Statistical Experimental Design for Live, Virtual, Constructive Simulation

Casey Haase, Capt, USAF
Raymond Hill, Ph.D

Overview

- Using LVC for Experiments
- Planning a Statistically Valid Experiment
- Case Study
  - Planning the Experiment
  - Compare Alternative Experimental Designs
  - Design Chosen for Experiment
- Additional Planning Considerations For LVC
- Summary

LVC for Experiments

- LVC is being considered for analytical purposes
  - Testing Systems in a Joint Environment
  - System of Systems context
  - Cost Savings over traditional Live test
- LVC can build large, complex test environments
  - Introduces new experimental design and analysis issues
  - Collecting quality data requires changes in the way users view LVC
- Experimental design techniques provide rigor needed to collect quality data from tests using LVC

Guidelines for Planning Statistically Valid Experiments

Planning a Statistically Valid Experiment

1. Recognition and Statement of the Problem
2. Choose Factors and Levels
3. Select Response Variable(s)
4. Choose an Experimental Design
5. Conduct Experiment
6. Analyze Data
7. Draw Conclusions and/or Make Recommendations

D. E. Coleman and D. C. Montgomery
A Systematic Approach to Planning for a Designed Industrial Experiment
Technometrics, Volume 35, No 1, 1993

Case Study

Case Study Background

- Aircraft are unable to transmit in denied access area
  - Transmitting makes them vulnerable to air defenses
  - Follow pre-planned routes to strike targets
    - Minimizes probability of detection
    - Limits ability to strike targets of opportunity
- Multifunctional Advanced Data Link (MADL)
  - Potentially allow friendly aircraft to talk in denied access area
  - Low probability of detection
- AF Simulation and Analysis Facility (SIMAF) tasked to assess MADL usefulness and suitability using LVC

Test Setup / Constraints

- Use LVC distributed simulation
  - SIMAF (virtual cockpit and constructive simulation host)
  - Second location (virtual cockpit)

- Test will be conducted in two phases (incremental tests)
  - Phase I – Assess MADL assuming perfect net capability
  - Phase II – Assess MADL with realistic capability degradation
  - Two weeks per phase

- Small incremental tests with one objective per test are preferred to one big test with multiple objectives

Test Setup / Constraints

- Two different aircrews will be used each week
  - 4 aircrew total
  - One pilot per virtual cockpit per week
- Human factors experts suggest aircrew can conduct 4-6 runs per day before fatigue affects data quality
- Potential for 12 – 16 runs per week

Test Setup / Constraints

- Test Environment
  - The test will be conducted using a typical operational strike environment
    - 4 different operational scenarios (vignettes)
    - Friendly strike aircraft
    - Friendly fighters
    - Enemy fighter aircraft
    - Various targets
    - Various bombing routes

Problem Statement

Experiment Planning

1. **Define and State Test Objectives**
2. Choose Factors and Levels
3. Select Response Variable(s)
4. Choose an Experimental Design

Stated Problem:

Assess the usefulness of information passed over the MADL link

- 5 months to define
- Difficulties:
  - Hard to focus on defining problem statement
    - Many people get involved – not in agreement
    - Easier to focus on building LVC, not what to study
  - Building consensus for conducting series of small experiments vice single big experiment

Experiment Factor Levels

Experiment Planning

1. Define and State Test Objectives
2. **Choose Factors and Levels**
3. Select Response Variable(s)
4. Choose an Experimental Design

- Potential Factor List with levels

Factors	Type	Levels
MADL	Categorical	3
Vignettes	Categorical	4
Route	Categorical	3
Target Position	Categorical	2
# of Red Air	Quantitative	2
# of Blue Air	Quantitative	2
Node Position	Quantitative	2
Quality of Service	Categorical	2

Response Variable

Experiment Planning

1. Define and State Test Objectives
2. Choose Factors and Levels
3. **Select Response Variable(s)**
4. Choose an Experimental Design

- No direct measurement response variable
  - Cannot directly measure "usefulness"
- Must use a surrogate measure
  - Aircrew Surveys
  - 5 point scale
    - Crew briefed about meaning of each level
    - Comments from aircrew cross check survey measurement
- Caution required
  - Must ensure response variable actually reflects stated objective

Compare Experimental Designs

Experiment Planning

1. Define and State Test Objectives
2. Choose Factors and Levels
3. Select Response Variable(s)
4. **Choose an Experimental Design**

Four different experimental designs were considered in the design process

1. 2-factor factorial design (12 runs) replicated
2. Split-Plot design (24 runs) un-replicated
3. Orthogonal Array (12 runs) replicated
4. Nearly Orthogonal Array (12 runs) replicated

Each has strengths and weakness when actually employed

Compare Experimental Designs

Experiment Planning

1. Define and State Test Objectives
2. Choose Factors and Levels
3. Select Response Variable(s)
4. **Choose an Experimental Design**

