STATISTICS FOR FOOD SCIENTISTS

MAKING SENSE OF THE NUMBERS

FRANK ROSSI
VIKTOR MIRTCHEV

STATISTICS FOR FOOD SCIENTISTS

This page intentionally left blank

STATISTICS FOR FOOD SCIENTISTS
Making Sense of the Numbers

FRANK ROSSI
VIKTOR MIRTCHEV
Kraft Foods, Illinois, USA

ELSEVIER

Amsterdam • Boston • Heidelberg • London
New York • Oxford • Paris • San Diego
San Francisco • Singapore • Sydney • Tokyo
Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, UK
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA
225 Wyman Street, Waltham, MA 02451, USA
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

Front and back cover design created by Kremena Rouhliadkova.

ISBN: 978-0-12-417179-4

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

Publisher: Nikki Levy
Acquisition Editor: Patricia Osborn
Editorial Project Manager: Jaclyn Truesdell
Production Project Manager: Caroline Johnson
Designer: Greg Harris

Typeset by TNQ Books and Journals
www.tnq.co.in

Printed and bound in the United States of America

CONTENTS

This page intentionally left blank

ACKNOWLEDGMENTS

From Viktor and Frank:

A project of this nature is not the work of just two people. Without the help of our friends, co-workers, business partners, and our families, we would have never been able to write this book. We have learned a lot about what analytical skills are desired from interacting with folks from different walks of life. Our intent with the book was to share a framework for problem solving rather than presenting a bag of tricks, and hopefully a comprehensive approach that would help busy people do well.

The folks at Rutgers University's Department for Continuing Professional Education asked us to create and teach a statistics workshop for food scientists several years ago. The interactive course we developed, "Making Sense of the Numbers," which we have now delivered for a number of years, has served as the basis for this book. We are thankful to the organizers and the participants for allowing us to develop and instruct the course and to "practice-write" the chapters of this book.

Just as important is the assistance and advice that we received from our friends and colleagues at Kraft Foods and Mondelēz International. Rob Glennon was an early reader of our text and provided critical feedback that makes this book more useful and readable. His experiences, strong grasp of reality, wonderful anecdotes, and the invaluable lessons he shared over the years have helped make the book far more readable and engaging. In addition to Rob, the statisticians who we have worked closely with over the years have taught us much and helped us to provide great examples. These statisticians include Keith Eberhardt, Sandra Echols, Addagatla Babu, Jeff Stagg, Gael Le Barzic, Fernando Rosa, Maria Teresa Lopez, Anja Schleppe, Marty Matzinger, Eric Lagergren, Boris Polsky, Yvonne Wang, Stanislav Zakharkin, Trevor Duguid Farrant, and Kamaljit Sagoo. In these pages, you will find the fruits of their work with us over the years. Shamsheer Mahammad is a talented Rheologist at Kraft Foods who helped us greatly with measurements from and advice about from the Brookfield viscometer. Any errors about that measurement system are ours alone even after his patient advice.

Two additional people played a very large role who we would like to specially thank. Donka Mirtcheva provided critical feedback to every chapter, and the book is better for her suggestions and creative ways in framing issues and addressing challenges. Kremena Rouhliadkova created the book cover design that is original, attractive, and enticing to potential readers, and this alone should increase the book's popularity.

And, above all, we would like to thank our families for letting us progress on this writing adventure.

From Viktor:

To my precious wife, Silvia Mirtchev, thank you for your encouragement to write this book. I could not have done it without your unconditional support and understanding; or the love of our children, Maxim and Viano. You are a treasure and without your comfort and cheer over the years and during countless several late nights, while I was writing my heart out, I would have never been able to complete this project.

From Frank:

To my wife, Aylesa Singley, thank you for your continued inspiration and encouragement throughout this project and my career. I'm sure I could not have done this without you!

CHAPTER 1

Introduction

"Maria accepted the job offer, and she will be joining your team in two weeks!" Joan had spent 7 months on a quest to find a food scientist, and this short voicemail from Human Resources put an end to his efforts. He was so excited that he stepped out of his office to share the great news with Steve, a senior food scientist, working for him. Steve had also enjoyed meeting Maria and shared Joan's high opinion of her from the recent university job fair.

Maria, a recent graduate holding a Master's degree in food science, is a bright, young professional who is not only excited, she has already posted the good news on Facebook and Twitter and has texted all of her friends about her decision to join the Ultimate BBQ company.

"Help is on the way," said Joan, jokingly, to Steve. "Maria has accepted, and her strong critical thinking and analytical skills will help us improve the quality of our BBQ sauce."

Steve: "BBQ!" BBQ sauce has been one of the company's core products, and it has been around for over 100 years. Sadly, in the most recent years, it has lost market share due to product quality degradation, a trend that the company hopes to reverse quickly to maintain a market leading position. "Will Maria be equipped for such a big first project?"

Joan: "I agree with you Steve, and that is why I would like to ask you to be Maria's manager. Help her understand the complex process of making BBQ sauce, the similarities between our pilot plant and plant production, and introduce her to the rest of the team. As you know 'it takes a village' to make good BBQ."

"There have been a lot of recent complaints indicating that our BBQ sauce is not thick enough. Our consumers care deeply about our product. They have given us a second chance through the gift of feedback, and now we owe it to them. Steve, I cannot think of a better person than you not only to pass on a lot of BBQ sauce institutional knowledge to Maria, but also to help her to focus on the areas that need the most attention. You have worked closely with James in Marketing Research to understand the specifics of the consumers' feedback from the BBQ users' focus group sessions."

Statistics for Food Scientists
http://dx.doi.org/10.1016/B978-0-12-417179-4.00001-9

Steve: “I would love to help Maria. Since this will be a highly analytical project, I’ll introduce her to Chad in Analytical Sciences, Louise in Quality, and the other subject matter experts that she will need to work closely with. I’ll also organize a visit to both of our BBQ plants after she is situated in the Research and Development center.”

In the spirit of good planning and to better prepare for a successful relationship with her, Steve decides to venture into the virtual space for traces of Maria. He revisits her LinkedIn profile and learns from it that Maria relishes challenges, never gives up, and most of all she enjoys working on teams. She seems to have some data analysis skills that Steve hopes he can learn from. At the same time, her profile suggests that she is interested in learning more about process capability analysis, which has been an area of focus for Steve in recent years. As Steve prepares to weave in the relationship between product concepts and analytical approaches, he realizes that he will need the help of others to enable Maria to see deeply into what is best for BBQ sauce.

Steve wants to help Maria try new approaches and statistical practices. Only after she sees some successes will she add them into her repertoire of statistical tools. These will enable her to develop critical practices to shape thought processes and transform the way scientists approach everyday quality and product development challenges. These practices will be relevant to many challenges that food scientists face, and Steve wants Maria to have a successful start into her corporate career. Now, onto Day 1…

CHAPTER 2

Descriptive Statistics and Graphical Analysis

One of the first objectives for Maria as a new employee is to develop an understanding of the BBQ sauce manufacturing process. Fortunately, the research and development center where she works is part of a larger facility that includes the main BBQ sauce production facility for her company. Steve arranges for her to visit the production facility in her second week. Lisa, the plant manager, will give her a tour.

Maria arrives at the facility a few minutes before the meeting is scheduled. Though only a 5-min walk from the main office building, she feels that she has entered another world. As she waits for Lisa to meet her, she can see the busy activity through the glass walls of the waiting room—bottles of sauce rattling down lines, forklifts moving pallets of materials, uniformed workers in hairnets, earplugs, and safety glasses bustling about. What a change from the quiet research labs where she spent her first week at her desk and in meetings!

When Lisa arrives, Maria is surprised to find that she appears to be quite young for someone who has responsibilities for a manufacturing facility. Short, trim, and completely unadorned, Lisa displays a confident manner.

"We are very happy to have you on board Maria," Lisa says as she vigorously shakes Maria's hand. "We were so excited when you accepted the position with us. Like you, I started in research right out of school. I never would have thought I would manage production one day."

"I see you are already wearing your safety shoes and lab coat. You will also need to remove all of your jewelry before we enter the plant," Lisa says as she guides Maria to a room with lockers and a dispensary for the gear required for entering the facility. She gives a hairnet and a set of earplugs to Maria and runs down a list of safety rules for the tour. "Keep your hands away from most everything and stay within the yellow lines on the floor and you will be fine," Lisa concludes.

When she enters the production floor, Maria is surprised at the noise level. Behind the glass in the waiting room, she saw the high level of activity, but she was unprepared for the accompanying noise.

Statistics for Food Scientists
http://dx.doi.org/10.1016/B978-0-12-417179-4.00002-0

"We do a lot of shouting around here," Lisa says as they start walking. "With the noise and the earplugs there is no other way!"

The next hour is a blur for Maria, as she encounters the many steps of BBQ sauce production. Ingredients arrive at a receiving dock and are moved by forklift to an internal storage facility within the plant. They are moved again to the manufacturing lines as they are needed. The sauce production starts with all ingredients mixed in an extremely large vat. Mixing continues as the temperature on the vat is increased. When the vat reaches a specific temperature, it is evacuated. The sauce moves through pipes to be filled into bottles. Then the bottles are capped. Although there are a number of production lines, they look similar, if not identical. The process seems straightforward enough, but Maria suspects that it may be more complicated than it appears.

Maria sees that on each line workers periodically pull bottles from the production line, and mark the date and time on the label. The bottles are placed on a table near the operator station.

Maria points and asks Lisa, "What is going on there?"

"The operators pull samples periodically to monitor product quality," Lisa responds. "Once an hour, the technician from the Quality Lab collects the bottles from all of the production lines and measures the viscosity. It is our most important quality measure. Consumers expect the sauce to cling to the meat as it cooks."

At the end of the tour, Lisa brings her to a stop and proudly exclaims, "I have saved the best for last! We just finished installation of our newest filling line a few weeks back. This is state-of-the-art technology that can fill at nearly twice the rate of our other lines. The viscosity should also be more consistent."

Maria looks on as Lisa points out the differences between the existing and new equipment. But one thing she sees again is the operator periodically pulling bottles from the line and marking the date and time, as on the other lines. Following Maria's eyes, Lisa comments, "We measure viscosity on the new line too. We collect samples as frequently as the other lines, but since the output from the line is much greater we may need to change our sampling strategy for this line."

As the tour concludes, Lisa mentions to Maria, "I hear that you have some data analysis skills from your university training. Can I have you look at the production data from the new and existing lines to verify the performance improvement on the new line?"

Maria eagerly agrees, "I did have terrific experience learning about statistics and data analysis working with my graduate advisor, Dr Wang. I'd be happy to take a look at the data."

Lisa agrees to organize and email the production data to Maria. "I'll take a look at the data and set up a time to discuss my findings," Maria promises.

Within a day, Maria receives an email from Lisa with a spreadsheet attached. The text of the email reads as follows:

> *It was nice meeting you yesterday. ...I have included viscosity data from the new line, line 6, and the best performing line from our existing production lines, line 4, for two days of production last week. Please let me know what you find.*

The first few rows of attached spreadsheet of data look like this:

Date and Time	Line 4 Viscosity	Date and Time	Line 6 Viscosity
4/28/2013 8:06	4776	4/28/2013 8:05	3746
4/28/2013 8:14	4318	4/28/2013 8:13	4424
4/28/2013 8:23	4363	4/28/2013 8:22	4284
4/28/2013 8:31	4447	4/28/2013 8:30	4241
4/28/2013 8:39	3832	4/28/2013 8:38	4132
4/28/2013 8:48	4426	4/28/2013 8:47	4189
4/28/2013 8:56	4516	4/28/2013 8:55	4650
4/28/2013 9:04	5085	4/28/2013 9:03	4303
4/28/2013 9:13	4585	4/28/2013 9:12	4252

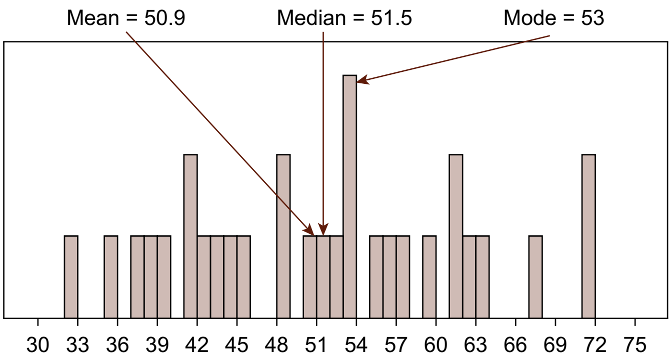
There are 116 data points for each of the two lines, each data point with a time stamp. For both lines, the data appear to come in roughly 8–9 min intervals.

Maria recalls from her work with Dr Wang a number of summary statistics that can be used to characterize the production data for lines 4 and 6. Having brought to the office most of her college textbooks, she locates and opens the introductory statistics textbook she used in Dr Wang's class. The book describes a list of summary statistics and includes illustrations. The list comprises two groups: measures of the center and measures of the spread of a set of numbers.

Three measures of the center are the mean, median, and mode:

1. Mean = This is the mathematical average; add up all of the values and divide by the number of observations.
2. Median = The median is the middle value of a set of numbers arranged in order from smallest to largest. This is the 50th percentile.
3. Mode = This is the most frequently occurring value in a dataset.

There is an illustration in the book that shows each of these measures for a given dataset. The illustration is a histogram, a graph depicting the distribution of the data values. Histograms group the data into a set of bins representing a subset within the range of the data. The frequency or percentage of occurrences in each bin is represented on the vertical axis. The illustration below identifies the three measures, which happen in this case to be close to one another.



Further reading describes the advantages and disadvantages of the measures. While the mean uses all of the individual values in its calculation, it can be influenced by extremely large or small values. The median and mode are not influenced by extreme values since they do not use each individual value. The mode may not be unique, and it may not reflect the center of the distribution very well.

Four measures of the spread are the variance, standard deviation, range, and interquartile range:

1. Variance $= \frac{\sum_{i=0}^{n}(x_i - \overline{x})^2}{(n-1)}$
2. The standard deviation is the square root of the variance.
3. Range = (largest value − smallest value).
4. Interquartile range (IQR) = the middle value of the second half of the data—the middle of the first half of the data when the data is arranged in order from smallest to largest. In other words, it can be represented as follows:

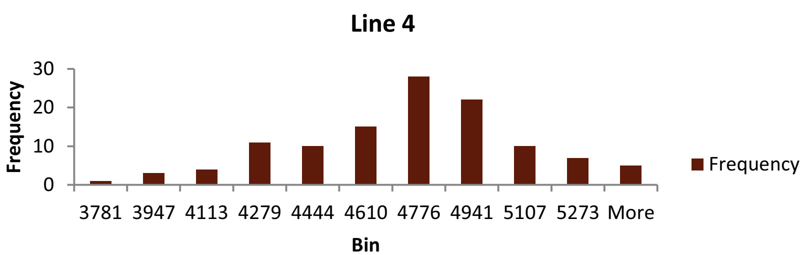
$$\text{IQR} = (\text{the value at the 75th percentile} - \text{value at the 25th percentile})$$

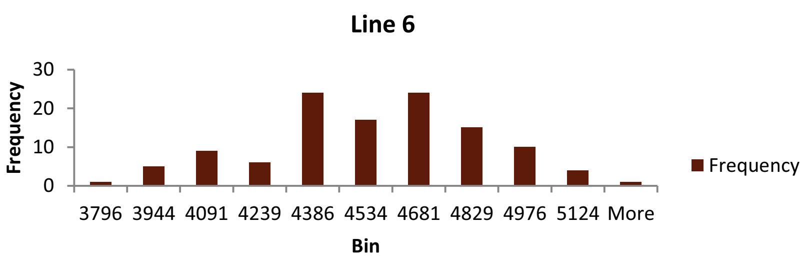
These measures of spread for the data in the illustration above are as follows: the variance is 112.36, the standard deviation is 10.6, the range is 39 (71–32), and the interquartile range is 17.75 (59.5–41.75).

These measures, too, have advantages and disadvantages. Like the mean, the variance and standard deviation use all of the individual values in their calculation, but they can be influenced by extreme values. The range is very easy to understand but does not use each value and is affected by extreme high or low values. The interquartile range ignores half of the data but is not affected by extreme values.

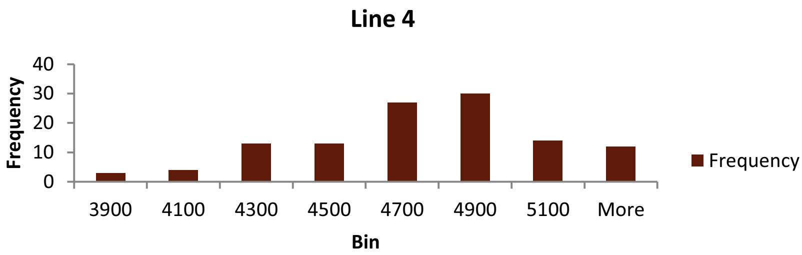
Maria had used the data analysis functions in Microsoft Excel to create summary statistics in Dr Wang's class. She recalls that she had installed the Analysis ToolPak add-in for Excel to accomplish this and will need to do so in preparation for creating descriptive statistics for the viscosity data. Dr Wang was a strong believer in looking at data in graphs before performing any statistical analyses. One graph she had used previously is a histogram like the one illustrated in her book. Comparing histograms for each of the product lines should help in understanding if they perform differently. Histograms are also created using the Analysis ToolPak in Microsoft Excel.

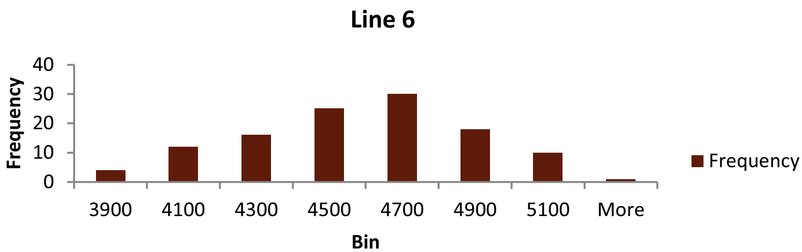
Maria arranges the histograms she creates in Excel to compare the lines. They do indeed look to be symmetric and bell-shaped. However, she quickly notes that Excel chose to use different bins for each line, so the scale is different on each graph.





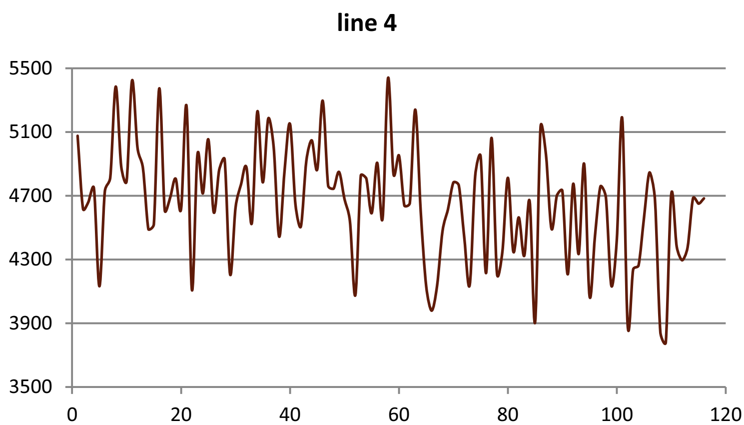
This makes it difficult to compare the lines. It is not apparent if the variability is similar for each line, or if they have similar centers. Excel is not making it easy for her to look at her data in the way that makes most sense to her. Looking back at the pop up menu from which she created the histograms, she sees that she has the ability to enter in the bin range, and with a bit of effort, she comes up with a range that represents the data from both lines reasonably well.

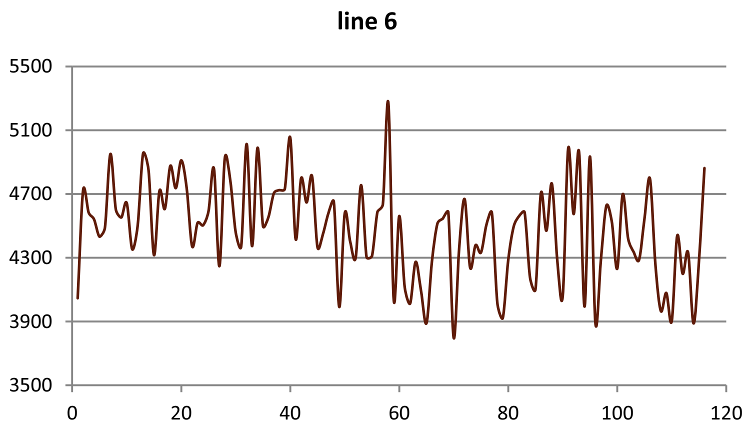




When Maria arranges the histograms to compare the lines, it does appear that the two lines perform differently. The center for line 6 seems somewhat lower in viscosity. The variability in line 6 seems to be smaller but not to a great degree.

It occurs to Maria that a different plot would indicate if there is some sort of increasing or decreasing trend in the data as it was collected over time. She uses Excel to graph the data in time sequence to see if there is an increasing or decreasing trend over time. She has difficulty using the date and time column for the horizontal axis and realizes that creating the plot without this will use the data point sequence number on the horizontal axis. This is sufficient for her purposes, since the data points from both lines are similarly spaced in time. Her graphs have viscosity on the vertical axis and time point sequence on the horizontal axis.





Maria does not observe an increasing or decreasing trend when the data is displayed in sequence, but the graphs do reinforce her earlier observation from the histograms that both the center and the variation in line 6 are smaller.

Maria has no problem creating the summary statistics for the two lines. The output from Excel not only includes the measures of the center and spread that she just read about, but it also gives her a number of additional statistics measures, as shown in a table that looks like the following:

Viscosity Line 4

Mean	4662.86
Standard error	32.95
Median	4686.60
Mode	#N/A
Standard deviation	354.90
Sample variance	125,955.02
Kurtosis	−0.09
Skewness	−0.20
Range	1657.19
Minimum	3781.39
Maximum	5438.59
Sum	540,891.86
Count	116

Viscosity Line 6

Mean	4480.46
Standard error	28.02
Median	4505.23
Mode	#N/A
Standard deviation	301.75
Sample variance	91,052.26
Kurtosis	−0.35
Skewness	−0.10
Range	1475.20
Minimum	3796.18
Maximum	5271.38
Sum	519,733.58
Count	116

Maria digs back into her statistics textbook again to re-familiarize herself with the additional measures included in the Excel output. For both lines, the mode is listed as N/A. She thinks this is because all of the data values are unique. She recalls that the variance is simply the square of the standard deviation. She is not sure what the sum of the values may tell her. She learns that the standard error is the standard deviation divided by the square root of the sample size, and it is an estimate of the standard deviation of the mean, though she is not sure how that may be useful to her. And while the range is listed in the output, she is disappointed that the interquartile range is not.

Two new terms for her are skewness and kurtosis. Checking back in her statistics text, she learns that these describe the shape of the distribution of the data.

Skewness quantifies the extent to which the data is not symmetric—zero indicates symmetry, a negative number indicates a skew to the left side, and a positive number indicates a skew to the right side.

Kurtosis indicates the flatness of the distribution of the data compared to a bell shape—numbers above zero indicate a flatter distribution, and numbers below zero indicate a more peaked distribution.

For the viscosity data the skewness and kurtosis values are quite small, indicating that the distributions of the viscosity data from both lines can reasonably be considered symmetric and bell-shaped.

Maria arranges her descriptive statistics summaries and histograms in a report to present to Lisa. Dr Wang always stressed making things clear and concise when putting a report together. Maria decides to remove some of the descriptive statistics that she does not think will be useful when speaking to Lisa, and she arranges the table of these statistics for a more direct comparison of the production lines. She also formats the values so that there are not too many values after the decimal point, which contain no information.

	Viscosity Line 4	Viscosity Line 6
Mean	4663	4480
Median	4687	4505
Standard deviation	355	302
Kurtosis	−0.09	−0.35
Skewness	−0.20	−0.10
Range	1657	1475
Minimum	3781	3796
Maximum	5439	5271

Maria feels confident that she has information that Lisa will find interesting. She sets up a meeting to discuss the results with Lisa for the following day.

"Take a look at the analyses that I created for you on the two lines," Maria says as she spreads her analysis sheets on the conference room table. "It seems clear that the two lines are operating differently."

Lisa examines the sheets before her. She looks puzzled, as she is sure that the new line is performing better. After a short time she comments:

"The means and medians for the two lines look to be pretty close; I'm not sure I agree that the lines are performing that differently in that respect. The mean for line 4 is 4663; for line 6, it is 4480. There does not seem to be much of a difference given the scale of the data. And the standard deviation of 302 for line 6 is definitely smaller than the 355 for line 4. But maybe I'm seeing these results as I want them to be. Is there a way to make a definitive statement about differences between the lines?"

Maria realizes that she could have done more in analyzing this data. She recalls from her work with Dr Wang that she can more definitively compare the lines and tells Lisa that she will perform some additional analyses to get an answer.

This page intentionally left blank

CHAPTER 3

Hypothesis Testing

Lisa is right! As a plant manager, she ought to be able to demonstrate that the investment made in a new line is a good investment. Alternatively, if the new asset brings no additional, or only marginal, benefit why should she upgrade the rest of the lines? Lisa needs more than a summary of line performances. She should be able to defend the case with confidence that the asset investment has merits. She ought to be able to clearly answer the question: Is there a significant difference between line 4 and line 6 viscosities?

Maria learned about statistical hypothesis testing in one of Dr Wang's classes. By using a statistical hypothesis test, she can test the hypothesis that line 4 and line 6 perform similarly in terms of finished product viscosity. Thus, she will be able to address Lisa's concerns if there are viscosity differences between the two lines.

Since Maria already has the data files, she decides to run a hypothesis test to answer Lisa's questions. She looks at the options in the data analysis add-in and is a bit uncertain which method is most appropriate. In Excel, she can choose the following:

1. *t*-test for two samples assuming equal variances
2. *t*-test for two samples assuming unequal variances
3. *t*-test paired for two-sample means
4. *F*-test for two-sample variances

She remembers the discussion of paired data in her class—the two production lines run independently, so test 3 can be eliminated. Since the difference between the remaining two *t*-tests hinges on the assumption of equal or unequal variances, it seems best to test the variances first, and then choose the appropriate *t*-test to use.

The test of the sample variances answers the question: Is there a difference in viscosity variability between the lines? It is easy to perform

Statistics for Food Scientists
http://dx.doi.org/10.1016/B978-0-12-417179-4.00003-2

this test in Excel; simply select the data points for each of the lines and elect the "*F*-test for two sample variances."

F-Test Two Sample for Variances

	Line 4	Line 6
Mean	4663	4480
Variance	125,955	91,052
Observations	116	116
df	115	115
F	1.38	
$P(F \leq f)$ one-tail	0.04	
F critical one-tail	1.36	

Maria recalls that the p-value (listed in the output as $P(F \leq f)$) is an important part of the output for her to focus on, and that small values indicate that the variances of the two lines are different. Dr Wang had used probabilities of 0.05 as a cutoff value. The p-value of 0.04 in the output is smaller than 0.05, so she can conclude that there is a statistically significant difference in the viscosity variance between line 4 and line 6. Since the variance is smaller for line 6, the data indicate that line 6 delivers a more consistent product. This finding is what Lisa had expected, so Maria is excited to deliver the news.

The t-test for sample means answers the question: Is there a statistical difference in the average viscosity between lines 4 and 6? Since the variances are different for the lines, Maria chooses test 2 that assumes unequal variances.

t-Test: Two Sample Assuming Unequal Variances

	Line 4	Line 6
Mean	4663	4480
Variance	125,955	91,052
Observations	116	116
Hypothesized mean difference	0	
df	224	
t stat	4.22	
$P(T \leq t)$ one-tail	0.00	
t critical one-tail	1.65	
$P(T \leq t)$ two-tail	0.00	
t critical two-tail	1.97	

Again, Maria looks at the p-value listed as $P(T \leq t)$ as one of the important output pieces for her to focus on. She sees that there are two p-values in the output: one-tail and two-tail. She does not recall the difference between the one- and two-tail tests, but because both values are very small, the conclusion is the same; the average viscosity of the lines is statistically significantly different. The average of line 6 is smaller, so Lisa will be alarmed as a more viscous product is desirable.

Maria is not confident that she has used the right statistical analyses. Did she read the results correctly? She has to be sure about her findings before she reaches out to Lisa. She pulls out a textbook from Dr Wang's class and puts together a cheat sheet with definitions for future reference.

Definitions Cheat Sheet:

1. Null hypothesis (H_0): By convention, H_0 is almost always a statement of no change or difference. It is assumed true until sufficient evidence is presented to reject it.
2. Alternative hypothesis (H_a): Statement of change or difference. This statement is considered true if H_0 is rejected.
3. Type I error: This is the error of concluding that there is a difference when, in fact, there is no difference.
4. Type II error: This is the error of concluding that there is no difference when there really is a difference.
5. Alpha (α) risk: This is the maximum risk, or probability, of making a type I error, sometimes also called a significance level. This probability is always greater than zero, and is often set at 0.05 (5%).
6. Beta (β) risk: This is the maximum risk or probability of making a type II error. This is often set at 0.20 (20%).
7. Confidence level ($1 - \alpha$): If the α-risk is 5%, then the confidence level for making the right decision is 95%.
8. Power ($1 - \beta$): This is the probability that a statistical test will detect a difference when there really is one.
9. Significant difference: This is the term used to describe the results of a statistical hypothesis test, where a difference is too large to be attributed to chance.
10. p-value: This is the probability that an observed difference is due to chance alone. Observed p-values equal to α or less are considered evidence to reject the null hypothesis.

A hypothesis test is a statistical means of answering a yes/no question. For many hypothesis tests, the yes/no question will be: *Is there a difference?* Maria is interested in ensuring that she is crystal clear on three important hypothesis-testing concepts.

The first key concept about hypothesis testing is that of comparison. Is the researcher comparing observed data to a standard (or benchmark) or directly comparing observed data from two populations (such as lines, processes, batches, etc.)?

1. A hypothesis test comparing observed data to a standard would look like this: Is the average viscosity for line 6 different than 4000?

   H_0: Line 6 average viscosity = 4000

   H_a: Line 6 average viscosity $\neq$ 4000

2. A hypothesis test comparing observed data from two populations would look like this: Is the average viscosity for line 4 different than the average viscosity for line 6?

   H_0: Line 4 average viscosity = Line 6 average viscosity

   H_a: Line 4 average viscosity $\neq$ Line 6 average viscosity

   Another way to specify these null and alternative hypotheses is as follows:

   H_0: Line 4 average viscosity − Line 6 average viscosity = 0

   H_a: Line 4 average viscosity − Line 6 average viscosity $\neq$ 0

The second key concept about hypothesis testing is that of direction. Are we only interested in one side of the comparison or both?

1. A hypothesis test with an one-sided comparison would look like this: Are line 4 average viscosities greater than line 6 viscosities?

   H_0: Line 4 average viscosity $\leq$ Line 6 average viscosity

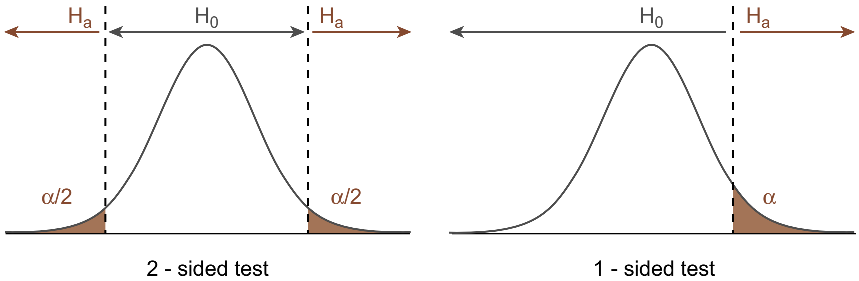
   H_a: Line 4 average viscosity $>$ Line 6 average viscosity

2. A hypothesis test with a two-sided comparison would look like this: Is the average viscosity for line 4 different than the average viscosity of line 6?

   H_0: Line 4 average viscosity = Line 6 average viscosity

   H_a: Line 4 average viscosity $\neq$ Line 6 average viscosity

In other words, the alternative hypothesis (H_a) could be specified as less than, not equal, or greater than, while the null hypothesis (H_0) is by convention equality. Many tests are set to check for any difference, either greater than or less than equality. These are called two-sided or two-tailed tests because the tests check for differences on both sides of the equality. When the alternative hypothesis is focused on only one side, the test is called one-sided or one-tailed test. In the two-sided test, the α-risk is split in half to account for both sides of the distribution. In the one-sided test, the entire α-risk is concentrated on a single side, which gives more statistical power than in the two-sided test to distinguish a difference on that side. However, any difference on the opposite side is indistinguishable from the null hypothesis. The graphs below illustrate how α-risk is divided to accommodate both sides of the equal sign.



The third key concept about hypothesis testing is whether the alternative hypothesis is specific or general. The examples so far have all been general—they do not specify the size of the difference in the alternative hypothesis. An example of a specific alternative hypothesis is as follows:

1. Is line 4 viscosity greater than line 6 viscosity by at least 400?

H_0: Line 4 viscosity − Line 6 viscosity = 0

H_a: Line 4 viscosity − Line 6 viscosity > 400

Failing to reject the null hypothesis does not necessarily mean that the alternative hypothesis is accepted. If an alternative hypothesis is not specified, then the conclusion drawn for a given sample size, is that there is not sufficient evidence to reject the null hypothesis and accept the alternative hypothesis.

Once appropriate null and alternative hypotheses and a one- or two-sided test, along with a specified significance level (α) are established, then a p-value is calculated from the collected data. Two possibilities can occur:

1. p-value $< \alpha$ => reject the null hypothesis.

2. p-value $\geq \alpha$ => fail to reject the null hypothesis. If an alternative is specified and sample size based on specified alpha and beta risks is determined, then we accept the alternative hypothesis.

Maria recalls doing the math to calculate a test statistic and estimating p-value from a statistical table while in school. Calculating an exact p-value is a lot simpler using Excel. Nevertheless, Maria expands her definitions sheet with some formulas for calculating the test statistics:

1. One sample t-test:

$$t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$$

2. Two sample t-test:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S_{X_1X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad S_{X_1X_2} = \sqrt{\frac{(n_1 - 1)S_{X_1}^2 + (n_2 - 1)S_{X_2}^2}{n_1 + n_2 - 2}}$$

3. Test for equal variances:

$$F = \frac{{S_1}^2}{{S_2}^2}$$

Depending on the question of interest, there is a variety of test statistics. Based on what is being compared and the data type (continuous or discrete), a number of different tests can be deployed.

In the case of continuous measures, for the two sample t-test and two variances F-test, there is an assumption that the data for both samples follows a normal distribution, that is, the histogram follows a bell shape. All the data that Maria has collected to date have followed a bell shape. When executing her tests, Maria had assumed that the data follow a normal distribution, and the histograms confirm that.

Armed with her findings, definitions sheet, and refreshed concepts, Maria feels confident reviewing the results with Lisa. She picks up the phone and calls Lisa.

Maria: "I'm very confident in stating that line 6 produces more consistent product. However, it is consistently lower in viscosity."

Lisa, who has also done some investigative work on the floor, has a hard time believing the results and thinks that the findings may be a fluke. She is convinced that there must have been a phenomenal day on line 4 or a really bad day on line 6 when the samples were collected.

Lisa: "What if we compare all plant lines to obtain a complete picture of the production?"

CHAPTER 4

Analysis of Variance (ANOVA)

Eager to continue with her work on BBQ sauce viscosity, Maria arrives at her office early the next day thinking she would look into statistical analysis techniques for comparing more than two samples before the additional data arrives. Her plan is to start with the textbook from Dr Wang's class. It had occurred to her that she could certainly perform *t*-tests as she had done before, one for each pair of lines. With five lines there are quite a few pairs to compare: line 2 versus line 3, line 2 versus line 4, and so on. She lists them all on a sheet of paper and counts 10 pairs. She has already performed the *t*-test for line 4 versus line 6, so nine more *t*-tests would be required. And then, she would have to compare the results of all 10 tests. It looks to be a lot of work even when using Excel. Maybe there is a more efficient way to tackle this problem.

She is surprised to find an email from Lisa with the data from three additional lines. The text of the email reads:

I have included viscosity data from lines 2, 3, and 5 to compare with the data you have for lines 4 and 6. I can't wait to see what you find!

"That woman must never sleep! With all of the things she is responsible for at the plant, I can't believe that she was able to get this information to me so quickly," Maria exclaims.

The data files look similar to the files she had received earlier. The first few lines in the data file look like this:

line	Date and Time	viscosity
2	4/28/2013 8:09	5109
2	4/28/2013 8:18	4978
2	4/28/2013 8:26	4929
2	4/28/2013 8:34	4916
2	4/28/2013 8:43	4430
2	4/28/2013 8:51	4822
2	4/28/2013 8:59	5006
2	4/28/2013 9:08	3987
2	4/28/2013 9:16	4487
2	4/28/2013 9:24	4219
2	4/28/2013 9:33	4494
2	4/28/2013 9:41	4633

Statistics for Food Scientists
http://dx.doi.org/10.1016/B978-0-12-417179-4.00004-4

There is now a column indicating the line where the product was produced. Maria checks and sees that there is the same number of data points from the same date as the previous files.

After taking care of a few minor tasks, Maria opens her textbook and scans the table of contents. She finds a chapter that contains the *t*-test that she performed. Two chapters after this, she finds another chapter titled "Comparing *k* Means—One-way Analysis of Variance (ANOVA)." Guessing that this may be what she is looking for, she turns to that chapter and begins reading.

Maria quickly learns that analysis of variance (often abbreviated as ANOVA) could be used for her comparison of the production lines. Its name relates to the concept that the variability in a set of data can be broken into different components. In the simplest form, there are two components:

1. Variability between the factor level means
2. Variability of the individual values within each factor level

Factors designate the group of things being compared; in her case, these are the production lines. Levels are the names of the elements in the factors; in her case, these are the line numbers 2 through 6. An illustration in the book demonstrates these concepts:

	Factor		
sample	**A**	**B**	**C**
1	90	94	100
2	92	96	115
3	88	98	97
mean score	90	96	104

Within Factor Level

Between Factor Levels

The statistical test for differences between the factor level means uses a ratio of these two variability components, with the variability between factor levels in the upper part of the ratio and the pooled within factor level variability in the lower part of the ratio. A larger ratio suggests treatment differences.

For her application, the factor levels would be the different production lines. Maria learns that if there are only two factor levels, ANOVA is identical to the *t*-test that she has performed. ANOVA is just an extension of the analysis she already completed. The technique will compare the variability between the lines and the variability within the lines.

Looking ahead in the book, she sees that the next chapter is titled "Multi-way Analysis of Variance." The ANOVA technique is quite flexible and can accommodate more complex problems, in particular additional factors. The example in the book describes a problem investigating different drugs (factor 1) for both male and female patients (factor 2).

Within Factor Level Combinations

		Factor 1			Mean Score
		A	B	C	
Factor 2	M	90	94	100	
		92	96	115	97
		88	98	97	
	F	95	99	108	
		92	95	123	101
		91	101	107	
Mean score		91	97	108	

Between Factor 2 Levels

Between Factor 1 Levels

This gets Maria thinking about the data Lisa provided her. It is from one production day, but from both the day and evening production shifts. There are different operators for the lines on each shift, and they may run the machinery differently. Perhaps she should consider the shift as an additional treatment in her analyses.

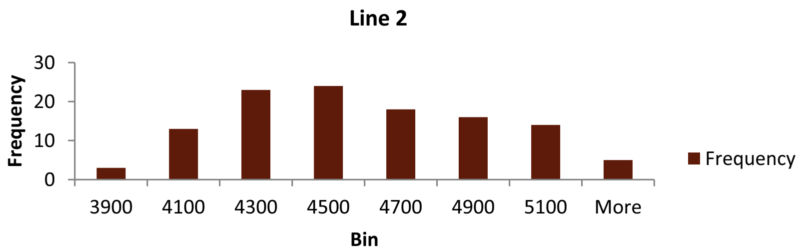
"Maybe I should not bite off more than I can chew," Maria thinks aloud. "Let me start with the simpler one-way ANOVA to compare the lines and see what I can learn from that analysis."

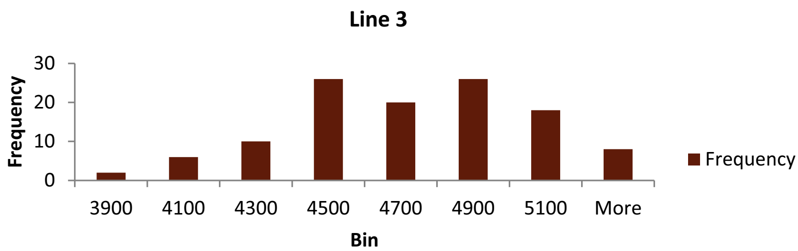
Maria sees that in the Analysis ToolPak add-in, there are three choices for ANOVA: Single Factor; Two-Factor with Replication; and Two-Factor without Replication.

