

STATeaser

ABOUT STAT-EASE® SOFTWARE, TRAINING, & CONSULTING FOR DOE

Workshop Schedule

Experiment Design Made Easy (EDME)

June 19-20, 2012: Minneapolis, MN
August 6-7, 2012: San Francisco, CA
\$1295 (\$1095 each, 3 or more)

Response Surface Methods for Process Optimization (RSM)

June 21-22, 2012: Minneapolis, MN
August 8-9, 2012: San Francisco, CA
\$1295 (\$1095 each, 3 or more)

Mixture Design for Optimal Formulations (MIX)

July 17-18, 2012: Minneapolis, MN
October 23-24, 2012: Minneapolis, MN
\$1295 (\$1095 each, 3 or more)

Advanced Formulations: Combining Mixture & Process Variables (MIX2)

October 25-26, 2012: Minneapolis, MN
\$1495 (\$1195 each, 3 or more)

PreDOE: Basic Statistics for Experimenters

Online Course

Free (a \$95 value). Learn more at:
http://www.statease.com/class_pre.html.

Free Webinar: Overview of Robust Design, Propagation of Error, & Tolerance Analysis

In this advanced-level webinar, Stat-Ease Consultant Pat Whitcomb will discuss robust design, propagation of error, and tolerance analysis.
www.statease.com/webinar.html.

★ Fourth European DOE User Meeting, Vienna, Austria

June 26-28, 2012—See page 4—Sign up today, space is limited!

Workshops limited to 16. Multiclass discounts are available. Contact Elicia Bechard at 612.746.2038 or workshops@statease.com.



Two Bean or Not Two Bean—Experimenting on Germination

Every so often I search the internet on “Design-Expert®,” “design of experiments” or some other term, such as a specific application that someone has asked me about. (Many people like the idea of DOE but need to be assured that it will work for them.) I generally net a few fish that I really wasn’t trying to catch, some that I find very fascinating. For example, Auburn University’s Samuel Ginn College of Engineering posted a report on a “DOE—Bean Experiment” by a team of students taking a chemical engineering lab.*

The reason this caught my eye is that I spent a number of years working for General Mills in Minneapolis as a seller and then buyer of guar beans, which work nicely to thicken salad dressings and the like. I learned a lot about what makes beans grow and what’s under their covers. So when I saw this report from Auburn talking about endosperms, cotyledons and so forth, that got me revved up for more.

This team of budding engineers (ha ha) focused their experiment on germination. They studied:

- Bean Type: Lima vs Kidney (pictured in Figure 1)
- Light Exposure: No vs Yes**
- Salt: 0 vs 4.7 mole % in water (tap vs saline solution)
- Temperature: 70 to 80 degrees F via a full, two-level factorial replicated

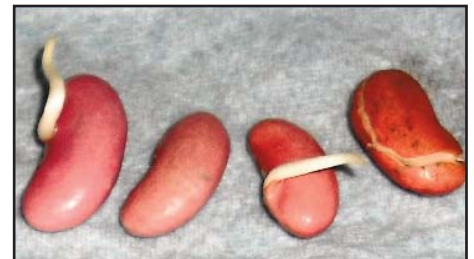


Figure 1: Sprouting kidney beans from Auburn University Experiment

in two blocks ($2^4 \times 2 = 32$ runs).

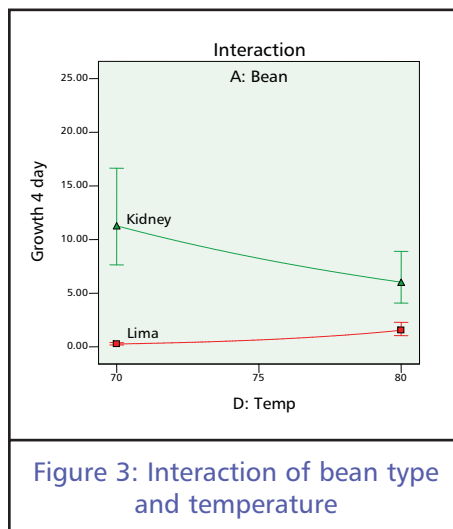
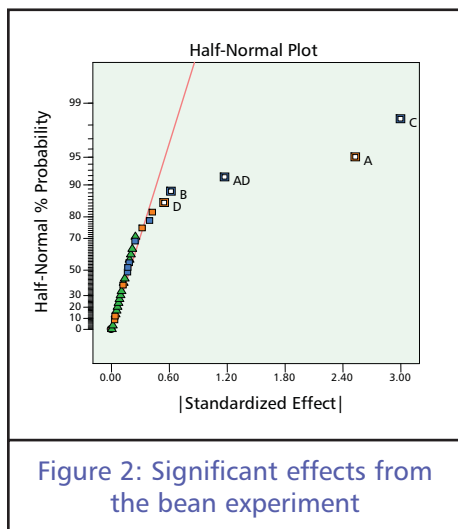
The results were very enlightening (I couldn’t resist). The half-normal plot in Figure 2 highlights the effects which stood out from the normally-distributed measures of error (green triangles) coming from replicate runs. The ones near the line, AC and BC, are significant only at $p < 0.1$ —their magnitude is of trivial importance.

