Pharmaceutical Quality by Design Using JMP®

Solving Product Development and Manufacturing Problems

Rob Lievense

The correct bibliographic citation for this manual is as follows: Lievense, Rob. 2018. *Pharmaceutical Quality by Design Using JMP®: Solving Product Development and Manufacturing Problems*. Cary, NC: SAS Institute Inc.

Pharmaceutical Quality by Design Using JMP®: Solving Product Development and Manufacturing Problems

978-1-62960-864-8 (Hardcopy)
978-1-63526-620-7 (Web PDF)
978-1-63526-618-4 (epub)
978-1-63526-619-1 (mobi)

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

September 2018

Contents

About This Book

What Does This Book Cover?

Regulatory agencies for the pharmaceutical and medical device industries have released several guidelines to promote the use of elements of Quality by Design (QbD). Technical professionals have great interest in QbD, but many are unsure of where to start. This book is a guide for using data visualization and statistical analyses as elements of QbD to solve problems and support improvement throughout the product life cycle.

The book includes three areas of general focus for the topics contained. The first several chapters focus on the type of data that is available for current commercial production of healthcare products. The book then focuses on the tools and techniques that are useful for product and process development. The final chapters are more specialized and deal with utilizing data visualization to solve complex problems, as well as special topics that are unique to healthcare products.

In chapters 1 through 5, technical professionals learn how to use JMP to obtain visualizations of their data by using the Distribution platform and the Graph Builder. The powerful, dynamic nature of the data visualizations is highlighted so that readers can easily extract meaningful information quickly. Techniques for including a time element for effective visualization and identification of trends is covered as well. Methods for comparing trends in the data to specification limits are covered, enabling you to diagnose the performance of a process and effectively communicate the findings to the stakeholders of an improvement project. The stream of topics moves on to the utilization of data from a random sample to make precise estimates (via statistical inference) on an entire population of units produced. Statistical inference is expanded to analyze for relationships and differences between two variables, utilizing the rich set of techniques available in the Fit Y by X platform.

Chapters 6 through 12 begin with applications that help the reader justify why structured, multivariate, experimental designs must be used to develop robust products and processes. Comparing designs created through the Design of Experiments (DOE) platform to the typical approach that uses one factor at a time (OFAT) clearly shows the advantages of structured, multivariate, experimental designs, especially in QbD era. Examples focus on effective techniques for analyzing measurement systems and quantifying how measurement variability may affect analysis results. Various modeling techniques are covered so that you know how to utilize available historical data to use resources efficiently in experimental designs. The DOE platform is extensively utilized to teach you how to create effective experimental designs for both materials and processes. The section is rounded out with analysis techniques for completed experiments as well as simulation tools that you can use to include known process variation and simulate likely results. Simulation can save a development team time, money, and increased credibility due to the potential to mitigate future mistakes.

The context in chapters 13 through 17 expand on the predictive modeling techniques presented in section two by including predictive models that can detect inputs that have subtle influences on outputs. Basic mixture designs are covered to help you effectively deal with three-component proportional mixes of materials. Many aspects of pharmaceutical products show trends in outputs that include rates of change as a function and that cannot be studied with typical linear modeling. Examples of non-linear modeling help you gain understanding about such applications. Analyses of measurement systems from the second section is expanded on with an example of how you can use a structured, multivariate, experimental design to support analytical method development. The section wraps up with the specialized topic of stability analysis vis a tool provided in the Degradation platform. The stability analysis techniques follow

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) guidelines regarding how to identify the likely shelf life of products. Using these techniques to dig deeper into the modeling details provides insight that is unparalleled.

The book does not offer a deep, theoretical understanding of the concepts or detail about the computational methods used by JMP to create the output. There are several references to the Help menu in JMP throughout the chapters so that you can find this detail if you are interested.

Is This Book for You?

I have read many instructional texts for data visualization and statistics. Most begin with the identification and discussion of a statistical topic or technique, followed by examples intended for readers to use to add practical ability. The typical flow of such textbooks creates barriers to technical professionals who want to efficiently apply the knowledge to solve problems involving data. They are often under time pressures and struggle most with trying to find the statistical technique that will work to extract the information they need from data. This book is written from a technical professional point of view to match the flow of work that occurs in the real world of the pharmaceutical and medical device industries. Each chapter involves a technical professional facing a problem that could benefit from the use of JMP.

Each chapter describes the problem at hand, followed by hands-on work in JMP. Examples include relevant screen shots of the JMP interface, along with figures, notes, and explanations of results. The data sets are based on actual problems in an attempt to make the examples reflect the real world. Many of the problems involve data preparation steps and table manipulation before analysis can be done, which is another issue that technical professionals encounter in the real world. Chapters culminate with practical conclusions that help the reader summarize the key points of the analysis. Most chapters include exercises for additional hands-on practice.

Scientists, engineers, and technicians involved throughout the pharmaceutical and medical device product lifecycles will find this book useful. The reliance upon principal science and professional experience for product development can combine to yield a batch that passes requirements. The use of JMP to apply data visualization and statistical modeling will create a product that robustly meets requirements for the entire life cycle. The trends in the inputs and outputs of processes are easy to explore from the creation of simple graphs to model analysis with simulations used to estimate the defect rate of a future product. The analysis completed in JMP provides a great foundation for regulatory submissions of products and processes. Submissions supported with robust statistics tend to have fewer deficiencies. Regulatory deficiencies that occur can be better answered with data visualizations and statistics, which tend to also increase the speed of product approvals.

JMP includes the versatility to be used to solve problems throughout the life cycle of a product. Quality control can monitor and assess processes through the use of control charting and capability studies. Filling processes can be optimized through the dynamic function of the distribution platform as well as predictive modeling, Stability studies are easy to create in JMP and offer the insight needed to predict the expected shelf life for multiple packaging configurations. Physical features of medical devices can be studied and optimized to ensure that variation in products is mitigated and customers are likely to enjoy consistency in the use of a product. The measurement systems used to quantify a physical or chemical attribute can be studied using JMP to ensure the highest levels of accuracy and precision in data obtained.

Products developed through the use of JMP DOE tools can reach the market in half the time required for development using principal science and experience alone. The resources required to get a product to

market are greatly reduced as models are utilized to find the optimum input settings to meet all product requirements simultaneously. Fewer developmental batches need to be run and the potential for making costly mistakes is greatly reduced. This book offers more than instruction on the use of JMP; it is also a guide for saving time and money.

What Are the Prerequisites for This Book?

This book makes a few assumptions about its readers. It is assumed that you possess a general understanding of the relevant scientific and technical concepts for the pharmaceutical or medical device industries. By following the examples, you will be able to fill in any details that you are not already familiar with. Some initial familiarity with JMP is helpful. You can use the JMP website to become familiar with JMP: https://www.jmp.com.

What Should You Know about the Examples?

This book includes relevant examples from the target industries for you to follow in order to gain hands-on experience with JMP.

Software Used to Develop the Book's Content

The book uses JMP 14.0 for the majority of content and JMP Pro 14.0 for a few high-level concepts. The screen shots used to demonstrate navigating the JMP menus are captured using JMP Pro, and most have the same look as what is seen with JMP. Other versions of JMP might not have the same options or have slightly different menu options.

Example Code and Data

It is intended that you work on the examples as you read through each chapter. The exercises at the end of most chapters provide an extension of this work by either expanding on the chapter examples or by using new data sets with similar problems. A set of additional materials including the data sets used throughout the book is available for download. You can access the example code and data for this book by visiting the author page at https://support.sas.com/lievense.

Where Are the Exercise Solutions?

A full set of solutions for the end-of-chapter exercises is available on the author page at https://support.sas.com/lievense.

About the Author

Rob Lievense is a Research Fellow of Global Statistics at Perrigo, as well as an active professor of statistics at Grand Valley State University (GVSU), located in Allendale, Michigan. At Perrigo, he leads a group that supports the consumer health care research and development department with statistical analysis, data visualization, advanced modeling, data-driven Quality by Design for product development, and structured experimental design planning. Rob has more than 20 years of experience in the applied statistics industry and 10 years of experience in the use of JMP. He has presented at major conferences including JMP Discovery Summit, where he served on the Steering Committee in 2017, and the annual conference of the American Association of Pharmaceutical Scientists. Rob has a BS in Applied Statistics and an MS in Biostatistics from GVSU. He currently serves as a member of the Biostatistics Curriculum Development Committee for GVSU and has his Six Sigma Black Belt Certification..

Learn more about this author by visiting his author page at support.sas.com/lievense. There you can download free book excerpts, access example code and data, read the latest reviews, get updates, and more.

We Want to Hear from You

SAS Press books are written *by* SAS Users *for* SAS Users. We welcome your participation in their development and your feedback on SAS Press books that you are using. Please visit sas.com/books to do the following:

- Sign up to review a book
- Recommend a topic
- Request information on how to become a SAS Press author
- Provide feedback on a book

Do you have questions about a SAS Press book that you are reading? Contact the author through saspress@sas.com or https://support.sas.com/author_feedback.

SAS has many resources to help you find answers and expand your knowledge. If you need additional help, see our list of resources: sas.com/books.

Chapter 11: Analysis of Experimental Results

Overview

The designs of chapter 10 were formed through detailed collaboration and communication with project stakeholders and subject matter experts. The designs developed in JMP were carefully executed in compliance with an experimental protocol. The protocol was clearly documented, approved by leadership, and supported by resources provided by the project team and the operations group conducting the granulations. The randomization plans were followed in the exact order noted in the data collection plans. The output values entered in the data sheet were confirmed to be precise and accurate. This chapter explains how to use JMP to efficiently analyze the experimental models to provide evidence in support of the goals for each project. You will see the comparisons between the information provided by a simple screening design and a high-resolution response surface model with multiple outputs.

The Problems: A Thermoforming Process and a Granulation Process, Each in Need of Improvement

Michelyne's team is charged with quickly identifying the process inputs that influence the thinning of the material thickness of surgical handle covers so that customer complaints can be corrected as soon as possible. The project stakeholders prefer speed and decisive action over the obtainment of detailed knowledge of the thermoforming process. The definitive screening design (DSD) used for the experiments met the goal of the stakeholders and even provides a glimpse into the amount that various factors affect the handle thickness.

The granulation process being studied by Emily's team is more mature than the thermoforming process being studied by Michelyne's team. The project stakeholders have run enough prior experiments to identify

the process inputs that have a significant effect on outputs. The problem involves multiple outputs that must be met robustly through control of four process inputs. The goal of the team is to quantify the effects of changing inputs so that the optimum settings can be identified. This information is crucial for the manufacturing team to have the best chance of robustly obtaining good product. The response surface model designed for the experiments provides a high resolution of information that will exceed the expectations of the team.

Execution of Experimental Designs

The team charged with addressing the incidence of cracked and loose surgical light handle covers created an experimental plan to screen for significant thermoforming process inputs in chapter 10. Michelyne found that the definitive screening design, exclusive to JMP, provides an extremely efficient set of experiments to screen to a limited number of process inputs. The main advantage of the DSD is a model that mitigates the typical correlation that is found between the individual inputs and the interactions. The small number of runs was very appealing to the project stakeholders because results can be obtained very quickly and with controlled expense. Additional advantages of the DSD is the detection of interactions and squared (non-linear) terms that might leverage the output.

The design of experiments (DOE) explained in chapter 10 is only a small part of the effort required to execute a set of multivariate structured experiments. The statistical analysis of an experimental model assumes that most of the variation in the output is due to changes in the inputs studied. All other potential sources of variation to the output must be discussed and appropriately controlled. The inclusion of controls to the process is easier said than done in most cases. The design team must include subject matter experts from the operations area to discuss the needed process controls before the experiments are executed. It is extremely important for the team to go into such meetings with healthy pessimism to ensure that all sources of potential variability are identified. These sources can include, but are not limited to, variations in the physical environment (temperature, humidity, air circulation), changes in people (operators, technicians, leadership), changes in the machines (warm up, continued wear and tear, automatic controls adjustments), and changes in materials (amount of incoming material, changes in physical characteristics, changes in lots).

Each potential influence should include a control as well as a team member responsible for ensuring that the control is in place as experimental runs are executed. There are some sources of variability that the team will decide to not control in order to ensure that the set of experimental runs adequately represent the population of process results being modeled. For instance, the operational facility might experience fluctuations in ambient temperature that range between 68 degrees F to 82 degrees F during typical processing. If the experimental environment were to be controlled to a rigid control of 70 degree F +/- 1 degree, the resulting model might not contain the amount of random error expected in ambient conditions and could provide misleading results. Operations would need to invest a significant expense to mimic a 70-degree F +/- 1-degree environment, which is highly unlikely.

