DFSS RELIABILITY SYSTEM THINKING ASSESSMENT

Forrest W. Breyfogle III
CEO, Founder
Smarter Solutions, Inc.
11044 Research Blvd., Suite B-400 Austin, TX 78759
512-918-0280
Forrest@SmarterSolutions.com
www.SmarterSolutions.com

SUMMARY

Product mean time between failures (MTBF) validation tests in Design for Six Sigma (DFSS) typically provide questionable results. The described alternative reliability assessment strategy can provide more meaningful information, while utilizing a much smaller sample size – perhaps only one.

INTRODUCTION

This paper uses an example to compare traditional reliability testing to a Smarter Six Sigma Solutions (S^4) alternative. In this S^4 alternative, a unique Design of Experiments (DOE) analysis technique is described, which can provide more insight to potential future product customer issues than does traditional reliability testing. These benefits are achieved with much less test effort and could be used as a component of an organization's DFSS development process.

EXAMPLE[1]

A power supply has sophisticated design requirements and a large MTBF criterion of 10×10^6 hours. A test organization is commissioned to evaluate the reliability and function of this non-repairable power supply.

Traditional reliability MTBF assessments involve testing the failure rate criterion by exercising units long enough to verify the failure rate criterion. Considering that the unit failure rate is constant with age, as the criterion implies, Table K in *Implementing Six Sigma*, 2nd edition[1] provides a factor for calculating total test hours. If we desire 90% confidence with a test design that allows no failures, then the tabled factor would be 2.303, which yields a total test time of

$$T = 2.303\ (10 \times 10^6) = 23.03 \times 10^6 \text{ hr}$$

The number of test units could range from 23.03×10^6 units tested for 1 hour or one unit tested for 23.03×10^6 hours. It is unlikely that many products would survive the single-unit test length without failure because of some type of mechanism wear-out. Actually, it is not very important that a single product last this long, because continual customer usage would require 114 years to accumulate 1×10^6 hr of usage For this test, a single product would need to survive over 2600 years without failure before passing the test. This test approach obviously makes no sense.

In addition, the wording of this criterion is deceptive from another point of view. Whenever a non-repairable device fails, it needs to be replaced. The wording for the above criterion implies a repairable device (mean time between failures). What is probably intended by the criterion is that the failure rate should not exceed 0.0000001 failures/hr (i.e., $l/[10 \times 10^6$ hr/failure]). From a customer point of view, where the annual usage is expected to be 5000 hr and there is a 5-year expected life, this would equate to about 0.05% of the assemblies' failing after 1 year's usage and 0.25% of the assemblies' failing after 5 year's usage. If the criterion were quoted with percentages of this type, there would be no confusion about the reliability objectives of the product.

Two percentage values are often adequate for this type of criterion: one giving maximum fail percentage within a warranty period, and the other giving maximum fail percentage during the expected product life.

Consider that the expected annual number of power-on hours for the power supply is 5000 hr, and each unit is tested to this expected usage. This test would require 4605 (i.e., $23.026 \times 10^6 /5000 = 4605$) units, while a 5-year test would require 921 units [i.e., $23.03 \times 10^6 /(5000)(5) = 921$]). For most scenarios involving complex assemblies, neither of these alternatives is reasonable because the unit costs would be prohibitive, the test facilities would be very large, the test would be too long, and information obtained late in the test would probably come too late for any value added. Accelerated test alternatives can help reduce test duration, but the same basic problems still exist with less magnitude.

Even if the time and resources are spent on running this test and no failure occurs, customer reliability problems can still exist. Two basic assumptions are often overlooked with the preceding test strategy. The first assumption is that the sample is a random sample of the population. If this test is performed early in the manufacturing process, the sample may be the first units built, which is not a random sample of future builds that will go to the customer. Problems can occur later in the manufacturing process and cause field problems that this test will not detect. The second assumption is that the test replicates customer usage. If a test does not closely replicate customer situations, real problems may not be detected. For example, if the customer turns off a system unit that contains the power supply each evening, while test units are run continuously, the test may miss thermal cycling component fatigue failures. Or the customer might put more electrical load on the system than is put on the test system, and the test might miss a failure mode caused by the additional load. A fractional factorial test strategy might better define how the power supplies should be loaded and run when trying to identify customer reliability problems.

