

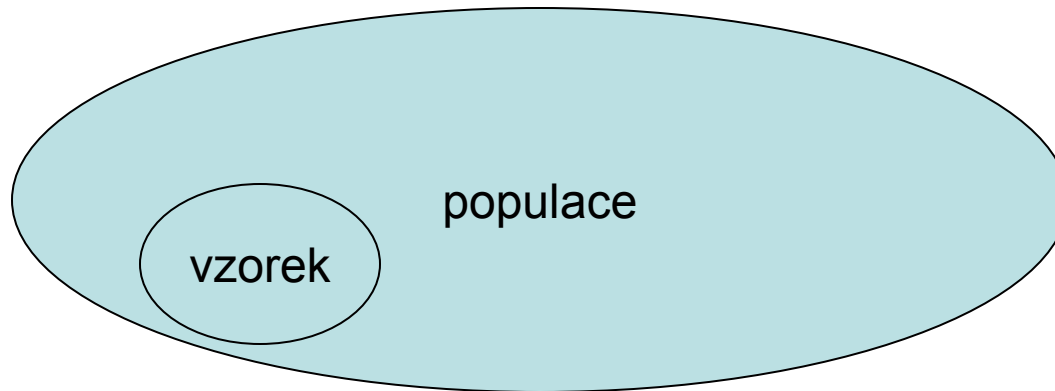
Výběr vzorku

nejen pro SPP118

by Miroslav Suchanec with many thanks to et al.

Co je vzorek?

....zkoumaná část populace

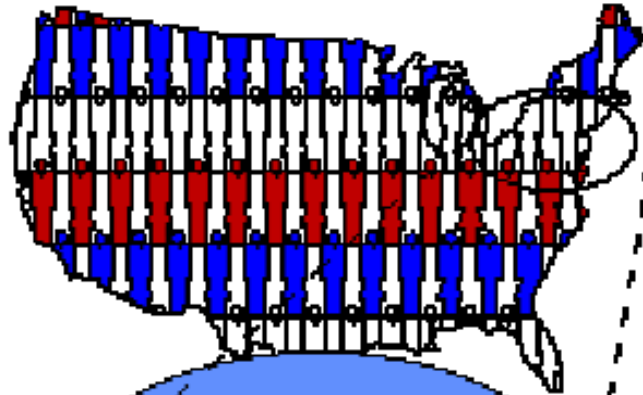


Proč vzorek?

....protože někdy je populace časově, prostorově, finančně nedostupná 😊

Klíčové pojmy

Na koho chceme zobecňovat?



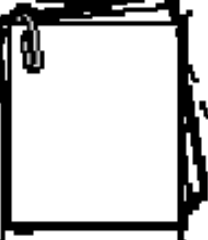
Teoretická populace

Jaká populace je nám přístupná?



Přístupná populace

Jakým způsobem k ní máme přístup?



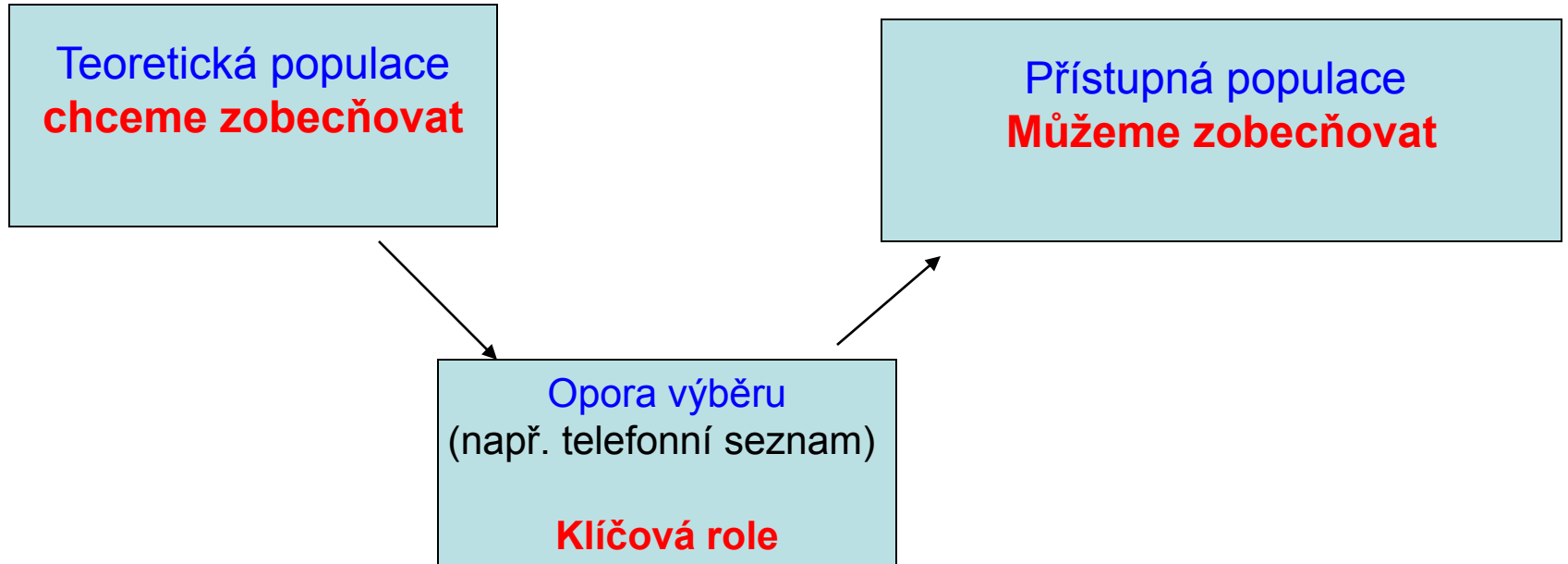
Vzorkovací
rámec
(opora výběru)

Koho studujeme?



