stateaser

ABOUT STAT-EASE® SOFTWARE, TRAINING, AND CONSULTING FOR DOE
Phone 612.378.9449 Fax 612.746.2069 E-mail info@statease.com Web Site www.statease.com

Workshop Schedule

Experiment Design Made Easy

August 18-20, 2009: Minneapolis, MN
November 3-5, 2009: Minneapolis, MN

Study the practical aspects of design of experiments (DOE). Learn about simple, but powerful, two-level factorial designs. $1595 ($1295 each, 3 or more)

Response Surface Methods for Process Optimization

July 7-9, 2009: Minneapolis, MN
December 1-3, 2009: Minneapolis, MN

Maximize profitability by discovering optimal process settings via RSM. $1595 ($1295 each, 3 or more)

Mixture Design for Optimal Formulations

August 11-13, 2009: Minneapolis, MN
October 27-29, 2009: Minneapolis, MN

Find the ideal recipes for your mixtures with high-powered statistical tools. $1595 ($1295 each, 3 or more)

DOE for DFSS: Variation by Design

November 10–11, 2009: Minneapolis, MN

Use DOE to create products and processes robust to varying conditions, and tolerance analysis to assure your specifications are met. A must for Design for Six Sigma (DFSS). $1195 ($995 each, 3 or more)

Designed Experiments for Life Sciences

Fall 2009: Cambridge, MA

Learn how to apply DOE to Life Science problems. $1495 ($1195 each, 3 or more)

PreDOE: Basic Statistics for Experimenters (Web-Based)

PreDOE is an entry-level course for those who need to go back to the basics. See *http://www.statease.com/clas_pre.html* for more information. $95

Attendance is limited to 16. Contact Elicia at 612.746.2038 or workshops@statease.com.

Stat-Ease statistics made easy®

Paper Clip Experiment Illustrates Statistical Design Principles

This is a reprint (updated only slightly) from a 1994 Stat-Teaser article by Mark Anderson. This second printing makes it available in the archive of fun and informative articles we provide on our web site via Stat-Teasers published after 1997.

Breaking paper clips provides marvelous therapy for any number of problems: boredom, frustration, anger, or whatever. So it comes as no surprise that students enjoy doing in-class tests of clip strength.

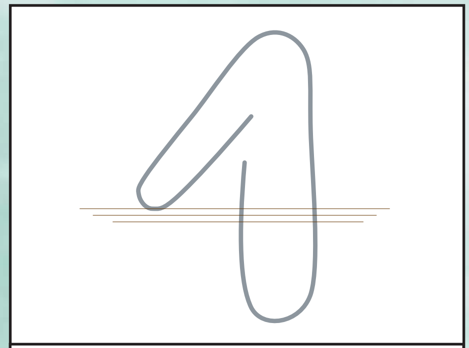

Fig. 1: Paper clip poised to bent

Years ago, before too many new tools crowded it out, Stat-Ease regularly presented this exercise in their DOE workshops. The experiment builds understanding of variation and how it can be handled with simple comparative designs. For teaching purposes it works best if each student breaks two brands of clips. This provides data for a paired t-test, blocking out variability due to the tester. Here is the procedure:

1. Randomly choose a clip.
2. Gently pull it apart with the big loop on the right. The angle affects performance so be precise.
3. Move the smaller loop of the clip to the edge of the table (the big loop should now overhang—see Figure 1).
4. Hold the small loop down firmly with your left thumb and forefinger. Then bend the big loop straight up and back. (This angle also affects performance, so be precise). Continue bending the big loop back and forth until it breaks. Record the count for each clip. (Each back and forth movement counts as 2 bends.)

In the process of developing this exercise, several brands were tested in a fully-randomized unblocked design (results shown in parentheses):

- Rogers #1 nickel (6, 8, 9, 12, 18, 3, 9, 13, 7, 4, 7, 9, 8, 11, 8)
- Baumgardens #1 Golden (8, 3, 8, 10, 7, 8, 8, 10, 10, 4, 11, 8, 9, 6, 11)
- Omni #1 Gem (4, 28, 12, 12, 27, 1, 13, 17, 8, 14, 1, 16, 10, 12, 6)
- ACCO #1 (21, 27, 24, 25, 26, 25, 20, 28, 27)

Analysis from Design-Expert® software confirms what you can see from the raw data: the ACCO clips clearly outperform all the others (see Figure 2 on page 2). If you have the software, you can verify this yourself by entering the data in a general factorial design and computing the

—Continued on page 2

—Continued from page 1

ANOVA. Post-ANOVA t-testing is also a very useful feature.)

