European Cooperation in
Legal Metrology (WELMEC) e. V.
Bundesallee 100
38116 Braunschweig
Germany

WELMEC Guide 8.10

Směrnice o měřidlech (2014/32/EU)

Příručka k plánům výběru vzorků pro statistické ověřování shody podle modulů F a F1

Verze 2023

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. je spolupráce mezi představiteli legální metrologie členských států Evropské unie a EFTA. Tento dokument je jednou z mnoha příruček vydávaných WELMEC e.V s cílem poskytnout vodítko výrobcům měřidel a oznámeným subjektům odpovědným za posuzování shody výrobků. Příručky mají výhradně poradenský charakter a neukládají žádná restriktivní opatření ani dodatečné technické požadavky oproti těm, které jsou obsaženy v příslušných směrnicích EU. Alternativní přístupy mohou být přijatelné, ale návody uvedené v tomto dokumentu jsou považovány za stanovisko WELMEC e.V. jako nejlepší možná praxe, která by měla být následována.

Vydal:
WELMEC Secretariat

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

1. Obecně

Několik modulů posuzování shody uvedených v příloze II směrnice o měřicích přístrojích 2014/32/EU (MID) umožňuje statistické ověřování shody buď implicitně (moduly A, C, D, D1, E, E1, H a H1) nebo explicitně (moduly A2, C2, F a F1). Všechny tyto moduly s výjimkou modulů F a F1 ponechávají volbu statistických metod zcela na výrobci nebo oznámeném subjektu, který může najít vhodné plány odběru vzorků např. v mezinárodních normách jako ISO řady 2859 "Postupy odběru vzorků pro kontrolu podle atributů" [1] nebo 3951 "Postupy odběru vzorků pro kontrolu podle proměnných" [2]. Moduly F a F1 však specifikují pod čísly 5.3 a 6.4 následující požadavky na statistické testování:

Statistická kontrola bude založena na atributech. Systém odběru vzorků musí zajistit:

(a) úroveň kvality odpovídající pravděpodobnosti přijetí 95 %, přičemž neshoda je menší než 1 %;

(b) hraniční kvalita odpovídající pravděpodobnosti přijetí 5 %, přičemž neshoda je menší než 7 %.

V současné době nejsou organizaci WELMEC známy normy obsahující plány odběru vzorků, které by byly přizpůsobeny této kombinaci požadavků. Kromě toho je třeba vyložit znění obou podmínek *(a)* a *(b)* a vyjádřit je přesnými matematickými termíny, aby bylo možné rozhodnout, zda jsou stávající plány odběru vzorků přípustné, nebo aby bylo možné vypočítat nové plány.

Tato příručka je určena k podpoře oznámených subjektů, které chtějí hodnotit stávající nebo vyvíjet nové plány odběru vzorků pro statistické posuzování shody v modulech F a F1; tato revidovaná verze bere v úvahu nedávný výzkum [3] [4]. Obecně průvodce dodržuje značení a definice ISO 3534 "Statistika – Slovník a symboly" [5] [6]. Oddíl 2 uvádí základní principy a definice statistického přejímacího odběru vzorků pro kontrolu podle atributů, jak jsou relevantní v kontextu modulů MID F a F1. Oddíly 3 a 4 interpretují výše uvedené podmínky MID a načrtávají obecnou metodu pro hodnocení stávajících nebo výpočet nových plánů odběru vzorků, které jsou přípustné pro moduly F a F1. Oddíly 5 a 6 vysvětlují, jak generovat optimální plány jednorázového odběru vzorků a poskytují příklad zjednodušeného schématu.

2. Výběr přejímacích vzorků pro kontrolu podle atributů - základní principy, pojmy a definice

2.1 Základní principy přejímacích odběrů vzorků a plánů odběrů vzorků

1. V modulech F a F1 budou zkoušky a testy k ověření shody měřicích přístrojů prováděny **podle volby výrobce** buď zkoušením a testováním každého přístroje, nebo zkoušením a testováním měřicích přístrojů na statistickém základě (viz odkaz [7]).

2. Pro statistické ověření shody musí výrobce přijmout všechna nezbytná opatření, aby výrobní proces zajistil homogenitu každé vyrobené šarže, a musí předložit své měřicí přístroje k ověření ve formě **homogenních sad.** [7]

3. Každá předložená sada představuje **populaci** (tj. souhrn posuzovaných položek) známé **konečné velikosti**: $N \in \mathbb{N} = \{1, 2, \ldots\}$.

