

# Úvod do statistiky

**Povinná literatura:** Mann (2016), Kapitola 1

## Z čeho studovat první lekci?

**Povinná literatura:** Mann (2016), kap. 01

**Příprava na cvičení:** Leaflet 01 a Leaflet 02  
Koncepty a procedury, cv.02, kap. 01

**Příprava na zkoušku:** Mann (2016), kap. 01  
Leaflet 01 a Leaflet 02  
Sbírka úloh, kap. 01  
Koncepty a procedury, cv.02, kap. 01

# Obsah přednášky

Statistika a typy statistik

Základní termíny

Typy proměnných

Typy datových sad

Populace vs. Vzorek

Šetření

## Co je to statistika? Jak ji dělíme?

**Statistika** je soubor metod používaných k sběru, analýze, prezentaci a interpretaci dat. Taktéž může sloužit jako soubor metod při různých typech (nejen) ekonomického rozhodování.

Typy statistiky:

1. **Popisná statistika** zahrnuje metody pro organizaci, zobrazení a popis dat pomocí tabulek, grafů a souhrnných ukazatelů.
2. **Induktivní (inferenční) statistika** sestává z metod, které využívají výsledky získané na nějakém vzorku k podpoře rozhodování nebo předpovídání faktů o celé populaci.

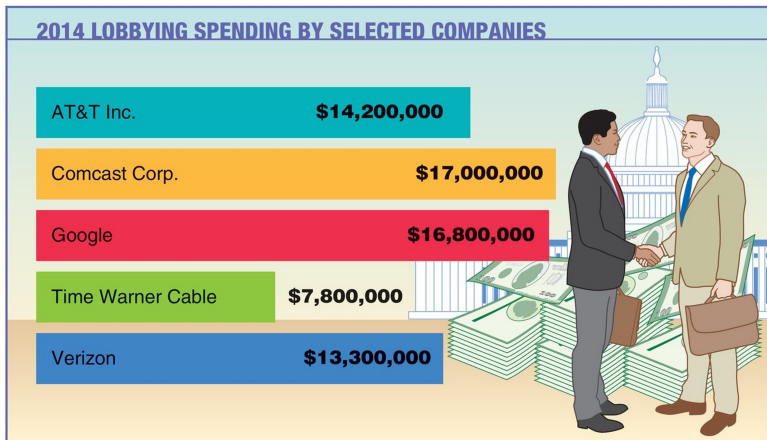
## Příklady využití statistiky: Popisná statistika

Následující graf ukazuje výdaje na lobbování pěti vybraných společností v roce 2014. Mnoho společností utrácí miliony dolarů, aby si získaly přízeň ve Washingtonu.

Podle časopisu Fortune z 1. června 2015 „Comcast zůstává jedním z největších korporátních lobbistů v zemi.“ V roce 2014 Comcast utratil 17 milionů dolarů, Google 16,8 milionu dolarů, AT&T 14,2 milionu dolarů, Verizon 13,3 milionu dolarů a Time Warner Cable 7,8 milionu dolarů na lobbování.

Tyto čísla jednoduše popisují celkové částky, které tyto společnosti utratily na lobbování. Z těchto dat nevyvozujeme žádné závěry, rozhodnutí ani předpovědi. Tento datový soubor a jeho prezentace jsou tedy příkladem popisné statistiky.

# Příklady využití statistiky: Popisná statistika



Data source: Fortune Magazine, June 1, 2015

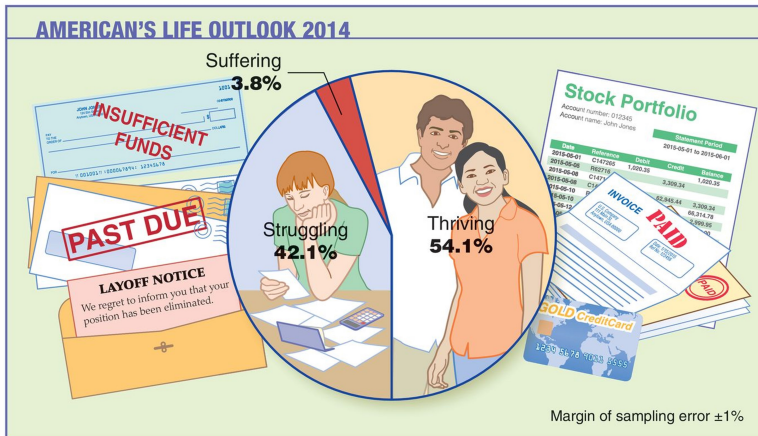
## Příklady využití statistiky: Inferenční statistika