Vzorek

Klíčová myšlenka



Obecné typy výběrů vzorku

Pravděpodobnostní

Náhodný

- 1. prostý
- 2. stratifikovaný
- 3. systematický
- 4. klastrový / shlukový
- 5. kombinovaný/
víceúrovňový

Nepravděpodobnostní (účelové neboli teoretické)

- Kvótní
- Extrémní / odchylové případy
 - intenzivní případy
- Maximální rozptyl (heterogenní)
 - Homogenní
 - Typický případ
 - Účelový stratifikovaný
 - Kritický případ
- Sněhová koule / řetězový
 - Kriteriální
 - Účelový náhodný
- Politicky důležité případy
 - Příležitostní
 - Nahodilý
 - na teorii založený

Pravděpodobnostní nebo nepravděpodobnostní?

Pravděpodobnostní

Obecně když...

...chci zobecňovat/usuzovat
ze vzorku na/o populaci

...chci používat statistickou analýzu
(kvantitativní (objektivistický) výzkum)

...mám velkou populaci

Nepravděpodobnostní (účelové neboli teoretické)

...mě zajímají pouze zkoumané případy

...nechci zobecňovat na populaci

...populace není známá / nedostupná

...když mě na základě výzkumného cíle
zajímají jen určití vybraní lidé

...když používám kvalitativní
(interpretativní / subjektivistické)
výzkumné metody

Pravděpodobností výběr

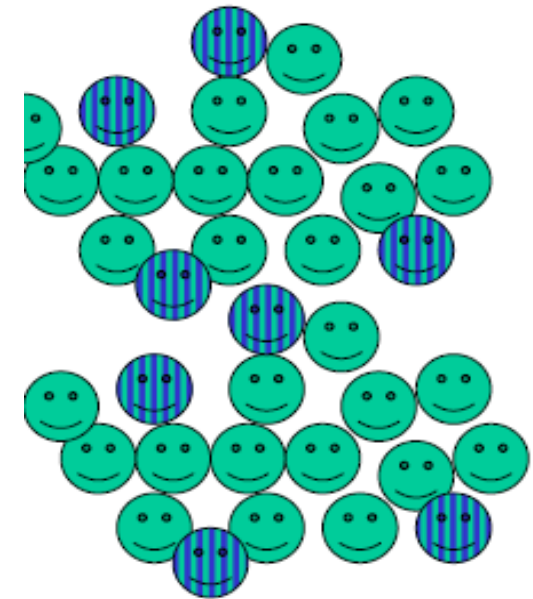
využívá matematickou teorii pravděpodobnosti

= každý člen populace má nenulovou a známou šanci dostat se do vzorku

Klíčová úloha opory výběru = není-li člen populace na seznamu, má nulovou pravděpodobnost být vybrán

1. Prostý náhodný

	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
00	22808	04391	45529	53968	57136	98228	85485	13801	68194	56382
01	49305	36965	44849	64987	59501	35141	50159	57369	76913	75739
02	81934	19920	73316	69243	69605	17022	53264	83417	55193	92929
03	10840	13508	48120	22467	54505	70536	91206	81038	22418	34800
04	99555	73289	59605	37105	24621	44100	72832	12268	97089	68112
05	32677	45709	62337	35132	45128	96761	08745	53388	98353	46724
06	09401	75407	27704	11569	52842	83543	44750	03177	50511	15301
07	73424	31711	65519	74869	56744	40864	75315	89866	96563	75142
08	37075	81378	59472	71858	86903	66860	03757	32723	54273	45477
09	02060	37158	55244	44812	45369	78939	08048	28036	40946	03898
10	94719	43565	40028	79866	43137	28063	52513	66405	71511	66135
11	70234	48272	59621	88778	16536	36505	41724	24776	63971	01685
12	07972	71752	92745	86465	01845	27416	50519	48458	68460	63113
13	58521	64882	26993	48104	61307	73933	17214	44827	88306	78177
14	32580	45202	21148	09684	39411	04892	02055	75276	51831	85686
15	88796	30829	35009	22695	23694	11220	71006	26720	39476	60538
16	31525	82746	78935	82980	61236	28940	96341	13790	66247	33839
17	02747	35989	70387	89571	34570	17002	79223	96817	31681	15207
18	46651	28987	20625	61347	63981	41085	67412	29053	00724	14841
19	43598	14436	33521	55637	39789	26560	66404	71802	18763	80560
20	30596	92319	11474	64546	60030	73795	60809	24016	29166	36059
21	56198	64370	85771	62633	78240	05766	32419	35769	14057	80674
22	68266	67544	06644	84956	18431	04015	89049	15098	12018	89338
23	31107	28597	65102	75599	17496	87590	68848	33021	69855	54015
24	37555	05069	38680	87274	55152	21792	77219	48732	03377	01160
25	90463	27249	43845	94391	12145	36882	48906	52336	00780	74407
26	99189	88731	93531	52638	54989	04237	32978	59902	05463	09245
27	37631	74016	89072	59598	55356	27346	80856	80875	52850	36548
28	73829	21651	50141	76142	72303	06694	61697	76662	23745	96282
29	15634	89428	47090	12094	42134	62381	87236	90118	53463	46969
30	00571	45172	78532	63863	98597	15742	41967	11821	91389	07476
31	83374	10184	56384	27050	77700	13875	96607	76479	80535	17454
32	78666	85645	13181	80700	08289	62956	64439	39150	95690	18555
33	47890	88197	21368	65254	35917	54035	83028	84636	38186	50581
34	56238	13559	79344	83198	94642	35165	40188	21456	67024	62771
35	36369	32234	38129	59963	99237	72648	66504	99065	61161	16186
36	42934	34578	28968	74028	42164	56647	76806	61023	33099	48293
37	09010	15226	43474	30174	26727	39317	48508	55438	85336	40762
38	83897	90073	72941	85673	85569	24183	08247	15946	02957	68504
39	82206	01230	93252	89045	25141	91943	75531	87420	99012	80751
40	14175	32992	49046	41272	94040	44929	98531	27712	05106	35242
41	58968	88367	70927	74765	18635	85122	27722	95388	61523	91745
42	62601	04595	76926	11007	67631	64641	07994	04639	93314	83126
43	97030	71165	47032	85021	65554	66774	21560	04121	57297	85415
44	89074	31587	21360	41673	71192	85795	82757	52928	62586	02179
45	07806	81312	81215	99858	26762	28993	74951	64680	50934	32011
46	91540	86466	13229	76624	44092	96604	08590	89705	03424	48033
47	99279	27334	33804	77988	33592	90708	56780	70097	39907	51006
48	63224	05074	83941	25034	43516	22840	35230	66048	80754	46302
49	98361	97513	27529	66419	35328	19738	82366	38573	50967	72754