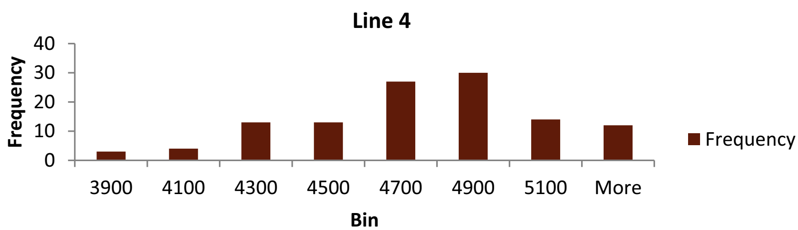
Maria arranges the data in Excel so that the viscosities from the five lines are in separate columns. The first few lines of the rearranged data sheet look like this:

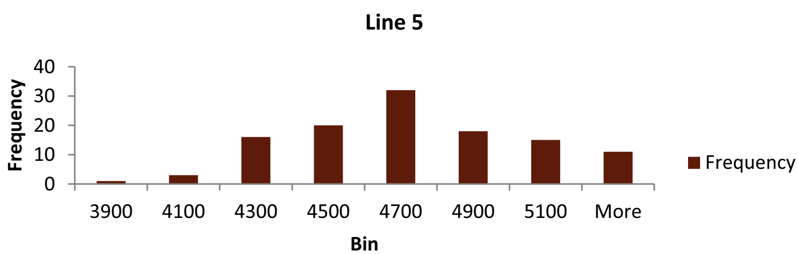
Line 2	Line 3	Line 4	Line 5	Line 6
5109	5172	5076	5289	4046
4978	5042	4618	5059	4724
4929	4796	4663	5125	4584
4916	4784	4747	4353	4541
4430	4840	4132	4807	4432
4822	5423	4726	4258	4489
5006	4918	4816	5113	4950
3987	4731	5385	4789	4603
4487	5070	4885	4801	4552
4219	4621	4793	4355	4643
4494	4423	5423	4260	4353
4633	4635	5001	5249	4500
4388	4473	4873	4784	4954
4425	4981	4490	4648	4849
4914	5066	4521	4565	4317
4607	4789	5374	4233	4719
4838	4586	4609	4688	4607
4179	5005	4689	4579	4876
4839	4948	4807	4028	4737
4269	4464	4617	4815	4911
4777	4666	5263	5421	4736
4221	4574	4108	4608	4372
5014	5048	4957	4776	4519

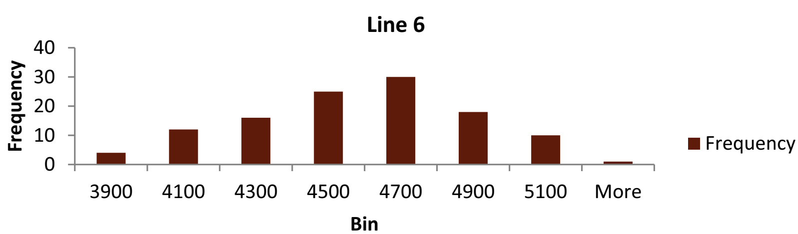
As she did when comparing only lines 4 and 6, Maria starts by making histograms of viscosity for each of the lines. Her graphs look like this:







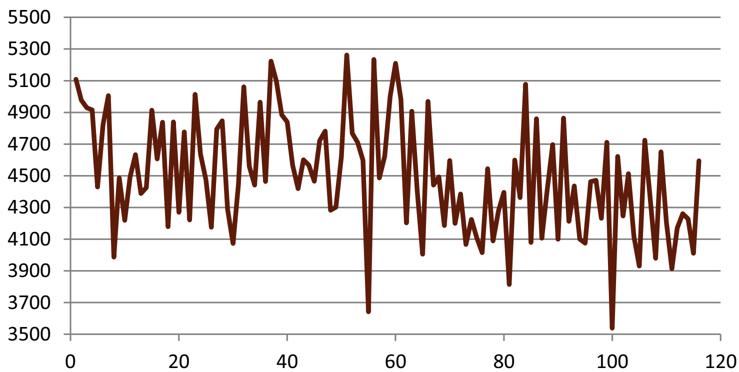




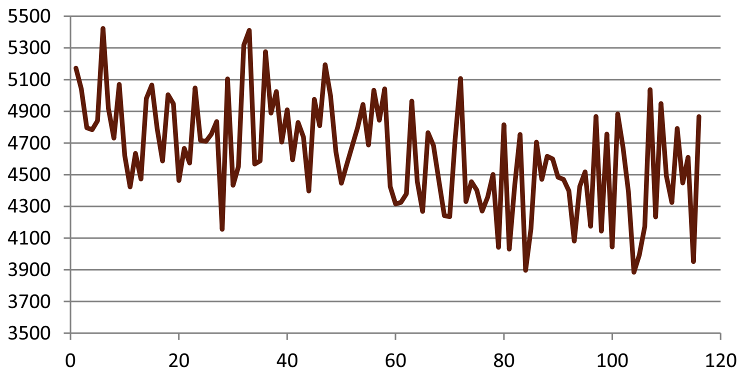
Maria notes that line 2 in addition to line 6 may have a lower average than the other lines. She also observes that the variability of line 6 still looks to be smaller than that of the other lines.

As she did when comparing only lines 4 and 6, Maria also makes plots of the viscosity data in time sequence for all five production lines. Her plots look like this:

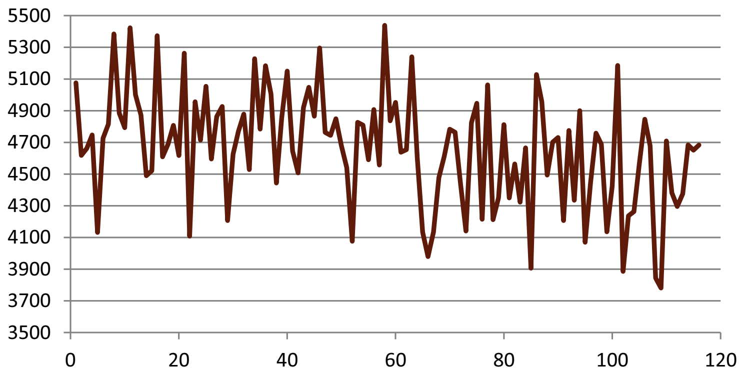
Line 2

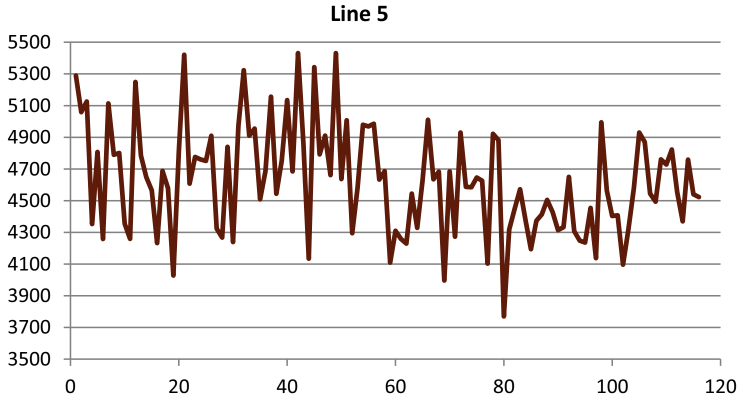


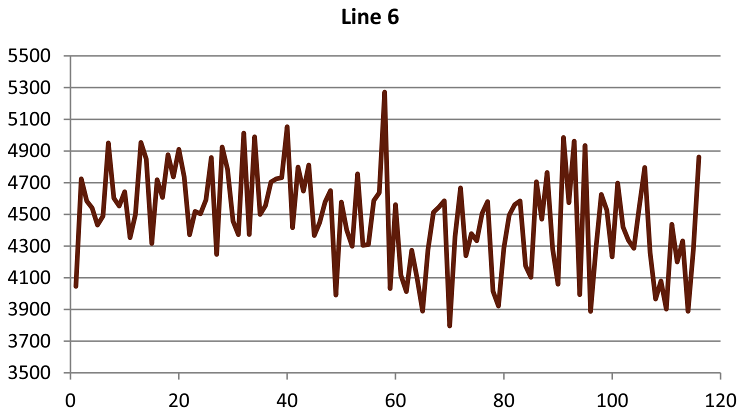
Line 3



Line 4







She does not see an increasing or decreasing trend for the lines except for line 3, which does seem to display diminishing viscosities over time. She makes a note to mention that to Lisa.

In the menu of her Analysis ToolPak add-in, she selects ANOVA: Single Factor. She selects the full input range and checks the box indicating that the column labels are in the first row. The analysis output that appears in a new work sheet looks like this:

ANOVA: Single Factor

Summary

Groups	Count	Sum	Average	Variance
Line 2	116	521,847	4499	129,823
Line 3	116	537,496	4634	111,617
Line 4	116	540,892	4663	125,955
Line 5	116	536,978	4629	113,420
Line 6	116	519,734	4480	91,052

ANOVA

Source of Variation	SS	df	MS	F	p-value	F crit
Between groups	3,325,273	4	831,318	7.27	0.00001	2.39
Within groups	65,764,690	575	114,373			
Total	69,089,962	579				

Most of the analysis details are easily understood. There is a table with summary statistics for each of the lines with the sum, average, and variance. The ANOVA table below the summary table contains a *p*-value for the between groups source of variation that seems to be the statistical test for the differences between the lines, exactly what she is looking for. She recalls that smaller *p*-values indicate significant differences. As the *p*-value here rounds down to zero, it would indicate the average viscosities are different between the lines.

"OK, now I'm getting somewhere!" Maria looks at the means for each of the five lines and sees that for line 3 and line 5, the means are pretty similar at 4634 and 4629. And line 2 and line 6 are also similar with means of 4499 and 4480. Line 4 has a mean of 4663, which is pretty close to the means for line 3 and line 5.

"Do I decide on my own what differences are significant? Lisa would want something more definitive I'm sure," Maria thinks out loud as she builds her analysis summary. "Maybe there is a way to determine this more objectively."

Maria goes back to her textbook to see if this question can be addressed. As she reads on, she comes across the concept of multiple paired comparison (MPC) procedures, which seems to be a way to determine individual treatment differences after an ANOVA analysis is performed. A number is calculated to determine the minimum significant difference between treatments. Any two treatments with a mean difference smaller than this number would not be considered different; any two treatments with a mean difference larger than this number would be considered different. There are several ways of calculating the number, and she is not sure which one to choose. The textbook mentions Fisher's least significant difference (LSD), Duncan's, Student-Newman-Keuls, Tukey's Honest Significant Difference (HSD), and Scheffe's as possible choices in certain situations. For some, the minimum significant difference is smaller and the significance test is more liberal. For others, the minimum significant difference is larger and the test is more conservative. But, she does see one statement that has relevance: If you are interested in all pairwise comparisons, you should use Tukey's HSD. She is interested in comparing all of the lines to each other, so she follows this advice.

Maria looks up MPC procedures in Excel and finds that none are available in the Analysis ToolPak add-in. "This is a problem; why not stop with the ANOVA analysis and try to determine the individual paired differences myself?" Maria grumbles in frustration. Fortunately the book describes the calculation for the minimum significant difference.

It uses a new tabled distribution that she is not familiar with: the studentized range. A value from this distribution is the q in this formula:

$$\frac{q_{k,\nu,\alpha/2}\, s\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}}{\sqrt{2}}$$

In this formula s is the square root of within groups variance from the ANOVA; $q_{k,\nu,\alpha/2}$ is the alpha upper significance level of the studentized range for k means; ν is the number of degrees of freedom for the within groups variance (labeled df in her ANOVA output); and n_i and n_j are the sample sizes for each of the treatments being compared.

Maria looks at the studentized range table in her textbook and sees that it stops at $\nu = 120$. The degrees of freedom for her analysis are 575, not even close to this! But, she also sees that there is a table entry for infinity, and that number is very close to the one for 120; so, for her degrees of freedom of 575, the entry for infinity should be OK to use. This value is 3.87 for an alpha = 0.05. Using this, she calculates the minimum significant difference to be 121.5.

Maria now examines the table of averages for each line. The two smallest averages are for line 2 and line 6 at 4499 and 4480, respectively. The difference between the two is smaller than 121.5, so the average viscosity of these two lines is not significantly different. Line 3, line 4, and line 5 have averages of 4634, 4663, and 4629, respectively. These three averages are all within 121.5 of each other and are then not significantly different. But, each of these three averages is greater than 121.5 from the averages of both line 2 and line 6. So, the lines form two groups: line 2 and line 6 have significantly smaller average viscosities than line 3, line 4, and line 5.

Maria adds these details to her analysis summary by creating a table ordering the lines by decreasing average viscosity. She feels confident now that she is prepared to meet with Lisa to deliver the results of her analyses. She is about to arrange a meeting for the following day when she remembers that she had thought about analyzing the data as a two-way

ANOVA with shift as the second treatment. The two-way ANOVA would add an additional dimension to the analysis that may be interesting.

Line	Average Viscosity	Group
4	4663	A
3	4634	A
5	4629	A
2	4499	B
6	4480	B
minimum significant difference	121.5	

Maria now starts working with the data to create a new column that indicates the shift that produced and collected the data based on the time stamp. All data from the first shift is from 8:00 AM to 4:00 PM. The second shift data is from 4:00 PM until midnight. Her data file now looks like this:

shift	Line 2	Line 3	Line 4	Line 5	Line 6
1	5109	5172	5076	5289	4046
1	4978	5042	4618	5059	4724
1	4929	4796	4663	5125	4584
1	4916	4784	4747	4353	4541
1	4430	4840	4132	4807	4432
1	4822	5423	4726	4258	4489
1	5006	4918	4816	5113	4950
1	3987	4731	5385	4789	4603
1	4487	5070	4885	4801	4552
1	4219	4621	4793	4355	4643
1	4494	4423	5423	4260	4353
…	…	…	…	…	…
2	5002	4426	4836	4110	4032
2	5209	4316	4953	4310	4561
2	4981	4325	4638	4260	4116
2	4203	4379	4654	4230	4013

In the menu of her Data Analysis add-in, she selects ANOVA: Two-Factor with Replication since she has repeated viscosity measurements for each line and shift combination. She chooses the full input range and indicates that there are 58 rows for each sample (Shift). The analysis output that appears in a new window looks like this:

ANOVA: Two-Factor with Replication

Summary	Line 2	Line 3	Line 4	Line 5	Line 6	Total
Shift 1						
Count	58	58	58	58	58	290
Sum	268,403	279,231	278,452	276,659	267,248	1,369,993
Average	4628	4814	4801	4770	4608	4724
Variance	109,425	70,285	95,764	119,372	61,583	97,848

Continued

ANOVA: Two-Factor with Replication—cont'd

Summary	Line 2	Line 3	Line 4	Line 5	Line 6	Total
Shift 2						
Count	58	58	58	58	58	290
Sum	253,444	258,265	262,439	260,319	252,486	1,286,953
Average	4370	4453	4525	4488	4353	4438
Variance	118,653	88,423	119,576	69,080	89,163	100,079
Total						
Count	116	116	116	116	116	
Sum	521,847	537,496	540,892	536,978	519,734	
Average	4499	4634	4663	4629	4480	
Variance	129,823	111,617	125,955	113,420	91,052	

ANOVA

Source of Variation	SS	d*f*	MS	*F*	*p*-value	*F* crit
Sample	11,889,085	1	11,889,085	126.30	0.000000	3.86
Columns	3,325,273	4	831,318	8.83	0.000001	2.39
Interaction	220,237	4	55,059	0.58	0.673685	2.39
Within	53,655,367	570	94,132			
Total	69,089,962	579				

In her analysis, the "Sample" source of variation is from the shifts and the "Column" source of variation is from the lines. This appears to be the standard naming convention in Excel. The analysis validates her earlier determination about the line differences since the *p*-value listed for "Columns" is smaller than 0.05. To her surprise, the analysis also indicates that there is a significant difference between the two shifts since the *p*-value for "Sample" is also smaller than 0.05. Shift 1 has a higher average viscosity than shift 2. She looks at the means for the two shifts for each of the production lines and sees that the differences between shift 1 and shift 2 for each of the lines seem to be much larger than the differences between the lines!

There is also a source of variation listed as "Interaction." She looks back at her textbook and finds out that a significant interaction is when the effect of one factor is not the same for different levels of another factor. In her analysis, the interaction source of variation is not significant. So this difference between shifts is consistent across the lines. Why would the viscosities be so different between the shifts?

Maria is now even more puzzled about the line performances at the plant. She is happy that she performed the additional analysis to account for shifts since that seems to be an important finding. She is certain she will need to defend this finding to Lisa when she shares the information.

Maria sets up a meeting with Lisa for the next morning. She carefully describes the steps in her analysis and how she has come to the conclusion that while lines do perform differently, there is an even bigger difference in viscosity for products produced between shift 1 and shift 2.

"I think that you have made an important discovery Maria," Lisa says after Maria presents her case. "I'm convinced that your analysis is thorough and correct from the details you have shown. But let's think this through some more. The data show clearly that the viscosities are consistently higher on shift 1 regardless of the production line. Although there are different operators on each line, the viscosity measurements for all lines are made in the quality lab by the same technician. The technician is different between the shifts, can we be sure that this is a true shift difference and not some sort of measurement problem?"

Maria realizes that Lisa is correct and that what she had assumed was a production difference could be a problem with the measurements. "I'll see what I can do to determine what is really going on here and get back to you with what I find," she promises Lisa.

This page intentionally left blank

CHAPTER 5

Measurement Systems Analysis (MSA)

Maria discovered large differences in performance of all production lines across the two production shifts, with generally consistent production within a shift. First shift production is consistently higher in viscosity. But, as Lisa pointed out, it is not clear whether the differences in viscosities between the shifts are real differences or due to the measurements being taken by different lab technicians.

Maria has never given any thought to the viscosity measurement process. She doesn't even know how the measurements are made. She just assumed that the measurements were accurate. She decides that an understanding of the measurement process is critical at this time.

She sets up a meeting with Chad, the lab supervisor, to have him explain the viscosity measurement process. He is happy to explain the measurement system, and he mentions that she is the first to demonstrate an interest. It turns out that the measurement process is somewhat involved.

Chad starts a lengthy description, "The measurement device is a Brookfield viscometer. This device is used for texture measurements in a wide variety of products. The device rotates a spindle that is submerged in a sample of the product. The force required to rotate the spindle at a specific speed, measured in centipoise (CP), is reported. Higher values indicate a thicker product. A large number of spindles can be used to perform measurements, since the device can measure equally well the texture of low viscosity products, such as salad dressing, and high viscosity products, such as mayonnaise. It is also used outside of the food industry for products like paints and oils."

Chad discusses with Maria the development of the measurement system using the Brookfield viscometer. A key step in developing a measurement process is determining the specific spindle to use and at which speed the measurements will be made. For BBQ sauce, Chad had selected a number of products that represent the range of viscosities (from thin to thick) for which the measurement system would need to apply. The company's BBQ sauce

Statistics for Food Scientists
http://dx.doi.org/10.1016/B978-0-12-417179-4.00005-6

portfolio contains several products with viscosities varying from thin to thick: Classic, the original product developed; Thick and Spicy, which contains higher levels of tomato paste; and Honey BBQ, the thickest in the set due to the addition of honey in the formulation. He had augmented the set of products with two competitor products that expanded this range, as future development efforts may involve the creation of products with viscosities similar to these competitors' products. One competitor's product is a Carolina-style sauce, which is quite thin. Another is a very thick sauce with a claim on the bottle that it "clings better to meat than any other sauce."

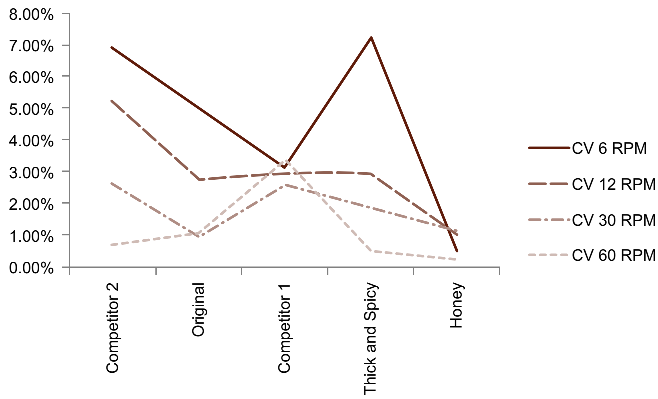
Chad's development of the measurement system involved subjecting these five products to the Brookfield device at a range of speeds, using several different spindles. The goal was to have a clear and consistent differentiation between the products. Consistency meant that repeated measurements of the same product were producing similar results. And a clear difference meant that the typical measurement of each product was not near that of another product.

Chad had started by using the Brookfield with different spindles on all of the sauces, at four different speeds. He had quickly determined that one spindle, the RV-3, worked well across the set of products since it could spin freely even in the thickest sauce. He then made repeated measurements of all of the products at four speeds: 6, 12, 30, and 60 rpm. The data he collected look like this:

Product	RPM	Test 1	Test 2	Test 3	Average	Stdev	CV
Competitor 2	6	36000	36267	32000	34756	2390	6.9%
Competitor 2	12	21333	21600	19600	20844	1086	5.2%
Competitor 2	30	10987	11093	10560	10880	282	2.6%
Competitor 2	60	6720	6720	6640	6693	46	0.7%
Original	6	23600	21733	23867	23067	1163	5.0%
Original	12	14067	13533	14267	13956	379	2.7%
Original	30	7387	7333	7467	7396	67	0.9%
Original	60	4600	4693	4667	4653	48	1.0%
Competitor 1	6	9600	9653	9120	9458	294	3.1%
Competitor 1	12	5600	5627	5333	5520	163	2.9%
Competitor 1	30	2830	2773	2688	2764	71	2.6%
Competitor 1	60	1720	1610	1685	1672	56	3.4%
Thick and Spicy	6	21867	24667	21733	22756	1657	7.3%
Thick and Spicy	12	13200	13800	13067	13356	391	2.9%
Thick and Spicy	30	6933	7120	6880	6978	126	1.8%
Thick and Spicy	60	4360	4387	4347	4365	20	0.5%
Honey	6	32933	32667	32667	32756	154	0.5%
Honey	12	19667	19867	19467	19667	200	1.0%
Honey	30	10160	10347	10133	10213	117	1.1%
Honey	60	6307	6333	6307	6316	15	0.2%

Chad had looked to determine a speed that provides the most stable and reliable measurement system. From the data he collected, he calculated the coefficient of variation, or CV, for each sample at each speed. CV is the standard deviation divided by the mean (and multiplied by 100 to show in %). Since measurements with higher values often display more variation, the CV gives a basis for comparing the measurement variability "corrected" for the part of the scale where the measurements reside.

Smaller CVs are most desirable, since they indicate a more repeatable measurement. The graph below displays the CVs for each sample by RPM. For nearly every sample, the lowest CVs are for measurements made with the highest speed, 60 rpm. Chad decided that the most reliable measurements would be made using this speed.



Having determined the best speed for making viscosity measurement using the Brookfield device, Chad then documented the procedure for making the measurements. The documentation included the volume of the sample being measured and the depth at which the spindle should be submerged. The analysts making the viscosity measurements were then given a brief training session on using the Brookfield viscometer to make viscosity measurements for BBQ sauce.

Maria shares with Chad her recent data analyses showing the consistently higher viscosities on the first shift. She mentions her discussion with Lisa about the measurements being made by different analysts between the two shifts. It has been a number of years since Chad developed the viscosity measurement system and trained the analysts. "Could something have changed since then?" Maria asks Chad. While he does not directly answer the question, he agrees that they should look into the situation.

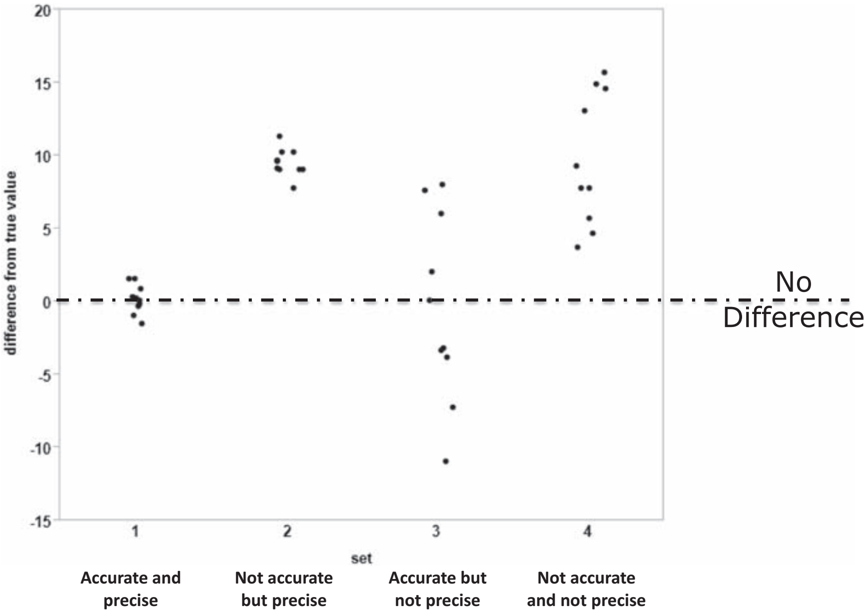
Maria starts looking for a way to assess the current state of the viscosity measurement process. She decides that a conversation with Dr Wang, her graduate advisor, would be beneficial as she seemed to be familiar with a wide variety of topics related to the food industry. She must have run into this type of problem before. She sends Dr Wang an email to schedule a call and is pleased when Dr Wang responds quickly with a suggested time for the next morning.

"Maria, how very nice to hear from you!" Dr Wang exclaims when she picks up the phone the next morning. "How is the BBQ sauce business?"

"I'm learning so much every day," Maria explains. "It seems like I will never know enough, and there is always a new challenge." Maria describes her recent work with the viscosity and her suspicion about the measurement process. "How can I determine if the measurements are valid?"

Dr Wang explains that engineers and scientists across all industries face similar questions and that the study of measurement processes has a specific name: measurement systems analysis. Methodologies were initially developed in the 1950s from the National Bureau of Standards, a US Government Agency (now called the National Institute of Standards and Technology). Methodologies were published by the Automotive Industry Action Group in the early 1990s. These have been refined over time, and a language protocol has been formalized.

Dr Wang goes on to explain that there are two components to measurement system variability: accuracy and precision. Accuracy is the difference between a sample's average measurement value and the sample's true value. Precision is the variation in measurements of the same sample measured repeatedly with the same measurement system. Dr Wang shares a diagram that illustrates these concepts. The measurements in set one are both accurate and precise. The measurements in set two are precise but not accurate. The measurements in set three are accurate but not precise. The measurements in set four are neither precise nor accurate.



Dr Wang further explains that there are two components to precision: repeatability and reproducibility. Repeatability is the variation in measurements obtained by one analyst using the same measurement device several times on the same sample. Reproducibility in its simplest form is the variation in the average of measurements made by different analysts using the same measurement device several times on the same sample. When laboratories have several of the same measurement devices, reproducibility may also include differences in the average measurements made by analysts across the set of devices.

"There is a very specific way in which measurement system precision is determined that I can share with you," Dr Wang concludes. I'll send to you a link that explains gage R&R studies."

"Have you been using Excel to analyze your data?" Dr Wang asks. "I know that we used Excel in class, but you are now getting into more sophisticated statistical procedures, so I think that you should invest in data analysis software that will enable you to do more. There are several good options. Any one of them will be very useful with your immediate measurement system task, and they will also have many more analysis tools that you will likely find helpful down the road."

"The gage R&R examples show output from one of the data analysis software packages, though the others will have a similar output format," adds Dr Wang. After a few more minutes of discussion the call concludes.

Not long after, Maria receives an email from Dr Wang with links to several sites for software packages that Dr Wang had mentioned and another site that describes gage R&R studies. She is excited after reading about the software options and the many data analysis tools she will have available and emails Steve requesting permission to purchase. Before long, his email response with permission arrives. A few clicks later, she has purchased and downloaded the data analysis software.

Maria clicks on the link describing gage R&R studies and starts reading:

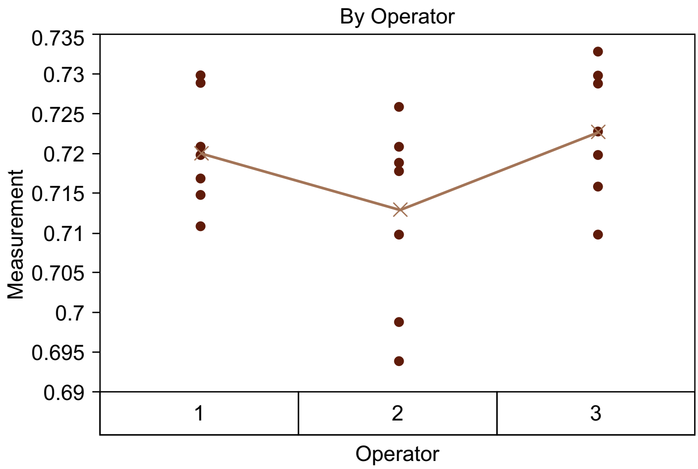
Gage R&R (repeatability and reproducibility) studies are the method for determining how much of the observed total variation is due to measurement system precision. A gage is any device or measurement method. The basic procedure for conducting a gage R&R study is very straightforward. A number of samples (N) are selected to represent the full range of long-term variation for a process. Each sample will be measured M times by K analysts. The samples are given generic codes to hide their identity, and typically, the first replicate measurements are made on the N samples, then the second replicate measurements are made on the N samples, with this procedure repeated until all M replicates are completed. An alternative to this is to completely randomize the M replicates of the N samples.

The data analysis plan for a gage R&R study is as follows:

1. Plot the data.
2. Examine the components of variation: sample-to-sample, R&R.
3. Calculate analysis of variance (ANOVA).
4. Evaluate gage R&R error.

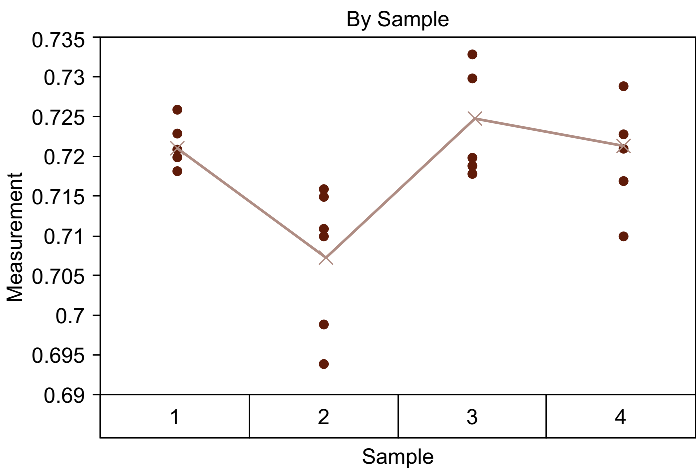
Maria is familiar with some of these steps, as she has been actively involved in analyzing data for her process improvement work. However, there are some specific outputs to a gage R&R study that she familiarizes herself with through an example. In this example, three operators perform two replicate measurements on four samples for water activity. There are several data plots that are typically used in these studies.

The "by Operator" graph shows differences in operators' averages (represented by the "x" on the graphs), while the individual sample measurements (represented by the dots) give insight into operator reproducibility.



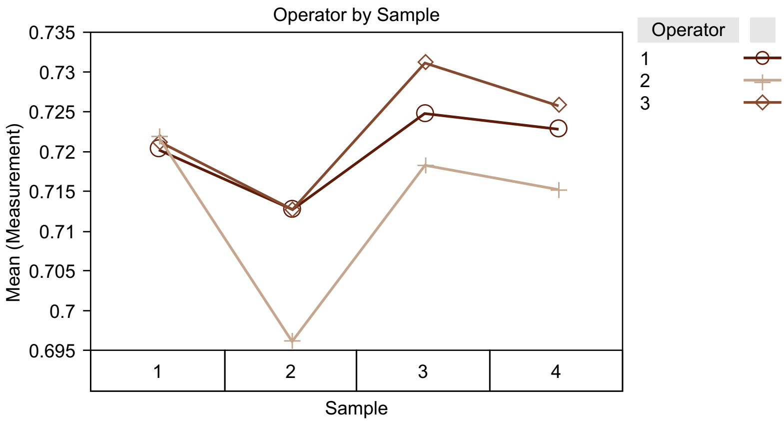
In this example, it appears that operator 2 gives lower results than the other two operators, due to its lower average value.

The “by Sample” graph shows differences in average between samples.

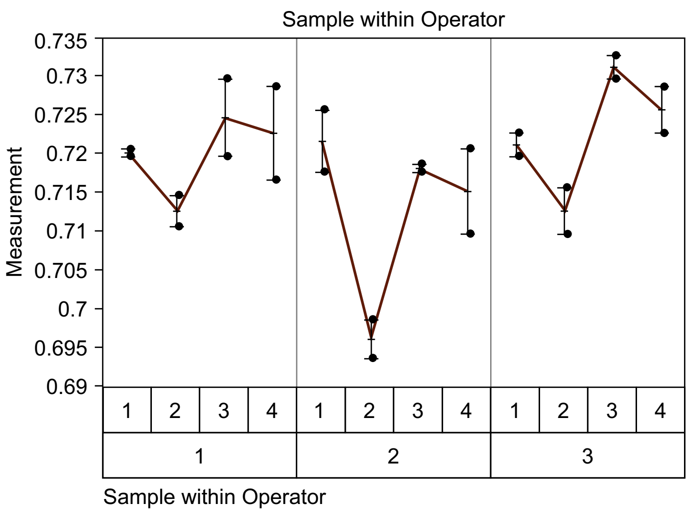


In this example, it appears that there are measured differences between the samples. This is expected since the samples were chosen to represent a wide range.

The “Operator by Sample” graph shows whether the operators measure the same sample consistently. In this example, operator 2 has lower measurements than the other operators for three of the four samples, and operators 1 and 3 differ on two of the samples (sample 3 and sample 4).



The "Sample within Operator" graphs gives insight into the consistency of the operators in measuring the samples. In the example below, each of the operators has a problem in repeating measurements for one or more samples. For example operator 1 has produced two different results, when measuring sample 3. The replicates are joined by the vertical line on the graph.



Maria is familiar with ANOVA and how the output from this analysis can help her determine the statistical significance. ANOVA can help quantify what has been perceived visually from the graphs.

Analysis of Variance

Source	DF	SS	Mean Square	F Ratio	Prob > F
Operator	2	0.000419	0.00021	5.39527	0.0456*
Sample	3	0.001079	0.00036	9.27175	0.0114*
Operator*Sample	6	0.000233	3.88e-5	1.65658	0.2150
Within	12	0.000281	2.34e-5		
Total	23	0.002011	8.74e-5		

Maria observes small *p*-values for both the sample and operator effects. These confirm the expected sample differences and also the operator differences she observed in the "by Operator" graph.

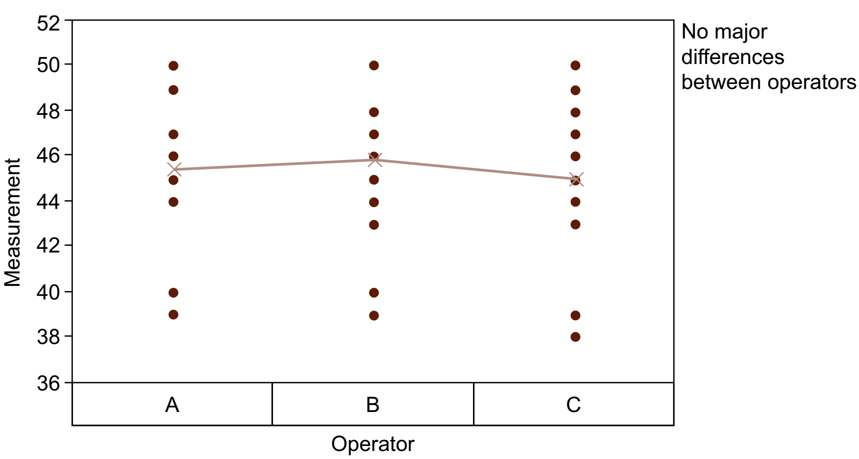
The components of variation are related to the ANOVA model and represent a partitioning of the total response variability into repeatability, reproducibility, and sample-to-sample.

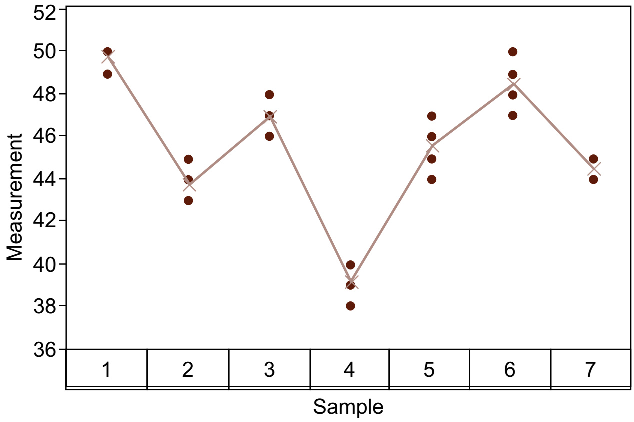
Variance Components for Gauge R&R

Component	Var Component	% of Total	20 40 60 80
Gauge R&R	0.00005242	49.50	
Repeatability	0.00002342	22.11	
Reproducibility	0.00002900	27.39	
Part-to-Part	0.00005348	50.50	

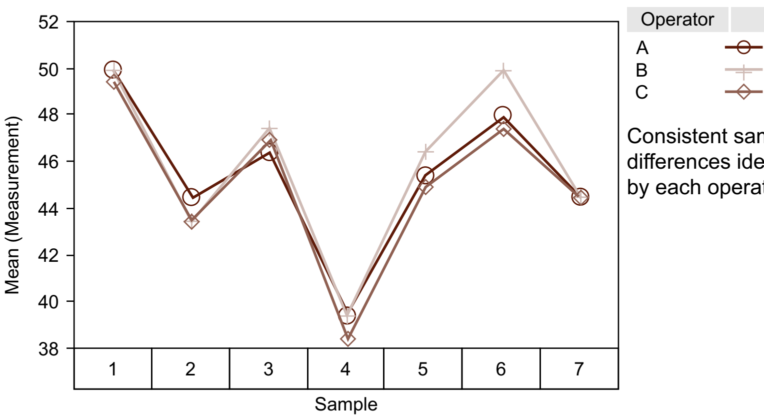
In the example, Maria sees that 50% of the variability is due to sample differences (denoted by part-to-part), while nearly 50% of the total variability is due to R&R error. She notes that guidelines for gage R&R error suggest that less than 10% would be considered acceptable, and that between 10% and 30% might be considered acceptable in some instances. In the example, the measurement system is obviously unacceptable.

A second example Maria has displays output for an acceptable measurement system. Below are "by Operator," "by Sample," "Operator by Sample," and "Sample within Operator" graphs for the acceptable measurement system.

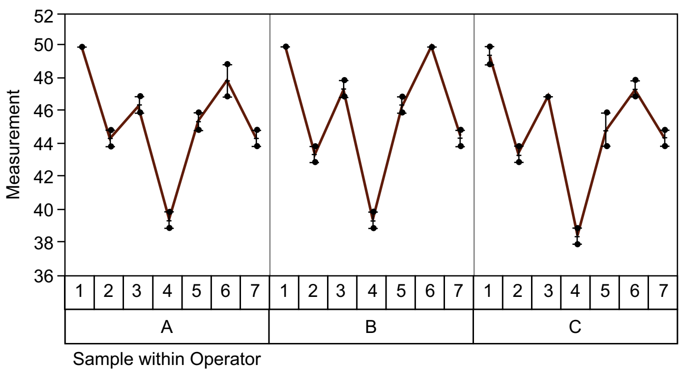




Consistent Repeatability when the same sample is measured multiple times.



Consistent sample differences identified by each operator.



General agreement on sample differences, good repeatability and good reproducibility.

Variance Components for Gauge R&R

Component	Var Component	% of Total	20 40 60 80
Gauge R&R	0.750000	5.79	
Repeatability	0.547619	4.23	
Reproducibility	0.202381	1.56	
Part-to-Part	12.210317	94.21	

She notes the obvious consistency of the data in each graph. The variance components summary shows that less than 6% of the total variability is due to R&R error, well below the 10% threshold of acceptability.

Maria discusses what she has learned about measurement systems analysis with Chad and shares the examples. They decide to conduct a gage R&R study with the analysts for each of the two production shifts, another lab technician who occasionally performs viscosity measurements when needed, and Chad himself. Chad points out that they will need to be clear that the "operators" in the study are really the different lab analysts and not the actual line operators.

Since the quality lab makes viscosity measurements for all of the BBQ sauce varieties produced by the company, Maria selects five products that span the range of viscosities from thinnest to thickest. The ideal gage R&R study has analysts making repeated measurements on the same sample. The Brookfield viscometer shears the sample as the spindle rotates during the measurement process. This means that the sample composition may change as a consequence of the measurement process. A true replicate measurement is not really possible since the additional measurement will be made on a sample that may now be different due to the process of making a measurement on it.

Maria and Chad consider the meaning of a replicate in this specific instance. Samples from the same bottle should be very similar. Bottles produced at the same time should also be similar: they have the same composition, were produced at the same time, and cooled in the bottle at the same rate. They decide to define a replicate as the same product formulation produced on the same line at the same time.

Maria selects a number of bottles of each variety produced at nearly the same time. She creates three samples for each of the five products for each of the four analysts and assigns them random three digit codes to blind them from the analysts. Her design worksheet looks like this:

sample	replicate 1	replicate 2	replicate 3
1	339	132	675
2	671	493	872
3	102	239	925
4	815	326	524
5	489	563	310

She also creates a separate data collection spreadsheet for 15 observations for each of the analysts, one of which looks like this:

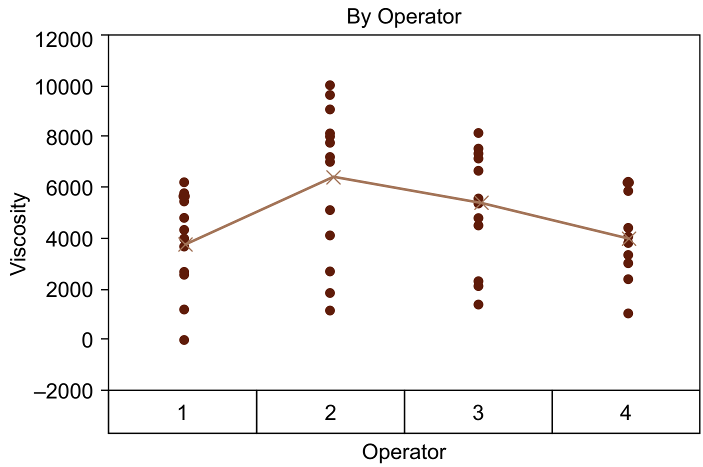
sample	viscosity
102	
489	
815	
339	
671	
239	
132	
563	
326	
493	
524	
872	
675	
925	
310	

For each analyst, she randomly orders the samples within the replicates so that there will be no bias due to the order in which the samples are measured within the replicates.

