How Acceptance Sampling Works

What can you say about the sampled lot?

Donald J. Wheeler

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, there are some very important aspects of Acceptance Sampling that are not included in the textbooks. These are the topic of this paper.

EXTRAPOLATION

In the interest of simplicity, the product that has been measured will be referred to as the "sample," while the product that has not been measured will be referred to as the "lot." Every time that we attempt to use a sample to characterize the quality of a lot we will be making an extrapolation from the product that has been measured to the product that has not been measured. So how can we extrapolate from the sample to the lot?

Extrapolation inevitably requires an assumption that the sample is *representative* of the lot. When this assumption is appropriate, the extrapolation will make sense. When this assumption is wrong, the extrapolation will also be wrong. No matter what computations we may use, the validity of our statements about the lot will always come back to the question of representativeness. The ability of any sample to represent a lot will be affected by the homogeneity of the lot itself and the manner in which the sample is selected from the lot.

We shall begin with the case of random samples drawn from uniform lots. Given a lot having uniform quality, and given a random sample obtained from this lot, we can use probability theory to develop a mathematical treatment of the uncertainty of the extrapolation from the sample to the lot.

ESTIMATES OF LOT QUALITY

Needless to say, the twin assumptions of uniform lot quality and a random sample remove a host of problems. A random sample may be broadly defined as one that is obtained in such a way that every item in the lot has the same chance of being included in the sample. When the assumptions of uniform lot quality and a random sample are satisfied, the following results will hold.

Whether we are measuring a property, testing for functionality, or counting blemished items, a sample of **n** items will contain some fixed number of nonconforming items, Y, and the sample fraction nonconforming will be:

$$\hat{p} = \frac{Y}{\mathbf{n}}$$

This value is known as the "binomial point estimate" for the lot fraction nonconforming. For hundreds of years it has been known to be the best single-number estimate of the lot fraction nonconforming. However, we need more than a single number estimate. We also need an interval estimate in order to incorporate the uncertainty of our extrapolation from the sample to the lot.

The best modern practice here is to use a 95% Agresti-Coull interval estimate. This interval estimate will be centered on the ratio.

$$\tilde{p} = \frac{Y + 2}{\mathbf{n} + 4}$$

And our 95% Agresti-Coull interval estimate may be found using the formula:

$$\tilde{p} \pm 1.96 \sqrt{\frac{\tilde{p}\ (1 - \tilde{p})}{\mathbf{n} + 4}}$$

This Agresti-Coull estimate is good all the way down to $Y = 0$ and all the way up to $Y = \mathbf{n}$.. A table of 95% Agresti-Coull interval estimates for various values of Y and **n** is given in Figure 1.

Y = 0	**Y = 1**		**Y = 2**		**Y = 3**		**Y = 4**		**Y = 5**		**Y = 6**			
n	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
5	0.0	49.4	2.5	64.1	12.0	76.9	23.1	88.0	35.9	97.5	50.6	100.		
10	0.0	32.6	0.0	42.9	4.9	52.2	10.6	60.8	16.9	68.8	23.8	76.2	31.2	83.1
15	0.0	24.3	0.0	32.2	2.7	39.4	6.5	46.1	10.7	52.5	15.2	58.5	19.9	64.3
20	0.0	19.4	0.0	25.7	1.8	31.6	4.6	37.1	7.7	42.3	11.0	47.4	14.5	52.2
25	0.0	16.1	0.0	21.4	1.2	26.3	3.5	31.0	5.9	35.4	8.6	39.7	11.3	43.9
30	0.0	13.8	0.0	18.4	0.9	22.6	2.8	26.6	4.8	30.5	7.0	34.2	9.3	37.8
35	0.0	12.1	0.0	16.1	0.7	19.8	2.3	23.3	4.1	26.7	5.9	30.0	7.8	33.2
40	0.0	10.7	0.0	14.3	0.6	17.6	2.0	20.7	3.5	23.8	5.1	26.7	6.8	29.6
45	0.0	9.6	0.0	12.8	0.5	15.8	1.7	18.7	3.1	21.4	4.5	24.1	6.0	26.7
50	0.0	8.7	0.0	11.7	0.4	14.4	1.5	17.0	2.7	19.5	4.0	21.9	5.3	24.3
60	0.0	7.4	0.0	9.9	0.3	12.2	1.2	14.4	2.2	16.5	3.3	18.6	4.4	20.6
70	0.0	6.4	0.0	8.5	0.3	10.6	1.0	12.5	1.9	14.3	2.8	16.1	3.7	17.9
80	0.0	5.6	0.0	7.5	0.2	9.3	0.9	11.0	1.6	12.7	2.4	14.2	3.2	15.8
100	0.0	4.6	0.0	6.1	0.2	7.5	0.7	8.9	1.3	10.3	1.9	11.5	2.6	12.8
120	0.0	3.8	0.0	5.1	0.1	6.3	0.6	7.5	1.1	8.6	1.6	9.7	2.1	10.8
150	0.0	3.1	0.0	4.1	0.1	5.1	0.4	6.0	0.8	7.0	1.3	7.8	1.7	8.7
175	0.0	2.7	0.0	3.6	0.1	4.4	0.4	5.2	0.7	6.0	1.1	6.8	1.4	7.5
200	0.0	2.3	0.0	3.1	0.1	3.9	0.3	4.6	0.6	5.3	0.9	5.9	1.3	6.6
250	0.0	1.9	0.0	2.5	0.0	3.1	0.3	3.7	0.5	4.2	0.7	4.8	1.0	5.3
300	0.0	1.6	0.0	2.1	0.0	2.6	0.2	3.1	0.4	3.5	0.6	4.0	0.8	4.4

