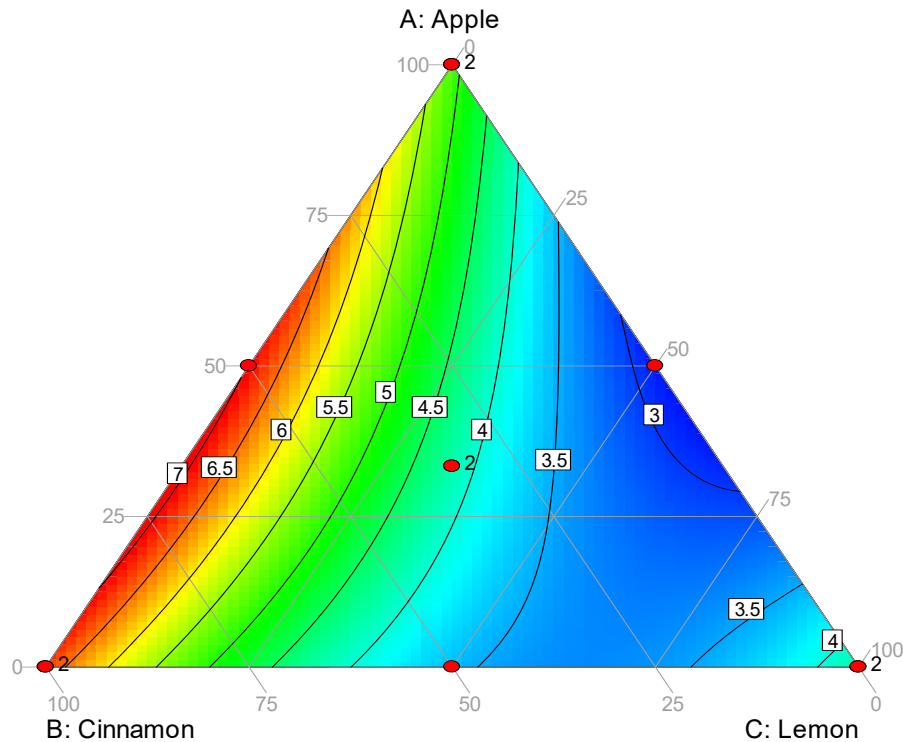# A Primer on Mixture Design: What's in it for Formulators?

Mark J. Anderson

Patrick J. Whitcomb

Martin A. Bezener

Chapter 1: Getting Your Toe into Mixtures

"Simplicity is the ultimate sophistication."
— Leonardo Da Vinci

Come on in – the water's fine! Ok, maybe you'd do best by first sampling the temperature of the pool with just your finger or toe. That's what we will try to do in this chapter – start with the simple stuff before getting too mathematical and statistical about mixture design and modeling for optimal formulation. Our two previous books, *DOE Simplified* and *RSM Simplified*, both feature chapters on mixture design that differentiate this tool from factorials and response surface methods; respectively. However, if you did not read these books, that's OK. We will start with an empty pool and fill it up for you!

It's natural to think of mixtures as liquids, such as the composition of chemicals a pool owner must monitor carefully to keep it sanitary. However, mixtures can be solids too, such as cement or pharmaceutical excipients – substances that bind active ingredients into pills. The following two definitions of mixtures leave the form of matter open:

- "Mixtures are combinations of ingredients (components) that together produce an end product having one or more properties of interest." – John Cornell & Greg Piepel (2008)
- "What makes a mixture?
  1. The factors are ingredients of a mixture.
  2. The response is a function of proportions, not amounts.
  ⇒ Given these two conditions, fixing the total (an equality constraint) facilitates modeling of the response as a function of component proportions."
  – Whitcomb (2009)

The first definition by Cornell and Piepel provides a practical focus on products and the interest that formulators will naturally develop for certain properties of their mixture (as demanded by their clients!). However, the second specification for a mixture provides more concise conditions that provide a better operational definition. Pat suggests that formulators ask themselves an easy question: "If I double everything, will I get a different result?" If the answer is no, such as it would be for a sip of sangria from the glass versus the carafe, for example (strictly for the purpose of tasting!), then mixture design will be the best approach to experimentation.

THE QUINTESSENTIALS OF MIXTURES

Mixture experiments date back to ancient times when it was thought that all matter came from four elements: water, earth, fire and air. For centuries, alchemists sought the magical fifth element, called the "quintessence", which would convert base metals to gold. Petrochemicals made from 'black gold' (oil) formed the focus for Henri Scheffé's pioneering article in the field of statistical design and analysis of mixtures – Experiments with Mixtures *(1958). Perhaps this interest was sparked by his work in World War II, which remains obscure (not revealed at the time) but described in general terms as having to do with "the effects of impact and explosion." (Daniel & Lehmann, 1979) Thus one assumes that Scheffé studied highly exothermic mixtures in his formative years as an industrial statistician.*

Later, Herman Sahrmann, John Cornell and Greg Piepel (1987) created a stir in the field for their mixture experiments on Harvey Wallbangers – a popular cocktail in the 1970s. A potential problem with mixture experiments on alcoholic drinks is that, unless the tasters are professional enough to refrain from drinking the little they sip, after several samples, the amount of alcohol ingested could matter – not just the proportions. Therefore, one must never permit sensory evaluators to consume the alcoholic beverage – only sip, spit and rinse afterward with water. Keep that in mind if you wish to apply the methods of Cornell's landmark book (2002) to such purpose (for example, if you become inspired by our case study in Chapter 2 on blending beers).

P.S. Of course, some mixtures are better liberally applied – for example primer paint – the more the better for hiding power. This would be a good candidate for a "mixture-amount" design of formulation experiment. They require more complicated approaches and modeling so let's set this aside for now. We will devote our full attention to mixture-amount experiments towards the end of the book.

"With mixtures the property studied depends on the proportions of the components present, but not on the amount." – *Henri Scheffé*

All that glitters is not gold

Let's now dive in on the shallow end of design and analysis of experiments with mixtures. This will go easiest with us leading by example via a case study.

Some years ago Mark enjoyed a wonderful exhibit on ancient gold at the Dallas Art Museum. It explained how goldsmiths adulterated gold with a small amount of copper to create a lower melt-point solder that allowed them to connect intricately designed filigrees to the backbone of bracelets and necklaces. This seemed very mysterious given that copper actually melts at a higher temperature than gold! However, when mixed together these two metals melt at a lower temperature than either one alone. This is a very compelling example of synergism – a surprisingly beneficial combination of ingredients that one could never predict without actually mixing them together for experimental purposes.

Figure 1-1: This exquisite necklace, now in the London's British Museum, came from a necropolis (burial site) on Rhodes. It features Artemis, the Greek goddess of hunting. (Bridgeman Art Library, London/New York)

A EUREKA MOMENT!

You may recall from studying Archimede's principle of buoyancy that this Greek mathematician, physicist, and inventor who lived from 287-212 B.C. was asked by his King (Hiero of Syracuse) to determine whether a crown was pure gold or was alloyed with a cheaper, lighter metal. Archimedes was confused as to how to prove this, until one day, observing the overflow of water from his bath tub, he suddenly realized that since gold is more dense, a given weight of gold represents a smaller volume than an equal weight of the cheap alloy and that a given weight of gold would therefore displace less water. Delighted at his discovery, Archimedes ran home without his clothes, shouting "Eureka," which means "I have found it." When you make your discovery with the aid of mixture design for optimal formulation, feel free to yell Eureka as well, but wait until you get dressed!

PS. If you have a copy of DOE Simplified, *see the Chapter 9 sidebar "Worth its weight in gold?" for a linear blending model we derived based on the individual densities of copper versus the much heavier (nearly double) gold.*

The ancient Greek and Roman goldsmiths mixed their solder by a simple recipe of 2 parts gold and 1 part copper (Humphrey, 1998). The use of "parts," while extremely convenient for formulators as a unit of measure, is very unwieldy for doing mathematical modeling of product performance. The reason is obvious, the more parts of one material that you add, the more diluted the other ingredients become, but there is no quantitative accounting for this. For example, some goldsmiths added 1 part of silver to the original recipe. That now brings the total to 4 parts and thus the gold becomes diluted further (2 parts out of 4 or 50 percent, versus the original concentration of 2/3rds or about 67 percent). Therefore, one of the first things we must do is wean formulators wanting to use modern tools of mixture design off the old-fashioned parts. In this case it will be convenient to specify the metal mixture by weight fraction – scaled from zero (0) to one (1). However, all that really matters is that the total be fixed, such as one for the weight fraction. So alternatively, if our goldsmith used a 50 milliliter crucible, then the ingredients could be specified by volume – provided that together they always added to 50 ml. You will see various units of measure used in mixture designs throughout this book, although perhaps the most common may be by weight. But the first thing we will always specify is the total.

Getting back to the task at hand, let's see the results for the temperature at which various mixtures of gold and copper begin to melt. Assume this was done in ancient times when measurements were not very accurate. (This is a pretend experiment!) We've covered the entire range from zero to one of each metal. However, we could have constrained the experimental region to only the blends with at least 50 percent of gold (0.5 weight fraction), assuming that jewelry buyers probably would not like any part of their precious purchase to be mostly base copper. Dealing with necessary constraints is a vital aspect of mixture design, but it introduces complications that had best be put off for now. (Remember we are trying to start off simply!)

ID	Point Type	Blend Type	Gold wt fraction	Copper wt fraction	Melt Point Deg C
1	Vertex	Pure	0.00	1.00	1063
2	"	"	0.00	1.00	1083
3	Axial Check Blend	Quarter	0.25	0.75	955
4	Centroid	Binary	0.50	0.50	926
5	"	"	0.50	0.50	921
6	Axial Check Blend	Quarter	0.75	0.25	952
7	Vertex	Pure	1.00	0.00	1049
8	"	"	1.00	0.00	1036

Table 1-1: Melt points of copper versus gold and mixtures of the two

Notice that the table sorts the blends by their purity of gold. The actual order for experimentation can be assumed to be random. As emphasized in both our previous books on statistical design, randomization provides insurance against lurking variables such as warm-up effects from the furnace, cross-contamination in the crucible, learning curves of operators and so forth. As the inventor of modern-day industrial statistics R. A. Fisher said

"Designing an experiment is like gambling with the devil: only a random strategy can defeat all his betting systems."

Another important element of this experiment design is the replication designated in the descriptor columns (point type and blend type) by ditto marks ("). We advise that at least three blends be replicated in the randomized plan, preferably four or more. These provide a measure of pure error, desirable for statistical purposes, but as a practical matter the replicates offer an easy way for formulators to get a feel for their inevitable variations in blending the materials and measuring the response(s) – simply look at the results from run-to-run made by the same recipe.

Generating a beautiful response surface – like a string of rubies on a gold strand!

Ok, perhaps we are getting carried away in our enthusiasm for using data from a well-designed mixture experiment to produce a very useful plot of predicted response at any given composition. Here is our equation, fitted from the experimental data by least squares regression, for modeling the melt point as a function of the two ingredients, gold and copper, symbolized by x_1 and x_2; respectively. These input values are expressed on a coded scale of zero to one, which statisticians prefer for modeling mixtures.

$$\text{Melt point} = 1044\, x_1 + 1071\, x_2 - 543\, x_1 x_2$$

This mixture model, developed by Henri Scheffé (1958), is derived from the usual second order polynomial for process response surface methods (RSM), called a quadratic equation. The mathematical details are spelled out by Cornell (2002). Two things distinguish Scheffé's polynomial from that used for RSM. First of all, there is no intercept. Normally this term

represents the response when factors are set to zero – set by standard coding to their midpoints for process modeling. However, a mixture would disappear entirely if all the components went to zero – we can't have that! The second aspect of this second order mixture model that differs from those used for RSM is that it lacks the squared terms. Again, refer to Cornell's book for the mathematical explanation, but suffice it to say for our purposes (keeping it simple) that the $x_1 x_2$ term captures the non-linear blending behavior – in this case one that is synergistic, that is – a desirable combination of two components.

THE JARGON OF DOE ON MIXTURES VERSUS PROCESS

Cornell and other experts are very particular on how one describes the elements of design and analysis for mixture experiments. For example, always refer to the manipulated variables as "components" – not factors. Those of you who are familiar with factorial DOE and response surface methods (RSM) will see other aspects that are closely paralleled in this book on mixture design, but named differently. One of the traps you may fall into is referring to the second order mixture term $x_i x_j$ as an interaction. If you say this in the presence of the real experts, you'd best duck and cover as school children were advised in the Cuban Missile Crisis (dating ourselves here!) – the proper descriptor is "nonlinear blending." Although it seems picky, there is a good reason for this: Curvature and interaction terms that appear in process models become partially confounded due to the mixture constraint that all components sum to a fixed total. Do not fight this – just don't say these words! To develop a high level of expertise in any technical field, one must learn the technical terms and express them with great care to maintain precision in communication. This can be a pain, but it provides great gain.
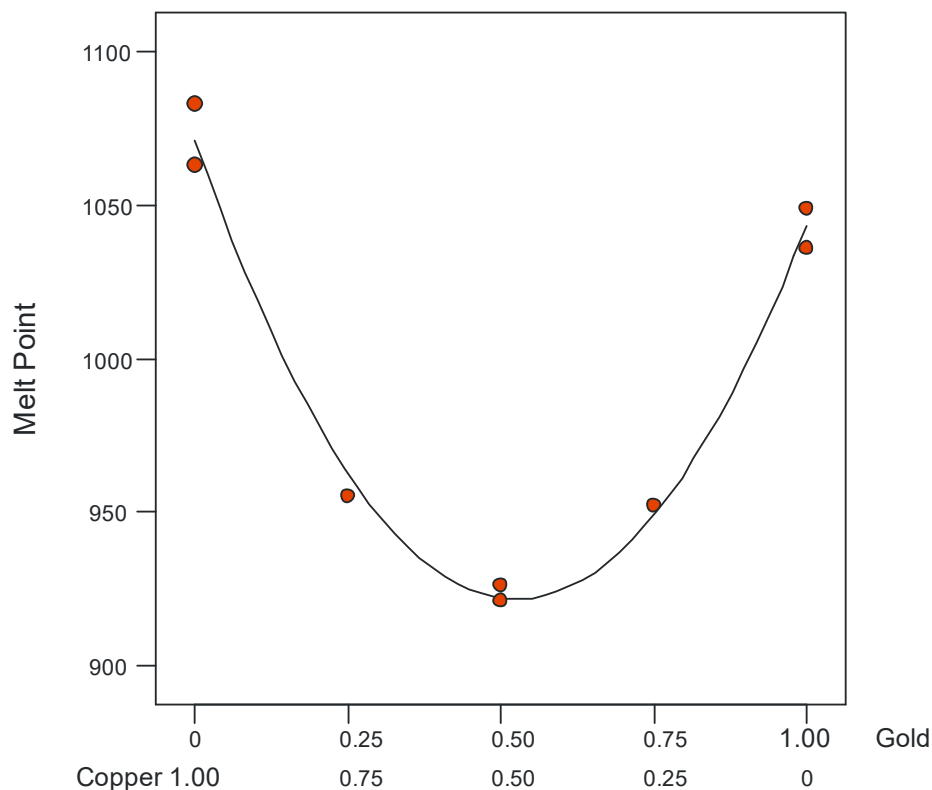

Figure 1-2: Response surface for melt point of copper versus gold and their mixtures

Observe that although this experiment requires the control of two inputs – gold versus copper, only one X axis is needed on the response surface graph. That is because of the complete inverse correlation of one component with the other – as one goes up the other goes down and vice-versa. In statistical terms this can be expressed as r = -1, where r symbolizes correlation and the minus sign indicates the inverse relationship.

Let's see how that model for melt point connects to the graph. First of all, the coefficient of 1044 for x_1 estimates that temperature in degrees Celsius at which pure gold melts. On the other hand, pure copper melts at a higher temperature – estimated from this experiment to be 1071 C. Always keep in mind that results will vary from any given experiment, which represents only a sampling of the true population of all possible results from your process – an unknown and unknowable value, but often easy to estimate. So if you are a metallurgist, please forgive the errors in these values for melt points. As a practical matter they will serve the purpose of the goldsmith for determining a good recipe using copper to produce a lower melt-point jewelry solder.