It is easy to fall into the trap of playing games when testing to verify a failure rate criterion. Instead, consider an S^4 alternative where reflection is given to the type of problems detected during previous tests and the customer's environment. Tests should be designed to give the *customer* the best possible product.

Single-lot initial production testing that treats the product as a black box can involve great effort but fail to detect important problems. To maximize test effectiveness, consider the following: Should more emphasis be placed on monitoring the component selection/design considerations as opposed to running them and counting the number of failures? Should more emphasis be given to monitoring the manufacturing process (i.e., control charting, process capability/performance metrics, etc.)? Should more fractional factorial experiments be used in the design process of the mechanism, with stress to failure considerations as an output?

Consider also the real purpose of a reliability test. For the product test criterion, what should really happen if a zero failure test plan had one, two, three . . .failures? It is hoped that the process or design would be fixed so that the failures would not occur again. It is doubtful that time would permit another sample of the "new design/process" every time a failure was detected. Typically, the stated objective may be to verify a criterion, but the real intent may be to determine and then fix problems.

Now, if the real objective is to identify and fix problems, instead of playing games with numbers, test efforts should be directed to ways of accomplishing this efficiently. An efficient test would not be to turn on 4605 units for a 5000-hr test and monitor them once a week for failures. An efficient approach could include querying experts for the types of failure expected and monitoring historical field data so that test efforts can be directed toward these considerations and new technology risks. For example, a test without power on/off switching and heating/cooling effects does not make sense if previous power supplies experience 70% of their field failures during the power-on cycling; e.g., a failure mode often experienced when turning on a light bulb. Another scenario is that if 10% of the times the power supplies do not start the first time, handling problems during shipment could be causing out-of-box failures.

S^4 considerations may indicate that more fractional factorial experiments should be used in the development process. For a preproduction test, the experimenter may decide to test only three units at an elevated temperature for as long as possible to determine if there are any wear-out surprises. The person who conducts the test may also plan to run some units in a thermal cycle chamber and a thermal shock chamber. Plans may also consider a shipping test and an out-of-box vibration stress to failure test for some units.

In addition, the experimenter should work with the manufacturing group to obtain time-of-failure information during production pre-shipment run-in tests. Data from these tests could be used to determine early-life characteristics of the product, which could possibly be projected into the customer environment. The experimenter would also like to ensure that any run-in test time in manufacturing is optimized.

A good reliability test strategy has a combination of test considerations that focus on efficiently capturing the types of failures that would be experienced by the customer. It is not a massive test effort that "plays games with numbers." In addition to reliability considerations, the experimenter needs also to address functional considerations in the customer environment. Fractional factorial testing is an efficient method for meeting these needs.

For the preceding test, one preproduction unit could be tested functionally at the extremes of its operating environment using a fractional factorial test strategy. The following is such a strategy where input factors are evaluated for their effect on the various important output characteristic requirements of the power supply; it is summarized in Table 1.

From Table M4[1], a 32-trial resolution IV design was chosen. With this design, the main effects would not be confounded with two-factor interactions, but there would be confounding of two-factor interactions with each other. The 11 contrast columns from Table M4 were assigned alphabetical factor designations from left to right (A-K). These test trials along with two of the experimental trial outputs (minus 12 and 3.4-output levels) are noted in Table 2.

