

The Truth About Acceptance Sampling: Part Two

How to avoid asking the wrong question

Donald J. Wheeler

One of the common tools of Quality Assurance is Acceptance Sampling. Acceptance Sampling uses the observed properties of a sample drawn from a lot or batch to make a decision about whether to accept or reject that lot or batch. While the textbooks are full of complex descriptions of various Acceptance Sampling plans, they never stop to tell you what you know about the lot you have just sampled. How to do this was the topic of Part One. Here we shall look at how Acceptance Sampling plans attempt to describe the quality in the warehouse.

THE QUALITY IN THE WAREHOUSE

A company making hardwood flooring used the following procedure as their quality assurance plan. After fabricating six-by-six parquet tiles, the tiles were glued together into 12-by-12 squares and boxed 25 squares to a box.

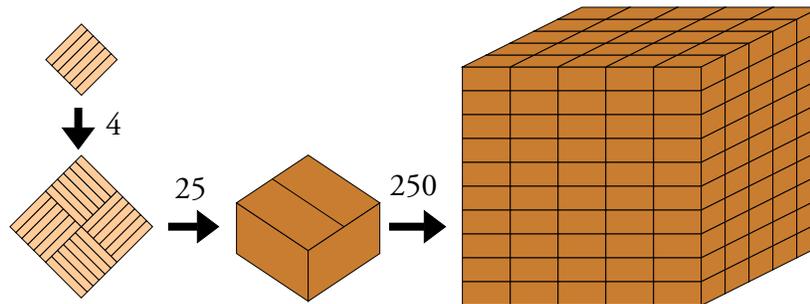


Figure 1: Pallets of Parquet Flooring

As each pallet of 250 boxes was ready to move to the warehouse the quality auditor would select seven boxes (175 squares) for inspection based upon a Dodge-Romig 5% Average Outgoing Quality Limit (AOQL) plan. If fifteen or more defective squares were found in these seven boxes the pallet would be rejected and subjected to 100% inspection. The pallet would be torn down, the remaining 243 boxes would be opened, all of the squares would be inspected, defective squares would be replaced with good squares, and then all would be reboxed and the boxes would be replaced on the pallet.

If the number of defective squares was fourteen or less the pallet would be accepted. Any defectives found in the sample would be replaced with good squares, the seven boxes would be closed up and replaced on the pallet, and the pallet would be moved to the warehouse.

Thus, the warehouse was filled with pallets of parquet floor tiles. Some pallets would contain material that had been subjected to 100% screening inspection. Other pallets would contain material substantially the same as it came from the production line. Thus, the warehouse

was a checkerboard of pallets having two different levels of defects.

To illustrate this point assume that 2.5 percent of the six-by-six tiles are defective. When four of these tiles are glued into 12 by 12 squares we will end up with 9.6 percent defective squares because:

$$\text{Fraction of Defective Squares} = [1 - (1 - 0.025)^4] = 0.0963$$

This means that the seven-box samples will average about 16.8 defective squares per sample. As a result, the AOQL plan described above will reject and screen about 71 percent of the pallets. The other 29 percent of the pallets will be accepted and shipped to the warehouse. In this scenario the warehouse will end up containing about 97 percent good squares overall, but while 71 percent of the pallets will be virtually perfect, the remaining 29 percent will have an average of 9.5 percent defective squares after the defectives in the seven sampled boxes have been replaced.

