Proper Sample Size Selection

CASE STUDY

AUGUST 2019

BY

BEN MIETHIG AND JOHN BURGERS

CENTRE FOR MECHATRONICS & HYBRID TECHNOLOGIES

224-200 LONGWOOD ROAD SOUTH

HAMILTON, ON

L8P 0A6, CANADA

Table of Contents

How to Select the Proper Sampling Size

Something that often goes unnoticed when people think of the automotive industry is the massive amounts of part testing required to validate part designs and a part's life cycle. Estimation of an adequate sample size drawn from the population is necessary to draw educated and meaningful conclusions on how durable a part will be over the course of its life. This case study uses the example of a glycol battery cooler to illustrate the collection and analysis of attribute data to determine the likelihood of cyclic loading causing a leak in the cooler's fluid fittings whose braze joints have variable quality. It will then explore how valid the conclusions drawn from this test are and what type of sample sizes would be needed to draw better conclusions of the expected failure rate of the population of parts from which the sample of parts have been taken from.

Initial Test Procedure and Results

The initial test on the battery cooling plates involved ten different cooling plates. X-rays of these plates have shown variable joint length where the fitting has been brazed together with the mirrored core plates to form the cooler. Examples of the variable joint lengths can be seen in Figure 1 where the dark object is the fitting and the gray object is the two core plates. Figure 2 shows a perspective view of the brazed assembly.

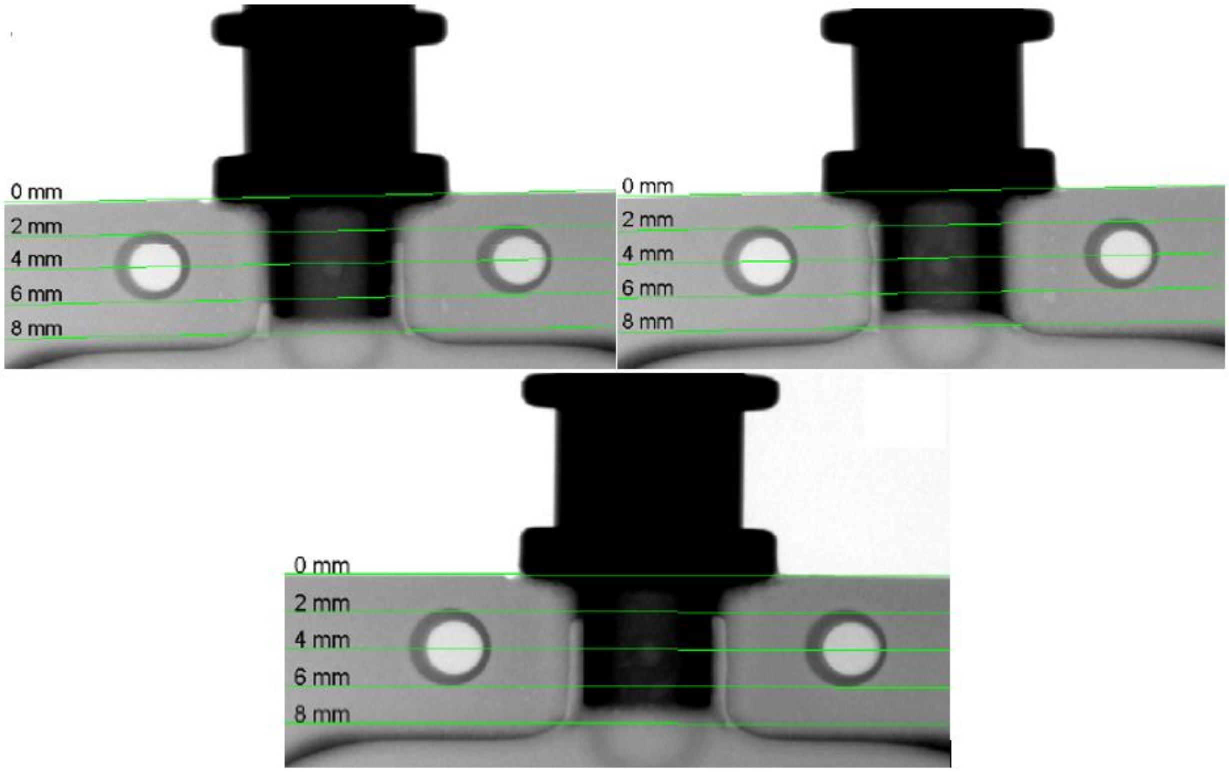

Figure 1:Example braze fills for three different fittings: top left: 6mm and 2.5mm braze fill; top right: <2mm and full braze fill; bottom: 2.25mm and 2mm braze fill

A core plate is a metal plate that forms one shell of the battery cooler. Brazing is a process used in many automotive production applications where parts are fused together in an oven, or another source of heat, by leveraging different melting points of materials. During heating, a thin layer of each core plate melts as that layer is an alloy with lower melt temperature than the remaining core plate. This thin layer of molten material flows together forming a braze seam to neighbouring parts, including the fitting which is inserted into the seam between the two core plates.