Každý člen / kombinace členů populace má STEJNOU šanci být vybrán

Např.

jména losovaná z klobouku kouzelníka Pokustóna

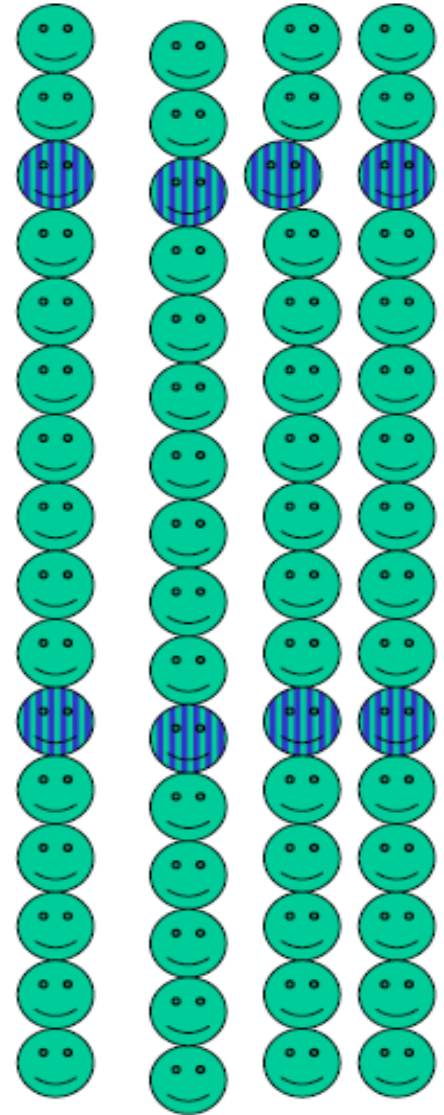
nebo

jména ze seznamu vybrána pomocí tabulky/generátoru náhodných čísel /kostky

3. Systematický náhodný (1)

- Každý člen populace má stejnou šanci být vybrán ale kombinace členů mají šance různé
- Velikost populace= N , požadovaná velikost vzorku= n , potom velikost kroku/intervalu $k=N/n$
- Postup: náhodně vybrat číslo j mezi 1 a k , pak vybrat toto číslo j a dále každý k -tý člen $j+k, j+2k, j+3k$ atd.

Př. Na obrázku: $N=64, n=8, k=64/8=8, j=3$

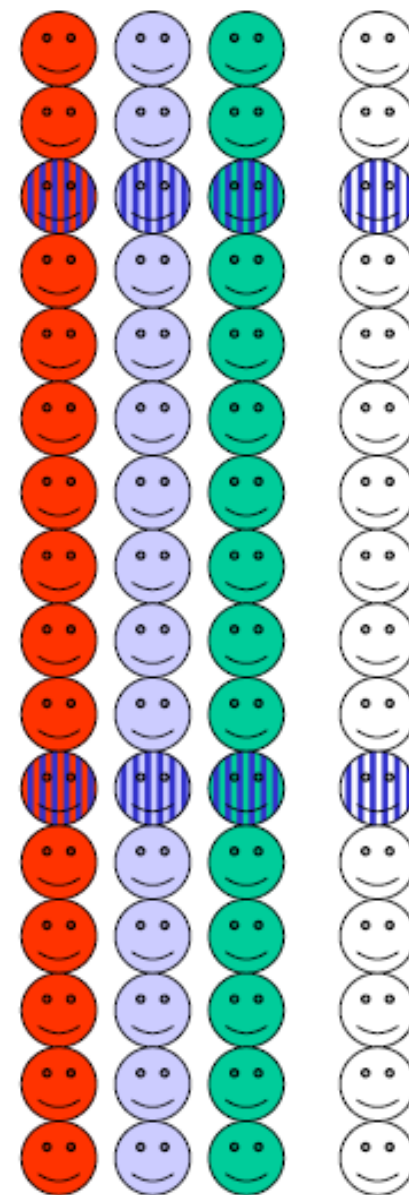


3. Systematický náhodný (2)

= vykazuje stejnou chybu
jako prostý náhodný
výběr, pokud jsou
členové seznamu
náhodně seřazeni

