Achieving more statistical power using orthogonal designs

Kedar M. Phadke (kedar@phadkeassociates.com)
Madhav S. Phadke, PhD (madhav@phadkeassociates.com)
Phadke Associates, Inc.

DoD Memorandum & Publications

- **Clear, visionary approach.**
- **Focus on improving fundamental processes to impact quality of testing at each system phase.**

Design of Experiments (DOE) in Test and Evaluation

At the request of the Service Operational Test Agency
hosted a meeting of OTA technical and executive age
common approach to utilizing DOE in operational test
Representatives from ATEC, OPTEVFOR, AFOTEC,
from the National Institute of Standards and Technolo
applicability of DOE principles to support test and eva

This group endorses the use of DOE as a discipline t
analysis, and reporting of integrated testing. DOE offe
approach to test and evaluation. DOE is appropriate f
when applied in a testing program,…

Dr. Charles E. McQueary **Director, Operational Test & Evaluation**	**David L. Reeves, Co** **USMC** **Director, MCOTEA**
Roger A. Nadeau, Major General, USA **Commander, ATEC**	**Stephen T. Sargeant** **General, USAF** **Commander, AFOTE**

DEPARTMENT OF DEFENSE

SCIENTIFIC TEST AND ANALYSIS TECHNIQUES IN TEST AND EVALUATION

IMPLEMENTATION PLAN

DOT&E GUIDANCE PRE – JUNE 28, 2013

Guidance Overview

- **Focus on the engineering operational requirements**
  - **Test parameters (“Factors”)**
  - **The ranges or values of the parameters (“Levels”)**
  - **The outputs and output requirements (“Responses”)**
    - **KPP’s, TPM’s, etc**
    - **Acceptable boundaries, confidence level**
- **Utilize statistical power and confidence as analytical measures of “goodness of test”**
  - **Single hypothesis framework**

DOT&E Guidance (pre June 28, 2013)

Statistical measures of merit (power and confidence) on the relevant response variables for which it makes sense. These statistical measures are important to understand "how much testing is enough?" and can be evaluated by decision-makers on a quantitative basis so they can trade off test resources for desired confidence in results.

- **Reference: *Guidance on the use of Design of Experiments (DOE) in Operational Test and Evaluation* 2010 Oct**
  - **http://www.dote.osd.mil/pub/reports/20101019GuidanceonuseofDOEinOT&E.pdf**

What data you need to compute statistical test power

- **Null Hypothesis**
- **Alternate Hypothesis**
- **α (acceptable type 1 error rate)**
- **Effect size to detect (difference between null and alternate)**
- **Distribution type of KPP's**
- **Std Dev of an observation (Standard error σ_s)**
- **Number of samples**

Issues of Single Hypothesis Framework

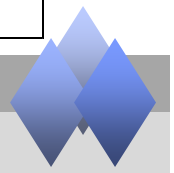
- **Distribution of tests is unspecified.**
- **Test case selection can be faulty**

		Threat Location			
		Within Range with Natural Barriers	Outside of range with natural barriers	Outside of range with natural barriers	within range with manmade barriers
50 Cal threats	1 present	X	X	X	X
	2 present				
	4 present	X			

		Threat Location			
		Within Range with Natural Barriers	Outside of range with natural barriers	Outside of range with natural barriers	within range with manmade barriers
ats	1 present	X X X		X X X	
					X X X

Mission Assurance implications are very different

They score similarly based on old guidelines



DOT&E GUIDANCE POST – JUNE 28, 2013

Guidance to use multiple hypothesis tests

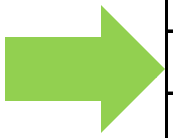
- **"In DOE we are interested multiple hypothesis tests, one for each model term considered." DOT&E July 23, 2013 Memo**
- **This means we need the following information for each test factor and for each output**
  - **Null Hypothesis**
  - **Alternate Hypothesis**
  - **α (acceptable type 1 error rate)**
  - **Effect size to detect (difference between null and alternate)**
  - **Distribution type of KPP's**
  - **Std Dev of an observation (Standard error σ_s)**
  - **Number of samples**
- **Do programs specify this by the parameter?**

