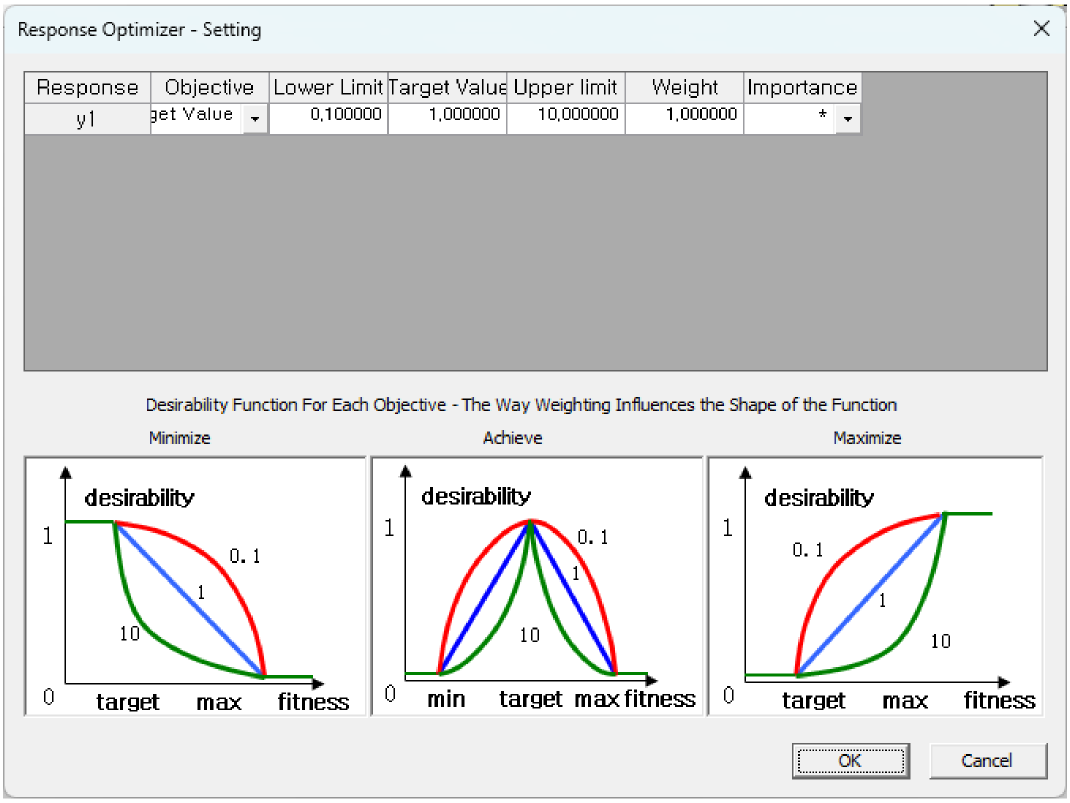




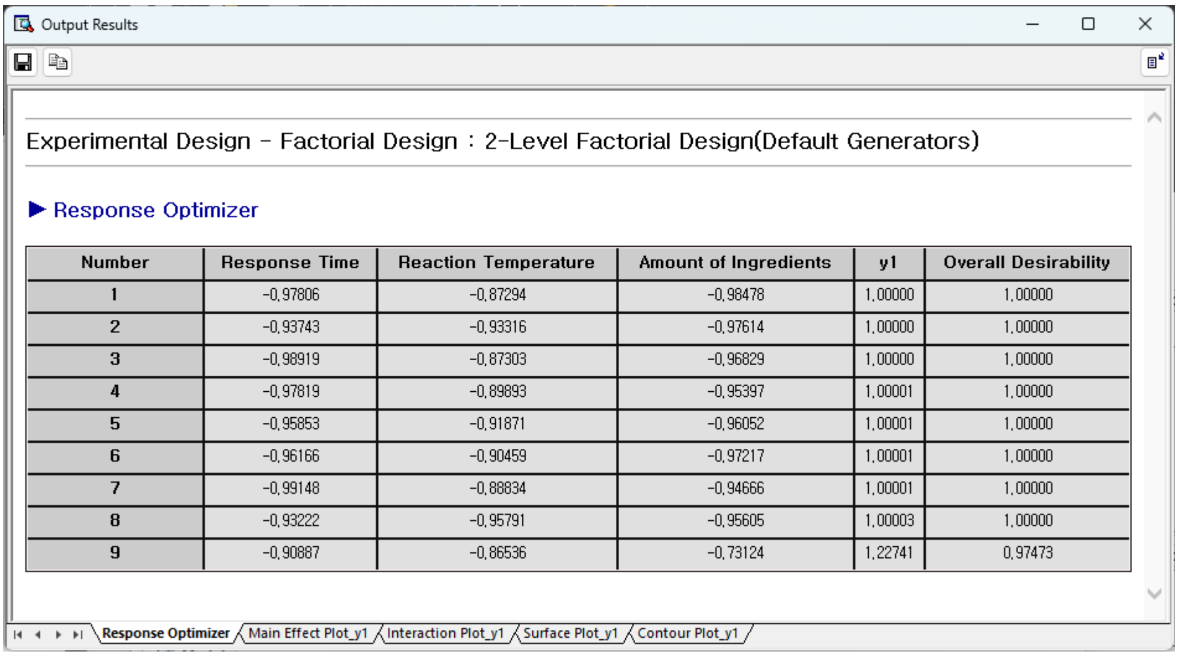




Response Optimizer helps identify the combination of factor settings that optimize the response. First, select the response variable to optimize and enter objective, lower limit, target value, upper limit, weight, and importance in the following settings window.



In the option window of the Response Optimizer, enter the initial value setting method and various settings used for optimization, and then click the **View Results** button on the main screen in Step 5 to obtain the following screen.



Number	Response Time	Reaction Temperature	Amount of Ingredients	y1	Overall Desirability
1	-0.97806	-0.87294	-0.98478	1.00000	1.00000
2	-0.93743	-0.93316	-0.97614	1.00000	1.00000
3	-0.98919	-0.87303	-0.96829	1.00000	1.00000
4	-0.97819	-0.89893	-0.95397	1.00001	1.00000
5	-0.95853	-0.91871	-0.96052	1.00001	1.00000
6	-0.96166	-0.90459	-0.97217	1.00001	1.00000
7	-0.99148	-0.88834	-0.94666	1.00001	1.00000
8	-0.93222	-0.95791	-0.95605	1.00003	1.00000
9	-0.90887	-0.86536	-0.73124	1.22741	0.97473

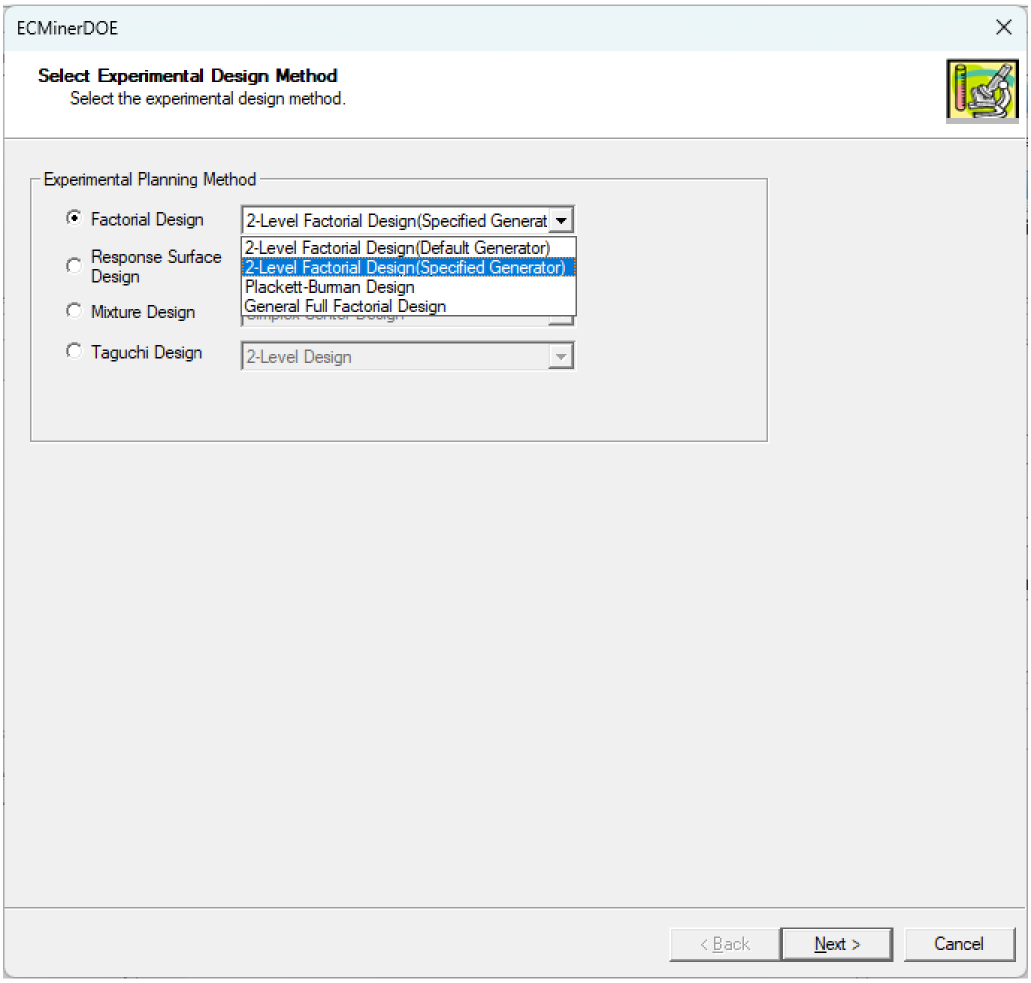
From the above process, you can find out how to determine the level of each factor (Response Time, Response Temperature, Amount) in order to maximize the response value (yield). (At this time, the level of each factor is displayed in coded units.)

6.1.3.1.2. 2-Level Factorial Design (Specified Generators)

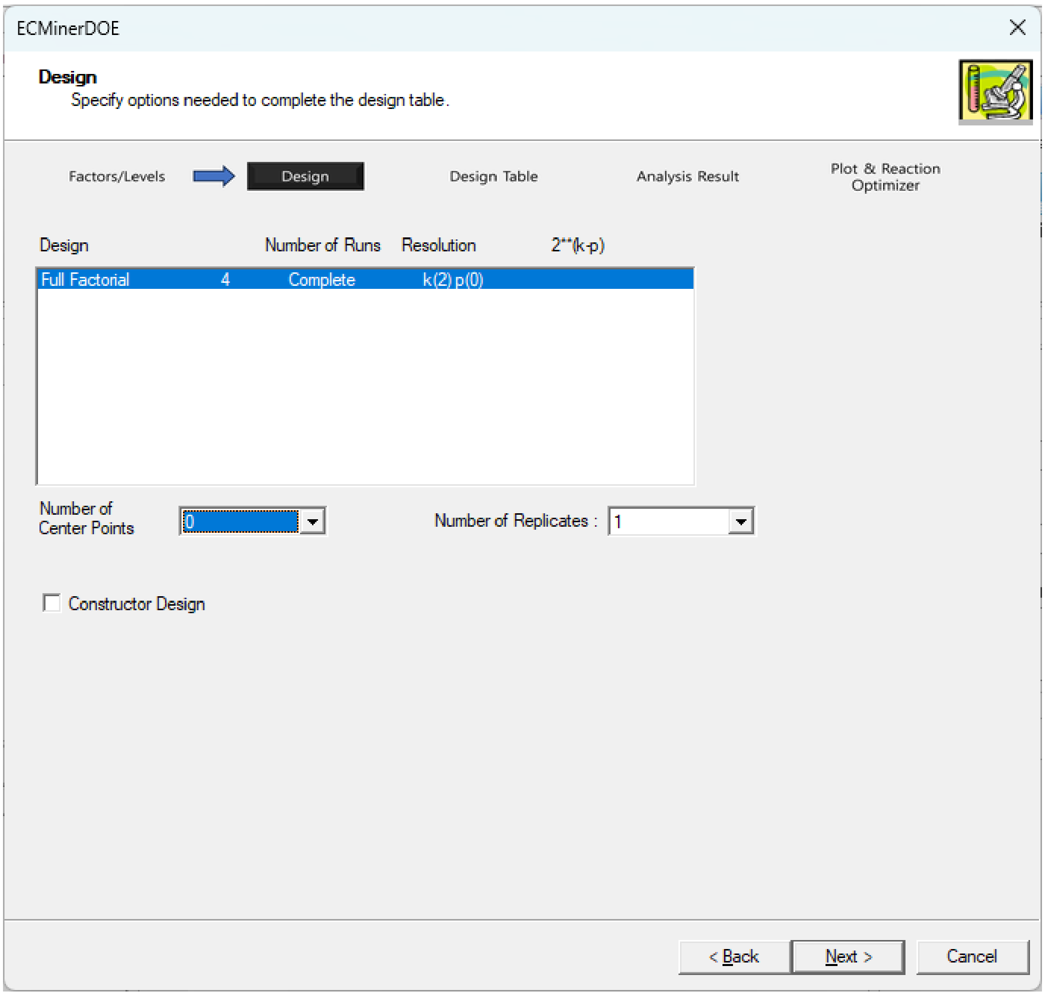
2-Level Factorial Design (Specified Generators) allows to manually define and create a custom factorial design by specifying its factors, levels, and interactions.

This design works when the default designs (e.g., full factorial or fractional factorial) are not appropriate for a specific experimental with specific factor combinations or constraints.

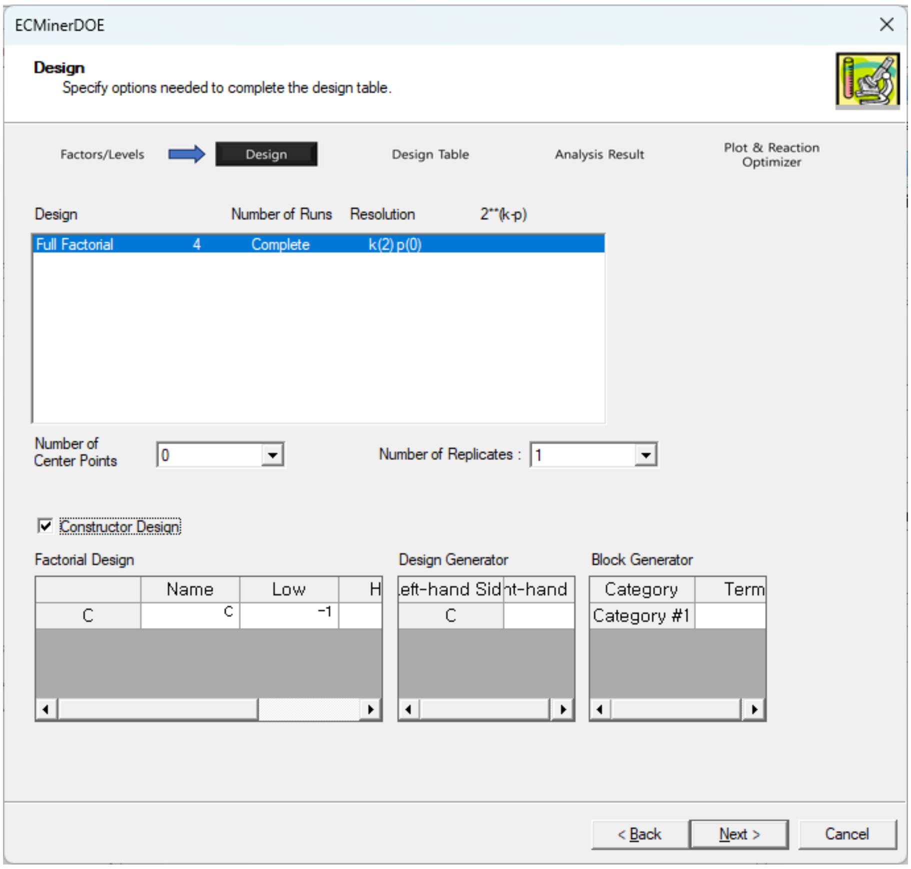
Select 2-Level Factorial Design (Specified Generators).

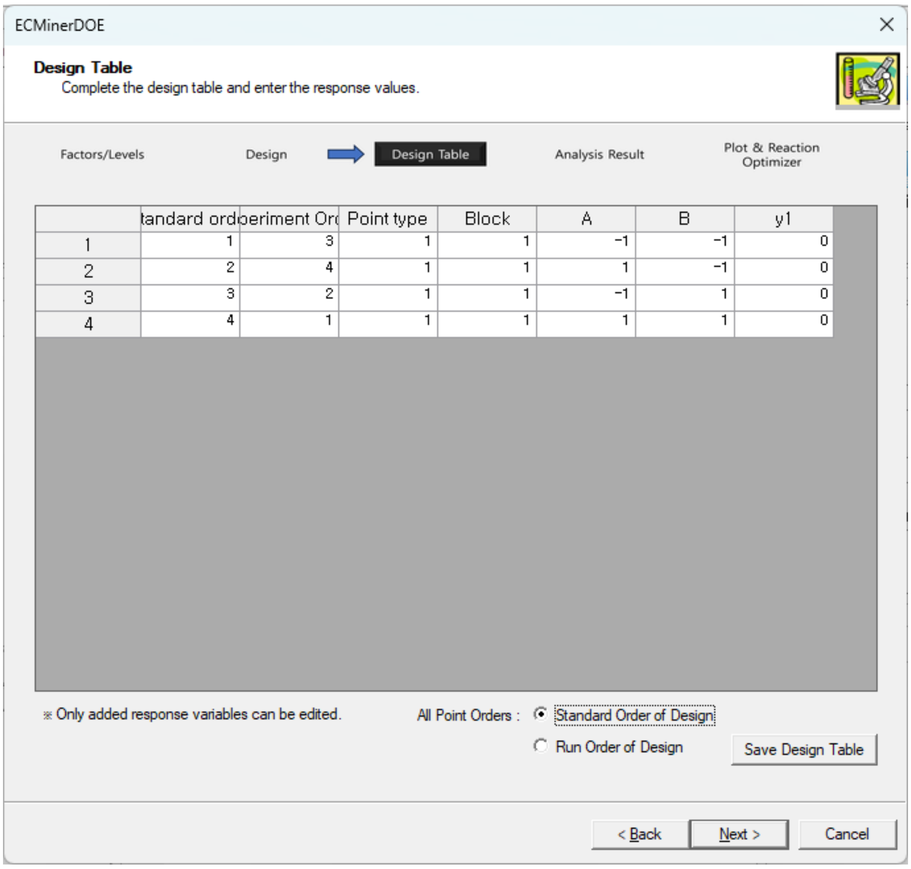


Step 1, enter the number of factors before increasing the design using the design generators. In the above setting, the number of factors is set to 2. Step 2 screen is as follows.



Click the Constructor Design to increase number of factors and complete the part corresponding to the design generators. You can enter block generators depending on your experiment. For example, if you want to increase the number of factors using the generator C=AB, enter C as the left term and AB as the right term as shown above and click the next button.


ECMinerDOE
Design
Specify options needed to complete the design table.
Factors/Levels
Design
Design Table
Analysis Result
Plot & Reaction Optimizer
Design
Number of Runs
Resolution
2**(k-p)
Full Factorial
4
Complete
k(2) p(0)
Number of Center Points
0
Number of Replicates :
1
Constructor Design
Factorial Design
Name
Low
H
C
C
-1
Design Generator
eft-hand Sid
nt-hand
C
Block Generator
Category
Term
Category #1
< Back
Next >
Cancel


ECMinerDOE
Design Table
Complete the design table and enter the response values.
Factors/Levels
Design
Design Table
Analysis Result
Plot & Reaction Optimizer
tandard ord
oeriment Or
Point type
Block
A
B
y1
1 1 3 1 1 -1 -1 0
2 2 4 1 1 1 -1 0
3 3 2 1 1 -1 1 0
4 4 1 1 1 1 1 0
※ Only added response variables can be edited.
All Point Orders :
Standard Order of Design
Run Order of Design
Save Design Table
< Back
Next >
Cancel

You can see that one more column corresponding to factor C has been created, and the value of this column is the same as the product of the value in Column A and the value in Column B. In this way, you can create design tables using different generators depending on your needs. The analysis process after this is the same as that of 2-Level Factorial Design (Default Generators).

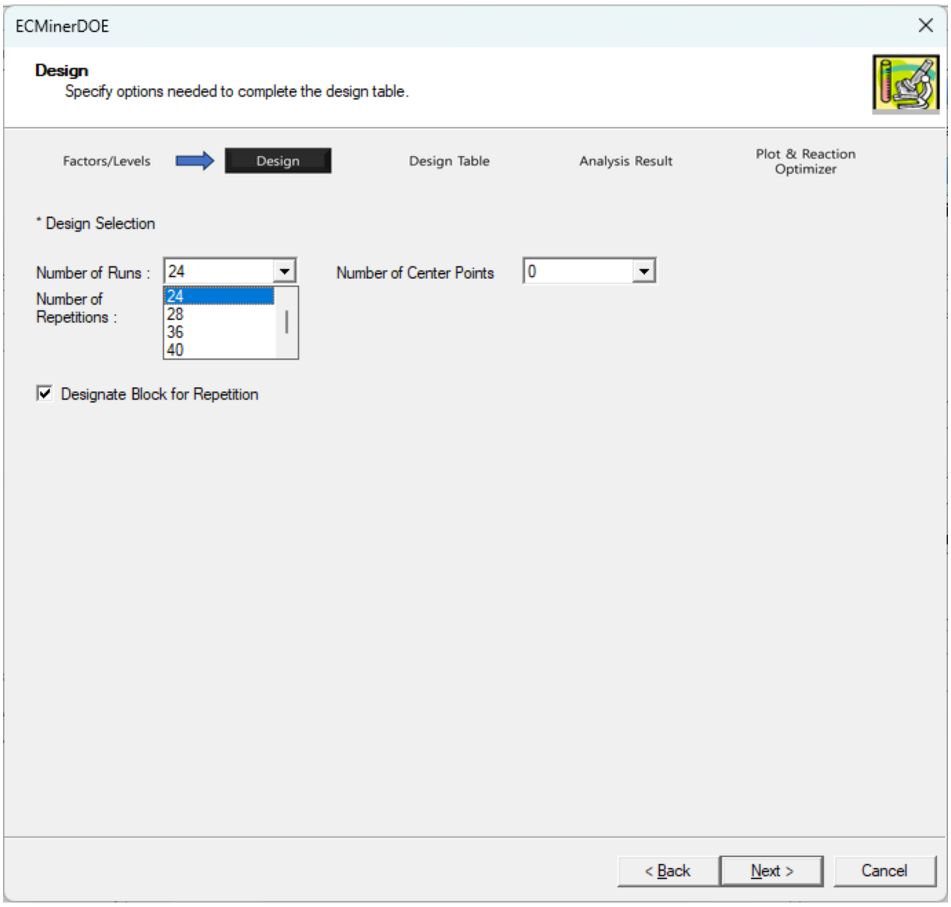
6.1.3.1.3. Plackett-Burman Design

This design is used when the number of factors is large. This design can be used when the user has prior knowledge that the only one factor the response value is the main effect. This prior knowledge helps to dramatically reduce the number of experiments.

Plackett-Burman design has similar post-analysis steps to 2-Level Factorial Design. Therefore, we will mainly explain the differences compared to 2-Level Factorial Design.

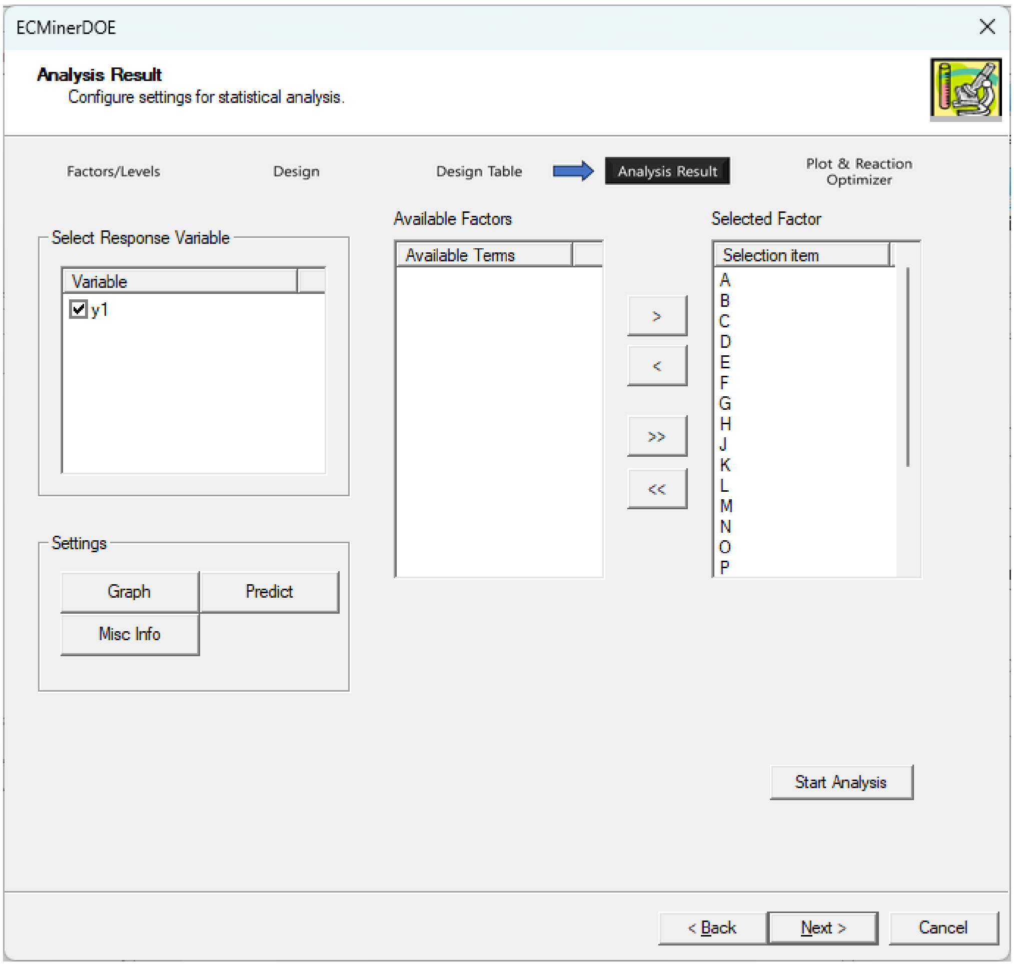
The characteristic of Plackett-Burman is that it dramatically reduces the number of experiments when the experimenter knows in advance that there is only a linear relationship between factors and response values. For example, the main effect can be detected with just 48 experiments for 47 factors. The reason we use the Plackett-Burman design is that it can dramatically reduce the need for experimentation.

Select Plackett-Burman design and select 20 factors on the Step 1 screen. Step 2 Afterwards, if you move to the Step 2 screen, you can select the following number of experiments.



If the experimenter aims to minimize the number of experiments, a design with 24 runs can be selected. The maximum number of runs can be selected up to 48, but 48 experiments for 20 is not particularly burdensome. Set the number of runs and the number of center points and repetition according to your purpose. Center point is used to check curvature, and by increasing repetition, pure error can be checked how much the difference in response value occurs at the same experimental point.

In Step 3, the design table is completed, and in Step 4, the following screen appears.



In the Plackett-Burman design, interactions other than the main effect cannot be detected. If you select a factor and proceed with analysis, the subsequent process is the same as 2-Level Factorial Design (Default Generators).

6.1.3.1.4. General Full Factorial Design

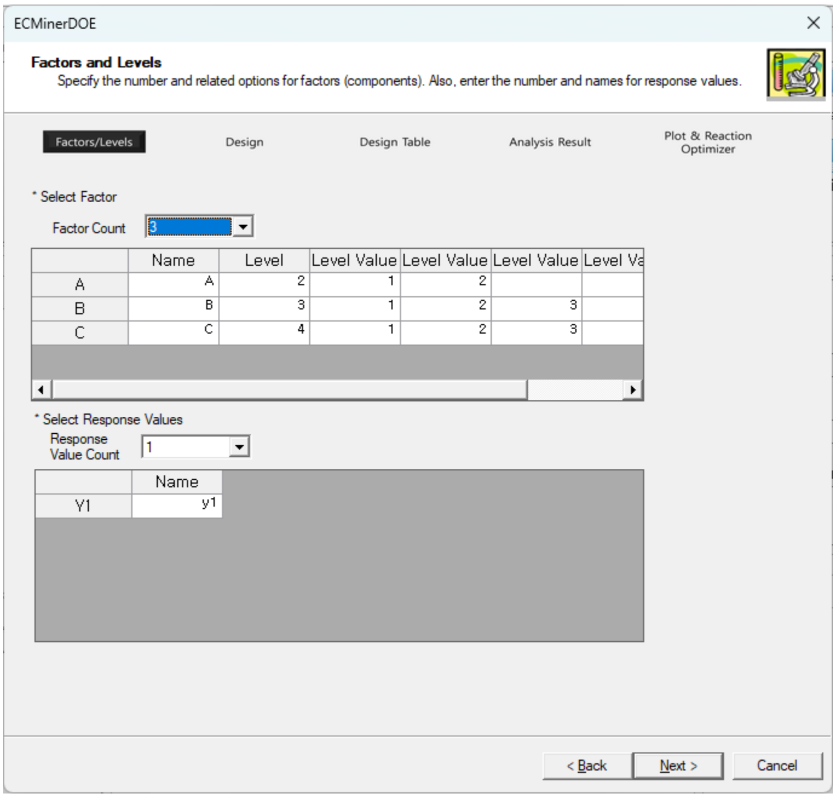
General Full Factorial Design overcomes the limitation that the previous three designs (2-Level Factorial Design (Default Generators), 2-Level Factorial Design (Specified Generators), and Plackett-Burman Design) can only conduct 2-level experiments. DOE is capable of. The level of each factor can have any value. For example, when the number of factors is 3, the number of levels can be 2, 3, and 4, respectively. In General, Full Factorial Design, all possible combinations are tested, so if the number of levels is 2, 3, or 4, you will get the following design table.

A	B	C
1	1	1
2	1	1
1	2	1
2	2	1
1	3	1
2	3	1
1	1	2
2	1	2
1	2	2
2	2	2
1	3	2
2	3	2
1	1	3
2	1	3
1	2	3
2	2	3
1	3	3
2	3	3
1	1	4
2	1	4
1	2	4
2	2	4
1	3	4

2	3	4

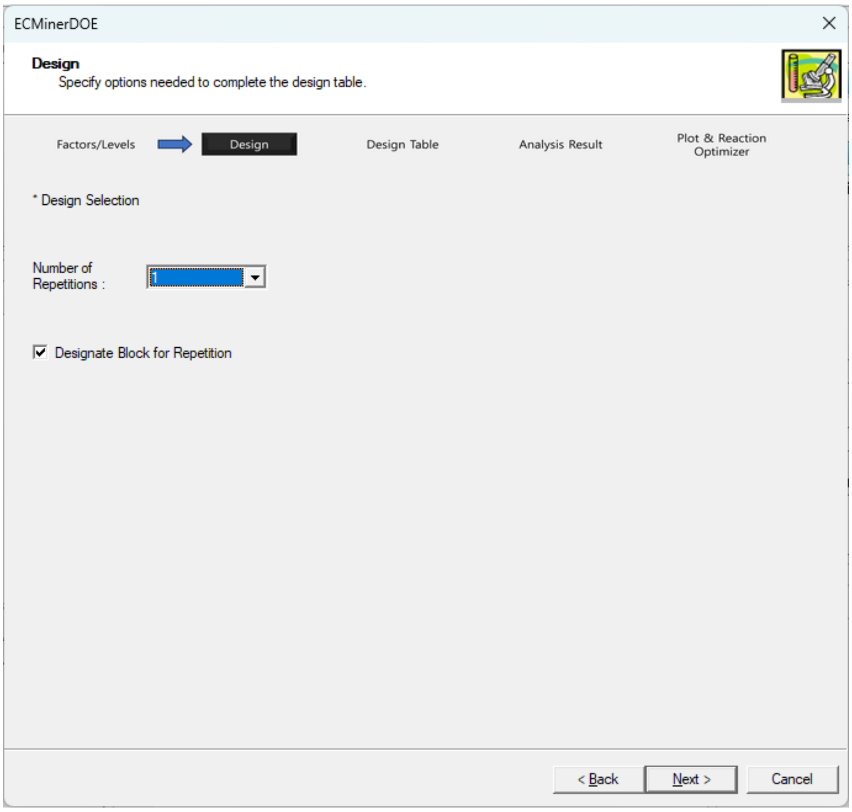
The following is a **Full Factorial Design**

- **Step 1: Factors and Levels**



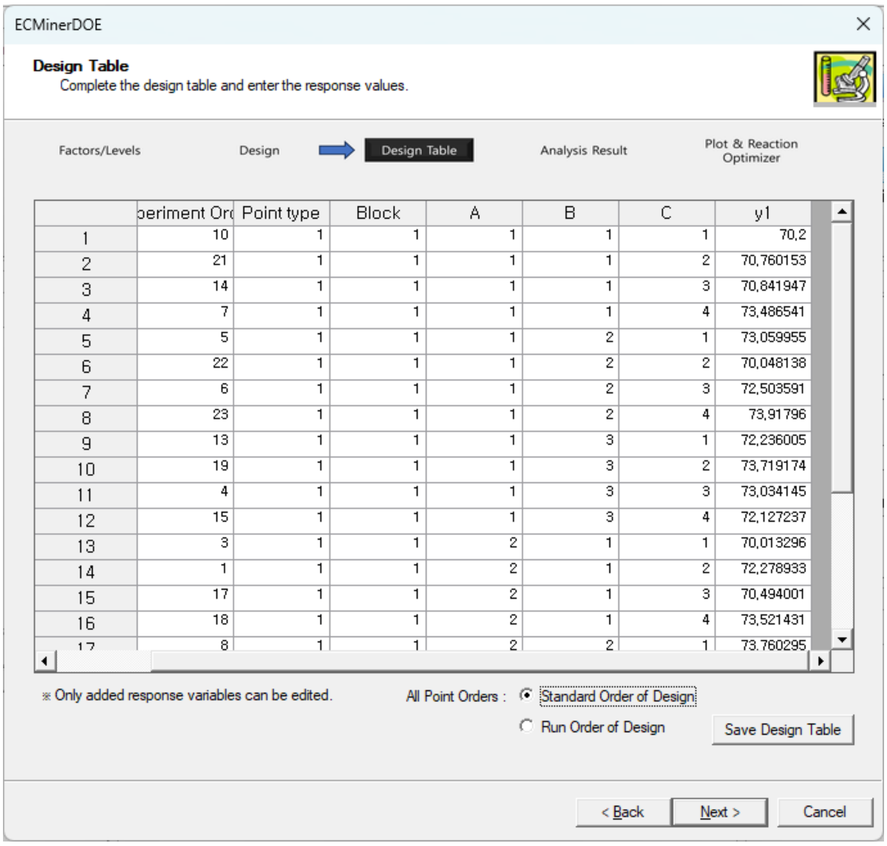
If there are three factors (A, B, C) that can have values of 2-Level, 3-Level, and 4-Level, respectively, enter the setting values as above.

- **Step 2: Design**



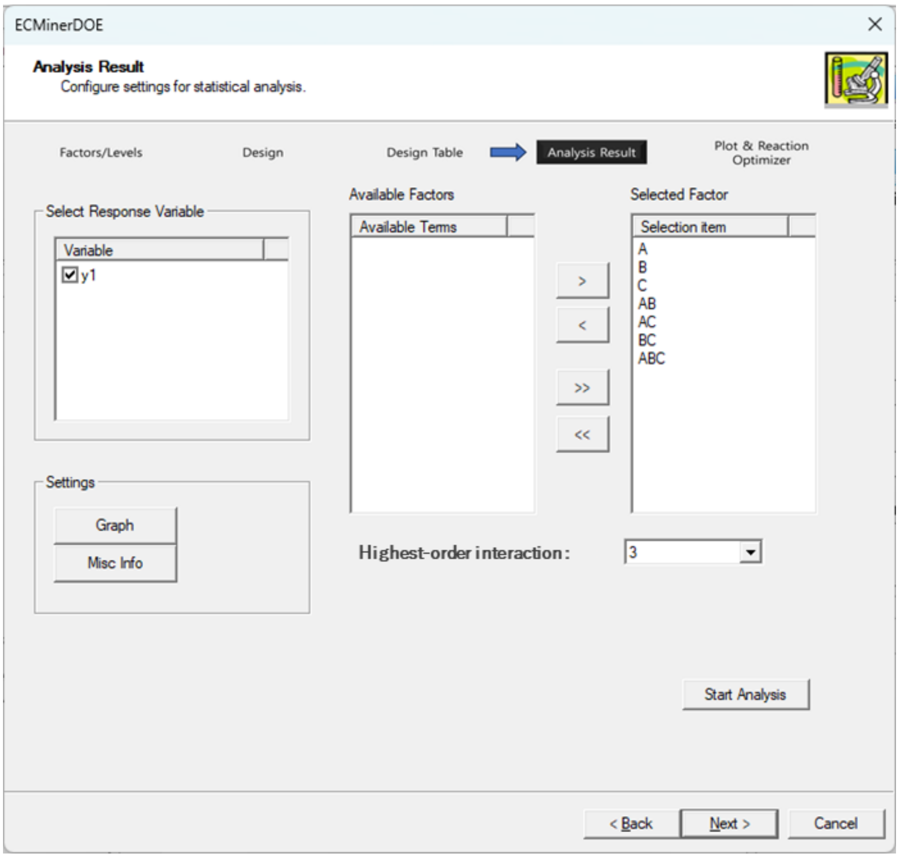
In Design Selection window you determine the number of repetitions and whether you want to assign blocks to repetitions.

- **Step 3: Design Table**

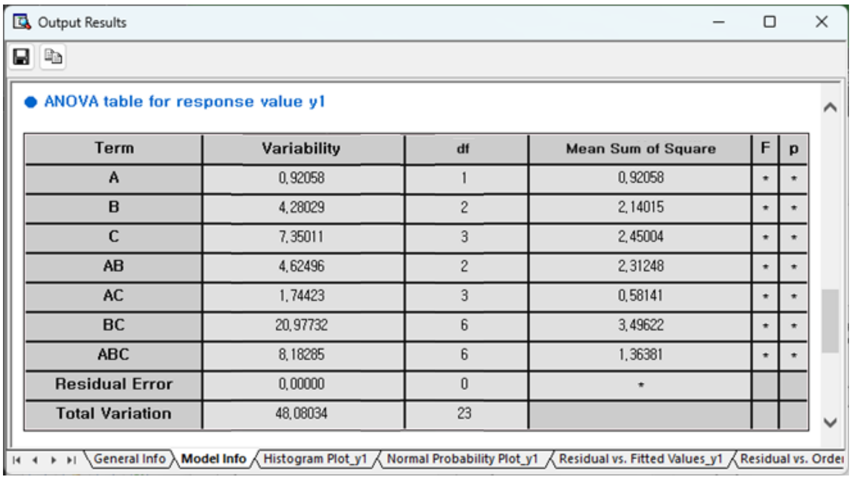


After the table is created, enter the response value y1 and click the Next button to proceed to Step 4.

- **Step 4: Analysis Result**



Set the Highest-order interaction to 3 and start the analysis. When you start analysis after completing settings, the following results screen will appear.

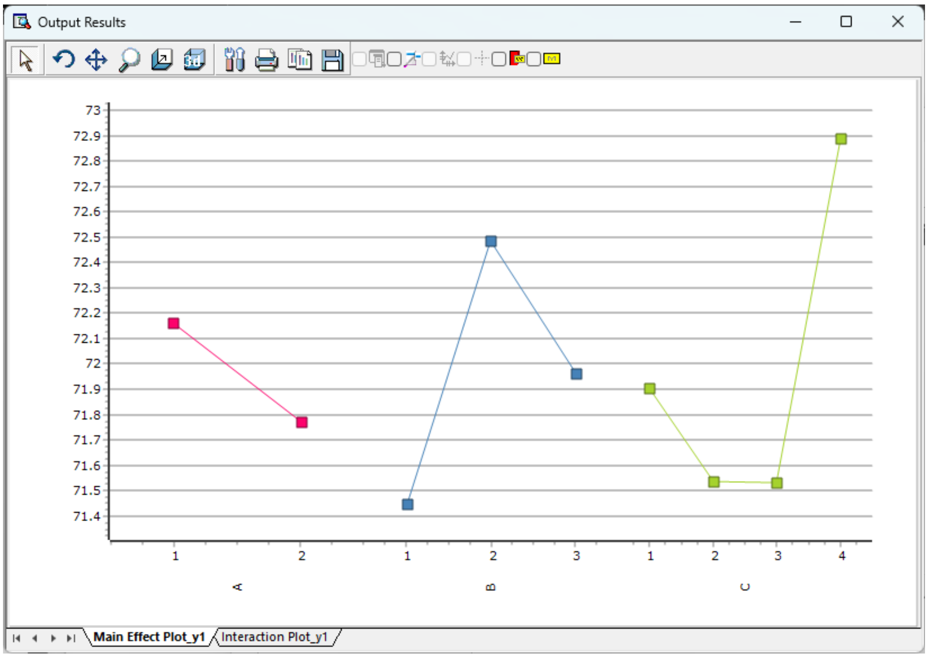


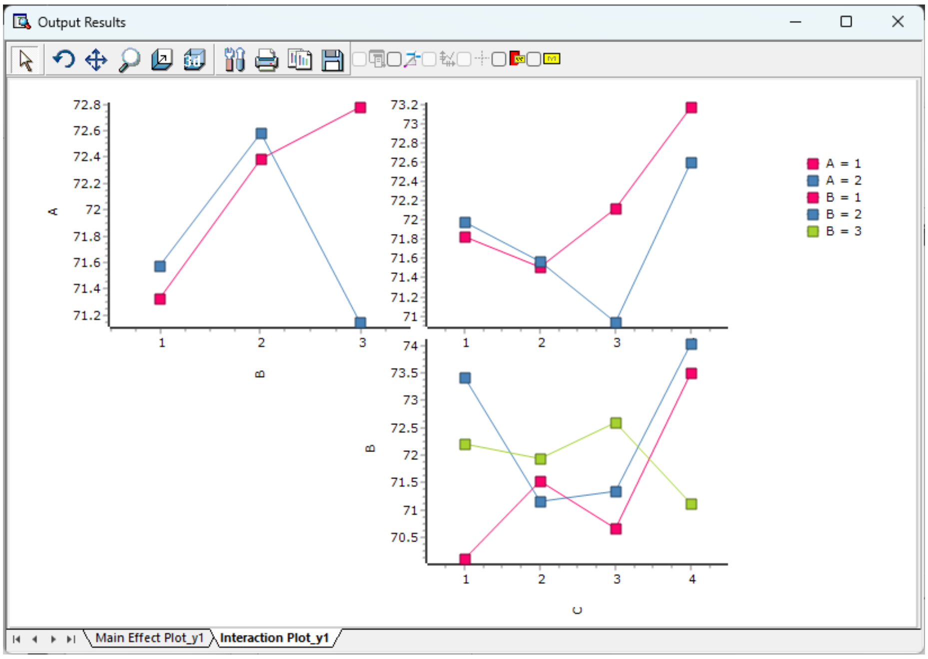
Term	Variability	df	Mean Sum of Square	F	p
A	0.92058	1	0.92058	*	*
B	4.28029	2	2.14015	*	*
C	7.35011	3	2.45004	*	*
AB	4.62496	2	2.31248	*	*
AC	1.74423	3	0.58141	*	*
BC	20.97732	6	3.49622	*	*
ABC	8.18285	6	1.36381	*	*
Residual Error	0.00000	0	*		
Total Variation	48.08034	23			

General Info, residual analysis and interpretation of other information are the same as 2-Level Factorial Design (Default Generators). The ANOVA table shows that the variance increases significantly when the degree of interaction is high. In this case, it is difficult to explain the response value with only the individual effects of A, B, and C, and the response value y1 is determined by complex effects.

- **Step 5: Plot and Response Optimizer**

  Main effect plot and interaction plot. (Surface plot, contour plot, and response optimizer functions are not provided.) Main effect plot and interaction plot, shows how the value y1 reacts according to each level of the factor.





6.1.3.2 Response Surface Design

Response Surface Design is a statistical analysis method that identifies the relationship between explanatory variables and response values when multiple explanatory variables (factors) interact in a complex manner to influence a certain response value.

For example, in a certain chemical reaction, the amount of reaction is said to change with temperature and time. At this time, if the temperature and time are x_1, x_2 and the reaction amount is y, will have a relationship of

$$y = f(x_1, x_2)$$

Response Surface Design allows the experimenter to identify the relationship between factors and responses with a small number of experiments when the experimenter has some prior knowledge about the factors and responses. ECMiner™ DOE's Response Surface Design offers two methods.

