

# CLOSED-FORM EQUATIONS TO DESIGN SINGLE SAMPLING PLANS FOR ISOLATED LOTS

*Giuseppe Cavone, Laura Fabbiano, Nicola Giaquinto*

Department of Electrics and Electronics, Polytechnic of Bari, 70125 Bari, Italy  
e-mail: [cavone, fabbiano, giaquinto]@misura.poliba.it

**Abstract** – In this paper, the problem of designing simple sampling plans by attributes for isolated lots is considered. The customary design procedure is based on numerical tables like those reported in the ISO 2859 norm, or on iterative algorithms implemented in a computer software, not suitable for inclusion in a norm. The paper proposes a procedure based on quite simple closed-form equations, yielding the required sample size and the acceptance number for any consistent set of specifications on the operating characteristic, in terms of AQL, RQL, PR, CR. The equations are derived using a Gaussian approximation of the hypergeometric distribution. It is shown that the proposed design procedure, besides being very simple, is accurate (i.e. the specification are accurately met) and universal (i.e. the method is valid for a wide range of specifications).

**Keywords:** sampling plans, statistical quality control, ISO 2859.

## 1. INTRODUCTION

A well-known and widely used tool for quality management in industry is statistical sampling of lots of items (produced or acquired). Particularly popular is quality control by attributes, which, with respect to control by variables, involves a smaller cost per item, and is less prone to mistakes and inaccuracies by the personnel in charge for the inspection. Besides, although double, multiple and sequential sampling plans are possible and often advantageous under some respects, the single-sample sampling plan is still the most common, due to its easy and clear implementation and management, leaving aside other advantages (for example, it makes possible a straightforward and uniform collection of statistical data about the quality of the items and the production process).

One fundamental reference about sampling plans by attributes is of course the ISO 2859 norm, revised quite recently [1]. The norm provides an extensive set of tables and procedures to design sampling plans suitable for a wide variety of situations (lot-by-lot, isolated lot, skip-lot sampling, etc.). The norm, on the other hand, is not suitable to carry out an optimal design of sampling plans, nor it allows flexibility in the design (the user, for example, has no freedom in choosing producer's and consumer's risks, and has other important constraints).

The user interested in an optimally designed sampling plan with customized characteristics must give up the tables in ISO 2859, and revert to using one of the many software packages available to this purpose (e.g. [2], [3]); or, if it is considered more satisfactory, one can choose a suitable algorithm in the literature (e.g. [4]) and implement it in the preferred software environment (C, Matlab, Microsoft Excel + Visual Basic, etc.). Using a computer is almost mandatory since the design technique is typically an iterative search.

The design procedures in ISO 2859, however, although rigid and sub-optimal, are still very widely used, because of their simplicity (a software package is not needed) and because one can easily understand and illustrate them to a third party, e.g. a quality inspector. Having design procedures with the simplicity and immediacy of the ISO 2859 tables, and *also* the flexibility and optimality achieved by a computer program is of course desirable.

In this paper it is shown that, for the case of single sampling plans by attributes, such procedures actually exists. One can design a nearly optimal sampling plan, without the tables in ISO 2859, using only a non-programmable pocket calculator and a table of the standard normal distribution (also available in some pocket calculators). In other words, the paper presents nearly exact closed-form design formulae, working for a very wide set of specifications.

For the sake of clarity and conciseness, the paper illustrate the design equations for the particular case of isolated lot sampling plans [5], which is ruled by the less treatable mathematic law, the hypergeometric distribution. Similar equations can be derived, with an analogous approach, for a number of cases covered by ISO 2859, and in particular those ruled by the binomial and Poisson distribution (lot-by-lot sampling in a continuous production [6]).

