(Excerpted from manuscript for Chapter 10 of *RSM Simplified* by Whitcomb and Anderson. See www.statease.com/rsm_simplified.html for information on this book.)

"I'm very well acquainted, too, with matters mathematical,
I understand equations, both the simple and quadratical,"

--A verse from "I Am the Very Model of a Modern Major General" from *The Pirates of Penzance* by Gilbert and Sullivan

Response Surface Methods for Experimenting on Computer Simulations

At this dawn of the 21st century, engineers are now taking on projects of unprecedented complexity. For example, consider the state of the art for aircraft now versus 100 years ago when the Wright brothers made their first flight at Kitty Hawk, North Carolina. They performed many experiments on wing design, propeller configuration, and the like. After all these preliminaries it still took a great deal of trial and error before the Wrights' aircraft finally got off the ground. Nowadays, much of the development for aircraft and other sophisticated equipment occurs via experiments on high-powered computer simulations.

One approach is simply to randomly sample some number of times within the experimental space. However to ensure a more systematic array of points, it makes sense to first segment the region into a given number of rows and columns. Then sampling can be done in such a way that, in each row and each column, 1 point appears – no more, no less. This result is called a Latin hypercube design or LHD (McKay, Beckman, and Conover, 1979).

HISTORY OF LATIN SQUARE

A Latin square is simply a grid in which each number appears only once in each row and column. Table 10-3 shows a Latin square of order 4.

Table 10-3: Latin square

	W	X	Y	Z
July	1	2	3	4
Aug	2	1	4	3
Sept	3	4	1	2
Oct	4	3	2	1

These arrays make good templates for design of experiments with multiple blocking. For example, suppose you wanted to check test four brands of tires: W, X, Y, and Z. The layout in Table 10-3 provides a sensible plan for rotating the tires month-by-month on varying wheels identified by numbers 1 through 4 by position on the car: 1 - left front, 2- right front, 3 - left back, 4 – right back (Peterson, 2000).

Why “Latin”? It turns out that in the late 18th century the mathematician Euler (pronounced “oiler”) laid out squares of this sort out using Latin letters. He postulated a famous problem that could not be solved with a Latin square:

“For since the fabric of the universe is most perfect and the work of a most wise Creator, nothing at all takes place in the universe in which some rule of maximum or minimum does not appear.”

--Leonhard Euler

Figure 10-4 shows an LHD cited by Myers and Montgomery (p. 484) for two factors, each ranging 0 to 16. The levels in each column of the design matrix are randomly arranged to construct the design.

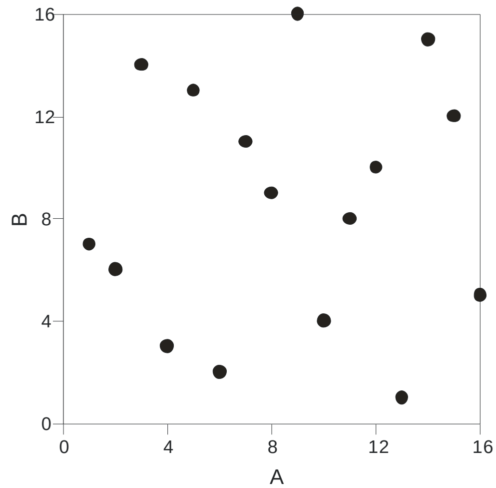

Figure 10-4: Latin hypercube on two factors

As you can see, this semi-random scattering of points leaves much to be desired for filling all the space defined by the experimental factors. We suggest better alternatives in the Appendix.

Within the past few decades, interest has shifted from the LHD to more traditional designs for response surface methods, such as the central composite design. For example, NASA engineers developed a computer simulation of a tetrahedral truss for a scientific space platform (Unal, Wu , and Stanley, 1997). They studied 9 factors in only 83 runs via a central composite design (CCD) with a fractional factorial core and rotatable star points (alpha 2.83).