	Y = 7		**Y = 8**		**Y = 9**		**Y = 10**		**Y = 11**		**Y = 12**		**Y = 13**	
n	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
10	39.2	89.4	47.8	95.1	57.1	100.	67.4	100.						
15	24.9	69.8	30.2	75.1	35.7	80.1	41.5	84.8	47.5	89.3	53.9	93.5	60.6	97.3
20	18.1	56.9	21.9	61.4	25.9	65.8	30.0	70.0	34.2	74.1	38.6	78.1	43.1	81.9
25	14.2	47.9	17.2	51.8	20.3	55.6	23.5	59.3	26.7	62.9	30.1	66.5	33.5	69.9
30	11.6	41.3	14.1	44.7	16.6	48.1	19.2	51.4	21.9	54.6	24.6	57.7	27.4	60.8
35	9.9	36.3	11.9	39.3	14.1	42.3	16.3	45.3	18.5	48.1	20.8	51.0	23.2	53.7
40	8.5	32.4	10.3	35.1	12.2	37.8	14.1	40.4	16.1	43.0	18.1	45.6	20.1	48.1
45	7.5	29.2	9.1	31.7	10.8	34.1	12.4	36.5	14.2	38.9	15.9	41.2	17.7	43.5
50	6.7	26.6	8.2	28.9	9.6	31.1	11.1	33.3	12.7	35.5	14.2	37.6	15.8	39.7
60	5.5	22.6	6.7	24.5	7.9	26.4	9.2	28.3	10.5	30.2	11.7	32.0	13.1	33.8
70	4.7	19.6	5.7	21.3	6.8	23.0	7.8	24.6	8.9	26.2	10.0	27.8	11.1	29.4
80	4.1	17.3	5.0	18.8	5.9	20.3	6.8	21.8	7.7	23.2	8.7	24.6	9.7	26.0
100	3.3	14.1	3.9	15.3	4.7	16.5	5.4	17.7	6.1	18.9	6.9	20.0	7.7	21.2
120	2.7	11.8	3.3	12.9	3.9	13.9	4.5	14.9	5.1	15.9	5.7	16.9	6.4	17.8
150	2.1	9.5	2.6	10.4	3.1	11.2	3.6	12.0	4.1	12.8	4.6	13.6	5.1	14.4
175	1.8	8.2	2.2	9.0	2.6	9.7	3.0	10.4	3.5	11.1	3.9	11.8	4.3	12.4
200	1.6	7.2	1.9	7.9	2.3	8.5	2.7	9.1	3.0	9.7	3.4	10.3	3.8	10.9
250	1.3	5.8	1.5	6.3	1.8	6.8	2.1	7.3	2.4	7.8	2.7	8.3	3.0	8.8
300	1.1	4.9	1.3	5.3	1.5	5.7	1.8	6.1	2.0	6.6	2.2	7.0	2.5	7.4