## 2. THE DESIGN PROBLEM

Consider lots of size  $N$  submitted to statistical inspection for acceptance. A single sampling plan consists in inspecting  $n$  randomly selected items from each lot, with the decision rule:

$$\begin{cases} X \leq c \Rightarrow \text{lot accepted} \\ X > c \Rightarrow \text{lot rejected} \end{cases} \quad (1)$$

where  $X$  is the number of nonconforming items (“defectives”) found among the  $n$  inspected items, and  $c$  is an integer threshold (acceptance number AN). The operating characteristic (OC) curve of the sampling plan is the probability of accepting a lot, expressed as a function of the actual number of defectives ( $K$ ) in the lot. Its exact expression is:

$$\beta(K) = P(X \leq c) = F_X(c) = \sum_{k=0}^c p_X(k) \quad (2)$$

where  $F_X(x)$  denotes the cumulative distribution function (cdf) of the random variable (RV)  $X$ , and  $p_X(k)$  denotes the probability mass function (pmf) of the same RV. Since  $X$  is hypergeometric, the pmf has the well-known and rather cumbersome expression:

$$p_X(k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} \quad (3)$$

Consequently, the cdf (2) is not representable in terms of elementary functions. However, for given values of  $N$ ,  $n$ ,  $c$ ,  $K$  the OC (2) is readily evaluated, and a plot of the kind depicted in Fig. 1 is obtained. Basically, the sampling plan acts like a low-pass electrical filter with respect to the variable “number of defectives in the lot” (instead of the variable “number of cycles per second in the signal”).

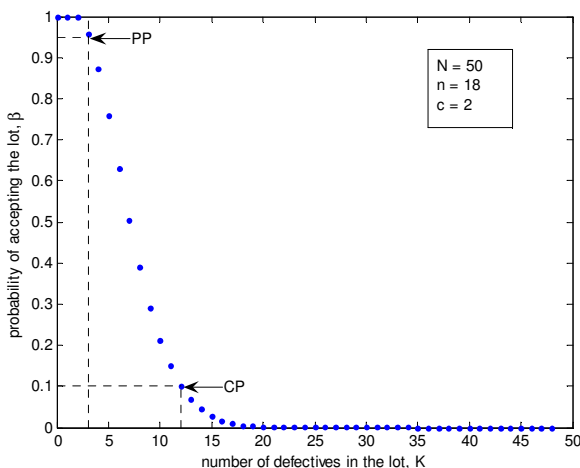


Fig. 1. Example of OC for a single-sample sampling plan of isolated lots. The Producer’s Point (PP) and the Consumer’s Point (CP) are highlighted.

Although one can choose  $n$  and  $c$  according to various different criteria, the most natural principle – sometimes called “two-point design” – is to comply with a specifications mask similar to that used in electrical filter

design. The mask is determined by the Producer’s Point (PP) and the Consumer’s Point (CP), with abscisses called Acceptable Quality Limit (AQL) and Rejectable Quality Level (RQL) (also called Lot Tolerance Percent Defective LTPD, or Limiting Quality LQ in [5]). If the AQL is  $K_0$  and the RQL is  $K_1$ , meeting the design mask means to meet simultaneously the conditions:

$$\begin{cases} \beta(K_0) = \beta_0 = 1 - \alpha_0 \\ \beta(K_1) = \beta_1 \end{cases} \quad (4)$$

where  $\alpha_0$  is the probability of rejecting an AQL lot (the Producer’s risk or PR), and  $\beta_1$  is the probability of accepting a RQL lot (the Consumer’s risk or CR).

Since  $\beta(K)$  is nonlinear in  $n$  and  $c$ , and is not even elementarily expressible, the solution is customarily found by means of numerical iterative methods.

If a design “by hand” (without the aid of a computer software) is desired, it is considered necessary to use specific pre-calculated tables, which is the kind of method described in all the regulatory documents on statistical quality control, ranging from the old MIL-STD105 to the ISO 2859 norm. A flexible two-point design, however, would lead to tables of unfeasible dimensions, and therefore some heavy constraints must be imposed on the design. In the part of ISO 2859 dedicated to isolated lots [5], the main design procedure (method A) allows the user to choose only RQL (among a set of 10 values), and leads to plans with CR within 10% and 13%, and with an arbitrarily prefixed PP (which cannot be chosen by the user; the PR is guaranteed to be within 5% and 0%). Of course, this is a one-point design – based on the CP only – with constrained choice of the CP. The alternative design procedure in [5] (method B) requires the user to choose an “inspection level”, which grossly means to choose  $n$ ; this is also, basically, a one-point design with constraints.

