STAT COE-Report-10-2015

Test Facility DOE Survey

Authored by: Francisco Ortiz, Ph.D.

24 April 2015

Revised 9 October 2018

The goal of the STAT COE is to assist in developing rigorous, defensible test strategies to more effectively quantify and characterize system performance and provide information that reduces risk. This and other COE products are available at www.afit.edu/STAT.

STAT Center of Excellence
2950 Hobson Way – Wright-Patterson AFB, OH 45433

Table of Contents

Revision 1, 9 Oct 2018, Formatting and minor typographical/grammatical edits.

Executive Summary

Department of Defense (DoD) instruction 5000.02 has mandated the use of design of experiments (DOE) to create an effective and efficient test program. DOE will be used to characterize system behavior across the operational space and to derive resources estimates to ensure defensible test results. The Scientific Test and Analysis Techniques (STAT) Center of Excellence (COE) has observed that while the use of designed experiments is increasing across test and evaluation (T&E) community there appears to be little focus on ensuring the correct experimental protocol and tenets are followed at facilities and test ranges. This can be due to a lack of understanding in the T&E community as to why these tenets are important. Another reason may be that, while test facilities and ranges boast great technology and capabilities to assess acquisition system performance, many were not created with tenets of DOE in mind. Therefore, while these tenets are simple in concept they can be daunting to implement on real tests on actual ranges. This paper will cover some of the key DOE tenets and why they are important followed by a survey that can be used to assess how well suited a facility is to correctly and effectively implement STAT/DOE.

Keywords: Randomization, replication, blocking, sequential testing, execution

Background

The goal of all designed experiments is to quantify the relationship between a set of input factors (i.e. system's technical, environmental and mission inputs) and a set of responses (the performance measures of interest). The experimental protocol aims to quantify this relationship as **clearly** and as **efficiently** as possible. Some key execution aspects of the experimental protocol are randomization, replication, blocking and sequential experimentation.

Randomization refers to the random order in which test runs are executed. Failing to randomize a test properly can lead to incorrect conclusions about which input factors are important and how much of an effect they have on the responses. When randomization is very difficult, or inefficient to implement, split plot designs may be employed (Anderson-Cook, 2007).

Replication refers to repeated test runs at a fixed set of input conditions within a test design. Replication allows us to estimate the experimental error and also provides better precision for estimating the effects/influence of input factors.

Blocking is a variance reduction technique used for dealing with nuisance/noise factors. These are factors that may exert some influence on the response being measured but are not of interest in the study. The Defense Acquisition University (DAU) glossary defines blocking as: *A planning and analysis technique that sets aside (blocks out) undesirable and known sources of nuisance variability so they do not mix with the estimation of other factors. Blocking is done when there is an expectation the nuisance variation may interfere with your experimental results. Examples of blocking schemes include grouping* *test points based on test schedule, equipment, test locations, or operators. Blocking reduces experimental error, increasing the power to detect impact of test conditions on system performance.*

Failure to follow proper protocol reduces the validity of experimental results. For a more detailed discussion on these tenets see Dr. Vining's paper on test protocols (2013).

Sequential experimentation is another aspect of the experimental protocol that should be considered when planning tests. Ideally, the most efficient way to obtain knowledge of a system is via the process of sequential experimentation, also known as the "test-analyze-test" process. This approach subdivides tests into multiple stages where analysis is conducted at the end of each stage to help influence the following stages. By using this sequential approach, insignificant factors can be identified quickly and therefore removed from consideration in subsequent tests. This reduces the dimension of the design space and can potentially reduce the total number of test runs needed in the latter stages. It is recommended that the initial set of tests consist of 25% of the run budget available (Montgomery, 2017). The last stage of tests is used to validate the statistical models created from prior test stages, which describe the systems performance across the operational space. More details on using sequential experimentation in a DoD environment are examined in Simpson's 2014 case study. In the following section we will present a survey that programs can use to assess how well suited a facility is to correctly and effectively implement STAT/DOE.

Survey with Scores

The survey below was created to aid both program and test facilities in assessing how "friendly" or well suited a venue is to employ DOE. Programs may find the survey useful for comparing potential venues under consideration for future tests and use it to help develop their testing strategy. Test facilities can use the survey to identify deficiencies or areas for improvement in order to make better use of DOE. Each survey question includes a brief explanation of its purpose along with multiple choice answers. No grading scale is presented; scores for each answer are to be used for comparative purposes only.

Questions	Score
1. How easy is it to change factor settings between test points?	
a. Automated - Changing any inputs into the system is done with no human interaction and points can be run in any order without an increasing execution time.	20
b. Little human intervention - Changing any inputs requires some human intervention but does not significantly increase execution time.	15
c. Some factors (1-2) are hard but not impossible to change – Altering hard to change factors resulting in a significant increase in execution time. Due to this a split plot design is usually considered.	10
d. Several factors (>2) are hard but not impossible to change.	5
e. Impossible – Some factors are impossible or so hard to change that they would require such a significant increase in test time as to make the test infeasible.	0
Reason: *Failing to randomize a test properly can lead to wrong conclusions about which factors are important. When randomization is very difficult, inefficient, or unsafe, then split plot designs may be employed.*	
2. How quickly is range data available for analysis?	
a. Real-time – Data is collected in such a way that would allow an analyst the ability to process data and report results in near real-time. This would allow for augmenting or modifying upcoming test designs. It was also allow for stopping testing early if test results are sufficient.	20
b. Some processing of data needed – Not exactly real-time but an analyst can augment the design before the end of testing.	15
c. Data can be analyzed between blocks- Some time for data processing is needed but it is fast enough to support analysis before starting a new block.	10
d. Entire test matrix must be run - There is time allocated afterwards to return to facility and run checkpoint/validation runs.	5
e. Data not available until entire test matrix is completed - There is **no** time to return to facility and run checkpoint/validation runs.	0
Reason: *Ideally, we want to employ sequential experimentation. A test-analyze-test process.*	
3. Do you have historic data on the variation of performance measures of similar systems?	
a. Data from past tests are readily available to program and can be used to assess system variation and measurement accuracy.	20
b. Data from past tests available but with limited application.	15
c. Data from past tests available but are not directly applicable.	10
d. Limited data available but are not be directly applicable.	5
e. Previous data not available.	0
Reason: *This would aid test planners to estimate the signal-to-noise ratio so they can more accurately determine an appropriate test size for the desired power.*	

