AIR ACADEMY ASSOCIATES

Simplify, Perfect, Innovate

Rules of Thumb for Test Design and Analysis

ITEA's 17th Test Instrumentation Workshop
21 May 2014
Las Vegas, NV

14-DOEROT-5A

Mark J. Kiemele, Ph.D.
President and Co-Founder
Air Academy Associates

Office: 719-531-0777
Cell: 719-337-0357
mkiemele@airacad.com
www.airacad.com

Questions to be Answered

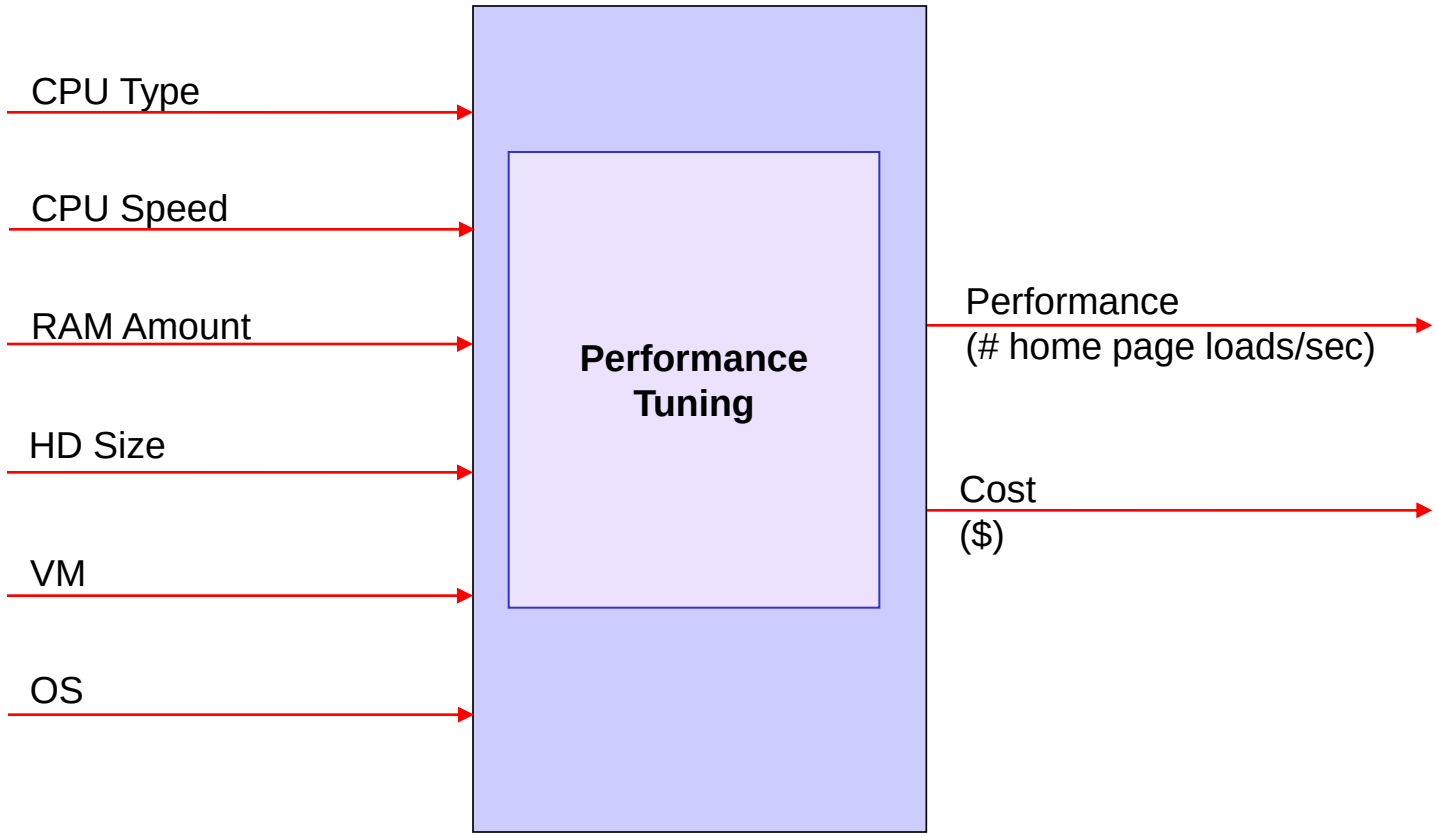
1. **What is an orthogonal test design and why is it important?**