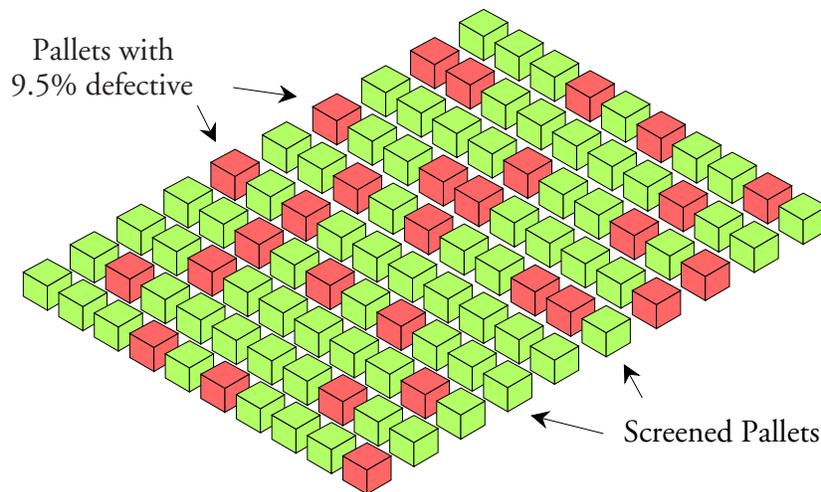


Figure 2: The Pallets in the Warehouse

One day I received a call from the quality control manager of this plant. He wanted to know if he could assure the president of the company that there was at least 95% good product in the warehouse. Since he was using a 5% AOQL plan the answer to his question was yes. In the scenario above it was actually 97% good stuff. However his question was the wrong question because the customers didn't buy the warehouse—they bought this product by the box.

A distributor gets a pallet of floor tiles and sells a few boxes at a time to the installers. When the distributor gets a screened pallet, the installers find the material to be satisfactory, and everyone builds up certain expectations regarding this product. When the distributor gets an unscreened pallet, the installers find more defective squares than they had expected. As a result, they may not have enough material to finish their job. They have to get more material. In the meantime the contractor and the homeowner are unhappy about the delay. Thus, the customers are unhappy, the installer is unhappy, and the distributor is unhappy about all the complaints and the returned material he has to handle. Yet the president of the flooring company and his quality control manager sleep soundly each night, comforted by the knowledge that there is at least 95% good stuff in the warehouse!

Hit-or-miss inspection based upon Acceptance Sampling will often make things worse by

creating an inconsistent product stream!

THE QUALITY OF THE LOT

All of the numbers used to describe an Acceptance Sampling plan apply to the warehouse. Characterizations such as AOQL, Lot Tolerance Percent Defective (LTPD), and Acceptable Quality Level (AQL) describe long-term properties of the product stream. None of them tell you anything about the lot you have just accepted or rejected. Yet the whole purpose of sampling is to discover something about the lot you have in hand.

Consider what we can say about the pallets above using the table of 95% Agresti-Coull interval estimates given in Part One. In the example above, the acceptance number was 14 out of 175 and the rejection number was 15 out of 175. Say the inspector ends up with 14 defective squares in his 175 piece sample. While this sample contains $14/175 = 0.08$ or 8 percent defective, the 95% Agresti-Coull interval estimate for the fraction nonconforming in this pallet is 4.8 percent to 13.1 percent. Since all the accepted pallets will have 14 or fewer defectives in the sample we can say that they are all likely to have fewer than 13.1 percent defective squares. Clearly, pallets with less than 13.1 percent defective deserve to be accepted!

If the inspector finds 15 defective squares in his 175 piece sample his sample fraction defective will be $15/175 = 0.086$ or 8.6 percent. The 95% Agresti Coull interval estimate for the fraction nonconforming in the pallet is 5.2 percent defective to 13.8 percent defective. Since all the rejected pallets will have 15 or more defectives in the sample we can say that they are all likely to have more than 5.2 percent defective squares. Clearly, pallets with more than 5.2 percent defective deserve to be rejected!

So, which is it? Do you want to accept those lots with less than 13.1 percent defective, or do you want to reject those lots with more than 5.2 percent defective? This AOQL plan lets you alternate between doing both!

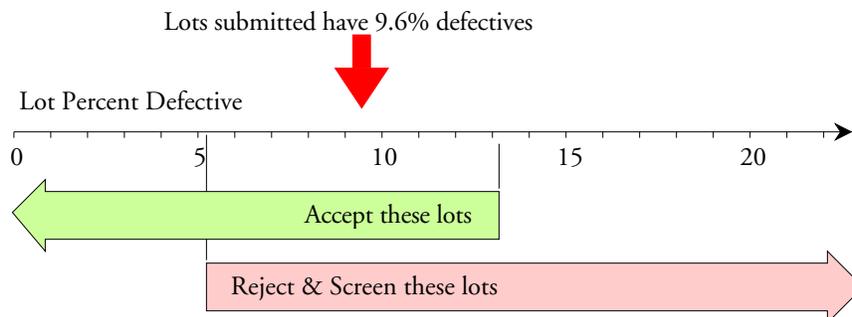


Figure 3: Actions Taken Based on the 5% AOQL Plan for Flooring

It is instructive to note that the interval estimates found above are consistent with the earlier estimate of 9.6 percent defective squares in the product stream. In fact, samples with anywhere from 9 to 24 defective squares out of 175 will yield interval estimates that contain 9.6 percent defectives. The formulas for an np-chart would suggest that with $n = 175$, sample values between 6 and 28 are consistent with a product stream having 9.6 percent defectives. Yet here we accept those pallets where we find up to 14 defectives and reject and screen those pallets where we find