- **Central Composite Design**
- **Box-Behnken Design**

6.1.3.2.1 Central Composite Design

For a quadratic regression model with k number of independent variables,

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i \le j}^{k} \beta_{ij} x_i x_j + \epsilon \quad \epsilon \sim N(0, \sigma^2)$$

the regression coefficient cannot be estimated using the 2-Level factor placement method. Because in a 2-level factorial experiment, the experiment is conducted only at two levels of each variable, it is not possible to detect the curved change in the response amount that occurs according to the change in the level of the variable, and it is impossible to estimate the coefficient of the square term in the quadratic regression model. In order to compensate for these shortcomings and estimate the curved surface with a small number of experiments, the DOE in which the center point and axis point are added to the 2-Level factor experiment as follows is called Central Composite Design.

In Central Composite Design, the number of center points is not limited and can be at least one, and the number of axis points is 2k. Here, the value of k can be a positive number. For example, if the factor is 3 and there are two central points, the central composite design in Full factorial design is as follows.

$$
\begin{bmatrix}
-1 & -1 & -1 \\
1 & -1 & -1 \\
-1 & 1 & -1 \\
1 & 1 & -1 \\
-1 & -1 & 1 \\
1 & -1 & 1 \\
-1 & 1 & 1 \\
1 & 1 & 1 \\
\alpha & 0 & 0 \\
-\alpha & 0 & 0 \\
0 & \alpha & 0 \\
0 & -\alpha & 0 \\
0 & 0 & \alpha \\
0 & 0 & -\alpha \\
0 & 0 & 0 \\
0 & 0 & 0
\end{bmatrix}
$$

Lines 1 to 8 are experiments on grid points. And lines 9 to 14 are the axis points. And the last two rows are the center points. This will result in a design that can sufficiently estimate the regression equation for the quadratic curve. This type of experimental design has the advantage of being able to find the regression equation for the quadratic curve with a much smaller number of experiments than 3-Level Factorial Design.

The number of experiments in this experiment is

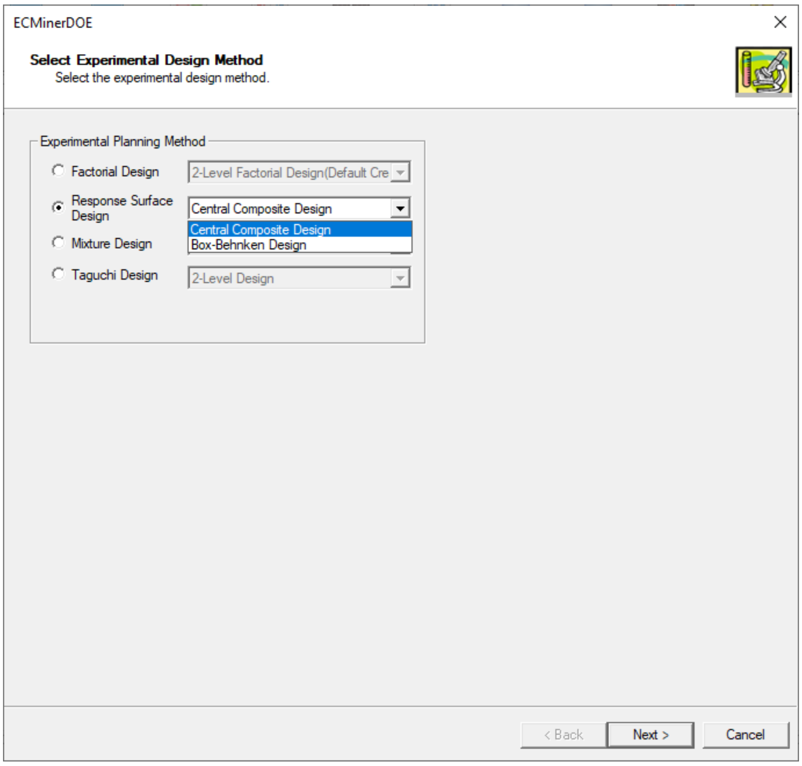
$$\boldsymbol{n} = 2^{\boldsymbol{k}} + \boldsymbol{2k} + \boldsymbol{n_0}, \quad \boldsymbol{n_0} \ \textbf{\textit{is a number of center points}}.$$

An advantage of Central Composite Design is its flexibility for sequential experiments. For instance, if a 2-Level Factorial Design with a first-order regression model proves inadequate, additional design points can be added to the center and axes to convert it into a Central Composite Design, without having to start a new DOE.

Introduction to Experiments

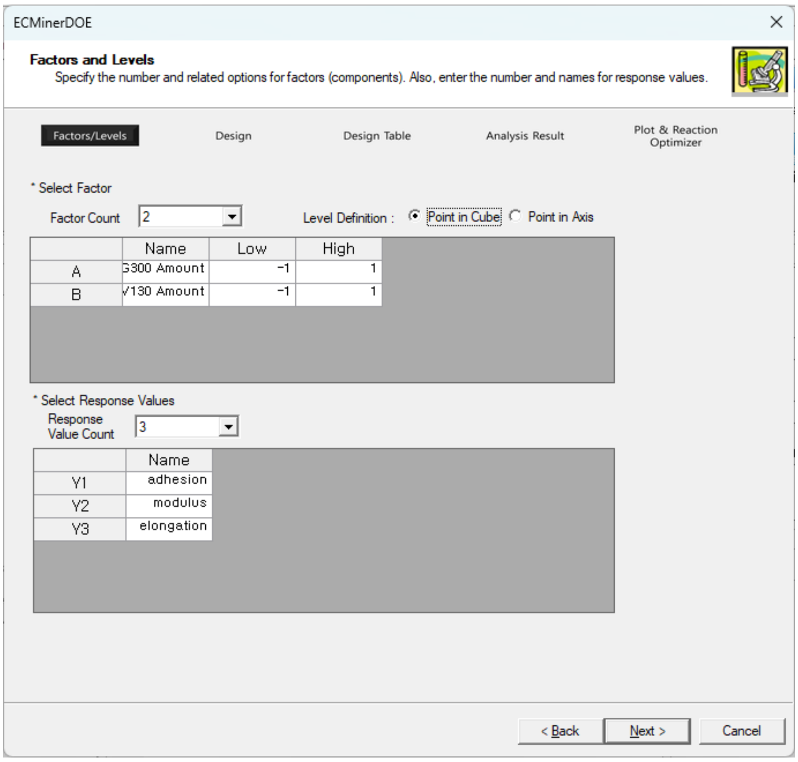
This experiment is an example of the application of the response surface experimental design method conducted by a tire company in 1979 to improve the driving performance of Monopoly radial tires. This is an experiment in which the amount of G300 and amount of V130 affect the response values of adhesion, modulus, and elongation.

Select **Central Composite Design** from Response Surface Design.



- **Step 1: Factors and Levels**

  First, name the factors as ‘G300 Amount’ and ‘V130 Amount,’ keeping their low and high units as initially defined. Next, set the **Number of Responses** to 3 and rename the responses to ‘adhesion,’ ‘modulus,’ and ‘elongation,’ as shown below.



- **Step 2: Design**

In the following screen, specify Select Design, Center Point Count, Repetition Count, and Alpha value.

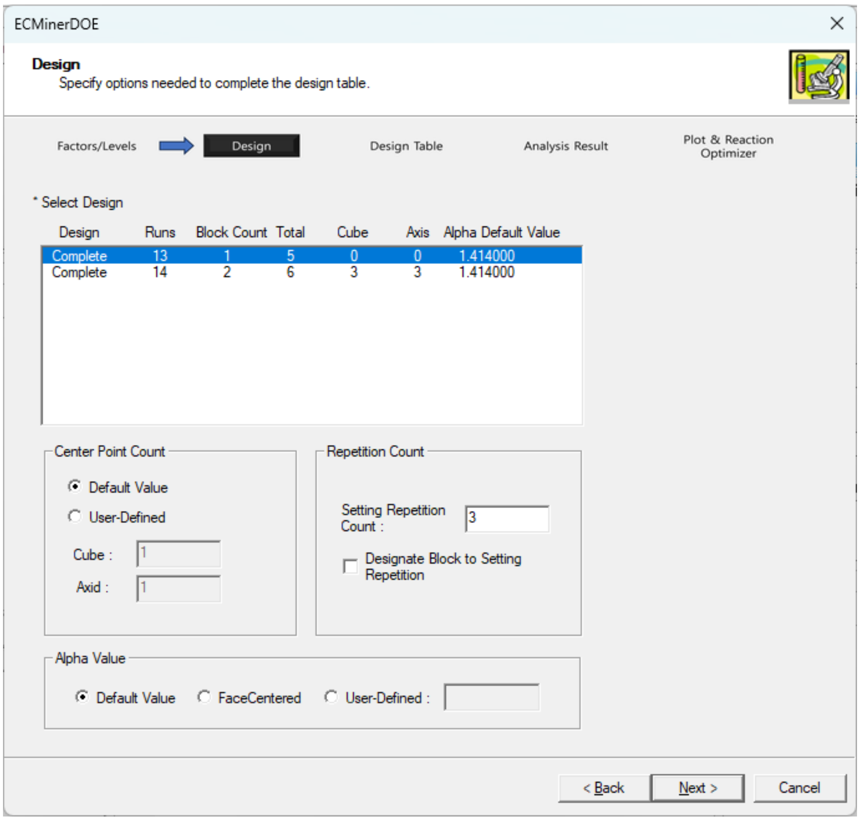
- **Select Design**: Choose one of the available design types. In a usable design, ‘complete’ means that the grid points contain all the points in the 2-Level Full factorial design, and ‘partial’ means that the grid points contain the points in the 2-Level Fractional factorial design.

- **Center Point Count**: Enter default or custom values.

- **Repetition Count**: The default repetition count is 1, but the user can enter the repetition count arbitrarily. Blocks can also be assigned to repetition as needed.

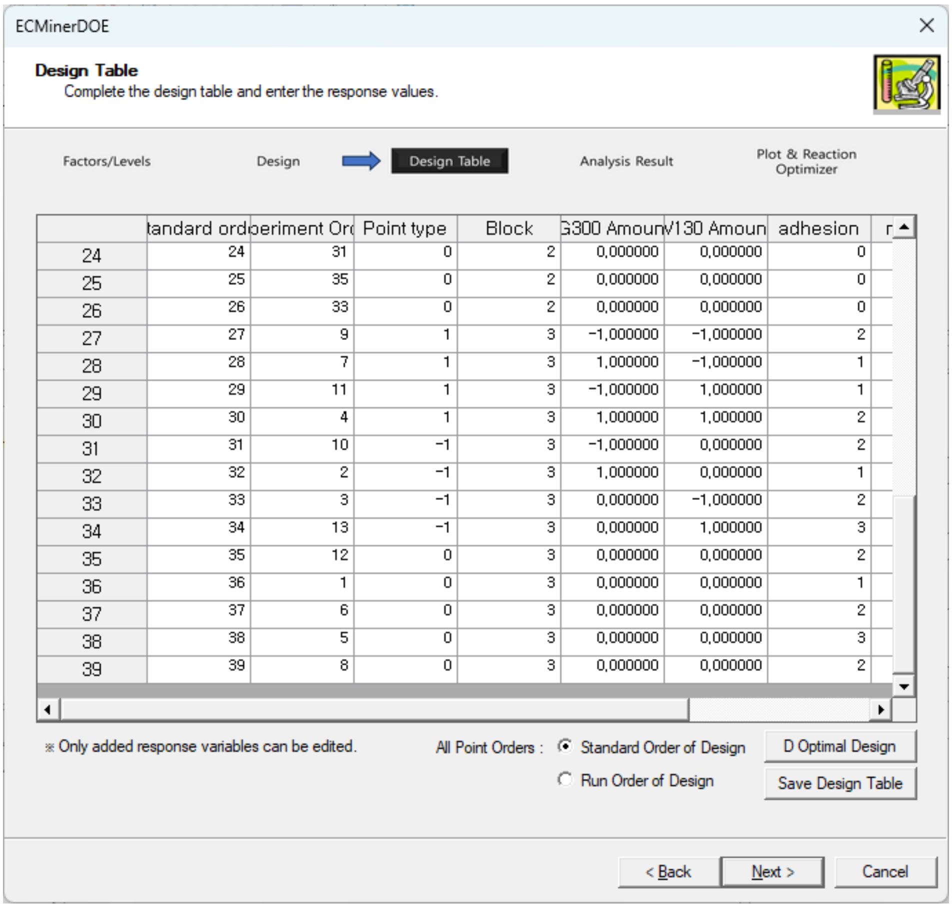
- **Alpha value**: Use the default value, Face Centered (Alpha value = 1), or enter a custom value.

Since we want to conduct the experiment three times at each factor point, axis point, and center point, set the center point count to 1, alpha value to 1 (Face Centered), and repetition count to 3, as shown in the following screen, and click the next button.



- **Step 3: Design Table**

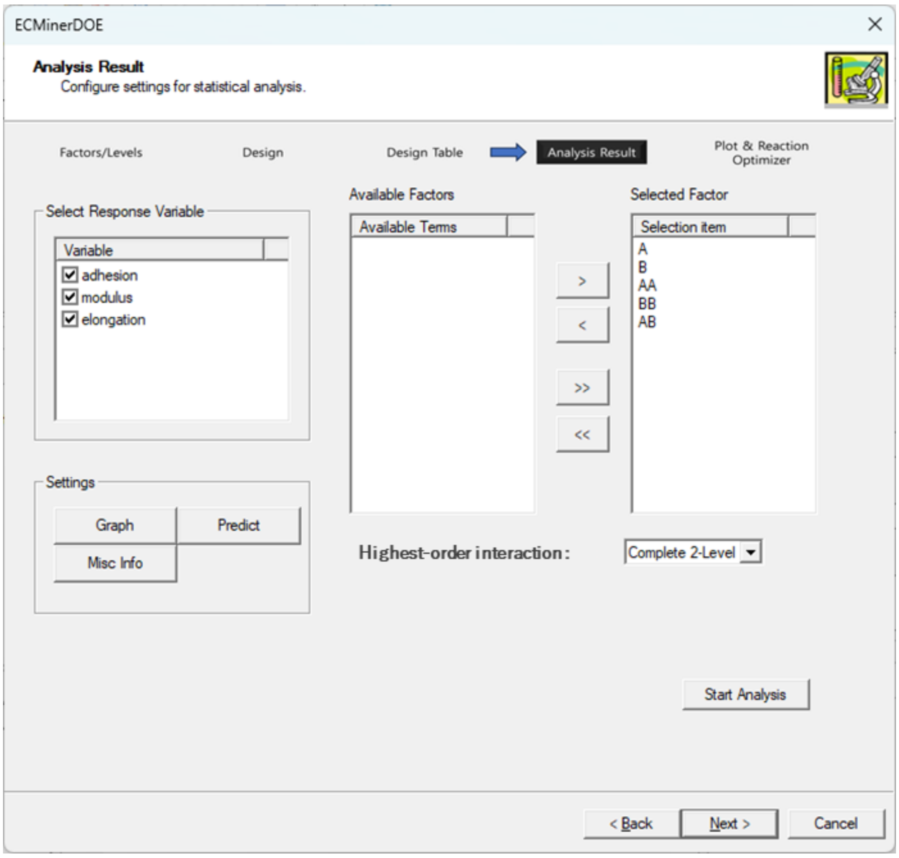
  At this stage, the design table is completed from the settings of Step 1 and Step 2. Through experiment, enter response values corresponding to adhesion, modulus, and elongation in the completed design table. At this time, in the case of the D Optimal Design option, it is a method to modify the created design table, and please refer to 6.1.3.2.3 Response Surface Design D Optimal Design.



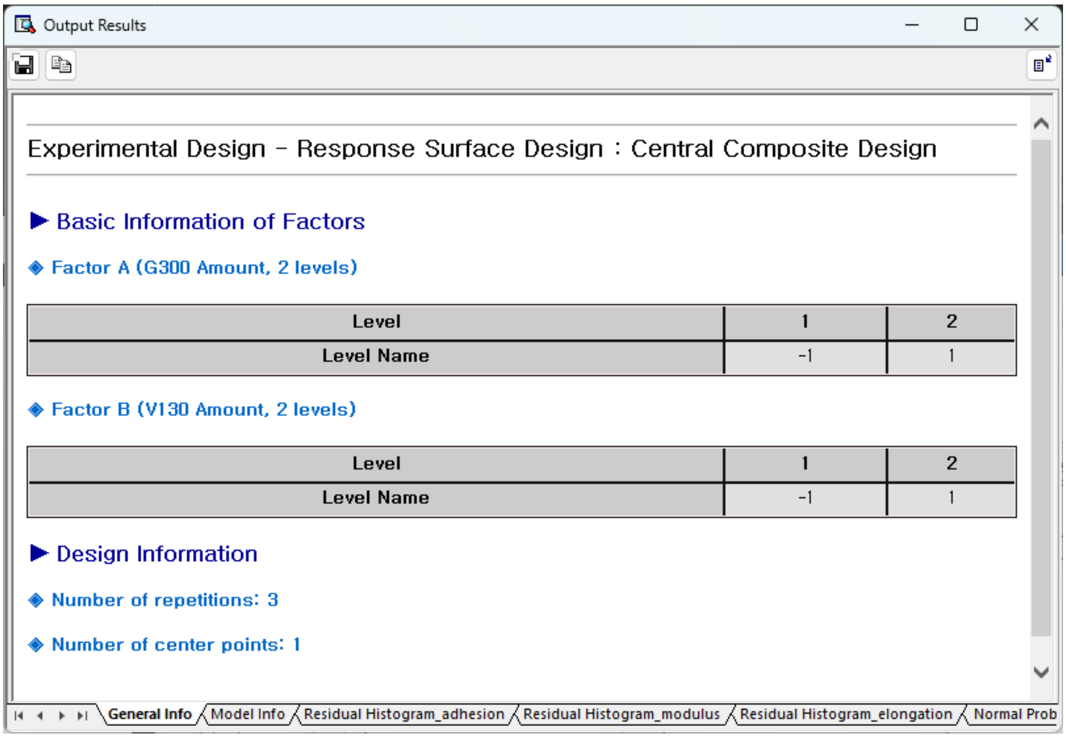
	tandard ord	periment Ord	Point type	Block	G300 Amoun	V130 Amoun	adhesion
24	24	31	0	2	0,000000	0,000000	0
25	25	35	0	2	0,000000	0,000000	0
26	26	33	0	2	0,000000	0,000000	0
27	27	9	1	3	−1,000000	−1,000000	2
28	28	7	1	3	1,000000	−1,000000	1
29	29	11	1	3	−1,000000	1,000000	1
30	30	4	1	3	1,000000	1,000000	2
31	31	10	−1	3	−1,000000	0,000000	2
32	32	2	−1	3	1,000000	0,000000	1
33	33	3	−1	3	0,000000	−1,000000	2
34	34	13	−1	3	0,000000	1,000000	3
35	35	12	0	3	0,000000	0,000000	2
36	36	1	0	3	0,000000	0,000000	1
37	37	6	0	3	0,000000	0,000000	2
38	38	5	0	3	0,000000	0,000000	3
39	39	8	0	3	0,000000	0,000000	2

- **Step 4: Analysis Result**

  This step is for analyzing results of the experiment. If we analyze all three response values, select all three response values in “Select Response Variable” and make the necessary settings in Graph, Predict, and Misc Info. If we currently want to do a full quadratic regression analysis, make the selection as follows and click the **Start Analysis** button.

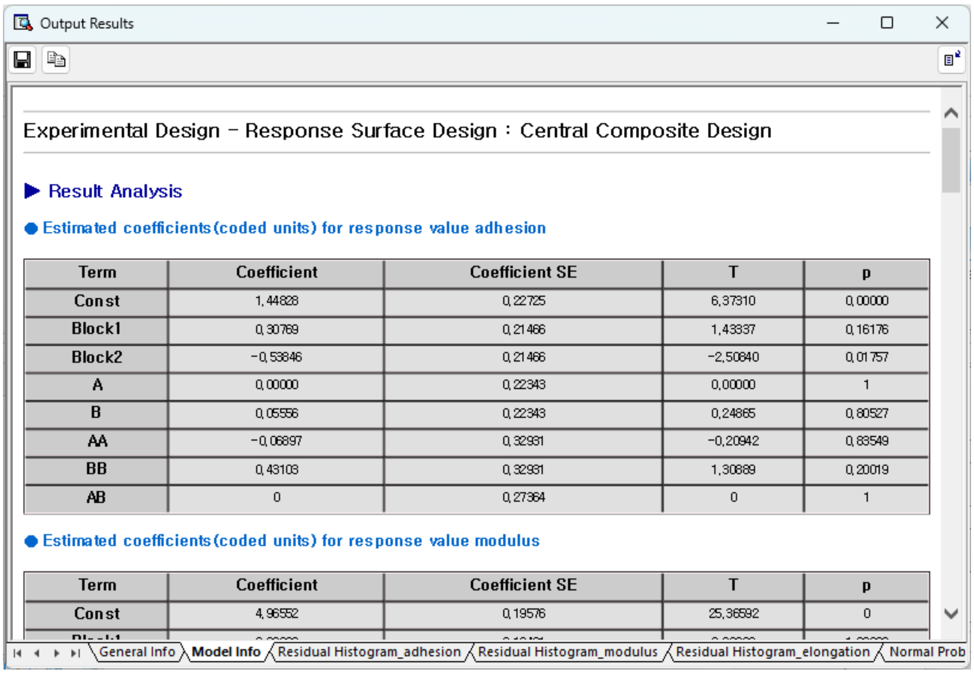


General Info: Shows general information about the design.



Model Info: Shows results of regression analysis, ANOVA, unusual observations (extreme leverage, standardized residual).

Regression Analysis Result



Output Results

Experimental Design – Response Surface Design : Central Composite Design

▶ Result Analysis

● Estimated coefficients(coded units) for response value adhesion

Term	Coefficient	Coefficient SE	T	p
Const	1,44828	0,22725	6,37310	0,00000
Block1	0,30769	0,21466	1,43337	0,16176
Block2	−0,53846	0,21466	−2,50840	0,01757
A	0,00000	0,22343	0,00000	1
B	0,05556	0,22343	0,24865	0,80527
AA	−0,06897	0,32931	−0,20942	0,83549
BB	0,43103	0,32931	1,30889	0,20019
AB	0	0,27364	0	1

● Estimated coefficients(coded units) for response value modulus

Term	Coefficient	Coefficient SE	T	p
Const	4,96552	0,19576	25,36592	0

General Info | Model Info | Residual Histogram_adhesion | Residual Histogram_modulus | Residual Histogram_elongation | Normal Prob

ANOVA Results

Output Results

● ANOVA table for response value adhesion

Term	Variability	df	Mean Sum of Square	F	p
Block	5,69231	2	2,84615	3,16744	0,05602
Linear	0,05556	2	0,02778	0,03091	0,96959
Square	1,62732	2	0,81366	0,90551	0,41476
Interaction	0	1	0	0	1
Residual Error	27,85559	31	0,89857		
Lack of Fit	23,85559	19	1,25556	3,76667	0,01135
PureError	4	12	0,33333		
Total Variation	35,23077	38			

● ANOVA table for response value modulus

Term	Variability	df	Mean Sum of Square	F	p
Block	0,15385	2	0,07692	0,11537	0,89142
Linear	0,88889	2	0,44444	0,66656	0,52068
Square	0,72325	2	0,36163	0,54236	0,58679
Interaction	0,33333	1	0,33333	0,49992	0,48482
Residual Error	20,66991	31	0,66677		
Lack of Fit	11,86991	19	0,62473	0,85191	0,63443

General Info | Model Info | Residual Histogram_adhesion | Residual Histogram_modulus | Residual Histogram_elongation | Normal Prob

Unusual Observations

Output Results

● Unusual observations for response value adhesion – Extreme Leverage

No abnormal observations (leverage criterion) for response value adhesion.

● Unusual observations for response value adhesion – Standardized Residual

Order	Actual Observation Value	Prediction	Residual	Standardized Residuals
16	3	1,32744	1,67256	2,13139
20	3	1,28529	1,71471	2,04299

● Unusual observations for response value modulus – Extreme Leverage

No abnormal observations (leverage criterion) for response value modulus.

● Unusual observations for response value modulus – Standardized Residual

No abnormal observations (standardized residual criterion) for response value modulus.

● Unusual observations for response value elongation – Extreme Leverage

No abnormal observations (leverage criterion) for response value elongation.

● Unusual observations for response value elongation – Standardized Residual

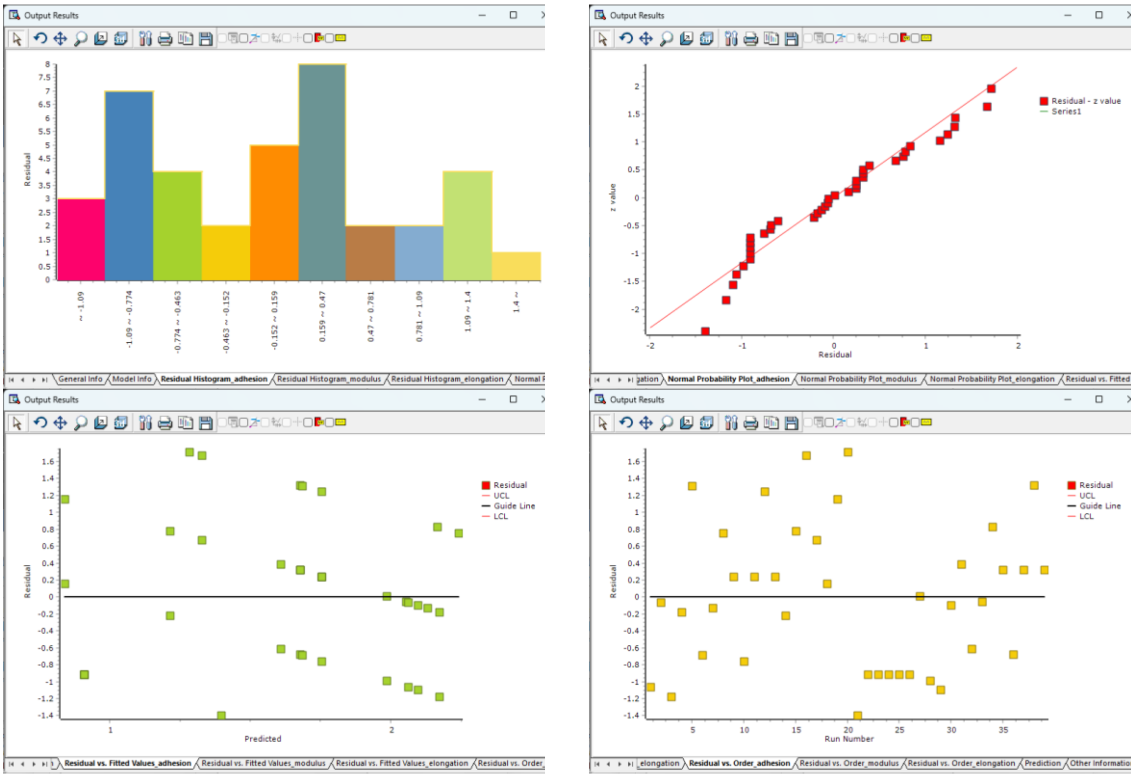
No abnormal observations (standardized residual criterion) for response value elongation.

General Info | Model Info | Residual Histogram_adhesion | Residual Histogram_modulus | Residual Histogram_elongation | Normal Prob

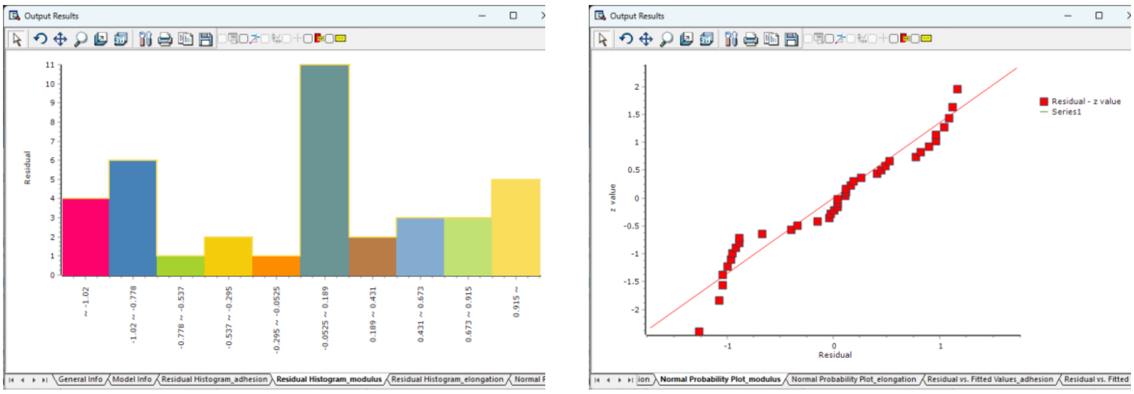
Residual-related Plot

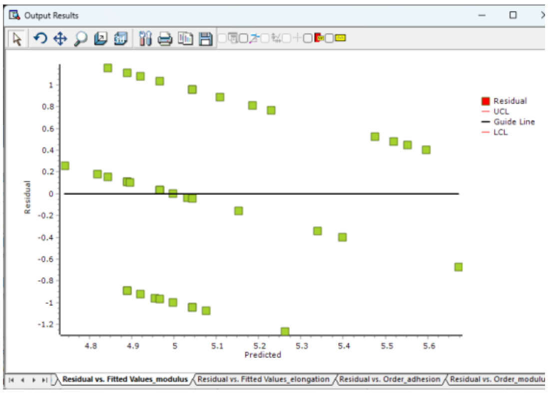
See residual histogram, residual normal probability plot, residuals versus ordinal, residuals vs. the fitted values.

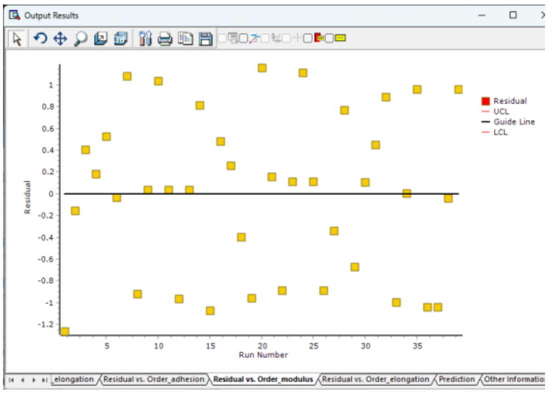
Residual-related plot for response value “Adhesion”



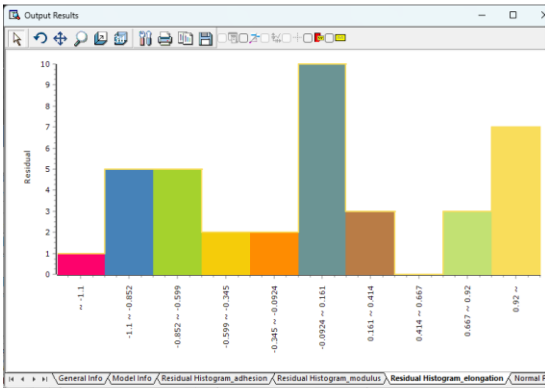
Residual-related plot for response value “Modulus”

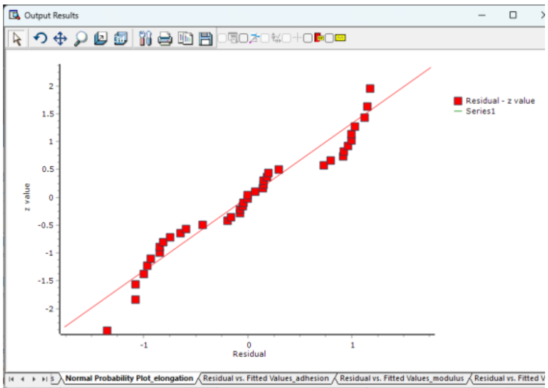


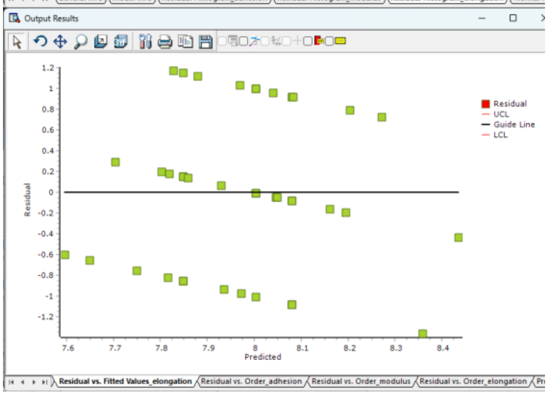


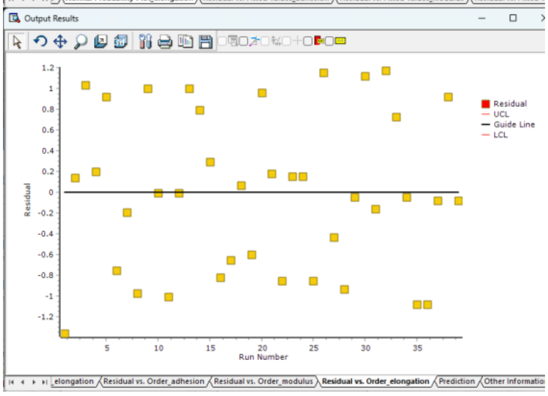


Residual-related plot for response value “Elongation”



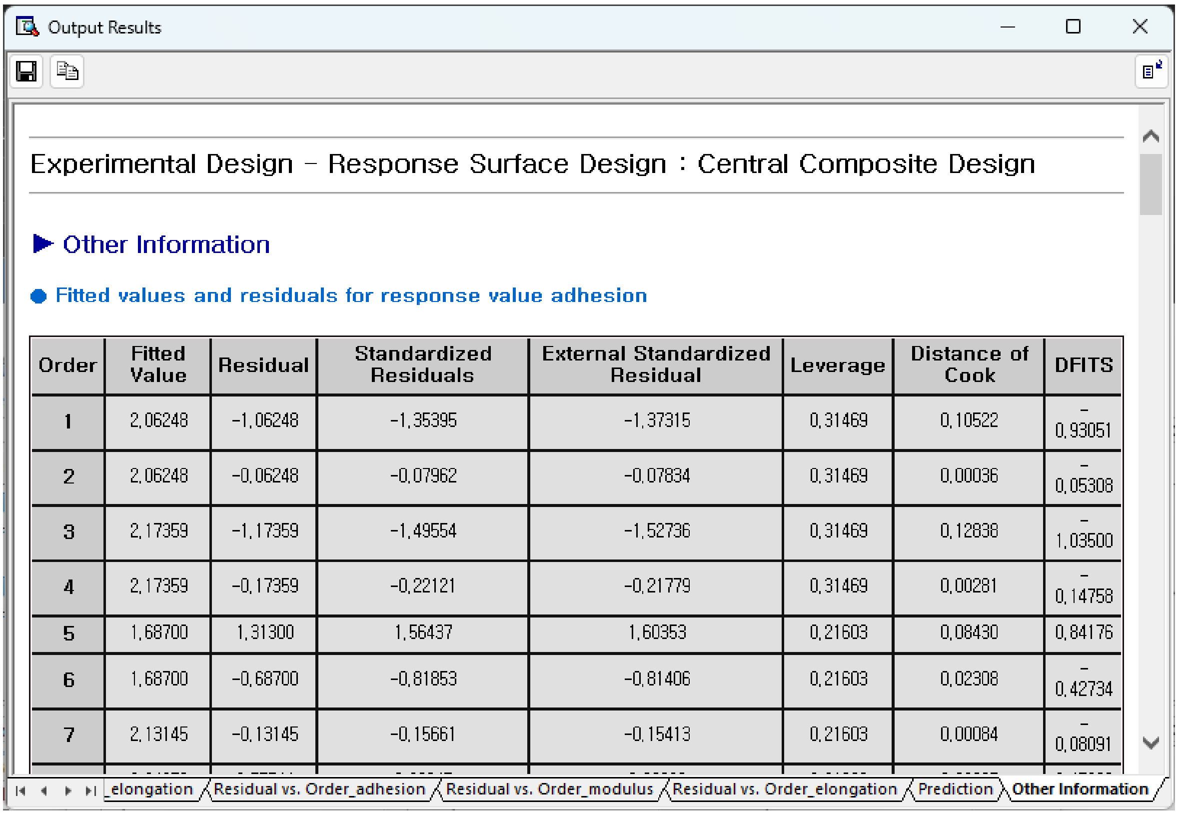






Other Information

Shows residual related statistics.



Output Results

Experimental Design – Response Surface Design : Central Composite Design

▶ Other Information

● Fitted values and residuals for response value adhesion

Order	Fitted Value	Residual	Standardized Residuals	External Standardized Residual	Leverage	Distance of Cook	DFITS
1	2,06248	-1,06248	-1,35395	-1,37315	0,31469	0,10522	-0,93051
2	2,06248	-0,06248	-0,07962	-0,07834	0,31469	0,00036	-0,05308
3	2,17359	-1,17359	-1,49554	-1,52736	0,31469	0,12838	-1,03500
4	2,17359	-0,17359	-0,22121	-0,21779	0,31469	0,00281	-0,14758
5	1,68700	1,31300	1,56437	1,60353	0,21603	0,08430	0,84176
6	1,68700	-0,68700	-0,81853	-0,81406	0,21603	0,02308	-0,42734
7	2,13145	-0,13145	-0,15661	-0,15413	0,21603	0,00084	-0,08091

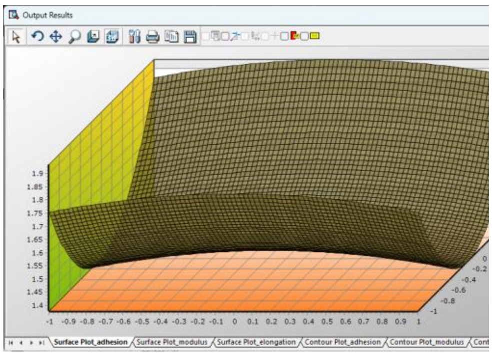
_elongation / Residual vs. Order_adhesion / Residual vs. Order_modulus / Residual vs. Order_elongation / Prediction / Other Information

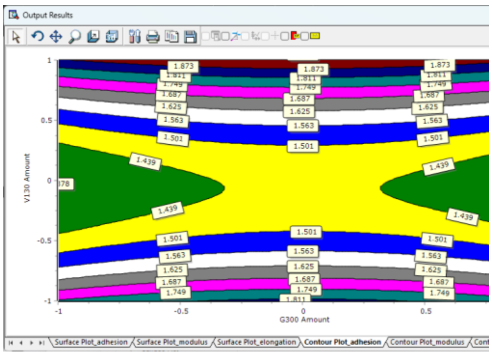
For detailed explanation, see 6.1.4. Settings and Analysis.

- **Step 5: Plot and Response Optimizer**

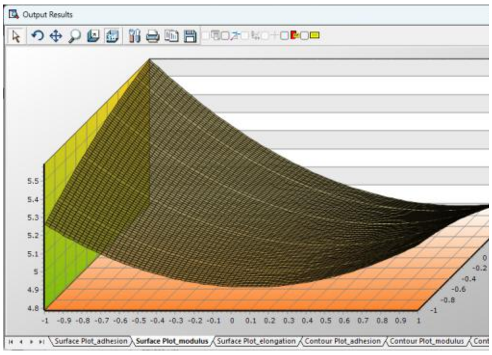
  In this step, you can draw a surface plot and contour plot with the Regression Model created in Step 4 and optimize the response according to the user's purpose. Since there are currently two factors, there is no need to enter a separate fixed value. The surface plot and contour plot according to each response value are as follows.

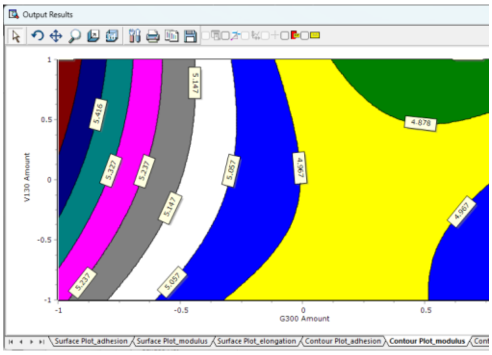
  Surface plot, contour plot for response value “Adhesion”



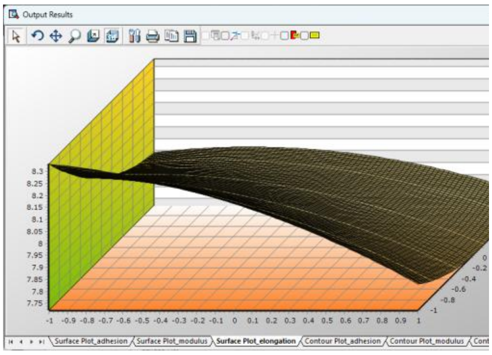


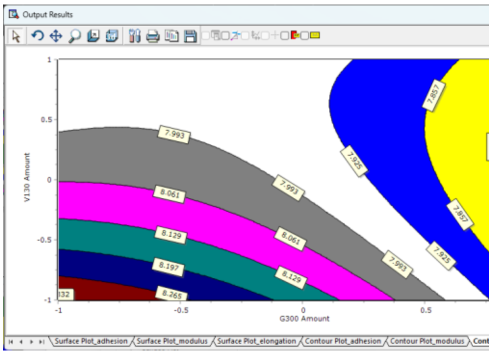
Surface plot, contour plot for response value “Modulus”



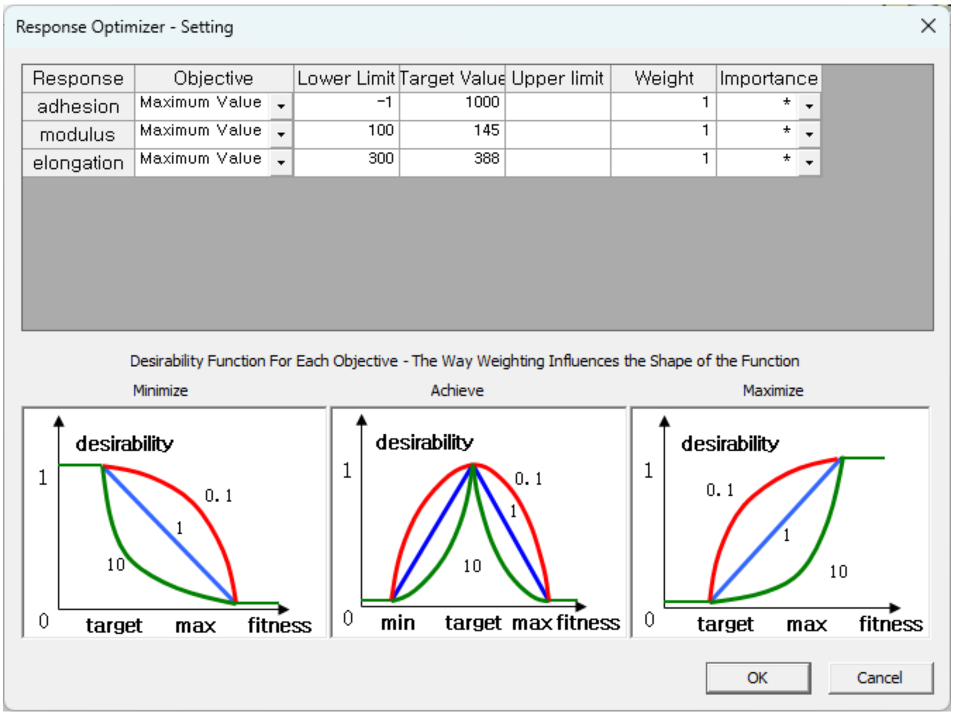


Surface plot, contour plot for response value "Elongation”





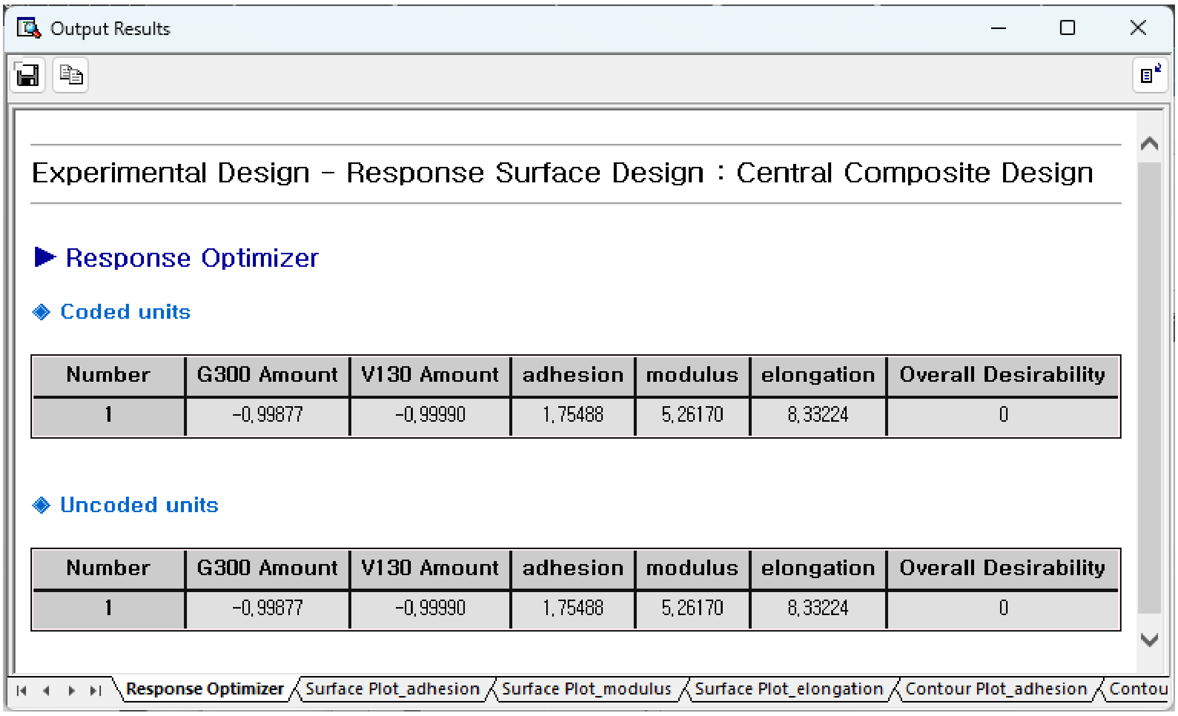
In this experiment, the experimenter's goal is to maximize the modulus value over 145, the elongation value over 388, and the adhesion value. To do this, set the following settings.



