How to Properly Size Response Surface Method Experiment (RSM) Designs for System Optimization

Mark J. Anderson, PE, CQE, MBA
Principal, Stat-Ease, Inc. Minneapolis, MN
mark@statease.com

Wayne F. Adams, MS
Consultant, Stat-Ease, Inc. Minneapolis, MN

Patrick J. Whitcomb, PE, MS
Principal, Stat-Ease, Inc. Minneapolis, MN

By sizing experiment designs properly, test and evaluation (T&E) engineers can assure they specify a sufficient number of runs to reveal any important effects on the system. For factorial designs laid out in an orthogonal matrix this can be done by calculating statistical power (Anderson and Whitcomb, 2014). However, when a defense system behaves in a nonlinear fashion, then response surface method experiment (RSM) designs must be employed (Anderson and Whitcomb, 2005). The test matrices for RSM generally do not exhibit orthogonality, thus the effect calculations become correlated and degrade the statistical power. This in turn leads to inflation in the number of test runs needed to detect important performance differences that may be generated by the experiment. A generally acceptable alternative to sizing designs makes use of fraction of design space (FDS) plots. This article details the FDS approach and explains why it works best to serve the purpose of RSM experiments done for T&E.

A Prime Case for Why Statistical Power Fails for Sizing RSM Experiments

To illustrate why statistical power will not work for response surface methods, consider the following case detailed in the RSM book by Anderson and Whitcomb (Ibid, p158 "A Real-Life Application for Computer-Generated Optimal Design"). It involves a proprietary mechanical system driven by two air pressures, designated as A and B. A differential pressure must be maintained to keep the air flowing in the proper direction. The operating constraints are detailed below:

- A from 5 to 8.
- B from 7 to 9.5.
- B must always be at least +1 greater A.

The last requirement introduces the following multifactor constraint:

- 1<= B – A.

Aided by software designed for this purpose, the test engineer set up the statistically-optimal design (D criterion) shown in Figure 1. The three extreme combinations were replicated, as indicated by the number "2" by those points.

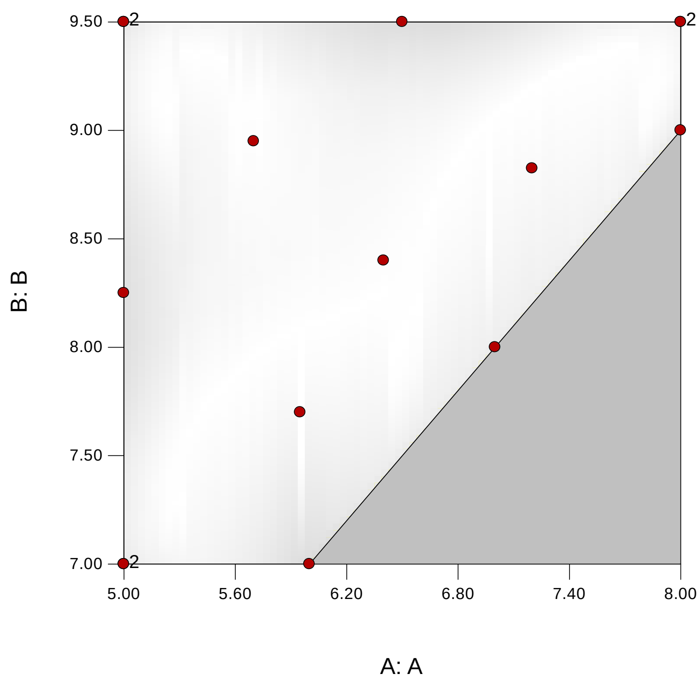

Figure 1: Optimal RSM design on an air-driven mechanical system

Table 1 shows some key measures for statistically evaluating the ability of this test plan to fit the designed-for quadratic polynomial:

- The multiple correlation coefficient (R_i^2), indicating how much the coefficient for each individual term is correlated to the others. (Ideally all these value will be zero, i.e., no correlation, and thus the matrix becomes orthogonal. However, this cannot be counted on for RSM experiments).
- Presuming that the effect of interest will be 1.5 standard deviations, power for the 15-run plan is shown as-is, then doubled to 30 runs and, finally, re-doubled to 60 runs. (Ideally a power of at least 80% will be achieved.)