2. **What are some of the key considerations when developing a test design?**

3. **What are some Rules of Thumb (or guidelines) that will allow us to test most effectively and efficiently and keep us safe in the T&E business?**

AIR ACADEMY ASSOCIATES
Simplify, Perfect, Innovate

Web-Based Application Testing Scenario



AIR ACADEMY ASSOCIATES
Simplify, Perfect, Innovate

Combinatorial Test/DoE Terminology

Y: Output, response variable, dependent variable

X: Input, factor, test parameter, independent variable (a measurable entity that is purposely changed during an experiment)

Level: A unique value or choice of a factor (X)

Run: An experimental combination of the levels of the X's

Replication: Doing or repeating an experimental combination

Effect: The difference or impact on Y when changing X

Interaction: When the effect of one factor depends on the level of another factor

Air Academy Associates
Simplify, Perfect, Innovate

Performance Tuning Terminology

Factors/Inputs (X's)	Levels (Choices)	Performance/Outputs (Y's)
CPU Type	Itanium, Xeon	# home page loads/sec
CPU Speed	1 GHz, 2.5 GHz	Cost
RAM Amount	256 MB, 1.5 GB	
HD Size	50 GB, 500 GB	
VM	J2EE, .NET	
OS	Windows, Linux	

Which factors are important? Which are not?
Which combination of factor choices will maximize performance?
Which combination of factor levels will create problems?
How do you know for sure? Show me the data.

Approaches to Testing Multiple Factors

- **Traditional Approaches**
  - One Factor at a Time (OFAT)
  - Oracle (Best Guess)
  - All possible combinations (full factorial)
- **Modern Approach**
  - Statistically designed experiments (DOE) … full factorial plus other selected DOE designs, depending on the situation

Air Academy Associates
Simplify, Perfect, Innovate

Oracle (Best Guess)

A = CPU Type (1=Itanium; 2=Xeon)

B = CPU Speed (1=1 GHz; 2=2.5 GHz)

C = RAM Amount (1=2 GB; 2=4 GB)

D = HD Size (1=50 GB; 2=500 GB)

E = VM (1=J2EE; 2=.NET)

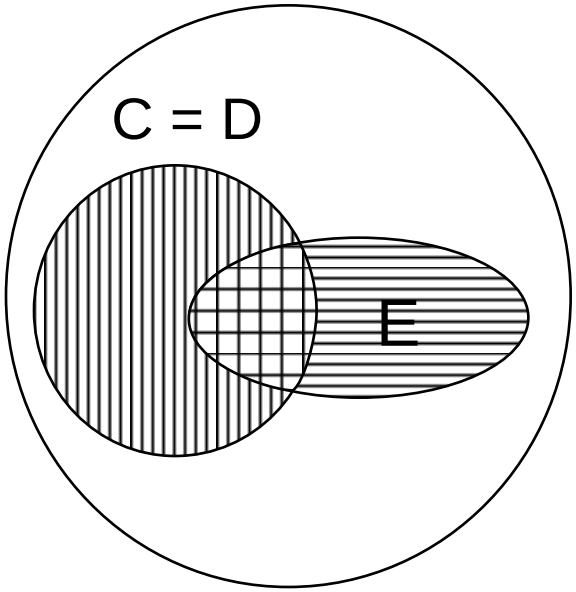
Y = # home page loads/sec

Run	A	B	C	D	E	Y
1	1	2	1	1	1	5
2	1	1	1	1	1	6
3	2	2	1	1	1	5
4	2	1	1	1	2	6
5	1	2	2	2	2	7
6	1	1	2	2	2	8
7	2	2	2	2	2	10
8	2	1	2	2	1	11

