

WELMEC Guide 8.10

Measuring Instruments Directive (2014/32/EU)

Guide to sampling plans for statistical verification of conformity according to Modules F and F1 Version 2024

For information:

This Guide is made available for the Working Group Measuring Instruments
(European Commission expert group E01349) for consideration for future
referencing on the Europa Website.



WELMEC e.V. is a cooperation between the legal metrology authorities of the Member States of the European Union and EFTA. This document is one of a number of Guides published by WELMEC e.V. to provide guidance to manufacturers of measuring instruments and to notified bodies responsible for conformity assessment of their products. The Guides are purely advisory and do not themselves impose any restrictions or additional technical requirements beyond those contained in relevant EU Directives. Alternative approaches may be acceptable, but the guidance provided in this document represents the considered view of WELMEC e.V as to the best practice to be followed.

Published by:
WELMEC Secretariat

E-mail: secretary@welmec.org
Website: www.welmec.org

1. General

Several conformity assessment modules listed in Annex II of the Measuring Instruments Directive 2014/32/EU (MID) allow for statistical verification of conformity, either implicitly (modules A, C, D, D1, E, E1, H, and H1) or explicitly (modules A2, C2, F and F1). All of these modules except F and F1 leave the choice of statistical methods entirely to the manufacturer or notified body, who can find appropriate sampling plans in, e.g., international standards such as the ISO series 2859 “Sampling procedures for inspection by attributes” [1] or 3951 “Sampling procedures for inspection by variables” [2]. The modules F and F1, however, specify under number 5.3 and 6.4, respectively, the following requirements for statistical testing:

The statistical control will be based on attributes. The sampling system shall ensure:

(a) a level of quality corresponding to a probability of acceptance of 95 %, with a non-conformity of less than 1 %;

(b) a limit quality corresponding to a probability of acceptance of 5 %, with a non-conformity of less than 7 %.

At present, WELMEC is not aware of standards containing sampling plans that are adapted to this combination of requirements. Moreover, the wording of the two conditions (a) and (b) has to be interpreted and cast in precise mathematical terms before one can decide whether existing sampling plans are admissible or before one can compute new plans.

This Guide is intended to support notified bodies wishing to evaluate existing or develop new sampling plans for the statistical conformity assessment in modules F and F1; this revised version takes into account recent research [3], [4]. Generally, the Guide adheres to the notations and definitions of ISO 3534 “Statistics – Vocabulary and symbols” [5], [6]. Section 2 recalls basic principles and definitions of statistical acceptance sampling for inspection by attributes, as relevant in the context of MID modules F and F1. Sections 3 and 4 interpret the above MID conditions and outline a general method to evaluate existing or compute new sampling plans that are admissible for modules F and F1. Sections 5 and 6 explain how to generate optimal, single-sampling plans, and provide an example for a simplified scheme.

2. Acceptance sampling for inspection by attributes – basic principles, terms and definitions

2.1 Basic principles of acceptance sampling and sampling plans

1. In modules F and F1, the examinations and tests to verify the conformity of the measuring instruments [...] shall be carried out, **at the choice of the manufacturer**, either by examination and testing of every instrument, or by examination and testing of the measuring instruments on a statistical basis (see reference [7]).

2. For a statistical verification of conformity, the manufacturer shall take all measures necessary so that the manufacturing process ensures the homogeneity of each lot produced, and shall present his measuring instruments for verification in the form of **homogeneous lots**. [7]
3. Each presented lot constitutes a **population** (i.e. totality of items under consideration) of known, **finite size**: $N \in \mathbb{N} = \{1, 2, \dots\}$.
4. Each **item** or **unit** (i.e. measuring instrument) of the lot, if examined and tested, is found to be either conforming or not conforming with the applicable requirements of the MID. The property “conforming/non-conforming” is the basic **attribute** at the basis of statistical control; equivalent notations include “1/0”, “pass/fail”, “true/false”.