The most intriguing feature of this mixture model is the large negative coefficient of 543 on the x_1x_2 term. The analysis of variance (ANOVA) shows the term to be significant at $p < 0.0001$ – a less than 1 out of 10 thousand chance of it being this large if the true effect were null. (For a primer on ANOVA and p-values, refer to *DOE Simplified*.) So together gold and copper melt at a lower temperature than either one alone – isn't that amazing!

OTHER DEPRESSING COMBINATIONS OF MATERIALS

Depressed melt points from mixtures of one material with another, such as gold with copper, are not that uncommon. The point at which a mixture of two such substances reaches the minimum melting temperature is called the "eutectic." For example, an ideal solder for electronic circuitry is made from 63 percent tin (m.p. 450 °F) and 37 percent lead (m.p. 621 °F) – together these metals melt at a lower temperature (361 °F) than either one in pure form. The constituents crystallize simultaneously at this temperature from molten liquid solution by what chemists call a eutectic reaction. The term comes from the Greek eutektos, meaning 'easily melted.'

The most prevalent eutectic reaction that we encounter in Minnesota occurs when our highway workers spread salt on roads to aid snow removal. The eutectic point for sodium chloride occurs at 23.3 weight percent in water at a freezing point of minus six degrees Fahrenheit. As salt is added to the mixture of water and ice on winter roads some of the ice melts due to the depression of the melt point. That causes heat to be absorbed from the asphalt or concrete surface, which is no big deal – it's got lots to give. However, in a well-insulated environment like the jacket of an old-fashioned ice-cream maker this effect becomes very chilling (and useful!).

Mathematically, due to the coding on a zero to one scale for each component, the maximum impact of this second-order x_1x_2 effect occurs at the 0.5-0.5 ("50/50") blend. Some quick figuring will help you see that this must be so. First multiply 0 by 1 and 1 by 0 to get the products at either end of the scale. If you do not compute zero in both cases, then perhaps you possess the street smarts to be a vendor like the one we quote in the sidebar below. Now things get a lot harder because fractions are involved. Multiply ¼ by ¾ and ¾ by ¼ to work out the result for the two axial check blends that this design specifies between the centroid and the

vertices. If you got by the first calculation, we trust you know that either way this product comes to three sixteenths. This is a little less than the $1/4^{th}$ result you get from multiplying 0.5 by 0.5 for the "50/50" blend at the centroid.

A FISHY WAY TO BLEND 50/50

A New Orleans street vendor was asked how he could sell Gulf shrimp-cakes so cheap. "Well," *he explained,* "I have to mix in some big old Mississippi catfish that the trawler dredges off the river bottom when it makes a shrimp run. But I mix them 50:50 – one shrimp, one catfish."

If you look closely at the curve in Figure 1-2, you may notice that the minimum actually occurs just a little to the right of the 0.5-0.5 point. This is due to the gold having a lower melt point than the copper, thus favoring a bit more of this noble metal. A computerized search for the minimum using a hill-climbing algorithm finds the minimum at 0.525 weight fraction gold, and thus 0.475 percent copper is required to make the two components total to 1.

Now for a major disclaimer: A mixture experiment like this one on gold and copper will only produce an approximation of the true response surface – it may not be accurate, particularly for the fine points such as the eutectic temperature. In the end you must ask yourself as a formulator whether the results can be useful for improving your recipe. In this case the next step would be to select a composition that meets the needs of solder for goldsmithing fine jewelry. Determine the predicted melt point from the graph or more precisely via the mathematical model. Then run a confirmation test to see how close you actually get. As a practical matter this might be off by some degrees and yet still be useful for improving your process.

TRIAL OF THE PYX

In Anglo Saxon times the debasing of gold coin was punished by the loss of the hand. In later *years the adulteration of precious metals was prohibited by the Goldsmiths' Company of London* *(founded 1180). The composition of gold sovereigns was eventually fixed at eleven-twelfths fine* *gold, and one-twelfth alloy (copper). So accurate became the composition and weight of the coin* *issued from the mint, that at the 1871 trial of the "Pyx" the jury reported that every piece they* *separately examined, representing many millions of pounds sterling, was found to be accurate* *for both weight and fineness. The term "Pyx," Greek in origin, refers to the wooden chest in* *which newly-minted coins are placed for presentation to the expert jury of assayers assembled* *once a year at the Hall of the Worshipful Company of Goldsmiths in the United Kingdom. This* *ceremony dates back to 1282.*

Source: Encyclopaedia Britannica, 10th Edition (1902).

Details on modeling the performance of a two-component mixture

Our example on blending glittery metals for jewelry provides a specific application of mixture design and modeling. Now that we've enticed you this far, it's time to consider some general guidelines for setting up a formulation experiment and analyzing the results. Let's start with the Scheffé equations for predicting the response from two components.

First order (linear): $\hat{y} = \beta_1 x_1 + \beta_2 x_2$

Second order (quadratic): $\hat{y} = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$

The hat (^), properly known as a circumflex, over the letter y symbolizes that we want to predict this response value. The β (beta) symbols represent coefficients to be fitted via regression.

We detail the third order (cubic), which you are unlikely to ever need, in the appendix to this chapter. There, for added measure, we also spell out the fourth order (quartic) Scheffé equation. By this stage, very complex behavior can be modeled for all practical purposes. However, this process of model-building could continue to infinite orders of the inputs x to approximate any true surface in what mathematicians refer to as a Taylor polynomial series.

The second order equation not only may suffice for your needs to characterize the two primary components in your formulation, but it also could reveal a surprising nonlinear blending effect. The possibilities are illustrated graphically in Figure 1-3, which presumes that the higher the response the better.
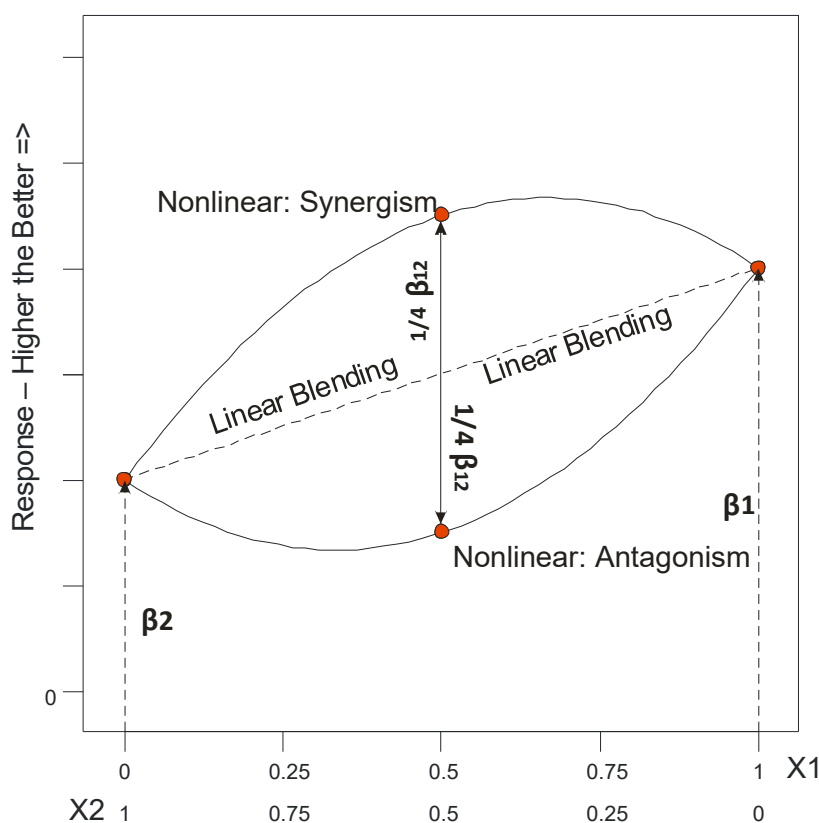

Figure 1-3: Graphical depiction of second-order mixture model

Notice that we tilted the linear blending line upwards, in other words, β_1 exceeds β_2. So this response surface predicts better performance for pure x_1 than for pure x_2. If together these two ingredients produce at the same rates as when working alone, then at the 0.5-0.5 midpoint the

response will fall on the linear blending line. However, you hope that they really hit it off and produce more than either one alone. Then the response will curve upwards – producing the maximum deflection at the midpoint. This synergistic (positive) nonlinear blending effect equals one-fourth (0.5 * 0.5) of the second-order coefficient. Unfortunately, some components just do not work very well together and things get antagonistic. Then the response curves downward and the β_{12} coefficient goes negative.

ISOBOLOGRAMS

In 1871 T.R. Fraser introduced a graphical tool called the "isobologram." It characterized departures from additivity between combinations of drugs. Although it differs a bit in shape from our graph in Figure 1-2, the isobologram is essentially equivalent – it plots the dose–response surface associated with the combination superimposed on a plot of the same contour under the assumption of additivity, that is – linear blending. The observed results are called the "isobol," generally produced for the combinations of individual drug dosages that produce a 50% response by the subjects. If the isobol falls below the line of additivity, a synergism is claimed, because less of the drugs will be needed. On the other hand, if the isobol rises above the line, then the drugs are presumed to be antagonistic. However, there are two major shortcomings associated with the use of isobolograms. They do not account for data variability and they are restricted to only a few components.

Source: Meadows, et al (2002)

In this example we made the response one where higher is better. Thus a positive β_{12} coefficient is desirable for this non-linear blending effect. However, in the first example – blending of copper into gold – the negative non-linear coefficient was what the jewelry maker hoped to see. Thus a synergistic deflection off the linear blending slope on the response surface could be positive or negative, depending on the goal being maximization or minimization.

When experimenting with mixtures, it pays to design an experiment that provides enough unique blends to fit the second order Scheffé polynomial model. Then you can detect possible non-linear blending effects. Hopefully these will prove as advantageous as the synergistic combination of copper with gold for soldering fine jewelry. But it's just as well to know about antagonistic ingredients so you can keep these separated!

AN ANALOGY FOR NONLINEAR BLENDING THAT WILL WORK FOR YOUR MANAGER

In the go-go years of the computer industry of the late 1990's (before the dot-com bust) a movement developed for extreme programming (XP) – a form of agile development. One of its premises, which many software executives found counter intuitive, is that two people working at a <u>single</u> computer are just as productive as they would be if kept in separate cubicles, but (this is the payoff!) this pair programming increases software quality. Thus better code emerges without impacting time to deliver. Wouldn't that be a wonderful case of synergistic (nonlinear) blending?

The idea of creating teams of two is nothing new. It's easy to imagine cave men pairing up to more effectively hunt down mastodons. However, an age old problem for managers of such tasks, for example a police chief setting up patrol cars, is choosing the right people to team together.

Unfortunately, some combinations turn out to be antagonistic, that is – they produce less as a partnership than either one alone. That creates a lot of headaches all around.

How does one know which elements will not only prove to be compatible, but more than that – they create a synergy? Experiment!

Practice Problems

To practice using the statistical techniques you learned in Chapter 1, work through the following problems.

Problem 1-1

To reinforce the basics of mixture modeling presented in this chapter, we will start you off with some obvious questions that stem from this imaginary, but commonplace, situation in our heartland of the USA.

The old truck on your hobby farm gets very poor gas mileage. Luckily you can purchase fuel from a wholesaler who serves the agricultural market. They have a low-grade gasoline that you've found will produce 10 miles to the gallon (mpg) when you must drive the old truck all the way back into the city where you normally dwell. It's cheap – only 3 dollars a gallon. Another possibility is to purchase the highly-refined premium gasoline that increases the engine efficiency to 14 mpg. However it costs 4 dollars a gallon.

Consider these questions:

1. Assuming you drive 1,000 miles per year going back and forth from your hobby farm, which grade of gasoline should you buy to minimize your annual fuel cost?
2. Now suppose the wholesaler offers to blend these two fuels 50/50 at $3.55 per gallon: How does this differ from the linear blend of prices?
3. Furthermore, you discover that your old truck gets 13 mpg with this blend of gasolines: Is this a synergism for fuel economy?
4. Should you buy the 50/50 blend of the two grades of gasoline? (Do not assume this will be so. Even if synergism is evident, the beneficial deflection off the linear blending point may not achieve the level of the best pure component. However, in this case the solution requires an economic analysis – look for the best bottom line on costs per year.)

INVERSE TRANSFORMATION PUTS MILEAGE COMPARISONS ON TRACK

When the price of gas went over 4 dollars a gallon, I started paying attention to which of my three cars went where. For example, my wife and her sister traveled 100 miles the other day to do some work at the home of their elderly parents. They had our old minivan loaded up, but, after thinking about it getting only about 15 miles per gallon (mpg), I moved all the stuff over to my newer Mazda 6 Sport Wagon, which gets 25 mpg. That meant no zoom-zoom for me that day going to work, but it was worth enduring the looks of scorn from the other road warriors.

National Public Radio's (NPR) All Things Considered *show on June 19, 2008 led off with this quiz: "Which saves more gas: trading in a 16-mile-a-gallon gas guzzler for a slightly more*

efficient car that gets 20 mpg? Or going from a gas-sipping sedan of 34-mpg to a hybrid that gets 50 mpg?" Of course the counter-intuitive answer is the one that's correct – the first choice.

This is a "math illusion" studied by Richard Larrick, a management professor at Duke University. He found it easy to fool college students into making the wrong choice in puzzlers like that posed by NPR. Larrick suggests that it makes far more sense to report fuel efficiency in terms of gallons per 10,000 miles -- an average distance driven per year by the typical USA car owner. Professor Larrick was inspired to promote "gpm" (vs mpg) after realizing in the end that he'd be better off trading in the family minivan and only gaining 10 miles per gallon with a station wagon; rather than swapping his second car, a small sedan, for a highly efficient hybrid.

Are you still not sure about the NPR puzzler? Imagine you and your spouse work at separate locations that require an annual commute of exactly 10,000 miles per year for both of you driving separately (two automobiles). Then your 16 mpg guzzler consumes 625 gallons (10,000/16). Trading that for a 20 mpg car you need only 500 gallons the next year – a savings of 125 gallons. On the other hand, your spouse drives the far more efficient 34 mpg sedan – it requires only 294 gallons of gas per year (10,000/34). Upgrading this to the 50 mpg hybrid saves only 94 gallons! We will let you do the math on this last bit.

It is surprising how something as simple as an inverse transformation makes things so much clearer.

Source: StatsMadeEasy (www.statsmadeeasy.net) blog of June 30, 2008 by Mark.

Problem 1-2

This exercise stems from an experiment done by Mark with help from his daughter Katie. To demonstrate an experiment on mixtures, they blew up a plastic film canister – not just once, but over a dozen times. The explosive power came from Alka Seltzer® -- an amalgam of citric acid, sodium bicarbonate (baking soda) and aspirin.

Figure 1-4: Apparatus for film-canister rocketry

You can see the experimental apparatus pictured: launching tube, container with water, the tablets, plastic film canister (Fuji's works best), a scale and stop-watch. Research via the internet produced many write-ups on making Alka Seltzer "rockets." These generally recommend using only a quarter of one tablet and they advocate experimentation on the amount of water, starting by filling the canister half way. Mark quickly discovered that the tablets break apart very easily, so he found it most convenient and least variable to simply put in a whole tablet every time (a constant). It then took a steady hand to quickly snap on the top of the canister, over which Katie placed the launching tube and Mark prepared to press his stop watch. (Subsequent research on this experiment indicated it would have been far less nerve-wracking to stick the tablet on the lid with chewing gum, put water in the container, put the lid on, and then tip it over – shooting the canister into the air.) After some seconds the explosion occurred – propelling the lid from the back porch to nearly the roof of his two-story home.

HEADS UP! DO NOT PICK PRANKSTERS AS YOUR ASSISTANT ON ROCKET SCIENCE.

