Reliability Test Planning Utility Tool Guide

Authored by: Michael Harman

31 December 2015

Revised 7 November 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

Executive Summary 2
Introduction 2
Use of the Tool 3
  Plot Format (see Figure 1) 3
  Data Entry (See Figure 2) 4
  Compare Sampling Plans 5
  Proportion Worksheet 6
Expectation Management 8
  Example 8
Conclusion 9
References 9

Executive Summary

The use of operating characteristic (OC) curves for sampling is not new however many test practitioners do not have access to statistical software to evaluate plans and make comparisons. This paper details the use of an Excel based tool and is intended for test and evaluation practitioners use in determining reliability test times, evaluating the associated risk, and determining resource requirements.

Keywords: test and evaluation, reliability planning, risk assessment

Introduction

The use of operating characteristic (OC) curves for sampling is not new. OC curves plot the probability of accepting a lot or system for a given amount of tests (samples) as a function of its true underlying performance metric (a mean time between failures or a proportion). This paper details the use of a STAT COE developed, Excel based, OC curve tool for comparing various plans. The tool is named **"STAT COE Reliability Test Planning Tool.xls"** and is available in the Best Practices folder at www.AFIT.edu/STAT or via email at COE@AFIT.edu. In addition, this paper covers the application of test time estimates to risk assessment for management use.

This paper covers the use of the Excel tool only. Details regarding the mathematics and use of OC curves can be found in many texts and references (at www.AFIT.edu/STAT) including Kensler (2014) and Truett (2013). Additionally, this tool should be used in place of the paper binomial nomograph for correct test planning. Some key points are worth restating outside of the details available in these references.

- OC calculations assume the sampling is done from a single population, meaning the test time should be applied to a single configuration. However, configurations usually change during developmental testing (DT). So, the test time planning that OC curves provide is just a baseline for initial resource planning.
- OC curves are a simple way to plan reliability testing and are based on a single hypothesis test. The calculations in this tool assume a null hypothesis that the system does not perform to threshold and the alternate (what we aim to prove) is that it does at least meet threshold. We will reject the null (the system passes) if the specified test time results in the allowed number of failures or less. Whereas, the null is not rejected (the system fails) if the prescribed number of failures is exceeded.
- A "point estimate" is the simple calculation of the proportion (number of successes divided by total number of trials) or mean-time (total time divided by number of failures) obtained via testing. Uncertainty in our estimate results in a parameter estimate (confidence) interval whose width is determined by the sampling plan parameters. Since the null hypothesis states that the system fails out testing must generate a point estimate that yields an interval lower bound that lies at or above the threshold in order to reject the null and claim the system passes.

- The inherently stochastic behavior of real-world systems admits the possibility that an acceptable system may fail during the limited test period. Similarly, a truly unacceptable system may pass. For this reason it is critical that the sampling plan be properly scoped to address these risks. Sizing an OC plan requires two risk variables, alpha and beta (numbers between 0 and 1). Alpha is also called consumer's risk and indicates the probability that the consumer observes the prescribed number of failures (or less) and accepts a system that is actually below threshold (1-alpha is called confidence). Beta is also called producer's risk and indicates the probability that the producer's truly acceptable system results in too many failures and is rejected (1-beta is called power). Balancing these risks on an OC curve requires both lower bound (typically an acquisition system threshold) and goal (usually acquisition system objective) values to properly determine a sampling plan. DoD acquisition policy defines the threshold (T) as "the minimum acceptable operational value" and objective (O) values as "the desired operational goal." In some cases the objective value may be the same as the threshold. Correctly scoping these risks requires them to be separated numbers so the contracted performance value can be used in lieu of an objective value. These details will be demonstrated later graphically.
- For a given alpha and beta, test time increases as T and O get closer together and decreases as they get farther apart. Realistic values with a meaningful difference between them will reduce the test time required to make an accurate assessment.

Use of the Tool

The Excel tool contains two spreadsheets which operate the same way. One is for a time metric like mean time between failure (MTBF). The other is for a proportion like probability of success. The data entry is similar for both and the following example refers to the MTBF sheet.