+ poskytuje přínos
implicitní stratifikace,
pokud jsou členové
seznamu seskupeni

(př. obrázek)



3. Systematický náhodný (3)

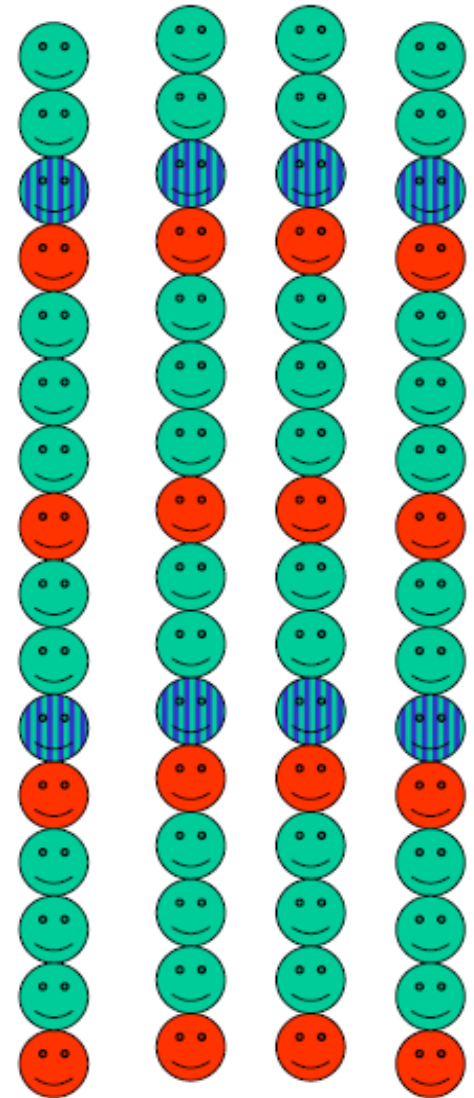
ALE , riziko velké výběrové chyby pokud periodicitu/uspořádání na seznamu kopíruje interval k

Př. Obrázek

Každý 4-tý člen je červený a červený nikdy nebude vybrán, protože $j=3$ a $k=8$.

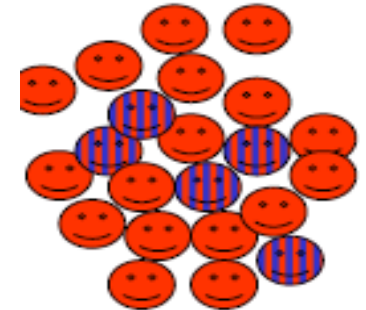
Jak by situace vypadala kdyby $j=4$?

Vybrání jen červení!



