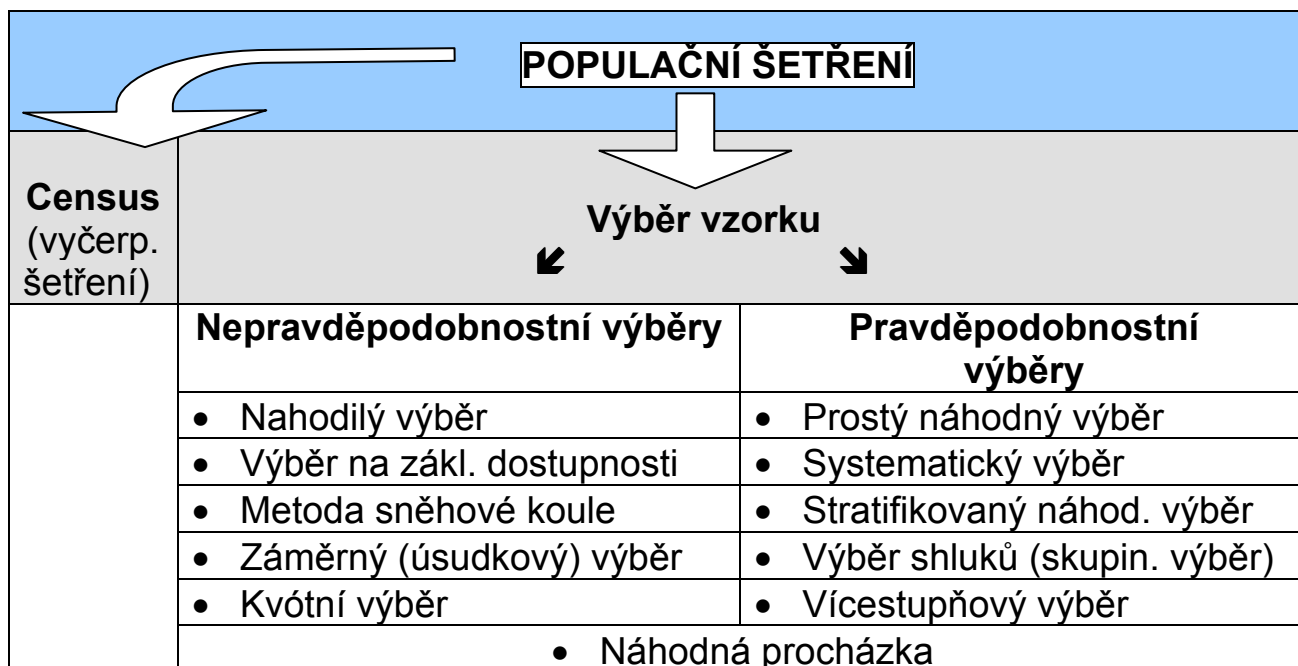


METODY VÝBĚRU VZORKU a VELIKOST VZORKU



Zdařilá práce se vzorkem:

Výsledky předvoleb. průzkumů a výsledek volby presidenta USA (1996) (%)

Agentura	Bill Clinton	Bob Dole	Ross Perot	Ostatní
Hotline/ Battleground	49	40	9	2
CBS/New York Times	54	35	9	2
Pew Research Center	52	38	8	2
Reuter/Zogby	49	41	8	2
Harris	51	39	9	1
ABC	52	39	7	2
NBC/Wall St. Journal	51	38	9	2
Gallup/CNN/USAToday	51	38	9	2
Výsledky voleb	49	41	8	2

Pramen: Babbie, E. 2001 *The Practise of Social Research*. Belmont: Wadsworth.

Neúspěšná práce se vzorkem:



Photo by United Press International, Inc.

