What's the Point of Design of Experiments?

Jason Martin

Test Design and Analysis Lead

Aviation and Missile Research, Development and Engineering Center

charles.j.martin1.civ@mail.mil

Phillip A. Farrington, Ph.D

SME Program Manager

Tri-Vector Services

phil.farrington@trivector.us

Professor Emeritus

The University of Alabama in Huntsville

SRE – RAM Training Summit 10 – Huntsville, Alabama - November 8-9, 2017

Purpose

- Explain the practical benefits of using Design of Experiments (DOE)
  - Quantitatively justify the amount of tests to be done (cost)
  - Quantitatively assess our chances of making incorrect conclusions when analyzing data (risk)
  - Quantitatively assess our ability to understand the effects of variables in our tests (information)
- At some point, most of us are required to trade cost, risk, and information
- Properly implementing DOE allows us to make this trade effectively
- Please do not expect
  - Silver bullet solutions for planning experimentation/testing
  - Simple solutions (our systems are not simple)
- Myths we will address
  - DOE is not flexible enough for my situation
  - Using DOE eliminates the need for system knowledge & engineering expertise

Types of Testing and Experimentation

- We will use the term “testing” very broadly for this discussion
- Testing: Changing something to understand the effect it has
- Examples
  - **Running digital or hardware-in-the-loop simulation to characterize how a component or system performs**
  - **Changing production parameters to assess their effect on product quality**
  - **Developmental or operational testing to assess system performance and to provide data for simulation validation**
  - **Gage repeatability and reproducibility studies**
  - Component level test to determine if a piece of hardware functions as expected or required
  - Changing system performance characteristics to understand the effect of specification changes in AoA or in trade studies
  - Changing training procedures to determine what method is most effective
  - Changing combinations of cockpit instruments and displays to determine their effect on pilot situational awareness
  - Testing software to identify bugs

Can DOE Be Useful for Me?

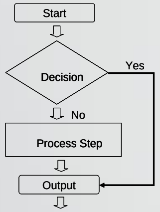
- DOE is not very useful if you are just trying to demonstrate something CAN work
  - This is different from WILL work
  - Single shot demonstration or proof of concept
- DOE is useful (I would say “is necessary”) if you find yourself answering (characterization) questions like:
  - Does this thing do what it is supposed to consistently across the requirements space? If not, where does it fail?
  - If I make these changes will things get better, worse or stay the same?
  - How closely does my simulation represent reality?
  - How much should I budget for experimentation/testing?
  - How much risk of making an incorrect conclusion am I taking?
  - If my budget is cut, how should I adjust my test plan?
  - What will make my software crash?
  - Is my measurement system adequate?

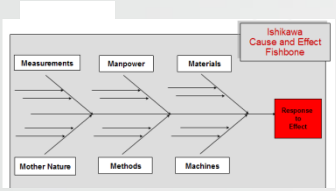
Planning Experiments

- To plan an effective experiment, several questions must be answered
  - What is the objective of the experiment? (i.e., What do I want to learn?)
  - What is the output variable(s) (response) you want to investigate and how will you measure it?
    - Do you want to response to be higher, lower, or do you want to hit a particular target value?
  - What input variables (factors) do you want to include in testing?
  - How many levels of those variables do you want to investigate?
- All tests/experiments are designed, but a good experimental design should also be able to address these questions
  - How likely am I to detect if an important variable matters?
  - How likely am I to think something is important when it isn't?
  - How many test scenarios and test assets will I need?
  - What scenarios should be included in the test plan?
- DOE is now being used to a greater extent than in the past to answer these questions.

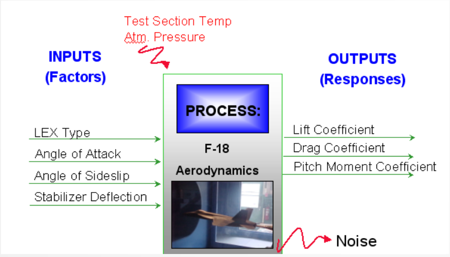
Design of Experiments Roadmap

Planning: Factors
Desirable and Nuisance

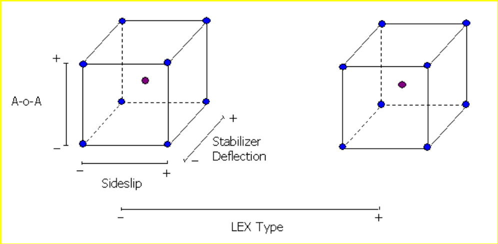




Desired Factors
and Responses



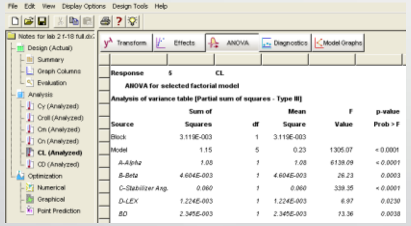
Design Points



Execute
Test Matrix

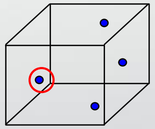
A-o-A	Sideslip	Stabilizer	LEX Type
2	0	5	-1
10	0	-5	1
10	8	5	-1
2	8	5	-1
2	8	-5	-1
2	0	-5	-1
10	8	-5	1
2	0	5	1
2	8	5	1
10	8	5	1
10	8	-5	-1
10	0	5	-1
10	0	-5	-1
2	8	-5	1
10	0	5	1
2	0	-5	1

Results and Analysis



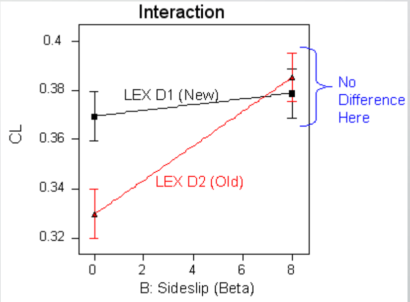
C_L = +0.38
+0.26 x A-o-A
+0.017 x Sideslip
+0.061 x Stabilizer Deflection
-.00875 x LEX Type
+0.012 x Sideslip x LEX Type

Validate the Model



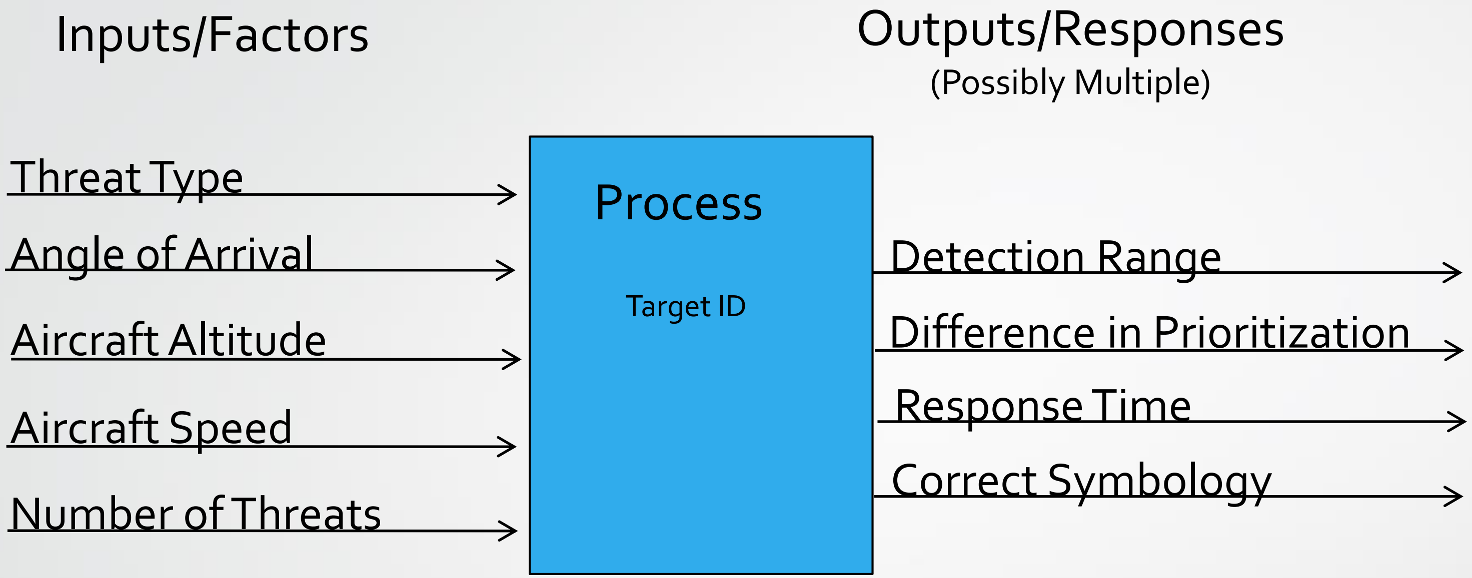
Actual	Predicted	Valid
0.315	(0.30 , .33)	✓

Discovery, Prediction



Source: Air Force DOE for Leaders Class Slides

Terminology



Y: Output, dependent variable, response variable

X: Input, independent variable (something we would like to measure and specify in an experiment)

Level: Unique values for the factor (X) that will be used in the experiment

Run or Scenario: In the experiment/test matrix, a combination of the levels of the Xs

Replication: A repeat of the experiment (if we do an experiment twice, we have two replications)

Effect: The difference in output (Y) when input X is changed

Interaction: When the effect of one input variable depends on the value of another input variable

The Basics:
What’s The Point?

What is Design Of Experiments (DOE)?