4. Každá **položka** nebo **jednotka** (tj. měřicí zařízení) šarže, pokud je prověřena a vyzkoušena, je shledána buď vyhovující, nebo nevyhovující příslušným požadavkům MID. Vlastnost „shoda/neshoda“ je základním **atributem**, který je základem statistické kontroly; ekvivalentní zápisy zahrnují „1/0“, „vyhovuje/nevyhovuje“, „pravda/nepravda“.

   Poznámka: Tento základní konečný atribut obecně zahrnuje shodu s několika jednotlivými požadavky, z nichž některé mohou být kvalitativní (např. přítomnost značení a nápisů), zatímco jiné mohou být kvantitativní (např. shoda s maximálními přípustnými chybami). Pro každou kontrolovanou jednotku jsou všechny tyto jednotlivé atributy spojeny logickou (Booleovskou) operací “AND” do konečného atributu; jinými slovy, pouze pokud měřicí přístroj splňuje všechny příslušné požadavky, je shledán celkově shodným.

5. Počet M neshodných jednotek $0 \leq M \leq N$ určuje **úroveň kvality** $p = M/N$ (míra neshodných jednotek) dané šarže, označovanou jako „neshoda“ podle MID. Možné hodnoty úrovně kvality $p = M/N$ jsou $0 = 0/N, 1/N, 2/N, \ldots, N/N = 1$. Nejlepší možná kvalita odpovídá M = 0 neshodných jednotek nebo p = 0. Nejhorší možná úroveň kvality je dosažena, pokud jsou všechny jednotky $M = N$ neshodné nebo $p = 1$.

   Poznámka: Před provedením kontroly není znám počet M nevyhovujících jednotek v posuzované šarži, a tedy ani její úroveň kvality p.

6. Pro kontrolu všech N jednotek šarže (**100% kontrola**) lze počet M neshodných jednotek a úroveň kvality $p = M/N$ určit s jistotou. Nicméně 100% kontrola je často nepraktická nebo ekonomicky neúnosná a někdy nemožná, například když je testování destruktivní.

7. Cílem statistického odběru vzorků je získat tolik informací, kolik je potřeba, z podmnožiny n jednotek ($0 < n \leq N$) odebraných z šarže, tedy **vzorku**. V modulech MID F a F1 musí být vzorek **odebrán náhodně** z dané šarže. Označme k počet neshodných jednotek nalezených ve vzorku $0 \leq k \leq min(M, n)$. Poměr k/n zjištěný ze vzorku je odhadem neznámé úrovně kvality p = M/N celé šarže.

8. **Přejímací vzorkování** je definováno jako proces vzorkování, po kterém se na základě výsledků vzorkování rozhoduje o přijetí nebo nepřijetí šarže (viz 1.3.17 v [6]). Pokud je předem stanoveno **rozhodovací pravidlo** přijmout šarži vždy, když je počet nevyhovujících jednotek ve vzorku $k \leq c$, a odmítnout šarži vždy, když je $k > c$, pak dvojice (n, c) s tímto rozhodovacím pravidlem definuje plán jedním výběrem. Číslo c *(kde* $0 \leq c \leq n$*)* je číslo přijetí, číslo $d = c + 1$ je číslo zamítnutí.

9. **Plány dvojího výběru** a obecněji **plány vícenásobného výběru** kombinují několik vzorků, které se mají postupně vybírat z téže šarže, s přejímacími a odmítacími čísly (s hodnotami většími než c + 1) v každé fázi; více indexový zápis:

$$(\mathrm{n}, \mathrm{c}, \mathrm{d}) = ((\mathrm{n}_1, \mathrm{n}_2, \dots, \mathrm{n}_f), (\mathrm{c}_1, \mathrm{c}_2, \dots, \mathrm{c}_f), (\mathrm{d}_1, \mathrm{d}_2, \dots, \mathrm{d}_f)) \qquad (1)$$

lze použít. V konečné fázi f je číslo odmítnutí $df = cf + 1$, aby se dospělo k definitivnímu rozhodnutí o přijetí nebo nepřijetí šarže. V předchozích fázích $1 \leq j \leq f - 1$ je šarže okamžitě přijata, pokud je celkový počet nevyhovujících jednotek $k \leq cj$, šarže je okamžitě odmítnuta, pokud $k \geq dj$, a v mezilehlém případě $cj < k < dj$ *(s* $dj > cj + 1$*)* se postupuje do další fáze losováním a testováním požadovaného počtu doplňkových položek, dokud se nedosáhne konečného rozhodnutí.