Plot Format (see Figure 1)

- True MTBF (X axis)
- Probability of Acceptance (Y axis)
- Red vertical line (threshold/rejection quality level: T/RQL)
- Green vertical line (objective/acceptance quality level: O/AQL)
- Blue vertical line (contract)
- Consumer's risk is read from the bottom up on the threshold line
- Producer's risk is read from the top down on the objective (or contract) line

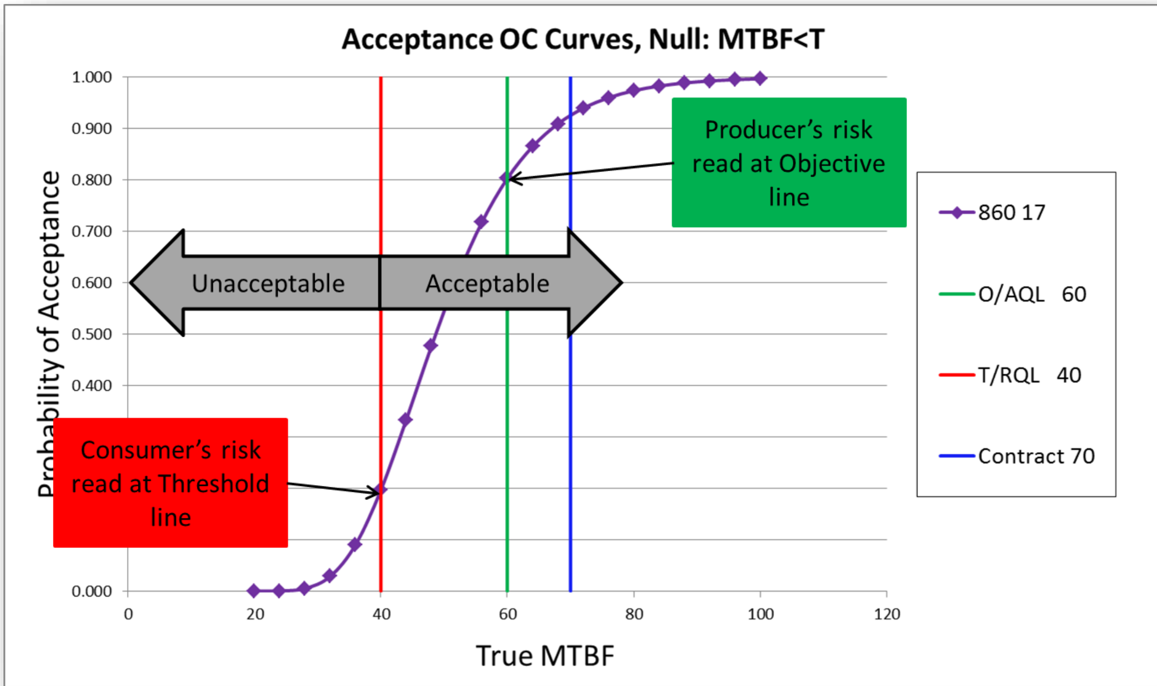

Figure 1: Tool Plot Area

Data Entry (See Figure 2)

1. Enter Threshold, Objective, and Contract goals (40, 60, 70 hours)
2. Enter Test Times and Allowed Fails (the source of these example numbers is described later)
   a. 860 time (hours) and 17 allowed fails
   b. 120 time and 1 fail (this is three times the threshold, also known as the three time rule)
   c. 320 time and 5 fails
   d. 200 time and 5 fails
3. MTBF and lower bound calculations are done for you

	T/RQL	O/AQL	Contract	
MTBF	40	60	70	
Test Time	860	120	320	200
Allowed Fails	17	1	5	5
MTBF point estimate	51	120	64	40
80% Conf Lower Bound	40	40	40	25

Figure 2: MTBF Data Entry Section

Note: creating a plan to meet alpha/beta risk goals is not trivial. See Kensler's MTBF Tool (2014) for further details. Software such as Minitab, JMP, or others may be used to generate the risk goals.