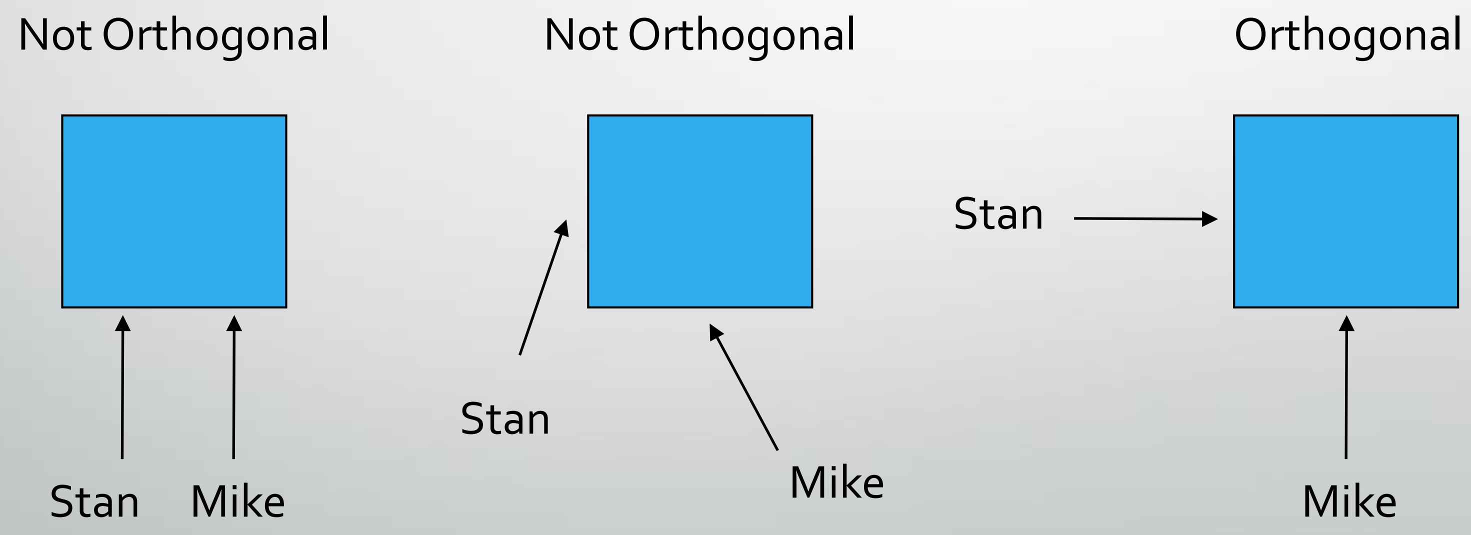
- "Every experiment is a designed experiment, some are poorly designed, some are well designed" – G.E.P. Box
- DOE generally refers to structured methods of developing experiments. It includes far more than the factorial or fractional factorial experiments that are taught in introductory DOE classes
  - Factorial experiments are very useful
  - Other designs are available when factorial designs do not meet our needs
  - Factor covering uses recently developed algorithms to efficiently test n-level combinations of variables. Good for interoperability and software testing.

Why use DOE?

- DOE is used to understand the effects of (usually multiple) input variables on an output or outputs
- DOE provides the maximum amount of information for a given number of tests/assets
- DOE allows us to evaluate effects of relationships between input variables (interactions)
- With DOE we can assess our likelihood of detecting significant variables (known as power)
- **DOE allows us to most efficiently assess performance over our requirements space**
- DOE assists in making intelligent trades between cost and information gained

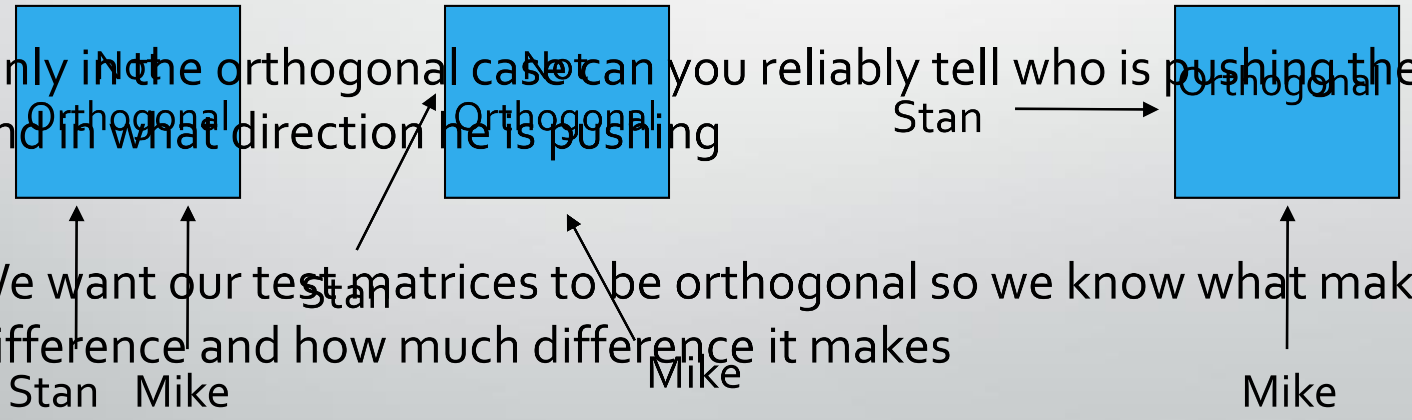
What Does DOE Do For Us?

- You will sometimes hear the term “orthogonal design”
- Orthogonality in test designs allows us to separate effects of factors
- DOE provides orthogonal (or nearly orthogonal) test designs
- One definition of orthogonal says that if two objects are orthogonal they are at 90 degree (normal) angles to each other



Why Be **Normal** (Orthogonal)?

Can you tell who is pushing these boxes?

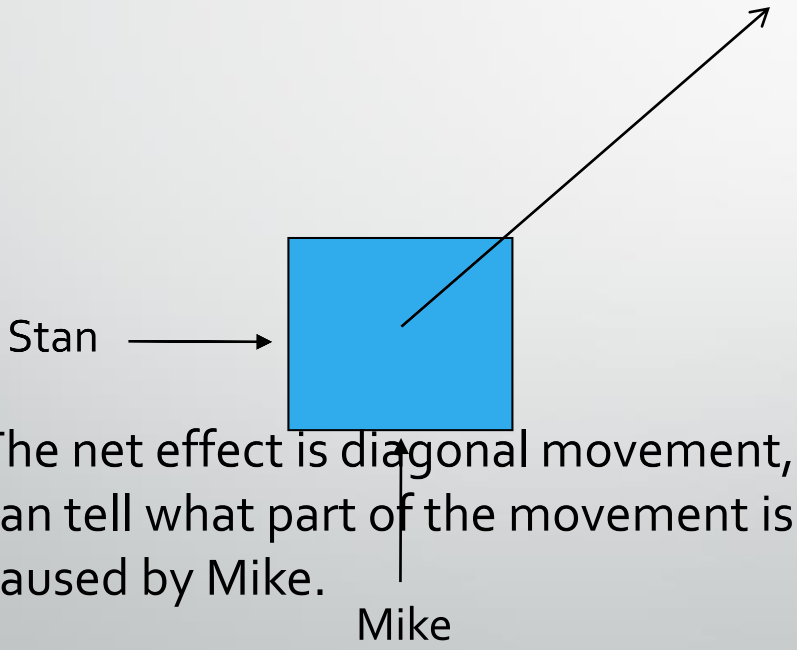


Only in the orthogonal case can you reliably tell who is pushing the box and in what direction he is pushing

We want our test matrices to be orthogonal so we know what makes a difference and how much difference it makes

What if both factors matter?

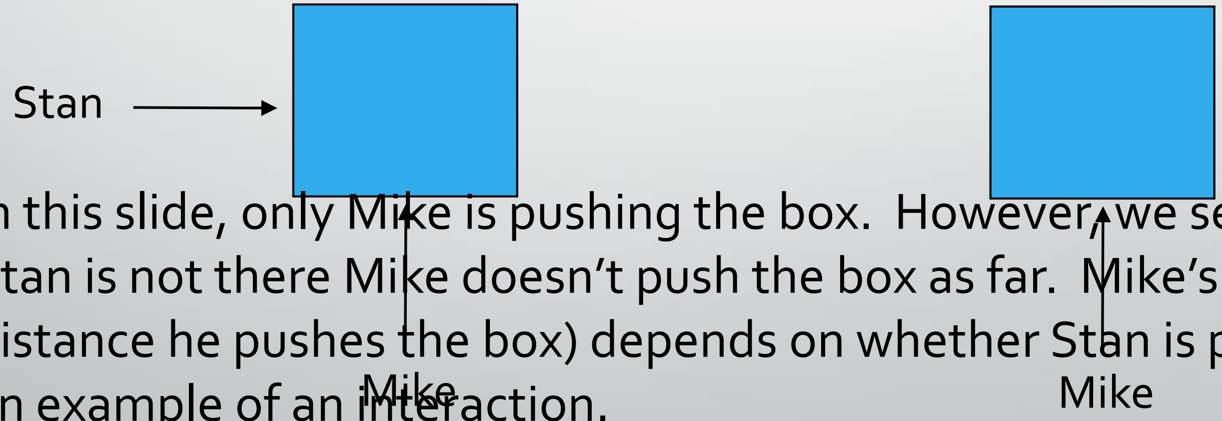
So what if Stan is also pushing the box?



The net effect is diagonal movement, but because of orthogonality we can tell what part of the movement is caused by Stan and what part is caused by Mike.

So What's An Interaction?

An interaction exists when the effect of one factor is dependent on the level of one or more other factors.



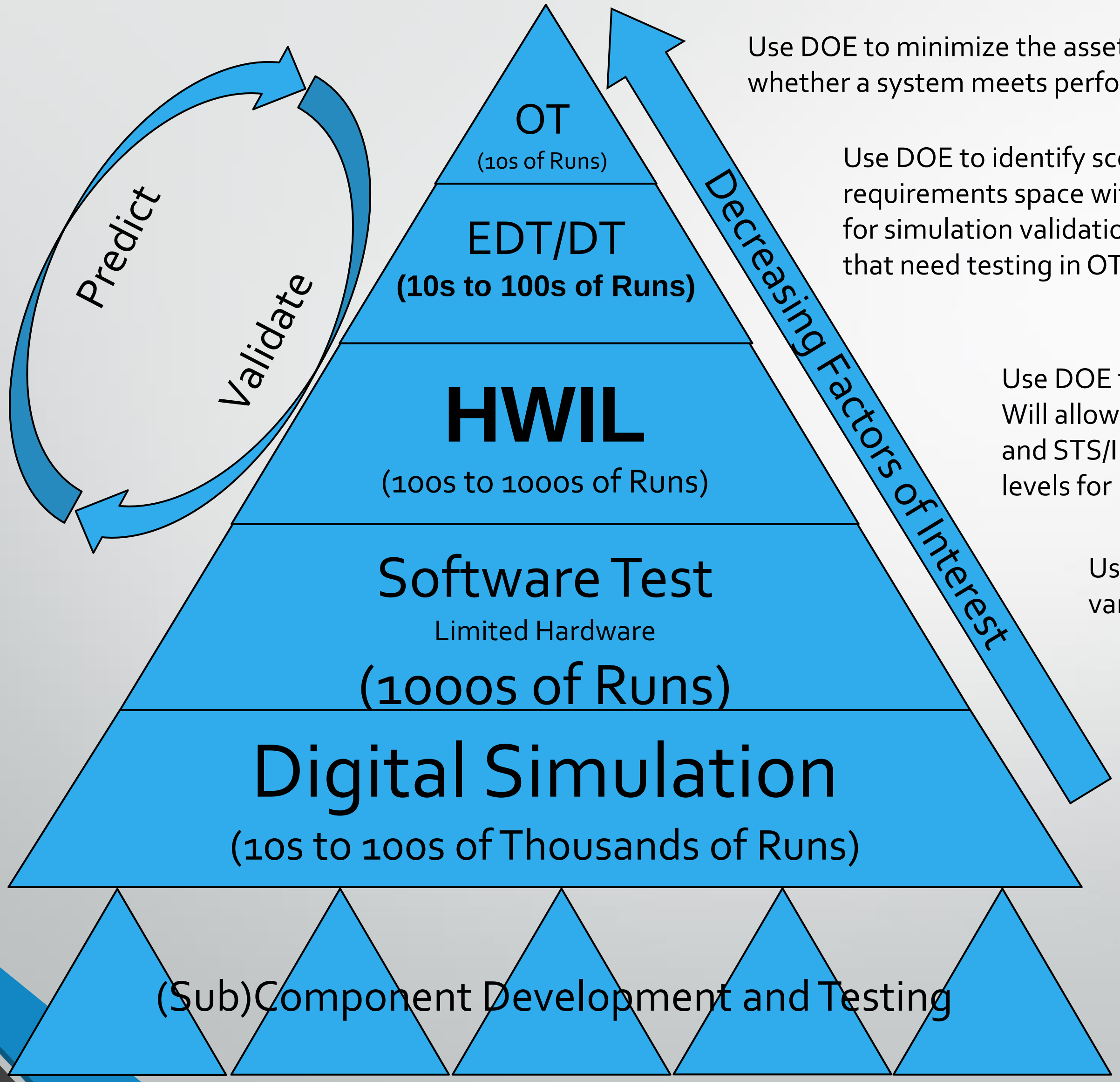
In this slide, only Mike is pushing the box. However, we see that when Stan is not there Mike doesn't push the box as far. Mike's effect (the distance he pushes the box) depends on whether Stan is present. This is an example of an interaction.