10. Statistický odběr vzorků může být obzvláště výhodný pro velké šarže ($N \gg 1$), kdy je nutné kontrolovat pouze mnohem menší vzorek *(*$n \ll N$*)*. Nevýhodou je, že omezené vzorkové údaje umožňují pouze odhadnout skutečnou úroveň kvality šarže, což se od jistoty liší, jakmile $n < N$. Proto rozhodnutí o shodě pro celou šarži založené na statistickém odběru vzorků mohou být pouze pravděpodobnostní povahy a přicházejí s **chybovými mírami** nebo **riziky**, která by měla být předem kvantifikována.

2.2 Rizika výběru vzorků, pravděpodobnost přijetí, operativní charakteristika

11. Plány přejímacího odběru vzorků jsou často hodnoceny na (nebo v blízkosti) dvou předem definovaných úrovní kvality:
    - **Limitní kvalita (LQ):** nedostatečná úroveň kvality pro spotřebitele s nízkou pravděpodobností přijetí při odběrové kontrole, což je pro výrobce nepřijatelné (po 4.6.13 v [6])
    - **Přijímací mezní kvalita (AQL-Acceptance quality limit):** dostatečná úroveň kvality pro spotřebitele s vysokou pravděpodobností přijetí při odběrové kontrole, což je pro výrobce přijatelné (po 4.6.15 v [6])
12. Pravděpodobnost přijetí šarže s úrovní kvality LQ, tj. riziko, že odběrová kontrola dospěje k nesprávnému rozhodnutí přijmout šarži s nedostatečnou kvalitou, je **riziko spotřebitele, odběratele**. Naopak, pravděpodobnost nesprávného odmítnutí šarže dostatečné kvality při AQL je **riziko výrobce, dodavatele** (po 4.6.2 a 4.6.4. v [6]).
13. $\mathrm{P_{ac}} = \mathrm{P_{ac}}\,(\mathrm{p}; \mathrm{N}, \mathrm{n}, \mathrm{c}, \mathrm{d})$ označuje **pravděpodobnost přijetí** dané šarže o N a úrovni kvality p podle plánu odběru vzorků $(\mathrm{n}, \mathrm{c}, \mathrm{d})$. **Pravděpodobnost odmítnutí** (tj. nepřijetí) je $1 - \mathrm{P_{ac}}$.
14. Pravděpodobnost přijetí plánu jednorázového odběru vzorků o velikosti n s přijímacím číslem c (a číslem odmítnutí $\mathrm{d} = \mathrm{c} + 1$) je dána **kumulativní distribuční funkcí**

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

pravděpodobnostní (hmotnostní) funkce $\mathrm{P}(\mathrm{k};\ \mathrm{n},\ \mathrm{p},\ \mathrm{N})$, tj. pravděpodobnosti nalezení *k* neshodných jednotek ve vzorku o velikosti n odebrané z šarže o velikosti N s úrovní kvality p.

Poznámka: Tato příručka sleduje častou interpretaci pravděpodobnosti. Alternativou by byla Bayesovská (stupně víry) interpretace [8].

15. Pokud má každá jednotka vzorku stejnou pravděpodobnost *p*, že bude nevyhovující (např. je vybrána ze spojitého procesu s úrovní kvality p ∈ [0, 1]), pak počet k nevyhovujících jednotek nalezených ve vzorku o velikosti n je **náhodná veličina** 0 ≤ k ≤ n s **binomickým rozdělením:**

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

Kde, $\binom{n}{k} = n!/(k!\,(n-k!))$ je binomický koeficient, tj. počet kombinací nebo různých výběrů k prvků mezi n prvky. Jako **pravděpodobnostní (hmotnostní) funkce**, $P(k;\, n,\, p)$ je normalizována na jednotku:

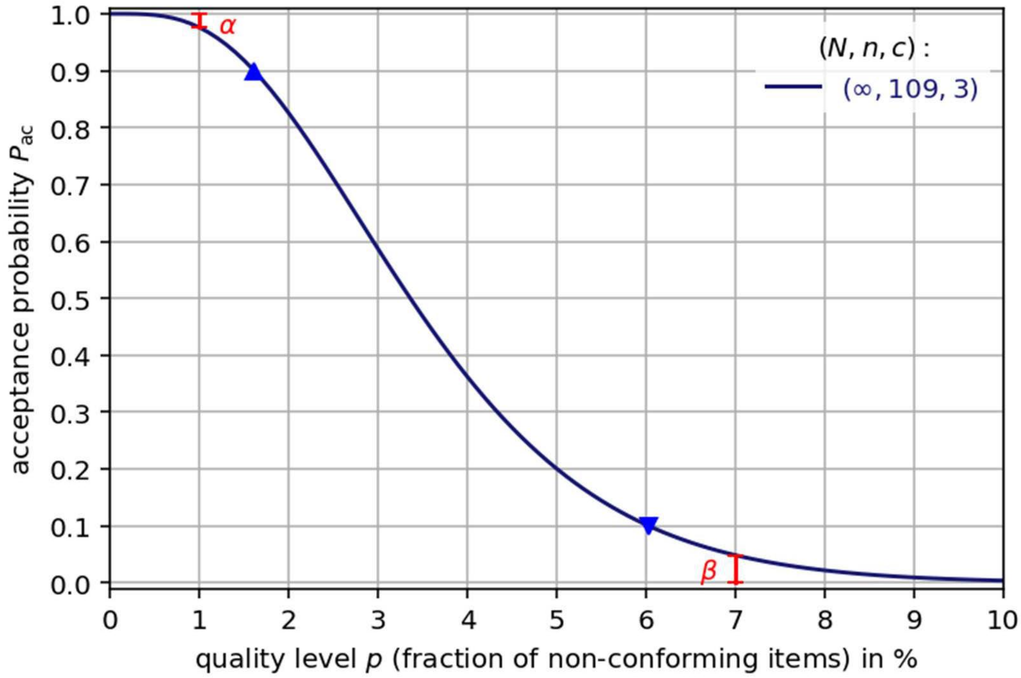
$$\sum_{k=0}^{n} P(k;n,p) = P(0;n,p) + P(1;n,p) + \cdots + P(n;n,p) = 1. \tag{4}$$

16. Pravděpodobnost přijetí zobrazená jako funkce (neznámé) úrovně kvality p je známá jako tzv. **operativní charakteristika (OC - Operating Characteristic)** a umožňuje vizualizovat výkonnost plánu výběru vzorků.

Poznámka: Pro šarže konečné velikosti s diskrétními úrovněmi kvality p = M/N je také operativní charakteristika souborem diskrétních bodů. Pouze pro velmi velké šarže (N → ∞) nebo sérii šarží vyráběných v kontinuálním procesu lze pravděpodobnost přijetí smysluplně vyhodnotit pro jakoukoli hodnotu p a operativní charakteristika se stává spojitou křivkou.

17. Pokud je operativní charakteristika spojitou křivkou, lze identifikovat **kvalitu rizika spotřebitele** jako úroveň kvality, která odpovídá danému riziku spotřebitele, stejně jako **kvalitu rizika výrobce** jako úroveň kvality, která odpovídá danému riziku výrobce.

18. Následující obrázek 1 ukazuje jako příklad operativní charakteristiku (modrá křivka) plánu odběru vzorků edním výběrem (n = 109, c = 3) pro velmi velikou šarži (N = ∞) tak, že platí binomické rozdělení (3). Riziko výrobce *α* při zvolené AQL (na ose x) lze odečíst z osy *y* jako 1 − OC, kde OC je hodnota operativní charakteristiky. Riziko spotřebitele β při zvolené LQ (na ose *x*) lze odečíst z osy *y* jako hodnotu operativní charakteristiky. Tento plán odběru vzorků vykazuje riziko výrobce α ≈ 2.4 % při AQL 1 % a riziko spotřebitele β ≈ 4.8 % při LQ 7 %, jak je uvedeno na grafu. Na takové spojité křivce lze také najít kvalitu rizika odpovídající danému riziku nebo pravděpodobnosti přijetí. Například riziko výrobce přesně 10 % je dosaženo při kvalitě rizika výrobce přibližně 1.6 % (označeno trojúhelníkem nahoru na grafu) a riziko spotřebitele přesně 10 % je dosaženo při kvalitě rizika spotřebitele přibližně 6 % (označeno trojúhelníkem dolů).



Obrázek 1: Operativní charakteristika (modrá křivka) plánu jedním výběrem (n = 109, c = 3) vypočtená pomocí kumulativní distribuční funkce (2) binomického rozdělení (3), použitelná pro velmi velké šarže a libovolné úrovně kvality p ∈ [0, 1]. α ≈ 2,4 % (červený sloupec) označuje riziko výrobce při AQL 1 %; β ≈ 4,8 % (červený sloupec) označuje riziko spotřebitele při LQ 7 %. Trojúhelník nahoru označuje kvalitu rizika výrobce při p ≈ 1,6 % odpovídající danému riziku výrobce přesně 10 % a trojúhelník dolů označuje kvalitu rizika spotřebitele při p ≈ 6 % odpovídající danému riziku spotřebitele přesně 10 %.

