Key Aspects of STAT in DoD Test and Evaluation

OSD STAT T&E COE
www.AFIT.edu/STAT
937-255-3636 x 4736

Outline

- STAT Frame of Reference
- Unique DoD Challenges
- The STAT in T&E COE Introduction
- STAT Focus Areas

STAT Frame of Reference

"Have a clear idea ***in advance*** of exactly ***what*** is to be studied, ***how*** the data are to be collected, and at least a qualitative understanding of how these data are to be ***analyzed***."

Montgomery ***Design and Analysis of Experiments*** 8th Ed.

What is STAT?

Scientific Test and Analysis Techniques (STAT) are the scientific and statistical methods and processes used to enable the development of efficient, rigorous test strategies that will yield defensible results.

The results of the **repeated use** of STAT is a more progressive sequential testing approach that carefully **leverages past test** information, along with **informing** the systems engineering process.

- STAT tools:
  - Design of Experiments
  - Quasi-Experimental designs
  - Natural Experiments
  - Hypothesis Testing
  - Regression Analysis
  - Data Analysis Techniques
  - Operating Characteristic curves
  - Optimization
  - Distribution fitting (historical data)
  - Reliability growth
  - Observational Studies
  - Surveys

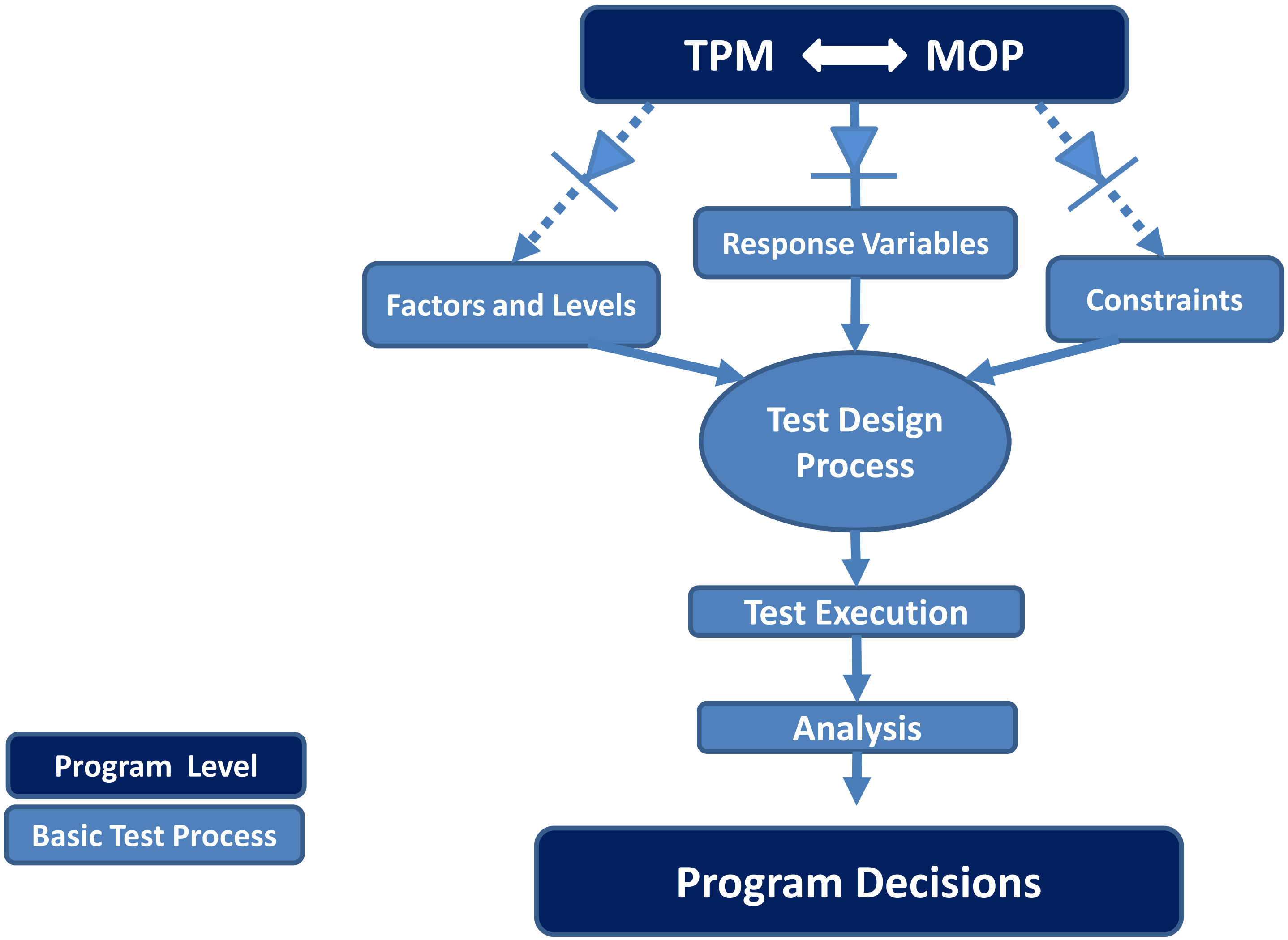
Recognized DoD Challenges

- Requirements
  - Often complex and/or ambiguous
  - Typically not well defined for the application of STAT
  - DOT&E/DASD(DT&E) working to improve req'ts definition
- Complex systems and conditions
  - Systems of systems, multiple missions, many scenarios
  - Operating conditions are broad and diverse
  - Many factors are uncontrollable or hard to change

A systematic approach is key to definitively and efficiently testing DoD programs