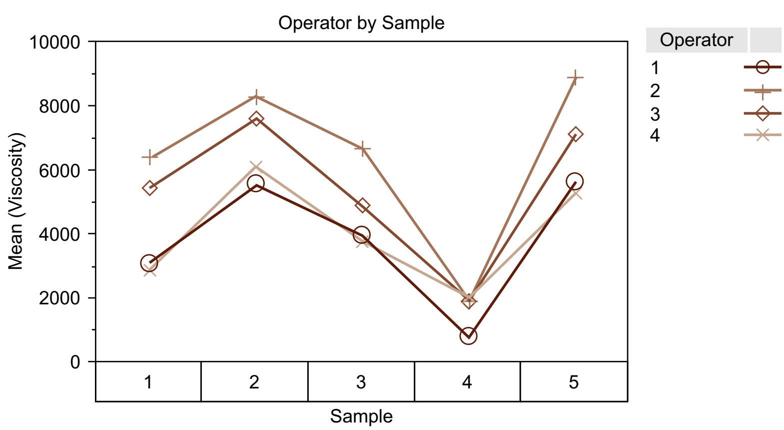
The data Maria receives from the operators look like this:

Sample	Analyst	Viscosity 1	Viscosity 2	Viscosity 3
1	1	2586	4074	2712
1	2	8152	6989	4150
1	3	5443	5475	5612
1	4	3041	3399	2407
2	1	4825	6266	5667
2	2	7829	7173	10064
2	3	7339	8151	7444
2	4	6097	6167	6213
3	1	4396	3715	3836
3	2	5150	7764	7230
3	3	4495	4868	5406
3	4	4438	3068	3993
4	1	31	1251	1187
4	2	1166	2768	1853
4	3	2127	2328	1404
4	4	1144	1017	4148
5	1	5501	5833	5739
5	2	9087	9646	8027
5	3	7195	6734	7552
5	4	6265	3862	5883

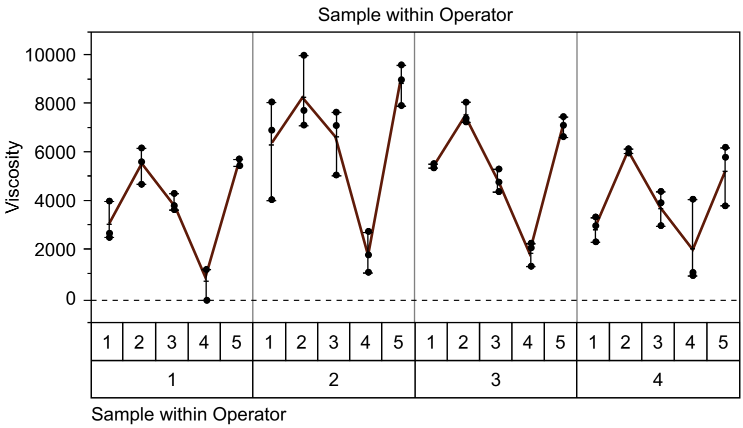
She repeats the data analysis steps described earlier. It very quickly becomes obvious that there are problems with the current measurement system. The "by Operator" plot shows noticeable operator differences.



The "Operator by Sample" graph shows clearly that the operators' measurements (of the same sample) are not consistent for any set of samples. It is desirable to observe sample-to-sample differences (indicating a selection of a good range of values); however, "within" sample (operator-to-operator) differences ought to be negligible.



Lack of repeatability seems to be a problem for all of the operators as demonstrated in the "Sample within Operator" graph. Operator 3 seems to have the best repeatability.



The variance components analysis results show a total gage R&R variance component of 36%, well above the suggested threshold of 10%. The component breakdown indicates both repeatability and reproducibility concerns.

Variance Components for Gauge R&R

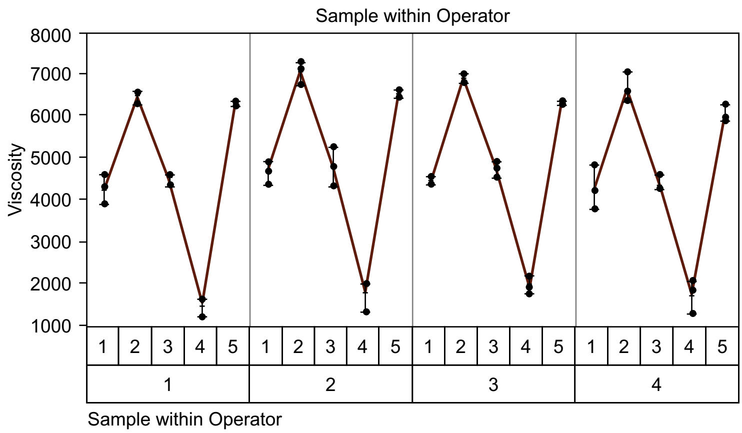
Component	Var Component	% of Total	20 40 60 80
Gauge R&R	2485664.3	36.13	
Repeatability	898607.1	13.06	
Reproducibility	1587057.2	23.07	
Part-to-Part	4393299.6	63.87	

Maria and Chad discuss potential ways to fix the R&R problems with the current measurement systems. It has been a while since the operators had received any training in the measurement system. Chad conducts training in the proper execution of the measurement process with each of the operators, highlighting the need for consistency in all aspects of the process.

Maria and Chad repeat the gage R&R study after the operator training. They use new samples of each of the products in the original study, and create new three-digit codes and randomizations for the study design. The data from the second study is captured in the following table:

Sample	Analyst	Viscosity 1	Viscosity 2	Viscosity 3
1	1	4649	4373	3947
1	2	4748	4984	4410
1	3	4404	4546	4606
1	4	4893	3830	4269
2	1	6347	6643	6593
2	2	6810	7355	7186
2	3	7053	6828	7055
2	4	6628	7086	6410
3	1	4369	4442	4668
3	2	4850	4378	5335
3	3	4570	4825	4958
3	4	4315	4647	4350
4	1	1718	1284	1670
4	2	1391	2062	2063
4	3	1803	1957	2238
4	4	2115	1334	1879
5	1	6358	6420	6288
5	2	6619	6669	6473
5	3	6413	6328	6331
5	4	6340	6020	5946

Maria repeats the gage R&R study analysis process with the new data. The "Sample within Operator" graph for the new data and the variance components are shown below.



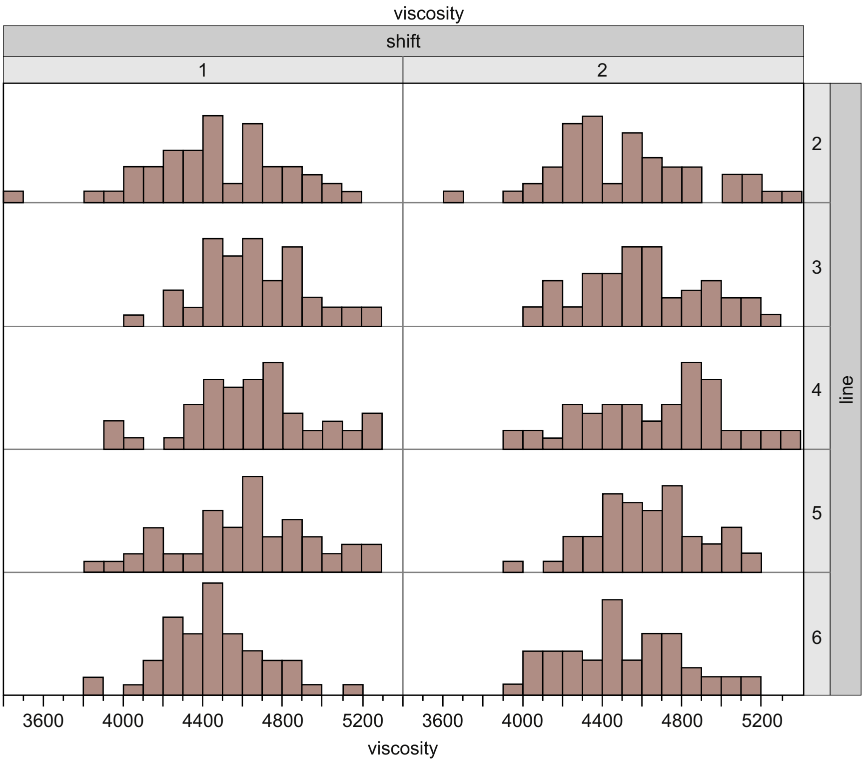
Variance Components for Gauge R&R

Component	Var Component	% of Total	20 40 60 80
Gauge R&R	83532.0	3.11	
Repeatability	63548.0	2.36	
Reproducibility	19984.0	0.74	
Part-to-Part	2606597.3	96.89	

Now, only slightly more than 3% of the variability is due to the gage R&R, and the graphs display a high degree of repeatability and better consistency across the operators.

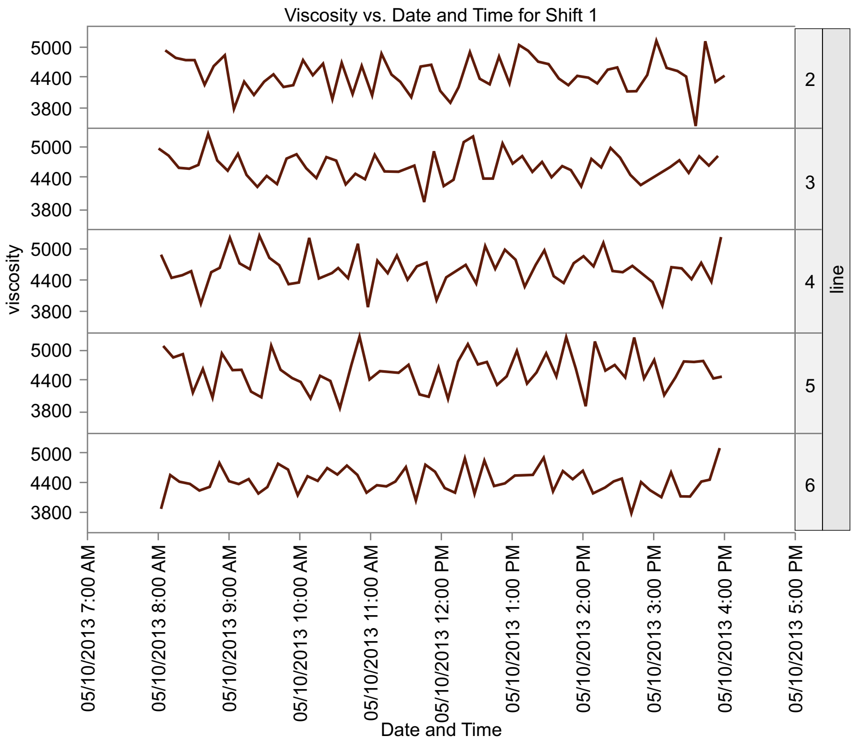
Maria and Chad are now satisfied with the measurement system and put plans into place for auditing and occasional retraining if necessary. Maria now feels confident that the measurements that she receives in the future will be useful in guiding her improvement efforts.

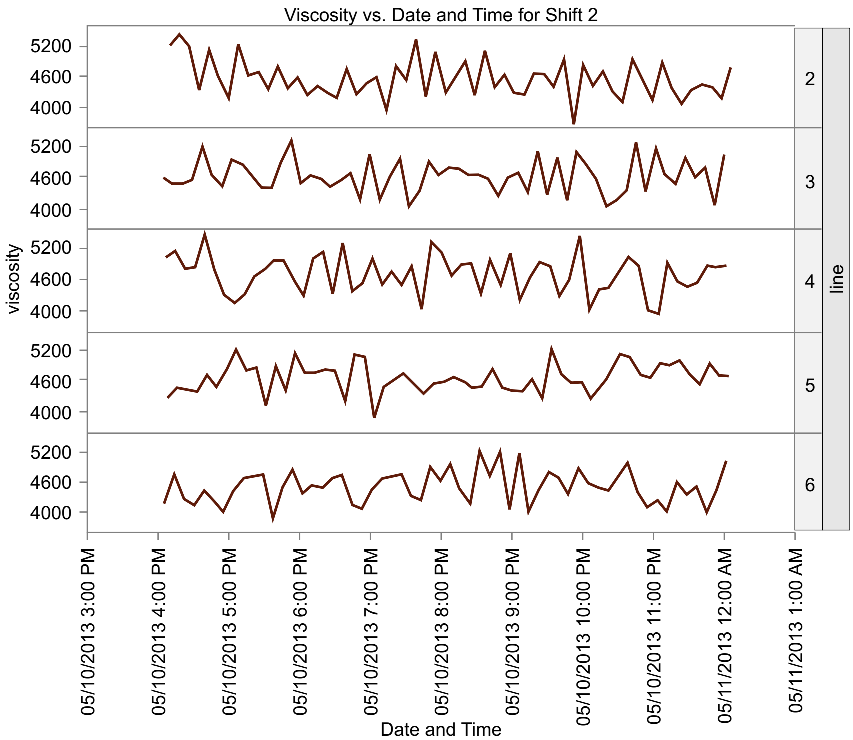
Now, Maria turns her attention back to the assessment of the viscosities for each of the lines. Has the lack of a good measurement system potentially masked the line-to-line differences? She decides to repeat her earlier work and collects, as earlier two shifts of data from each of the lines. Before using her software to perform the ANOVA modeling, she sees that she can create graphs to visualize the data much more easily that she could in Excel. She first creates histograms for each line and shift combination. She can create these at once and have them put together in one presentation with the same scale as the default. Her histograms look like this:



The histograms are laid out so that each row is a shift and each column is a production line. They are presented on their sides to provide a better comparison across the production lines. Maria notes the similarity in shape of the distributions and that the center of these distributions seem smaller for lines 2 and 6.

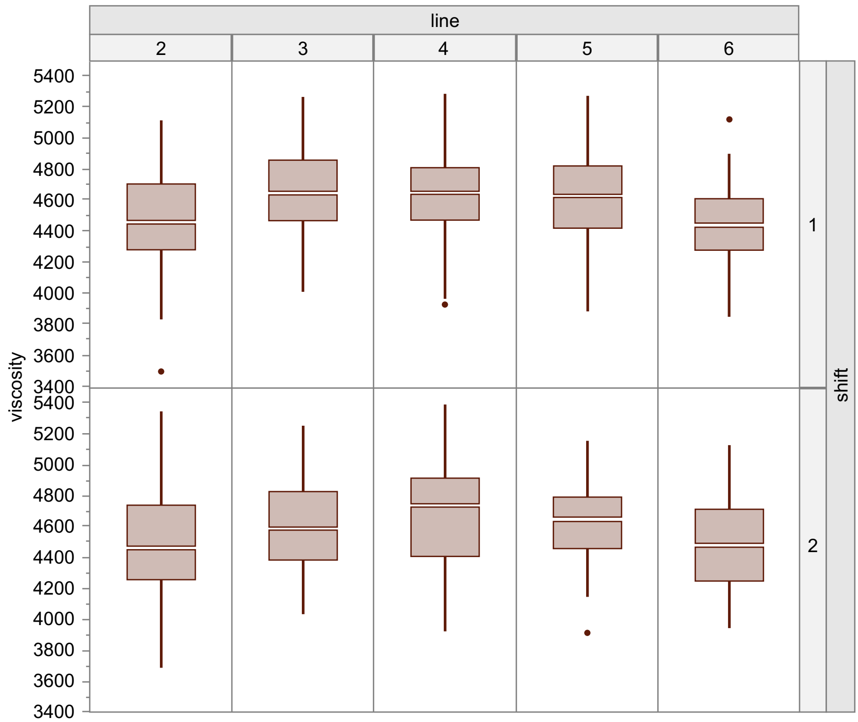
Maria also creates the time series plots for the new data using her software. She creates a separate set of graphs for shifts 1 and 2 since they cover different time frames and stacks them by line for easier comparison. The graphs look like this:





She sees that there are no increasing or decreasing trends in viscosity for any of the lines for both shifts.

Maria finds another graph that she finds useful for visualizing the viscosity data. A box plot shows the distribution of the data somewhat like the histogram but can sometimes make it easier to compare the center and spread. She arranges box plots of the viscosity data by line and shift.



The box in a box plot is the middle 50% of the data, and the line in the middle of the box is the median. The lines extending from the boxes either represent the range of the data or, in some instances, individual data points are left in the box plot to indicate that they stand out from the rest of the data distribution. From Maria's box plots, she can more readily see that lines 2 and 6 are centered lower than the rest for both shifts, and that perhaps line 6, the recently installed line, does have less variation in viscosity.

Maria also sees that the ANOVA modeling, to compare lines and shifts, is easy to create and that the Tukey's method for comparing means is an easy option to include in the analysis. Her results look like this:

Summary of Fit

RSquare	0.062346
RSquare Adj	0.047541
Root Mean Square Error	306.8063
Mean of Response	4580.975
Observations (or Sum Wgts)	580

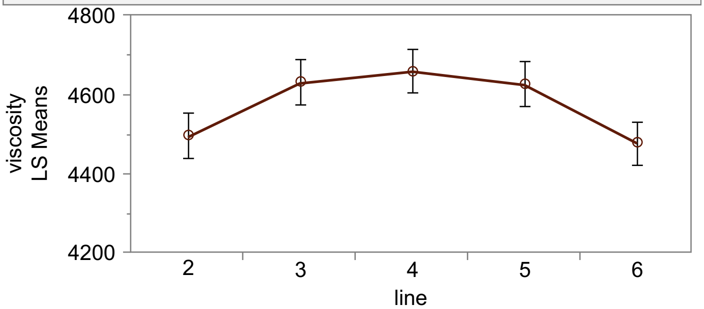
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	9	3567544	396394	4.2111
Error	570	53654151	94130	**Prob > F**
C. Total	579	57221696		<.0001*

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Shift	1	1	27265.9	0.2897	0.5906
Line	4	4	3320601.6	8.8192	<.0001*
Shift*line	4	4	219676.8	0.5834	0.6748

LS Means Plot



LSMeans Differences Tukey HSD

α= 0.050 Q= 2.73655

Level			Least Sq Mean
4	A		4662.7433
3	A		4633.6256
5	A		4629.1960
2		B	4498.7070
6		B	4480.6035

Levels not connected by same letter are significantly different.

With no significant shift differences, it is clear now how the lines differ in their performance. The Tukey's means comparison lists the mean viscosities for the lines in descending order. In the last output table, it states that "levels not connected by the same letter are significantly different." It is clear that lines 2 and 6 have significantly lower viscosity levels than the other three lines. Though these findings are no different from what she observed before, Maria now has confidence in her performance assessment and can develop a strategy for improving performance across the lines. "Lesson learned," Maria notes to herself. One needs to be confident that the measurement system is reliable before any analyses are performed. A bad measurement system will lead to bad conclusions!

CHAPTER 6

Regression and Correlation

Maria now better understands the BBQ sauce process and how the individual lines perform. However, she still does not have an understanding of why viscosity varies. There are many things that may account for the variability in viscosity levels, including processing and ingredient variation.

Maria gathers a small team of experts to help determine the potential factors that may affect BBQ sauce viscosity. The team includes her manager, Steve; Louise, the quality manager; and several of the line operators. The team produces a long list! The factors fall into two groups: processing factors such as pump pressure, steam pressure, and line speed; and formulation factors such as the amounts of sugar, tomato paste, and fruit paste. Maria suggests continuing with the viscosity sampling throughout the upcoming shift, but also to have one of the line operators record the levels of these processing and formulation factors at the same time as taking the viscosity measurements.

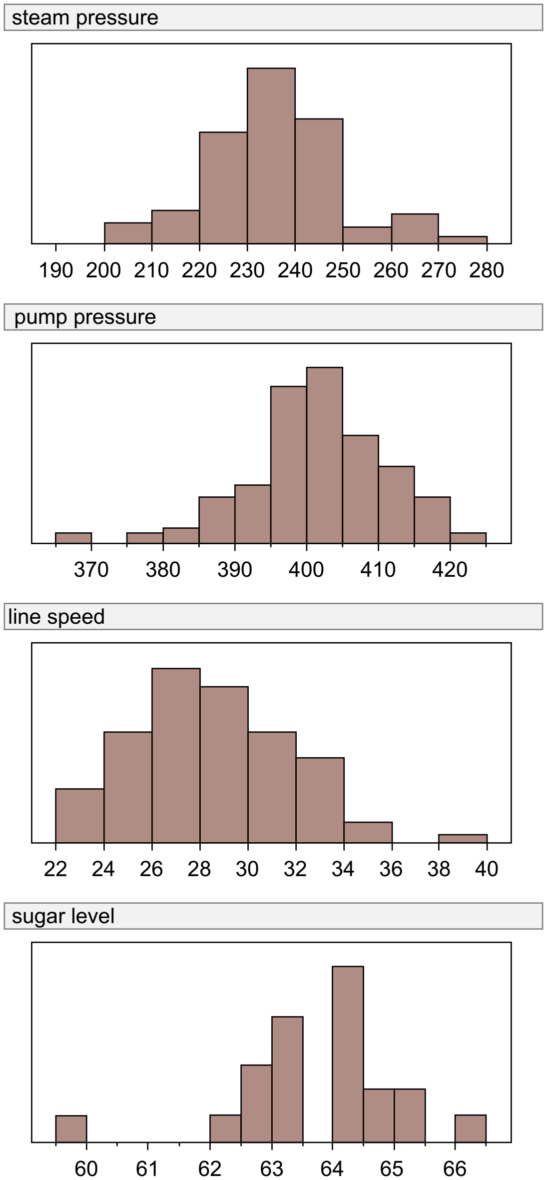
BBQ sauce is made in batches, with the ingredients mixed in large vats that are then pumped to the bottling lines. Consequently, the ingredients only vary batch to batch, whereas the pressures and line speed can vary as the batch is pumped and bottled. Once the ingredients are mixed, it takes around half an hour to process the batch into bottles. Maria suggests taking measurements five times as each batch is bottled and proceeds to create a data collection template for the operator.

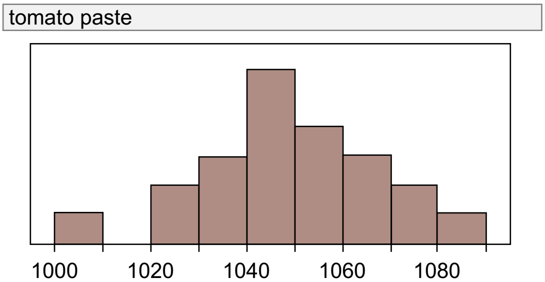
The next morning, the collected data arrive in Maria's inbox. The sample data for the first two batches look like this:

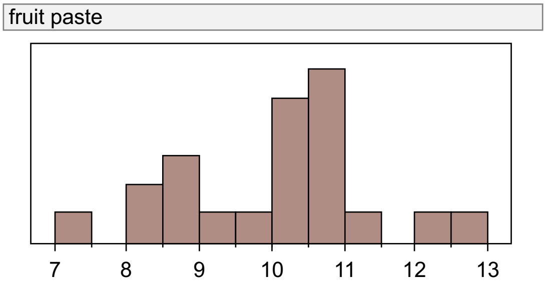
date and time	batch	steam pressure	pump pressure	line speed	sugar level	tomato paste	fruit paste	viscosity
5/23/2013 8:14	1	246.95	395.38	26.90	62.81	1032.17	10.14	4125
5/23/2013 8:23	1	248.82	391.80	32.16	62.81	1032.17	10.14	4126
5/23/2013 8:31	1	244.60	397.82	31.31	62.81	1032.17	10.14	4130
5/23/2013 8:39	1	243.89	401.87	28.44	62.81	1032.17	10.14	4135
5/23/2013 8:48	1	245.07	387.82	32.00	62.81	1032.17	10.14	4142
5/23/2013 8:56	2	237.32	399.06	22.60	63.12	1042.06	10.96	4216
5/23/2013 9:04	2	240.51	404.06	32.63	63.12	1042.06	10.96	4221
5/23/2013 9:13	2	241.20	399.29	31.59	63.12	1042.06	10.96	4221
5/23/2013 9:21	2	238.68	396.43	29.81	63.12	1042.06	10.96	4225
5/23/2013 9:29	2	241.07	398.42	34.39	63.12	1042.06	10.96	4225

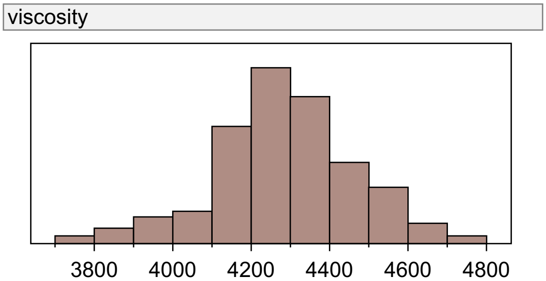
Maria first creates histograms for viscosity and each of the six factors.

Statistics for Food Scientists
http://dx.doi.org/10.1016/B978-0-12-417179-4.00006-8


steam pressure
190 200 210 220 230 240 250 260 270 280
pump pressure
370 380 390 400 410 420
line speed
22 24 26 28 30 32 34 36 38 40
sugar level
60 61 62 63 64 65 66

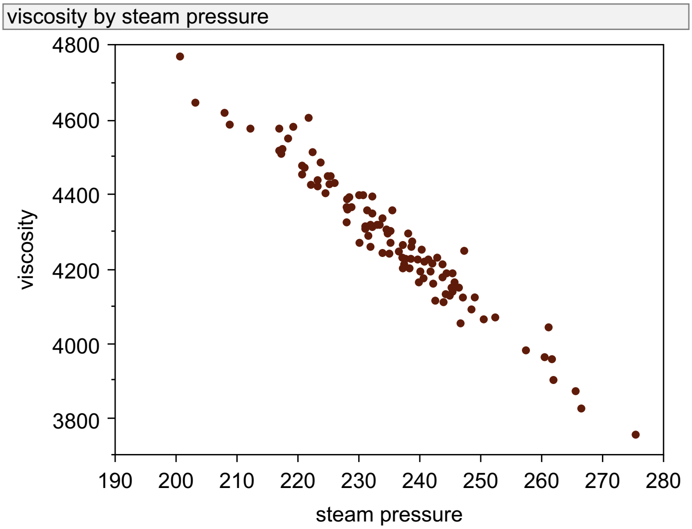


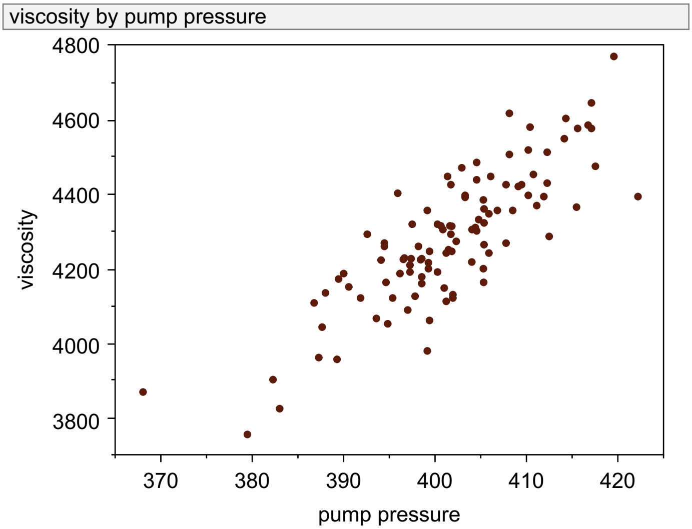


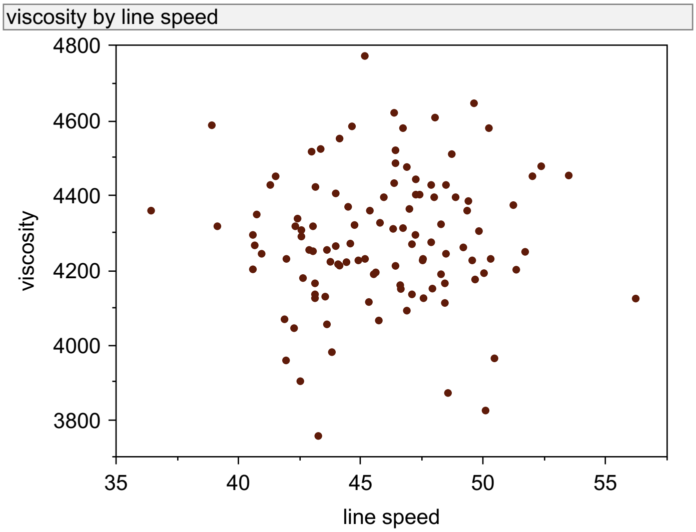


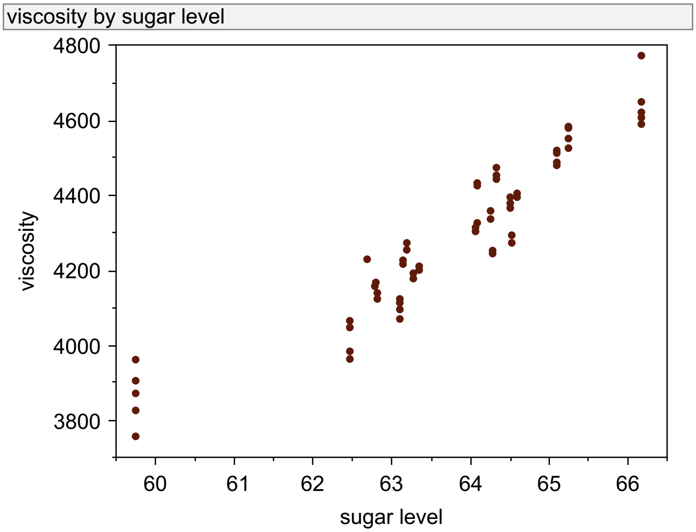
She observes no major anomalies and sees that each of the data sets seem to follow approximately a bell shape.

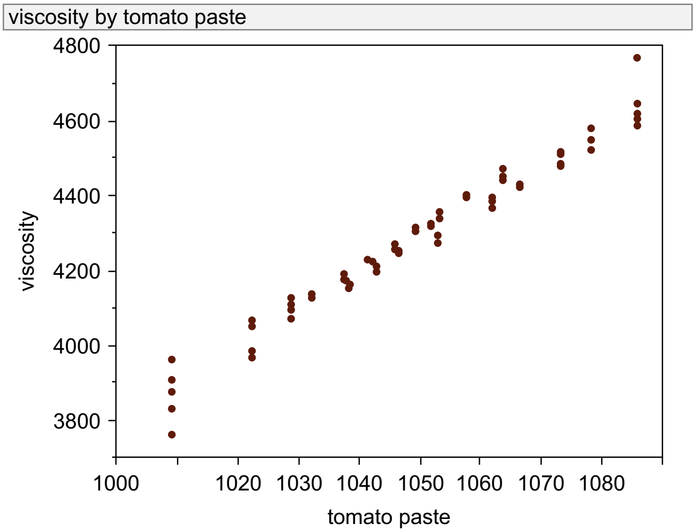
To understand the potential relationships between the individual factors and viscosity, Maria creates scatter plots of each factor with viscosity, with the factors on the horizontal axis and viscosity on the vertical axis. Her scatter plots look like this:

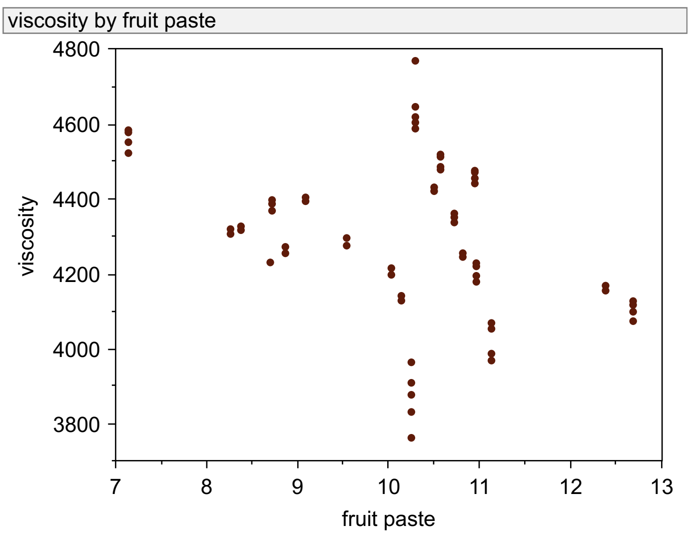





viscosity by line speed
4800
4600
4400
4200
4000
3800
viscosity
35
40
45
50
55
line speed


viscosity by sugar level
4800
4600
4400
4200
4000
3800
viscosity
60
61
62
63
64
65
66
sugar level





Some of the scatterplots seem to show stronger relationships than others, though it is difficult to objectively determine the strength of some of the relationships based on the plots alone.

Maria recalls from her notes of data analysis sessions with Dr Wang that the correlation coefficient is a measure of the strength of the relationship between two measures. This measure is represented by the letter r, with a value between -1 and 1. The sign of the correlation coefficient indicates the direction of the relationship. An $r = 1$ indicates a perfect positive linear relationship, while $r = -1$ indicates a perfect negative linear relationship. Her data analysis software lets her easily create a matrix of correlation coefficients, with all possible pairs of her input variables. She has correlations not only between the factors and viscosity, but also of the factors with each other. Her correlation matrix looks like this:

	steam pressure	pump pressure	line speed	sugar level	tomato paste	fruit paste	viscosity
steam pressure	1.00						
pump pressure	-0.82	1.00					
line speed	0.53	-0.44	1.00				
sugar level	-0.90	0.83	-0.50	1.00			
tomato paste	-0.96	0.83	-0.49	0.92	1.00		
fruit paste	0.36	-0.28	0.22	-0.28	-0.38	1.00	
viscosity	-0.97	0.83	-0.50	0.92	0.98	-0.37	1.00

Maria notes both positive and negative correlation coefficients, and some with values close to ± 1. The correlation between tomato paste and viscosity is 0.98, indicating that as tomato paste level increases, viscosity increases. This makes sense to her as more paste would seem to make a thicker product. The correlation between steam pressure and viscosity is -0.97, indicating that as steam pressure increases, viscosity decreases. This too makes sense since the higher steam pressure would introduce more water into the product, making it thinner.

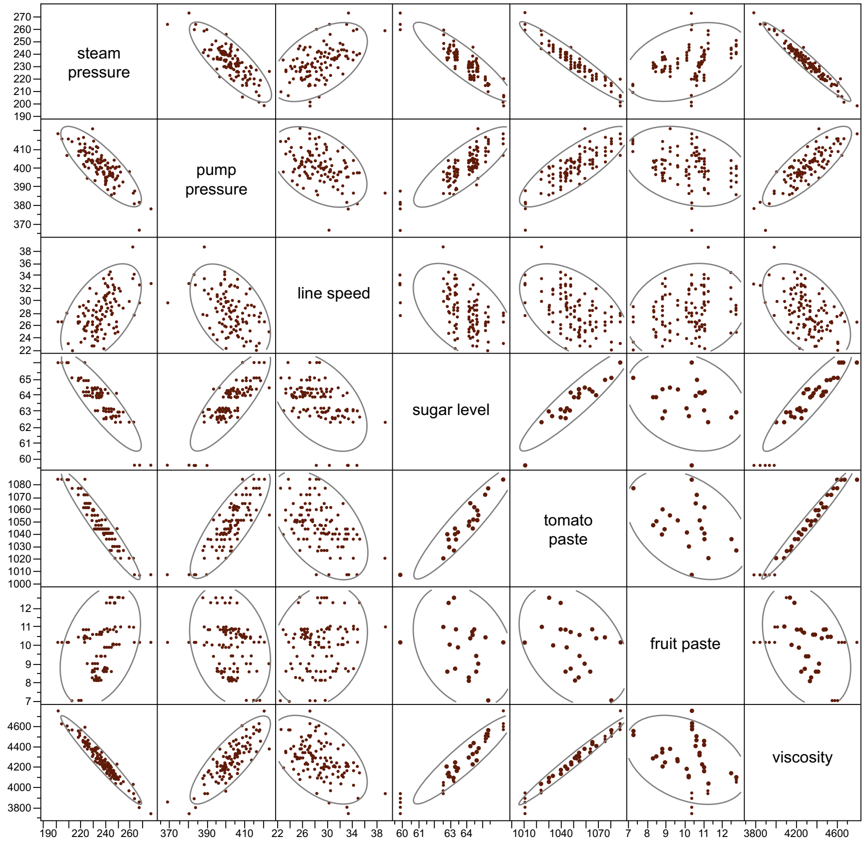
The other correlations with viscosity are smaller, but they may still be meaningful. Also she sees that processing and formulation factors are correlated with each other in varying degrees. Maria recalls that Dr Wang used the following guidelines for interpreting correlations:

$|r| \geq 0.7 \rightarrow$ Strong correlations

$0.4 < |r| < 0.7 \rightarrow$ Moderate correlation

$0 < |r| \leq 0.4 \rightarrow$ Weak correlation

Her data analysis software also provides her with a matrix of scatter plots for each of the pairs of variables. That matrix of plots looks like this:



Each plot has an ellipse added that contains most of the data points. Ellipses provide a guide in understanding the relationships between the measures. The more circular ellipse shapes, the weaker the relationships are between the measures.

Maria recalls that Dr Wang had stressed that correlation is not the same as causation. For example, she cannot state that the slower line speed is causing higher viscosity. It may be that a more viscous product causes the line to run slower or that there is something else that is causing both viscosity and line speed to increase or decrease at the same time. Maria also recalls that correlation is a measure of a linear relationship, and if the relationship is not linear, the correlation coefficient would likely show a weak relationship. This is another reason it is useful to visualize the relationships.

The correlation coefficient can be calculated using the following formula:

$$r = \frac{\sum_i (X_i - \overline{X}) \sum_i (Y_i - \overline{Y})}{\sqrt{\sum_i (X_i - \overline{X})^2 \sum_i (Y_i - \overline{Y})^2}}$$

Maria by now knows that graphing the data each and every time is the first step in analyzing any data set. This time she made no exception and is highly rewarded for it. She sees that line speed and fruit paste seem to have weak relationships with viscosity, as also indicated by their corresponding correlation coefficients. At the same time, she notices that there is a strong linear relationship between viscosity and steam pressure as well as between viscosity and tomato paste level. Maybe these are the keys to controlling viscosity.

A related statistical analysis technique is linear regression. It expands on correlation by fitting a line relating a predictor variable to a response. Maria considers viscosity a response and all of these factors as potential predictors. Estimating the regression line could serve as an additional control to check if the fit of the relationship is strong.

In its simplest form, linear regression relates one factor X to a continuous response variable Y. The fitted line has the following equation:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where

Y = Output or response (dependent variable)

β_0 = Intercept (the point at which the line crosses the Y-axis)

β_1 = Slope

X = Input factor or predictor (independent variable)

ε = Error or noise

Multiple regression expands this simple form to include more than one input factor. The fitted line has the following equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n + \varepsilon$$

where

$Y =$ Output or response (dependent variable)

$\beta_i =$ Model parameters to be estimated

$X_i =$ Inputs or predictors (independent variables)

$\varepsilon =$ Error or noise

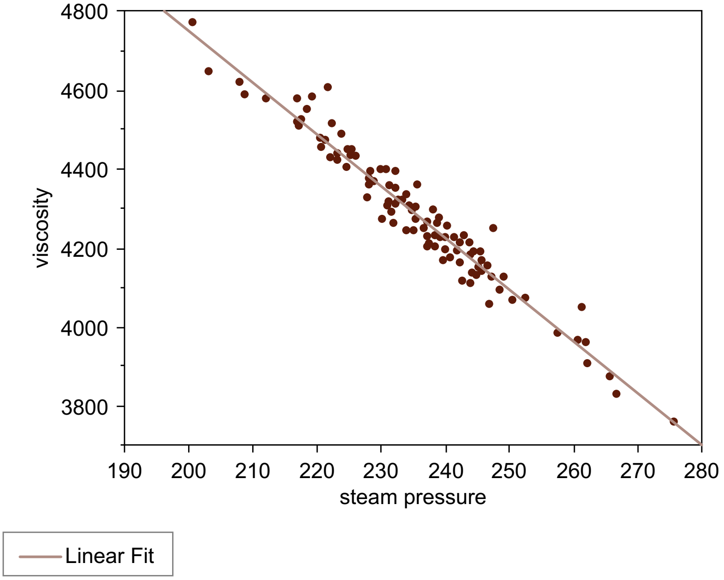
As per Maria's notes, the main purpose of regression analysis is to find the line that best fits the data. The equation for the line relates the quantitative predictor variable to a response variable that can be used to make predictions. With a multiple regression equation, the relationships between Y and the Xs is still linear. The purpose of multiple regression is also to make predictions about a response variable.

Regression can be used to determine if a factor is a true predictor or if the relationship is observed by chance. With one factor, there is a p-value associated with a hypothesis test of whether the slope of the line is significantly different from zero. While a $p < 0.05$ does indicate a significant relationship between the factor and response, it does not alone determine if the relationship is strong enough to accurately predict the response from the factor. Another measure from the regression, R^2, describes the strength of the relationship between the response and the factor. It measures how close the individual data points are to the fitted regression line. Deviations between the points and the line can be considered error. The smaller the errors, the closer R^2 is to 1. If the line does not fit well, then R^2, approaches zero.

R^2 is also called the coefficient of determination. It represents the percentage of variability in the response that can be explained by the factor(s). R^2 is the square of the correlation coefficient. So, $R^2 = r^2$.

