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

Workshop Schedule

DOE Simplified
October 4: MN ASQ Conf., Minneapolis, MN

An overview of Design of Experiments (DOE) from A to Z, based on the popular book. $295* ($195 each, 3 or more)

Statistics for Technical Professionals
October 5–6: Minneapolis, MN

Revitalize the statistical skills you need to stay competitive. $995* ($795 each, 3 or more)

Experiment Design Made Easy
June 8–10: San Jose, CA SOLD OUT!
July 13–15: Minneapolis, MN
August 17–19: Philadelphia, PA
September 21–23: Minneapolis, MN

Study the practical aspects of DOE. Learn about simple, but powerful, two-level factorial designs. $1495* ($1195 each, 3 or more)

Response Surface Methods for Process Optimization
June 22–24: Minneapolis, MN
October 12–14: Minneapolis, MN

Maximize profitability by discovering optimal process settings. $1495* ($1195 each, 3 or more)

Mixture Design for Optimal Formulations
August 3–5: Minneapolis, MN
November 9–11: Minneapolis, MN

Find the ideal recipes for your mixtures with high-powered statistical tools. $1495* ($1195 each, 3 or more)

Robust Design: DOE Tools for Reducing Variability
September 14–16: Minneapolis, MN

Use DOE to create products and processes robust to varying conditions. A must for Six Sigma. *Factorial and RSM proficiency are required.* $1495* ($1195 each, 3 or more)

PreDOE: Basic Statistics for Experimenters

Six-hour web-based training. This course or the equivalent is a prerequisite for all workshops—**www.statease.net**. $95

Attendance is limited to 20. Contact Sherry at **800.801.7191 x18** or **sherry@statease.com**.

**Includes a $95 student materials charge which is subject to state and local taxes.*

Stat-Ease
statistics made easy™

Katie's Coke® vs Pepsi® DOE

This past year my youngest daughter Katie challenged me to a blind taste-test of Coca-Cola® versus Pepsi-Cola® soft drinks. Her 7th grade social studies teacher had done something similar in her class that day to illustrate how advertising influences consumer perceptions. I was tired after a day of teaching DOE and figured a good dose of caffeine might help me stay awake to moonlight on a new book titled, "RSM Simplified" (due out 2004). Katie helped me experiment on spring toys for a general factorial design featured in the first volume, "DOE Simplified" (Productivity, 2000). She also participated in a similar DOE on flying disks that we published in one of my prior columns for this newsletter (September 2002, Stat-Teaser). Therefore, Katie fancies herself quite the experimenter and wanted to dig into the business herself.

While Katie prepared in our kitchen for the cola challenge, I settled at my home office (really just a broken-down old roll-top desk) and sat staring at my computer screen. I chuckled to myself when Katie paraded in with two glasses —one made of white foam and the other of blue plastic. She explained that, although I and other taste testers in the household should not be told which cola was which, she needed to keep them straight. Before we get to the "lessons learned" part, let's give her credit for replicating the test, rather than just doing it once on me!

"Katie," I said, "Do you realize that you've made the fundamental error of confounding my treatment of Coke versus Pepsi with my preference for white versus blue and Styrofoam versus plastic?" She gave me this look. I've seen it before from my other daughters and sons and even more powerfully from my wife. It's not good, but it was worth taking a hit to defend the sanctity of unaliased experimental designs. If one cola is always in white styrofoam and the other cola is always in blue plastic, there is no way to tell if my preference for one over the other is due to the type of soft drink or the type of cup—these factors are aliased. For instance, perhaps in general people prefer drinking pop in a plastic cup rather than a Styrofoam cup. Then whichever cola is in the plastic cup would get the most favorable reviews.

In the end, both Katie and I came out of this experience with positive thoughts. My daughter got her revenge later that evening when she confronted me with three cups of cola all in foam containers this time. "OK Dad," she said with a gloating look, "Taste these and tell me if

—Continued on page 2.

FAQ - Interpreting Lack of Fit

A statistically significant lack of fit (LOF) value often worries experimenters. This article is meant to give experimenters a better understanding of this statistic and what could cause it to be significant. The formula is:

$$\text{Lack of Fit F - test} = \frac{\text{Lack of fit MS}}{\text{Pure Error MS}}$$

where MS = Mean Square. The numerator in this equation is the variation between the actual values and the values predicted from the model. The denominator is the variation within any replicates. The variation between the replicates is meant to be an estimate of the normal process variation. Significant lack of fit means that the variation of the replicates about their mean values is less than the variation of the design points about their predicted values. Either the runs replicate well and their variance is small, or the model doesn't predict well, or some combination of the two.

Figure 1 is an illustration of a data set that had a statistically significant model, but also had a statistically significant lack of fit. Believe it or not, there are actually six center points on the graph. They are so close together that they overlap and some are hidden below the response plane. Now look at the other model points. Many are positioned off the response plane. Although the predicted response surface fits the model points well (providing the significant model fit), the differences between the actual data points and the response plane are **greater** than the differences between the center points. This is what triggers the significant LOF statistic. The center points are fitting better than the model points. Does this significant LOF require us to declare the model unusable?