1. 2-factor factorial design (12 runs) replicated
   - MADL and Vignettes are only factors considered
   - **Strengths:**
     - Simple
     - Easy to analyze
   - **Weakness:**
     - Too Simplistic
     - Ignores environmental factors
     - Doesn't account for human variability
     - Could introduce learning bias

Compare Experimental Designs

Experiment Planning

1. Define and State Test Objectives
2. Choose Factors and Levels
3. Select Response Variable(s)
4. **Choose an Experimental Design**

2. Split-Plot Design (24 runs) un-replicated

- 3 Factors considered
  - MADL, Vignette, Crew
- MADL factor changed less frequently than other factors to prevent crew confusion
- **Strengths:**
  - Efficient experiment
  - Only way to design for restricted run order
- **Weakness:**
  - More difficult to analyze
  - 2 error terms
  - Whole plot factors get less precise estimates
  - Still ignores some environment effects

Compare Experimental Designs

Experiment Planning

1. Define and State Test Objectives
2. Choose Factors and Levels
3. Select Response Variable(s)
4. **Choose an Experimental Design**

3. Orthogonal Array (12 runs) replicated

- 4 Factors considered
  - MADL, Vignette, Target Location, Route
- **Strengths:**
  - Environmental factors included in the design
  - Can accommodate up to 12 factors and still estimate the main effects
  - Replication allows for more precise estimate of error
  - Some interaction effects can still be estimated
  - Straightforward to analyze main effects
- **Weakness:**
  - Analysis becomes complicated if interactions present
  - Lose ability to estimate all high order interactions

Experiment Planning

1. Define and State Test Objectives
2. Choose Factors and Levels
3. Select Response Variable(s)
4. **Choose an Experimental Design**

4. Nearly Orthogonal Array (12 runs)

- 6 Factors considered

  MADL, Vignette, Target Location, Route, Node Position, Quality of Service

- **Strengths:**

  Environmental factors included in the design

  Accommodates more factors than Orthogonal Array

  Replication allows for more precise estimate of error

  Some interaction effects can still be estimated

  Straightforward to analyze main effects

- **Weakness:**

  Analysis more complicated – correlated estimates

  Correlation creates less precise estimates

Chosen Experimental Design

- Phase I
  - Orthogonal Array (3 x 4 x 2^2)
    - 4 factors
    - Allows for replication
    - Accommodates environmental factors
    - Can obtain more precise estimates of effects and error
- Phase II
  - Nearly Orthogonal Array
    - 6 factors (3 x 4 x 2^4)

      Quality of Service and Node Position added to design
    - 2 additional factors make orthogonal design impossible
    - Allows for replication
    - Caution when analyzing, error and effects may contain bias

Lessons Learned from LVC Case Study

Additional Planning Considerations for LVC

- LVC was originally intended for training purposes
  - Little analytical rigor necessary for training
  - Little up-front planning required for **post-ops analysis**
  - High-fidelity, complex, noisy environments preferred
- T&E shares many resource requirements with training
  - Hence, LVC becoming central to DoD test strategy
- Requires new paradigm to use LVC effectively
  - Extensive up-front planning required
  - Excessive fidelity & complexity can & will ruin experiment

Additional Planning Considerations for LVC

- Defining clearly stated objectives more difficult with LVC
  - As ability to create test environment gets better the number of potential test objectives get larger
  - There is a lure toward complexity
    - Requires discipline to scope test with realistic goals
    - Requires discipline to keep experiment within scope of study
    - Just because you can do something doesn't mean you should

- Response variables are not always obvious with LVC
  - Objectives are often qualitative
  - Response Variables are not obvious
  - Surrogate measures need to accurately reflect objectives

Additional Planning Considerations for LVC

- Testing in Joint Environments using LVC
  - Joint mission environments contains copious noise
  - Experimenters must:
    - Be aware of noise
    - Account for noise by using statistical noise control methods
  - Human Operators are one of the largest sources of noise

- LVC experiments produce abundance of data
  - Extra effort required to plan
  - Else effort wasted collecting, sifting, and analyzing data

Summary

- DoD seeks use of LVC for analytical purposes (T&E)
- Experimental design creates solid framework for conducting experiments that result in valid conclusions for LVC experiments
- LVC introduces additional considerations into the experimental design process
- Case study illustrates benefits of using statistical experimental design methods for LVC

References

- D. E. Coleman and D. C. Montgomery, *A Systematic Approach to Planning for a Designed Industrial Experiment,* Technometrics, Volume 35, No 1, 1993

- Department of Defense, *Testing in a Joint Environment Roadmap,* https://extranet.dote.osd.mil/pub/otherrep.html, *2004.*