After going through her notes, Maria feels confident about fitting regression lines to the data collected from the production line. She fits a line for each of the processing factors with viscosity.

viscosity by steam pressure



Linear Fit

viscosity = 7388.1034 – 13.194431*steam pressure

Summary of Fit

R Square	0.950291
R Square Adj	0.949831
Root Mean Square Error	39.79612
Mean of Response	4286.281
Observations (or Sum Wgts	110

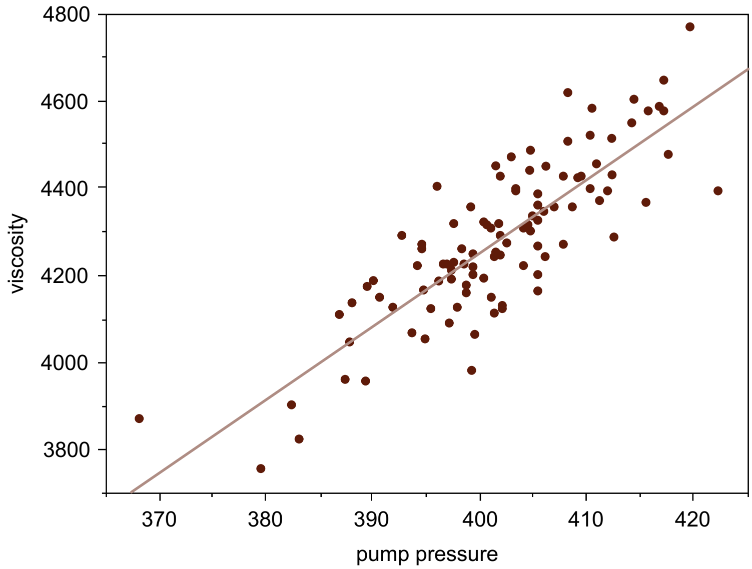
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	3269859.3	3269859	2064.656
Error	108	171043.0	1584	**Prob > F**
C. Total	109	3440902.3		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	7388.1034	68.3696	108.06	<.0001*
steam pressure	–13.19443	0.29038	–45.44	<.0001*

viscosity by pump pressure



Linear Fit

Linear Fit

viscosity = –2449.113 + 16.769609*pump pressure

Summary of Fit

R Square	0.692845
R Square Adj	0.690001
Root Mean Square Error	98.9244
Mean of Response	4286.281
Observations (or Sum Wgts	110

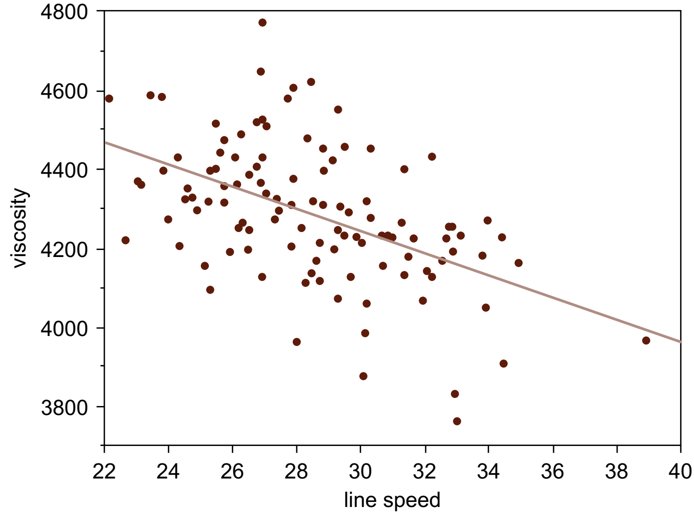
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2384010.3	2384010	243.6135
Error	108	1056891.9	9786	**Prob > F**
C. Total	109	3440902.3		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	–2449.113	431.6344	–5.67	<.0001*
pump pressure	16.769609	1.074416	15.61	<.0001*

viscosity by line speed



Linear Fit

Linear Fit

viscosity = 5091.8942 – 28.291634*line speed

Summary of Fit

R Square	0.253509
R Square Adj	0.246597
Root Mean Square Error	154.2185
Mean of Response	4286.281
Observations (or Sum Wgts	110

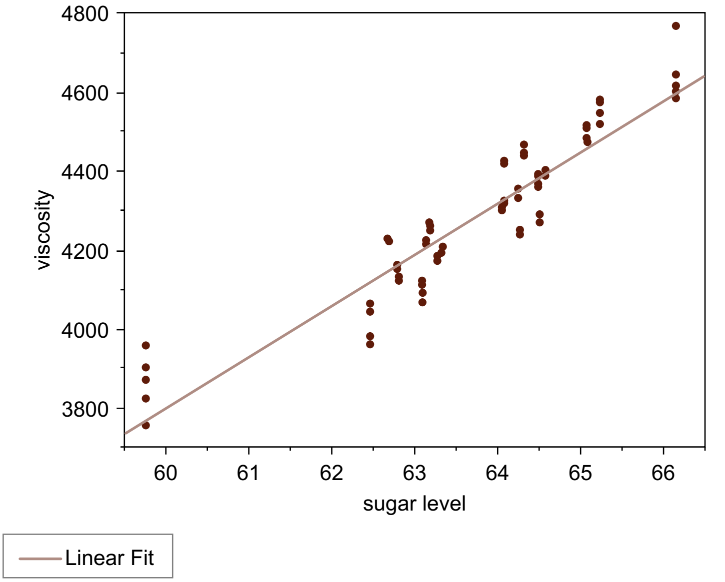
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	872300.6	872301	36.6769
Error	108	2568601.7	23783	**Prob > F**
C. Total	109	3440902.3		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	5091.8942	133.8343	38.05	<.0001*
line speed	–28.29163	4.671555	–6.06	<.0001*

viscosity by sugar level



Linear Fit

viscosity = –3952.097 + 129.32095*sugar level

Summary of Fit

R Square	0.848672
R Square Adj	0.84727
Root Mean Square Error	69.43596
Mean of Response	4286.281
Observations (or Sum Wgts	110

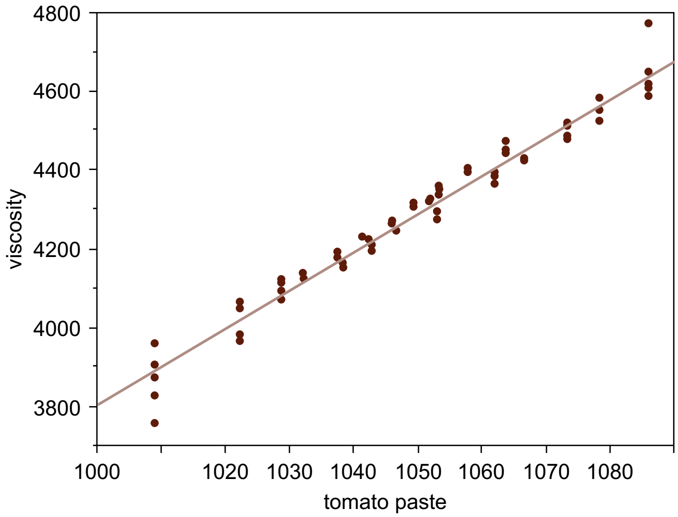
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2920196.2	2920196	605.6798
Error	108	520706.1	4821	**Prob > F**
C. Total	109	3440902.3		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	–3952.097	334.8151	–11.80	<.0001*
sugar level	129.32095	5.254693	24.61	<.0001*

viscosity by tomato paste



Linear Fit

Linear Fit

viscosity = –5933.165 + 9.7413042*tomato paste

Summary of Fit

R Square	0.968854
R Square Adj	0.968566
Root Mean Square Error	31.50089
Mean of Response	4286.281
Observations (or Sum Wgts	110

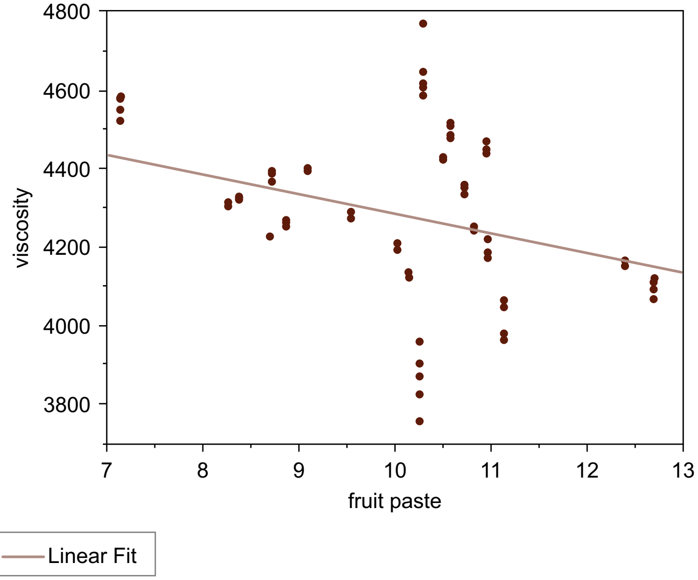
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	3333733.2	3333733	3359.581
Error	108	107169.1	992	**Prob > F**
C. Total	109	3440902.3		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	–5933.165	176.3388	–33.65	<.0001*
tomato paste	9.7413042	0.168064	57.96	<.0001*

viscosity by fruit paste



Linear Fit

viscosity = 4787.9859 – 49.963258*fruit paste

Summary of Fit

R Square	0.13958
R Square Adj	0.131613
Root Mean Square Error	165.5692
Mean of Response	4286.281
Observations (or Sum Wgts	110

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	480281.4	480281	17.5201
Error	108	2960620.9	27413	**Prob > F**
C. Total	109	3440902.3		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	4787.9859	120.8967	39.60	<.0001*
fruit paste	–49.96326	11.93665	–4.19	<.0001*

There is a lot in the output from her statistics software! She is familiar with the analysis of variance (ANOVA) table since that was in the output from her ANOVA models. She sees in the parameter estimates table the intercept and slope for each of the lines fitted with the individual factors. The Prob $> |t|$ column lists the p-values for the intercept and slope. For each of her processing and ingredient factors, the p-values for the slope are very small, so all of the factors are significant. But she can see from the graphs that some relationships are stronger than others and the R^2 values differ greatly. Maria makes the following observations:

1. Strong relationships exist between tomato paste and viscosity ($R^2 = 0.97$) and steam pressure and viscosity ($R^2 = 0.95$).
2. Moderate relationships are seen between sugar level and viscosity ($R^2 = 0.85$) and pump pressure and viscosity ($R^2 = 0.69$).
3. Weak relationships are present between line speed and viscosity ($R^2 = 0.25$) and fruit paste and viscosity ($R^2 = 0.14$).

Maria brings her expert team back together to discuss the results. Since both tomato paste level and steam pressure have strong relationships with viscosity, she suggests that these two need to be better controlled to keep viscosity on target and to reduce variation.

Connie, one of the line operators, mentions that the ingredient factors are adjusted for each batch to meet food safety control points. "Acid and sugar levels can vary with the input tomato paste. We add the ingredients to the batch tank based on the formula. We measure the acid level of each batch and often need to adjust sugar or tomato paste levels to meet the acid target. We cannot use them to control viscosity too. But, we can adjust steam pressure if we need to" adds Connie.

The regression line between steam pressure and viscosity will be the basis of the control plan Maria puts in place. The slope of the regression line is -13.2. The units for steam pressure are millibars (mb). So, an increase in steam pressure of 1 mb will decrease viscosity about 13 cP. Maria uses this finding to document the control strategy as a standard operating procedure for the line operators to use going forward.

Steve is pleased with Maria's work on developing the control strategy. "This is the first time we have used actual data to develop controls for the plant. Your data analysis skills are proving to be incredibly valuable!"

Maria blushes when she hears this praise from her boss.

This page intentionally left blank

CHAPTER 7

Statistical Process Control (SPC)

Now that the team has determined a way to adjust the process to change viscosity by adjusting the steam pressure, Maria has been asked to determine a strategy for monitoring and controlling viscosity during production. She has seen that there is not much consistency in what is currently being done and that each line operator has their own practice for how often they monitor viscosity and how they adjust process settings based on what they see.

She remembers that Dr Wang had discussed statistical process control (often abbreviated SPC) in one of her classes as one way to monitor and control processes, and Maria decides to set up a call with her. Before the call Maria does a bit of homework on this topic, so that she can make best use of their time. A search on Google gives enough background to make her feel smart!

Process capability is a methodology to quantify process performance with respect to specifications. SPC is a methodology for monitoring and adjusting processes to keep them on target and within specification limits. They are both quality tools that use statistical methods.

SPC emphasizes early detection and prevention of quality defects rather than correcting defects after they occur. It can provide real-time monitoring and control of process variability and operator adjustments to keep a process on target. A key tool in SPC is a control chart.

For quantitative measures like viscosity, the definition of a target is required. The target is the most desirable value for the process measure. Often when a target is determined, a tolerance around that target is also determined. The tolerance, or specification range, is the range of the measure on the product that is still considered acceptable. This concept acknowledges that some level of variation in the product is considered acceptable. This target is used in the SPC charts, and the tolerance is used in determining process capability.

As Maria starts thinking about BBQ sauce viscosity, she realizes that though she has been told the ideal value, she has no idea how this was determined and if an appropriate specification range exists! Is what she now believes to be the target level well defined? What level of viscosity below

Statistics for Food Scientists
http://dx.doi.org/10.1016/B978-0-12-417179-4.00007-X

or above the target would consumers even notice? Furthermore, even if they noticed it, would they care? Maybe consumers have a wide tolerance to viscosity variation. Would higher or lower viscosity product affect production efficiencies?

Maria contacts the person responsible for understanding consumer perceptions of the company's products, a marketing professional named James. She asks: "Do we know how consumer preferences for our BBQ sauces relate to our analytical viscosity measurements?"

James's answer is both surprising and puzzling: "No, not really. We have never tried to relate viscosity measurements to consumer acceptance of the product. We don't really know the ideal viscosity level and how much tolerance consumers have to variation in viscosity."

"I need to understand these to create a viscosity control strategy," Maria explains. "Is there research that we can conduct to better understand this?"

"Can you create or provide products of the same variety with different viscosities?" James asks. "I could then design a consumer test to evaluate these products and determine the acceptable range and also verify your target."

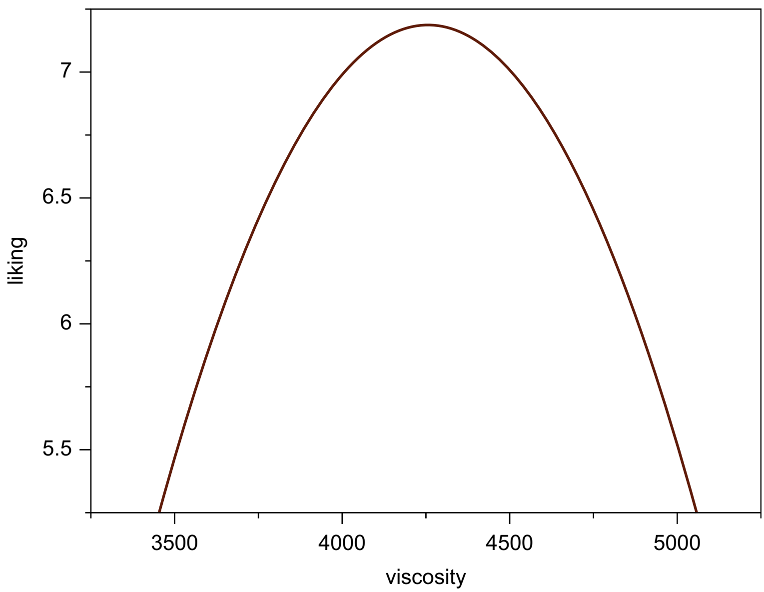
"I think I can do this. I'll get back to you soon with some ideas," Maria quickly replies, though she really does not know how she will approach James's request.

Maria starts thinking about the product and how to determine a viscosity specification. Ideally, she would need to have products that span a wide range of viscosity to provide to James for his consumer test. Maybe she can start evaluating product lots to identify those with different viscosities and select a range. That would be time-consuming and could potentially add other sources of variability like ingredient variation and product age. Maybe there is a way to force viscosity differences?

"Wait, I know how to do this!" Maria exclaims. "Since steam pressure affects viscosity, I can make systematic changes so that viscosity will vary too!"

Maria starts thinking about how she would design a trial to produce products with different viscosities by changing levels of steam pressure. She checks back in with James to see how many products he can handle in the consumer test. The consumers will rate their liking of each product on a 1–9 scale that is commonly used in product testing and also answer a number of other questions on the products. Since this process takes some time, he is not comfortable giving a consumer more than four samples in the same testing day, but he can recruit the consumers to return for an additional day if needed.

Maria envisions how she and James can relate the product viscosities to the consumer liking measurement. She will have measured viscosity on each of the products being tested. Since she now knows there is variability across bottles, she will measure viscosity on more than one bottle for each of the products. James will have individual consumer ratings for each of the products. By relating the average consumer ratings with the average viscosity measurements for each product, they should be able to understand consumer tolerance to viscosity differences and determine a viscosity range where the product is well liked by consumers. A plot would visualize this relationship, and a fitted curve could be used to determine the specific lower and upper viscosity limits for acceptance. She expects the relationship will look something like this:



The real relationship may be more or less peaked; that is what she is really after. She expects that to get a good idea of the relationship between consumer liking and viscosity, she will need more than four products. So, Maria decides that she will create and provide eight products to James for the consumer test. She sets aside time in the pilot plant to create these products.

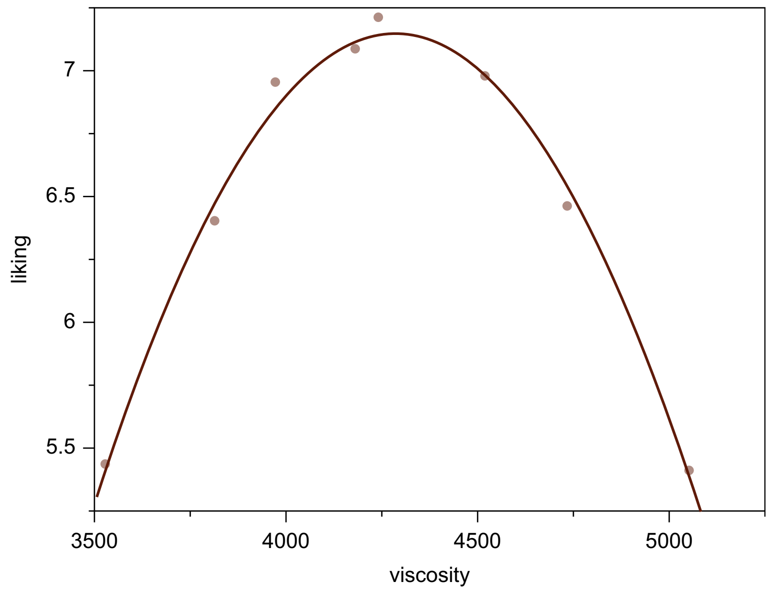
Maria produces the products for the consumer across the anticipated viscosity range of interest by varying the steam pressure in seven equally spaced levels. As a test, she runs a duplicate of the middle steam pressure setting to see how similar these samples will be. She expects their measured viscosities to be similar. If they are very different, then something unexpected happened when she produced the products, and she will need to determine what happened, correct it, and make the samples again. She recalls that Dr Wang had repeatedly mentioned how inadvertent bias can influence research results, so she does not produce the products by varying the steam pressure in an increasing or decreasing order, but shuffles these levels in a random way to minimize any effect on viscosity over the production order.

She measures the viscosity on these eight samples, and they do indeed have quite a large range. The samples from the two runs with the same target steam pressure have similar viscosities, so she feels good about her execution of the trial. The extremes of this range are well beyond what she thinks consumers would tolerate, but she is guessing somewhat, since she really does not know how the consumers will react. If she confidently knew the consumer tolerance ranges, she would not need this consumer study!

James is very specific on how he has consumers evaluate the eight product samples. He conducts the consumer evaluation at a time when the products Maria has produced are of an age (shelf life) typical to that of when consumers buy and use the product. Each consumer evaluates the samples in a different order based on a pattern James has created to minimize tasting order bias. The average acceptance scores for each of the eight products are calculated and presented to Maria. She has already had viscosity measurements made on five bottles of each of the eight samples and sets up a table with the average viscosity for each of the samples in one column and the average acceptance score in a second column.

viscosity	liking
3528	5.44
3812	6.41
3971	6.96
4178	7.09
4240	7.22
4516	6.98
4728	6.47
5048	5.42

Maria creates a graph of these two measures and sees the anticipated inverted U relationship. She knows that to fit a curvilinear relationship between viscosity and acceptance, she will need to perform a regression analysis as she had done before, but now include the squared term for viscosity as well in order to describe the curved nature of the relationship.



Polynomial Fit Degree=2

liking = 6.6279473 + 0.0001238*viscosity - 3.0274e-6*(viscosity-4261)^2

Summary of Fit

RSquare	0.991681
RSquare Adj	0.988354
Root Mean Square Error	0.077472
Mean of Response	6.49875
Observations (or Sum Wgts)	8

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	3.5774777	1.78874	298.0256
Error	5	0.0300098	0.00600	**Prob > F**
C. Total	7	3.6074875		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	6.6279473	0.251913	26.31	<.0001*
viscosity	0.0001238	5.94e-5	2.08	0.0915
(viscosity-4261)^2	–3.027e-6	1.241e-7	–24.38	<.0001*

Maria is familiar with the analysis output, since she has had recent experience with linear regression models. She observes in the graph that her fitted line is close to the data points. The R-Square for the regression model is very close to 1, another indication that the model represents the individual data points well. And, her model has a very small p-value for both the overall model and the parameter estimate for the squared term, which indicates the curvilinear relationship between viscosity and liking.

Maria shares her analysis with James, and they both agree that the results show a strong relationship between viscosity and liking and that the model can be used to develop a target and specification limits. The peak of the curve is for a liking of around 7.15 and corresponds to a viscosity close to 4300; this will serve as the target. He indicates that liking mean differences less than 0.50 are not likely to be important. Using the model to predict liking, the viscosity range of 3900 to 4700 ($\pm$ 400 from the target) should have a predicted liking not less than 6.65 (0.50 less than the ideal value of 7.15 target).

Now that she feels confident with the specification she has developed, Maria turns her focus on developing a monitoring and control strategy. She arranges a call with Dr Wang to discuss her basic understanding of SPC and get some direction.

Dr Wang is her usual cheery self. “How is my favorite BBQ sauce developer?” is her greeting to Maria when the call connects.

Maria catches her up on her recent work developing the viscosity specification and her need to now create a monitoring and control strategy. Dr Wang recommends a specific SPC textbook that Maria can use as guidance, but advises that Maria alone will be able to gather the necessary information on how to best choose and implement the tools for her project. Maria purchases the textbook that Dr Wang recommends and dives in.

The primary SPC tool is a control chart. The chart displays a measure over time. For Maria, the measure of interest is viscosity. There are control charts to monitor individual viscosity measurements or averages of small groups to understand if a process is centered on the target. Variability should also be monitored over time in a separate control chart. Tracking variability may be just as important as tracking viscosity, since if a process becomes more variable, even if it is still centered at the target, then a defective product (a product with viscosity outside the specification range) is more likely to be produced. There are also control charts for discrete characteristics, such as number of defective samples or number of defects per unit.

In the ideal world, every bottle produced would be measured for viscosity. With this perfect information, an operator would be able to adjust the process if it was shifting to lower or higher viscosity levels. Measuring viscosity for every bottle produced would be prohibitively expensive, so Maria will have to choose both how often samples will be taken for measurement and how many samples to take each sampling period. Both sampling frequency and sampling amount affect the resources that will be needed for this monitoring program, as well as the effectiveness of the program, so Maria wants to get it right!

She reads that it is best to collect a small number of samples at each sampling time. A control chart of the sample averages (called an X-bar chart) is used to evaluate process centering. A control chart of either the sample standard deviations (called an s-chart) or the sample range (called an R-chart) is used to evaluate process variation. There are two reasons that this is best:

1. Short-term process variation is best understood when the samples are taken at the same time, rather than accumulated over a number of sampling times.
2. The control charts operate best when the data follows a normal distribution, and even if the individual data points do not follow a normal distribution, sample averages do based on the central limit theorem in statistics.

Sometimes, the measurement process is so involved or costly that only one sample can be measured at each sampling time. In this case, a control chart of the individual measurements (I-chart) is used to monitor how the process is performing with respect to the target. To monitor variation, a moving range chart (MR chart) is used, with the range calculated over a number of sampling time points. If the individual data points do not follow a normal distribution, then other types of charts may be needed.

Maria thinks about her BBQ production. It is easy enough to measure a number of samples taken at the same time since the measurement process is relatively easy (and reliable now!) and takes no more than a minute.

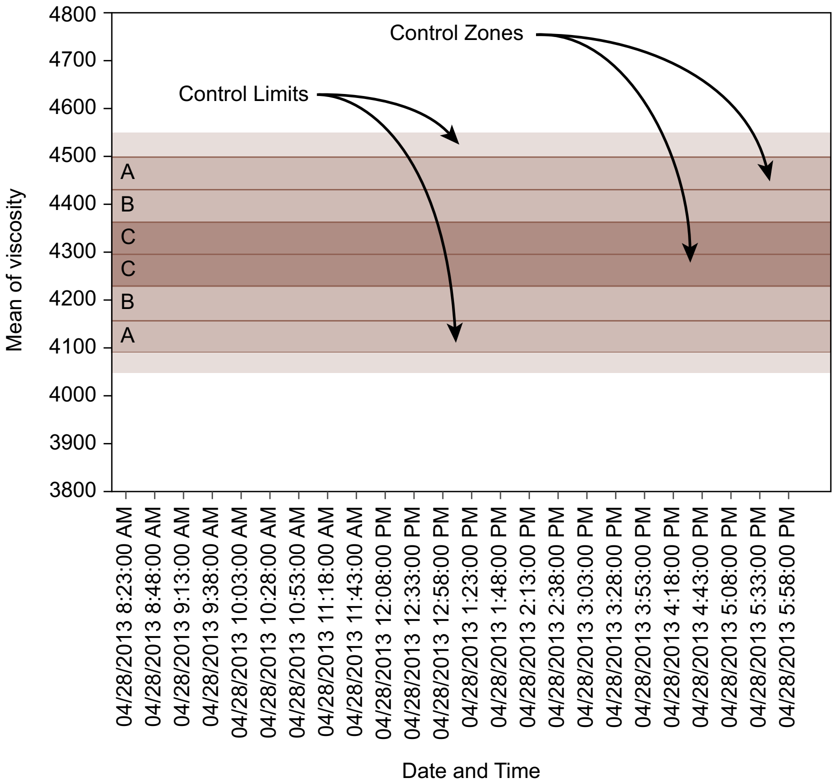
She would like to have samples taken relatively frequently so that if the process is shifting up or down, an adjustment can be made before too much product is produced. Since the line operator will be collecting the samples and then those samples will be taken to the Quality Lab to make the measurements, she cannot have the line operators performing this task more frequently than every half hour, considering the number of lines and lab resources. There are five filler heads in each bottling line, and it occurs to Maria that collecting samples from each filler head would be prudent—as there maybe some differences between them. So, her plan is to propose that five samples, one from each filler head, be taken and measured every half hour.

Maria arranges to discuss her plan with the line operators at their next weekly meeting. Also present at this meeting are the Quality Lab staff members. She talks through her proposed monitoring plan and asks for questions. The operators see no problems with pulling the samples at specified times, and the lab staff can handle the measurements if the lines stagger their sampling times. The measurement values will be transmitted back to each line operator through the data management system, and each line operator will see the results within 10 min of the sampling time.

Maria is feeling good about it all until Connie, one of the line operators, says "So, it is still up to us to decide when and how to adjust the process, right?"

Maria was so focused on the data collection that she had forgotten about establishing clear guidelines for process adjustments! She asks the group how and when they currently make adjustments. She recalls that she had been told that each operator has his/her own strategy, and those adjustment strategies differ widely. She tells the group that she wants them to operate consistently and will need to do some additional work to determine what that practice will be.

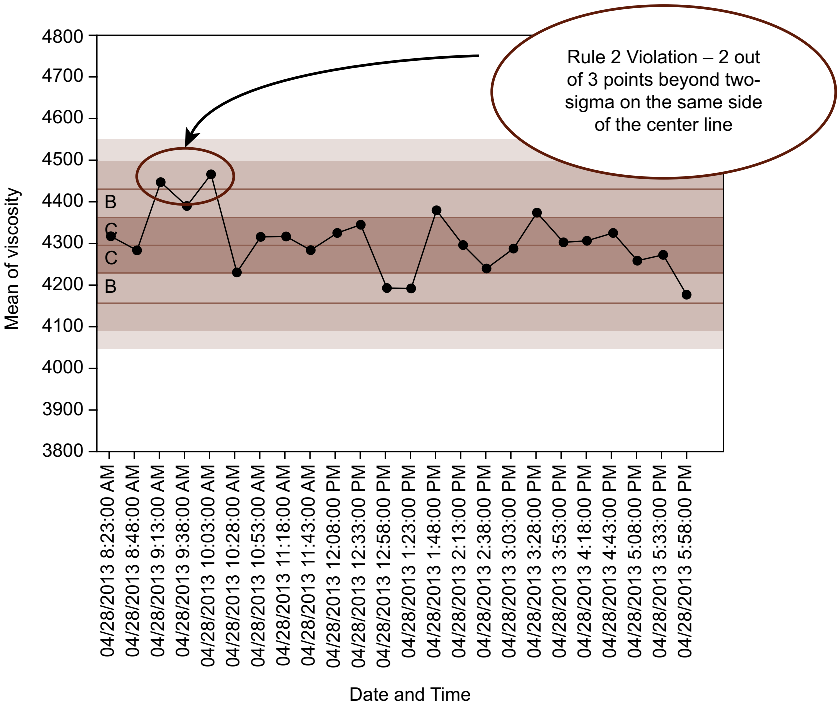
Maria remembers that she had seen a chapter in her SPC text about run rules and thinks that it may give her some idea on how to react to the control chart output. She looks into this chapter and finds out that there are indeed ways to determine when to make an adjustment to the process based on what is viewed in the control chart. There are several different sets of rules (Western Electric, Nelson, Montgomery, and so on), and as she reads through them, she sees that some of them are the same or similar. They all seem to work in a consistent way. Control zones are created on the charts based on the process standard deviation. Control limits are defined as ±3 standard deviations from the target or center line. A blank control chart with zones looks like this:



Further in the chapter, she sees a summary: the top five run rules for SPC charts to detect process shifts are as follow:

1. One point outside of the control limits
2. Two out of three points beyond two-sigma on the same side of the target
3. Four out of five points beyond one-sigma on the same side of the target
4. Eight consecutive points on either side of the center line
5. A steadily increasing or decreasing pattern of six points

These are all situations that are unlikely to occur if the process is still in control and not deviating from the target and has not changed in some way. A control chart with a run rule violation on rule 2 would look like this:



Now that she has some rules for determining when a process is shifting away from the target, she needs to determine a strategy for process adjustment. It seems best to her to use the information in the data that triggers the rule—that is telling her something about how the process may be shifting. She is also aware that bold adjustments to the process will likely be an overreaction. Whatever she decides, it needs to be easily understood and implemented by the operators.

To use the information from the data, she decides that she will simply average the data points that triggered the rule. For rule 2, this method means averaging the three relevant points in "the two out of three." For rule 3, this means averaging the five data points in "the four out of five." For rules 4 and 5 this means the eight or six points, respectively, that create the pattern. And for rule 1, there is no averaging; the information is contained in the single data point.

Maria feels confident about the regression relationship that she had determined between steam pressure and viscosity in her earlier work, so she decides to use that regression to help her determine how to adjust the process. The equation for the regression line is as follows:

$$\text{Viscosity} = 7388.1 - 13.19 \times \text{Steam Pressure}$$

The average viscosity change above or below the target will determine the increase or decrease in steam pressure. For example, an average viscosity of 4500 determined from the run rule violation is 200 above the target of 4300. Using the equation, Maria calculates that an increase in steam pressure of 15 would bring the viscosity back to the target. It may well be that different actions may be needed for different run rule violations, so as this strategy is implemented, Maria will evaluate process performance and make changes to the correction strategy if the performance is not adequate.

Maria writes up, as a work procedure, the data collection strategy, run rules, and adjustment method using the process data. With the work procedure in place, the current and any future line operators can consistently monitor and make process adjustments when needed.

This page intentionally left blank

CHAPTER 8

Sampling

Maria arrives at the plant one morning, and she becomes painfully aware that her statistical process control (SPC) program and the processes that she implemented just a few weeks ago are having an impact. During the previous night's production, many run rule violations were triggered. Although process adjustments were made according to the SPC strategy, the viscosity could not be brought back to target. The production manager placed the entire shift's production on hold. Maria is excited to see that the SPC process is well embraced by the plant management but also sad as she realizes that a large amount of product is put on hold as a direct result of the SPC program. She is wondering what could have caused the run rule violations and the product hold. She goes to the Quality Lab and bumps into the analytical lab manager, Chad, and the plant quality manager, Louise.

Maria: "Chad and Louise, I heard about the product hold!"

Louise: "I'm so glad to see you here today. We really need your help with this situation."

Maria: "Do you know what happened last night? Why is the product on hold?"

Chad: "The Brookfield measurement device in the Quality Lab failed verification at the end of the shift. That means it stopped measuring viscosity correctly at some point after the verification at the end of the previous shift. The operator made process adjustments based on these inaccurate measures and may have inadvertently caused the viscosity to go off target. I just fixed the Brookfield measurement device and verified it with the device in the R&D analytical lab. So this will no longer be a problem, and hopefully today's production will be fine."

Louise: "I'm getting a lot of pressure from Corporate to release the product on hold. I received four phone calls, asking me about our last night's BBQ production, including threats for not loading the product on the trucks this morning. So, I need your help to develop a sampling plan to determine if last night's production is of acceptable quality and can be shipped!"

Maria: "I understand. Our sales team must be feeling pressure from our customers, especially now that the BBQ grilling season has begun.

Statistics for Food Scientists
http://dx.doi.org/10.1016/B978-0-12-417179-4.00008-1

The plant has been running at full capacity for a month to prepare for the peak-season, and now that the demand has arrived, we cannot afford not to fulfill orders. But we cannot send bad quality product out the door either. Do you have any more specifics on the product on hold?"

Louise: "Could we sample 10 bottles and release the product to get Corporate off my back?"

Maria: "We need to do what is right for the customer. That includes not releasing the product if it does not pass our quality standards. I would not think that this could be considered a physical hazard to consumers?"

Louise: "No. The product viscosity during our night shift was not in line with our specifications. The device was reading higher than it should have so we are worried that the product produced is not thick enough. I already pulled around 10 bottles to measure the viscosity. Do you think that those will be enough to make a statistically sound decision whether or not to release the product?"

Maria: "Are you able to isolate a time frame when the instrument was working properly?"

Chad: "We had not thought about that, but you make a good point. Potentially we could release the first half of the production lot, since the line operator started making adjustments to the process around the middle of the shift based on the bad measurements from the Brookfield viscosity meter. Product viscosity was likely under control before the operator started making adjustments."

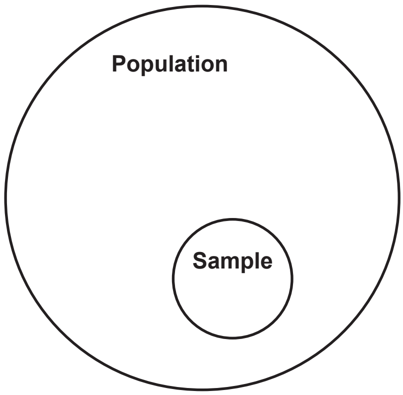
Maria: "If it is true that during the second half of the shift we are uncertain about the results and during the first half our process was still in control, we may consider evaluating the product on hold in two separate phases. We will do our due diligence with the first half of the shift to confirm our hypothesis and sample more frequently from the second half of the shift where we suspect the quality issue may have occurred due to potentially incorrect adjustments. For starters, could you please find out how much product was made during the first and the second halves of the shifts? What is the overall product dollar value? What kind of sampling resources do we have? How many samples can be evaluated per hour?"

Louise: "I will collect that information right away."

Maria: "Thanks Louise, if you find out anything new during your snooping around, please call or email me."

Maria rushes out of the lab to the cafeteria. In order for her to function properly in this crisis, she needs a double espresso. As she waits for the coffee brewer to fill her cup, she thinks about how she may be able to determine the number of samples needed and how to collect the samples. But the answers to both of these questions are based on many considerations, including scope, risk, measurement capability, and the specific production line.

Maria is interested in evaluating a certain number of samples from last night's production (the population) and drawing conclusions about what last night's production defect rate likely is.



She recalls what she learned about sampling at school. Dr. Wang had always said that it is not just how many samples, but also how they are collected that is important. She digs out her notes from the class where sampling had been discussed.

HOW MANY SAMPLES?

There are two common approaches to determining the number of samples evaluated:

1. **Acceptance sampling plan**: As the name suggests, this approach is typically used to accept or reject the lot of raw materials or finished product based on predetermined acceptable and unacceptable defect levels.
2. **Minimum sample size sampling plan**: The approach is used to quantify the maximum level of a defect with a specified level of confidence.

Acceptance sampling plans are based on a statistical hypothesis test set up to answer the question: Should a lot be accepted and shipped based on the observed number of defects. There are two types of incorrect decisions that can be made, and risks associated with each:

On one hand, an unacceptable lot can be falsely accepted; while on the other hand, an acceptable lot can be falsely rejected. The hypothesis test has certain components. First, the null and alternative hypotheses are defined based on the proportion of defectives. The null hypothesis is shown next:

$$H_0 : p_{\text{defective}} = p_1$$

p_1 is called the acceptable quality level (AQL). For example, a 1% defect level could be considered acceptable.

The alternative hypothesis is this:

$$H_1 : p_{\text{defective}} = p_2$$

p_2 is called the rejectable quality level (RQL). For example, a 5% defect level could be considered unacceptable.

The choices for p_1 and p_2 are specific to each sampling situation and should be dictated by business considerations.

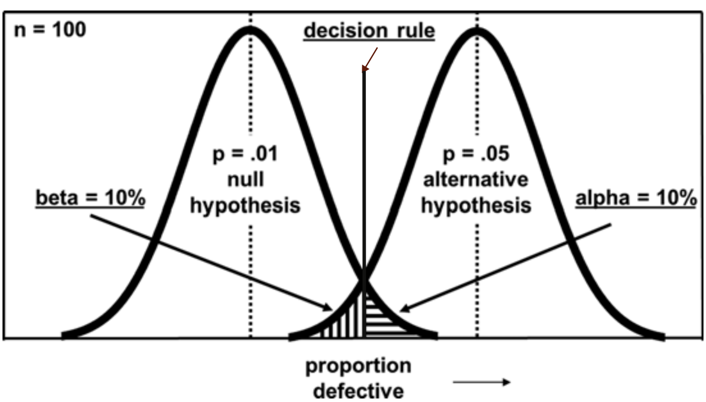
The acceptance sampling plan can be designed to hold the probability of accepting a "bad" lot and rejecting a "good" lot based on specific levels. These probabilities are as follows:

Alpha: This is the probability of rejecting a lot with a defect level of p_1. For example, this probability could be set at 10%.

Beta: This is the probability of accepting a lot with a defect level of p_2. For example, this probability could be set at 10%.

A sampling plan with AQL = 1%, RQL = 5%, alpha = 10%, and beta = 10% would have a sample size of 100 and a decision rule to reject the lot if three or more defectives were found. This example has a balanced risk profile where alpha = beta. There is willingness to tolerate the same amount of risk of not releasing a good lot (a cost to the business), as of releasing a bad lot (a risk to the business). As with the choices for p_1 and p_2, the choices of alpha and beta are specific to each sampling situation and should be dictated by business considerations.

A visualization of the hypothesis test looks like this:



If the decision rule is moved to the right, alpha will decrease and beta will increase. If the decision rule is moved to the left, beta will decrease and alpha will increase. If the sample size is increased, both curves will become narrower and more peaked, reducing both risks equally.

Minimum sample size plans allow for the smallest possible sample to be taken to answer the question: Is the percent defective products no greater than a predetermined maximum allowable defect level (MADL), and at what level of confidence? The table here is used to determine the number of samples:

	Number of Defective Units in Sample (**c**)					
Confidence Level	0	1	2	3	4	5
80	160/n	299/n	428/n	552/n	673/n	791/n
90	230/n	389/n	532/n	668/n	799/n	927/n
95	300/n	474/n	630/n	775/n	915/n	1051/n
99	461/n	664/n	841/n	1005/n	1160/n	1311/n

To estimate the number of samples needed, a maximum allowable defect level and the desired level of confidence is chosen. Then, a minimum number of samples can be determined using the table above.

For example, in a sample of $n = 300$:

- If zero defects are found, we can be 95% confident that the true defect rate is no greater than 1% (300/300 = 1%).
- If one defect is found (out of 300 samples), we can be 80% confident that the true defect rate is no greater than approximately 1% (299/300 = 0.997%).

For example, if a 95% level of confidence is desired and no more than 1% defect rate is acceptable, then a minimum of 300 samples would be evaluated, and no defects would be allowed to accept the production lot. Alternatively, 474 samples could be evaluated, and one defect would be acceptable to yield the same conclusion of 95% level of confidence and no more than 1% defect rate estimate.

HOW TO COLLECT SAMPLES?

How samples are collected is just as important (and perhaps even more important) than the number of samples. It is important to collect samples without any form of bias. A well-chosen smaller number of samples is better than a poorly chosen large number of samples. Randomly choosing units for inspection from the lot will ensure that there is no bias, but this can be difficult to implement with a production lot of units in multiple cases and cases stacked into many pallets. If a defect can appear anywhere in a lot, then a stratified sampling approach with strategically chosen subgroups may ensure that units selected represent the production lot without bias.

Maria's incoming email signal chimes. The email from Louise with information about the product on hold has finally arrived. The lot on hold consists of 120 pallets, 12,000 cases, or 120,000 bottles (10 bottles per case and 100 cases per pallet, with 10 cases on each of 10 layers) of BBQ produced over an 8-hours shift. The pallets are numbered and arranged so that the order in which they were produced is easy to follow. The last 4 hours are more highly suspect, while the first 4 hours carry a lower risk. There is no consumer safety hazard, but some consumers may express dissatisfaction from a less viscous, or thinner, product.