	Y = 14		**Y = 15**		**Y = 16**		**Y = 17**		**Y = 18**		**Y = 19**		**Y = 20**	
n	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
15	67.8	100.	75.7	100.										
20	47.8	85.5	52.6	89.0	57.7	92.3	62.9	95.4	68.4	98.2	74.3	100.	80.6	100.
25	37.1	73.3	40.7	76.5	44.4	79.7	48.2	82.8	52.1	85.8	56.1	88.7	60.3	91.4
30	30.3	63.8	33.2	66.8	36.2	69.7	39.2	72.6	42.3	75.4	45.4	78.1	48.6	80.8
35	25.6	56.5	28.0	59.2	30.5	61.8	33.0	64.4	35.6	67.0	38.2	69.5	40.8	72.0
40	22.1	50.6	24.2	53.0	26.4	55.4	28.5	57.8	30.7	60.2	33.0	62.5	35.2	64.8
45	19.5	45.8	21.4	48.0	23.2	50.2	25.1	52.4	27.1	54.6	29.0	56.7	31.0	58.8
50	17.5	41.8	19.1	43.9	20.8	45.9	22.4	47.9	24.2	49.9	25.9	51.9	27.6	53.8
60	14.4	35.6	15.7	37.4	17.1	39.1	18.5	40.9	19.9	42.6	21.3	44.3	22.7	46.0
70	12.2	31.0	13.4	32.6	14.5	34.1	15.7	35.6	16.9	37.1	18.1	38.7	19.3	40.1
80	10.7	27.4	11.6	28.8	12.7	30.2	13.7	31.6	14.7	32.9	15.7	34.3	16.8	35.6
100	8.5	22.3	9.2	23.5	10.0	24.6	10.8	25.7	11.7	26.8	12.5	27.9	13.3	29.0
120	7.0	18.8	7.7	19.8	8.3	20.7	9.0	21.7	9.7	22.6	10.3	23.5	11.0	24.5
150	5.6	15.2	6.1	16.0	6.6	16.8	7.1	17.5	7.7	18.3	8.2	19.1	8.8	19.8
175	4.8	13.1	5.2	13.8	5.7	14.5	6.1	15.1	6.6	15.8	7.0	16.4	7.5	17.1
200	4.2	11.5	4.5	12.1	4.9	12.7	5.3	13.3	5.7	13.9	6.1	14.5	6.5	15.0
250	3.3	9.3	3.6	9.8	3.9	10.2	4.2	10.7	4.6	11.2	4.9	11.7	5.2	12.1
300	2.8	7.8	3.0	8.2	3.3	8.6	3.5	9.0	3.8	9.4	4.1	9.8	4.3	10.1

Figure 1: 95% Agresti-Coull Interval Estimates for Percent Nonconforming in Lot

Say, for example, that a random sample of **n** = 175 parts has been selected from a lot containing 6,250 parts, and when measured we find that $Y = 3$ of these parts are nonconforming. Then our best point estimate for the lot fraction nonconforming would be (3/175) = 1.7 percent, while from Figure 1 we find our 95% Agresti-Coull interval estimate for the lot fraction nonconforming to be 0.4 percent to 5.2 percent. This means that we can be reasonably certain that the lot fraction nonconforming is no more than 5.2%, and also that it is probably greater than 0.4%. With these data we cannot really narrow it down more than this.

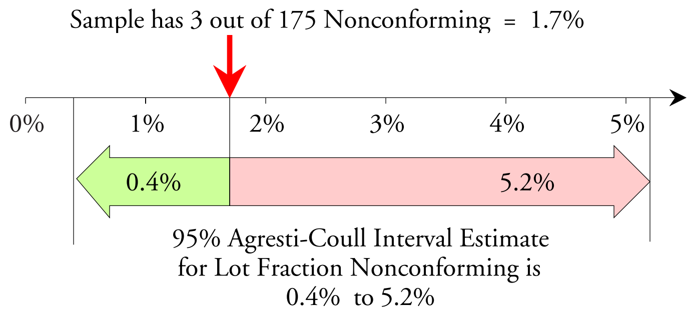

Figure 2: What We Know About the Lot when $Y = 3$ and n = 175

If knowing that there is more than 0.4% nonconforming in this lot is bad enough to justify rejecting the lot, then reject this lot.

On the other hand, if knowing that there is less than 5.2% nonconforming in this lot is good enough to justify accepting the lot, then accept this lot.

But if 0.4% nonconforming is not bad enough to justify rejecting the lot, or if 5.2% nonconforming is not good enough to justify accepting the lot, *then you do not have sufficient evidence to take either action*.

Figure 3 illustrates this truth about Acceptance Sampling. Think about this very carefully.