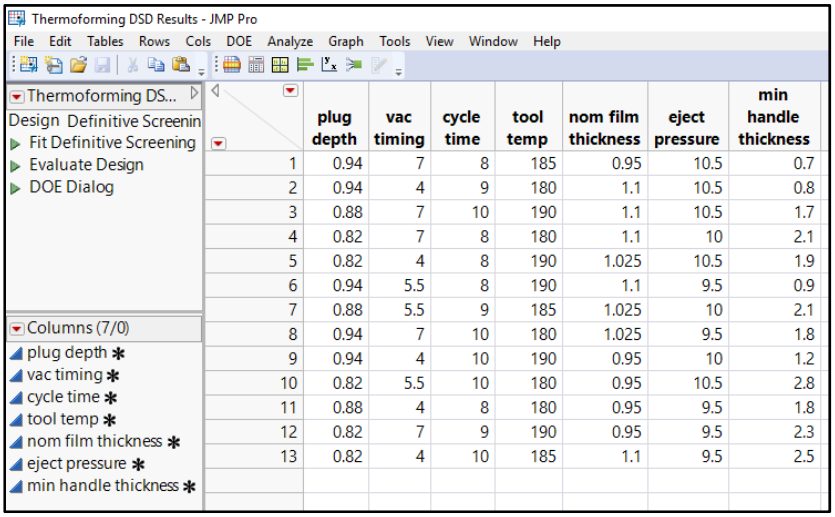
There is a very high likelihood that the execution of a set of experimental runs will not go perfectly. An experimental model can extract the highest-value information when levels chosen for the design are pushed out as far as possible to the extreme low and high ends of the operational range. With wide process levels, a risk exists for the inability of the process to create a sufficient output with the most extreme combinations of input levels. This can be especially problematic if the process is unable to create an output several runs into the design plan. It is highly recommended that the design team discuss the plan with subject matter experts (SMEs) in order to identify any runs that are at high risk for an inability to create an output. For

instance, a surgical handle might not adequately form with the shallowest plug depth (0.82), shortest vacuum time (4), and coolest tool temperature (180). The DSD plan from chapter 10 does not included a run with the noted extremes. However, one of the runs is very close. It is good practice to make the highest-risk run the first run executed in the plan. This violates the concept of a completely randomized design slightly. If the process is unable to produce a viable output, the design team can quickly adjust the levels in the design and add a run to replace the failed first run. Front-loading the highest-risk run can help maintain the intended number of runs. The movement of one run does not interfere with the assumption of randomness of runs enough to be practically relevant in most cases. If a failed processing attempt occurs late in the design plan, the team will need to exclude the run or scrap the effort and develop a new plan. Either option is very disruptive and costly, not to mention the loss of credibility that can occur with stakeholders of the project.

Analysis of a Screening Design

Michelyne assigned a member of the team to upload the outputs collected from the set of experiments to the design plan that they created in chapter 10. There are great advantages to using the original design plan created in JMP; scripts are created to make the job of analysis much easier for the analyst. Open *Thermoforming DSD Results.jmp* to view the data table shown in Figure 11.1.

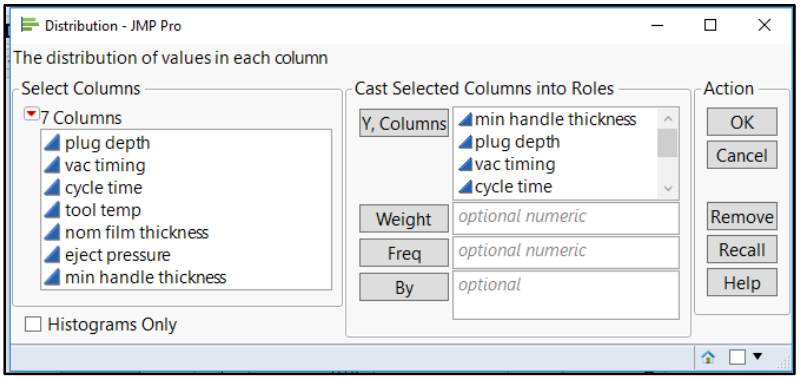
Figure 11.1: Thermoforming Experiments Results



	plug depth	vac timing	cycle time	tool temp	nom film thickness	eject pressure	min handle thickness
1	0.94	7	8	185	0.95	10.5	0.7
2	0.94	4	9	180	1.1	10.5	0.8
3	0.88	7	10	190	1.1	10.5	1.7
4	0.82	7	8	180	1.1	10	2.1
5	0.82	4	8	190	1.025	10.5	1.9
6	0.94	5.5	8	190	1.1	9.5	0.9
7	0.88	5.5	9	185	1.025	10	2.1
8	0.94	7	10	180	1.025	9.5	1.8
9	0.94	4	10	190	0.95	10	1.2
10	0.82	5.5	10	180	0.95	10.5	2.8
11	0.88	4	8	180	0.95	9.5	1.8
12	0.82	7	9	190	0.95	9.5	2.3
13	0.82	4	10	185	1.1	9.5	2.5

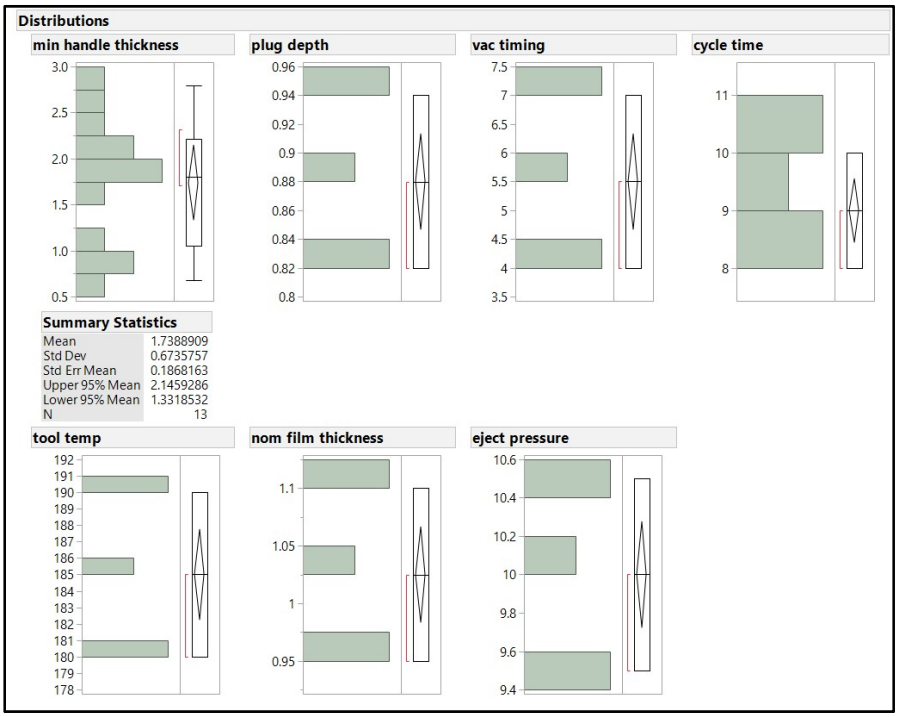
The fit definitive screening script in Figure 11.1 will be used to execute the analysis. However, it is good practice to visualize the data at a high level before completing a detailed model analysis. Select *Analyze* ▶ *Distributions* and move the variables to the *Y,Columns* box with the response on top and the six inputs underneath, as shown in Figure 11.2. Click *OK* to get the output.

Figure 11.2: Distribution Setup Window



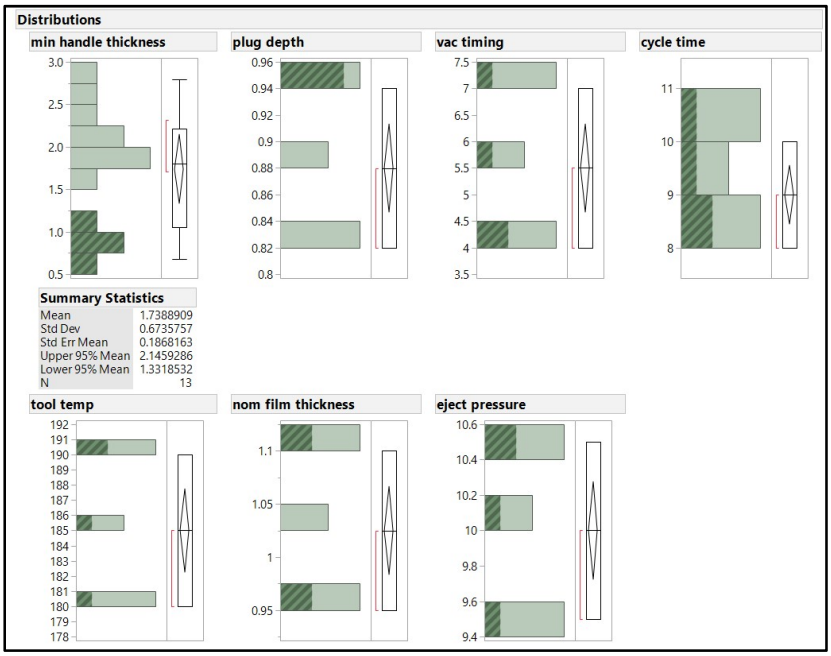
The output for the seven variables takes up a large amount of space and includes information that is not important. Press the Ctrl key and click the gray arrow next to the *Quantiles* header underneath *min handle thickness* to hide the quantiles summary for each variable. Press the Ctrl key and click the gray arrow next to the *Summary Statistics* header underneath *min handle thickness* to hide the summary statistics for each variable. Click on the gray arrow next to *Summary Statistics* underneath *min handle thickness* to unhide *Summary Statistics* for only the output. The summary statistics of the inputs are irrelevant since fixed levels were chosen when the model was designed. Lastly, use the red triangle menu next to *Distributions* to select *Arrange in Rows*, and enter 4 to display the output in two rows, as shown in Figure 11.3.

Figure 11.3: Distribution Output



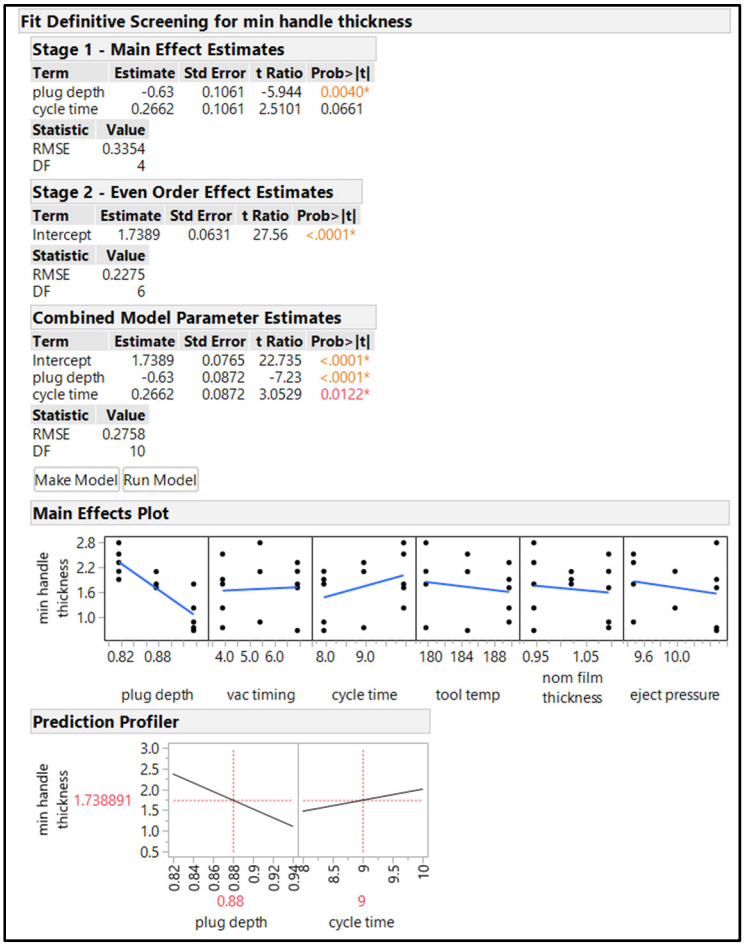
The average minimum handle thickness for the set of experiments is 1.74 mm and the standard deviation is 0.67 mm. The reasonable expectation for average minimal thickness for a population of handles made with the ranges of input changes is noted by the 95% confidence interval of 1.33 mm to 2.15 mm. The three levels explored for each process input are obvious since each distribution plot includes three bars. The dynamic functionality of distributions is used to determine whether any relationships are obvious. Notice that the *min handle thickness* plot seems to illustrate two groups of results. Press the Shift key and click on the lower group bars of *min handle thickness* to get the output shown in Figure 11.4.