Those of you who are fans of Gary Larsen's Far Side *series of cartoons may recall a classic on depicting a white-coated scientist putting the last nail on the nosecone of a big rocket. In the background you see his assistant sneaking up with an inflated paper bag – poised to pop it!. Mark's rocketry assistant Katie discovered that enough fizz remained in the canister to precipitate a second blow up. On randomly chosen runs she would sneak up on her father while he recorded the first shot's results and blast away. The only saving grace for Mark was the ready availability of Alka Seltzer for the ensuing headache.*

Before designing this experiment, Mark did some range finding to discover that only 4 cubic centimeters (cc) of water in the 34 cc canister would produce a very satisfactory explosion. However, it would not do to fill the container completely because the Alka Seltzer effervesced too quickly and prevented placement of the lid. After some further fiddling, Mark found that a reasonable maximum of water would be 20 cc's – more than half full. He then set up a two-component mixture design that provided the extreme vertices (4 to 20 cc of water), the centroid (12 cc) and axial check blends at 8 and 16 cc's. Mark replicated the vertices and centroid to provide measures of pure error for testing lack of fit.

Blend #	Run	Type	A: Water (cc)	B: Air (cc)	Flight time (Sec.)
1	2	Vertex	4	30	1.88
2	6	Vertex	4	30	1.87
3	4	AxialCB	8	26	1.75
4	3	Center	12	22	1.60
5	8	Center	12	22	1.72
6	5	AxialCB	16	18	1.75
7	1	Vertex	20	14	1.47
8	7	Vertex	20	14	1.53

Table 1-2: Results from film-canister rocket experiment

Just for fun, Mark asked several masters-level engineers, albeit not rocket scientists, but plenty smart, what they predicted – the majority guessed it would make no difference how much water given a minimum to wet the tablet and not so full it would prevent the top going on. This becomes the null hypothesis for statistical testing – assume no effect due to changing the mix of air and water in the film canister.

Table 1-2 shows the results of flight time in seconds for various blends of water versus air. Looking over the data, sorted by amount of water, do you agree with these engineers that this component makes no difference? You may be somewhat uncertain with only an 'intraocular test' – statistical analysis would be far more definitive to assess the significance of the spread in flight times relative to the variation due to blending errors and the perilous process of launching the rockets. Now would be a good time to fire up your favorite statistical software, assuming it provides the capability for mixture design, modeling, analysis, response surface graphics, and multiple-response optimization. In case you have no such program readily available, we offer one via the Internet – see the About the Software section for the website location and instructions for downloading. To get started with the software, try reproducing the outputs embedded in the answer to this problem posted at the same site (in portable document format – .pdf). In the following chapters we will lead you to more detailed tutorials on using this particular DOE program.

<div style="border:1px solid">

BLASTING OFF FROM TUCSON, ARIZONA

After touring the Titan Missile Museum south of Tucson, Arizona, Mark found the toy pictured in their gift shop. This product, made by a local inventor (CSC Toys LLC), improves the aerodynamics of the seltzer-powered rocket by the addition of a nose cone and fins.

Figure 1-5 – The MIGHTY Seltzer Rocket pictured from a launch pad in Tucson

Like these film canister rockets, the thrust of the Titan missile depended on two components, albeit many orders of magnitude more powerful – a precisely controlled combination of nitrogen tetroxide (oxidizer) and hydrazine (fuel) that spontaneously ignited upon contact. This extreme exothermic chemical behavior is characterized as "hypergolic." The fuels were stable only at 58-62 degrees, which meant that temperature control was critical. In 1980 a worker dropped a 9 pound socket from his wrench down a silo and punctured the fuel tank. Fortunately the 8,000 pound nuclear warhead, more destructive than all the bombs exploded in all of World War II, landed harmlessly several hundred feet away. Some years later the Titans were replaced with MX "Peacekeeper" rockets that used solid fuel.

</div>

Chapter 1 Appendix: Cubic Equations for Mixture Modeling (and Beyond)

The full cubic (third order) equation for modeling a two-component mixture is shown below:

$$\hat{y} = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \delta_{12} x_1 x_2 (x_1 - x_2)$$

Notice that the coefficient on the highest order non-linear blending term is distinguished by the Greek letter delta. Think of the letter "d" (delta) as a symbol for the differences ("d" for difference) pictured in Figure 1-6. It depicts a very unusual response surface for two components with only first and third order behavior – the second order coefficient was zeroed out to provide a clearer view of how the new term superimposes a wave around the linear blending line. Also, to add another wrinkle (pun intended) into this surface the coefficient is negative.

$$\hat{y} = \beta_1 x_1 + \beta_2 x_2 - \delta_{12} x_1 x_2 (x_1 - x_2)$$

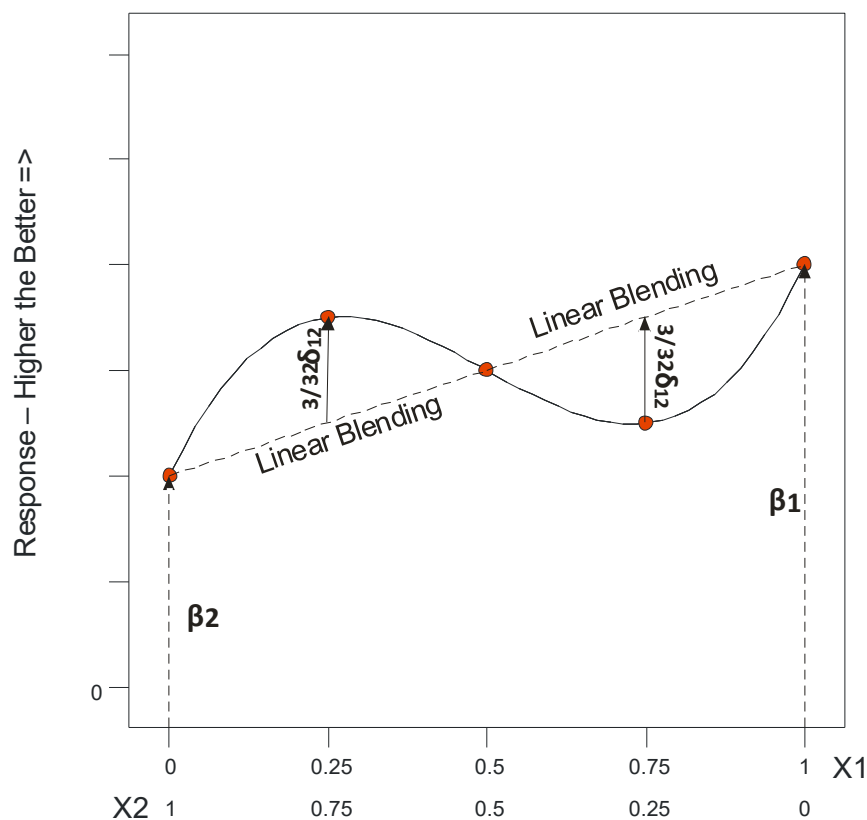See if can bend your brain around this complex mixture model: It's challenging!

Figure 1-6: Graphical depiction of cubic mixture model

At the 50-50 blend point the components are equal, so the offset is zeroed ($x_1 - x_2 = 0$). When x_1 exceeds x_2 to the right of the midpoint, the difference δ (delta) is positive; thus (due to the negative coefficient) the curve deflects downward. To the left the difference goes negative causing the response to go up (a negative times a negative makes a positive!). The maximum wave height from linear blending of 3/32 delta occurs at one-fourth and three-fourth blend points ($\frac{3}{4} * \frac{1}{4} * (\frac{3}{4} - \frac{1}{4}) = 3/32$).

FOR THOSE OF YOU MORE FAMILIAR WITH RSM MODELS

You may wonder why the third order equation for mixtures is more complex than that used to model similar behavior in a process. Remember that the coding for process models goes from –1 to +1. When you cube these quantities, the positive stays positive (+1*+1*+1=+1), and the negative stays negative (-1*-1*-1=-1). Thus you get wavy, up-and-down, behavior in the surface. But in mixtures the coded units are 0 to 1, which will always be positive, so you model waviness via a difference of components.

PS. Since this chapter has kept things simple by focusing only on two components, the cubic mixture model is missing one general term that involves three components: $x_i x_j x_k$. You will see this term highlighted in future chapters that delve into a "special" cubic model, which shortcuts some unnecessary complexities and thus makes things a lot easier for formulators.

For illustrative purposes only, we dramatized the impact of the cubic term in Figure 1-6. Usually it creates a far more subtle "shaping" of the surface such you see illustrated in Figure 1-7, which shows a cubic model (solid line) fitting noticeably better than the simpler quadratic (dotted).
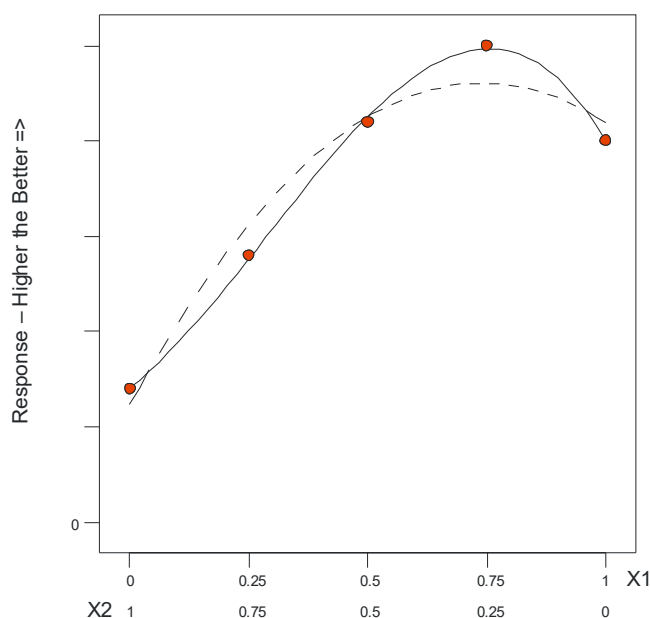

Figure 1-7 – A more subtle surface fitted best by a cubic model

Think of polynomial terms as shape parameters, becoming more subtle in their effect as they increase by order. Linear (first order) terms define the slope. As shown in the blending case of gold and copper we went through earlier in this first chapter of *Formulation Simplified*, the second order (quadratic) fits curvature. The cubic order that we've just introduced in the Appendix fits any asymmetry in the response surface.

It's good at this stage to simply consider mixture design as a special form of response surface methods (RSM), which relies on empirical, not mechanistic, model building. In other words, it's best that you not try relating specific model parameters to the underlying chemistry and physics of your formulation behavior. However, asymmetry is more prevalent in mixture experiments than it is in process experiments. Thus, if the data suffices to fit a surface very precisely to the third order model, it might capture subtle non-linearity that the quadratic would not; that is, this lower equation would exhibit a significant lack of fit.

AN IDIOM FOR NON-LINEAR BLENDING: "CHEMISTRY"

Chemistry is the study of matter and its interactions. It amazes us by unexpected reactions between particular substances. The word "chemistry" is often used to expressively describe a potently positive pairing, such an irresistible attraction between two lovers. Less often one may hear of a "bad chemistry" building up in a group that includes antagonistic elements. Thus this term "chemistry" has become a word that generally describes non-linear blending effects.

"We have very good team chemistry this year."

-- Phil Housley, Minnesota hockey great, assessing his 2009 Stillwater Area High School team

---

In any case, to fit this cubic equation one must design an experiment with at least four unique blends, whereas three suffices (at the bare minimum) to fit the quadratic. The more complex the behavior you want to model, the more work you will have to do as a formulator. You get what you pay for.

If you have plenty of materials and time to mix them together – not to mention the capability for making many response measurements, you could design an experiment that fits a quartic Scheffé polynomial model. Here it is in general terms for however many components ("q") you care to experiment on:

$$\hat{y} = \sum_{i=1}^{q} \beta_i x_i + \sum_{i<j}^{q-1} \sum_{j}^{q} \beta_{ij} x_i x_j + \sum_{i<j}^{q-1} \sum_{j}^{q} \delta_{ij} x_i x_j \left(x_i - x_j \right) + \sum_{i<j}^{q-1} \sum_{j}^{q} \gamma_{ij} x_i x_j \left(x_i - x_j \right)^2 + \sum_{i<j}^{q-2} \sum_{j<k}^{q-1} \sum_{k}^{q} \beta_{iijk} x_i^2 x_j x_k$$

$$+ \sum_{i<j}^{q-2} \sum_{j<k}^{q-1} \sum_{k}^{q} \beta_{ijjk} x_i x_j^2 x_k + \sum_{i<j}^{q-2} \sum_{j<k}^{q-1} \sum_{k}^{q} \beta_{ijkk} x_i x_j x_k^2 + \sum_{i<j}^{q-3} \sum_{j<k}^{q-2} \sum_{k<l}^{q-1} \sum_{l}^{q} \beta_{ijkl} x_i x_j x_k x_l$$

Notice that squared terms now appear. Although statistical software (such as the one we provide to you readers) will handle the design and analysis of a mixture experiment geared to this fourth order, it is very unlikely that this will provide any practical gain over the fit you get from cubic or quadratic models. For response surface modeling it's good to keep in mind the principle of parsimony, which advises that when confronted with many equally accurate explanations of a scientific phenomenon it's best to chose the simplest one (Anderson & Whitcomb, 2005, Chapter 1, sidebar "How Statisticians Keep Things Simple").

OUT OF ORDER?

Back in the days when computer-aided mixture modeling was limited to cubic, an industrial statistician cornered Mark at a conference and complained that he needed quartic to fit a formulation over the entire experimental region. Quadratic fit fine for most of the results but fell short where the performance fell off very rapidly. Mark tried a trick that his doctor told him after he injured his shoulder playing softball. "When does it hurt," the medico asked. "Only when I throw a softball," said Mark. "Just don't do that," the doctor advised. In similar fashion, Mark – being ever practical – suggested that one could simply not look at the response surface where it drops off and gets fit inaccurately, because no one cares at that point.

This is not as unhelpful as you might think. If you can apply your subject matter knowledge and do some pre-experimentation to restrict the focus of the mixture design to a desirable region, the degree of Scheffé polynomial required to approximate the response surface will likely be less, thus reducing the number of blends required by simplifying the modeling needed for adequate prediction power. For example, why model all of the Rocky Mountains when you are really interested only in exploring one of the peaks?

"Some might say that this question is academic, but that's OK because I am an academic."

-- Kevin Dorfman, Professor, University of Minnesota (speaking on esoteric research in his specialty – chemical engineering on the macromolecular scale)

Chapter 2: Triangulating Your Region of Formulation

"If you don't know where you are going, you will wind up somewhere else."
— Yogi Berra

In this chapter we build up from the simplicity of dealing only with two components to experiments on three or more. The biggest step will be in recognizing that if you lay this out in rectangular coordinates then you really do not know where you are going and you will wind up somewhere else (to paraphrase baseball guru Yogi). You need to get yourself into the triangular space depicted in Figure 2-1.
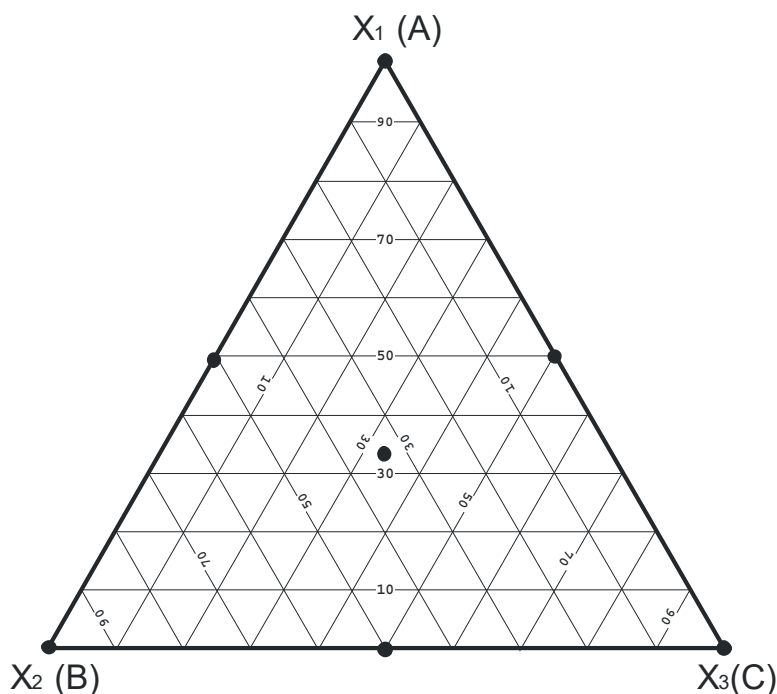
Figure 2-1: Trilinear graph paper for mixtures with points to mark pure components, binary blends and overall centroid

The levels of three ingredients can be represented on this two-dimensional graph paper, also known as "trilinear" for the way it's ruled. It allows for blends of one, two or three materials:
1. Vertices are the pure components. For example, pure X_1 (or ingredient "A") is the point plotted at the top. For the sake of formulators this paper is marked off on a zero to one-hundred scale, which can be easily translated to a more mathematically convenient range of zero to one.
2. Sides are binary blends. The midpoints are 50/50 blends of the components at each end of the side. For example, the point between A and C represents exactly half of each (and none of material B!).
3. Mixtures of three components are in the center area. For example, the point located precisely in the middle of the triangle, called the "centroid," represents a blend of one-third each of all three ingredients.