Note: This basic, final attribute generally encompasses the conformance with several individual requirements, some of which may be qualitative (e.g., presence of markings and inscriptions) while others may be quantitative (e.g., compliance with maximum permissible errors). For each inspected unit, all these individual attributes are combined by the logical (Boolean) operation AND into the final attribute; in other words, only if a measuring instrument conforms with all applicable requirements, it is found conforming on the whole.

5. The number M of non-conforming units, $0 \leq M \leq N$, determines the **quality level** $p = M/N$ (rate of non-conforming units) of the given lot, termed “non-conformity” by the MID. The possible values of the quality level $p = M/N$ are $0 = 0/N, 1/N, 2/N, \dots, N/N = 1$. The best possible quality corresponds to $M = 0$ non-conforming units, or $p = 0$. The worst possible quality level is reached if all $M = N$ units are non-conforming, or $p = 1$.

Note: Before inspection, the number M of non-conforming units in the lot under consideration and thus its quality level p are unknown.

6. By inspecting all N units of the lot (**100 % inspection**), the number M of non-conforming units and the quality level $p = M/N$ can be determined with certainty. However, 100 % inspection is often neither practical nor economical and sometimes impossible, for instance when testing is destructive.
7. The goal of statistical sampling is to obtain as much information as needed from a subset of n units ($0 < n \leq N$) taken from the lot, the **sample**. In MID modules F and F1, the sample shall be **taken at random** from the lot under consideration. Let k denote the number of non-conforming units found in the sample, $0 \leq k \leq \min(M, n)$. The ratio k/n found from the sample is an **estimate** for the unknown quality level $p = M/N$ of the entire lot.
8. **Acceptance sampling** is defined as a process of sampling after which decisions are made to accept or not to accept a lot based on sample results (see 1.3.17 in [6]). If one fixes in advance the **decision rule** to accept the lot whenever the number of non-conforming units in the sample is $k \leq c$ and to reject the lot whenever $k > c$, then the pair (n, c) with this decision rule defines a **single-sampling plan**. The number c (with $0 \leq c \leq n$) is the **acceptance number**, the number $d = c + 1$ is the **rejection number**.
9. **Double-sampling plans** and, more generally, **multiple-sampling plans** combine several samples to be drawn consecutively from the same lot, with acceptance and rejection numbers (with values larger than $c + 1$) at each stage; a multi-index notation:

$$(n, c, d) = ((n_1, n_2, \dots, n_f), (c_1, c_2, \dots, c_f), (d_1, d_2, \dots, d_f)) \quad (1)$$

can be used. At the final stage f , the rejection number is $d_f = c_f + 1$ in order to arrive at a definite decision to accept or not accept the lot. In the preceding stages $1 \leq j \leq f - 1$, the lot is accepted immediately if the total number of non-conforming units is $k \leq c_j$, the lot is rejected immediately if $k \geq d_j$, and in the intermediate case $c_j < k < d_j$ (with $d_j > c_j + 1$) one proceeds to the next stage by drawing and testing the required number of supplementary items, until the final decision is reached.

10. Statistical sampling can become particularly advantageous for large lots ($N \gg 1$) when only a much smaller sample ($n \ll N$) needs to be inspected. The disadvantage is that the limited sample data only allow to estimate the true quality level of the lot, falling short of certainty as soon as $n < N$. Therefore, conformity decisions for the entire lot based on statistical sampling can only be probabilistic in nature and come with **error rates** or **risks** that should be quantified in advance.

