WHERE 3458A TEST LIMITS COME FROM

Clint Snyder

June 1988

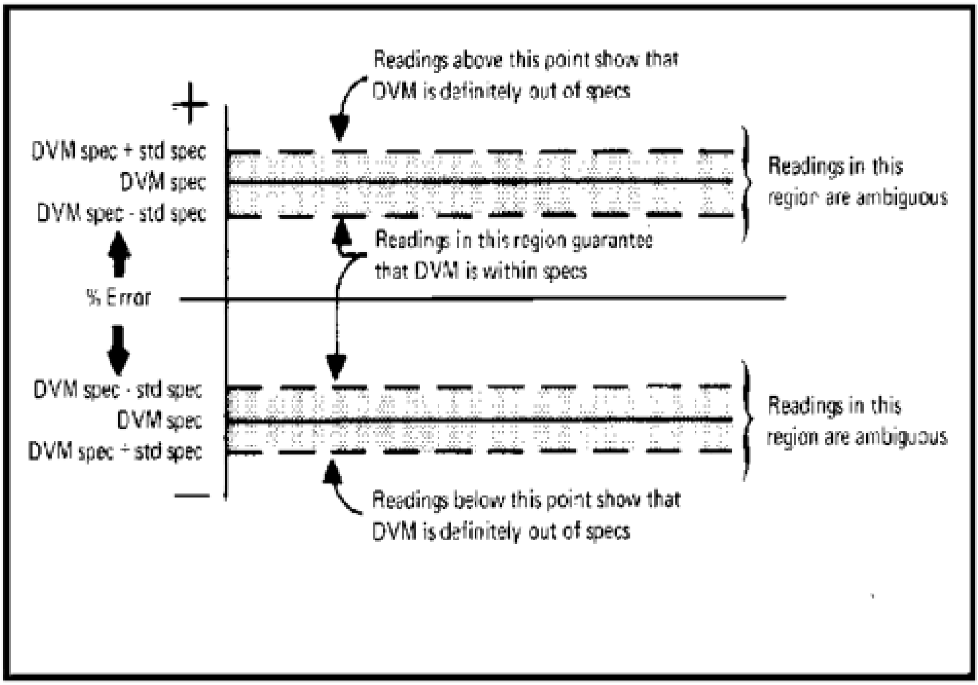
Abstract:

This paper gives a brief description of the statistics involved in setting appropriate test limits for verifying 3458A specifications. The case for verifying a 3458A to 1-year specifications using a 90-day verified 3458A is examined. This paper does not discuss the issues of calibration definitions, adjustments, etc.

PROBLEM:

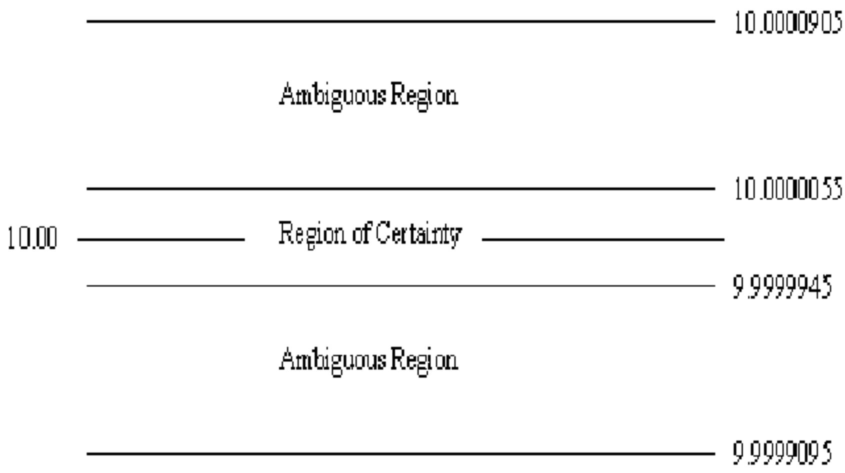
The 3458A has one real "problem"--it's different than our previous multimeters in two major areas. First, its accuracy is much better in relation to currently available calibration standards, and secondly, its specifications are written in relative terms instead of absolute values. While it is pretty easy to understand relative specs, the issue of high accuracy in relation to available calibration sources causes test limit derivation to be a little tricky.