TRIANGLE SPACE IS TRICKY, BUT IMAGINE GRAPHING ON A MOBIUS STRIP!

Triangular coordinates are also known as barycentric coordinates. (The point at which an object can be balanced is called the "barycenter," derived from the Greek word "barus" for heavy.) This trilinear graphing technique was introduced in 1827 by August Ferdinand Mobius, known for his space-bending Mobius strip – a two dimensional surface with only one side. It never ends!

Source: Yu, C. H., & Stockford, S., "Evaluating spatial- and temporal-oriented multi-dimensional visualization techniques for research and instruction," Practical Assessment, Research & Evaluation*, 8(17), 2003.*

The really neat thing about mapping mixtures to this triangular space is that once you know two component fractions, the third is determined by the total. To illustrate how this works, consider the stainless steel flatware – knives, forks, spoons, etc. – that you keep handy in your kitchen drawer for everyday eating. A very common metallurgical formulation for this purpose is 18 percent chromium (Cr) and 8 percent nickel (Ni) by weight – the remainder being iron (Fe), of course. Let's plot this on the trilinear paper.
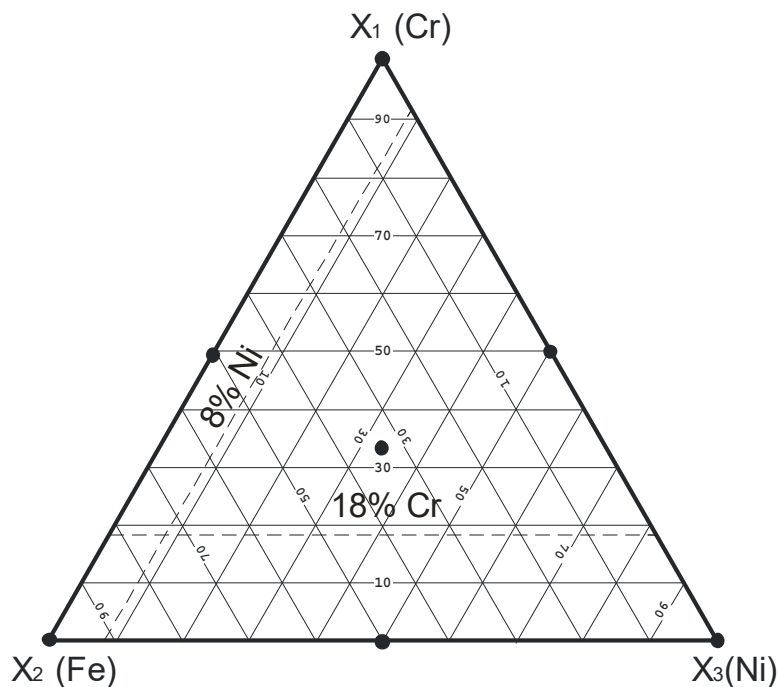
Figure 2-2: Locating the 18-8 composition of stainless steel (for flatware)

In this case the metallurgist conveniently chose chromium as the first component to make the starting point easy – simply draw a line horizontally 18 percent of the way from the bottom (0% Cr) to the top (100% Cr). Now things get tricky because you must rotate the graph so the nickel (Ni) comes out on top. (Think of these triangular graphs as being "turnery" paper!) Then it will be easy to draw the 8 percent line for this metal, designated as the third component. Now turn the graph so iron (Fe) is at the top. Notice that the two lines intersect 74 percent of the way up from

the zero base of iron to its pure component apex. The three ingredients now add to 100 percent! This feature of the ternary graph is very convenient for formulators.

The simplex centroid design

The pattern of points depicted in Figure 2-1 forms a textbook design called a "simplex-centroid" (Scheffé, 1963, Cornell, 2002). We will introduce a more sophisticated design variation called a "simplex-lattice" later on, but let's not get ahead of ourselves. The term "simplex" relates to the geometry—the simplest figure with one more vertex than the number of dimensions. In this case only two dimensions are needed to graph the three components on to an equilateral triangle. However, a four-component mixture experiment requires another dimension in simplex geometry—a tetrahedron (like a pyramid, but with three sides, not four). To show how easy it is to create a simplex centroid, here is how you'd lay it out for four components:

1. Four points for the pure components (A, B, C, D) plotted at the corners of the tetrahedron).
2. Six points at the edges for the 50/50 binary blends (AB, AC, AD, BC, BD, CD).
3. Four three-component blend points at the centroids of the triangular faces of the tetrahedron.
4. The one blend with equal parts of all ingredients at the overall centroid of the tetrahedron.

This totals to 15 unique compositions from the four components. See these depicted in Figure 2-3.
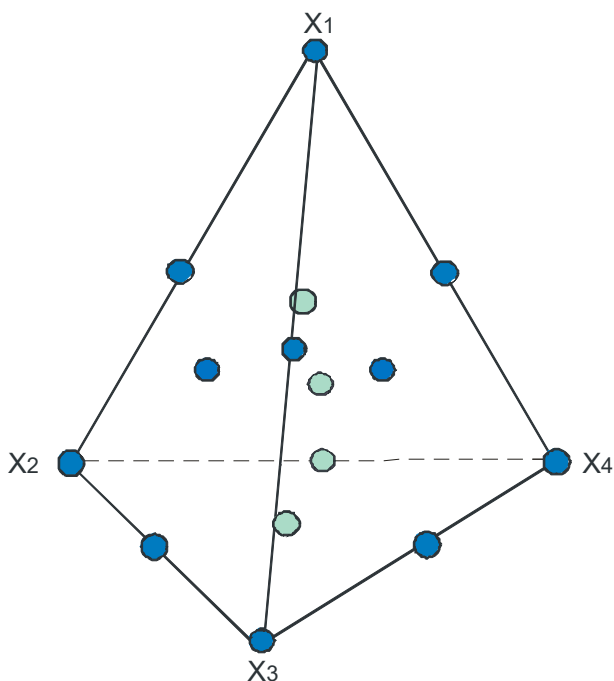
Figure 2-3: Four-component simplex centroid design

Although the simplex-centroid is not a very sophisticated design, it does mix things up very well by always providing a blend of all the components (the overall centroid). Let's work through an example of a simplex centroid for three components, illustrating an entire cycle of mixture

experimentation from design, through actual execution of the runs, statistical analysis and, finally, optimization of multiple responses with cost taken into account.

The Black and Blue Moon beer cocktail

A "cocktail" generally refers to a mixture of hard (high alcohol) liquor. Few drinkers nowadays would think of blending beers. However, one summer Mark was hit by a sudden intuition that it might be very tasty to combine a black lager with a wheat beer – possibly this might produce a synergistic sensation. Furthermore, to provide a contrast to this premium pairing and possibly enable a cost savings, he decided to mix in a cheap lager.

Once this idea took hold, Mark knew it must be tested by unbiased tasters with a talent for drinking beer, and that the experiment itself had to be conducted in a way that would prevent preconceived notions from contaminating the results. Let's see what can be learned via this case study about the application of multicomponent mixture design aimed at discovering a sweet spot of taste versus cost.

DO NOT ASSUME A DIRECT CORRELATION OF COST WITH QUALITY

In the late 1970's Mark took an evening course in marketing en route to his Masters in Business Administration (MBA). He worked through a case study showing how, although the national beer brands in the USA differed very little in their brews, their marketing campaigns divided drinkers into distinct segments. For example, Miller advertised their high-priced product as the 'champagne' of bottled beer while Old Milwaukee went for the working man and took the low road on price. Meanwhile, on Mark's day job as a chemical process development engineer, an R&D colleague made a big deal over how one got what one paid for in beer: The cheap stuff was simply swill in his opinion. At this time Mark was gaining a great appreciation for experiments based on statistical principles, such as use of the null hypothesis for reducing prejudice. Here was an opportunity to put the beer snob to the test via a blind, randomized, statistically-planned experiment. You can guess the outcome: He rated the Old "Swill"waukee (his misnomer) number 1!

"It will come to pass that every braggart shall be found an ass."

 - William Shakespeare (from "All's Well that Ends Well)

"He was a wise man who invented beer." *-- Plato*

Here are the beer-cocktail ingredients (prices per 12 ounce serving shown in parentheses):

   A. Coor's brand Blue Moon Belgian-style wheat ale ($1.16)
   B. Anheuser-Busch brand Budweiser American lager ($0.84)
   C. Samuel Adams brand Black Lager($1.24)

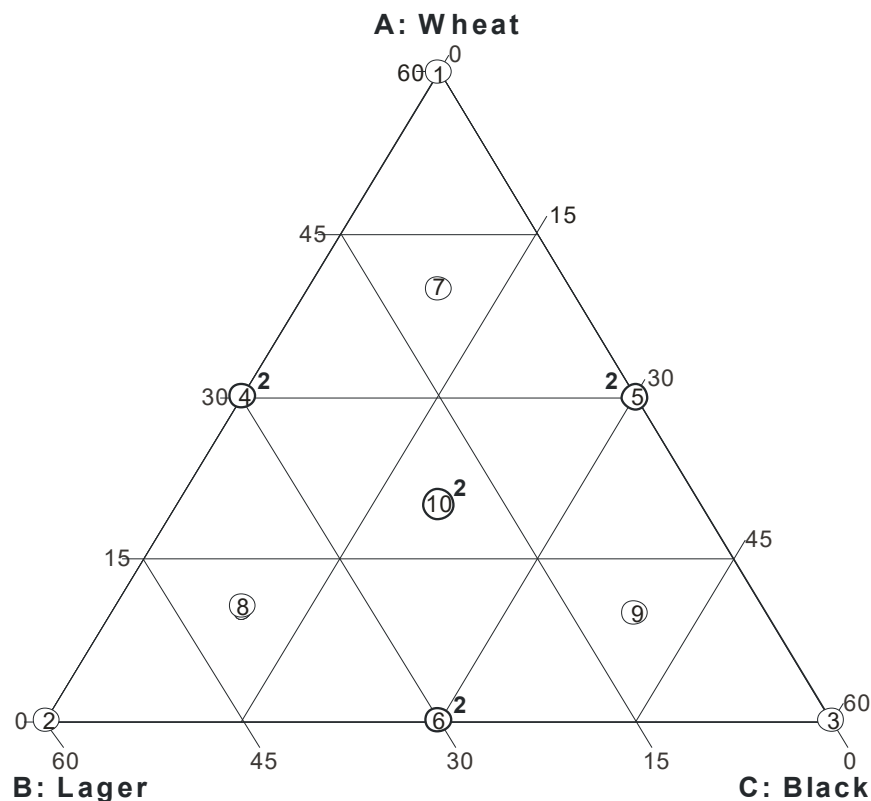The design of experiment, based on a simplex centroid, is laid out in Figure 2-4.

Figure 2-4: Simplex centroid bolstered with replicates and check blends

Mark bolstered this design in three ways:

- Three added blends midway between the centroid and each of the vertices (pure beers). In the jargon of mixture design these are called "axial check blends." They fill otherwise empty spaces in the experimental region. The addition of points to a textbook layout like the simplex centroid is called "design augmentation."
- Four point replicates (designated by "2"s) -- the three binary blends (midpoints of sides) plus the centroid. This provides four measures ("degrees of freedom" in statistical lingo) of pure error. By establishing a benchmark against which the deviations of actual points from the fitted line can be assessed, pure error enables the testing of lack of fit – useful for assessing model adequacy.
- Three replications of the entire design by taster. Although the three subjects were chosen carefully on the basis of their good taste in beer, they differed in their generosity of rating; that is, tending to score every brew higher or lower. These individual biases were corrected via a statistical technique called "blocking."

The 14 blends per person (blocked) were provided in random run order for these three sensory responses:

Y1. Taste (without looking at the cocktail) on hedonic scale of 1 (worst) to 9 (best)
Y2. Appearance (1-9)
Y3. Overall liking (1-9)

To keep things simple for educational purposes, we will only look at the overall liking (Y3) and the response of cost, which is determined completely by the blend's composition and, of course, the current cost of each ingredient (already noted).

Figure 2-5 - Precisely mixing a beer cocktail behind the screen (to keep tasters blind)

Mark owns a very accurate kitchen scale (pictured in Figure 2-5) that he uses to weigh out green coffee beans for roasting (another story!) so it was convenient for him to set the total for each blend by weight rather than volume – to 60 grams (roughly two fluid ounces). That kept the total beer consumption per person to a reasonable level – about two bottles worth. (Mark admits that during the experiments he managed to drink about the same amount – in the name of science, naturally.) Each drinker kept his own beer-shot glass. The plan was for excess material, beyond what was required for measurement purposes, to be discarded, but Mark found this very difficult to enforce on the hot summer day of the experiment – done on his back porch.

EARLY THREE-COMPONENT BEER MIXTURES

According to popular bar lore, 18th century Londoners developed a liking for a beverage called "three threads" – made by blending a third of a pint each of ale, beer and "twopenny" (the strongest of these brews, costing two cents a quart). Another three-component recipe dictated stale (aged for up to two years), mild and country (pale) ale. It seems likely that pub keepers experimented on mixtures in the hopes of finding something both tasty and cheap (by diluting costly brews with less expensive ones).

"It is not entirely clear as to why a fashion for mixing beers arose…, other than a desire to match palate and pocket."

Source: "BEER BEFORE PORTER".

P.S. Many Americans say "cheers" to a binary blend of beers from the British Isles – thick, dark stout poured on a pale ale. This is commonly called a "Black and Tan." What makes this

---

combination interesting is that, with the right order of addition – first the ale and then the stout, and a steady hand on the pouring, the drink exhibits a distinct layering of black on tan. Evidently the stout is less dense than the ale, but the reason remains mysterious – a matter for more research, no doubt!

"Fear not the beer cocktail." -- *Stephen Beaumont (from his book* World of Beer*.)*

See Table 2-1 for the text matrix, laid out by blend type and location, and the overall liking ratings for the three tasters. The actual order of presentation was randomized, thus decoupling the cocktail type from possibly lurking variables such as degrading taste (related to admissions above), dehydration from exposure to the summertime elements, etc.