Term	**R_i^2**	**Power (1.5σ effect) (15 run design)**	**Power (1.5σ effect) (Doubled to 30 runs)**	**Power (1.5σ effect) (Re-doubled to 60 runs)**
A	0.778	15.1 %	28.6 %	52.3 %
B	0.764	15.8 %	30.2 %	55.0 %
AB	0.783	12.0 %	21.5 %	39.3 %
A^2	0.465	39.0 %	72.2 %	95.8 %
B^2	0.313	45.3 %	79.9 %	98.2 %

Table 1: Power calculations

Notice that the terms exhibiting the highest correlations—A, B and AB—provide the least statistical power. Even after a four-time increase (doubling twice over) the power remains inadequate, especially for the important two-factor interaction term AB: it lagging behind at only 39.3% power.

To get AB powered up properly above 80% requires more than 200 test runs. That cannot fly for the costly environment of test and evaluation for defense systems.

It turns out, fortuitously, that power is not the right metric for design sizing when the goal is optimization, as is the case for RSM. Instead, the emphasis should be placed on producing a fitted surface as precisely as needed. For this purpose, the fraction of design space (FDS) plot is the tool to use (Whitcomb, 2008). An FDS above 80% suffices according to generally accepted statistical guidelines. In the case of the air-driven mechanical system, the FDS for a 1.5 standard deviation precision comes to ~95% in the original 15-run test plan. That would be great news to a test engineer who by the mismatched measure of statistical power would be forced into completing 200 runs.

By the way, as you might expect, the more constrained the experimental region becomes, the more difficult it gets to size for power. For example, consider the 15-point optimal design shown in Figure 2 within a more-restricted space, the upper left corner now being cut off in addition to the lower right region.

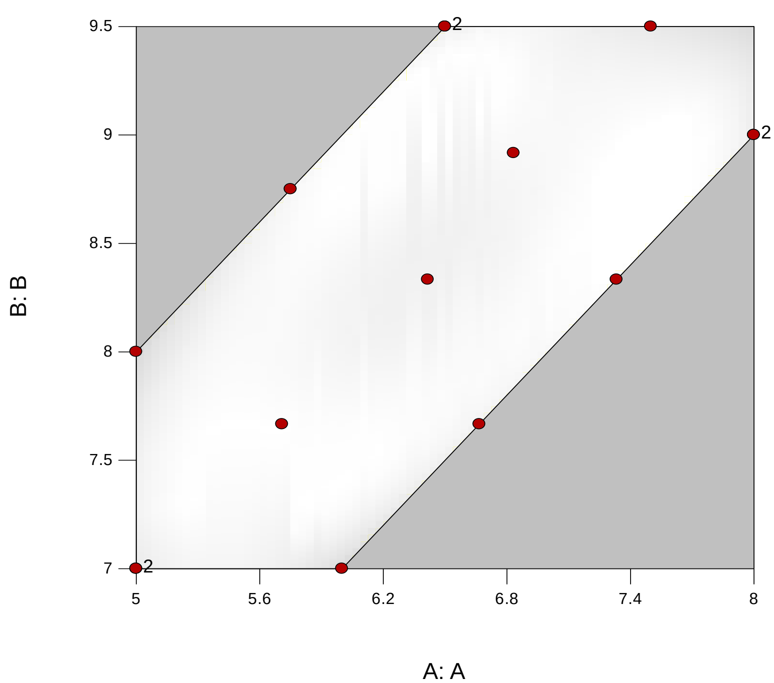

Figure 2: Upper constraint added to design on an air-driven mechanical system

As detailed in Table 2, the power drops substantially for the second-order terms (AB, A^2, B^2) due to this squeeze play on the design space. (Because the ranges tested remain intact at 5 to 8 and 7 to 9.5; respectively, the power on the first-order terms A and B remains undamaged by this change in constraints; in fact they go up slightly, but only as an artifact of the particular point placements.)