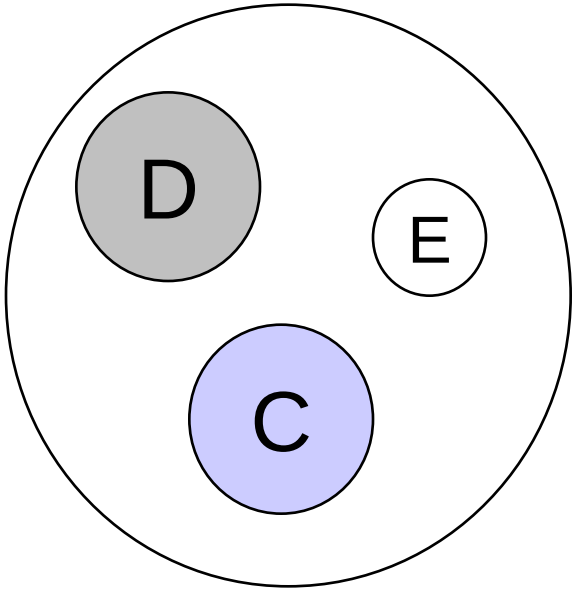
Does factor D shift the average of Y?

Evaluating the Effects of Variables on Y

What we have is:



What we need is a design to provide independent estimates of effects:



How do we obtain this independence of variables?

Statistically Designed Experiments (DOE):
Orthogonal or Nearly Orthogonal Designs

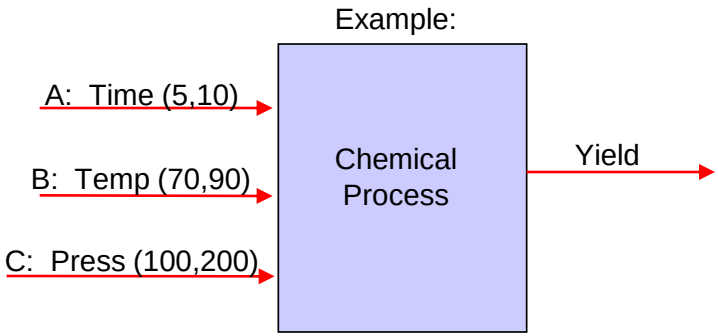
- FULL FACTORIALS (for small numbers of factors)
- FRACTIONAL FACTORIALS
- PLACKETT - BURMAN
- LATIN SQUARES
- HADAMARD MATRICES

Taguchi Designs

- BOX - BEHNKEN DESIGNS
- CENTRAL COMPOSITE DESIGNS

Response Surface Designs

- HIGH THROUGHPUT TESTING (ALL PAIRS)
- NEARLY ORTHOGONAL LATIN HYPERCUBE DESIGNS



SIMPLE DEFINITION OF A TWO-LEVEL ORTHOGONAL DESIGN

Run	Actual Settings			Coded Matrix			Interactions	
	(5, 10) A: Time	(70, 90) B: Temp	(100, 200) C: Press	(A) Time	(B) Temp	(C) Press	(AB) Uncoded	(AB) Coded
1	5	70	100	-1	-1	-1	350	+1
2	5	70	200	-1	-1	+1	350	+1
3	5	90	100	-1	+1	-1	450	-1
4	5	90	200	-1	+1	+1	450	-1
5	10	70	100	+1	-1	-1	700	-1
6	10	70	200	+1	-1	+1	700	-1
7	10	90	100	+1	+1	-1	900	+1
8	10	90	200	+1	+1	+1	900	+1

AIR ACADEMY ASSOCIATES
Simplify, Perfect, Innovate

The Beauty of Orthogonality:
independent evaluation of effects

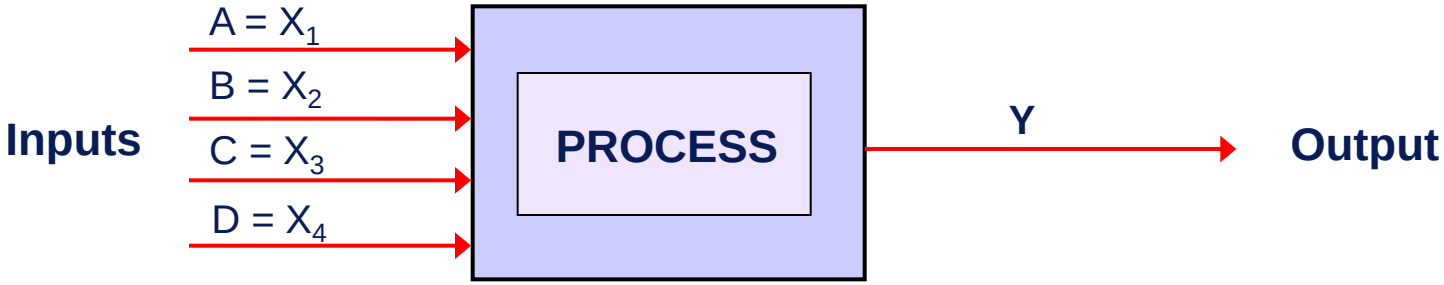
A Full Factorial Design for 3 Factors, Each at 2 Levels