Allowing much flexibility and precision in the “hand-made” design requires a mathematical study of equations (4). In the next section, it is shown that, by introducing proper approximation, it is derived a fully flexible two-point design, suitable for inclusion in a concise written document – much more concise than the tables in ISO 2859 – and with performances comparable to those of computer software design algorithms.

### 3. THE PROPOSED SOLUTION OF THE DESIGN PROBLEM

Gaussian approximation of complicated distributions is a well-known tool in statistics. Approximating the hypergeometric distribution with the Gaussian one is not a new proposal (see, e.g. [7]); however, in a recent paper [8], the possibility of approximating the hypergeometric law with a normal one has been studied with particular care, including the case of very little sampling fraction  $n/N$ . This provides theoretical basis to a solution of (4) using the normal approximation.

The hypergeometric RV with pmf (3) has mean  $\mu = np$

and variance  $\sigma^2 = npq(N-n)/(N-1)$ , being  $p = K/N$  the fraction of defectives, and  $q = 1-p$  the fraction of non-defectives. Therefore, under proper conditions (thoroughly discussed in [8]) it can be approximated by a normal RV  $Y \sim N(\mu, \sigma^2)$ .

In writing down an approximate expression of the OC it must be taken in due account that it involves the cdf, not the pmf. Since a discrete distribution is here approximated by a continuous one, the continuity correction is dutiful. The approximated expression of the OC  $\beta(K)$  is, therefore

$$\begin{aligned} \beta(K) &= F_X(c) \cong F_Y(c+0.5) = \\ &= \Phi\left(\frac{c+0.5-np}{\sqrt{npq(N-n)/(N-1)}}\right) \end{aligned} \quad (5)$$

where  $\Phi(\cdot)$  denotes the standard normal cdf.

By substituting the approximate expression (5) in (4), and applying the biunivocal transformation  $\Phi^{-1}(\cdot)$  to both sides, the design equations become:

$$\begin{cases} \frac{c+0.5-np_0}{\sqrt{p_0q_0n(N-n)/(N-1)}} = z_{\beta_0} \\ \frac{c+0.5-np_1}{\sqrt{p_1q_1n(N-n)/(N-1)}} = z_{\beta_1} \end{cases} \quad (6)$$

where  $z_{\beta_0}$  and  $z_{\beta_1}$  are the standard normal quantiles  $\beta_0$  and  $\beta_1$ , and  $p_0 = K_0/N$ ,  $p_1 = K_1/N$ ,  $q_0 = 1-p_0$ ,  $q_1 = 1-p_1$  are the fractions of defectives and non-defectives corresponding the AQL and RQL.

The simultaneous equations (6) are solvable with respect to sample size  $n$  and the acceptance number  $c$  with elementary algebraic manipulations. The results, although having an elementary expression, deserve some specific comment.

### 3.1. Sample size

As regards the sample size, it is convenient to express the solution in terms of two relationships. The *sample size with item replacement* is

$$n' = \left( \frac{z_{\beta_1} \sqrt{p_1q_1} - z_{\beta_0} \sqrt{p_0q_0}}{p_1 - p_0} \right)^2, \quad (7)$$

and the *sample size without item replacement* (or, simply, the sample size) is

$$n = \frac{Nn'}{N-1+n'} \quad (8)$$

The final sample size  $n$  is expressed in terms of the intermediate quantity  $n'$  because of the special meaning of the latter. This is, indeed, the sample size that would be obtained by omitting the term  $(N-n)/(N-1)$ , or, in other terms, the size that would be obtained by assuming for the

mean  $\mu$  and the variance  $\sigma^2$  the values relevant to the binomial distribution ( $\mu = np$ ,  $\sigma^2 = npq$ ). Therefore, (7) is the solution of the design problem if the number of extracted defectives is binomial, instead of hypergeometric. This happens when sampling with replacement (or, better, re-introduction) of the inspected items, but also in the much more practical and well-known case of lot-by-lot inspection of an entire production (covered in part 1 of ISO 2859 [6]).

Summing up, solving the case of isolated lot sampling leads also to a solution of the case of lot-by-lot sampling for the number of defectives in a production (binomial distribution). This case is not discussed further in the present paper, but will be included in an extended version discussing also the lot-by-lot sampling for the number of defects (Poisson distribution).