Response	Objective	Lower Limit	Target Value	Upper limit	Weight	Importance
adhesion	Maximum Value	-1	1000		1	*
modulus	Maximum Value	100	145		1	*
elongation	Maximum Value	300	388		1	*

In the window above, the target values of modulus and elongation are set to 145 and 388, it means that there is no need to increase the values any more, once these values are reached. And the reason why the target value of adhesion was set as large as possible is because in the case of adhesion, the larger the better. (In fact, it seems reasonable to set the lower limit of modulus and elongation to 145 and 388. However, in this case, if the modulus value is less than 145 or the elongation value is less than 388, the desirability function becomes 0. In fact, this area is so large that it is often difficult to find the optimal value when performing optimization. For this purpose, the performance of optimization was improved by setting the lower limit a little smaller.)

After configuring this and performing optimization, you can get the following results.



Output Results

Experimental Design - Response Surface Design : Central Composite Design

▶ Response Optimizer

◈ Coded units

Number	G300 Amount	V130 Amount	adhesion	modulus	elongation	Overall Desirability
1	-0.99877	-0.99990	1.75488	5.26170	8.33224	0

◈ Uncoded units

Number	G300 Amount	V130 Amount	adhesion	modulus	elongation	Overall Desirability
1	-0.99877	-0.99990	1.75488	5.26170	8.33224	0

Response Optimizer / Surface Plot_adhesion / Surface Plot_modulus / Surface Plot_elongation / Contour Plot_adhesion / Contou

From this, we can see that if the G300 content is -1.414 and the V130 content is 0.211, the adhesion can be maximized to 73.652 while satisfying the conditions of modulus and elongation.

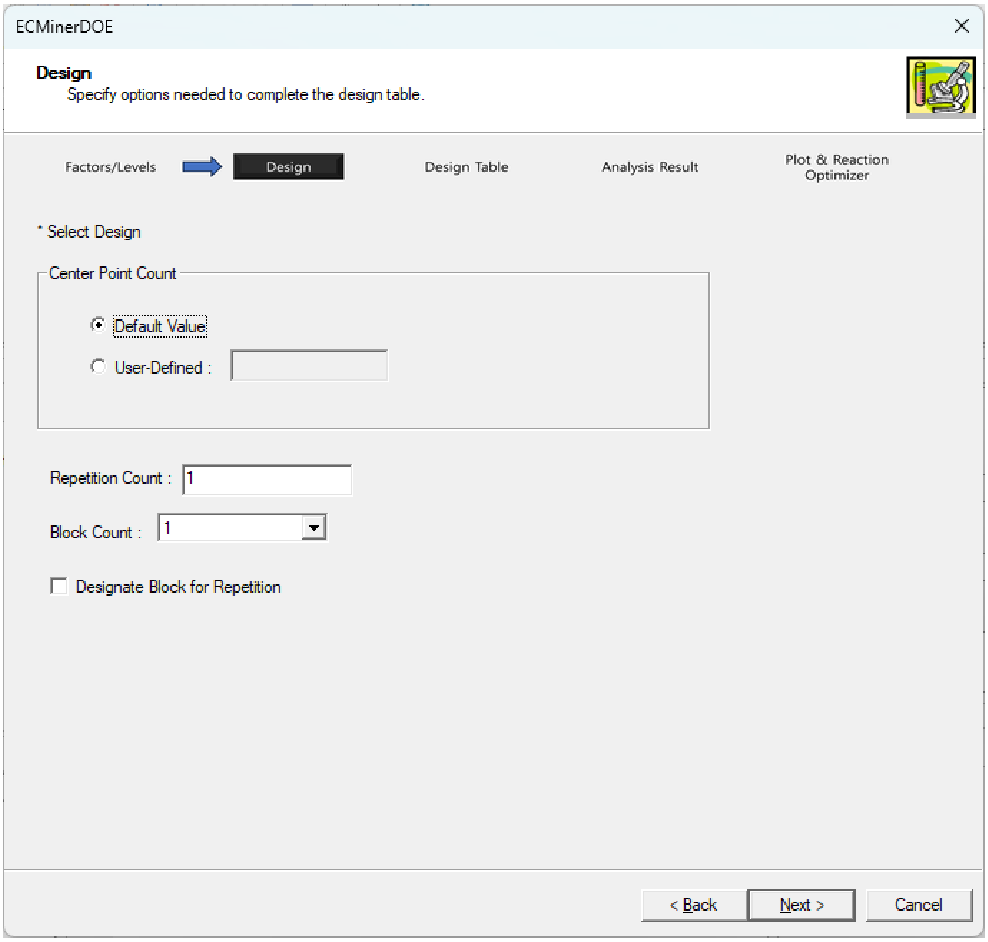
6.1.3.2.2. Box-Behnken Design

Box Behnken's design was created to complement the shortcomings of 2-Level design and 3-Level design. In general, p-level designs are suitable for polynomial fits of order p-1. Therefore, it is difficult to fit a second order polynomial with a 2-level design, and although it is possible with a 3-level design, it has the disadvantage of requiring a large number of experiments. Box Behnken's design is an experiment that selects a part of the 3-level design and allows for more efficient second order polynomial fit.

Box Behnken design is also a design created for the same purpose as central composite design. Therefore, rather than explaining this in detail with an example, we will explain the differences in central composite design. (For basic methods of Response Surface Design, please refer to 6.1.3.2.2. Central composite design.)

Select Box Behnken Design from Response Surface Design. In the Step 1 screen, enter the number of factors, factor name, and upper and lower limits, and enter the number of response values and names according to the response value.

The screen for Step 2 is as follows.



As shown, simply select the options, and the subsequent process will be the same as central composite design.

6.1.3.2.3. Response Surface Design D Optimal Design

D-Optimal Design is a method that adjusts the experimental design to best suit future statistical analysis, allowing customization based on the user's needs. There are the following indicators to judge the excellence of design.

D Optimality (Determinant)

D Optimality is the most commonly used criterion and is used to find a design that maximizes the Determinant of the $X^T X$ inverse matrix. When obtaining a Design Matrix, create it by selecting the necessary candidate points from a set of several potential points. The design table that makes the Determinant of the $X^T X$ inverse matrix the largest is called D-Optimal Design.

A Optimality (Trace)

When obtaining Design Matrix create by gathering the necessary candidate points from a set of several candidate points, the design table that creates the largest TRACE of the $X^T X$ inverse matrix is called A-Optimal Design.

However, in the case of A-Optimality, it is not often used due to computational difficulties.

G Optimality (Average Leverage / Maximum Leverage)

G Optimality means average leverage divided by maximum leverage. Here, leverage refers to the Generalized Linear Model Matrix as X, when H Matrix is

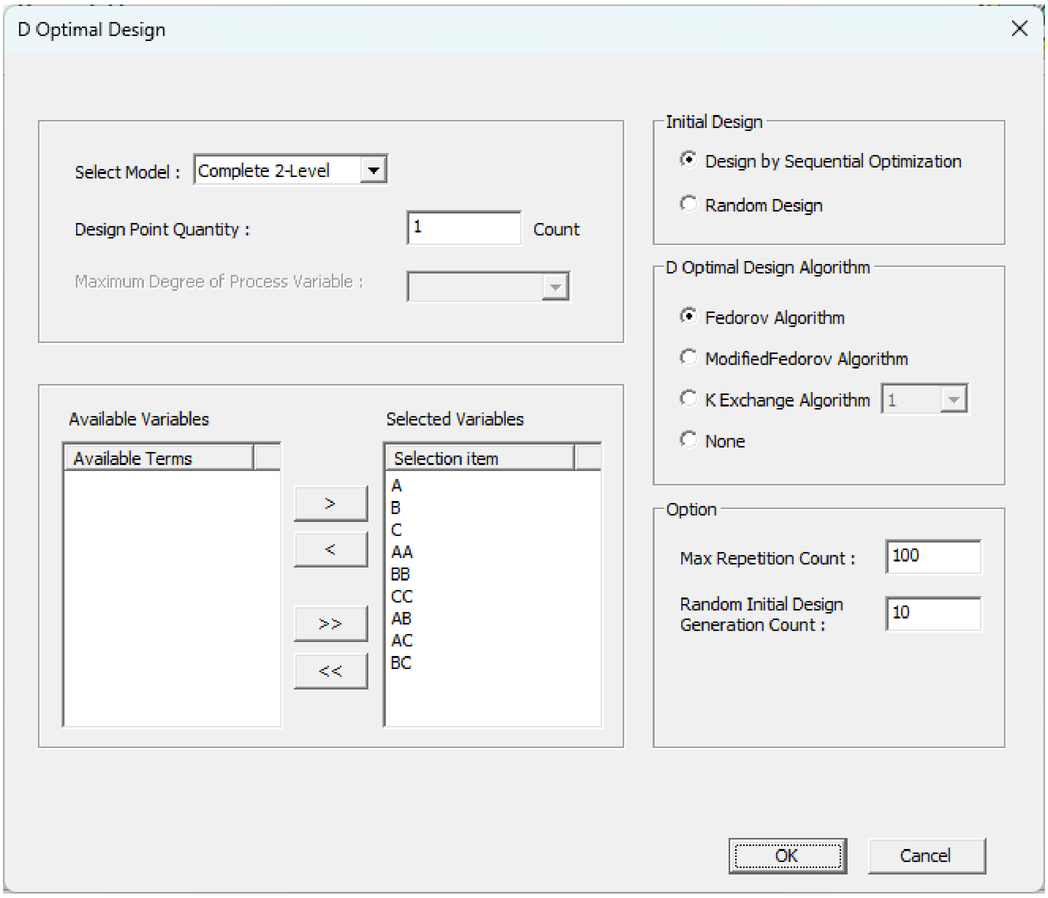
$$H = X\left(X^T X\right)^{-1} X^T$$

It refers to their diagonal components. Dividing the average of these leverages by the maximum value is G-Optimality.

V Optimality

The average of leverage is V Optimality.

Among these indicators, D Optimal Design is used to maximize D Optimality. The D Optimal Design screen of Response Surface Design is composed as follows.



- **Model Selection**

  First, select a model. Depending on the selected model, the selection items in the selection window below will change, and you can use only some of them.

- **Number of design points for D Optimal Design**

  There is the modified design table to determine how many designs points

- **Initial Design**

  To perform Optimal Design, you must select initial design. Set how to select the initial design with the same number of design points as the number of design points in D Optimal Design.

- **D Optimal Design Algorithm**

  Decide which algorithm will be used to improve the initial design to obtain D-Optimal Design. ECMiner™ DOE provides Fedorov Algorithm, Modified Fedorov Algorithm, and K Exchange Algorithm. Among these, Modified Fedorov Algorithm and K Exchange Algorithm are known to have good performance.

- **Options**

  Maximum number of repetitions means the maximum number of repetitions to be performed in D Optimal Design Algorithm. Number of creations in random initial design means how many random designs will be created when designing the initial design randomly. Among the various random designs created at this time, the algorithm with the highest D Optimality is used as the initial design to perform the algorithm.

6.1.3.3 Mixture Design

The main aim of experimental designs to discover whether one or two or more factors like $(x_1, x_2, \ldots, x_k)$ have a significant effect on the response value y of interest, or furthermore, to find optimal conditions for x_i that maximize or minimize y. These include factorial design and response surface design, and these experimental design methods have no constraints on the ratios or sums that factors can take to each other. However, the Mixture Design method provides an experimental design where the sum of the factors (factors are called components in Mixture Design) is constant.

$$\sum_{i=1}^{k} x_i = constant \; x_1 \geq 0, x_2 \geq 0, \ldots x_k \geq 0$$

There are three Mixture Design methods provided by ECMiner™ DOE as follows.

- **Simplex Center Design**
- **Simplex Lattice Design**
- **Vertex Design**

In addition, Mixture Design is divided into three categories depending on the type of experiment.

- **Experimental design using only mixture components**
- **Experimental design adding process variables**
- **Mixture volume experimental design**

In the end, there are three types of experiments like the above in Simplex Center Design, three types of experiments like the above in Simplex Lattice Design, and three types of experiments like the above in Vertex Design, ultimately providing a total of nine methods. To explain this, the experimental design using only the mixture components can be explained using Simplex Center Design, the mixture amount design can be explained using Simplex Lattice Design, and the experiment adding process variables can be explained using Vertex Design. Through this, you can gain a general understanding of Mixture Design.

6.1.3.3.1. Simplex Center Design

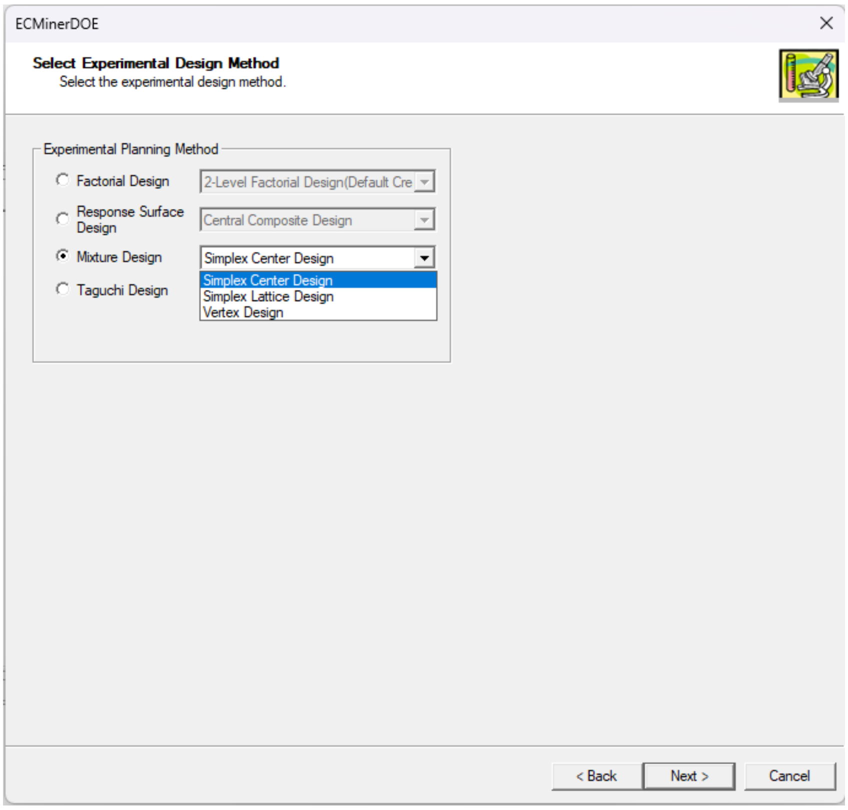
Simplex Center Design, which consists of q components, consists of $2^q - 1$ different experimental points. These points correspond to all possible permutations. This design can be said to be suitable for the following polynomial.

$$\boldsymbol{\eta} = \sum_{1}^{q} \boldsymbol{\beta_i x_i} + \sum_{i<j}^{q} \boldsymbol{\beta_{ij} x_i x_j} + \sum_{i<j<k}^{q} \boldsymbol{\beta_{ijk} x_i x_j x_k} + \cdots + \boldsymbol{\beta_{12\ldots q} x_1 x_2 \ldots x_q}$$

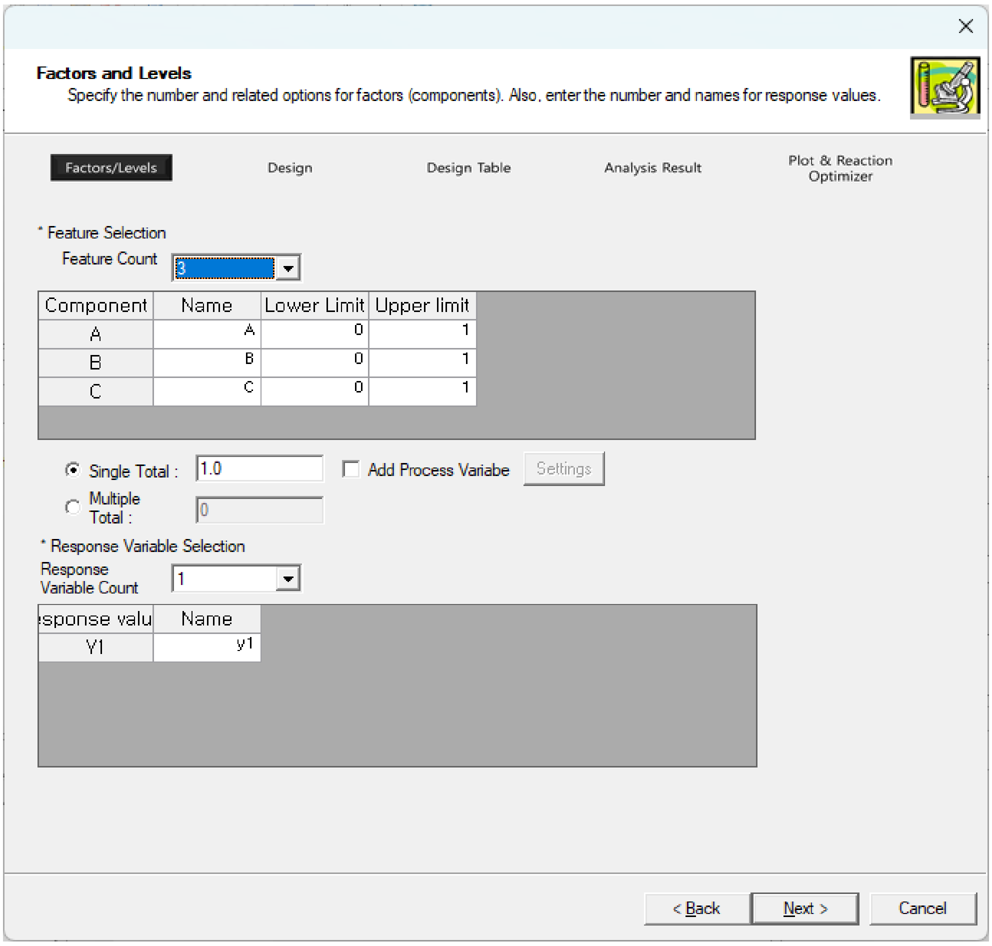
From this, we can see that as the number of components increases, the degree of the appropriate polynomial also increases.

Introduction to experiments In this experiment, we investigate the extent to which three components affect the firmness of the response product. The components have default values (0,1) without any special upper or lower limit conditions, and the experimenter tries to find out how each component affects the response value.

To perform **Simplex Center Design**, select Mixture Design -> Simplex Center Design on the following screen.



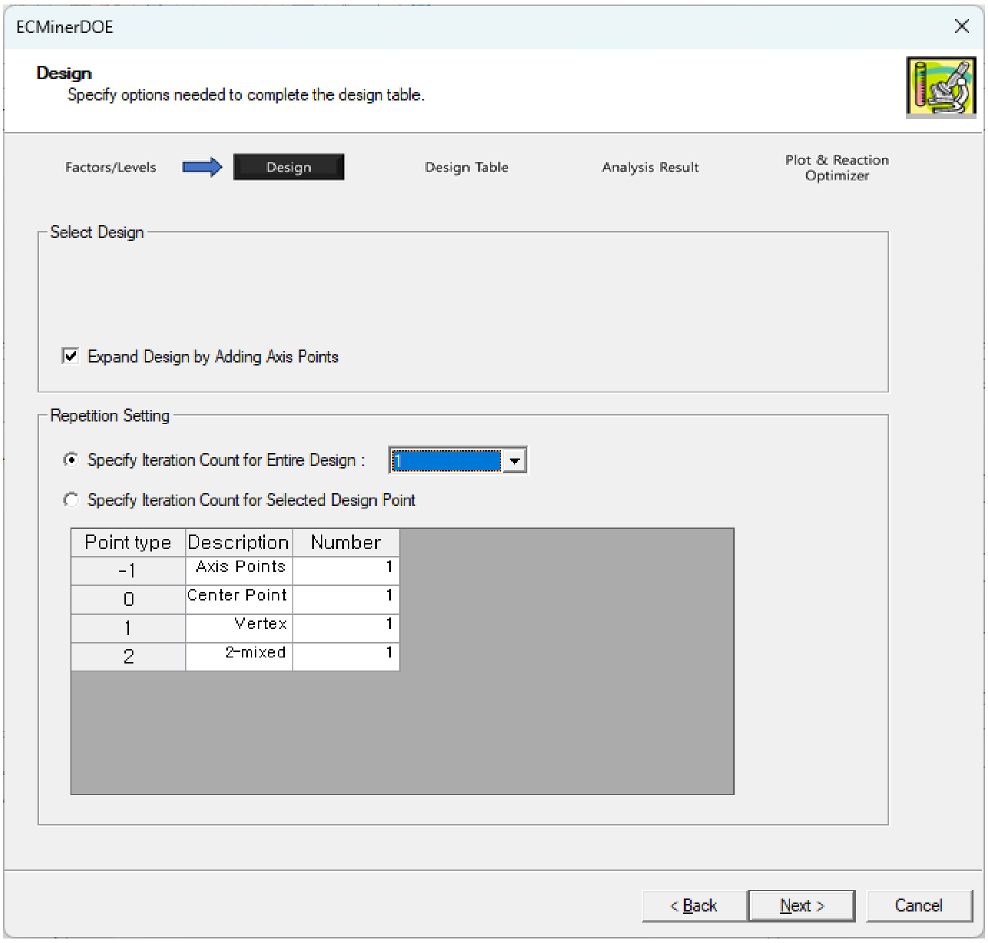
- **Step 1: Factors and Levels**



In this screen, the number of components is 3, the number of response values is 1, and the name of the response value is firmness. The options that appear on this screen are explained as follows.

- **Single Total**: Specifies the sum of each component. The default value is 1.
- **Multiple Total**: Enter the values for the mixture amount design. All values must be positive numbers and should be separated by spaces.
- **Add Process Variable**: Use when adding a Process Variable to Mixture Design.

- **Step 2: Design**



On this screen, set several options to confirm the design.

Expand Design by Adding Axis Points

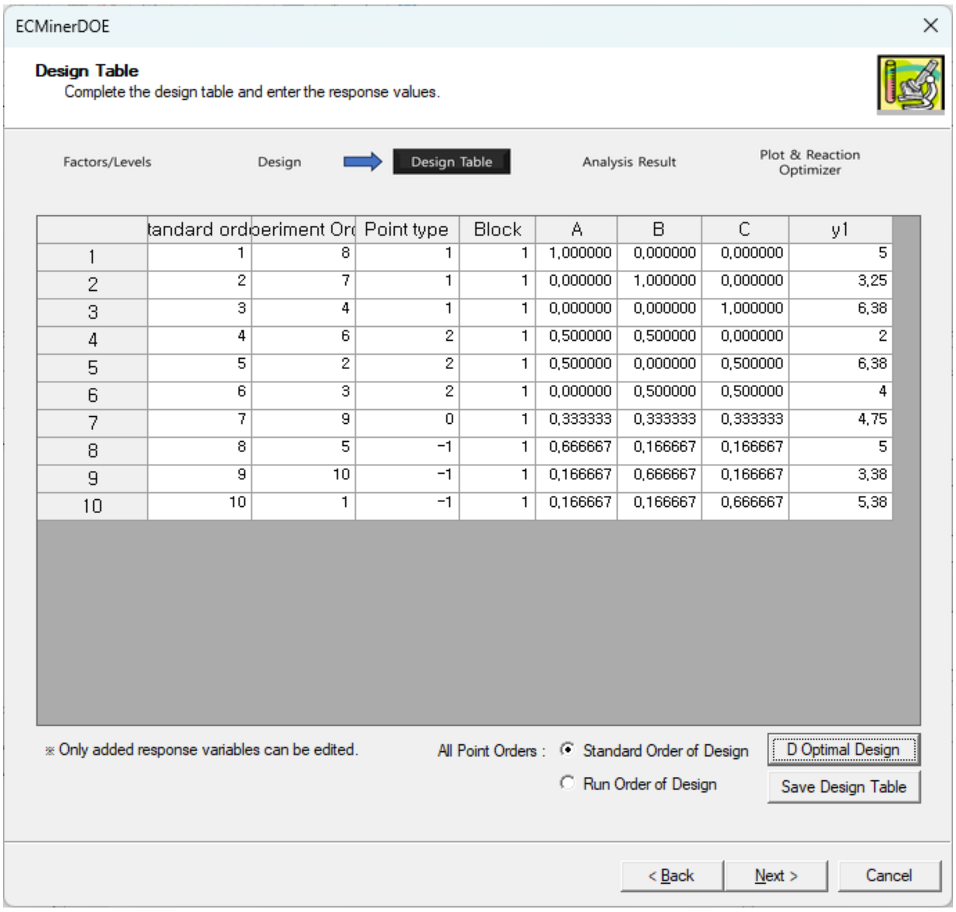
Adds axis points between the center point and the points corresponding to each vertex.

Repetition Setting

- **Specify Iteration Count for Entire Design**: Specify how many times to repeat the design table created through this option.

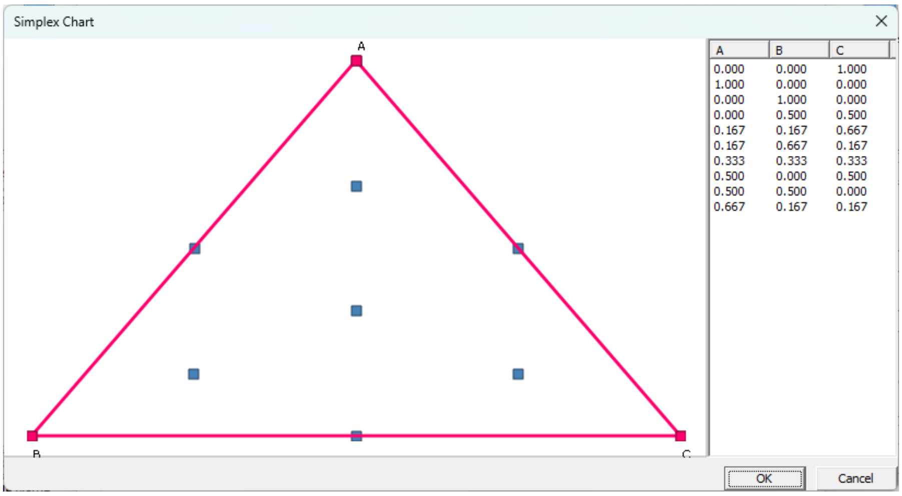
- **Specify Iteration Count for Selected Design Point**: This option allows the user to set the number of repetitions depending on the point type.

- **Step 3: Design Table**

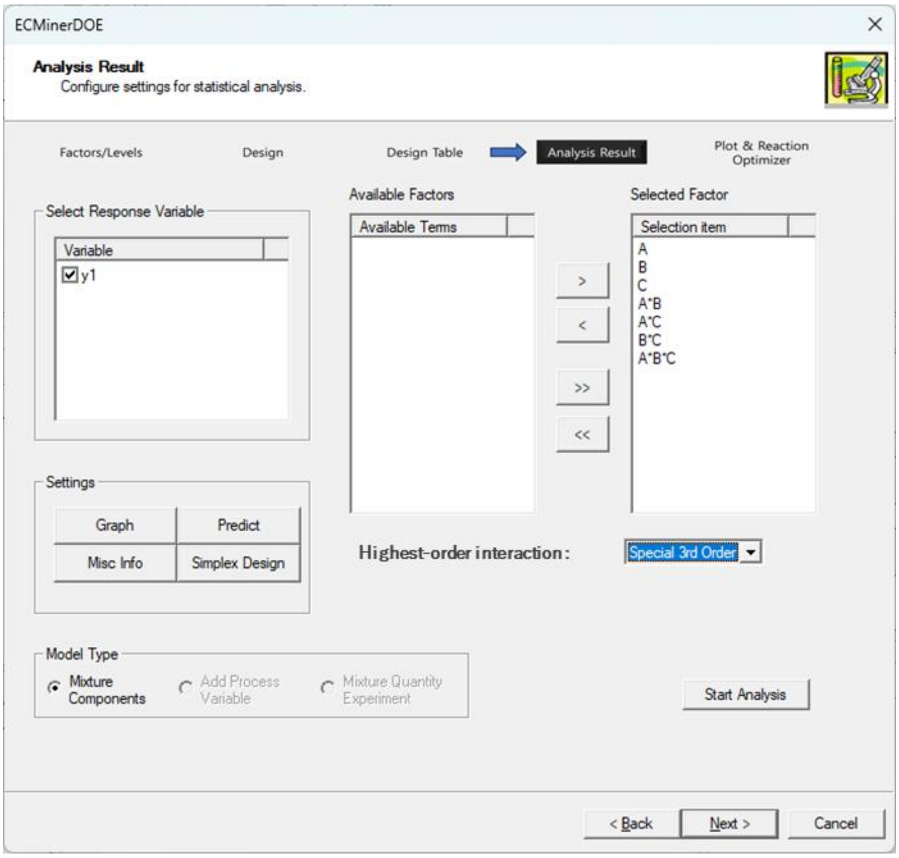


At this stage, the design table is completed, and the experimenter conducts the experiment according to the created table and inputs the resulting response values. This completes everything you need for analysis. Click the **Next** button to proceed to Step 4.

- **Step 4: Analysis Result**



Before making analysis-related settings, you can use the Simplex Design plot to see how the design points created in the design table are arranged. The simplex design plot above shows that the points are evenly spaced



Set the model in the main screen of Step 4 and start the analysis. This allows regression analysis, analysis of variance, and residual analysis. If you made a prediction on the main screen in Step 4, you can also see the predicted value.

General Info: Shows general information about the design.

Model Info: View results of regression analysis, ANOVA, and unusual observations (extreme leverage, standardized residuals).

Output Results

Experimental Design - Mixture Design : Simplex Center Design

▶ Result Analysis

● Estimated coefficients(ratio) for response value y1

Term	Coefficient	Coefficient SE	T	p
A	5.09206	0.27237	*	*
B	3.27570	0.27237	*	*
C	6.26842	0.27237	*	*
A*B	-8.26449	1.37105	-6.02786	0.00915
A*C	2.72096	1.37105	1.98459	0.14143
B*C	-3.43176	1.37105	-2.50302	0.08747
A*B*C	23.95059	9.03888	2.64973	0.07701

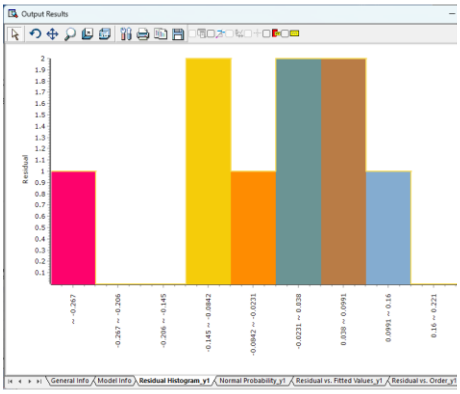
● Estimated coefficients(quantity) for response value y1

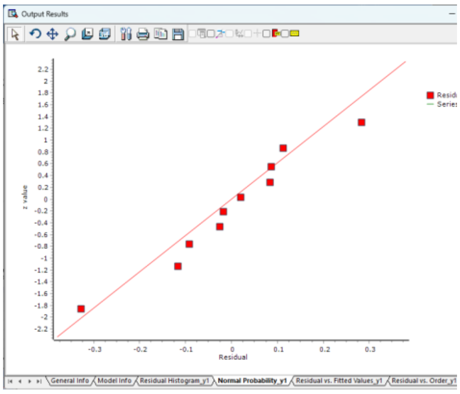
Term	Coefficient	Coefficient SE	T	p
A	5.09206	0.27237	*	*
B	3.27570	0.27237	*	*
C	6.26842	0.27237	*	*
A*B	-8.26449	1.37105	-6.02786	0.00915

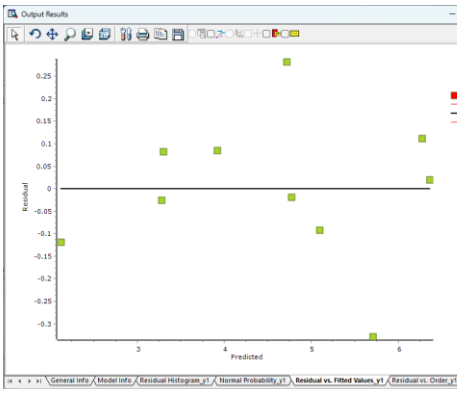
General Info | Model Info | Residual Histogram_y1 | Normal Probability_y1 | Residual vs. Fitted Values_y1 | Residual vs. Order_y1 | Prediction | Othe

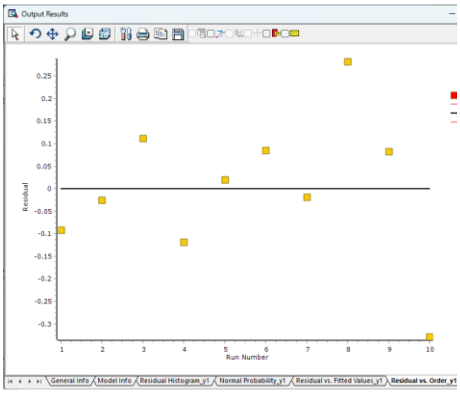
Regression analysis and analysis of variance shows that the model has high suitability, and residual analysis below shows that the normality assumption is reasonable.

Residual-related plots: View residual histograms, residual normal probability plots, residuals versus ordinal, and residuals versus fitted values.









Other Information

Shows residual related statistics.

Output Results

Experimental Design – Mixture Design : Simplex Center Design

▶ Other Information

● Fitted values and residuals for response value y1

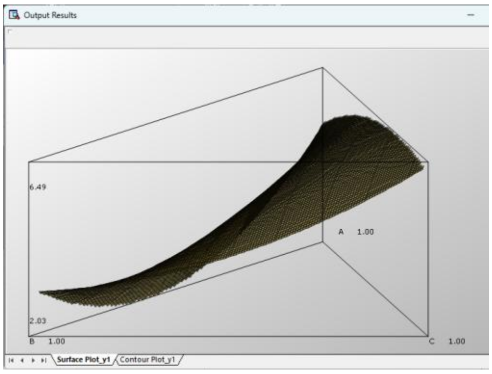
Order	Fitted Value	Residual	Standardized Residuals	External Standardized Residual	Leverage	Distance of Cook	DFITS
1	5,09206	-0,09206	-1,27658	-1,54223	0,93449	3,32108	-5,82492
2	3,27570	-0,02570	-0,35632	-0,29729	0,93449	0,25873	-1,12284
3	6,26842	0,11158	1,54725	2,81080	0,93449	4,87868	10,61622
4	2,11775	-0,11775	-1,47564	-2,30109	0,91979	3,56699	-7,79206
5	6,36048	0,01952	0,24460	0,20174	0,91979	0,09801	0,68313
6	3,91412	0,08588	1,07624	1,12154	0,91979	1,89740	3,79781
7	4,76853	-0,01853	-0,10437	-0,08537	0,60294	0,00236	-0,10520
8	4,71765	0,28235	1,17944	1,31500	0,27807	0,07655	0,81613
9	3,29674	0,08326	0,34780	0,28988	0,27807	0,00666	0,17991
10	5,70856	-0,32856	-1,37244	-1,83695	0,27807	0,10365	-1,14007

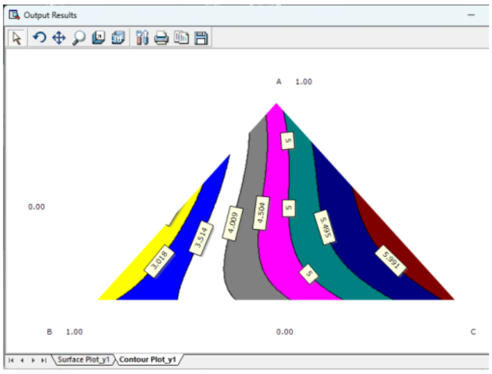
Model Info | Residual Histogram_y1 | Normal Probability_y1 | Residual vs. Fitted Values_y1 | Residual vs. Order_y1 | Prediction | Other Information

For detailed explanation, see 6.1.4. See Settings and Analysis.

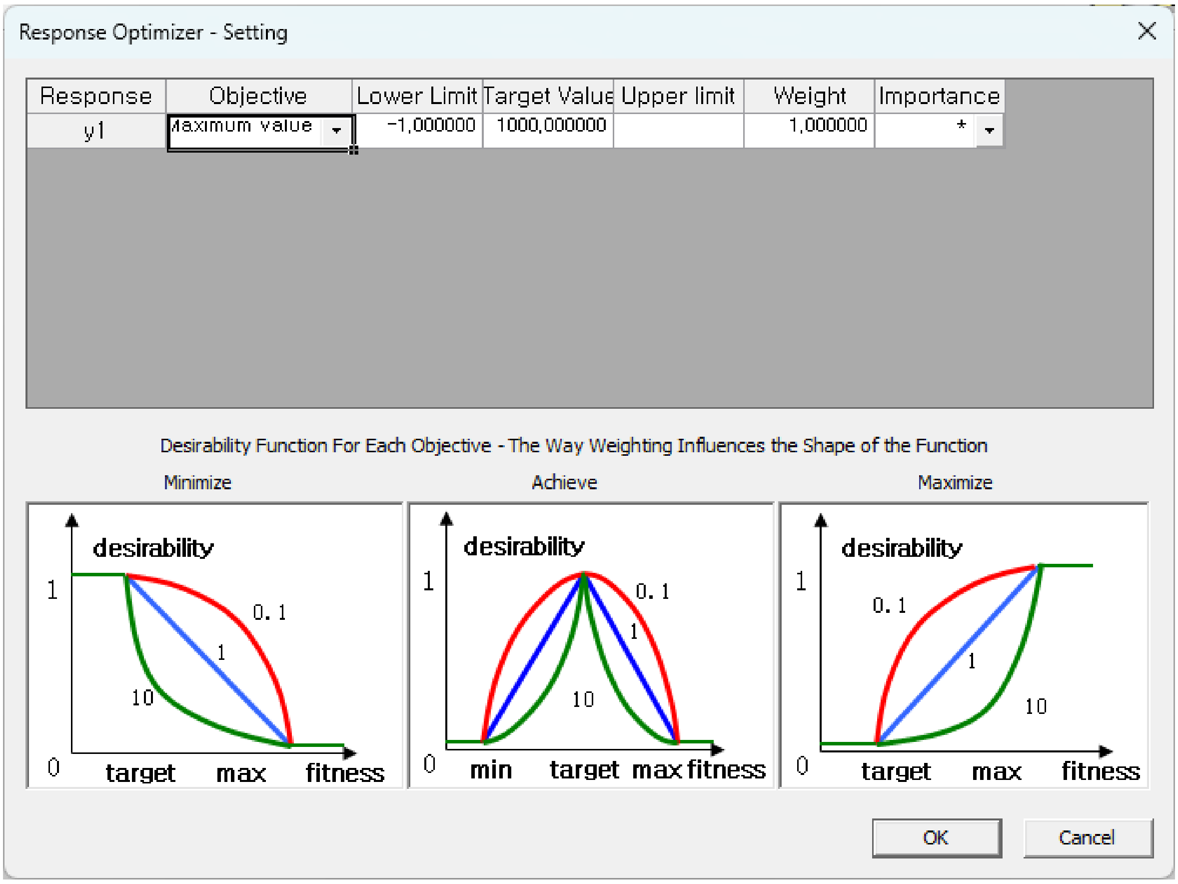
- **Step 5: Plot and Response Optimizer**

  Through Step 5 of Mixture Design, you can view surface plots and contour plots, and perform response optimizer tailored to the user's purpose. Below are the surface plot and contour plot for the Regression Model created in Step 4.

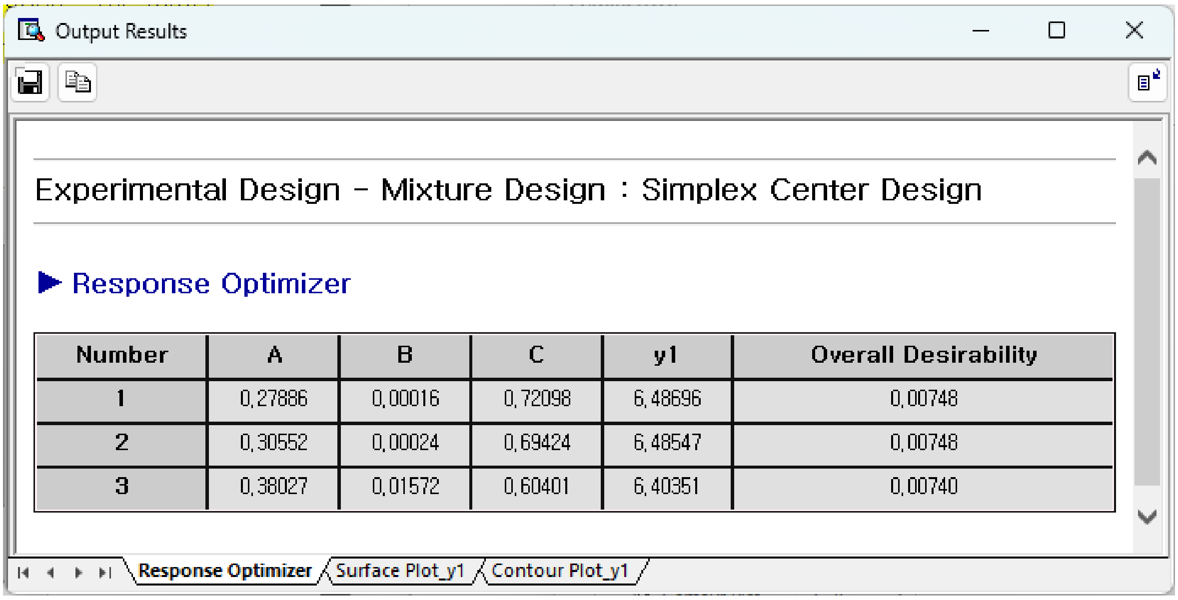




If the purpose of this experiment is to know which combination of ingredients maximizes the response value (firmness), enter the following in the setting window. Setting the lower limit at -1 means that any response value below -1 is considered equally 'not good.' The target value is set at a very large value, such as 1000, to indicate that we want to maximize the response value as much as possible.



After making the settings as above and completing the simple option settings, click **View Results** and you will get the following results.



Number	A	B	C	y1	Overall Desirability
1	0,27886	0,00016	0,72098	6,48696	0,00748
2	0,30552	0,00024	0,69424	6,48547	0,00748
3	0,38027	0,01572	0,60401	6,40351	0,00740

In other words, through response optimizer, it can be seen that the firmness value increases to 6.52244 when component A is 0.00044, component B is 0.00104, and component C is 0.99853. The experimenter checks whether the firmness value increases to a satisfactory level in this combination of components. You can end the experiment or resume the experiment.