Inputs			
		Levels	
Factors		-	+
A:	Ambient Temperature	25 deg C	47 deg C
B:	Input ac voltage range	110V	220 V
C:	Mode of programmable output	3.4 V	5.1 V
D:	ac line voltage (within range in B)	Min	Max
E:	Frequency at ac input	Min	Max
F:	Load on -12 V output	Min	Max
G:	Load on -5 V Output	Min	Max
H:	Load on 12 V output	Min	Max
I:	Load on 5.1 V output	Min	Max
J:	Load on 3.4 V output	Min	Max
K:	Load on programmable output	Min	Max
Outputs			
Output voltage on each output (-12V, -5V, 12V, 5.1V, 3.4V, programmable volt output			
Ripple/noise			
Noise			
Input (power factor)			
Efficiency			
Line current			
Line power			

Table 1

The effect of the minus12-V loading (factor F), for example, on the minus12-V output level is simply the difference between average output response for the trials at the high load and those at low load:

$$\begin{array}{l}\text{Average effect} \\ \text{on } -12\text{-V output} \\ \text{by } -12\text{-V load} \\ (F \text{ effect})\end{array} = \frac{(-11.755 - 11.702, \cdots)}{16} - \frac{(12.202 - 12.200, \cdots)}{16}$$

$$= -0.43 \text{ V}$$

A probability plot of the main effect and interaction considerations, given the confounding noted in Table N2[1], is shown in Figure 1, where the standardized effect of F-effect of 103 in the plot translates to the -0.43 voltage value. Since the 0.43-V effect is a large outlier from any linear plot relationship, it is concluded that the load on the –12-V power supply output significantly affects the –12-V output; i.e., a best estimate value of 0.43 V. Other less statistically significant effects are also highlighted.

Trial	A	B	C	D	E	F	G	H	I	J	K	-12V	3.4V
1	+	-	-	-	+	+	+	+	-	+	-	-11.755	3.1465
2	+	+	-	-	-	+	+	-	+	-	-	-11.702	3.3965
3	+	+	+	-	-	-	+	+	-	+	+	-12.202	3.147
4	+	+	+	+	-	+	-	+	+	-	-	-11.813	3.4038
5	+	+	+	+	+	+	+	+	+	+	+	-11.761	3.1537
6	-	+	+	+	+	-	+	-	+	+	-	-12.2	3.1861
7	-	-	+	+	+	-	-	+	-	+	+	-12.325	3.1902
8	+	-	-	+	+	-	-	+	+	-	+	-12.292	3.398
9	+	+	-	-	+	+	-	-	+	+	+	-11.872	3.1498
10	-	+	+	-	-	+	+	-	-	+	-	-11.819	3.1914
11	+	-	+	+	-	+	+	-	-	-	+	-11.685	3.4084
12	-	+	-	+	+	+	+	+	-	-	+	-11.763	3.4217
13	-	-	+	-	+	+	+	+	+	-	-	-11.78	3.4249
14	+	-	-	+	-	-	+	+	+	+	-	-12.223	3.1403
15	-	+	-	-	+	-	-	+	+	+	-	-12.344	3.1782
16	-	-	+	-	-	+	-	+	+	+	+	-11.909	3.1972
17	-	-	-	+	-	+	+	-	+	+	+	-11.834	3.1902
18	-	-	-	-	+	-	+	-	-	+	+	-12.181	3.1847
19	+	-	-	-	-	+	-	+	-	-	+	-11.801	3.4063
20	-	+	-	-	-	-	+	+	+	-	+	-12.146	3.4184
21	+	-	+	-	-	-	-	-	+	+	-	-12.355	3.1401
22	-	+	-	+	-	+	-	+	-	+	-	-11.891	3.1826
23	+	-	+	-	+	-	+	-	+	-	+	-12.146	3.4044
24	+	+	-	+	-	-	-	-	-	+	+	-12.337	3.1435
25	+	+	+	-	+	-	-	+	-	-	-	-12.28	3.3975
26	-	+	+	+	-	-	-	-	+	-	+	-12.275	3.423
27	+	-	+	+	+	+	-	-	-	+	-	-11.852	3.1459
28	+	+	-	+	+	-	+	-	-	-	-	-12.131	3.39
29	-	+	+	-	+	+	-	-	-	-	+	-11.819	3.4281
30	-	-	+	+	-	-	+	+	-	-	-	-12.134	3.4193
31	-	-	-	+	+	+	-	-	+	-	-	-11.846	3.4226
32	-	-	-	-	-	-	-	-	-	-	-	-12.261	3.4203