To verify that the joints of variable length are sufficiently strong compared to service and assembly loads, a test has been developed to apply a cyclical bending moment to these fittings. The data is classified as attribute data since the plates will either pass or fail the test. The attribute is whether there is a leak or not after application of the load. By doing this, the structural integrity of these plates can be analyzed and their use in the field can be validated. During assembly, the fittings on the cooler plates can be accidentally bent and the vehicle manufacturer wants to know what the potential consequences are if fittings are accidentally bent and then purposefully bent back. The bending moment applied to these fittings during assembly has been estimated to be approximately one Newton-metre.

In brief, the test procedure is to cyclically apply the bending moment to the fittings in two orthogonal directions as illustrated by the blue and orange arrows in Figure 2. All plates will be leak tested with a nitrogen-air mix before and after cyclically loading them. Any plate that exhibits a leak rate of more than 0.4 cubic centimetres per minute will be deemed to have failed. Any leak rate less than this will be considered to have passed the leak test and is therefore a part which will retain the water-glycol coolant. This is a destructive test procedure as these plates cannot be used in the field afterwards, so a sampling technique with statistical inference to the population is required.

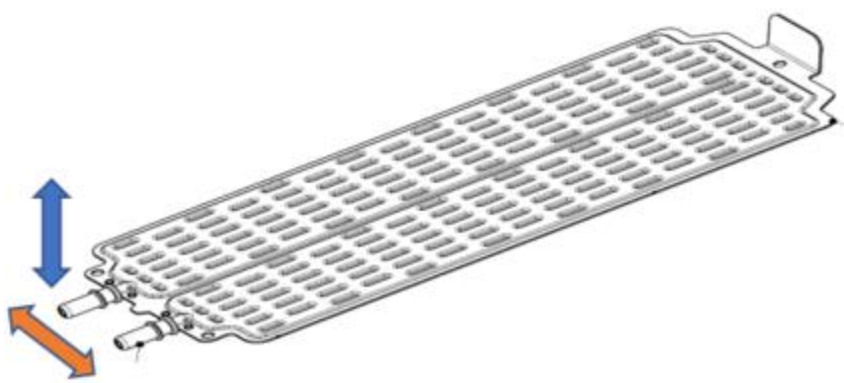

Figure 2: Depiction of battery cooler and bending moment directions

After applying 2000 cycles in both orange and blue directions (Figure 2) to each fitting on all ten plates, no failures were seen after leak testing. On the surface, this is a great result, but we need to understand the level of statistical significance for the entire production population of cooler plates.

Statistical Relevance

In order to yield statistically relevant information from data with attribute data which take on discrete values such as pass/fail, discrete data distributions must be used. Examples of these include the binomial distribution and the Poisson distribution. A classical example of something that can be represented by a binomial distribution is a calculation of the odds of rolling a certain number on a six-sided dice in a given number of rolls. For example, what are the chances that a six is rolled five consecutive times? What about on three of five rolls? These types of calculations require that the probability of one event occurring is independent of any other events occurring. For example, a six being rolled on the first roll will have no impact on a six being rolled on the second roll.

The discrete data distributions seen here can be extrapolated to a manufacturing environment in order to estimate the number of parts that will be free of defects in a given lot of parts. By knowing the distribution of defective parts, the supplier and customer can predict how many parts they order will be acceptable for production. In this case, one part being defective must not impact the quality of another part.

The binomial distribution formula can be used when the sample size is small. The following formula represents the binomial distribution:

$$P(a, n) = \binom{n}{a} * p^a * q^{n-a}$$

$$where\ p = probabilty\ of\ an\ event\ occurring;$$

$$q = probability\ of\ an\ event\ not\ occurring = 1 - p;$$

$$n = number\ of\ samples;$$

$$a = number\ of\ events\ that\ occurred;$$

When a given sample size is large, and p (the probability of failure) is small, the Poisson distribution can be calculated relatively easily. The Poisson distribution is a distribution which takes into account the average number of occurrences of an event in a given dataset. In this context, the probability of a given event occurring $\boldsymbol{a}$ times given the Poisson distribution is:

$$P(a, \mu) = \frac{e^{-\mu}\mu^a}{a!} \quad (1)$$

$$where\ \mu = mean\ number\ of\ events\ in\ a\ given\ quantity = n * p;$$

$$e = base\ of\ natural\ log;$$

For a Poisson distribution, since p is small then q is almost one, the variance, σ^2, becomes:

$$\sigma^2 = n * p * q = n * p = \mu \quad (2)$$

<u>Example 1:</u>

In this case, 10 battery coolers with varying braze fills were tested. All 10 of these parts passed the requirements of the bending and deflection tests as they did not leak. But what does this mean statistically? 10 is a very small sample size which means conclusions for the parts contained in the population of these parts cannot be made with much confidence. In order to draw conclusions for predicted failure rates using the binomial and Poisson distributions,