The hold product value is substantial, but if a large product inspection is required, then the sampling costs will outweigh the benefits due to the large labor costs.

Maria thinks about how to approach the situation at hand. The defect levels in the acceptance sampling plan example from her class notes seem to be much larger than would be tolerable in this situation. And she is not sure that Louise or anyone else would easily be able to determine an AQL and RQL for this situation. The minimum sample size plan approach seems like it would be easier to use in this situation since the concepts of maximum allowable defect level and confidence level are more easily understood.

Maria emails Louise and explains MADL and confidence level and suggests that since the first half of the shift is less suspect separate plans, with different MADL levels, should be created.

Louise states in the email that for the more suspect being in the second half of the shift, she is willing to accept at most a 0.5% defect rate at 95% level of certainty. Since the first half of the shift is less suspect, the business is willing to have reduced sampling and looks to a recommendation from Maria.

Based on the information Louise provides, Maria recommends starting by sampling from the first 4 hours of production so some customer orders can be fulfilled quickly, since this part of the lot has a higher chance of being released. Maria proposes the following sampling strategy:

1. For the first 4 hours of production, sample 60 bottles. If no defects are found, then the team will be 95% confident that the defect rate is no larger than 5%, and this portion of production can be shipped. This sample plan may not seem to be very rigorous, but the sample size is higher than what is collected in the current SPC program. Since the

team feels that the production during this timeframe is likely of acceptable quality, the proposed plan is essentially repeating the measurements now that the Brookfield has been fixed.

2. For the second 4 hours of production, sample 600 bottles. If no defects are found, then the conclusion will be the defect rate is no larger than .5% with 95% confidence. Then, the second 4 hours of the production lot can be shipped as the defect rate would fall in line with the plant historical tolerance levels for viscosity. However, if one, even one, defect is identified, then this portion of production will not be shipped.
3. How will the 60 or 600 samples be chosen? Equal numbers of bottles should be from each hour of production. Since the production is continuous, a sample chosen from every n^{th} bottle to get to the target sample size would be desired. However, to minimize sampling effort, Maria recommends sampling cases and then taking a sample of bottles from each case.
   a. For the first 4 hours of production, 60 bottles are required. Since there are 120 pallets, a case can be taken from every other pallet; this will result in 60. Select one bottle from each case to provide the 60 bottles. Since the potential for a defect is not related to the layer or position on the pallet, any case from the top layer in the pallet can be chosen, as this will be easiest to pull without taking apart the pallet.
   b. For the second 4 hours of production, 600 bottles are required. Again there are 120 pallets, so sampling one case from each pallet and selecting five bottles will provide the 600 bottles. Again selecting any case from the top layer is easiest.

Maria's sampling plan is quickly put into effect. She joins the team to help select the bottles and helps Chad record the viscosity levels for each of the selected BBQ bottles. Some of the viscosity levels are high and some are very low, but none of the samples seem to be outside of the identified specification limits. The first 4 hours of production is quickly approved for release. A few hours later, the second 4 hours of production is also approved for release.

Maria and the team are pleased with the results. Since this is the first time she has constructed a sampling plan, she decides to set up a call with Dr Wang for a quick debrief on the situation. It is a few days before they are able to connect.

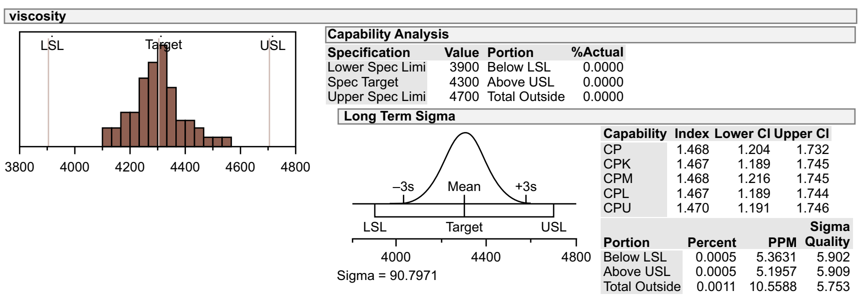
After listening to Maria's update, Dr Wang says, "Maria, you have done a great job and I'm sure your company appreciates your efforts. But I think that your sampling plan could have been more efficient. When you have a continuous response like viscosity, you often need far fewer samples than with a binary (or pass/fail) response. You may be able to use as few as 30 representative samples to be able to estimate the defect rate if the data is normally distributed. There is a great benefit in using continuous response data if at all possible and exploiting that continuity."

"Wow, I didn't even think of that," responds Maria. "There was such a push to have an answer that I was happy to find the table I used from your class notes!"

Dr Wang continues, "If you create a histogram of the viscosity data, you should be able to estimate this by adding the target and specification limits as if you were calculating process capability measures. Then, it will be a part of the standard output."

Maria is excited, "I think I know how to do this in my software. I wish I had called you before we started, we could have finished much faster. But I'll be better prepared next time thanks to you!"

After hanging up, Maria performs the analysis Dr Wang had suggested. She creates a histogram using the 60 viscosity measurements from the first 4 hours of production. She adds the current viscosity target and upper and lower specification limits.



Capability Analysis

Specification	Value	Portion	%Actual
Lower Spec Limi	3900	Below LSL	0.0000
Spec Target	4300	Above USL	0.0000
Upper Spec Limi	4700	Total Outside	0.0000

Capability	Index	Lower CI	Upper CI
CP	1.468	1.204	1.732
CPK	1.467	1.189	1.745
CPM	1.468	1.216	1.745
CPL	1.467	1.189	1.744
CPU	1.470	1.191	1.746

Portion	Percent	PPM	Sigma Quality
Below LSL	0.0005	5.3631	5.902
Above USL	0.0005	5.1957	5.909
Total Outside	0.0011	10.5588	5.753

From examining the histogram, Maria notes that the data seem to follow a normal distribution and confirms that there are no actual values both below and above the specification limits. The long-term projection based on the normal distribution suggests only a tiny percentage (0.0005%) of product produced would be below the specification limit. She repeats this analysis with the 600 viscosity measurements from the second 4 hours of production with similar results. She certainly would not have needed this many measurements to understand the viscosity distribution for the second half of the held product. This is a valuable lesson she will not forget!

Maria now feels much more confident in her sampling skills and hopes that she will not need to use them again any time soon!

CHAPTER 9

Process Capability

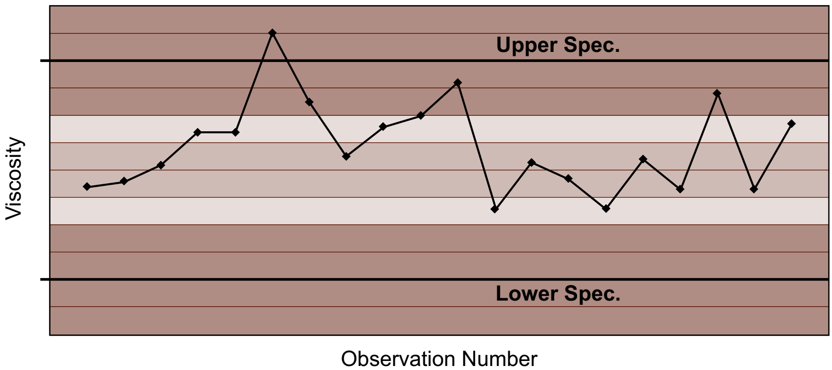
With the issue of product dispensation past, the team focus on delivering the best product on consistent basis has resumed. Lisa's recent feedback had been that there was a positive impact on the BBQ lines since the statistical process control (SPC) system was implemented. She asks Maria to present the dynamic SPC monitoring system implemented earlier on the BBQ lines to the operations management team.

Lisa: "This will be a great opportunity to demonstrate to the operations management team all of the improvements we have made. The operations Vice President, my boss, along with the managers of the other production facilities will be present along with the research vice president, Steve's boss. You should make sure that Steve can attend. Their next meeting is next Tuesday; I can make sure you are added to the agenda. Let's plan on a half hour, with a 20 min presentation and time afterward for questions."

The rest of the week went by quickly and the weekend past even faster, as Maria spent most of it setting up tables, decorating them with flowers, and arranging seating for her Grandma's 90th birthday celebration. Most importantly, everyone at the party had a good time, and Grandma Maria (who Maria is named after) had the time of her life on the dance floor. When Maria arrived at work Monday morning, her immediate task was to create her presentation for the senior management team. She decided that a live demonstration of the SPC monitoring system would be most impactful and made arrangements to have that as the focus of her presentation. She spent most of the day preparing for the presentation, and by the time she left the office, she felt confident.

Maria was a bit nervous as she did not personally know most of the meeting participants. Armed with her SPC knowledge and a smile, she entered the room and set up the projector, so that everyone could see her screen. Then she projected an example of the "live" BBQ viscosity SPC monitoring system for one of the production lines. In the demonstration she showed how a run rule violation would trigger an alarm and how the system would then advise on the corrective action.

Statistics for Food Scientists
http://dx.doi.org/10.1016/B978-0-12-417179-4.00009-3



Maria began with defining SPC as a process that helps operators monitor the production performance as well as a system that helps operators determine when to adjust a process and to what degree. Most plant managers were eager to be early SPC adopters and to implement SPC on their respective production lines. At that moment, something unexpected happened: Joe, the operations Vice President, asked Maria, a question in a soft voice.

Joe: "I'm puzzled that your system only looks at immediate control. I assume that the data you are capturing is saved on the plant server. Can't you summarize this data to produce a report that Lisa can use to demonstrate overall performance versus specifications on each line? We can use this report to determine which lines perform better or worse and focus improvement efforts on those that are performing poorly. Also, we monitor other quality characteristics like pH, salt, and sugar levels. Shouldn't we use this SPC system to monitor and control these too? The performance reporting would include these measurements."

Maria is stunned that she had not considered performance reporting! But before she can respond, Lisa comes to her rescue.

Lisa: "We had considered this of course. We are so excited about what we have started that we did not want to wait to present it to you and the team. And since this will take considerable resources to put in place, we wanted to make sure everyone is in agreement."

Steve jumped in: "I have pretty good knowledge of process capability measures and had planned to suggest these to Maria. I wanted her to keep focused on the dynamic monitoring to make sure that was correct before moving to the reporting stage."

Lisa: "Maria, Steve, and I will partner to implement a performance reporting system for the monthly plant performance meetings. We will focus on viscosity for now, and once we have your and the team's approval, we can replicate the approach for the other quality measures. We'll present this at next months' meeting."

The very next day, Maria meets with Steve. He shares some of the positive feedback he had received on Maria since she started working with the team at the plant. Maria acknowledges the feedback and relates the plant team's desire to implement SPC and adopt it across the floor.

Steve: "Maria, I'm really proud with what you have been able to achieve in such a short time."

Maria: "Actually I was blindsided by Joe's comments yesterday. I feel that I should have anticipated that there is so much more to do. You and Lisa saved my butt for sure! If I remember correctly, last month you led a seminar on process capability."

Steve nods his head in an agreement, as Maria continues.

Maria: "However, at the time of the presentation, process capability did not mean much to me, and honestly, some of the information went way over my head. Now, my absent-mindedness during your presentation is haunting me. Can you give me a quick summary of what it is about? I want to be able to put it to practice soon. I know I can do this using my statistics software. Dr Wang had suggested to look at the distribution of the viscosity measurements of the held product to project the proportion of a lot that would be out of specification. Lisa said we will partner to implement this, but she is so busy that I think that it will be mostly up to you and me."

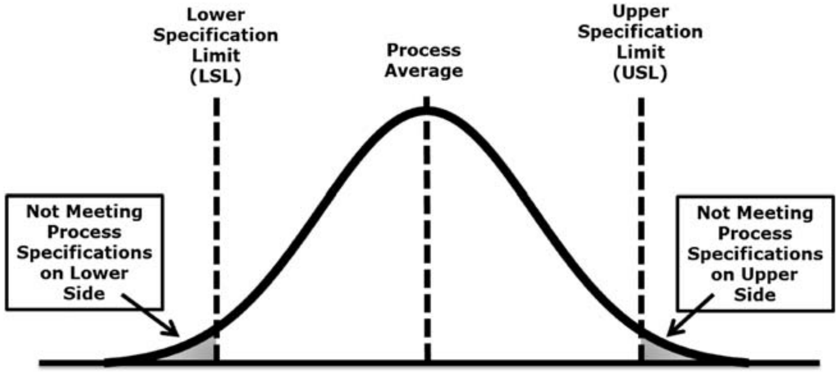
Steve responded, "I'll be glad to help you out with this. Actually, I have a lot of passion for process capability measures. They are great measurement tools that are underutilized in our plants. Since you have established a great rapport with Lisa and her teams, it looks like now is the time for a successful implementation. Let me introduce the concepts to you, and you can bring forward process capability measures to your friends at the plant."

Luckily, although the name may be perplexing, the process capability approach is not. Process capability analysis is a standard output from many statistical packages, and the concepts are very simple.

Process capability analysis allows researchers to find the linkage between the voice of the customer (what the customer wants, based on defined specification limits) and the voice of the process (what the process can deliver). To quantify the relationship, a capability index is calculated:

$$\text{Capbility Index} = \frac{\text{Voice of the Customer}}{\text{Voice of the Process}}$$

Capability indices help us focus on the nature of the problem: Can we provide what the customer is asking us to deliver? In the example below, the customer tolerance (upper specification limit—lower specification limit) is smaller than what the process is capable to deliver, and some proportion of the finished product is expected to meet customer expectations.



For process capability evaluation to be meaningful, the specifications used must be customer based. There are four groups of potential customers:

1. Consumers
2. Customers that sell a company's products (i.e., a grocery store chain)
3. The government (through regulations)
4. The company producing the product (i.e., manufacturing efficiencies, food safety requirements)

If a specification is not based on a customer need, then two adverse conditions may occur:

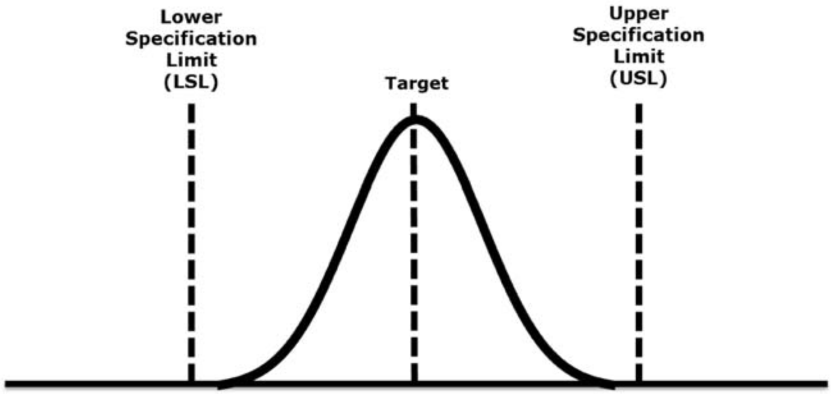
- Under-specification: We think we are producing good products, but in fact, we are failing to meet the customer needs.
- Over-specification: We are operating to overly tight requirements that were not generated by the customer and have no impact on the customer experience.

The approach for estimating if a process is capable is straightforward. Three things are needed to measure process capability:

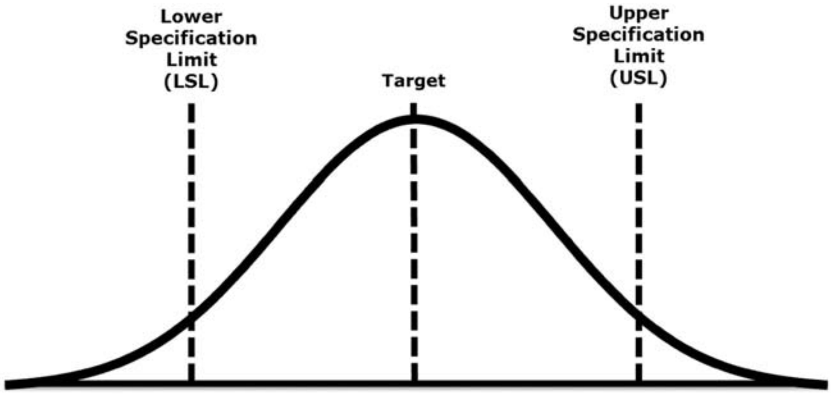
- The customer specifications
- The process typical values—what the process is centered on (measured by the process average)
- The process variation—how much variation there is around the center (measured by the process standard variation)

A process researcher compares the data representative of the process to the predefined customer tolerances (specifications). Four possible outcomes from comparing the process performance with the customer specifications are shown below:

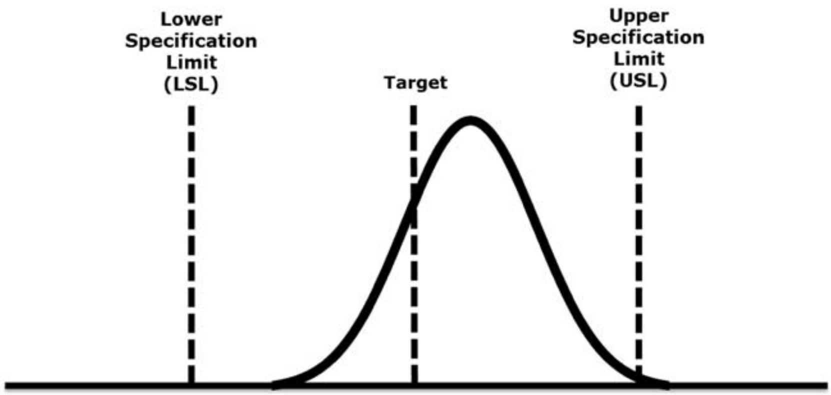
On Target And Capable

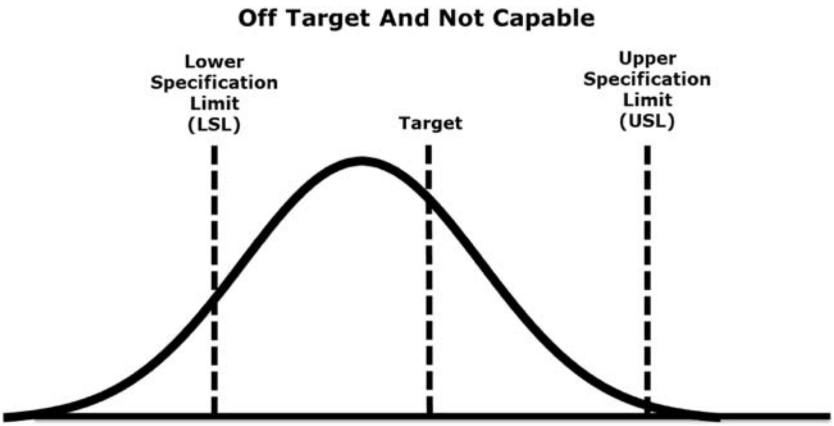


On Target But Not Capable



Off Target But Capable





Process capability gives us visibility to the breadth of the process—a measure of the process performance in relation to the customer specifications, while also providing a perspective on the process center—a measure of the process performance in relation to a target (often the middle point between the lower and upper specification limits).

If more than a visual interpretation is needed, process capability indices are used to further quantify the process ability to meet the specifications. These indices allow the researcher to predict defect levels.

To properly set up the analysis and to measure capability, one needs the following:

1. A clearly defined customer specification. Upper Specification Limit (USL) - the maximum acceptable level and Lower Specification Limit (LSL) - the minimum acceptable level. Both of them ought to be meaningfully defined by the customer.
2. Data collected as part of a SPC plan—collected at time intervals over an extended period of time that comes from a stable process and follows a normal distribution

Two capability indices frequently used are C_p and C_{pk} (and similarly P_p and P_{pk}).

C_p and P_p relate customer tolerance (USL minus LSL) to the process variability. However, they do not take into consideration process centering.

$$C_p = \frac{\text{USL} - \text{LSL}}{6 \times \sigma_{\text{Within}}}$$

$$P_p = \frac{\text{USL} - \text{LSL}}{6 \times \sigma_{\text{Overall}}}$$

In these formulas the σ's represent process standard deviation. C_{pk} and P_{pk} take into account the location of the mean (process center), and a performance index is calculated relative to the specification limit nearest to the mean.

$$C_{PK} = \min(C_{PL}, C_{PU})$$

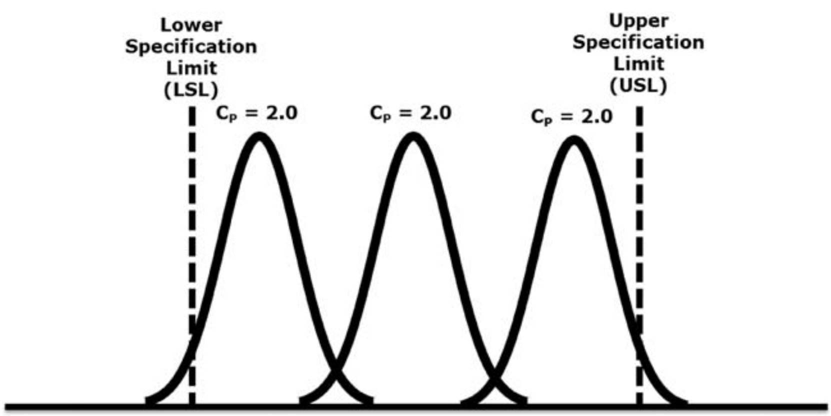
$$C_{PU} = \frac{USL - \overline{X}}{3 \times \sigma_{Within}} \quad C_{PL} = \frac{\overline{X} - LSL}{3 \times \sigma_{Within}}$$

$$P_{PK} = \min(P_{PL}, P_{PU})$$

$$P_{PU} = \frac{USL - \overline{X}}{3 \times \sigma_{Overall}} \quad P_{PL} = \frac{\overline{X} - LSL}{3 \times \sigma_{Overall}}$$

The difference between C_p and P_p and C_{pk} and P_{pk} is in the way the standard deviation used in the measure is calculated. C_p and C_{pk} represent short-term performance; P_p and P_{pk} represent long-term performance.

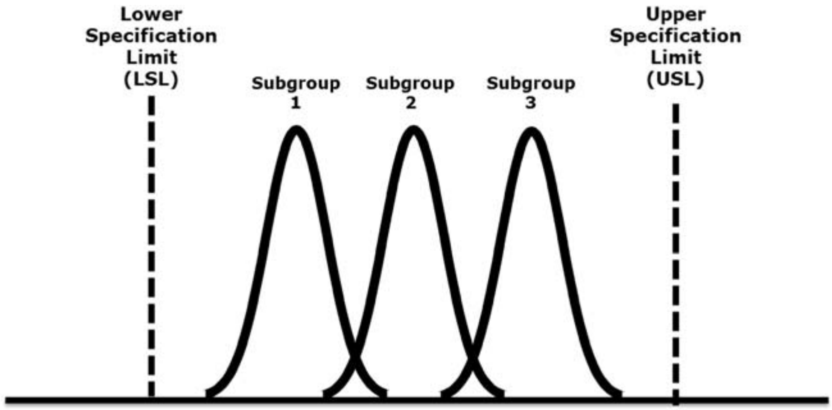
C_{pk} and P_{pk} quantify how many defects are being produced (product out of specification). C_p and P_p quantify the spread of the process in relation to the consumer-defined specification range, a measure of process *potential*, but they do not quantify how many defects are actually produced. This is evident in the illustration here:



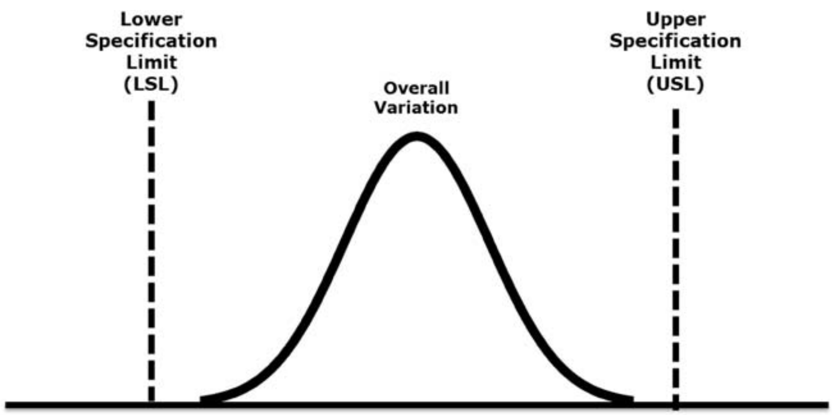
Short-term variation is sometimes thought of in very small time intervals. Process control data are often collected in subgroups in frequent time intervals, for example, five samples every 15 min. In these instances, C_{pk} uses a pooled standard deviation based on these subgroups; P_{pk} uses one overall standard deviation across the time intervals without regard for subgroupings. Subgroups can be determined by a number of factors:

- Time period: Five samples taken every 15 min.
- Lot numbers: Six samples taken from each ingredient blend.
- Tooling: Eight samples each taken from a different filler head.

Processes can vary even between short periods of time, due to ingredient changes or variation or physical sources such as heat build-up. Each process can have its own unique structure that can be a source of variation, and the data used to evaluate process capability should be collected to adequately reflect this structure. Examples of elements of structure that should be considered include filler heads, locations in ovens and driers, mixing tanks, and physical tooling such as dies.



C_{PK} is calculated using the pooled standard deviation of the individual subgroups



P_{PK} is calculated using the overall standard deviation across the subgroups

P_{pk} is also sometimes used to estimate performance over much longer periods of time to account for seasonal variation in ingredient streams and atmospheric conditions like temperature and humidity in production facilities that have little control of these elements.

C_p can be considered as a measure of process entitlement—how well we can potentially perform if our process is centered and all sources of long-term and structural variation are eliminated.

General guidelines for all of these capability measures are the following:

- Excellent process capability: measures > 1.67 ($P_{pk} = 1.67$ corresponds to an expectation of observing one defect per million units manufactured)
- Good process capability: measures between 1.33 and 1.67 ($P_{pk} = 1.33$ corresponds to an expectation of observing 63 defects per million units manufactured)
- Fair process capability: measures between 1.0 and 1.33 ($P_{pk} = 1.00$ corresponds to an expectation of observing 2700 defects per million units manufactured)
- Poor process capability: measures < 1.0

The larger the C_p index, the "tighter" the process is running compared to the specification limits. For example, a manufacturer may have a good control of the production process, or customers have a wide tolerance in their definition of acceptable product. In such situations, there may be an opportunity for a manufacturer to operate in the part of the specification range that requires a minimal cost or effort.

After Steve finishes his lecture to Maria, they consider how they could use capability measures in the performance reporting system.

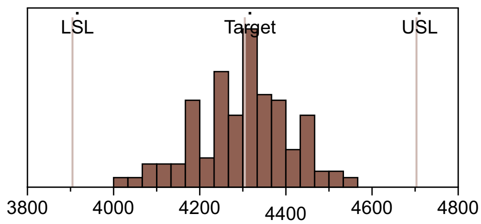
Maria: "Our current SPC data collection plan is already well designed for understanding process capability. I had not thought of the idea of structure when we came up with it. It just felt right to include data from all filler heads each time, and to collect samples at roughly equal time intervals. We can easily look at long- and short-term variation."

Steve: "I think that for each line we can use data from within a shift. We can use C_{pk} to understand the performance of that specific shift and P_{pk} to report performance for the month, which will include variation between the shifts as they have different operators and also long-term variation sources like ingredient batches. C_p calculated within each shift can tell us how far we are from perfection; we can look at the individual measures and may be average them in some way to give us an overall picture."

Maria: "I think my data analysis software will let me calculate these measures. I'll design a report and show it to Lisa first to see what she thinks before we respond to Joe."

Maria pulls the most recent shift of SPC data for viscosity from one of the lines to use as a sample data set. This data has five samples at roughly equal intervals of around 25 min. Her software does indeed calculate the capability measures. She sees that she can look at measures for both short-term and long-term variations.

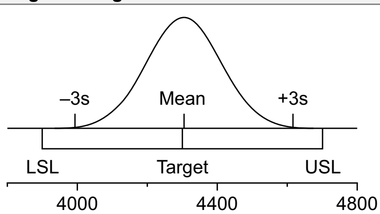
viscosity



Capability Analysis

Specification	Value	Portion	% Actual
Lower Spec Limit	3900	Below LSL	0.0000
Spec Target	4300	Above USL	0.0000
Upper Spec Limit	4700	Total Outside	0.0000

Long Term Sigma

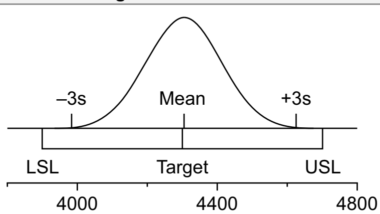


Capability	Index	Lower CI	Upper CI
CP	1.28	1.12	1.44
CPK	1.27	1.10	1.45
CPM	1.28	1.12	1.45
CPL	1.29	1.11	1.46
CPU	1.27	1.10	1.45

Portion	Percent	PPM	Sigma Quality
Below LSL	0.0057	56.8933	5.359
Above USL	0.0066	65.7882	5.323
Total Outside	0.0123	122.6815	5.167

Sigma = 104.131

Short Term Sigma



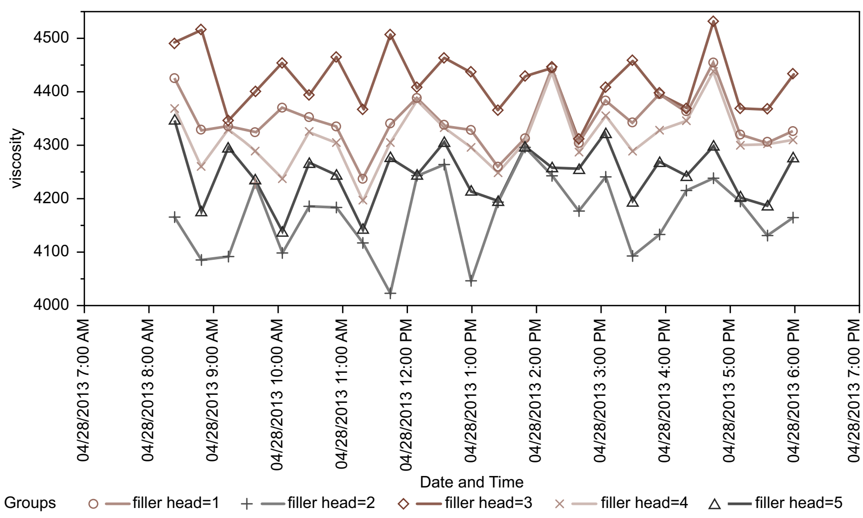
Capability	Index
CP	1.25
CPK	1.24
CPM	1.25
CPL	1.26
CPU	1.24

Portion	Percent	PPM	Sigma Quality
Below LSL	0.0082	82.0890	5.269
Above USL	0.0094	94.3186	5.234
Total Outside	0.0176	176.4077	5.073

Sigma = 106.633

Maria likes the visual display in the output. She sees that her individual samples all fall within the viscosity specification; both in the histogram with the specification limits added and in the "% Actual" for both below and above, the specification limit is zero. She also sees that the percent projected in both the long-term sigma and short-term sigma portions of the output is not zero but very small. Steve had explained that since the capability indices are calculated using the sample average and standard deviation, this could occur—in this case, the indices suggest that even though none of the individual bottles measured were outside of the specification limits, there is a small proportion of the production that likely is. Maria sees that the output uses C_p and C_{pk} in the long-term sigma and short-term sigma parts of the output, a difference from Steve's terminology. This is not a problem since the calculations are correct, but she should make sure this terminology difference does not confuse everyone moving forward.

The biggest surprise for Maria is that there is little difference between short-term and long-term performance, at least within this shift. Sigma, the estimate of the standard deviation, is slightly larger short-term than long-term. That would suggest that the variation in viscosity is due to the filler heads rather than over time. She creates a graph of the data over time to see if this is truly the case. Her graph looks like this:



From her graph, it is obvious that the filler heads are delivering different viscosities. Filler head 3 with the diamond symbol is consistently highest, and filler head 2 with the plus symbol is consistently the lowest. There is some overlap between the other filler heads, but it seems obvious that filler head is a large source of variation, at least for the shift of data that Maria had selected for her process capability analyses.

Maria pulls data from a longer time period for that line and repeats the capability analyses and graphs the data with similar results. She then uses ANOVA to confirm her findings.

Summary of Fit

RSquare	0.847981
RSquare Adj	0.803367
Root Mean Square Error	46.17518
Mean of Response	4301.856
Observations (or Sum Wgts)	120

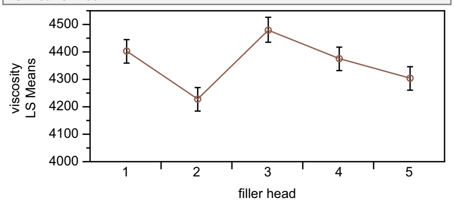
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	27	1094191.4	40525.6	19.0
Error	92	196157.5	2132.1	Prob > F
C. Total	119	1290348.9		<.0001*

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
filler head	4	4	895426	105.0	<.0001*
Date and Time	23	23	198766	4.1	<.0001*

LS Means Plot



LSMeans Differences Tukey HSD

α= 0.050 Q= 2.78263

Level					Least Sq Mean
3	A				4484
1		B			4407
4		B			4377
5			C		4308
2				D	4231

Levels not connected by same letter are significantly different.

The ANOVA analysis clearly shows that though viscosity does vary significantly over time, much of the viscosity variation is due to the filler heads (the sum of squares due to filler head is much greater than that for data and time). In the Tukey's analysis, she can clearly see how the filler heads viscosity averages differ; confirming that filler head 3 delivers the thickest product and filler head 2 the thinnest.

Maria is excited both with her findings on the filler head differences and the potential performance assessment report that she can provide to the operations management team. She sets up a meeting with Lisa to review.

Next morning, as Lisa walks through the plant doors with a hot cup of coffee, the receptionist tells her that she has a visitor waiting in her office.

Maria: "Good morning Lisa, I hope that you do not mind me coming in so early and waiting in your office. I have been putting together the monthly performance summary using the capability measures with Steve's guidance. And, I found something important that may help us to reduce viscosity variation. I was so eager to share the results with you that I could barely sleep last night."

"Slow down, Maria," said Lisa. "Catch your breath, or you will get both of us in trouble. Then, tell me how was Granny Maria's birthday last weekend. Was she dancing like no one is watching?"

Maria: "So sweet of you to ask about my Grandma. I think that Granny had one too many martinis with her friends, and later that night she had the time of her life on the dance floor. Rumor has it that her party this year was the best yet!"

Lisa smiles and continues the conversation: "So what have you discovered, Maria? Will it help our performance for viscosity?"

Maria shows Lisa her graph of viscosity by filler head over time and her ANOVA analysis and describes how it demonstrates that the filler heads deliver different viscosities. Lisa agrees with the interpretation.

Lisa: "I can have the maintenance crew look into the filler heads on this line to determine why this is happening. However, before I do so, it would be a good idea to see if this is the case for all of the lines. You are becoming quite a process detective!"

Maria: "I'll perform similar analyses with data from the other lines. But now, let me show you what I have been working on for the performance summary. Steve schooled me in process capability measures, and they are a perfect complement to our SPC system. SPC is typically used to establish control through increasing visibility to short-term process variability and establishing directives for corrective actions. Process capability measures allow us to monitor long-term process compliance to specifications. I think this is what Joe had asked about at the meeting the other day."

Lisa: "Yes, and it sounds like you have something for the management team. Have you had a chance to look at our viscosity data using these measures?"

Maria: "Take a look at my laptop's screen, and let me explain what you are looking at."

Maria describes the process capability output she had generated the day before and how the individual capability measures are interpreted. Lisa too likes the visual aspect of the analysis.

Maria: "It was this analysis that led me to look into the filler heads. I would have thought that short-term variation would have been much smaller than the longer term variation. When it was not, I recalled that we collected data from each of the filler heads at each time point for the SPC data collection. Fortunately, the filler head is part of that data record."

Maria continues after taking a sip from her water bottle: "First take a look at the histogram, this shows the distribution of last month's viscosity measurements for two production days exported from our SPC daily data. The LSL of 3900 and the USL of 4700 represent the specification limits that we determined from our consumer research. The target is simply the midpoint between the LSL and the USL. What do you think of our performance?"

Lisa: "I know that last month did not have any quality concerns with viscosity, and your chart suggests that we did well; nothing is outside of the specification limits."

Maria: "That is correct; we had no individual bottles with measured viscosity outside of the tolerance. But, the chart under "Long-Term Sigma" shows what we would expect from the process in a long run. Based on our current process variation, we would expect that around 120 bottles out of each million would fall outside of the specification range. Slightly more above the upper limit than the lower limit. The larger the process capability measures C_p and C_{pk} are, the less likely it is that the process will produce defects."

Lisa: "Are you saying that we shipped out of spec product?"

Maria: "The data analysis suggests that based on the measured data collected, a small proportion of production could have been out of specification. Maybe this is acceptable? If not, then we may have to do some work to reduce the viscosity variation and ultimately the proportion of product produced out of specification."

Lisa: "Maria, these measures are exactly what Joe and the management team is looking for. I love your analytical and practical approach to problem solving. I think that for this line the risk of producing product out of specification is low. I'm not sure about the other lines. Can you produce similar analyses for the rest of the lines so that we can identify where our biggest opportunities are?"

Maria: "Yes! With these analyses we will be able to make some intelligent decisions as to where to put our resources for maximum impact."

Lisa: "If we can implement these in a report that we can call up for any line for any time period, then we will have fulfilled Joe's request. I cannot describe to you how happy I am with the approach that you shared with me. It seems that process capability is a quality professional's 'best friend.' I wish that I knew about it years ago."

Maria: "Thank you for your time, Lisa, and for the opportunity to implement these measures at the plant. I'll work with Steve and our plant IT applications developer to create the reporting structure."

CHAPTER 10

Design of Experiments Foundation

Maria's process control strategy has been in place for a number of months now with mixed results. In many instances, when viscosity starts to drift, the process changes in steam pressure based on her regression equation are effective. But in other instances, this strategy does not seem to work, and the operators are making adjustments to the steam pressure without improving performance.

Maria recalls that when she had process and ingredient data collected to determine her process control strategy, several of the measures were related. Maybe controlling viscosity is not as simple as monitoring and adjusting steam pressure. She repeats the data collection exercise she performed when she first developed regression relationships and reaches similar conclusions. Steam pressure has the strongest relationship with viscosity, and the ingredient and process measures are correlated with each other.

Maria sets up a call with Dr Wang to discuss her dilemma. Dr Wang is as usual accommodating to her "favorite BBQ sauce student."

Dr Wang listens patiently to Maria's update, and then says, "I'm impressed with how much you have been using your data analysis skills!"

Maria thanks her and continues, "While I have made some great advances in our understanding of viscosity, I think I may have run into a wall. The ingredient and processing variable relationships are probably more complicated than I thought."

"I agree that the variable relationships are probably quite complicated," said Dr Wang. "The best way to understand how they work is to study them independently and see how they work together. How familiar are you with the approach of statistical design of experiments?"

"You may have mentioned this in one of our classes," Maria tried to search her memory, "but otherwise, I know nothing."

"I'm really not an expert in this area," said Dr Wang. "You would be better served taking a class so that you will have a proper understanding. I'll send you links to some that I know to be good."

Maria receives information on classes in statistical design of experiments from Dr Wang and sees that one is being offered nearby in the next few

Statistics for Food Scientists
http://dx.doi.org/10.1016/B978-0-12-417179-4.00010-X

weeks. In the class description, she sees that they will use the same statistical software that she has been using for her recent data analysis work. She checks with her manager Steve, who readily agrees that this will be time and money well spent.

Maria returns from her three-day class full of excitement. She creates another cheat sheet that summarizes what she learned from the course to help prepare her for a discussion with the team.

Maria summarizes her key takeaways from class as follows:

- In a statistically designed experiment, factors (usually ingredient levels and/or processing conditions) are systematically varied so that their effects can be quantified in the most efficient way.
- Statistical analysis of designed studies focuses on the development of models that relate how formulation and/or process variable changes affect the responses.
- These models can provide a wide variety of information, such as the following:
  - An ordering of important formulation/processing variables
  - Optimal product formulations (subject to cost and production constraints, if applicable)
  - Consumer-based specifications
- There are a number of experimental types each with a different objective:

  Screening designs are used to collect data to quantify the effect of factors, with a goal of either reducing the number of factors under investigation or identifying factor ranges for further study.

  Response surface designs are used to collect data to develop a detailed model of the space defined by the factor ranges. These models can be used to determine optimal factor settings.

  Mixture designs are special forms of screening or response surface designs where the factor levels are constrained to a sum total.

  Robust (Taguchi) designs are used to determine the factor settings that reduce end product variability due to uncontrollable noise.

In the class, she saw specific examples of each of these design types. The instructor led the class through the process of creating the designs for each of the examples using the statistical software and then provided data to follow through the data analysis for each example. Maria is eager to use the design of experiments approach to better understand how elements of the BBQ production process affect viscosity.

She recalls the list of formulation and processing variables that her team had come up with in their first attempt to better control viscosity. There were three ingredient variables: sugar level, tomato paste level, and fruit paste level. There were also three processing variables: pump pressure, steam pressure, and line speed. When she had analyzed the process data collected over a period of time, there were correlations between many of these variables. But they can be independently changed for her experiment, which is an important consideration in the statistical design of experiments approach.

Maria ultimately wants to find the combination of these factors that will produce the target viscosity and also develop a control strategy that will work consistently. In other words, she realizes that she wants to create a response surface design and use the resulting model to determine the optimal factor settings. At this point, she does not want to rule out any of the six ingredients and process variables as potential factors in any experiments she conducts on the BBQ process.

Using her software, she looks at her options for a six-factor response surface design. The smallest design she sees is very large with 46 runs—this means 46 different combinations of the six ingredient and processing variables. It seems like a lot of work, and she is not even sure if it can be done. In particular, she doubts that production can be stopped for experimentation for very long. She decides to discuss the matter with the team at the morning production meeting the next day.