New Metrics for "Goodness of Test"

- **Correlation (aka Pearson Correlation)**
  - **Describes degree of linear relationship between individual factors. 0 is the ideal value.**
- **Variance Inflation Factor**
  - **A one number summary describing collinearity with other factors in the model. 1 is the ideal value.**
- **Scaled Prediction Variance**
  - **Variance of prediction model at a specified location in the design space.**
- **Reference: "DOT&E 7-23-13 Best Practices for Assessing the Statistical Adequacy of Experimental Designs Used in Operational Test and Evaluation (6866)"**

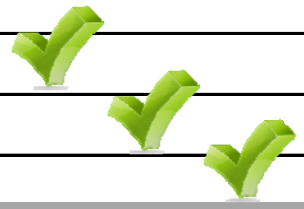
What type of DOE designs score better? Same factors, same levels, one orthogonal and other correlated

A	B	C
A1	B1	C1
A1	B1	C1
A1	B1	C1
A2	B2	C2
A2	B2	C2
A2	B2	C2
A3	B3	C3
A3	B3	C3
A3	B3	C3

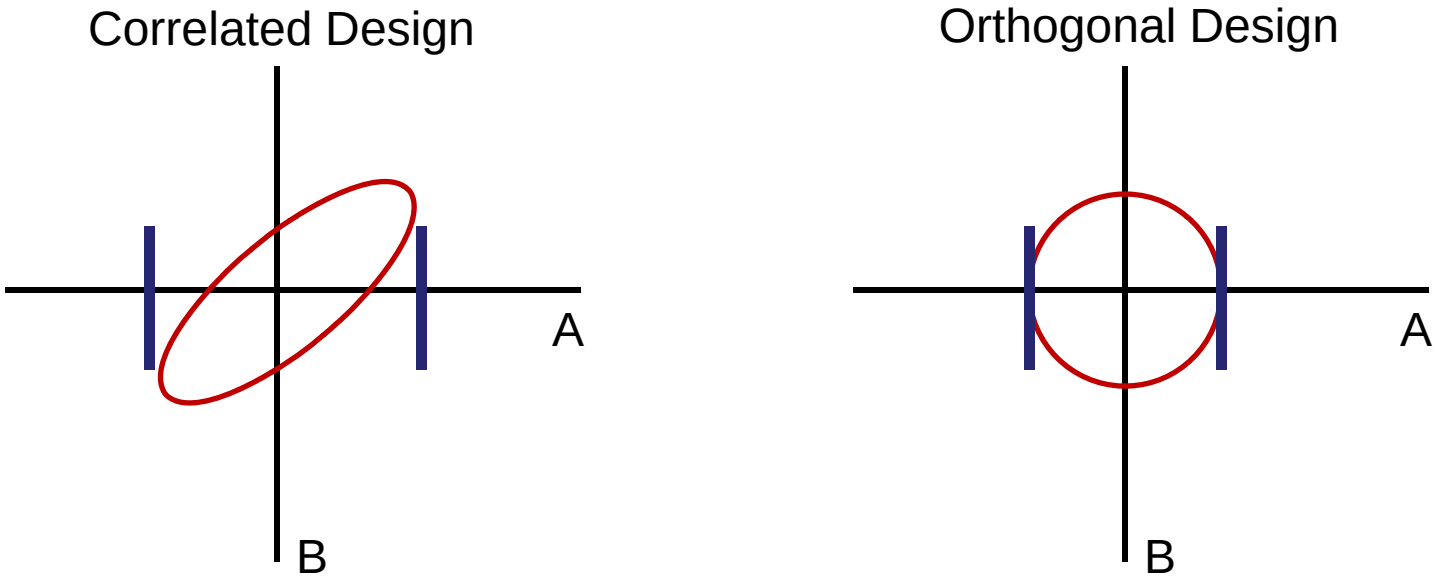


A	B	C
A1	B1	C1
A1	B2	C2
A1	B3	C3
A2	B1	C2
A2	B2	C3
A2	B3	C1
A3	B1	C3
A3	B2	C1
A3	B3	C2

Design Type	Correlated	Orthogonal
Power		
Confidence		
Correlation		
VIF		
SPV		



What about power and confidence? Closer look at effects of correlation on standard error



- **Correlations cause an increase in the standard error for estimate of effects**
- **Many common designs including d-optimal, i-optimal, space filling, and n-way can be correlated.**

What does an increase in standard error mean for test power?