15 or more defectives. Thus, by using this AOQL plan we end up taking actions that depend more upon the luck of the draw than anything else. Thus, Acceptance Sampling is just another way to play roulette!

RANDOM SAMPLES AND CONVENIENCE SAMPLES

The computations behind the 95% Agresti-Coull interval estimates assume that you have a random sample from a uniform lot. A random sample is defined as one where each of the items in a lot has the same chance of being included in the sample. This is usually expressed by saying that a random sample has to give “equal and complete coverage” to the product in the lot. The objective of this requirement of a random sample is to assure that the sample is, in some logical sense, representative of the lot.

However, in practice, the samples are rarely random. Given the 250 boxes of parquet squares as shown in Figure 4, where would you select your sample of seven boxes? In industrial practice the samples are almost always drawn from the end of the roll, the top of the basket, and the outer layer of boxes on the pallet.

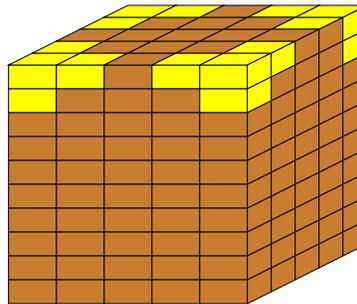


Figure 4: Where a Pallet of Boxes Gets Sampled

Such samples can only be called “convenience samples.” Since these samples do not have the property that every portion of the lot has the same chance of being included, we cannot consider them to be “random samples,” and the whole house of cards known as sampling theory comes tumbling down.

Given the complexities of breaking down a pallet, opening every box, inspecting all 6075 remaining squares, and repacking them all, how long do you think it will be before the workers begin to load “specially selected” boxes on the outer corners of the pallets? By putting 16 good boxes on the corners they can greatly affect the outcome of the Acceptance Sampling plan. When this happens the percentage of rejected pallets will drop, and the quality in the warehouse will drift back toward 9.5% defective throughout, in spite of the use of a 5% AOQL plan.

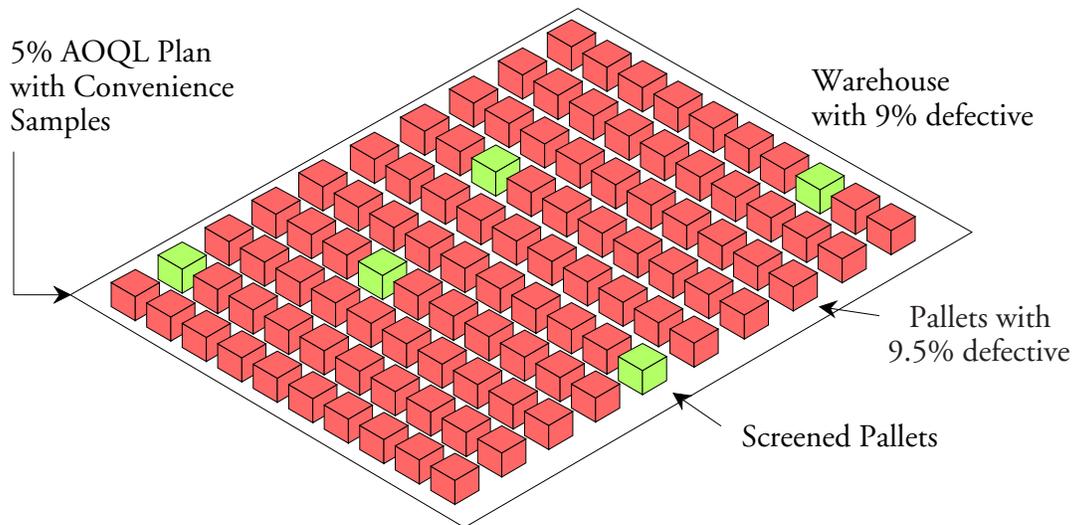


Figure 5: Pallets in the Warehouse After Workers Place “Specially Selected Boxes” on Corners