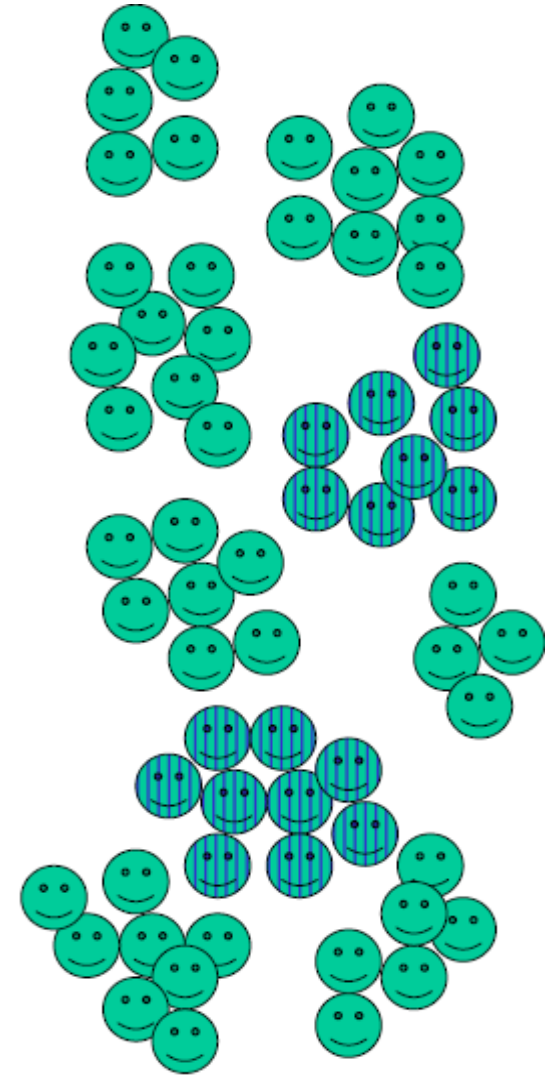
2. Stratifikovaný náhodný

- Rozdělíme populaci do skupin, které se liší pro výzkumný cíl (závisle proměnná) podstatným způsobem (např. podle pohlaví, etnika, věku)
- Dělicí proměnná=stratifikační p.
- Poté vybereme náhodný vzorek z každé skupiny
- *Př. Jaká je spokojenost se stavem v armádě? (Souvisí s hodností člena, dělíme na skupiny velitelů, podplukovníků, pěšáků, z každé vybereme náhodně vzorek)*
- *Předpokládejme, že seznam je seřazen podle rot, proč ne systematický výběr? (vybrání jen rotmistři – zkresluje spokojenost)*
- + menší chyba než prostý náhodný, pokud se skupiny liší
- Kompromis mezi zvýšenými náklady na stratifikaci a menší velikostí vzorku potřebnou ke stejné velké chybě
- Přesamplování malých skupin zlepšuje meziskupinové srovnání



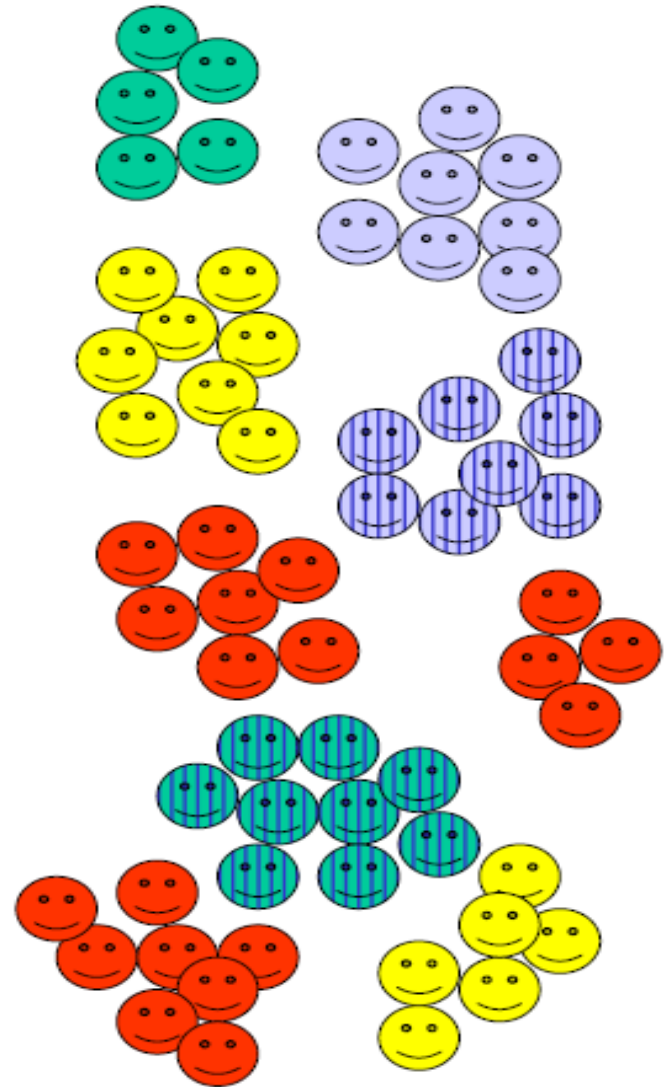
4. Klastrový / shlukový náhodný výběr (1)

- Populace je rozdělená do srovnatelných skupin, obvykle geografických nebo organizačních (školy, města)
- Některé CELÉ skupiny jsou náhodně vybrány do vzorku
- Vykazuje větší chybu než prostý náhodný
- Relativně levný, pouze některé oblasti/organizace, umožňuje velký vzorek
- Chyba je menší čím více jsou klastry vzájemně podobné (obrázek)



4. Klastrový / shlukový náhodný výběr (2)

- ALE chyba je velká, pokud se klastry liší (viz obrázek) – klastrování nevhodné!
- Zcela opomeneme některé členy populace



5a. Kombinovaný (stratifikovaný-klastrový)