Given 4 days for growth, it turns out that:

- In general, kidney beans far outgrew lima beans.
- It mattered little for either bean whether they were kept in the dark or exposed to light.
- Salt halts germination for all practical purposes (kidney beans overcame this slightly).
- The effect of temperature depends on the type of bean. Kidney beans grew better at 70 degrees, whereas lima beans liked the warmer environment. See Figure 3.

—Continued on page 2

—Continued from page 1



By the way, I found from the Box-Cox plot in the diagnostics provided by Design-Expert that the log transformation gave a significantly better model-fit. I favor the base-10 log due to it being easier to anti-log. However, in this case I thought it best to apply the natural log (joke intended). In any case (makes no difference which log) this transformation causes the wider least-significant-difference (LSD) bars at the higher growth.

All in all I am very impressed with the work done by this team of student

experimenters. I have no doubt that they will grow far (wink) in their careers as chemical engineers.

—Mark J. Anderson,
mark@statease.com

*Search the Internet for “DOE Bean” to find the report by Jaki Fleming, Emily McMenemy, Andy Todd, Adam Warnke presented on 12/2/09.

**I reversed the order on this factor so “no” is the low (minus) level and “yes” the high (plus). This created less confusion when interpreting the results.

Academic Support

Dave Mills, Lab Manager for the Chemical Engineering department at Auburn University’s Samuel Ginn College of Engineering, told me that every semester he requires their senior class to do a “capstone” lab project using Design-Expert software. He says that this project always gets positive student feedback. They enjoy using the software, being ready to embrace statistical methods after a good deal of lecturing along those lines. Other engineering departments at Auburn teach DOE with more general statistical software, but Dave feels that Design-Expert is “much superior” for experimental work done by chemical engineers.

Stat-Ease provides a great deal of support for students and graduate researchers. Economical licensing can be purchased for PC labs or for ongoing research. For more details, contact Heidi Hansel Wolfe, Director of Academic Support, via heidi@statease.com.

MONEY in Baseball—Does More Money Equal More Wins?

It’s that time of year again. Spring is in the air, things are greening up, and the start of another baseball season is upon us. I love this time of year. Hope springs eternal, and everything is made new after a long winter (though this extremely mild winter wasn’t so bad). Our hometown nine, the Minnesota Twins, are going into their third year in the still new Target Field. So, when the Twins cut the payroll from \$115 million to \$100 million, many of the fans, myself included, were quite frustrated. “How could you cut the pay-

roll after having one of the worst years in recent memory and with season tickets still nearly sold out?” “We’re spending our money,” say the fans, “so use it to put a good team on the field!”

Well, this got me to thinking about the correlation between payroll and winning. Maybe they are smart to pull in the reigns a little. Is there really a correlation between a major league baseball team’s payroll and its number of wins? Design-Expert software provides a good

tool to delve into this question.

I first looked at the data for just last season: team payrolls and team wins. The data I found included total team payroll, average player salary, and median player salary, so I included all three of these as factors. The only response was “team wins.” Unfortunately, this one season didn’t provide a good model for any of the factors, though team payroll was slightly significant. The adjusted and predicted R^2 values were very low

MONEY in BaseBALL—Does More Money Equal More Wins? (Cont. from page 2)

(0.1375 and 0.0652 respectively). This means that only about 13% of the variation in the data is explained by the model. The rest is unexplained noise.

From there, I decided to look at a decade of data and average out the payroll and wins for each team. The averaging helped eliminate all of the year-to-year abnormalities, such as an injury-plagued season (like the Twins had last year). These abnormalities can cause a breakdown in the relationship between payroll and wins, and show up as statistical outliers. The data averaged over a decade did provide a much better model (0.48 adj. R^2 and 0.43 pred. R^2).

Then, the model “diagnostics” highlighted some interesting aspects. Because the Yankees spend much more money than any other team, their data point is far away from the other data points (see Fig. 1, where it is highlighted with a white center). This creates a great potential for influence on the model, that is, high leverage (see Fig. 2). In the figure, the leverage is above the red cutoff line, which indicates extremely high leverage. Leverages range from 0 to 1. A leverage of 1 means the model must fit that data point perfectly (it has ultimate influence). The Yankees data point has a leverage more than twice the average (0.46), so it has great potential to influence the model.