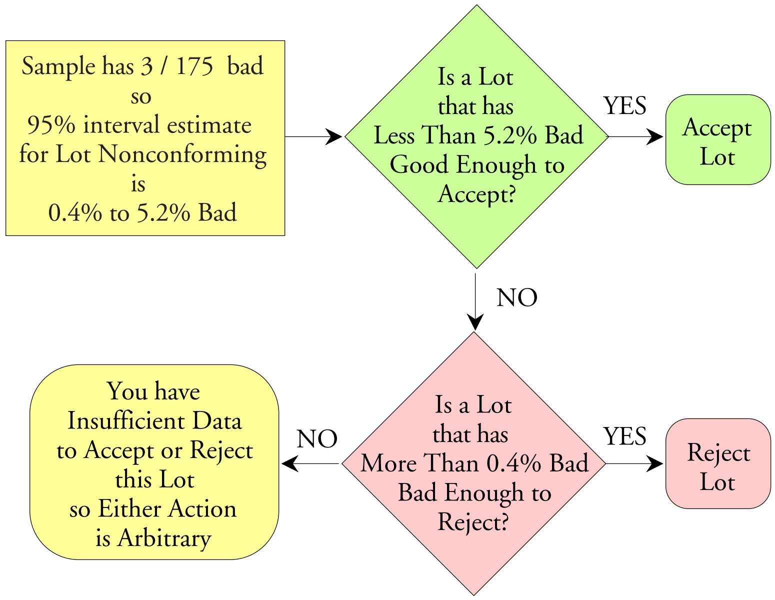

Figure 3: The Truth About Acceptance Sampling

In addition to forcing a decision based on inadequate information, Acceptance Sampling also implicitly assumes that there is some non-zero level of nonconforming product that is *acceptable*. Thus, Acceptance Sampling always aims at less than perfection.

Rather than using an interval estimate to characterize what the sample reveals about the lot being sampled, Acceptance Sampling plans are commonly characterized by quantities such as the Acceptable Quality Level (AQL); the Lot Tolerance Percent Defective (LTPD); or the Average Outgoing Quality Limit (AOQL). All of these numbers describe long-term properties of the product stream, and as such they apply to the warehouse as a whole.

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.

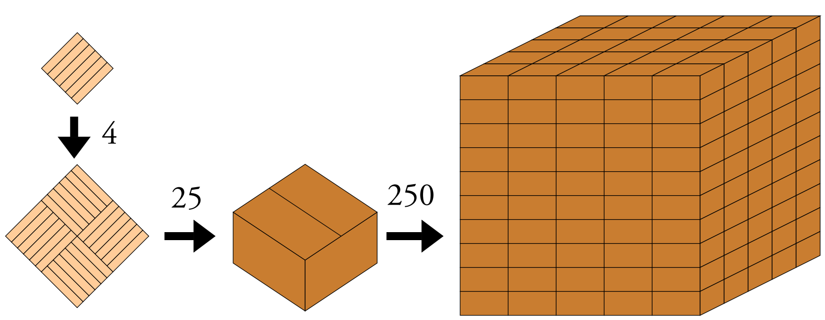

Figure 4: Pallets of Parquet Flooring

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 defective product.

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.

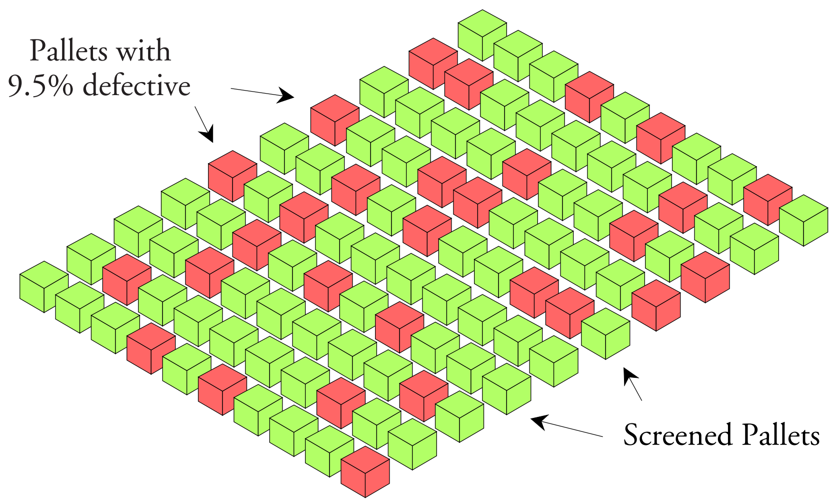

Figure 5: The Pallets in the Warehouse

At this point the quality manager 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 the warehouse actually had 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 manager sleep soundly each night, comforted by the knowledge that the warehouse contains at least 95% good stuff!

