This paper explains the research conducted by Minitab statisticians to develop the methods and data checks used in the Assistant in Minitab Statistical Software.

Design of Experiments (DOE)

Overview

The Assistant DOE includes a subset of the DOE features available in core Minitab and uses a sequential experimentation process that simplifies the process of creating and analyzing designs. The process begins with screening designs to identify the most important factors. Then, we provide higher-resolution designs to look for curvature and determine a final model that can be used to identify factor settings that optimize the response.

In this paper, we outline the steps of the experimentation process. We provide information on how we determined which designs to offer in the Assistant, including the role of power. We also discuss the process of detecting and fitting curvature in the data. The paper also describes the method used for analyzing the data and identifying the best model.

The paper also provides additional information about the following data checks in the Assistant Report Card:

- Blocks
- Unusual data
- Detection Ability

Method

Sequential experimentation process

The DOE features in the Assistant guide users through a sequential process to design and analyze one or more experiments to identify the most important factors and find the factor settings that optimize a response. A sequential experimentation approach uses a set of smaller experiments where the results at each stage guide the experimentation at the next stage. An advantage of the sequential approach is that at each stage, only a small number of experimental trials are run so that you are less likely to waste resources on unproductive trials.

The Assistant provides a subset of the DOE features available in core Minitab in a structured format that simplifies the process of creating and analyzing designs. The steps in the process are:

1. Create a screening design for 6 to 15 factors.
2. Fit a screening model that includes the main effects and analyze the results to find the most important factors.
3. Create a modeling design based on the results of step 2 that includes the 2–5 most important factors.
4. Fit a linear model that includes main effects and 2-way interactions and analyze the results and look for evidence of curvature in the relationship between the factors and the response.
5. If curvature is not detected in step 4, use that model to identify factor settings that optimize the response.
6. If curvature is detected in step 4, the Assistant recommends that you add points for curvature to the design.
7. Fit a quadratic model that includes square terms to model the curvature and analyze the results.
8. Using the final model, identify factor settings that optimize the response.

The following sections provide more detailed information on these aspects of Assistant DOE:

- Screening designs
- Modeling designs
- Fitting the model

Screening designs

Typically, you begin the sequential experimentation process with a large number of potential factors and then you eliminate the ones with little effect on the response. Screening designs are experimental designs that are intended to identify the few most important factors from a larger set of factors. In the Assistant, screening designs are offered for 6 to 15 factors.

Type of design

The Assistant screening designs are Plackett-Burman designs, a special type of Resolution III 2-level designs. There are two primary advantages of Plackett-Burman designs:

- They allow for the estimation of the main effects of the factors using very few experimental runs (as few as 12). Because experimental runs can be expensive to perform, this makes these designs more cost-effective.
- There is only partial or fractional confounding between the main effects and two-factor interactions. Effects that cannot be estimated separately from one another are said to be confounded. In Plackett-Burman designs, the confounding is considered partial because the contribution of each effect is only a fraction of the full size of the interaction effect.

We determined that, for the purposes of screening, it was a reasonable approach to use the Plackett-Burman designs that estimate main effects only and do not estimate interaction terms. Screening designs are intended to include a large number of factors. Because at least one run is required for each term in the model, and the number of interaction terms increases faster than the number of main effects, for most situations it is not practical or cost-effective to fit a model with interactions. Additionally, in most cases, only a small number of factors explain most of the effects on the response. The goal of a screening design is to identify these factors, and Plackett-Burman designs enable users to identify these important main effects. Furthermore, as stated earlier, because the confounding between terms in Plackett-Burman designs is only partial, it is less likely that a significant main effect is in reality a significant 2-factor interaction.

Power and folding

When we created the design catalog, our goal was to only make available designs that had adequate power. We calculated the power for all the designs and eliminated certain designs due to low power, including the 12-run Plackett-Burman design for 10 or 11 factors. For designs with 10 or 11 factors, only the 20-run Plackett-Burman design is available. We also eliminated the designs for 16, 17, and 18 factors because of low power and the higher number of runs. For more information on the specific power for the designs, see the section on Detection ability.

For designs with 6 to 9 factors, we allow folding, which adds runs to the experiment, increasing the precision and power of the design. In some cases, it may be desirable to add runs to a design to increase the likelihood of detecting important effects. With folding, new runs are added to the design in which some or all the factor levels are reversed by switching low and high levels of the factors. Folding also eliminates the partial confounding between main effects and two-factor interactions, which reduces the bias of the main effect estimates due to confounding. The Detection Ability section in the Create Screening Design Summary Report provides information to help users determine whether the design has enough power to detect effects of adequate size.