Figure 11.4: Distribution Output with Dynamic Selection



The lower group of *min handle thickness* results seem to relate to the high level of plug depth since the majority of the high bar is highlighted. Low cycle time has a larger portion of the bar highlighted, but is not as convincing a pattern as plug depth. The four other variables do not illustrate as much of a pattern. Try other dynamic selections to look for potential patterns in the model data. Patterns in distribution plots offer a high-level view of possible relationships. However, the analysis of the model includes detail needed to quantify the level of significance for process inputs that influence *min handle thickness*. Go back to the *Thermoform DSD Results.jmp* table shown in Figure 11.1, and click the green arrow beside the *Fit Definitive Screening* script to get the output shown in Figure 11.5.

Figure 11.6: Thermoforming DSD Initial Analyses



The analysis relies on a unique model selection algorithm, which detects for significant inputs in two stages (1-main effects, 2-interactions). The *Stage 1 – Main Effects Estimates* in the *Fit Definitive Screening for min handle thickness* window provides information about the inputs that have significant evidence of influence. The root mean square error of the model is 0.3334, which is relatively high compared to the overall range of thickness results of 0.7 mm to 2.8 mm. The 4 degrees of freedom relates to the number of inputs that were not selected as a main effect for a simple DSD without replicate center points or a blocking term.

The plug depth is highly significant (p=.004) in reducing the minimum thickness output. The estimate of the effect (-0.63) means that a 0.63 mm average reduction in thickness for every percent increase in plug depth can be expected. This estimate of change in minimum thickness due to plug depth is only 4/10ths of a percent likely to have come from random variation. The cycle time has a minimally significant evidence of influence on minimum thickness (p=0.066). This means that the estimate of a 0.27 mm average increase in thickness for every second of increase in cycle time identified by the model is 6.6% likely to have come from random variation. The DSD algorithm utilizes a p-value threshold for selecting main effects that is based on the degrees of freedom in the error term of the model. This is why it is possible for selected main

effects to have a p-value higher than the 0.05 level of significance. (You can find additional technical details about selecting main effects for a DSD in the Design of Experiments Guide, available in the Help menu.). None of the remaining four inputs seem to have significant influence on minimum thickness.

The second stage is used to detect for significant interactions. The minimum resource DSD with no replicate runs, no blocks, and no extra center runs does not offer enough runs to gain the highest level of robustness in sensitivity to higher order factors. The results indicate a significant intercept, which means only that the response value of linear model is not zero when the explanatory value is at zero. There is no useful practical interpretation of the intercept value. The lack of any other terms in the Stage 2 analysis indicates that the influence on minimum thickness is limited to main effects.

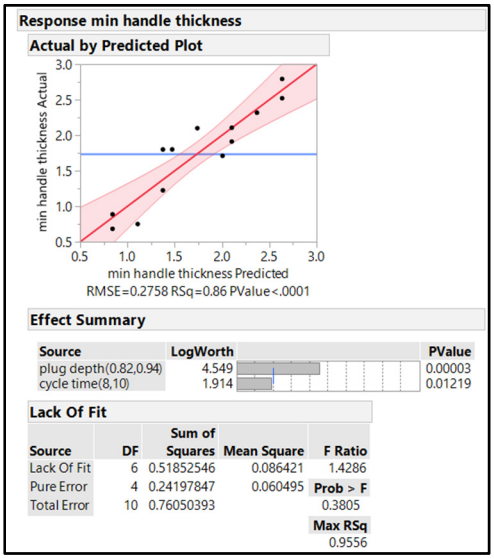
Combined model parameter estimates are derived from the typical least squares analysis of the model. With the 10 degrees of freedom used for model selection, the significance of the plug depth and cycle time is stronger than the results of the specialized DSD main effects modeling. The slope of the blue model lines in the main effects plots of plug depth and cycle time indicate evidence of significant influence on *min handle thickness* because they are the two with the steepest slopes. In addition, the small distance between the observation points and the plug depth effect line also emphasize the strength of the significant relationship. The observations located about the cycle time effect line illustrate a similar pattern of a strong relationship. However, the greater distances between the points and the line illustrates the marginal significance of the effect.

The Prediction Profiler provides for a dynamic view of the effects of plug depth and cycle time on minimum thickness. Click on the vertical red segmented line of *plug depth*, and drag horizontally along the range of values to estimate the minimum thickness. Repeat the dynamic estimation of *min handle thickness* with *cycle time*. To estimate the minimum handle thickness with an exact numeric value, click on the red input value for *plug depth* and enter 0.9. Repeat with other values for plug depth and cycle time to see how the estimates change. It is clear that minor changes in plug depth have a larger effect on minimum thickness than changes in cycle time. Basically, plug depths that are 0.9 or greater estimate a minimum thickness that is less than the 1.5 mm low specification. In addition, excessive plug depth (>=90%) with longer cycle times exacerbate the risk of producing handles that have minimum wall thickness that is below the minimum specification.

Detailed Analysis of the DSD Model

The next step is to run the model to get additional details of the analysis. Click *Run Model* in the *Fit Definitive Screening for min handle thickness* window to get the model output shown in Figure 11.7.

Figure 11.7: Thermoforming Reduced Model

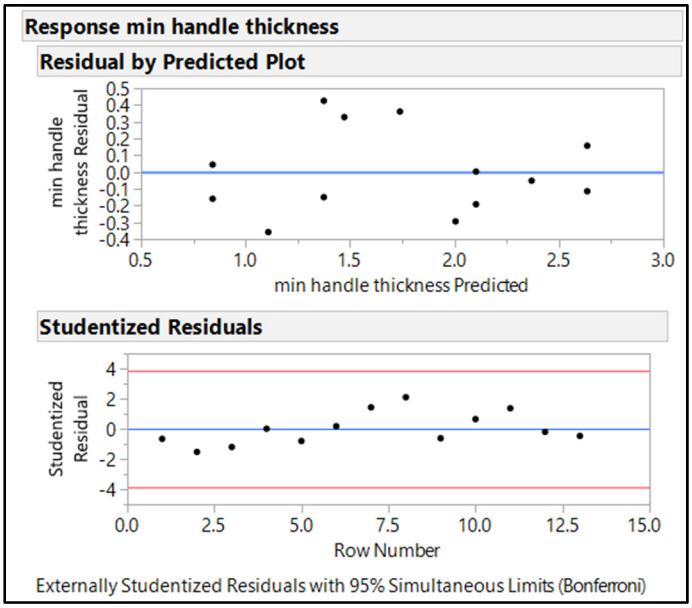


The Actual by Predicted plot enables you to see the relationship between actual and predicted values, indicated with a solid red line. The 95% confidence interval for the mean effect is illustrated with a shaded red area about the model line. The pattern of the predicted results by actual results of the individual runs is shown by the black circular markers. The relationship seems reasonable due to the evidence that the linear model is highly significant ($p<0.0001$). The model explains 86% of the changes in minimum thickness (r-square = 0.86), and has a relatively small amount of random error (RMSE=0.2758). The effect summary lists plug depth and cycle time as significant effects, ($p=0.00003$) and ($p=0.01219$) respectively.

The lack of fit tests are used to detect whether there are observations that have a poor fit to the model, even if the overall trend is of high significance. Notice that there are a limited number of observations that are outside of the shaded confidence interval region. If the observations were within the trend of the interval, the maximum r-square fit of 0.9556 would be achieved. The lack of evidence that poor fitting observations are within the experimental results is noted by the p-value of 0.3805. In general, the model for the two significant thermoforming process variables is very robust.

The use of residual analysis is shown in Figure 11.8.

Figure 11.8: Thermoforming Reduced Model Residual Diagnostics



Residual analysis adds detail to the diagnostics of the robustness of the model. The random pattern of the residuals illustrated in the Residual by Predicted plot is highly desirable. If a cone pattern or other non-random pattern were evident in the plot, the conclusions of the model might be suspect due to the potential for error. The Studentized Residuals plot provides a reference for the user vis the red decision lines at the studentized residual values of +/- 4. None of the observations are beyond the decision lines. Therefore, the residual analysis provides further evidence of a robust model.

The output provided by JMP has justified that the model formed through the two significant process variables of plug depth and cycle time is robust. The goal of the set of experiments has been met. The analysis was used to narrow the improvement efforts in controlling the minimum handle thickness by 67% (two of the six inputs can be manipulated to better control the output). The set of structured, multivariate experiments enabled the team to achieve the goals quickly and efficiently, but even more useful information is available. Effects and predictions can be made from the model to provide additional direction for the team. Michelyne needs to use the information with caution because there is typically an inadequate number of runs included in a screening design to precisely quantify the effects. The parameter estimates are shown in Figure 11.9.

Figure 11.9: Thermoforming Reduced Model Analysis Interpretation

Response min handle thickness

Parameter Estimates

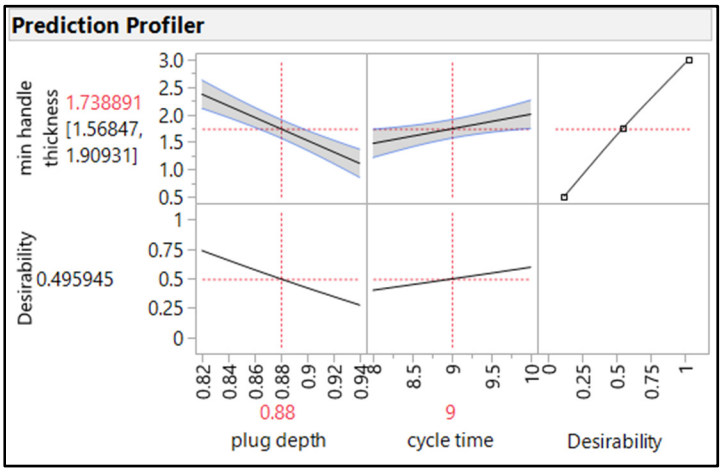
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1.7388909	0.076485	22.73	<.0001*
plug depth(0.82,0.94)	-0.630489	0.087207	-7.23	<.0001*
cycle time(8,10)	0.2662302	0.087207	3.05	0.0122*

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
plug depth(0.82,0.94)	1	1	3.9751607	52.2701	<.0001*
cycle time(8,10)	1	1	0.7087854	9.3199	0.0122*

The parameter estimates indicate that an approximate 0.63 mm decrease in minimum thickness is expected as plug depth is increased one unit of standardized increase (standardized to the design space). This unit of increase is either between the low limit (0.82) and center point (0.88), or between the center point and the high limit (0.94). A one-unit of standardized increase in the design space of cycle time results in an approximate 0.27 mm increase in minimum thickness. The Prediction Profiler in Figure 11.10 is a dynamic illustration of how changes in the two process inputs of the model relate to changes in *min handle thickness*.

Figure 11.10: Thermoforming Reduced Model Prediction Profiler



The Prediction Profiler provides a dynamic plot in which the analyst can try different values of plug depth and cycle time by sliding the vertical segmented red slider lines. You should consider general relationships because prediction accuracy might be lacking. The practical interpretation of the limits on prediction accuracy is evident in the 95% confidence limits of the minimum thickness prediction at the center point of the design space, as shown in the profiler. The average minimum thickness predicted is shown in red as 1.74 mm. However, the value could be as low as 1.57 mm or as high as 1.91 mm, as indicated by the black values for the confidence interval. Michelyne can guard against inaccurate conclusions made by project stakeholders by sticking with the interval estimates in all communication. The report on the analysis might include the statement "For the process set at a plug depth of 88% and a cycle time of 9 seconds, we can expect that the handles produced will have an average minimum thickness of between 1.57 and 1.91 mm". If the stakeholders want increased precision in the estimates, they will need to support the project with more resources for further study.

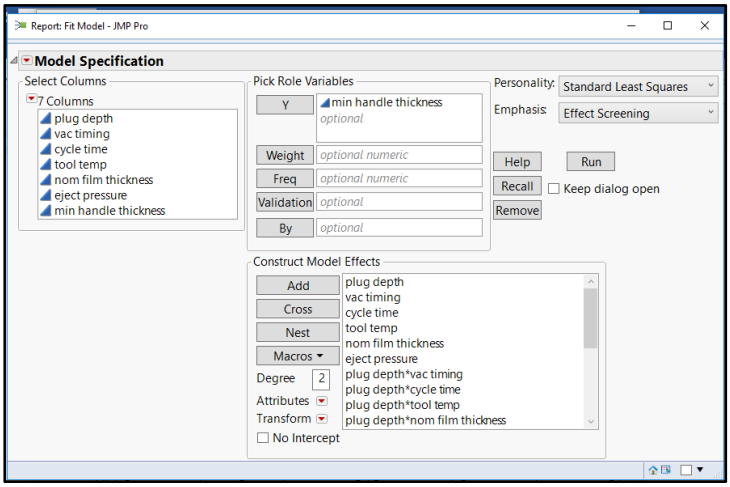
Use of the Fit Model Analysis Menu Option

The previous example went through the model analysis that is run from scripts that are automatically created when you make a table for a DSD model. There might be times when the data from the experiments was not collected via the JMP table, or the person doing the analysis has a JMP license that is a version

prior to 13.0. In such cases, you can complete the analysis of experimental data from a DSD by using the Fit Model platform. Complete the following steps to set up the model analysis:

1. Open *Thermoform DSD Results.jmp*, and select *Analyze* ▶ *Fit Model*.
2. In the *Model Specification* window, make sure that *min handle thickness* is selected as *Y*, *Response*.
3. Select all six process inputs in the *Columns* box, click *Macros*, and then select *Response Surface*, from the options shown in Figure 11.11.