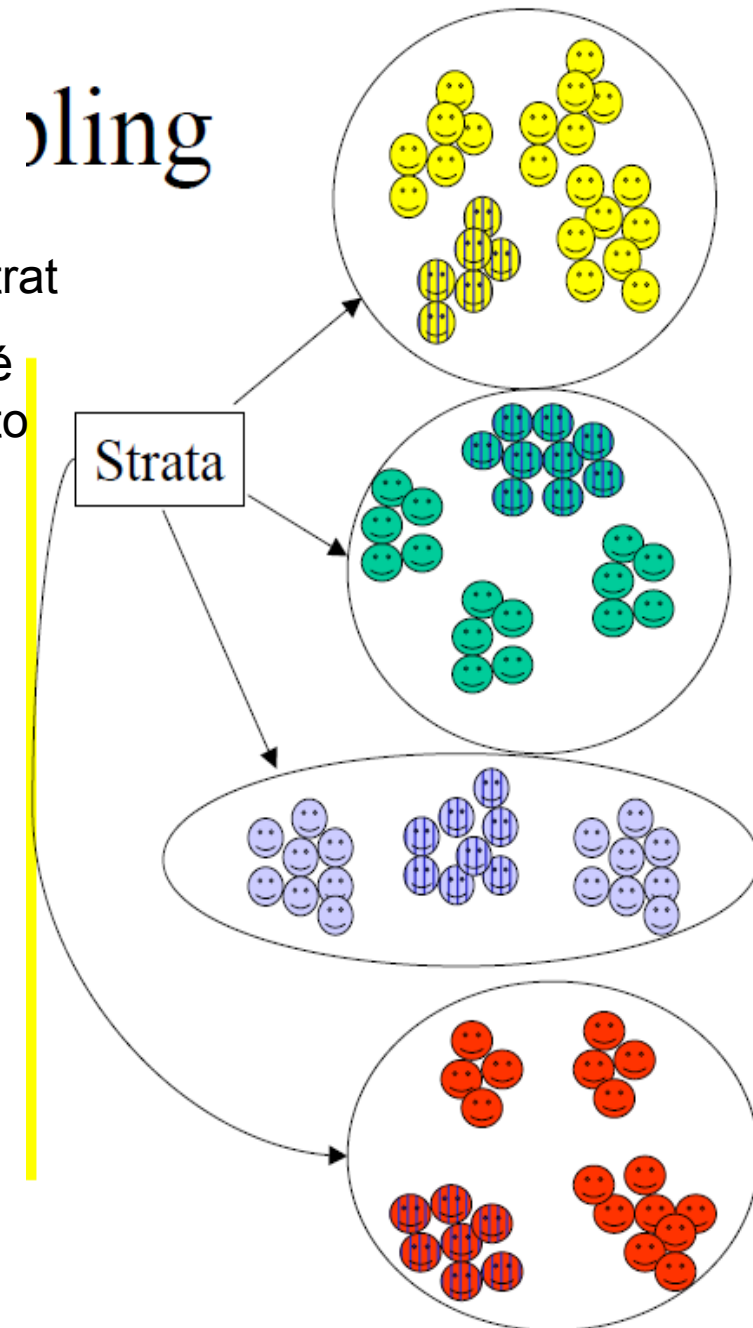
Postup:

1. Definujeme klastry
2. rozdělíme klastry podle podobnosti do skupin / strat
3. náhodně vybereme 1/více celých klastrů z každé straty (nebo pouze náhodně vybereme členy takto vybraných klastrů)

Vlastnosti:

1. Využívá výhody stratifikovaného a klastrového výběru = Snižuje chybu klastrového výběru vytvořením strat a zároveň zachovává levnost klastrového výběru

oling



5b. Kombinovaný víceúrovňový stratifikovaný-klastrový

- Postup:
- 1. země je rozdělena na klastry (např. regiony)
- 2. některé konurbace jsou vybrány s jistotou, zbytek je rozdělen do strat podle velikosti (např. středně velká města, menší města, venkov)
- 3. náhodně vybereme klastry z každé straty
- 4. uvnitř každého vybraného území znovu definujeme klastry atd. až se dostaneme k domácnostem
- 5. Kombinovaný víceúrovňový náhodný

5c. Kombinovaný víceúrovňový stratifikovaný-klastrový

Postup:

1. náhodně vybereme regiony, z vybraných regionů města, z měst čtvrtě atd. až k domácnostem

Problematika návratnosti

- Obecně čím vyšší návratnost, tím nižší chyba
- Návratnost 90% OK, 50% ne OK -ohrožení zobecnitelnosti
- Pokud neodpověděli náhodně, pak OK, pokud systematicky (určitý typ lidí), pak zkreslení

Velikost vzorku

- Vyšší heterogenita populace (rozptyl), větší vzorek
- Vyšší přesnost (spolehlivost/výběrová chyba), větší vzorek
- Design vzorku (klastrový větší vzorek, stratifikovaný menší vzorek)
- Komplexnější analýza (více proměnných), větší vzorek
- pozor: přesnost vzorku závisí na jeho velikosti, nikoli na poměru vzorek / populace

Obecné typy výběrů vzorku

Pravděpodobnostní

Náhodný

- 1. prostý
- 2. stratifikovaný
- 3. systematický
- 4. klastrový / shlukový
- 5. kombinovaný/
víceúrovňový

Nepravděpodobnostní (účelové neboli teoretické)

-Kvótní

- Extrémní / odchylní se případy
 - intenzivní případy
- Maximální rozptyl (heterogenní)
 - Homogenní
 - Typický případ
 - Účelový stratifikovaný
 - Kritický případ
- Sněhová koule / řetězový
 - Kriteriální
 - Účelový náhodný
- Politicky důležité případy
 - Oportunistický
- Nahodilý/příležitostní
- Na teorii založený

Nepravděpodobnostní výběry

Nevyužívá matematickou teorii pravděpodobnosti, tudíž nelze usuzovat ze vzorku na populaci (na základě teorie pravděpodobnosti)