Modeling designs

Once 2 to 5 important factors are identified, Minitab recommends creating a modeling design to obtain a model that can be used to identify factor settings that optimize the response.

Type of design

The modeling designs for 2 to 5 factors are all full-factorial or resolution V designs. These designs can be used to fit all main effect and 2-factor interaction terms without any confounding between the terms. Some or all of the higher-order terms (e.g., 3-factor interactions) may be confounded with the terms in the model. However, higher-order terms can often be assumed to be negligible compared to main effect and 2-factor interaction terms.

When we created the design catalog, our goal was to only make available designs that have adequate power. As a result, we eliminated the 2-factor design with 4 runs and instead we use a replicated 4-run design for 2 factors.

Center points and modeling curvature

The modeling designs in Assistant also include center points to check for the presence of curvature in the data. These are points where all the continuous factors are set midway between the low and high settings. If there is no curvature, then the mean response at the center point equals the average of the mean response of the factors at their low and high settings (the corners of the design space). Curvature is detected when the average mean response at the center points is significantly greater or less than the average mean response of the factors at their low and high settings.

Although center points can detect curvature, they don’t provide enough information to model the curvature. To model the curvature, square terms are needed, which requires adding more points to the design. These additional points convert the design to a face-centered central composite design. This is a form of response surface design, which makes it possible to fit a quadratic model that has linear main effects, all 2-factor interactions, and square terms of all continuous factors.

Fitting models using backward selection

We explored several methods of fitting the models and determined that backward selection using an α of 0.10 was the best approach. When you fit a model, Minitab starts by including all possible terms. Then, one by one, Minitab removes the least significant term, while maintaining the hierarchy of the model. Hierarchy means that, if an interaction term is significant, then the linear terms of both factors that form the interaction must also be included in the model. This is a form of backward selection and is intended to automate a process of model selection that is typically done by hand. In all the designs in Assistant DOE, the terms are independent or (in the case of square terms) nearly so. Therefore, multicollinearity, which indicates factors are correlated with one another, is not likely to occur. Multicollinearity can cause stepwise procedures to miss the best model. Using $\alpha = 0.10$ rather than the commonly used $\alpha = 0.05$ helps to increase the power of the tests, which increases the likelihood that important terms remain in the model.

Data Checks

Blocks

Blocks are used in experimental designs to minimize bias and error due to external influences on the experiment, such as noise variables, missing variables, or differences in how the runs in each block were performed. By placing experimental runs performed together in blocks, you can determine whether differences exist between blocks and account for these differences in the model.

Objective

In Assistant DOE, you can include replicates of the design when creating a modeling design and can add axial points to a modeling design to fit curvature in the model. Often, replicates and axial points are performed at different times or under different conditions than runs in the base design. When runs are performed at different times or conditions, it is best practice to account for possible effects due to different conditions.

Method

To account for possible differences in experimental conditions for replicates or axial points and the base design, Minitab places replicates and axial points in separate blocks. Specifically, in modeling designs, Minitab places replicates of the base design in separate blocks in the model. In quadratic designs, Minitab places the axial points used to detect curvature in the design in a separate block.

Results

To be consistent with the treatment of other terms in the model, blocks are evaluated using the backward elimination method. The report card states whether the block term is statistically significant, indicating that there are differences across the blocks. If a difference exists between blocks, consider investigating the cause to determine if there were any inconsistencies in the experimental conditions or procedures.

Status	Condition
	Blocks are in the final model Blocks are significant. Because the runs in each block are typically performed at different times, the significant difference in blocks indicates that conditions may have changed over time. This difference could be due to external influences on the experiment, such as noise variables, missing variables that should have been included in the experiment, or differences in how the runs in each block were performed. Consider investigating the cause of the difference between the blocks. **Blocks are not in the final model** Blocks are not significant. Because the runs in each block are typically performed at different times, this result indicates that there is no evidence of differences in the experimental conditions over time.

Unusual Data

In the Assistant DOE procedures, we define unusual data as observations with large standardized residuals, a common measure used to identify unusual data in model-fitting procedures (Neter et al., 1996). Because unusual data can have a strong influence on the results, you may need to correct the data to make the analysis valid.

Objective

We wanted to determine how large the standardized residuals need to be to signal that a data point is unusual.