Compare Sampling Plans

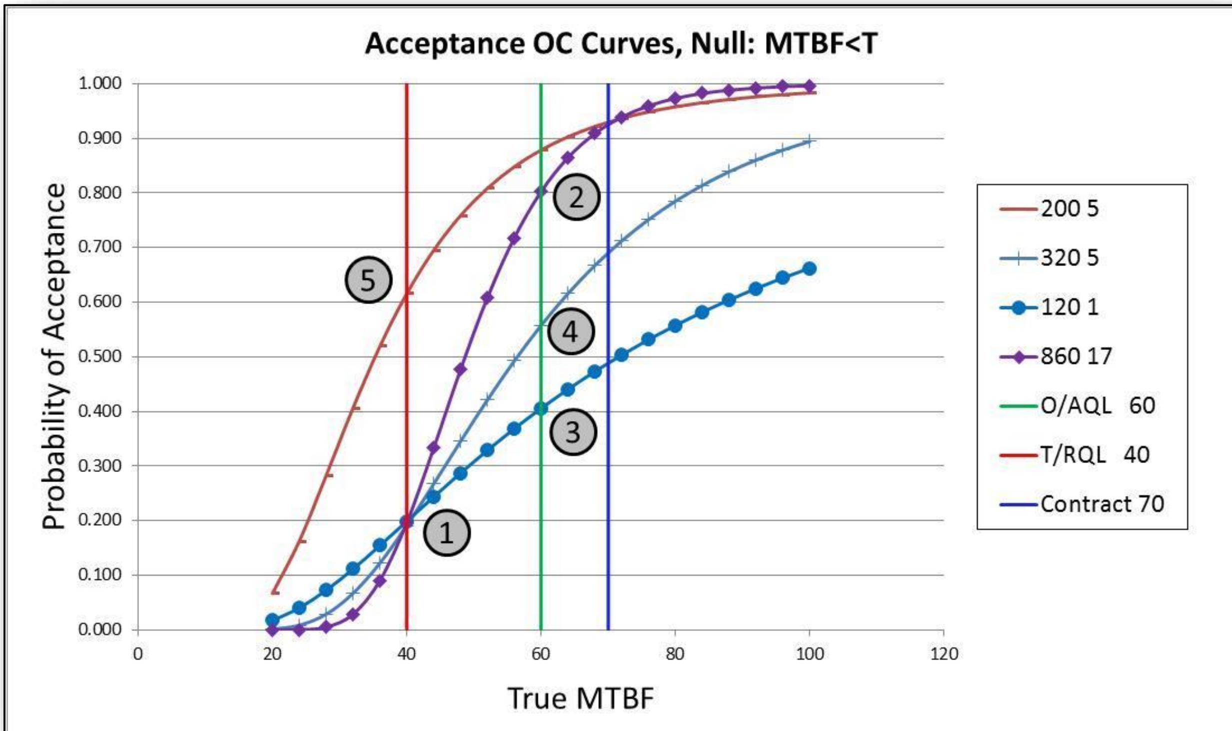

Figure 3: MTBF Plot

Test Time 860 hours, 17 allowed Failures (860/17): Example of a test designed to meet risk goals

Referring to Figure 3, this plan was specifically designed to meet a consumer's risk of 0.20 (1.0-0.2=80% confidence) at point (1) and a producer's risk of 0.20 (1-0.20=80% power) at point (2). This indicates that the consumer has a maximum 20% risk of accepting a system at or below threshold. Similarly, the producer has an 80% probability that his system will be accepted (20% chance of rejection) if he delivers it with objective performance. If he delivers it at the contract level of performance he has a 93% probability of acceptance and only 7% risk of rejection. **This is the preferred amount of testing for this system**.

Test Time 120/1 Fail: Example of the "Three Times Rule"

This plan is known as the "three times rule" meaning the test time is determined as three times the threshold and allowing only 1 failure. Note that while the consumer's risk remains at 20% at point (1) the producers risk is significantly higher, despite the lower bound being at threshold. Point (3) shows about 43% probability of acceptance meaning his risk of rejection is 57%. This improves slightly to 49% probability of acceptance if he delivers at the contract level. One can project the line to estimate that (for this amount of testing) he would have to produce a system with an MTBF above 120 to have an 80% probability of passing. This stems from the low test and time and singular failure. If one expects a failure every 40 hours (given a threshold system) then 120 test hours should produce 3 failures so the probability of rejection is very high. Note that the point estimate is very high if the system actually passes the testing! But, more test time is needed for an accurate assessment and most probably the system will fail the short test time (by observing more than 1 failure) even if the system is actually at objective. *This test is not recommended due to high producer's risk as compared to the 860/17 plan.*