Term	R_i^2	Power (1.5σ effect) (15 run design)
A	0.63	18.3 %
B	0.68	18.2 %
AB	0.89	6.2 %
A^2	0.81	15.9 %
B^2	0.70	21.7 %

Table 2: Power calculations for more-restricted experimental region

Amazingly, despite this far greater restriction on the experiment, the FDS for a 1.5 standard deviation precision remains at a very satisfactory level of ~95%, that is, 15 runs holds up for a sound test plan. Sizing for power becomes completely impractical for this restricted region.

Now let's dig a bit into the statistical details on FDS.

Sizing for Precision via Fraction of Design Space (FDS)

To keep things really simple statistically let's consider a hypothetical case where a test engineer is fitting a straight line to response data as a function of one factor. Naturally as more runs are made the more precise this fit becomes. However, due to practical limits on the run budget, it becomes necessary to be realistic in establishing the desired precision "d"—the half-width of the red-lined interval depicted in Figure 3.

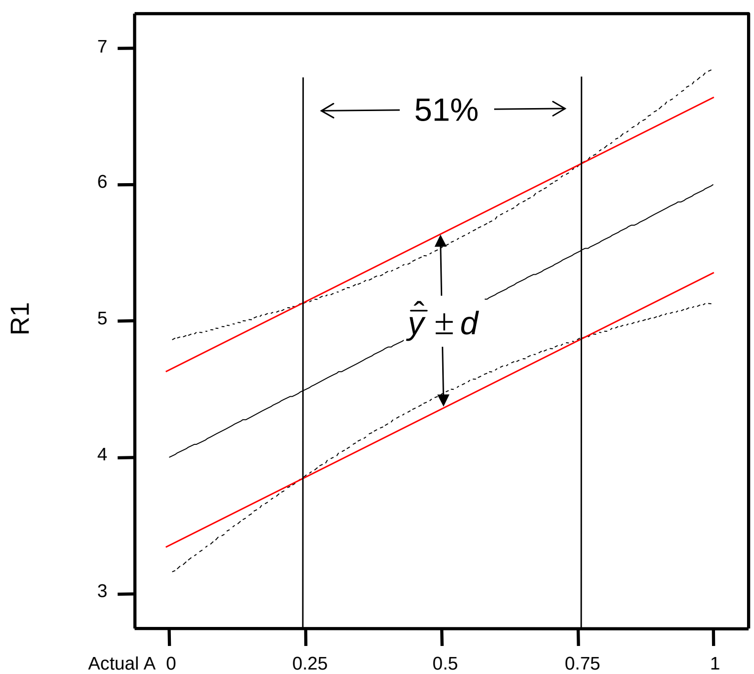

Figure 3: Fraction of design space illustrated on a simple one-factor experiment

This simple case simulates 5 runs—two at each end and one in the middle of the 0 to 1 range—from which a confidence interval (CI) can be calculated. The CI is displayed by the dotted lines that flare out characteristically at the extremes of the factor. Unfortunately only 51% of the experimental region provides mean response predictions (symbolized by the "y" covered by the bar (—) with the hat (^) on top) within the desired precision plus-or-minus d. This is the fraction of design space (FDS), reported as 0.51 on a scale of zero to one. More runs, only a few in this case, are needed to push FDS above the generally acceptable level of 80%.

Now that you have seen what FDS measures for one factor, let's move up one dimension to a two-dimensional view by revisiting the case of the air-driven mechanical part. Figure 4 illustrates an FDS graph with the crosshair set at the level of 0.5 on the standard error (SE) of the mean (y-axis). It tells us that about one-third (FDS=0.34) of the experimental region will provide predictions within this specified SE with 95% confidence (alpha of 0.05).