Hit-or-miss inspection based upon Acceptance Sampling will often make things worse by creating an inconsistent product stream.

WHAT CAN BE SAID ABOUT THE PALLETS?

Consider what the 95% Agresti-Coull interval estimates can tell us about the pallets above. All of the pallets that were accepted had 14 or fewer defective squares in their 175 square sample. For those pallets with 14 defectives the upper 95% Agresti-Coull interval estimate is 13.1%. So this 5% AOQL plan will accept some pallets having up to 13.1% nonconforming.

Those pallets that were rejected and screened all had at least 15 defectives in their 175 square sample. For those pallets with 15 defectives the lower 95% Agresti-Coull interval estimate is 5.2%. Those pallets that were rejected are likely to have had at least 5.2% defective squares before screening.

So this 5% AOQL plan will reject some of the pallets having more than 5.2% nonconforming while accepting others having up to 13% nonconforming.

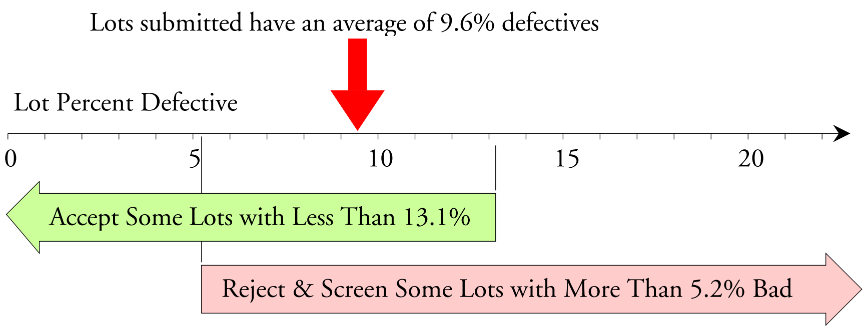

Figure 6: Actions Taken Based on the 5% AOQL Plan for Flooring

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 5% AOQL plan lets you alternate between doing both. Here our actions depend more upon the luck of the draw than anything else. Acceptance Sampling is just another way to play roulette!

RANDOM SAMPLES AND CONVENIENCE SAMPLES

The computations behind all Acceptance Sampling plans 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 assumption is the basis for the extrapolation from the sample to the lot. However, in practice, the samples are rarely random. Given the 250 boxes of parquet squares as shown in Figure 7, 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.

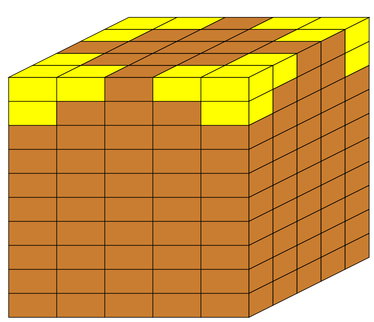

Figure 7: Where a Pallet of Boxes Gets Sampled

Given the complexities of breaking down a pallet, opening every box, inspecting the 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.