Průzkum mezi 176 903 americkými dospělými ve věku 18 let a více byl proveden od 2. ledna do 30. prosince 2014 jako součást průzkumu Gallup-Healthways Well-Being Index. Gallup a Healthways sledují „hodnocení života Američanů každý den“ od roku 2008.

Dle průzkumu byl v roce 2014 pohled Američanů na život nejlepší za posledních sedm let, protože 54.1% hodnotilo svůj život natolik pozitivně, že byli považováni za prosperující, 42.1% uvedlo, že bojují, a 3.8% zmínilo, že trpí. Uvedená výběrová chyba byla  $\pm 1\%$

V kapitole 8 se budeme zabývat konceptem výběrové chyby, kterou lze využít při vyvozování závěrů. Jak si všímáme, výsledky popsané v grafu byly získány z průzkumu mezi 176 903 dospělými. V pozdějších kapitolách se naučíme, jak aplikovat tyto výsledky na celou populaci dospělých. Takové rozhodování o populaci na základě výsledků vzorku se nazývá inferenční statistika.

# Příklady využití statistiky: Inferenční statistika



Data source: Gallup-Healthways Well-Being Index



# Obsah přednášky

Statistika a typy statistik

Základní termíny

Typy proměnných

Typy datových sad

Populace vs. Vzorek

Šetření

## Základní termíny:

**Prvek** nebo **člen** vzorku nebo populace je konkrétní subjekt nebo objekt (například osoba, firma, položka, stát nebo země), ke kterému se sbírají informace.

**Proměnná** je zkoumaná charakteristika, která může nabývat různých hodnot u různých prvků. **Konstanta** na rozdíl od proměnné je charakteristická tím, že každý prvek výběru nebo populace má přiřazenu stejnou hodnotu.

Hodnota proměnné pro prvek se nazývá **pozorování** nebo **měření**.

**Datová sada** je sbírka pozorování na jedné nebo více proměnných.

# Základní termíny

	<b>Name</b>	<b>Total Wealth (billions of dollars)</b>	← Variable
	Bill Gates	79.2	
	Carlos Slim Helu	77.1	
An element or member } →	<b>Warren Buffett</b>	<b>72.7</b>	← { An observation or measurement
	Amancio Ortega	64.5	
	Larry Ellison	54.3	
	Charles Koch	42.9	
	David Koch	42.9	
	Christy Walton	41.7	

# Obsah přednášky

Statistika a typy statistik

Základní termíny

Typy proměnných

Typy datových sad

Populace vs. Vzorek

Šetření

# Typy proměnných

## Kvantitativní proměnná

- Proměnná, která může být měřena numericky, se nazývá kvantitativní proměnná. Data shromážděná v kvantitativní proměnné se nazývají kvantitativní data.
- Rozlišujeme: **Diskrétní proměnné**, **Spojité proměnné**

## Kvalitativní nebo kategorické proměnná

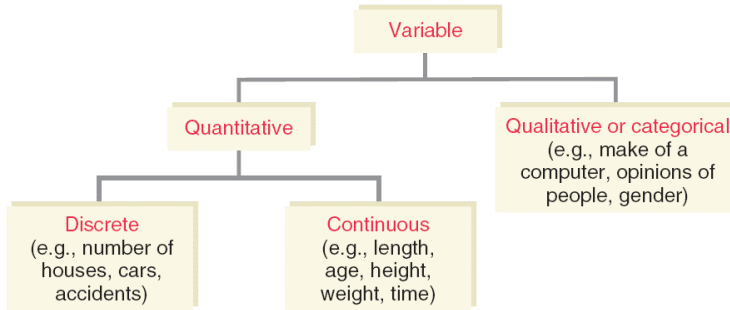
- Proměnná, která nemůže nabývat numerické hodnoty, ale lze ji zařadit do dvou nebo více nečíselných kategorií, se nazývá kvalitativní nebo kategorická proměnná. Data shromážděná v této proměnné se nazývají kvalitativní data.