Let's look at an example. Suppose our multimeter accuracy is +/- .0025%, and our calibration standard's accuracy is +/- .00025x. (ten times better). If I dial up 10.00000 V on the standard and the multimeter reads 10.00026, is it in spec? I'm not sure! If the standard is putting out 10.00002 V (within spec) then the meter is fine (within .0024%). However, if the standard is putting out 9.99998 V (still within spec), then the multimeter is out of tolerance (.0028% off). There are ambiguous regions as shown in the following figure.



By the way, this example was for the 3456A...so this problem is not new!

Let's look at an example using a 3458A Option 002 and a Datron 4000A, each at 90-day specs on the 10V range using a calibration uncertainty of 2 ppm. for each. The uncertainties will be 4.8 ppm. for the 3458A and 4.25 ppm. for the Datron 4000A.



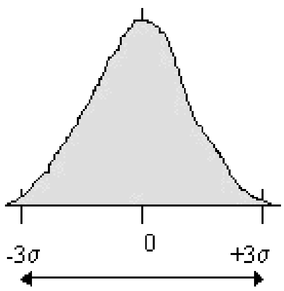
The region of certainty is 5.5 uV on each side of a perfect 10 volts, while the region of uncertainty is 85 uV tacked on to each side of that. The ambiguous region is about 15 times larger than the region where the outcome is sure!

What's the point? There are always ambiguous regions when comparing a multimeter to a calibration standard. When the relative accuracy's are on the same order of magnitude for both instruments then the regions of uncertainty are quite large. Consequently, more readings are apt to fall in the ambiguous regions. Is there any way to make a meaningful statement about readings that fall in these regions? Statistics can help.

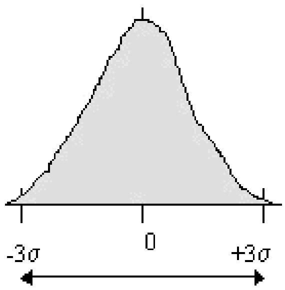
<u>STATISTICS:</u>

Everyone remembers from their statistics course that if the area under a Normal distribution is equal to one, then the area between limits set on the curve can be viewed as the probability of an event having a result that falls between those limits. Right ?!

We can consider that there is a Normal distribution for every 3458A specification. Every 3458A produced (that's not broken) will return a value that falls somewhere under the distribution curve. While we do not have an infinite sample of 3458A's, it is obvious that adding or removing a few units is not going to change the shape of the distribution greatly. We can also consider that the mean of this distribution is zero--that is, Box A is as likely to read 7ppm high as read 7ppm low. I believe our data sheet is written to reflect 3 standard deviations (3) of the distribution for each specification.

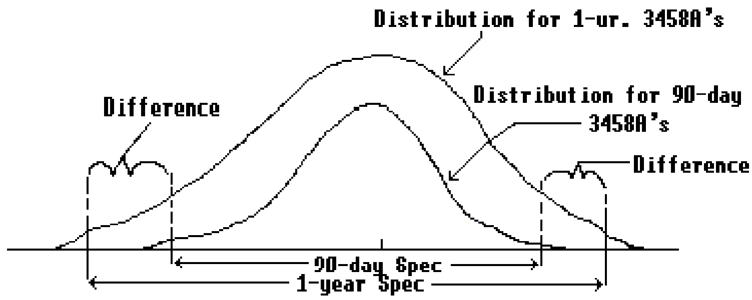


10Vdc spec



10mA DCi spec

Remembering that our problem is finding a source accurate enough to verify the 3458A specs, statistically how much more accurate does this source have to be? Would 2 times better be good enough to give us a meaningful result? How about 1.2 times better? Most people like to see a 4:1 accuracy ratio but that s pretty tough for the 3458A. Because of current hardware constraints in the field, we decided to look at using a 3458A verified to 90-day specs to verify another 3458A to 1-year specs. Consider these distributions for a given spec:



The question is "how much confidence can I get out of the difference in accuracy specs between the two distributions?".