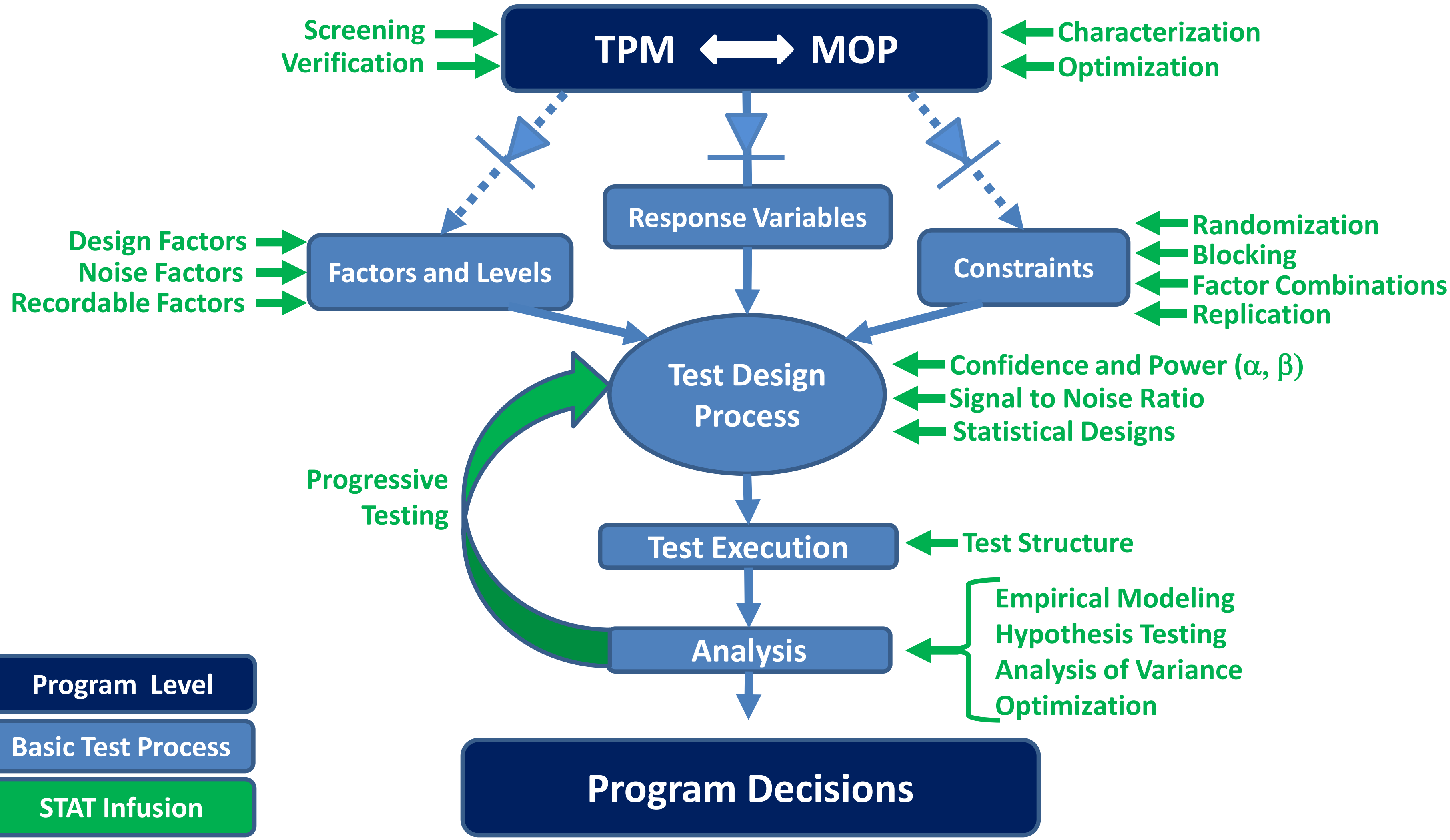
STAT in the Test Process

STAT T&E COE: Scientia Prudentia et Valor



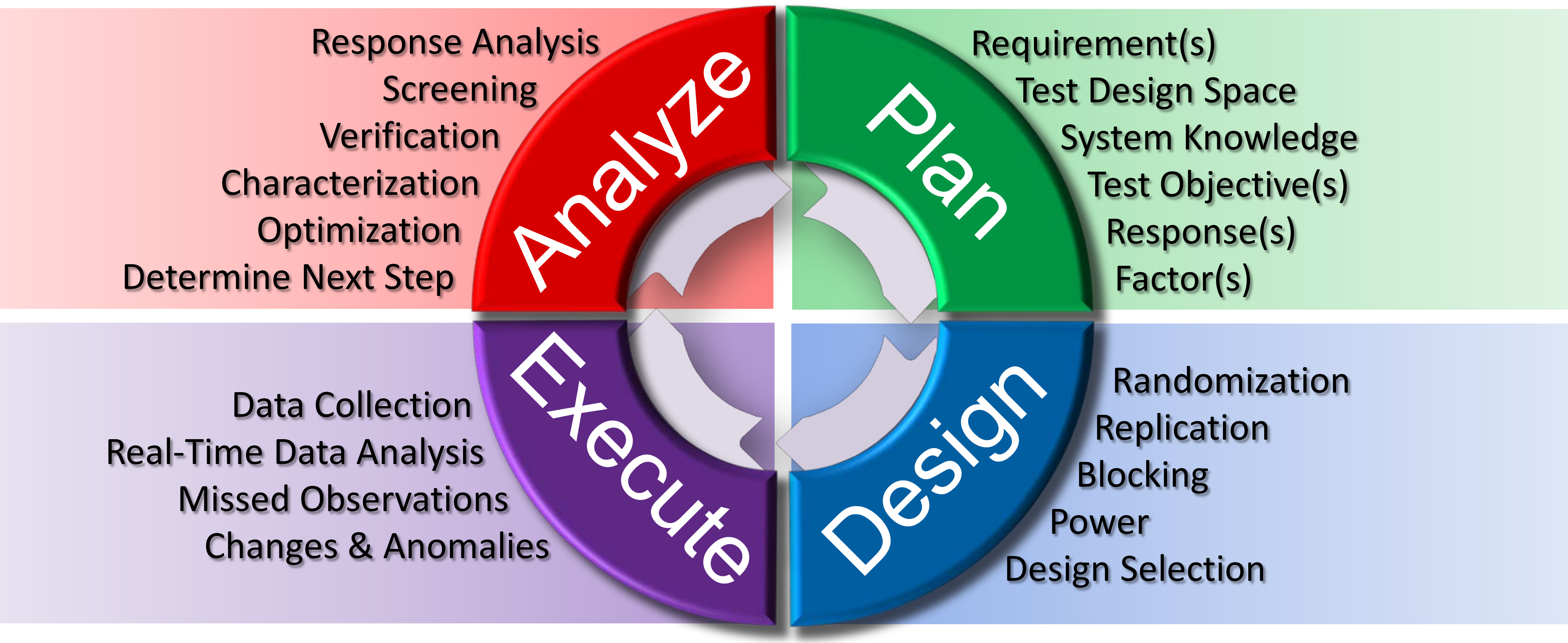
STAT in the Test Process

STAT T&E COE: Scientia Prudentia et Valor



The Test Design Cycle

STAT T&E COE: Scientia Prudentia et Valor



Focus Areas

- A Basis in Statistics
- From Requirements to Test Space
- Categorical/Continuous Factors & Responses
- End-to-End Mission Decomposition
- Blocking and Randomization