Figure 11.11: Thermoforming Full Model Creation

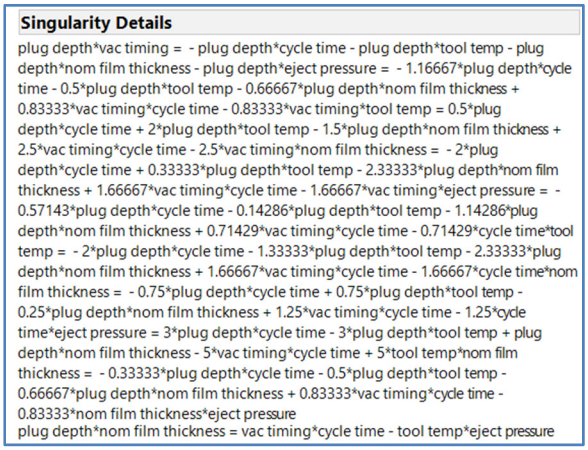


4. Make sure that the *Degree* box includes the default value 2 so that only the two-way interactions and squared terms are included in the model.
5. Click *Run* to get the model output.

Singularity

The thirteen runs of the model are not enough to be able to detect all six individual inputs, all interactions, and all squared terms. When there are more terms than runs in a model, the output includes singularity details. The singularity details in Figure 11.12 indicate that the response surface model does not have enough degrees of freedom for estimating the effects of the 21 factors.

Figure 11.12: Thermoforming Full Model Singularity



The model must be updated to reduce terms down to the number of factors that can be estimated with the degrees of freedom available from 13 runs. The Effect Summary information in Figure 11.13 lists the factors with singularity without the LogWorth or PValue information. You could go back to the *Model Specification* window for a trial and error set of model selections. However, JMP provides an easy tool to reduce the model within the analysis output. Work from the bottom of the list upwards by selecting *plug depth*vac timing* and clicking *Remove* to eliminate it from the model. The model is reduced by the selected factor, and the analysis is redone automatically. Select multiple factors and click *Remove* until the remaining factors match the six that are included in Figure 11.14. (A more efficient method for reducing multiple factors from a model uses stepwise regression. Details about this method are more complex than the manual method. The topic is covered in chapter 13.)

Figure 11.13: Thermoforming Full Model Effect Summary

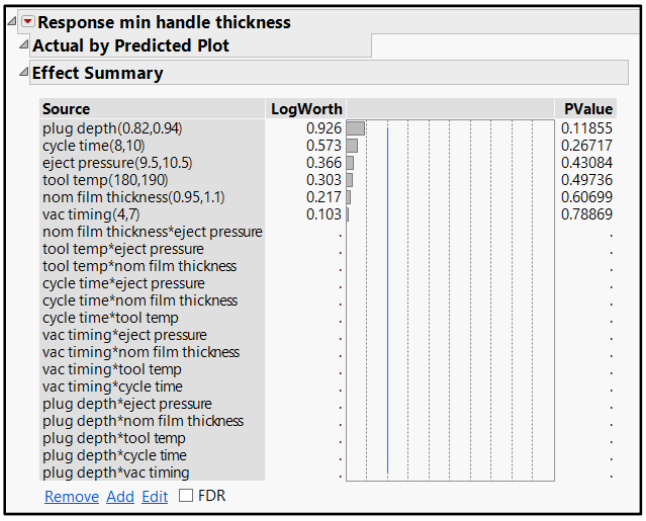
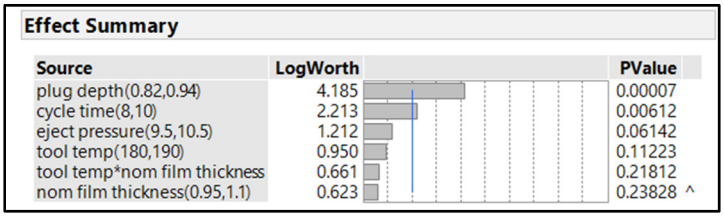


Figure 11.14: Thermoforming Partially Reduced Model

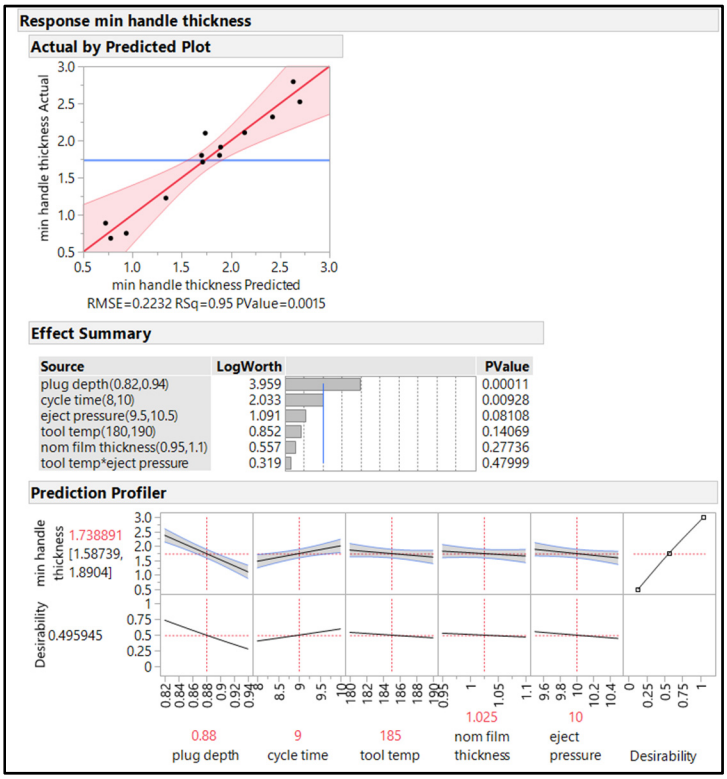


The Effect Summary includes a threshold value for detecting important factors, which is shown as a vertical blue line on the LogWorth horizontal bar chart. The effect summary horizontal bar chart illustrates that the plug depth and cycle time are the only factors that have enough evidence of influence to be considered significant to changes in minimum thickness. The model with the six factors also indicates that interactions are not likely to be of significant influence; the *tool temp*nominal film thickness* interaction has a LogWorth value that is very small (0.661) with an insignificant PValue of 0.218.

Analysis of a Partially Reduced Model

The model summary shown in Figure 11.15 includes additional factors as compared with the Fit Definitive Screening analysis.

Figure 11.15: Thermoforming Partially Reduced Model Fit



Notice that the model fit of Rsquare =0.95 is better than the fit of the 2-factor DSD reduced model fit of Rsquare = 0.86. In addition, the amount of random error in the model decreased from an RMSE of 0.2758 to 0.2232 by including four additional factors. The basic conclusion has not changed; the plug depth is the most significant factor and cycle time the only other significant factor. The four additional factors explain more of the random error, but the signals are too weak to be significant. The analysis provides some hints that further study could offer additional value for the team to optimize the process. It is also possible that the model with additional factors is overfit, so adding runs might not provide useable information.

The team must keep in mind that the DSD provides the best results in defining which of the six process inputs has significant influence on *min handle thickness*. Even though the analysis tools allow for some interpretation of how much influence is exerted, the model lacks enough runs to provide robust estimates. Runs can be augmented easily for more detailed focus on plug depth and cycle time so that the amount of influence can be robustly quantified.

The results of the model analysis are enough for leadership to decide to take quick action. The process controls are changed so that plug depth does not exceed 90% and the maximum cycle time does not exceed 9.5 seconds. The action will immediately reduce the risk of producing thin handles, but the team will not be able to robustly estimate the effectiveness of the actions prospectively. Enhanced process monitoring must be initiated for a period that is sufficient to represent the population of all commercial production. The minimum thickness data collected from in-process checks is to be assessed for capability (chapter 3) and tested for significance to the distribution of data from parts produced prior to the improvement (chapter 4) to ensure that the changes have been effective.

Management has the option for further study of the process at any time to obtain additional improvement by augmenting the DSD model. Augmentation of an existing model requires that the process environment is equivalent to that which was in place for the initial experiments. Adding runs to the model can be a cost-effective way to improve the predictive nature of the model. The risk involved for augmentation is that more resources are added with no additional information extracted from the model.

Analysis of a Response Surface Model with Multiple Outputs

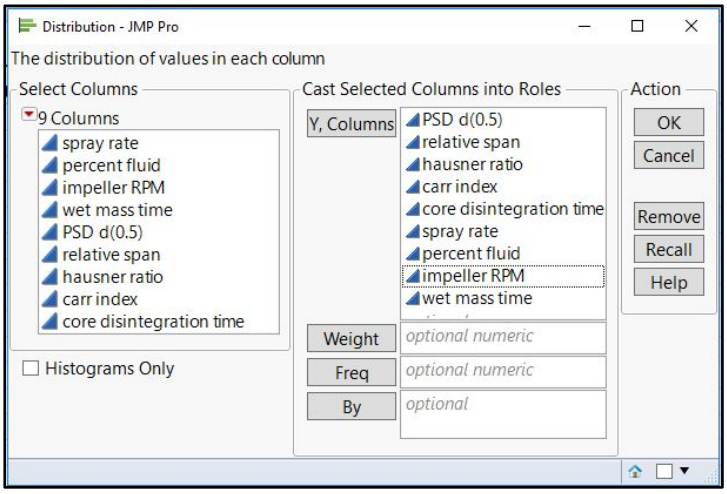
The stakeholders in the next set of experiments have a different set of goals. The high-shear granulation process, which is the subject of the study, has four process inputs that were found to exert significant influence on outputs of the process. Erica, the team leader, faces the challenge that the process must perform well for multiple outputs. JMP is an invaluable tool for the exploration of a process involving multiple inputs as well as multiple outputs. More resources are required for such experiments since the analysis will be used to quantify the levels of inputs that will be included in the manufacturing order protocols to get optimal results. Erica gained the support of leadership to include an adequate number of runs and robustly quantify the effects of the inputs. The response surface design from chapter 10 is complete and the data is ready for analysis. Open *Granulation Process Experiment Results.jmp*, shown in Figure 11.16.

Figure 11.16: Granulation Process RSM Data

	spray rate	percent fluid	impeller RPM	wet mass time	PSD d(0.5)	relative span	hausner ratio	carr index	core disintegration time
1	8	0.3	50	7	125.3	1.78	0.20	16.7	45.6
2	6	0.2	40	2	127.1	1.03	2.30	24.3	38.2
3	8	0.3	40	7	137.5	0.20	3.54	17.3	40.9
4	8	0.2	60	2	101.4	1.33	1.92	23.8	55.2
5	8	0.3	50	7	125.4	1.61	1.92	11.4	48.7
6	6	0.4	60	2	105.4	0.15	1.20	39.0	47.0
7	8	0.3	50	2	116.7	0.65	0.25	28.0	45.2
8	8	0.4	40	2	127.6	0.15	0.58	26.7	36.7
9	8	0.4	50	7	122.3	3.33	2.94	23.9	49.6
10	6	0.2	60	12	101.1	1.64	2.09	7.1	50.0
11	10	0.4	60	2	111.6	1.78	3.17	36.9	65.5
12	6	0.3	40	12	101.2	2.75	2.67	45.5	52.8
13	6	0.4	50	12	103.7	2.01	3.38	24.4	55.2
14	8	0.4	60	12	121.9	1.56	1.49	38.1	48.3
15	10	0.4	40	12	140.3	2.60	1.45	30.2	46.0
16	6	0.2	50	7	115.6	1.30	0.75	14.5	46.9
17	10	0.2	60	8.5	127.7	3.18	0.12	0.6	50.6
18	10	0.3	50	12	137.0	2.95	0.02	24.6	45.7
19	10	0.2	40	2	131.0	0.43	0.18	11.5	44.6
20	8	0.2	40	12	120.0	0.11	1.86	26.4	46.9
21	6	0.3	60	7	111.7	0.57	0.15	30.6	53.1

