STAT COE-Report-09-2015

STAT Planning Tool

Authored by: Francisco Ortiz, PhD

24 April 2015

Revised 31 Aug 2018

The goal of the STAT COE is to assist in developing rigorous, defensible test strategies to more effectively quantify and characterize system performance and provide information that reduces risk. This and other COE products are available at www.AFIT.edu/STAT.

STAT Center of Excellence
2950 Hobson Way – Wright-Patterson AFB, OH 45433

Table of Contents

Revision 1, 31 Aug 2018, Formatting and minor typographical/grammatical edits.

Executive Summary

This best practice serves as help documentation detailing how to use the Scientific Test and Analysis Technique (STAT) planning tool available at www.afit.edu/STAT. The STAT planning tool was created to aid test teams in following the STAT process and better ensures the development of a defensible and rigorous test strategy that creates traceability from requirements to analysis. For best results, the tool should be filled out in a collaborative environment by a STAT working group consisting of system and test engineers, lab/field representatives and STAT experts. Using this tool will:

- Address key questions for each specific STAT phase.
- Achieve working group consensus on the objectives of test.
- Define a clear operational/test space to be explored.
- Obtain key inputs for TEMP and test plans (e.g. responses, factors, levels, experimental design considerations, etc.)

In the following sections we will provide a brief summary of the STAT process, details for each section within the tool as well as an example demonstrating how to fill out the tool.

Keywords: Design of Experiments, TEMP, test planning, test strategy, Scientific Test and Analysis Technique

Background

The STAT process prescribes an iterative procedure that begins with the requirement and proceeds through the generation of test objectives, designs, and analysis plans, all of which may be directly traced back to requirements. Figure 1 provides a process flow diagram that summarizes the application of STAT to the test and evaluation process.