A Basis in Statistics

The ***Fun***damentals

Random Variation

STAT T&E COE: *Scientia Prudentia et Valor*

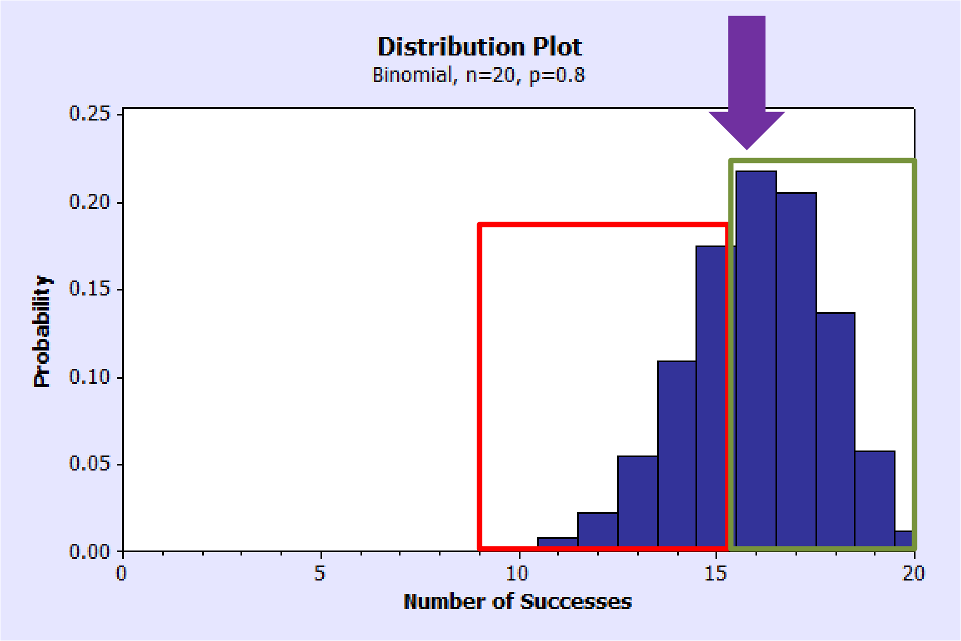
- The outcome of a trial classified as "success" or "failure."
- Suppose the true (***unknown***) probability of success is 80%.

80% = 16 successes in 20

63% chance of at least 16

22% chance of exactly 16

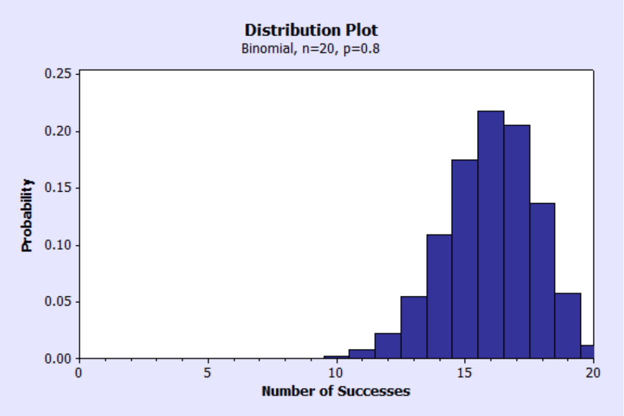
37% chance of 15 or less



Probability distribution of successes if 20 trials are completed.

Random Variation in Testing

- "Statistics is the science of variation." Douglas M. Bates
- Variation is at work in **every system.**
- In reality, we **do not know** the underlying probability.
- **Testing is sampling** from the true (unknown) population
- Suppose we **observed** 15/20 successes (75%)
  - 75% **does not prove** it at 75%, **nor fail** a true 80% system



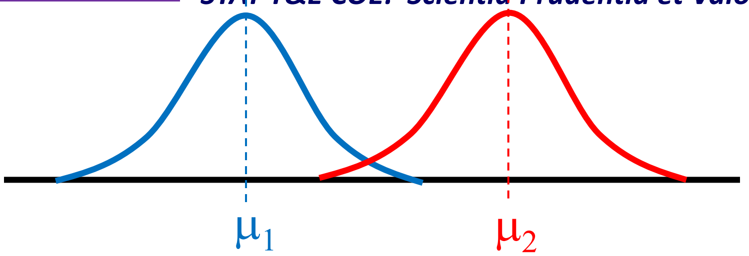
Accounting for variation through statistical analysis techniques will significantly improve test designs

From Requirements to Test Space

Where Are You Going?

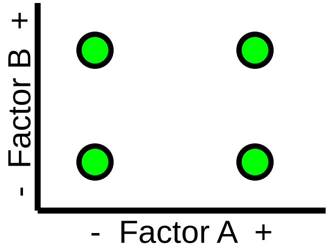
What Is the Goal of the Experiment?