The next morning Maria shares with the team, "I have an approach for determining how to better keep viscosity under control. It will require us to stop production and perform an experiment where we systematically change six factors in a very specific pattern. The factors are: sugar level, tomato paste level, fruit paste level, line speed, pump pressure, and steam pressure. There will be 46 different combinations of these six factors. Is this something we can do?"

Alex is the production supervisor for the daytime shift. His eyes widen considerably when Maria mentions 46 combinations.

He says, "I guess anything is possible, but what you are suggesting would mean stopping production 46 times to make changes. Then we will need to run the line at the new levels long enough for the line to stabilize before we collect the samples to measure for viscosity. I would guess that each run would take around 30 min. That is a lot of production interruption! Will the product you produce with all of these changes be something that we can sell? Will viscosity be in specification?"

"I don't know for sure," answers Maria, "So we should probably assume that much of it will not be within specification. So there is a cost to this work for sure. But what we learn from this experiment will mean that future production will be better, so there is a cost advantage in that."

"I will check with Lisa, the plant manager," says Alex. "I'm sure she will see the benefit to some experimentation, but I think she will ask if something less costly will get us the information we need."

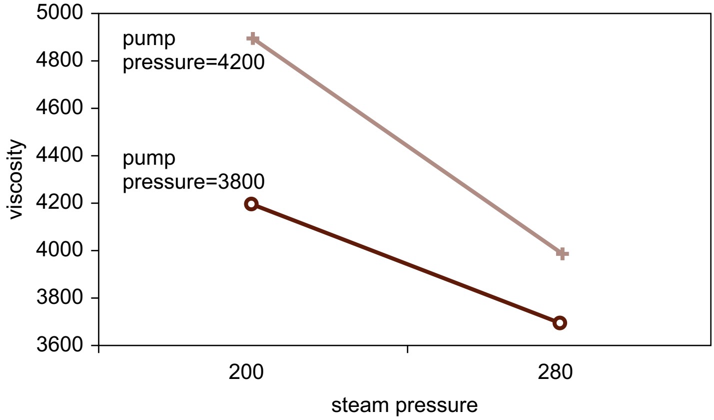
Maria hears back from Alex later in the day. He confirms what he had thought the reaction from Lisa would be. She is willing to devote some resources to experimentation, but would like some additional options from Maria.

Maria had already been thinking that this was a possibility. She had recalled from her class that screening designs were often used as a step before a response surface design to eliminate factors that had little impact. She uses her software to look into screening design options for six factors.

Screening designs typically have two levels for each factor. The designs are factorial combinations of these levels. A full factorial design will have all possible combinations for each factor. For Maria's six factors this means $2^6 = 64$ combinations or design runs. But many screening designs are fractions of the set of all possible combinations. A specially determined fraction may provide sufficient information about the individual factors. These fractions are most always in powers of two: $^1/_2$, $^1/_4$, $^1/_8$, and so on. For Maria's six factors, a half fraction would be 32 combinations or runs, a quarter fraction would be 16 runs, and an eighth fraction would be eight runs.

Maria recalls from her class that the concept of design resolution was important in determining design fractions. This concept considered the idea of factor interactions. A factor interaction means that the effect of one factor is different for different levels of another factor. For Maria, an example would be that the effect of steam pressure would be different for different levels of pump pressure. She tries to imagine what this would look like.

From her earlier work, Maria had determined that increasing steam pressure caused a decrease in viscosity. An interaction between steam pressure and pump pressure would mean that the rate of decrease would be different for different pump pressures. A graph of this might look like this:



Maria considers that each factor could interact with any other factor. With six factors, this means that there are 15 potential pairwise interactions! Maria thinks that though there is the potential for this to happen for some of the factor pairs, the goal of her screening stage is to identify the most important factors. If there are interactions between these important factors, then they would be identified in the subsequent response surface design.

The screening design options are identified by their resolution in her software. Resolution is listed as a number: 3, 4, 5, or greater. She looks back at her notes from her class and finds these definitions:

- Resolution 3 designs: Main effects are not confounded with other main effects but are confounded with pairwise interactions.
- Resolution 4 designs: Main effects are not confounded with other main effects or pairwise interactions; pairwise interactions are confounded with each other.
- Resolution 5 designs: There is no confounding between main effects, between pairwise interactions, or between main effects and pairwise interactions.

Confounding is when effects cannot be separated from each other. To her, it seems that a resolution 4 design would be best, since she is most interested in the main effects to reduce the number of factors for her second design. At least the main effects would not be confounded with any pairwise interactions.

Her software output shows that there is a 16-run resolution 4 screening design for her six factors. This seems like a good option for her first design. She also looks into her options for response surface designs with fewer than six factors. A design with just two factors could be as small as 10 runs. A design with three factors could be as small as 16 runs. If with screening design she can eliminate half of the factors for the next phase, then she would have reduced the total number of runs for experimentation from 46 runs to 32 or less. This two-step approach seems like a better strategy.

At the next morning's production team meeting, she sees that Lisa, the plant manager, has joined the meeting. Maria presents her new experimentation strategy to the team.

Maria explains, "If we conduct our experimentation in two stages, we will likely need fewer resources and have less interruption in production. Stage one would consist of a screening design with all six factors, each at a low and high level. This design would have 16 runs and the goal would be to identify the factors that are most important in affecting viscosity. Stage two would consist of a response surface design with only the important factors identified in stage one. I'm hoping that we can focus on three or fewer at this stage. If so, the second design could be as few as 16 runs. So our total effort will be around 2/3 of the 46 runs I had originally proposed."

"I like this approach," Lisa says. "We do need to fix this viscosity problem, and this is a much more efficient plan. According to Alex, these two experiments would each take 8 h to execute. I'm willing to devote one of the production lines for a full shift for each experiment. Alex, would you please work with Maria to set this in motion."

Maria is pleased that she can proceed. She and Alex agree to meet later in the day to work out the details of the design and make a schedule for the execution.

CHAPTER 11

Screening Experimental Designs

Maria and Alex meet the next morning to plan the screening design. Maria had already decided on a 16-run resolution 4 design for her six factors. She and Alex will also need to decide on the ranges for each factor.

Maria: "A while back, I had one of the operators collect both ingredient and process variable measurements during production. The variation in each of these might help us set a range for these factors in the design."

Alex: "Over how much time was that data collected? Do you think it represents production overall?"

Maria: "It was just one production shift. It would be good to know how representative that is. For the processing variation, it may be representative, but I don't have a good idea about ingredient variation."

Alex: "I think the operators would have some insight. Let's ask a few of them."

Maria and Alex walk onto the plant floor and visit with a few of the operators for the production lines. There is consistency to what they hear:

1. For ingredients: Tomato paste, fruit paste, and sugar levels can all vary quite a bit, even within a single shift of production. Actually, the operators change these levels based on how they think a batch looks in the mix tank.
2. For processing factors: The operators change the pump pressure as they see fit to make a good product. Lately they were instructed to adjust steam pressure according to the SPC plan. But, prior to that they had used their intuition to adjust the steam pressure. Line speed varies based on line performance, but higher line speeds are desirable to meet production targets.

Armed with this information Maria and Alex again discuss the factor ranges. They both agree that the data that Maria had already collected probably represents typical variation for these factors. If anything, the

Statistics for Food Scientists
http://dx.doi.org/10.1016/B978-0-12-417179-4.00011-1

variation may be more due to the operator adjustments than anything else!

Maria then looks back at the process data she worked with earlier. She calculates the ranges of each of the measures to come up with these ranges for the design:

- Steam pressure: 200–280
- Pump pressure: 370–420
- Line speed: 22–40
- Sugar level: 60–66
- Tomato paste: 1000–1100
- Fruit paste: 7–13

Maria creates her design based on these levels. It looks like this:

steam pressure	pump pressure	line speed	sugar level	tomato paste	fruit paste
200	370	22	60	1000	7
200	370	22	66	1100	13
200	370	40	60	1100	13
200	370	40	66	1000	7
200	420	22	60	1100	7
200	420	22	66	1000	13
200	420	40	60	1000	13
200	420	40	66	1100	7
280	370	22	60	1000	13
280	370	22	66	1100	7
280	370	40	60	1100	7
280	370	40	66	1000	13
280	420	22	60	1100	13
280	420	22	66	1000	7
280	420	40	60	1000	7
280	420	40	66	1100	13

There are two levels for each of the factors. For steam pressure, the first eight runs are at the low level of 200, and the next eight runs are at the high level of 280. For pump pressure, the first four runs are at the low level of 370, and the next four are at the high level of 420; then this pattern repeats for the remaining eight runs. There are also repeating patterns for line speed and sugar level, and somewhat odd patterns for tomato paste and fruit paste.

Maria recalls from her class the importance of randomizing the experimental design runs. This ensures that any unknown factors that may influence the viscosity will not be mixed up with the factors in the design. If for some unknown reason the viscosity increases or decreases over the time the design runs are executed, then with the design above the increase or decrease would look like an effect of steam pressure. Maria is fairly sure that the design runs could be executed in any order but thinks it is best to confirm this with Alex.

Another concept that Maria recalls from her class is replication, or repeating certain design runs. This is an important check—the resulting data from these runs should be similar, and if they are not, this would indicate a problem in the execution of the design runs or possibly with the measurement system. Often the repeated runs are at the center of the design space so that the design is still balanced. This factor level combination is called the center point, the middle level of each factor. This will certainly add to the size of the design, but Maria feels that this is important enough to devote additional resources.

She discusses both randomization and replication with Alex. He agrees that replication is important here since this is the first time that any experimentation has been done at the production facility.

Alex: "I'll need to confirm with Lisa that we can make commit the additional resources, but I think that she will understand the importance too. It will add an hour to the shift when we run the experiment, so I'll ask the operators to stay on to complete the experiment. Executing the experiment in a randomized order should not be a problem for these factors. Fortunately, we are not changing the cook temperature in this design, since we would want to make as few changes as possible to that. Any time we change cook temperature, it takes nearly 30 min to reach the new set point. That would virtually double the execution time for each design run."

Maria: "I think in class the instructor mentioned that randomization can sometimes be restricted for these types of situations. I'm glad we don't have that complication for our very first designed experiment!"

A short time later, Alex emails Maria confirming that Lisa has agreed to the added center point replicates. He adds that after consulting the production schedule, the experiment can be executed the following week. Maria creates a modified version of her earlier design with the added runs and randomizes the run order.

steam pressure	pump pressure	line speed	sugar level	tomato paste	fruit paste
200	370	22	66	1100	13
280	370	40	66	1000	13
280	420	22	60	1100	13
240	395	31	63	1050	10
200	420	40	66	1100	7
280	370	22	66	1100	7
200	420	40	60	1000	13
200	370	40	60	1100	13
200	370	40	66	1000	7
280	420	40	66	1100	13
280	420	22	66	1000	7
240	395	31	63	1050	10
280	370	22	60	1000	13
200	370	22	60	1000	7
280	420	40	60	1000	7
200	420	22	60	1100	7
280	370	40	60	1100	7
200	420	22	66	1000	13

Maria wants to leave nothing to chance for her first experiment. She is confident that Alex will ensure that the execution on the line will be conducted properly. She wants the same for the viscosity measurements that will be the response in the experiment, so she sets up time to discuss this with Chad, the analytical scientist she had worked with a few months back on the viscosity measurement system. She describes the upcoming experiment and the upcoming execution process.

Chad: "I can see why you will want to ensure that you have reliable data since a considerable amount of resources will be spent on the execution. If Lisa is willing to devote these resources, she is expecting that the resulting information will be useful. I can make the needed viscosity measurements for your study rather than interrupting the daily operations of the plant analytical lab. We should discuss the experimental execution and how many samples you will need to have measured."

Maria: "I had not thought much about the number of samples we would be collecting. Since Alex is coordinating the execution of the design runs we should include him in this discussion. I'll set up a meeting where we can go through the details."

Alex responds to Maria's email request suggesting a meeting the following day, mentioning that he needs more time to work out his execution plan. When Maria, Chad, and Alex meet late the following morning, Alex has a notepad with several sheets filled.

Alex: "It took some work to figure out how best to execute the design runs, but I like the plan I've worked out. Connie is probably our best operator, so I want to use her and her production line. Her line is scheduled for sanitation next Tuesday night, so we can execute the experiment on Wednesday during her shift. I want us to start with a clean line since we are changing ingredient levels along with the processing conditions."

Alex continues: "We will create a separate batch for each design run with the ingredient levels specified in the design. The mixing and cook cycle should take about 15 min for each batch. We will pump each vat to be bottled with the steam pressure, pump pressure, and line speed specified in the design. That too will take about 15 min. Since each vat is emptied to a holding tank before bottling, we can rinse the batch tank and fill it with the ingredients for the next design run while the previous design run is being bottled. While the next batch is mixed and cooked, we will rinse the bottling line and set it at the steam pressure, pump pressure, and line speed for the next design run. We will repeat this sequence until we have completed all of the design runs."

Maria: "Wow you have really thought this through! It sounds like a great plan, but how does Connie feel? She may have a very long day if this does not go as planned."

Alex: "Connie is committed to seeing the execution through. She will likely get some overtime pay and also knows that if we can better understand how to control the production lines it will make her life easier in the future."

Chad: "I too am committed to seeing this design through regardless of how long I have to stay to make the measurements. We need to determine how many samples and how to collect them during the bottling process. I'm sure you are expecting more than one bottle measured for each design run."

Maria: "I'm sure there is some level of variation in viscosity between the bottles no matter how well we control the mixing, cooking, and bottling. For our process monitoring, we collect samples from each of the five filler heads each time. Since it takes around 15 min to bottle a batch, maybe we can collect 10 bottles, one sample from each filler head at two times during the bottling and have Chad measure them. Is that a reasonable ask?"

Alex feels that this will not be a burden for Connie to pull the samples, and Chad has no objections to the number of measurements he will perform, so all are agreed to the data collection plan. Next Wednesday looks to be a very busy day!

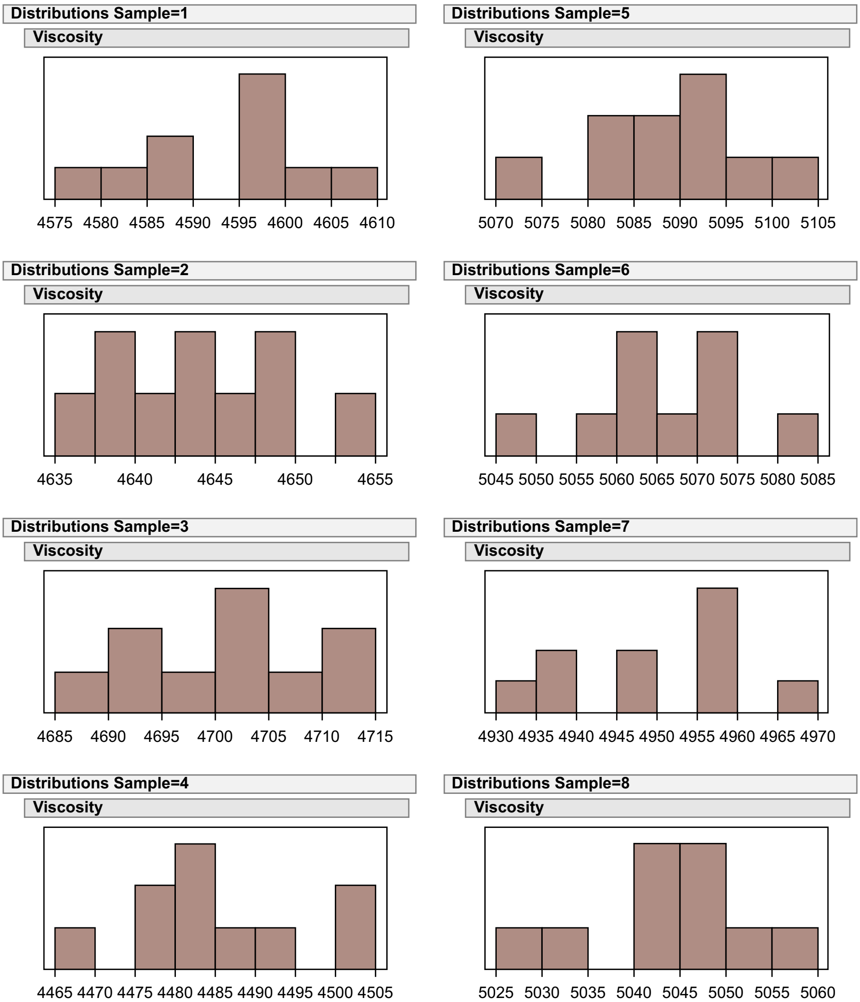
Maria: "I'll plan to spend the entire day with you all. If something unexpected comes up, I can lend a hand."

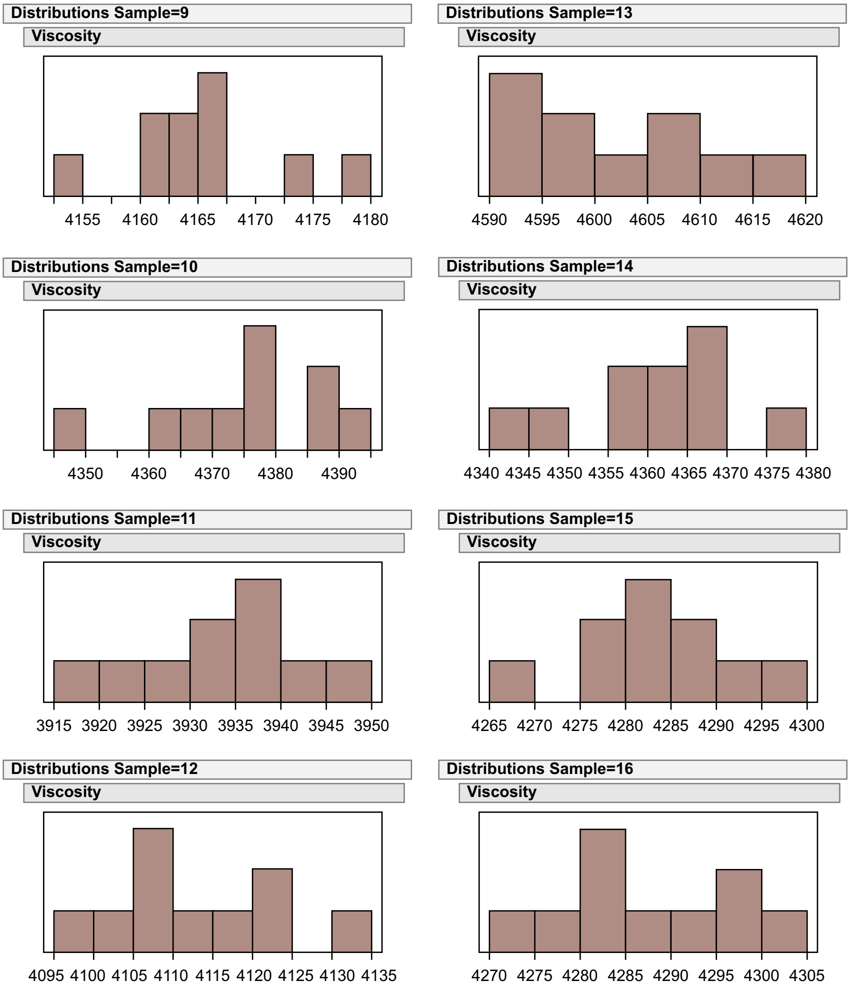
The design execution day turns out to be very intense, with a lot of work. And while the execution did not go perfectly and a few times they had to stop the line to make the factor adjustments, there were no major issues. Connie, Alex, Chad, and Maria felt good about the day overall and hoped that they will have useful information.

Maria transfers Chad's viscosity measurements for each of the design runs to her spreadsheet. She has 10 viscosity measurements for each of her 16 design runs. The raw data looks like this:

Sample	V 01	V 02	V 03	V 04	V 05	V 06	V 07	V 08	V 09	V 10
1	4588	4585	4599	4607	4595	4581	4601	4596	4577	4595
2	4639	4641	4648	4639	4654	4637	4647	4644	4644	4648
3	4711	4694	4710	4715	4694	4700	4703	4704	4687	4700
4	4500	4483	4487	4483	4493	4478	4468	4481	4477	4504
5	5091	5093	5101	5099	5086	5083	5084	5088	5090	5071
6	5063	5074	5047	5071	5058	5072	5061	5082	5062	5065
7	4939	4934	4947	4959	4958	4967	4960	4938	4958	4948
8	5046	5052	5042	5047	5049	5034	5040	5029	5055	5042
9	4164	4163	4174	4160	4167	4165	4161	4155	4166	4179
10	4389	4379	4374	4387	4393	4378	4369	4375	4363	4346
11	3936	3931	3927	3944	3937	3947	3924	3918	3936	3934
12	4098	4132	4116	4123	4109	4100	4115	4106	4110	4122
13	4606	4598	4590	4615	4608	4604	4598	4615	4593	4591
14	4340	4356	4364	4368	4350	4367	4360	4356	4379	4367
15	4285	4280	4296	4280	4279	4282	4281	4286	4292	4267
16	4283	4282	4291	4282	4286	4279	4301	4275	4296	4297

She again starts by visualizing the individual viscosity measurements for each of her 16 design runs. She creates histograms of the 10 viscosity measurements by run.





Maria looks at these histograms to see if there is any indication of skewness in the distributions. She is also hoping to see that the variation in measurements for the samples is similar. The graphs have very different ranges since they also have very different centers (due to the factors that were varied in her experiment she hopes!), so it is difficult to see if this is true. She quickly calculates the average, range, and standard deviation in viscosity measurements for the 16 samples and arranges them in a table. Her table looks like this:

Sample	Average Viscosity	Range	Standard Deviaton
1	4592	30	9
2	4644	16	5
3	4702	28	9
4	4485	36	11
5	5089	30	9
6	5065	35	10
7	4951	33	11
8	5044	26	8
9	4165	24	7
10	4375	47	14
11	3933	29	9
12	4113	34	11
13	4602	25	9
14	4361	38	11
15	4283	29	8
16	4287	26	9

The ranges and standard deviations for each of the 16 design runs are similar enough that she feels comfortable in thinking that there are no substantial differences in variability between the runs. The averages are considerably different; these averages are what she intends to use as a response for modeling the factor effects. Maria is happy to see these differences; otherwise, the factors in her design do not likely affect viscosity!

For her resolution 4 design, the intended model for Maria's design has main effects for each of the six factors but no pairwise interactions. Maria fits this model to her experimental data.

Summary of Fit

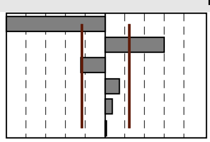
RSquare	0.93
RSquare Adj	0.89
Root Mean Square Error	120.28
Mean of Response	4543.25
Observations (or Sum Wgts)	16.00

Analysis of Variance

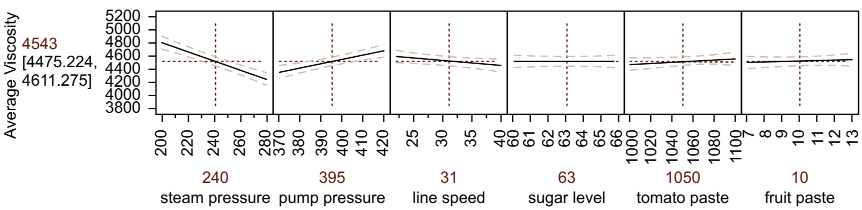
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	1795714	299286	20.7
Error	9	130215	14468	**Prob > F**
C. Total	15	1925929		<.0001*

Sorted Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
steam pressure(200,280)	–278.3	30.1	–9.25	<.0001*
pump pressure(370,420)	166.9	30.1	5.55	0.0004*
line speed(22,40)	–68.5	30.1	–2.28	0.0488*
tomato paste(1000, 1100)	41.2	30.1	1.37	0.2036
fruit paste(7,13)	22.9	30.1	0.76	0.4651
sugar level(60,66)	3.6	30.1	0.12	0.9067



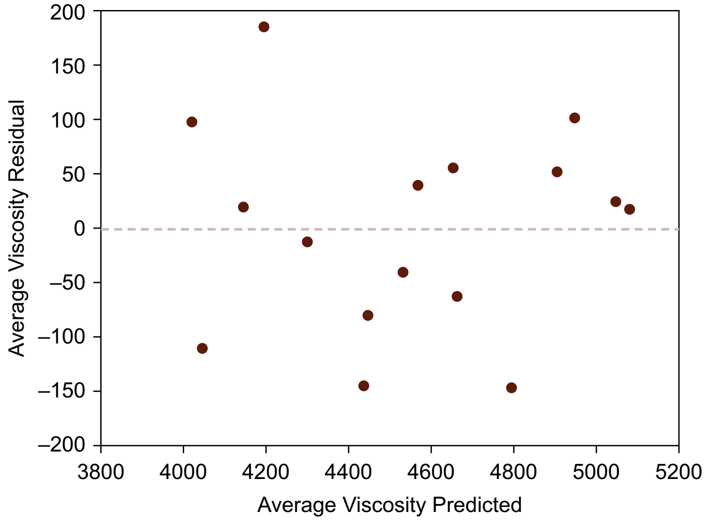
Maria's software orders the factor effects in decreasing levels of significance. These factor effects have been standardized to the range of each of the factors so that the size of their effects can be directly compared. She sees that steam pressure is the most significant effect, followed by pump pressure, and then line speed. The ingredient factors of tomato paste, fruit paste, and sugar level all have effects that are not significant as their *p*-values are all much greater than 0.05, the level commonly used as a cutoff to determine significance. The direction of the factor effects can be seen in the signs of the estimates, but Maria produces a plot to better visualize not just the direction of the effect but also the relative impact of each factor.



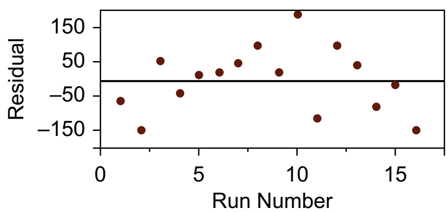
She can see from the slopes of the lines for each factor the different impacts. Steam pressure and pump pressure have pronounced effects. The effect of line speed is somewhat modest. The ingredient level effects are not significant. This is a very important finding for Maria since the ingredient levels only change for each batch, but the processing factor levels can be adjusted as each batch is filled. This suggests more opportunities to control the process to deliver the desired viscosity.

There is one more step in the analysis that Maria recalls from her class—verifying that her model is not missing anything critical by analyzing the model's residuals. Residuals are the difference between the predicted viscosity from her model and the actual viscosity measurements of her design runs. Though these model predictions may not be very accurate for a screening design like hers, there should be no pattern in them if she plots them against the order of the design runs and the predicted values from the model. They should follow a normal distribution centered at zero. Maria's software allows her to easily create these plots since they are an important step in the analysis process.

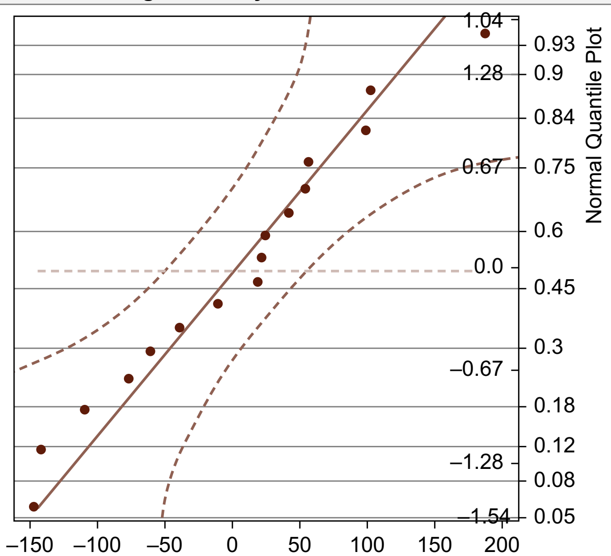
Residual by Predicted Plot



Residual by Run Plot



Residual Average Viscosity Normal Quantile Plot



If in these plots Maria was to see any kind of pattern, it would indicate that something in addition to the factor effects might have influenced the viscosity data from her design. A pattern could be a steady increasing or decreasing of the residuals over the ranges, or more or less variation in them across the ranges. She sees no pattern in either the residual by predicted or residual by run number plot. Maria also creates a normal quantile plot of the residuals. This is a special type of plot to assess if a set of data follows a normal distribution, with the data values on the horizontal axis and normal quantiles on the vertical axis. The points should fall on a straight line if they do follow a normal distribution. Maria sees this is the case in her plot, and that the residuals are more or less centered at zero. Her residual analyses leave her feeling confident that the viscosity differences in her experiment are due to her experiment's factors and nothing else.

Maria collects her data analysis output to create a report for the team. She also starts thinking about her subsequent design to optimize her process and determine the factor levels that deliver the target viscosity. She is not sure whether this new design she should vary only steam pressure and pump pressure or should also include line speed. It would certainly be simpler to vary only the first two. She is thinking that this will be her suggestion for the team. She sets up a meeting with Alex, Chad, and Connie for the next morning to discuss the results.

Maria shares her analysis results with the team and concludes by stating:

"Our design has given us some clear direction on what strongly influences viscosity. Steam pressure, pump pressure, and line speed all have significant effects on viscosity. All the ingredient factors have minimal effects. The effect of line speed is not nearly as large as the other two, so maybe we do not need to vary that in the next experiment. I can design a follow-up experiment using the three or just the two. What do you all think?"

Alex and Chad are quick to point out that eliminating line speed will make the next experiment much smaller. "Lisa will be much happier with less interruption," Alex points out.

> Connie speaks up: "I have noticed that line speed sometimes seems to be important on certain days when I'm operating the line. Also, we sometimes want to run at higher speeds to meet product demand, especially in the warmer weather months when people are barbecuing more. I think we should continue with this factor in the next design so that we will better understand if it interacts with pump pressure and steam pressure. And I'm willing to do the extra work for a larger design."

Alex: "Connie makes a good point. I think if I explain this reasoning to Lisa, she will commit the resources for the larger design. How many runs would we need for three factors versus two?"

Maria: "16 runs would be the minimum for three factors versus 10 runs for two factors. When I first proposed the approach of screening and then the response surface design, I had suggested that if we could get down to three factors for the second design we would still be better off than a 46-run response surface design with all six factors. Please remind her that what we are asking her for now is no different than that!"

Alex emails the team later that day stating that Lisa has agreed to move forward with the three-factor, 16-run design and that he will set up time with Maria to plan this next step.

This page intentionally left blank

CHAPTER 12

Response Surface (Optimization) Experimental Designs

Maria and Alex meet to discuss the response surface design. Again, the first step is to determine the factor ranges.

Alex: "I think we can keep the same ranges from the screening design, since the viscosity data was more or less centered around our target of 4300."

Maria: "I think we will have some choices to make in the design. Let me work with my software and see what I can come up with and we can meet to discuss the options."

Maria starts by looking at the design possibilities for three factors. She sees that the two choices with the smallest number of runs have 15 or 16 runs. She recalls that when she had first looked at this in designing the overall experimental approach, the designs had different names. At that point, she had not looked into the difference between the two, but she had suggested 16 as the minimum number of runs needed since one more run would not make much of a difference in the overall effort. Now, she needs to better understand her options to make the best choice of a design.

Two typically used response surface designs are the central composite design (CCD) and the Box-Behnken design. Both can be used to fit the quadratic response surface model, which has linear and curvilinear (quadratic) effects for each of the factors along with all pairwise interaction effects. But the designs look different, so she wants to understand the differences. And with the CCD, there are some choices that she also needs to understand.

Maria creates a three-factor Box-Behnken design and a three-factor CCD in their generic form with −1, 0, and 1 coding for the low, medium, and high levels. She has the designs created in the standard order so that she can more easily compare them.

Statistics for Food Scientists
http://dx.doi.org/10.1016/B978-0-12-417179-4.00012-3

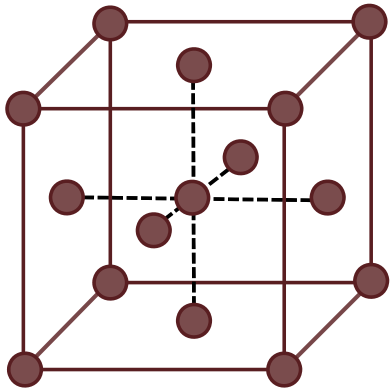
The Box-Behnken design looks like this:

Run	X1	X2	X3
1	-1	-1	0
2	-1	1	0
3	1	-1	0
4	1	1	0
5	0	-1	-1
6	0	-1	1
7	0	1	-1
8	0	1	1
9	-1	0	-1
10	1	0	-1
11	-1	0	1
12	1	0	1
13	0	0	0
14	0	0	0
15	0	0	0

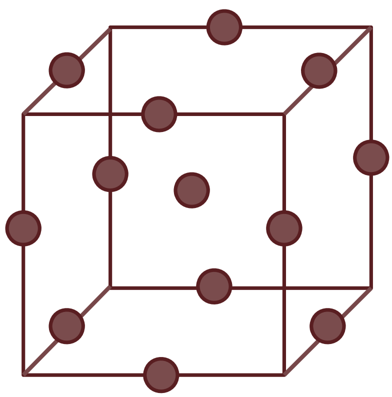
The CCD looks like this:

Run	X1	X2	X3
1	-1	-1	-1
2	-1	-1	1
3	-1	1	-1
4	-1	1	1
5	1	-1	-1
6	1	-1	1
7	1	1	-1
8	1	1	1
9	-1	0	0
10	1	0	0
11	0	-1	0
12	0	1	0
13	0	0	-1
14	0	0	1
15	0	0	0
16	0	0	0

She immediately notices some differences in the two designs. First is that the number of center points is different, three for the Box-Behnken versus two for the CCD. But the biggest difference is how the other design runs are structured. In the CCD, there are runs that have all low or high levels for all three factors. For example, in row 1 the levels for all three factors are set to -1. There are eight runs like this—they are the factorial combinations of the low and high levels for each of the factors. For the Box-Behnken design, every run contains the middle level for at least one of the factors. Maria thinks about the space that the design covers are represented as a cube. She draws a cube and then draws circles where the design points fall in the cube for both design types. She compares the two drawings and sees that for the CCD, the design reaches to the corners of the cube, but the Box-Behnken design does not.



Central Composite Design



Box-Behnken Design

"I think I would want my design to cover as much area as possible," Maria thinks aloud. She decides to create a CCD for her factors.

Maria sees in her software that she has a choice to make with the construction of her CCD. There are three types of points in the design:

Factorial points that are high and low-level combinations for each of the factors, which are like in her earlier screening design. These are the first eight runs in her design above.

Axial points, where the levels of two of the three factors are at the middle, and the third factor is set to either a low or high level. These are runs 9–14 in her design above.

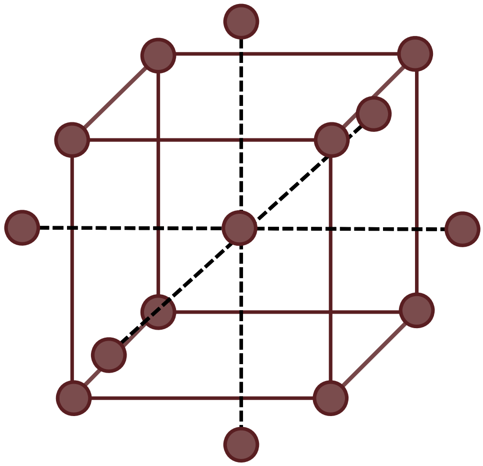
Center points that are the middle level of all factors at the same time, which are also like in her earlier screening design. These are runs 15 and 16 in her design above.

There is a choice to be made as to the level of the axial points, how low or high that the level is set. Her choices are on face, rotatable, and orthogonal. The design she has already created has the axial points on face. This means they are in the middle of the face of the six sides of the cube that defines the design region. She creates designs with the rotatable and orthogonal axial values to compare.

Run	X1	X2	X3
1	-1	-1	-1
2	-1	-1	1
3	-1	1	-1
4	-1	1	1
5	1	-1	-1
6	1	-1	1
7	1	1	-1
8	1	1	1
9	-1.68	0	0
10	1.68	0	0
11	0	-1.68	0
12	0	1.68	0
13	0	0	-1.68
14	0	0	1.68
15	0	0	0
16	0	0	0

The CCD with the orthogonal axial values looks like this:

Run	X1	X2	X3
1	-1	-1	-1
2	-1	-1	1
3	-1	1	-1
4	-1	1	1
5	1	-1	-1
6	1	-1	1
7	1	1	-1
8	1	1	1
9	-1.29	0	0
10	1.29	0	0
11	0	-1.29	0
12	0	1.29	0
13	0	0	-1.29
14	0	0	1.29
15	0	0	0
16	0	0	0



Central Composite Design with axial points off the face of the cube

Maria again draws a cube and places the design points in the space and sees how the axial points would now be pulled away from the six faces of the cube. Maria sees that for the rotatable design, the axial levels are set at ± 1.68, and for the orthogonal design, the axial levels are set at ± 1.29. The help feature in her software describes what is special about these level choices. The rotatable axial levels are set so that each point in the design is equidistant from the center point and that the increase in the precision of any model predictions from the center of the design is the same in all directions. The orthogonal axial levels are set so that each of the response surface model effects is orthogonal, or independent from the others.

Maria: "I get that these are both desirable properties. I like the idea of increasing the size of the design space without increasing the number of design runs. But I'm not sure how far we can go with the ranges. I'll create each of the designs and talk them over with Alex."

The design with the rotatable axial levels looks like this:

run	steam pressure	pump pressure	line speed
1	200	370	22
2	200	370	40
3	200	420	22
4	200	420	40
5	280	370	22
6	280	370	40
7	280	420	22
8	280	420	40
9	173	395	31
10	307	395	31
11	240	353	31
12	240	437	31
13	240	395	16
14	240	395	46
15	240	395	31
16	240	395	31

To illustrate the differences in the axial level choices, she arranges them in a table that looks like this:

	steam pressure		pump pressure		line speed	
axial type	low	high	low	high	low	high
face	200	280	370	420	22	40
orthogonal	189	291	363	427	19	43
rotatable	173	307	353	437	16	46

Maria and Alex discuss the design and the potential axial-level choices. They both agree that expanding the design space by using the widest axial levels would be best. But Alex has a concern.

Alex: "I'm sure that like in the screening design we will be making some product that we can't sell. Lisa had agreed that would be OK for this stage of experimentation too. But I don't know if we can operate at the extremes. I think we should see what Connie thinks since she has been a line operator for many years. She may know if this will be a problem. I'll check with her at the end of her shift today. I'll email you with her response."

Maria: "You can point out that these extreme levels for each factor are run at the middle level of all of the other factors. That may influence her thinking."

Maria receives the email from Alex just before shutting down her computer for the evening. In it, Alex explains that Connie's only concern with the most extreme axial levels is the line speed level of 46. She can't recall ever running the line at such a high speed. She will make a small test run at the start of her shift tomorrow at that speed to confirm that it would work. Alex will join Connie for the test and get back to Maria with a decision.

By the middle of the next morning, Maria again receives an email from Alex confirming that the line can be run at a speed of 46. Maria can now finalize the design creation. As with her screening design, she randomizes run order. Her final design looks like this:

run	steam pressure	pump pressure	line speed
1	280	420	22
2	240	395	31
3	200	420	40
4	200	370	40
5	280	370	22
6	240	395	16
7	280	370	40
8	240	353	31
9	307	395	31
10	200	420	22
11	240	395	46
12	280	420	40
13	200	370	22
14	173	395	31
15	240	395	31
16	240	437	31

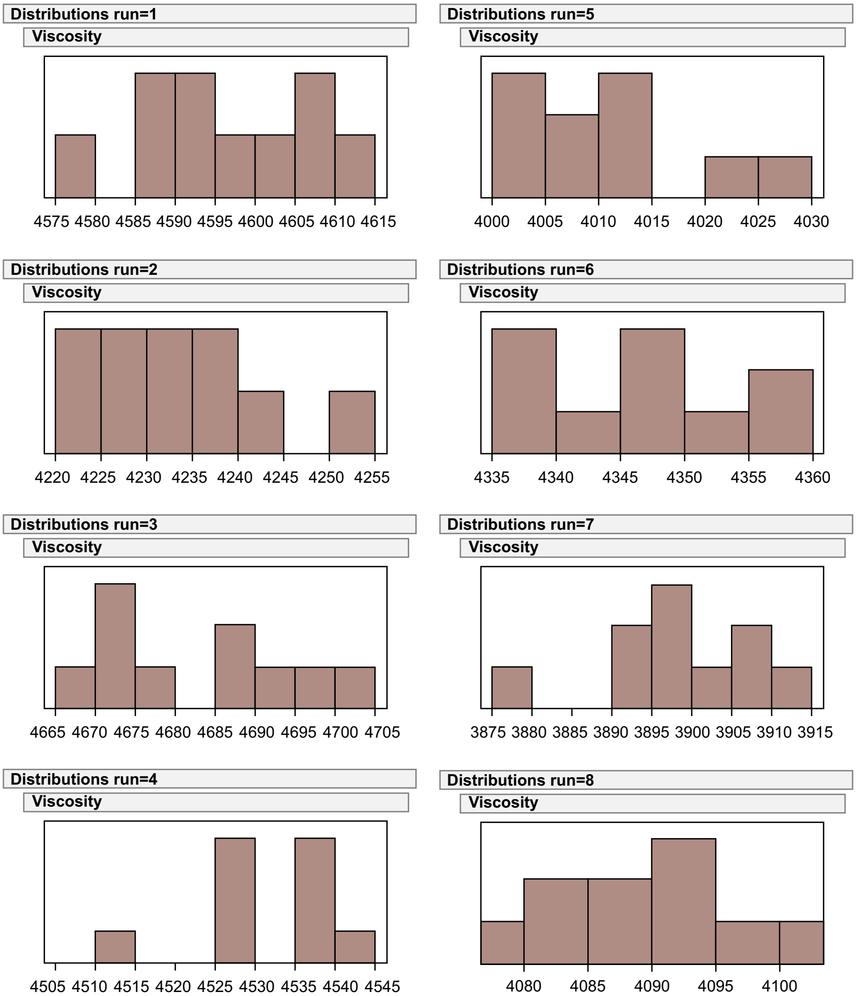
All of the same team members are on board for this second design. Alex will work with Connie executing the runs. They will again collect 10 bottles while running the line at each of the factor settings. Chad will perform the viscosity measurements himself so that their experiment does not interfere in the Quality Lab's daily operations. And, Maria will be at the line to assist in any way needed.