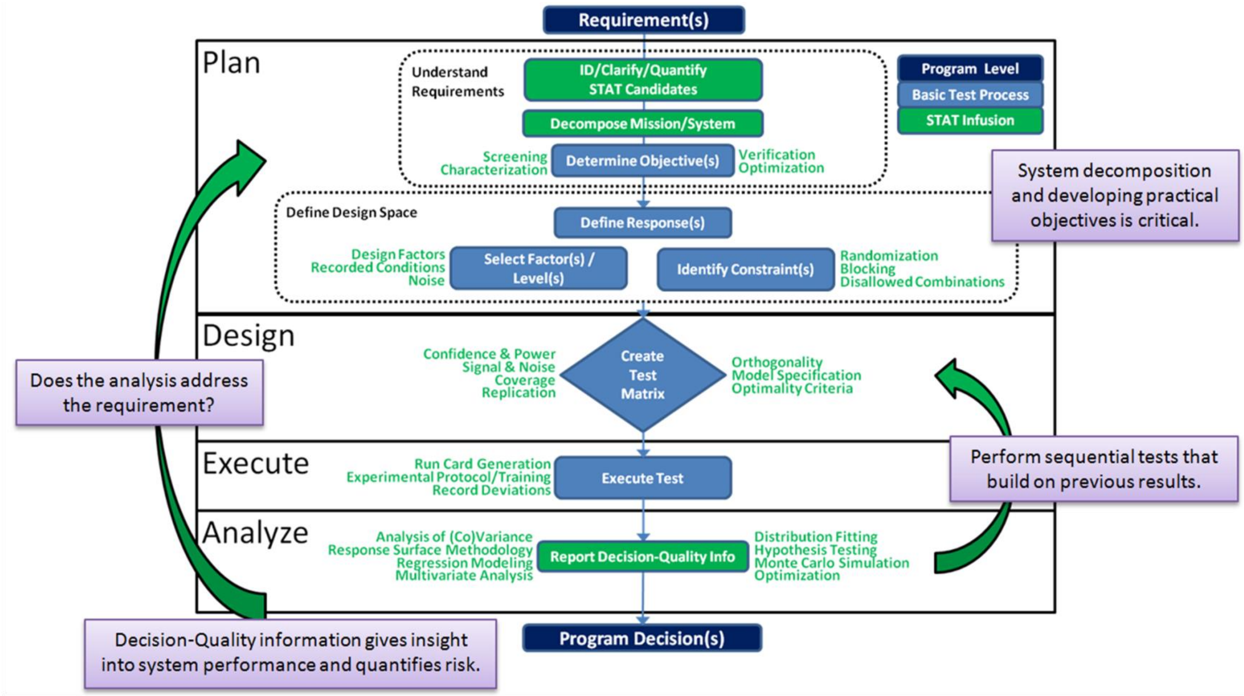

Figure 1: STAT Process Flow Chart

The process is broken up into four phases (Plan, Design, Execute, and Analyze). The following is a brief description of what takes place at each phase:

- Plan:
  - Examine and try to understand the requirements.
  - Decompose system or mission down into smaller more manageable segments.
  - Set the objectives of specific test events.
  - Identify responses (performance measures of interest).
  - Define the operational/test space by listing potential factors and their levels.
  - Identify constraints that limit the test space.
- Design
  - Create a test matrix that efficient and effectively uses available resources and achieves objectives laid out in the “Plan” stage.
  - Provide the tester with an exact roadmap of what and how much to test.
  - Provide the framework for future statistical tests that will focus on identified significant factors.
- Execute
  - Run design experiment according to the appropriate protocol.
- Analyze
  - Obtain conclusions on the efficacy of the system.

- Document, organize, and then report to decision makers.

The STAT planning tool is focused on collecting the information needed in the “Plan” phase but it is necessary to consider the other phases of the process in order to ensure a cohesive strategy is created that will result in decision-quality information. For more details on the STAT process please see the STAT Planning Guide, also available at www.afit.edu/STAT

Method

Test Planning Tool

In the following sections we will walk through the STAT planning tool. We will use a generic hardware-in-the-loop (HWIL) test event for a counter measure (CM) system as an example. **All information provided is notional and for demonstrative purposes only.**

Section: General Information

Record some basic information regarding the test event and the makeup of the STAT working group. The STAT working group should consist of system and test engineers, lab/field representatives and the STAT experts.

Field Descriptions

- **Point of Contact:** This should be the owner of the document, either the Test Lead or team STAT Expert.
- **Organization:** Self-explanatory.
- **Project title:** Self-explanatory.
- **Test Event Name:** Self-explanatory.
- **STAT Working Group Members:** List STAT working group members by Name, Organization, and Title.

Section: Test Event Description and Relevant Background

Provide background on the test event. The information here should motivate discussion amongst the STAT working group about relative test event knowledge. In this section briefly describe:

- The test event
- Previous relevant test entries
- Computational results prior to experiment
- What new knowledge is to be obtained
- Uncover possible regions of interest or regions that should be avoided
- Expert knowledge or other experiences

SAMPLE OUTPUT

Test Event Description and Relevant Background

The HWIL simulation test equipment includes a flight motion simulator, simulation computer, hardware interface adapters, a target/scene projector system and associated data capture equipment. This architecture will allow the stimulation of a missile threat in a closed-loop environment. *The HWIL has been used to test past legacy systems. Data from legacy system are available and can be used to assess expect system variation and therefore will be used to estimate signal-to-noise ratios for the test.* *Testing will evaluate CM jamming effectiveness for selected missile threats.* Specifically, the system *needs to be able to output enough laser energy onto the threat, with the correct jam code output required to be able to achieve optical break lock.* *Calculation of probability of optical break lock (P_{OBL}) is a function of multiple responses.* *This test will focus on operationally relevant region of the total potential test space. Test will focus on highest priority threats.* *Previous tests of similar systems suggest we can expect to obtain approximately 200 data points per week. Lab is booked for 10 weeks.*

Figure 2: Test Event Description and Relevant Background Example

Section: Requirements

List the specific requirements that will be addressed either fully or partially by this test event. The information, purpose, and intent contained in the requirements drive the entire process. All subsequent steps of the STAT process support the selection of test points that will provide sufficient data to definitively address the original requirement.