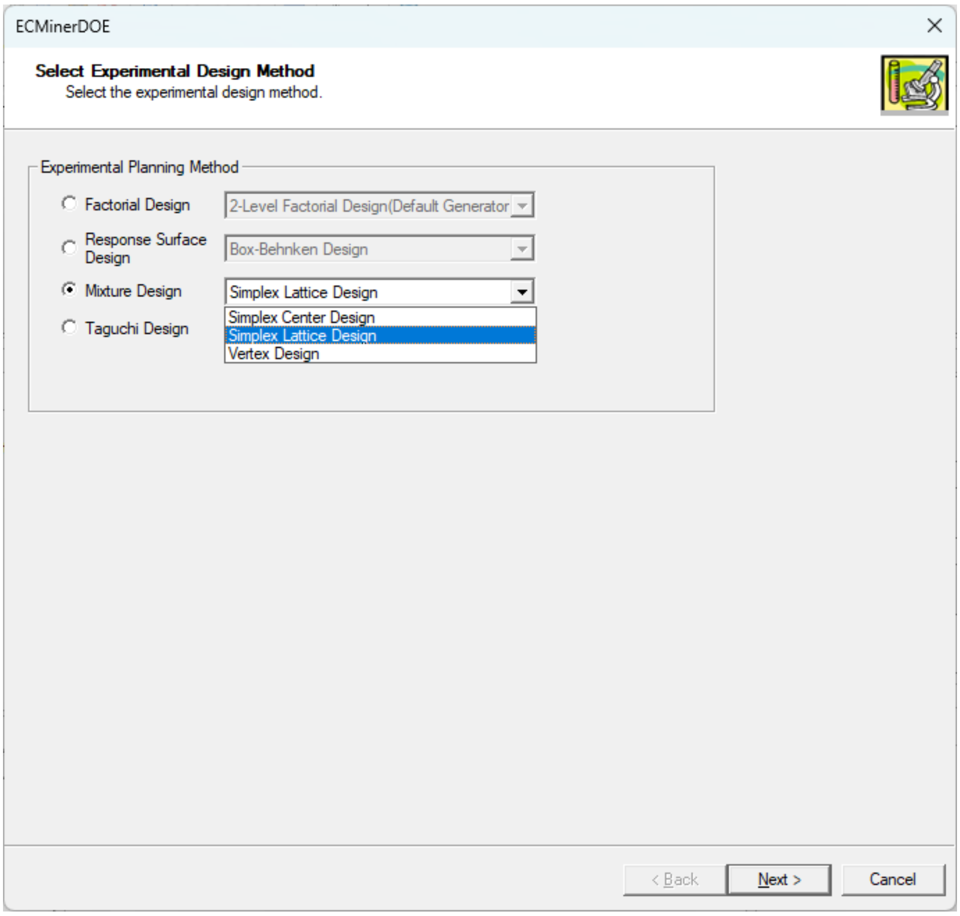
6.1.3.3.2. Simplex Lattice Design

If the goal is to conduct an experiment in a simplex region and create a regression model based on the experiment results, it would be desirable for the experiment points to be evenly distributed throughout the simplex region. The design that meets this purpose is Simplex Lattice Design. Simplex Lattice Design is a useful design for fitting m-order polynomials when there are n components.

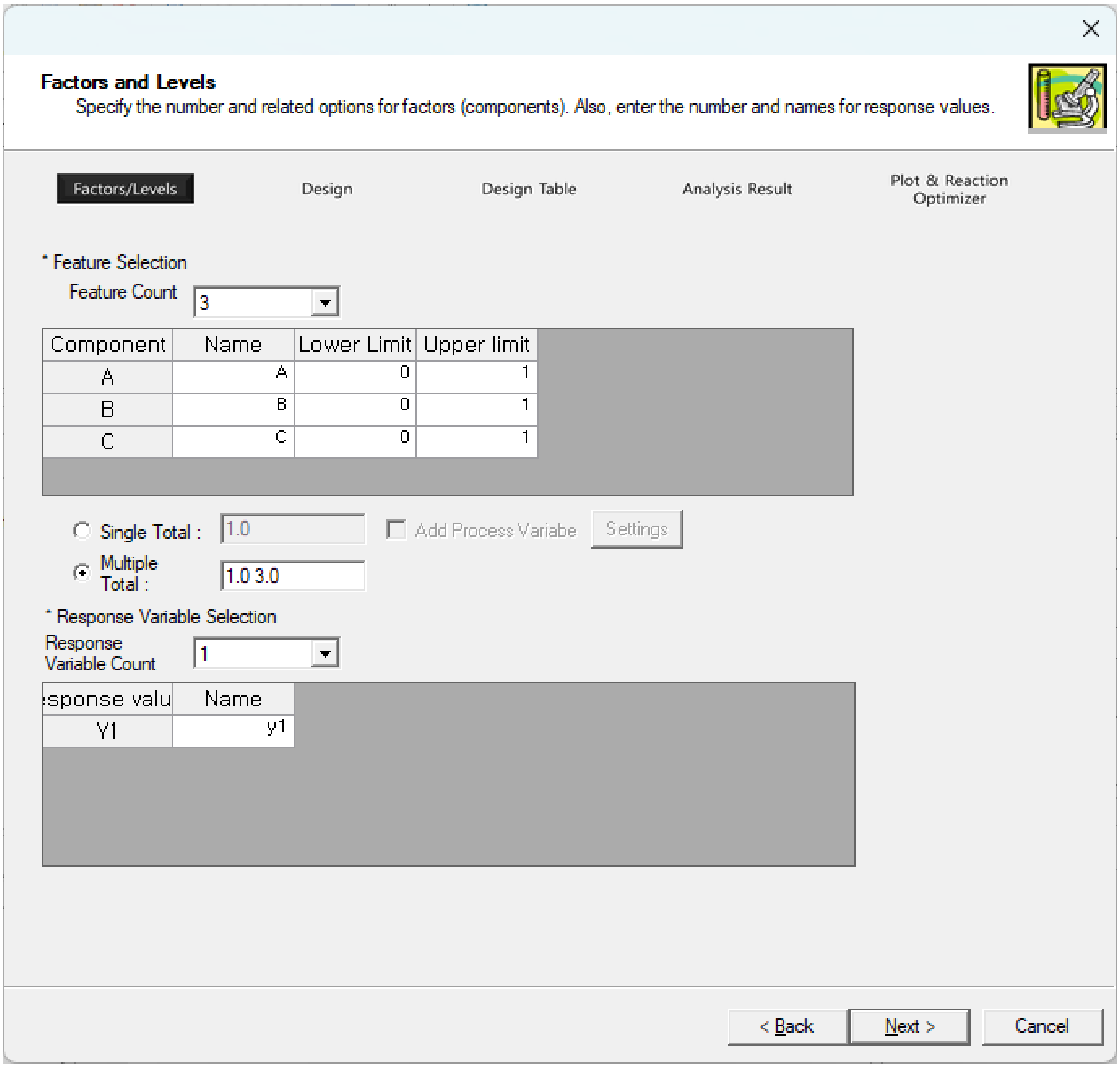
Introduction to experiments

It is said that the strength of a metal is affected not only by the ratio but also by the amount of its components. Then, simply looking at the proportions of the components does not reveal the relationship between the components and the strength of the metal. Therefore, a mixture volume design can be used that can be used in these situations.

Through Simplex Lattice Design, we will explain mixture volume design. First, **select Simplex Lattice Design**.



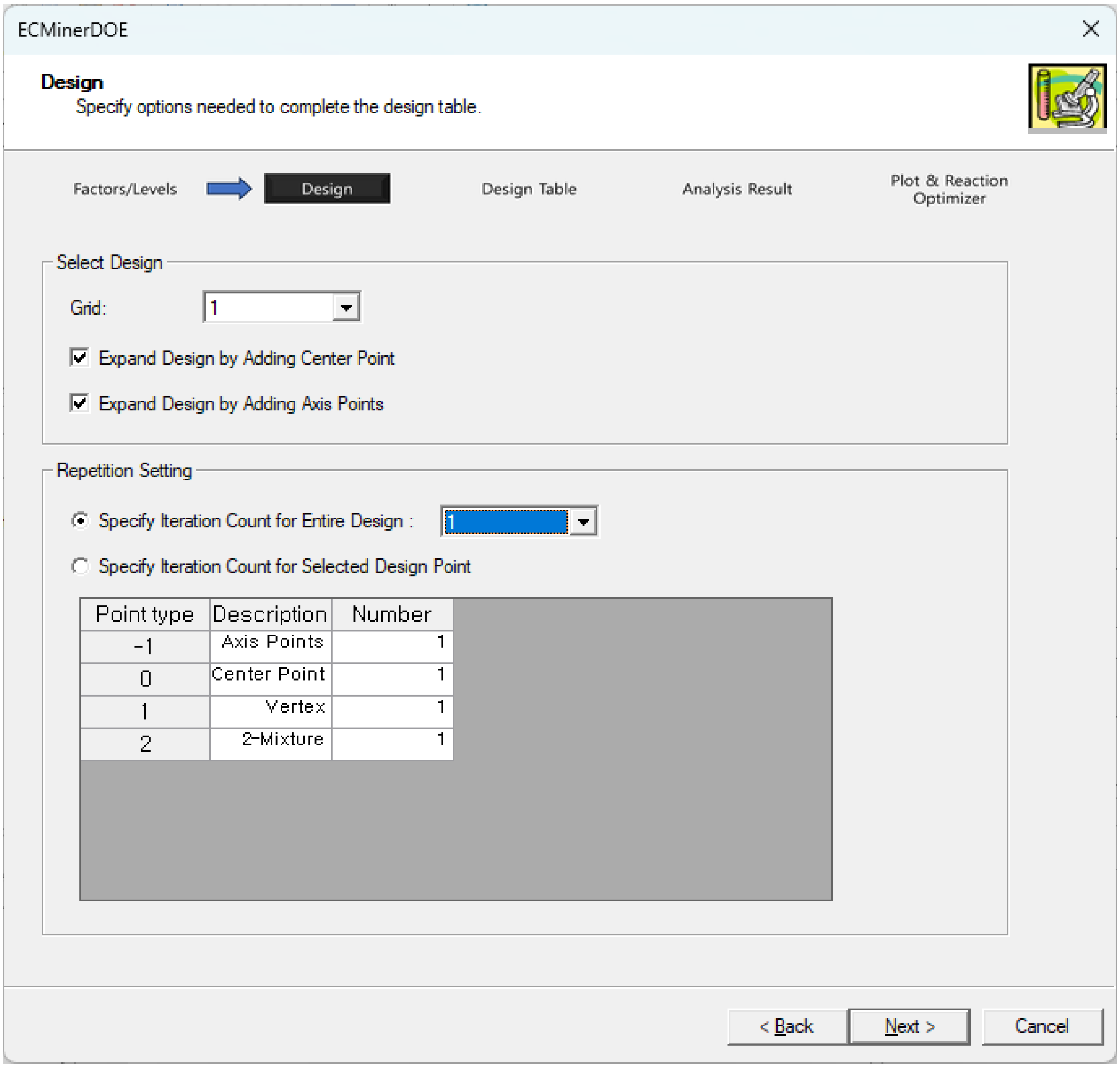
- **Step 1: Factors and Levels**



In this step, we set up the basic settings for the ingredients used in the experiment. Since this experiment is a mixture amount design, multiple total amounts are selected rather than a single total. If you think the quantity will have a quadratic or greater impact on the response value, experiment with three or more positive values. If you are sure that the quantity will have a linear effect on the response value, two quantity values are sufficient. This requires careful consideration because the smaller the number of positive values, the fewer the number of experiments.

However, as a result of the analysis of the first experiment conducted in this way, if it is confirmed that the effect of the amount of mixture is quadratic or higher, there is no need to perform the experiment perfectly from the beginning, as you only need to do a few additional experiments. After completing the settings, click the Next button to proceed to the next screen.

- **Step 2: Design**

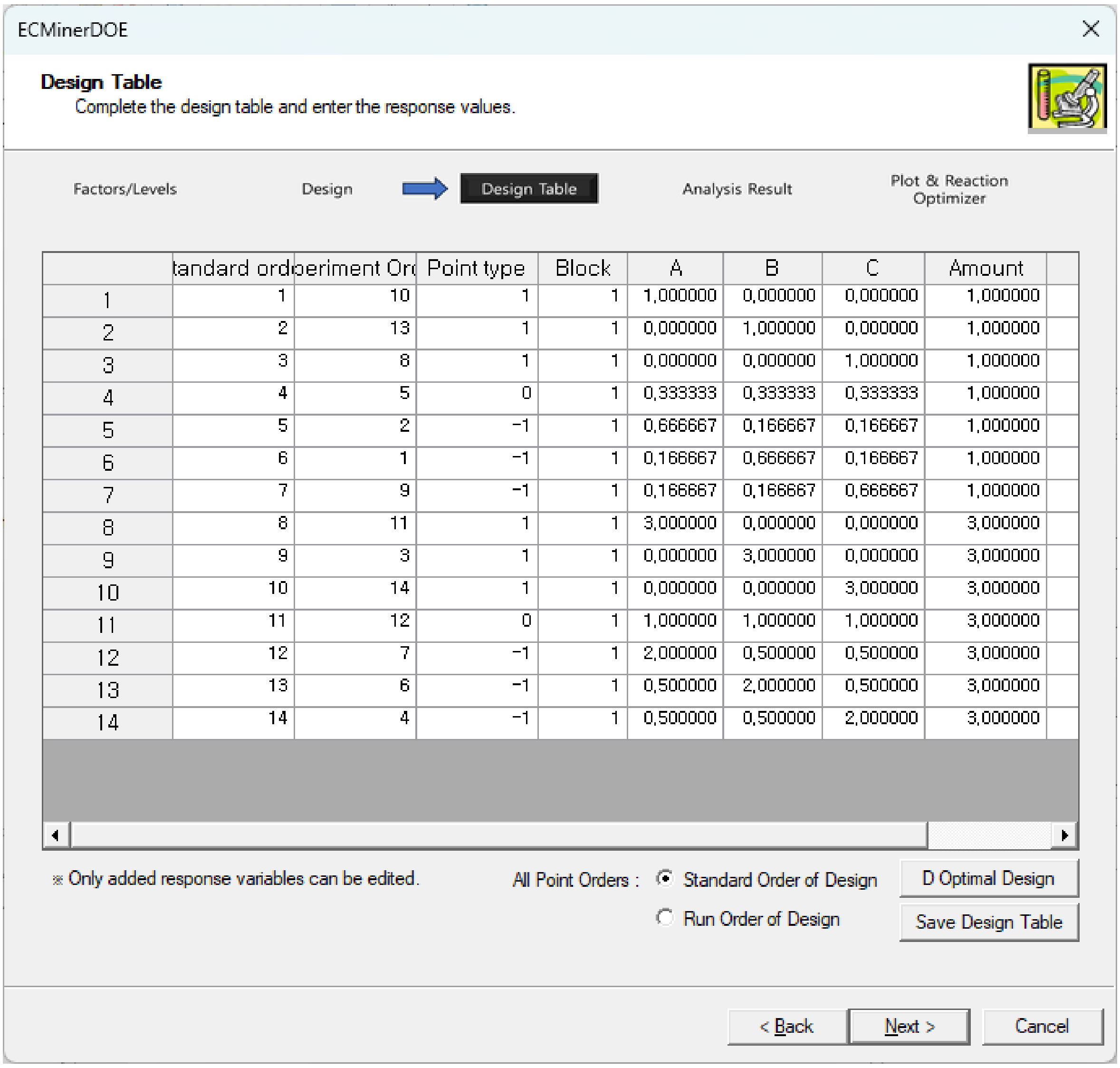


In this step, several detailed options are set to complete the design.

Grid: It is very important how you select this grid because the degree of model that can be suitable is determined depending on the grid. If you want to do the first fit, you need to set the grid to 1 or higher, and if you want to do the second fit, you need to set the grid to 2 or more. Since the current composition is expected to have a secondary effect on the strength of the metal, the grid is set to 3 to allow some margin. (As the number of Grids increases, the number of experiments increases, so be sure to keep this in mind when choosing options.)

For repeat settings, it is the same as Simplex Center Design.

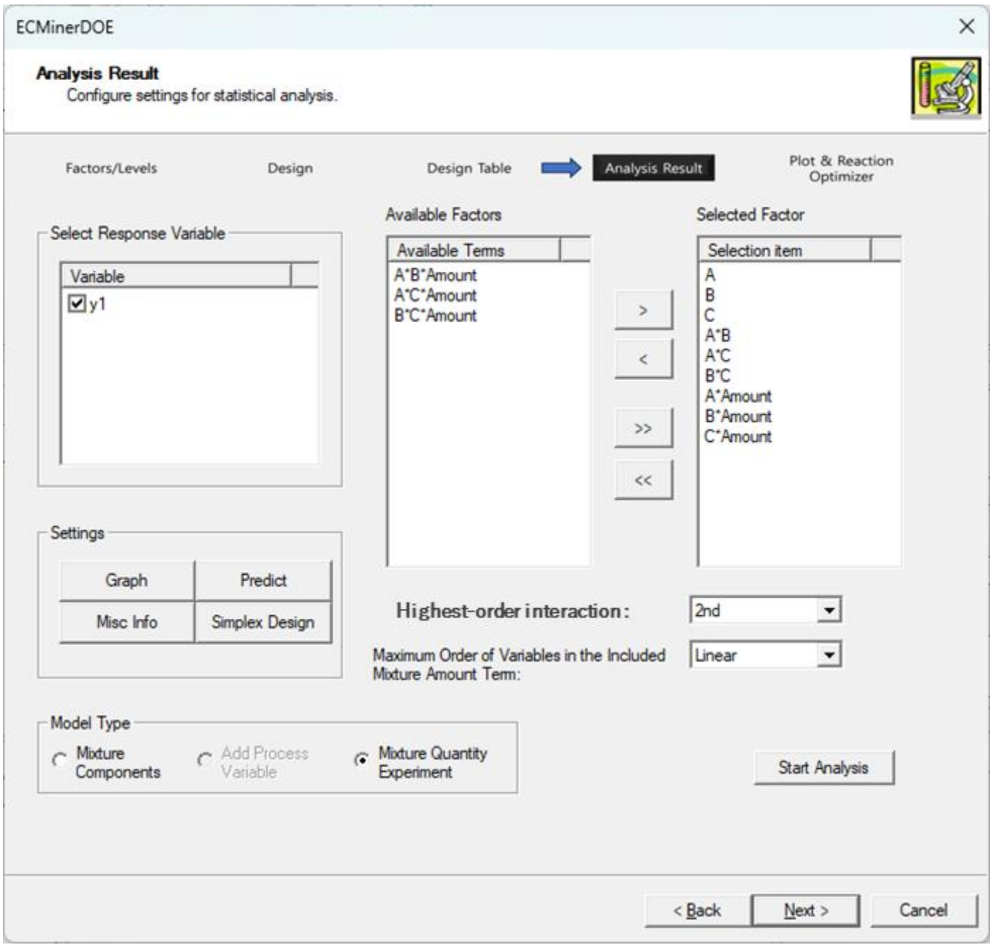
- **Step 3: Design Table**



	tandard ord	periment Or	Point type	Block	A	B	C	Amount
1	1	10	1	1	1,000000	0,000000	0,000000	1,000000
2	2	13	1	1	0,000000	1,000000	0,000000	1,000000
3	3	8	1	1	0,000000	0,000000	1,000000	1,000000
4	4	5	0	1	0,333333	0,333333	0,333333	1,000000
5	5	2	-1	1	0,666667	0,166667	0,166667	1,000000
6	6	1	-1	1	0,166667	0,666667	0,166667	1,000000
7	7	9	-1	1	0,166667	0,166667	0,666667	1,000000
8	8	11	1	1	3,000000	0,000000	0,000000	3,000000
9	9	3	1	1	0,000000	3,000000	0,000000	3,000000
10	10	14	1	1	0,000000	0,000000	3,000000	3,000000
11	11	12	0	1	1,000000	1,000000	1,000000	3,000000
12	12	7	-1	1	2,000000	0,500000	0,500000	3,000000
13	13	6	-1	1	0,500000	2,000000	0,500000	3,000000
14	14	4	-1	1	0,500000	0,500000	2,000000	3,000000

Through Step 3, the Design Table is completed, and the experimenter performs the experiment according to the given Design Table and then inputs the response value (strength). After completing your input, click the Next button to proceed to the analysis step.

- **Step 4: Analysis Result**



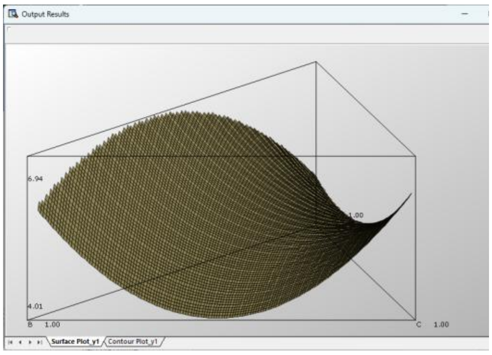
In Step 4, you can select Mixture Quantity Experiment in Model Type. If mixture components are selected, even if the amounts are different, if the proportions are the same, they are considered the same and analyzed. However, since it is clear that the actual experiment given here is one that requires consideration of the amount of mixture, we choose the experiment of amount of mixture. The maximum order of the term to be included and the maximum order of both terms of the mixture to be included are set to quadratic and linear, respectively, and terms that are known in advance to have no effect are removed as above. (Alternatively, it is a good idea to analyze by including all possible terms and then look at the analysis results and remove terms that are deemed meaningless.) The analysis details are 6.1.3.3.1. Please refer to Simplex Center Design.

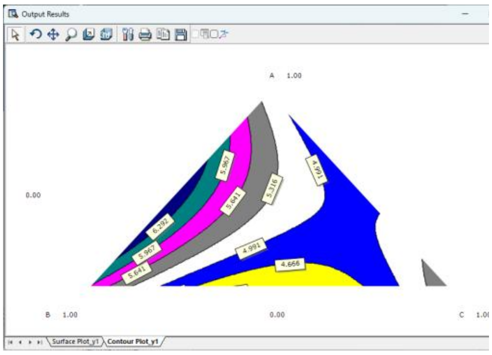
However, the Regression Model created above has the following form.

$$y = b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{1,Amount}x_1(Amount - \bar{A}) + b_{2,Amount}(Amount - \bar{A}) + b_{3,Amount}(Amount - \bar{A})$$

- **Step 5: Plot and Response Optimizer**

  In this step, the regression model created in Step 4 is expressed as a surface plot and contour plot, and a response optimizer function is provided to meet the experimenter's purpose. When the amount of mixture is 2, the plot looks like this.





The experimenter's goal is to maximize the response value (strength), so the goal is maximization and the lower limit and target values are entered as appropriate values. Again, for the same reason as in most examples, we set the lower limit to -1 and the upper limit to 1000.

Output Results

Experimental Design - Mixture Design : Simplex Lattice Design

▶ Response Optimizer

Number	A	B	C	Amount	y1	Overall Desirability
1	0.38300	0.61679	0.00021	2.96683	7.55405	0.00855
2	0.56610	0.43303	0.00087	1.83496	6.76532	0.00776
3	0.58218	0.41734	0.00048	1.20739	6.59607	0.00759

Response Optimizer

From the above results, we can see that the maximum strength of the metal can be obtained at (A, B, C) = 3*(0, 0, 1).

6.1.3.3.3. Vertex Design

Vertex Design refers to finding the vertices of the Convex Set that satisfy the upper and lower limit conditions and linear constraints when these conditions are added, augmenting the experiment with these vertices, and then performing experiments on these points. Intuitively, we can see that when there are only upper and lower limit conditions and linear constraints in an n-dimensional space, it is not an easy process to find the vertices of the space created by these conditions. Therefore, this cannot be obtained through a simple process, but Piepel (1988) presented an algorithm to obtain these vertices.

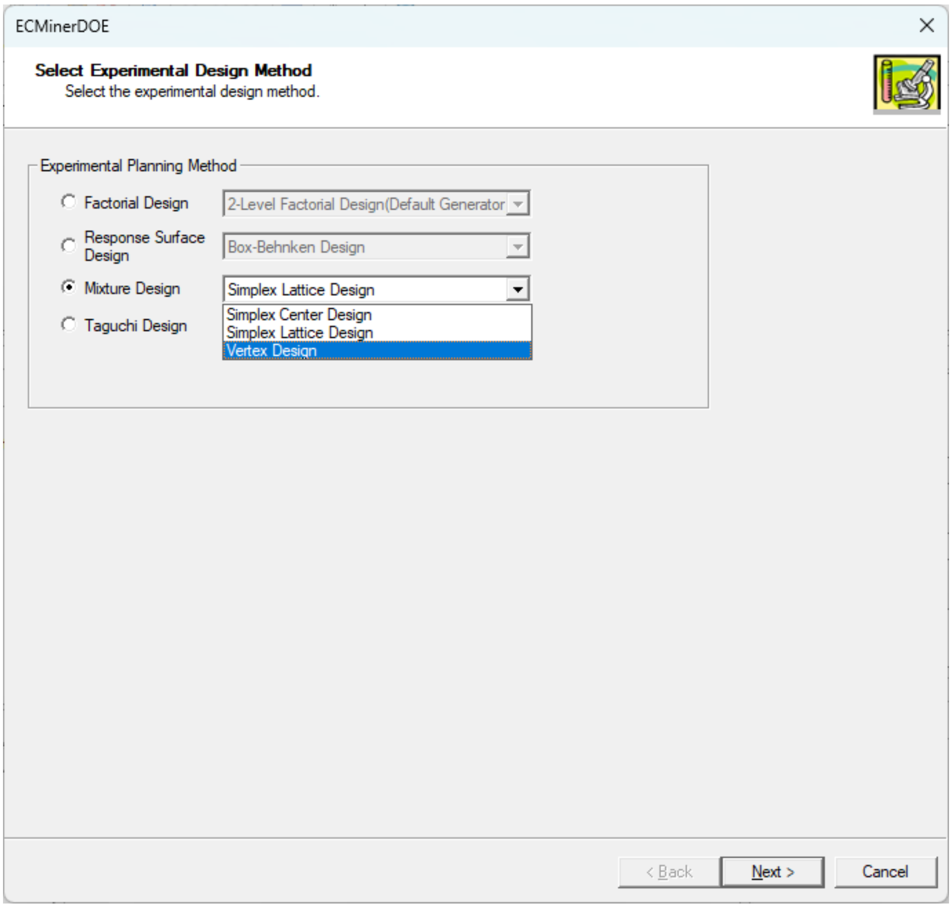
We will explain how to add process variables to Mixture Design through Vertex Design.

Introduction to experiments

This experiment aims to find out the relationship between three ingredients and how process variables (working temperature) affect the quality of the product. At this time, the limiting conditions for the three ingredients are as follows.

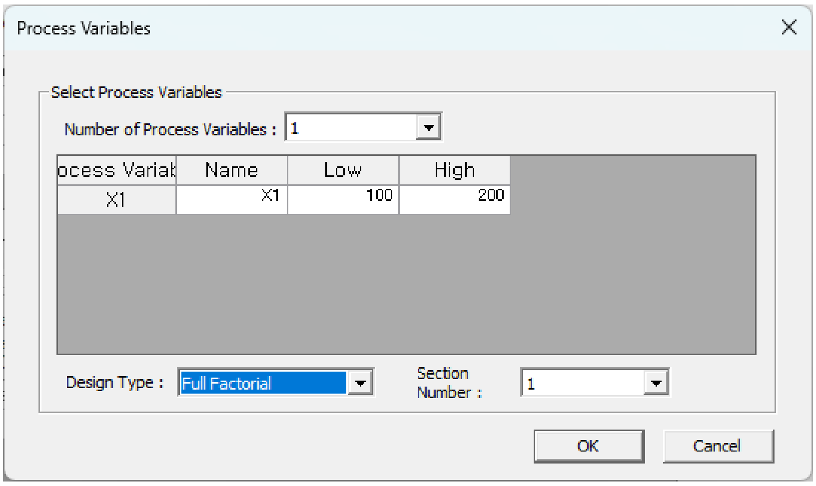
$$0.1 \le x_1 \le 0.8, \qquad 0.0 \le x_2 \le 0.6, \qquad 0.0 \le x_3 \le 0.5$$

First, select Vertex Design.

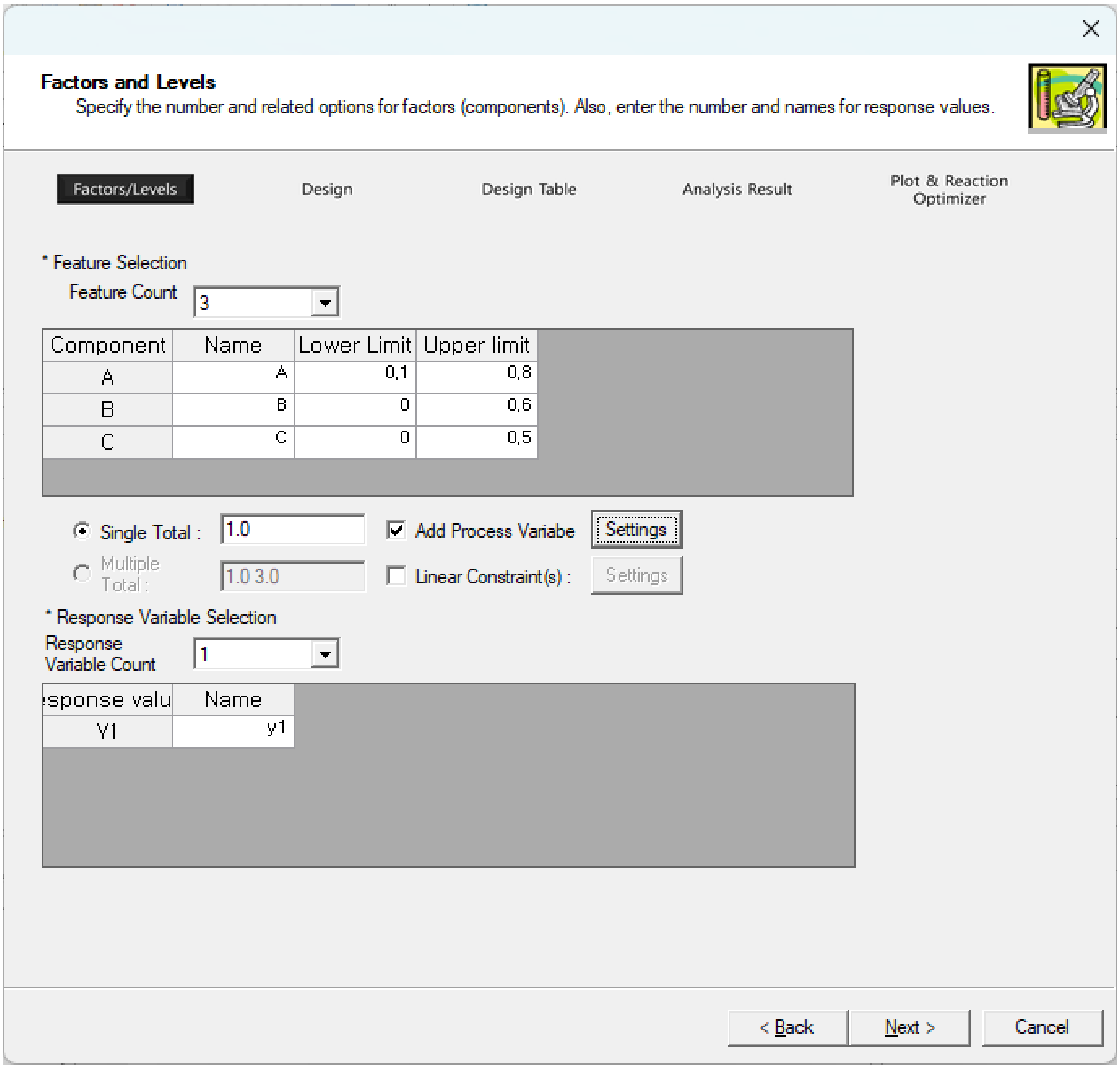


- **Step 1: Factors and Levels**

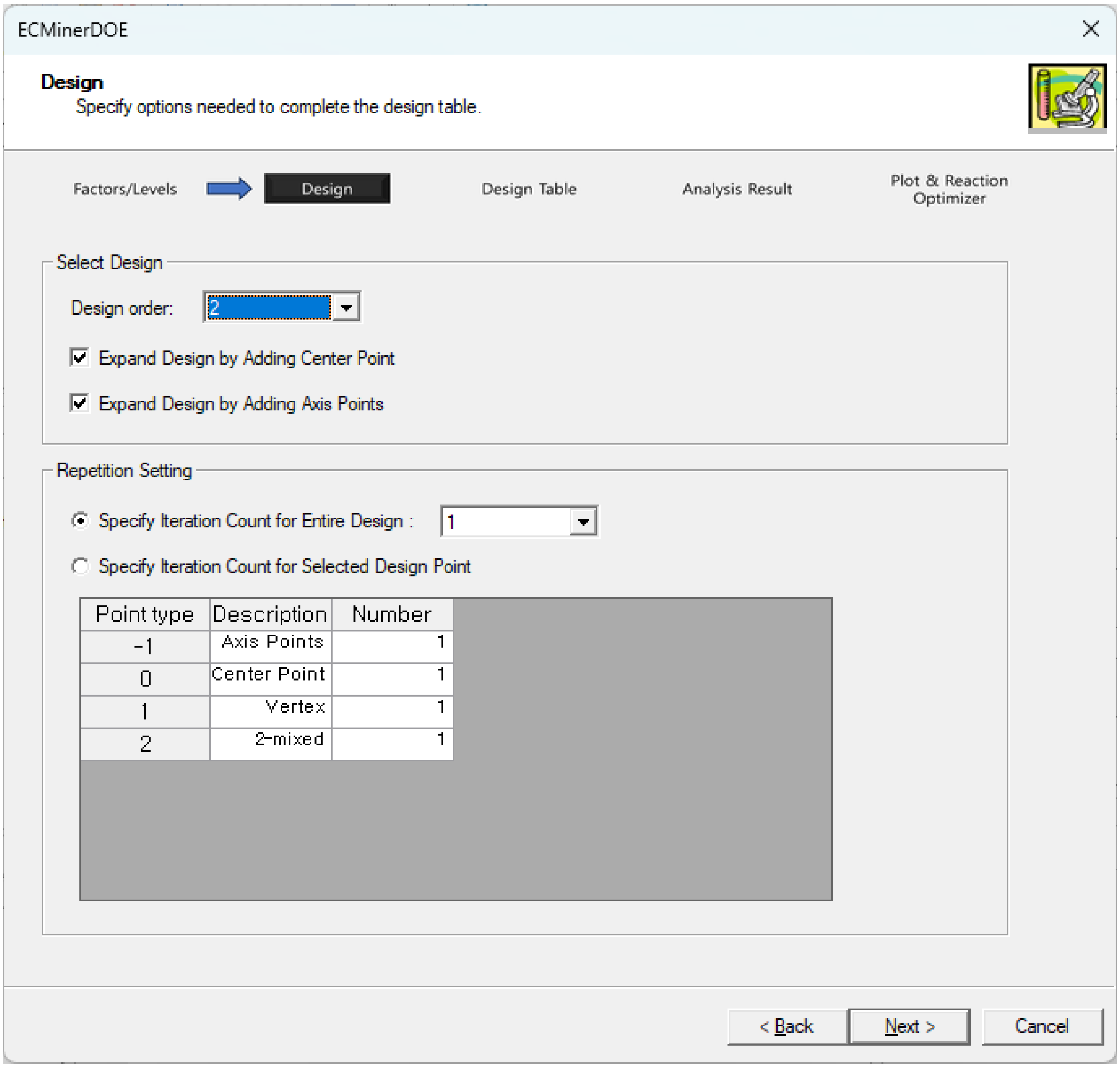
  First, add a process variable by selecting the Add Process Variable checkbox. Then click **Setting** button to view the pop-up window as shown below.



Then, enter the constraints for the ingredients in the main screen of Step 1 and click the **Next** button.



- **Step 2: Design**



Several options can be specified in the Design selection step.

Design order: Take the current vertex and decide to what degree the point should be augmented.

Expand Design by Adding Center Point: Determine whether to add a center point

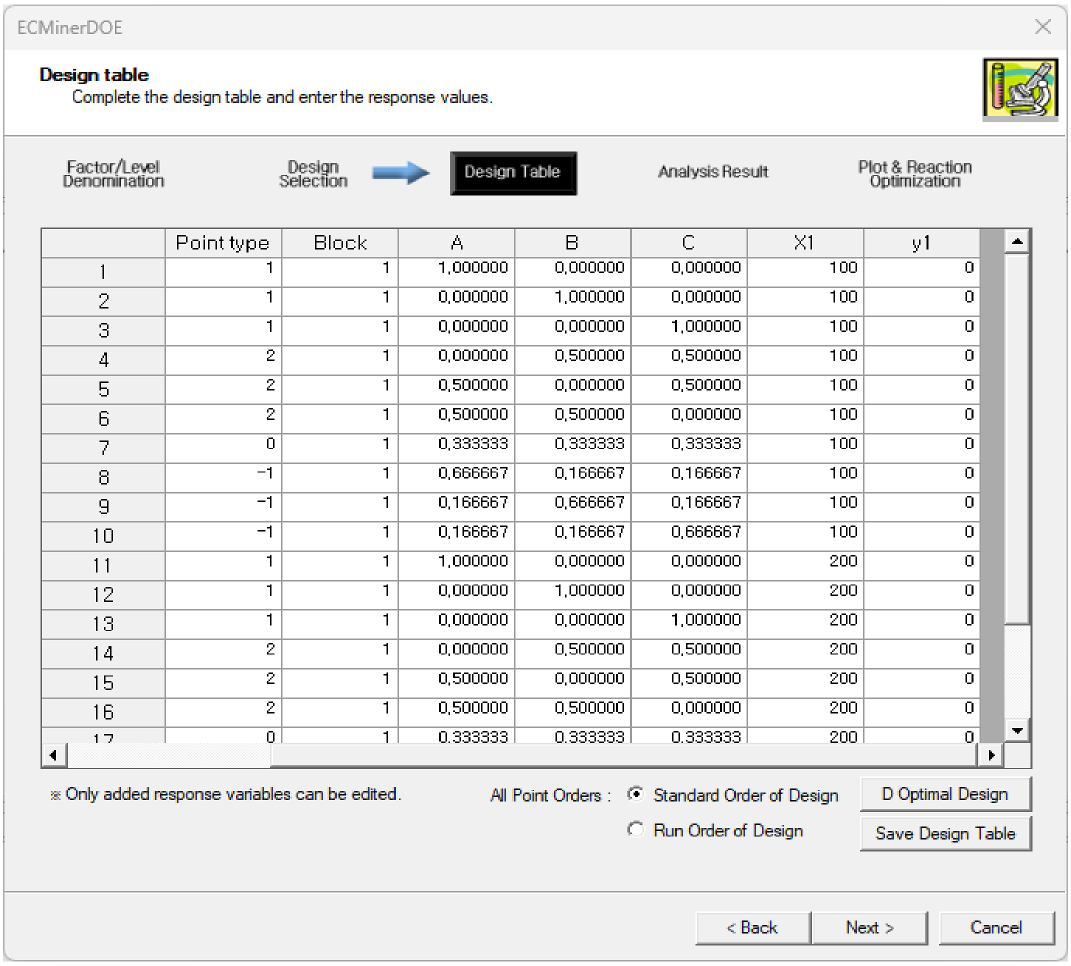
Expand Design by Adding Axis Points: Determine whether to add an axis point, which is the midpoint between the center point and each vertex

Repetition Setting

- **Specify iteration Count for Entire Design**: Specifies how many times to repeat the same entire experiment

- Specify iteration Count for Selected Design Point: Depending on the type of point, the number of repetitions varies.

- **Step 3: Design Table**

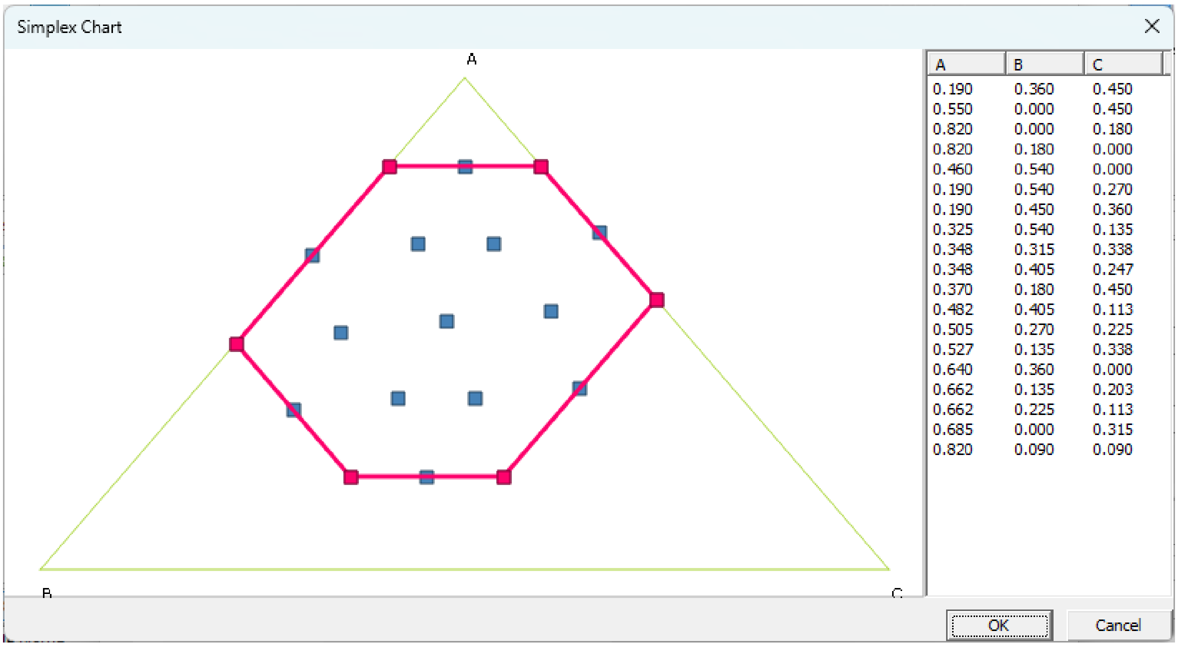


	Point type	Block	A	B	C	X1	y1
1	1	1	1,000000	0,000000	0,000000	100	0
2	1	1	0,000000	1,000000	0,000000	100	0
3	1	1	0,000000	0,000000	1,000000	100	0
4	2	1	0,000000	0,500000	0,500000	100	0
5	2	1	0,500000	0,000000	0,500000	100	0
6	2	1	0,500000	0,500000	0,000000	100	0
7	0	1	0,333333	0,333333	0,333333	100	0
8	-1	1	0,666667	0,166667	0,166667	100	0
9	-1	1	0,166667	0,666667	0,166667	100	0
10	-1	1	0,166667	0,166667	0,666667	100	0
11	1	1	1,000000	0,000000	0,000000	200	0
12	1	1	0,000000	1,000000	0,000000	200	0
13	1	1	0,000000	0,000000	1,000000	200	0
14	2	1	0,000000	0,500000	0,500000	200	0
15	2	1	0,500000	0,000000	0,500000	200	0
16	2	1	0,500000	0,500000	0,000000	200	0
17	0	1	0,333333	0,333333	0,333333	200	0

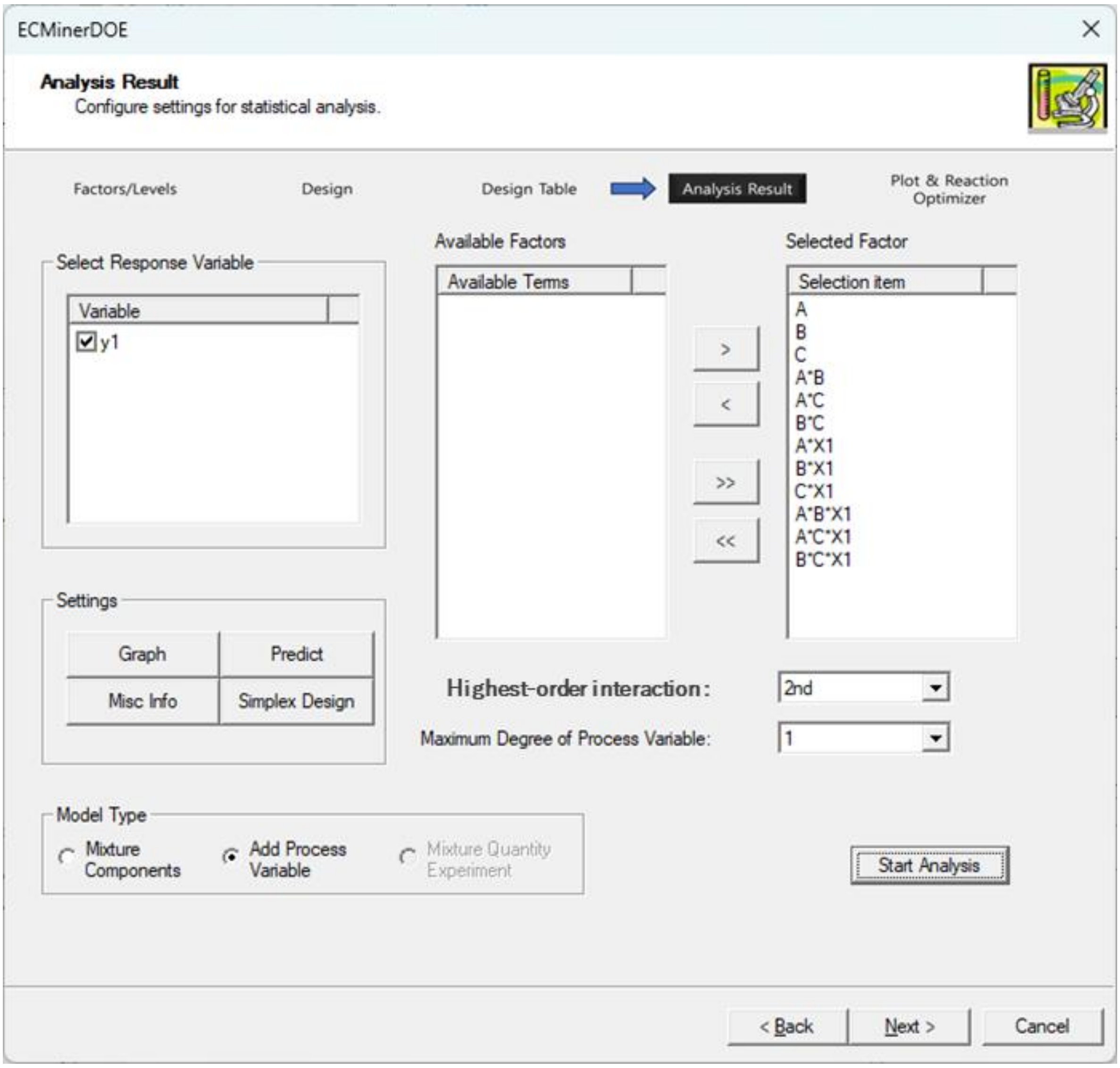
At the Design Table stage, a Design Table is created, the experimenter conducts an experiment according to the created design, and then inputs the response values. After completing the input, click the **Next** button to proceed to Step 4.

