AdeptPackaging
PACKAGING ENGINEERS & CONSULTANTS

Selecting a Sample Size

A Model for Calculating Sample Size to Allow Development Decisions

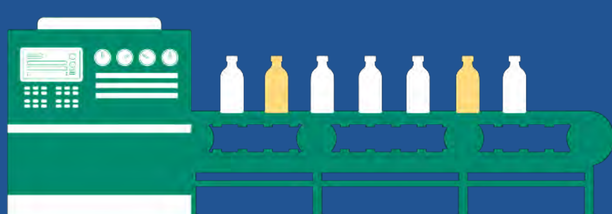
An Adept Packaging White Paper

Selecting a Sample Size

A Model for Calculating Sample Size to Allow Development Decisions

Sample Size Definition for Packaging:

A sample size refers to the number of items used to collect data for inference of information in a population.



I. Introduction

One essential step during the development of a new package is validation. During the data collection phase of validation, a common challenge engineers face is how to determine what sample size should be used to provide a result that they can apply to the population with a high level of confidence that it is accurate.

Mishandling this challenge often leads to project delays, retesting or even failed projects due to lack of accurate data.

As a fallback, packaging engineers often resort to using the acceptable quality limit (AQL) to determine sample size. For circumstances in which there is an established operating process and defect rate (operating characteristic curves), the AQL can prove effective. In circumstances where these have yet to be established, it can be misleading. The AQL is intended for acceptance sampling and does not provide data on the actual defect rate; only whether to accept or reject a lot based on a performance history.

The reality is that sample size is not difficult to determine, but does require an estimation of how the package might perform to establish a typical or expected defect rate. Often, when dealing with new package development, there is no previous defect rate data, so informed assumptions must be made. Consider making these assumptions based on previous experience with similar designs or a small sample preliminary test to estimate the defect rate.

Before getting started, it is important to understand a few statistical principles as they ensure the correct mathematical approach is utilized to determine the defect rate. This is critical since these assumptions, along with many other factors will impact the sample size determination, such as:

- Level of confidence
- Sampling technique
- Data type
- Sampling objective

For the purposes of this paper we have simplified some of the terms, phrases and concepts for easy explanations and to introduce novices to the subject of statistics.

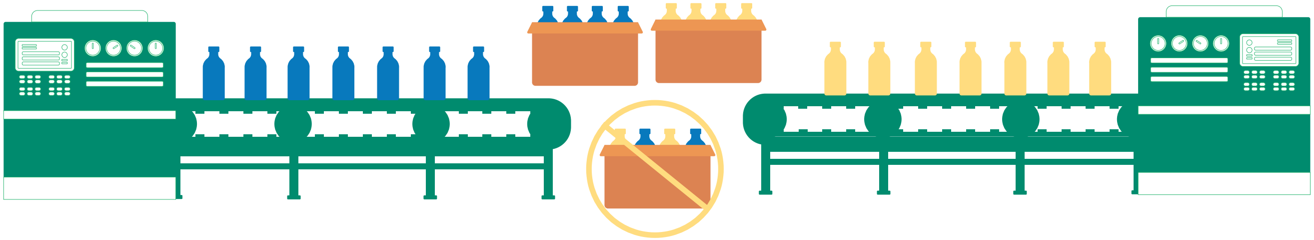
II. Principles

The following set of principles need to be clearly defined and understood to achieve representative results:

Population and Sample

A population is the entire set of entities that are under study. The sample is a subset of the population. For example, the population might be all products manufactured in a batch and the sample might be a group of products selected from that batch.

The population must be properly defined. For example, if two machines are used to make a product but the product is mixed at the end of a production line, using the mixed product as a population may be problematic. It would be preferable to treat the study as two separate populations.



Types of Data

There are generally two types of data encountered in packaging; Variables Data (aka continuous data) and Attribute data (aka discrete data).

Variables data are things that can be measured, i.e. The number of possible outcomes of each measurement is **infinite**. Example: (length of a pencil) Generally, if "variables data" can be collected, the statistics become somewhat straight forward enabling the use of mean, standard deviation, and the normal curve. This is particularly important when developing a dimensional or performance specification and is fundamental to Statistical Process Control.

Attribute data is when there are **specific** outcomes (generally two); such as good/bad, yes/no, leaking/not leaking, and are often reported in percentages.

Random

Because defects occur randomly, the samples need to be selected randomly, so that each item has an equal probability of being a part of the sample group. An entirely random sampling cannot always be achieved. The rule of thumb is to conduct it as randomly as possible given time and resources and encompass the entire population of interest.

Margin of Error and Confidence Level

The selection of confidence level and Margin of error can have an enormous impact on sample size. The more precise we need the answer to be directly correlates to the amount we need to sample. In other words, the higher confidence level we need, the more we need to sample.

Margin of Error

Margin of Error is the range in which a defect rate value falls within.

Confidence Level

Confidence level is how accurate it is that the true value falls within the Margin of Error.