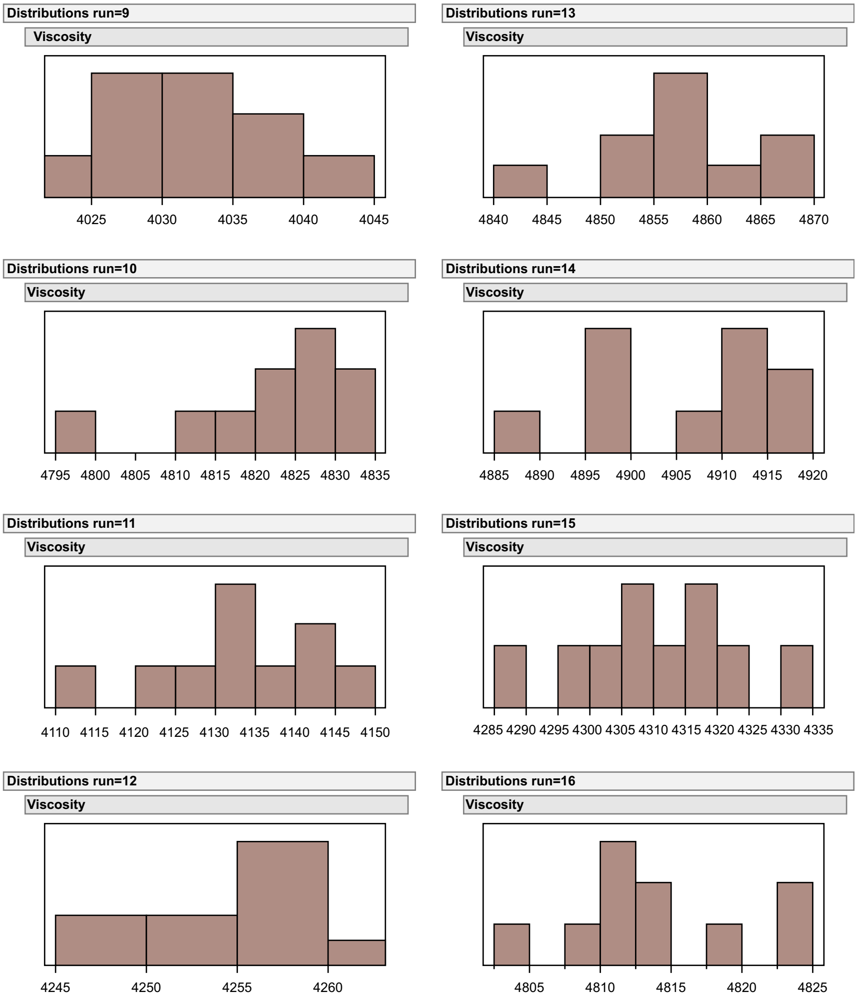
The design execution day is again very intense but runs more smoothly overall than their first design execution. Their experience from the first execution helped greatly, and they did not need to stop the line to make the factor level adjustments. Maria and the team are pleased with themselves.

Maria's spreadsheet with the viscosity measurements for the design runs looks very similar to the sheet for her first design.

Sample	V 01	V 02	V 03	V 04	V 05	V 06	V 07	V 08	V 09	V 10
1	4597	4606	4593	4593	4604	4613	4589	4607	4586	4579
2	4253	4225	4224	4229	4242	4238	4232	4235	4223	4233
3	4699	4693	4688	4689	4673	4678	4674	4673	4701	4668
4	4540	4528	4528	4510	4529	4541	4525	4539	4536	4537
5	4006	4028	4011	4005	4024	4012	4002	4004	4011	4004
6	4360	4349	4351	4336	4344	4346	4338	4339	4346	4359
7	3905	3897	3897	3879	3911	3904	3896	3892	3894	3908
8	4084	4085	4081	4090	4092	4091	4090	4101	4079	4099
9	4032	4023	4039	4032	4029	4025	4044	4026	4033	4037
10	4824	4830	4828	4821	4829	4798	4819	4834	4828	4811
11	4129	4124	4143	4142	4137	4132	4113	4133	4132	4145
12	4254	4257	4256	4245	4256	4245	4259	4253	4257	4262
13	4867	4856	4854	4866	4859	4853	4841	4862	4858	4857
14	4908	4916	4913	4899	4899	4899	4919	4913	4889	4913
15	4297	4322	4309	4317	4301	4331	4308	4319	4289	4315
16	4811	4808	4824	4813	4823	4810	4803	4812	4813	4819

Maria again starts her analysis process by visualizing the distributions of the individual viscosities for each of the design runs. She creates histograms of the 10 viscosity measurements by run.


Distributions run=1
Viscosity
4575 4580 4585 4590 4595 4600 4605 4610 4615
Distributions run=5
Viscosity
4000 4005 4010 4015 4020 4025 4030
Distributions run=2
Viscosity
4220 4225 4230 4235 4240 4245 4250 4255
Distributions run=6
Viscosity
4335 4340 4345 4350 4355 4360
Distributions run=3
Viscosity
4665 4670 4675 4680 4685 4690 4695 4700 4705
Distributions run=7
Viscosity
3875 3880 3885 3890 3895 3900 3905 3910 3915
Distributions run=4
Viscosity
4505 4510 4515 4520 4525 4530 4535 4540 4545
Distributions run=8
Viscosity
4080 4085 4090 4095 4100



Maria looks at these histograms to see if there is any indication of skewness in the distributions. She is also hoping to see that the variation in measurements for the samples is similar. As with her screening design, the graphs have very different ranges since they also have very different centers, so it is difficult to see if this is true. So again, she calculates the average, range, and standard deviation in viscosity measurements for the 16 samples and arranges them in a table.

Sample	Average Viscosity	Range	Standard Deviaton
1	4597	34	11
2	4233	30	9
3	4684	33	12
4	4531	31	9
5	4011	26	9
6	4347	24	8
7	3898	32	9
8	4089	22	7
9	4032	21	7
10	4822	36	11
11	4133	32	10
12	4254	17	6
13	4857	26	7
14	4907	30	10
15	4311	42	13
16	4814	21	7

The ranges and standard deviations for each of the 16 design runs are similar enough that she feels comfortable in thinking that there are no differences in variability between the runs. As with her screening design, the averages are substantially different, and these are what she will use as a response for modeling the factor effects.

The model that typically fits to data from a response surface design has main effects, pairwise interactions, and squared terms. The squared terms are to accommodate any curvature in the factor effects. For three factors, this means three main effects, three pairwise interactions, and three squared terms. Maria fits this model to her experimental data.

Summary of Fit

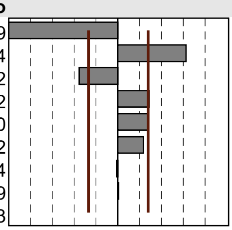
RSquare	0.96
RSquare Adj	0.90
Root Mean Square Error	107.30
Mean of Response	4407.53
Observations (or Sum Wgts)	16.00

Analysis of Variance

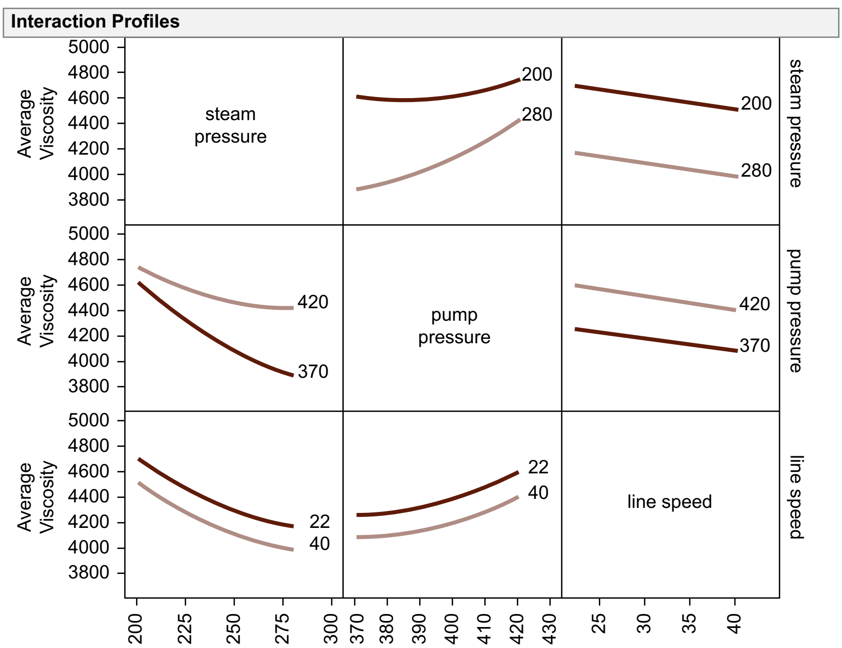
Source	DF	Sum of Squares	Mean Sqaure	F Ratio
Model	9	1647863	183096	15.90
Error	6	69074	11512	**Prob > F**
C. Total	15	1716937		0.0016*

Sorted Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
steam pressure(200,280)	–264.0	29.0	–9.09	<.0001*
pump pressure(370,420)	166.8	29.0	5.74	0.0012*
line speed(22,40)	–93.6	29.0	–3.22	0.0180*
steam pressure*pump pressure	103.1	37.9	2.72	0.0347*
steam pressure*steam pressure	84.5	35.3	2.40	0.0536
pump pressure*pump pressure	78.1	35.3	2.22	0.0686
pump pressure*line speed	–5.3	37.9	–0.14	0.8927
line speed*line speed	3.3	35.3	0.09	0.9283
steam pressure*line speed	1.3	37.9	0.03	0.9739



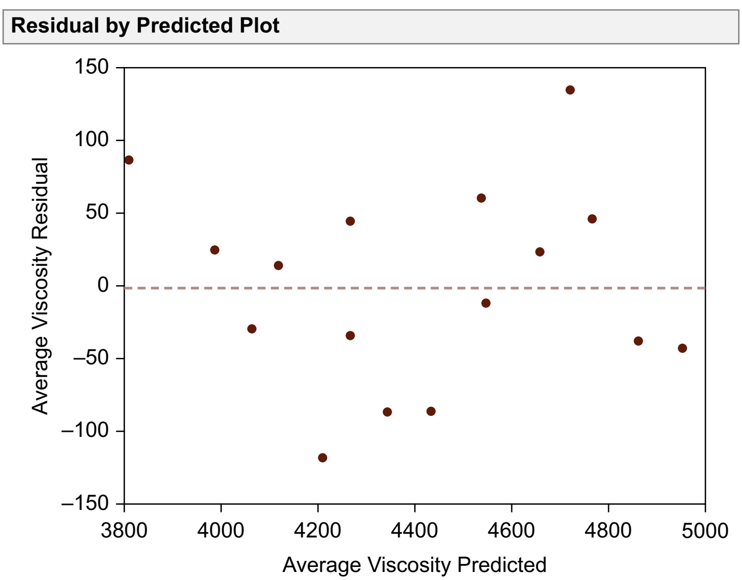
The output from Maria's software is similar to what she saw for the screening design. The factor effects are again ordered in decreasing levels of significance and have been standardized to the range of each of the factors so that the size of their effects can be directly compared. She sees that steam pressure is the most significant effect, followed by pump pressure, and then line speed. All of these have *p*-values smaller than 0.05, the level commonly used as a cutoff for significance. In addition, the interaction of steam pressure and pump pressure also has a *p*-value smaller than 0.05, and the squared terms for both steam pressure and pump pressure have *p*-values only slightly larger than 0.05. With the more complicated model with interactions and squared terms, it is not obvious how all of these factors affect viscosity; so Maria now creates interaction plots to visualize these effects.

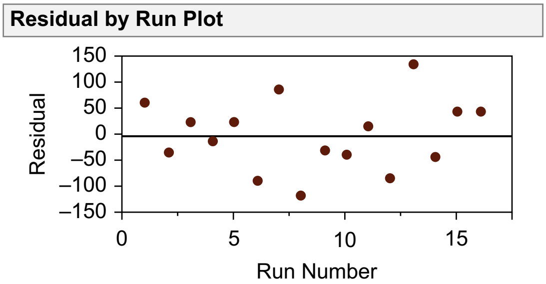


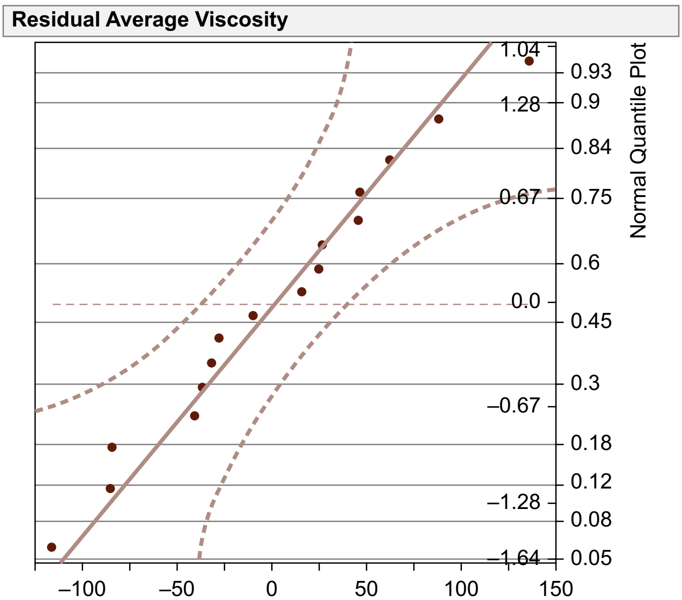
These graphs show all of the potential factor effects and their interactions. The plots in the first column all have steam pressure on the horizontal axis; those in the second column have pump pressure on the horizontal axis; and those in the third column have line speed on the horizontal axis. There are separate lines in each plot for the levels of the factor listed for each row. Maria examines the plot in the first column, second row.

This plot shows the effect of steam pressure at two different levels of pump pressure. There is a dramatic decrease in viscosity as steam pressure increases at the lower pump pressure level of 370, but only a modest decrease at the higher pump pressure level of 420. This clearly shows the significant interaction of steam pressure and pump pressure. The decreases have a curve to them rather than being strictly linear. The graph in the second column, first row shows the same interaction effect, with the pump pressure on the horizontal axis and curves for two different steam pressures. The graphs in column three show that the effect of line speed appears strictly linear and is consistent regardless of the levels of both steam pressure and pump pressure. Though the lines for the different levels of steam pressure and pump pressure are far apart, the slopes of the lines are nearly identical. This confirms the lack of significance of the interaction effects between line speed and steam pressure and line speed and pump pressure.

As with the screening design, Maria verifies that her model is not missing anything critical by analyzing the residuals. As with the screening design, there should be no pattern in them if she plots them against the order of the design runs and the predicted values from the model, and they should follow a normal distribution centered at zero. She again creates plots to perform this assessment.







Maria sees nothing in these graphs to suggest that her model is not adequately describing the factor effects on viscosity.

Maria recalls from her class that often response surface models are refined to remove the effects that are not significant. This simplifies the models somewhat prior to using them to make predictions. In her analysis, she had seen that three of the effects had very large *p*-values: the interaction of steam pressure by line speed; the interaction of pump pressure by line speed; and the squared term for line speed. She will remove these terms from her model. The squared terms for steam pressure and pump pressure were close to the 0.05 level of significance, so she will keep these in the model, at least for this initial refinement.

Summary of Fit	
RSquare	0.96
RSquare Adj	0.93
Root Mean Square Error	87.82
Mean of Response	4407.53
Observations (or Sum Wgts)	16.00

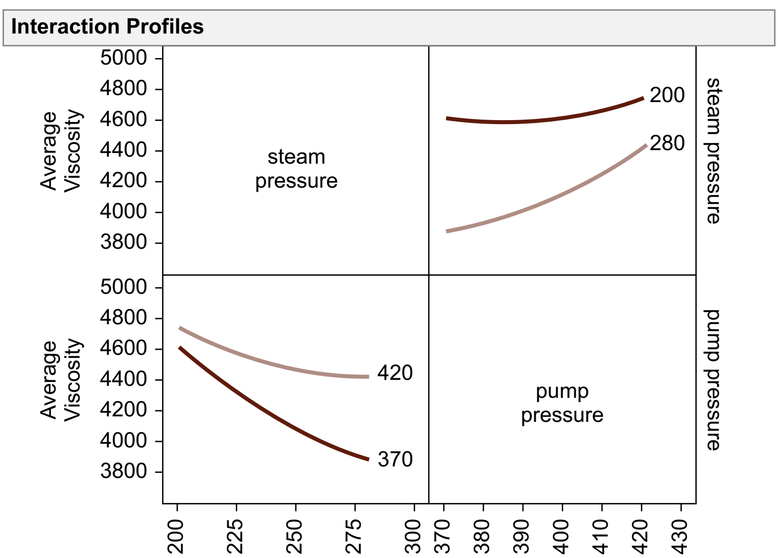
Analysis of Variance

Source	DF	Sum of Squares	Mean Sqaure	F Ratio
Model	6	1647520	274587	35.60
Error	9	69417	7713	Prob > F
C. Total	15	1716937		<.0001*

Sorted Parameter Estimates

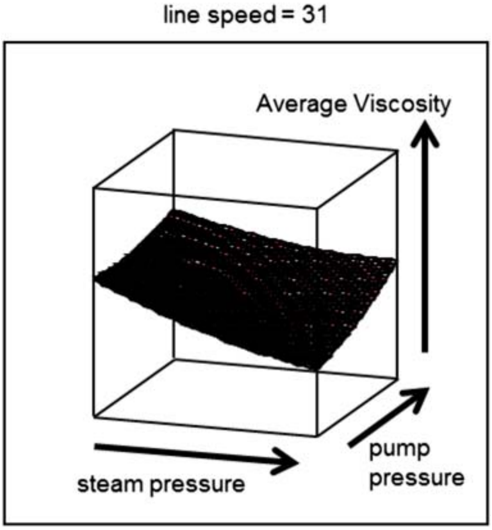
Term	Estimate	Std Error	t Ratio	Prob>|t|
steam pressure(200,280)	–264.0	23.8	–11.11	<.0001*
pump pressure(370,420)	166.8	23.8	7.02	<.0001*
line speed(22,40)	–93.6	23.8	–3.94	0.0034*
steam pressure*pump pressure	103.1	31.1	3.32	0.0089*
steam pressure*steam pressure	83.1	26.2	3.17	0.0113*
pump pressure*pump pressure	76.7	26.2	2.93	0.0167*

Maria notices that much in this refined model output is similar or identical to her initial model output. The estimated effects for steam pressure, pump pressure, line speed, and the interaction of steam pressure and pump pressure are the same. The *p*-values are mostly smaller, and those for the squared terms for steam pressure and pump pressure are well below the 0.05 level of significance, suggesting that they do belong in the model. She notices that the R Square for the model is the same at 0.96, meaning that her model explains 96% of the total variability in the response. The root mean square error is also somewhat smaller. This is essentially the standard deviation of her model's predictions, so smaller is better since she will use the model to make predictions. She again creates interaction plots.



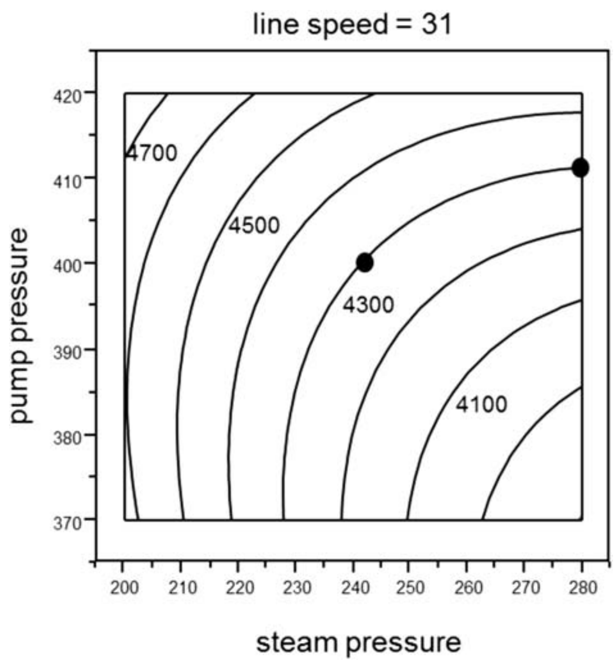
This is a simpler visualization now since the interaction effects of steam pressure and pump pressure with line speed have been removed from the model. The plots look nearly identical to those in her initial visualization, so her interpretation of how steam pressure and pump pressure affect viscosity is still the same. And since the estimate of the effect of line speed is the same as in her initial model, her interpretation of that factor effect is also the same. The refined model has not changed her understanding of the factor effects but made the model easier to understand and better able to predict.

Maria creates additional visualizations to better understand the factor effects. Her software enables her to create a three-dimensional representation of the model for any two factors and the viscosity response for set levels of any additional factors. Since line speed does not interact with the other factors, she chooses to set this at the middle level of 31. Her visualization looks like this:



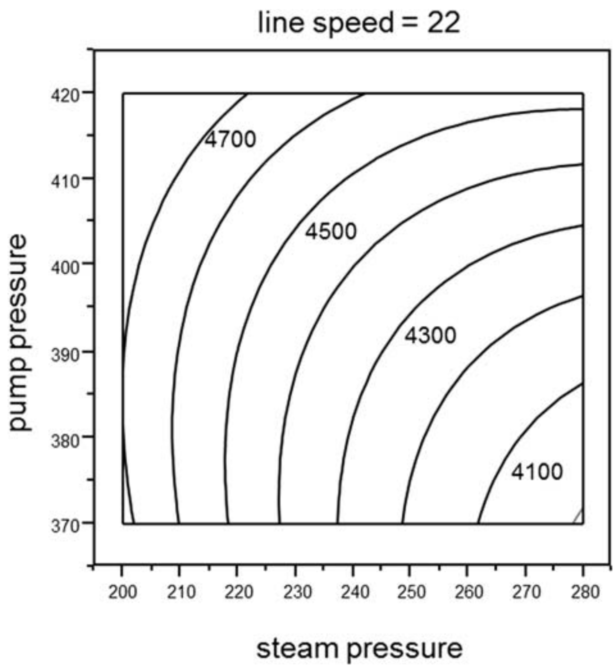
In this graph, steam pressure increases from left to right in the cube, and pump pressure increases from front to back in the cube. Viscosity increases from top to bottom. She can easily see that the viscosity is highest at low steam pressure and high pump pressure, and that viscosity is lowest at high steam pressure and low pump pressure. Since the effect of line speed is linear and negative, this shape would shift up for a lower line speed and down for a higher line speed. There are no axis labels, but even with them, it would be difficult to determine factor levels that would produce a specific viscosity.

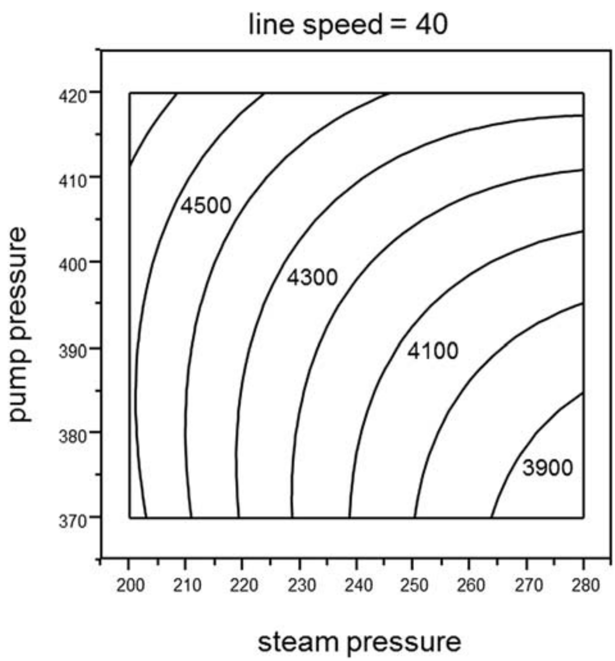
A second visualization is better for that determination. It again presents the model for any two factors, but now the viscosity response is represented as contours. She also creates this visualization for a line speed of 31.



Maria can now easily see what combinations of steam pressure and pump pressure are predicted to deliver a viscosity of 4300, her target. There are many, and they fall along the line labeled 4300. A steam pressure of 240 and pump pressure of 400 near the middle of the plot looks to come very close to this target. Also, at the far right, a steam pressure of 280 and a pump pressure of 410 also predicts close to the target. Maria notes these potential factor level combinations on her plot with a heavy dot.

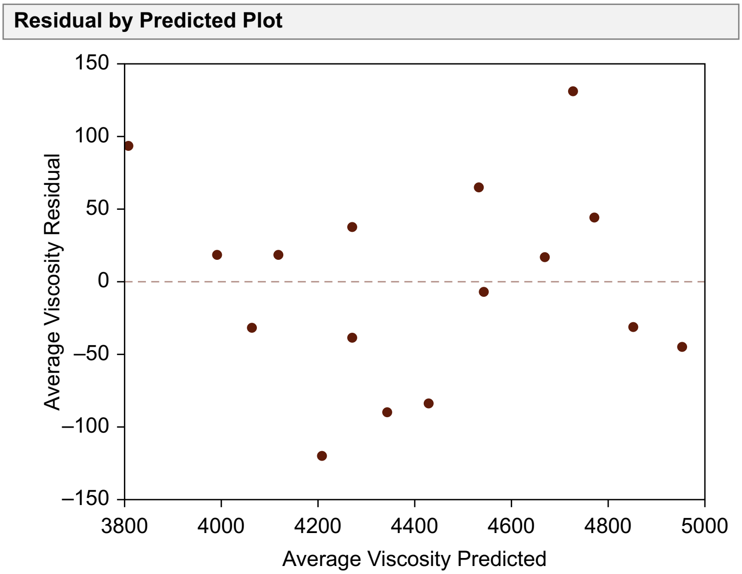
Maria creates similar plots for line speeds of 22 and 40.

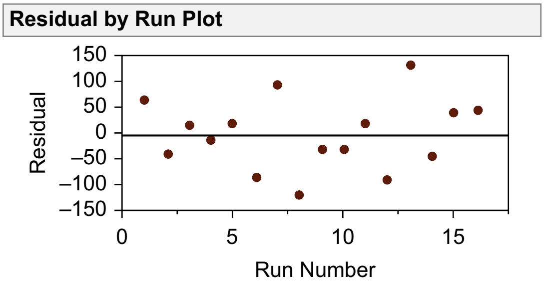

line speed = 22
pump pressure
steam pressure
4700
4500
4300
4100
420
410
400
390
380
370
200
210
220
230
240
250
260
270
280

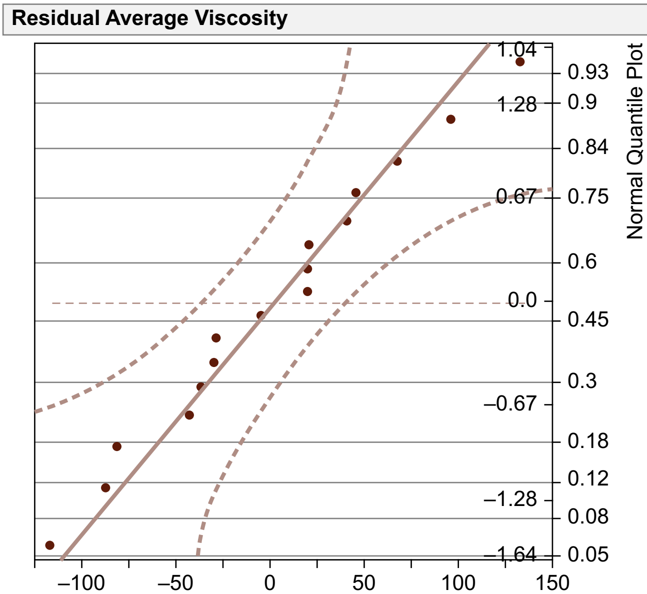

line speed = 40
pump pressure
steam pressure
4500
4300
4100
3900
420
410
400
390
380
370
200
210
220
230
240
250
260
270
280

From these graphs, Maria can see now the curve at her target of 4300 moves from the lower left to the upper right as line speed increases. Her software will allow her to save the model formula so that she can use it to predict the viscosity anywhere within her design space.

Maria has many things to share with the team. Since her refined model is so similar to the initial model, she does not expect to see anything in her residual plots. She again creates them just to be sure. They look like this:







As expected, Maria sees no patterns in the residuals. She is confident with her model and prepares to share the results. She arranges a meeting for the next day with Alex, Connie, Chad, and most importantly, Lisa.

Maria shares the analysis results of the second phase of the designed experiment to the team the following day. All are excited with the results and are eager to implement them. Lisa questions the team.

Lisa: "We now have a model that relates how steam pressure, pump pressure, and line speed affect viscosity. Is there anything more we need to do before we implement it in our operations? And how do we do that?"

Maria: "While I'm confident with our model, it is important that we do not neglect the important step of validating the model. At this point, we only have paper predictions."

Lisa: "What do we need to do to validate the model? Would we need to make some additional runs?"

Maria: "Yes, that is the only way to truly validate the model. It would not need to be very many."

Lisa: "How do we determine what to do here?"

Alex: "I would like to see if the model predictions are valid over a range of line speeds. We often run faster or slower depending on demand for the product. If the model predicts that certain steam pressure and pump pressure combinations will deliver the target viscosity, I would like to verify that for slow, moderate, and fast line speeds. This way, we know that our process adjustments will deliver the desired response."

Maria: "I think that this is a good approach, we can validate the model where it matters the most to us."

Chad: "I would like for us to consider something else in the validation. The process settings that we are most interested in are not likely to be identical to design runs from our experiment. We should also include some validation runs that are identical to design runs. That way, we will be sure that nothing else has changed substantially with this second set of runs."

Lisa: "This all makes sense to me. I'm willing to commit resources for the validation runs. I like Alex's suggestion that we validate at process settings that should deliver the target at low, medium, and high line speeds. And I agree with Chad that we would want to include some repeated runs from the design. What should they be?"

Maria: "Since the other runs are all at our target of 4300 with different line speeds, I think these other runs should be at predictions below and above our target. This way, we will know that our model will still be accurate at a different target."

Lisa: "I see a plan forming here. If we have six validation runs, we can use three to validate the process settings at low, medium, and high line speeds. And then have three repeats of design points that delivered low, medium, and high viscosity levels. Maria, can you put that together for us? Alex, please help Maria with this."

Maria: "I'll send out a plan to the team before the end of the day. Alex, let's find a time to sit down and work out the details of the validation."

Maria and Alex meet later in the day to work out the specifics of the validation plan. They look at the observed data from their design and notice that there are three design points that have both different process settings and substantially different observed viscosities. They use the prediction equation from the model to determine the predicted viscosity for these runs. It is easy enough to determine runs that will meet the target by examining the model's contour plots for each of the line speeds and using the model's prediction equation.

Maria and Alex decide on the following runs for validation:

run	steam pressure	pump pressure	line speed	average viscosity from design	predicted viscosity
1	200	420.0	22	4822	4852
2	240	395.0	31	4311	4271
3	280	370.0	40	3899	3803
4	260	400.0	22		4300
5	228	380.0	31		4304
6	248	412.5	40		4297

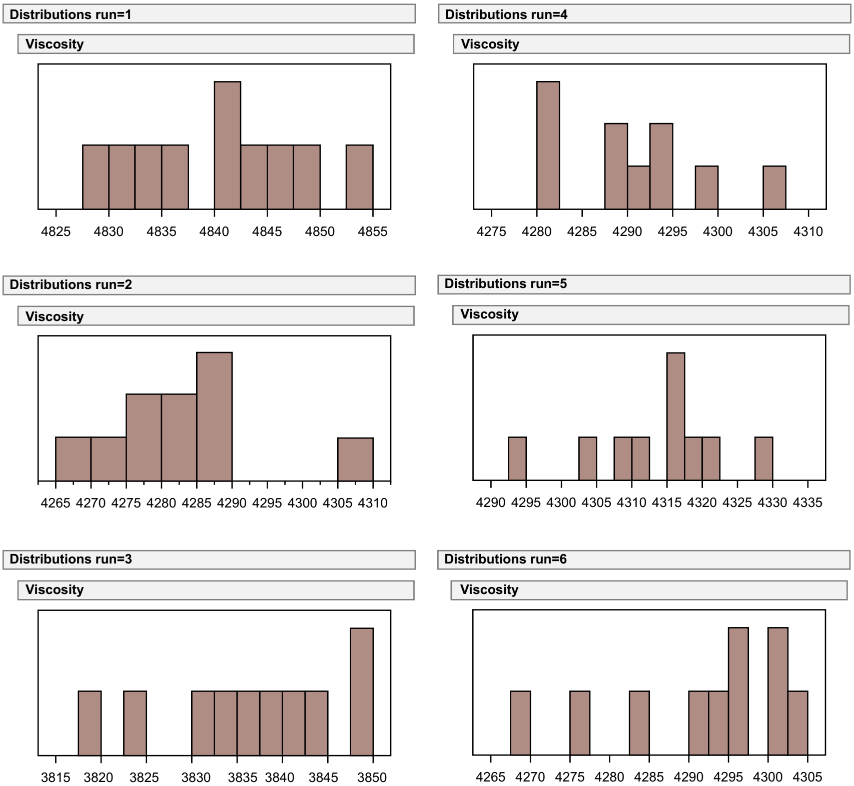
The first three runs in the validation set are repeats from the experimental design. The second three are model predictions that are near the target of 4300 at the low, middle, and high line speeds from the design.

Maria shares the recommendations for validation that she and Alex have determined with the team. All are in agreement, so Maria and Alex schedule the validation runs for the following week.

The execution plan for the validation runs is the same as what has been used for both designs. Chad again makes all of the measurements for runs. There are again 10 measurements for each of the six runs. The validation data looks like this:

run	V 01	V 02	V 03	V 04	V 05	V 06	V 07	V 08	V 09	V 10
1	4846	4842	4831	4837	4853	4828	4845	4841	4833	4850
2	4309	4289	4280	4289	4268	4276	4283	4273	4280	4288
3	3849	3848	3843	3824	3842	3835	3837	3839	3818	3831
4	4289	4298	4293	4288	4280	4293	4280	4306	4291	4282
5	4317	4330	4311	4294	4320	4303	4317	4308	4318	4316
6	4295	4295	4283	4269	4301	4291	4302	4303	4296	4277

Maria again starts by creating distributions of the data by run. Her histograms look like this:



She sees no substantial differences in the distributions for the runs. Like before, she also calculates the range and standard deviations for the runs. These summary statistics look like this:

Run	Average Viscosity	Range	Standard Deviation
1	4841	25	8
2	4284	41	11
3	3837	31	10
4	4290	26	8
5	4313	36	10
6	4291	34	11

Maria sees no substantial differences in these summary measures of variation and now feels confident that she can compare the run averages to the model predictions and for the first three the observed values from the design. She adds these means to her earlier table; it now looks like this:

run	steam pressure	pump pressure	line speed	average viscosity from design	predicted viscosity	average validation viscosity
1	200	420.0	22	4822	4852	4841
2	240	395.0	31	4311	4271	4284
3	280	370.0	40	3899	3803	3837
4	260	400.0	22		4300	4290
5	228	380.0	31		4304	4313
6	248	412.5	40		4297	4291

For the first three validation runs, Maria has three values to compare: the average viscosity from the design; the predicted viscosity from the model; and the average viscosity from the validation. For all three runs, the differences between any of these is not more than 40, which is similar to the range of the 10 individual bottle measurements for the design runs in both her screening and response surface designs. The last three validation runs have predictions near her target of 4300, and the average viscosity of the validation runs is within 10 units, even closer than in the first three validation runs. Based on these observations, Maria feels very confident that her model predictions are accurate enough for her purpose of determining steam pressure and pump pressure levels that will deliver the target viscosity across a range of line speeds. Maria shares the validation results with the team at the next morning's production meeting.

Maria: "The average viscosities from the validation runs were very close to the model predictions. I'm confident that we can use our model as a basis for a control strategy in production going forward."

Lisa: "This is great work! Besides improving our control strategy, I feel we now have a much better understanding of our production process. I will share this work with our management so that they will not only understand the progress that we have made but also to demonstrate the effectiveness of the systematic experimental approach. I can already think of a few other places where we can apply this approach."

This page intentionally left blank

CHAPTER 13

Mixture Experimental Design

Steve opens up the research and development leadership team meeting bragging about the value of the DOE work that Maria has just finished. Design of experiments has presented a powerful approach that had not been used before. Maria's work has helped the plant not only to deliver more consistent BBQ viscosity, but also the systematic approach has led to a better understanding of how factors such as steam pressure and line speed affect viscosity. Steve has several improvement opportunities in mind, and he hopes that by applying the design of experiments methodology to them, the results will be similarly positive.

A few days later, Maria receives a call from Adam, one of the procurement managers. The cost of one of the three starches used in the base BBQ sauce has increased dramatically because of a material shortage. The price is now nearly three times its historical average.

Adam: "We are hoping to use a different starch in the product, so that we don't have a cost increase. Joe, our starch expert, has identified a potential alternative. Can you replace the current high-cost starch with this alternative?"

Maria: "We will need to perform some experiments to answer this question. I'll engage the development team to create a research plan and will get back to you."

Maria starts by discussing the problem with Steve. He mentions that there had been some research done with the alternative starch in the past which indicated it was not a direct replacement in the current starch blend. Prior development work showed that finished product would not be thick enough if the starches are simply swapped on a one-to-one basis. There may be a blend of this new starch with the others that could deliver the desired product viscosity.

Maria: "So, we need to figure out what the best three-starch blend is."

Steve: "Can we use the design of experiments approach that we used to optimize viscosity in the past?"

Maria: "That is exactly what I am thinking. But, there is a big difference here, the three starches sum to a fixed percentage in the formula. I remember in my design of experiments class that mixture

Statistics for Food Scientists
http://dx.doi.org/10.1016/B978-0-12-417179-4.00013-5

designs are used in these situations. I will have to look at my class notes to recall more details about this approach."

Maria has time to look at her notes later in the day and then finds Steve at his desk.

Maria: "I think I have some ideas on how to approach this. We are really interested in the best proportional mixture of the three starches. This is similar to mixing a cocktail. I'll use a Margarita as an example."

Steve: "That is one of my favorites!"

Maria: "The classic recipe for a Margarita is two parts tequila, one part lime juice, and one part triple sec. This is the same as 50% tequila, 25% lime juice, and 25% triple sec. But maybe, there is a different mixture of these three that will taste even better—maybe a little less tequila and a little more triple sec. But, the proportions will always add to 100%. Each mixture will be a different design run."

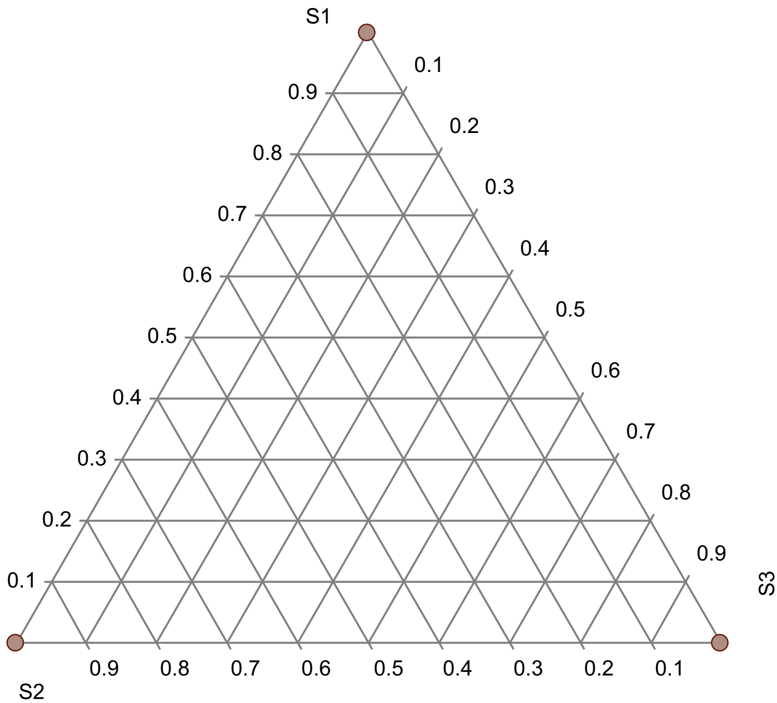
Steve: "I'm following so far."

Maria: "In a mixture design, we could have a design run with 100% one component and 0% of the others. This would make a terrible tasting Margarita. So, with a Margarita, we would have some constraints, like tequila needs to be between 40% and 60%, lime juice between 20% and 30%, and the same for triple sec. Maybe there are no constraints for our starch blend. Would we ever use only one of the starches?"

Steve: "That is a great question. I am not sure if our BBQ formula needs three starches; this is how we have been making it for as long as I can remember. Since we are changing to a new starch, we don't likely know. If we can make a good product with only two or even one starch, that would simplify the process and the ingredient line and also make Adam in Procurement happy. Maybe there are some constraints with the starches, like one of them cannot exceed 50% or another can be no less than 30%, but no more than 50%. Joe would be better able to tell us this."

