

Výběr vzorku v kvantitativním výzkumu

Metodologie ISK, 8/11/2013

Terminologie

Populace / základní soubor

- Soubor jednotek, které chceme zkoumat – předpokládáme, že naše výroky jsou pro tento soubor platné
- Soubor jednotek, ze kterých vybíráme vzorek

Vzorek (výběrový soubor):

- Množina subjektů, které ve výzkumu zastupují náš výběrový soubor
- Jednotky, které skutečně zkoumáme, pozorujeme

Terminologie

Reprezentativnost

- Zobecnitelnost vzorku na populaci
- Jde to, co bylo vyzkoumáno, vztáhnout i na další prvky populace?

Cenzus

- Úplný/totální/vyčerpávající výběr
- Speciální případy (ČSÚ) nebo pokud je populace tak malá, že není efektivní dělat výběr (desítky až stovky prvků)



Typy výběrů

Reprezentativní výběry		Nereprezentativní výběry
Pravděpodobnostní výběry	Nepravděpodobnostní výběry	
Prostý náhodný výběr	Kvótní výběr	Snowball technika
Systematický výběr		Teoretický výběr
Náhodný stratifikovaný výběr		Výběr typických případů
Náhodný skupinkový výběr		Výběr kritických případů
		Účelový výběr



Jak vybírat?

Co brát v úvahu při výběru?

- Velikost populace
- Složení populace
(homogenní/heterogenní)
- Složitost zkoumaného problému
- Zamýšlenou míru statistické
pravděpodobnosti
- Jaké informace o populaci máme k
dispozici?
- Zdroje (ekonomickou stránku)



Pravděpodobnostní výběry

Pravděpodobnostní výběry

- Každý prvek (jednotka) základního souboru má **stejnou šanci** dostat se do výběrového souboru (vzorku)
- **Princip náhody**

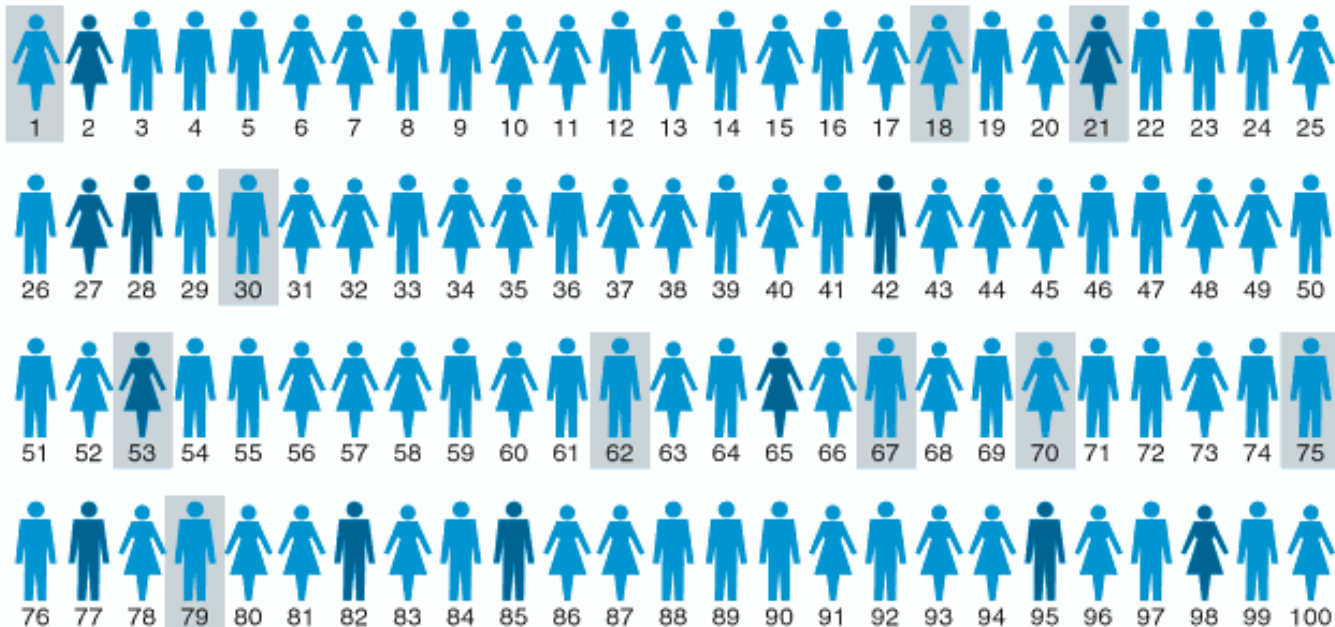
Prostý náhodný výběr

- Vyžaduje **sampling frame** (oporu výběru)
- „losování“
- Vyjadřuje **všechny známé i neznámé** vlastnosti populace (největší výhoda!)
- Jsme schopni odhadnout, jak se liší od populace (výběrová chyba)
- Úplná **eliminace možnosti ovlivnit podobu vzorku** ze strany výzkumníka

Náhodný výběr ≠ jak vás zrovna napadne

Náhodný výběr ≠ kdo jde zrovna kolem

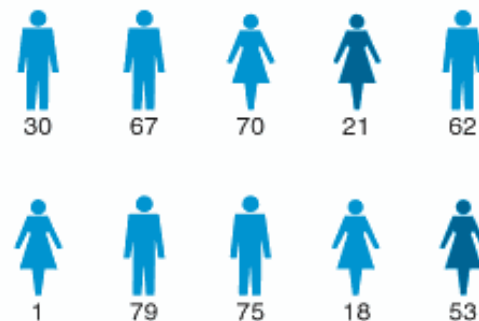
Mikropopulace



Tabulka náhodných čísel

10480	15011	01536
22368	46573	25595
241 30	48360	22527
421 67	93093	06243
375 70	39975	81837
779 21	06907	11008
995 62	72905	56420
963 01	91977	05463
895 79	14342	63661
854 75	36857	53342
289 18	69578	88231
635 53	40961	48235
09429	93969	52636

Vzorek



Systematický náhodný výběr

- Vybíráme každý n -tý případ:
 - Nejprve je náhodně (losem) stanoven první prvek a poté každý x -tý (x = podíl velikosti základního a výběrového souboru)
- Pozor na **organizační klíč** při řazení seznamů! Seznam musí být v tomto případě řazen náhodně!