When to Use DOE

- **Screening designs**
  - Factorial and fractional factorial designs
  - Used to determine what main effects (no interactions) have a significant impact
- **Process optimization**
  - Could be used in modeling or hardware
  - Useful when nonlinear responses are suspected
- **Performance characterization**
  - Several types of design depending on circumstances: Factorial, fractional factorial, optimal, response surface
  - Hardware performance or using simulation to determine sensitivities
- **Combinatorial testing or space filling designs for debugging software or for interoperability testing**
  - All combination of variables/levels up to n-levels
  - Example: All 4-way combinations of variable levels

From Scientific Test and Evaluation Design Working Group Kick-off Meeting Slides

DOE Throughout Development



Use DOE to minimize the assets while gaining information required to verify whether a system meets performance requirements.

Use DOE to identify scenarios that most completely cover the requirements space with limited assets and to provide the most useful data for simulation validation. Use to identify with confidence those variables that need testing in OT and those that can be excluded.

Use DOE to most effectively use lab and hardware assets. Will allow comparison of interaction results between HWIL and STS/ILS. Us information to help finalize factors and levels for DT.

Use DOE to understand effects of component variation and sensitivities to many variables.

Use DOE for simulation to determine software sensitivities to many variables. Also used to develop a test matrix that will give a "quick look". Use factor covering arrays to identify software bugs.

Design for Six Sigma and DOE used during subcomponent development leads to confidence in performance and reliability. Helps identify and correct problems early.

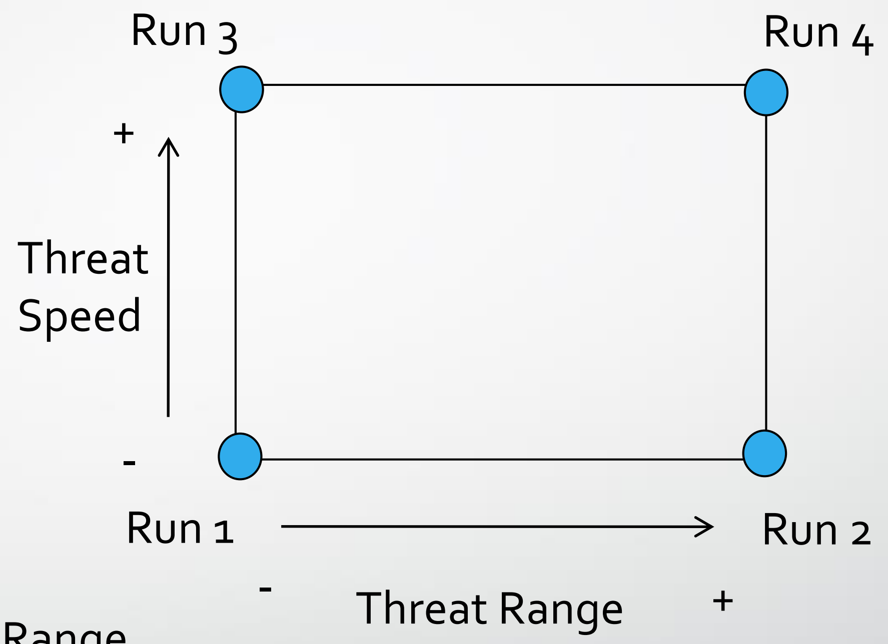
Types of Designs

Full Factorial Design

Contains all possible combinations of factors and levels

For two independent variables, threat speed and threat range, we want to know their effect on ability to hit the threat:

Run	Aircraft Altitude (A)	Aircraft Speed (B)	AB
1	-	-	+
2	-	+	-
3	+	-	-
4	+	+	+



Interaction between Threat Speed and Threat Range
We don't get this from changing one factor at a time

Design Nomenclature:

- The columns contain the factors that are being controlled in the test
- For two level factors, "-" or "-1" is typically used for the low level and "+" or "+1" for the high level
- Each row in a matrix is called a run

Full Factorial (cont)

For three independent variables:

	Main Effects			Interactions			
Run	**Aircraft Altitude (A)**	**Aircraft Speed (B)**	**Threat Type (C)**	**AB**	**AC**	**BC**	**ABC**
1	-	-	-	+	+	+	-
2	-	-	+	+	-	-	+
3	-	+	-	-	+	-	+
4	-	+	+	-	-	+	-
5	+	-	-	-	-	+	+
6	+	-	+	-	+	-	-
7	+	+	-	+	-	-	-
8	+	+	+	+	+	+	+

Advantages

Tests all possible combinations

Allows evaluation of interactions

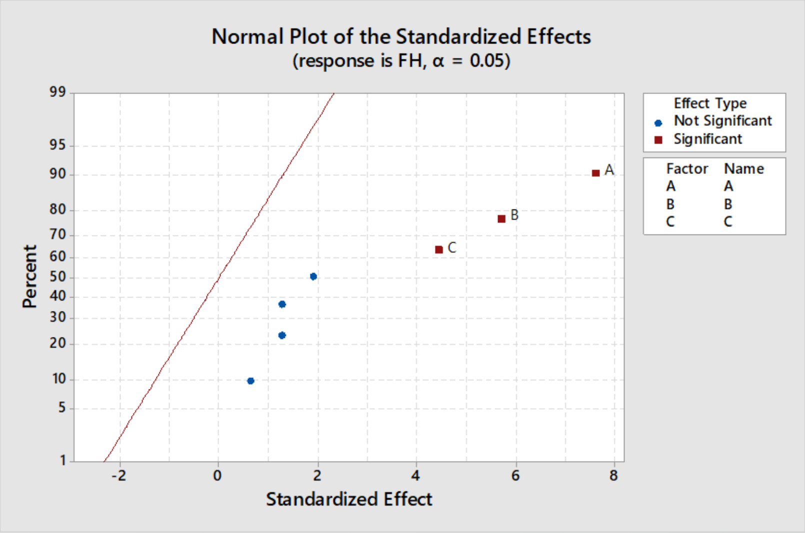
Disadvantages

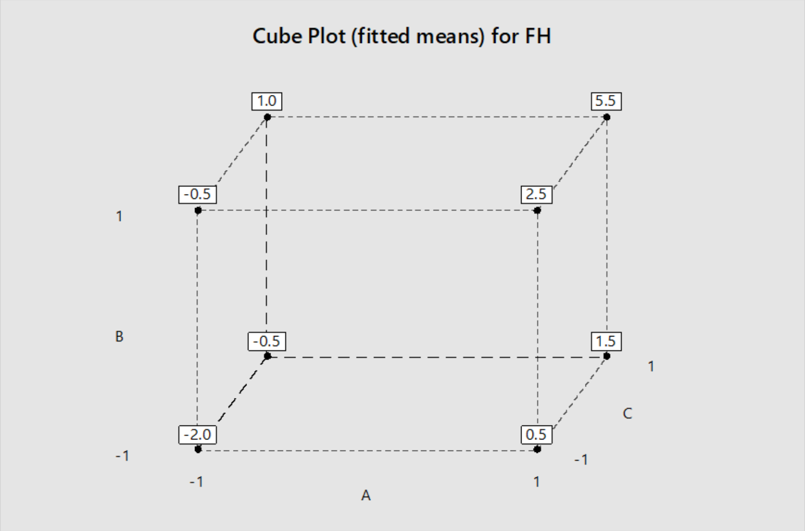
For more than a few variables, requires many runs (2^k)

Full Factorial Case Study

A 2^3 factorial design was used to study the effect of percentage carbonation (A), operating pressure (B), and line speed (C) on the fill height (FH) of a carbonated beverage.

A:PC	B:OP	C:LS	FH
-1	-1	-1	-3
1	-1	-1	0
-1	1	-1	-1
1	1	-1	2
-1	-1	1	-1
1	-1	1	2
-1	1	1	1
1	1	1	6
-1	-1	-1	-1
1	-1	-1	1
-1	1	-1	0
1	1	-1	3
-1	-1	1	0





Source: D.C. Montgomery, *Design & Analysis of Experiments*, 8th Ed, Wiley, 2001, p. 232-233.

Fractional Factorial Experiments

- Only runs a "fraction" of the runs in a full factorial

	# Runs in 1 Replication of the Experment				
	Full Factoral	**1/2 Fraction**	**1/4 Fraction**	**1/8 Fraction**	**1/16 Fraction**
# Factors (k)	2^k	2^{k-1}	2^{k-2}	2^{k-3}	2^{k-4}
4	16	8	4		
5	32	16	8	4	
6	64	32	16	8	4
7	128	64	32	16	8
8	256	128	64	32	16

- Based on the principal that main effects and 2-way interactions have the largest effect on the response variable
- Advantages
  - Significantly reduces the number of runs required
  - Allows Sequential Experimentation which can save time, effort, and resources
  - Good for screening a large number of factors to a smaller number
- Disadvantages
  - Higher order interactions can be missed
  - Lower power for factors tested

Designing a Fractional Factorial Experiment

Full Factorial (2 Factors)

Run	Aircraft Altitude (A)	Aircraft Speed (B)	Threat Type (C)
1	-	-	+
2	-	+	-
3	+	-	-
4	+	+	+

AB Interaction

Main effect replaces AB Interaction

C=C+AB

Start with Full 2 Level Factorial

Replace the Interaction with a Main Effect

Allows us to see 3 Factors in half the runs

Confounded with 2 Level Interactions

By replacing the interaction AB with the main effect Network Traffic we have assumed that the interaction AB has negligible effect. Otherwise, we will not be able to accurately estimate the effects of Network Traffic.