Be careful about drawing too many conclusions and extrapolating these very far. However, like all experiments, this one may produce some useful findings. Let's see what can be made of it.

| Blend #	Type	Location (Coded)	A: Wheat grams	B: Lager grams	C: Black grams	Liking 1-9	Cost $/12 oz
1	Pure	Vertex (1,0,0)	60	0	0	5,5,5	1.16
2	Pure	Vertex (0,1,0)	0	60	0	4,4,3	0.84
3	Pure	Vertex (0,0,1)	0	0	60	7,6,5	1.24
4a	Binary	Center Edge (0.5,0.5,0)	30	30	0	5,5,4	1.00
4b	"	"	30	30	0	5,4,5	"
5a	Binary	Center Edge (0.5,0.5,0)	30	0	30	8,7,6	1.20
5b	"	"	30	0	30	7,8,7	"
6a	Binary	Center Edge (0.5,0.5,0)	0	30	30	4,4,3	1.04
6b	"	"	0	30	30	4,4,2	"
7	Check	2/3rd,1/6th,1/6th Axial $(0.6\overline{6}, 0.1\overline{6}, 0.1\overline{6})$	40	10	10	6,7,5	1.12
8	Check	1/6th,2/3rd,1/6th Axial $(0.1\overline{6}, 0.6\overline{6}, 0.1\overline{6})$	10	40	10	5,5,4	0.96
9	Check	1/6th,1/6th,2/3rd Axial $(0.1\overline{6}, 0.1\overline{6}, 0.6\overline{6})$	10	10	40	7,7,6	1.16
10a	Ternary	Centroid	20	20	20	5,7,4	1.08
10b	"	"	20	20	20	6,6,4	"

Table 2-1: Results of beer-cocktail experiment (ratings by three tasters by blend)

Go ahead and look over the results – as Yogi Berra said "you can see a lot just by looking." For example, is it possible that some combinations of beers might be perceived as being unexpectedly tasty? Or, perhaps, the opposite may be true: Putting certain beers together may not be such a good idea. Keep in mind that this experiment represents only a sampling of possible reactions by these particular tasters, who may or may not represent a particular segment of the

beer-drinking market. Does it appear as if any of the three tasters may have been tougher than the others (hint!)? If so, do not worry, so long as this individual remains consistent with the others in his or her relative rankings by blend, then this consistent bias can be easily (and appropriately) blocked out mathematically, thus eliminating this easily-anticipated source of variation (person-to-person).

Diving under the response surface to detail the underlying predictive model

The results are very interesting. As you can see by the location of the peak region in the 3D response plot (Figure 2-5), a blend of Blue Moon wheat ale (A) and Sam Adams Black Lager (C) really hit the spot for overall liking! Taste and appearance ratings also favored this binary blend. The tasters all preferred this combination, which Mark deemed the "Black and Blue Moon" beer cocktail.
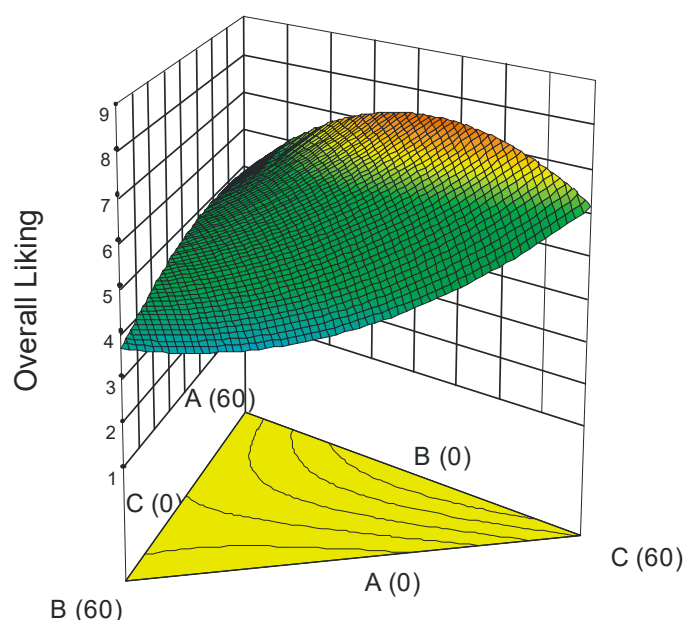
Figure 2-5: Response surface shows peak taste with synergistic blend of two beers

The fitted model in coded units (0 to 1) that produced this surface is:

Overall Liking = 4.92A + 3.68B + 6.13C + 2.01AB + 7.35AC – 4.65BC

Notice from the coefficients on the pure-component terms (conveniently labeled alphabetically – A, B & C, rather than mathematically – x_1, x_2, x_3) that these beer aficionados liked the Sam Adams Black Lager (C) best and Budweiser (B) the least: 6.13C > 4.92A > 3.68B. From the second-order non-linear blending terms one can see the synergism of beer A (wheat) with beer C (black) by the large positive coefficient: 7.35AC. Do not get too excited by the very high coefficient: Remember that this value must be multiplied by one-fourth to calculate the 'kicker' for the binary blend. Nevertheless, this product provides nearly a two-point gain on the hedonic scale: 7.35 * 1/4[th] = 1.84.

On the other hand, these tasters' buds were antagonized by combining the Budweiser with the Sam Adams Black Lager as evidenced by the negative coefficient (–4.65) on model term BC.

You can see this downturn in the response surface along the BC edge. It is less of a deviation from linear blending than is observed for the AC binary blend (BC < AC).

That leaves one coefficient to be interpreted – that of term AB. It turns out that the p-value for the statistical test on this coefficient (2.01) exceeds 0.1, that is, there is more than a ten percent risk that it could truly be zero. (In contrast, the coefficients for terms AC and BC were both significant at the 0.01 level.) This time around we did not bother to exclude the insignificant term (AB) from the model. Removing it would make little difference in the response surface – just a straight edge between the wheat beer (A) and lager (B), rather than a slightly upwards curve. We will revisit the issue of model reduction later. As the number of components increase and modeling gets more complex, it will become cumbersome to retain insignificant terms.

WOULDN'T IT BE EASIER TO MODEL MIXTURES IN ACTUAL UNITS?

Statistical software that can fit formulation results to the Scheffé mixture models may offer these to users in either coded or actual form, or both. In this case the actual equation is:

> *Overall Liking   = 0.082031 Wheat +0.061396 Lager +0.10214 Black*
> *+ 0.000559187 Wheat*Lager + 0.00204067 Wheat*Black – 0.00129267 Lager*Black*

The one advantage you get from this format is being able to plug in the actual blend weights and chug out the predicted response for overall liking. This gets more intense as the order of terms goes up due to the exponential impact on coefficients – they get really small or very large, depending on whether your actual inputs are greater than one (as in this case) or less than one (for example if you were serving beer to ants – they would be happy with very tiny amounts).

Now, look back at the coded equation we provided in the main text and consider how easy it is to interpret. For example, one can see immediately what the predicted sensory result will be for each of the pure components (A, B and C) – these are the coefficients – no calculating required.

So, here's the bottom line: For interpretation purposes, always use the coded equation as your predictive model.

PS. In case you were wondering (?), neither the coded nor the actual equation features a coefficient for the block effect. These models are intended for predicting how an elite beer drinker will react to these three types and their blends. True, some of these individuals will feel compelled to be snobby and look down on all beers, but this cannot be anticipated by the formulator, nor controlled once a product goes up for sale. Thus the block effect provides no value for predicting future behavior – only to explain what happened during the experiment.

As discussed in Chapter 1, statisticians like to keep models as parsimonious as possible, so let's see if we could get by with only the linear model. Figure 2-6 provides an enlightening view of the BC edge (two-component) after the least-squares fit without the non-linear blending terms (AB, AC, BC).
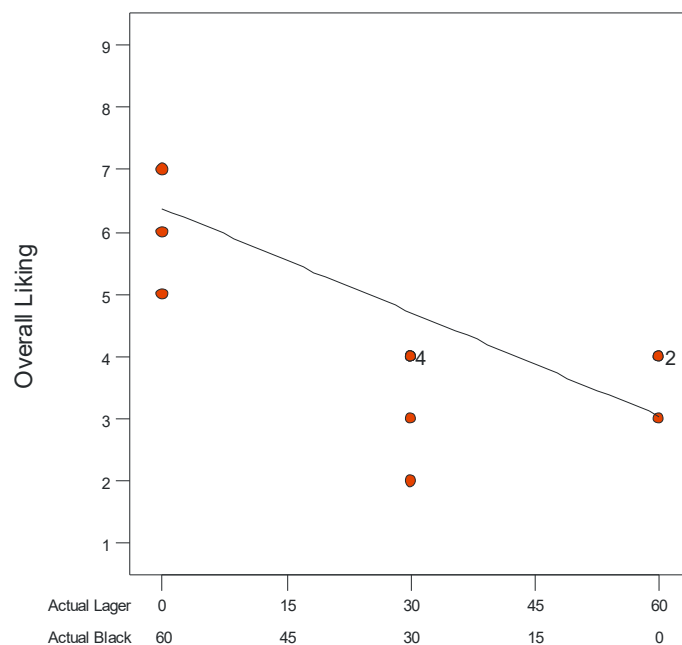
Figure 2-6: View of BC (Lager-Black) edge after misfit with linear model

Notice that all 6 of the actual results (4 of which are at the same point) fall below the predicted value from an over-simplified linear blending model. The cumulative impact of these deviations far exceeds the value expected on the basis of the pure error pooled from the four replicate blends tasted by each tester, thus this linear model exhibits a significant ($p<0.1$) lack of fit. Clearly the surface needs to dip down at the binary blend of B (lager beer) and C (black), which it does when fit with the quadratic model (look back at Figure 2.5). Not surprisingly, this second-order model for nonlinear blending does not exhibit a significant lack of fit ($p>0.3$), that is, it fits!

Taking cost into account

Given the expense per 12 ounce serving of each of the three beers as detailed at the outset of this case, it was a simple matter mathematically to compute the blended costs shown in the last column of Table 2-1. In fact, the software we steer you to for doing the practice problems will do this calculation for you very easily. If one wanted to reduce expense, mixing in a cheap amber-lager like Budweiser (component B) would help as you can see in the response surface on cost in Figure 2-7.
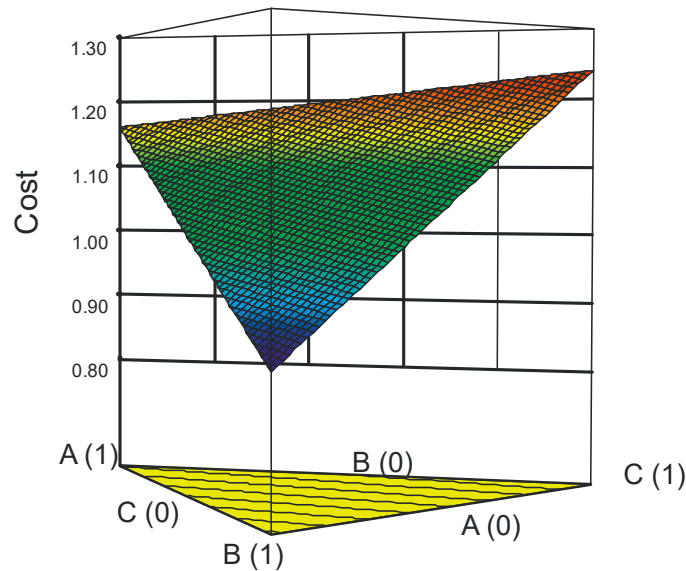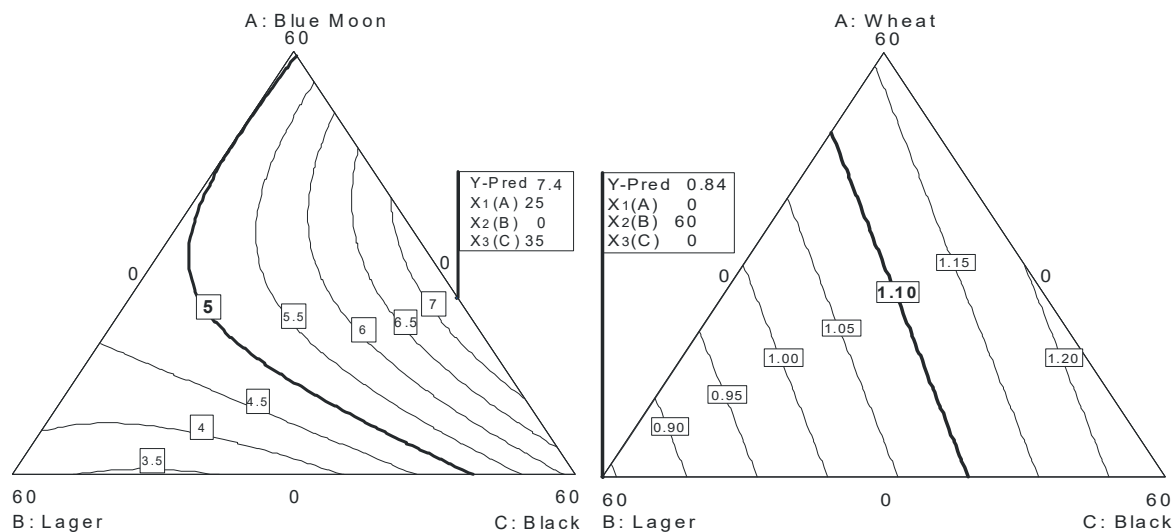
Figure 2-7: Cost as a function of the beer cocktail composition

However, in this case you get what you pay for – the cheaper the blend the less likable it becomes. You can see this in the side by side comparison of contour plots for overall liking versus cost in Figures 2-8a and 2-8b. To make it even more obvious, we flagged the optimums of maximum liking versus minimum cost.
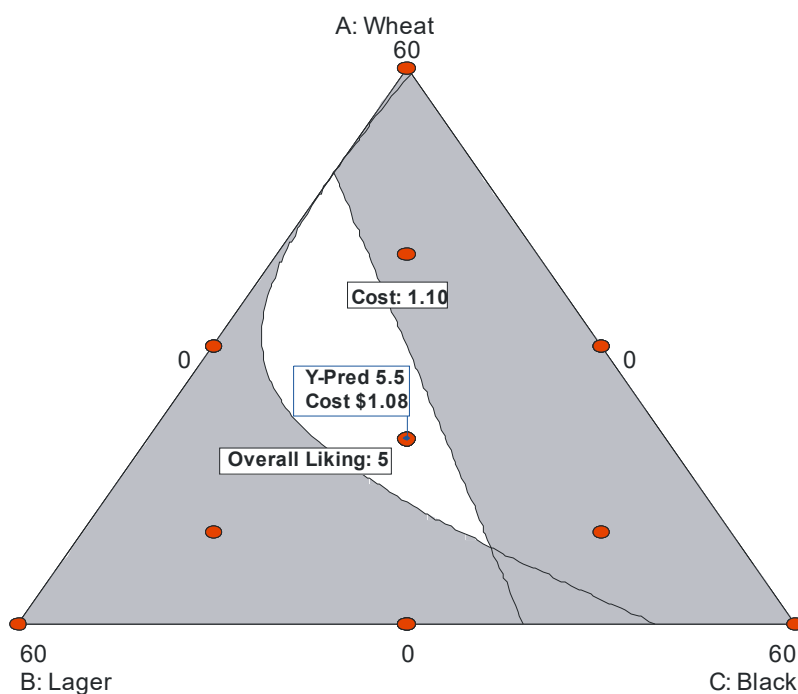

Figures 2-8a and 2-8b – Contour plots for overall liking (left) and cost of blended beers (right)

SEARCHING OUT THE OPTIMUM NUMERICALLY

Figuring out which blend of beer is cheapest is a 'no brainer' – even a six-pack drinker could figure out it's the Budweiser American lager, pure and simple. That's because the cost is a simple linear function. (Also, it's deterministic, that is, not derived empirically.) However, it would take some calculus to find the optimum of the second-order model for overall liking – perhaps a bit beyond the average beer drinker. As we will discuss later, things get a lot more

complicated when searching out the most desirable blends on the basis of multiple responses. It turns out that for computational purposes a hill-climbing algorithm generally works best for this purpose. That's how we came to the flagged optimum of 25 parts of Blue Moon to 35 parts of Black Lager for the (theoretically) most likable Black & Blue Moon cocktail. As a practical matter, Mark blends these two beers half and half (give or take) in two identical glass mugs for the drinking pleasure of him and his better half (the wife).