The question is, does this high leverage point have an undue influence on the model, or does it still fit well with the rest of the data? This can be answered by looking at another diagnostic plot: the Cook's Distance. The Cook's distance is a measure of how much the regression would change if the case were deleted. When I looked at this plot, it was clear that the Yankees data point had a low Cook's distance, so its

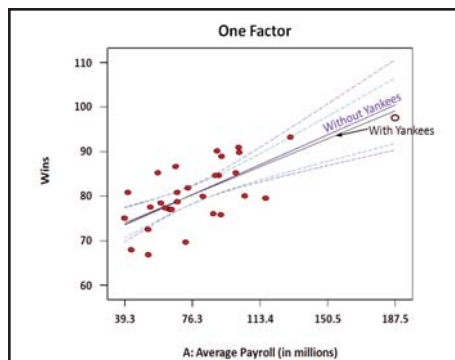


Figure 1: Overlay of the Average Payroll vs. Average Wins

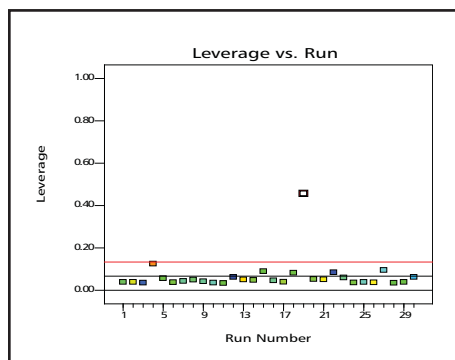


Figure 2: Leverage of data points

actual effect on the model is minimal. This is further illustrated in Fig. 1. Here, I overlaid two models for comparison. One model has the Yankees point included (black line and cyan dashed confidence intervals) and the other has the point ignored (blue line and blue dashed confidence intervals). As predicted by the low Cook's Distance, the two models match very closely, despite the high leverage caused by the Yankees exorbitant spending.

Even though I've been well-ingrained with the stock investing mantra that “past performance is not indicative of future results,” I can't help but try to predict by using this “historical data.” How will the Twins do this year according to our model? Well, the prediction for a team with a \$100 million payroll like the Twins

is for 84.3 wins, which is well above the line set by Vegas (over/under of 74), but how confident can we be in this prediction?

The prediction interval (PI) tells us what we'd expect to get for one run of the DOE at the settings we're predicting. The 95% PI for a \$100 million payroll team is 73.2 - 96.1 wins. The over/under of 74 is right near the lower edge of that interval. However, keep in mind that the data in our model are averaged over a decade. So, technically, the PI is for average wins per year for a team that has an average payroll of \$100 million for an entire decade. We would expect more variation over just one season, resulting in a wider prediction interval. Based on this average payroll vs. average wins model, I'd definitely take the over on the bet of the Twins winning 74 games this year, but I wouldn't be shocked if they only won 73) 😊. 74 wins would still be a big improvement from last year's injury-plagued 63-win campaign.

As we've seen, there is some correlation between team payroll and wins. While I don't think the small difference between a \$100 million and \$115 million payroll (2.5 wins according to our model) is likely to be the difference between missing and making the playoffs, the cut in payroll is a tough sell to the fans coming off an extremely disappointing season. If I were them, I'd spend the money on a solid player just so they could sell more hope (and tickets). Apparently, the Twins' front office doesn't agree. Even with the cut, the Twins payroll is much higher than it was before Target Field was opened. Hopefully, they can stay healthy this year and get the 84 wins predicted by our model. This would give them a decent chance of making the playoffs. I know I'll be watching anxiously, with a little extra hope provided by this analysis.

—Brooks Henderson, brooks@statease.com

4th European DOE User Meeting in Vienna

05/12

Stat-Ease, Inc. and our German partner, Statcon, are pleased to invite you to the *4th European DOE User Meeting* in Vienna, Austria this coming June 27-28, 2012. This meeting is only held once every two years and is always greatly enjoyed by all who attend.

The meeting will focus on DOE, with a special emphasis on Design-Expert® software. Both the theoretical and practical aspects of DOE will be addressed, including some of the latest developments in the field. The two meeting days will include lectures by keynote speakers and other DOE experts, case study presentations by DOE practitioners, and an opportunity to consult with instructors about your own DOE applications. Stat-Ease is also offering two optional pre-meeting workshops on June 26, 2012 to help you expand your DOE knowledge.



Hotel Bristol, Vienna, Austria

Don't miss this incredible opportunity to learn more about DOE while visiting the beautiful, cultural city of Vienna! Your meeting registration fee includes a dinner boat cruise on the first night of the conference. Registration is limited. Sign up soon at http://www.statcon.de/4th_european_doe_user_meeting_in_vienna_70_en.html to ensure your spot! For more information, please see our web site at <http://www.statease.com/4theuropeanmeeting.html>.

Address Service Requested

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

statistics made easy®
Stat-Ease®



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