Table 2

The results of statistical analyses are commonly presented as significance statements. However, a practitioner may be interested in the overall effects relative to specification limits. A DOE Collective Response Capability Analysis (DCRCA) plot addresses this desire. Figure 2 illustrates a DCRCA probability plot of the 32-experimental trial −12-V source experimental trial outputs, with specification limits. Note that this plot is not a true random sample plot of a population. In this plot the magnitude of the loading effect is noticeable as a discontinuity in the line. This plot reflects the variability of one machine given various worst-case loading scenarios. Because the distribution tails are within the specification limits of 12 ± 1.2 V, it might be concluded that there are no major problems if the variability from machine to machine is not large and there is not degradation with usage.

DOE-collective-response-capability-assessment (DCRCA):

Consider a DOE where the factors were chosen to be the tolerance extremes for a new process, and the response was the output of the process. Consider also that there were no historical data that could be used to make a capability/performance statement for the process. A probability plot of the DOE responses can give an overall picture of how we expect the process to perform later, relative to specification limits or other desired conformance targets. This type of plot can be very useful when attempting to project how a new process would perform relative to specification limits; i.e., a DCRCA assessment. Obviously, the percentage of occurrence would provide only a very rough picture of what might occur in the future since the data that were plotted are not random future data from the process.

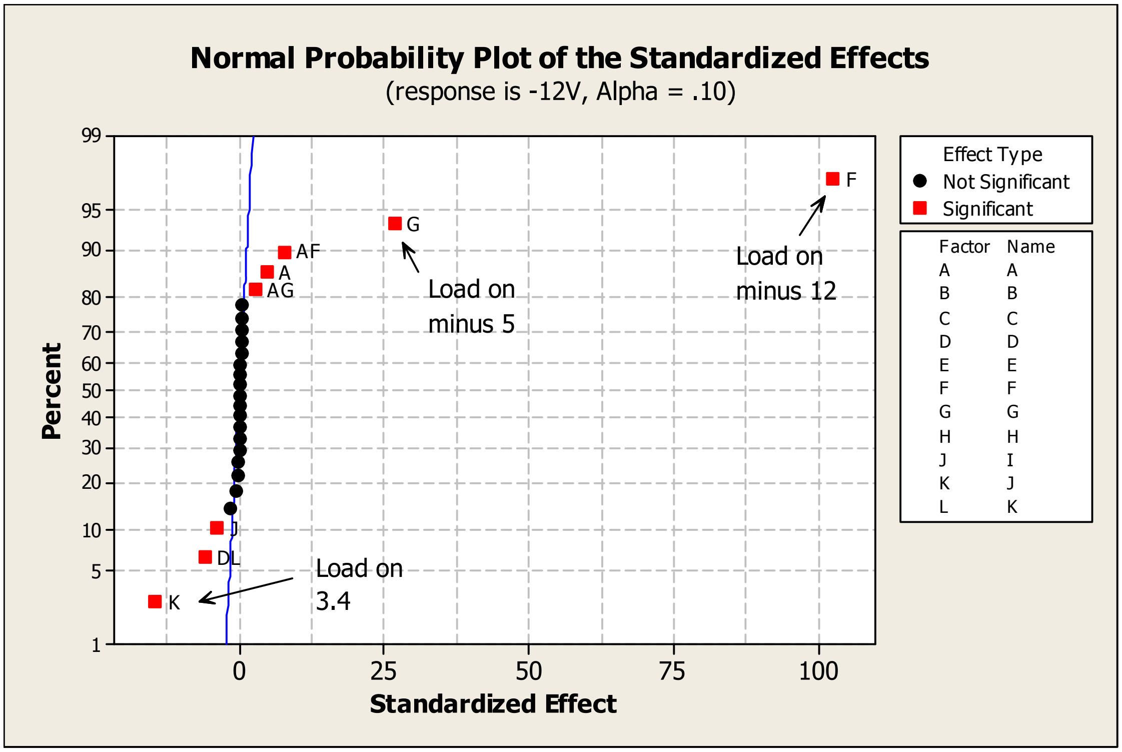

Figure 1

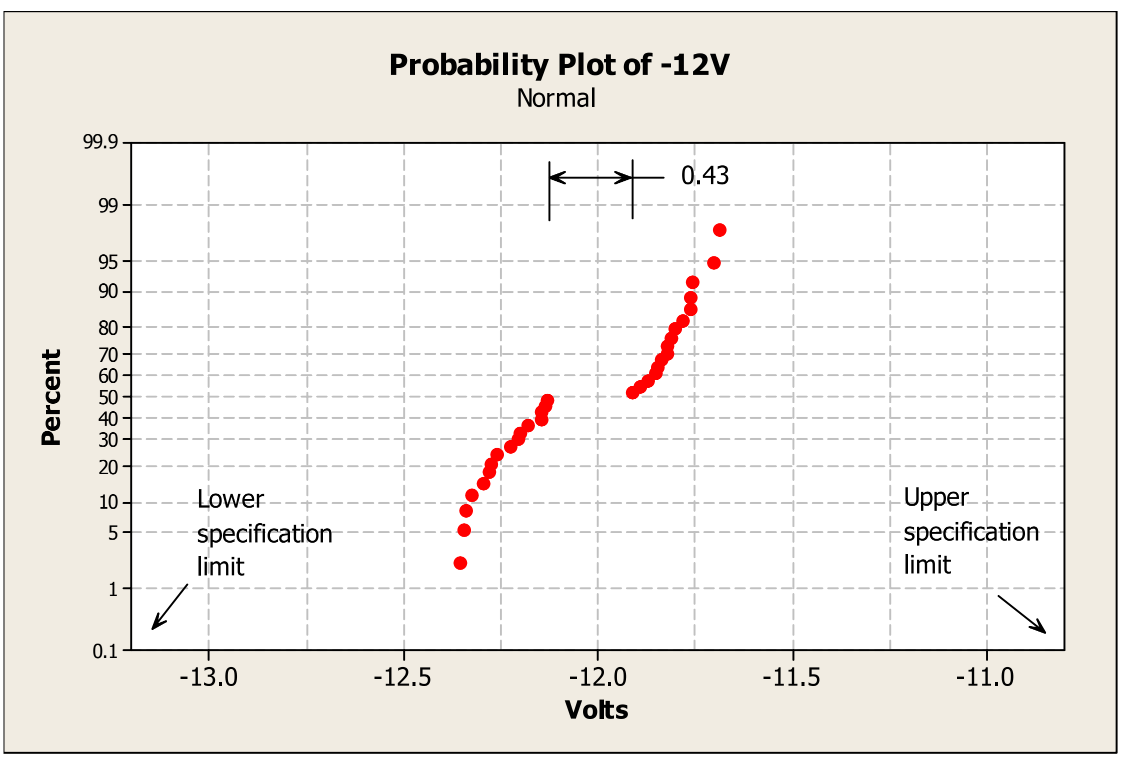

Figure 2

Consider now the 3.4-V output. A normal probability plot of the 3.4-V effects is shown in Figure 3. The 3.4-V loading effect appears most statistically significant when followed by the temperature effect. The third most statistically significant effect is an interaction between loading and temperature.

Since there is two-factor interaction confounding in this resolution IV experiment design, this interaction effect is considered to be a suspect. This probability plotted point is contrast column 15 from the design matrix shown in Table M4[1]. Table N2[1] indicates that contrast 15 represents the aliased two-factor interactions of *EF, GH, AJ,* and *BK,* which physically are frequency at ac input/load on −12-V output (*EF*), load on −5-V output/load on 12-V output (*GH*), ambient temperature/load on 3.4-V output (*AJ*), and input ac voltage range/load on programmable output (*BK*). A technical assessment can then be made as to which interaction is most likely to affect the −12-V output level. For this situation, it was concluded that the most likely interaction was temperature/load on minus12-V output (*AJ*); i.e., an interaction between the two main effect factors that were found significant. Obviously, if it is important to be certain about this contrast column effect, a confirmation experiment would need to be conducted.