2.2 Acceptance sampling risks, acceptance probability, operating characteristic

11. Acceptance sampling plans are often evaluated at (or near) two predefined quality levels:
- **Limiting quality (LQ)**: insufficient quality level for the consumer, with a low probability of acceptance in a sampling inspection that is intolerable for the manufacturer (after 4.6.13 in [6])
 - **Acceptance quality limit (AQL)**: sufficient quality level for the consumer, with a high probability of acceptance in a sampling inspection that is tolerable for the manufacturer (after 4.6.15 in [6])
12. The probability of acceptance of a lot with quality level LQ, i.e. the risk that the sampling inspection arrives at the wrong decision to accept a lot with insufficient quality, is the **consumer's risk**. Conversely, the probability of wrongly rejecting a lot of sufficient quality at the AQL is the **manufacturer's risk**. (after 4.6.2 and 4.6.4. in [6])
13. $P_{ac} = P_{ac}(p; N, n, c, d)$ denotes the **probability of acceptance** of a given lot with size N and quality level p under the sampling plan (n, c, d) . The **probability of rejection** (i.e. non-acceptance) is $1 - P_{ac}$.
14. The acceptance probability of a single-sampling plan of size n with acceptance number c (and rejection number $d = c + 1$) is given by the **cumulative distribution function**

$$P_{ac}(c; n, p, N) = \sum_{k=0}^c P(k; n, p, N) = P(0; n, p, N) + P(1; n, p, N) + \dots + P(c; n, p, N) \quad (2)$$

of the probability mass function $P(k; n, p, N)$, i.e. the probability to find k non-conforming units in a sample of size n drawn from a lot of size N with quality level p .

Note: This guide follows a frequentist interpretation of probability. An alternative would be the Bayesian (degree of belief) interpretation [8].

15. If each unit of a sample has the same probability p to be non-conforming (e.g., drawn from a continuous process with quality level $p \in [0, 1]$), then the number k of non-conforming units found in a sample of size n is a **random variable**, $0 \leq k \leq n$, with **binomial distribution**:

$$P(k; n, p) = \binom{n}{k} p^k (1 - p)^{n-k}. \quad (3)$$

Here, $\binom{n}{k} = n!/(k!(n-k!))$ is the binomial coefficient, i.e. the number of combinations, or different choices of k elements among n elements. As a **probability mass function**, $P(k; n, p)$ is normalised to unity:

$$\sum_{k=0}^n P(k; n, p) = P(0; n, p) + P(1; n, p) + \dots + P(n; n, p) = 1. \quad (4)$$

16. The acceptance probability plotted as function of the (unknown) quality level p is known as the **operating characteristic (OC)** and permits to visualize the performance of the sampling plan.

Note: For finite-size lots with discrete quality levels $p = M/N$, also the operating characteristic is a set of discrete points. Only for very large lots ($N \rightarrow \infty$) or a series of lots produced in a continuous process, the acceptance probability can be meaningfully evaluated at any value of p , and the operating characteristic becomes a continuous curve.

17. If the operating characteristic is a continuous curve, one can identify a **consumer's risk quality** as the quality level that corresponds to a given consumer's risk, as well as a **manufacturer's risk quality** as the quality level that corresponds to a given manufacturer's risk.
18. Figure 1 shows, as an example, the operating characteristic (blue curve) of the single-sampling plan ($n = 109, c = 3$) for a very large lot size ($N = \infty$) such that the binomial distribution (3) applies. The manufacturer's risk α at a chosen AQL (on the x -axis) can be read from the y -axis as $1 - OC$, where OC is the value of the operating characteristic. The consumer's risk β at the chosen LQ (on the x -axis) can be read from the y -axis as the value of the operating characteristic. This sampling plan features a manufacturer's risk $\alpha \approx 2.4\%$ at an AQL of 1% and a consumer's risk $\beta \approx 4.8\%$ at an LQ of 7% , as indicated in the plot. On such a continuous curve, one can also find the risk quality corresponding to a given risk or acceptance probability. For example, a manufacturer's risk of exactly 10% is reached at the manufacturer's risk quality of approximately 1.6% (marked by a triangle up in the plot), and a consumer's risk of exactly 10% is reached at the consumer's risk quality of approximately 6% (marked by a triangle down).