Run	A	B	C	AB	AC	BC	ABC
1	-	-	-	+	+	+	-
2	-	-	+	+	-	-	+
3	-	+	-	-	+	-	+
4	-	+	+	-	-	+	-
5	+	-	-	-	-	+	+
6	+	-	+	-	+	-	-
7	+	+	-	+	-	-	-
8	+	+	+	+	+	+	+

Air Academy Associates
Simplify, Perfect, Innovate

What is a Designed Experiment?

Purposeful, systematic changes of the inputs (factors) in order to observe corresponding changes in the output (response).



	Factors				Replications			
Run	X_1	X_2	X_3	X_4	Y_1	Y_2	. . . . .	$\bar{Y}$	S_Y
1 **2** **3** . .	**Inner array**: each row or run in this test design matrix is a combination of levels of the different factors.				**Outer array:** where the response data is entered				

Air Academy Associates
Simplify, Perfect, Innovate

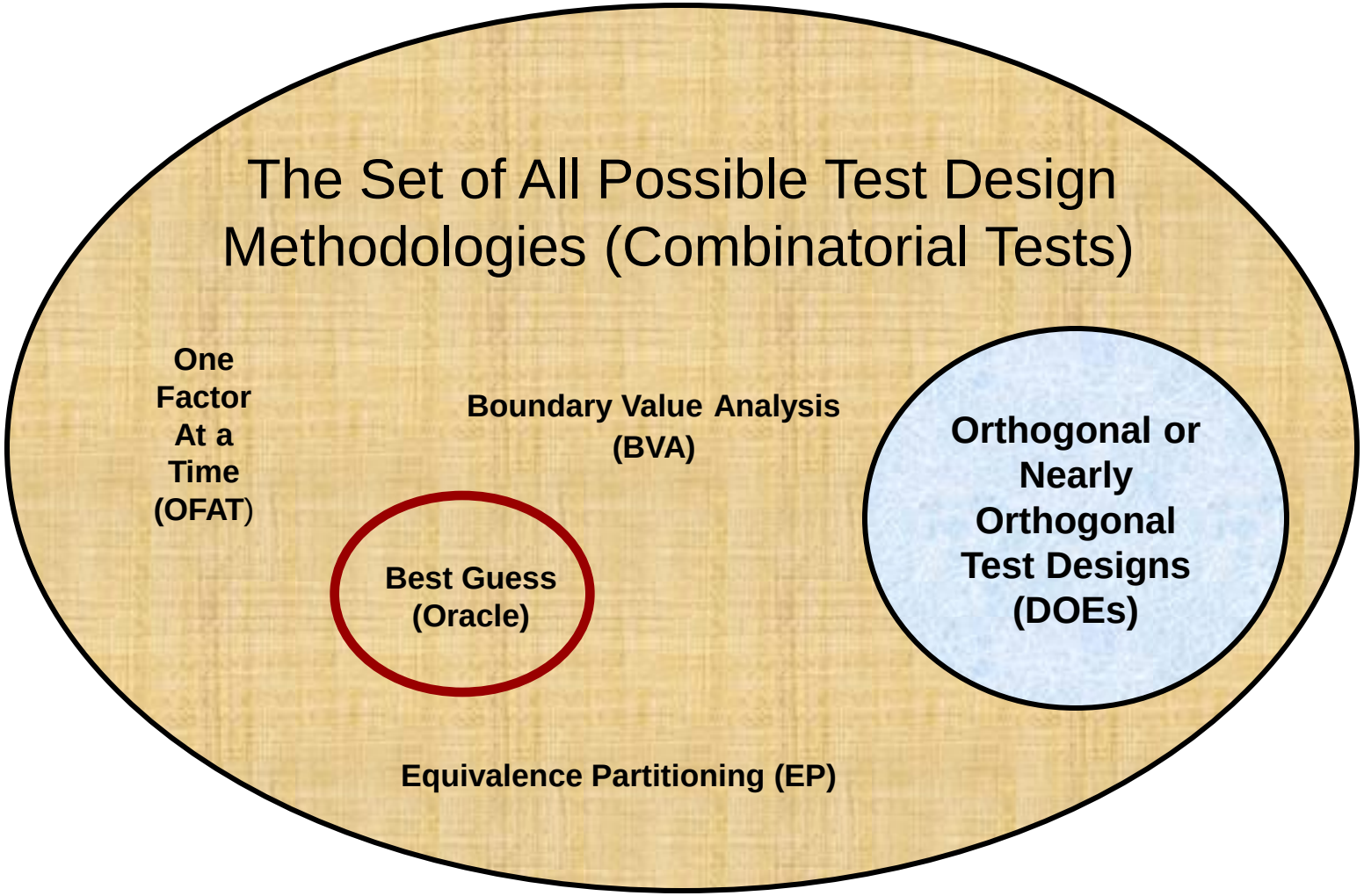
Famous Quote

“All experiments are designed experiments; some are poorly designed, some are well designed.”

George Box (1919-2013), Professor of Statistics, DOE Guru

Air Academy Associates
Simplify, Perfect, Innovate

Design of Experiments (DOEs): A Subset of All Possible Test Design Methodologies



AIR ACADEMY ASSOCIATES
Simplify, Perfect, Innovate

What can DOE do for us?

- An optimal data collection methodology
- “Interrogates” the process
- Used to identify important relationships between inputs and outputs
- Identifies important interactions between process variables
- Can be used to optimize a process and assess risk
- Changes “I think” to “I know”

Air Academy Associates
Simplify, Perfect, Innovate

Three Major Reasons for Using a DOE

- **Screening**