When there is significant lack of fit, check how the replicates were run—were they independent process conditions run from scratch, or were they simply replicated measurements on a single setup of that condition? Replicates that come from independent setups of the process are likely to contain more of the natural process variation. Look at the response measurements from the replicates and ask yourself if this amount of variation is similar to what you would normally expect from the process. If the "replicates" were actually run more like repeated measurements, it is likely that the pure error has been underestimated (making the LOF denominator artificially small). In this case, the lack of fit statistic is no longer a valid test and decisions about using the model will have to be made based on other statistical criteria.

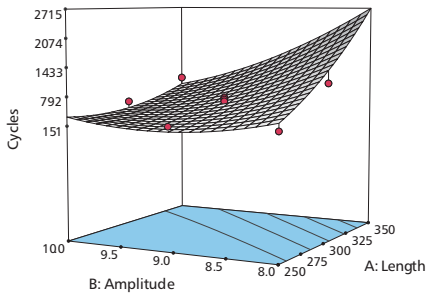


Figure 1—The centerpoints have less variation than the model points, causing significant lack of fit.

If the replicates have been run correctly, then the significant LOF indicates that perhaps the model is not fitting all the design points well. Consider transformations (check the Box Cox diagnostic plot). Check for outliers. It may be that a higher-order model would fit the data better. In that case, the design probably needs to be augmented with more runs to estimate the additional terms. This may be the case when, for instance, the Fit Summary screen shows that the linear and quadratic models have significant lack of fit, but the aliased cubic model has insignificant lack of fit.

If nothing can be done to improve the fit of the model, it may be necessary to use the model as is and then rely on confirmation runs to validate the experimental results. In this case, be alert to the possibility that the model may not be a very good predictor of the process under certain conditions. *(To learn more about lack of fit and basic DOE, sign up for Stat-Ease's Experiment Design Made Easy workshop.)*

Shari@StatEase.com

Continued from page 1

there's any difference." I knew it must be a trick, but I tossed all three caffeinated beverages down the hatch. Since I'd already drank so much of the stuff, this new round of drinks all tasted the same—sickly sweet and bubbly. I said as much to Katie, "You are trying to fool me by putting the same brand in all the glasses." "Ha!!! I did fool you Dad—one glass is Pepsi, the other is Coke and the third is a mixture of the two!" "You win," I said and preserved the self-esteem of my daughter and avoided the marital discord that would've resulted otherwise by Dad being a statistical bully around the home.

Why was I happy? Because Katie's poor DOE on Coke versus Pepsi provided an ideal example for illustrating the concept of aliasing, especially for people who don't normally perform experiments. I presented this case at a gathering of non-manufacturing trainees for Six Sigma at the Ohio State University's Fisher School of Management and they drank it up (pun intended).

Mark@StatEase.com

Optimize Mediums with Mixture DOE

The following study by the cell culture R&D group at Sigma-Aldrich (St. Louis, MO) is an example of a unique application of design of experiments (DOE) to optimize a biological process. We've described the case study here in general terms for those of us who are not biologists. For technical details, please refer to the article, "An Efficient Approach to Cell Culture Medium Optimization—a Statistical Method to Medium Mixing," on Stat-Ease's web site at **www.statease.com/pubs/cellculture.pdf**.

Chinese Hamster Ovary (CHO) cells are used extensively by researchers in industry and academia for recombinant protein production. Grown in media containing all of their nutritional requirements, these cells vary greatly in the specific amounts of amino acids, vitamins, salts, etc. they need for maximum cell growth and recombinant protein production. This presents a challenge for researchers because each particular CHO cell line is unique.

As more and more recombinant CHO clones have been developed, it has become increasingly important to streamline the medium optimization process. Sigma-Aldrich has taken this process to a new level with their CHO Medium Optimization Kit 1. This kit allows the researcher to quickly screen six different media formulations (which represent the wide range of nutritional requirements found in CHO clones) for cell growth and recombinant protein production. If optimization is desirable, it then leads researchers to the development of an optimized medium through mixture design and analysis.

For this case study, Sigma-Aldrich researchers started their creation of a better medium by screening the six primary media in their kit against the target CHO cell line. Using Design-Expert® (DX) software, they then took the top three of these six and conducted a mixing assay via a simple mixture design. Almost all of the new formulations were found to improve cell growth and produce high levels of recombinant protein production.

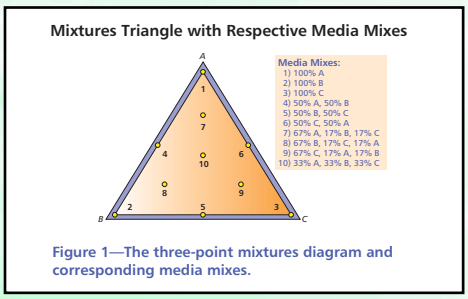


Figure 1—The three-point mixtures diagram and corresponding media mixes.

