Sampling for Validation Testing

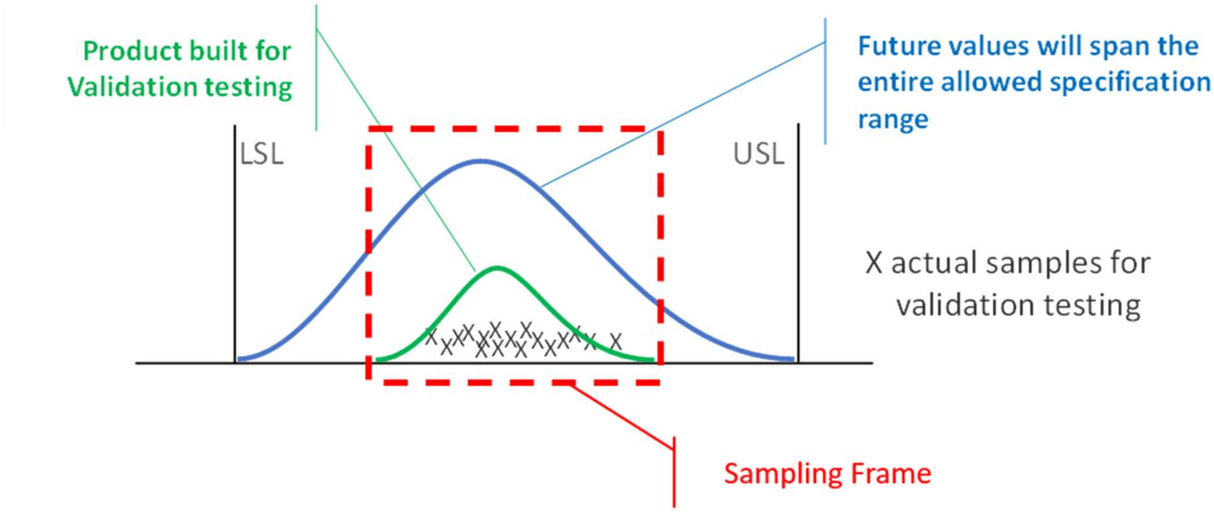
It's important to understand that Validation is not development. The development process is intended to understand and improve the process and product such that it meets requirements. Critical characteristics should have been determined and realistic tolerances established through experimentation that will guarantee that future product will meet the performance specifications and requirements. Validation is intended to be a confirmation that the product that meets the specifications will in fact meet requirements.

Often the requirements include some acceptable defect rate for the users or Customers. This is where people begin to get confused about what kind of 'sampling plan' they should use. The two default sample plans are the AQL inspection plans and the Confidence/Reliability plans. Neither of these plans are particularly useful for validation.

The AQL and Confidence/Reliability plans are based on inferential statistics and are not predictive of future of events. They can only tell you that the population (or frame) that you sampled from might not exceed some defect level. And for AQL plans that defect level isn't even the AQL value, it's the LTPD (or RQL) level that is rarely specified or checked. These plans are intended for lot to lot acceptance sampling: You have a lot in front of you, you need to determine if it is too many defects to ship or if it has a minimal number of defects and can be shipped without screening (or scrapping). Statistically, these types of plans are not very effective at differentiating defective lots from acceptable lots[i].

The biggest problem with using sampling plans for validation is shown in figure 1 below[ii]. A single statistical sample is only as good as the actual population from which the samples are drawn. This is called the frame. Remember validation is confirming that ALL parts made in the future will meet requirements. Unfortunately, validation typically occurs at the end of development before manufacturing begins in earnest and so there are only a few parts made to run validation testing on. If the parts that are made for validation are near the target – or at least do not include parts at the specification limits – no amount of zero defects found, sample size, confidence level or wishful thinking will provide you with any reliable assurance that parts made in the future at the allowable min and max specification limits will not fail, or will only fail at some small defect rate. It is simply impossible. The parts that could fail must be included in the actual parts from which the sample is taken for any conclusion regarding the failure rate can be reliably made.