  - For testing many factors in order to **separate** the critical factors from the trivial many.
- **Modeling**
  - For **building functions** that can be used to predict outcomes, assess risk, and optimize performance. These include the ability to evaluate interaction and higher order effects.
- **Performance Verification and Validation**
  - For **confirming** that a system performs in accordance with its specifications/requirements.

Air Academy Associates
Simplify, Perfect, Innovate

Key Considerations for Determining the Test Design

- The Purpose of the Test
  (Screening, Modeling, Performance Validation)
- Number of Factors (k)
- Number of Levels each factor is to be tested at
- Number of replications (sample size), which will be dependent upon the desired confidence and power of the test

Two Types of Risk in Evaluating the Result of a Test

- α Risk = **P(false detection)** means we falsely concluded that a factor is important
  - **P(false detection) = p-value** (software calculates this value from the data)
  - **Confidence = [1 – p-value] x 100%**
  - Rule of Thumb (ROT) for "highly significant" result: Confidence ≥ 95%

- β Risk = **P(missed detection)** means we failed to detect something important
  - **Power = [1 – P(missed detection)] x 100%**
  - Rule of Thumb (ROT) for sufficient power: Power ≥ 75%
  - A Priori (prior to the test) power calculations are good for test planning purposes, and sample size is the way we can control the power of the test.