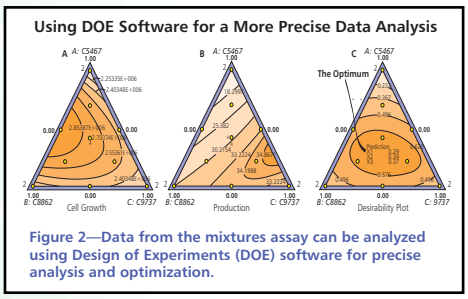


Figure 2—Data from the mixtures assay can be analyzed using Design of Experiments (DOE) software for precise analysis and optimization.

Assay data can be analyzed by one of two methods: 1. Separately analyzing the data for each response [cell growth (Fig. 2, left) and recombinant protein production (Fig. 2, center)] and ranking it in order, or 2. Doing an in-depth analysis using Design-Expert software (which allows the optimal media mixture for a particular CHO clone to rapidly be found). Using Design-Expert software, Sigma-Aldrich discovered the particular combination of the three media that was the most desirable (Fig. 2, right). The result was a medium that provided their CHO cell line with the optimal nutritional requirements for maximum cell growth *and* recombinant protein production. Tests of the recommended media mixture verified the results from Design-Expert software.

Design of experiments is tailor-made for optimization applications such as this. Traditional one-factor-at-a-time analysis can be useful, but it is unlikely that the best combination of media would be found in this manner. DOE creates predictive models that can interpolate to all of the combinations of media that have not yet been tested. Based on this interpolation, the best formulation can be predicted, and then validated via a confirmation run. With DOE, Sigma-Aldrich can develop an optimized medium for any given recombinant CHO clone.

For further information, contact Sigma-Aldrich Biotechnology, 3050 Spruce St., St. Louis, MO 63103, USA.

(Figures courtesy of Sigma-Aldrich Biotechnology.)

Where can you find us?

June 9-10 — Quality Expo, Detroit, MI, Booth 339

June 23-25 — MedEdge 2004, Minneapolis, MN, Booth 710

July 26-29 — Society for Industrial Microbiology (SIM) 2004, Anaheim, CA, Talks by Mark Anderson:

"Statistical Design of Experiments (DOE) for Making Breakthroughs" and "Response Surface Methods (RSM) for Optimization of Products and Processes"

August 8-11 — Joint Statistical Meetings, Toronto, CANADA, Booth 405

Roundtable Discussion by Shari Kraber:

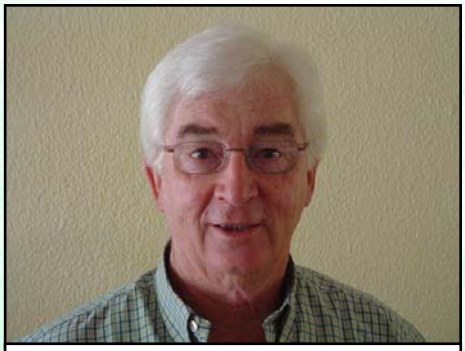
"Design of Experiments Trials and Tribulations"

October 4-5 — MN ASQ Conference, Minneapolis, MN

October 14-15 — 48th Annual Fall Technical Conference, Roanoke, VA

Spotlight on South Africa...

05/04



Nico Laubscher, InduStat Pro

Stat-Ease is fortunate to be represented by many knowledgeable and experienced distributors around the world. In this article we are pleased to introduce to you our representative in South Africa, Nico Laubscher, D. Sc., of InduStat Pro.

Nico started his industrial statistics consulting business, InduStat Pro, after retiring from his position as Company Statistician at SANS Fibres (a manufacturer of synthetic fibres) in 1996. His main expertise is in the areas of data mining, experimental design, and statistical process control.

Nico was first introduced to Design-Expert (DX) software in 1999 while attending a two-day seminar on experimental design by Dr. Douglas Montgomery (author of the book, Design & Analysis of Experiments, 5th Edition (J. Wiley & Sons, 2001)). So impressed was he with the software and recommendation from Dr. Montgomery, Nico approached Stat-Ease with a proposal to be our distributor in South Africa.

Since then Nico has successfully marketed Stat-Ease software from his hometown of Stellenbosch and many top production companies in South Africa now use DX. He has also followed up sales to the majority of his customers with a two-day training course in experimental design, using DX, of course!

Photo by Nico (Kruger National Park)

Nico has many hobbies including reading (he tries to read one non-technical book a month, especially in Afrikaans), golf, gardening with a particular interest in succulents (with which South Africa is abundantly blessed), and wildlife (he & his wife visit Kruger National Park about twice a year). He is also a very accomplished photographer (see one of his photos above).

If you would like to contact Nico, you can e-mail him at **industat.pro@mweb.co.za**.

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

Address Service Requested

Stat-Ease® statistics made easy™

Presorted
Standard
U.S. POSTAGE PAID
Minneapolis, MN
Permit No. 28684