3. Přijímací plány odběru vzorků přípustné pro moduly MID F a F1

19. V rámci obecných definic 11 a 12 výše podmínky MID *(a)* a *(b)* zřejmě vyžadují AQL lepší (tj. menší) než $p_a = 1\ \%$ a LQ lepší (tj. menší) než $p_b = 7\ \%$, s rizikem spotřebitele (maximálně) $P_b = 5\ \%$ a rizikem výrobce (maximálně) $1 - P_a = 1 - 95\ \% = 5\ \%$.

20. Protože možné úrovně kvality uvažovaných konečných velikostí šarží v modulech F a F1 jsou diskrétní, jsou diskrétní i související pravděpodobnosti přijetí. Proto je také operativní charakteristika daného plánu vzorkování souborem diskrétních bodů a obecně neexistuje hodnota $p = M/N$, která by „odpovídala“ pravděpodobnostem přijetí $P_a = 95\%$ and $P_b = 5\%$ stanoveným MID. Proto je znění podmínek MID *(a)* a *(b)* nejednoznačné; je třeba je interpretovat a vyjádřit přesnějšími matematickými termíny.

21. Postup výběru vzorků při výběru při přijetí by měl respektovat oprávněné zájmy výrobců i spotřebitelů. Podmínky MID uvedené v písmenech *(a)* a *(b)* by měly s vysokou pravděpodobností zajistit přijetí vysoce kvalitních šarží a odmítnutí nekvalitních šarží. Rovněž riziko výrobce i spotřebitele by mělo být shora omezeno. Kromě toho by výběr vzorků pro přijetí MID měl být v souladu s řádným statistickým rámcem testování hypotéz.

S ohledem na toto zdůvodnění lze podmínky MID upřesnit následovně:

Statistická kontrola bude založena na atributech. Systém výběru vzorků zajistí:

(a') pravděpodobnost přijetí ***nejméně*** *95 % pro úrovně kvality 1 % neshod* ***a méně****;*

(b') pravděpodobnost přijetí ***nejvýše*** *5 % pro mezní úroveň kvality 7 % neshod* ***a více****.*

Matematicky vyjádřeno

$$(a')\ \mathrm{p} \le \mathrm{p_a} = 1\% \Rightarrow \mathrm{P_{ac}(p)} \ge \mathrm{P_a} = 95\%;$$

$$(b')\ \mathrm{p} \ge \mathrm{p_b} = 7\% \Rightarrow \mathrm{P_{ac}(p)} \le \mathrm{P_b} = 5\%. \tag{5}$$

Šedě oblasti na Obrázku 2 níže znázorňují oblasti, kde je některá z těchto podmínek porušena; žádný přípustný vzorkovací plán nesmí mít v těchto oblastech body OC.

22. Pokud velikost šarže N není násobkem 100, nemají $\mathrm{p_a} = 1\%$ ani $\mathrm{p_b} = 7\%$ samy o sobě operativní význam. Tyto úrovně kvality $\mathrm{p_a}$ a $\mathrm{p_b}$ totiž nemohou být v šarži realizovány, protože $\mathrm{M_a} = \mathrm{p_a N}$ a $\mathrm{M_b} = \mathrm{p_b N}$ nejsou celá čísla. Nechť $\mathrm{M_\alpha} = [\mathrm{M_a}]$ je nejbližší celočíselný počet neshodných položek pod $\mathrm{M_a}$, a $\mathrm{M_\beta} = [\mathrm{M_b}]$ nejbližší celočíselný počet neshodných položek nad $\mathrm{M_b}$. Pak

$$p_\alpha = \frac{M_\alpha}{N} \text{ and } \qquad p_\beta = \frac{M_\beta}{N} \tag{6}$$

Jsou statisticky významné úrovně kvality, které hrají roli AQL, resp. LQ pro danou šarži.

23. Pro daný plán vzorkování $(\mathrm{n, c, d})$, je **riziko výrobce, dodavatele** (pravděpodobnost vyřazení šarže při AQL) následující

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

Riziko spotřebitele, odběratele (pravděpodobnost přijetí šarže za LQ) je následující

$$\beta = \mathrm{P_{ac}(p_\beta; N, n, c, d)}. \tag{8}$$

24. Pro kontrolu výrobků v šaržích v modulech F a F1 se vzorek náhodně a bez náhrady vybere ze šarže velikosti N obsahující M = pN nevyhovujících položek. Náhodná veličina k nevyhovujících položek ve vzorku velikosti n se řídí **hypergeometrickým rozdělením**:

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

V modulech F a F1 bude hypergeometrické rozdělení obecně vhodné, protože šarže jsou konečné a testované položky se před dalším výběrem v šarži nevyměňují. Číselné hodnoty pravděpodobnostní (hmotnostní) funkce P(k; n, p, N) lze snadno vypočítat pomocí počítačových aplikací, jako je MS Excel (HYPERGEOM.DIST), OpenOffice Calc (HYPGEOMDIST), vědecký Python (scipy.stats.hypergeom) atd.