Test Time 320/5 Fails: Example of a test scoped on lower bound alone

This plan adds additional test time and failures in an attempt to keep the lower bound at threshold. However, while the consumer's risk remains 20% at point (1) the producer's risk only decreases to 44% at point (4) (56% probability of acceptance). One can project the line to estimate that he would have to produce a system with an MTBF about 85 to have an 80% probability of passing. *This test is not recommended due to high producer's risk as compared to the 860/17 plan.*

Test Time 200/5 Fails: Example of a test scoped on the point estimate alone

This plan demonstrates testing incorrectly designed around the point estimate without including risk metrics and considering the lower bound. The probability of acceptance at threshold is 62% at point (5) and the producer's risk is only 7% (93% probability of acceptance) at the contract level. This may seem like a win-win for both parties but consider a system delivered below threshold and the high probability of acceptance (still 40% at MTBF of 30). The point of testing is to accurately assess system performance in support of a fair program decision so this plan falls short. *This test is not recommended due to high producer's risk as compared to the 860/17 plan.*

Proportion Worksheet

The second worksheet in the tool is for requirements expressed as proportions (a number between 0 and 1). An example would be a system with a probability of detection threshold at 70% (0.70) and an objective at 85% (0.85). Enter data as shown in Figure 4.

	T/RQL	O/AQL	Contract	
Proportion	0.70	0.85	0.95	
Test Time	21	8	12	4
Allowed Fails	4	1	2	1
Proportion point estimate	0.81	0.88	0.83	0.75
80% Conf Lower Bound	0.71	0.71	0.70	0.44

Figure 4: Proportion Data Entry Section

Like the MTBF examples previously, these time/failure combinations represent a risk-based (80% confidence/80% power) plan (21/4), a “lower-bound-only” plan (8/1), a slightly improved “lower-bound-only” plan (12/2), and a point-estimate-only (4/1) plan. Note there is no “three-times” rule for proportions. It is left to the reader to similarly identify the risk levels as seen in Figure 5.