- **"Power is a function of the statistical confidence level, the effect size of interest, the variability in the outcomes [standard error], and the number of tests." DOT&E 7-23-2013 Memo**

Design Type	Correlated	Orthogonal
Confidence Level	=	=
Effect Size	=	=
Standard Error	LARGER	 More Power
Number of Tests	=	=

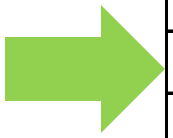
What can I do to increase the power of a correlated design?

Design Type	Correlated	Orthogonal
Confidence Level	=	=
Effect Size	=	=
Standard Error	=	=
Number of Tests	LARGER	SMALLER

If your design is correlated, you need more tests to get the same power as an orthogonal design.

What type of DOE designs score better? Same factors, same levels, one orthogonal and other correlated

A	B	C
A1	B1	C1
A1	B1	C1
A1	B1	C1
A2	B2	C2
A2	B2	C2
A2	B2	C2
A3	B3	C3
A3	B3	C3
A3	B3	C3



A	B	C
A1	B1	C1
A1	B2	C2
A1	B3	C3
A2	B1	C2
A2	B2	C3
A2	B3	C1
A3	B1	C3
A3	B2	C1
A3	B3	C2

Design Type	Correlated	Orthogonal
Power		✓
Confidence		✓
Correlation		✓
VIF		✓
SPV		✓

Orthogonal Designs score best for all DOT&E metrics

How to score best with DOT&E guidelines

- **Use an orthogonal design to score better for all DOT&E metrics**
  - **Correlation**
  - **Variance Inflation Factor**
  - **Scaled Prediction Variance**
  - **Power : Get more Power per Test**
  - **Confidence : Get more Confidence per Test**
- **If you need to improve test precision, be cautious while using designs such as d-optimal, i-optimal, space filling, n-way, and others that can correlate effects.**

Facts regarding Orthogonal Designs

- **Orthogonal Designs have been widely used in Design of Experiments for decades.**
- **Fractional Factorial and Factorial Designs are orthogonal designs**

Misconception

However, it should be noted that Taguchi designs [orthogonal designs]] … are inappropriate for characterization because they provide low power for detecting differences in performance across the operational envelope.
- DOT&E memo July 23, 2014

Conclusion

- **Within the same budget, you can better meet DOT&E Design of Experiments objectives by using orthogonal designs.**
  - **More statistical power per test!**
- **To maximize test power and confidence, be cautious while using designs such as d-optimal, i-optimal, space filling, n-way, etc. that can correlate effects**

About the presenters

- **Kedar Phadke is Vice President of Phadke Associates, a global consultancy and software company specializing in statistical tools for improving testing and design productivity. Kedar has led numerous deployments for improving test and design effectiveness. He has a MS in Statistics, MS in Management, and a BS in Economics from the Wharton School, University of Pennsylvania.**
- **Dr. Madhav S. Phadke is the Founder and President of Phadke Associates, Inc. He is an ASQ Fellow and the author of the first engineering textbook on Robust Design Methods in the US, "Quality Engineering Using Robust Design". He is a recipient of the Technological Innovation Award from IEEE Region 1. He holds a PhD in Mechanical Engineering and MS in Statistics from the University of Wisconsin – Madison, MS in Aerospace Engineering from the University of Rochester, and a BTech in Mechanical Engineering from the Indian Institute of Technology – Mumbai. Prior to founding Phadke Associates, Dr. Phadke was a manager in AT&T Bell Labs, a visiting scientist at the IBM Watson Research Center, and a Research Associate in Statistics Department and the Army Math Research Center at the University of Wisconsin – Madison.**

Contact info

- **Phone: 908-770-7400**
- **Email: kedar@phadkeassociates.com**
- **Web: www.phadkeassociates.com**