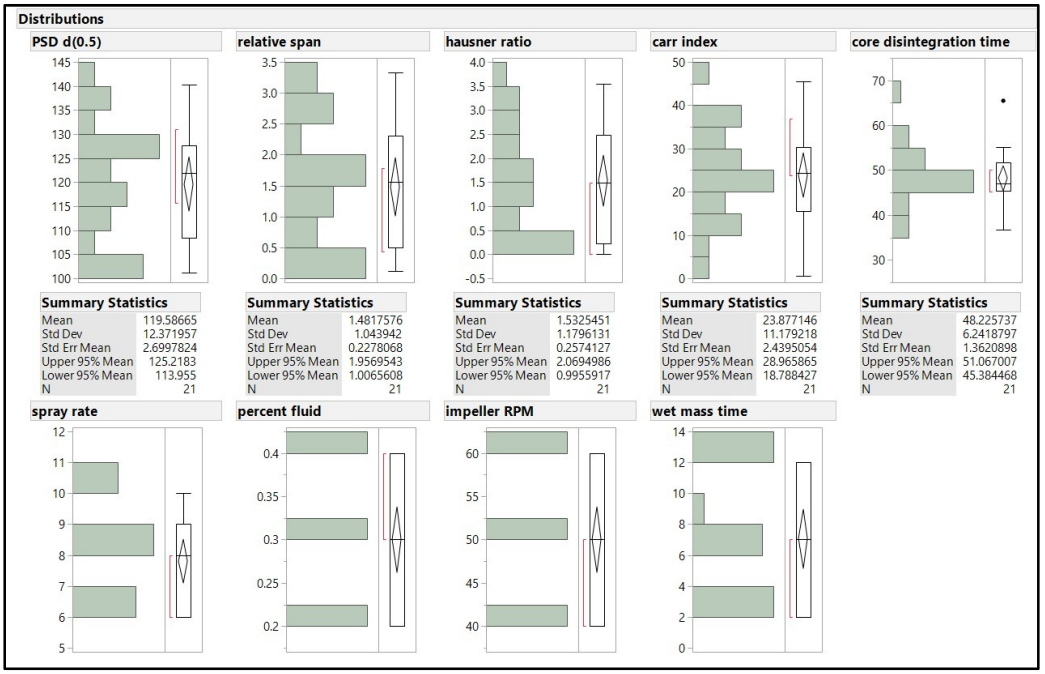
Prior to initiating the model analysis, a high-level look at the data using the distributions platform is appropriate. Select *Analyze ► Distributions* and move the variables to the *Y,Columns* box with the five responses on top and the four inputs underneath, as shown in Figure 11.17. Click *OK* to get the output.

Figure 11.17: Distribution Platform Window



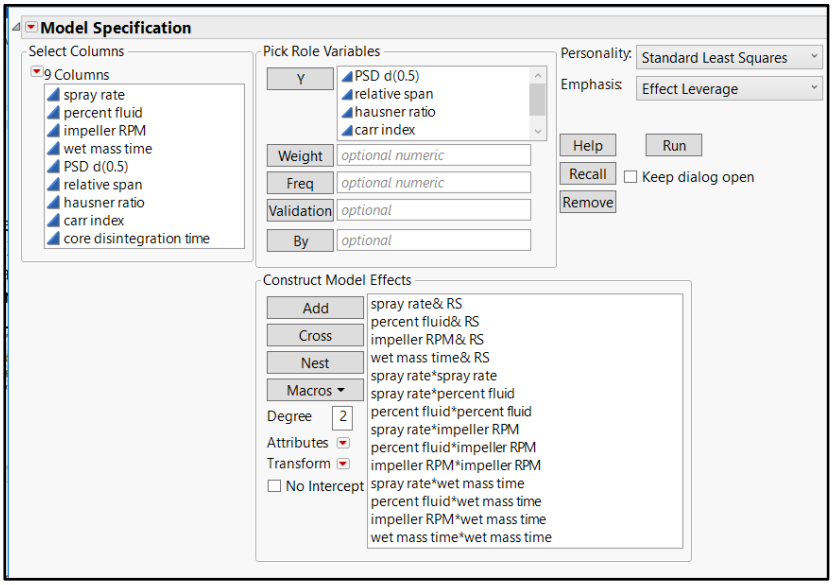
The output needs to be formatted for the best visualization of the experimental data. Press the Ctrl key and click on the gray arrow next to the *Quantiles* header underneath *PSD d(0.5)* to hide the quantiles summary for each variable. Click on the gray arrow next to each of the four inputs to hide the summary statistics. If capability studies are automatically added due to the response limits, use the red triangle menu of each to deselect the *Capability Analysis* option and remove the output. Lastly, use the red triangle menu next to *Distributions* to select *Arrange in Rows*. Enter 5 to arrange the output into two rows, as shown in Figure 11.18.

Figure 11.18: Distributions Output



Use the dynamic features of the Distributions platform to look for high-level patterns of potential relationships among the inputs and outputs. The visualization is basic because interactions and squared terms cannot be viewed. Note any non-random pattern that you observe so that you can explore it with the detailed model analysis. Multivariate analysis can be used as an efficient way to look for relationships among several variables and is explained in a later chapter. With *Granulation Process Experiment Results.jmp* open and in view, click on the green arrow beside the *Model* script to run it. The *Model Specification* window appears, as shown in Figure 11.19. The column properties were previously set in the data table to identify the five outputs and to include response limits for each. Notice that the outputs are selected automatically and moved into the *Pick Role Variables* box in the *Model Specification* window due to the column property settings.

Figure 11.19: Model Specification for Granulation Process RSM

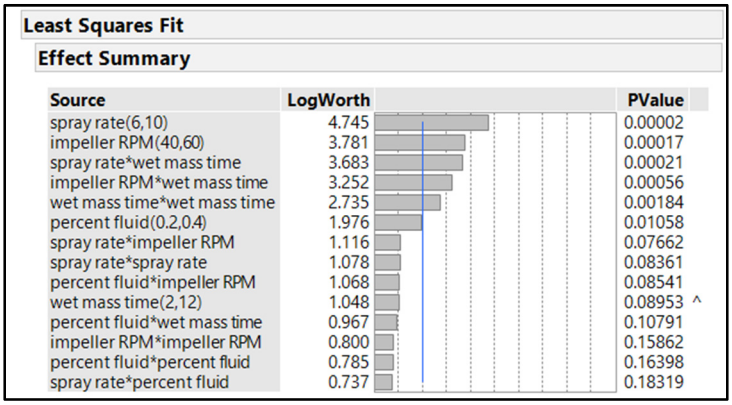


You can manually select the outputs if the data was not collected into the JMP design table or if you used a JMP version earlier than JMP 13:

1. Select *Analyze* ▶ *Fit Model* to open the *Fit Model* window.
2. Press the Shift key and select the five output variables. Drag the selections to the *Y* box in the *Pick Role Variables* section.
3. Press the Shift key, select the four input variables, and click *Macros* (keep the default value of 2 in the *Degree* box). Select *Response Surface* to create the model effects.
4. The *Construct Model Effects* section lists all individual inputs, two-way interactions, and squared terms, as shown in Figure 11.19

The *Model Specification* window is complete with the desired outputs and inputs to study a response surface model (RSM). The default *Personality* value *Standard Least Squares* is used. However, you should change the *Emphasis* to *Effect Leverage* for an RSM. Click *Run* to get the analyses output shown in Figure 11.20.

Figure 11.20: Effect Summary for Granulation Process RSM



Least Squares Fit

Effect Summary

Source	LogWorth	PValue
spray rate(6,10)	4.745	0.00002
impeller RPM(40,60)	3.781	0.00017
spray rate*wet mass time	3.683	0.00021
impeller RPM*wet mass time	3.252	0.00056
wet mass time*wet mass time	2.735	0.00184
percent fluid(0.2,0.4)	1.976	0.01058
spray rate*impeller RPM	1.116	0.07662
spray rate*spray rate	1.078	0.08361
percent fluid*impeller RPM	1.068	0.08541
wet mass time(2,12)	1.048	0.08953 ^
percent fluid*wet mass time	0.967	0.10791
impeller RPM*impeller RPM	0.800	0.15862
percent fluid*percent fluid	0.785	0.16398
spray rate*percent fluid	0.737	0.18319

The Effect Summary is at the top of the analysis output, which examines the significance of the factors regarding influence for all five outputs combined. A significance threshold represented by the vertical blue line adjusts for the number of factors being compared. Five experimental factors have strong enough influence to be significant for all outputs:

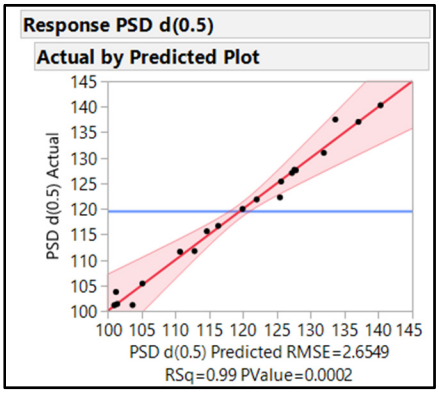
- *Spray rate* with a LogWorth of 4.745 and highly significant PValue of 0.00002
- *Impeller RPM* with a LogWorth of 3.781 and highly significant PValue of 0.00017
- The interaction of *spray rate*wet mass time* with a LogWorth of 3.683 and highly significant PValue of 0.00021
- The interaction of *impeller RPM*wet mass time* with a LogWorth of 3.252 and highly significant PValue of 0.00056
- The squared term of *wet mass time*wet mass time* with a LogWorth of 2.735 and significant PValue of 0.00184

It is clear that the team needs to ensure that the manufacturing limits for spray rate, impeller RPM, and wet mass time are set to obtain robust results. Other factors might be important to specific outputs and will be explored within the experimental models for each output. The analyses of the individual models must be completed next.

Examination of Fit Statistics for Individual Models

Figure 11.21 provides detailed information about the model of the physical attribute particle size d(0.5).

Figure 11.21: Model Fit for Granulation Process RSM



The RSM model exerts a high level of significant influence on particle size (p=0.0002). Changes in the factors of the RSM model explain 99% of the changes in particle size (r-square=0.99). The amount of random variation in the model is limited to 2.65 microns (RMSE=2.6549), which is very small relative to the scale of changes in the outputs approximately 45 microns. The pattern of observations in the predicted plot tightly follows the red model line, with a narrow 95% confidence interval indicated by the red shaded area about the model line.

The next evaluation is the lack of fit test, shown in Figure 11.22.

Figure 11.22: Lack of Fit for Granulation Process RSM

Response PSD d(0.5)

Lack Of Fit

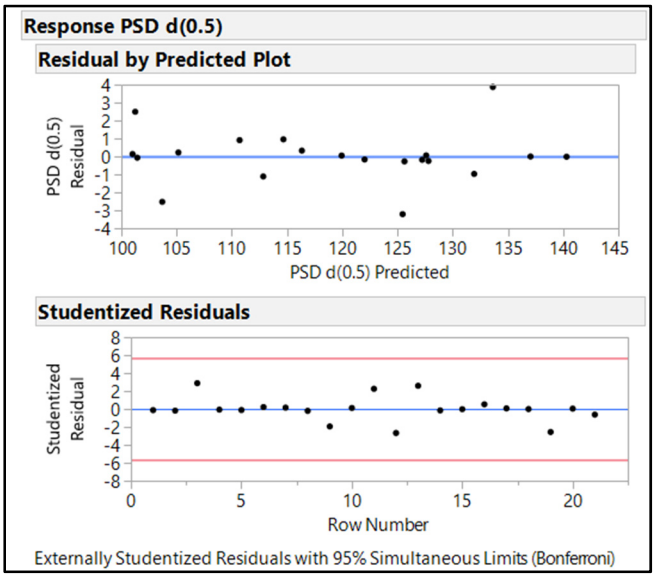
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	5	42.289390	8.45788	17624.66
Pure Error	1	0.000480	0.00048	Prob > F
Total Error	6	42.289870		0.0057*
				Max RSq
				1.0000

The test indicates that a significant lack of fit is present within the model (Prob > F = 0.0057). It is important to keep in mind that the fit of the model is extremely good. Therefore, individual observations that are a small distance from the model line can result in significant lack of fit. The Actual by Predicted plot in Figure 11.21 has one observation that is just below the 95% confidence interval at the approximate prediction of 126 and actual value of approximately 123. Residual analysis can help the analyst better visualize the points contributing to lack of fit.

Model Diagnostics through Residual Analysis

Figure 11.23 includes the residual plots used to further analyze and diagnose the fit of the model.

Figure 11.23: Model Residual Diagnostics Granulation Process RSM



The Residual by Predicted plot illustrates a random pattern that is desired for a robust model. However, there are a few points that have residual values that are more extreme than the other observations. An observation with a predicted value of 126 microns has a residual of -3, indicating that the model overpredicted. An observation with a predicted value of 134 has a residual of 4, indicating that the model underpredicted.

The Studentized Residuals plot has no observations that are outside of the decision limits, which are illustrated by red horizontal lines above and below the studentized residual average of 0. The decision limits are adjusted for the number of factor comparisons included in the model. The Studentized Residual plot including all points that are well within the statistical limits provides evidence that the significant result for the lack of fit test is not likely to create prediction error large enough to be of practical relevance for the information gained.