Method

We developed our guidelines for identifying unusual observations based on the standard DOE procedure in Minitab (**Stat** > **DOE** > **Factorial** > **Analyze Factorial Design and Stat** > **DOE** > **Response Surface** > **Analyze Response Surface Design**).

Results

The standardized residual equals the value of a residual, e_i, divided by an estimate of its standard deviation. In general, an observation is considered unusual if the absolute value of the standardized residual is greater than 2. However, this guideline is somewhat conservative. You would expect approximately 5% of all observations to meet this criterion by chance with large data sets (if the errors are normally distributed). However, with small experimental data sets, few if any observations will be flagged by chance, and it is good practice to investigate the cause of unusual values.

When checking for unusual data, the Assistant Report Card displays the following status indicators:

Status	Condition
	**No standardized residuals ≥ 2** There are no unusual data points. Unusual data can have a strong influence on the results.
	**One standardized residual ≥ 2** One data point has a large residual and is not well fit by the model. This point is marked in red on the Diagnostic Report and is in row X of the worksheet. Because unusual data can have a strong influence on the results, try to identify the cause for its unusual nature. Correct any data entry or measurement errors. Consider performing trials associated with special causes again and redoing the analysis. **More than one standardized residual ≥ 2** x data points have large residuals and are not well fit by the model. These points are marked in red on the Diagnostic Report. You can hover over a point or use Minitab's brushing feature to identify the worksheet rows. Because unusual data can have a strong influence on the results, try to identify the cause for their unusual nature. Correct any data entry or measurement errors. Consider performing trials associated with special causes again and redoing the analysis.

Detection ability

When performing designed experiments, it is useful to know what effect size a design is likely to detect prior to collecting data. If the design isn't powerful enough to detect the desired effect size, it may be appropriate to include more runs in the design. However, because including more runs in a design can be expensive, it is important to determine whether the additional power is necessary.

Objective

We wanted to provide users with information about the effect size their design can detect at 60% and 80% power levels. We also wanted to provide users with information about the effect sizes for design that include additional runs when available. For screening designs with 6 to 9 factors, users can choose to include 12 or 24 runs in their design. For modeling designs, users can include replicates of the base design, increasing the total number of runs in the design.

Method

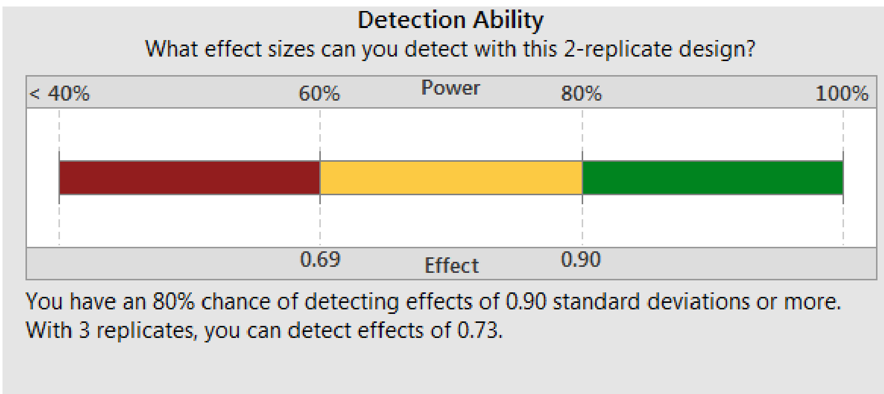
We computed the power and the effect size that can be detected for each design in the Assistant. The power is the probability of finding a factor effect to be statistically significant. The effect sizes are in standard deviation units.

Results

The Summary report displays the effect sizes you can detect with your design with 60% power and 80% power. For screening designs where a larger (folded) design is available, the report also specifies what size effect can be detected with 80% power in the larger design. For modeling designs where more replicates are available, the report specifies what effect size can be detected with 80% power with additional replication. Then, the user can judge whether the selected design is appropriate and weigh the advantages of using a design with more runs when available.

See Appendix A for specific information about the effect sizes each design can detect at 60% power and 80% power.

The following image is an example of the power information provided in the Summary report.



References

Neter, J., Kutner, M.H., Nachtsheim, C.J., & Wasserman, W. (1996). *Applied linear statistical models*. Chicago: Irwin.

Appendix A: Detection ability

We computed the power and the effect size that can be detected for each design in the Assistant. The power is the probability of finding a factor effect to be statistically significant. The effect sizes are in standard deviation units.

The effect size associated with a model term is twice the term's coefficient in the true model equation. In a screening model, the effect size has a simple interpretation: the change in the mean response when a factor changes from its low level to its high level.