## Typy Kvantitativní proměnné

**Diskrétní proměnná** je proměnná, jejíž hodnoty jsou spočetné. Jinými slovy, diskrétní proměnná může nabývat pouze určitých hodnot bez mezihodnot.

**Spojité proměnná** je proměnná, která může nabývat jakékoli numerické hodnoty v určitém intervalu nebo intervalech.

# Typy proměnných



## Typy proměnných

Rozhodněte, zda jsou následující proměnné kvalitativní nebo kvantitativní, pokud jsou kvantitativní rozdělte je na diskrétní nebo spojité.

Proměnná	Typ proměnné
Počet přihlášek na MUNI sbíraný každý rok	
Hmotnost balení rýže prodaných trhovcem dne 1.1.2025	
Velikost trička (S,M,...) u studujících na této přednášce	
Typ telefonu, který studenti dneska používají	
Hrubý domácí produkt všech zemí v dolarech v 2021-2025	
Energetická úspornost (A, B, ...) budov v Brně k dnešku	



# Obsah přednášky

Statistika a typy statistik

Základní termíny

Typy proměnných

Typy datových sad

Populace vs. Vzorek

Šetření

## Typy datových sad

**Průřezová data** jsou data shromážděná na různých prvcích ve stejný časový okamžik nebo po čas stejného období.

**Časové řady** jsou data shromážděná pro jeden prvek na stejné proměnné v různých časových okamžicích nebo po dobu různých časových období.

**Panelová data** jsou data, které kombinují jak časové řady, tak i průřezová data. Jsou to data, která sledují stejné prvky (například osoby, firmy, regiony) opakovaně v průběhu času.

# Typy proměnných

**Table 1.2** Total Wealth of World's Eight Richest Persons

Name	Total Wealth (billions of dollars)
Bill Gates	79.2
Carlos Slim Helu	77.1
Warren Buffett	72.7
Amancio Ortega	64.5
Larry Ellison	54.3
Charles Koch	42.9
David Koch	42.9
Christy Walton	41.7

Source: *Forbes*, March 23, 2015.

Obrázek: Průřezová data

**Table 1.3** Average Tuition and Fees in 2014 Dollars at Four-Year Public Institutions

Years	Tuition and Fee (Dollars)
1974–75	2469
1984–85	2810
1994–95	4343
2004–05	6448
2014–15	9139

Source: The College Board.

Obrázek: Časová řada

# Počet obyvatel ve třech největších městech v ČR

Rok	Praha (tis.)	Brno (tis.)	Ostrava (tis.)
2020	1 324	378	290
2021	1 345	382	291,5
2022	1 360	385	293
2023	1 375	388	294,5
2024	1 390	392	296

## Typy datových sad

Rozhodněte, zda se jedná o průřezová data či časové řady.

Proměnná	Typ proměnné
Počet přihlášek na MUNI sbíraný každý rok	
Hmotnost balení rýže prodaných trhovcem dne 1.1.2025	
Velikost trička (S,M,...) u studujících na této přednášce	
Typ telefonu, který studenti dneska používají	
Hrubý domácí produkt všech zemí v dolarech v 1990-2025	
Energetická úspornost (A, B, ...) budov v Brně k dnešku	

# Obsah přednášky

Statistika a typy statistik

Základní termíny

Typy proměnných

Typy datových sad

Populace vs. Vzorek

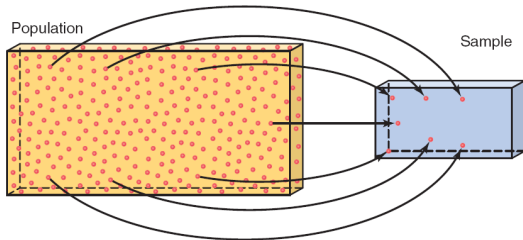
Šetření

## Populace vs. Vzorek

**Populace** zahrnuje všechny prvky, jejichž charakteristiky jsou zkoumány (například: jednotlivce, položky nebo objekty).