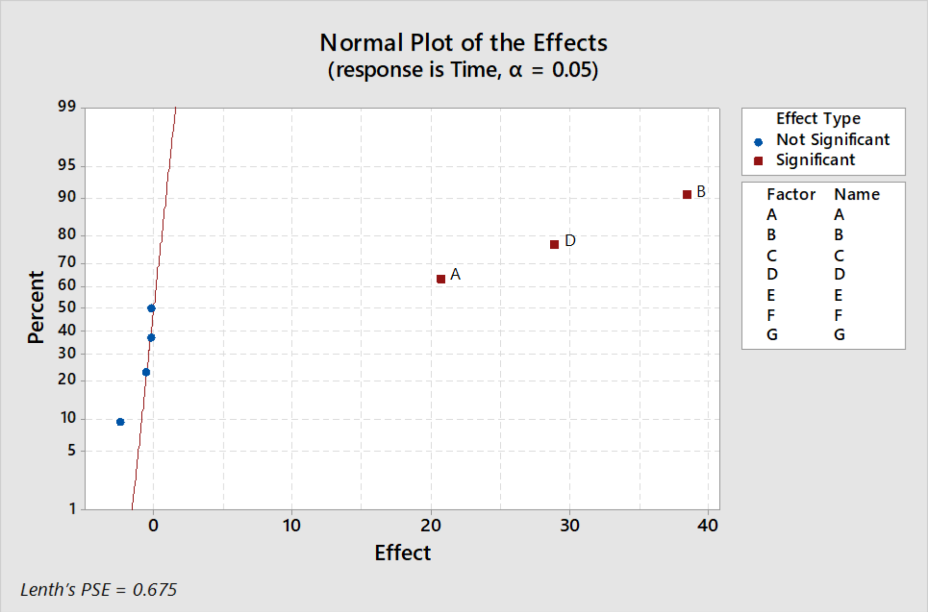
Fractional Factorial Case Study

(2^{7-4} Fractional Factorial Experiment)

Objective: Identify factors that impact eye focus time

Factors: Visual acuity (A), Distance from target to eye (B), Target shape (C), Illumination level (D), Target size (E), Target density (F), and Subject (G)

A	B	C	D	E	F	G	Time
-1	-1	-1	1	1	1	-1	85.5
1	-1	-1	-1	-1	1	1	75.1
-1	1	-1	-1	1	-1	1	93.2
1	1	-1	1	-1	-1	-1	145.4
-1	-1	1	1	-1	-1	1	83.7
1	-1	1	-1	1	-1	-1	77.6
-1	1	1	-1	-1	1	-1	95
1	1	1	1	1	1	1	141.8



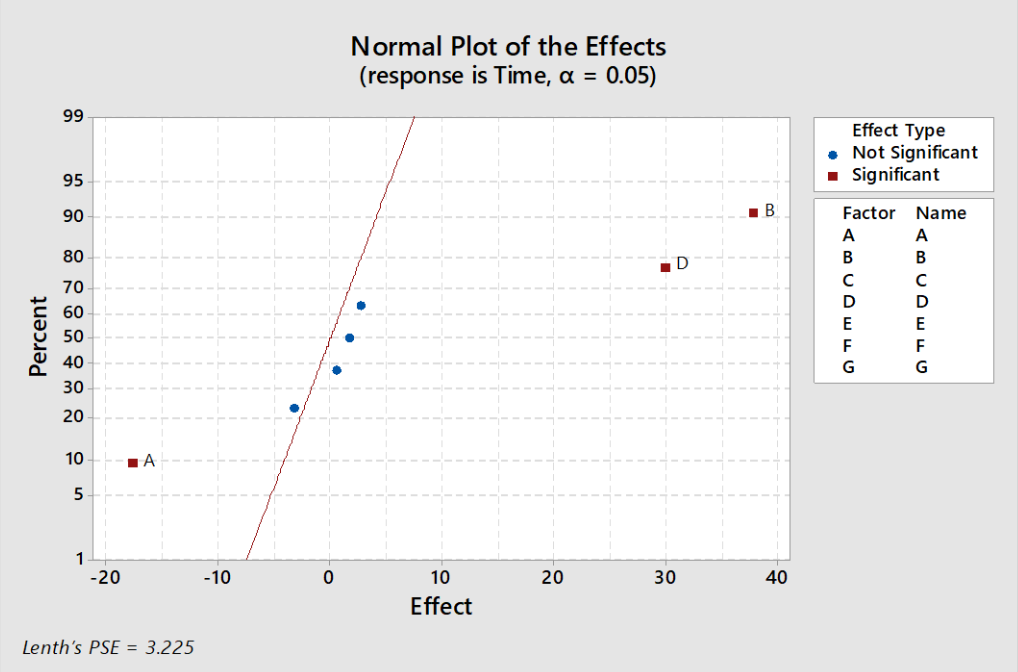
$$[A] = 20.63 \rightarrow A + BD + CE + FG$$
$$[B] = 38.38 \rightarrow B + AD + CF + EG$$
$$[C] = -0.28 \rightarrow C + AE + BF + DG$$
$$[D] = 28.88 \rightarrow D + AB + CG + EF$$
$$[E] = -0.28 \rightarrow E + AC + BG + DF$$
$$[F] = -0.63 \rightarrow F + BC + AG + DE$$
$$[G] = -2.43 \rightarrow G + CD + BE + AF$$

Source: D.C. Montgomery, *Design & Analysis of Experiments*, 8th Ed, Wiley, 2013, p. 355.

Fractional Factorial Case (cont)

Alternative fraction run to gain more insight into the relationships

A	B	C	D	E	F	G	Time
1	1	1	-1	-1	-1	1	91.3
-1	1	1	1	1	-1	-1	136.7
1	-1	1	1	-1	1	-1	82.4
-1	-1	1	-1	1	1	1	73.4
1	1	-1	-1	1	1	-1	94.1
-1	1	-1	1	-1	1	1	143.8
1	-1	-1	1	1	-1	1	87.3
-1	-1	-1	-1	-1	-1	-1	71.9



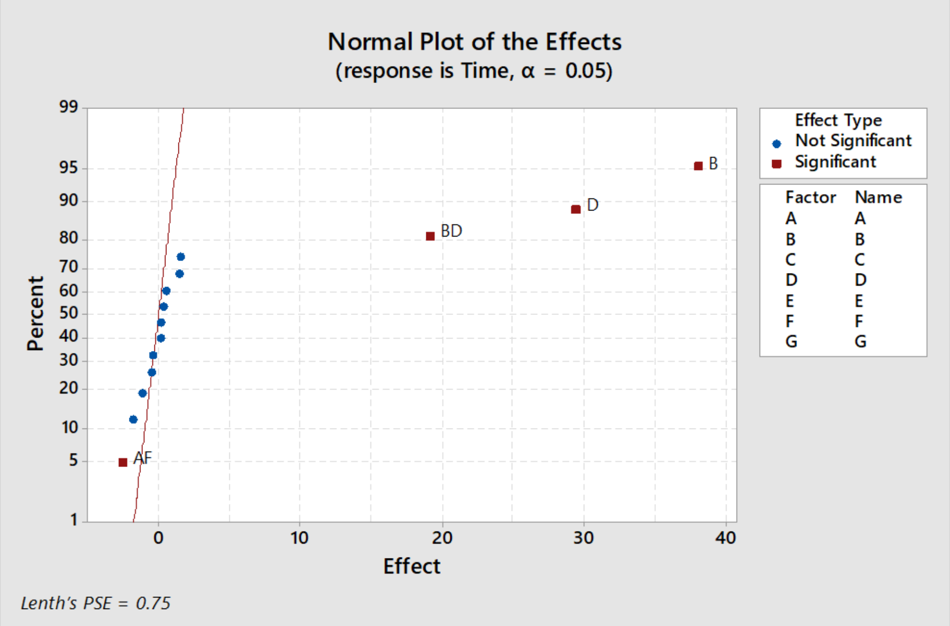
$$[A]' = -17.68 \rightarrow A - BD - CE - FG$$
$$[B]' = 37.73 \rightarrow B - AD - CF - EG$$
$$[C]' = -3.33 \rightarrow C - AE - BF - DG$$
$$[D]' = 29.88 \rightarrow D - AB - CG - EF$$
$$[E]' = 0.53 \rightarrow E - AC - BG - DF$$
$$[F]' = 1.63 \rightarrow F - BC - AG - DE$$
$$[G]' = 2.68 \rightarrow G - CD - BE - AF$$

Source: D.C. Montgomery, *Design & Analysis of Experiments*, 8th Edition, Wiley, 2013, p.356.

Fractional Factorial Case (cont)

Data for the two experiments can be combined to break apart alias structure and understanding which factors have the strongest relationship to the response (Eye Focus Time)

Block	A	B	C	D	E	F	G	Time
1	-1	-1	-1	1	1	1	-1	85.5
1	1	-1	-1	-1	-1	1	1	75.1
1	-1	1	-1	-1	1	-1	1	93.2
1	1	1	-1	1	-1	-1	-1	145.4
1	-1	-1	1	1	-1	-1	1	83.7
1	1	-1	1	-1	1	-1	-1	77.6
1	-1	1	1	-1	-1	1	-1	95
1	1	1	1	1	1	1	1	141.8
2	1	1	1	-1	-1	-1	1	91.3
2	-1	1	1	1	1	-1	-1	136.7
2	1	-1	1	1	-1	1	-1	82.4
2	-1	-1	1	-1	1	1	1	73.4
2	1	1	-1	-1	1	1	-1	94.1
2	-1	1	-1	1	-1	1	1	143.8
2	1	-1	-1	1	1	-1	1	87.3
2	-1	-1	-1	-1	-1	-1	-1	71.9



i	From $\frac{1}{2}([i]+[i]')$	From $\frac{1}{2}([i]-[i]')$
A	$A = 1.48$	$BD + CE + FG = 19.15$
B	$B = 38.05$	$AD + CF + EG = 0.33$
C	$C = -1.80$	$AE + BF + DG = 1.53$
D	$D = 29.38$	$AB + CG + EF = -0.50$
E	$E = 0.13$	$AC + BG + DF = -0.40$
F	$F = 0.50$	$BC + AG + DE = -1.13$
G	$G = 0.13$	$CD + BE + AF = -2.55$