Now suppose as a beer-tender you'd be satisfied to serve a cocktail with an overall liking at or above 5, midway on the hedonic scale of 1 ☹ to 9 ☺. We highlighted this contour in Figure 2-8a. Furthermore, assume that you need to hold the cost to $1.10 per 12 ounce serving – the highlighted contour on Figure 2-9b. Years ago before the onslaught of presentation programs like Microsoft Powerpoint, statisticians would transfer contour graphs to individual transparencies, shade out the undesirable regions and overlay all the transparencies (also known as "view foils") on an overhead projector. Ideally that left a window of opportunity, or "sweet spot," like that shown in Figure 2-9 – produced directly by modern computer software.


Figure 2-9 – Contour plots for overall liking and cost overlaid

The flag marks the centroid blend with equal measures of all three beers, which falls inside the window of overall likability at 5 or higher and cost less than $1.10 per serving. Mark cannot build up much enthusiasm for this – too much work and not first-class. He is an elitist when it comes to the finer things in life, such as a cold beer on a hot evening sitting on the back porch after a long, hard day at the office. Thus, the binary blend of the "Black and Blue Moon" cocktail gets his nod.

Do not put a square peg into a triangular hole

We hope that by now you see the merit of mapping mixtures to the triangular space, unless the experiment involves only two components, in which case a simple number line suffices, as illustrated in Chapter 1. Although you are convinced, it may be hard to convert your fellow formulators who remain 'square' by sticking to the factorial space used by process developers. Here's a postscript to the beer cocktail case that may help you turn the tide to the triangle.

At this stage the "Black and Blue Moon" cocktail has achieved a foothold in the drink recipes for those who enjoy a relaxing beverage now and then. However, some dispute remains on the precise proportions of the Black Lager versus the Blue wheat ale. An experimenter who has mastered two-level factorials, but remains ignorant of mixture design, creates the experiment laid out in Figure 2-10.
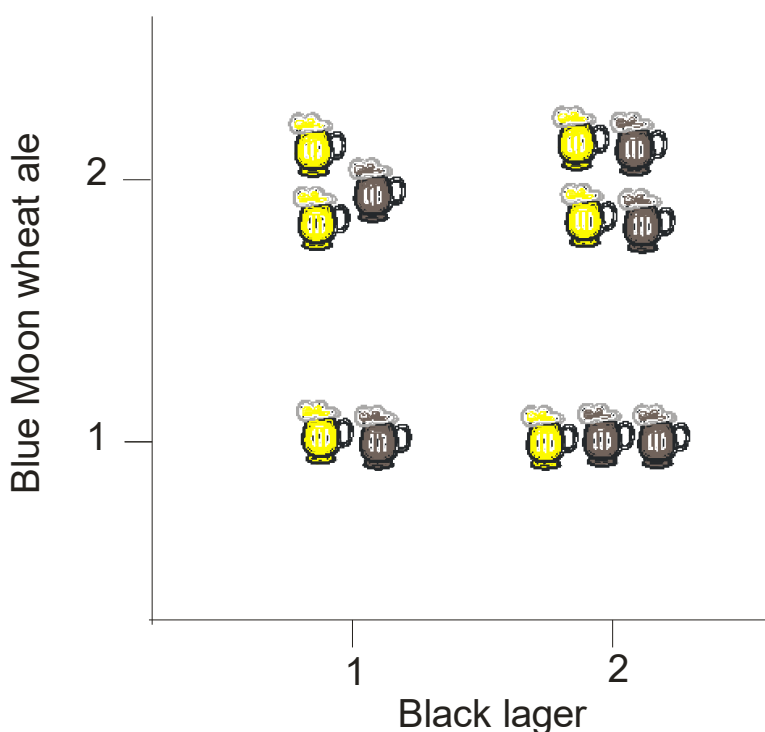
Figure 2-10: Factorial design on formulation of Black and Blue Moon beer cocktail

If this graph makes you thirsty, then take a brief break from your reading to pour a little something. OK, now concentrate! What's wrong with this picture given that taste is a function of proportions and not amounts of a beverage? The problem occurs along the diagonal running from lower left to upper right in this square experimental region. One of each of the beers versus two of each makes no difference once blended – a sip tastes the same. All that's happening in this direction is a scale up of the recipe. That would be a waste of good beer!

WHEN YOU KNOW THAT THE BEER DRINKING HAS GOTTEN OUT OF HAND

Mark and Pat are big fans of baseball and their hometown club – the Minnesota Twins. Mark shares season tickets with a relative, but he did not hold seats for a critical year-end game, thus

he was forced to purchase a spot in the outfield bleachers where the really rabid boosters congregate. As he worked his way down the row, Mark observed that the portly fellow in the adjoining seat had lined up 9 beers under foot – one for each inning, thus avoiding the need to fight the lines at the concession stand. To avoid such abuse of alcohol, the policy of the Twins is two beers per purchase, so this guy must have come quite early to stock up so! Such behavior goes beyond the pale of good taste in our opinion. We urge you to be more moderate if you try to replicate our beer-blending experiments.

Practice Problems

To practice using the statistical techniques you learned in Chapter 2, work through the following problems.

Problem 2-1

Normally, we do not recommend the simplex centroid design, especially if done 'by the book,' that is – without check blends. However, it can be useful for three components that cannot conveniently be broken down too far into fractions. A case in point is the "Teany Beany Experiment" we detailed in *DOE Simplified, 2nd Edition* (Chapter 9). According to a randomized plan, a number of tasters were each required to taste small jelly candies encased in an ablate spheroid of hard sugar. Actually this is not as difficult as it sounds from this overly detailed description of the confection, at least not for a single bean. However, the tastings get increasingly tricky as one tries blends of two beans and finally a three-bean mixture. As you can see from Table 2-2, we decided to draw the line at this point, rather than adding check blends necessitating the cutting of beans into smaller segments.

ID #	Blend	A: Apple %	B: Cinnamon %	C: Lemon %	Taste Rating
1	Pure	100	0	0	5.1
2	"	"	"	"	5.2
3	Pure	0	100	0	6.5
4	"	"	"	"	7.0
5	Pure	0	0	100	4.0
6	"	"	"	"	4.5
7	Binary	50	50	0	6.9
8	Binary	50	0	50	2.8
9	Binary	0	50	50	3.5
10	Centroid	33.3…	33.3…	33.3…	4.2
11	"	"	"	"	4.3

Table 2-2: Teeny-beany mixture design – a three-component simplex centroid

This table translates the relative proportions to a percent scale. Beware: software that supports mixture design may require you to make each row add to the specified total, in this case one hundred. If so, blend number 10 will fail unless you 'plug' it by entering 33.34 for one of the three components.

Note from the response column that each of the three pure-component blends was replicated, as well as the overall centroid (the three 'beaner'). The results come from the averaging of a number of tasters, done for simplicity sake. Analyze this data. What hits the sweet spot for this candy-loving crowd?

MORE SOPHISTICATED APPROACHES FOR TASTING TINY JELLY BEANS

To keep things simple, we provide only the average taste rating for this case study on candy. However, a more precise analysis blocked out the rating shifts caused by personal preferences – some scoring high across the board (people with a 'sweet tooth') and others low. We found a remarkable consistency – all but one of the tasters liked cinnamon best and lemon least. However, one individual loved lemon far and away. We excluded her from the average. But do not fear, she remains free to fill her candy jar with these yellow beans.

"I'm not an outlier; I just haven't found my distribution yet." -- Ronan M Conroy

---

Problem 2-2

This is a case where materials can be freely blended, thus the formulators could augment the simplex centroid with check blends. The experimenters measured the effects of three solvents known to dissolve a particular family of complex organic chemicals (Del Vecchio, 1997). They had previously discovered a new compound in this family. It needed to be dissolved for purification purposes, so they needed to find the optimal blend of solvents.

Table 2-3 shows the experimental design and results. Remember that the actual run order for experiments like this should always be randomized to counteract any time-related effects due to aging of material, etc. Also, we recommend that you always replicate at least four blends to get a measure of error. In this case, it would have been helpful to do two each of the pure materials and also replicate the centroid.

ID#	A MEK	B Toluene	C Hexane	Blend Type	Solubility (g/l)
1	100	0	0	Pure A	121
2	0	100	0	Pure B	164
3	0	0	100	Pure C	179
4	50	50	0	Binary AB	140
5	0	50	50	Binary BC	180
6	50	0	50	Binary AC	185
7	33.3	33.3	33.4	Centroid	199
8	66.6	16.7	16.7	Check	175
9	16.7	66.6	16.7	Check	186
10	16.7	16.7	66.6	Check	201

*Table 2-3. Design Matrix and Data for Solvent Study*

Notice that for the sake of formulating convenience, the interior points (centroid and check blends) were rounded to the nearest tenth of a percent so that they always added to one hundred. Which chemical, or blend of two or three, will work best as a solvent? (Hint: Read the Appendix before finalizing your answer.) For extra credit on this problem, determine the relative costs of each chemical (MEK is methyl ethyl ketone) and work the material expense into your choice.

Chapter 2 Appendix: The Special Cubic (& Advice on Interpreting Coefficients)

In the Appendix to Chapter 1 we detailed a very sophisticated third-degree mixture model called the "full cubic." As we illustrated, this equation captures wavy response behavior that would be unusual to see in a typical formulation experiment. As a practical matter (remember the principle of parsimony) it often suffices to apply a simpler "special cubic" model, such as this one for three components:

$$\hat{y} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3$$

As you may've already surmised, this modeling option is not applicable for two components. For three components, the special form saves three model terms over the full cubic and it's easier to interpret. Figure 2-11 provides an example of three components of a solid rocket fuel blended for a peak in the measured response – elasticity (Cornell & Piepel, 2008, p.55).
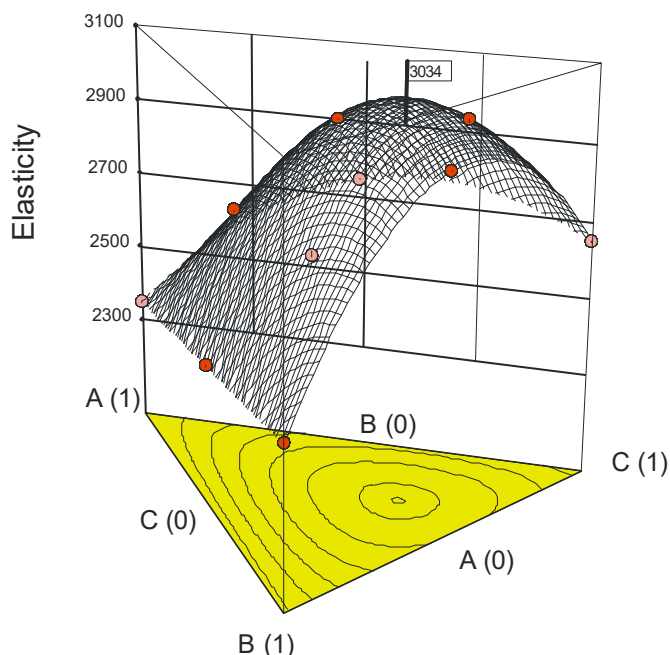
Figure 2-11 – Response surface from special cubic model

Notice the dramatic upward curve along the BC edge – a clear case of synergism between these two materials (oxidizer and fuel). However, a further increase occurs with the addition of component A (binder) – a move from the center of the BC edge to the flagged peak in the interior of the triangular mixture space. This represents a beneficial non-linear blending involving three components.

The fitted special cubic model (coded) that generated this response surface is:

$$\text{Elasticity} = 2351A + 2446B + 2653C - 6AB + 1008AC + 1597BC + 6141ABC$$

The coefficient on ABC seems surprisingly large unless you remember that the components are on a scale of 0 to 1. For example, recall from the last chapter that the maximum deflection from linear blending occurs at the 1/2-1/2 ("50/50") point, thus you must multiply the two-component terms, such as BC, by one-fourth to assess the synergism (or antagonism). However, for the three-component term the maximum non-linear effect occurs at the $1/3^{rd}$-$1/3^{rd}$-$1/3^{rd}$ point (centroid) with a magnitude of $1/27^{th}$ the coefficient of that term. Thus in this case the maximum effect for BC of approximately 400 ($1/4^{th}$ of 1597) almost doubles the greatest impact of ABC ($1/27^{th}$ of 6141).

How do higher order terms compare to the linear ones in this case (2351, 2446 and 2653)? Here again you must be careful not to jump to conclusions without first considering the meaning of linear coefficients in Scheffé polynomial mixture models – the difference is what counts, not the absolute magnitude. The range of linear coefficients is only a bit over 300 (2653 for C minus 2351 for A), so the tilt in the plane of response (upwards to component C) is actually exceeded by the synergism of B and C.

Have you had enough of trying to interpret coefficients in these higher order mixture models? We hope so, because it's really not worth belaboring – simply look at the response surfaces to get a feel for things. Then with the aid of computer tools, use the model to numerically pinpoint the most desirable recipe for your formulation.

A FURTHER WRINKLE: THE SPECIAL QUARTIC

Another variation on mixture models is the "special quartic" – shown here for three components (Smith, 2005):

$$\hat{y} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{1123} x_1^2 x_2 x_3 + \beta_{1223} x_1 x_2^2 x_3 + \beta_{1233} x_1 x_2 x_3^2$$

This model provides additional terms that captures more complex non-linear blending than the special cubic. However, be forewarned that the number of unique blends in the mixture design must always exceed the count of terms in the model you want to fit. As spelled out in Table 2-4, the special quartic model requires considerably more blends at four or more components, which may make the experiment unaffordable.

Components (q)	Linear	Quadratic	Special Cubic	Cubic	Special Quartic
2	2	3	NA	4	NA
3	3	6	7	10	9
4	4	10	14	20	22
5	5	15	25	35	45

*Table 2-4 – Number of terms in full versus special Scheffé polynomials*

Chapter 3: Simplex Lattice Design

"My beloved is like a gazelle or a young stag. Behold, he is standing behind our wall,
He is looking through the windows, He is peering through the lattice."
— Song of Solomon 2:9

In this chapter we will detail the design of lattices within a simplex geometry. The procedure will work for any number of components, but we will illustrate simplex lattice design only up to four ingredients, which form a three-dimensional tetrahedron. That will provide enough of a challenge for now.

Working with four components in tetrahedral space

To get a feel for a tetrahedron, make one out of a piece of paper (preferably ruled for graphs) according to the following procedure (Box & Draper, 2007, page 512). Start by drawing an equilateral triangle as best you can (perfectionists see the sidebar). Then using the midpoints of the edges as vertices, inscribe a second equilateral triangle as pictured in Figure 3-1a.
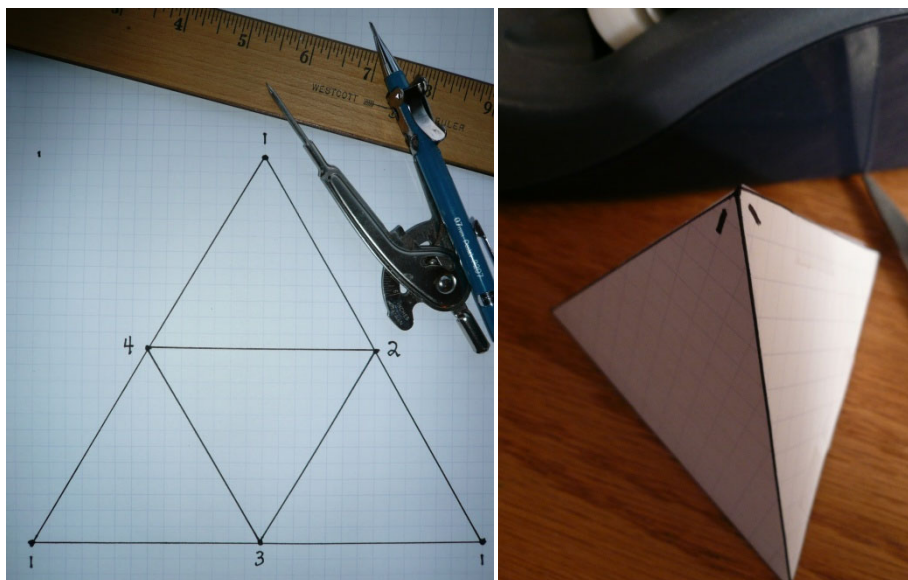
Figures 3-1a and 3-1b – Building a tetrahedron

Label the vertices of the larger triangle "1" – these represent the first ingredient. Identify the corners of the smaller triangle as 2, 3 and 4 to pinpoint three more components (thus allowing for four, in all). Now cut out the large triangle and fold the 1's along lines 4-2, 2-3 and 3-4 to a point as shown in Figure 3-1b. There – you've made a tetrahedron! Keep this handy to help you visualize our illustrations of four-component mixture design.