Vítěz presidentských voleb v USA (1948), Harry Truman, ukazuje mylnou předpověď výsledků voleb, otištěnou v *Chicago Daily Tribune*. Předvolební průzkum (George Gallup) nesprávně určil jako vítěze Thomase Deweyho. Neúspěch Gallupovy předpovědi zapříčinil (také) nepřesně postavený vzorek respondentů.

Kroky vytváření vzorku:

- (a) stanovení cílové populace (=základního souboru)
- (b) získání opory výběru

(1) Nepravděpodobnostní výběry (*Nonprobability sampling*)

Typické: v některé fázi tvorby vzorku obsahují osobní úsudek

- nemáme plnou kontrolu nad kvalitou vzorku
- nevíme zdali, nakolik a kterým směrem je vychýlen

Kvótní výběr (*quota sampling*)

do praxe zavedl:
George Gallup
zakladatel *American Institute of Public Opinion* (1936)

Kvóty = předem specifikované skupiny
zdroje kvót – např. SLDB, VŠPS

Kvótní kritéria (zpravidla)

- věkové skupiny
- pohlaví
- vzdělání
- region
- velikosti obce

Předpoklad: shoda složení výběrového souboru se souborem základním z hlediska vybraných identifikačních znaků, zajistí i shodu v rozdělení ostatních znaků

(2) Pravděpodobnostní výběry (*probability sampling*)

= základní metoda výběru velkých reprezentativních vzorků populace

- Vylučují vliv úsudku na proces výběru (=objektivita)
- Princip : každá jednotka má stejnou šanci dostat se do vzorku

Stratifikovaný náhodný výběr (*stratified random sampling*)

„Strata“ (většinou) = „subpopulace“

Postup: 2 kroky: **krok A.** základní soubor rozdělíme do „strat“
(strata musí soubor úplně pokrývat)

krok B. v každém stratu = prostý náhodný výběr
(zpravidla stejnou proporci jednotek)

Příklad: „Šetření univerzitních studentů“ (*Postoje ke konzumaci alkoholu*)

Fakulty	Populace	Stratifikovaný vzorek (1/20 z popul.)	Pravděpodobný vzorek při užití prostého náhod. výběru
<i>Humanities</i>	1 800	90	85
<i>Social Sciences</i>	1 200	60	70
<i>Pure sciences</i>	2 000	100	120
<i>Applied sciences</i>	1 800	90	84
<i>Engineering</i>	2 200	110	91
TOTAL	9 000	450	450

Převzato z: Babbie, E. 2001 *The Practise of Social Research*. Belmont, Wadsworth. (str. 90)

Stratifikovaný výběr → vyšší míru reprezentativity, než prostý náhodný výběr
= vybíráme správné počty případů z homogenních podsouborů populace

Při větší homogenitě souboru: klesá velikost „výběrové chyby“

Výběr shluků (skupinkový výběr) *cluster sampling*

Základní soubor sestává ze „shluků“

= vnitřně heterogenních, ale navzájem podobných skupinek

Náhodně vybíráme skupinky / shluky → kompletně šetříme

„Ideální shluk“ = odráží strukturu názorů celé (šetřené) populace

Vícestupňový výběr (*multistage sample*)

Postupně vybíráme jednotky (např. územní) od velkých k menším
Na posledním stupni výběru → případy

příklad: okresy – obce - části obce / ulice – dům – domácnost / osoby

Technika znamená opakování 2 základních kroků
(1) soupis jednotek **(2)** výběr jednotek

Náhodná procházka (*random route, random walk*)

Charakteristika: metoda výběru adres pro terénní šetření, v situacích, kdy není k dispozici seznam adres pro výběr. Ač se hojně užívá, vede se spor, nakolik metoda splňuje kriteria pravděpodobnostních výběrů. Některá šetření ji neakceptují (např. *European Social Survey*). V přesné definici jejího postupu bývají rozdíly.

Postup: tazatel vyjde z předem (náhodně) určeného „startovního bodu“.