Random start



The Sample



O výběru, který byl příliš systematický

Bylo, nebylo, kdesi existovalo malé království, které se jmenovalo Org. Bylo to království, kde všechno bylo velice dobře zorganizováno, a přesto byl každý šťastný a spokojený. Každý, až na vojáky základní služby. Ti si stěžovali na plat, na stravu, na zacházení od představených, na všechno, a protože vše bylo dobře zorganizováno, vláda pozvala zahraničního odborníka, profesora P.I. Tomu, aby provedl výzkum postojů v armádě.

P.I. Toma přijel, zkonstruoval výborný dotazník a vyzkoušel jeho validitu. Protože to království bylo tak malé, že se tam ani počítač nevešel a místní knihovny neměly tabulku náhodných čísel, rozhodl se použít pro konstrukci vzorku techniku systematického výběru. Armáda toho malého království byla taky malá, důstojníci, poddůstojníci i mužstvo dohromady jen 12.000 osob. Profesor P.I. Toma odhadl, že vzorek 200 osob mu poskytne přijatelný interval spolehlivosti a zvolil tedy krok 60. Náhodně vybral prvního jedince. Byla to osoba č. 31 a pak vybíral každého dalšího šedesátého vojáka. Výsledky výzkumu byly prostě náramné. Ještě nikdo nikde nezkoumal tak spokojenou armádu. Každý byl šťastný v tom malém šťastném království - až do příštího jara, kdy začalo krvavé povstání vojáků základní služby.

Ale vy už víte, co se stalo: Prostě, v království Org vše bylo dobře organizováno. I seznamy členů armády byly uspořádány po četách, v každé četě nejdříve dva důstojníci, pak tři poddůstojníci, pak mužstvo základní služby a každá četa měla ne více, ne méně než 30 osob, a náš profesor měl smůlu, protože zvolený krok se shodoval přesně nejen s dvojnásobkem velikostí čety, ale také proto, že první náhodně vybraná osoba byl důstojník a tedy každá následující osoba musela být také důstojník. Poddůstojníci a vojáci základní služby nebyli zahrnuti do vzorku vůbec.

Stratifikovaný výběr

1. Vytvoříme **straty** (skupiny homogenní vzhledem ke konkrétnímu kritériu – pohlaví věk – záleží na hypotézách)
2. Ze strat vybereme náhodně jednotky
3. Může být **proporcionální** i **neporcionální**

Příklad: předpokládáme, že vnímání kognitivní autority u studentů KISK je odlišné dle věku, délky studia, pohlaví. Vytvoříme straty dle daných kritérií a z nich vybíráme náhodně.

Skupinkový náhodný výběr

Podobný jako stratifikovaný, ale **skupinky** jsou přirozeného původu a jsou uvnitř heterogenní

1. Identifikujeme skupiny (clustery)
2. Vybereme náhodně clustery
3. Vybereme náhodně prvky z clusterů

Příklad vícestupňového výběru: Chceme zkoumat registrované uživatele v knihovnách v ČR. Vybereme nejprve náhodně kraje, poté knihovny, poté z nich náhodně čtenáře.

Orientační přehled - výběr vzorku

Velikost populace	Velikost vzorku (pravděpodobnostní výběry)
Do 100 jednotek	80 %
Do 1000 jednotek	40 %
Do 10 000 jednotek	7,5 %
Do 100 000 jednotek	1,5 %
Do 1 000 000 jednotek	0,25 %
Do 10 000 000 jednotek	0,045 %

Kvótní výběr

- Imituje **známé vlastnosti** ve struktuře populace (vytváříme model populace)
- Lze použít jen pro **dobře zmapované populace** (známe podíly zastoupení kvót)
- **Skládání kvót znesnadňuje výběr**
- **Nelze stanovit výběrovou chybu(!)**

Zdroje informací o populaci: např. Český statistický úřad (Sčítání lidu), seznam čtenářů knihovny atd.

River sampling (real-time)

- Charakteristický pro online dotazování
- Reaguje na nedostatky panelových výzkumů (vzorek zatížený opětovným dotazováním)
- Rekrutace za pomoci bannerů, pop-upů, reklamy

Výběry založené na dobrovolnosti/dostupnosti

- Experimenty
- **Ankety** (samovýběr), výběr namátkou („kdo jde zrovna kolem“)

Výsledky nejsou zobecnitelné na populaci! „Reprezentativita není určována jen počtem jedinců, ale i mechanismem jejich výběru“
(Reichel, 2009)



Sběr dat z pohledu respondenta

Druhy sběru dat

- PAPI (Paper and Pencil Interviewing)
- CAPI (Computer Assisted Personal Interviewing)
- CAWI (Computer Assisted Web Interviewing) CATI (telefonické dotazování respondentů za pomoci počítačového programu)
- CAMI (Computer Assisted Mobile Interviewing)



Cvičení: výběr vzorku a sběr dat