Poznámka: Pro vzorky, které jsou ve srovnání s velikostí šarže velmi malé, $n \ll N$, zůstává pravděpodobnost zařazení nevyhovujícího vzorku téměř konstantní a hypergeometrické rozdělení je dobře odhadnutelné binomickým rozdělením (3).

25. Pro plán jedním výběrem o velikosti *n* s akceptačním číslem *c* je pravděpodobnost akceptace dána vztahem (2), který se rovná kumulativní distribuční funkci hypergeometrického rozdělení, jež je rovněž snadno dostupná v běžných počítačových aplikacích.

26. Pro plán dvojím výběrem $(\mathrm{n}, \mathrm{c}, \mathrm{d}) = ((\mathrm{n}_1, \mathrm{n}_2), (\mathrm{c}_1, \mathrm{c}_2), (\mathrm{d}_1, \mathrm{d}_2))$, se v první fázi vybere a otestuje vzorek o velikosti n_1 je s okamžitým přejímacím číslem c_1 a okamžitým vyřazovacím číslem d_1 pro počet k_1 nevyhovujících jednotek v prvním vzorku. Pouze v případě kde $\mathrm{c}_1 < \mathrm{k}_1 < \mathrm{d}_1$, se vybere a otestuje druhý vzorek o velikosti n_2 s konečným přejímacím číslem c_2 a konečným zamítacím číslem $\mathrm{d}_2 = \mathrm{c}_2 + 1$. Pravděpodobnost přijetí pak zní

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

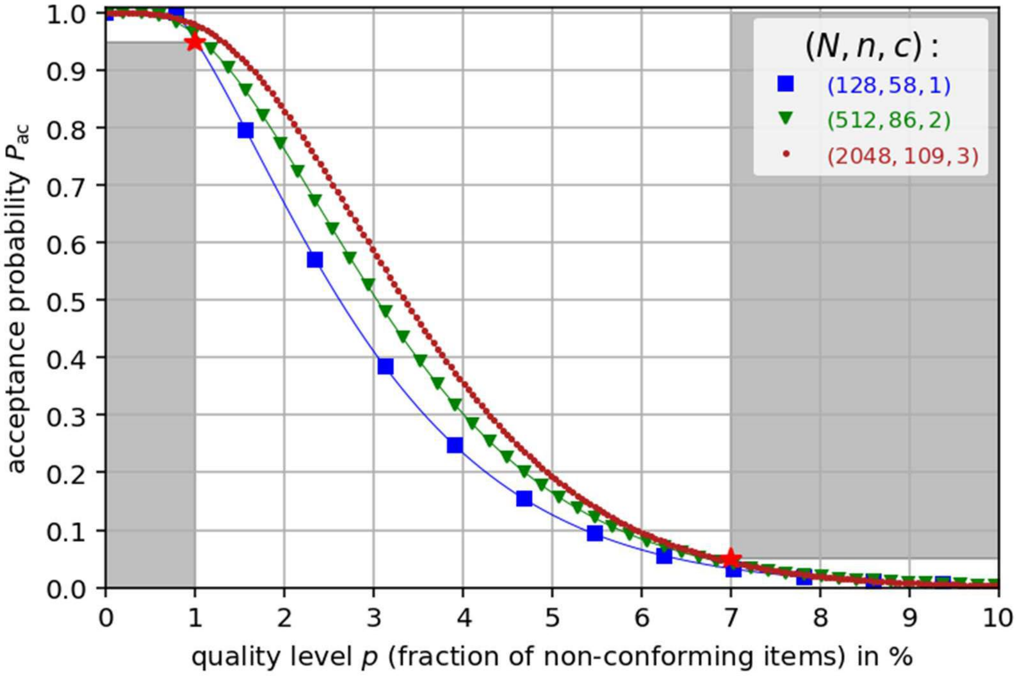
kde $\mathrm{N}_2 = \mathrm{N} - \mathrm{n}_1$ je velikost zbývající šarže, která stále obsahuje $\mathrm{M}_2 = \mathrm{M} - \mathrm{k}_1$ nevyhovujících položek, takže její úroveň kvality je $\mathrm{p}_2 = \mathrm{M}_2/\mathrm{N}_2$.