If you are especially concerned about reductions in the precision of estimates due to lack of fit, you can filter to exclude the runs that contribute to lack of fit, and then run the model analysis with the remaining data. If the conclusions from the analysis of filtered data do not change enough to be practically relevant, the original model with the lack of fit is further supported. Running models with excluded observations should be done with caution. It is possible that the exclusions might not clear up the lack of fit, and the results might change the conclusions that come from the analysis. Other issues for filtered models include lack of power, changes in the structure that might increase correlation and confounding of factors, and reductions in statistical power. In general, it is typically best to note the lack of fit in assumptions for models that have a good overall fit to ensure that details are fully disclosed to the consumers of the information.

Parameter Estimates

The summary values of model fit and residual analysis help you determine that the model is adequate to explain changes that occur in PSD d(0.5). Parameter estimates shown in Figure 11.24 are the part of the analysis that provide a great deal of information about the design space.

Figure 11.24: Parameter Estimates for Granulation Process RSM

Response PSD d(0.5)

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	125.63	1.309209	95.96	<.0001*
spray rate(6,10)	9.906923	0.808325	12.26	<.0001*
percent fluid(0.2,0.4)	1.4491105	0.764072	1.90	0.1067
impeller RPM(40,60)	-6.343936	0.764072	-8.30	0.0002*
wet mass time(2,12)	1.4187354	0.75404	1.88	0.1089
spray rate*spray rate	0.6152473	1.276877	0.48	0.6470
spray rate*percent fluid	1.3973299	0.928854	1.50	0.1832
percent fluid*percent fluid	-1.61861	1.425257	-1.14	0.2994
spray rate*impeller RPM	-1.186929	0.928854	-1.28	0.2485
percent fluid*impeller RPM	1.7606584	0.855945	2.06	0.0854
impeller RPM*impeller RPM	1.660268	1.425257	1.16	0.2883
spray rate*wet mass time	7.3441552	0.921117	7.97	0.0002*
percent fluid*wet mass time	0.2716136	0.851917	0.32	0.7607
impeller RPM*wet mass time	5.6633563	0.851917	6.65	0.0006*
wet mass time*wet mass time	-7.872113	1.486879	-5.29	0.0018*

The first term of the parameter estimate list is the intercept. There is significant evidence ($p<0.0001$) that the intercept of the model is not zero (intercept =125.63), which is information of no practical value for the context of the granulation study.

It is good practice to evaluate the effects of the complex factors before the main effects. There are two interactions and one squared term that have significant leverage on changes in PSD d(0.5):

- The interaction of *spray rate*wet mass time* is estimated to significantly increase PSD d(0.5) by an average of 7.3 microns for each unit increase of the factor ($p=0.0002$).
- The interaction of *impeller RPM*wet mass time* is estimated to significantly increase PSD d(0.5) by an average of 5.7 microns for each unit increase of the factor ($p=0.0006$).
- The curved effect of *wet mass time*wet mass time* is estimated to significantly decrease PSD d(0.5) by an average of 7.9 microns for each unit increase of the factor ($p=0.0018$).

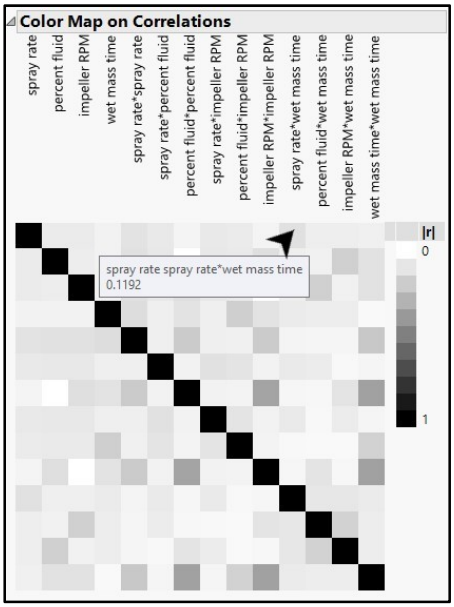
The presence of highly significant two-way interactions and a curved effect provide evidence of a complex design space. The stakeholders of the developmental drug product could not have identified the real cause of changes in various responses without the use of the multivariate, structured experiments. The results of this example emphasize the quintessential reason why international regulatory agencies have emphasized Quality by Design (QbD). The knowledge of the cause and effect model relationships are used to set robust process controls. Such controls are not possible when teams use the one factor at a time (OFAT) methods.

The main effects are the last to be evaluated in an RSM. There are two significant main effects that also have a part in the interactions noted above:

- A unit increase in *spray rate* significantly increases PSD d(0.5) by 9.9 microns ($p<0.0001$).
- A unit increase in *impeller RPM* significantly decreases PSD d(0.5) by 6.3 microns ($p=0.0002$).

Refer to the model evaluation to determine the amount of correlation present between significant interactions and significant main effects to determine the practical relevance of the individual inputs. You can quickly explore the model evaluation by using the scripts that have been automatically added to the data table during the design phase. With *Granulation Process Experiment Results.jmp* open and in view, click on the green arrow beside the *Evaluate Design* script. *Color Map on Correlations* appears. Place your pointer over each cell of the matrix and move it slightly until a note appears with the amount of correlation present. Figure 11.25 shows the highest amount of correlation between a significant individual input and a significant two-way interaction. *Spray rate* is 12% correlated with the interaction of *spray rate*wet mass time*.

Figure 11.25: Color Map Analysis

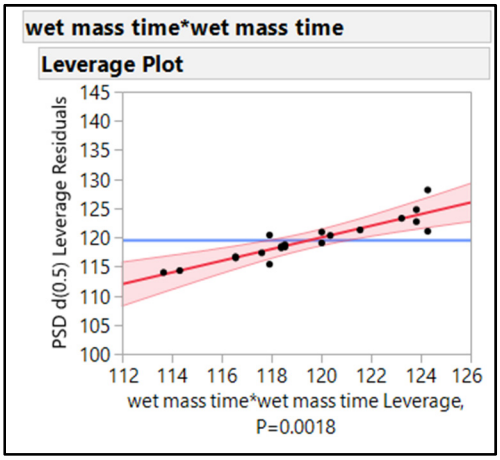


The low correlations of 6% to 12% are not of great concern. The evidence of a significant effect from the two individual inputs, two interactions, and the squared term can be included in the conclusions of the analysis. A mention of the low correlation between factors is advised to ensure that consumers of the analysis have all the important details.

Detailed Analyses of Significant Factors with Leverage Plots

The goal for the set of experiments is to obtain high-resolution information so that robust predictions can be made for how changes in the process inputs affect the outputs. The experimental design with a response surface model includes individual inputs as well as complex terms to aid the goal. You might have noticed that the model output for each response includes a set of Leverage plots, located to the left of the Prediction plot for the full model. Leverage plots isolate the effect of each factor on the output in the presence of the influences of all other factors. The Leverage plot shown in Figure 11.26 provides detail about the amount of leverage that results from the curved effect of wet mass time.

Figure 11.26: Leverage Plot of (Wet Mass Time)² for PSD d(0.5)



The squared term factor exerts highly significant (P=0.0018) leverage on PSD d(0.5). The power for the curved wet mass time factor exerting significant influence on PSD d(0.5) is 32.5%, as estimated during the design phase (Figure 11.27). The estimate is based on one unit of experimental error (RMSE) and a one-unit anticipated coefficient for the factor.

Figure 11.27: Estimated Model Power from the Experimental Design (from Chapter 10)

Design Evaluation

Power Analysis

Significance Level 0.05

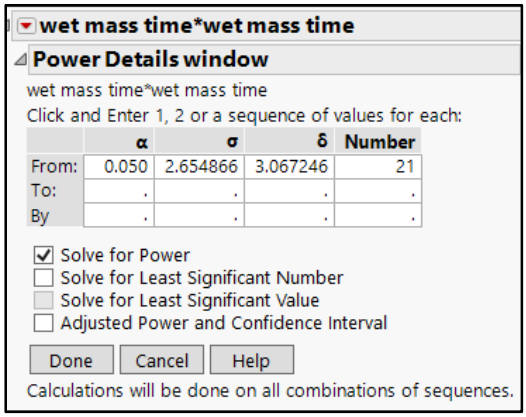
Anticipated RMSE 1

Term	Anticipated Coefficient	Power
Intercept	1	0.4
spray rate	1	0.781
percent fluid	1	0.823
impeller RPM	1	0.823
wet mass time	1	0.833
spray rate*spray rate	1	0.416
spray rate*percent fluid	1	0.666
percent fluid*percent fluid	1	0.348
spray rate*impeller RPM	1	0.666
percent fluid*impeller RPM	1	0.735
impeller RPM*impeller RPM	1	0.348
spray rate*wet mass time	1	0.673
percent fluid*wet mass time	1	0.738
impeller RPM*wet mass time	1	0.738
wet mass time*wet mass time	1	0.325

The analysis of the PSD d(0.5) model is complete. It is good practice to obtain the actual statistical power for the significant factors of the model and compare that with the estimates used to justify the design chosen. The power is easily calculated from the leverage plots by completing the following steps:

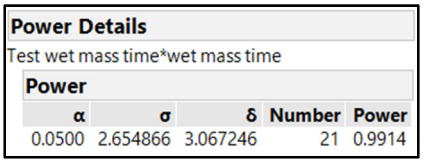
1. Click on the red triangle menu next to the *wet mass time*wet mass time* plot header.
2. Select the only available option for the plot, *Power Analysis*. The *Power Details* window shown in Figure 11.28 appears.

Figure 11.28: Power Details Window



3. The significance value α, standard deviation σ, the size of effect to be detected δ, and sample size *Number* have been automatically included by JMP. There is no need to alter the given values for power to be calculated.
4. Select the *Solve for Power* check box.
5. Click *Done* to add the *Power Details* underneath the leverage plot, as shown in Figure 11.29.

Figure 11.29: Power of the (wet mass time)² Significant Leverage on PSD d(0.5)



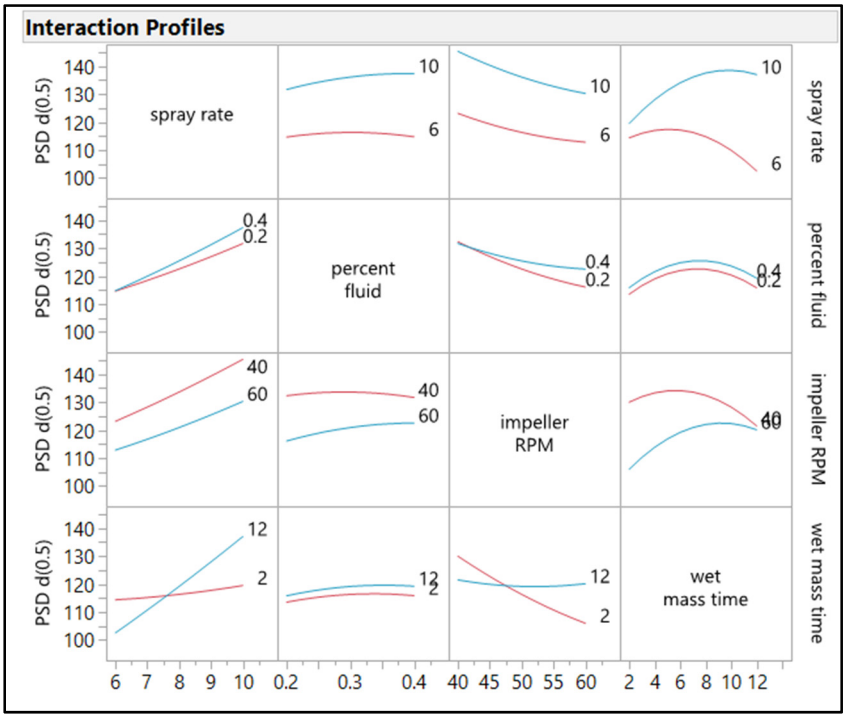
The results shown in Figure 11.29 indicate that the significant effect of *wet mass time*wet mass time* on PSD d(0.5) is more than 99% likely to be a real effect for the population of granulations represented by the design space explored by the set of experiments. Power increased dramatically from the design estimates largely because the magnitude of the amount of change in the output levered by the factor is larger than the random error represented by the standard deviation. Put simply, the signal is large as compared with the small amount of noise. The significance level of 0.05 and the sample size could affect power but neither changed between the design phase and the analysis of the experiments. There are four other significant effects that should be analyzed for statistical power and included somewhere in the details of an analysis report provided to the project stakeholders. There is no need to analyze power for the factors that lack evidence of significance; the definition of power is a value that explains the likelihood that a significant effect will be real for the population represented by the experimental sample.