Other aerospace engineers made use of a face-centered central composite (FCD) to improve the wing design on a lightweight fighter jet (Zink, et al, 1999). They wanted to assess a new active aeroelastic wing technology (see Figure 10-5) that could only be simulated via physics-based finite element analysis on high-powered computers.

Figure 10-5: NASA's Active Aeroelastic Wing (AAW) F/A-18A research aircraft

The simulator generated an estimate of wing weight, which the engineers hoped to minimize as a function of three key factors:

A. Aspect Ratio, 3–5.

B. Taper Ratio, 0.2–0.4.

C. Thickness Ratio, 0.03–0.06.

Table 10-4: FCD on wing design for jet fighter

Run	A: Aspect	B: Taper	C: Thick-ness	Wing Weight (pounds)
1	3	0.2	0.03	334.6
2	3	0.2	0.06	126
3	3	0.4	0.03	407.3
4	3	0.4	0.06	161.2
5	5	0.2	0.03	833.7
6	5	0.2	0.06	392.1
7	5	0.4	0.03	1070.2
8	5	0.4	0.06	326.5
9	3	0.3	0.045	226
10	5	0.3	0.045	460.9
11	4	0.2	0.045	286.9
12	4	0.4	0.045	408.3
13	4	0.3	0.03	608.5
14	4	0.3	0.06	236.7
15	4	0.3	0.045	380.5

This design contains no replicates because they would generate identical responses from the deterministic computer simulation. Therefore the ANOVA, shown in Table 10-5, does not include any pure error, nor does it provide a test on lack of fit.

Table 10-5: ANOVA on wing design

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F
Model	867623	6	144603	47.4	< 0.0001
A	*334268*	*1*	*334268*	*109.5*	*< 0.0001*
B	*16016*	*1*	*16016*	*5.2*	*0.0512*
C	*404733*	*1*	*404733*	*132.6*	*< 0.0001*
AC	*66722*	*1*	*66722*	*21.9*	*0.0016*
BC	*14416*	*1*	*14416*	*4.7*	*0.0615*
C^2	*31467*	*1*	*31467*	*10.3*	*0.0124*
Residual	24416	8	3052		
Cor Total	892039	14			

When analyzing data from simulations, keep in mind that the true computer model will be only approximated by RSM. The RSM metamodel will not only fall short in form of model, but also in the number of factors. Therefore, predictions will exhibit systematic error, or bias. This is what will be measured in the residual–not the normal variations observed from experiments on physical processes. Despite these circumstances, much of the standard statistical analyses remain relevant, including measures of model-fit such as PRESS and $R^2_{Predicted}$. However, the p values will not be accurate estimates of risks associated with the overall model or any of its specific terms.

The fit of predicted versus actual data (from the simulation) looks very good as you can see in Figure 10-6.

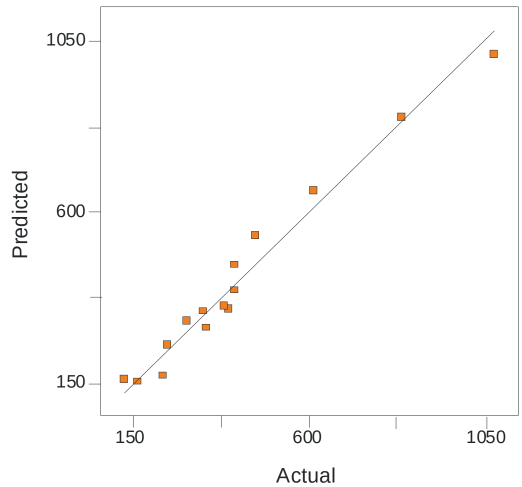

Figure 10-6: Actual response from simulator versus predicted

However, the Box-Cox plot (see sidebar "The Box-Cox Plot") shows that residuals can be reduced significantly via a log transformation, which is not surprising considering the broad range of response–from 150 to 1,050. Furthermore, as shown in Table 10-6, the resulting model with the response in log scale is more parsimonious (what a great word for saying "simpler"!). Even though it contains only linear terms (A, B, and C), the log model fits the actual data more precisely.