A DCRCA probability plot of the 32 trial outputs, similar to the −12-V analyses, is shown in Figure 4. The grouping of data in this plot illustrates the previously suspected two-factor interaction. When this probability plot is compared to the specifications of 3.15 to 3.65, we note that many of the data points fall outside the lower specification limit. It appears that the supplier adjusts the power supply to a 3.4-V output under a low-load condition. With additional load, the output decreases to a value close to the lower specification limit. One potential temporary resolution to the problem is to set the non-load voltage to the high end of the specification limits.

The next question of concern is whether any other parameters should be considered. One possible addition to the variability conditions is circuitry drift with component age. Another is variability between power supply assembly units.

To address the between-assemblies condition, multiple units can be used for these analyses. However, the test duration and resource requirements could be much larger. Other alternatives are to capture additional information derived from historical information or evaluate a sample of parts held at constant conditions. This can also serve as a confirmation experiment to assess the validity of the fractional factorial experiment conclusions.

Voltage responses from 10 power supply assemblies were plotted on a probability plot. The test data consist of two points for each assembly taken at the two extreme interaction conditions. The plot indicates that approximately 99.8% of the population variability is within a 0.1-V range.

The information gleaned from the power supply qualification test process can be used to determine which parameters need to be monitored or controlled in the manufacturing process. If this problem escaped development tests, manufacturing would probably report it as a no-trouble-found conduction when failing units were returned by the customer. These no-trouble-found issues can have a major financial impact on both the computer manufacturer and their customers.