Předmět: informačně-bohaté případy

Účel: dovědět se do hloubky spíše než zobecňovat

1. Nahodilý / příležitostní / anketa

Jednotky jsou vybrány protože k nim máme snadný přístup

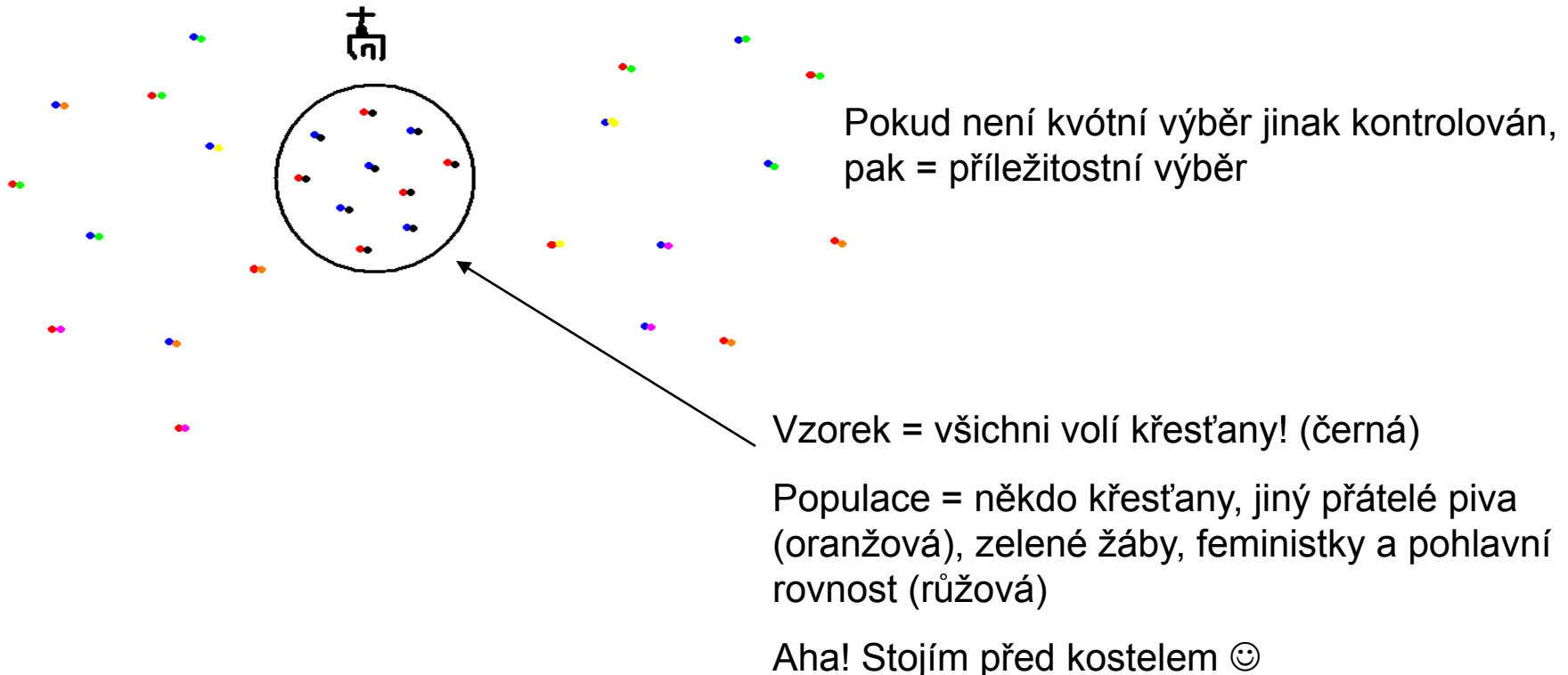
Př. Spolužáci ve třídě, lidé na ulici, kamarádi

2. Kvótní

Předem stanovím počet subjektů v jednotlivých kategoriích, tak aby kopírovaly rozložení v populaci (např. 100 mužů a 100 žen, když poměr M a Ž v populaci je 50/50) a tyto kvóty pak při výběru naplňuji

Problém?

Př. Předvolební průzkum: Víme že v populaci M/Ž 15/15 (M=modrá, Ž=červená), proto kvóta pro výběr = 5 M + 5 Ž, kvóta OK reprezentativní



3. Extrémní / odchylovící se případy

= Výběr případů, které jsou neobvyklé / speciální (odchylky) nějakým způsobem,
(např. velký úspěch / neúspěch)

Např.

2 Školy - s vysokou vs. Nízkou úrovní násilí

2 programy – s vysokou vs. Nízkou úspěšností

Výzkum excelence průmyslových podniků (64 nejlepších podniků)

Obzvláště efektivní programy pro snižování chudoby

Deviantní chování u extrémních/specifických případů

Bestseller „7 zvyků vysoce efektivních lidí“

Výzkum na malém vzorku lidí u nichž výjimečně HIV nepropuklo v AIDS

Evaluační studie „za jakých podmínek program selže / exceluje.“

Důvod: Studie extrémních případů pomůže zlepšit více obvyklé (běžné) případy

Logika: Extrémní případy jsou informačně-bohaté, neboť tím, že jsou neobvyklé, mohou osvětlit obé – neobvyklé a typické.