Table 10-6: Statistics for models on wing simulation, with and without log transformation

Statistic	No Transform	Log Transform
Model	Quadratic	Linear
Root MSE	55.25	0.05
$R^2_{Predicted}$	0.8627	0.9362

THE BOX-COX PLOT

We already detailed the Box-Cox plot in the Appendix to chapter 5, but we thought it would be worth reviewing. As you may recall, this graphical method, named after its originators, shows how dimensionless residuals change as a function of varying powers of response transformation. Figure 10-7 shows the results for fitting the wing simulation.

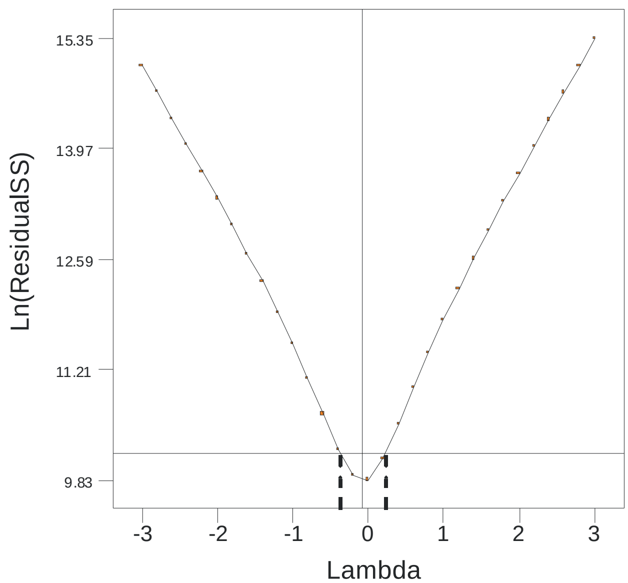

Figure 10-7: Box-Cox plot for linear model for wing performance

The y-axis displays the residuals after transforming the response by a range of powers from -3 (inverse cubed) to +3 (cubed). The powers, labeled "Lambda" (Greek letter λ), are displayed on the x-axis of the Box-Cox plot. The option of not doing any transformation is represented by the value 1. This specifies that all responses y be taken to the power lambda of 1, but y^1 equals y, so this makes no difference.

In this case, the plot shows the power of 1 being outside of the 95% confidence zone for minimal residuals (indicated via the dotted lines). Notice that the minimum residual occurs near the value of 0. This is another special power because it makes no sense to compute y^0, which of course would create all 1's. It actually represents taking the logarithm of the response, which proves to be very effective for modeling the performance of the wing on this particular rocket ship.

Which is more daunting–rocket science or statistics?

"A human being is the best computer available to place in a spacecraft. . . It is also the only one that can be mass produced with unskilled labor."

--Werner von Braun

The 3D surface in Figure 10-7 shows how the predicted response varies for the two factors creating the most impact on the wing–aspect (A) and taper (C).

