STAT – Observational Example

D.1 Test Design Overview

The purpose of this appendix is to provide a framework for the Operational Test Activity's (OTA's) observational analysis of aircraft carrier flight deck operations. Testing will support an assessment of changes to Nimitz-class aircraft carrier flight decks that are designed to increase the number of aircraft sorties per a day from 120 to 135 in sustained operations. The test design is not based on a traditional approach in which specific factors and levels are controlled. Rather, the test design focuses on an observational analysis in which the test team collects and analyzes data from flight deck operations over which they exert limited control. In this case, the ship's crew will execute an operationally realistic air plan based on the Design Reference Mission (DRM) documented in the system's requirements. The air plan will include 6 days of sustained operations (12-hours per day). The OTA will collect data during flight operations to assess performance. The discussion below shows that sufficient data will be collected to analyze performance across relevant factors and to determine whether the number of sorties per day is increased.

The primary metric is Sortie Generation Rate (SGR), which measures the number of aircraft launched in a day. While there are factors that affect SGR, the OTA cannot control them. For example, the type of aircraft (e.g., helicopter versus F/A-18) and the mission assigned to the aircraft (e.g., tanker versus Combat Air Patrol) are likely to affect SGR. However, the missions assigned to aircraft and the order in which they are launched are highly constrained (e.g., the rescue helicopter will be launched first). The OTA cannot control this during a test. Furthermore, during flight operations, the flight deck crew makes numerous real-time decisions that affect SGR. For example, Nimitz-class carriers have four catapults for launching aircraft. During a typical launch cycle, multiple aircraft are launched and spare aircraft are available. If a problem occurs during a launch cycle, such as a catapult or an aircraft breaking, the flight deck crew will consider numerous factors. How long will the repair take? How critical is the aircraft to the mission? Where are they in the launch cycle? Based on this and other information, the flight deck crew may wait until the repair is completed, move the aircraft to a different catapult, use a spare aircraft, cancel the launch, or select another option. Artificially constraining the flight deck crew's options would not be realistic. Consequently, the test design does not control these factors and instead the test design is based on an operationally realistic air plan that the crew will execute during the test. The crew will be allowed to make real-time changes, as appropriate, following standard Navy procedures with the goal of achieving mission success for the scenario. The OTA will collect data on various aspects of flight deck operations; see Table D-1. The distributions of the times listed in the table and other metrics to be defined in the test plan will be analyzed to determine whether the flight deck changes improve flight deck operations and whether there are any bottlenecks in the process.

Table D-1. Primary Metrics for Observational Analysis

Metric	Definition
Sortie Generation Rate	Number of aircraft sorties launched in a flight day. Measured in a notional 12-hour day during sustained operations per the Design Reference Mission.
Recovery to Engine Shutdown Time	Measured from the time the aircraft is recovered until its engines are shut down.
Turnaround Time	Measured from engine shutdown until engine start and includes refueling, rearming, and maintenance time.
Engine Start to Taxi Time	Measured from engine start until the aircraft starts to taxi.
Taxi Time to Launch	Measured from start of aircraft taxi until aircraft launch.

Historically, Nimitz-class carriers are able to conduct 120 sorties per day in sustained operations. Over the planned 6 days of sustained operations, the test is expected to have a minimum of 720 sorties.

Overall, the flight deck design changes are expected to improve SGR performance from 120 to 135, and proposed testing has high power to detect SGR improvements of this magnitude. Figure D-1 shows the power of planned testing as a function of true SGR performance of the redesigned flight deck. As true SGR increases, the power to detect the improvement increases. At 95 percent confidence, power is 80 percent of detecting a difference in performance when the true SGR of the aircraft carrier is above 135 sorties per day in sustained operations.

The analysis should be able to detect relatively small differences in time on the order of a few minutes (e.g., a seven minute difference in turnaround time). To understand variability in the data, the OTA examined the results from the Navy's model of flight deck operations. This model was used to develop the changes to the Nimitz-class flight deck. As an example, Figure D-2 shows model results for turnaround time for aircraft that land, are refueled and rearmed, and are relaunched during the next launch cycle. Turnaround time is measured from engine shutdown to engine start and has a mean of 35.7 minutes, with a standard deviation of 7.0 minutes. Figure D-3 shows the power for an independent means test as a function of sample size, which can be used to compare test results to model predictions. As the number of turnaround times measured during the test increases, the power of the test increases. Measuring turnaround time in 35 events provides 90 percent power at 95 percent confidence for evaluating performance against expectations (effect size = standard deviation).