AIR ACADEMY ASSOCIATES
Simplify, Perfect, Innovate

Sample Size (n_{reps}) for Desired Level of Confidence and Power

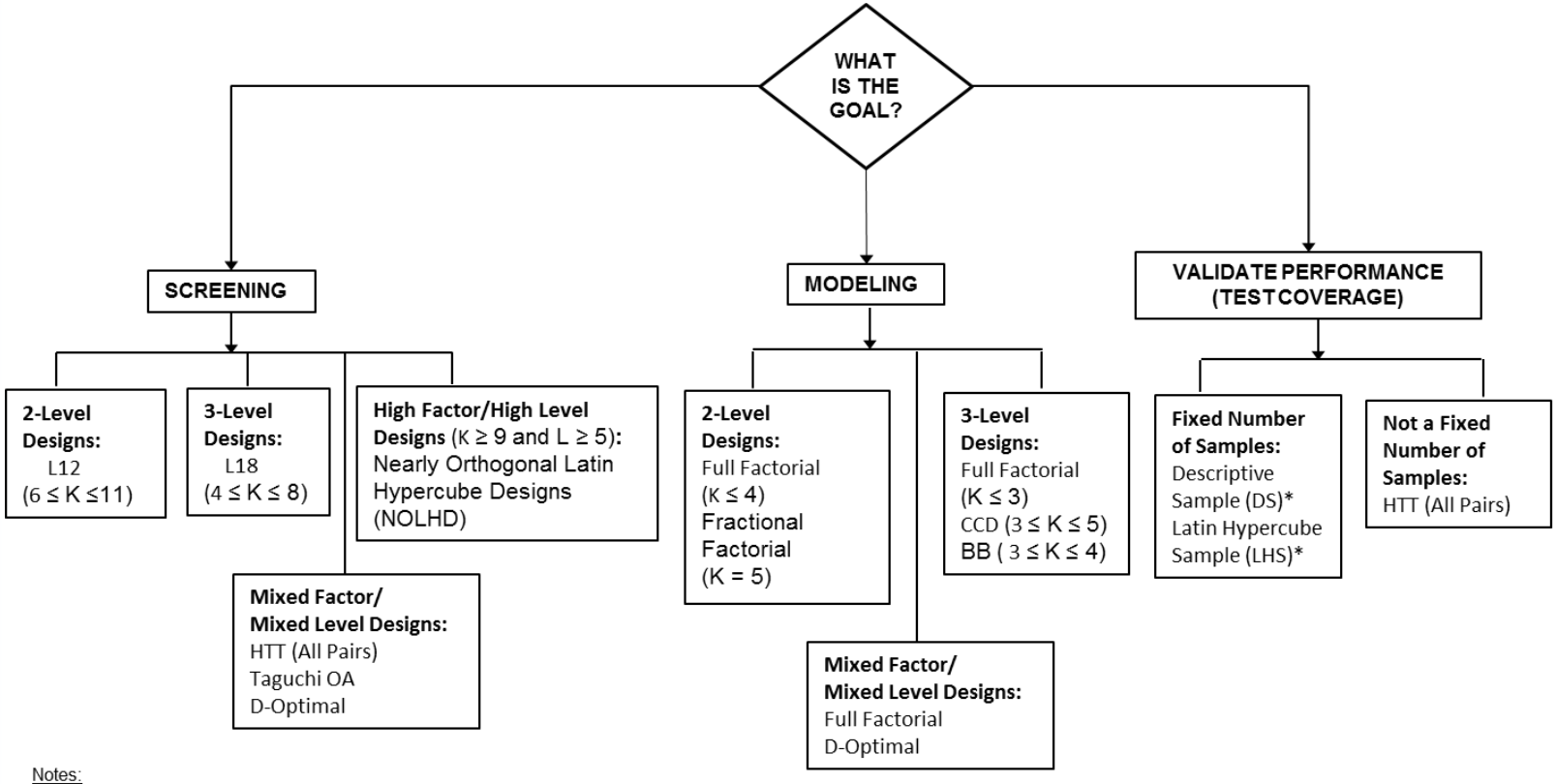
Simplified Table for Determining Sample Size Based on Confidence and Power*

Percent **confidence** that a term identified as significant truly does shift the variance [mean]	Probability in percent (or **power**) of detecting a truly significant variance [mean] shifting term	Number of Runs in 2 Level Portion of the Design				
		2	4	8	12	16
		Sample Size per Experimental Condition				
95% (α = .05)	40% (β = .60)	5 [3]	3 [2]	2 [1]	N/A	N/A
95% (α = .05)	75% (β = .25)	9 [5]	5 [3]	3 [2]	2 [1]	2 [1]
95% (α = .05)	90% (β = .10)	13 [7]	7 [4]	4 [2]	3 [2]	N/A
95% (α = .05)	95% (β = .05)	17 [9]	9 [5]	5 [3]	4 [2]	3 [2]
95% (α = .05)	99% (β = .01)	21[11]	11 [6]	6 [3]	5 [3]	4 [2]

* This table assumes an effect size for a shift in mean as small as 1.5 standard deviations and a shift in standard deviation as small as a factor of 2.7.