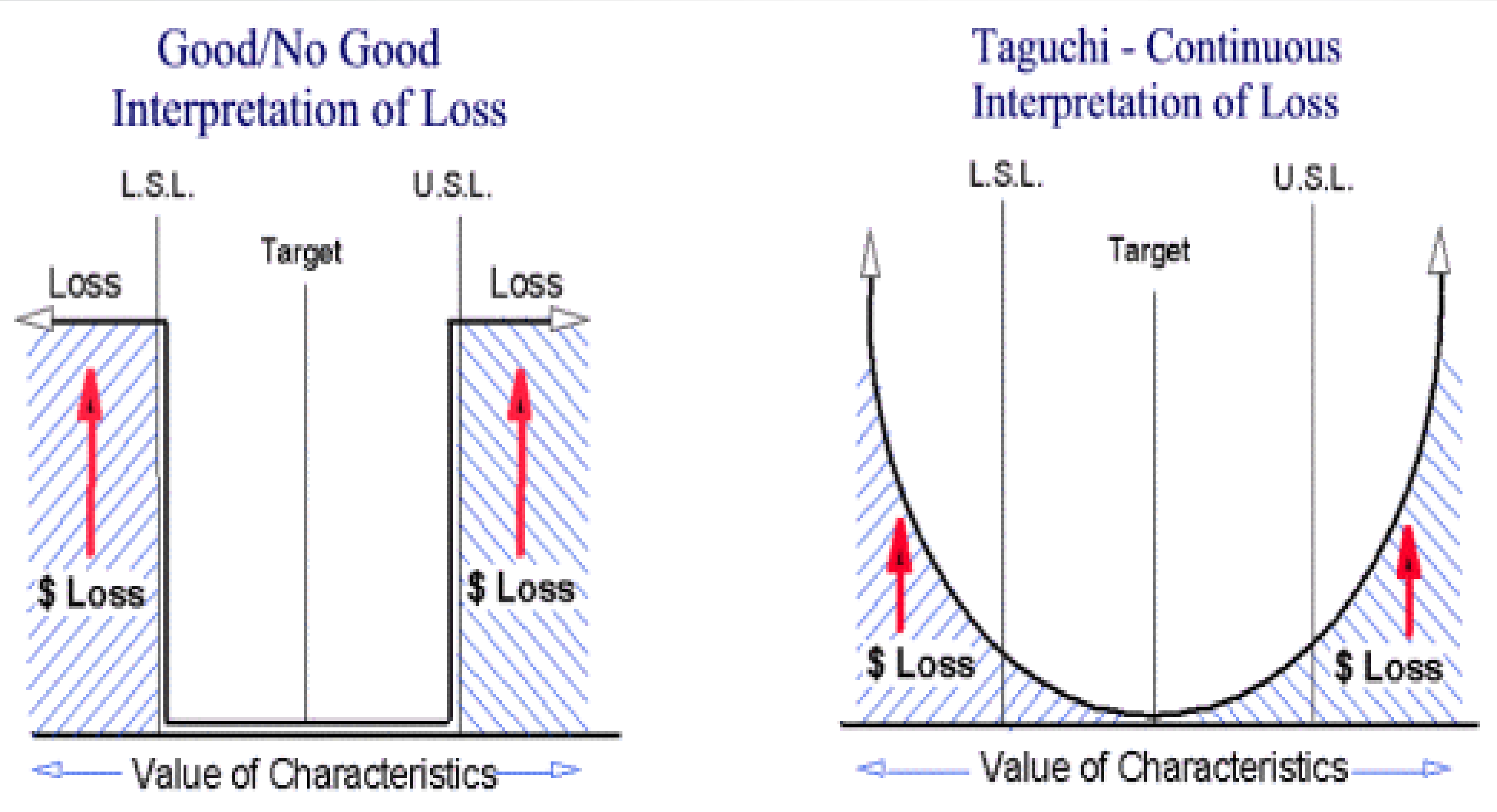
Adapted from D.C. Montgomery, *Design & Analysis of Experiments*, 8th Ed, Wiley, 2013, p. 356.

Taguchi's Approach to Design of Experiments

Source for Orthogonal Arrays & Linear Graphs: Taguchi, G., Chowdhury, S. and Wu, Y. (2004) Appendix C: Orthogonal Arrays and Linear Graphs for Chapter 38, in *Taguchi's Quality Engineering Handbook*, John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/9780470258354.app3 http://onlinelibrary.wiley.com/doi/10.1002/9780470258354.app3/pdf

Two Views of Quality

Traditional View of Quality | **Taguchi's View of Quality**



Source: http://leansixsigmadefinition.com/glossary/taguchi-loss-function/

Some Orthogonal Arrays

Taguchi Nomenclature: 1 is low & 2 is high

$L_4(2^3)$ **2^3 Full Factorial Design**

No.	1	2	3
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1
	a	*b*	*a*
			b
	1	2	

$L_8(2^7)$ **2^{7-4} Fractional Factorial Design**

No.	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2
	a	*b*	*a*	*c*	*a*	*b*	*a*
			b		*c*	*c*	*b*
							c
	1	2		3			

$L_{16}(2^{15})$ **2^{15-11} Fractional Factorial Design**

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
3	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
4	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1
5	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2
6	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1
7	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1
8	1	2	2	2	2	1	1	2	2	1	1	1	1	2	2
9	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
10	2	1	2	1	2	1	2	2	1	2	1	2	1	2	1
11	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1
12	2	1	2	2	1	2	1	2	1	2	1	1	2	1	2
13	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1
14	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2
15	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2
16	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1
	a	*b*	*a*	*c*	*a*	*b*	*a*	*d*	*a*	*b*	*a*	*c*	*a*	*b*	*a*
			b		*c*	*c*	*b*		*d*	*d*	*b*	*d*	*c*	*c*	*b*
							c				*d*		*d*	*d*	*c*
															d
	1	2		3				4							

DOE using L8 Orthogonal Array

$L_8(2^7)$

No.	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2
	a	b	a b	c	a c	b c	a b c
	1	2		3			

Basic Design Process:
- Determine number of factors to include in design
  - L8 Array can be used for up to 7 factors
- Assign factors to columns
  - Taking into consideration aliasing

Three Experiments Designed using L8 Array

Experiment 1 (2^3 Full Factorial)						
Column Number						
1	2	3	4	5	6	7
A	B	AxB	C	AxC	BxC	AxBxC

Experiment 2 (2^{4-1} Fractional Factorial)						
Column Number						
1	2	3	4	5	6	7
A	B	AxB	C	AxC	BxC	AxBxC
BxCxD	AxCxD	CxD	AxBxD	BxD	AxD	D

Experiment 3 (2^{5-2} Fractional Factorial)						
Column Number						
1	2	3	4	5	6	7
A	B	AxB	C	AxC	BxC	AxBxC
BxCxD	AxCxD	CxD	AxBxD	BxD	AxD	D
BxE	AxE	E	DxE	AxDxE	BxDxE	CxE
				BxCxE	AxCxE	

Source: P.J. Ross, *Taguchi Techniques for Quality Engineering*, 2nd Ed, McGraw Hill, 1996, p. 56-57.

Taguchi's Approach to DOE

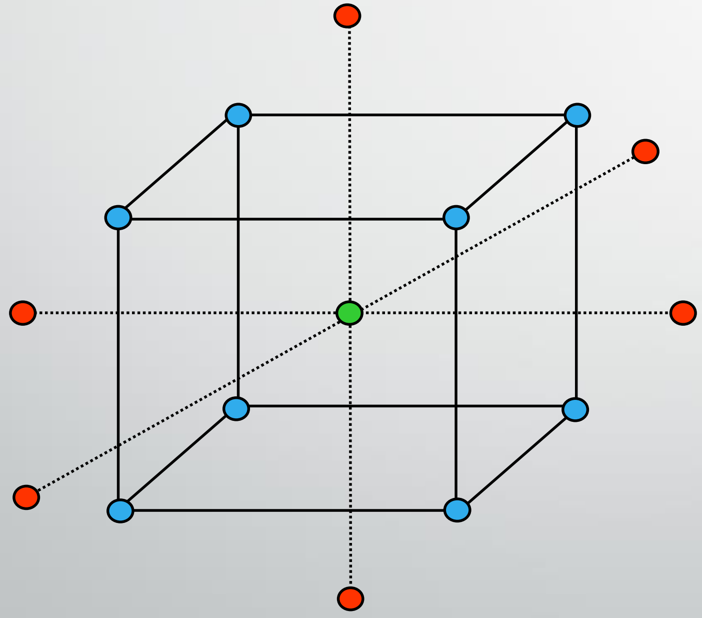
Observations:

- Taguchi raised awareness that we really want the process on-target AND we want to reduce variability
- Orthogonal Arrays provides a set of pre-established designs
  - Some of these designs are traditional fractional factorial designs
  - Some of these designs are Plackett Burman designs
    - Issue: These designs have complex aliasing that we may not be able to break apart using sequential experimentation strategy
- Potentially provides information on more factors with fewer runs
  - The trade-off is that we have to give up some information
  - If we don't know the alias structure of a design we cannot run an alternative fraction of the design to break apart the aliases to determine what is actually "driving" the process
- A better approach would be to use fractional factorial designs with a sequential strategy to purposefully identify factors that are "driving" the process.

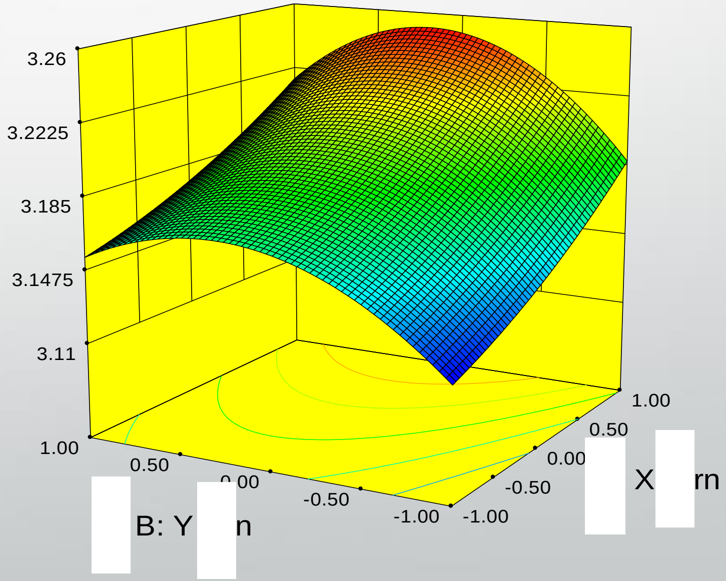
Response Surface Design

- Adds center points and additional runs to the factorial design
- Allows evaluation of curvature
- Mostly for quantitative variables

