

Výběr vzorků – rychlý exkurz

Workshop ČIIA Hradec Králové

17. 4. 2024

Ministerstvo financí
České republiky



Výběry vzorků

○ **Statistické metody**

- **Umožňují vybrat reprezentativní vzorek (na základě statistických metod)**
- **Umožňují extrapolovat zjištěné chyby na celý soubor (populaci)**
- **=> umožňují určit, jak velké (významné) nesprávnosti soubor obsahuje**
- **Vyžadují dostatečně velkou (početnou) populaci**

○ **Nestatistické metody**

- **Nelze zajistit reprezentativnost vzorku (velikost se stanovuje konsensuálně)**
- **Nelze říci, s jakou pravděpodobností je zjištěná chyba spolehlivá**
- **Používají se na menší populace (kde nelze použít statistické metody)**



Výběr z metod

- **Statistický výběr vzorků**
 - **Pravděpodobnost úměrná velikosti**
 - Nejčastěji Výběr vzorků podle peněžních jednotek (Monetary Unit Sampling – MUS)
- **Nestatistický výběr vzorků**
 - **Náhodný výběr**
 - **Pravděpodobnost úměrná velikosti (Probability Proportional to Size – PPS)**
- **V čem se liší?**
 - **Ve stanovení velikosti vzorku a možnostech vyhodnocení**
- **V čem se shodují?**
 - **V postupu při výběru vzorků z populace**



Velikost vzorku u nestatistické metody (PPS)

- **Stanovena konsensuálně**
 - tedy ze zkušenosti nebo závazným předpisem apod.
- **Pro ESIF 21+ je v obecném nařízení, článek 79 bod 2**
 - Pokud základní soubor zahrnuje méně než 300 jednotek vzorku, může auditní orgán ... použít metodu nestatistického výběru vzorků. ... Metoda nestatistického výběru vzorků musí zahrnovat minimálně 10 % náhodně vybraných jednotek vzorku v základním souboru za účetní rok
 - Je stanovené minimum, v odůvodněných případech (např. vyšší chybovost v minulosti) může být vzorek větší



Nestatistická metoda – jak na to

- **Podmínka nutná, nikoli však dostačující**
 - **Populace (soubor), ze které je vzorek vybírán, by měla být homogenní**
- **Stanovení populace (souboru)**
 - **Období (rok, jak je třeba u ESIF)**
 - **Stejně aktivity**
- **Jak se dobrat homogenní populace**
 - **Mám stejné aktivity – nic už neřeším**
 - **Nemám stejné aktivity – provedu stratifikaci = vytvořím dílčí homogenní populace**



Stratifikace

- **Nelze vybrat hodnověrný vzorek z koše, ve kterém mám jablka, hrušky a ještě švestky**
 - **Co tedy s hromadou, v níž jsou projekty stavební, projekty na nákupy IT techniky, projekty na školení apod.?**
- **Hromadu je potřeba roztržít**
 - **Pod-hromada A) stavební projekty**
 - **Pod-hromada B) nákupy IT techniky**
 - **Pod-hromada C) školení**
- **Výhoda stratifikace**
 - **Vzorek se rozdělí mezi pod-hromady úměrně jejich velikosti (významnosti)**



Stratifikace prakticky

- **Hromada obsahuje 130 projektů v celkové hodnotě 800 mil. Kč, z toho**
 - 20 stavebních za 400 mil. Kč
 - 40 IT nákupů za 300 mil. Kč
 - 70 školení za 100 mil. Kč
- **Velikost vzorku = 10 % = 13 projektů**
 - Stavební => $(400/800)*13=6,5$ zaokrouhлено 7
 - IT => $(300/800)*13=4,875$ zaokrouhлено 5
 - Školení => $(100/800)*13=1,625$ zaokrouhлено 2



Význam stratifikace

- **Vybírám 10 % z každé hromady (všechny projekty jsou stejně velké)**

- | | | |
|--------------------------------|-------------|------------|
| • 20 stavebních za 400 mil. Kč | 2 projekty | 40 mil. Kč |
| • 40 IT nákupů za 300 mil. Kč | 4 projekty | 30 mil. Kč |
| • 70 školení za 100 mil. Kč | 7 projektů | 10 mil. Kč |
| • Sumárně | 13 projektů | 80 mil. Kč |

- **Výběr ze stratifikované populace**

- | | | |
|--------------------------------|-------------|-------------------------|
| • 20 stavebních za 400 mil. Kč | 7 projektů | 140 mil. Kč |
| • 40 IT nákupů za 300 mil. Kč | 5 projektů | 37.5 mil. Kč |
| • 70 školení za 100 mil. Kč | 2 projekty | 2,9 mil. Kč |
| • Sumárně | 14 projektů | 180,4 mil. Kč (22,55 %) |



Principy výběru

- **Náhodný výběr**
 - Náhodné seřazení prvků populace
 - Náhodný počátek výběru
- **Výběrový interval**
 - Podíl celkové hodnoty populace a počtu vybíraných prvků
 - Slouží k vyhledání těch prvků, u kterých dojde k průniku kumulativní hodnoty náhodně seřazených prvků s 1 až n násobkem výběrového intervalu, kde n je počet vybíraných prvků => vzorek
- **Horní vrstva**
 - Prvky, které jsou větší než výběrový interval, automaticky spadají do vzorku
 - Sníží se počet prvků vybraných náhodně a zmenší se výběrový interval
 - Ověřuje se opakovaně, zda neexistuje prvek větší než nový výběrový interval



Schéma náhodného výběru

Výchozí parametry:

- 48 položek v celkové hodnotě 820 mil. Kč
- Velikost vzorku = 10 %
- Počet vybraných položek = 5
- Výběrový interval -> $820 : 5 = 164$
- ! Schéma platí pouze v případě, že všechny položky jsou menší než 164, tzn. neexistuje horní vrstva !