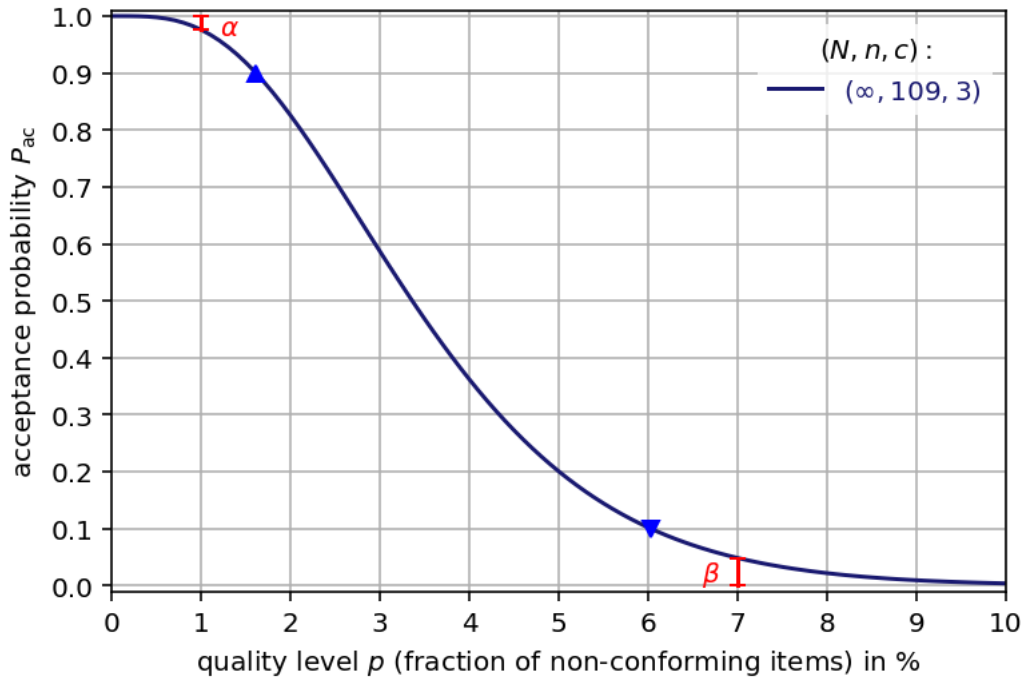


Figure 1: Operating characteristic (blue curve) of the single-sampling plan ($n = 109, c = 3$) calculated with the cumulative distribution function (2) of the binomial distribution (3), applicable for very large lot sizes and arbitrary quality levels $p \in [0, 1]$. $\alpha \approx 2.4\%$ (red bar) indicates the manufacturer's risk at the AQL of 1 %; $\beta \approx 4.8\%$ (red bar) indicates the consumer's risk at the LQ of 7 %. The triangle up marks the manufacturer's risk quality at $p \approx 1.6\%$ corresponding to a given manufacturer's risk of exactly 10 %, and the triangle down marks the consumer's risk quality at $p \approx 6\%$ corresponding to a given consumer's risk of exactly 10 %.

3. Acceptance sampling plans admissible for MID modules F and F1

19. In terms of the general definitions 11 and 12 above, the MID conditions (a) and (b) plausibly require an AQL better (i.e. smaller) than $p_a = 1\%$ and an LQ better (i.e. smaller) than $p_b = 7\%$, with a consumer's risk of (at most) $P_b = 5\%$ and a manufacturer's risk of (at most) $1 - P_a = 1 - 95\% = 5\%$.
20. Since the possible quality levels of the finite-size lots under consideration in modules F and F1 are discrete, the associated acceptance probabilities are discrete. Thus also the operating characteristic of a given sampling plan is a set of discrete points, and there is generally no value $p = M/N$ that "corresponds" to the acceptance probabilities $P_a = 95\%$ and $P_b = 5\%$ specified by the MID. Therefore, the wording of the MID conditions (a) and (b) is ambiguous; they need to be interpreted and cast in more precise mathematical terms.
21. The procedure of acceptance sampling should respect the legitimate interests of manufacturers and consumers alike. Reasonably, the MID conditions (a) and (b) should ensure the acceptance of high-quality lots and the rejection of low-quality lots with high probability. Equivalently, both the manufacturer's and the consumer's risk should be bounded from above. Moreover, MID acceptance sampling should agree with the sound statistical framework of hypothesis testing.