For three independent variables

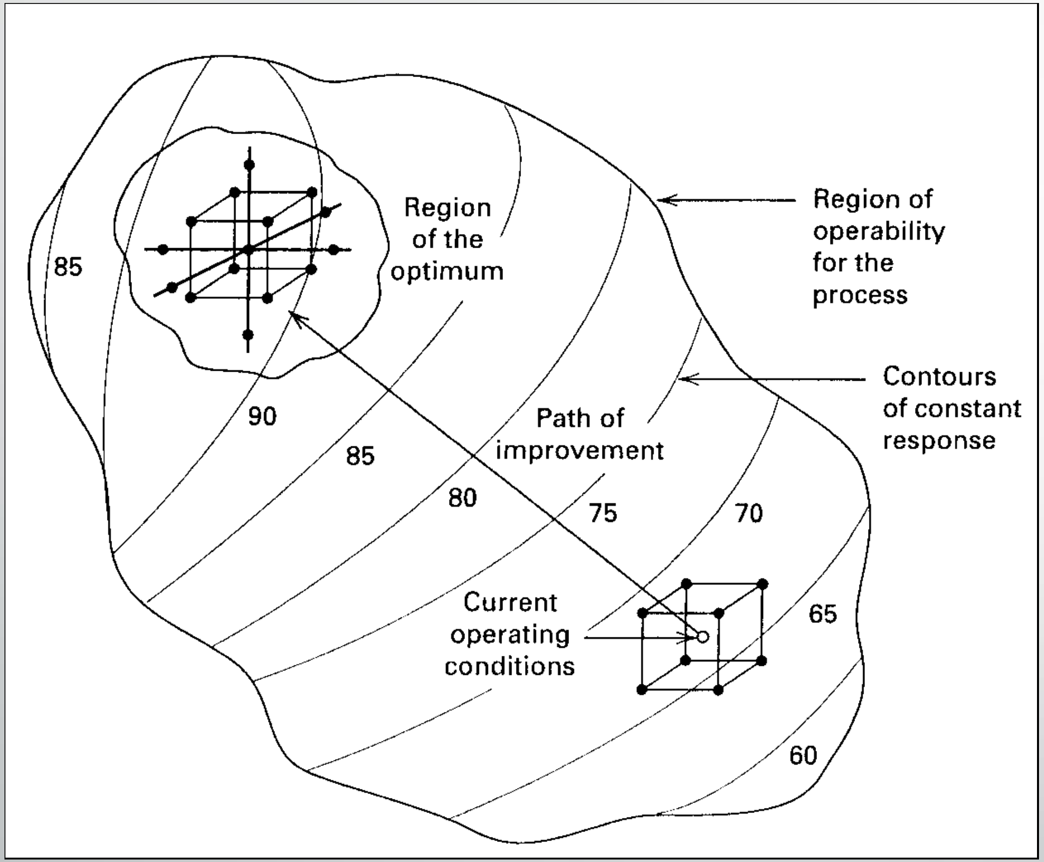


Source: Air Force DOE Class Slides



Source: Air Force DOE Class Slides

Sequential Nature of RSM



Initial Objective:
Lead Experimenter rapidly and efficiently along a path of improvement toward the general vicinity of the optimal

Final Objective:
To determine optimum operating conditions for system or where operating requirements are satisfied

Source: D.C. Montgomery, *Design & Analysis of Experiments*, 8th Ed, Wiley, 2013, Chapter 11 (Fig 11-3, p. 419)

Response Surface Design Case

Objective: Determine operating conditions that maximize the process yield (y)

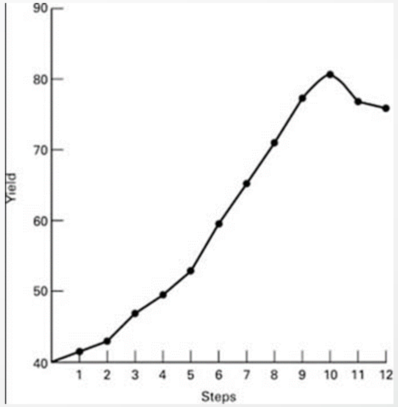
Factors: Reaction time (x1), Reaction temperature (x2)

Initial 2^2 Design

x1	x2	y
-1	-1	39.3
-1	1	40.0
1	-1	40.9
1	1	41.5
0	0	40.3
0	0	40.5
0	0	40.7
0	0	40.2
0	0	40.6

Results indicate A&B Significant

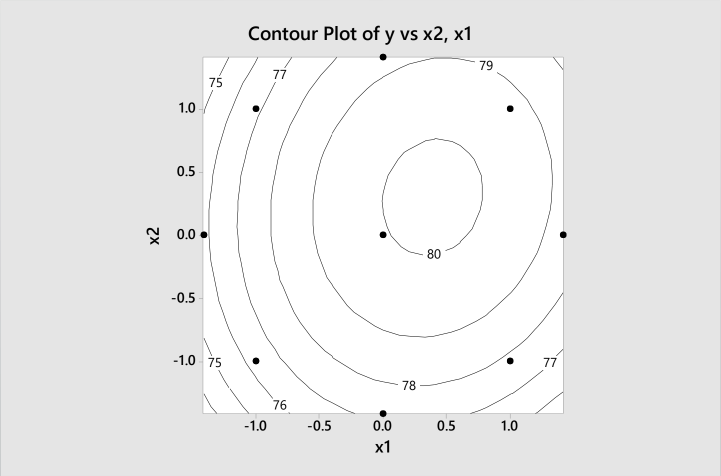
Traverse Path of Steepest Ascent



RSM Design & Peak

x1	x2	y
-1	-1	76.5
1	-1	78.0
-1	1	77.0
1	1	79.5
-1.414	0	75.6
1.414	0	78.4
0	-1.414	77.0
0	1.414	78.5
0	0	79.9
0	0	80.3
0	0	80.0
0	0	79.7
0	0	79.8

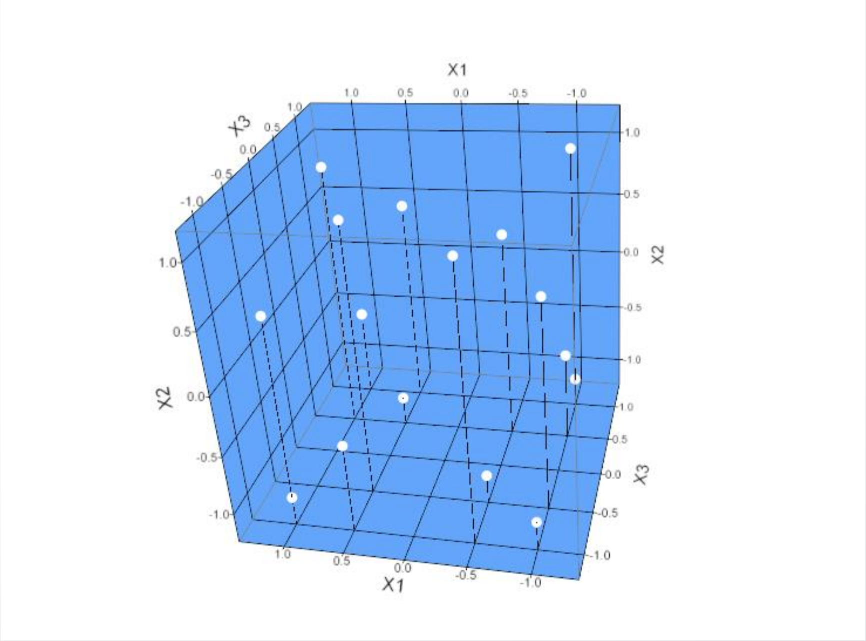
Results Show Optimal Region



Adapted from D.C. Montgomery, *Design & Analysis of Experiments*, 8th Ed, Wiley, 2013, Chapter 11.

Optimal Designs

- Computer generated designs based on SME inputs
- Two primary uses
  - Creating designs with significant constraints
  - Repairing existing designs
- Purpose is to optimize a specific test design parameter
  - D-optimal: Minimizes variance of model coefficients
  - I-optimal: Minimizes average prediction variance
- There are many types of optimal designs
- Can be used to back into a design when assets available are known and limited



Source: Air Force DOE Class Slides

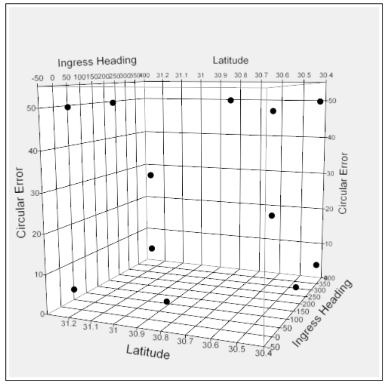
Software Testing

- Testing for software is a newer field than traditional testing
- Because software will give repeatable answers for the same inputs, the variability normally addressed by designed experiments does not exist
- Testing for software is based on probabilities
- Two primary types of designs
  - Space filling
  - Factor covering
- Much study still ongoing in this area
- There is not a well-developed set of principles and methods
- Not used to quantify the effects of inputs on an output

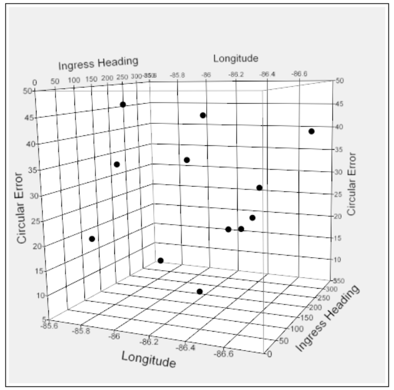
Space Filling Designs

- 3 Popular Algorithms:
  - Sphere-Packing
    - Maximize the smallest distance between neighbors
    - Effect: Moves points out to boundaries
  - Uniform
    - Minimize discrepancy from a uniform distribution
    - Effect: Spreads points within interior
  - Latin Hypercube
    - Assign n congruent levels and minimize covariance
    - Effect: Combination of the above

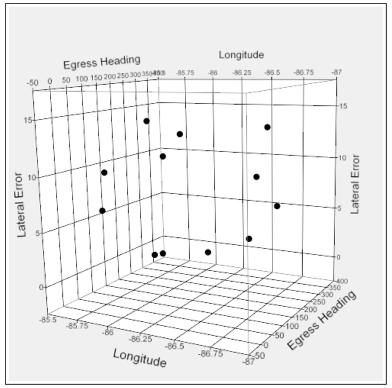
Source: Air Force DOE I Class Slides



Source: Air Force DOE Class Slides



Source: Air Force DOE Class Slides



Source: Air Force DOE Class Slides

Factor Covering

- Also known as high throughput testing
- Goal is to cover some number (n) of combinations of factors and levels
  - Does not quantify effects
  - Uses many (sometimes phenomenally) fewer runs than traditional designs
  - Good for quantitative and qualitative variables
- Example

Type of Radar	Quantity of Radar	Target Speed
Patriot	1	Slow
Sentinel	2	Mid
JLENS	3	Fast

- All possible combinations would require 27 runs
- All possible pairs can be covered in 9 runs
- Design is software generated and is not unique

Confidence and Power

Confidence and Power?