SAMPLE OUTPUT

REQUIREMENTS

Requirement #	Description
1	CM must achieve optical break lock for individual threats with threshold and objective P_{OBL} of 0.7 and 0.8, respectively.

Figure 3: Requirement Example

Section: Test Design Objectives

The objectives are derived from the requirements and serve to focus resources and test designs toward generating sufficient information to address the requirement in a clear, quantitative, and unambiguous way. Common types of objectives are available in a drop down:

Field Descriptions

- **Characterization**: Model performance across an operational envelope.
- **Screening**: Identify the key factors affecting performance.
- **Optimization**: Find settings of controllable factors that achieve ideal performance.
- **Comparison**: Compare performance to a baseline, standard, or goal.
- **Other**: Provide an explanation.

SAMPLE OUTPUT

TEST DESIGN OBJECTIVES

Objective #	Nature	Description
1	Screening	Identify significant factors that affect optical breaklock (OBL) for the CM systems.
2	Characterization	Characterize various responses of the CM systems across test space.

Figure 4: Test Design Objectives Example

Section: Responses

Responses are the measured output (performance measures) of a test event. There may be several responses measured for a given test and/or in support of a requirement. Process maps are a good tool for determining the various performance measures (responses) that can be recorded. See Simpson et. al. 2013.

Field Descriptions

- **Name:** Name of performance measure/response.
- **Data Type:** Specify the response data type. Options are:
    - **Continuous:** Data take a value based on a measurement at any point along a continuum. The value given to an observation for a continuous variable can include values as small as the instrument of measurement allows.
    - **Discrete:** Data take a value based on a count from a set of distinct whole values. A discrete variable cannot take the value of a fraction between one value and the next closest value.

  - **Nominal:** Data take on values that are not able to be organized in a logical sequence. Binary data (0-1, on/off, hit/miss) is an example of a nominal data.
  - **Ordinal:** Data take on values that can be logically ordered or ranked. The categories associated with ordinal variables can be ranked higher or lower than another but do not necessarily establish a numeric difference between the each category.
  - **Other:** Provide an explanation.

Note: The data type chosen to represent responses in an experiment can have a major effect on the resources needed to conduct an experiment and the quality of its respective analysis. In the case of responses, nominal data types contain a relatively poor amount of information compared to continuous data types. See Ortiz, 2014.

- **Unit of Measure:** Self-explanatory.
- **Range**: Expected interval values for a response across the test space.
- **Accuracy:** Precision of measurements. The smallest measurement allowed by instrumentation. This is used to verify that the measurements have the precision and accuracy required of the experiment's objective.
- **Difference to detect (δ):** The desired detectable change in the response. This will be used to calculate signal-to-noise ratio. System subject matter experts are the best source of information in determining δ.
- **Estimated standard deviation (σ):** Estimate of the natural variation in the system**.** Pilot studies or historical data of similar systems under like conditions usually lead to sufficient noise (σ) estimates.
- **Signal-to-Noise:** Calculation (δ/σ).
- **Priority:** Use to rank importance among the responses considered. High rated responses should have a higher influence on what factors should be considered and what regions of the test space should be explored.

SAMPLE OUTPUT

RESPONSES

Number	Name	Data Type	Unit of Measure	Range	Accuracy (+/-)	Diff to detect (Delta/Signal)	Est Std Dev (Sigma/Noise)	Signal to Noise Ratio	Priority
1	Acquire Time	Continuous	seconds	0-10	0.1	1	0.5	2	Low
2	Acquire Range	Continuous	meters	1000-3000	5	50	50	1	Medium
3	Track Time	Continuous	seconds	10-15	0.1	1	0.5	2	High