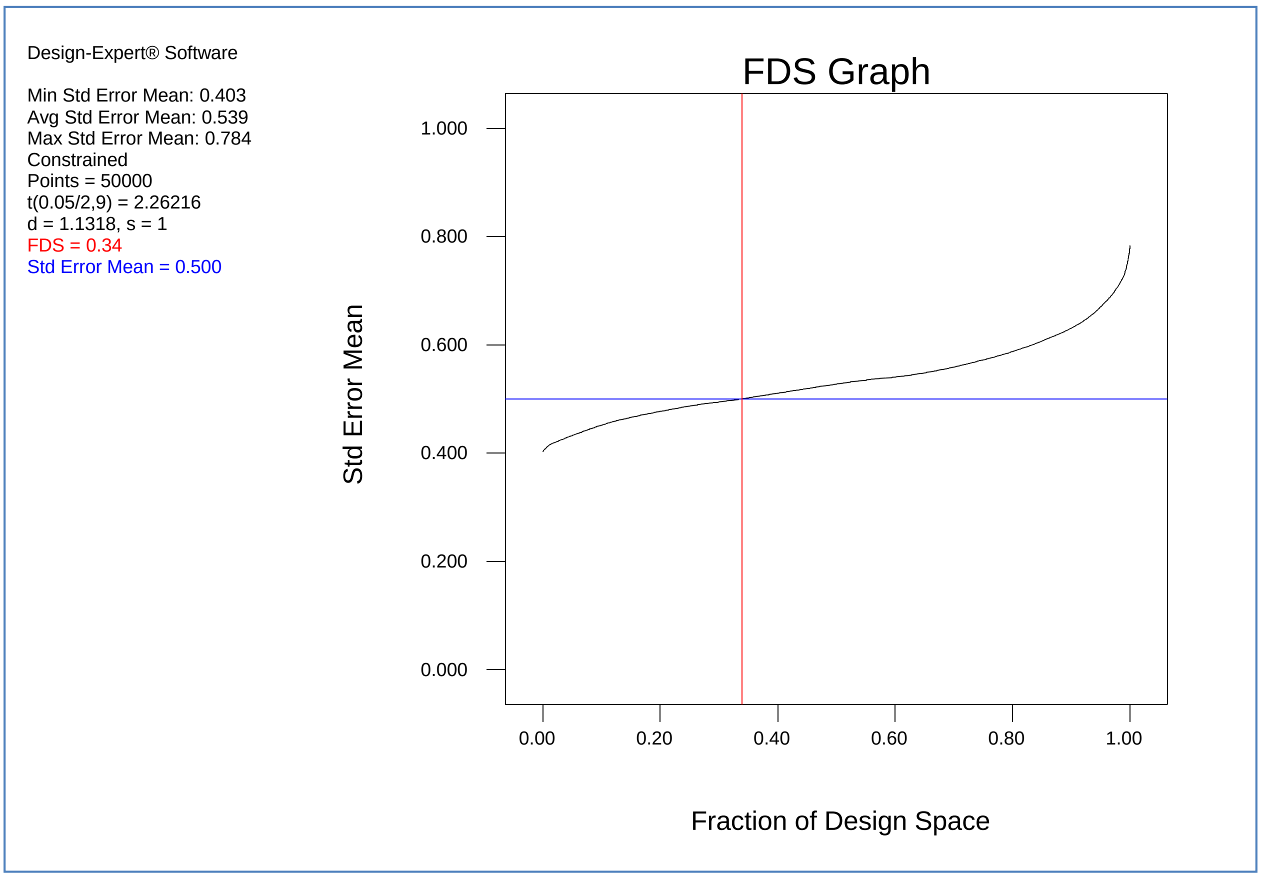

Figure 4: Fraction of design space illustrated for the two-factor experiment

Figure 5, a contour plot of SE, shows which regions fall below the 0.5 level.