In addition to the problem of using a convenience sample, there is the fact that the inspector selects boxes rather than individual squares. While this is much easier than selecting individual squares, it changes the nature of the sampling from simple random sampling to cluster sampling. Of course the AOQL plan was designed with simple random sampling in mind. This slight change in how the sample is obtained introduces considerable complexities into the mathematical theory, and yet these complexities are rarely considered in practice. We usually assume our convenience sample is equivalent to a random sample and proceed to interpret our results using the standard theory. But what do these violations of the theory behind Acceptance Sampling do to the interpretation of the interval estimate?

While the 95% Agresti-Coull interval estimates may characterize the uncertainties associated with drawing a random sample from a uniform lot, they cannot characterize the non-sampling errors associated with using convenience samples from non-uniform lots. Since these non-sampling errors are likely to increase the uncertainties associated with the extrapolation from the sample to the lot we may be worse off than the interval estimate might lead us to believe. We certainly are unlikely to ever be better off. Thus, with convenience samples the interval estimates become a “best-case scenario.”

UNIFORM LOTS

However, there is a remedy for the problem introduced by the use of convenience samples. If the lot is known to be reasonably uniform, any convenience sample will be equivalent to a random sample. This equivalence is actually a by-product of the mathematics, but it is not a trivial point.

The mathematical reason for the assumption of a uniform lot is to justify the use of a single statistic to characterize a single property for the lot as a whole. If the lot is made up of several sublots, and if these sublots have different properties, then how can a single number computed from the sample describe the multiple properties of the lot? The whole theory of descriptive statistics, probability theory, and statistical inference assume that the sample is drawn from a

uniform lot.

To understand the imperative nature of having uniform lots consider the fact that between 1935 and 2012 there was an average of 2.56 major North Atlantic hurricanes per year (Category 3 or larger). This average is descriptive of this time period. An approximate 95% interval estimate for the average number of major hurricanes per year is 2.2 to 3.0. But what do these values represent? This 78 year interval is divided by NOAA into four periods: two periods with low hurricane activity and two periods with high hurricane activity. The 38 years from the periods of low activity only averaged 1.49 major hurricanes per year. The 40 years from the periods of high activity averaged 3.55 major hurricane per year. So which weather pattern is characterized by the average of 2.56: The two periods of low activity, or the two periods of high activity? Of course the answer is neither one.

Before a descriptive statistic can be generalized to characterize something beyond the sample itself you need to have a uniform universe from which the sample was obtained. Without this uniformity the statistic cannot be used for an extrapolation. Yet extrapolation is the heart of the Quality Assurance question.

So how can we know if the lot is uniform? This requires process data collected while the lot is being produced. If we have this type of data, we can place them on a process behavior chart. If this process behavior chart shows no evidence of a lack of homogeneity, then we can conclude that the lots are reasonably uniform. Moreover, these process data will often allow us to fully characterize each lot so that Acceptance Sampling is not needed. The convenience samples used to construct the process behavior charts may also serve to answer the Quality Assurance question.

When the production process is operated predictably we can use the samples obtained for the process behavior chart to also characterize the product stream. With a predictable process the product stream will be uniform and consistent. This predictability will justify the extrapolation from the samples to the product not measured, and the consistency justifies the assumption that the samples are representative of the product stream.

So how can we justify the extrapolation from the sample to the lot? We can either use carefully defined random samples from lots that we assume are uniform, or we can use convenience samples for lots produced by a predictable process. Anything else is just wishes and hopes. And the only way to know if the lot was produced by a predictable process is to use a process behavior chart.

UNIFORMITY LOT TO LOT

Now consider the second part of Acceptance Sampling schemes—the decision to accept some lots and reject others. The implicit assumption of Acceptance Sampling is that the lot quality is highly variable from batch to batch. If this were not so, why would we need to accept some batches and reject others? However, if the lot quality is varying widely from batch to batch, then how can we be assured regarding the uniformity of the product within each batch? The acceptance or rejection of each batch will depend upon the extrapolation from the product measured to that product which has not been measured, and the assumption of uniformity within each batch will be essential for this extrapolation to make sense with convenience samples. So using Acceptance Sampling is rather like wanting to eat your cake and have it too; you must

assume that the product quality is highly variable from batch to batch, but that, at the same time, it is very uniform within each batch.