Figure 5: Test Design Objectives Example

Section: Factors and Levels

Factors are inputs into or conditions for a test event that may influence the variability in the response. Factors can be derived from prior testing, system knowledge, or insight into the underlying physics of the problem. Levels are the settings for each factor in a design. A fishbone diagram is a good tool for brainstorming potential factors of interest. See the Guide to Developing an Effective Test Strategy (2017), also available at www.afit.edu/STAT

Field Descriptions

- **Factor Name:** Name of potential input factor.
- **Data Type:** Specify the factors data type. Options are:
    - **Continuous:** Data take a value based on a measurement at any point along a continuum. The value given to an observation for a continuous variable can include values as small as the instrument of measurement allows.
    - **Discrete:** Data take a value based on a count from a set of distinct whole values. A discrete variable cannot take the value of a fraction between one value and the next closest value.
    - **Nominal:** Data take on values that are not able to be organized in a logical sequence. Binary data (0-1, on/off, hit/miss) is an example of a nominal data.
    - **Ordinal:** Data take on values that can be logically ordered or ranked. The categories associated with ordinal variables can be ranked higher or lower than another but do not necessarily establish a numeric difference between the each category.
    - **Other:** Provide an explanation.

Note: The data type chosen to represent factors in an experiment can have a major effect on the resources needed to conduct an experiment and the quality of its respective analysis. Using categorical data types to describe factors may also have a detrimental effect on the overall size of the test and the quality of the analysis. See Ortiz, 2014.

- **Unit of Measure:** Self-explanatory.
- **Range:** List levels of interest for each factor.
- **Desired Number of Levels:** Recommend just two levels (e.g. high and low) especially for initial screening experiments.
- **Factor Change:** An important step in the test planning process is to identify any restrictions on the test design or execution. This identification is critical because any restrictions will influence the design choice and analysis. Common constraints include the budget, the experimental region, difficulty changing factor levels, and restrictions on randomization. If there are hard-to-change factors and the runs cannot be completely randomized, then one can create a split-plot design that accounts for this restriction on randomization. Each factor is also classified as easy, hard, or very hard to change.

  - **Easy**: Factor levels can be adjusted with little to no increase in execution time.
  - **Hard:** Factor levels can be adjusted but there is a significant increase in execution time.
  - **Very Hard:** Factor levels are extremely difficult to adjust and therefore lead to a significant increase in execution time.
  - **Impossible:** Factor levels cannot be adjusted or changing them would make test infeasible.
- **Factor Type:** The test team will determine if a factor is to be controlled, held constant, or treated as noise. Options are:
  - **Vary**
  - **Hold Constant**
  - **Noise**
- **Priority:** Used to rank importance among the factors considered. High rated factors should candidates to be controlled and systematically varied during testing. Lower rated factors could be uncontrolled and allowed to vary.

SAMPLE OUTPUT

FACTORS

Number	Factor Name	Data Type	Units	Range	Desired # of Levels	Factor Change	Factor Type	Priority
1	Aircraft	Categorical	NA	Aircraft 1, Aircraft 2	2	Easy	Vary	Medium
2	Velocity	Continuous	knots	0-100	3	Easy	Vary	Medium
3	Bank Angle	Continuous	degrees	0-60	3	Easy	Vary	Medium
4	Altitude	Continuous	meters	500-1500	5	Easy	Vary	Medium
5	Ground Range	Continuous	meters	1000-4000	3	Easy	Vary	Medium
6	AOL	Continuous	degrees	0-180	5	Easy	Vary	Medium
7	Threat Type	Categorical	NA	10 different threats	10	Impossible	Vary	Medium

Figure 6: Test Design Factors Example

Section: Resource and Design Information

The intent of this section is to record additional information that will influence the test design approach, creation and selection.

Field Descriptions

- **Total Run Budget:** Provide a conservative estimate of the total number of test points that can be executed within the scheduled facility time.
- **Full Factorial Size:** The product of factor levels being considered. This is a baseline value that can be used to assess test design efficiency.