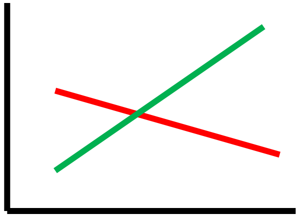
Comparison



- t-test
- Sampled at same factor levels
- Compare 2 or to requirement
- "**Point**" test

Characterization

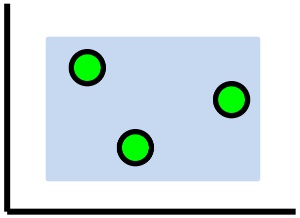


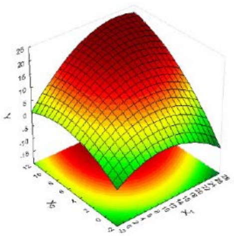


Interaction plot

- Significant factors
- "**Space**" test

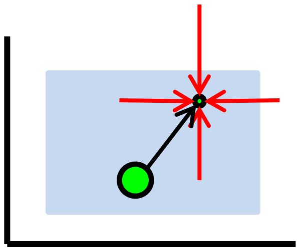
Modeling





- Regression equation
- Prediction/verification
- Error assessment

Optimization



- Small region of interest
- Optimize a response
- Select best factor levels

Point and Space: What is the Difference?

- Compare (Point)
  - A single point in the envelope (**same conditions**)
  - Expect repeated responses to be **close in magnitude**
  - ***Example:*** Highway MPG test (55 mph, paved, level)
- Characterize (Space)
  - Performance across the spectrum (**varying conditions**)
  - Expect responses will **vary across the space**
  - ***Example:*** MPG across spectrum (varied weight, altitude, speed, configuration, road surface)

Determine if your test requirements demand Point and/or Space information

Modeling and Optimization

- Modeling (prediction)
  - Regression equation derived for each response (software)
  - Can be built upon from “space” testing designs
  - Performance prediction for use in OT scenario validation
  - Perf models inform decisions on long term programs
- Optimization (best factor settings)
  - Maximize/minimize a desired response (look at sensitivity)
  - Explores a smaller region of interest
  - Supports development of tactics and techniques
  - Can support engineering decisions

Categorical/Continuous Factors and Responses

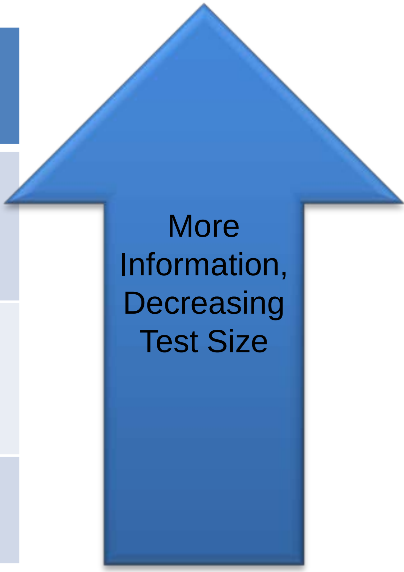
Ensuring Data Granularity

Data Type

- Metric choice alone can increase test resources by 50% or more*

** Gilmore, M. "The Role of Test and Evaluation in Defense System requirements"*

Data Type	Definition	Examples	Information Content
Continuous	Data can take on an infinite number of values	Detection range, Time until event	Most Information
Ordinal	Data with discrete values that imply an ordering relationship	Rank order of preferences on a Scale of 1-5, Order in Races, Letter Grades	More Information
Binary	Data can only assume one of two values	Pass/Fail, Hit/Miss, Detect/Non-Detect	Less Information

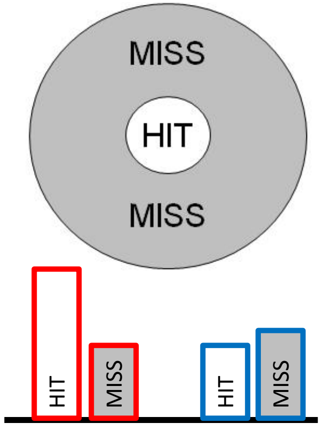


Continuous variables enable the most efficient use of resources

Data Type (Responses)

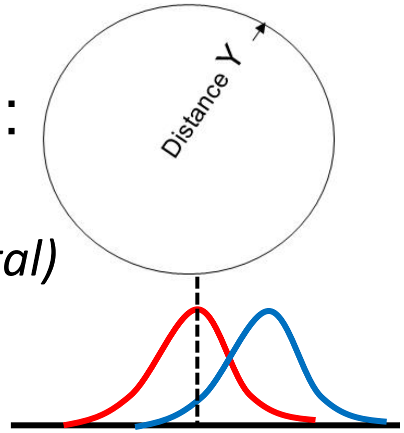
Darts Example: Compare 2 player's HIT performance

- 95% confidence (5% α) and 80%+ power (20% β)



- ***Binary*** response:
  - Hit/Miss
  - *Requires **96** runs each (192 total)*

