What Do We Mean by "Zero Defects"?

Part 3 of 3: Finite Population Sampling

BY STEPHEN N. LUKO AND DEAN V. NEUBAUER

Q: How do we handle zero defects or non-conformities when sampling a finite population of objects?

A. In this the third article on the topic of zero defects, we discuss the scenario wherein sampling is done from a finite lot of objects. In the context of acceptance sampling, this is referred to as sampling an isolated lot.

We denote the lot size using N and the sample size using n. Sampling is done randomly and without replacement. Let D be the unknown possible number of non-conforming objects originally in the lot. When the sample has been collected and we find $x = 0$ non-conforming objects, the question is: What is the largest value that D could be with some stated confidence C? This upper bound value is designated D_u. Having found this value, we then may state that $D \leq D_u$ at confidence level C. Then D_u is the upper 100C% confidence bound for the unknown D.

We may also have a consumer's risk of, say, β. This means that if a non-conforming object is inspected, there is a probability equal to β that the object will be incorrectly classified as conforming and hence a risk to the consumer of β that the non-conforming object will not be detected during inspection. We further assume in this discussion that the producer's risk of classifying a conforming object as non-conforming is α = 0. (See ASTM E2334, Practice for Setting an Upper Confidence Bound for a Fraction or Number of Non-Conforming Items, or a Rate of Occurrence for Non-Conformities, Using Attribute Data, Where There Is a Zero Response in the Sample, for further details on this.)

The statistical distribution that governs this type of sampling is the hypergeometric distribution (see Reference). Using the parameters described above, assuming that β = 0, and substituting $x = 0$ in the hypergeometric formula, the following expression may be written for the probability of $x = 0$, following simplification.

$$P(x = 0) = \frac{(N - D)!(N - n)!}{(N - D - n)!N!} \qquad (1)$$

As with the cases in Parts 1 and 2 of this DataPoints article series, we want this probability to be a small but reasononable value (such as 0.05 or 0.01). We refer to this using $P(x = 0) = 1 - C$ for a chosen confidence level C. In Equation 1, replace D with D_u and set it equal to $1 - C$. Then Equation 1 takes the following form:

$$P(x = 0) = \frac{(N - D_u)!(N - n)!}{(N - D_u - n)!N!} = 1 - C \qquad (2)$$

Equation 2 must be solved numerically for D_u.

EXAMPLE 1

Suppose $N = 60$ and $n = 15$, and we want to state that $D \leq D_u$ with confidence at least 95 percent. Substituting in Equation 2 using $N = 60$ and $n = 15$, and incrementing D_u until $C \geq 0.95$ is just met, gives $D_u = 10$. The actual confidence achieved in this example is 95.8 percent. We may state with this confidence that $D \leq 10$. The value "10" is the largest value that D could have been in finding $x = 0$ in the sample and using the specified confidence. There is an approximately 4.2 percent probability that D could have been larger than 10, and therefore 95.8 percent confidence that D is not more than 10.

When there is a non-zero consumer's risk, β, at play, the event $x = 0$ can occur in two ways: 1) the sample can truly contain 0 non-conforming objects, and 2) the sample may contain some non-conforming objects, but these have been misclassified as conforming due to the consumer's risk probability β. This complicates the solution to $P(x = 0) = 1 - C$.

Essentially, $P(x = 0)$ includes the event that the sample contains all true conforming objects, the event of exactly 1 misclassified non-conforming object, exactly 2 misclassified non-conforming objects, etc., through all possible cases. Let $P(r)$ be the probability that exactly r non-conforming objects show up in the sample without a consumer's risk, i.e., $\beta = 0$. The variable r can vary between 0 and the maximum of D and n. $P(r)$ is the expression for the hypergeometric distribution:

$$P(r) = \frac{D!(N - D)!(N - n)!n!}{r!(D - r)!(N - D - n + r)!(n - r)!N!} \quad (3)$$

Note that when $r = 0$, Equation 3 equals Equation 1. The expression for $P(x = 0)$ may be shown to be:

$$P(x = 0) = P(0) + \sum_{r=1}^{n} P(r)\beta^{r} \quad (4)$$

Set Equation 4 equal to $1 - C$ and solve numerically by incrementing D until the equality is just met. This process yields the upper bound D_u for D.

EXAMPLE 2

Given: The same N and n as in Example 1, but this time with a consumer's risk of $\beta = 0.1$. Using Equation 4 with 95 percent confidence, and incrementing D until $1 - C = 0.05$ is just met, shows that $D_u = 12$. The actual confidence level achieved in this example is 95.8 percent. We may state with this confidence that $D \leq 12$.

These calculations can become computationally intense and almost always require a computer program. For modest lot and sample sizes, a spreadsheet type program works well; otherwise other tools may be required. It is sometimes disconcerting to find out that a small lot can still contain a few non-conforming units despite a relatively large sample size. The final example illustrates this occurrence.

EXAMPLE 3

In a lot of $N = 10$ objects, $n = 5$ are inspected using an electronic device for a critical attribute. No non-conformances are found, i.e., $x = 0$. The inspection device was determined to have an approximate 5 percent consumer's risk. What is the largest number of non-conforming objects that might still be outstanding in the remaining 5 uninspected objects? Using 90 percent confidence, and the methodology outlined above, this number is 3. The actual confidence achieved is 90.3 percent. Thus the upper 90.3 percent confidence bound on the remaining number of non-conforming is 3. For comparative purposes, if the consumer's risk were 0, the upper bound would still be 3 but with 91.7 percent confidence.

REFERENCE

Hogg, R. V. and Tannis, E., *Probability and Statistical Inference*, 7th edition, Upper Saddle River, NJ, Prentice Hall, 2005.

STEPHEN N. LUKO, *Pratt & Whitney Aircraft, is the immediate past chairman of Committee E11 on Quality and Statistics and a fellow of ASTM International.*

DEAN V. NEUBAUER, *Corning Inc., is vice chairman of Committee E11 on Quality and Statistics, chairman of Subcommittee E11.30 on Statistical Quality Control and chairman of E11.90.03 on Publications; he also coordinates the DataPoints column and is an ASTM International fellow.*

Statistics play an important role in the ASTM International standards you write, and a panel of experts is ready to answer your questions about how to use statistical principles in ASTM standards. Please send your questions to SN *Editor in Chief Maryann Gorman at mgorman@astm.org or ASTM International, 100 Barr Harbor Drive, P.O. Box C700, West Conshohocken, PA 19428-2959.*