With this rationale in mind, the MID conditions can be stated more precisely as follows:

The statistical control will be based on attributes. The sampling system shall ensure:

*(a') a probability of acceptance of **at least** 95 %, for quality levels of 1% non-conformity **and less**;*

*(b') a probability of acceptance of **at most** 5 %, for a limit quality of 7 % non-conformity **and more**.*

In mathematical terms

$$\begin{aligned} (a') \quad p \leq p_a = 1\% &\Rightarrow P_{ac}(p) \geq P_a = 95\%; \\ (b') \quad p \geq p_b = 7\% &\Rightarrow P_{ac}(p) \leq P_b = 5\%. \end{aligned} \quad (5)$$

The grey shaded areas in Figure 2 below show the regions where either of these conditions is violated; no admissible sample plan shall have OC points in these regions.

22. Unless the lot size N is a multiple of 100, neither $p_a = 1\%$ nor $p_b = 7\%$ by themselves have an operational meaning. Indeed, these quality levels p_a and p_b cannot be realised in the lot because $M_a = p_a N$ and $M_b = p_b N$ are not integer. Let $M_a = \lfloor M_a \rfloor$ be the nearest integer number of non-conforming items below M_a , and $M_b = \lceil M_b \rceil$ the nearest integer number of non-conforming items above M_b . Then

$$\begin{aligned} p_a &= \frac{M_a}{N} \text{ and} \\ p_b &= \frac{M_b}{N} \end{aligned} \quad (6)$$

are operationally meaningful quality levels playing the role of AQL and LQ, respectively, for the lot under consideration.

23. For a given sampling plan (n, c, d) , the **manufacturer's risk** (probability of rejecting a lot at the AQL) is

$$\alpha = 1 - P_{ac}(p_a; N, n, c, d). \quad (7)$$

The **consumer's risk** (probability of accepting a lot at the LQ) is

$$\beta = P_{ac}(p_b; N, n, c, d). \quad (8)$$

24. For product inspection of lots in modules F and F1, the sample is drawn at random without replacement from a lot of size N containing $M = pN$ non-conforming items. The random variable k of non-conforming items in the sample of size n follows a **hypergeometric distribution**:

$$P(k; n, p, N) = \frac{\binom{pN}{k} \binom{N-pN}{n-k}}{\binom{N}{n}}. \quad (9)$$

In modules F and F1, the hypergeometric distribution will generally be appropriate since the lots are finite and tested items are not replaced into the lot before further sampling. Numerical values

of the probability mass function $P(k; n, p, N)$ can be readily computed by computer applications such as MS Excel (HYPERGEOM.DIST), OpenOffice Calc (HYPGEOMDIST) scientific Python (scipy.stats.hypergeom), etc.

Note: For samples that are very small compared to the size of the lot, $n \ll N$, the probability to draw a non-conforming item into the sample stays nearly constant, and the hypergeometric distribution is well approximated by the binomial distribution (3).

25. For a single-sampling plan of size n with acceptance number c , the acceptance probability is given by (2), which amounts to the cumulative distribution function of the hypergeometric distribution, also readily available in standard computer applications.
26. For a double-sampling plan $(n, c, d) = ((n_1, n_2), (c_1, c_2), (d_1, d_2))$, in the first stage a sample size of n_1 is drawn and tested, with an immediate acceptance number of c_1 and immediate rejection number d_1 for the number k_1 of non-conforming units in the first sample. Only if $c_1 < k_1 < d_1$, then a second sample of size n_2 is drawn and tested, with final acceptance number c_2 and final rejection number $d_2 = c_2 + 1$. The acceptance probability then reads

$$P_{ac}(c; n, d, p, N) = \sum_{k_1=0}^{c_1} P(k_1; n_1, p, N) + \sum_{k_1=c_1+1}^{d_1-1} P(k_1; n_1, p, N) \sum_{k_2=0}^{c_2} P(k_2; n_2, p_2, N_2) \quad (10)$$

where $N_2 = N - n_1$ is the size of the remaining lot, which still contains $M_2 = M - k_1$ non-conforming items such that its quality level is $p_2 = M_2/N_2$.

