stat teaser

ABOUT STAT-EASE SOFTWARE, TRAINING, AND CONSULTING FOR DOE
Phone 612.378.9449 Fax 612.378.2152 E-mail info@statease.com Web Site www.statease.com

University of MN Duluth Dogs

An irresistable tale of a DOE frenzy. By Bill Pederson, a professor at the University of Minnesota, Duluth campus. Enjoy!

A group of young soon-to-be engineers and a number of their unsuspecting friends at the University of Minnesota Duluth (UMD) Department of Mechanical and Industrial Engineering recently solved a question that has plagued man since the discovery of propane, "Which is better for cooking, gas or charcoal?" During the annual welcome-back barbeque for our students and faculty, the student section of the Institute of Industrial Engineers and the American Society of Mechanical Engineers collaborated to finally put this discussion to rest… well, we gave it a shot anyhow. The goal of the experiment was to not only determine which tasted better, gas- or charcoal-cooked food, but also to determine the best type of hot dog amongst beef, pork or turkey. The color of napkin was also included in the study as an additional variable for instructional purposes.

Before going any further, I need to warn you. Don't try this at home. There was significant danger to the individuals performing this experiment. If you were to stand between a couple dozen hungry college students and a pile of free hot dogs, you would understand. I don't recommend it. All kidding aside, conducting this test properly presented significant challenges in order to not bias the results and still allow production to proceed at an acceptable pace.

Because our students had not had the opportunity to take the "Experiment Design Made Easy" course, there were several people in our group who did not believe it was possible to compare taste results using a subjective measurement system. Additionally, there was concern regarding how we could compare the results from one person to another since personal taste preferences vary.

Eager UMD Hot Dog Testers

Let me address the concerns about a subjective measurement system first. On a different project, I was working with a

—*Continued on page 2.*

—*Continued from page 1.*

small injection molding company that was trying to rapidly bring a new product to market. (No one has ever tried to do that before!) Unfortunately, the product was geometrically complex, was made of a polymer new to the company, and was being made on a new machine. The company had no experience with DOE and typically relied on their very capable technician to pull them through this type of situation. Unfortunately, there were simply too many variables for that to work out. I could talk about this project for awhile, but let me just say that DOE was the right tool. It resulted in a nearly optimized system in an afternoon of work, while the trial and error method had produced no significant results in nearly two weeks of testing.

One of the issues with the part was the amount of flash on the product (or squeeze-out of polymer between different parts of the die set). The company had no means of measuring flash, so we came up with a subjective rating from one to ten. I prefer that my graphs point up to the right as an indication of a good result (something my stock portfolio rarely does), so I assigned one to be the worst flash and ten to be the best. Having designed the experiment and given explicit test performance instructions, I left to work on a different project that was burning hot. Later that day, I received an e-mail with the resulting flash ratings. I excitedly started to analyze them, only to be quickly disappointed with the results. The analysis in Design-Expert showed no significant effects on flash from a couple of factors I expected to have a physical effect on the flash of the product.

After a series of questions about test conduct, it became apparent what had happened. The technician had performed the test exactly as I had asked, but not as I had intended. He had diligently evaluated flash on a one to ten scale as the parts had come off the press and immediately recorded them. The problem was that his measurement system changed over the course of making 64 parts! What he had considered to be a five at the beginning had become an eight by the end. Fortunately, he had labeled all the parts appropriately by sequence number, so they could be re-evaluated using a better methodology. The new system used ten bins for parts. We sorted ALL the parts AT ONCE according to flash and then assigned a score. Analysis then revealed a greatly reduced level of variation in the measurement system which allowed several other significant factors and interactions to reveal themselves. In effect, we reduced the noise in the system so that smaller effects could be determined.

Back to our hot dog experiment. A taste measurement system is a destructive measurement system (I supposed re-using the parts is an option, although not a good one!). Therefore, it is important to minimize the amount of time between rating different parts so that the measurement system does not change over the course of evaluating all parts. With only six hot dog bites to rate, we determined this to be an acceptable risk for our experiment. Unfortunately, it also means that these results can never numerically be compared with future experiments, but this is also acceptable for this case.