Our general test procedure is to measure a decent source (like a Datron 4000A) with the 3458A-90 day unit, and then to measure it with the 3458A unit being tested. We are making the reasonable assumption that the source will not drift appreciably over a couple of minutes. The value of interest is the absolute value of the difference between the two readings. Let's call it delta. Obviously, if delta is less than the 1-year spec minus the 90-day spec (shown as "difference" on the graph above) then things are great. But what if delta is greater than this but still less than the 1-year spec? LID Statistician Greg Larsen did a computer simulation of this problem and generated some tables like the following:

Our Test Says …

		PASS	FAIL
Unit is actually …	GOOD	95.97% 99.74%	4.03% 75.62%
	BAD	16.34% 0.26%	83.66% 24.38%

Before I explain what these numbers mean, let s stop a second and think about what we want a performance test to do.

1. A customer has their box tested by Agilent. If it passes they should be very confident that it is indeed good.
2. It's not right for bad boxes to pass our verification tests.
3. We don t want good boxes to fail our verification tests.

Needless to say, we can t have 100% of all of these.

Now let's take a look at some of those numbers.

- 95.97% of all boxes tested by our procedure will PASS the given test.
- 99.74% of all boxes that PASS the test for a spec are indeed good on that spec.
- 75.62% of the 4.03% that FAIL are actually good on that spec.
- We have an 83.66% chance of catching bad boxes with our test.
- If we test a box and say it PASSes then there is a .26% chance that it is actually bad for that spec.

Box confidence:

If I 'm 99.76% sure that a box which passes the 10V DC test is really good, how sure am I that the entire unit is good? Not 99.76%! The statistical answer is (.9976)^N, where N is the number of independent tests done on the unit. N is a very subjective number, but I feel it is not unreasonable to use the number of functions the box has. I don't feel that several tests of a given function are truly independent. So, (.9976)^7 is .9833. Am I 98% certain that the box is totally good? No. Many specifications are not tested by our current Performance Test.

I feel like 95% is a fair guesstimate of total box confidence after a unit passes all the Performance Tests. Remember, there are no hard and fast (or any at all, for that matter) rules regarding the confidence level a Performance Test should provide. It is largely an exercise in Human Engineering. The important point is--don t tell a customer that passing our Performance Test makes them 99.76% confident that the whole 3458A is good. Each test makes them 99.76% confident about that spec only. If they are truly concerned about more points, or a particular point, the formula in Appendix A can be used to calculate the appropriate limits.

Appendix A
Limit Derivation Formula

In the general sense the formula for this type of test is:

$$\text{Delta} = Z * \sqrt{\left(\frac{\text{Spec1}}{\text{Tspec1}}\right)^2 + \left(\frac{\text{Spec2}}{\text{Tspec2}}\right)^2}$$

where:

Delta = Absolute value of difference of values returned by the test

Z = Confidence factor from a Normal Table

Tspec1 = Number of Standard Deviations at which Spec 1 is written

Tspec2 = Number of Standard Deviations at which Spec 2 is written

The recommended formula for 3458A limits is:

$$\text{Delta} = 1.96 * \sqrt{\left(\frac{\text{90-day spec}}{2.5}\right)^2 + \left(\frac{\text{1-yr. spec}}{2.5}\right)^2}$$

Note that Z = 1.96 and Tspec's equal to 2.5 result in the confidence levels discussed earlier in this paper. The 3458A specs are actually written at 3 Standard Deviations, but I took a conservative approach. The difference isn't great. Again, note that this is only good for the testing method described in this paper. If you have a calibration source that is three or four times more accurate than the spec being tested, then a straight comparison is probably adequate. Another fine point is that I algebraically added the errors to arrive at the 3458A specs. Many people, like Datron, RSS (Root Sum Square) these errors to make the end result look better.