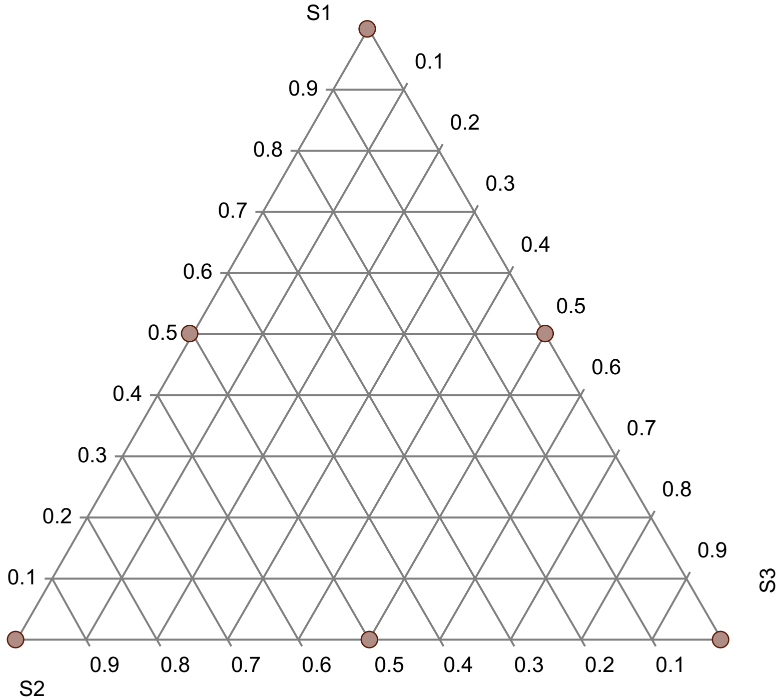
Maria moves closer to the white board in Steve's office and starts drawing.

Maria: "Let me help you visualize how the three starch design space would look, and how it would change if we add constraints. The mixture of three components can be represented by a ternary plot. A ternary plot looks like this:"

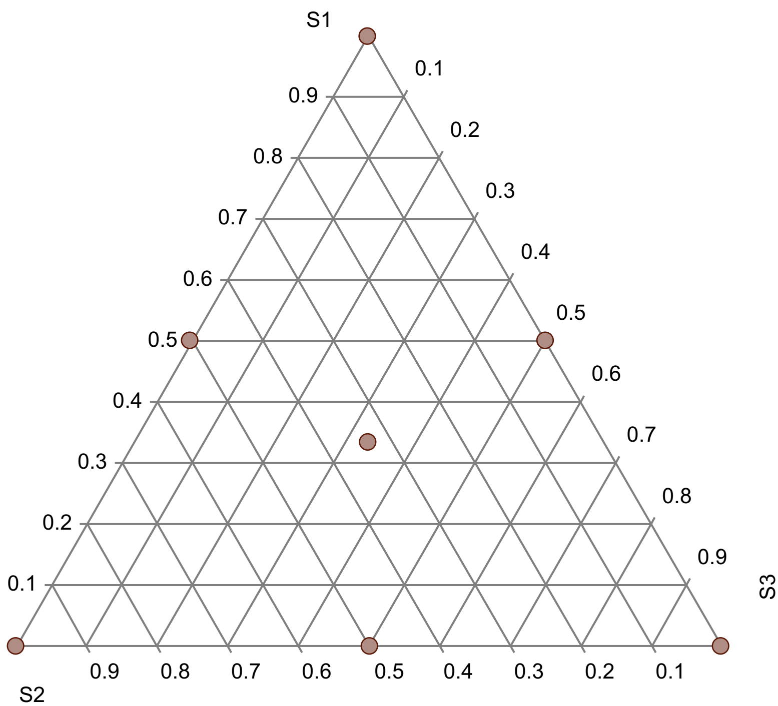


In the plot, any blend of the three starches can be represented, and the sum of the three starches will be equal to 100%. The corners of the triangle represent a sample with only one of the starches. The one at the top is 100% of starch 1. The one in the left corner is 100% of starch 2. The one in the right corner is 100% of starch 3. The tick marks on the plot face in specific directions indicate the level of each component in the starch blend.

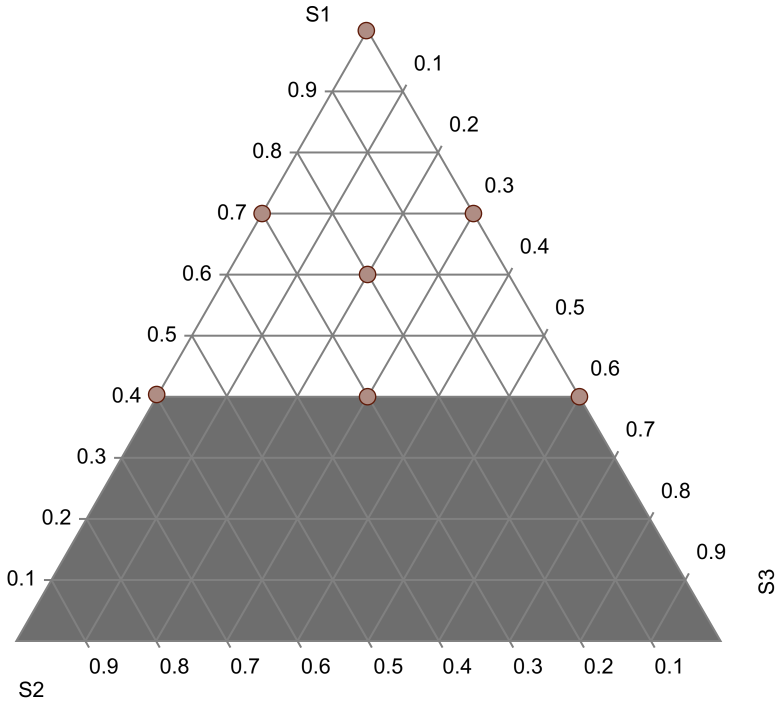
The middle of the edge of the triangle represents the 50/50 blends of the two components on that edge of the triangle. In the next graph, points in the middle of the edges in the triangle have been added. The point in the middle of this right edge of the triangle is the 50/50 blend of starch 1 and starch 3. The point in the middle of the bottom edge is the 50/50 blend of starch 2 and starch 3. The point in the middle of the left edge is the 50/50 blend of starch 1 and starch 2.



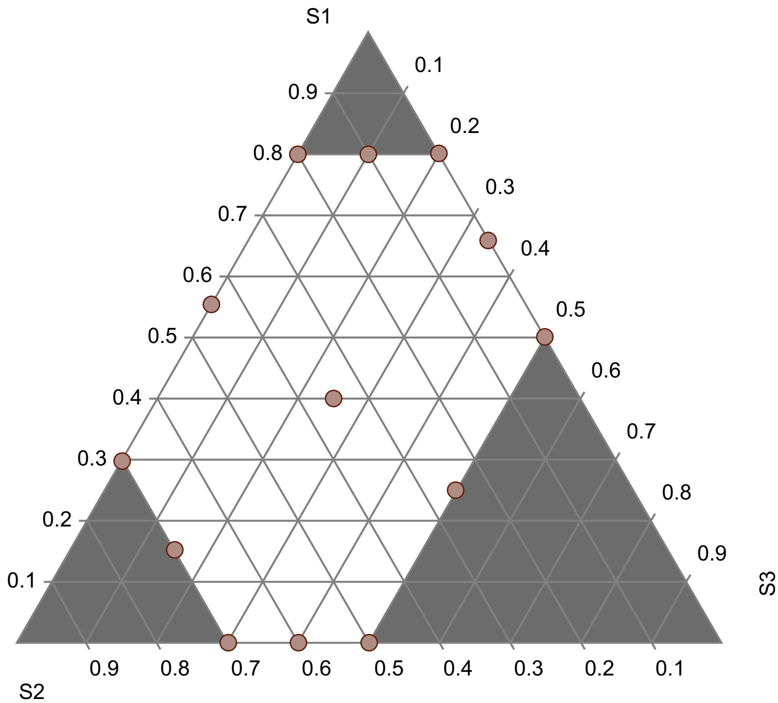
Finally in the next graph, the point in the middle of the triangle is the 33/33/33 blend of the three starches. Like in an optimization design, it is the center point in the design space.



If we add a constraint on any one of the starches, the design space changes. The next graph illustrates the situation where starch 1 cannot be less than 40%. This means that the other starches cannot be greater than 60%. The area outside of the design space is a darkened.



If all three starches have constraints, so that each cannot take values above or below certain parameters, the design space will be constrained. In the next graph, starch 1 must be less than 80%, starch 2 must be less than 70%, and starch 3 must be less than 50%.



Steve: "Wow, I never thought that a simple triangle could help me visualize the mixture space and any potentially complex constraints. We need to find out if we need all three starches and if there are constraints on any one of them."
Maria: "Let's call Joe and see what he says."
Joe picks up before the third ring.
Joe: "Hi Steve, how are you doing?"
Steve: "Greetings Joe. I am doing great, thank you for asking. I have Maria with me in my office, and we have been talking about our BBQ starch mix."
Maria: "Joe, you suggested a new starch for BBQ sauce to replace the starch with the increased cost in our formula, and we need your help with answering a few questions. Do you think that one or two starches may be sufficient to match the current viscosity instead of using all three starches? And do you have any concerns in terms of starch ranges? Are there any constraints that we should be aware of, specific ranges for any of the starches that we will need to operate within?"

Joe: "To tell you the truth, anything is possible. A two-starch mix may be sufficient, though the work we did a few years ago indicated that all three types of starches may be needed. However, based on that research and since we have not worked with this new starch, I would not eliminate the opportunity to simplify our ingredient line. As far as starch constraints, I would look at all possibilities, as there is no reason we should have more or less of any one starch. Please let me know what you learn, and feel free to reach out again if you have any other questions."

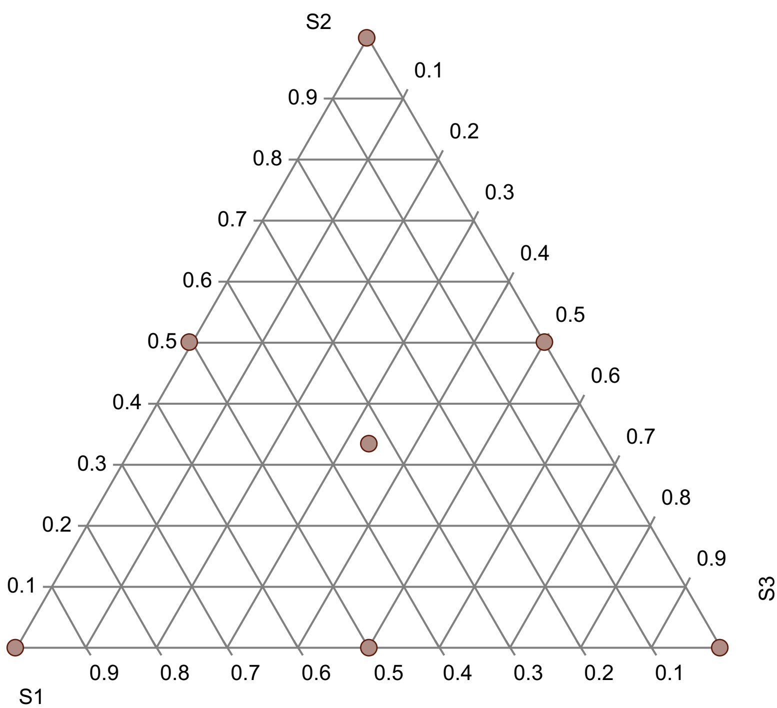
Steve: "Thanks Joe. Have a great day."

Joe: "You too. Have a productive rest of the week."

Maria: "It sounds like we should look at all possibilities. I will set up the design across the entire 0–1 range."

Steve: "Do you mean 0–100% range?"

Maria: "Yes. The range could be referred to both ways: 0–1 if a proportion or 0–100% if a percentage. I'll use proportions. The design we can use will look like this:"



Steve: "I recall that you replicated the design run at the center in your earlier experiment. If you do that here, we would need to create just eight samples to evaluate all three starches? That seems like a very small number! Is it possible that we are forgetting something? Your earlier design had 16 runs to fully understand the linear and curvilinear effects of the ingredients and their interactions."

Maria: "You are correct; your memory is quite good. With a mixture design where two of the three factor levels are known, the last is known too, since they sum to 100%. The model we use is a different form than what we used when the factors were completely independent, and this smaller number of runs is sufficient."

Steve: "The mixture design is similar to the response surface design, yet so different. On one side, the thought process is the same, but on the other side, the structure of the design and the visualization of the design space is completely different. Fortunately, instead of 16 samples to evaluate the three starches, we will only need to produce and evaluate eight since we are interested in the optimal blend of the three starches."

Maria sets up the design runs. As with her earlier design, she decides to produce the runs in a random order. Her design looks like this:

Sample	S1	S2	S3	Cost
1	0.33	0.33	0.33	0.32
2	1.00	0.00	0.00	0.25
3	0.00	0.00	1.00	0.30
4	0.50	0.50	0.00	0.33
5	0.00	1.00	0.00	0.40
6	0.33	0.33	0.33	0.32
7	0.00	0.50	0.50	0.35
8	0.50	0.00	0.50	0.28

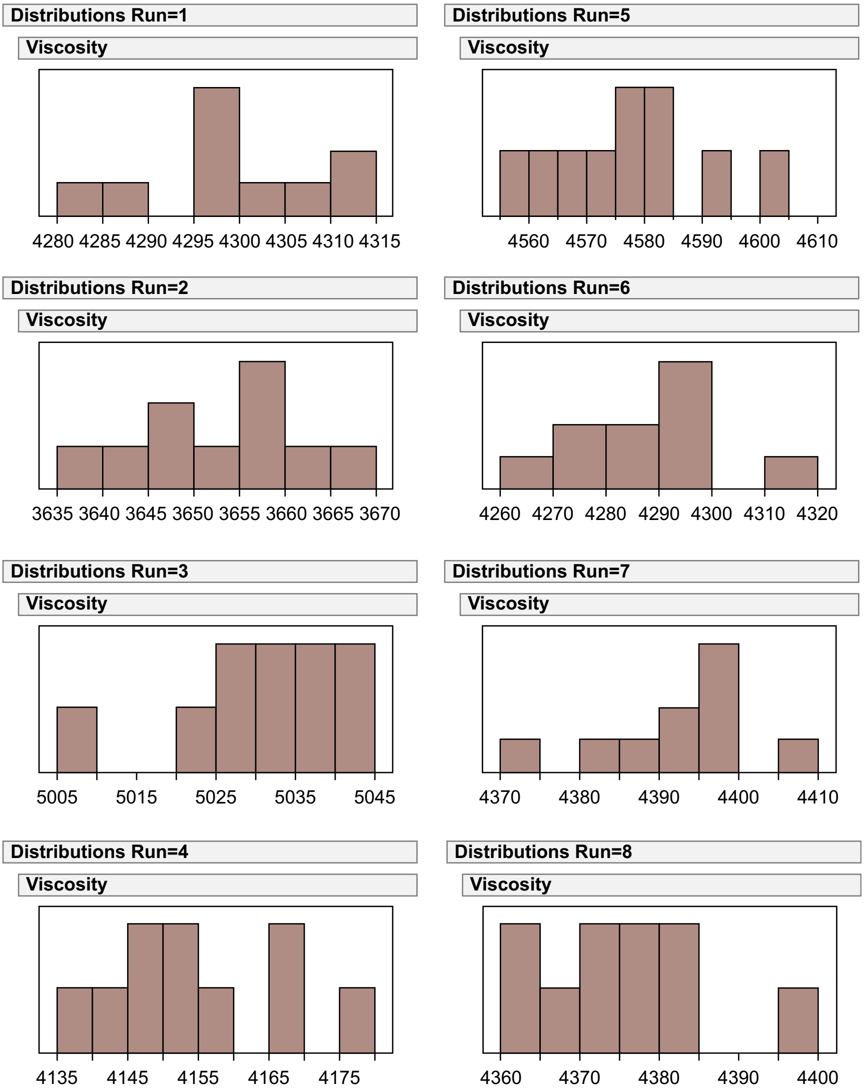
Joe had provided Maria with the cost of the individual starches, and she easily determines the cost of any starch blend. This information will be helpful in determining the best blend if she has more than one that meets the viscosity target.

All of the same team members are on board for this new design. Connie will run the line, but since there are no changes in the operating conditions, she will need no additional help. Maria will make sure that the different starch blends will be incorporated during the ingredient mixing stage of the process. Alex will again collect 10 bottles while running the line at each of the factor settings just as in the earlier experiments, two pulls from each of the five filler heads. Chad will again perform the viscosity measurements himself so that their experiment does not interfere with the Quality Lab's daily operations.

The design execution day is again very busy but runs smoothly as the team is now even more experienced in execution. When Chad has completed the measurement of all of the samples, he provides the data to Maria.

Sample	V 01	V 02	V 03	V 04	V 05	V 06	V 07	V 08	V 09	V 10
1	4311	4311	4295	4299	4298	4302	4306	4285	4296	4281
2	3654	3666	3639	3662	3645	3648	3655	3659	3656	3645
3	5026	5036	5026	5041	5007	5043	5024	5038	5031	5031
4	4177	4167	4146	4153	4167	4154	4147	4158	4145	4138
5	4576	4605	4593	4584	4581	4562	4572	4559	4567	4579
6	4278	4315	4292	4293	4286	4262	4298	4278	4291	4287
7	4371	4390	4383	4396	4395	4391	4399	4396	4395	4405
8	4369	4365	4383	4374	4381	4375	4371	4364	4379	4395

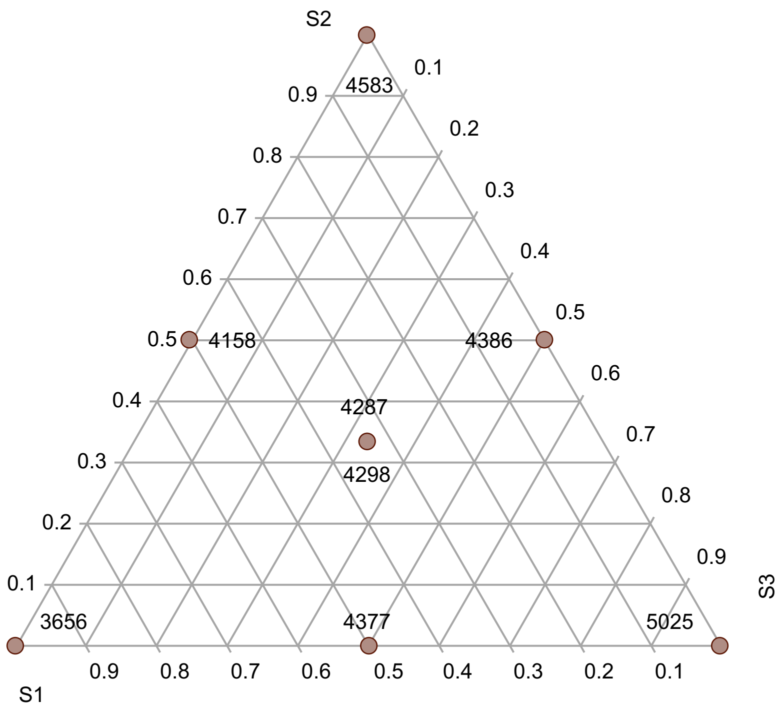
Maria again starts the data analysis process by creating histograms of the 10 individual viscosities for each of the design runs.



Maria sees that the variation in measurements for the samples is similar, and there is no indication of skewness or unusually high or low measurements. As before, she calculates the average, range, and standard deviation in viscosity measurements for the eight design runs and arranges them in a table.

Sample	Average Viscosity	Range	Standard Deviation
1	4298	30	10
2	3653	27	8
3	5030	36	10
4	4155	39	12
5	4578	46	14
6	4288	53	14
7	4392	34	9
8	4376	31	9

Being satisfied that the ranges and standard deviations are not substantially different for the design runs and seeing the average viscosities, she is eager to start the modeling process. She first creates a ternary plot with the average viscosities for the design runs at the design points.



In the plot, it is clear that starch 3 alone delivers the thickest product since the highest viscosity average is highest for that design point. Similarly, starch 2 alone delivers the thinnest product since the viscosity average is lowest for that product. None of the average viscosities are at her target of 4300 exactly, but the two at the middle come very close, and several others are not too far off of the target. Maria is hoping that the modeling will reveal a number of starch blends predicted to meet the target.

Maria recalls from her design of experiments class notes that the model for a mixture design like hers is a bit different than the typical regression models used with the screening and response surface designs. For her mixture, the model includes effects for the three individual starches and also interaction terms for the three pairs of starches. The intercept is dropped from the model and squared terms are not included, this is also called a Scheffe model. Maria fits this model to her experimental data.

Summary of Fit

RSquare	1.00
RSquare Adj	1.00
Root Mean Square Error	16.03
Mean of Response	4346.25
Observations (or Sum Wgts)	8.00

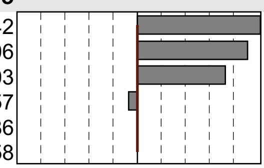
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	1046580	209316	814.8
Error	2	514	257	**Prob > F**
C. Total	7	1047094		0.0012*

Tested against reduced model: Y=mean

Sorted Parameter Estimates

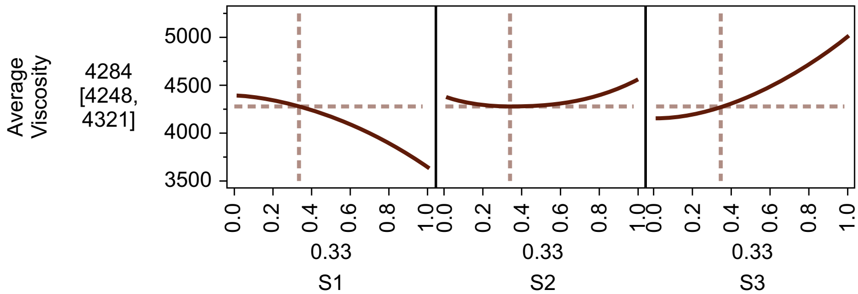
Term	Estimate	Std Error	t Ratio	Prob>|t|
S3	5027.75	15.94	315.42	<.0001*
S2	4575.75	15.94	287.06	<.0001*
S1	3650.75	15.94	229.03	<.0001*
S2*S3	–1602.97	71.01	–22.57	0.0020*
S1*S2	203.03	71.01	2.86	0.1037
S1*S3	183.03	71.01	2.58	0.1233



The output from her software is again similar to the output from her earlier designs. Maria is surprised to see that the R-square for her model is 1.00—a perfect fit! Looking closer, she sees that this is really 0.9996, and her software has rounded it up for the output. It seems that the execution of this design was very good and likely due to the teams experience with the process and that changing ingredient blends is easier to control than changing the process settings.

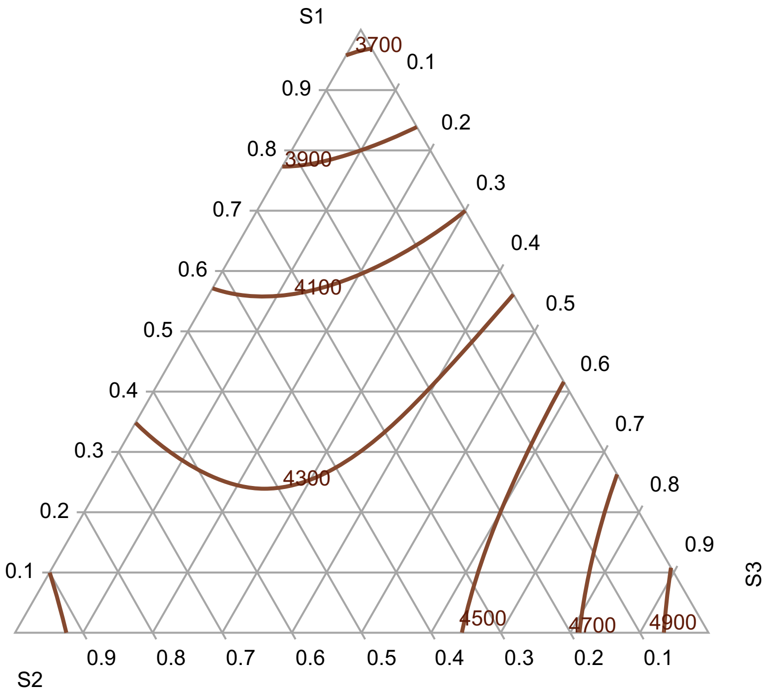
The individual starch effects in the model correspond to the viscosity levels where each is the only starch used. The interaction of starch 2 and starch 3 is highly significant with a p-value much smaller than 0.05. Maria is not sure how to interpret that without some type of visualization. The other two interactions have p-values higher than 0.05. Since the model already fits so well, it does not seem that refining the model by removing these terms will affect her interpretation and predictions, so she does not make that effort for this design.

Her software has visualizations of her model that help her to understand the effects of the three starches. One is a profiler that shows the effects of increasing or decreasing any one of the starches while keeping the other two in the same proportions. The profiler looks like this:



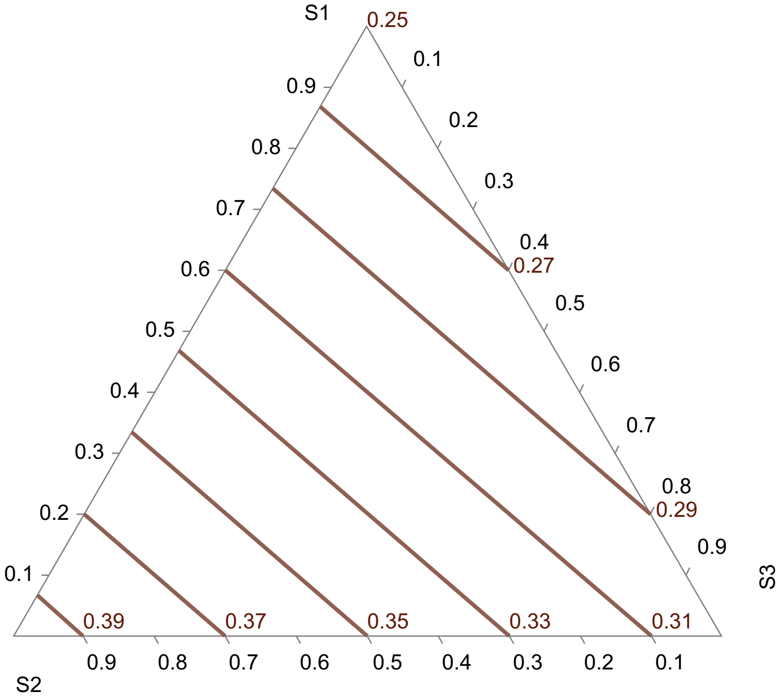
In this plot, each of the starches is in the blend in equal proportions—0.33 for each. The predicted viscosity for this blend is 4284, and a 95% confidence interval for this prediction is also displayed. Maria can see that by increasing starch 1, while keeping the other two in equal proportions, the viscosity will decrease. By increasing starch 3, while keeping the other two in equal proportions, the viscosity will increase. The effect is similar for starch 2, though the increase in viscosity is much smaller.

The second visualization is a contour plot of the ternary space. Maria creates the plot so that it displays the contour for her viscosity target of 4300 with other contours in increments of 200. The plot looks like this:



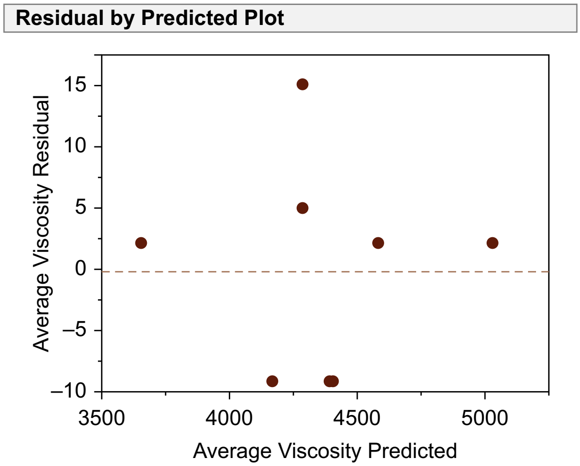
Maria sees that there are many blends that will result in a predicted viscosity of 4300, including two at the edges of the triangle that would be blends of only two starches. On the left side of the triangle, the blend of 35% starch 1 and 65% of starch 2 is predicted to meet this target. On the right side of the triangle, the blend of 56% starch 1 and 44% of starch 3 is also predicted to meet this target.

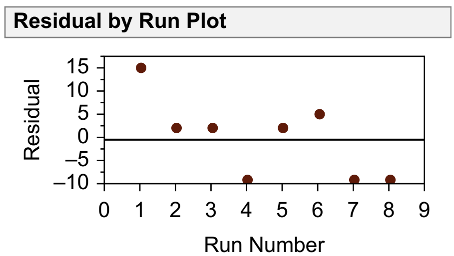
Maria now creates a visualization of the cost of the blends in the ternary space. She has contours displayed at intervals of 0.02. Since these contours are straight lines, she removes the reference lines on the graph so that the cost contours can be more easily seen. The plot with formula cost contours looks like this:

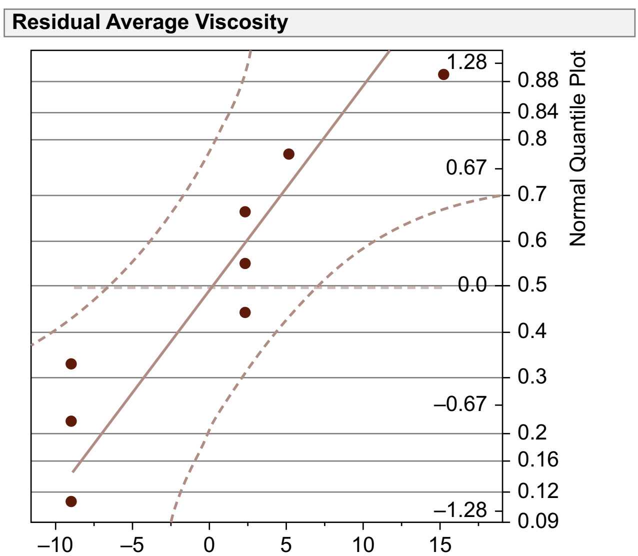


Maria sees that cost decreases moving from the lower corner of the triangle to the top. This makes sense since starch 2 is most expensive, and moving in this direction decreases the amount of that starch in the blend. Putting this information together with the viscosity prediction, she determines that the blend of 56% starch 1 and 44% of starch 3 is the lowest cost blend to meet her target.

As with her other designs, Maria verifies that her model is not missing anything critical by analyzing the residuals. There should be no pattern in the residuals if she plots them against the order of the design runs and the predicted values from the model, and they should follow a normal distribution centered at zero. She again creates plots to perform this assessment.


Residual by Predicted Plot
Average Viscosity Residual
15
10
5
0
–5
–10
3500
4000
4500
5000
Average Viscosity Predicted


Residual by Run Plot
Residual
15
10
5
0
–5
–10
0 1 2 3 4 5 6 7 8 9
Run Number


Residual Average Viscosity
1.28
0.67
0.0
–0.67
–1.28
0.88
0.84
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.16
0.12
0.09
Normal Quantile Plot
–10
–5
0
5
10
15

Maria sees nothing in these graphs to suggest that her model is not adequately describing the factor effects on viscosity. She conveniently finds Steve at his desk and shares the modeling results with him.

Steve: "This is a great result. Not only have we reduced the cost from the current formula, but also using only two starches in the blend simplifies production."

Maria: "While I'm confident with our model, like with our response surface model, it is important that we validate the model predictions. I think that we should make four validation runs, two blends at the target of 4300 and repeat two design runs from the experiment. This is the approach we took before."

Steve: "I agree. Let's use the low-cost two-starch blend and a three-starch blend with a cost similar to the current for the two predicted to meet the target. Pick two from the design that span a wider range of viscosity."

Using this guidance Maria comes up with the following runs for the validation study:

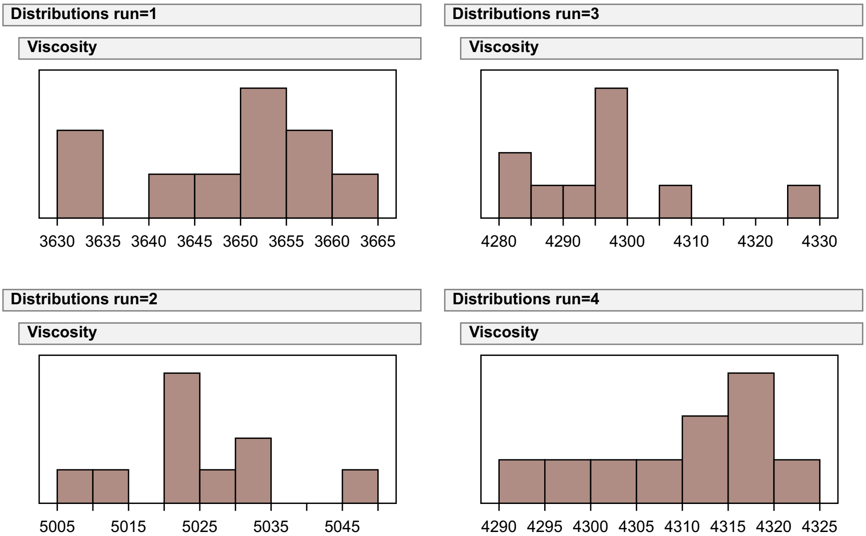
run	S1	S2	S3	Cost	observed viscosity from design	predicted viscosity
1	1.00	0.00	0.00	0.25	3656	3651
2	0.00	0.00	1.00	0.30	5025	5028
3	0.56	0.00	0.44	0.27		4302
4	0.40	0.20	0.40	0.30		4303

Steve agrees with the plan and arranges with Connie, Alex, and Chad the time to execute these runs the following week.

Maria receives the viscosity data from Chad shortly after the execution.

run	V 01	V 02	V 03	V 04	V 05	V 06	V 07	V 08	V 09	V 10
1	3658	3634	3652	3634	3653	3656	3642	3663	3652	3648
2	5026	5007	5012	5022	5022	5034	5025	5048	5024	5030
3	4292	4326	4300	4282	4287	4300	4295	4309	4284	4298
4	4306	4297	4312	4317	4300	4317	4311	4323	4292	4317

Maria again starts by creating distributions of the data by run.



She sees no substantial differences in the distributions for the runs. Like before, she also calculates the range and standard deviations for the runs. Thus the summary statistics look like this:

run	Average Viscosity	Range	Standard Deviation
1	3649	29	10
2	5025	41	11
3	4297	44	13
4	4309	31	10

Maria sees no substantial differences in these summary measures of variation and now feels confident that she can compare the run averages to the model predictions and for the first two observed values from the design. She adds these means to her earlier table, and it now looks like this:

run	S1	S2	S3	Cost	observed viscosity from design	predicted viscosity	observed validation viscosity
1	1.00	0.00	0.00	0.25	3656	3651	3649
2	0.00	0.00	1.00	0.30	5025	5028	5025
3	0.56	0.00	0.44	0.27		4302	4297
4	0.40	0.20	0.40	0.30		4303	4309

It is clear that the results validate the predictions and that the blend of 56% starch 1 and 44% starch 3 can be used. Maria shares the validation results with Steve.

Steve: "Maria, this is great news. If we had to continue to use the existing starch blend our formula would have more than doubled to $0.63.

We would have to pass this larger ingredient cost onto the consumers, or severely reduce our margins. As a result of your team's great work, we can now share that we will save a few cents on a bottle instead of facing the unpleasant moment of increasing BBQ sauce retail shelf price. Your mixture design is a lovely thing!"

This page intentionally left blank

CHAPTER 14

Wrapping It All Up!

As Maria approaches her end of year performance review at the end of her first year at Ultimate BBQ, Maria tries to list all of the things she has learned. Of course, she had never been in a food production facility before, and all of the sanitation and safety rules that have become second nature were new at first. Her day-to-day work has been different than what she had expected—she did not realize how much time she would be spending in the production facility. And, she is most proud to have developed her statistics and data analysis skills and applied them to solve problems and make a real impact.

Maria is planning to visit her university in a few weeks for homecoming weekend. She emails Dr Wang, mentioning her visit and inviting her to lunch to celebrate her first successful year as a food scientist. She is pleased with Dr Wang's enthusiastic response, and after exchanging a few more brief emails, they settle on a time and place. A few weeks pass, and on the agreed upon date, Maria arrives only a moment before Dr Wang at Little Smokey's BBQ Shack.

Dr Wang greets Maria "I laughed out loud when you suggested this place in your email! I would think by now that you are tired of BBQ!"

Maria chuckles "I guess I've become obsessed by BBQ!"

The host leads them to a table, and soon after, they place their orders with a waiter. Maria insists that they order two dishes with very different sauces so that she can check them both out.

While they are waiting for their order to arrive, Maria says "I'm amazed how much I've learned in just one year. You prepared me well here at the university and away from the classroom throughout my first year. Thank you so much!"

"I've been happy to do so. "Dr Wang responds." It is wonderful to see you grow into a solid food scientist. What was your biggest surprise?"

Maria had already been thinking about this and is quick to respond. "I would not have thought that so much of the work would involve production. However, an even bigger surprise is how much I use statistical tools."

Statistics for Food Scientists
http://dx.doi.org/10.1016/B978-0-12-417179-4.00014-7

Dr Wang smiles and says "From your phone calls and emails, I know some of the tools you have been using. And I think you have picked up some more from colleagues. Let's see if we can make a list of these skills and how you have used them."

Maria starts her list by saying "You taught me about summary statistics and the importance of visualizing data rather than just crunching numbers. And we had performed some hypothesis tests in your class, but I did not really see how they would be used to answer real-life questions. Early on, when I used these tools to compare the production lines, hypothesis tests were important in helping us make decisions about how the lines were performing. This led us to focus on where the real problems were."

Dr Wang adds "I remember too that you had to determine if the viscosity measurements were reliable."

"Oh yes, that was so important!" Maria agrees "I just assumed that the numbers coming from an expensive instrument are correct. But it is people who are operating the instruments, so we had to look at the entire system, including operators and sample preparation. We now repeat that assessment of the measurement system every quarter since if the measurements are not reliable, everything else we are doing is a mess."

Dr Wang adds "So, once you ensured that the measurement process was good then you made a control plan for viscosity, right?"

Maria quickly agrees "I thought I had determined that it was steam pressure that most affects viscosity and thought that a control plan based on adjusting that alone would be sufficient. But it was not as simple as that since other things in addition to steam pressure affect viscosity."

Dr Wang nods in agreement "I think that was when I suggested that you learn more about experimental design."

"Yes" Maria agrees "but even before that I had the situation where we had a whole shift of production on hold because of the problem with the measurement system. I had thought of the product as being defective or not defective, a binary result, and came up with a sampling plan based on this consideration. You gave me some great advice on how to think about exploiting that the viscosity measurements are continuous in order to come up with a more efficient sampling scheme in the future."

"I remember that! "says Dr Wang." We all learn how to do things better based on hindsight. But, you were still able to give your colleagues good advice on how to determine what to do with the product you had on hold."

Maria agrees and continues "Even though I had created the control plan and also a monthly reporting system, our process still was not producing at the needed levels. The design of experiments approach was instrumental in understanding the factors influencing this and ultimately understanding how to better control viscosity."

Dr Wang smiles "I think you now know more about designing experiments than I do! Not only did you go through the typical screening and optimization steps, but you also used a mixture design at some point, right?"

"Yes "Maris says enthusiastically" we needed to change the starch blend when one of the ingredients became unavailable because of a shortage. I just finished up that work a few weeks ago."

Dr Wang concludes "Your statistical skills have certainly developed greatly in your first year working as a food scientist. Your questions made me realize the need for a resource list for my students. I consulted several professors at the University's Statistics department. I'll email you the list. And I'll be sure to mention your experiences in my classes in hopes of stressing to my students the importance of these skills in becoming effective food scientists."

"That is so flattering to me! "Maria gratefully responds" I was planning to pick up the check anyway; now, I'm certainly going to do so!"

When Maria returns from her trip, she finds the email from Dr Wang with the resource list. She will discuss purchasing these texts with Steve at her next development meeting.

Dr Wang's Resource List:

David S. Moors, George P. McCabe: Introduction to the Practice of Statistics.

Douglas C. Montgomery: Introduction to Statistical Quality Control, 7th Edition.

George E. P. Box, J. Stuart Hunter, William G. Hunter: Statistics for Experimenters: Design, Innovation, and Discovery, 2nd Edition.

Douglas C. Montgomery: Design and Analysis of Experiments, 8th Edition.

This page intentionally left blank

INDEX

Note: Page numbers followed by "f" and "t" indicate figures and tables, respectively.

Q

R

S

T

THE PRACTICAL APPROACHES CHAMPIONED IN THIS BOOK have led to increasing the quality on many successful products through providing a better understanding of consumer needs, current product and process performance, and a desired future state. In 2009, Frank Rossi and Viktor Mirtchev brought their practical statistical thinking forward and created the course "Statistics for Food Scientists." The intent of the course was to help product and process developers increase the probability of their projects' success through the incorporation of practical statistical thinking in their challenges. The course has since grown and has become the basis of this book.

FRANK ROSSI is Associate Director of Statistics at Kraft Foods Group. Much of his career as an industrial statistician has been in the food industry, including positions atthe Campbell Soup Company and General Foods Corporation. A frequent presenter at scientific symposiums related to the food industry, he is also a regular presenter at short courses for the consumer products industry. He is the author and co-author of a number of publications on statistics in sensory science and consumer research. He obtained a BS in Mathematics and an MA in Statistics from The Pennsylvania State University.

VIKTOR MIRTCHEV is an Analytics Manager in the Consumer Insights and Strategy at Kraft Foods. During Viktor's professional tenure, he has held positions in Statistics, Marketing Analytics, Quality and Operations in the CPG, pharmaceutical, and logistics industries while working at Kraft Foods, Baxter Healthcare and UPS. Over the years Viktor has shared his passion about statistics while teaching fundamentals, solving interesting problems, and empowering teams with analytical capabilities. He is a Six Sigma Master Black Belt and holds a BS in Statistics and Computer Science, and an MS in Applied Statistics and an MBA.

ELSEVIER

ACADEMIC PRESS

An imprint of Elsevier
store.elsevier.com

ISBN 978-0-12-417179-4

9 780124 171794