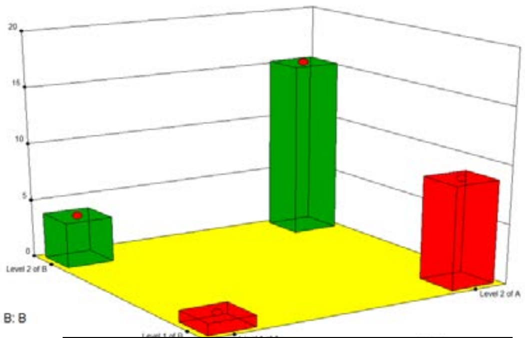
- ***Continuous*** response:
  - Radial Distance
  - *Requires **17** runs each (34 total)*



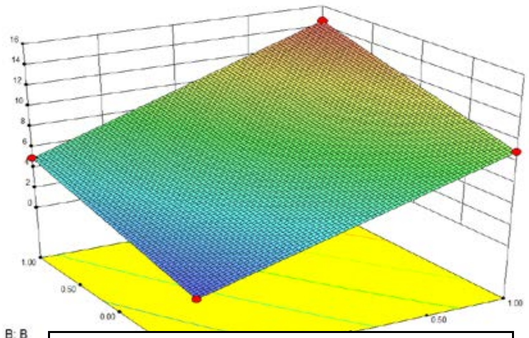
Just because the requirement is a Probability of Hit (P_H) does not mandate how the data is collected!

Data Type (Factors)

- Example categorical factor: Speed Lo/Hi
  - Data will be lost if speed is a *factor of interest*
  - Convert to continuous where practical: MPH
- Categorical masks information (imprecise)
  - What is “low” speed: 5, 10, 15 MPH?
- Cannot interpolate b/w categorical points
- **Continuous factors require fewer runs!**



Categorical Response Plot



Continuous Response Plot

More effort defining continuous factors up front will greatly improve the quality of the analysis

Decomposition and Mission Relation for Test Planning

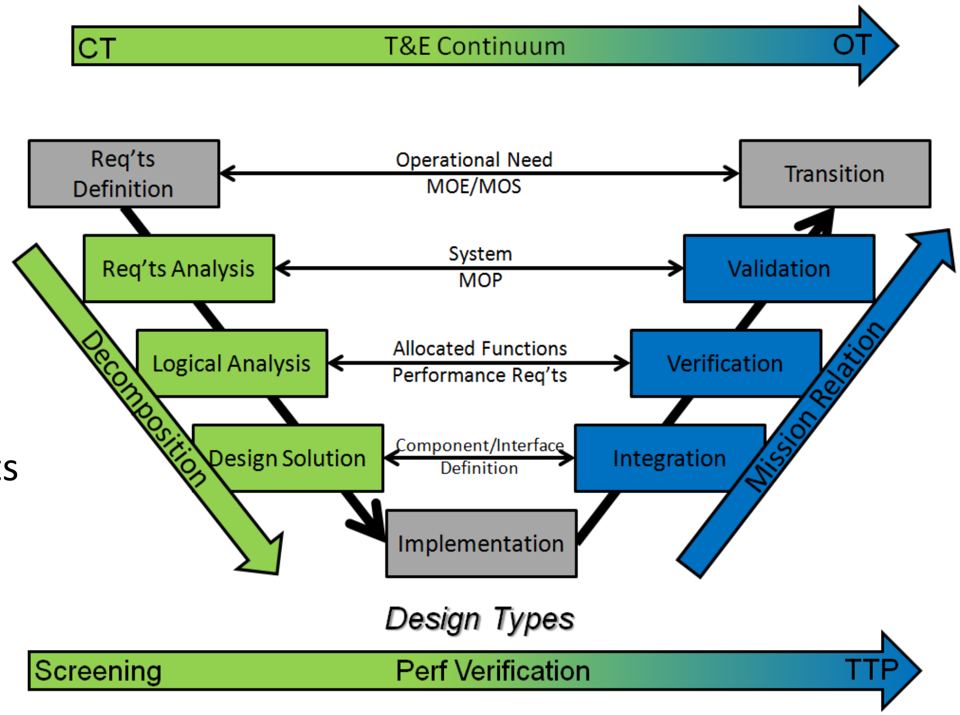
The System Engineering Process Informs Good Testing: Keys to Defining Testable and Measurable Responses

Decomposition and Mission Relation for Test Planning

STAT T&E COE: Scientia Prudentia et Valor

Process Principles

- Leverage **system components**
  - Defines low level testable elements
  - Screen factors early in DT
  - Develop models in DT (not just point check)
- Understand **mission segments**
  - Assign mission-related (derived) requirements
  - Include significant DT factors
- Tailor test space for OT
  - Critical operational regions or DT risk areas
  - OT events can verify DT model predictions



Decomposition helps define DT scope and allows assignment of **mission related requirements** for OT

Decomposition and Mission Relation for Test Planning

STAT T&E COE: Scientia Prudentia et Valor

- Decomposition (from **mission level** to **component**)
  - Decomposition facilitates better test planning **granularity**.
  - At low level, **responses are clearly defined**, factors/levels easily assigned.
  - Helps DT information to directly support OT planning and analysis.