The first three brands (Rogers, Baumgardens, and Omni) all incorporate a metal plating that looks nice but reduces strength. (Non-skid clips also were considered, but the notches weaken them so much that they usually break on the first bend.)

After reviewing the data, Stat-Ease settled on the Rogers (nickel) and the Baumgardens (gold) clips for its in-class, paired experiment. (The Omni clips varied too much. The ACCO clips performed too well—it takes too much work to break them*.)

Several series of paired tests were conducted with 10 to 20 students each. The results consistently favored the Rogers clip by a slight margin over the Baumgardens. The difference was so small that it would have been obscured if each student had tested only one clip. But by having each person test both clips, an example of blocking, a clear difference often emerged.

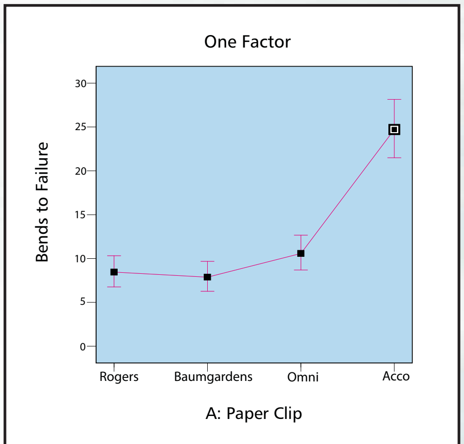

Fig. 2: Clips: bends to failure

Stat-Ease found the paper clips exercise to be a useful element in it's DOE workshops. It nicely illustrated how statistical design and analysis of experiments can overcome variation. Plus the materials fit in a pocket!

Is all this data on lowly paper clips making you tense and irritable? That's easily remedied. Grab a clip and break it into little pieces. By the way, could you get a count on that?
—Mark Anderson, mark@statease.com

*Notice that I gave up on the ACCO clips after breaking only 9 of them (versus 15 for all the others). Can I invoke a seven-year statute of limitations for slacking off like this? I hope so!"

P.S. In a Winter 2002 article published by the *American Journal of Pharmaceutical Education (V66)* a University of New Mexico professor details how he used paper clips to run students through a mock clinical trial. The hypothesis was that breaking them provides stress relief. This exercise revealed many pitfalls to avoid in experimental design, such as not performing a power study beforehand to establish the required sample size. (See the article below for help with this.)

Power Calculator for Binomial Responses

A clip on a medical device

Let's say that your company makes a metal clip for a medical device, such as the one pictured in Figure 1 for a constant-pressure air mask.

It works very well as is. However, it would be good to make it robust to the bending that naturally occurs during use in the field.

The engineers consider a number of factors for a two-level experiment, such as changing to a new material, making the part wider and/or thicker, and so on. After much brainstorming and tough decision-making, the team settles on 9 factors to be tested by a minimum-run resolution IV (MR4) plus 2 design—requiring only 20 combinations (experimental runs).* The true measure of success will be a reduction in the rate of returns from customers, but this occurs so rarely that, to create a more rapid rate of failure, an accelerated test is developed.

Fig. 1: Metal clip on medical device

Now the stage is all set, except for one critical detail: How many clips must be made for each of the runs? Fortunately for all concerned with quality and reliability issues, the statistical staff at Stat-Ease (primarily via work by Consultant Wayne Adams) developed a power calculator for binomial responses such as fraction defect. This handy spreadsheet (Microsoft Excel) calculates sample sizes needed for tests involving multiple factors with two levels each. It's posted at http://www.statease.com/powercalc.html.

Statistics made easy

First of all, do not be put off by the term "binomial." This is simply a statistical term that describes an experiment that consists of repeated trials with only two

—*Continued from page 2*

possible outcomes, such as success or failure.

Secondly, be forewarned that Greek letters will be presented in the sample size calculator. Don't let this scare you off. For example the template displays a field labeled "Δ" (delta) to enter the change (think "d" for difference) that you want to accomplish in your response, such as a decrease of 0.02 (2 percent) in the proportion of failures. The "α" (alpha) and "β" (beta) symbolize two types of error that must be managed for purposes of designing your experiment. One other Greek letter that appears on the spreadsheet, "δ" (small case delta), represents a technicality that one finds detailed in the original reference (noted below the calculations in the Excel file). Don't worry—be happy. When you plug in the required numbers, the answers you receive will be fairly obvious in their meaning.