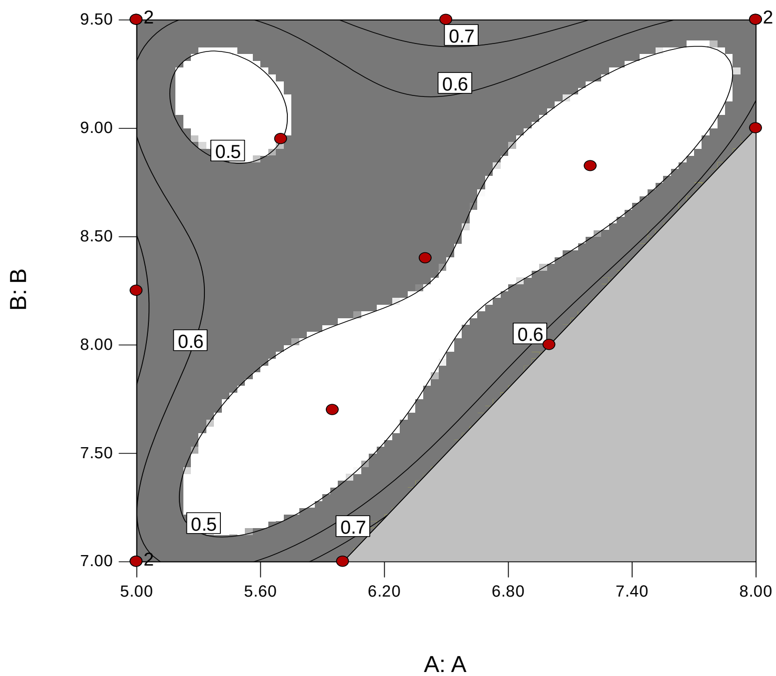

Figure 5: Contour plot of standard error with regions above 0.5 SE shaded out

Notice that only about a third of the triangular experimental region falls within the two areas bordered by the 0.5 SE contours. This corresponds directly to the 0.34 FDS measure. The FDS for 0.6 SE comes to 0.83, in other words 83% of the experimental region can be estimated to within this level of precision. The standard error derived from the 1.5 standard deviation requirement on precision for our case study is 0.663. (To be specific, it equals 1.5 divided by the two-tailed t-value for alpha of 0.05 with N minus p degrees of freedom, where N is the number of runs and p is the number of terms including the intercept.) As displayed by the FDS plot in Figure 6, a healthy 94% of the experiment-design space falls within this SE.