Visualization of the Higher-Order Terms with the Interaction Plots

The analysis of the effects provides robust numeric evidence of how the factors affect the PSD d(0.5). Numeric estimates are very useful but interpretation is confusing with relationships that are known to be complex. JMP always includes excellent graphics to help people visualize results and interpret trends with ease. The Interaction Profiler is a tool that breaks down complex relationships for sound understanding,

even for people who are not well versed in statistics. Click on the red triangle menu next to the ***Response PSD d(0.5)*** header in the analysis output. Select ***Factor Profiling*** ▶ ***Interaction Plots*** option to get the plot shown in Figure 11.30.

Figure 11.30: PSD d(0.5) Interaction Profiles



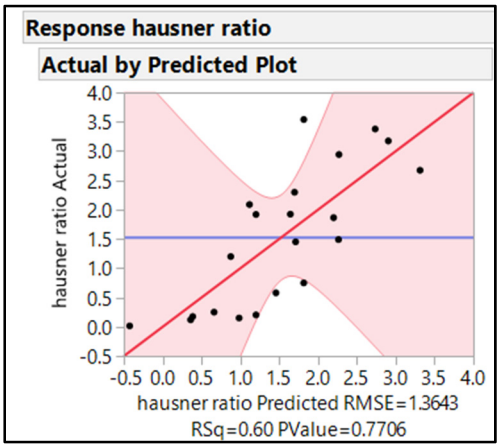
Trend lines are illustrated with blue and red colors for the noted levels, which is not evident in the plot converted to black and white for publishing. This text references the colors for ease of understanding.

The complex relationship for ***spray rate*** and ***wet mass time*** appears in the top right plot of the matrix. The curvature of the profiles illustrates the significance of the ***wet mass time*wet mass time*** factor. The blue profile shows that PSD d(0.5) increases at a steep rate of growth for the highest spray rate (10). The red profile illustrates that the PSD d(0.5) decreases at an increasing rate as ***wet mass time*** is increased and ***spray rate*** is at the lowest level (6). The differing profiles provide robust graphical evidence of the presence of the highly significant interaction. The plot in the lower left corner of the matrix illustrates the same interaction. However, ***wet mass time*** is shown in the fixed low and high values, with the ***spray rate*** indicated as the explaining variable along the X axis. The plot just below the ***wet mass time*spray rate*** plot examines the effect of ***wet mass time*percent fluid***. Notice that the curved profiles are parallel to each other. The parallel profiles illustrate the lack of a significant interaction. There can be no argument that the model created from the set of experiments provides a great deal of useful information about the PSD d(0.5) response. Each of the five outputs will have a unique model included as outline headers in the analysis output. The model for the Hausner ratio is examined next.

Examination of an Insignificant Model

Figure 11.31 provides information to assess the fit of the model for the Hausner ratio.

Figure 11.31: Model Fit for Hausner Ratio



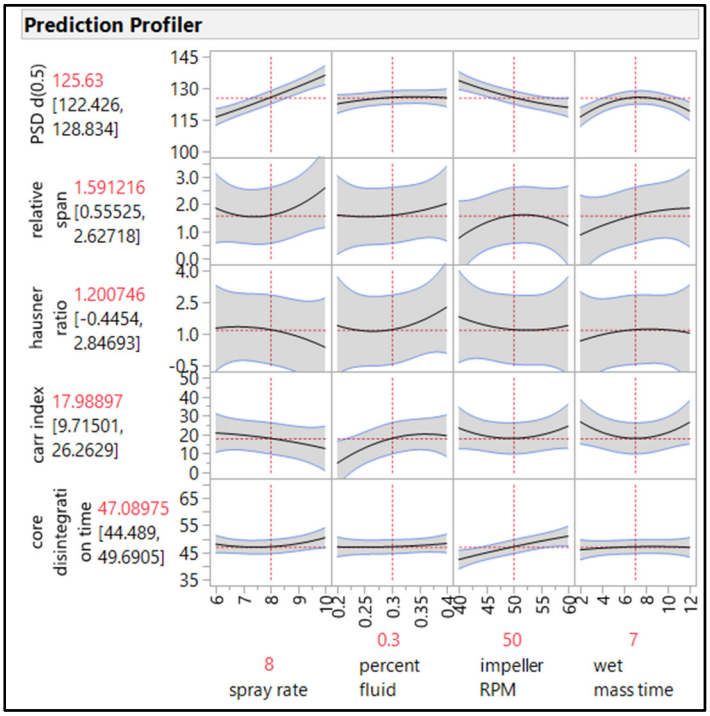
The model is insignificant for providing information about the Hausner ratio (PValue=0.77). Changes to the inputs of the model explain 60% of the changes in Hausner ratio (Rsquare = 0.60). The RMSE of 1.36 is compared with the overall range in output values of approximately 3.7 and is sizeable. The shaded area of 95% confidence interval widens dramatically about the center point of the predicted values, also known as centroid. The actual observations are scattered widely about the model prediction line

The information provided by the Actual by Predicted plot tells you that the effect on the Hausner ratio is very subtle or insignificant. Therefore, the detailed analysis of the Hausner ratio is pointless. JMP calculates values for residual analysis and model effects, but the detail is not needed for an insignificant model.

Dynamic Visualization of a Design Space with the Prediction Profiler

The detailed analysis of the remaining three outputs is left up to you to explore. The Prediction Profiler is an extremely useful tool for examining the dynamic relationships among all the inputs and the outputs simultaneously. This powerful tool typically becomes available by default when you run the model script. If the profiler is not initially visible, you can easily add it to the output. Click on the red triangle menu located to the left of the *Least Squares Fit* header, and select *Profilers* ▶ *Profiler* to add it to the bottom of the analysis output, as shown in Figure 11.32.

Figure 11.32: Granulation Process RSM Model Profiler

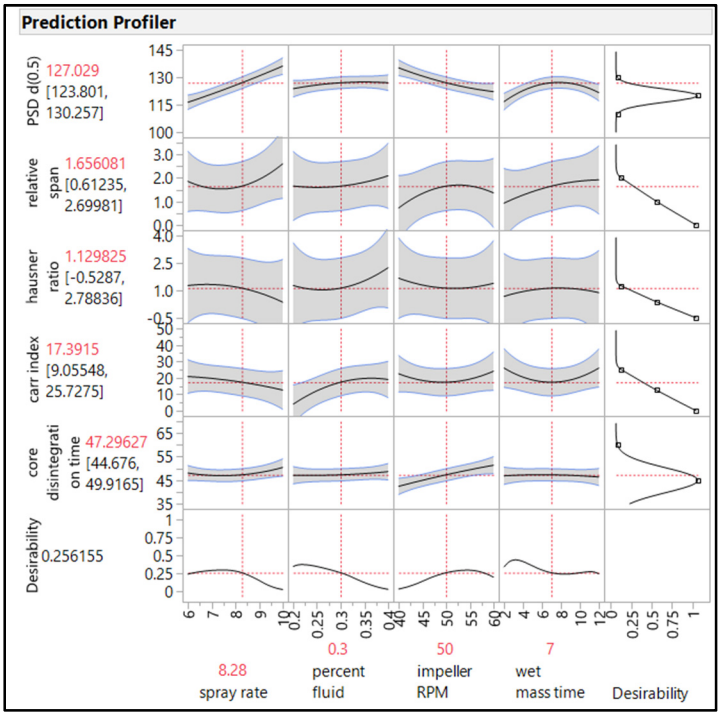


The profiler is a dynamic matrix plot of all inputs as columns and outputs as rows. Recall that the curved term of *wet mass time*wet mass time* is significant for several of the output models. The curvature of the model profile lines illustrates the non-linear responses of the model effects. The gray regions about the profiler lines show the 95% confidence interval for the average response and are illustrative of prediction error. The horizontal, segmented red lines indicate the average response for the chosen levels of the inputs. The vertical, segmented red lines are dynamic sliders that you can click and drag to explore the dynamic changes that take place in the outputs.

Notice that the slope of some profiles change as you manipulate the slider of an input between low and high values. Move the slider to represent changes in *wet mass time* to watch the slope change in *impeller RPM* for various outputs. Changing slopes indicate that significant interactions are present; vertical shifts indicate independence among inputs. You can also click on the red numeric value of an input to enter a specific value. The red numeric value for the output is the average prediction, and the black numeric values are the low and high 95% confidence limits for the prediction. It is best practice to focus on the 95% confidence intervals for the most precise predictions.

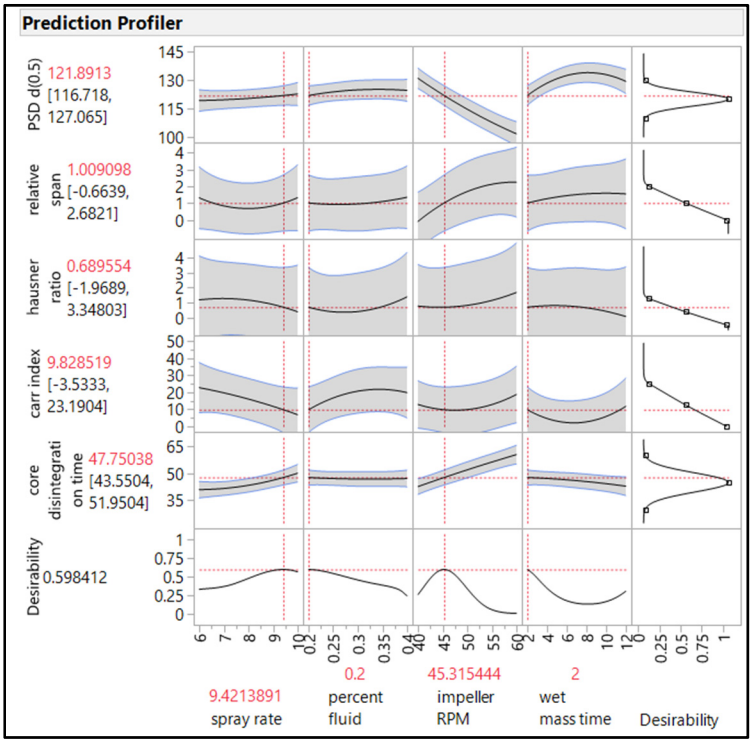
Recall that the responses modeled include both a goal and specified limits. The goal can be to match a target, or to minimize or maximize results. The profiler includes functionality to indicate how well the models meet the goals for all responses. The desirability function accounts for the goals of all responses. The range of desirability is between a minimum of 0 and a maximum of 1 to explain how well the models will meet goals. The higher the value, the more likely that all goals will be met satisfactorily. Click on the red triangle menu next to the *Prediction Profiler* header, and select *Optimization and Desirability* ▶ *Desirability Functions* to get the plot shown in Figure 11.33.

Figure 11.33: Granulation Process RSM Profiler with Desirability Function



The peaked desirability for PSD d(0.5) and core disintegration time reflects the goal to match a target. The small squares located on the lower and higher tails of the function indicate the limits; the square on the peak is the target. The downward sloping desirability functions for relative span, Hausner ratio, and Carr index reflect the goal to minimize. The small square located at the highest point of the function is the upper limit, and the middle and lower squares reflect the median and minimum predicted values for the output. The mid-point settings for the four process inputs reflected in the profiler yield a relatively low desirability of 0.256. The analyst can manipulate each of the inputs to obtain higher desirability, but the function shape changes dynamically as inputs change. JMP provides an algorithm that quickly finds the optimum settings to get the highest possible level of desirability automatically. Click on the red triangle menu next to the *Prediction Profiler* header, and select *Optimization and Desirability* ▶ *Maximize Desirability* to get the plot shown in Figure 11.34.

Figure 11.34: Granulation Process RSM Profiler Optimized



The highest possible desirability is 0.598, which is more than double the value obtained by the mid-point settings. The team can obtain the best possible results with a high spray rate (9.4), least amount of percent fluid (0.2), moderately low impeller RPM (45.3), and least amount of wet mass time (2). Stakeholders can expect the following outputs:

- PSD d(0.5) of between 116.7 and 127.1 microns
- Relative span maximum of 2.7*
- Hausner ratio maximum of 3.3*
- Carr index maximum of 23.2
- Core disintegration time of between 43.6 and 52.0 seconds

**Model not significant*

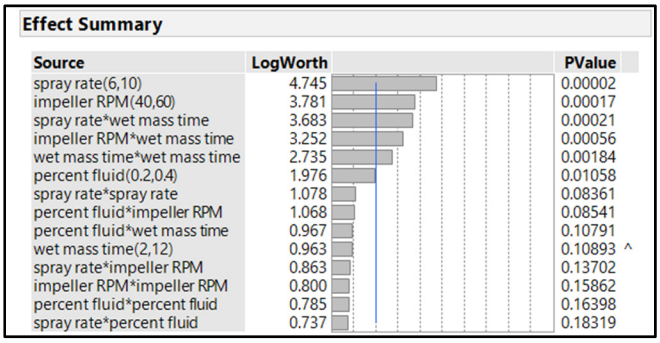
Elimination of Insignificant Models to Enhance Interpretation

It is very important to present a clear message to the project stakeholders. There is no requirement for the number of words in business; in fact, less is always more. There is little point in providing detailed results and predictions for models that you know to be insignificant. The analysis of all five outputs is an important step in the process of discovery. The analyst should save the analysis as a script so that it can be pulled up as needed at any time. Complete the following steps to create a new set of analyses limited to the significant models. The result provides the clarity needed to share results with the stakeholders.