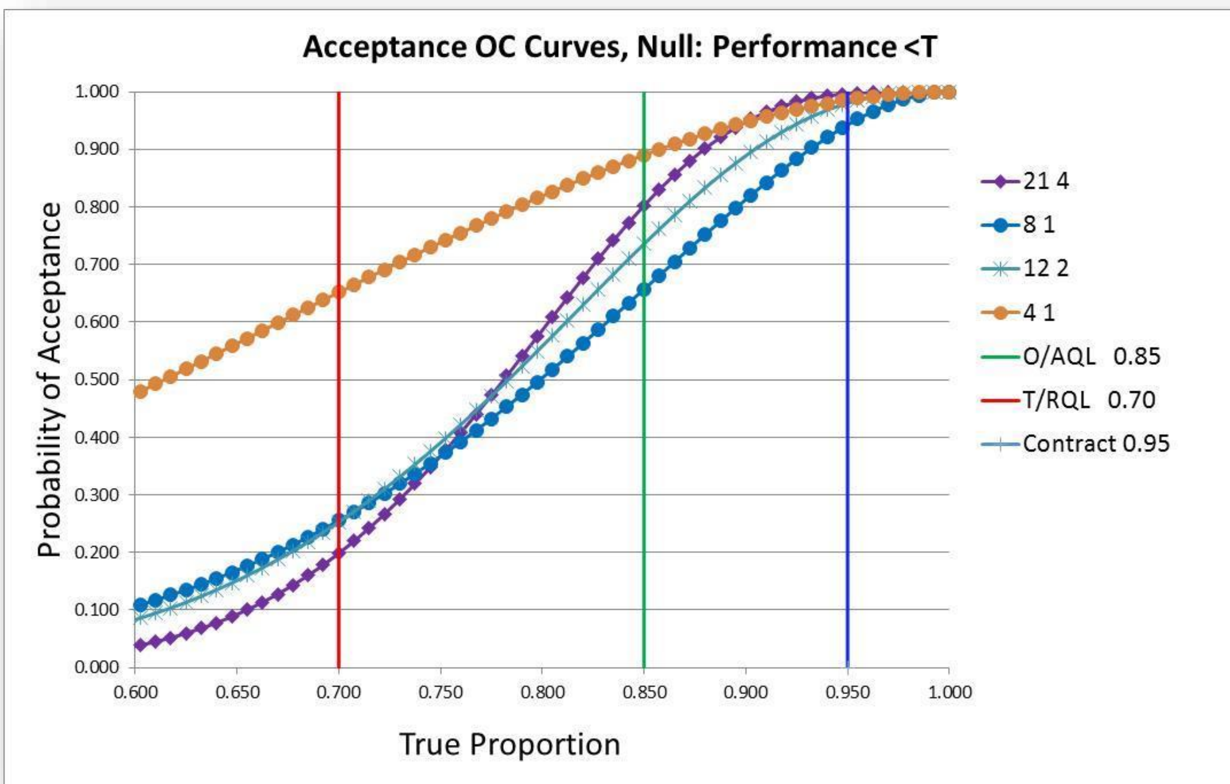

Figure 5: Proportion Plot

Note that for proportions, the lines all converge on 100% probability of acceptance for a 100% system, unlike the MTBF plot. Similarly, test time still goes up as T and O get closer together however, for proportions, there is a finite upper bound at 100% that does not exist for time-based metrics. And again, realistic proportional T and O with a meaningful difference between them will reduce the test time required to make an accurate assessment. Commercial claims of 99.99% or 99.999% reliability are almost impossible to prove, especially if the threshold and objective are close in value. Use the tool to explore these potential sampling plans.

Expectation Management

As previously mentioned, configuration changes in DT will expectedly violate the "single population" assumption built into the OC curve and OT will only have time to execute a subset of the total test time. Aside from expectations derived from reliability growth, failure analysis, corrective actions and other supporting analysis, OT will draw conclusions based on their operationally realistic configuration observed during their test time. Even given "three times rule" test plans, there is a possibility in OT that the system will pass statistically (very few failures observed and lower bound above threshold). A more probable outcome is that the system's point estimate is above threshold but the lower bound lies below threshold. In this case the system fails statistically but still may result in a fielding decision. The decision maker must assess this risk to moving forward and when a better assessment can be made. The sampling plan information can be used to set an event-based follow up date to reassess program progress.

Example

One needs to assess how much test time remains to complete the well scoped plan and apply some assumptions regarding how long it will take to gain that time on the system.

Referring to Figure 2 let's assume the desired sampling plan was 860 hours and 17 allowed failures (for 80% confidence, 80% power). But, OT has planned their resources using the three times rule and will only execute 120 (3 x threshold 40 hours = 120) hours and allow only 1 failure. Let us assume the system experiences more than 1 failure during OT but it is approved for limited fielding assuming it can prove its reliability goals in the near future. Program leadership wants to know when the equivalent of the full test plan (860 total hours) will be obtained in order to more rigorously support or refute their decision. We will calculate the calendar time required to accumulate the remaining test time.

740 hours remain (860-120) to complete the plan in the OT configuration.

According to the fielding plan 3 units will be deployed initially and they will each operate for 10 hours per week (10 hours/unit/week).

(740 hours) / (10 hours/unit/week) / (3 units) = 24.6 weeks

Providing this information to leadership helps bound the decision process and set realistic expectations.

Conclusion

The use of operating characteristic (OC) curves for sampling is an effective way to scope reliability test plans and resources. In addition, the test plan information can be used to scope future decision risk for systems experiencing OT fails.

References

Kensler, Jennifer, "Reliability Test Planning for Mean Time Between Failures-Best Practice." Scientific Test and Analysis Techniques Center of Excellence (STAT COE), www.AFIT.edu.STAT, 2014.

Kensler, Jennifer, "MTBF Test Planner Tool." Scientific Test and Analysis Techniques Center of Excellence (STAT COE), www.AFIT.edu.STAT, 2014.

Truett, Lenny, "Using Operating Characteristic (OC) Curves to Balance Cost and Risk-Best Practice." Scientific Test and Analysis Techniques Center of Excellence (STAT COE), www.AFIT.edu/STAT, 2013.