Concerns regarding the personal preference of taste were addressed by using a randomized block design in which all hot dog scientists would evaluate all factor combinations, each in their own random order. Hot dog bites were evaluated rather than full hot dogs—so that we could afford the test. There wasn't a problem with college students happily eating six hot dogs, but we had to draw the line somewhere! By blocking on the evaluator, we were able to include multiple opinions about tastes of hot dogs,

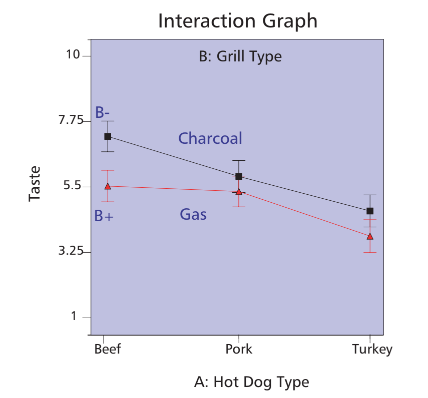

Figure 1: Interaction Graph

while still getting meaningful data between what amounts to different measurement devices.

The factors studied were: A. Hot Dog Type—Beef, Pork, or Turkey, B. Grill—Charcoal or Gas, and C. Napkin Color—White, Red, Yellow, or Blue. Factor C was chosen to see whether or not the color of napkin used could have biased the results in one way or another. Because even college students might complain about tasting 24 combinations of hot dogs, we opted to randomly assign napkin colors within a block of test conditions. While this resulted in a small amount of correlation between our regression coefficients, there were enough samples from 24 people as to make this not a concern as verified in the design evaluation section of Design-Expert software. Correlation was minimized by randomly choosing a color for each sample. This greatly simplified the actual means of conducting the test as well.

When tasting hot dogs, there are obviously many more factors that could be tested such as ketchup, mustard, relish, etc. While these might be interesting, it was decided that the primary goal was to evaluate the type of hot dog and grill, and allow these factors to all be included in the analysis error. The only stipulation to minimize the error was that

—*Continued on page 3.*

Three Factors Interacting vs. the 3FI Term

Most of you have likely heard the DOE adage, "three-factor interactions are highly unlikely to be real." But perhaps it's time to explain why this is true, and under what circumstances a three-factor interaction term (3FI) could exist.

Starting with some basic definitions, a main effect in 3D is the slope of a flat plane in that factor direction. A two-factor interaction (2FI) adds the amount of twist in that plane.

When there are two 2FI's that have a letter in common, such as AC and AD, then these two terms model the behavior of the interacting factors. (See Figure 1.) A 3FI term, such as ACD, implies that the behavior of two factors is opposite (inverted) at the extremes of the third factor. (See Figure 2.)

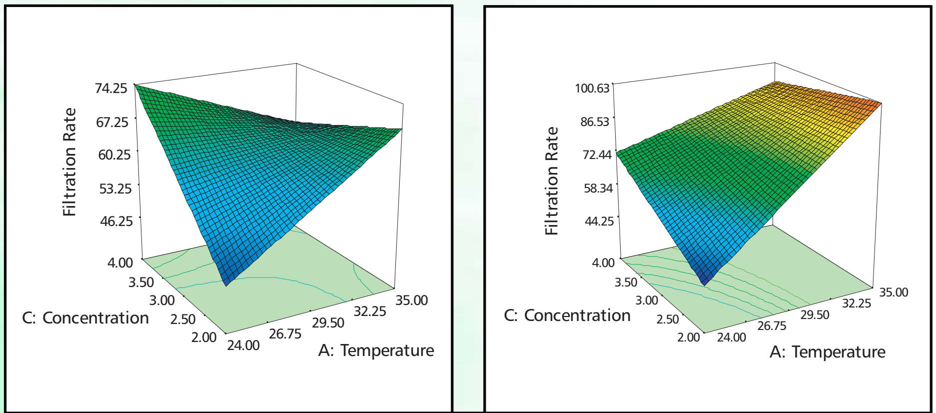

Figure 1: AC interaction at low D (left) and high D (right)

The presence of 2FI's with a letter in common means that three factors are dependent on each other. The shape of the AC surface depends on the level of D and, likewise, the shape of the AD surface depends on the level of C. So, the three factors do interact. This is not the same as saying that there is a significant 3FI. If ACD is significant, then changing D would produce the exact opposite behavior in AC. The surface flip-flops in shape.