27. Vzhledem k pravděpodobnosti přijetí P_{ac} jako funkce úrovně kvality p, lze snadno rozhodnout, zda je plán výběru vzorků přípustný za podmínek MID *(a')* a *(b')*. Vzhledem k tomu, že dobře definovaná pravděpodobnost přijetí je nerostoucí funkcí p $(\mathrm{p}_1 < \mathrm{p}_2 \Rightarrow \mathrm{P}_{\mathrm{ac}}(\mathrm{p}_1) \geq \mathrm{P}_{\mathrm{ac}}(\mathrm{p}_2))$, je dostačující vypočítat riziko výrobce a spotřebitele, (7) a (8). Pokud

$$\alpha \leq 5\% \text{ a } \beta \leq 5\% \qquad (11)$$

pak je plán přípustný za podmínek MID *(a')* a *(b')*.

28. Obrázek 2 ukazuje jako příklad operativní charakteristiku tří přípustných vzorkovacích plánů pro tři různé velikosti šarží $\mathrm{N} = 128, 512, 2048$.



Obrázek 2: Operativní charakteristiky plánů jedním výběrem o velikosti n a přejímacím čísle c přípustných za podmínek MID *(a')* a *(b')* pro tři různé velikosti šarží N. Pravděpodobnost výběru je dána kumulativním hypergeometrickým rozdělením (rovnice (2) a (9)) pro možné úrovně kvality $p = M/N$ a zobrazena jako diskrétní datové body; čáry jsou vodítkem pro oko. Šedě oblasti znázorňují dvě oblasti, kde by riziko výrobce nebo spotřebitele překročilo meze MID a' a b'.

4. Přijímací odběr vzorků podle MID jako statistické testování hypotéz

29. MID akceptační výběr lze chápat jako příklad **testování statistických hypotéz**, což je formální rámec pro hodnocení tvrzení nebo prohlášení na základě omezeného počtu pozorování. Test hypotéz je formulován stanovením dvou vzájemně se doplňujících hypotéz, H_0 a H_A. **Nulová hypotéza** H_0 je obvykle výrok, který lze zamítnout pouze na základě dostatečných důkazů o opaku. V tomto kontextu přejímací kontroly v modulech MID F a F1 lze předpokládat, že šarže má dostatečnou kvalitu, tj. obsahuje podíl neshodných položek lepší než AQL, $p < p_a$. **Alternativní hypotéza** H_A obsahuje zamítnutí nulové hypotézy, kdy je třeba zjistit, zda je jmenovitě podíl neshodných položek horší než LQ, $p > p_b$.

30. Kvalitu testu hypotézy lze vyjádřit pomocí poměru chyb typu I a typu II. K **chybě typu I** (falešně pozitivní) dochází, když je H_0 zamítnuta na základě nekvalitního vzorku, ačkoli je H_0 pravdivá. V tomto kontextu je poměr chyb typu I α omezen rizikem výrobce zde. **Chyba typu II** (falešně negativní) nastane, pokud H_0 není zamítnuta, ačkoli H_A je pravdivá, tj. pokud je šarže přijata díky vzorku dobré kvality, ačkoli šarže je ve skutečnosti špatné kvality. Poměr chyb typu II β je zde omezen rizikem spotřebitele.

31. Vzhledem k výše uvedeným definicím jsou podmínky MID (a') a (b') ekvivalentní testu hypotézy

$$H_0:\ p \leq 1\%,\ \ H_A:\ p \geq 7\%, \text{s poměrem chyb typu I a II } \alpha, \beta \leq 5\%. \qquad (12)$$

Maximální poměry chyb pro plán výběru vzorků jsou dány rizikem výrobce, resp. spotřebitele, definovaným výše v rovnicích (7) a (8). Podle této interpretace podmínek MID založené na hypotéze jsou obě rizika shora symetricky omezena, což se zdá být přiměřené a zajišťuje to, že plány výběru vzorků jsou dobře zvládnutelnými funkcemi svých vstupních parametrů.

5. Optimální plány odběru vzorků jedním výběrem pro moduly F a F1

32. V případě plánů jedním výběrem s danou velikostí vzorku *n*, existuje maximální přípustné číslo c_n aby nebyla překročena povolená rizika nebo poměry chyb. Při testu hypotézy (12) nebo ekvivalentně za podmínek (a') a (b'), je přípustných více výběrů (n, c, d). Vzhledem k tomu, že velikosti šarže N, se za optimální plán výběru vzorků považuje dvojice (n^*, c^*) s minimální velikostí vzorku, t.j. $n^* \leq n$ pro všechny přípustné plány, a $c^* = c_{n^*}$. Pro tento optimální plán odběru vzorků se pravděpodobnost přijetí jako funkce nevyhovujícího poměru p a velikosti šarže N označí jako

$$P^*_{\mathrm{ac}}(p, N) \coloneqq P_{\mathrm{ac}}(c^*; n^*, p, N).$$

33. Pro velké šarže s $N > 14286$, se ukazuje, že optimální plán vzorkování je $(n^*, c^*) = (109, 3)$ s rizikem ohraničeným hodnotami $\alpha = 1 - P^*_{ac}(0.01, \infty) = 2.4311\ \%$ a $\beta = P^*_{ac}(0.07, \infty) = 4.85\ \%$, jak se předpokládá přibližně podle binomického modelu v limitě $N \to \infty$.