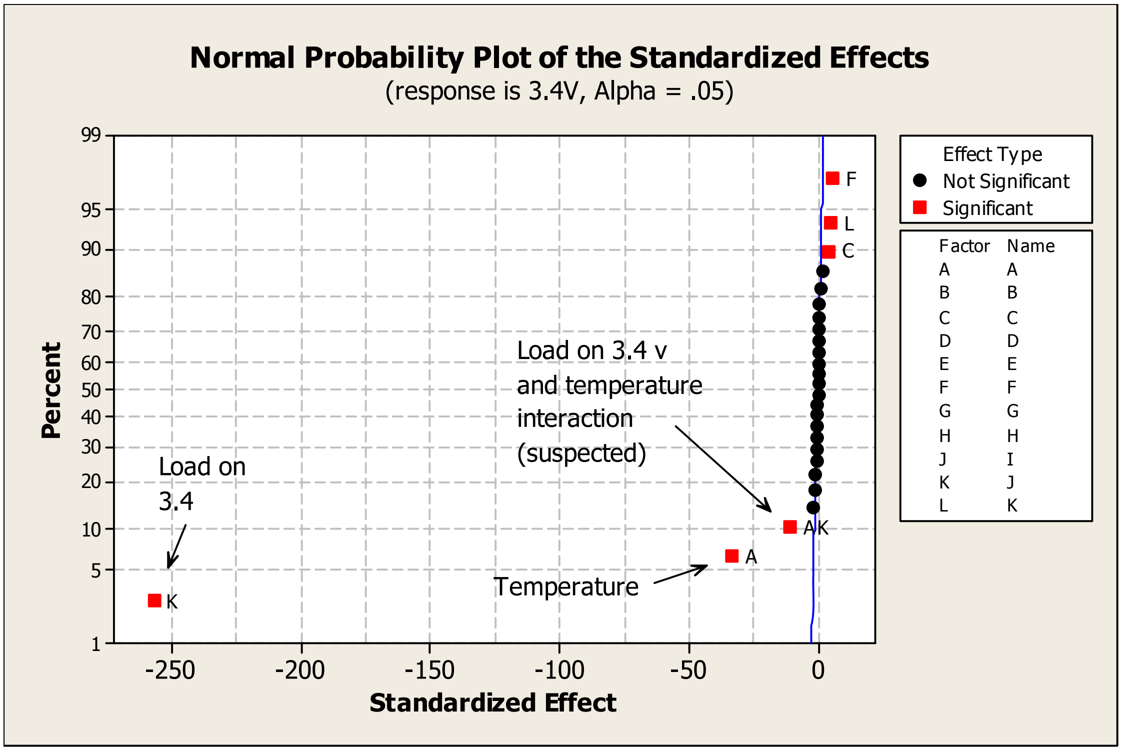

Figure 3

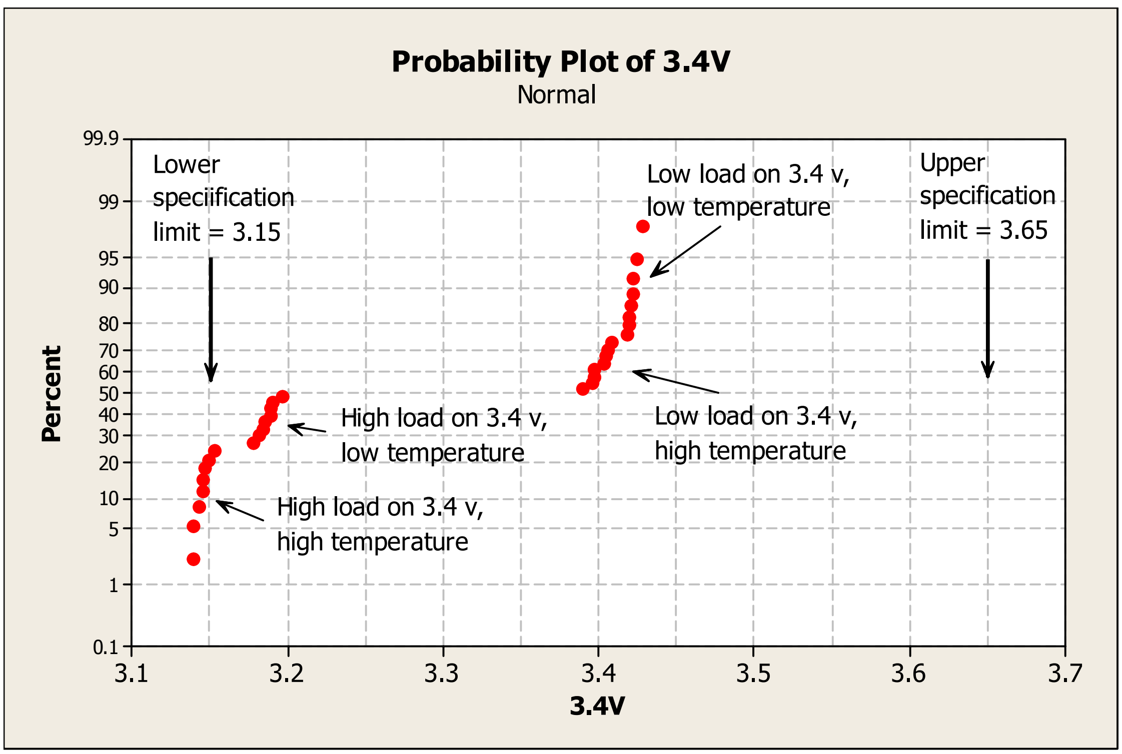

Figure 4

CONCLUSIONS

The described methodology can help organizations improve their enterprise design process efficiency and effectiveness. In a computer-industry DFSS Design-Measure-Analyze-Design-Verify (DMADV) process, a verify-phase application of this procedure can help reduce the number of future post-ship no-trouble founds.

REFERENCE

1. Breyfogle, Forrest *Implementing Six Sigma*, 2nd edition, Wiley, Hoboken, NJ, 2003.