- **Confidence Goal:** The probability of concluding a factor has an effect on the response when in fact it does (true positive rate).
- **Power Goal:** The probability of concluding a factor has no effect on the response, when it does not have an effect (true negative rate).
- **Can we perform sequential experimentation:** Ideally, the most efficient way to obtain knowledge of a system is via the process of sequential experimentation. The approach subdivides tests into multiple stages where analysis will be conducted at the end of each stage to help influence the following stages. By using this sequential approach, insignificant factors can be identified more quickly and therefore removed from consideration in subsequent tests. This reduces the dimensions of the test space and can potentially reduce the total number of test runs needed in the latter stages. It is recommended that the initial set of tests consist of 25% of the run budget available (Montgomery, 2009).
- **Trial runs:** Can some preliminary runs be executed in order to get a better estimates of signal-to-noise and to identify any issues with execution protocol that may affect the run budget and test space.

SAMPLE OUTPUT

RESOURCES AND DESIGN INFORMATION

Total Run Budget	2000	200 data points and 10 weeks of testing
Full Factorial Size	13500	Product of all levels for each factor (Column: "# of Levels")
Confidence	0.95	Default = 0.95 (95% Confidence)
Power	0.80	Default = 0.8 (80% Power)
Can we build design sequentially?	Yes	
Trial Runs?	Yes	

Section: Design Approach

Document all discussions on approaches that will be considered to generate the experimental design. Some considerations to discuss are:

- What underlying model needs to be supported (whether we are looking at main effects and two factor interactions, or our we assuming a higher order model)
- Prediction quality
- Aliasing
- Design execution, whether blocking or a split-plot design is necessary
- Sequential experimentation strategy

SAMPLE OUTPUT

Based on subject matter expert comments nonlinear effects on the responses are expected.

Since design can be built sequentially we will start off with a Resolution V fractional factorial design to screen for significant factors.

There will be a separate model per threat type since threat type is an impossible to change factor.

Estimate total runs for initial design = 740 runs.

Based on analysis we may want to remove insignificant terms and then augment the design (using I-optimality criteria) to allow estimation on nonlinear effects.

Figure 7: Test Design Approach

Section: Analysis Approach

The STAT working group should record notes on different analysis approaches that will be considered. The requirement will guide us to the type of analysis that must be performed. Analysis should facilitate and inform acquisition decisions.

SAMPLE OUTPUT

Predictive models for each response will be created.

Statistical intervals (i.e. tolerance intervals) will be calculated to determine POBL across the attack surface.

Areas where system fails to meet threshold objective will be focused on for corrective action.

Figure 8: Test Design Analysis Approach

Conclusion

Thorough planning is fundamental to ensuring that sufficient rigor and traceability from requirements to analysis are incorporated into the test design. The STAT planning tool was created to aid test teams in addressing the critical questions posed by each phase of the STAT process. Using this tool will result in clear and agreed upon objectives for a test event, a mapping out of the operational space that will be covered, and assurance that the analysis produce will result in decision-quality information that address the requirement.

References

Coleman, David E. and Montgomery, Douglas C., “A Systematic Approach to Planning for a Designed Industrial Experiment,” *Technometrics*, vol. 35, no. 1, 1993, pp. 1-12.

Freeman, Ryan, Kensler, J. Dickinson, R. and Vining, G. “A Tutorial on the Planning of Experiments,” *Quality Engineering*, vol. 25, 2013, pp. 315-332.

Simpson, James, Listak, Charles, and Hutto, Gregory. “Guidelines for Planning and Evidence for Assessing a Well-Designed Experiment,” *Quality Engineering*, vol. 25, no. 4, 2013, pp. 1-23

Burke et al. “Guide to Developing an Effective STAT Test Strategy V5.0.” Scientific Test and Analysis Techniques Center of Excellence (STAT COE), 2017.

Ortiz, Francisco. “Categorical Data in a Designed Experiment Part 1: Avoiding Categorical Data”. Scientific Test and Analysis Techniques Center of Excellence (STAT COE), 2014.