- **Step 4: Analysis Result**

Before starting Analysis Result, check how the design points created through the current algorithm are distributed in space through a Simplex Design plot.



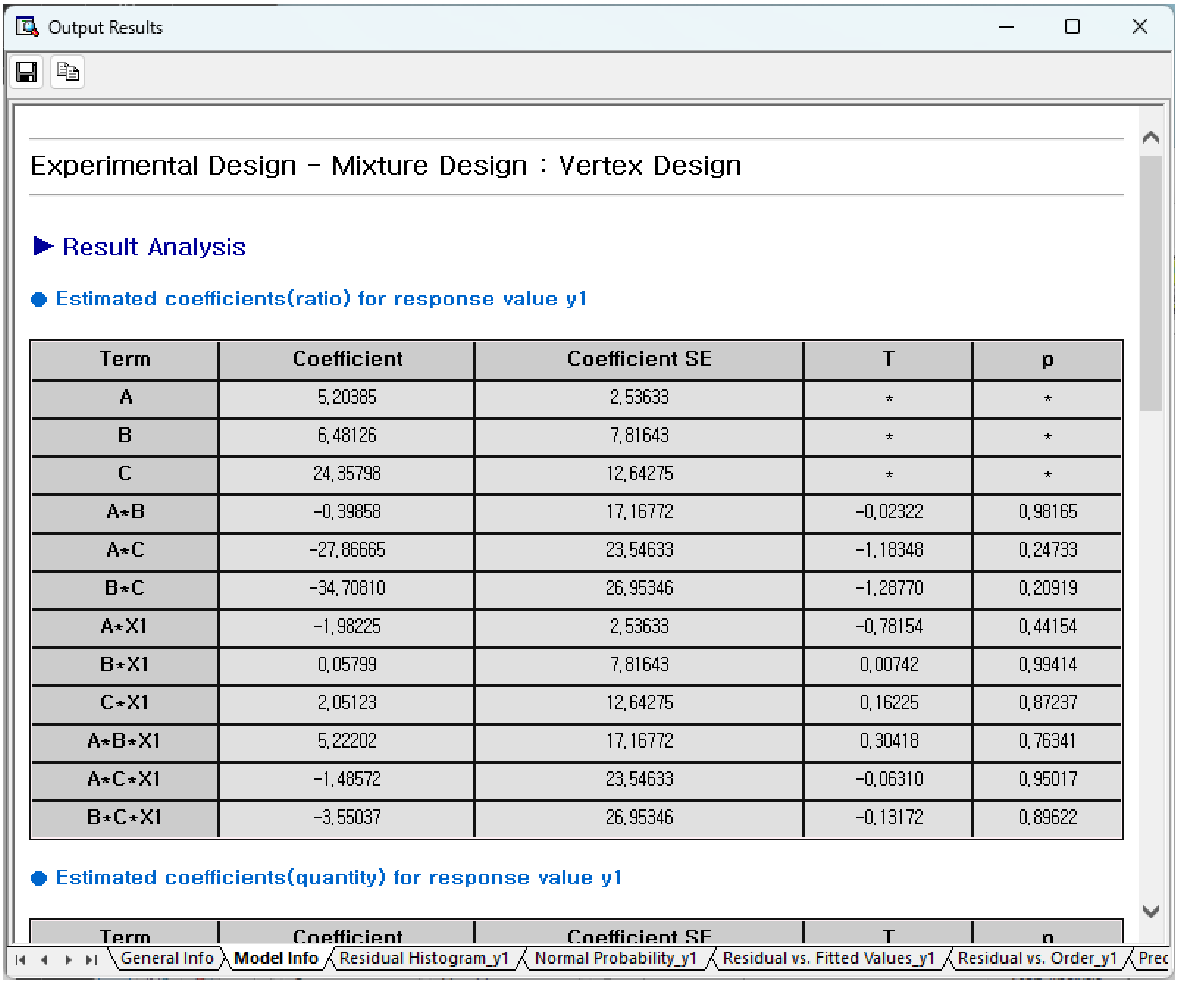
Although the arrangement of the design points is not triangular due to the limiting conditions, they are arranged evenly in the limited space.



For Model Type, select Add process variable and enter the maximum order of terms to include and the maximum order of the process variable terms to be included as required. Then you can get the following result:

General Info: Provides basic information about design.

Model Info: Provides information on regression, analysis of variance, and unusual observations (extreme leverage, standardized residuals).



Output Results

Experimental Design - Mixture Design : Vertex Design

Result Analysis

Estimated coefficients(ratio) for response value y1

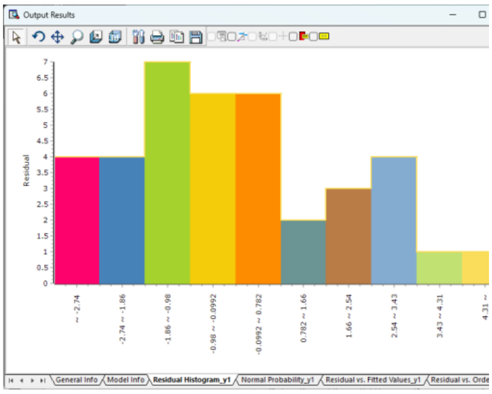
Term	Coefficient	Coefficient SE	T	p
A	5.20385	2.53633	*	*
B	6.48126	7.81643	*	*
C	24.35798	12.64275	*	*
A*B	-0.39858	17.16772	-0.02322	0.98165
A*C	-27.86665	23.54633	-1.18348	0.24733
B*C	-34.70810	26.95346	-1.28770	0.20919
A*X1	-1.98225	2.53633	-0.78154	0.44154
B*X1	0.05799	7.81643	0.00742	0.99414
C*X1	2.05123	12.64275	0.16225	0.87237
A*B*X1	5.22202	17.16772	0.30418	0.76341
A*C*X1	-1.48572	23.54633	-0.06310	0.95017
B*C*X1	-3.55037	26.95346	-0.13172	0.89622

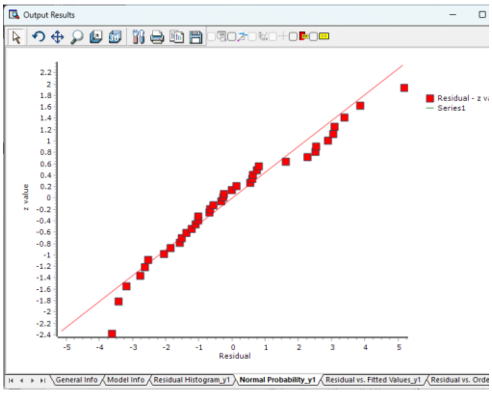
Estimated coefficients(quantity) for response value y1

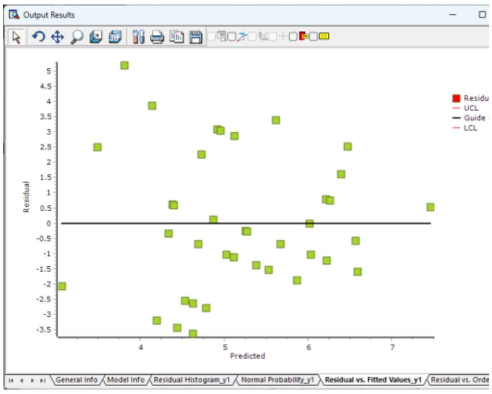
Term	Coefficient	Coefficient SE	T	p

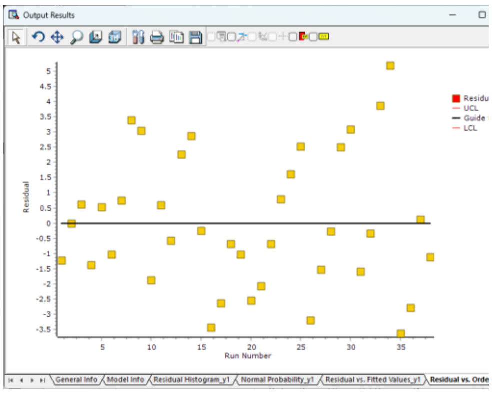
General Info | Model Info | Residual Histogram_y1 | Normal Probability_y1 | Residual vs. Fitted Values_y1 | Residual vs. Order_y1 | Prec

Residual related (residual histogram, residual normal probability plot, residual vs. ordered, residual vs. fitted Value)









Other Information

organize and displays other residual-related information.

Output Results

Experimental Design – Mixture Design : Vertex Design

▶ Other Information

● Fitted values and residuals for response value y1

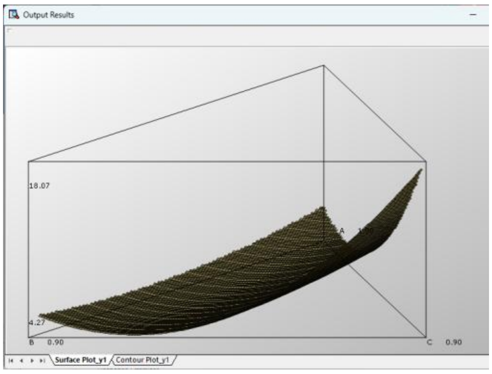
Order	Fitted Value	Residual	Standardized Residuals	External Standardized Residual	Leverage	Distance of Cook	DFITS
1	6,21919	-1,21919	-0,63761	-0,63017	0,47187	0,03027	-0,59567
2	6,01400	-0,01400	-0,00697	-0,00683	0,41744	0,00000	-0,00579
3	4,38393	0,61607	0,30677	0,30136	0,41744	0,00562	0,25510
4	5,37801	-1,37801	-0,79106	-0,78520	0,56168	0,06682	-0,88886
5	7,46112	0,53888	0,30935	0,30390	0,56168	0,01022	0,34402
6	6,02820	-1,02820	-0,53773	-0,53024	0,47187	0,02153	-0,50121
7	6,25677	0,74323	0,33898	0,33314	0,30562	0,00421	0,22101
8	5,61650	3,38350	1,56187	1,60887	0,32213	0,09660	1,10908
9	4,95369	3,04631	1,37191	1,39677	0,28780	0,06338	0,88791

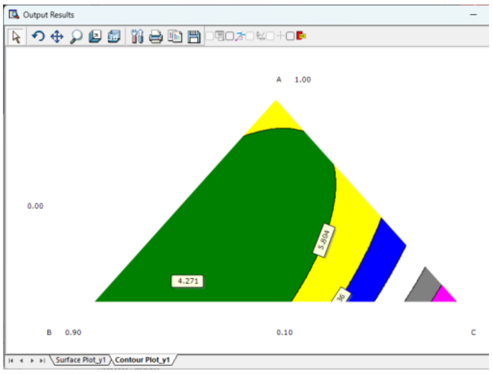
Residual Histogram_y1 / Normal Probability_y1 / Residual vs. Fitted Values_y1 / Residual vs. Order_y1 / Prediction / **Other Information**

For more information, see 6.4. See Settings and Analysis.

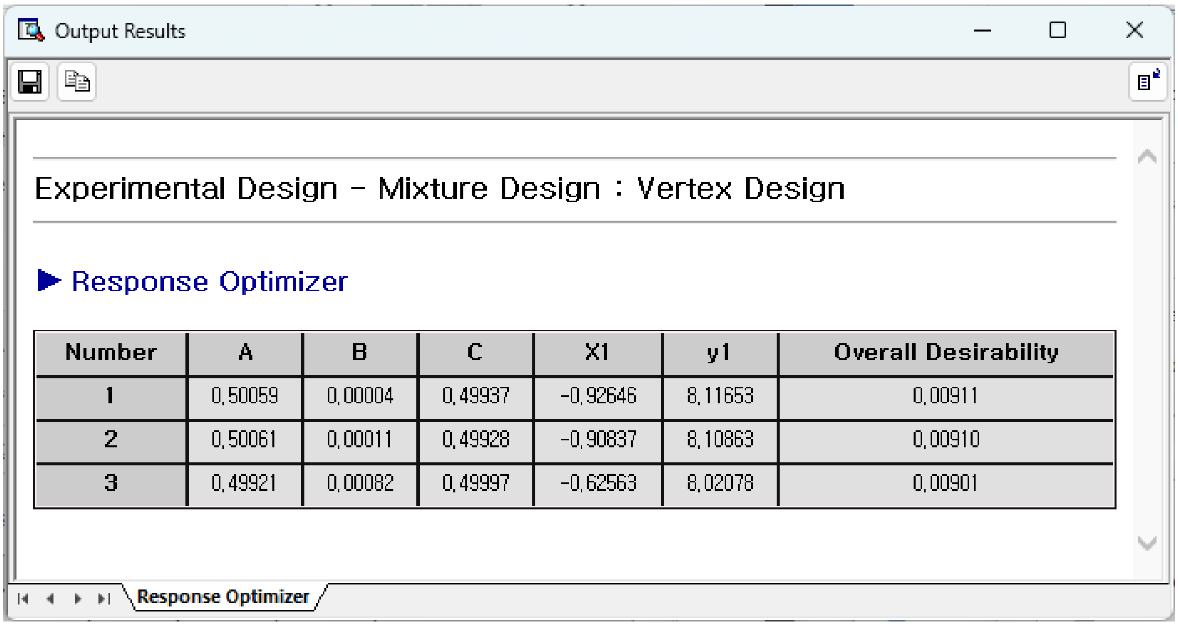
- **Step 5: Plot and Response Optimizer**

  In the Step 5, it shows what the Regression Model created in Step 4 looks like through surface plots and contour plots.





Response optimizer is performed according to the experimenter's objectives. The goal of the current experiment is to maximize quality, so if the goal is maximized and the optimization is performed with the lower limit and target value as -1 and 1000, respectively, the following results can be obtained.


Output Results

Experimental Design - Mixture Design : Vertex Design

▶ Response Optimizer

Number	A	B	C	X1	y1	Overall Desirability
1	0.50059	0.00004	0.49937	-0.92646	8.11653	0.00911
2	0.50061	0.00011	0.49928	-0.90837	8.10863	0.00910
3	0.49921	0.00082	0.49997	-0.62563	8.02078	0.00901

Response Optimizer

In other words, you can see that the maximum quality of 36.05873 can be obtained when Components (A, B, C) = (0.5, 0, 0.5) and Process Variable = 1 (working temperature = 100 degrees).

6.1.3.3.4. Mixture Design and D Optimal Design

Mixture Design and D Optimal Design is a method that optimizes the design for the best process (e.g., food ingredients, chemical blends, or formulations). D-Optimal Mixture Design combines the principles of mixture designs and D-optimality to create an efficient experimental plan for mixture experiments.

D Optimality (Determinant)

D Optimality is the most commonly used criterion and is used to find a design that maximizes the Determinant of the $X^T X$ inverse matrix. When we look for the Design Matrix X created by gathering the necessary candidate points from a set of several candidate points, the design table that makes the Determinant of the $X^T X$ inverse matrix the largest is called D-Optimal Design.

A Optimality (Trace)

When we look for the Design Matrix X created by gathering the necessary candidate points from a set of several candidate points, the design table that creates the largest TRACE of the $X^T X$ inverse matrix is called A-Optimal Design.

A-Optimality is not widely used due to computational difficulties.

G Optimality (Average Leverage / Maximum Leverage)

G Optimality means average leverage divided by maximum leverage. Leverage refers to the Generalized Linear Model Matrix as X, and the H Matrix as

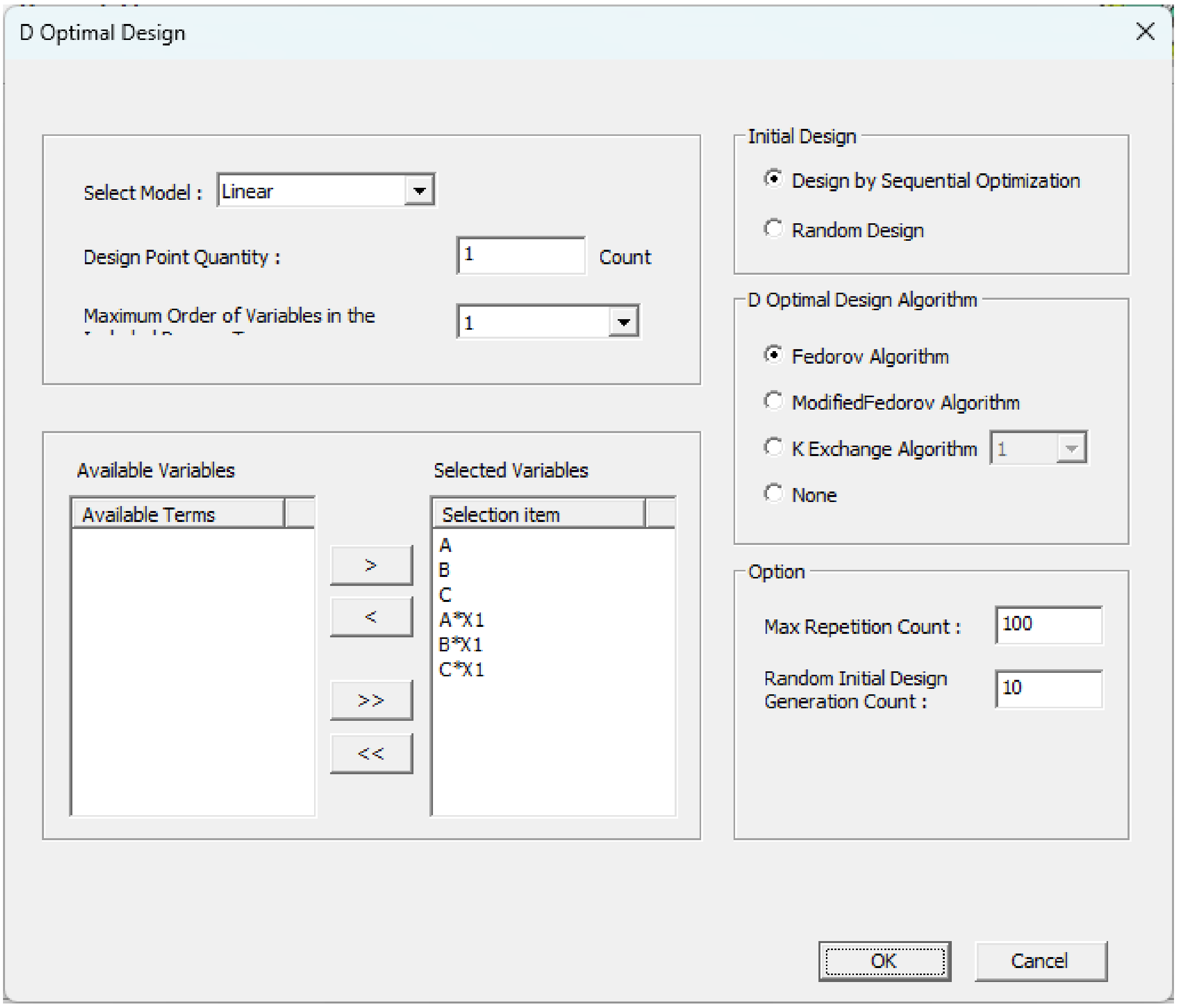
$$H = X(X^T X)^{-1} X^T$$

this refers to their diagonal components. Dividing the average of these leverages by the maximum value is G-Optimality.

V Optimality

The average of leverage is V Optimality.

D Optimal Design is used to maximize D Optimality. The D Optimal Design of response surface design is composed as follows.



- **Select Model**

  First select a model. Depending on the model selected, the selection terms in the selection window below will change, and only some of them may be used.

- **Design Point Quantity at D Optimal Design**

  Determine how many design points there will be in the modified design table.

- **Maximum Degree of Process Variable Terms Included**

  Use the option to experiment with adding a Process Variable in Mixture Design Select up to the 2nd order if there are 2 or more Process Variables, or select the 1st order if there is only 1 Process Variable.

- **Initial Design**

Select an initial design to perform Optimal Design. Set how to choose an initial design with the same number of design points as in Optimal Design.

- **D Optimal Design Algorithm**

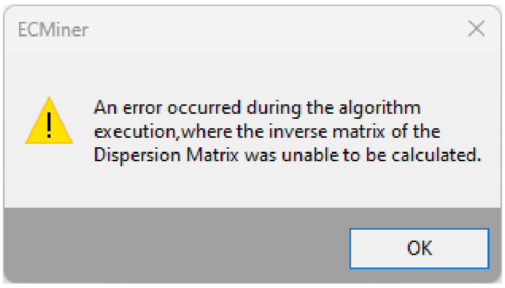
  Decide which algorithm will be used to improve the initial design to obtain D Optimal Design. ECMiner™ DOE provides Fedorov Algorithm, Modified Fedorov Algorithm, and K Exchange Algorithm. Among these, Modified Fedorov Algorithm and K Exchange Algorithm are known to have good performance.

- **Option**

  Max Repetition Count refers to the maximum number of repetitions to be performed in D Optimal Design Algorithm. Random Initial Design Generation Count refers to how many random designs are created when randomly designing the initial design. At this time, the algorithm is performed using the one with the largest D Optimality as the initial design among the various random designs created.

- **Caution**

  Unlike D Optimal Design of response surface design, in D Optimal Design of Mixture Design, the following message often appears during algorithm execution. This means that the Determinant is near 0 when the inverse matrix of the dispersion matrix is obtained during the algorithm execution process. In situations like this, if you select a design model by lowering the order of the model, this error will not appear.



6.1.3.4 Taguchi Design

The Taguchi design of experiment has several roles, especially as an expanded one compared to the conventional design of experiment.

Previously, it was difficult to evaluate the degree of influence on data by causes such as uncontrollable environmental conditions, difficult-to-control production conditions, and process conditions (these are collectively referred to as noise factors), but it is gradually becoming possible to evaluate this objectively and quantitatively. In Taguchi's experimental design, this is evaluated using the SNR. In other words, the Taguchi experiment plan finds the optimal experiment combination that can reach the operator's desired conditions in an experiment environment that includes noise. Additionally, the number of experiments is dramatically reduced through the use of orthogonal array tables.

In addition to the above features, the Taguchi design has many advantages. The Taguchi method provided by ECMiner™ DOE is as follows:

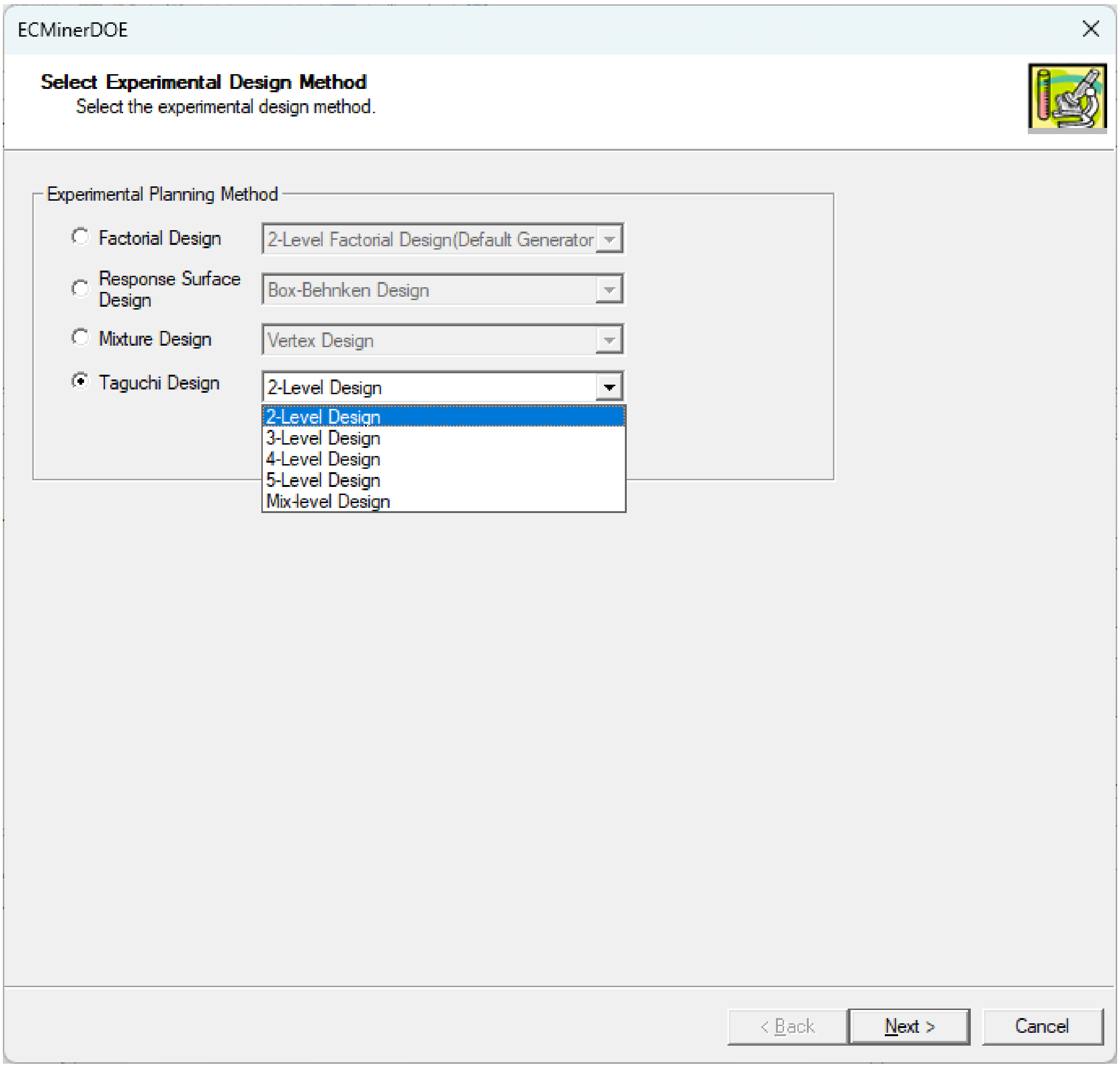
- **2- Level Design**
- **3- Level Design**
- **4- Level Design**
- **5- Level Design**
- **Mix-Level Design**

6.1.3.4.1. 2- Level Design

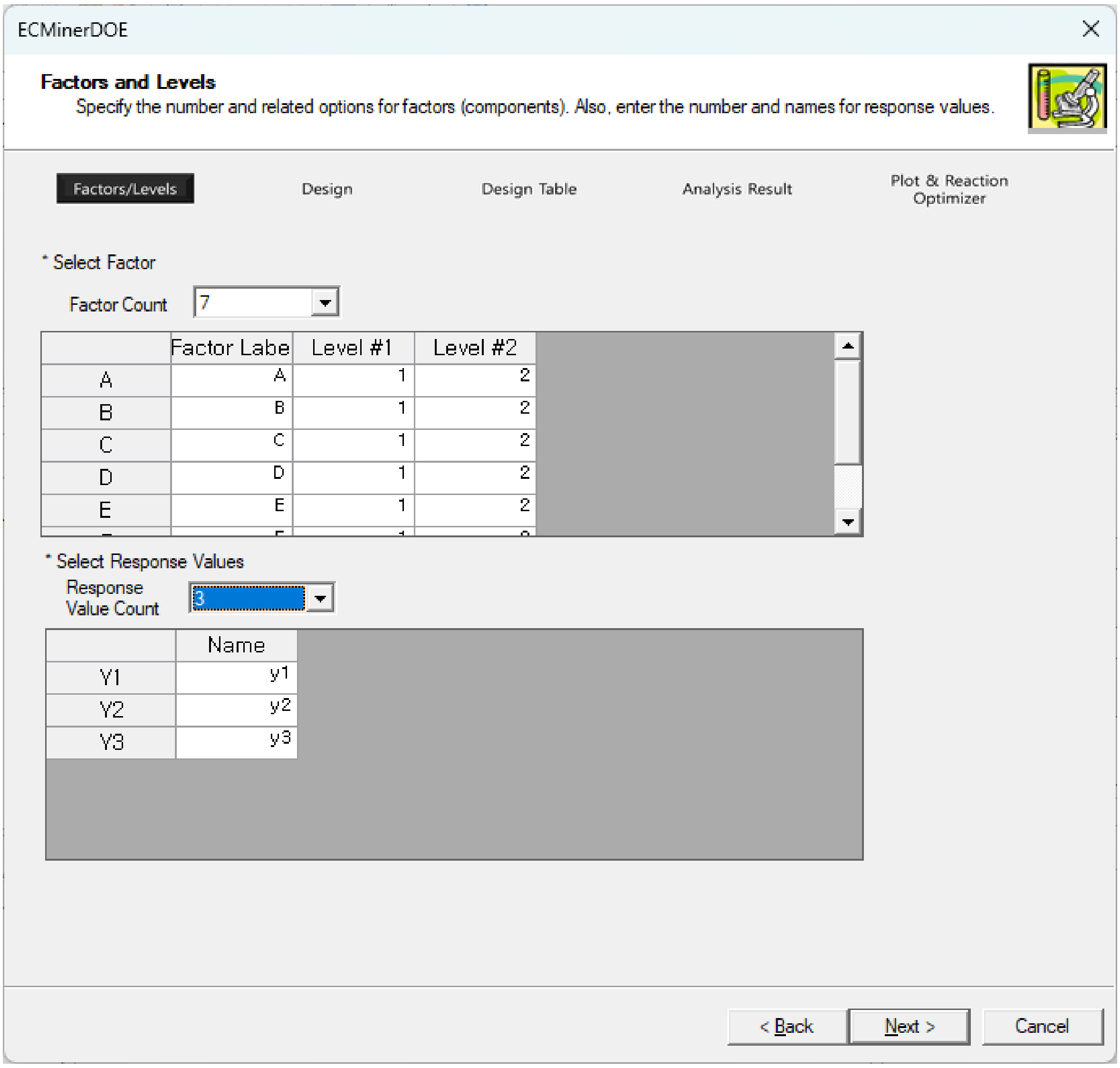
2-Level Design is a design used when there are two levels per factor. It is said that there is an experiment as follows.

Introduction to experiments In order to reduce the amount of CO in the exhaust gas, the related factors A-H are arranged in an orthogonal array table, and the noise factors are used to generate the minimum CO when driven in a certain way on three types of roads (R1, R2, and R3). Find the conditions.

Select **Taguchi 2-Level Design**.

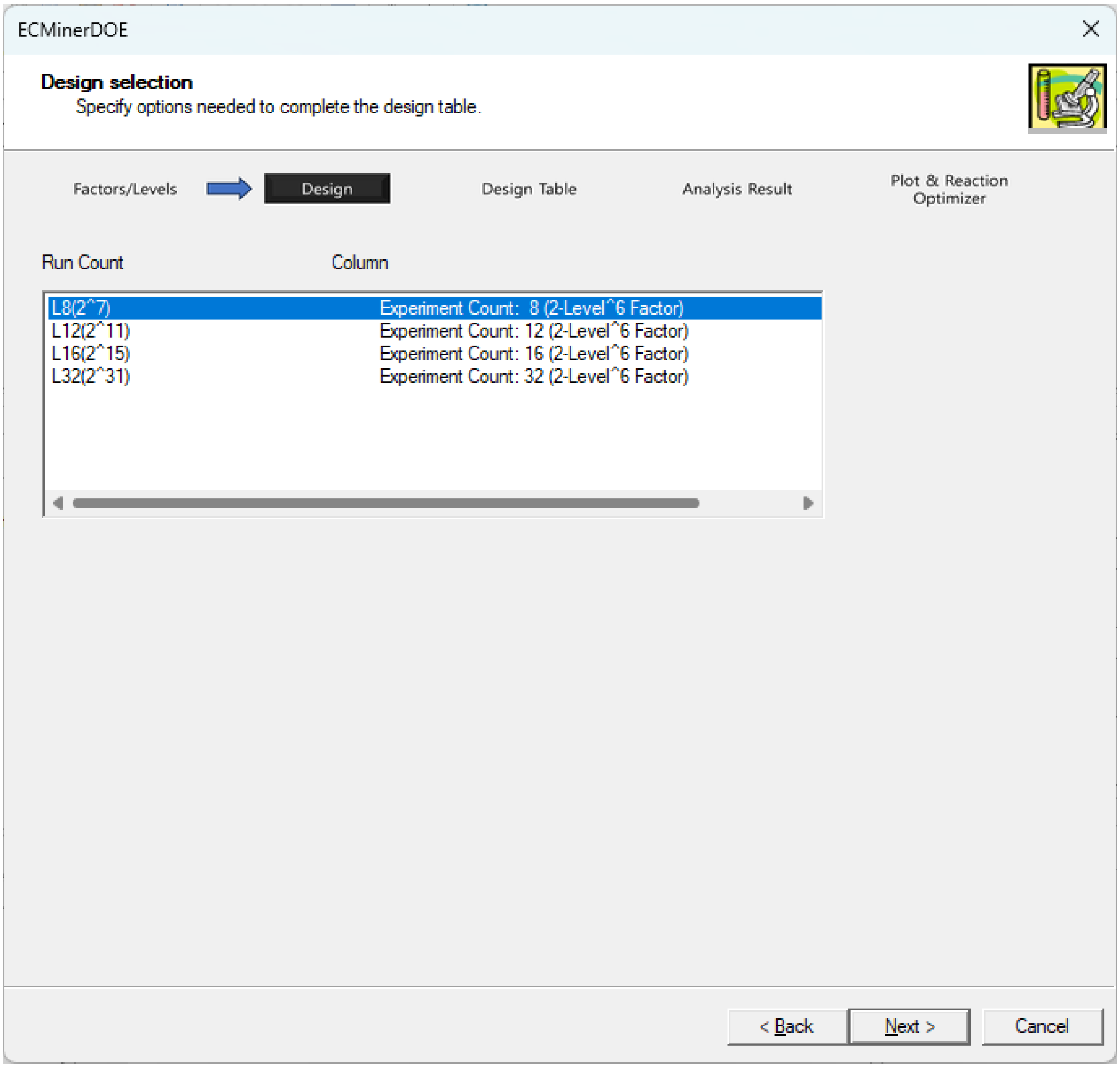


- **Step 1: Factors and Levels**



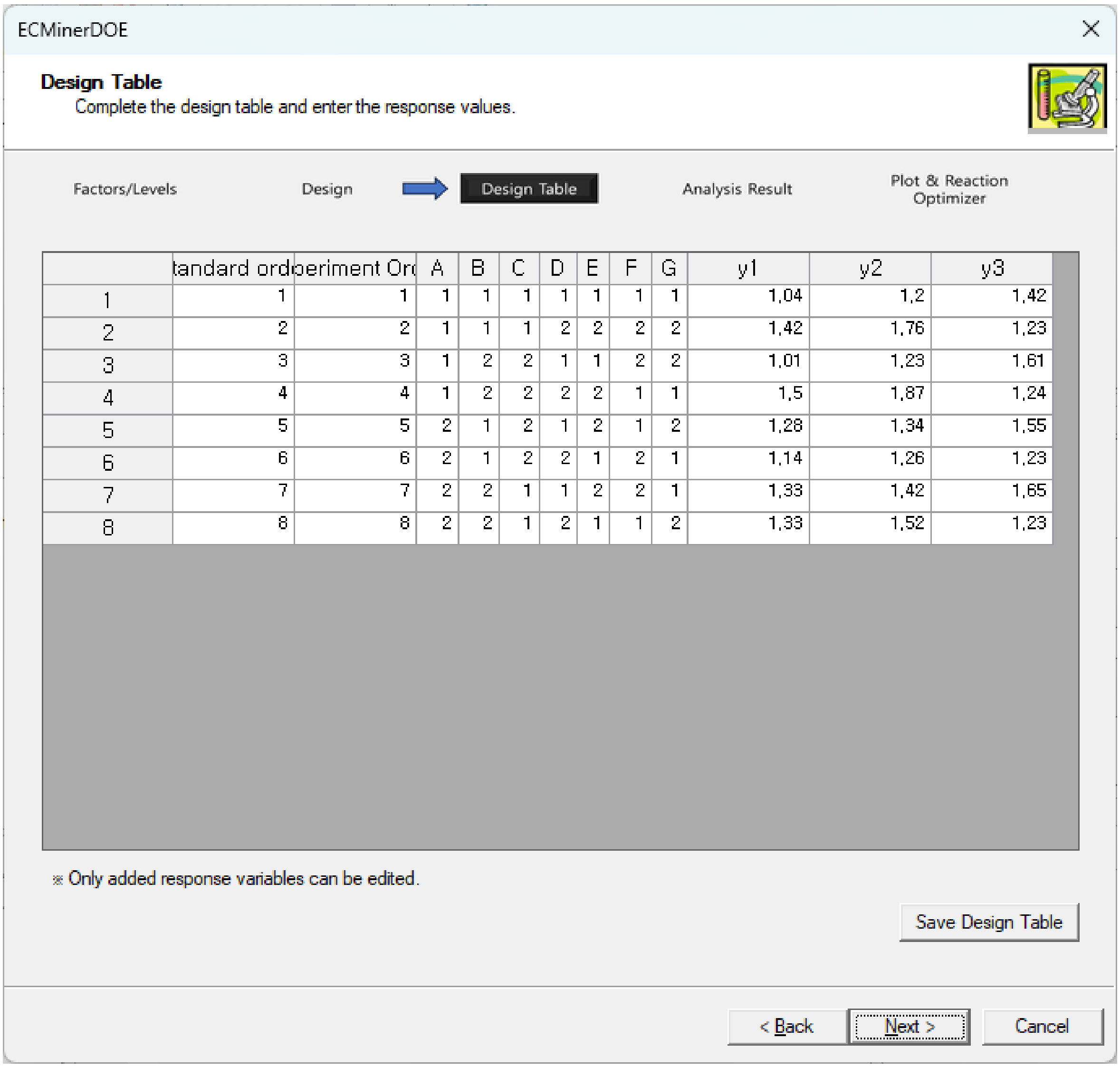
Select 7 factors and select 3 as the number of response values. In this case, the number of response values 3 means three types of roads.

- **Step 2: Design**



There are many designs to choose from, of which we choose the L8 design, which has the fewest number of experiments.

- **Step 3: Design Table**

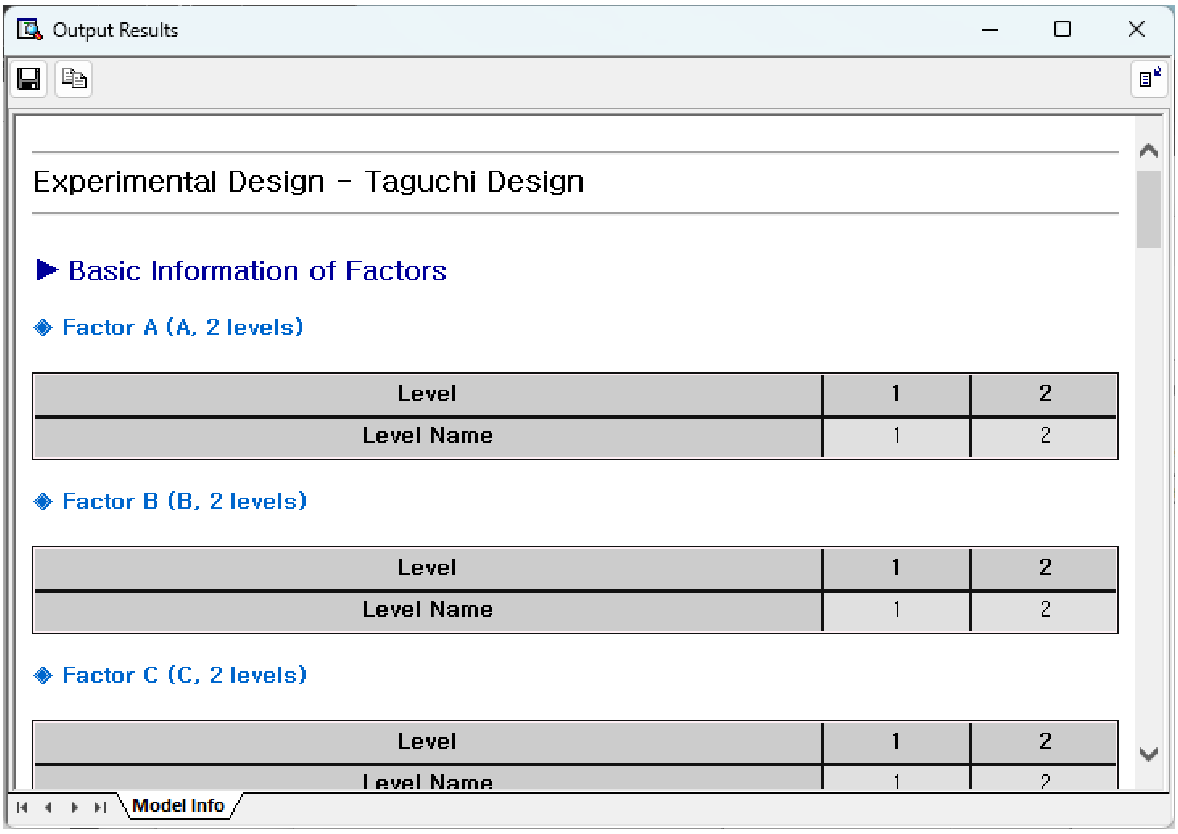


	tandard ord	periment Ord	A	B	C	D	E	F	G	y1	y2	y3
1	1	1	1	1	1	1	1	1	1	1,04	1,2	1,42
2	2	2	1	1	1	2	2	2	2	1,42	1,76	1,23
3	3	3	1	2	2	1	1	2	2	1,01	1,23	1,61
4	4	4	1	2	2	2	2	1	1	1,5	1,87	1,24
5	5	5	2	1	2	1	2	1	2	1,28	1,34	1,55
6	6	6	2	1	2	2	1	2	1	1,14	1,26	1,23
7	7	7	2	2	1	1	2	2	1	1,33	1,42	1,65
8	8	8	2	2	1	2	1	1	2	1,33	1,52	1,23

Enter the amount of CO from the three types of roads in the completed Design Table.

The values entered in this way will be used for later statistical analysis.