Confidence Limit Tables for Binomial and Poisson Distributions

Table 1: Poisson Distributions: Upper Confidence Limits for No Failures in Sample (a=0)

a	n	$p = \frac{a}{n}$	Confidence limits for $\mu = np$ corresponding to probabilities of:		
			Upper limit		
			0.1	0.025	0.005
0	5	0	1.84	2.61	3.27
	10		2.06	3.09	4.11
	20		2.17	3.37	4.65
	∞		2.30	3.69	5.30

Table 2 can be used (found on the Page 9 of this report). This table requires the user to know a given confidence interval and sample size to derive the number of expected or acceptable failures within a tested sample to meet the failure criteria in the overall population. Using the table with a 90% confidence interval and the sample size of 10, the only statistically significant conclusion that can be drawn from this test is that, on average, 2.06 parts could fail in a new sample size of 10 parts from the same lot. This is a failure rate of 20.6% which is obviously far too high for production and likely too high based on the number of failures seen in the initial test set of 10 parts. In order to bring this predicted failure rate down, the only way to accomplish this with a Poisson distribution is to increase the number of samples included in the test. For example, if the number of samples tested were to increase to 20 with zero failures, then, from

Confidence Limit Tables for Binomial and Poisson Distributions

Table 1: Poisson Distributions: Upper Confidence Limits for No Failures in Sample (a=0)

a	n	$p = \frac{a}{n}$	Confidence limits for $\mu = np$ corresponding to probabilities of:		
			Upper limit		
			0.1	0.025	0.005
0	5	0	1.84	2.61	3.27
	10		2.06	3.09	4.11
	20		2.17	3.37	4.65
	∞		2.30	3.69	5.30

Table 2 the expected number of parts that will fail in a new set of 20 tested parts would be 2.17, or 10.85%.

To further understand how this table can be used in production, the concepts of producer and consumer risk should be known. Producer risk is the risk that a difference in defects between two lots – or populations – is found when the lots have the same defect rate. This means that producers could risk throwing away a perfectly good lot of parts since the sampled parts misrepresent the defect rate in the population. Consumer risk is the risk that there are more defects in a lot despite sampling results. For consumers, a risk arises when two lots are assumed to have the same defect rate when they actually have different amounts of defects. Consumer risk decreases as the sampling size increases.

In an ideal world, the exact number of defective parts in a population would be known precisely. Since it is very costly to sample parts, and sometimes sampling is destructive to the parts, making them unusable, producers and consumers must agree on sampling plans with a certain degree of confidence that all parts within a population will meet a certain level of quality. This is where the confidence interval and number of required samples in

Confidence Limit Tables for Binomial and Poisson Distributions

Table 1: Poisson Distributions: Upper Confidence Limits for No Failures in Sample (a=0)

a	n	$p = \frac{a}{n}$	Confidence limits for $\mu = np$ corresponding to probabilities of:		
			Upper limit		
			0.1	0.025	0.005
0	5	0	1.84	2.61	3.27
	10		2.06	3.09	4.11
	20		2.17	3.37	4.65
	∞		2.30	3.69	5.30

Table 2 and Table 2 come in.

<u>Example 2:</u>

Assume that a car manufacturer is willing to accept a 10% chance they will receive a lot of parts with defects in 1% of them. What sample size would be required by the consumer to confirm that this criteria has been met?

Assuming one defect is found and the sample size of parts is large, then from Table 2, the upper confidence limit is 3.89 (meaning 3.89 parts or less in the sample will be found defective), which can then be used with the formula:

$$\mu = p * n \quad (3)$$

Where μ is the mean number of failures, p is the probability of a defect being present, and n is the sample size. Therefore,

$$n = \frac{\mu}{p} = \frac{3.89}{0.01} = 389$$

Therefore, the sample size the manufacturer will be required to inspect to verify that the lot has been sent with the specified allowable number of defects is 389 parts. In contrast, the producer of the part will need to inspect a sample from the same lot of parts to ensure that there are no more than one defect in a given sample with a 90% confidence level. This means that if they find 2 defects in a given sample, the given lot will need to be scrapped. For this, the lower confidence limit for finding 2 defects with a probability of 10% (100% minus 90%) is 0.532 (from Table 2). Rearranging Equation 3:

$$p = \frac{\mu}{n} = \frac{0.532}{389} = 0.14\%$$

This 0.14% represents what is called the producer risk in scrapping a given lot of parts when the population has a failure rate of 1% and 2 or more failures are found in inspection lot sizes of 389 parts. The following table shows the number of required samples taken from a lot in order to verify that part defect criteria have been met and the corresponding producer risk associated with those samples.