Both formulae (7) and (8) yield, in general, a non-integer value, while a final integer answer is needed. A given double-point specification, on the other hand, is in general not exactly met by an integer sample size. The obvious rule to follow is rounding  $n$  by excess if one wants risks not higher than the specified one, and to round by defect if slightly higher risks are tolerated. In section 4 it is shown that, if the specifications can be met by an integer sample size, rounding (8) to the nearest integer yields, in most cases, the exact size.

### 3.2. Acceptance number

By solving (6), two expressions for the AN are obtained, that is:

$$c = np_0 + z_{\beta_0} \sqrt{np_0q_0} - 0.5 \quad (9)$$

and

$$c = np_1 + z_{\beta_1} \sqrt{np_1q_1} - 0.5 \quad (10)$$

The equations are equivalent if the non-integer value of  $n$  given by (8) is used, but yield, in general, different values using a rounded value of  $n$ . This is simply due to the fact that an integer pair  $(n, c)$  is in general not able to meet exactly a two-point specification. In choosing between the two formulae one must consider that, with an integer  $n$  and a non-integer  $c$  given by (9), the approximated OC (5) will pass exactly through the PP; on the contrary, with the value given by (10) the approximated OC will pass exactly through the CP. Therefore, the provided equations for  $c$  can be readily used to design with a one-point specification, which is required when the sample size chosen on the basis of different considerations (e.g. costs). In a two-point design using one or the other equation hardly makes a difference, since the obtained value must be rounded in its turn.

It is useful to note that the term in the equations 0.5 is due to the continuity correction (which instead does not affect the sample size). Omitting the continuity correction would induce in many cases an error of one unit, which would make the design much less satisfactory, especially when the AN is low.

#### 4. PERFORMANCE OF THE DESIGN FORMULAE

In order to show the accuracy of the proposed design formulae the following procedure has been followed.

First, the exact OC for a set of sampling plans (determined by the values of  $N$ ,  $n$  and  $c$ ) has been evaluated using (2)-(3). Then, the exact position of two points in the OC, serving as PP and CP, has been evaluated for each plan. This data are reported in Tab. 1.

Table 1. Data of the sampling plans used to test the design formulae.

plan #	$N$	$n$	$c$	$K_0$ (AQL)	$\beta_0$ (1-PR)	$K_1$ (RQL)	$\beta_1$ (CR)
1	10	2	0	1	0,8	7	0,06667
2	10	2	1	3	0,93333	9	0,2
3	20	4	0	1	0,80003	10	0,04334
4	20	4	2	6	0,9391	17	0,08772
5	50	8	2	6	0,95556	26	0,09961
6	50	15	5	11	0,9463	25	0,1083
7	100	8	2	12	0,947	53	0,0991
8	100	15	5	20	0,9539	52	0,09844
9	200	15	5	39	0,9522	105	0,1007
10	200	35	15	64	0,9548	108	0,1023

Afterwards, the two points  $PP = (K_0, \beta_0)$  and  $CP = (K_1, \beta_1)$  has been used to design the sampling plans, so that the obtained values of  $n$  and  $c$  can be compared with the actual values that meet exactly the specifications.

For the sake of completeness, the results has also been compared with the only other closed-form two-points design known to the authors, which is described in [9]. This design formula, derived by a nonlinear transformation on the hypergeometric distribution, is:

$$n = \left( \frac{z_{\beta_1} - z_{\beta_0}}{2 \arcsin \sqrt{p_1} - 2 \arcsin \sqrt{p_0}} \right)^2 \quad (11)$$

It is interesting to note that the structure of (11) resembles that of (7). No closed formula for  $c$  is provided in [9].

Table 2. Results of sample size designs using (8) and using (11), compared with the exact solution.

plan #	$N$	$n$ , exact	$n$ , (8)	$n$ , (11)
1	10	2	2,1441	3,0619
2	10	2	2,1441	3,0619
3	20	4	4,3889	5,2059
4	20	4	3,9688	5,9788
5	50	8	7,6958	10,9203
6	50	15	15,0265	22,9200
7	100	8	7,5701	9,8856
8	100	15	14,6343	18,9368
9	200	15	14,4331	17,3863
10	200	35	34,7079	43,6399

Tab. 2 reports the results as concerns the sample size. The accuracy of (8) as a design formula is apparent. In all the examined cases, including those with very small  $N$ ,  $n$  or  $c$ , the obtained value is very close to the exact one. In all the cases but #9, by rounding the result of the formula the exact sample size is obtained. In case #9 the result,  $n = 14,43$ , is all the same very near to the exact solution  $n = 15$ , and rounding leads to the very small error of one unit. The performance of (11) is, compared to the results of (8), uniformly and meaningfully worse.