34. Optimální plány vzorkování pro všechny konečné velikosti šarží N jsou zobrazeny na Obrázku 3 a lze je vypočítat systematickým ověřováním podmínek (7) a (8). An R Shiny app je dostupná online která vypočítá optimální plán vzorkování MID pro danou velikost šarže (https://klauenberg.shinyapps.io/MIDSamplingPlans/, see [9])

6. Zjednodušený plán odběru vzorků jedním výběrem pro moduly F a F1

35. Nejmenší přípustná velikost výběru $n*$ není rostoucí funkcí velikosti šarže N, ale má kolísavý průběh v důsledku rozdělení úrovní kvality na p = M/N a bodových kritérií přijatelnosti. Za účelem dosažení kompaktnějšího plánu výběru lze definovat větší intervaly velikostí šarží a dospět k jednodušším, i když částečně suboptimálním plánům výběrů.

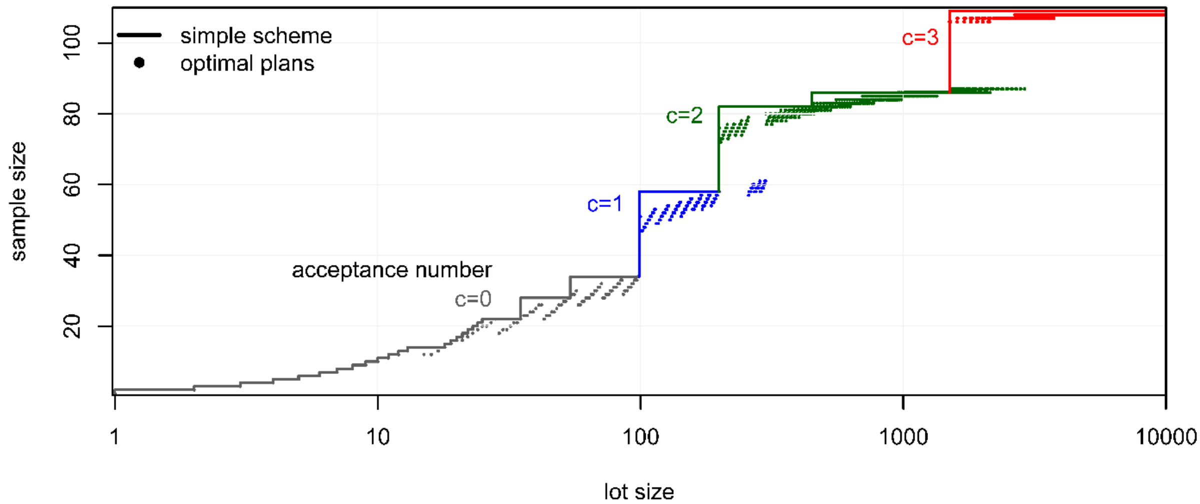
36. Jeden z možných návrhů rozumně zjednodušeného, ale přesto téměř optimálního plánu výběru vzorků, jehož velikost vzorků se nezmenšuje s rostoucí velikostí šarže, je uveden v následující tabulce:

Velikost šarže *N*		Vzorky		Riziko výrobců α [%]		Riziko spotřebitelů β [%]	
od	do	*n*	*c*	od	do	od	do
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

Tabulka 1: Zjednodušené, téměř optimální výběrové schéma pro výběr vzorků na základě hypotézy MID. Podle koncepce riziko výrobce a spotřebitele nikdy nepřekročí 5 %.

Navrhované schéma je kompaktní, téměř optimální a zaručuje riziko spotřebitele i výrobce nižší než 5 %.

37. Obrázek 3 ukazuje velikosti vzorků v závislosti na velikosti šarže v logaritmickém měřítku (převzato z [3]). Zjednodušený plán s jedním výběrem je uveden jako přímka; optimální plány vzorkování představené v kapitole 5 jsou znázorněny jako tečky. Barvy jsou použity k rozlišení plánů podle přejímacího čísla.



7. Reference

[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.