When are 3FI's more likely? They are more likely when one or more of the factors is categoric in nature. In this case, you might be better off to split your data into separate sets, one for each level of the categoric factor, and model them separately. Three (or more) factors interacting is not uncommon, but this behavior is usually adequately modeled with 2FI's with a factor (or factors) in common. Significant 3FI's are more often a sign of problems with the data than they are legitimate model terms.

By Pat Whitcomb, pat@statease.com

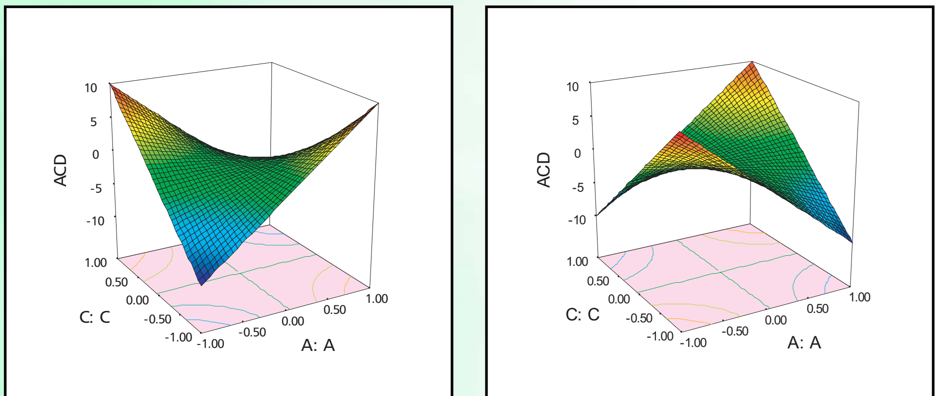

Figure 2: ACD interaction at low D (left) and high D (right)

—Continued from page 2.

whatever condiments were used on one sample should be used on all samples for that person, and in the same quantities. Interestingly many opted for none, in the name of science, I guess. It is important to remember that an experiment should be made as simple as possible, but no simpler.

As shown in Figure 1, we determined that beef hot dogs cooked on a charcoal grill had the best overall flavor rating. Napkin color did not have a significant effect (as hoped, although engineers are obviously well known for their sense of style). The model was significant with a Prob>F of <0.0001 despite the variation from different personal taste preferences.

Of course, like all experiments, our results are only valid for the conditions we studied. There are already plans for continued experimentation to narrow down some other factors not included in this study. Some of our scientists have proposed more elaborate means of ensuring product consistency. At any rate, we have made an important step for mankind. Go dogs!

(Thanks to Bill Pederson for this "DOE in Action" story! Comments can be sent to shari@statease.com and will be forwarded to Bill.)

New Network Pricing & DX 6.0.10 Update

12/04

New Network Pricing

Beginning February 1st, 2005, Stat-Ease, Inc. will institute a new pricing scheme for network purchases. Instead of selling perpetual licenses as we currently do, Stat-Ease will now be offering annual licenses for all new network purchases. The advantages to you, the customer, are multiple.

1. Annual licenses require a smaller initial investment, in effect spreading the cost of the software out over several years.
2. Annual licenses include free updates, free upgrades, and free technical support as long as your license is current.
3. Annual licenses are easy to budget for and administer, thus making them the preferred choice for many corporations.

For those customers who prefer not to be on an annual license plan, single-user licenses will still be sold on a perpetual basis and include free maintenance fixes and technical support.

Below you will find pricing for Design-Expert annual licenses purchased on or after February 1st next year.

3-seat = $1050.00/year
5-seat = $1625.00/year
10-seat = $3000.00/year
15-seat = $4200.00/year
20-seat = $5600.00/year
25-seat = $6500.00/year

Please contact Stat-Ease for pricing information on higher quantities.

Download a FREE DX6.0.10 Update!

If you haven't already, update your individually licensed Design-Expert 6.0.x to 6.0.10 at **http://www.statease.com**. Browse to Software, Downloads and follow the instructions.

Companies with Network Licenses can obtain a free update CD by calling 1.612.378.9449 and providing their product serial number.

Address Service Requested

Stat-Ease, Inc., Hennepin Square
Suite 480, 2021 E. Hennepin Ave.
Minneapolis, MN 55413-2726

Stat-Ease statistics made easy