III. Determining Sample Size

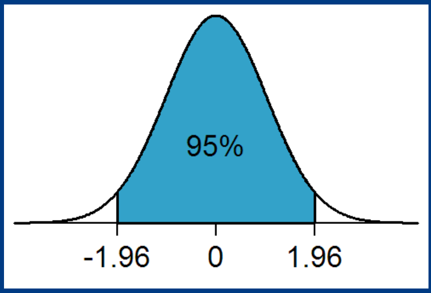
The first rule of sample size is, "more is always better than less." Larger sample sizes always provide more insight into the true defect rate. Keep in mind that statistics yield an **"estimate"** of the population data combined with a confidence level. If you want 100% confidence in the defect rate, you must sample 100% of the population. However, it is understood that in order to save time and money, it is necessary to use as small a sample size as possible to allow decisions to be made.

The majority of variables data follows the normal distribution curve. This allows us to use standard deviation and mean to characterize a population. If we have an idea of what the standard deviation should be, we can choose a preferred Margin of Error and calculate the number of samples to take using the tool below.

Adept Packaging has developed calculators to make this process easier. These tools are provided as an example of how to calculate sample size for both variables and attribute data.

Margin of Error Example:

A statement might be: there is a 95% Level of Confidence that the actual value falls between 0.85 and 0.95 (Margin of Error). This means that if we sampled 100 times, the actual value theoretically would fall between 0.85 and 0.95 in 95 out of 100 samples.



Variables Data Example:

A company has developed a new closure system and is interested in sampling bottles for removal torques. In previous testing the average torques have been 11.5 inch lbs. with a standard deviation 0.88 inch lb.

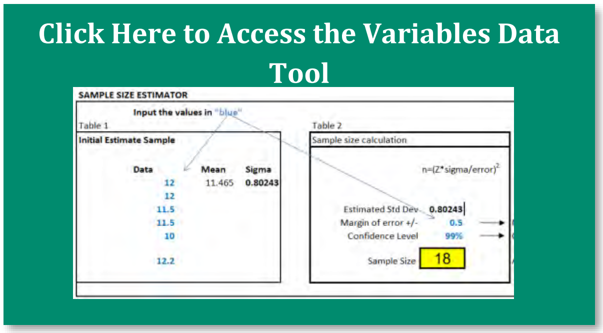
Additionally, they want to be sure that their sample mean reflects the actual population mean within +/- 0.5 inch pounds with 99% confidence.

The inputs would be:

- Estimated Standard Deviation = 0.88
- Margin of Error = 0.5
- Confidence Level = 99%
- The sample size = 21 samples

Variables Data Tool and Instructions

In order to calculate a sample size for variables data (things that can be measured), click on the Variables Data Tool below:



Instructions:

1. Insert the expected **Standard Deviation** for the data. If you know it from previous testing or experience, you can insert it in Variables Data Tool; if you don't know it, you can collect 5 or 10 samples and calculate a standard deviation and use that in the tool.

2. Insert the **Margin of Error**. This is the numerical distance that you want to have confidence that your estimate population mean falls within.

4. Lastly insert the **Confidence Level**.

5. The Sample size will appear in the **yellow box.**

Attribute Data Tool and Instructions

As mentioned, often production data is in percentages or proportions in the form of; yes/no, good/bad, leaking/not leaking, etc. This type of data follows a different distribution (hypergeometric) and is calculated differently. To calculate a sample size for attribute data, click on the Attribute Data Tool below:

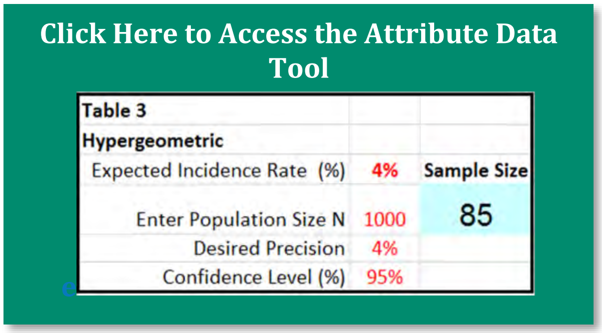


Table 3		
Hypergeometric		
Expected Incidence Rate (%)	4%	Sample Size
Enter Population Size N	1000	85
Desired Precision	4%	
Confidence Level (%)	95%	

Instructions:

1. Insert in the:
   - **Expected Incident rate (defect rate)**
   - **Population size**
   - **Desired Precision (think Confidence Interval ....+/- around the defect rate)**

- **Confidence Desired**

2. The Sample size appears in the **blue box**.

IV. Conclusion

Sampling size and testing is an iterative process. You begin with assumptions to determine a sample size and then use the data provided about the defect to calculate a new sample size; each time learning and becoming more confident in the data.

In simple situations, the statistics are relatively straight forward. When more complicated situations are encountered, the statistics can become more involved and it is recommended that a Statistician be involved.

To partner with a team that is well versed in even the most complicated circumstances and has a depth of experience calculating sample size, as well as redesigning packaging when the situation calls for it, contact us at Adept Packaging.

www.adeptpackaging.com

Appendix:

A. Formulas used for calculating sample size for Variables Data

$$\text{Sample size } n = (Z\sigma / E)^2$$

Where:

Z is the Z score

σ is the standard deviation of the population

E is Margin of error

B. Formulas used for calculating sample size for Attribute Data

where:

N is the Population Size

Z is the Z score

p is the expected population proportion defective.

q is 1-p

E is the Margin of error

$$\text{Sample size } n = \frac{N Z^2 pq}{E_2(n-1) + Z^2 pq}$$