Identifikátory	Kumulativní hodnoty	Výběrové intervaly	Vybrané položky
položky 1 – 9 (9)	0	Náhodný začátek - třeba od 53	položka 4
	160	←	
položky 10 – 16 (7)	161	$53 + 164 = 217$	položka 12
	302	←	
položky 17 – 27 (11)	303	$217 + 164 = 381$	položka 17
	492	←	
položky 28 – 32 (5)	493	$381 + 164 = 545$	položka 30
	597	←	
položky 33 – 48 (16)	598	$545 + 164 = 709$	položka 44
	820	←	

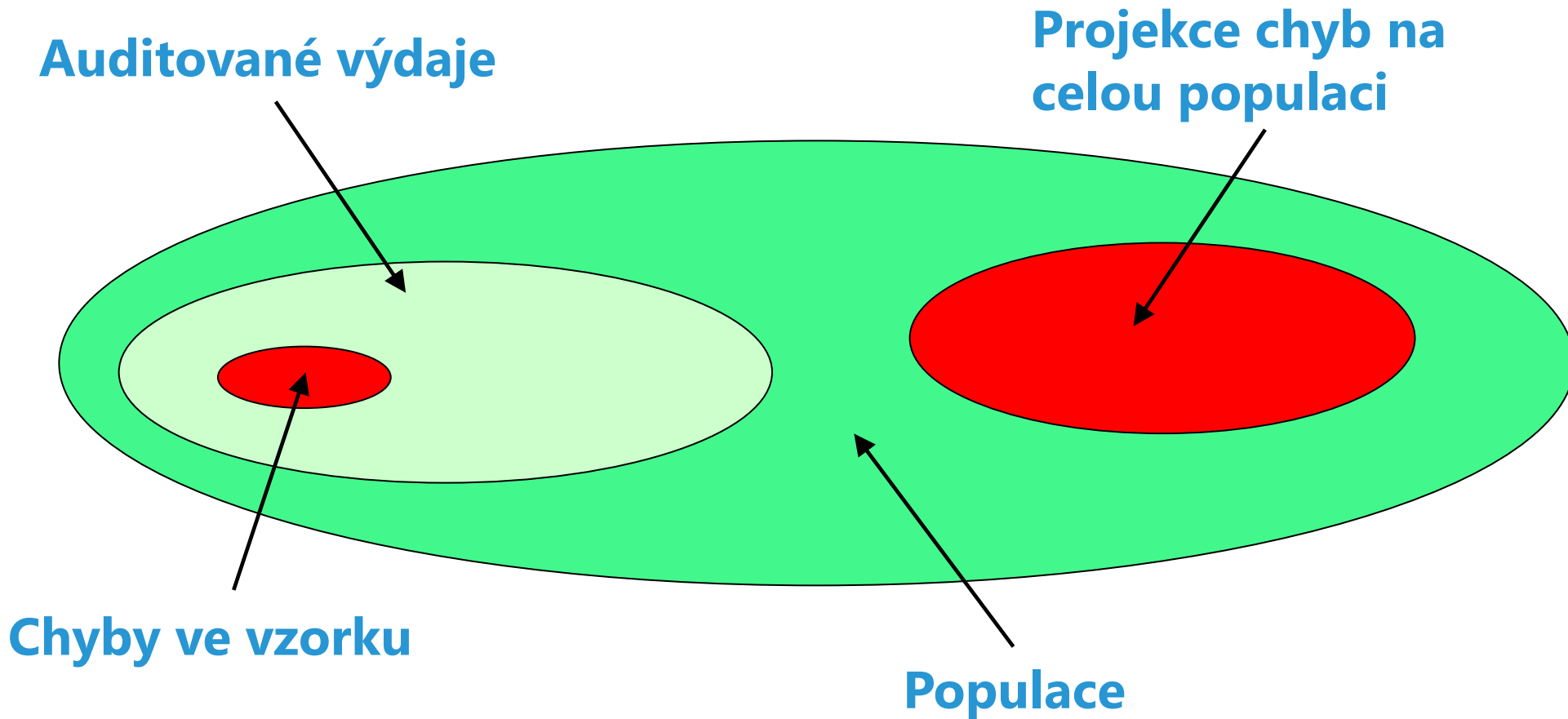


Vyhodnocení výsledků

- **Jakou výši chyb lze tolerovat**
 - Kdo nic nedělá, nic nezkaží => chyba se prostě může stát
 - Obecně platí, že do 2 % je chyba tolerovatelná (ESI fondy, NKÚ ...)
- **Jak se chyba zjistí**
 - U vzorku nelze použít prostou chybovost (suma chyb/hodnota populace)
 - Chyby v horní vrstvě se vezmou v plné výši
 - Chyby v dolní vrstvě se extrapolují
- **Výpočet chybovosti**
 - $\sum \text{chyb HV} + (\sum \text{prosté chybovosti prvků dolní vrstvy} \times \text{výběrový interval})$
= zjištěná chyba
 - Chybovost = zjištěná chyba / hodnota populace



Extrapolace (projekce) chyb



A teď si to ukážeme prakticky

- **Populace a co s ní**
 - 33 položek v hodnotě 11 026 672 360,22 Kč
 - Ověřit existenci horní vrstvy
 - Výběr z dolní vrstvy
- **Pojďme na to**
 - ❖ [výběr_vzorku_příklad.xlsx](#)



Děkuji za pozornost.😊

Dušan Víték

**Ministerstvo financí
České republiky**