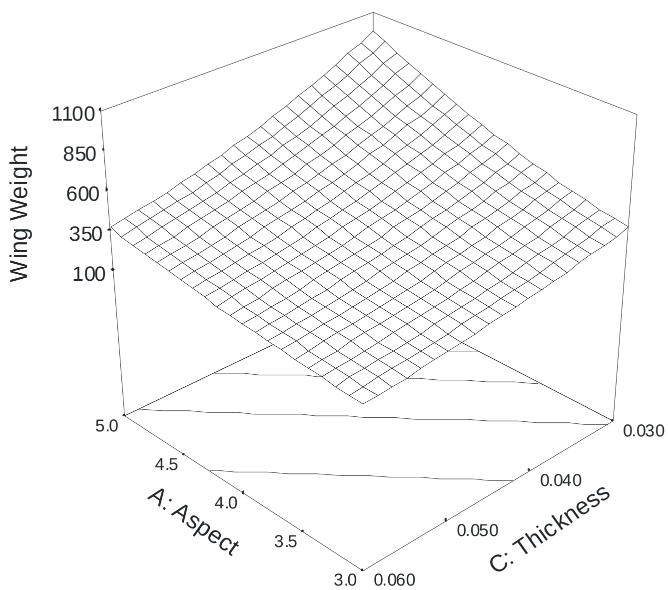

Figure 10-7: 3D response surface for wing simulation

Pictures like this can far outweigh the untold numbers potentially generated by a computer simulation, oftentimes at great expense.

Appendix: Alternative RSM Designs for Experiments on Simulations

In chapter 7 we provided a checklist for gauging the quality of designs for RSM. Many of the line items can be carried forward for application to deterministic computer simulations, but, as shown in the following list, others (struck out in this list) can be ignored–mainly due to the lack variability in the inputs and outputs:

- ☐ Generate information throughout the region of interest.

- ☐ Ensure that the fitted value, ŷ, be as close as possible to the true value.
- ☐ ~~Provide an internal estimate of error~~.
- ☐ Give good detectability of lack of fit.
- ☐ Allow experiments to be performed in blocks.
- ☐ Allow designs of increasing order to be built up sequentially.
- ☐ ~~Be insensitive to wild observations~~.
- ☐ ~~Behave well when errors occur in the settings of factors (the “x’s”)~~.
- ☐ ~~Not require an impractically large number of factor (x) levels~~.
- ☐ Require a minimum number of runs.
- ☐ Unique design points in excess of the number of coefficients in the model chosen by the experimenter.
- ☐ ~~Provide an estimate of pure error~~.
- ☐ Remain insensitive to influential values and bias from model misspecification.

Santner, Williams, and Notz (2003), who wrote the book on computer experiments, offer a much shorter list. They advise use of designs that:

- Provide information about all portions of the experimental region.
- Allow one to fit a variety of models.

Surprisingly, these authors suggest that at least one observation in any set of inputs be replicated to guard against unannounced changes in computer code.

Let’s revisit the simple case of two factors for which we showed application of the Latin Hypercube design (LHD). Figure 10-4 displays an LHD with 16 runs–far more than needed to fit a quadratic polynomial, which normally would be adequate for purposes of RSM, but may not be for complex computer simulations.

One alternative to consider is a distance-based design, which can be readily constructed via software that supports response surface methods. Simply put, the distance criterion selects a given number of points from a candidate set in such a way that they will be spaced as far apart as possible. For example, the following 17-point distance-based design (black circles in Figure 10-8) results from a candidate set consisting of 29 factorial combinations laid out concentrically by Design-Expert software (open circles for points not chosen).