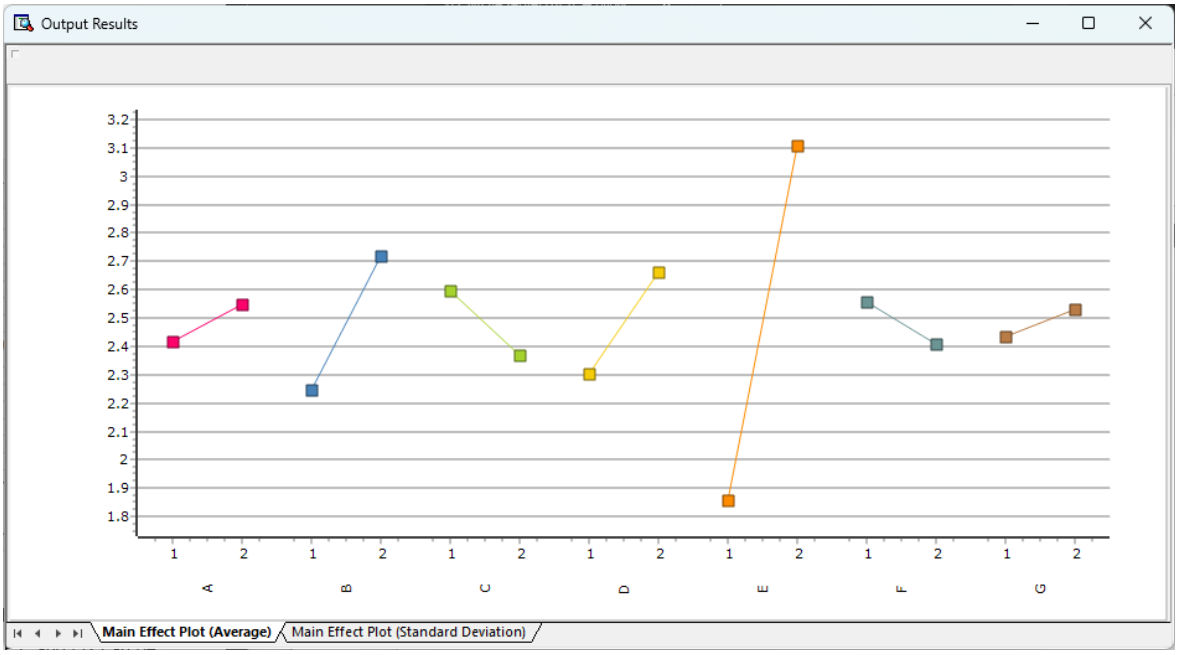
- **Step 4: Analysis Result**



Since a smaller the current CO value is better, select the Mesh characteristic and start the analysis. As a result of the analysis, you can obtain a variance analysis table along with the SN ratio Factors with a negligible small SS can be excluded from the initial experiment. Since the SS of E is substantially large, this factor is considered a decisive factor in the SN ratio of CO.

For a detailed explanation, see 6.1.4. See Settings and Analysis.

- **Step 5: Plot and Response Optimizer**



At this stage, the optimal conditions can be found through the main effects plot for the SN ratio. Since a larger SN ratio is better, the optimal conditions are A2, B2, C1, D2, E2, F1, and G2.

However, if the levels of only a few factors can be determined, B and E should first be set to levels 1 for both factors. And for the remaining factors, the level is determined by considering economic and cost effectiveness.

The experimental plan is terminated after determining reproducibility at the optimally determined level through re-experimentation.

6.1.3.4.2. 3- Level Design

3-Level Design is a design used when there are three levels per factor. Suppose we have the following experiment.

Introduction to experiments

A chemical company that produces a certain resin wants to conduct an experiment to reduce the content of impurities contained in this resin. The upper limit of the specification is 4.0%, and if this specification is not met, a loss of 50,000 won per 10 kg occurs. Four control factors expected to affect impurities were taken as follows.

$$\mathrm{A}:\ Bond\ mixing\ ratio\ level\ 3(\mathrm{A}_0, \mathrm{A}_1, \mathrm{A}_2)$$

$$\mathrm{B}:\ Bonding\ method\ level\ 3(\mathrm{B}_0, \mathrm{B}_1, \mathrm{B}_2)$$

$$\mathrm{C}:\ Surface\ treatment\ method\ level\ 3(\mathrm{C}_0, \mathrm{C}_1, \mathrm{C}_2)$$

$$\mathrm{D}:\ Heat\ treatment\ method\ level\ 3(\mathrm{D}_0, \mathrm{D}_1, \mathrm{D}_2)$$

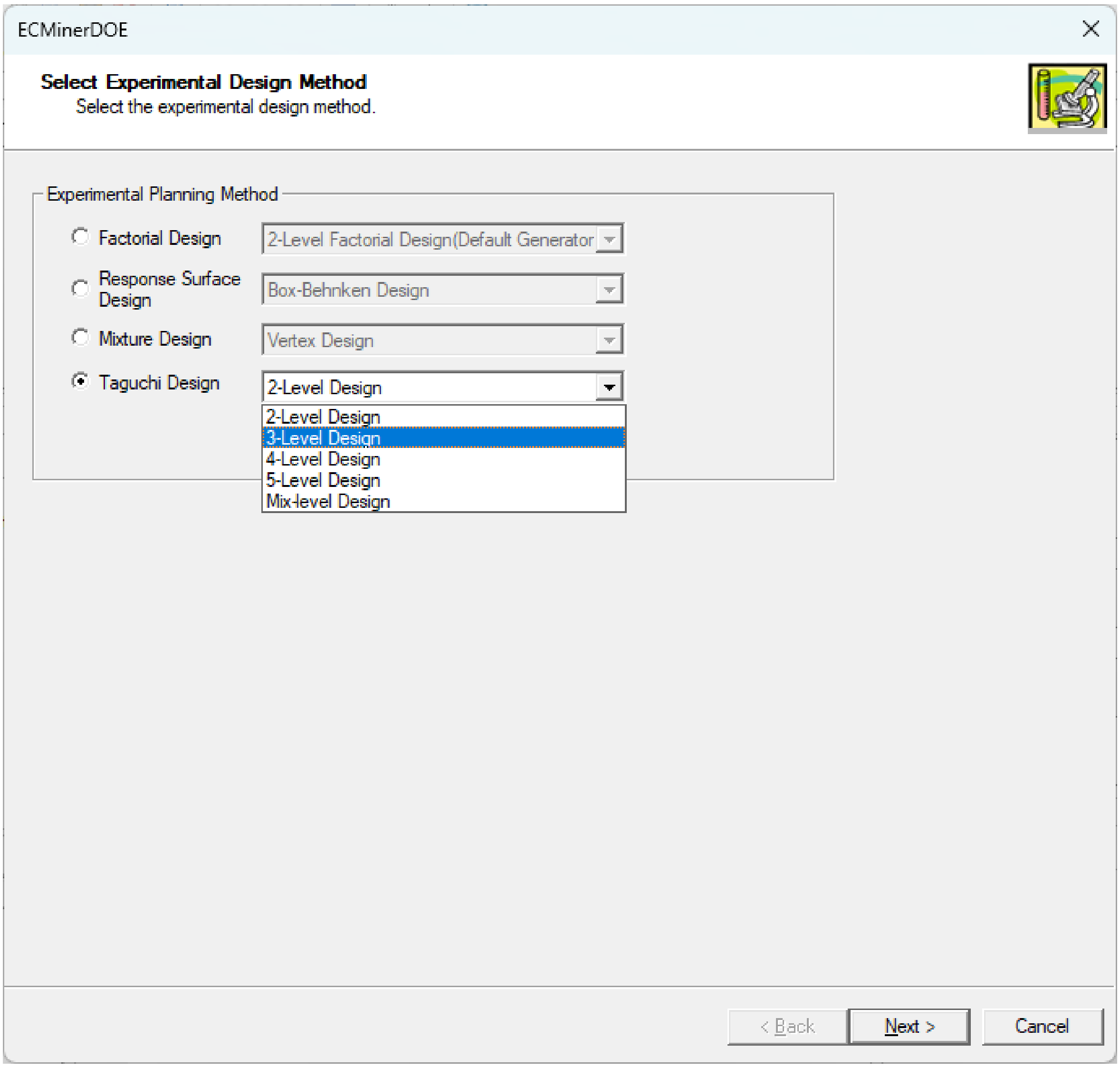
As a non-controlling factor

$$\mathrm{U}:\ 2nd\ level\ of\ worker(unskilled\ worker, skilled\ worker)$$

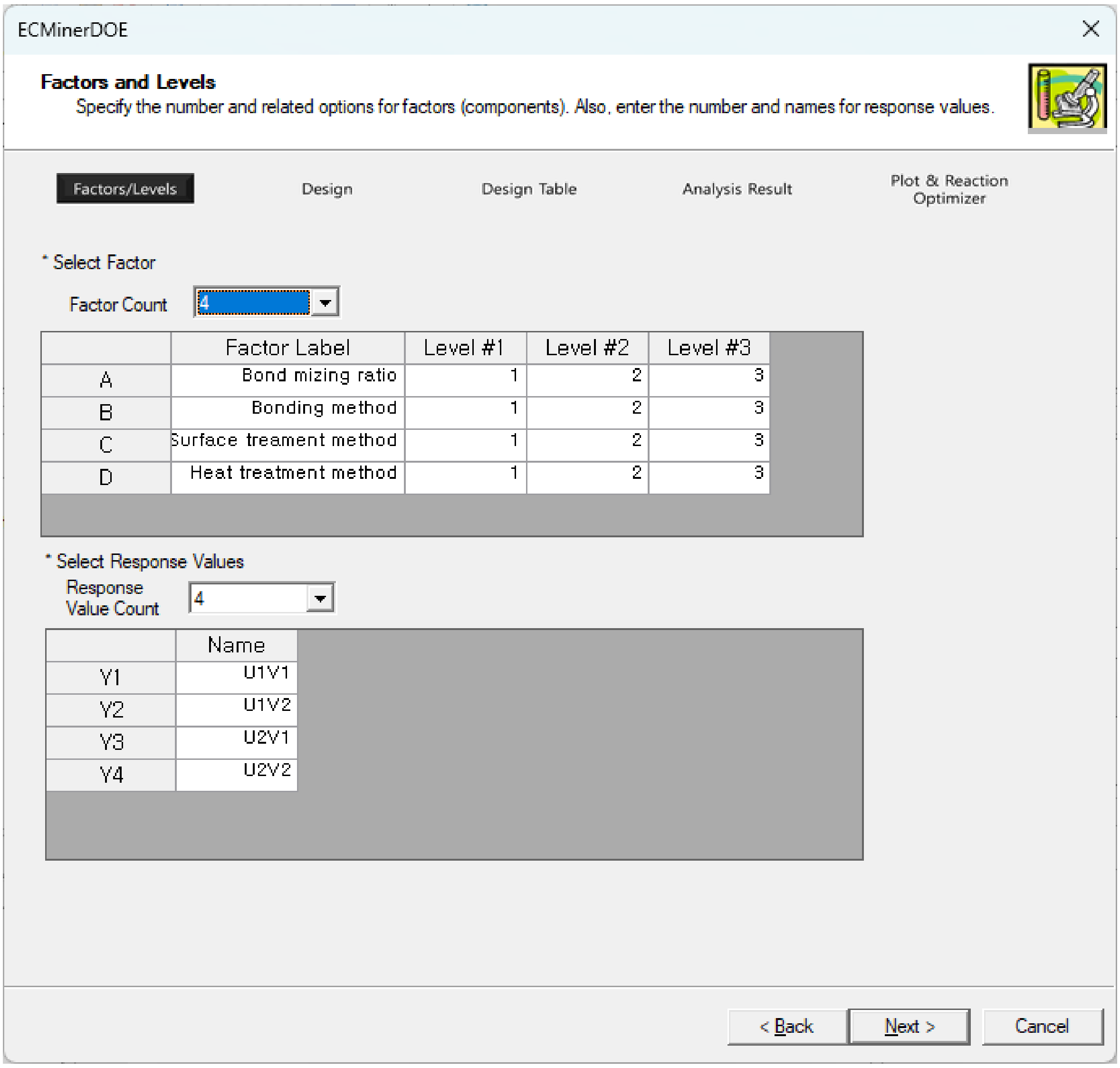
$$\mathrm{V}:Resin\ production\ line\ level\ 2$$

was selected and the produced resin was analyzed in the laboratory to obtain the content of impurities.

Select Taguchi Design 3-Level Design.

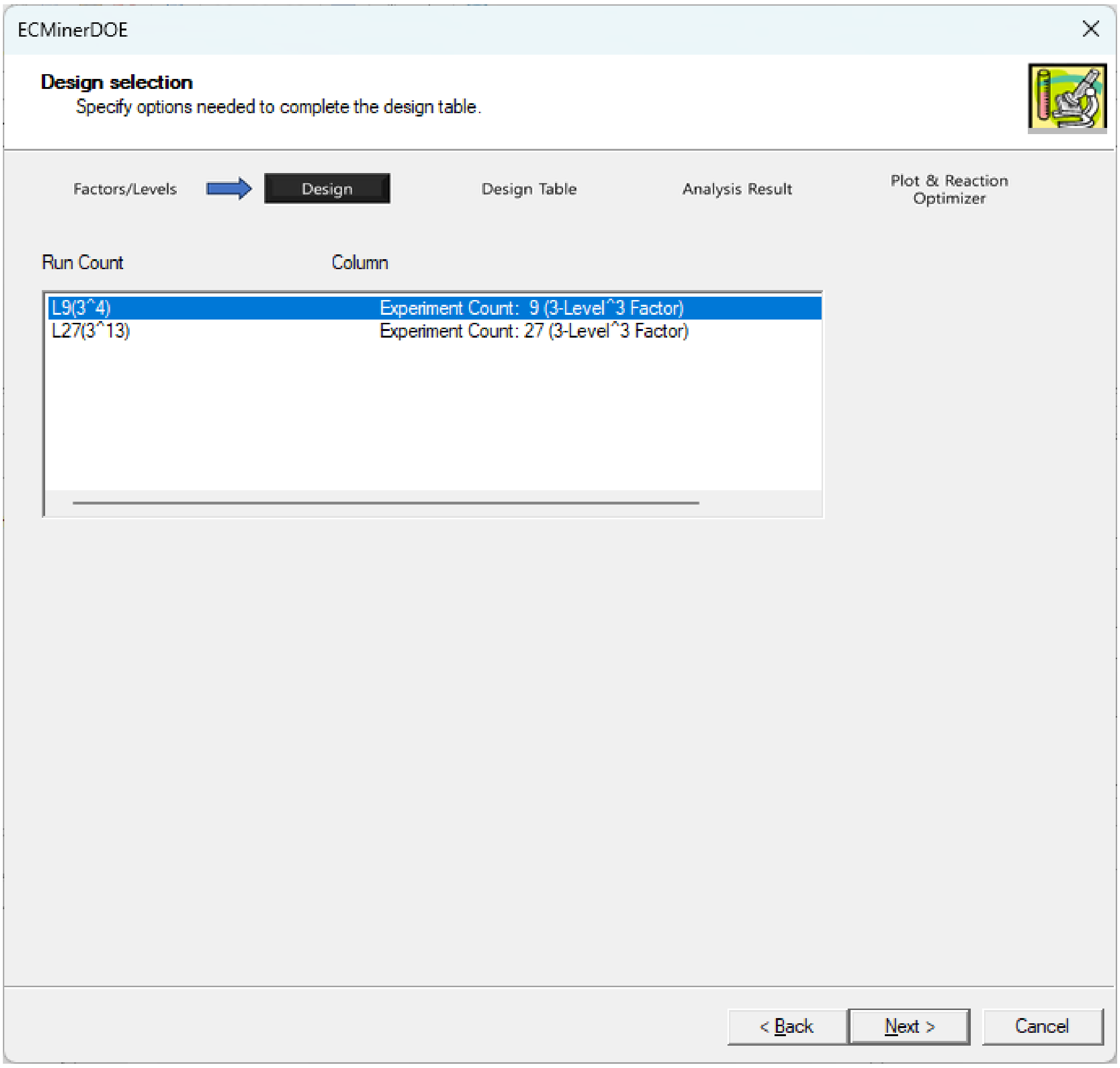


- **Step 1: Factors and Levels**



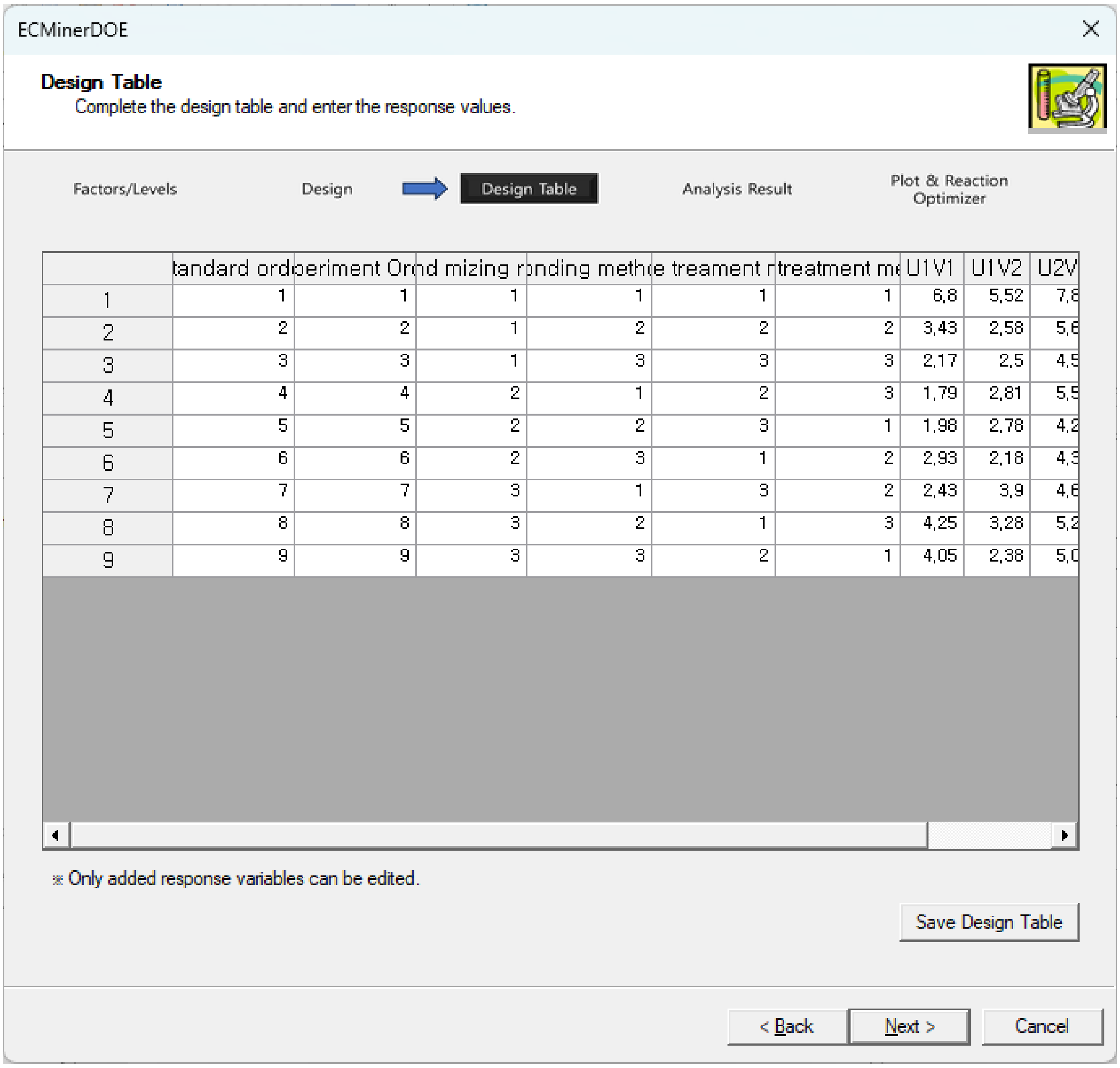
Name the factors as above, select 4 response values, and give them names appropriate for the level of each noise factor.

- **Step 2: Design**



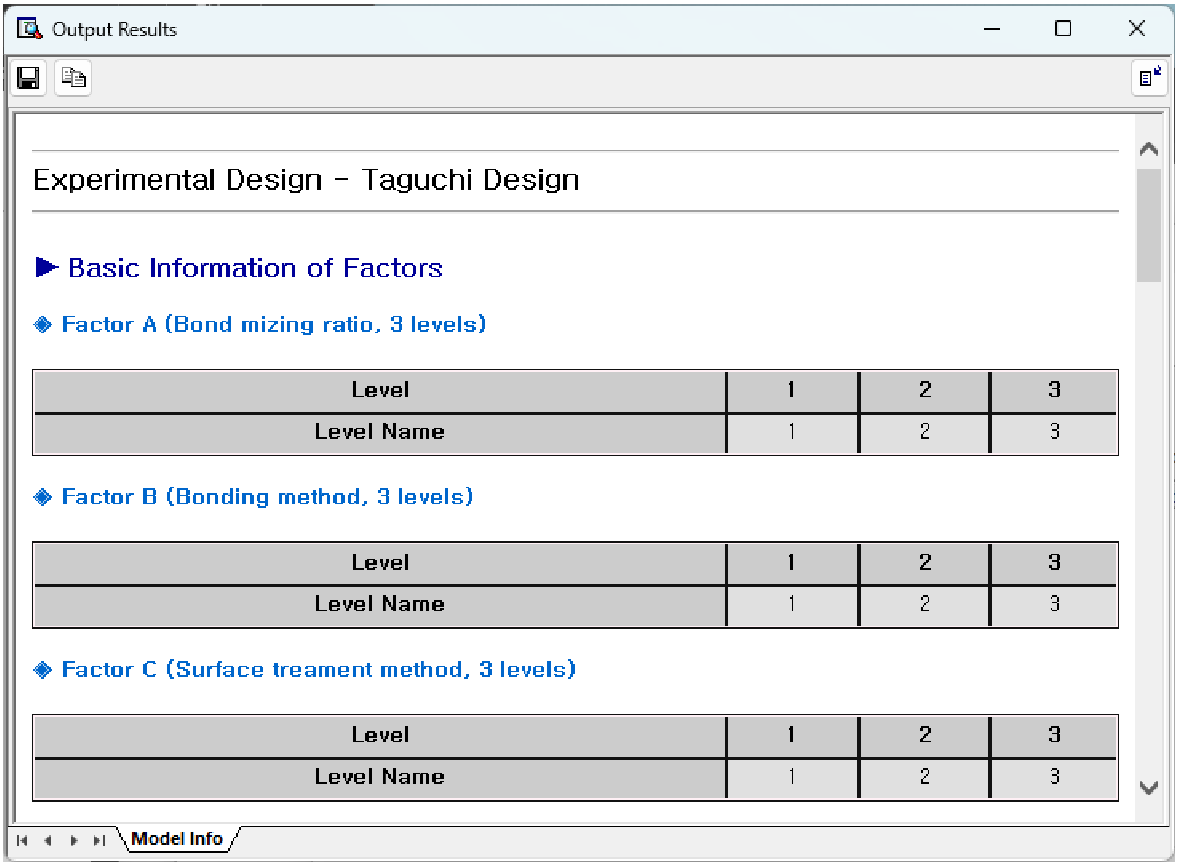
There are currently two designs to choose from, of which I choose the L9.

- **Step 3: Design Table**



Once the design table is completed, enter the experimental characteristic values at each level of the non-controlling factors.

- **Step 4: Analysis Result**


Output Results

Experimental Design - Taguchi Design

▶ Basic Information of Factors

◈ Factor A (Bond mizing ratio, 3 levels)

Level	1	2	3
Level Name	1	2	3

◈ Factor B (Bonding method, 3 levels)

Level	1	2	3
Level Name	1	2	3

◈ Factor C (Surface treament method, 3 levels)

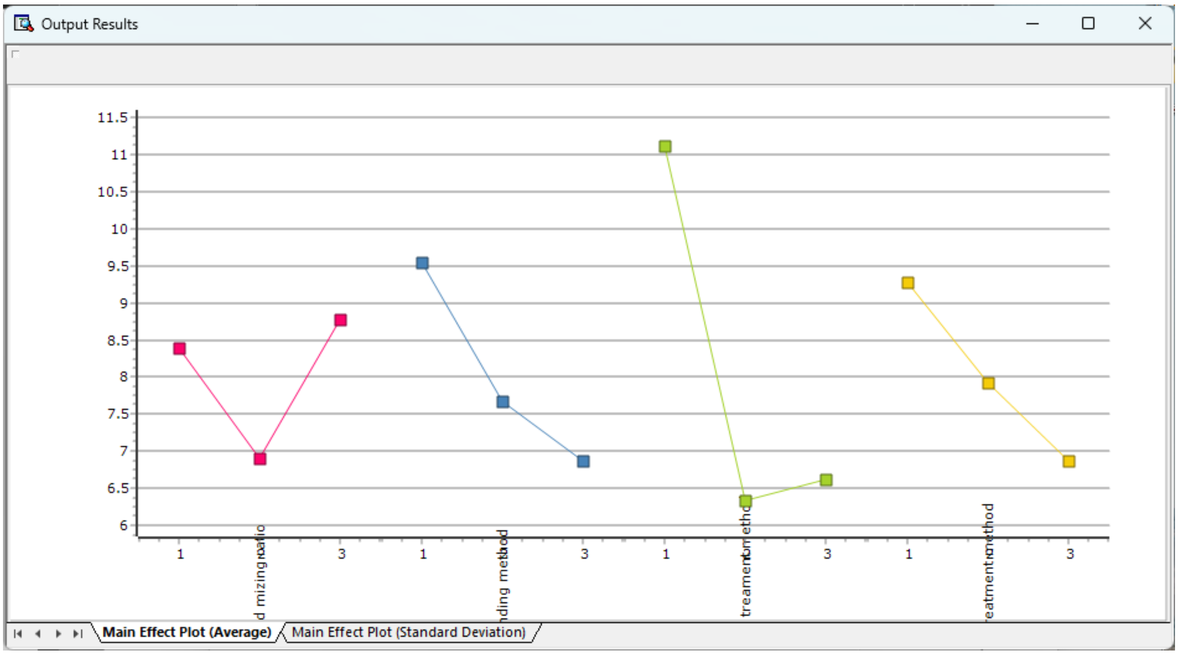
Level	1	2	3
Level Name	1	2	3

Model Info

Since the current characteristic value represents the content of impurities, a smaller value indicates better performance. So, select the mesh feature and start the analysis. The analysis of variance table based on the SN ratio shows that B's SS (Sum of Square) is extremely small. From this, we can see that factors A, C, and D all well explain the variance in SN ratio.

For detailed explanation, see 6.1.4. See Settings and Analysis.

- **Step 5: Plot and Response Optimizer**



In this step, the optimal experimental conditions can be found through the main effects plot for the SN ratio. Currently, the SN ratio is the highest for A at level 2, B at level 3, C at level 3, and D at level 3, so A1, B2, C1, and D1 are the best conditions, assuming there is no interaction between A, B, and C. We conduct a confirmation experiment to confirm reproducibility and complete the experiment.

6.1.3.4.3. 4 - Level Design, 5- Level Design, Mix-Level Design

---

The analysis method of all Taguchi Designs is the same as 2 and 3-Level Design. Therefore, we will only mention the characteristics of 4-Level Design, 5-Level Design, and Mix-Level Design.

- **4 - Level Design**

  A 4-Level Design is a design in which the number of factor levels is 4.

- **5- Level Design**

  A 5-Level Design is a design in which the number of factor levels is 5.

- **Mix-Level Design**

  Mix - Level Design is a design where the number of levels of a factor varies depending on the factor. ECMiner™ DOE offers a $2^1 \times 3^7$ design.

For detailed analysis methods, see 6.1.3.4.1. 2-Level Design, 6.1.3.4.2. 3-Level Design, 6.1.4.

See Settings and Analysis.

6.1.4 Settings and Analysis

6.1.4.1 Settings

6.1.4.1.1. Predict Settings

Predict is used when you want to predict the response value at a specific factor (component) level using the Regression Model created in Step 4. The input values that must be set in Predict are as follows.

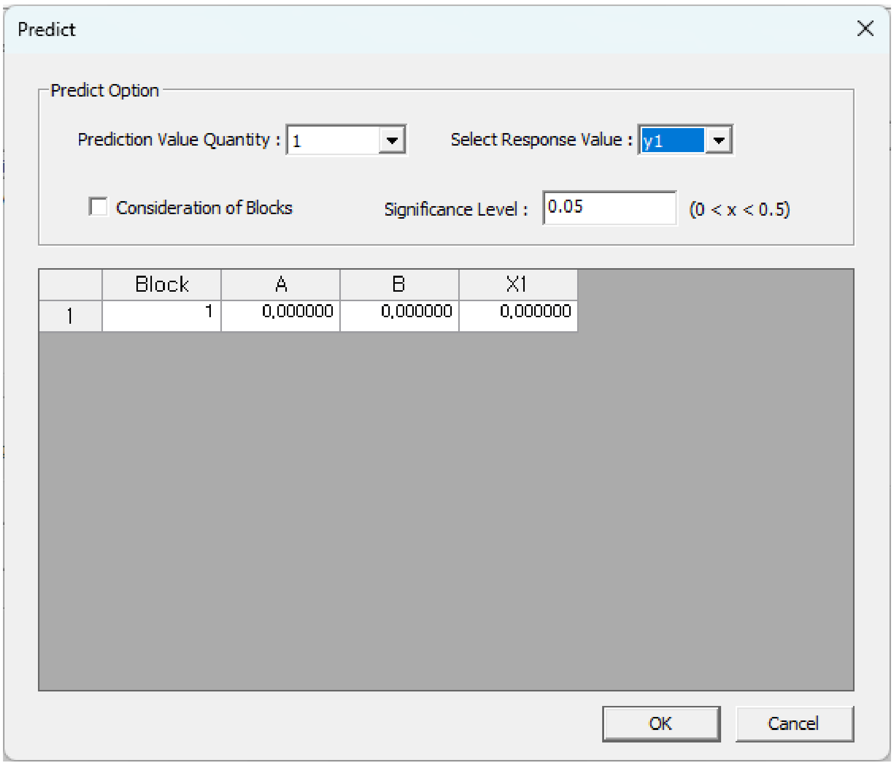
- **Prediction Value Quantity**

  Enter how many points you want to predict.
- **Select Response Value**

  Select the desired value to predict for multiple response values.
- **Confidence Level**

  Use when calculating the confidence interval of the response value.
- **Consideration of blocks**

  Decide whether to take the block into consideration.



6.1.4.1.2. Graph Settings

Through Graph Settings, settings are made for the graph related to various statistics obtained through the Regression Model in Step 4.

- **Residual Histogram**

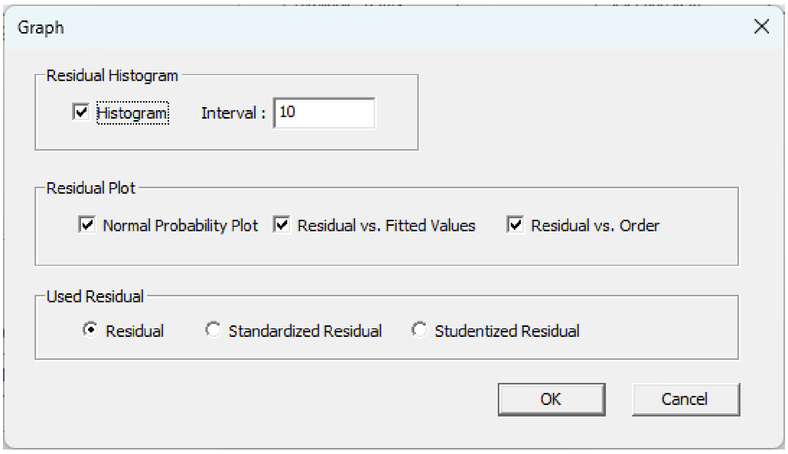
  Check whether to draw a residual histogram and how many sections to divide it into.

- **Residual Plot**

  Choose which residual plot to draw.

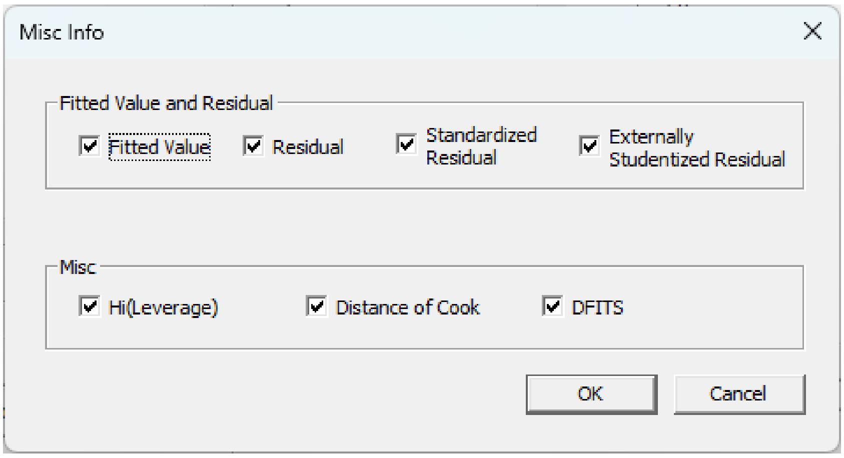
- **Used Residual**

  Select which residuals to use when drawing residual-related plots.



6.1.4.1.3. Other Information

Through Other Information, you can select which statistics to show in the analysis results. Selectable statistics include fitted values, residuals, standardized residuals, externally Studentized residuals, leverage, distance of Cook, and DFITS.



6.1.4.2 Analysis

6.1.4.2.1. Regression Analysis

Once the design table is created and response values are obtained according to the created design, full-scale analysis can be started. ECMiner^{TM} DOE provides functions such as regression analysis, analysis of variance, residual analysis, prediction, graph analysis, and response optimizer. The most basic of these is regression analysis. The regression analysis model can be expressed mathematically as follows.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \quad \epsilon \sim N(0, \sigma^2)$$

$$y = X\beta + \epsilon \quad \epsilon \sim N(0, \sigma^2 I),\ I \textit{ is the identity matrix.}$$

$$where\ y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_N \end{bmatrix}, X = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ x_{21} & \cdots & x_{2k} \\ \cdots & \cdots & \cdots \\ x_{N1} & \cdots & x_{Nk} \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdots \\ \epsilon_N \end{bmatrix}$$

In this case, the estimator of β using the least squares method is calculated as follows.

$$\widehat{\beta} = (X^T X)^{-1} X^T y \quad Var(\widehat{\beta}) = \sigma^2 (X^T X)^{-1}$$

Using these facts, we can estimate the fitted model that best explains the data as follows.

$$\widehat{y} = X\widehat{\beta}$$

***Note**: Mixture Design excludes constant terms from the Regression Model above. Due to the constraint that the sum of the mixture components must always be constant, the constant term disappears.*

Factorial Design regression analysis results

● Estimated effects and coefficients for response value y1

Term	Effect	Coefficient	Coefficient SE	T	P
Const	*	12,87500	10,44536	1,23260	0,30551
Block1	*	2,85000	10,44536	0,27285	0,80267
A	-4,30000	-2,15000	10,44536	-0,20583	0,85010
B	7,20000	3,60000	10,44536	0,34465	0,75310
C	6,80000	3,40000	10,44536	0,32550	0,76618

● Standard error for response value y1

R Square	0.10221
Adjusted R Square	0
RMSE	29.54395

Response Surface Design regression analysis results

● Estimated coefficients(coded units) for response value y1

Term	Coefficient	Coefficient SE	T	P
Const	0.78004	0.33302	2.34231	0.05167
A	0.04268	0.26330	0.16210	0.87580
B	0.06769	0.26330	0.25707	0.80452
AA	-0.20257	0.28239	-0.71733	0.49641
BB	-0.15255	0.28239	-0.54021	0.60580
AB	0	0.37233	0	1

● Estimated coefficients(coded units) for response value y2

Term	Coefficient	Coefficient SE	T	P
Const	11.74600	8.08988	1.45194	0.18982
A	0.03838	6.39609	0.00600	0.99538
B	0.14572	6.39609	0.02278	0.98246
AA	-5.65510	6.86001	-0.82436	0.43692
BB	-5.78013	6.86001	-0.84258	0.42731
AB	-0.05000	9.04476	-0.00553	0.99574

● Standard error for response value y1

R Square	0.10371
Adjusted R Square	0
RMSE	0.74466

Mixture Design regression analysis results

● Estimated coefficients(ratio) for response value y1

Term	Coefficient	Coefficient SE	T	P
A	-1,32726	13,23721	*	*
B	1,09637	13,23721	*	*
C	-0,97272	13,23721	*	*
A*B	34,03747	61,00849	0,55791	0,60665
A*C	37,89929	61,00849	0,62121	0,56811
B*C	26,34657	61,00849	0,43185	0,68811

● Estimated coefficients(quantity) for response value y1

Term	Coefficient	Coefficient SE	T	P
A	-1,32726	13,23721	*	*
B	1,09637	13,23721	*	*
C	-0,97272	13,23721	*	*
A*B	34,03747	61,00849	0,55791	0,60665
A*C	37,89929	61,00849	0,62121	0,56811
B*C	26,34657	61,00849	0,43185	0,68811

● Standard error for response value y1

R Square	0,18109
Adjusted R Square	0
RMSE	13,72535

6.1.4.2.2. Dispersion Analysis (Analysis of Variance, ANOVA)

Analysis of Variance allows you to determine whether each term affects the response value by determining the degree of variance caused by each term in the total variance. Distributed Analysis can be easily performed through the Regressor Matrix used in regression analysis.

First, if there is a block, we need to find the variance for the block.

If we call the Submatrix consist of Regressor Matrix X's first to ith Column as $\mathrm{X_i}$, and when the total number of blocks is bn (>=2) it is

$$\boldsymbol{Sum\ of\ squares\ of\ block} = \boldsymbol{y}^{\boldsymbol{T}}\left(\boldsymbol{X_{bn}}\left(\boldsymbol{X_{bn}^T X_{bn}}\right)^{-1}\boldsymbol{X_{bn}^T} - \boldsymbol{X_1}\left(\boldsymbol{X_1^T X_1}\right)^{-1}\boldsymbol{X_1}^{\boldsymbol{T}}\right)\boldsymbol{y}$$

If the number of blocks is 1, the above process is not executed. And the degree of freedom at this time is bn-1.

Now we need to find the Sum of Squares for each term. One term corresponds to one column of X. If a term corresponds to the jth column, the sum of square of that term is

$$\boldsymbol{y^T}\left(\boldsymbol{X_i}\left(\boldsymbol{X_i^T X_i}\right)^{-1}\boldsymbol{X_i} - \boldsymbol{X_{i\_1}}\left(\boldsymbol{X_{i\_1}^T X_{i\_1}}\right)^{-1}\boldsymbol{X_{i\_1}}\right)\boldsymbol{y}$$

At this time, the degree of freedom of this term is 1. However, ECMiner™ DOE does not calculate the Sum of Square for each term, but instead calculates the Sum of Square by grouping it by the main effect, two-way interaction, three-way interaction, or other characteristics. First, if the columns due to the main effect are from i to j, use the

$$\boldsymbol{y^T}\left(\boldsymbol{X_j}\left(\boldsymbol{X_j^T X_j}\right)^{-1}\boldsymbol{X_j} - \boldsymbol{X_{i\_1}}\left(\boldsymbol{X_{i\_1}^T X_{i\_1}}\right)^{-1}\boldsymbol{X_{i\_1}}\right)\boldsymbol{y}$$

formula to find the Sum of Square. This is ultimately equal to the sum of the Sum of Squares for each column that makes up the main effect. Its degrees of freedom are i-j. In the same way, k-way interactions are also grouped to obtain the Sum of Square.

If there is a center point in the design, the last column corresponds to it. The variation about the center point is called Curvature. When the number of columns of X is Ncols, it is

$$\begin{aligned}\boldsymbol{Sum\ of\ Squares\ of\ Curvature}\\ &= \boldsymbol{y^T}\left(\boldsymbol{X_{Ncols}}\left(\boldsymbol{X_{Ncols}^T X_{Ncols}}\right)^{-1}\boldsymbol{X_{Ncols}^T}\right.\\ &\left.- \boldsymbol{X_{Ncols\_1}}\left(\boldsymbol{X_{Ncols\_1}^T X_{Ncols\_1}}\right)^{-1}\boldsymbol{X_{Ncols\_1}}\right)\boldsymbol{y}\end{aligned}$$

and the degree of freedom is 1.

In the case of SSE and SST, they are obtained as follows.

$$\boldsymbol{SSE = SST - Sum\ of\ Squares\ of\ all\ items}.$$

$$\boldsymbol{SST} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}\left(\boldsymbol{y_i} - \overline{\boldsymbol{y}}\right)^{\boldsymbol{2}}$$

The degree of freedom at this time is ‘total number of data – 1’. Blocks, Curvature, main effects, two-way interaction. The F value and P value of are obtained through

$$F\_value = \frac{SS/df}{SST/(n-1)} \qquad p\_value = P(F > Fvalue) \text{ where } F \sim F(df, n-1)$$

Another thing to consider is Lack of Fit. When we calculate SSE, it can be said that SSE consists of Pure Error and Lack of Fit. Pure Error can be obtained when experiments are conducted on the same point with multiple times (twice or more).

(However, the same point here does not mean the actual same point. Here, the same point is said if the rows of the Regressor Matrix are the same. Even if the actual experimental points are different, the rows of the Regressor Matrix may be the same depending on how the selection term is selected.)

If ak(1<=k<=m) experiment point (when Nk experiments are performed) Pure Error is

$$Pure\ Error = \sum_{k=1}^{m} SST\ at\ a_k$$

$$where\ SST\ at\ a_k = \sum_{i \in A} y_i^2 - N_{a_k} * \left(\frac{\sum_{i \in A} y_i}{N_{a_k}}\right)^2 \ The\ set\ A\ consists\ of\ the\ rows\ of\ a_k$$

And the degrees of freedom of Pure Error are:

$$DF\ of\ Pure\ Error = \sum_{k=1}^{m} (N_{a_k} - 1)$$

If the Pure Error and its degrees of freedom are obtained in this way, the SS and degrees of freedom of the Lack of Fit can also be obtained as follows.

$$Sum\ of\ Squares\ of\ \ Lack\ of\ Fit' = SSE - Pure\ Error$$
$$DF\ of\ Lack\ of\ Fit = DF\ of\ SSE - DF\ of\ Pure\ Error$$

At this time, the F value and P value of Lack of Fit obtained are

$$Fvalue = \frac{\frac{SS\ of\ Lack\ of\ Fit}{DF\ of\ Lack\ of\ Fit}}{\frac{Pure\ Error}{DF\ of\ Pure\ Error}}\ and$$

$$p\_value = P(F > Fvalue)\ where\ F \sim F(DF\ of\ Lack\ of\ Fit,\ DF\ of\ Pure\ Error)$$

Through this, you can statistically check whether there is a lack of suitability of this model.

Results of Factorial Design

● ANOVA table for response value y1

Term	Variability	df	Mean Sum of Square	F	P
Block	56.71125	1	56.71125	0.39178	0.57575
Main Effect	252.08375	3	84.02792	0.58049	0.66697
Residual Error	434.26375	3	144.75458		
Total Variation	743.05875	7			

Results of Response Surface

● ANOVA table for response value y1

Term	Variability	df	Mean Sum of Square	F	P
Linear	0.01844	2	0.00922	0.02249	0.97783
Square	0.74226	2	0.37113	0.90565	0.44689
Interaction	0	1	0	0	1
Residual Error	2.86854	7	0.40979		
Lack of Fit	0.13654	3	0.04551	0.06664	0.97478
PureError	2.73200	4	0.68300		
Total Variation	3.62923	12			

Results of Mixture Design

● ANOVA table for response value y1

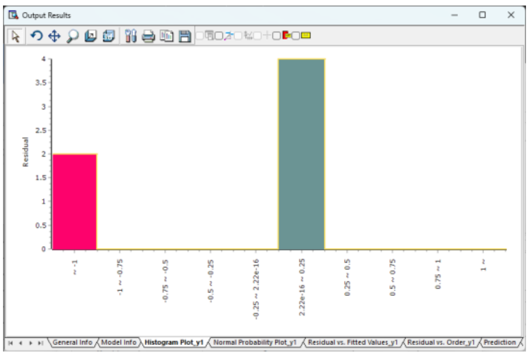
Term	Variability	Freedom Degree	Average Variability	F	P
Linear	5,35684	2	2,67842	1,79914	0,27713
A*B	0,29682	1	0,29682	0,19938	0,67833
A*C	5,95258	1	5,95258	3,99845	0,11617
B*C	3,16048	1	3,16048	2,12295	0,21883
Residual Error	5,95488	4	1,48872		
Total Variation	20,72161	9			

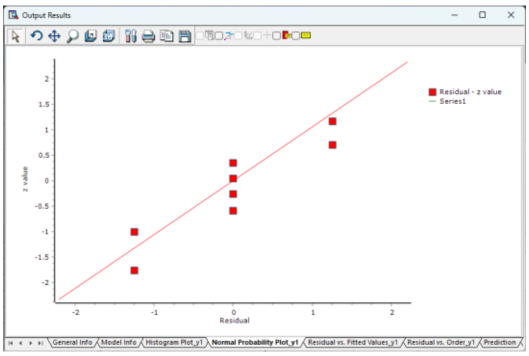
6.1.4.2.3. Residual Analysis

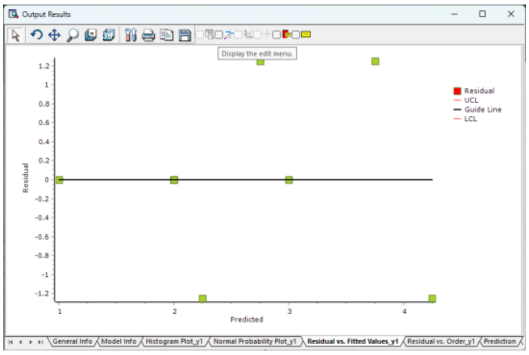
For Residual Analysis, several statistics are used. (This refers not only to the residual but also to all statistics related to the residual. ECMiner™ DOE provides the following statistics.