Questions			Score
4.	**What proportion of range personnel has received some STAT training?**		
	a.	~100% - All personnel that are a part of planning, designing, executing, and analyzing designed experiments have taken some DOE courses.	20
	b.	~ 75% of personnel.	15
	c.	~ 50% of personnel.	10
	d.	~ 25% of personnel.	5
	e.	~0% - No personnel that are a part of planning, designing, executing, and analyzing designed experiments have taken some DOE courses.	0
***Reason:** It is important that personnel have an understanding of the tenets of DOE and why any deviation from them needs to documented and discussed.*			
5.	**How well are nuisance factors controlled?**		
	a.	All nuisance factors are completely controlled.	20
	b.	Nuisance factors are not controlled. They are blocked and monitored.	15
	c.	Nuisance factors are not controlled. They are blocked, but not monitored.	10
	d.	Nuisance factors are not controlled. They cannot be blocked, but are monitored.	5
	e.	Nuisance factors are not controlled and cannot be blocked or monitored.	0
***Reason:** Blocking is done when there is an expectation the nuisance variation may interfere with your experimental results. Monitoring/recording nuisance factor values could aid with analysis.*			

Conclusion

The purpose of this survey is to lead facilities and programs to seek ways to improve their test venues and make them more conducive for effective DOE. Testers should look for ways to perform sequential experimentation and thus afforded more than a single test session to conduct their full matrix of tests. Programs and testers should seek out ways to improve data acquisition methods that facilitate quick data turnaround in support of system performance analysis. Doing so not only improves the fidelity of analysis, and thusly decisions making, but it may reduce the resources needed to complete successful testing.

References

Anderson-Cook, Christine. “When Should You Consider a Split-Plot Design?” *Quality Progress*, Oct. 2007, pp. 57-59

Montgomery, Douglas C. *Design and Analysis of Experiments*. 9th ed., John Wiley & Sons, Inc., 2017.

Simpson, James. “Testing via Sequential Experiments.” Scientific Test and Analysis Techniques Center of Excellence (STAT COE), 2014.

Vining, Geoff. “Technical Advice: Experimental Protocol and the Basic Principles of Experimental Design.” Quality Engineering, vol. 25, no. 3, 2013, pp. 307–311., doi:10.1080/08982112.2013.789278.

Appendix: Survey

Questions
1. **How easy is it to change factor settings between test points?**
a. Automated - Changing any inputs into the system is done with no human interaction and points can be run in any order without an increasing execution time.
b. Little human intervention - Changing any inputs requires some human intervention but does not significantly increase execution time.
c. Some factors (1-2) are hard but not impossible to change – Altering hard to change factors resulting in a significant increase in execution time. Due to this a split plot design is usually considered.
d. Several factors (>2) are hard but not impossible to change
e. Impossible – Some factors are impossible or so hard to change that they would require such significant increases in test time as to make the test infeasible.
2. **How quickly is range data available for analysis?**
a. Real-time – Data is collected in such a way that would allow an analyst the ability to process data and report results in near real-time. This would allow for augmenting or modifying upcoming test designs. It was also allow for stopping testing early if test results are sufficient.
b. Some processing of data needed – Not exactly real-time but an analyst can augment the design before the end of testing.
c. Data can be analyzed between blocks- Some time for data processing is needed but it is fast enough to support analysis before starting a new block.
d. Entire test matrix must be run - There is time allocated afterwards to return to facility and run checkpoint/validation runs.
e. Data not available until entire test matrix is completed - There is **no** time to return to facility and run checkpoint/validation runs.
3. **Do you have historic data on variation of performance measures of similar systems?**
a. Data from past tests are readily available to program and can be used to assess system variation and measurement accuracy.
a. Data from past tests available but with limited application.
b. Data from past tests available but are not directly applicable.
c. Limited data available and but are not be directly applicable.
d. Previous data not available.
4. **What proportion of range personnel has received some STAT training?**
b. ~100% - All personnel that are a part of planning, designing, executing, and analyzing designed experiments have taken some DOE courses.
a. ~ 75% of personnel.
b. ~ 50% of personnel.
c. ~ 25% of personnel.
d. ~0% - No personnel that are a part of planning, designing, executing, and analyzing designed experiments have taken some DOE courses.
5. **How well are nuisance factors controlled?**
c. All nuisance factors are completely controlled.
d. Nuisance factors are not controlled. They are blocked and monitored.
e. Nuisance factors are not controlled. They are blocked, but not monitored.
f. Nuisance factors are not controlled. They cannot be blocked, but are monitored.
g. Nuisance factors are not controlled and cannot be blocked or monitored.