Table 3. Results of the acceptance number designs using (10) with the noninteger  $n$  given by (8), and with the same value rounded to the next integer.

plan #	$N$	$c$ (exact)	$c$ (noninteger $n$ )	$c$ (rounded $n$ )
1	10	0	0,0598	0,0222
2	10	1	1,0843	0,9778
3	20	0	0,0678	0,0175
4	20	2	1,9881	2,0077
5	50	2	1,8486	1,9414
6	50	5	4,99	4,9803
7	100	2	1,8049	1,9358
8	100	5	4,8196	4,9525
9	200	5	4,7368	4,5797
10	200	15	14,8476	14,9768

Tab. 3 reports the results as concerns the acceptance number calculated using (10) (that is, taking the RQL and CR as a reference, like in the ISO 2859-2 [5]). The formula has been calculated using both the noninteger sample size given directly by (8), and using the same value rounded to the next integer. In both cases all the obtained values are so close to the exact solution, that a simple rounding achieves it.

The obtained results show the performance of the design formulae for the more critical case of low values of  $N$ ,  $n$  and  $c$ . It can be easily verified that the design is even more accurate for higher values of these parameters.

#### 5. CONCLUSIONS

A set of simple closed-form equations to design single sampling plans by attributes for isolated lots have been presented. The formulae can be calculated by hand with a non-programmable pocket calculator and a table of the standard normal distribution. They allow one to calculate easily the sample size and acceptance number required to meet accurately a two-point specification mask. Also an accurate single-point design can be readily obtained. The ability of the formulae to derive the right parameters has been shown numerically in a set of cases, chosen among the typically critical ones.

The formulae allows one to avoid the use of complicated iterative algorithms, and are able to improve distinctly the procedures in the ISO 2859 norm, in which special tables are required, and a much less flexible design is allowed. The formulae, therefore, appear clearly suitable for inclusion in a

norm or a guide about sampling plans by attributes.

## REFERENCES

- [1] International Organization for Standardization, *ISO 2859-10:2006 – Sampling procedures for inspection by attributes – Part 10: Introduction to the ISO 2859 series of standards for sampling for inspection by attributes*. ISO, 2006.
- [2] W. A. Levinson, “Acceptance sampling plan designer.” [Online]. Available: <http://www.ct-yankee.com/sampplan/>
- [3] Servicco Corp., “Acceptance sampling software based on the OC curve.” [Online]. Available: [http://www.samplingplans.com/software\\_oc.htm](http://www.samplingplans.com/software_oc.htm)
- [4] T. P. McWilliams, E. M. Saniga, and D. J. Davis, “On the design of single sample acceptance sampling plans,” *Economic Quality Control*, vol. 16, no. 2, pp. 193–198, 2001.
- [5] International Organization for Standardization, *ISO 2859-2:1985 – Sampling procedures for inspection by attributes – Part 2: Sampling plans indexed by limiting quality (LQ) for isolated lot inspection*. ISO, 1985.
- [6] —, *ISO 2859-1:1999 – Sampling procedures for inspection by attributes – Part 1: Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection*. ISO, 1999.
- [7] W. L. Nicholson, “On the normal approximation to the hypergeometric distribution,” *The Annals of Mathematical Statistics*, vol. 27, pp. 471–483, 1956.
- [8] S. Lahiri, A. Chatterjee, and T. Maiti, “Normal approximation to the hypergeometric distribution in nonstandard cases and a sub-Gaussian Berry–Esseen theorem,” *Journal of Statistical Planning and Inference*, vol. 137, no. 11, pp. 3570–3590, Nov. 2007, special Issue: In Celebration of the Centennial of The Birth of Samarendra Nath Roy (1906-1964).
- [9] W. Caspary and G. Joos, *Spatial Data Quality*. CRC Press, 2002, ch. 7, pp. 106–115.