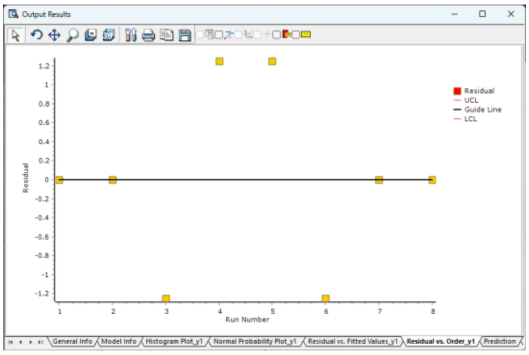
- **Residual**
- **Standardized Residual**
- **External Standardized Residual**
- **Leverage**
- **Cook's Distance**
- **DFITTS**

And as graphs related to residual, we provide Residual Histogram, Normal Probability Plot, Residual vs. Order, and Residual vs. Fitted Value Graph.









6.1.4.2.4. Statistical Analysis on Taguchi

6.1.4.2.4.1. Loss function and SN ratio

Loss function

Quality characteristics are classified into three categories as follows and loss functions for each are defined separately.

- **Nominal the best characteristics**

  This is a case where a specific target value is given, such as length, weight, thickness, etc. If the measured value is y and the target value is m, the loss function is defined as

$$L(y) = k(y - m)^2$$

  If the consumer's loss at the consumer tolerance threshold $m+\Delta, m-\Delta$ of the characteristic value is A won, this is determined by the following equation.

$$A = k\Delta^2$$

  In other words, in the case of network characteristics, the loss function becomes

$$L(y) = \frac{A}{\Delta^2}(y - m)^2$$

- **The smaller the better characteristics**

  The smaller the characteristic value, the better, such as wear, vibration, and defect rate. In this case, the network characteristic can be considered as m=0, so the loss function is as follows.

$$L(y) = ky^2 = \frac{A}{\Delta^2} y^2$$

- **Large the better characteristics**

  The larger the characteristic value, the better—for example, strength, lifespan, and fuel efficiency. The loss function at this time is as follows.

$$L(y) = \frac{A}{\Delta^2}\left(\frac{1}{y^2}\right)$$

SN ratio

It is expressed as

$$SN\ ratio = \frac{power\ of\ signal}{power\ of\ noise}$$

which is defined differently for each type.

- **nominal the best characteristics case**

$$\frac{power\ of\ signal}{power\ of\ noise} = \frac{Estimated\ value\ of\ population\ mean\ \mu\ squared\ \mu^2}{Estimated\ of\ variance\ \sigma^2}$$

  At this time, the estimated value of variance is

$$\widehat{\sigma^2} = V = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{n-1}$$

$$S_m = \frac{(y_1 + y_2 + \cdots + y_n)}{n}$$

  In this case,

$$E(S_m) = \sigma^2 + n^2$$

is established, so it becomes

$$S_m = \widehat{\sigma^2} = n\,\widehat{\mu^2}$$

$$\widehat{\mu}^2 = \frac{1}{n}(S_m - V)$$

For common logs in

$$SN\ ratio = \frac{\frac{1}{n}(S_m - V)}{V}$$

you can get the following values.

$$SN\ ratio = 10\ log\left[\frac{\frac{1}{n}(S_m - V)}{V}\right]$$

The larger this value, the greater the power of the signal and the smaller the power of noise, and the condition that makes this SN value the largest becomes the optimal condition. However, since it is $S_m = n(\bar{y})^2$, the SN ratio becomes

$$SN = 10\ log\left[\frac{(\bar{y})^2 - \frac{V}{n}}{V}\right]$$

When n is large enough, V/n becomes small enough to be ignored, so when $V = s^2$ is true it becomes

$$SN = 10\ log\left[\frac{(\bar{y})^2}{V}\right] = 10\ log\left[\frac{(\bar{y})^2}{s^2}\right] = 20\ log\left(\frac{\bar{y}}{s}\right)$$

- **The smaller the better characteristics case**

  In the case of network properties, the SN ratio that minimizes the expected value of the loss function is considered. f repeated measurement data $y_1, y_2, \ldots, y_n$

obtained, the estimated value of E(y) can be viewed as

$$MSD = \frac{1}{n}\sum_{i=1}^{n}(y_i - 0)^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2$$

Using this, the SN ratio is calculated as follows.

$$SN = -10\,log\left[\frac{1}{n}\sum_{i=1}^{n} y_i^2\right]$$

▪ **Large the better characteristics case**

In order to make the expected loss L smaller, as in the case of the smaller the better characteristics case, the estimated value of E(1/y) is used as

$$MSD = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{y_i} - 0\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i^2}$$

and the SN ratio is set as

$$SN\,ratio = -10\,log\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i^2}\right]$$

To summarize, it is as follows.

Types of characteristic values	Loss function of one data y	Average loss function when n pieces of data are obtained	SN ratio
nominal the best characteristics	$\frac{A}{\Delta^2}(y - m)^2$	$\frac{A}{\Delta^2}\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - m)^2\right]$	$10\log\left[\frac{\frac{1}{n}(S_m - V)}{V}\right] = 20\log$

the smaller the better characteristics	$\frac{A}{\Delta^2}y^2$	$\frac{A}{\Delta^2}\left[\frac{1}{n}\sum_{i=1}^{n} y_i^2\right]$	$-10\log\left[\frac{1}{n}\sum_{i=1}^{n} y_i^2\right]$
large the better characteristics	$A\Delta^2\left(\frac{1}{y^2}\right)$	$A\Delta^2\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i^2}\right]$	$-10\log\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i^2}\right]$

6.1.4.2.4.2. Parametric design of quantitative values

Parametric design is the core of Taguchi's experimental design method, which is useful in product design and process design. Parameters refer to controllable factors that affect the characteristics of product performance, and parameter design refers to determining the optimal levels of these factors. These parameters are also called design variables, and in parametric design, the optimal conditions for design variables are obtained so that the product can achieve target quality while being insensitive to noise.

Parametric designs typically have several important characteristics:

It is mainly designed using an orthogonal array table, and two or more characteristic values are obtained under one experimental condition of control factors. The reason for obtaining multiple characteristic values under one experimental condition of control factors is to understand the influence of variable factors that are difficult to control, and there are two ways to repeatedly obtain characteristic values.

Repeatedly measuring characteristic values while leaving noise factors are unchanged

Placing noise factors using an orthogonal array as an outer array

During distributed analysis, the performance characteristics are not analyzed, but the SN ratio is analyzed.

Include all control factors that are expected to affect the characteristics of the system that is the subject of product design or process design, and place noise factors, block factors, etc. as non-control factors, but do not place too many of them.

Parameter design method for smaller the better / larger the better characteristics
Set up an experiment with control factors. (using the intersection table)

The SN ratio is calculated from repeated measurements for each experimental condition.

Through distributed analysis of the SN ratio, we find control factors that affect the SN ratio.

The level combination that maximizes the SN ratio is the optimal level combination. For control factors that do not have a significant effect on the SN ratio, an appropriate level is selected considering economic feasibility, workability, etc.

We estimate the population average of characteristic values at the optimal level combination and conduct a confirmation experiment to check whether there is reproducibility.

Parameter design method for nominal the best characteristics
Set up an experiment with control factors. (using orthogonal array table)

SN ratio and Sn are calculated from repeated measurements for each experimental condition. Here, Sn is a quantity that represents sensitivity and is a statistic defined to find a significant factor in the average of y.

Through analysis of variance of the SN ratio, we find control factors that have a significant impact on the SN ratio.

Find control factors that affect the average of y through variance analysis of Sn. Through analysis of variance of the SN ratio and Sn, control factors can be classified into three categories as follows.

dispersion control factor: Factors that have a significant impact on SN ratio

mean adjustment factor: Factors that have a significant effect only on the mean of y

Other control factors: Factors that do not simultaneously affect the SN ratio or the average of y

If one control factor simultaneously affects the SN ratio and the average of y, it is classified as a dispersion control factor.

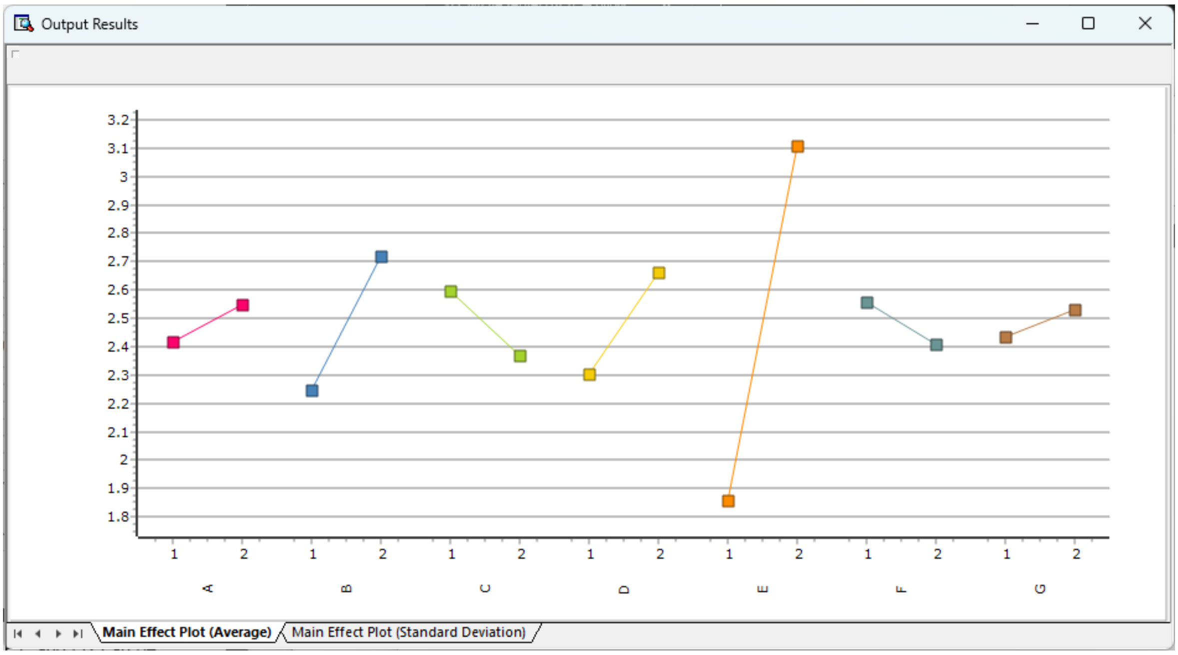
The dispersion control factor is set at a level that maximizes the SN ratio, and the level of the average adjustment factor is adjusted so that the average of y approaches the target value. For other control factors, select appropriate levels considering economic feasibility, workability, etc.

The population average of the characteristic values in the optimal level combination obtained above is estimated and a confirmation experiment is performed to determine whether the reproducibility is sufficient.

6.1.4.3 Plot

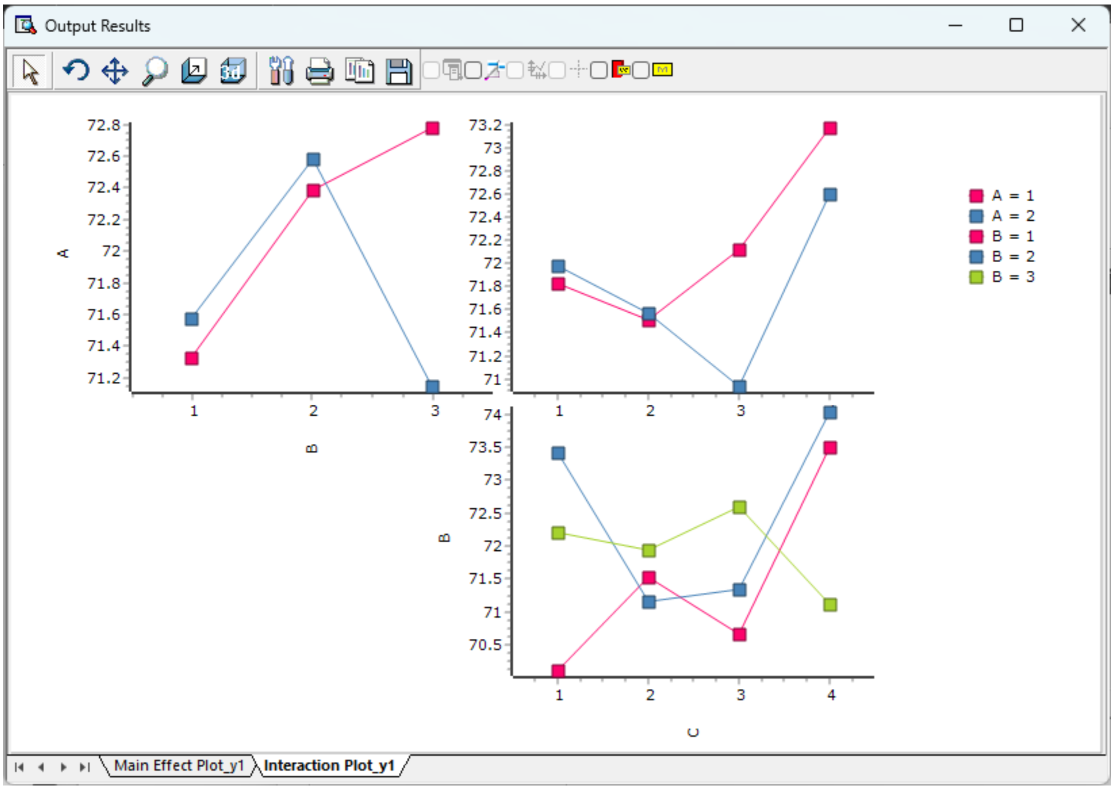
6.1.4.3.1. Main Effects Plot

The main effects plot is a plot that appears when we create a design with Process Variable added in factorial design (2-Level Factorial Design, Plackett-Burman Design, General Full Factorial Design) and Mixture Design. Using the main effects plot, you can intuitively understand what the response value or the fitted value in a regression model is at a specific level. In factorial design, you can select the data mean (using the response value directly) or the fitted mean (using the fitted value from the Regression Model), and in the design with Process Variable added in Mixture Design, only the data mean can be used.



6.1.4.3.2. Interaction Plot

This is a plot to display the interaction of two factors, such as AB and BC. For example, if factor A can have -1,1 and factor B can have -1,1 the average of the response values (or fitted values) of all experiments when factor A is -1 and B is -1, and the average of the response values (or fitted values) of all experiments when factor A is -1 and factor B is 1. Then draw one line. When factor A is 1 and factor B is -1 another line is drawn by connecting the average of the response values of all experiments and the average of the response values (or fitted values) of all experiments when factor A is 1 and factor B is 1. If we draw this on one screen, we can see the interaction of the ABs. If there are three factors, a plot can be drawn for AB, AC, and BC, and if the number of factors increases, an additional plot must be drawn accordingly.

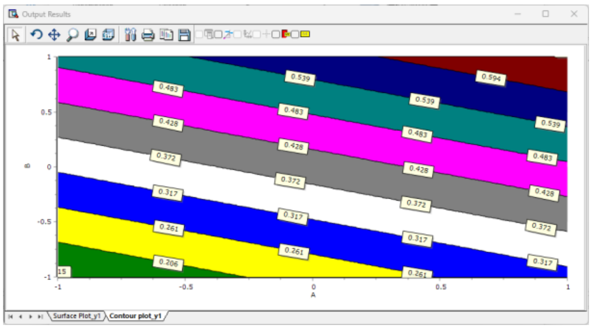


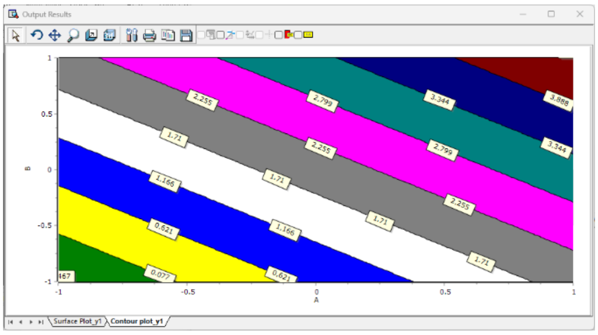
6.1.4.3.3. Contour Plot

It is said that the following regression model was estimated through the regression.

$$y = f(x_1, x_2, \ldots, x_n)^{(t)}$$

In order to express the shape of such a model on a two-dimensional plane, n-2 factors (n-3 components in Mixture Design) are fixed and then the area of values that the remaining 2 factors (3 components in Mixture Design) can have. The y value is obtained from and the larger the y value, the darker the color, and the smaller the area, the lighter the color. This provides an intuitive understanding of the shape of the response surface.



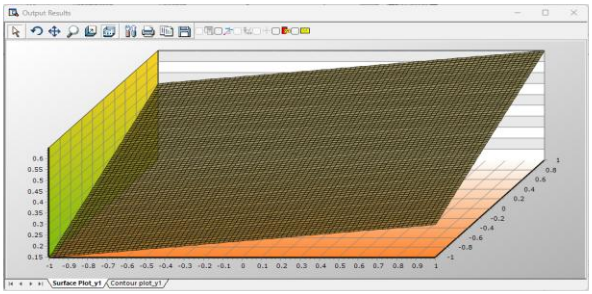


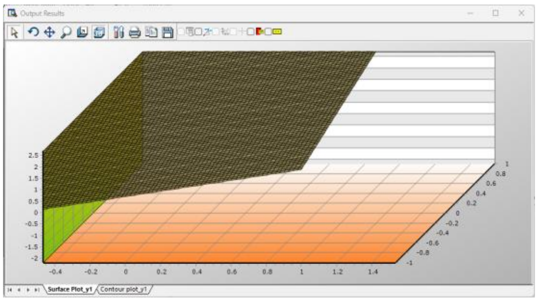
6.1.4.3.4. Surface Plot

It is said that the following regression model was estimated through the regression.

$$y = f(x_1, x_2, \ldots, x_n)$$

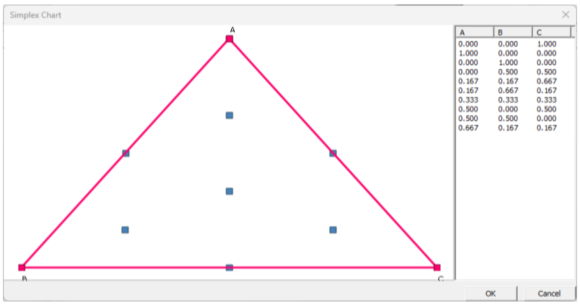
In order to express this model in three dimensions, n-2 factors (n-3 components in Mixture Design) are fixed, and then the y-values are calculated from the range of values that the remaining 2 factors (3 components in Mixture Design) can have. Find and draw a plane by considering it as the height of the 3D graph. This can be said to be the easiest tool to understand the response surface obtained through Regression along with the contour plot.

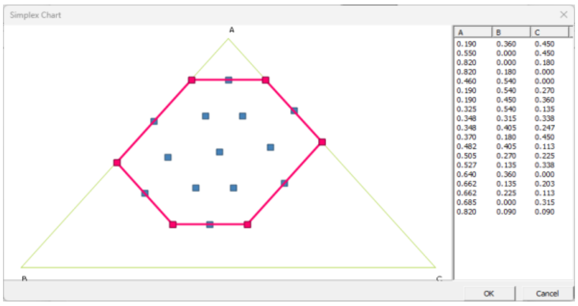




6.1.4.3.5. Simplex Design Plot

While the main effects plot, interaction plot, surface plot, and contour plot are all plots that are interested in response values, the simplex design plot provided only by Mixture Design is a plot that shows how the experimental points are arranged. DOE's Mixture Design method presents a method of placing experimental points with components where the sum of the components is constant. Mixture Design includes Simplex Center Design, Simplex Lattice Design, and Vertex Design. In each of these designs, you can check whether the experimental points are appropriately placed in space by showing how the experimental points are determined and arranged in space.





6.1.4.4 Response Optimizer

Basically, DOE's primary goal is to design an experiment and create a regression model using the results of the experiment. This is because meaningful results can be obtained with just this regression model. To add a little more, residual analysis, analysis for variance, etc. are performed to determine the significance of the regression model. This somewhat concludes the DOE's process.

However, for user convenience and ease of interpretation, a plot can be drawn to express the results more visually and intuitively. For this purpose, in case of factorial design, main effects plot, interaction plot, surface plot, and contour plot are provided, and in case of response surface design, surface plot and contour plot are provided. In the case of Mixture Design, surface plots and contour plots are provided, and when Process Variable is added, main effect plots and interaction plots for process variables are provided.

However, the final step of DOE can be said to be the response optimizer. Response optimizer is not simply aimed at maximizing or maximizing the response value of a regression model. Of course, this partial purpose can be fully achieved by adjusting various input values. Response optimizer, as DOE calls it, is a useful tool for meeting more general purposes. For example, if there are two or more response values in an experiment and you want to make one response value larger and one response value smaller, this can be solved by creating a new function and addressing the problem of increasing this function. Of course, this function will grow when the first response value is large, and it will grow when the second response value is small. After this, we will explain how to mathematically formalize this to reach the goal.

6.1.4.4.1. Desirability Function

There is one desirability function for one reaction column. If there is only one reaction column, you need to find a combination of the levels of factors that optimize this one desirability function. If there are multiple reaction heats and you want to find the optimal factor level combination considering all multiple reaction heats, you can create a single overall desirability function that considers multiple desirability functions and optimize it. If there are m reaction heats and the desirability functions for them are$d_1, d_2, \ldots, d_m$ respectively, and the importance of each reaction heat is $r_1, r_2, \ldots, r_m$ respectively, then the comprehensive desirability function that takes all of them into consideration. Is

$$\boldsymbol{D} = \left(\boldsymbol{d}_1^{r_1}\boldsymbol{d}_2^{r_2} \ldots \boldsymbol{d}_m^{r_m}\right)^{\frac{1}{\sum_{i=1}^{m} r_i}}$$

However, since each d_i is a function of factor's $(x_1, x_2, \ldots' s)$, D can also be called a function of $(x_1, x_2, \ldots' s)$

The final goal is to find the combination of $x_1, x_2, \ldots$ to maximize D.

Minimization

As explained above, one reaction value corresponds to one desirability function. The goal is to maximize this desirability function, whether you want to minimize the response, maximize it, or hit the target value. Therefore, the desirability function will depend on the kind of optimization you want. When you select Minimize, you can set the upper limit and target value. When the upper limit is U and the target value is T, the desirability function in minimization is

$$d = 1 \quad if \quad y < T$$

$$d = \left(\frac{U - y}{U - T}\right)^{\lambda} \quad if\ T \leq y \leq U$$

$$d = 0 \quad y > U$$

$$\lambda\ means\ the\ weight$$

Maximization

When you select Maximize, you can select a lower limit and target value. When the lower limit is L and the target value is T, the desirability function in maximization is

$$d = 0 \quad if \quad y < L$$

$$d = \left(\frac{y - L}{T - L}\right)^{\lambda} \quad L \leq y \leq T$$

$$d = 1 \quad y > T$$

$$\lambda\ means\ the\ weight$$

Hit target value

When hit the target value is selected, you can select the upper and lower limits and the target value. when lower limit is L, upper value is U and the target value is T,

$$d = 0 \quad y < L$$

$$d = \left(\frac{y - L}{T - L}\right)^{\lambda} \quad L \leq y \leq T$$

$$d = \left(\frac{U - y}{U - T}\right)^{\lambda} \quad T \leq y \leq U$$

$$d = 0 \quad y > U$$

$$\lambda\ means\ the\ weight$$

The y that commonly appears above is a regression equation created through a Regression Model. Therefore, the following equation is established

$$y = f(x_1, x_2, \dots, x_k)$$

and is also a function of $x_1, x_2, \dots 's$ Through this, each d can be obtained, and by calculating it considering the weight, the overall desirability function can be obtained. Summarizing this process, it is as follows.

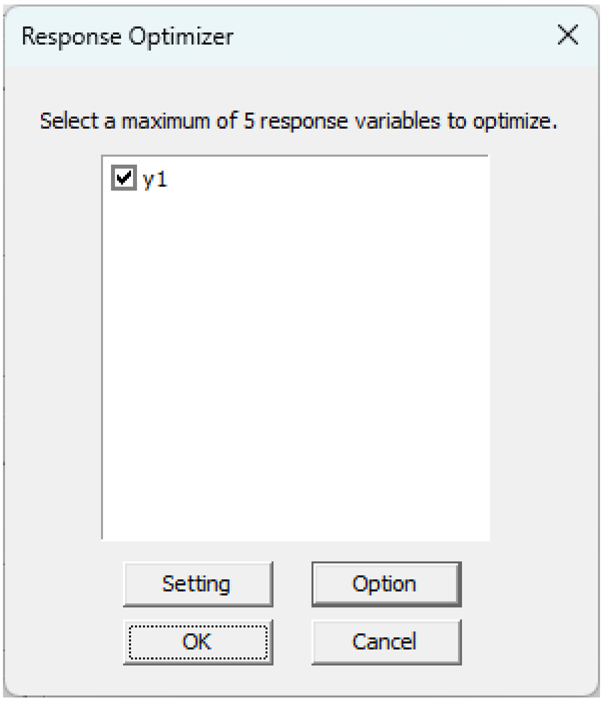
$$\begin{tikzcd}
\mathbf{x}=[x_1\ \ x_2\ \ \cdots\ \ x_n]^T \arrow[d, bend right, "\text{Regression}"] & & D = D(\mathbf{x}) \\
\mathbf{y}(\mathbf{x})=[y_1(\mathbf{x})\ \ y_2(\mathbf{x})\cdots\ \ y_m(\mathbf{x})]^T \arrow[r, "\text{Mapping}"] & \mathbf{d}(\mathbf{x})=[d_1(\mathbf{x})\ \ d_2(\mathbf{x})\cdots\ \ d_m(\mathbf{x})]^T \arrow[ur, bend right, "\text{Composite}"] &
\end{tikzcd}$$

Now we need to introduce an optimization algorithm to maximize this overall desirability function. This overall desirability function has a different shape from the functions you commonly encounter. Differentiation is impossible in many respects, so the Derivative Based Optimization Algorithm cannot be used. in ECMiner^{TM} DOE, Box. M.J's Constrained Simplex Algorithm is used. Through this algorithm, the overall desirability function can be maximized.

6.1.4.4.2. ECMiner^{TM} DOE Response Optimizer

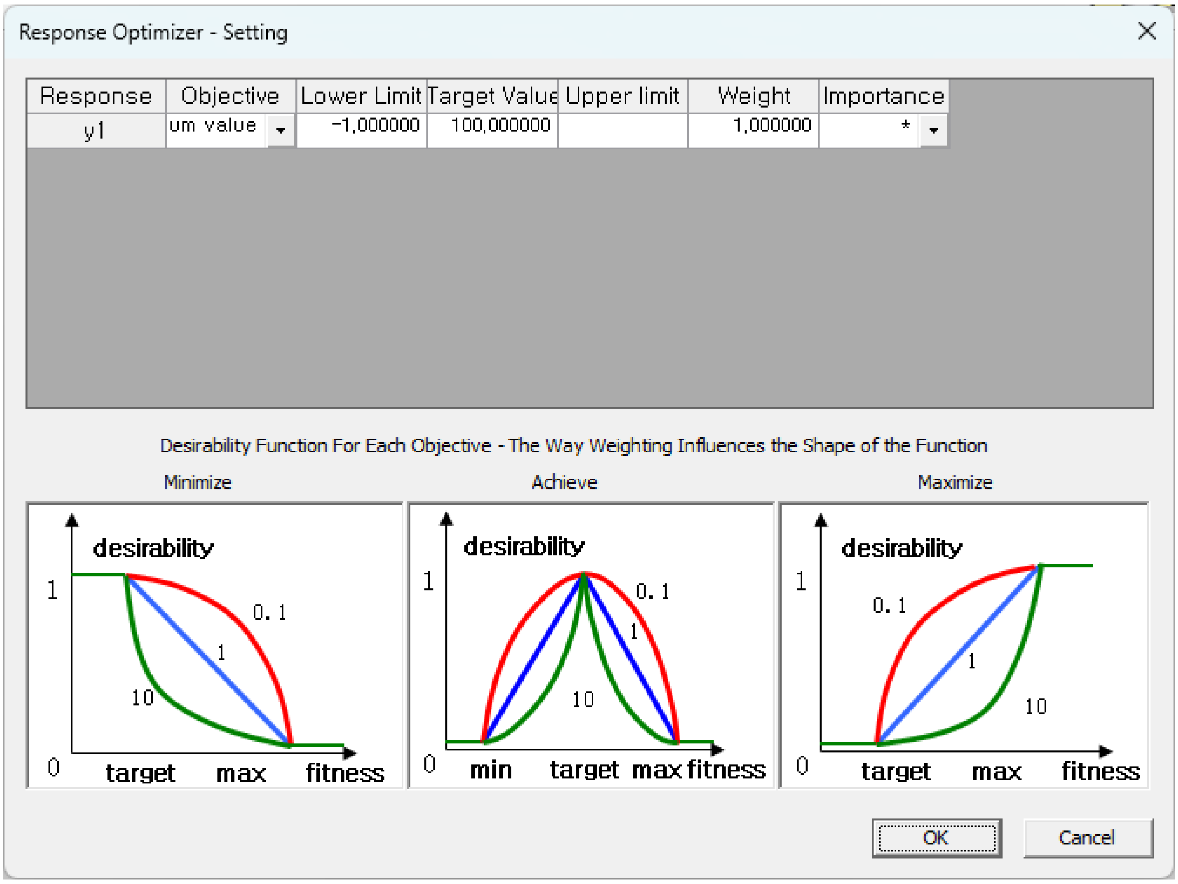
Response Variation Selection

ECMiner™ DOE response optimizer begins with selecting the response variables to be optimized. For multivariate response optimizer, select two response variables.

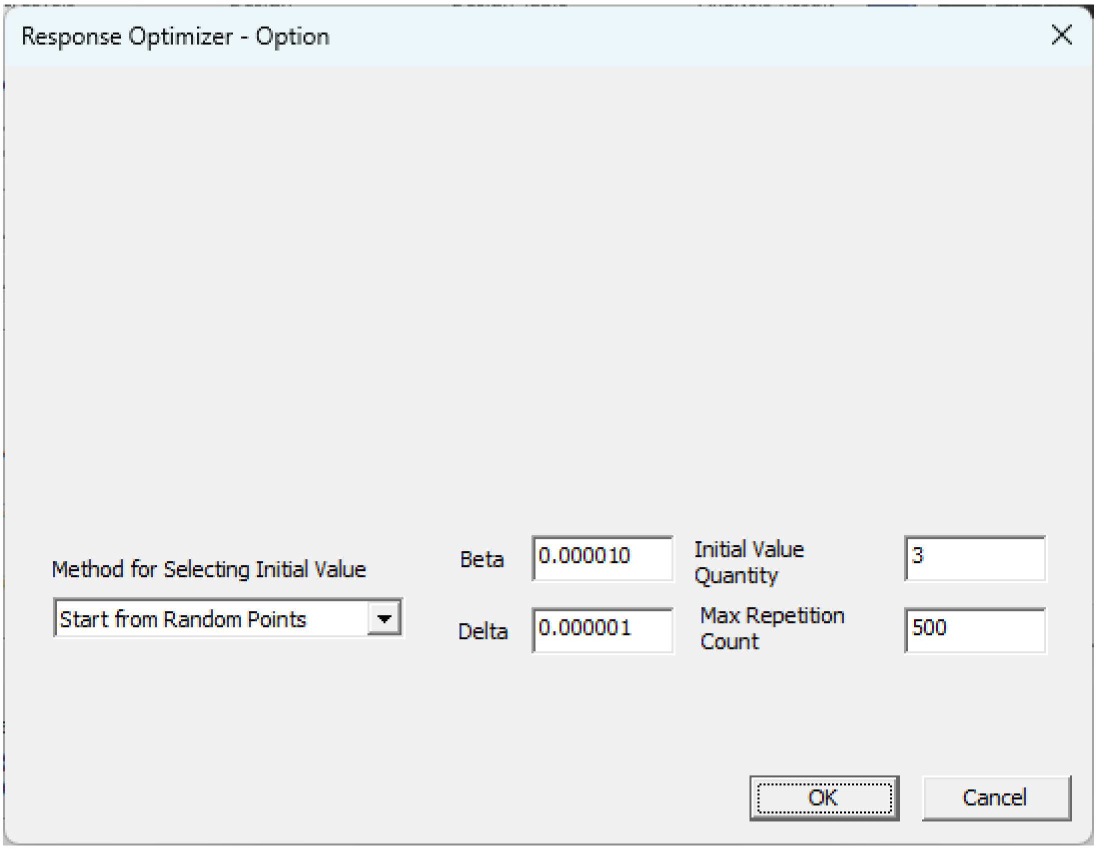


Responsive Optimization Settings

When you press the **Setting** button, the following screen appears. If you want to maximize yield 1 and minimize yield 2, enter the following things. At this time, the weight determines the curvature of the curve as shown in the desirability function figure, and importance is an indicator of the degree of importance of each response variable.



Responsive Optimization Option



Select the options required by the optimization algorithm.

- **Method for Selecting Initial Value**

  Initial values play a crucial role in all optimization algorithms. This is because the initial values can determine whether the algorithm finds a local optimum or a global optimum. ECMinerTM DOE provides the selection methods of factorial design, response surface design, starting from a grid point, starting from a random point, and starting from a user-defined point, and Mixture Design provides a method that starts with a random point selection.

- **Beta**

  Due to the nature of the algorithm, Overall Desirability obtained from multiple points is stored, and the convergence condition is that the maximum difference between Overall Desirability obtained from multiple points must be smaller than Beta. Therefore, for users who want a more accurate value, the desired goal can be achieved by making this Beta value minimal.

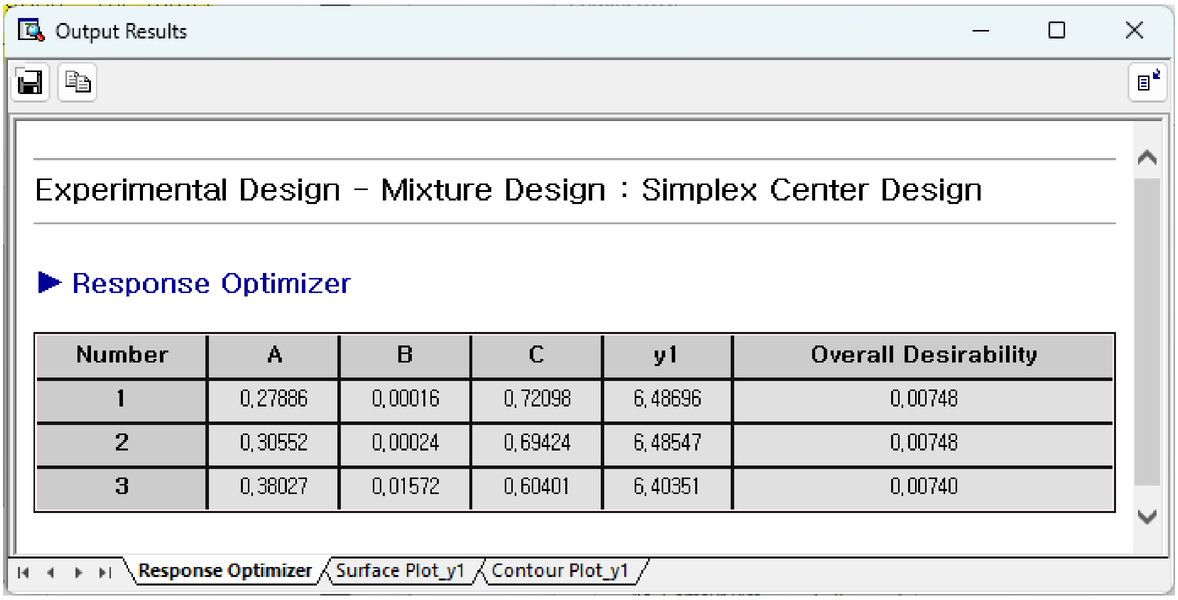
- **Delta**

  Delta is a measure that determines how far a point should be brought back into the restricted area when it goes out of the restricted area. However, empirically, the measure does not have a decisive effect on the performance of the algorithm. Initial Value Quantity: It is a decision to perform the algorithm on a number of different initial values.

- **Max Repetition Count**

  It determines the maximum number of times to repeat the algorithm. The more iterations you have, the more likely you are to find a better solution.

Output Results



Number	A	B	C	y1	Overall Desirability
1	0.27886	0.00016	0.72098	6.48696	0.00748
2	0.30552	0.00024	0.69424	6.48547	0.00748
3	0.38027	0.01572	0.60401	6.40351	0.00740

ECMiner™ DOE shows the optimization results as above. The user can achieve the goal by showing at what value of each factor (or component) the Overall Desirability is maximized.

6.2 Probability Distribution

- Beta
- Binominal
- Chi-squared
- Exponential
- F
- Normal
- Poisson
- T
- Discrete Uniform
- Continuous Uniform
- Weibull

6.2.1 Beta distribution

The Beta distribution is a continuous probability distribution on (0,1) defined with two parameters α and β. It is used to model proportions and the distribution of probabilities. As the parameters α and β vary, the Beta distribution takes on many shapes.

PDF of Beta distribution

$$f(x|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$
$$0 \le x \le 1, \alpha > 0, \beta > 0$$

CDF of Beta distribution

$$F(x|a,b) = \frac{1}{B(\alpha,\beta)} \int_0^x t^{\alpha-1}(1-t)^{\beta-1} dt$$

Mean and variance of Beta distribution

$$E(X) = \frac{\alpha}{\alpha+\beta}$$

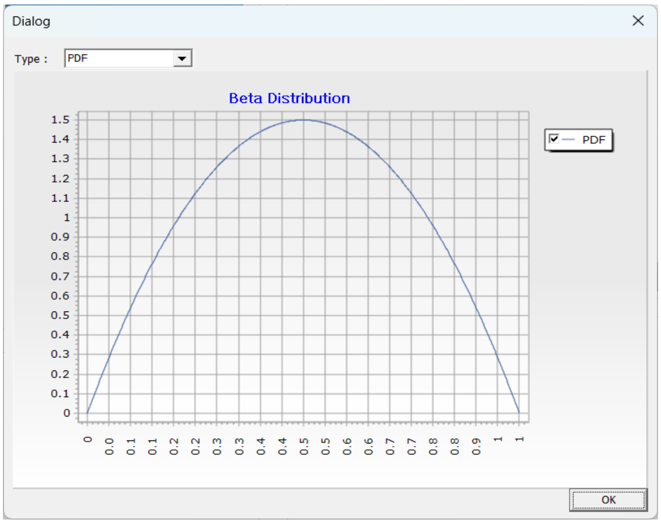
$$Var(X) = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

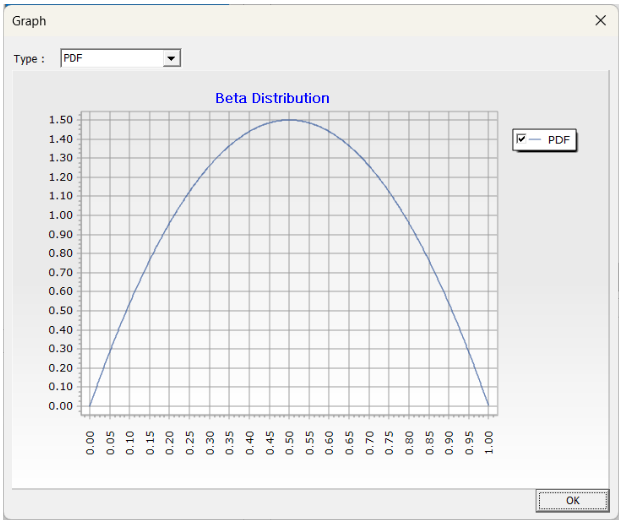
Inverse of cumulative distribution function of Beta distribution

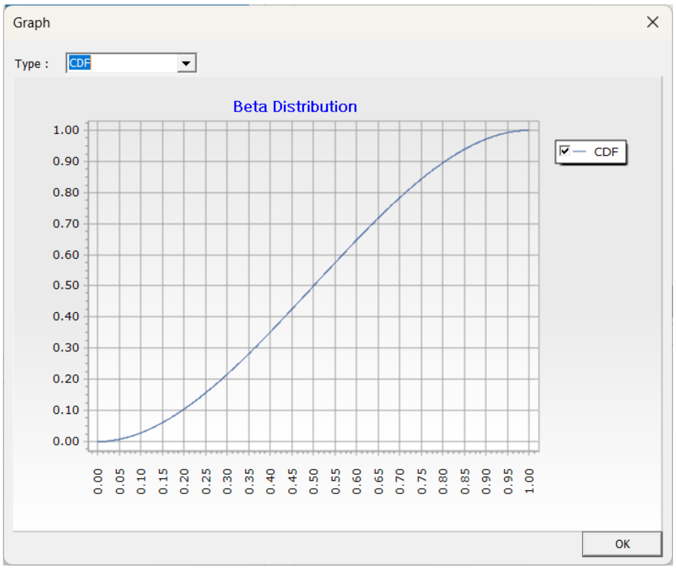
$$x = F^{-1}(p|\alpha,\beta) = \{x : F(x|\alpha,\beta) = p\}$$

Example

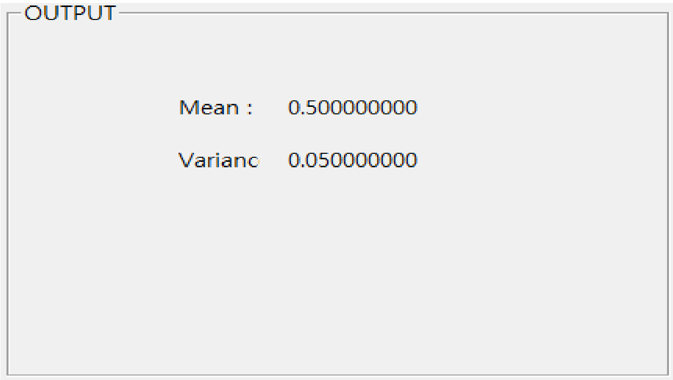
- PDF and CDF graph when $\alpha = 2$ and $\beta = 2$







- Mean and variance when $\alpha = 2$ and $\beta = 2$



6.2.2 Binomial distribution

Binomial distribution is derived from Bernoulli distribution. Bernoulli random variables have the following properties:

$$X_k = \begin{cases} 1 & \textit{with probability p} \\ 0 & \textit{with probability 1-p} \end{cases}$$

The binomial distribution represents the total number of successes in n independent Bernoulli trials, each with the same probability of success p. The total number of successes X in n trials can be expressed as:

$$\boldsymbol{X = X_1 + X_2 + \ldots + X_n}$$

The binomial distribution represents the probability of achieving exactly k successes and n−k failures in n independent trials, where the probability of success, 1 in each trial is p, and the probability of failure, 0 is 1−p.