Hlavní idea je: pravidlo „pravé ruky“ (popř. levé ruky) a výběr „každého x-tého domu“.

Potenciální problémy: nutno vyloučit zkreslení, např. ignorování bočních uliček.

	<p>Vyjděte ze startovního bodu. Běžte po pravém chodníku a vyhledejte každý třetí vchod ve vašem směru, znázorněného tučnou šipkou. Narazíte-li na křižovatku, zatočte vpravo, držte se pravého chodníku a pokračujte ve vyhledávání. Naleznete-li v obytném bloku, který vám byl přiřazen, jen malý počet domů, pak poté, co znova dorazíte ke startovnímu bodu, pokračujte v chůzi ve směru tečkované šipky (<i>Moon street</i>). Pokud byste (což je málo pravděpodobné) nenaplnil určenou kvótu respondentů, i přesto že jste využil všech směrů od startovního bodu, jděte opět původní cestou, tentokrát vyberte PRVNÍ vchod a dále každý třetí. Nepomůže-li to, požádejte svého instruktora o další startovní bod.</p>
--	--

Organizační kroky:

- Náhodný výběr lokalit
- Výběr startovních adres v každé lokalitě
- Zpracování instrukcí pro „procházku“ a výběr domů
- Kontrola (např. podle kvót a podle výskytu specifických respondentů)

VŠECHNY UVEDENÉ TECHNIKY VÝBĚRU SE NĚKDY KOMBINUJÍ S CÍLEM VYVÁŽIT JEJICH VÝHODY A NEVÝHODY.

Jak je vybírán konkrétní respondent ve vícečlenné domácnosti?

Metoda nejbližších narozenin:

z několikačlenné domácnosti je vybrán respondent, jehož narozeniny přijdou nejdříve po datu interview.
(alternativa: „Metoda posledních narozenin“)

VELIKOST VZORKU (Sample Size)

U pravděpodobnostních výběrů

Determinanty velikosti vzorku:

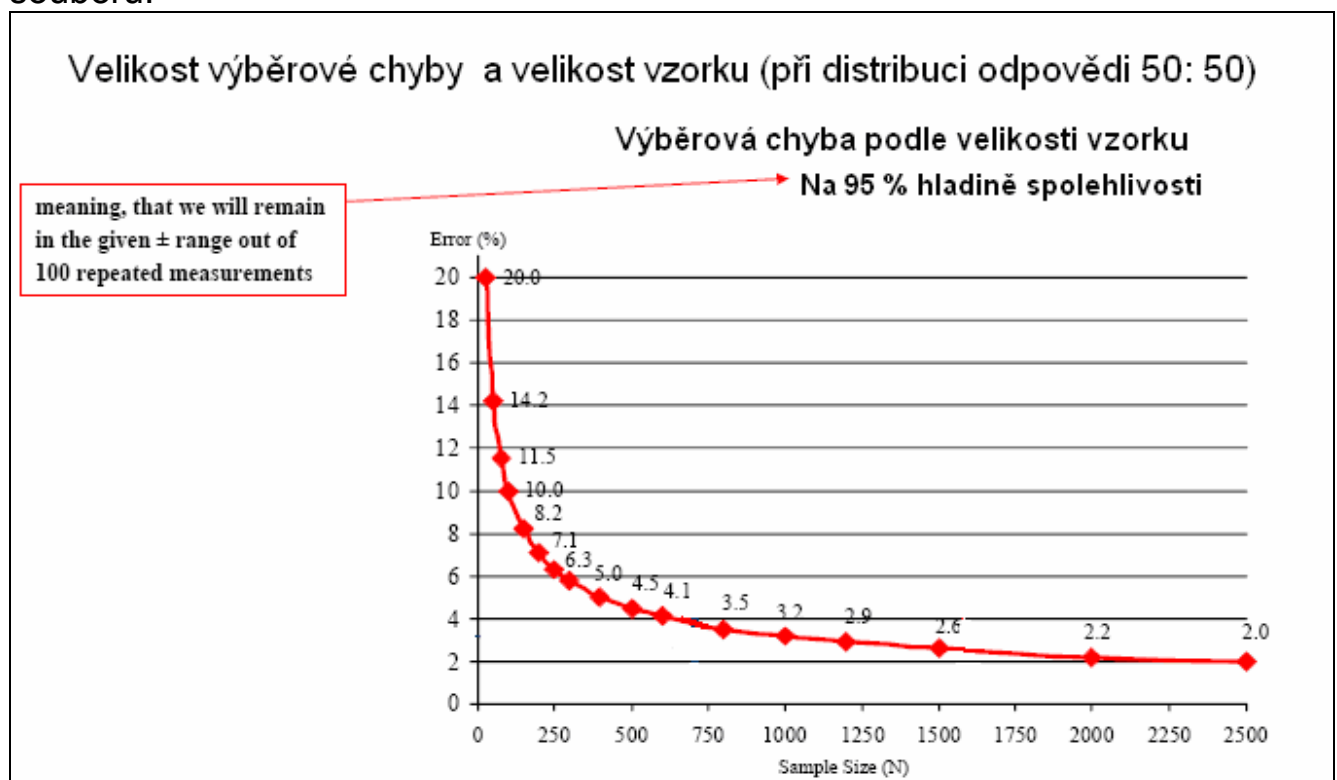
- Jak velkou výběrovou chybu hodláme tolerovat
- Jak heterogenní bude sledovaná populace
- Jakého typu budou klíčové proměnné
- Jaké podskupiny vzorku hodláme analyzovat
- Kolik to bude vše stát