4. Intenzivní případy

Logika: stejná jako u výběru extrémních případů, ale méně důrazu na extrémy, neboť extrémy mohou být tak extrémní, že mohou být ojedinělé a nejsou použitelné na typičtější případy

5. Maximální rozptyl

Hledání společného mezi různými případy

Logika: cokoli společného v jinak odlišných případech je pozoruhodné

Postup:

1. Napřed si stanovím kritéria (různí lidi podle vzdělání, programy podle velikosti regionu atd.)
2. Všímám si jedinečností
3. Všímám si společného

Př. Evaluační studie úspěšnosti inkluze studentů s postižením

1. *Kritérium: různá velikost a charakter sídla (venkov, malé město, velkoměsto, předměstí)*
2. *Jedinečnost: schopnost oslovit a udržet kvalitní personál*
3. *Společné: zvýšená interakce mezi studenty s/bez postižení*

6. Homogenní vzorek

Opak k maximálnímu rozptylu

Účel: popsat určitou podskupinu do hloubky

Př. Program vzdělávání rodičů

Vedoucí programu chtějí informace o matkách samoživitelkách, neboť to je těžko dosažitelná a udržitelná skupina v programu

Doporučená strategie pro focus-group výzkum, neboť pokud jsou lidé heterogenní, m
dostat prostor dominantnější lidé a převážit „dominantní“ pohled na věc
(např. focus-group s matkami s postiženými dětmi a administrátory programu zároveň)

7. Typický případ

Výběr typického / průměrného případu

Účel: podat informaci o typických případech / např. typických zkušenostech s programem

Postup: definovat kritéria typičnosti (př. věk, vzdělání, ale také výkon programu, spokojenost klientů)

Př. Evaluační studie: Klíčoví rozhodovatelé o programu chtěli informace o programech, které jsou tak obyčejné, že se o nich neví...

8. Kritický případ

Logika: „pokud se to stane zde, pak se to stane kdekoli“
nebo „pokud se to nestane v tomto případě, pak se to nestane nikde“
Nebo „pokud má tato skupina problém, pak si můžeme být jisti, že všichni mají problém“

Např. Politici chtějí zahrnout místní obyvatele do rozhodování o způsobu fungování lokálního programu – budou lidé rozumět pravidlům rozhodování?

Kritérium pro kritický případ: úroveň vzdělání

Pokud vysoce vzdělaní lidé nebudou rozumět, pak méně vzdělaní vůbec.

Pokud málo vzdělaní rozumí, pak všichni budou rozumět

Galileo Galilei: Teorie gravitace.

Ovlivňuje váha objektu rychlost s jakou padá k zemi?

Kritický případ: porovnal pírko a minci zda padají stejnou rychlostí

9. Metoda sněhové koule / řetězový

Proces nabalování lidí, kteří jsou posuzováni okolím za největší experty na daný problém

Postup: Ptám se lidí, o kterých si myslím, že mají přehled (klíčoví informanti), kdo podle nich ví nejmíc. Dostávám seznam jmen. Ptám se lidí ze seznamu kdo podle nich víc nejmíc. Dostávám další jména – koule se nabaluje. Postupně se jména opakují až se opakují zcela – kruh se uzavírá.

Vyberu jména, která jsou opakovaně zmiňována různými lidmi.

Př.

Já: „Koho bych se měl zeptat? Kdo o tom podle vás ví nejmíc?“

Odpověď: Lojza.

Já: „Lojzo, kdo o tom podle tebe ví nejmíc?“

Lojza: Adam.

Já: „Adame, kdo o tom podle tebe ví nejmíc?“

Adam: „Lojza“.

Já pro sebe: „Hm, koule se ještě nezačala pořádně nabalovat a už se kruh uzavírá, do vzorku jde Lojza a Adam.“

10. Kriteriaální

Postup: stanovím kritéria. Lidi kteří splní kritéria jdou do vzorku.

Př. Očekávaná délka pobytu v léčebně je 4 až 26 týdnů.

Pacienti kteří překročí 28t jsou prozkoumání, aby se zjistil důvod.

Př. Dramaticky stoupl počet žádostí o přestup ze základní na speciální školu. Výzkumník chce zjistit důvod (Odešel někdo z personálu ze ZŠ? Přišel nějaký expert na SŠ?)

Kritérium výběru: Studenti kteří byli odesláni na speciální školu v posledních 6 měsících.

Př. Všichni klienti psychiatrické léčebny, kteří spáchali sebevraždu do 3 měsíců od propuštění.

Př. Všichni studenti s absencí nad 25procent.

Účel: Porozumění informačně-bohatým případům, kteří splňují kritéria, může odhalit chybu v programu a pomoci k nápravě.

11. Na teorii založený

Vychází z teorie (a její operacionalizace), kterou chci ověřovat.

Př. Studium významu úzkosti – sociální faktory, které způsobují úzkost – nezaměstnanost – vybírám nezaměstnané

12. Oportunistický

V kvalitativním / interpretativním výzkumu

Doplňování vzorku **v průběhu studie**, reakce na situaci / příležitost.

13. Stratifikovaný účelový

Kombinace typického případu a maximálního rozptylu

Př. Vzorek nadprůměrných, průměrných a podprůměrných případů.

(méně než maximální variance, ale více než typický případ)

Účel: zachytit hlavní rozdíly, spíše než společné jádro.

14. Účelový náhodný

účel: ne reprezentativnost, ale **důvěryhodnost**

Př. Sexuální zneužívání na rezidenční škole pro hluché studenty

Účel: důvěryhodnost - studenti vybráni náhodně, ne na základě doporučení ředitele, který by mohl mít zájem vybrat studenty, kteří by školu uvedli v „dobrém světle“

15. Výběr politicky důležitých případů

Účel: politická vhodnost

Př. Program byl implementován v různých regionech. Evaluační studií chceme zjistit jeho efektivitu. Náhodný výběr může opomenout region, v kterém bydlí člověk, který „sedí na penězích“ pro program, což by nebylo „politicky prozíravé“.