- From test data we want to determine which of our test variables impact performance (and how much) and which variables don't
- Ideally, no test variable will have an unfavorable impact on performance when set within the requirements space
  - This would mean that the system performs equally well anywhere in the space
  - Not likely
- To determine if variables have an impact and the magnitude of the impact we perform various analyses on our test data
  - Hypothesis testing
  - Analysis of Variance (ANOVA)
  - Regression Analysis
- It is important to know how much we can rely on the results of our analysis
  - Confidence tells us the likelihood that we will not falsely identify a factor as significant (avoiding a false positive)
  - Power tells us the likelihood that analysis will not fail to identify a significant variable (avoiding a false negative)

Example

- Say we are testing the effect of threat type on the distance at which we detect the target. Specifically, is our detection distance for Target A greater than for Target B?
- For the difference in detection distance, we measure our distance from the target when we detected it.
- We can use a hypothesis test for differences in means to evaluate whether our average detection distance changed with target type.
- But first we have to answer a few questions
  - How much of an effect (difference in detection range) do we need to detect?
  - How sure do we want to be that target type has a given effect on detection range if analysis says it does? This is our confidence (related to α or Type I error).
  - How sure do we want to be that we will successfully determine that target type has a given effect on detection range if it really does? This is our power (related to β or Type II error).
- These questions help determine the required sample size

What Does Confidence Mean to Me?

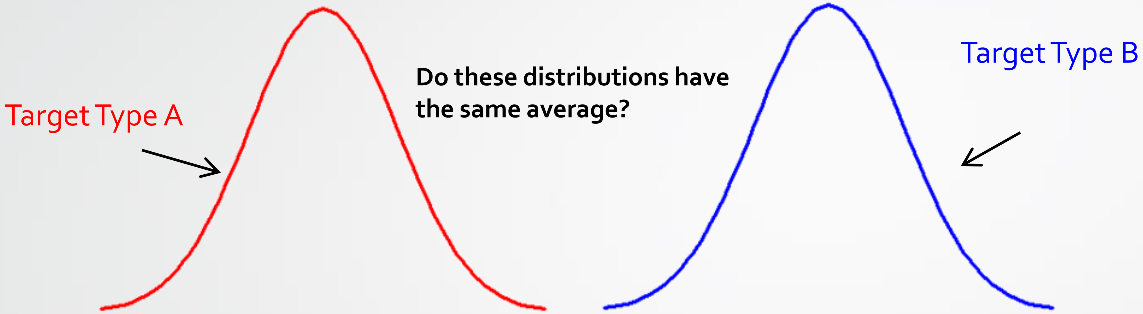
- The risk of falsely identifying an effect is α
- Confidence is equal to $1-\alpha$
- Using our detection range example, let's say that we want to show that target type (within spec limits) does not have an effect significant enough to cause us to fail to detect a target at the required range. If it does, then say the radar system is not acceptable.
- Those trying to show that the system is acceptable want to plan the test with a high confidence level
  - This will give confidence that we do not falsely identify a problem
  - False identification will unnecessarily add cost for the producer
- Bottom line: A high confidence level is most important to those who want to avoid finding an issue that doesn't really exist
- Effect on cost: All other things equal, a larger sample size (increased cost) is required to increase confidence

What Does Power Mean to Me?

- The risk of failing to identify a significant effect is β
- Power is equal to $1-\beta$
- Using the detection distance example, those who are using the radar system to make sure that any significant effects of target type on detection range are known.
- Users want to make sure that tests are planned with a high power
  - This increases the likelihood that we do not fail to identify a problem
  - Failure to identify a problem may result in field failures or poor employment strategy
- Bottom line: High power benefits those who are interested in ensuring that issues are detected
- Impact on cost: All other things equal, to correctly identify a given effect with given confidence, higher power requires a larger sample size (higher cost)
- Power is only meaningful prior to performing a test

In General, What's Our Goal

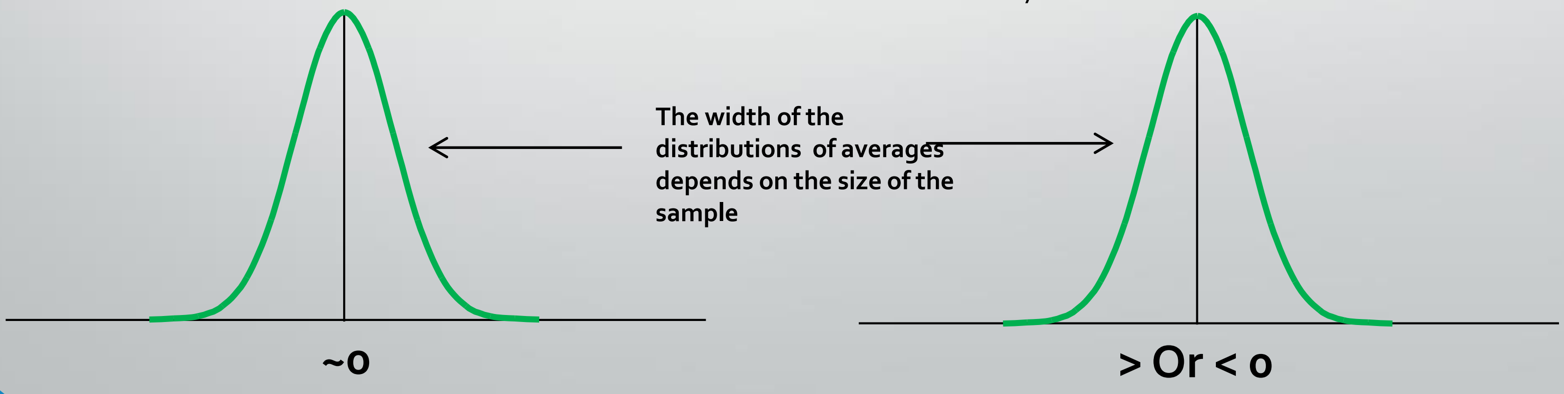
We want to know if the average detection range is affected by our test factors. For example, does threat type affect our miss distance?



If we 1) take a sample from each distribution 2) calculate the average of each sample 3) subtract the average of one from the average of the other and 4) do this many times…

If the averages are equal then the distribution will be centered near zero.

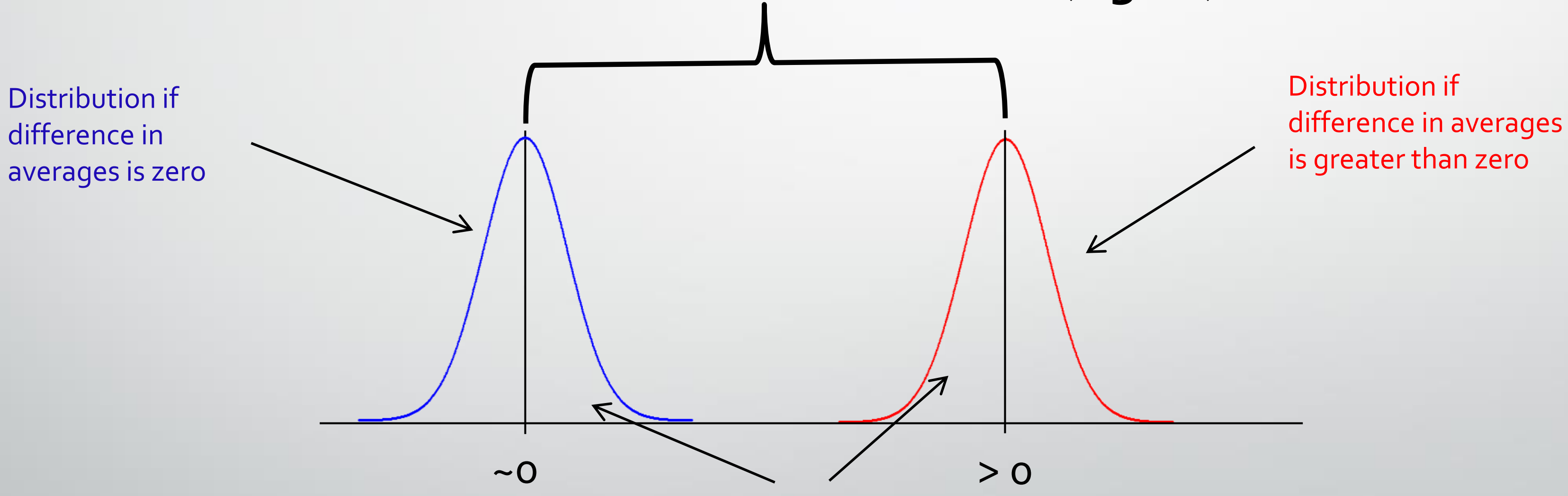
If the averages are not equal then the distribution will be centered away from zero.



More Specific Goal

- Let's say we want to design a test to determine what effect our test factors (Aircraft Altitude, Aircraft Speed, Target Type) and their interactions have on our detection range.
- We'll need to answer a few questions to help size the matrix.

How much of an effect do we care about (signal)?



How wide do we think these distributions might be (noise)?

Signal, Noise, Risk, and Cost

Expected distribution if averages are equal

Expected distribution if averages are not equal

α: The risk of mistakenly concluding there is a difference in the averages. Confidence is equal to 1- α.