(1) Jak velkou *výběrovou chybu* hodláme tolerovat (Sampling Error)

Výběrová chyba = nakolik se může výsledek zjištěný ve výběrovém souboru odchylovat od skutečnosti v základním souboru.

Udávána v % na zvolené *hladině spolehlivosti*
(zpravidla 95 %)

Stanovíme-li si přípustnou výběrovou chybu, můžeme určit velikost výběrového souboru.



Pramen: Gallup Organization Europe (upraveno)

(2) Jak heterogenní bude sledovaná populace (a tedy získané odpovědi)

Velikost vzorku podle homogenity odpovědí a požadované přesnosti

Přijatelná výběr. chyba	homogenita			heterogenita		
	5 : 95	10 : 90	20 : 80	30 : 70	40 : 60	50 : 50
1 %	1 900	3 600	6 400	8 400	9 600	10 000
2 %	479	900	1 600	2 100	2 400	2 500
3 %	211	400	711	933	1 066	1 100
4 %	119	225	400	525	600	625
5 %	76	144	256	336	370	400
6 %	-	100	178	233	267	277
7 %	-	73	131	171	192	204
8 %	-	-	100	131	150	156
9 %	-	-	79	104	117	123
10 %	-	-	-	84	96	100

Pramen: Blaikie, N. 2000 *Designing Social Research*. Cambridge: Polity. (str. 210)

(3) Jakého typu budou klíčové proměnné

(intervalové, ordinální, nominální)

- Čím nižší úroveň měření, tím větší vzorek je potřeba

např.

při užití neparametrických testů by vzorek měl být cca 10x větší, než je počet políček v kontingenční tabulce

- Velikost vzorku nutno podřídít té nejnižší úrovni měření

(4) Jaké podskupiny vzorku hodláme analyzovat (souvisí s body 2+3)

- Chceme-li analyzovat podskupiny, rozšiřujeme vzorek
- Důležitá = velikost nejmenší podskupiny
- Obecné pravidlo: nejmenší podskupina by měla mít alespoň 50 jedinců

(5) Kolik peněz to vše má stát

Faktická velikost vzorku bývá **kompromisem** mezi metodologickými požadavky a dostupnými (finančními) zdroji.