Figure 1

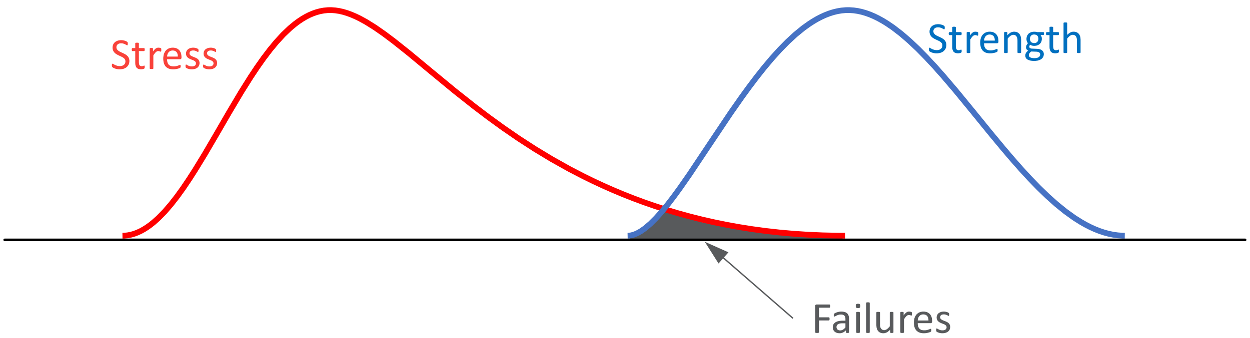


Statistical sampling plans might provide some degree of prediction regarding the future IF the population of manufactured parts that will be sampled are built from across the entire range of allowable future parts. Of course, we must ensure that the sample of tested parts includes parts made at the outer limits and that the distribution of the sampled parts is representative of the distribution that will be manufactured in the future. Without this the estimates of how defective we will be in the future will be wrong. This is the drawback of random sampling and inferential statistics. One could use conditional probability to determine the actual expected defect rate if the distribution of samples does not match the distribution of the population, but that is a complex task.

An easier and far more predictive way is to not use statistical sampling at all, but to leverage your knowledge of the physics. If you test only one part at the min and one at the max limits and neither one fails, you can be reasonably sure that your future failure rate (based on the characteristic(s) for which you tested) will be zero. Of course, it is possible that there are characteristics that are critical and you didn't determine them in development and they aren't active in the parts used in validation testing; those might affect your failure rate at some future time. We always have that possibility so that kind of miss is no excuse for not testing appropriately in validation.

A second physics situation that often occurs is where a defect also requires a condition for failure. This is best described as the classical stress-strength interaction shown in figure 2.

Figure 2



In the case of a condition for failure we must include the worst-case conditions in our testing scheme in validation or again we have excluded the real opportunity for a failure to occur as it will in the future. Conditions for failure are also part of the sampling frame. We can again reduce sample sizes by testing the worst-case parts that are allowed by specification at the worst-case conditions. If there are no failures, then we can be assured that the failure rate will be reasonably near zero.

The important mind shift here is that when we use this approach, we are moving from inferential statistics to pure probability.

---

[i] Wheeler, Donald J., "The Truth About Acceptance Sampling, Part 1, What can you say about this lot?", Quality Digest, July 2014
https://www.qualitydigest.com/inside/six-sigma-column/truth-about-acceptance-sampling-part-1-070214.html

Wheeler, Donald J., "The Truth About Acceptance Sampling, Part 2, How to avoid asking the wrong question?", Quality Digest, August 2014
https://www.qualitydigest.com/inside/six-sigma-column/truth-about-acceptance-sampling-part-2-080414.html

W. E. Deming, "*On Probability as a Basis for Action*", American Statistician, November 1975, Vol. 29, No. 4, pp. 146-152 https://www.deming.org/media/pdf/145.pdf

[ii] Please note that while the figures use bell shaped curves they are for convenience only. The Normal distribution doesn't exist in real life.

Pyzdek, Thomas, "Non-Normal Distributions in the Real World", *Quality Digest*, December 1999, https://www.qualitydigest.com/magazine/1999/dec/article/non-normal-distributions-real-world.html)