1. Click on the red triangle menu next to the *Least Squares Fit* header.
2. Select *Save Script* ▶ *To data table* to save the 5-output script.
3. In the *Save Script As* window, click to the right of the *Fit Least Squares Name* and enter "5 outputs" so that the name of the script is clear.
4. Click *OK* to save the script to the data table.
5. Notice that other options exist for saving scripts. Explore these options to find the one the best meets the needs of your organization.
6. Click again on the red triangle menu next to the *Least Squares Fit* header.
7. Select *Redo* ▶ *Relaunch analysis* to get to the *Model Specification* dialog box.
8. Press the Ctrl key or the Shift key, and click *Relative Span* and *Hausner Ratio* to highlight them in blue in the *Pick Role Variables* box.
9. Click *Remove*, and then click *Run* to get the analysis output for three responses.
10. Click on the red triangle menu next to the *Least Squares Fit* header.
11. Select *Profilers* ▶ *Profiler* to add it to the bottom of the analysis output.

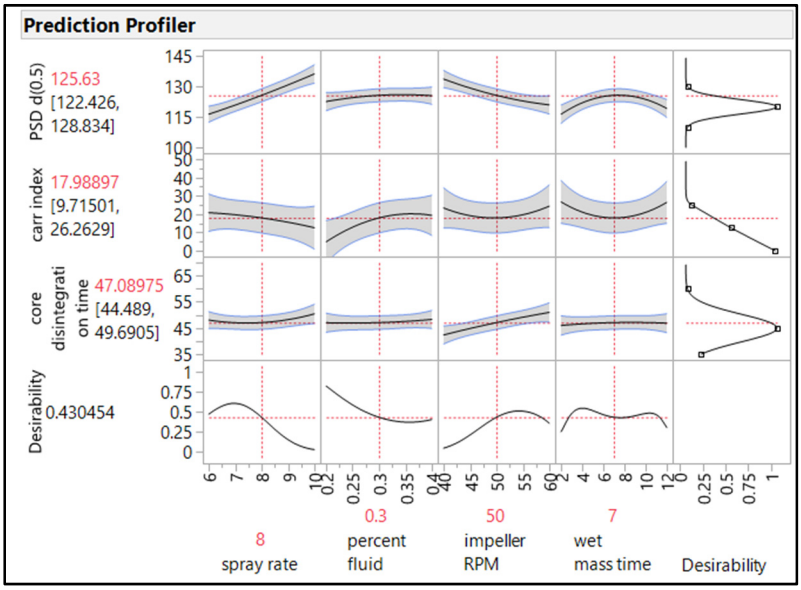
None of the models change. However, the effect summary shown in Figure 11.35 changes slightly because the LogWorth values are calculated for three models instead of for the original five.

Figure 11.35: Granulation Process RSM Profiler Optimized (Significant Models Only)



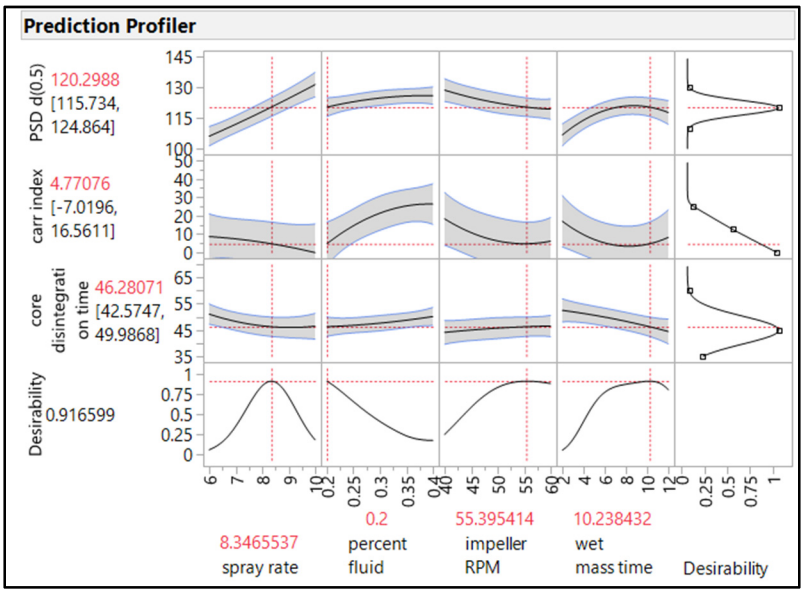
The profiler in Figure 11.36 illustrates an increased desirability for the inputs set to the mid-point values (desirability = 0.43). This makes sense because it is much easier to meet the goals for a reduced number of outputs (responses).

Figure 11.36: Granulation Process RSM Profiler Optimized (Significant Models Only)



Use the red triangle menu options of the profiler to maximize the desirability of the three output models. Figure 11.37 indicates that the analysis can achieve an excellent desirability of 0.917.

Figure 11.37: Granulation Process RSM Profiler Reduced and Optimized



The settings listed in Table 11.1 are the most likely values needed for the granulation process to achieve all the experimental goals.

Table 11.1 Granulation Process RSM Profiler Reduced and Optimized

Input Settings	Outputs	95% Low	95% High
Spray rate at 8.3	PSD d(0.5)	115.7	125.0
Percent fluid at 0.2	Carr Index		16.6
Impeller RPM at 55.4	Core Disintegration Time	42.6	50.0
Wet mass time at 10.2			

The results for the RSM have provided a great deal of information. It is good practice to save the analyses for the reduced set of models to the data file as a final step. Click on the red triangle menu next to the *Least Squares Fit* header, and select *Save Script ► To data table* to save the 5-output script. In the *Save Script As* window, click to the right of the *Fit Least Squares Name* and enter "3 outputs" so that the name of the script is clear. Click *OK* to save it to the data table.

Practical Conclusions

The stakeholders for both projects enjoy a rich amount of information, providing detail about how the experimental goals have been met through the use of structured, multivariate experimentation. Michelyne's team tasked with the problems of surgical light handle covers with thin walls quickly found that they can focus on changing only two of the six process inputs. They quickly confirmed the results of the new input requirements with a few confirmation runs and immediately incorporated them into the manufacturing order process controls. Erica's team, studying the granulation process, used the detailed analysis outputs to create a set of process input controls that have the highest likelihood of producing robust results for three important outputs. The two outputs that had insignificant models are not critical. The team decided to monitor the uncontrolled outputs and determine whether further study might be necessary to determine whether controls are possible. Structured, multivariate experimentation provides the highest value information to incorporate as elements of QbD for the planned submission package in order to gain regulatory approval and achieve the highest level of quality in the drug product produced.

The resources required to design and execute the sets of experiments might have seemed daunting at first as compared with utilizing hierarchical or OFAT experimentation. However, the amount of information provides for great confidence in decisions made to ensure the highest level of robust output from each process. This chapter might paint a somewhat unrealistic picture, since a single set of experiments was executed for each problem. Use of these techniques can best be described as a journey of enlightenment about the process studied. There will be times when a single, well-designed set of experiments provides the information needed by the stakeholders. It is more likely that one set of experiments answers many

questions yet initiates more detailed questions as layers of understanding are achieved through analysis of results. The wonderful aspect of the utilization of the techniques is the structure provided, which you can use throughout the lifespan of a process. If problems arise in commercial production, teams can use the models to focus efforts on the most likely contributing factors for unsatisfactory results. JMP provides a nearly infinite amount of design and analysis options within the DOE and Analysis platforms, which you can easily employ to facilitate the study of all types of processes.

Exercises

E11.1—The project to improve the film seal on the surgical trays is progressing quickly since the team started using JMP. Predictive modeling narrowed the scope from nine process inputs to three: line speed, dwell time, and head energy. Leadership has asked the team to define how each of the influential predictors is affecting the seal strength of the film. Each of the inputs can be easily changed for each run, and the project team was able to get agreement to study the process with a response surface design. The custom designer in the DOE platform included 16 runs as the default size of the experiment, which the team executed.

1. Open *Burst Testing Experiments 3F 16R CP.jmp*.
2. Use the *Model* script to launch the analysis. This script is available since the responses were added to the JMP data table that was created during the experimental design phase of the project.
3. In the *Model Specification* dialog box, change *Emphasis* to *Effect Leverage* before you run the analysis.
4. Summarize the results into a report , and share only the most important information Be sure to quantify the random error for the model.
   - Is the model fit strong enough to suggest a robust model?
   - Is the model significant?
   - Is there an issue with lack of fit?
   - Are the residuals indicating a random pattern?
   - Which are the significant parameter estimates and how much does each affect seal strength?
   - Use the red triangle menu options for each Leverage plot to get the actual power for the significant effects. Is there enough sample size for a robust model?
5. Use the red triangle menu next to the *Response burst pressure (in Hg)* and select *Factor Profiling* ▶ *Interaction Profiler* to visualize the interactions. Explain them in the report.
6. How important was it to include squared terms?
   - Go to the *Effect Summary* header near the top of the analysis and select the three squared terms to remove them from the model.
   - Compare and contrast the fit statistics, model significance, and random error to determine the importance of including the squared terms.
   - Click *Undo* to add the squared terms back to the model.
7. Use the red triangle menu next to the *Response burst pressure (in Hg)* and select *Factor Profiling* ▶ *Profiler* to use the model profiler.
   - Use the red triangle menu options to maximize the model.
   - Make a table of the settings required to get the best seal strength.

E11.2—A minimal screening study with six experimental runs was developed for the tablet compression process in the previous chapter. A pressing concern is to determine whether any of the process inputs are related to lowered tablet hardness and changes in variability. The technical operations team utilized pilot scale equipment to execute the set of six experiments and collected the results of mean tablet hardness (SCU) and the variability in tablet hardness (%RSD). The information is available in the Excel spreadsheet *tablet compression study results.xlsx*.

1. Open the Excel data sheet in JMP and utilize the effect screening emphasis of fit model to analyze the data.
2. Use the output to summarize the fit of the model and the evidence of model significance.
3. Utilize the residual plots to diagnose the health of the model.
4. What conclusion would you provide to project stakeholders?

E11.3—The injection molding process was modeled with a definitive screening design due to the large number of process inputs and the likelihood that influence on the outputs of part width and part weight is limited to a few of the ten inputs. The molded part is a trigger for the inhaler assembly and must have a width of between 11.75 mm and 12.25 mm in order to have the needed clearance to operate as designed. The weight of the molded trigger is designed to be between 3.05 g and 3.35 g in order for the return spring to work properly. Operations and engineering worked together to execute the set of experiments and collect the data from the parts made. The project stakeholders need to know the inputs that affect the outputs so that the team can quickly optimize the process and immediately improve the quality of parts made.

1. Open *inhaler molded component process study DSD 26R.jmp*.
2. Use the *Fit Definitive Screening Design* script to get a summary of the important factors.
3. Click *Run Design* to obtain the detailed analysis for each of the outputs.
4. Create a summary report with practical conclusions that you would present to the project stakeholders. Be sure to include a table of the optimized process with expected results in the summary report.
    - What is the fit of each model? Do you expect that the models will product robust estimates for the two outputs?
    - Which of the process inputs have the most influence? Is there a presence of combined effects (interactions) or complex (non-linear effects) of inputs?
    - Are there any risks of parts that might not function properly for the combination of inputs studied?
    - What can be expected for the average part width and average part weight of the population of all parts that are made at the optimized input settings?

E11.4—A materials study design was initiated in the chapter 9 exercises involving three materials inputs and one slack variable to make up the total weight of a dose. Mixes were made and test batches completed for the 12 runs of the experiments. The team collected data on six outputs: tablet assay, content uniformity (acceptance value of 10 tablets), average dissolution at 1 hour, average dissolution at 4 hours, and the variance within the six dissolution tablets tested for both 1 and 4 hours. Project stakeholders are interested in the effects of the changes in materials as well as any combined effects (interactions).

Goals for the outputs are as follows:

- Assay 90% to 110%
- AV10 (content uniformity) NMT 15
- Dissolution at 1 hour 25% to 50%
- Dissolution at 4 hours 60% to 85%
- Variance in 1-hour dissolution NMT 10
- Variance in 2-hour dissolution NMT 10

1. Open *formulation materials study results.jmp*.
2. Run the analysis of the models for the six outputs.
3. Create a summary report of the analysis to present to the project stakeholders. Can the goals be met robustly?
    a. Be sure to comment on any limitations to the design.