> *HOW TO DRAW AN EQUILATERAL TRIANGLE*
>
> *It's easy to draw a triangle, but surprising hard to make it equilateral, that is, with sides of precisely equal length. The trick is to start with the proper tool – a compass – and knowing how to use it for drawing arcs and setting fixed distances. We could detail this here but it will make a lot more sense for you to see it demonstrated by The Math Open Reference Project at the internet web page for "Constructing an Equilateral Triangle" at*
> *www.mathopenref.com/constequilateral.html.*

Building a simplex lattice design

Henry Scheffé's pioneering publication "Experiments with Mixtures" (noted in a Chapter 1 sidebar) introduced simplex lattice design to industrial statisticians and their client formulators. The name of this mixture design provides two big clues about its construction:

- A <u>simplex</u> is a geometric figure having a number of vertexes or corners equal to one more than the number of dimensions of the variable space for n dimensions. For example, when n equals two, the simplex has three corners – an equilateral triangle.
- A <u>lattice</u> is an arrangement in space of isolated points in a regular pattern, such as atoms in a crystalline solid.

The simplex lattice design is comprised of m+1 equally spaced values from 0 to 1, thus

x_i = 0, 1/m, 2/m, ..., 1

where x is the component in coded levels and m represents the degree of polynomial that the formulator feels will be needed to fit the experimental response. The number of blends in a simplex lattice design depends on both the number of components (q) and the degree of the polynomial. The simplest (overly so!) simplex lattice is this one designed for two components at a degree of one (m=1):

x_i = 0, 1/m, ~~2/m, ...~~

x_i = 0, 1/1

x_i = 0, 1

There are only two points, 0 and 1! Going beyond 1 is not allowed, so the design must stop there. It's designated as "(2, 1)" on the basis of the number of components and degree; respectively. We do not recommend this (2, 1) simplex lattice as-is – too few points for any appreciable power needed for model-fitting. However, it serves as a launching pad to designs for three components that are not that bad – for example, the two pictured in Figures 3-2a and 3-2b for second and third degree modeling.
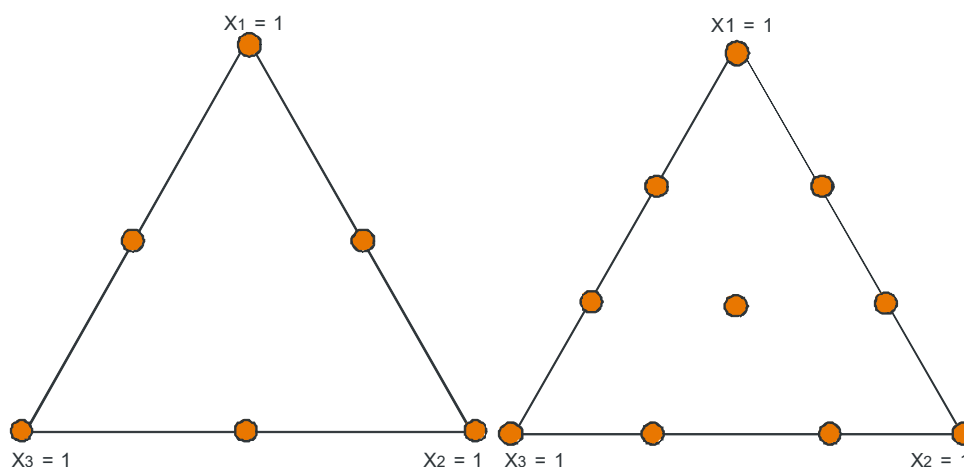
Figure 3-2a & 3-2b: Three-component simplex lattices of 2nd (left) and 3rd (right) degree

If you are a formulator, it will seem odd to do an experiment on several components but never a complete blend, which is exactly what happens with the (3, 2) design depicted in Figure 3-2a – its interior remains devoid of points. However, keep in mind that you need not adhere to this textbook template – strongly consider adding the centroid and, if not impractical, additional check blends inside the simplex formulation space. This will be demonstrated by example later in this chapter.

Two more simplex lattice designs are shown in Figures 3-3a and 3-3b. Notice by their geometry – tetrahedral – that these encompass four components. Also evident at-a-glance is the increase in degree of the lattice from left to right of two to three; respectively.
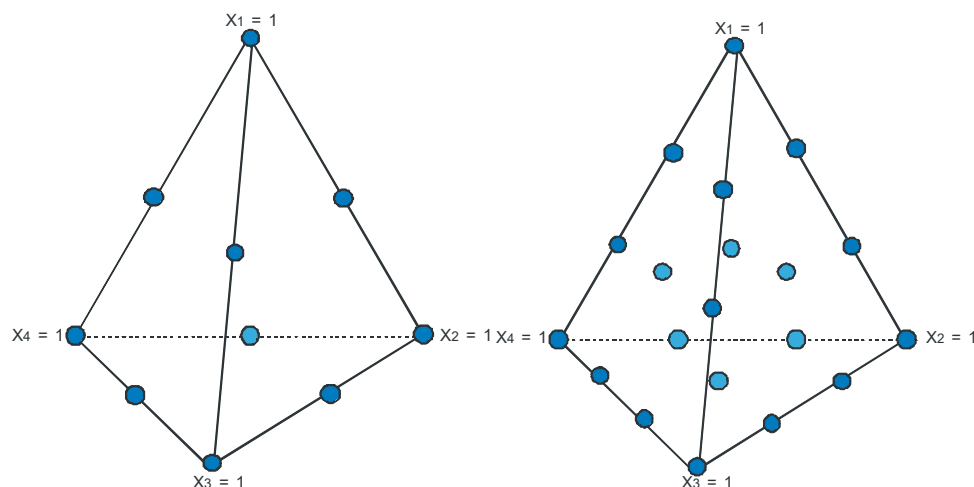

Figure 3-3a & 3-3b: Four-component simplex lattices of 2nd (left) and 3rd (right) degree

An easy way to infer the degree is by the number of design points along the edges; when broken in half the degree is two – whereas a fragmentation by thirds indicates a third degree lattice.

AN EXCITING (!) FORMULA FOR THE NUMBER OF DESIGN POINTS IN A SIMPLEX LATTICE

The number of design points N in a simplex lattice depends only on the number of components q and the degree of polynomial m. The equation is:

$$N = (q + m - 1)! / (m! (q - 1)!)$$

You may recall from math and/or stats class that the exclamation marks denote a factorial notation. This will come back to you more quickly by following this example calculation on the number of design points for a four-component simplex lattice designed to the third degree – a (q, m) of (4,3), which computes as:

$$= (4+3-1)!/(3!(4-1)!) = (6!)/(3!3!) = (6 \times 5 \times 4 \times \cancel{3 \times 2 \times 1})/[(3 \times 2 \times 1)(\cancel{3 \times 2 \times 1})]$$

Notice that the multiplication of 3x2x1 appears both in the numerator (top) and denominator (bottom) of this equation. Thus these numbers cancel and the equation simplifies to:

$$N = (6 \times 5 \times 4)/(3 \times 2 \times 1) = (\cancel{6} \times 5 \times 4)/\cancel{6} = 5 \times 4 = 20 \text{ points}$$

Check this calculation by counting the number of points in Figure 3-3b, comprised of 4 pure blends at the four vertices, 12 two-part blends along the six edges and 4 three-part blends at the centers of the four faces. That comes to 20. Yes?

P.S. For purposes of building simplex lattice designs, the special model forms are considered to be full degree. So, for example, the degree (m) for the special cubic is 3 – the same as it would be for the full cubic Scheffé polynomial.

Augmenting raw simplex lattices to make them more practical for formulators

As we've shown, the simplex lattice design may not provide a complete mixture – in fact this glaring deficiency occurs whenever the number of components exceeds the degree (q > m). Thus the first step for a practical augmentation will be to add the overall centroid – a complete blend with equal amounts (1/q) of each component. However, there's a less obvious drawback to running the raw simplex lattice as spelled out by textbook – they often use up all the degrees of freedom for model-fitting, with none left over to provide check blends for a lack of fit test. For example, notice in Table 3-1 how the number of points in the second-degree (m=2) simplex-lattice designs equals the number of terms in the quadratic mixture model (second-order Scheffé polynomial). These are said to be "saturated" designs. If the formulator can be certain of the degree, a saturated experiment would be most efficient. But that level of knowledge generally comes after all the research, not before. Otherwise why bother experimenting?

Components (q)	Simplex Centroid	Simplex Lattice (m=2)	Quadratic Model Terms
3	7	6	6
4	15	10	10
5	31	15	15
6	63	21	21

Table 3-1: Number of points in textbook (raw) simplex designs versus what's required by model

In case you may be tempted to fall back on the less sophisticated simplex centroid, which does guarantee inclusion of a complete blend (by definition), notice how the number of points increase exponentially – going far beyond the minimum required for the quadratic mixture model. Thus, we advise the simplex centroid only for very special cases for three components only, such as the taste-testing experiment presented in Problem 2-1.

Instead we recommend an augmented simplex-lattice (ASL) that incorporates these added points:
- 1 centroid (if needed)
- q axial check blends
- At least 3 replicates.

We detailed axial check blends via the beer blending case study presented in Chapter 2, so you're aware of how these points fill the gaps between the centroid and each of the q components. This design also featured replicates and how they provide measures of pure error, which, in conjunction with check blends, facilitate testing for lack of fit. Now we are establishing this augmentation as the standard procedure for simplex designs – centroid or lattice.

HOW MANY POINTS TO REPLICATE AND WHICH ONES

We're trying to be sensible in advising how many points need to be replicated. Having none provides zero degrees of freedom (df) for estimating pure error, so adding one replicate is infinitely better from a mathematical perspective (one divided by zero). However, when computing the lack-of-fit test, a very large ratio of variance from the benchmark pure error would be required. For example, going back to the gold-copper mixture experiment in Chapter 1, here are the critical values of this ratio at a p-value of 0.05 for df's of 1,2,3,4 and 5, respectively: 199.5, 19.0, 9.5, 6.9 and 5.8.* Notice how quickly this value falls but then it starts leveling off after 3 df. By 5 df the critical F becomes stable. Thus we recommend that five points be replicated in any mixture design.

Now which points will be best to replicate? We suggest picking points that exhibit the greatest leverage. You may be familiar with this statistical term: It's very simply a measure of potential influence for any given point. Consider again, for example, the gold-copper blending experiment. If the goldsmith only checked the melt point on one sample each of pure gold and pure copper, these two results would be totally influential on a linear model, that is, both design points exhibit a leverage of one. However, replicating these points – increasing the experimental runs from two to four – reduces the leverage to 0.5 for each.

Computing leverage becomes more complicated for higher-order models,** but to give you some feel for how it works, here are the measures for the gold-copper experiment without any

replicates: 0.89 at the vertices (pure metal), 0.49 at the centroid ("50-50" gold-copper) and 0.37 for the axial check blends (1/4th-3/4th and 3/4th-1/4th). The three highest leverage points are the two vertices and the centroid, so these were the ones chosen for replication.

* (From 5% table provided in Appendix 1-3 of DOE Simplified *with 2 df in numerator of the F-ratio – starting from top of the second column and looking down from 1 df in the denominator – the first row.)*

**(For more details on leverage, see the sidebar titled "Assessing the Potential Influence of Input Data via Statistical Leverage" provided in Chapter 1 of RSM Simplified*.)*

Second-order designs augmented according to the guidelines we've provided are considered to be suitable for producing a "simplex response-surface" (Smith, p. 49). This leads to an important insight: Mixture design for optimal formulation is a close cousin to response surface methods (RSM) for process optimization.

Using ASL mixture design to optimize an extra virgin formulation of olive oils

Olive oil, an important commodity of the Mediterranean region and a main ingredient of their world-renowned diet (see sidebar), must meet stringent European guidelines to achieve the coveted status of "extra virgin." Oils made from single cultivars (a particular cultivated variety of the olive tree) will at times fall into the lower "virgin" category due to seasonal variation. Then it becomes advantageous to blend in one or more superior oils based on a mixture design for optimal formulation. For example, a team of formulators (Vojnovic, D., et al., 1995) experimented on four Croatian olive oils – Buza, Bianchera (pronounced "be an kay ra"), Leccino (pronounced "la chee no") and Karbonaca – to achieve an overall sensory rating of at least 6.5 on a 9-point hedonic scale, thus easily exceeding the cut off for "extra virgin" (5.5 considered to be "virgin"). The ratings were done by ten assessors trained on fundamental tastes (sweet, salt, sour and bitter) and defects of virgin olive oils, such as rancidity.

Figure 3-4: Olive oil (source: royalty-free post at www.sxc.hu/photo/670878)

We've adapted the original study a bit to simplify it and make it more educational, while capturing the essence of how these formulators made use of mixture design and what they discovered as a result.

The assessors can discern very tiny differences in sensory attributes that may depend on subtle non-linear blending of two or more oils. Therefore the four component (q = 4) simplex lattice is

set up to the third degree (m = 3). That produces 20 unique blends as shown in Figure 3b. To augment this lattice, the formulators add 4 axial check blends and the overall centroid. They then specify that the four vertices (chosen for their high leverage) and centroid be replicated (for added pure error measure) at random intervals. (Always randomize!) Assume that the formulators use a 1 liter blender to mix the oils – 30 blends in total after the augmentation. This ASL (augmented simplex lattice) design and the end results for overall sensory ratings are shown in Table 3-2. (Note that, for the sake of simplicity, the one-third and two-third levels are rounded to 0.333 and 0.667; respectively – thus adding to the total of 1.)

#	Point Type	A: Buza	B: Bianchera	C: Leccino	D: Karbonaca	Sensory Rating
1	Vertex	1	0	0	0	6.98
2	"	1	0	0	0	6.84
3	Vertex	0	1	0	0	6.49
4	"	0	1	0	0	6.45
5	Vertex	0	0	1	0	7.25
6	"	0	0	1	0	7.30
7	Vertex	0	0	0	1	5.88
8	"	0	0	0	1	5.95
9	Third Edge	0.667	0.333	0	0	7.38
10	Third Edge	0.333	0.667	0	0	7.12
11	Third Edge	0.667	0	0.333	0	6.87
12	Third Edge	0	0.667	0.333	0	6.84
13	Third Edge	0.333	0	0.667	0	6.95
14	Third Edge	0	0.333	0.667	0	7.17
15	Third Edge	0.667	0	0	0.333	7.36
16	Third Edge	0	0.667	0	0.333	7.14
17	Third Edge	0	0	0.667	0.333	7.5
18	Third Edge	0.333	0	0	0.667	7.16
19	Third Edge	0	0.333	0	0.667	6.95
20	Third Edge	0	0	0.333	0.667	7.00
21	Triple Blend	0.333	0.333	0.333	0	7.56
22	Triple Blend	0.333	0.333	0	0.333	7.53
23	Triple Blend	0.333	0	0.333	0.333	7.29
24	Triple Blend	0	0.333	0.333	0.333	7.28
25	Axial CB	0.625	0.125	0.125	0.125	7.41
26	Axial CB	0.125	0.625	0.125	0.125	7.37
27	Axial CB	0.125	0.125	0.625	0.125	7.50
28	Axial CB	0.125	0.125	0.125	0.625	7.19
29	Centroid	0.25	0.25	0.25	0.25	7.58
30	"	0.25	0.25	0.25	0.25	7.55

Table 3-2: ASL design for blending four olive oils and their sensory results

The statistical analysis of this data will be detailed via Problem 3-2. The chosen model is a reduced special cubic:

Sensory Rating = 6.91 A + 6.47 B + 7.29 C + 5.93 D
+ 2.51 AB – 0.91 AC + 3.70 AD + 0.54 BC + 3.75 BD + 2.78 CD + 11.65 ABC

The presence of the ABC non-linear blending term supports the choice of a third-degree lattice design. The other three special-cubic terms (ABD, ACD, BCD) were insignificant (p>0.1) so we chose to remove them from the final model. Rather than laboriously dissecting the model by its remaining terms, let's focus on the response surface graphics: The pictures will tell the story.