- Mission relation (from **component** to **mission level**)
  - End-to-end OT events typically **resource limited**; mission has many factors
    - Factors may be constrained, hard to change, or uncontrollable.
  - Utilize info gleaned in DT: **significant** factors, performance **predictions**
    - Helps to identify how factors influence each component of the entire process
  - OT design should build from a series of focused & well controlled DT events
  - Assign mission-related MOP/MOE responses to mission segments
    - Segment responses are clearly relatable to top level requirements
    - Factors/levels are easily selected to facilitate assessment of segment (e.g. Detect, Track, Attack)

Notional Awesome Example

Characterize the Batmobile!

- Start with mission level requirements**:** Speed, Stealth, Detect (P_D), Lethality (P_K)
- Decompose system into **detailed** & **functional** components for good DT design
  - Brakes, engine, steering, suspension, interior, missiles, Bat-a-rang, etc.
- Determine **responses/factors** for each component and DT methodology
  - Brakes – friction, heat, distortion. Bench test-> test rig-> mock-up
  - Engine – power, heat, noise, vibration, etc
  - Screen factors, characterize and predict performance
- Decompose OT **mission** segments: Stealth Drive, Detect Perp, Shoot, Paddy-Wagon
  - Assign **mission related responses** to segments; max counter-detect range, P_K, # seats
  - Integrate applicable significant factors into OT (screened in DT).
  - Define mission test space to **verify** integrated performance and/or **examine risk**

STAT helps reduce and quantify risk to determine a "right sized" design.
Notice the ***design matrix is an output*** of the whole process.

Blocking and Randomization

Why It Matters

Randomization

- Randomization
  - A **fundamental principle** in test design
  - Prevents potential bias (*cannot remove it afterwards*)
  - Prevents confounding of factors with *lurking variables*

- Example: Range at which an operator identifies a target
  - Two types of targets (10 presentations each)
  - Test strategies:
    - Randomized
    - 1 target type (A/B) at a time

Randomization Example: Designation Range

STAT T&E COE: *Scientia Prudentia et Valor*

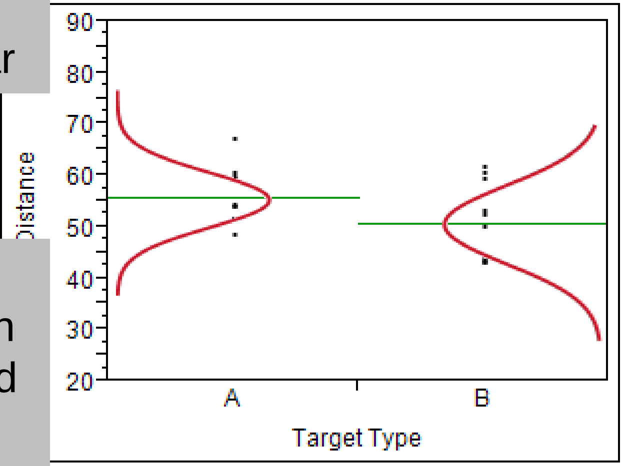
What if there is a learning effect?

- Why did the average range increase?
- Is the ***difference significant***?

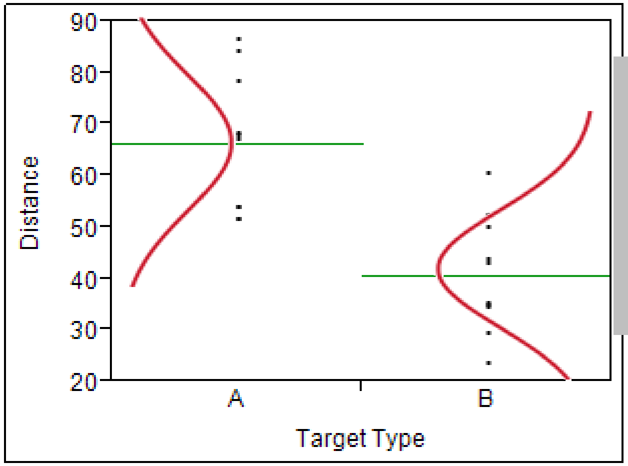
Performance ***appears*** similar

In reality, performance on B was improved ***after*** doing A

Not Randomized
(Type A then B)