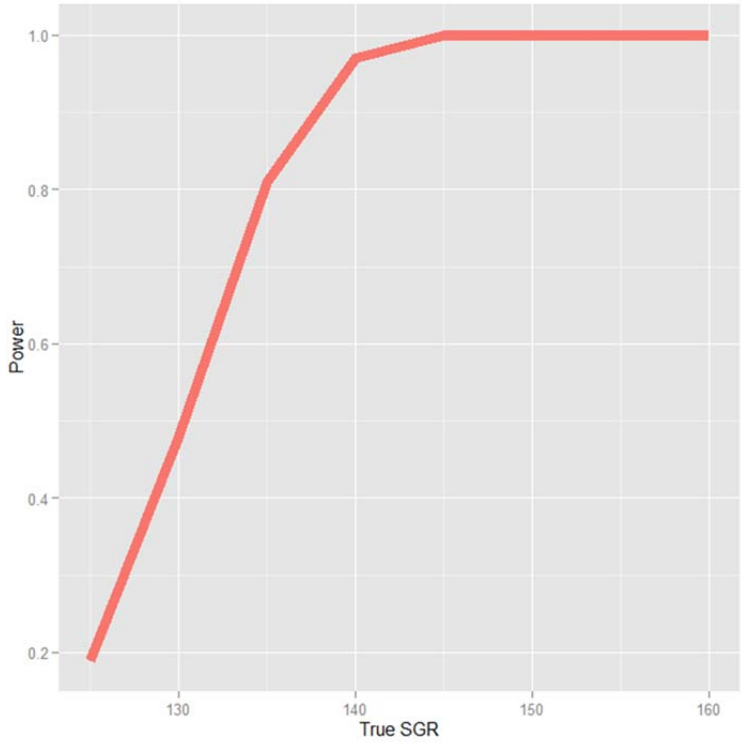

Figure D-1. Power estimate at 95 percent confidence for six days of sustained operations

As discussed above, planned testing is expected to have over 720 sorties based on historical Nimitz-class performance. Based on the planned airwing composition, these sorties will be distributed across six types of aircraft (e.g., E-2Cs, EA-18Gs, and F/A-18s) and six missions (e.g., Combat Air Patrol, Strike Support, and Interdiction). The proposed air plan, based on the Design Reference Mission, has a minimum of 35 sorties per aircraft type and minimum of 35 sorties per mission type. Consequently, the test should provide sufficient data to analyze results by aircraft type and mission.

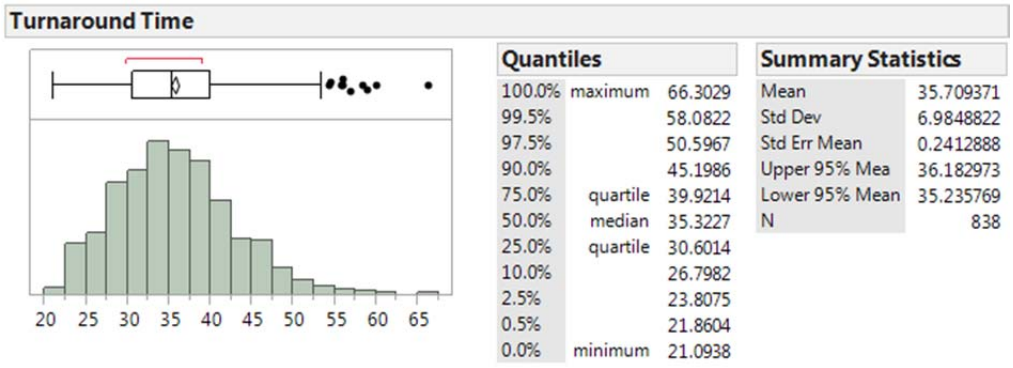

Figure D-2. Distribution of Turnaround Times for Aircraft that Land, are Refueled and Rearmed, and are Immediately Relaunched

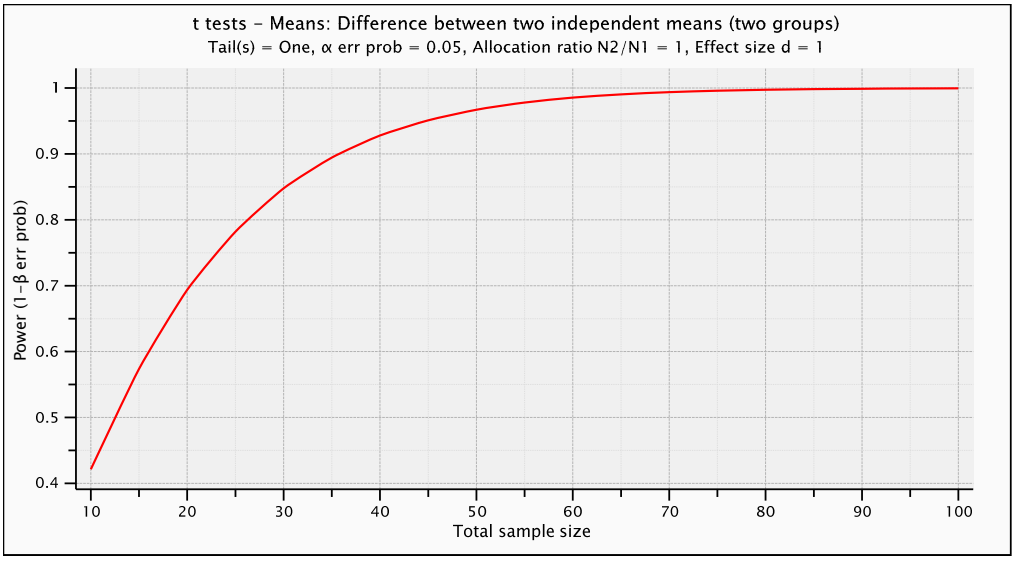

Figure D-3. Power for Independent Means Test

Overall, planned testing will provide sufficient power to determine whether flight deck changes improve SGR performance to 135 sorties per day and to determine whether there are any major deviations from expected turnaround times and other metrics.