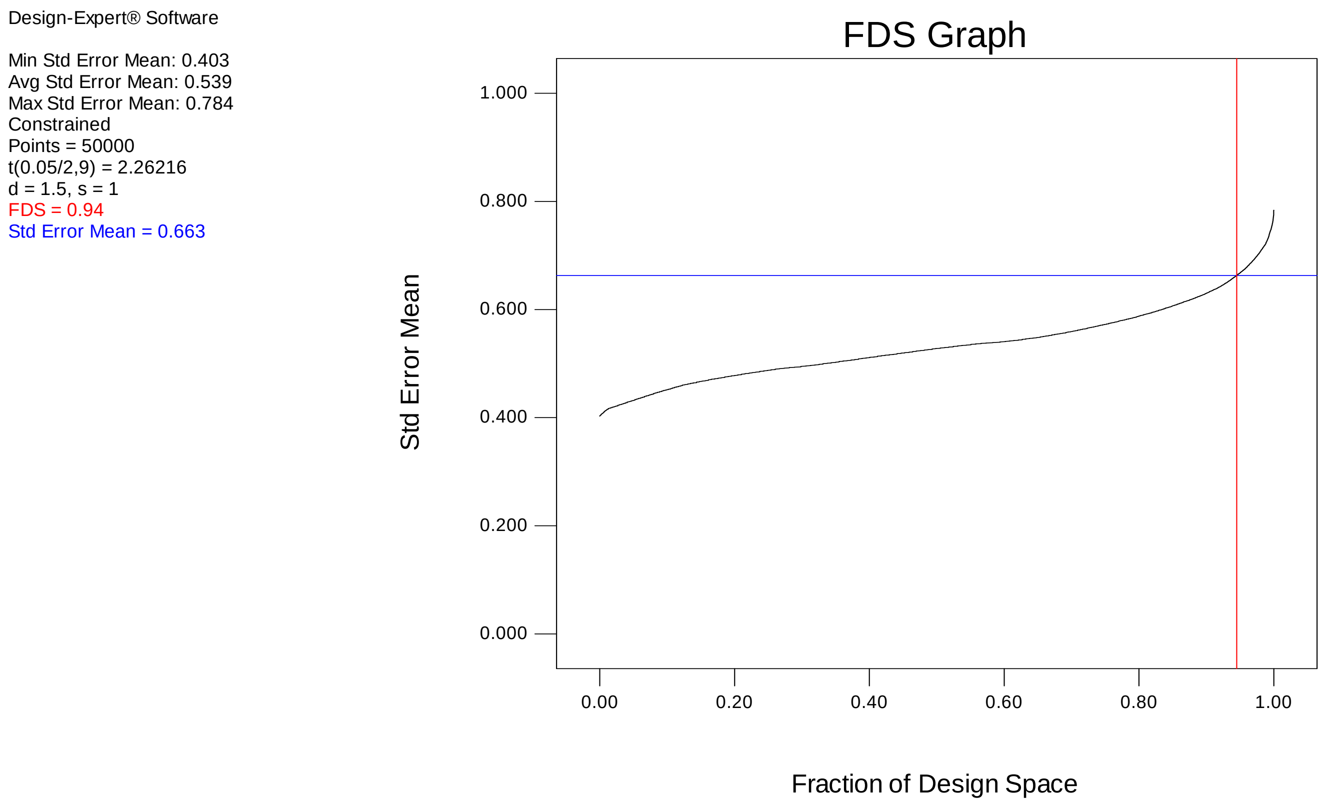

Figure 6: FDS based on actual requirements for RSM experiment on the air-driven mechanical system

This far exceeds the acceptable FDS of 80%. Therefore, all systems are go for the T&E engineer on this 15-run experiment-design that initially came out as a fail on statistical power. Thank goodness for applying the proper statistical tools for sizing, that is, fraction of design space.

Conclusion

For non-orthogonal designs, such as response surface methods (RSM) along the lines of the example in this article, fraction of design space (FDS) plotting works much better and more appropriately for the purpose of optimization than statistical power for 'right-sizing' experiment designs. They generally lead to a smaller number of runs while providing the assurance needed that, whether or not anything emerges significant, the results will be statistically defensible. That is the key to good T&E.

Mark J. Anderson, PE, CQE, MBA is a principal and general manager of Stat-Ease, Inc. He co-authored "DOE Simplified: Practical Tools for Effective Experimentation" and "RSM Simplified: Optimizing Processes Using Response Surface Methods for Design of Experiments" Mark has also published numerous articles on design of experiments (DOE). He is also a guest lecturer at the University of MN Chemical Engineering & Materials Science department and the Ohio State University Fisher College of Business.

Wayne F. Adams, MS is a master statistician who in his career as a consultant with Stat-Ease, Inc has worked extensively to deploy DOE for national defense T&E. His prior work experience included application of statistical techniques for the U.S. Army. Wayne also taught business statistics at Western Michigan University.

Patrick J. Whitcomb, PE, MS is the founding principal and president of Stat-Ease, Inc. Before starting his own business, he worked as a chemical engineer, quality assurance manager, and plant manager. Pat co-authored Design-Ease® software, an easy-to-use program for design of two-level and general factorial experiments and Design-Expert® software, an advanced user's program for response surface, mixture, and combined designs. He's provided consulting on the application of design of experiments (DOE) and other statistical methods for several decades. In addition, Pat is co-author of the books, "DOE Simplified: Practical Tools for Effective Experimentation" and "RSM Simplified: Optimizing Processes Using Response Surface Methods for Design of Experiments," and has published many articles on design of experiments (DOE).

References

Anderson, M. J. and P.J. Whitcomb. 2014. Employing Power to 'Right-Size' Design of Experiments. *ITEA Journal*, 35: 40-44.

Anderson, M. J. and P.J. Whitcomb. 2005. *RSM Simplified, Optimizing Processes Using Response Surface Methods for Design of Experiments*. New York, NY. Productivity Press.

Whitcomb, P.J. September 2008. FDS—A Power Tool for Designers of Optimization Experiments. *Stat-Teaser Newsletter*, Stat-Ease, Inc.