27. Given the acceptance probability P_{ac} as function of the quality level p , one can easily decide whether a sampling plan is admissible under the MID conditions (a') and (b'). Since a well-defined acceptance probability is a non-increasing function of p ($p_1 < p_2 \Rightarrow P_{ac}(p_1) \geq P_{ac}(p_2)$), it is sufficient to compute the manufacturer's and consumer's risk, (7) and (8). If both

$$\alpha \leq 5\% \text{ and } \beta \leq 5\% \quad (11)$$

then the plan is admissible under the MID conditions (a') and (b').

28. Figure 2 shows, as an example, the operating characteristic of three admissible sampling plans for three different lot sizes $N = 128, 512, 2048$.

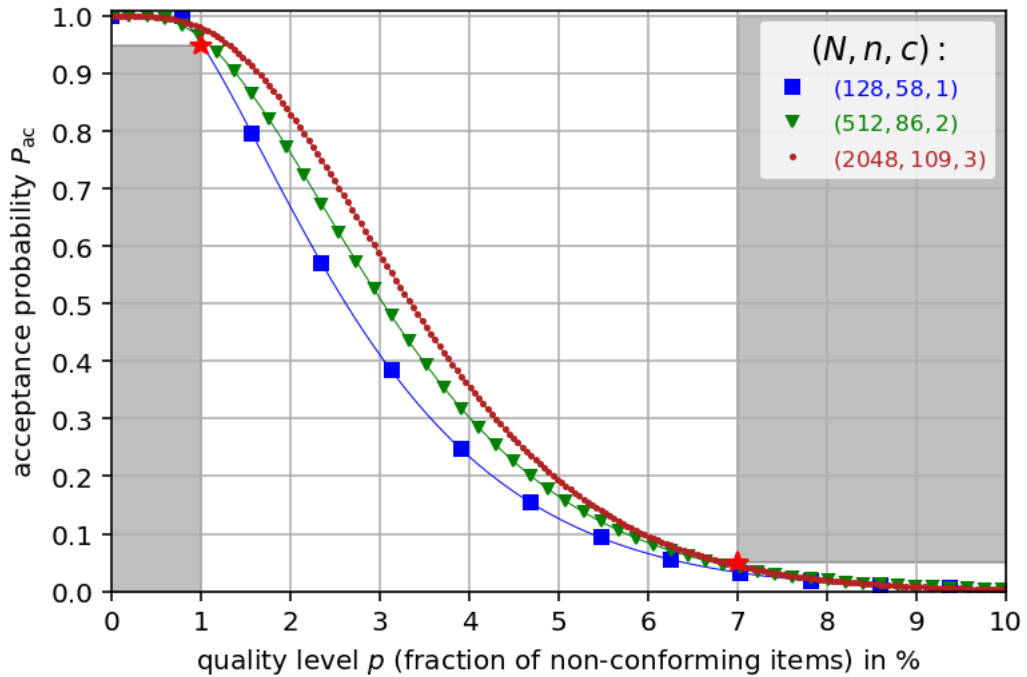


Figure 2: Operating characteristics of single-sampling plans with size n and acceptance number c admissible under the MID conditions (a') and (b') for three different lot sizes N . The acceptance probability is given by the cumulative hypergeometric distribution (Eq. (2) and (9)) for possible quality levels $p = M/N$ and shown as discrete data points; lines are a guide to the eye. Grey shaded areas show the two regions where either the manufacturer's or the consumer's risk would exceed the MID limits a' and b' .