Table 1 The following table shows the effect sizes for the Screening designs available in the Assistant.

Factors	Number of runs	Effect with power 60%	Effect with power 80%
6	12	1.27325	1.67693
6	24	0.80721	1.05805
7	12	1.32820	1.75498
7	24	0.80936	1.06092
8	12	1.43101	1.90493
8	24	0.81180	1.06420
9	12	1.68682	2.29728
9	24	0.81462	1.06797
10	20	0.919135	1.20607
11	20	0.928949	1.21945
12	20	0.941923	1.23725
13	20	0.959863	1.26206
14	20	0.986258	1.29895
15	20	1.02882	1.35940

Table 2 The following table shows the effect sizes for the Modeling designs available in the Assistant.

Total Factors	Categorical Factors	Replicates	Effect with power 60%	Effect with power 80%
2	0	2	1.517676	1.995488
2	0	4	0.991255	1.299356
2	0	6	0.795584	1.042572
2	1	2	1.489312	1.956272
2	1	4	0.986308	1.292747
2	1	6	0.793360	1.039623
2	2	2	1.626711	2.149402
2	2	4	1.012132	1.327420
2	2	6	0.805317	1.055524
3	0	1	1.752624	2.333048
3	0	2	1.001723	1.313394
3	0	3	0.798410	1.046325
3	0	4	0.685061	0.897680
3	0	5	0.609738	0.798946
3	1	1	1.626711	2.149402
3	1	2	0.994252	1.303368
3	1	3	0.795584	1.042572
3	1	4	0.683497	0.895612
3	1	5	0.608716	0.797597
3	2	1	1.468798	1.928128
3	2	2	0.977848	1.281481
3	2	3	0.788844	1.033647
3	2	4	0.679641	0.890522
3	2	5	0.606149	0.794214

Total Factors	Categorical Factors	Replicates	Effect with power 60%	Effect with power 80%
3	3	1	3.804252	5.792800
3	3	2	1.038597	1.363392
3	3	3	0.811803	1.064195
3	3	4	0.692413	0.907434
3	3	5	0.614534	0.805288
4	0	1	1.053102	1.383293
4	0	2	0.689744	0.903887
4	0	3	0.556612	0.729334
4	0	4	0.479760	0.628615
4	0	5	0.428010	0.560802
4	1	1	1.038597	1.363392
4	1	2	0.688304	0.901977
4	1	3	0.556027	0.728562
4	1	4	0.479427	0.628176
4	1	5	0.427789	0.560511
4	2	1	1.006462	1.319772
4	2	2	0.684233	0.896585
4	2	3	0.554302	0.726288
4	2	4	0.478427	0.626861
4	2	5	0.427119	0.559631
4	3	1	0.982394	1.287529
4	3	2	0.679988	0.890980
4	3	3	0.552383	0.723762
4	3	4	0.477284	0.625358
4	3	5	0.426341	0.558609

Total Factors	Categorical Factors	Replicates	Effect with power 60%	Effect with power 80%
4	4	1	1.102670	1.452267
4	4	2	0.694658	0.910421
4	4	3	0.558674	0.732059
4	4	4	0.480955	0.630190
4	4	5	0.428812	0.561858
5	0	1	1.460831	1.989497
5	0	2	0.694658	0.910421
5	0	3	0.557797	0.730899
5	0	4	0.480244	0.629252
5	0	5	0.428261	0.561133
5	1	1	1.239292	1.649714
5	1	2	0.692413	0.907434
5	1	3	0.557051	0.729913
5	1	4	0.479850	0.628733
5	1	5	0.428010	0.560802
5	2	1	1.053102	1.383293
5	2	2	0.686516	0.899606
5	2	3	0.554925	0.727108
5	2	4	0.478694	0.627212
5	2	5	0.427261	0.559817
5	3	1	0.994252	1.303368
5	3	2	0.680992	0.892303
5	3	3	0.552683	0.724156
5	3	4	0.477418	0.625533
5	3	5	0.426414	0.558704

Total Factors	Categorical Factors	Replicates	Effect with power 60%	Effect with power 80%
5	4	1	0.970149	1.271267
5	4	2	0.676819	0.886805
5	4	3	0.550801	0.721681
5	4	4	0.476297	0.624062
5	4	5	0.425652	0.557704
5	5	2	0.703042	0.921620
5	5	3	0.560538	0.734525
5	5	4	0.481695	0.631166
5	5	5	0.429191	0.562356