AIR ACADEMY ASSOCIATES
Simplify, Perfect, Innovate

KISS (Keep It Simple Statistically) Guidelines for Choosing a Test Design



Notes:

1. "Mixed" factors means a combination of quantitative and qualitative (categorical)
2. "Mixed" levels means that not all factors have the same number of levels (settings)
3. K = Number of Factors and L = Number of Levels
4. "OA" stands for Orthogonal Array
5. Software such as DOE Pro™, HD Tools™, rdExpert™ Lite, Pro-Test™ and Quantum XL™ generate some or all of these designs

* DS and LHS are sampling techniques to generate representative samples according to a specified distribution and a specified sample size

* Representative samples do not give orthogonal designs. They are often used for getting test coverage, validating performance/ determining capability, or creating noise combinations for test

AIR ACADEMY ASSOCIATES
Simplify, Perfect, Innovate

Command & Control (Software)Test Example

(15 factors each at various levels)
Total Combinations: 20,155,392

Variable or Factor	Levels	(# of levels)
Mission Snapshots	Entry, Operations, Consolidation	(3)
Network Size	10 Nodes, 50 Nodes, 100 Nodes	(3)
Network Loading	Nominal, 2X, 4X	(3)
Movement Posture	ATH, OTM1, OTM2	(3)
SATCOM Band	Ku, Ka, Combo	(3)
SATCOM Look Angle	0, 45, 75	(3)
Link Degradation	0%, 5%, 10%, 20%	(4)
Node Degradation	0%, 5%, 10%, 20%	(4)
EW	None, Terrestrial, GPS	(3)
Interoperability	Joint Services, NATO	(2)
IA	None, Spoofing, Hacking, Flooding	(4)
Security	NIPR, SIPIR	(2)
Message Type	Data, Voice, Video	(3)
Message Size	Small, Medium, Large, Mega	(4)
Distance Between Nodes	Short, Average, Long	(3)

AIR ACADEMY ASSOCIATES
Simplify, Perfect, Innovate

Command & Control Test Example

(All Pairs Testing generates 26 test cases)