Now let's see how this works by way of our example.

Generating a suggested sample size for the proposed experiment

Let's say that the medical-clip manufacturer sets a goal to reduce failure by at least 25 percent. When undergoing the accelerated test, the current design generates a 0.2 failure rate, so it must be decreased by at least 0.05 (in proportional scale) to accomplish the technical mission (0.05/0.2 = 0.25 or 25 %).

To generate the sample size, the binomial power calculator only needs three numbers:

1. The current average proportion ("p(bar)"): **0.2** in this case.
2. The minimally-desired difference delta (Δ): **0.05**, for example.
3. The total combinations being tested ("Design size N"): **20** runs.

The inputs (in blue), default parameters

Sample size to detect a Δ change in a binomial proportion using a two-level factorial:

p(bar) =	0.2	current proportion		
Δ =	0.05	minimum change in proportion		
alpha (α is typically 0.05 or 0.10) =	0.05	z-alpha/2 =	1.96	
Power (1-β is typically 0.80 or more) =	0.80	z-beta =	0.84	
δ =	0.0626101	change in transformed scale		
Design size N =	20	user input (number of runs)		
Units per run for power =	101			
Units per run to avoid too many zeros =	25			
Recommended Units per run; n =	**101**			
Total units =	**2020**			

Fig. 2: Screen shot of sample-size calculator after entering required inputs (in blue)

(green) and outputs (gray or black) for the proposed experiment are shown in Figure 2.

Keep in mind that the 0.05 improvement (Δ) in fraction defect represents the minimal change that will earn the engineers a "job well-done." (Obviously a larger decrease in failure rate would be even better!) It would be a shame if this positive outcome was missed due to testing too few parts. This underscores the value of doing a sample-size calculation.

Some statistical details

The value for alpha can be left untouched at the generally-acceptable default of 0.05 which allows a 5 percent risk of a false positive; that is, declaring that an effect is real when it's really caused by chance. (Statisticians deem this a Type I error.) The power is set at 0.80 to achieve at least an 80 percent chance of seeing the difference delta in the response that's been entered. Try not to get thrown off by the parenthetical note that power is actually computed as a function of beta (β), which is an error in detection. This again is fairly standard for design of experiments—an error in detection; that is, not seeing a real effect of some importance. (This is known as a Type 2 error.) The "z" scores are a function of the alpha and power. These values are required for the calculations. (If you took a basic class in statistics, recall that the normal bell-shaped curve can be divided into units of standard deviation that are expressed as "z scores.")

The answer for sample size

In our illustration, the recommended sample size for each of the 20 runs is calculated to be 101; for a total of roughly 2,000 (rounding n to 100 won't weaken the design appreciably!). Assuming you like to experiment (or you wouldn't have read this article!), why not go ahead and try this sample size calculator? For example, increase the power to 0.95, which might be required for a verification test to pass FDA standards. Notice the increase in units per run and total number. Feel free to fiddle around with other parameters as well. For example, what if a much smaller difference is considered to be important—say 0.01? You may be surprised at how much that affects the sample size!

—Mark Anderson, mark@statease.com

*MR4 designs are available exclusively in later versions of Design-Ease® and Design-Expert® software, which also offer MR5 for minimum-run resolution V experiments.

Introducing Boston Software Group

06/09

Stat-Ease is pleased to announce our new partnership with Boston Software Group of Boston, MA to market Stat-Ease software and training in the Northeast region (extending as far West as MI and OH) of the United States. Boston Software Group has over 40 years of combined sales, marketing and technical experience in the experimental design and exploratory data analysis markets, making them well-equipped for this new role. Boston Software Group's customers include worldwide market leaders in numerous industries such as Merck, ExxonMobil, IBM, Genentech, United Technologies, and American Express.

Fig. 1: Left to right—Pat Whitcomb, Katrina Hauser, Peter Trogos, Mark Anderson

If you are a Stat-Ease customer in this region, we think you will find it both convenient and beneficial to work with the knowledgeable Boston Software Group team. They look forward to meeting you and providing local support.

Be assured that free technical support and limited free statistical support is still provided to you by Stat-Ease. As always, you are welcome to come to us directly for any reason. We greatly appreciate your business!

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

Address Service Requested

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