4. MID acceptance sampling as statistical hypothesis testing

29. MID acceptance sampling can be understood as an instance of **statistical hypothesis testing**, a formal framework to evaluate claims or statements on the basis of limited observations. A hypothesis test is formalized by stating two complementary hypotheses, H_0 and H_A . The **null hypothesis** H_0 is typically the proposition that can only be rejected by sufficient evidence to the contrary. In the present context of acceptance inspection in MID modules F and F1, one can hypothesize that the lot is of sufficient quality, i.e. it contains a proportion of non-conforming items better than the AQL, $p < p_a$. The **alternative hypothesis** H_A contains the violations of the null hypothesis that shall be detected, namely a non-conforming proportion worse than the LQ, $p > p_b$.
30. The quality of a hypothesis test can be quantified by its type I and type II error rates. A **type I error** (false positive) occurs when H_0 is rejected based on a bad-quality sample, although H_0 is true. In the present context, the type I error rate α is limited by the manufacturer's risk here. A **type II error** (false negative) occurs if H_0 is not rejected, although H_A is true, i.e. when a lot is accepted due to a good-quality sample although the lot is really of bad quality. The type II error rate β is limited by the consumer's risk here.
31. Given the above definitions, the MID conditions (a') and (b') are equivalent to the hypothesis test

$$H_0: p \leq 1\%, H_A: p \geq 7\%, \text{ with type I and II error rates } \alpha, \beta \leq 5\%. \quad (12)$$

The maximum error rates for a sampling plan are given by the manufacturer's and consumer's risk, respectively, defined above in Eq. (7) and (8). Under this hypothesis-based interpretation of the MID conditions, both risks are bounded symmetrically from above, which appears rather reasonable and ensures that sampling plans are well-behaved functions of their input parameters.

5. Optimal single-sampling plans for Modules F and F1

32. In a single-sampling setting with a given sample size n , there is a maximum acceptance number c_n such that the allowed risks or error rates are not exceeded. Under the hypothesis test (12) or equivalently under conditions (a') and (b'), many sampling plans (n, c, d) are admissible. Given the lot size N , the optimal sampling plan is taken to be the pair (n^*, c^*) with minimal sample size, i.e. $n^* \leq n$ for all admissible plans, and $c^* = c_{n^*}$. For this optimal sampling plan, the acceptance probability as a function of the non-conforming proportion p and lot size N shall be denoted by $P_{ac}^*(p, N) := P_{ac}(c^*; n^*, p, N)$.
33. For large lots with $N > 14286$, the optimal sampling plan turns out to be $(n^*, c^*) = (109, 3)$ with risks bounded by $\alpha = 1 - P_{ac}^*(0.01, \infty) = 2.4311\%$ and $\beta = P_{ac}^*(0.07, \infty) = 4.85\%$, as predicted approximatively by the binomial model in the limit $N \rightarrow \infty$.
34. The optimal sampling plans for all finite lot sizes N are displayed in Figure 3 and can be computed by systematically checking conditions (7) and (8). An R Shiny app is available online that calculates the optimal MID sampling plan for a given lot size (<https://klauenberg.shinyapps.io/MIDSamplingPlans/>, see [9])

6. Simplified single-sampling plan for Modules F and F1

35. The smallest admissible sample size n^* is not an increasing function of the lot size N , but follows a seesaw pattern due to the discretization of quality levels to $p = M/N$ and the pointwise admissibility criteria. In order to arrive at a more compact sampling plan, one may define larger intervals of lot sizes and arrive at simpler, albeit partially sub-optimal sampling plans.