Randomized



Randomized testing reveals the ***true difference***

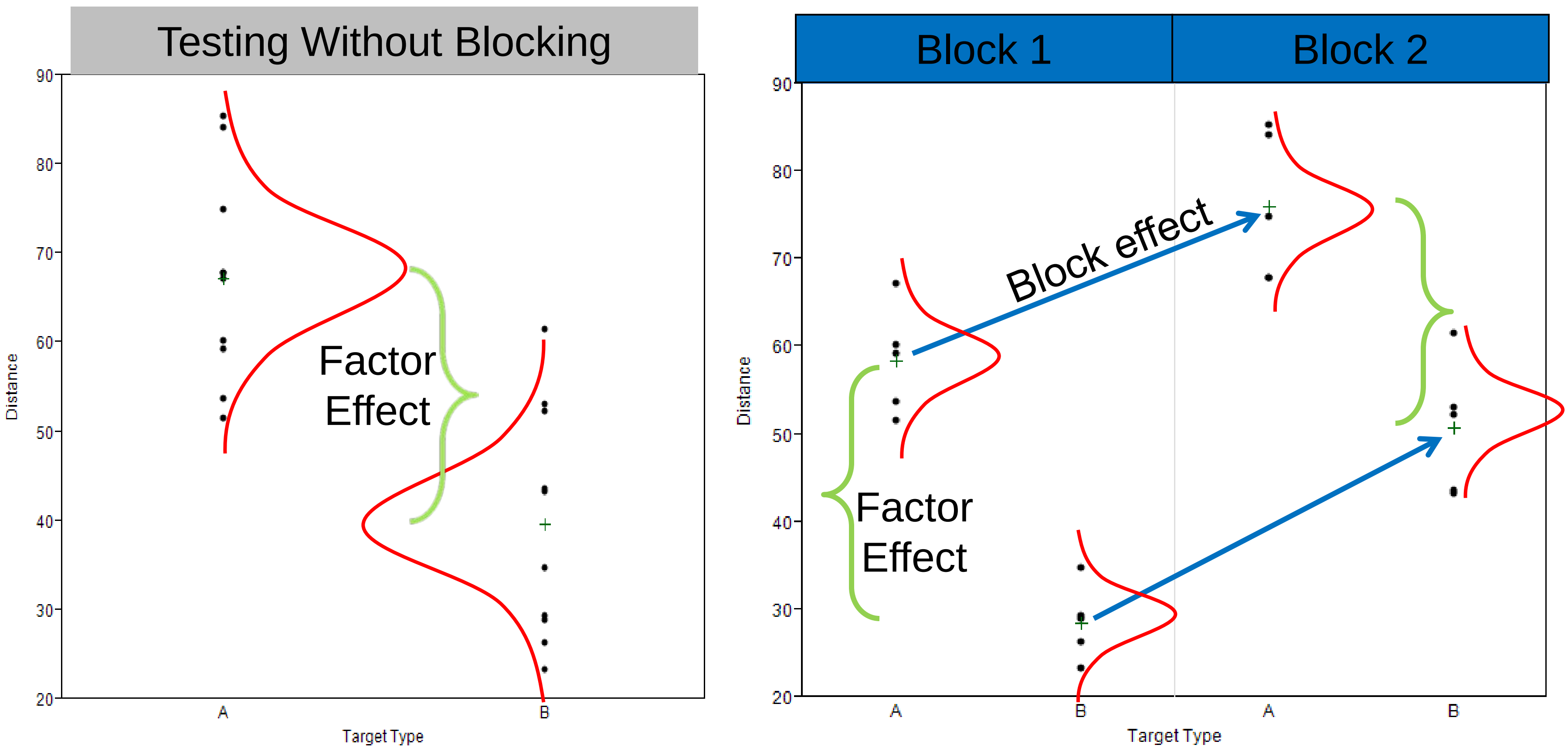
Randomization helps balance out lurking variables, making it ***easier to detect*** the effect of target type

Blocking

- Blocking
  - A **fundamental principle** in test design.
  - Reduces response variability due to nuisance factors.
  - Increases the ***power*** for detecting a factor effect.
- Identification range example:
  - Concerned there may be a learning effect
    - Which we *are not interested* in
  - Run experiment in blocks, randomize within each

Blocking Example: Identification Range

STAT T&E COE: Scientia Prudentia et Valor



Blocking helps to **isolate** the effect of target type

Conclusions

- STAT rigor improves test efficiency throughout
- Identify available STAT resources!
- Know where you need to go with your analysis
- Decompose the problem into small pieces
- Let STAT help with DOD system complexity

"To call in the statistician after the experiment is done may be no more than asking him to perform a postmortem examination: he may be able to say what the experiment died of." R. A. Fisher

Questions?

OSD STAT T&E COE

www.AFIT.edu/STAT
937-255-3636 x 4736