Unfortunately, now that we've gone to the third dimension the imaging gets trickier – only three out of the four components can be depicted on a contour plot, for example. This complication provides the perfect opportunity to present the "trace" plot – a way to view the relative effects of any number of components. A trace plot for the olive-oil mixture experiment is shown in Figure 3-5.
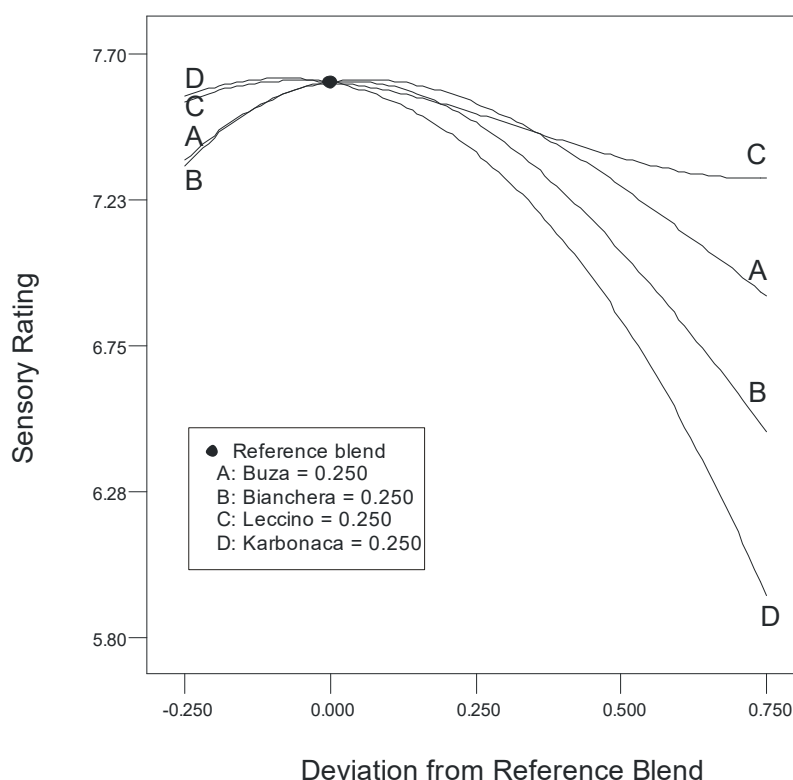

Figure 3-5: Trace plot for olive-oil mixture experiment

The traces are drawn from the overall centroid – all components at equal volume within the 1 liter vessel. This is called the "reference blend." Each component alone is then mathematically varied while holding all others in constant proportion. This reveals, for example, that the predicted sensory evaluation falls off dramatically as the Karbonaca oil (D) is increased relative to the three alternatives.

3-8

To give you a better feel for how the trace plot is produced, consider the simpler case of only three components. Figure 3-6 shows the paths of the three traces.
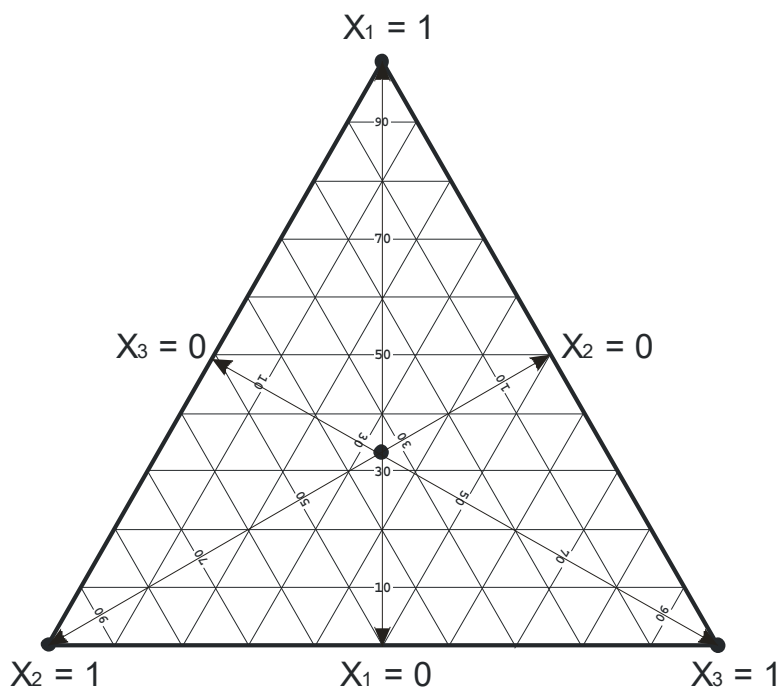

Figure 3-6: Traces for three components only

The trace for x_1 starts at the overall centroid where it amounts to one-third of the three-component blend. The other two components are also at one-third, thus their ratio is one-to-one. Tracing x_1 from the centroid down to the base of the ternary diagram reduces the amount of this individual component to zero. At this point the amounts of the other two components become one half each – thus their ratio remains one-to-one. In fact if you pick any point along any of the three traces, the other two components remain at constant proportions! Try working this out for yourself – it will be good practice for reading off coordinates on the ternary diagram.

TRACE VERSUS PERTURBATION

Those of you who are knowledgeable about response surface methods (RSM) for process optimization may be familiar with the "perturbation" plot – the RSM equivalent of the mixture trace. This plot predicts what will happen if you perturb your process by changing only one factor at a time, for example by first varying time and then temperature of a chemical reaction. The perturbation plot generally emanates from the center point – all process factors at their middle level. It looks the same as a trace but the difference is that, given a fixed total on amount, no single component can change without one or more of the others taking up the slack. Creating a trace as we've detailed is a work-around that provides the same benefits as a perturbation plot for RSM, that is, graphically depicting the relative effects of individual components as they become more or less concentrated in your formulation. However, keep in mind the one-dimensional nature of the trace, which cannot substitute for contour plots or 3D views of the surface as a function of any two components. Only then will you see an accurate picture of non-linear blending effects.

Now that we've been provided with clues on the non-linear blending behavior of the four olive oils, it seems sensible to study the response surfaces of the three good components 'sliced' at varying levels of the inferior fourth component. For example, Figures 3-7a and 3-7b show the sensory results at the overall centroid (all components, including D, at 0.25 concentration) versus no Karbonaca oil (D = 0). If anything, it's the Bianchera (B) oil that fails to excite the taste buds – at least from these slices with D held fixed at two specific levels (0 and 0.25).
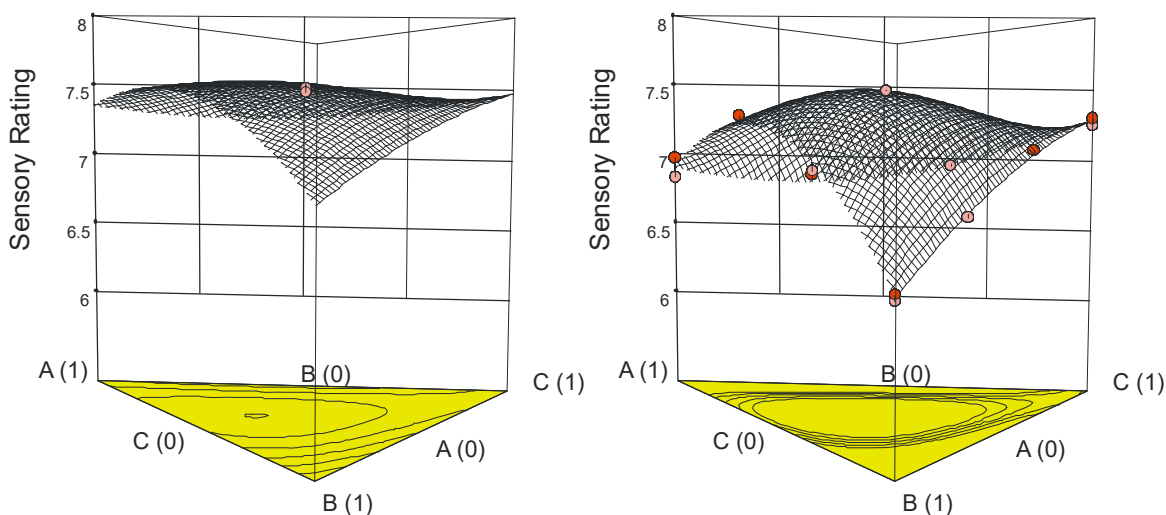
Figures 3-7a and 3-7b: Sensory results at the overall centroid versus no Karbonaca oil (D = 0)

These response surface graphs are very illuminating! It appears as if the complete four-part blend at the centroid, shown on the left (Figure 3-7a), will be most robust to variations in olive oil concentrations and deliver a superior sensory rating for the most part. A more comprehensive computer-aided search of the entire tetrahedral formulation space produced that optimal blend depicted in Figure 3-8:

- A. 0.333 Buza
- B. 0.299 Bianchera
- C. 0.189 Leccino
- D. 0.179 Karbonaca

This is predicted to produce a sensory rating of 7.63 – higher than any of the actual test results. However, any such prediction must be subject to verification.
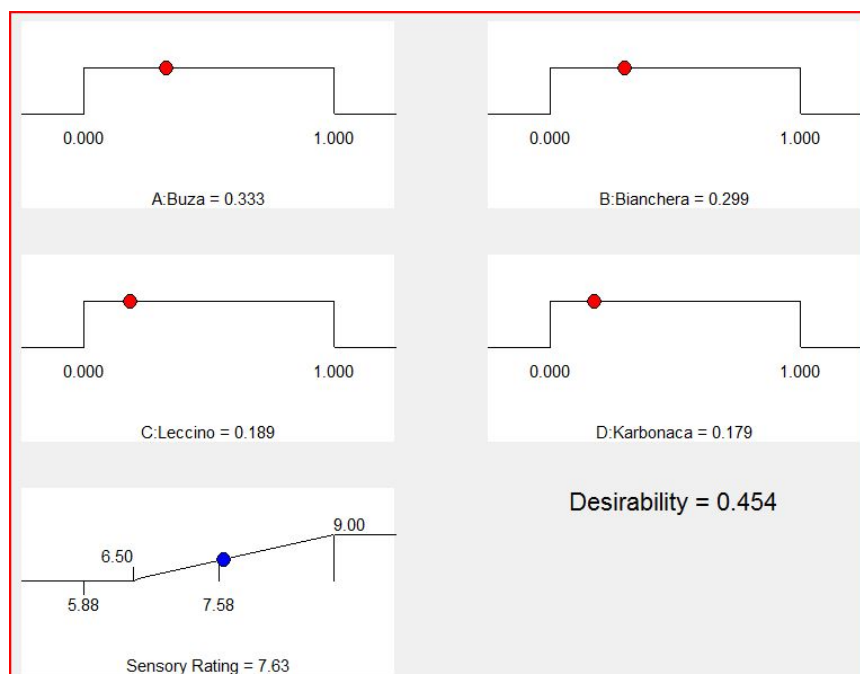
Figure 3-8: Most desirable blend via computer-aided numerical search

The upward ramp for the sensory response shows how a rating below 6.5 will be completely undesirable, whereas a rating of 9 represents the peak of desirability (prima!). The predicted optimum falls 0.454 up the scale from zero to one on desirability. This may be the best the blenders can do with these particular varietals. A sensory outcome of 9 on the hedonic scale remains in the province of the gods who enjoy only the best of the best.

Figure 3-9 flags the attainable optimal blend of these four olive oils. It displays the prediction interval (7.51 to 7.76) based on 95 percent confidence. Ideally the verification blend will be rated by the sensory panel within this range. A result outside of the prediction interval would cast doubt on the validity of the model.
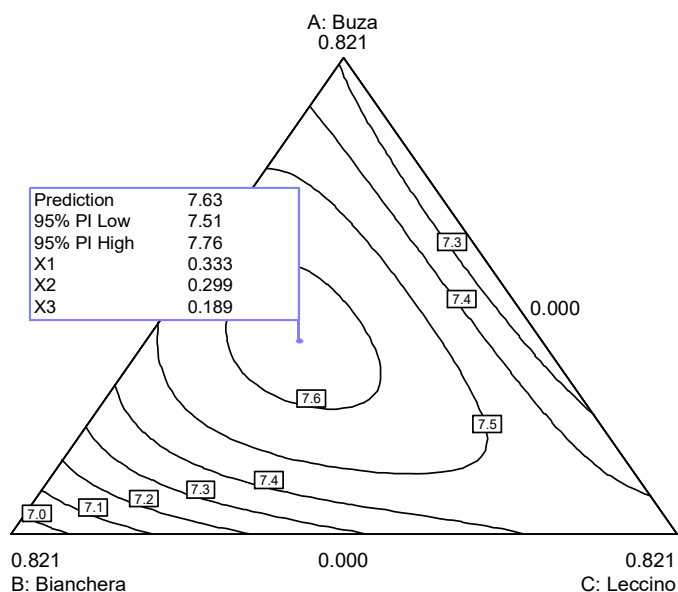
Figure 3-9: Most desirable blend flagged on contour plot (D sliced at 0.179)

Remember that these plots are derived from the Scheffé-polynomial predictive-model fitted to the actual experimental data and validated statistically. However, only by producing a confirmatory blend and subjecting it to sensory evaluation will this be verified for all practical purposes.

The optimum sensory comes from a blend of the three better olive oils A, B & C. But what if the inferior oil D can be bought very cheaply? Then a blend with it being maximized could be made at a sensory just good enough to pass the panel at Extra Virgin level. This might be worth a try!

MEDITERRANEAN DIET

According to the various sources the key components of the Mediterranean diet include:
- *Olive oil as an important source of monounsaturated fat*
- *Generous amount of fruits and vegetables*
- *Red wine in moderation*
- *Fish on a regular basis*
- *Very little red meat*

Many benefits have been attributed to this diet, including reduced rate of coronary events and weigh loss.

"Two themes characterize people who have lost a significant amount of weight and kept it off long-term: 1) they don't eat as much as in the past, and 2) they exercise more. Look for these when you search for effective weight-loss programs."

- Dr. Steve Parker, author of The Advanced Mediterranean Diet *(Vanguard Press, 2008).*

By employing mixture design these sophisticated olive oil formulators obtained "a knowledge of the whole experimental dominion with the advantage to be able to find various mixtures bearing the same qualities." Furthermore, "in this way, in spite of the presence of external limits such as olive oil availability or other economic aspects, a variety of optimal blendings can be selected according to market preference."

Practice Problem

To practice using the statistical techniques you learned in Chapter 3, work through the following problem.

Problem 3-1

To drive out any fears of being overwhelmed by the methodology, see how easy it will be to design a mixture experiment and analyze the results using a computer tool specialized for this purpose – Design-Expert® or Stat-Ease® 360. Download a free fully functional trial via the Software link at www.statease.com. Then complete the first tutorial on mixture design found under the program Help. Via a case study on a detergent, this exercise provides practical aspects on how to experiment on a portion of the entire formulation by setting constraints on each of the individual components.

If you need a little help with mixture design, feel free to contact the staff of statistical consultants at Stat-Ease by emailing stathelp@statease.com. Please mention that your interest stems from reading *A Primer on Mixture Design*.