Number of defects found (a)	Upper confidence limit	n	Number of defects found	Lower confidence limit	Producer risk
0	2.30	230	1	0.105	0.046%
1	3.89	389	2	0.532	0.14%
2	5.32	532	3	1.10	0.21%

3	6.68	668	4	1.74	0.26%
4	7.99	799	5	2.43	0.30%
10	15.4	1540	11	7.02	0.46%

We conclude therefore that the sampling effort (sample – load test – leak test) should continue for 389 parts to verify that one or fewer failures are observed. This ensures the consumer is 90% confident that lots will have less than 1% failures.

Conclusion

Having derived the measured joint length from each of the samples, a correlation could also be drawn between joint length associate with the consumer risk. Since X-ray can be implemented as an in-line process measure, in which images are taken and automatically processed to find the joint lengths, one can derive a continuous variable control chart by which the process quality can be charted in order to ensure that the process remains in control.

The two brief examples show that it is not as simple as testing an arbitrary number of parts from a population for failures and making an assumption on the population based on this. In order to reliably verify that parts made in manufacturing processes are of the desired quality, rigid sampling plans must be followed. By using structured sampling plans, suppliers and consumers can have transparent expectations on the quality of the parts they receive and accountability over part quality can be taken throughout the manufacturing process. Part sampling for testing purposes is a critical aspect in quality assurance processes in manufacturing environments. While it may not play a glamorous role when considering manufacturing, it fills a vital role in the manufacturing process that ensures all parties involved accept and maintain standard measures of quality.

Confidence Limit Tables for Binomial and Poisson Distributions

Table 1: Poisson Distributions: Upper Confidence Limits for No Failures in Sample (a=0) [1]

a	n	$p = \frac{a}{n}$	Confidence limits for $\mu = np$ corresponding to probabilities of:		
			Upper limit		
			0.1	0.025	0.005
0	5	0	1.84	2.61	3.27
	10		2.06	3.09	4.11
	20		2.17	3.37	4.65
	∞		2.30	3.69	5.30

Table 2: Poisson Distributions: Confidence Limits for Large Sample Sizes (a/n ≈ 0) [1]

a	Confidence limits for $\mu = np$ corresponding to probabilities of:					
	Lower limit			Upper limit		
	0.005	0.025	0.1	0.1	0.025	0.005
0	0	0	0	2.30	3.69	5.30
1	0.005	0.025	0.105	3.89	5.57	7.43
2	0.103	0.242	0.532	5.32	7.22	9.27
3	0.338	0.619	1.10	6.68	8.77	11.0
4	0.672	1.09	1.74	7.99	10.2	12.6
5	1.08	1.62	2.43	9.27	11.7	14.1
6	1.54	2.20	3.15	10.5	13.1	15.7
7	2.04	2.81	3.89	11.8	14.4	17.1
8	2.57	3.45	4.66	13.0	15.8	18.6
9	3.13	4.12	5.43	14.2	17.1	20.0
10	3.72	4.80	6.22	15.4	18.4	21.4
11	4.32	5.49	7.02	16.6	19.7	22.8

12	4.94	6.20	7.83	17.8	21.0	24.1
13	5.58	6.92	8.65	19.0	22.2	25.5
14	6.23	7.65	9.47	20.1	23.5	26.8

All theory and tables from this case study can be found in [1]. Please refer to this book for more information on statistical sampling and how it can be completed in a practical application. The derivations for the values seen in Table 1 and Table 2 can be found in the attached MATLAB script. This script can be used to yield lower and upper confidence limits not stated in these tables as well.

A special thank you must be extended to John Burgers and Dana Canada Corporation. Without their help and support, this case study would not have been possible. Funding from the NSERC CREATE program is also acknowledged in making this case study possible.

References

[1] O. L. Davies and P. L. Goldsmith, Statistical Methods in Research and Production, 4 ed., Edinburgh, Scotland: Longman Group Ltd., 1977.

[2] Stat Trek, "Poisson Distribution," Stat Trek, 2019. [Online]. Available: https://stattrek.com/probability-distributions/poisson.aspx. [Accessed 1 August 2019].