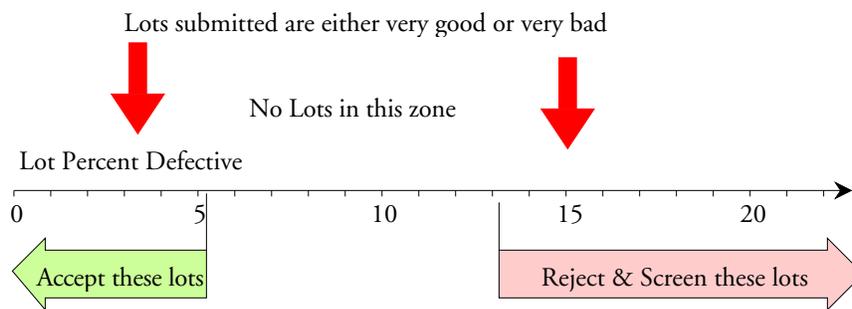


Figure 6: When the 5% AOQL Plan for Flooring Would Work Well

If the product quality is changing from batch to batch, is it not likely that it is also changing within each batch? And if this is the case, then how likely is it that our convenience samples will properly characterize each batch?

If the product quality is uniform from batch to batch, then why do we need to accept some batches and reject others?

Thus, the only situation where Acceptance Sampling schemes will work with convenience samples is one where the lots are known to be internally uniform, but where the lot quality is highly variable from batch to batch. This is the case in Figure 2. Thus, the condition assumed by the use of Acceptance Sampling is often the condition created by the use of Acceptance Sampling.

INSPECTION

Hit-or-miss inspection based upon Acceptance Sampling will not minimize the cost of any operation. The role of inspection is to improve the economics of production. Inspection upstream will reduce subsequent costs by reducing the waste when the defectives are found downstream. Three times in my brief conversation with the quality manager of the flooring company he commented on how painful it was to throw away the three good six-by-six tiles every time they found a square with one defective tile. Consider the impact of the 5% AOQL plan they were using. With 2.5 percent defective six-by-six tiles they have 9.6 percent defective squares. Their plan would result in screening 71 percent of the pallets. With replacement of defective squares this means that they will have to produce material for 114.3 pallets in order to actually get 100 pallets in the warehouse.

Contrast this with what they could do with 100% inspection for the six-by-six tiles prior to gluing them into squares. Say this inspection is 90% effective and cuts the fraction of defective tiles from 2.5% to 0.25%. This would result in one percent of the 12 by 12 squares being defective. Boxing these and sending them straight to the warehouse without any further inspection would result in 99% good stuff on each pallet, and 99% good stuff throughout the warehouse as well. To get 100 pallets in the warehouse would require the production of six by six tiles sufficient for 102.3 pallets. Given the complexities of screening 71% of the pallets, this 100% inspection upstream would be cheaper to implement than the Acceptance Sampling plan, would result in higher process yields, and would produce a consistent stream of higher quality product.

SUMMARY

The whole thrust of Acceptance Sampling is to appear to have scraped the burnt toast. By focusing on the quality in the warehouse we can distract people from thinking about the quality of the lot in hand. By sampling we can give the appearance of having done something about quality without having to actually perform a 100% inspection. By accepting or rejecting each lot, regardless of whether or not we have sufficient evidence to do so, we pretend to know things that we do not know. When this ignorance catches up with us, we simply blame it on a bad sample, and go on as before. Thus, Acceptance Sampling is nothing more than a band-aid on the problem of nonconforming product.

Using interval estimates will allow you to approximate what you do not know about the lot in hand. They may only be a best-case scenario, but they at least quantify some of the uncertainty in your extrapolation from the sample to the lot.

Finally, to use an Acceptance Sampling plan you have to simultaneously believe two contradictory things about successive lots: Each lot is uniform while successive lots can be very different in quality. This means that about the only time Acceptance Sampling might be appropriate is when you have to clean up the mess left when someone upstream used Acceptance Sampling as their quality assurance plan.

When it is a matter of rectifying defects, the only economic levels of inspection are all or nothing. Hit-or-miss inspection based upon Acceptance Sampling will not minimize the cost of any operation. In fact, it can actually make things worse!