	Factor_A	Factor_B	Factor_C	Factor_D	Factor_E	Factor_F	Factor_G	Factor_H	Factor_I	Factor_J	Factor_K	Factor_L	Factor_M	Factor_N	Factor_O
Factor Name	Mission	Network Size	Network Load	Movement	SATCOM Band	SATCOM Angle	Link Degradation	Node Degradation	EW	Interoperability	IA	Security	Message Type	Size of Message	Node Distance
Case 1	Entry	100 nodes	4X	OTM2	Combo	0	0%	0%	None	NATO	None	SIPIR	Voice	Medium	Short
Case 2	Consolidation	10 nodes	Normal	ATH	Ka	45	5%	5%	GPS	NATO	Spoofing	NIPR	Video	Large	Normal
Case 3	Operation	50 nodes	2X	OTM1	Ku	75	20%	20%	Terrestrial	Joint Serv	Hacking	NIPR	Voice	Small	Long
Case 4	Entry	50 nodes	2X	ATH	Ku	45	10%	10%	None	NATO	Flooding	NIPR	Data	Mega	Short
Case 5	Operation	100 nodes	Normal	OTM1	Combo	75	10%	10%	GPS	NATO	Spoofing	SIPIR	Data	Small	Normal
Case 6	Operation	10 nodes	4X	OTM2	Combo	45	0%	5%	Terrestrial	Joint Serv	None	NIPR	Video	Mega	Long
Case 7	Consolidation	100 nodes	4X	ATH	Ka	75	20%	10%	Terrestrial	NATO	Hacking	SIPIR	Video	Medium	Long
Case 8	Operation	10 nodes	Normal	ATH	Ka	0	20%	0%	Terrestrial	Joint Serv	Flooding	NIPR	Data	Large	Short
Case 9	Consolidation	10 nodes	2X	OTM2	Ku	45	5%	20%	None	Joint Serv	Flooding	SIPIR	Voice	Medium	Normal
Case 10	Consolidation	50 nodes	2X	OTM1	Combo	0	0%	20%	GPS	NATO	None	NIPR	Data	Mega	Normal
Case 11	Entry	50 nodes	Normal	OTM2	Ka	75	10%	5%	GPS	Joint Serv	Hacking	SIPIR	Voice	Large	Long
Case 12	Entry	50 nodes	4X	OTM1	Ku	0	5%	0%	None	Joint Serv	Spoofing	SIPIR	Video	Small	Long
Case 13	Consolidation	100 nodes	4X	OTM2	Ku	45	20%	5%	GPS	Joint Serv	Flooding	NIPR	Data	Small	Short
Case 14	Entry	10 nodes	2X	OTM1	Ka	75	5%	0%	None	Joint Serv	Hacking	SIPIR	Data	Mega	Normal
Case 15	Entry	50 nodes	2X	ATH	Ka	75	0%	20%	Terrestrial	NATO	Spoofing	NIPR	Video	Large	Short
Case 16	Consolidation	10 nodes	4X	ATH	Ku	0	10%	20%	Terrestrial	NATO	None	NIPR	Video	Small	Normal
Case 17	Operation	50 nodes	Normal	OTM1	Ku	75	0%	5%	None	Joint Serv	Flooding	NIPR	Data	Medium	Short
Case 18	Operation	10 nodes	Normal	OTM1	Ka	75	20%	10%	None	Joint Serv	None	SIPIR	Video	Large	Normal
Case 19	Operation	100 nodes	2X	OTM2	Combo	0	5%	10%	Terrestrial	NATO	Hacking	SIPIR	Data	Large	Short
Case 20	Consolidation	100 nodes	Normal	ATH	Combo	0	20%	20%	Terrestrial	Joint Serv	Spoofing	NIPR	Voice	Mega	Short
Case 21	Consolidation	50 nodes	2X	OTM1	Ka	45	10%	0%	GPS	Joint Serv	Spoofing	SIPIR	Data	Medium	Normal
Case 22	Entry	100 nodes	Normal	OTM1	Combo	0	20%	5%	GPS	NATO	Flooding	NIPR	Video	Medium	Long
Case 23	Operation	10 nodes	Normal	ATH	Ka	45	0%	10%	None	NATO	Hacking	SIPIR	Voice	Small	Normal
Case 24	Entry	50 nodes	4X	ATH	Ku	45	5%	20%	None	NATO	None	NIPR	Video	Large	Long
Case 25	Consolidation	10 nodes	2X	ATH	Ku	75	10%	5%	None	Joint Serv	Spoofing	NIPR	Data	Large	Long
Case 26	Consolidation	100 nodes	Normal	OTM2	Combo	45	5%	20%	GPS	Joint Serv	Spoofing	NIPR	Voice	Mega	Normal

AIR ACADEMY ASSOCIATES

Simplify, Perfect, Innovate

Questions That Should Have Been Answered

1. **What is an orthogonal test design and why is it important?**
   - It is a test design that is balanced both vertically and horizontally.
   - It makes the subsequent analysis of the test data much easier because we can evaluate each of the factors/effects independently.

2. **What are some of the key considerations when developing a test design?**
   - Purpose of the test, the number of factors to be tested, the number of levels for each factor, and the desired confidence and power of the test.

3. **What are some Rules of Thumb (or guidelines) that will allow us to test most effectively and efficiently and keep us safe in the T&E business?**
   - Always use an orthogonal or nearly orthogonal design, i.e., use DOE.
   - Guidelines for choosing a test design, depending on key considerations.
   - Guidelines for sample size (for desired confidence and power).

Thank You

Colorado Springs, Colorado

AIR ACADEMY ASSOCIATES
Simplify, Perfect, Innovate

For More Information, Please Contact

Air Academy Associates, LLC
1650 Telstar Drive, Ste 110
Colorado Springs, CO 80920

Toll Free: (800) 748-1277 or (719) 531-0777
Facsimile: (719) 531-0778
Email: aaa@airacad.com
Website: www.airacad.com

AIR ACADEMY ASSOCIATES
Simplify, Perfect, Innovate