PMF (Probability Mass Function) of Binomial distribution

$$\boldsymbol{f(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}}$$

CDF (Probability Density Function) of Binomial distribution

$$\boldsymbol{F(x|n,p) = \sum_{i=1}^{x} \binom{n}{i} p^i (1-p)^{n-i}}$$

Mean and variance of Binomial distribution

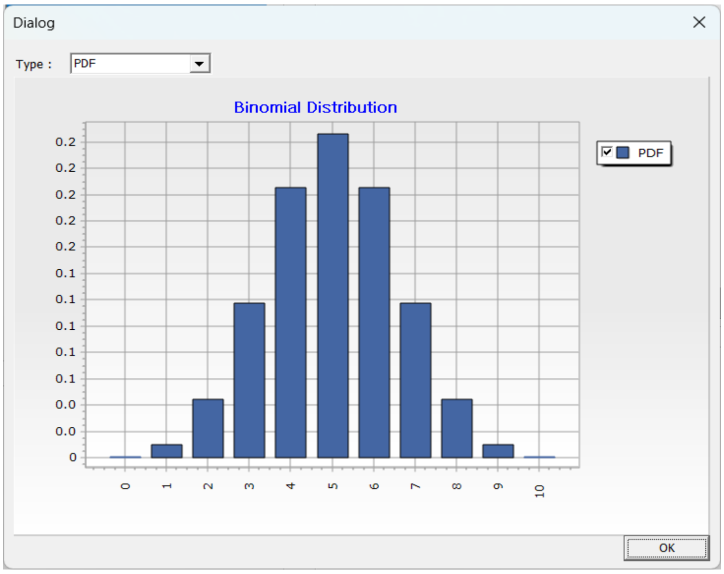
$$\boldsymbol{E(X) = np}$$
$$\boldsymbol{Var(X) = np(1-p)}$$

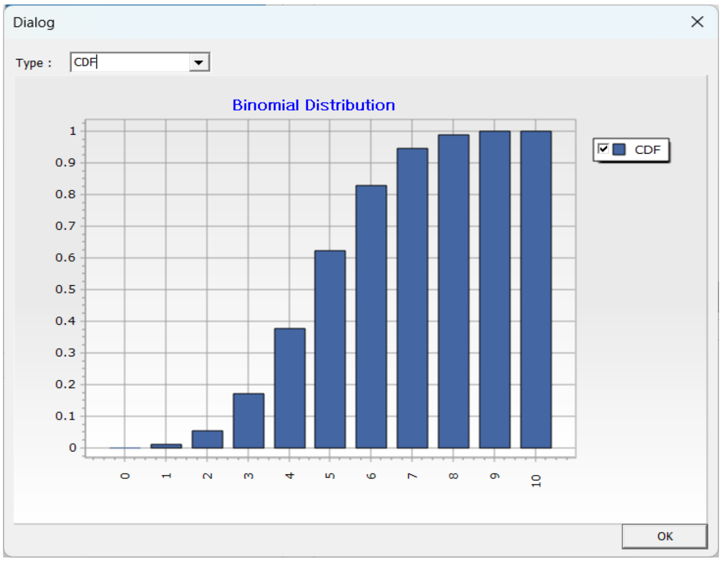
Inverse of cumulative distribution function of Binomial distribution

$$\boldsymbol{x = F^{-1}(p|n,p)}$$

Example

- PMF, CDF graph when $n = 10$, $p = 0.5$, and $x = 0.5$





- Mean, variance, PMF, and CDF when $n = 10$, $p = 0.5$, and $x = 0.5$

OUTPUT

Mean : 5.000000000

Variance : 2.500000000

pmf : 0.246093750

cdf : 0.623046875

6.2.3 Chi-squared distribution

The chi-squared distribution is the distribution of the sum of the squares of *n* independent random variables $Z_1, Z_2, Z_3, \ldots, Z_n$, each following a standard normal distribution $N(0,1)$. Chi-squared distribution defined as:

$$\chi^2 = \sum_{i=1}^{n} Z_i^2$$

The degree of freedom of the chi-square distribution is typically denoted as υ (called 'nu'). The chi-square distribution is used in statistical testing and analysis, including the Goodness of Fit Test and Likelihood Ratio Test.

PDF of Chi-squared distribution

$$f(x|\nu) = \frac{x^{\frac{\nu-2}{2}} e^{-\frac{x}{2}}}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})}$$

CDF of Chi-squared distribution

$$F(x|\upsilon) = \int_0^x \frac{t^{\frac{\upsilon-2}{2}} e^{-\frac{t}{2}}}{2^{\frac{\upsilon}{2}} \Gamma(\frac{\upsilon}{2})} dt$$

Mean and variance of Chi-squared distribution

$$E(X) = \int_0^\infty t \,. \frac{t^{\frac{\upsilon-2}{2}} e^{-\frac{t}{2}}}{2^{\frac{\upsilon}{2}} \Gamma(\frac{\upsilon}{2})} dt = \upsilon$$

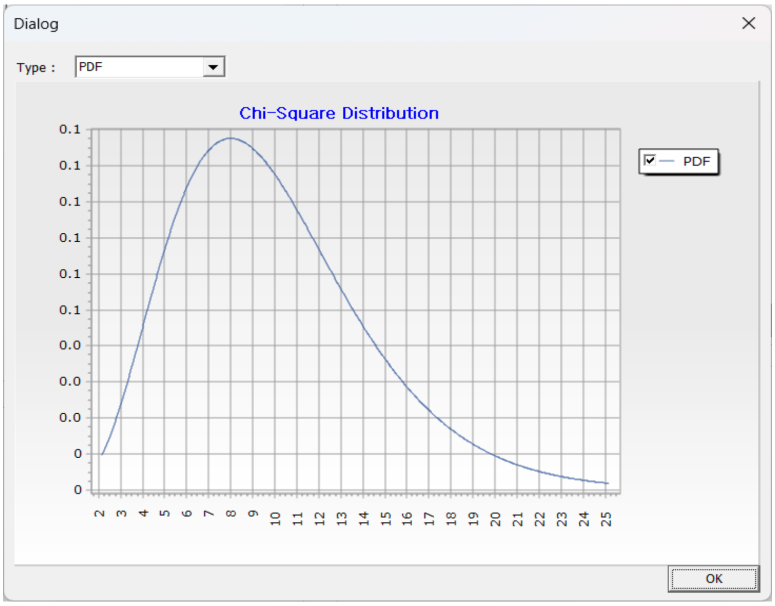
$$Var(X) = E(X^2) - (E(X))^2 = \int_0^\infty t^2 \cdot \frac{t^{\frac{v-2}{2}} e^{-\frac{t}{2}}}{2^{\frac{v}{2}} \Gamma(\frac{v}{2})} dt - v^2 = 2v$$

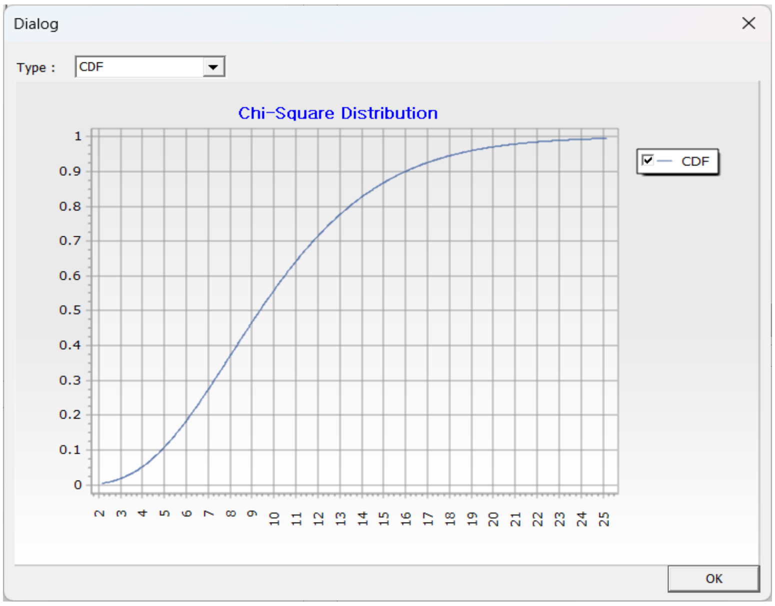
Inverse of cumulative distribution function of Chi-squared distribution

$$x = F^{-1}(p|v)$$

Example

- PDF, CDF graph when $v = 10$ and $x = 5$





- Mean, variance, PDF, and CDF when $v = 10$ and $x = 5$

OUTPUT

Mean :	10.000000000
Varianc	20.000000000
pdf :	0.066800943
cdf :	0.108821981

- Inverse of cumulative distribution function when $v = 10$ and $p = 0.5$

OUTPUT

Mean :	10.000000000
Varianc	20.000000000
inv :	9.341817766

6.2.4 Exponential distribution

The exponential distribution is a type of continuous probability distribution defined by a single parameter, λ, which represents the rate at which events occur. The exponential distribution is used to model the time intervals between occurrences of a specific event. It models the time intervals between independently occurring events in a Poisson process and has the characteristic that the probability of an event occurring remains constant over time.

A higher λ indicates events occur more frequently, while a lower λ indicates they are less frequent.

PDF of exponential distribution

$$f(x|\mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$$

$$f(x|\mu) = \lambda e^{-\lambda x}$$

$$x > 0, \mu > 0$$

CDF of exponential distribution

$$F(x|\mu) = \int_0^x \frac{1}{\mu} e^{-\frac{x}{\mu}} dt = 1 - e^{-\frac{x}{\mu}}$$

$$F(x|\mu) = \int_0^x \lambda e^{-\lambda x} dt = 1 - e^{-\lambda x}$$

$$x > 0, \mu > 0$$

Mean and variance of exponential distribution

$$E(X) = \int_0^\infty x f(x|\mu) dx = \mu$$

$$E(X) = \int_0^\infty x f(x|\mu) dx = \mu = \frac{1}{\lambda}$$

$$Var(X) = E(X^2) - (E(X))^2 = \mu^2$$

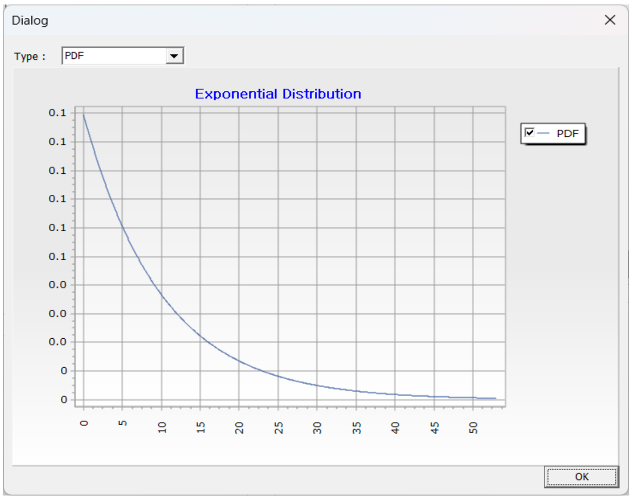
$$Var(X) = E(X^2) - (E(X))^2 = \mu^2 = \frac{1}{\lambda^2}$$

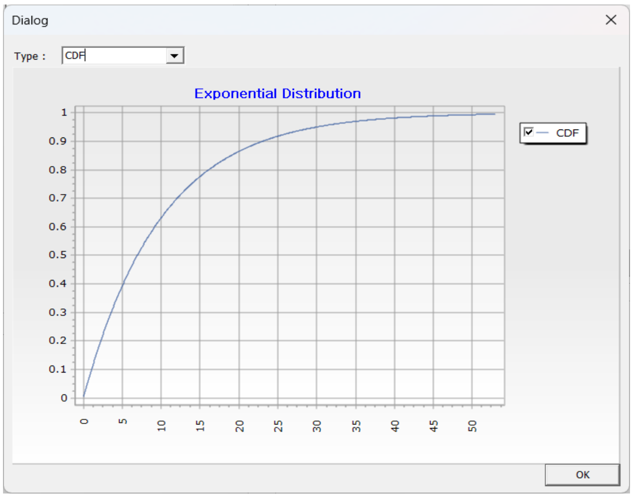
Inverse of cumulative distribution function of exponential distribution

$$x = F^{-1}(p|\mu) = -\mu \ln(1 - p)$$

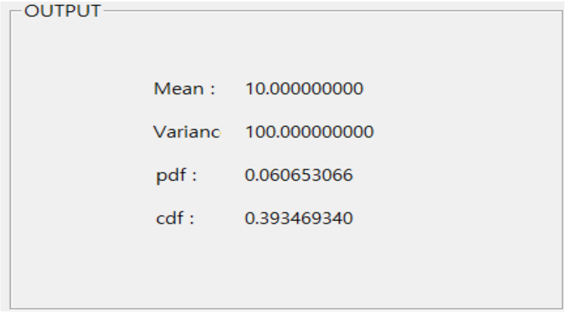
Example

- PDF, CDF graph when $\mu = 10$ and $x = 5$

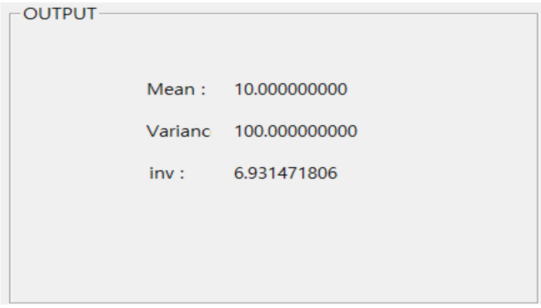




- Mean, variance, PDF, and CDF when $\mu = 10$ and $x = 5$



- Inverse of cumulative distribution function when $\mu = 10$ and $p = 0.5$



6.2.5 F-distribution

The F-distribution is a type of continuous probability distribution defined as the ratio of two independent chi-squared distributions with degrees of freedom $v_1$and v_2. The F-distribution is used in statistical hypothesis testing, in analysis of variance (ANOVA) and regression analysis, and is particularly useful for testing whether the variances between two groups are significantly different.

PDF of F-distribution

$$f(x|\nu_1,\nu_2) = \frac{\Gamma[(\nu_1+\nu_2)/2]}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)}(\nu_1/\nu_2)^{\nu_1/2}\frac{x^{(\nu_1-2)/2}}{[1+(\nu_1/\nu_2)x]^{(\nu_1+\nu_2)/2}}$$

$$\nu_1,\nu_2 > 0, x > 0$$

CDF of F-distribution

$$F(x|\nu_1,\nu_2) = \int_0^x \frac{\Gamma[(\nu_1+\nu_2)/2]}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} (\nu_1/\nu_2)^{\nu_1/2} \frac{t^{(\nu_1-2)/2}}{[1+(\nu_1/\nu_2)t]^{(\nu_1+\nu_2)/2}} dt$$

$$\nu_1, \nu_2 > 0,\ x > 0$$

Mean and variance of F-distribution

$$E(X) = \frac{\nu_2}{\nu_2-2} \quad \nu_2 > 2$$

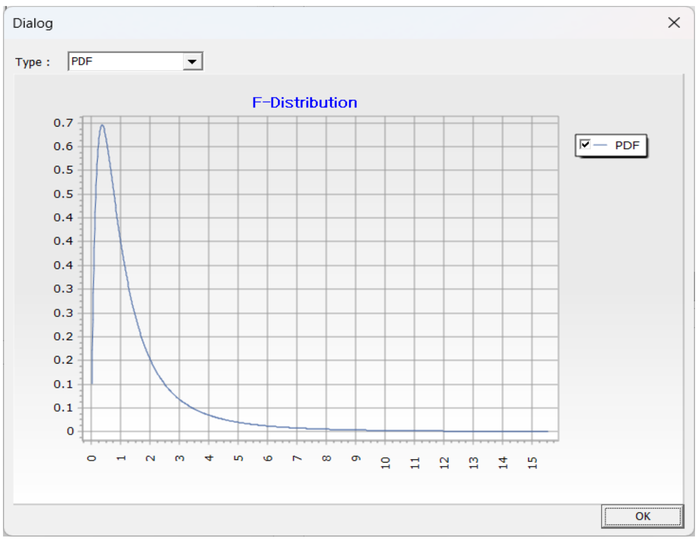
$$Var(X) = 2\left(\frac{\nu_2}{\nu_2-2}\right)^2 \frac{(\nu_1+\nu_2-2)}{\nu_1(\nu_2-4)} \quad \nu_2 > 4$$

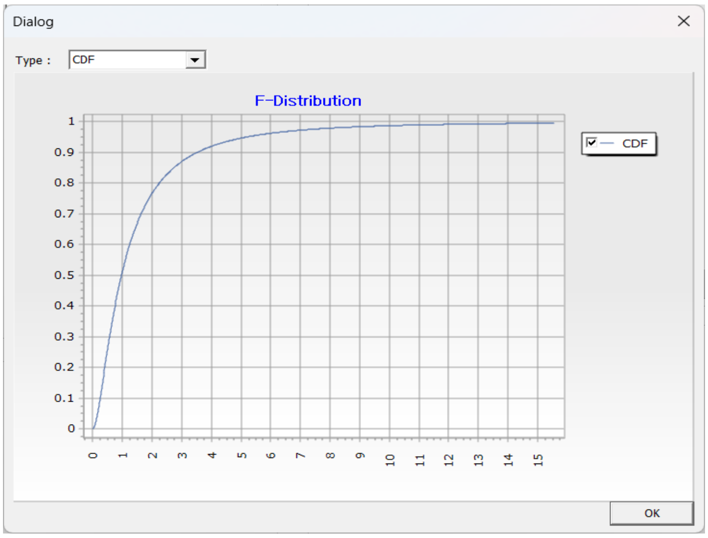
Inverse of cumulative distribution function of F-distribution

$$x = F^{-1}(p|\nu_1,\nu_2)$$

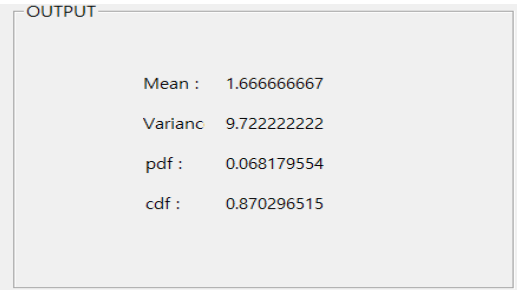
Example

- PDF and CDF graph when $\nu_1 = 4$, $\nu_2 = 5$, and $x = 3$

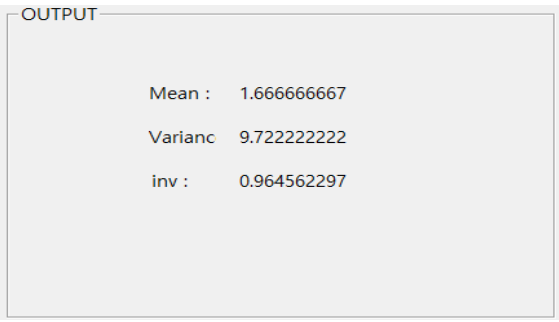




- Mean, variance, PDF, and CDF when $v_1 = 4, v_2 = 5$, and $\boldsymbol{x = 3.}$



- Inverse of cumulative distribution function when $v_1 = 4,\ v_2 = 5$, and $p = 0.5$



6.2.6 Normal distribution

Normal distribution is a type of continuous probability distribution with a bell-shaped, symmetric curve. It shows how data is distributed around the mean μ and variance σ^2. According to the Central Limit Theorem, when the sample size is sufficiently large, the average of multiple distributions approaches a normal distribution.

PDF of normal distribution

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x-\mu}{\sigma^2}\right)^2}$$
$$\sigma > 0$$

CDF of normal distribution

$$F(x|\mu,\sigma) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{t-\mu}{\sigma^2}\right)^2} dt = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sqrt{2\sigma^2}}\right)\right]$$
$$\sigma > 0$$

$erf()$ *is error function*

Mean and variance of normal distribution

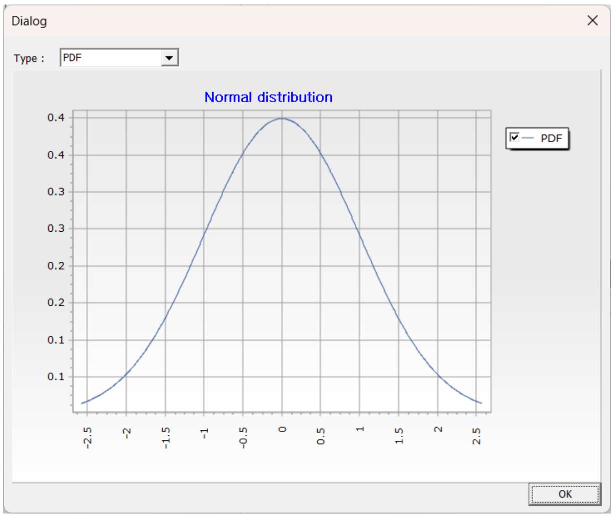
$$E(X) = \mu$$
$$Var(X) = \sigma^2$$

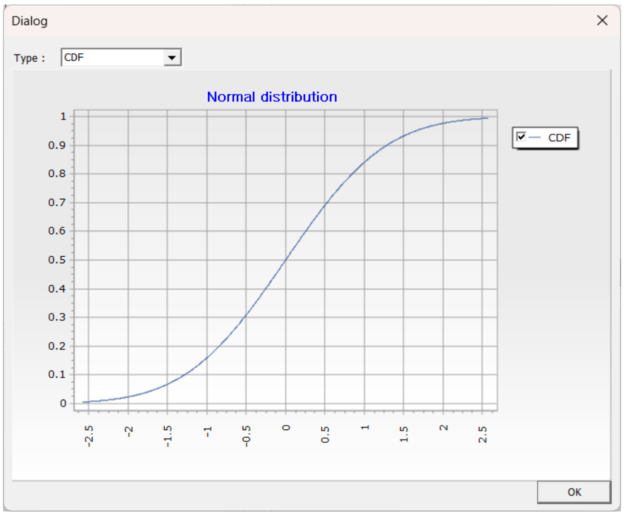
Inverse of cumulative distribution function of normal distribution

$$x = F^{-1}(p|\mu,\sigma)$$

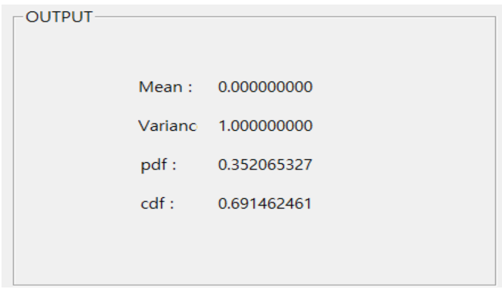
Example

- PDF, CDF graph when $\mu = 0$, $\sigma = 1$, and $x = 0.5$.

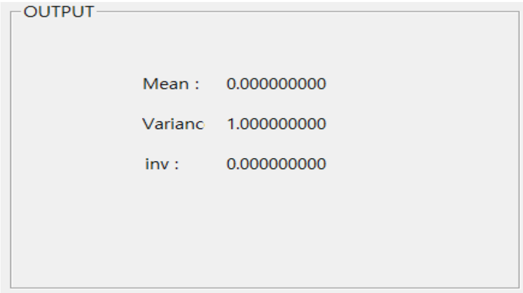




- Mean, variance, PDF, and CDF when $\mu = 0$, $\sigma = 1$, and $x = 0.5$.



- Inverse of cumulative distribution function when $\mu = 0, \sigma = 1$, and $p = 0.5$



6.2.7 Poisson distribution

The Poisson distribution is a discrete probability distribution defined by a single parameter, λ, which represents the average rate at which events occur. The Poisson distribution is used to model the number of occurrences of a specific event within a given time period. It is suitable for situations where events occur independently and at a constant average rate, making it ideal for modeling the number of events in a Poisson process.

PMF of Poisson distribution

$$f(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$x = 0, 1, 2, \ldots \quad \lambda > 0$$

CDF of Poisson distribution

$$F(x|\lambda) = e^{-\lambda} \sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!}$$

$$x = 0, 1, 2, \ldots \quad \lambda > 0$$

MGF of Poisson distribution

$$E(e^{tX}) = e^{\lambda(e^t - 1)}$$

Mean and variance of Poisson distribution

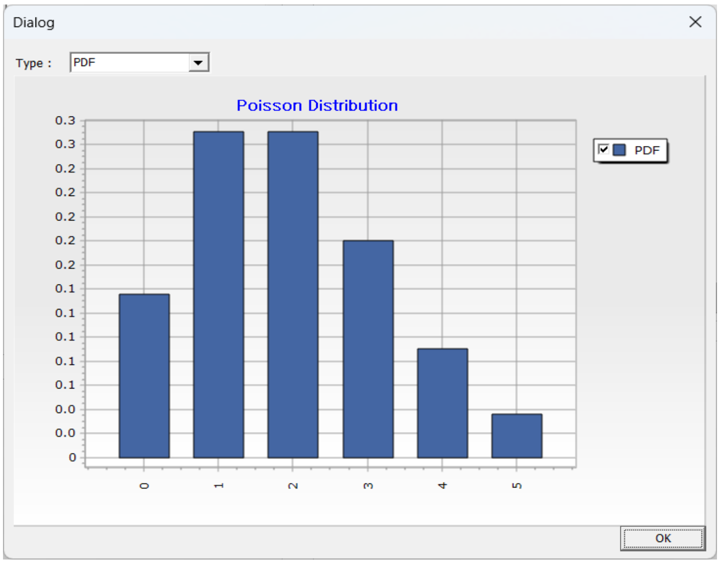
$$E(X) = \lambda$$
$$Var(X) = \lambda$$

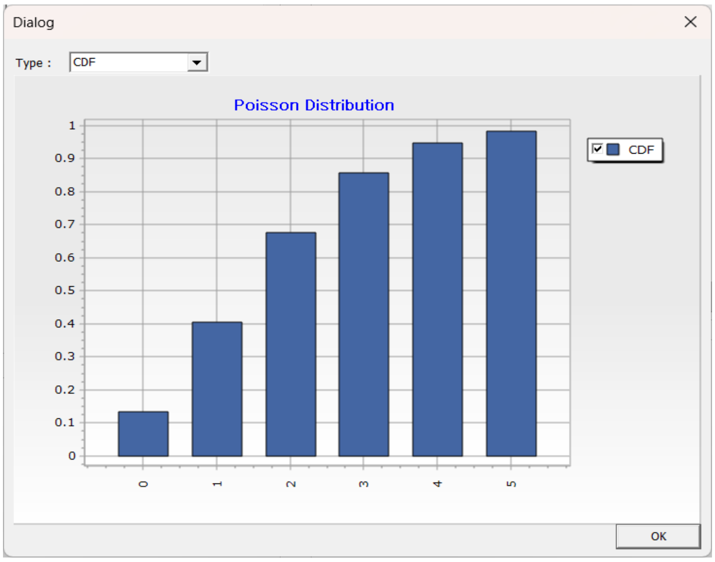
Inverse of cumulative distribution function of Poisson distribution

$$x = F^{-1}(p|\lambda)$$

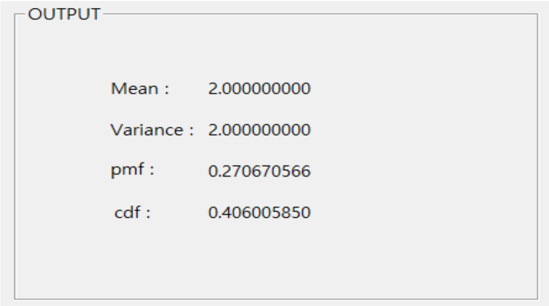
Example

- PMF, CDF graph when $\lambda = 2$ and $x = 1$.





- Mean, variance, PMF, and CDF when $\lambda = 2$ and $x = 1$.



- Inverse of cumulative distribution function when $\lambda = 2$ and $p = 0.5$

6.2.8 T-distribution

The t-
distribution is a continuous probability distribution defined by degrees of freedom v. It is used for estimating confidence intervals for sample means and hypothesis testin
g.

PDF of T-distribution

$$\boldsymbol{f}(\boldsymbol{x}|\boldsymbol{\nu}) = \frac{\boldsymbol{\Gamma}\left(\frac{\boldsymbol{\nu}+\mathbf{1}}{\mathbf{2}}\right)}{\sqrt{\boldsymbol{\nu\pi}}\,\boldsymbol{\Gamma}\left(\frac{\boldsymbol{\nu}}{\mathbf{2}}\right)} \frac{\mathbf{1}}{\left(\mathbf{1}+\frac{\boldsymbol{t}^{\mathbf{2}}}{\boldsymbol{\nu}}\right)^{(\nu+1)/2}}$$

CDF of T-distribution

$$F(x|\nu) = \int_{-\infty}^{x} f(t|\nu)\, dt = I_x\left(\frac{\nu}{2}, \frac{\nu}{2}\right) = \frac{B\left(x; \frac{\nu}{2}, \frac{\nu}{2}\right)}{B\left(\frac{\nu}{2}, \frac{\nu}{2}\right)}$$

$B\left(x; \frac{\nu}{2}, \frac{\nu}{2}\right)$ *is incomplete beta function*

Mean and variance of T-distribution

$$E(X) = 0 \qquad if\ v > 1$$

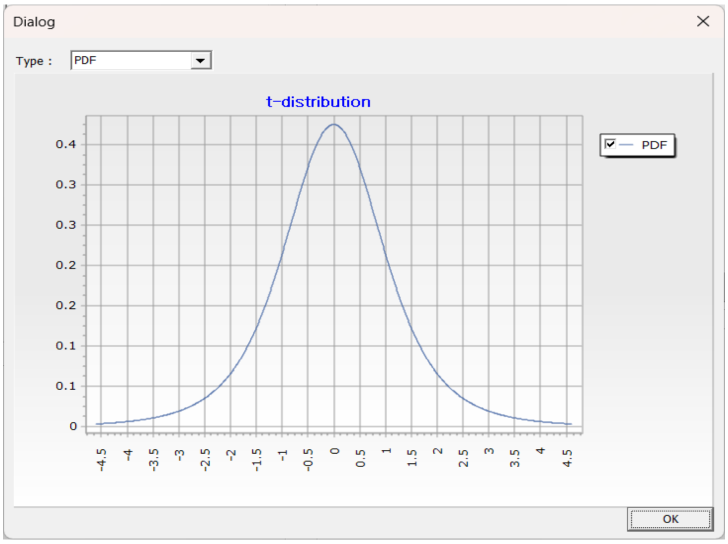
$$Var(X) = \frac{v}{v-2} \qquad if\ v > 2$$

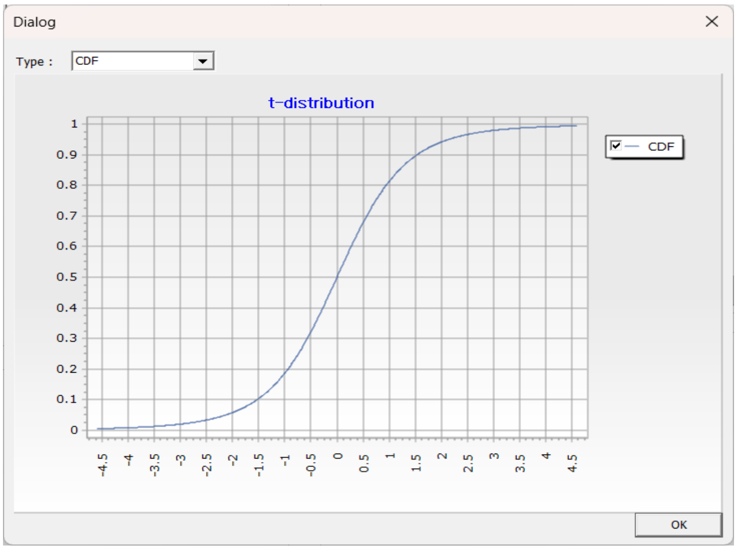
Inverse of cumulative distribution function of T-distribution

$$x = F^{-1}(p|\nu)$$

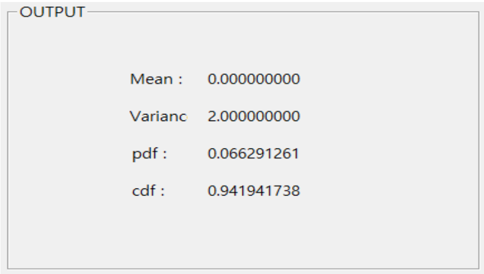
Example

- PDF, CDF graph when $\nu = 4$ and $x = 2$.

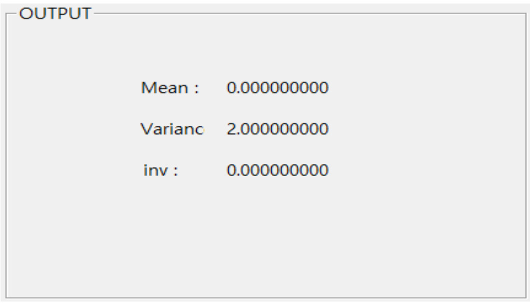




- Mean, variance, PDF, and CDF when $\nu = 4$ and $x = 2$.



- Inverse of cumulative distribution function when $\nu = 4, p = 0.5$



6.2.9 Discrete uniform distribution

Discrete uniform distribution is a discrete probability distribution where all possible values of the random variable have the same probability.

PMF of discrete uniform distribution

$$\boldsymbol{f(x|N) = \frac{1}{N}}$$

CDF of discrete uniform distribution

$$\boldsymbol{F(x|N) = \frac{\lfloor x \rfloor}{N}}$$
$$\boldsymbol{x = 0, 1, 2, \ldots}$$

Mean and variance of discrete uniform distribution

$$\boldsymbol{E(X) = \frac{N+1}{2}}$$

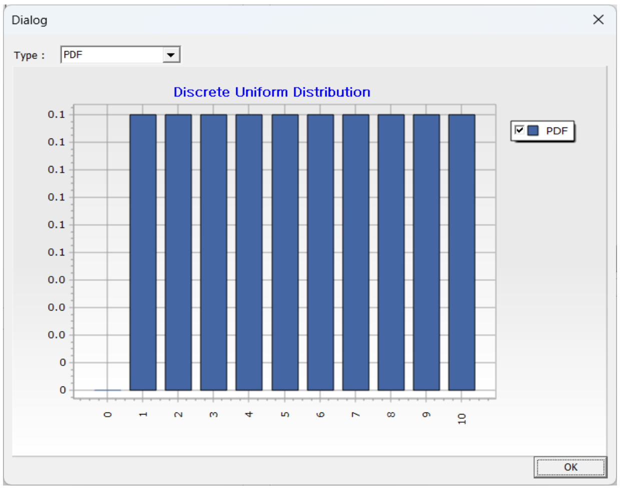
$$\boldsymbol{Var(X) = \frac{N^2 - 1}{12}}$$

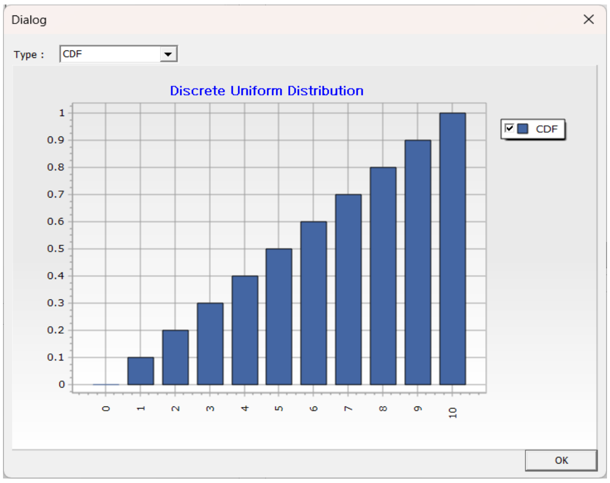
Inverse of cumulative distribution function of discrete uniform distribution

$$\boldsymbol{x = F^{-1}(p|N)}$$

Example

- PMF, CDF graph when $n = 10$



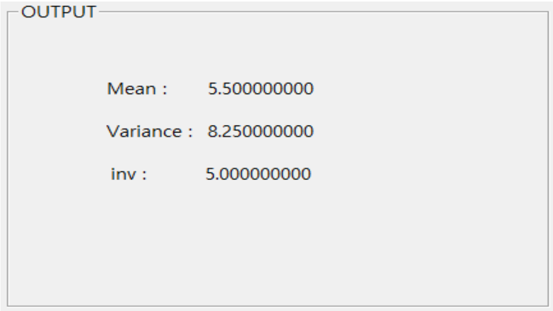


- Mean, variance, PMF, and CDF when $n = 10$ and $x = 1$.

OUTPUT

Mean :	5.500000000
Variance :	8.250000000
pmf :	0.100000000
cdf :	0.100000000

- Inverse of cumulative distribution function when $n = 10$ and $p = 0.5$



6.2.10 Continuous uniform distribution

Continuous uniform distribution is a type of continuous probability distribution where all values within a given interval have an equal probability of occurring. This distribution is used when any value within the interval is equally likely to be chosen at random.

PDF of continuous uniform distribution

$$f(x|a,b) = \frac{1}{b-a}$$
$$a \leq x \leq b$$

CDF of continuous uniform distribution

$$F(x|a,b) = \begin{cases} 0 & if\, x \leq a \\ \frac{x-a}{b-a} & if\, a < x < b \\ 1 & if\, x \geq b \end{cases}$$

Mean and variance of continuous uniform distribution

$$E(X) = \frac{a+b}{2}$$

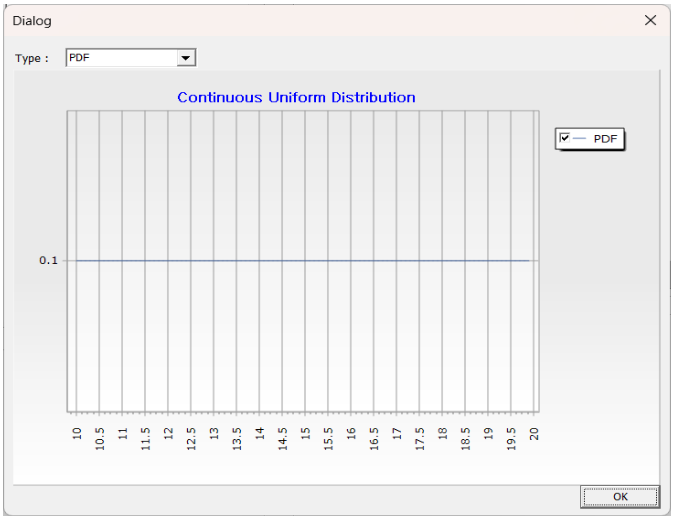
$$Var(X) = \frac{(b-a)^2}{12}$$

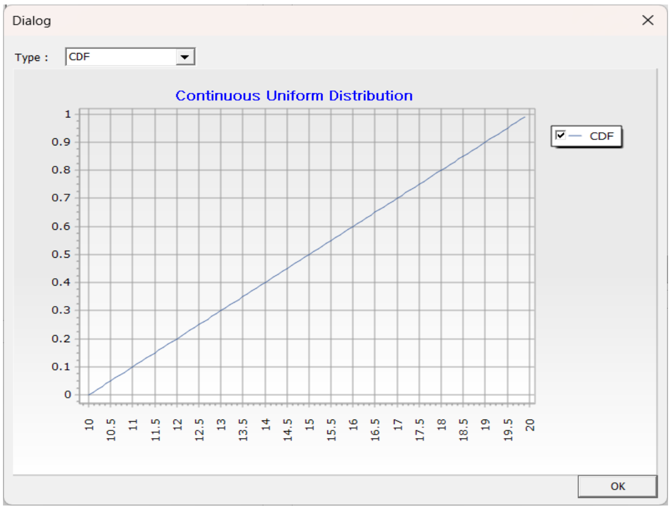
Inverse of cumulative distribution function of continuous uniform distribution

$$x = F^{-1}(p|a, b) = a + p(b - a)$$

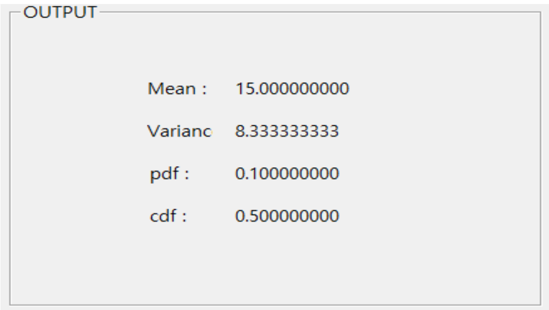
Example

- PDF, CDF graph when $a = 10$, $b = 20$ and $x = 15$.

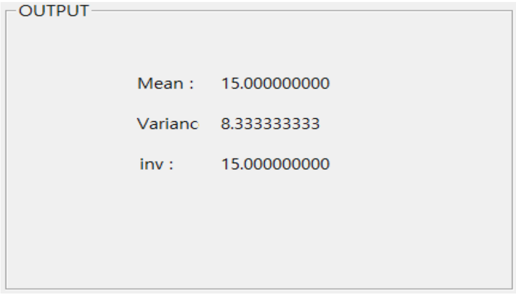




- Mean, variance, PDF, and CDF when $a = 10$, $b = 20$, and $x = 15$.



▪ Inverse of cumulative distribution function when $a = 10, b = 20$, and $p = 0.5$



6.2.11 Weibull distribution

The Weibull distribution is a continuous probability distribution defined by two parameters: shape, α and scale, β. It is primarily used in survival analysis, reliability engineering, and extreme value analysis.

PDF of Weibull distribution

$$\boldsymbol{f(x|\alpha,\beta) = \left(\frac{\beta}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^{\beta}}}$$
$$\boldsymbol{x \geq 0, \alpha > 0, \beta > 0}$$

CDF of Weibull distribution

$$\boldsymbol{F(x|\alpha,\beta) = 1 - e^{-\left(\frac{x}{\alpha}\right)^{\beta}}}$$
$$\boldsymbol{x \geq 0, \alpha > 0, \beta > 0}$$

Mean and variance of Weibull distribution

$$\boldsymbol{E(X) = \beta^{\frac{1}{\alpha}}\Gamma\left(1 + \frac{1}{\alpha}\right)}$$

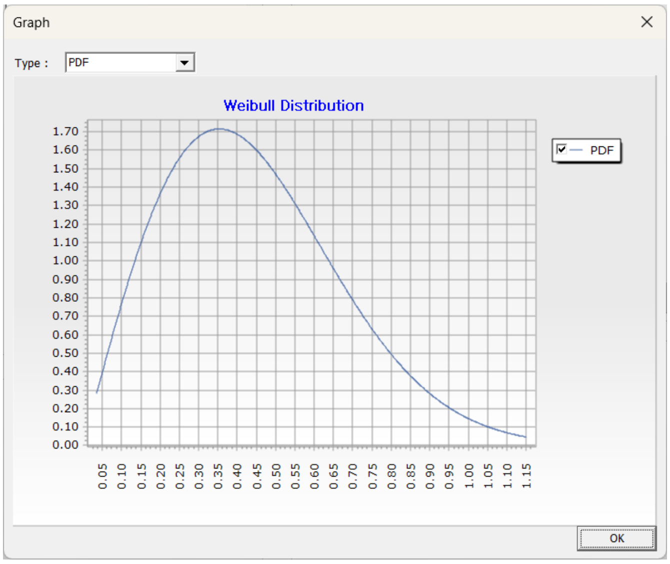
$$Var(X) = \beta_{\alpha}^{2}[\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^{2}\left(1+\frac{1}{\alpha}\right)]$$

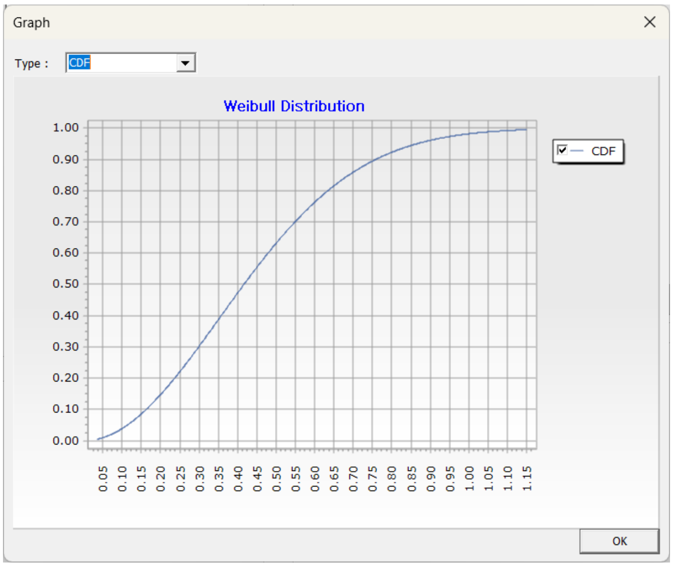
Inverse of cumulative distribution function of Weibull distribution

$$x = F^{-1}(p|\alpha,\beta) = \begin{cases} \infty & if\ p = 1 \\ \alpha(-\ln(1-p))^{1/\beta} & otherwise \end{cases}$$

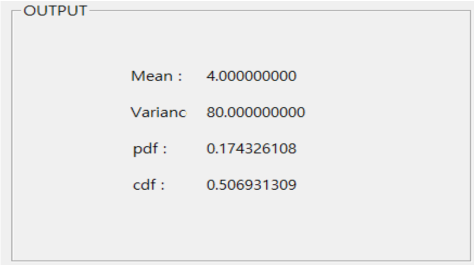
Example

- PDF, CDF graph when $\alpha = 2$, $\beta = 0.5$, and $x = 1$.

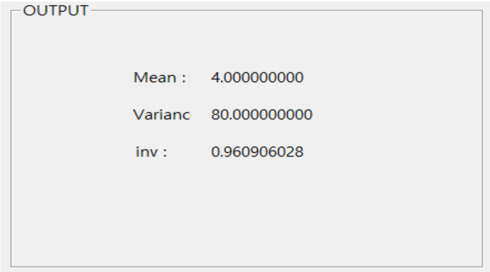




▪ Mean, variance, PDF, and CDF when $\alpha = 2$, $\beta = 0.5$, and $x = 1$.



▪ Inverse of cumulative distribution function when $\alpha = 2$, $\beta = 0.5$, and $, p = 0.5$



ECMiner™ allows for easy and convenient analytical model by connecting **Nodes** into a **Stream**. **The stream consists of nodes** with data input, preprocessing, modeling, and data output. ECMiner™ project file can be saved as an .ecm in the ECMiner™ Project window.