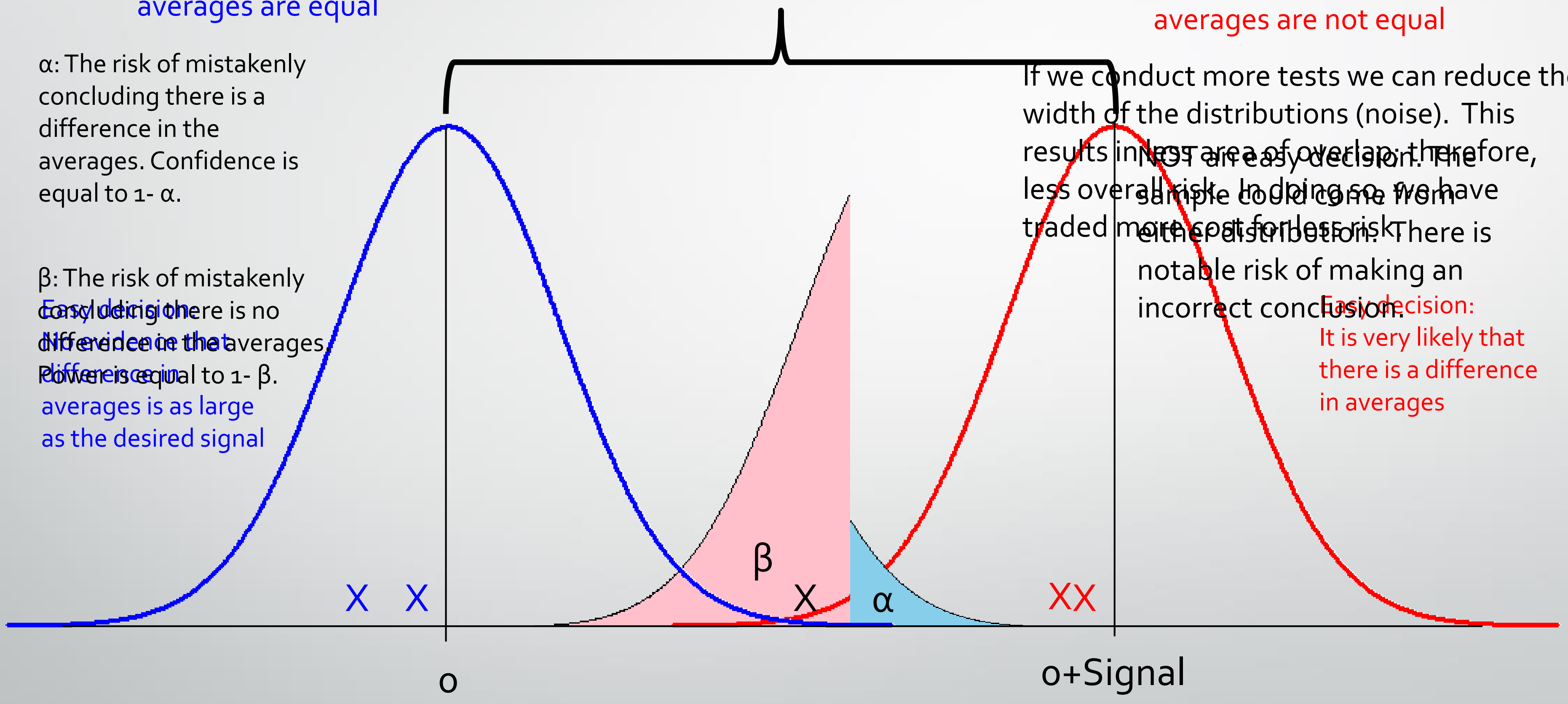
β: The risk of mistakenly concluding there is no difference in the averages. Power is equal to 1- β.

Easy decision: No evidence that difference in averages is as large as the desired signal

If we conduct more tests we can reduce the width of the distributions (noise). This results in less area of overlap; therefore, less overall risk. In doing so, we have traded more cost for less risk.

NOT an easy decision. The sample could come from either distribution. There is notable risk of making an incorrect conclusion.

Easy decision: It is very likely that there is a difference in averages



Summary (1 of 2)

- DOE does not make obsolete the need for technical expertise. It allows us to get the most information from technical knowledge and testing.
- DOE encompasses a very wide variety of methods and is much more than the factorial matrices commonly taught
- DOE allows us to efficiently identify the effects of input variables on an output
    - Box pushing example
    - Allows us to evaluate interactions between input variables
- DOE is applicable to testing that involves hardware, software, and simulation
    - Most effective when test designs for these activities support each other
    - Not just at the system level
    - The earlier we use DOE in subsystem and component design and testing, the greater the benefit

Summary (2 of 2)

- We can evaluate the likelihood that a test will allow us to identify variables with significant effect (known as power)
- In test planning, we can trade power, confidence, effect size, and sample size
- For a given sample size increasing confidence decreases power
- Increasing sample size will result in lower variation which has an effect on power and confidence
- Non-orthogonal variables result in artificially inflated variances when examining multiple variables
- Inflated variances result in lower power (more likely to miss a significant effect)
- Design of experiments ensures that the variance is not artificially inflated.

Backup Slides

Two-Level Orthogonal Array Factor Assignment

OA	# Factors	Use Column Numbers	Resolution
L4	1-2	1,2	High
	3	1,2,3	Low
L8	1-3	1,2,4	High
	4	1,2,4,7	Moderate
	5-7	1,2,4,7,(3,5,6)*	Low
L12	1-11	1-11	Low
L16	1-4	1,2,4,8	High
	5	1,2,4,8,15	Moderate
	6-8	1,2,4,7,8,(11,13,14)	Moderate
	9-15	1,2,4,7,8,11,13,14,(3,5,6,9,10,12,15)	Low
L32	1-5	1,2,4,8,16	High
	6	1,2,4,8,16,31	Moderate
	7-16	1,2,4,8,16,31,(7,11,13,14,19,21,22,25,26,28)	Moderate
	17-31	1,2,4,7,8,11,13,14,16,19,21,22,25,26,28,31,(3,5,6,9,10,12,15,17,18,20,23,24,27,29,30)	Low

Resolution is and indciation of the amount of confounding in a column

*Column numbers in parentheses may be assigned in any order to achieve the indicated resolution; column numbers not in parentheses must be used first.

Source: P.J. Ross, *Taguchi Techniques for Quality Engineering*, 2nd Ed, McGraw Hill, 1996, p. 284

Three-Level Orthogonal Array Factor Assignment

OA	# Factors	Use Column Numbers	Resolution
L9	1-2	1,2	High
	3-4	(1,2,3,4)	Low
L18	1-8	1-8	Low
L27	1-3	1,2,5	High
	4	1,2,5,(9,10,12,13)	Moderate
	5-13	1,2,3,4,5,(6-13)	Low

Resolution is an indiciation of the amount of confounding in a column

*Column numbers in parentheses may be assigned in any order to achieve the indicated resolution; column numbers not in parentheses must be used first.

Source: P.J. Ross, *Taguchi Techniques for Quality Engineering*, 2nd Ed, McGraw Hill, 1996, p. 285

Example of Robust Design Analysis

							J	-	+	+	-		
							I	-	+	-	+		
							H	-	-	+	+		
Run	A	B	C	D	E	F	G					$\bar{y}_r$	s_r
1	-	-	-	-	-	-	-	2.6	2.7	2.8	2.8	2.725	0.096
2	-	-	-	+	+	+	+	0.8	3	0.8	3.2	1.950	1.330
3	-	+	+	-	-	+	+	3.6	1	3.3	0.9	2.200	1.449
4	-	+	+	+	+	-	-	2.5	2.4	2.5	2.3	2.425	0.096
5	+	-	+	-	+	-	+	3.5	3.6	3.5	3.5	3.525	0.050
6	+	-	+	+	-	+	-	2.6	4.7	3.6	1.5	3.100	1.369
7	+	+	-	-	+	+	-	4.5	2.4	2.7	5.1	3.675	1.328
8	+	+	-	+	-	-	+	2.4	2.5	2.3	2.4	2.400	0.082

Source: S.R. Schmidt and R.G. Launsby, *Understanding Industrial Designed Experiments*, Air Academy Press, Colorado Springs, Colorado, 1992, p. 5-31

Additional Orthogonal Arrays

L12 / L9 / L18 / L27

Source for Orthogonal Arrays & Linear Graphs: Taguchi, G., Chowdhury, S. and Wu, Y. (2004) Appendix C: Orthogonal Arrays and Linear Graphs for Chapter 38, in *Taguchi's Quality Engineering Handbook*, John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/9780470258354.app3 http://onlinelibrary.wiley.com/doi/10.1002/9780470258354.app3/pdf

$L_{12}(2^{11})$

No.	1	2	3	4	5	6	7	8	9	10	11
1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	2	2	2	2	2	2
3	1	1	2	2	2	1	1	1	2	2	2
4	1	2	1	2	2	1	2	2	1	1	2
5	1	2	2	1	2	2	1	2	1	2	1
6	1	2	2	2	1	2	2	1	2	1	1
7	2	1	2	2	1	1	2	2	1	2	1
8	2	1	2	1	2	2	2	1	1	1	2
9	2	1	1	2	2	2	1	2	2	1	1
10	2	2	2	1	1	1	1	2	2	1	2
11	2	2	1	2	1	2	1	1	1	2	2
12	2	2	1	1	2	1	2	1	2	2	1
	1	2									

$L_9(3^4)$

No.	1	2	3	4
1	1	1	1	1
2	1	2	2	2
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1
	a	b	a	a
			b	b^2
	1	2		

$L_{18}(2^1 \times 3^7)$

No.	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	1	1
2	1	1	2	2	2	2	2	2
3	1	1	3	3	3	3	3	3
4	1	2	1	1	2	2	3	3
5	1	2	2	2	3	3	1	1
6	1	2	3	3	1	1	2	2
7	1	3	1	2	1	3	2	3
8	1	3	2	3	2	1	3	1
9	1	3	3	1	3	2	1	2
10	2	1	1	3	3	2	2	1
11	2	1	2	1	1	3	3	2
12	2	1	3	2	2	1	1	3
13	2	2	1	2	3	1	3	2
14	2	2	2	3	1	2	1	3
15	2	2	3	1	2	3	2	1
16	2	3	1	3	2	3	1	2
17	2	3	2	1	3	1	2	3
18	2	3	3	2	1	2	3	1
	1	2	3					

$L_{27}(3^{13})$

No.	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	2	2	2	2	2	2	2	2	2
3	1	1	1	1	3	3	3	3	3	3	3	3	3
4	1	2	2	2	1	1	1	2	2	2	3	3	3
5	1	2	2	2	2	2	2	3	3	3	1	1	1
6	1	2	2	2	3	3	3	1	1	1	2	2	2
7	1	3	3	3	1	1	1	3	3	3	2	2	2
8	1	3	3	3	2	2	2	1	1	1	3	3	3
9	1	3	3	3	3	3	3	2	2	2	1	1	1
10	2	1	2	3	1	2	3	1	2	3	1	2	3
11	2	1	2	3	2	3	1	2	3	1	2	3	1
12	2	1	2	3	3	1	2	3	1	2	3	1	2
13	2	2	3	1	1	2	3	2	3	1	3	1	2
14	2	2	3	1	2	3	1	3	1	2	1	2	3
15	2	2	3	1	3	1	2	1	2	3	2	3	1
16	2	3	1	2	1	2	3	3	1	2	2	3	1
17	2	3	1	2	2	3	1	1	2	3	3	1	2
18	2	3	1	2	3	1	2	2	3	1	1	2	3
19	3	1	3	2	1	3	2	1	3	2	1	3	2
20	3	1	3	2	2	1	3	2	1	3	2	1	3
21	3	1	3	2	3	2	1	3	2	1	3	2	1
22	3	2	1	3	1	3	2	2	1	3	3	2	1
23	3	2	1	3	2	1	3	3	2	1	1	3	2
24	3	2	1	3	3	2	1	1	3	2	2	1	3
25	3	3	2	1	1	3	2	3	2	1	2	1	3
26	3	3	2	1	2	1	3	1	3	2	3	2	1
27	3	3	2	1	3	2	1	2	1	3	1	3	2
	a	b	a	a	c	a	a	b	a	a	b	a	a
			b	b^2		c	c^2	c	b	b^2	c^2	b^2	b
									c	c^2		c	c^2
	1	2			3								