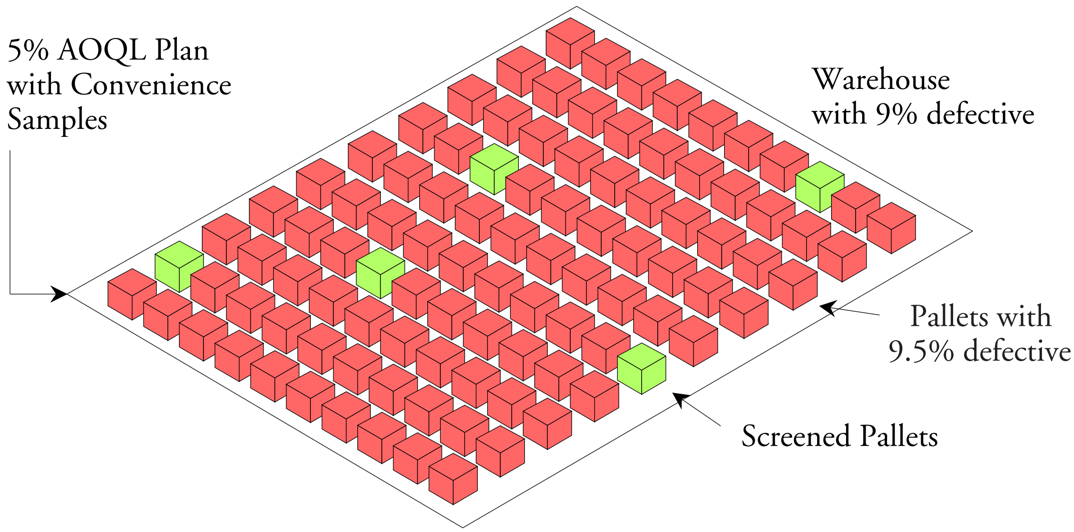

Figure 8: Pallets in the Warehouse After Workers Place "Specially Selected Boxes" on Corners

In addition to the problem of using a convenience sample, there is the fact that the inspector selects boxes rather than individual squares. While the 95% Agresti-Coull interval estimates may characterize the uncertainties associated with drawing a random sample from a uniform lot, they cannot begin to characterize the non-sampling errors associated with using convenience samples or selecting boxes rather than individual squares. 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 "best-case scenarios."

TWO ASSUMPTIONS OF ACCEPTANCE SAMPLING

Acceptance Sampling implicitly makes two contradictory assumptions about your product stream. The first of these is that the product quality is uniform throughout each lot. If a lot is internally uniform, then the extrapolation from the sample to the lot will be reasonable. But how can this assumption be justified in the absence of data? Without a process behavior chart that shows a predictable product stream any assumption about within-lot uniformity is simply wishful thinking.

The second assumption of Acceptance Sampling is that the lot quality is highly variable from lot to lot. If this were not so, why would we need to accept some lots and reject others? Acceptance Sampling plans cannot reliably separate good lots from bad lots unless the lot qualities are substantially different. In the example given here, when each lot is internally uniform, the 5% AOQL plan will reliably sort lots with more than 13% defective from those with less than 5% defective. But as shown back in Figure 6, it will do a poor job with uniform lots of intermediate quality.

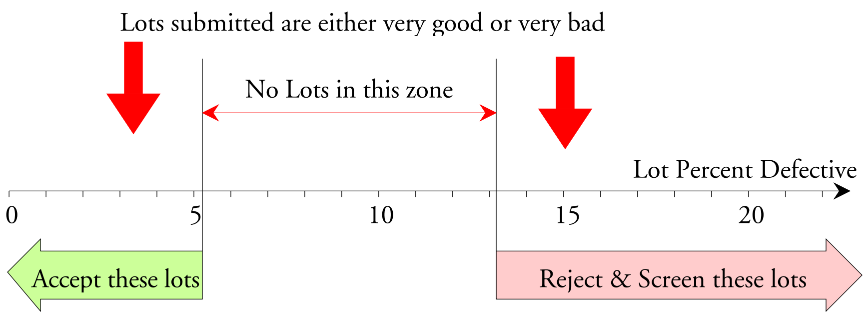

Figure 9: When the 5% AOQL Plan for Flooring Would Work Well

So where will Acceptance Sampling work? You will have to have some reason to believe that each lot in internally uniform, and yet also believe that the lot quality is substantially different from lot to lot.

But if your process is being operated predictably, then the lots are likely to be internally uniform and also uniform from lot to lot. Any use of Acceptance Sampling here will arbitrarily reject and accept lots having essentially the same quality level.

On the other hand, when your process is being operated unpredictably, process changes are unlikely to only occur between lots. Each lot may be non-uniform, which will undermine the extrapolation from your convenience sample to the lot as a whole. Any use of Acceptance Sampling here will fail to separate the good stuff from the bad stuff since each lot is likely to have some of both.

So when you use Acceptance Sampling you must assume that the product quality is highly variable from lot to lot, but that, at the same time, it is very uniform within each lot.

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

If the product quality is uniform from lot to lot, 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 lot to lot. This is the case in

Figure 5. Thus, the condition assumed by the use of Acceptance Sampling is often the condition created by the use of Acceptance Sampling.

INSPECTION VERSUS ACCEPTANCE SAMPLING

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 strike a compromise with imperfection. It begins with the idea that there is some non-zero level of nonconforming product that is acceptable. Next it uses hit or miss inspection to give the appearance of having scraped the burnt toast. And by focusing on the quality in the warehouse it distracts people from thinking about the quality of the lot in hand.

Thus, by using Acceptance 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 actually know about the lot in hand. These interval estimates 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!