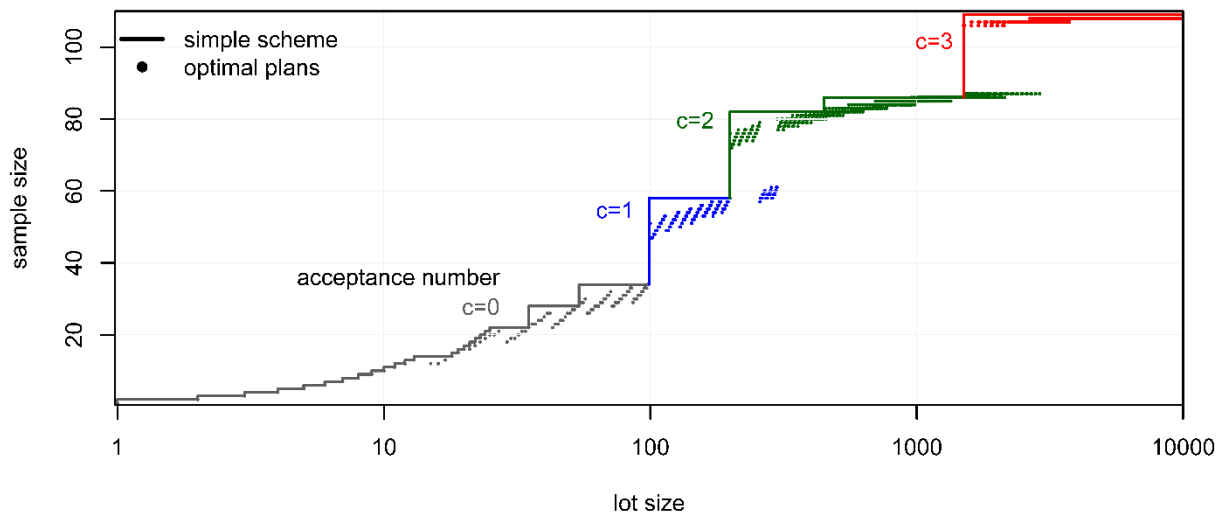
36. One possible proposal for a reasonably simplified, yet nearly optimal sampling plan whose sample sizes do not decrease with increasing lot size is listed in the following table:

Lot size N		Sample		Manufacturers' risk α [%]		Consumers' risk β [%]	
from	to	n	c	from	to	from	to
1	14	N	0	0	0	0	0
15	18	14	0	0	0	0	3.92
19	25	$N-4$	0	0	0	2.00	3.51
26	35	22	0	0	0	0.96	4.37
36	54	28	0	0	0	0.78	4.73
55	99	34	0	0	0	0.93	4.68
100	199	58	1	0	0	1.00	4.84
200	449	82	2	0	2.85	1.97	4.96
450	1499	86	2	1.74	4.98	3.36	4.99
1500	∞	109	3	1.55	2.43	4.07	4.85

Table 1: Simplified, nearly optimal sampling scheme for hypothesis-based MID acceptance sampling. By construction, the manufacturer's and consumer's risk never exceed 5%.

The proposed scheme is compact, nearly optimal and guarantees both consumer's and manufacturer's risks below 5 %.

37. Figure 3 shows the sample sizes as function of lot size on a logarithmic scale (taken from[3]). The simplified single-sampling plan is given as a line; the optimal sampling plans introduced in chapter 5 are shown as dots. Colors are used to distinguish plans by acceptance number.



7. References

- [1] ISO 2859 series "Sampling procedures for inspection by attributes".
- [2] ISO 3951 series "Sampling procedures for inspection by variables".
- [3] K. Klauenberg, C. A. Müller and C. Elster, "Hypothesis-based acceptance sampling for modules F and F1 of the European Measuring Instruments Directive," *Statistics and Public Policy*, vol. 8, no. 1, pp. 9-17, 2021.
- [4] C. A. Müller, "Optimal Acceptance Sampling for Modules F and F1 of the European Measuring Instruments Directive," *Journal of Applied Statistics*, vol. 46, p. 2338–2356, 2019.
- [5] ISO 3534-1:2006, "Statistics – Vocabulary and symbols – Part 1: General statistical terms and terms used in probability", 2006.
- [6] ISO 3534-2:2006, "Statistics – Vocabulary and symbols – Part 2: Applied statistics", 2006.
- [7] "Directive 2014/32/EU of the European Parliament and of the Council of 26 February 2014 on the harmonisation of the laws of the Member States Relating to the Making Available on the Market of Measuring Instruments (recast).", 2014.
- [8] A. Gelman, J. Carlin, H. Stern and D. Rubin, *Bayesian Data Analysis* (1st ed.), Chapman and Hall/CRC., 1995.
- [9] K. Klauenberg, "MID Acceptance Sampling Plans," url: <https://klauenberg.shinyapps.io/MIDSamplingPlans/> (accessed September 2023), 2021.