

An Engineering Alternative to Statistically Designed Acceptance Tests

Bryan L. Dodson
SKF

Rene Klerx
SKF AB

ABSTRACT

There are several fundamental problems with statistically designed acceptance tests. Two of the largest problems are the inability to obtain a random sample and the tendency of these tests to reject designs that do not significantly exceed reliability requirements. This paper explores the problems with statistically designed acceptance tests and offers an alternative.

CITATION: Dodson, B. and Klerx, R., "An Engineering Alternative to Statistically Designed Acceptance Tests," *SAE Int. J. Passeng. Cars - Mech. Syst.* 5(1):2012, doi:10.4271/2012-01-0205.

INTRODUCTION

If there were no limitations on time and money, acceptance tests could be designed to expose all risks. Thousands of parts could be tested for long periods of time under all possible conditions encountered in the real world allowing all failure modes to be discovered. Because real products must be tested in a reasonable time frame with a realistic budget, statistically designed tests are employed to allow inferences regarding entire populations to be made from samples.

Statistically designed tests balance the risk of releasing a design that does not meet customer requirements (type II error) against the costs (type I error) of failing a design that meets the customer requirements. It is important to understand the probability of passing a statistically designed test as a function of the degree the design meets the requirements, the sample size and the test duration. If a design fails a statistically designed test it does not mean the design does not meet the customer's requirements, it simply means that there is not enough evidence to accept that the design meets the customer's requirements based on the number of samples tested.

Consider a basketball player shooting free throws. If a player makes 3 consecutive free throws, what is your perception of the player's free throw ability? Most people would be cautiously optimistic; the player appears to be a good free throw shooter, but many more samples are required before you are confident in your decision. Now suppose you

are the coach of the basketball team, and the game will be decided on two free throw attempts. You have one player who has attempted 1000 free throws and was successful on 900 of those attempts. You have another player who has attempted 3 free throws and was successful on all 3 attempts. Which player do you choose to decide the game? Most people would choose the player with 90% free throw percentage over the player with the 100% percentage because they have more confidence in the player with the 90% free throw percentage. Statistically we are 90% confident that the player with 1000 attempts has a true free throw percentage between 88.3% and 91.5%. We are 90% confident that the player with 3 attempts has a true free throw percentage between 36.8% and 100%. This is the same dilemma faced in engineering design. How can statistical confidence be increased in an efficient and timely manner?

One of the most important considerations when using statistically designed tests is the ability to obtain a random sample. Statistically designed tests assume the sample has been randomly selected from the population. This means the sample represents the variability in the population including variability from the following sources:

- variation from multiple production operators,
- variation from multiple lots of raw materials,
- variation from tool wear,
- variation from machine maintenance, and

- variation from supplier changes.

It can be difficult to obtain a random sample from the population for testing purposes, and it is important to understand the effects of a non-random sample on the statistical calculations of the test.

STATICALLY DESIGNED TESTS

Bogey testing has been used for reliability verification testing for many years (Kapur and Lamberson, 1977). In 1991, Wang (1991) introduced a method of bogey testing to automotive electronic components, which was later refined by Wang and Lu (1992). Bogey is a numerical standard of performance set-up as a mark to be aimed at especially in competition. Bogey testing requires the testing duration be equal to the required life. For example, if 95% reliability is required at 200,000 kilometers of service then units being tested will be removed from testing when they fail or when they complete the equivalent of 200,000 kilometers of testing. The sample size required to demonstrate reliability of R with a confidence level of C is

$$n = \frac{\ln(1 - C)}{\ln R} \quad (1)$$

Because of the large sample sizes required for bogey testing and the desire to run parts to failure Bayesian testing is another type of statistically designed test which was introduced by Weir (1960). Nelson (1976), Grohowski (1976) and Dodson (2006) have all expanded on this technique. Bayesian testing requires the test duration to extend beyond the required life, requiring assumptions to be made concerning the shape of the distribution of the time-to-fail. This is done by assuming a Weibull distribution for time-to-fail, and assuming a shape parameter. For a sample of n units the required test duration to demonstrate a reliability of R at time t with a confidence level of C assuming no failures is

$$T = \left(\frac{t}{[-\ln R]^{1/\beta}} \right) \left(\frac{-\ln(1 - C)}{n} \right)^{1/\beta} \quad (2)$$

The performance of both bogey tests and Bayesian tests are extremely sensitive to how the samples are selected. If the samples are not randomly chosen from the population, then the statistical properties of the tests are not valid. Stated another way, statistically designed tests are only valid if all sources of variation have a chance of being represented in the parts tested because of the procedure used to select the parts.

PROBABILITY OF PASSING STATICALLY DESIGNED TESTS

Dodson and Schwab (2006) note the difficulty of passing a statistically designed durability test unless the reliability of the design greatly exceeds requirements. For example, if a bogey test is used to verify that a basketball player has a 90%

or better free throw percentage with a confidence level of 85%, the basketball player would be required to make 18 consecutive free throws. As shown in the computation below, the probability of a player with a true free throw percentage of exactly 90% would pass this test only 15% of the time.

$$p = 0.9^{18} = 0.15$$

If it is known that the time-to-fail distribution for a system is Weibull with a shape parameter of 2.5, then a sample of 4 systems would have to be tested for a duration of 26.3 years without failure to demonstrate a reliability of 95% at 10 years with a confidence level of 90%. This is shown below.

$$T = \left(\frac{10}{[-\ln(0.95)]^{1/2.5}} \right) \left(\frac{-\ln(1 - 0.9)}{4} \right)^{1/2.5} = 26.3$$

If the system has a true reliability of 95% after 10 years, and 4 systems are randomly selected for this demonstration test, the probability that this test is completed successfully (all 4 systems complete 10 years of testing without failure) is 10% as shown below.

$$\theta = \frac{10}{[-\ln(0.95)]^{1/2.5}} = 32.81$$

$$R(26.3) = e^{-\left(\frac{26.3}{32.81}\right)^{2.5}} = 0.562$$

$$P(\text{Passing Test}) = (0.562)^4 = 0.1$$

In the previous examples the probability of failing the statistically designed test was equal to the confidence level of the test. This is true in every case with statistically designed tests requiring no failures. To have a reasonable chance of passing a statistically designed test the requirements must be substantially exceeded. Consider the example above the required L_5 life (the time at which 95% reliability is required) is 10 years. Figure 1 shows the probability of passing the test as a function of the L_5 life. An L_5 life of 45 years is required to have a 95% chance of passing the test even though the required L_5 life is only 10 years.

Figure 2 shows the performance of a 4 sample Bayesian test for a variety of Weibull shape parameters. This figure shows the tests require less excessive performance to pass, but even with a Weibull shape parameter of 5, an L_5 of 18.5 years (85% greater than the required L_5) is necessary to have a 90% chance of completing the test successfully.

EFFECTS OF NOT OBTAINING A RANDOM SAMPLE

The easiest way to defeat a statistically designed test is to select the parts in a non-random manner. By purposely selecting parts that are close to the nominal dimension the variability is reduced. In addition, a positive bias is introduced because the parts close to nominal will perform