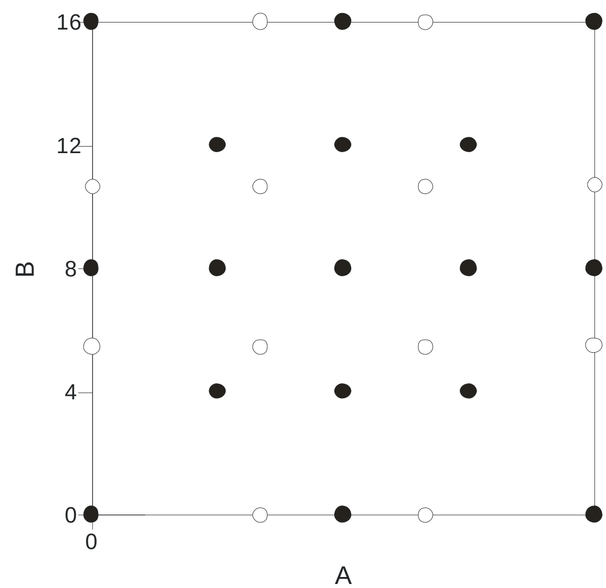

Figure 10-8: 17-point distance-based design for two factors

When computer time is very costly, a 16-17 point design may be excessive for only two factors. One school of thought by experienced practitioners of RSM on simulation is that CCDs work well for up to five factors (Unal, Lepsch, Jr., and McMillin, 1998). This design choice reduces the number of runs considerably, for example, to only 9 points for the two-factor case (for a CCD with no replicates). However, these same experts (Unal, et al) suggest that for six or more factors D-optimal design geared to the quadratic model is a good choice <u>if</u> overdetermined by 50% or so. For example, NASA scientists performed 45 runs on a simulator for a rocket-powered single-stage launch vehicle (Unal, et al). Only 29 runs were needed at a minimum for fitting the quadratic model on six factors related to wings and the like on this spacecraft. Therefore, they overdid the design by 55%.

Hitchhiking off this idea, we propose using D-optimal criterion to pick the minimal set of candidate points needed for the quadratic polynomial. Then add half that many points based on distance-based criterion. These augmented design points plug the remaining gaps and thus achieve the objective for computer simulations to be space-filling.

GOING THE DISTANCE

Here's how the distance-based criterion works:

1. Specify n points such that $p < n < c$, where p is the number of parameters in the chosen model and c is the candidate set.
2. Choose an initial design point at a vertex in the experimental space.
3. Add the next candidate point whose minimum Euclidean distance* from points already in the design is as large as possible.
4. Repeat step 2 until all n points are chosen.

This criterion may not provide sufficient points to fit the chosen model. However, the distance algorithm can be modified so that it only picks points that increase the rank of the design matrix until it equals the number of model coefficients. We do not recommend use of this modified distance approach because D-optimal selection is far superior for picking model points.

*Defined by U.S. National Institute of Standards (NIST) as follows:

"The straight line distance between two points. In a plane with p_1 at (x_1, y_1) and p_2 at (x_2, y_2), it is $\sqrt{((x_1 - x_2)^2 + (y_1 - y_2)^2)}$. *In N dimensions, the Euclidean distance between two points p and q is* $\sqrt{(\Sigma_{i=1}^{N} (p_i-q_i)^2)}$ *where* p_i *(or* q_i*) is the coordinate of p (or q) in dimension i.*" (Source: http://www.nist.gov/dads/HTML/euclidndstnc.html.)

NIST goes on to suggest that you see also "Manhattan distance." This is the distance between two points measured along axes at right angles, which some people call the "taxicab metric." Unless you are a pigeon, this sounds much more practical for city dwellers than Euclidean distance!

References

McKay, M., Beckman, R. and W. Conover. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code." *Technometrics* 21 (1979): 239-245.

Myers, Raymond, and Douglas Montgomery. *Response Surface Methodology*. 2nd ed. New York: John Wiley and Sons, 2002.

Peterson, Ivars. "Completing Latin Squares." *Science News Online*, Week of May 6, 2000, Vol. 157, No. 19.

Santner, Thomas, Brian Williams and William Notz. *The Design and Analysis of Computer Experiments*. New York: Springer-Verlag, 2003.

Unal, Resit; Wu , K. Chauncey, and Douglas O. Stanley. "Structural Design Optimization for a Space Truss Platform Using Response Surface Methods." *Quality Engineering* 9, no. 3 (1997): 441-447.

Unal, Resit, Lepsch, Jr, Roger A. and Mark L. McMillin. "Response Surface Model Building and Multidisciplinary Optimization Using Overdetermined D-Optimal Designs." *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, St. Louis, Missouri, September 2-4, 1998.

Zink, P. Scott, et al. "Impact of Active Aeroelastic Wing Technology on Wing Geometry Using Response Surface Methodology." Talk at *Langley International Forum on Aeroelasticity and Structual Dynamics*, Williamsburg, VA, June 22-25, 1999.