Populace, která je předmětem výzkumu, se nazývá **cílová populace**.

Část populace vybraná ke studiu (nebo na ní prováděno určité šetření) se nazývá **vzorek**.



## Populace vs. Vzorek

Průzkum, který zahrnuje každého člena populace, se nazývá **census (sčítání)**.

Technika sběru informací z části populace se nazývá **průzkum na vzorku (sample survey)**.

Vzorek, který co nejvěrněji reprezentuje charakteristiky populace, se nazývá **reprezentativní výběr**.



## Proč vzorek?

Existuje několik důvodů, proč provádět analýzu na vzorku místo na celém sčítání (censu).

**Časová a finanční efektivita:** Sčítání celé populace může být časově náročné a finančně nákladné. Používání vzorků umožňuje provádět analýzy a získávat odhady za mnohem kratší dobu a s menšími náklady.

**Praktická proveditelnost:** V některých případech je prakticky nemožné získat úplný přehled o celé populaci. Použití vzorku umožňuje praktičtější a proveditelnější získání informací o charakteristikách celé populace.

## Náhodný výběr

**Náhodný výběr** je výběr prvků z populace tak, že každý prvek má při výběru šanci být zahrnut, a to bez systematického zkreslení. Jinými slovy každý člen populace má nějakou šanci být vybrán.

**Nenáhodný výběr** je výběr prvků z populace, ve kterém nejsou dodržovány zásady náhodnosti. Tento typ výběru může být ovlivněn subjektivními nebo systematickými faktory, což může vést k zkreslení výsledků a ztížit zobecňování poznatků na celou populaci. V případě nenáhodně vybraného vzorku prvků z populace může dojít k tomu, že někteří členové populace nemusí mít žádnou šanci být vybráni do vzorku.

## Příklady nenáhodného výběru

**Judgment sample** je typ výběru vzorku, který není náhodný, ale závisí na odborném posouzení nebo úsudku jednotlivce provádějícího výzkum. V tomto případě je vzorek vybrán na základě odborného uvážení kvalifikované osoby, která má znalosti nebo zkušenosti týkající se dané oblasti. Tento typ výběru se často používá v případech, kde náhodný výběr není prakticky proveditelný nebo kde je důležitá expertní znalost při výběru reprezentativních prvků pro studii.

### Příklad

**Hodnocení dopadu politiky:** Při hodnocení dopadu nové vzdělávací politiky může výzkumník vybrat ředitele škol z nejlépe hodnocených škol (protože tito jedinci mohou poskytnout nejrelevantnější informace pro studii), aby získal názor na to, jak politika ovlivnila jejich provoz.

## Příklady nenáhodného výběru

**Convenience sample** je typ výběru vzorku, kdy jsou jednotky do vzorku zařazeny na základě jejich snadné dostupnosti nebo přístupnosti pro výzkumníka. Při použití této metody jsou vybírány prvky, které jsou snadno dosažitelné nebo k dispozici, aniž by byl uplatněn nějaký specifický náhodný postup. Tento přístup může být rychlý a levný, ale může způsobit zkreslení výsledků, protože vzorek nemusí správně reprezentovat celou populaci.

### Příklad

**Studie v obchodním centru:** Výzkumníci chtějí zkoumat nákupní chování, a proto se rozhodnou oslovit lidi v nákupním centru. Dotazují kolemjdoucí, protože jsou na místě a ochotni odpovědět.

## Příklady nenáhodného výběru

**Pseudo polls** se obvykle označují za "falešné nebo klamné průzkumy veřejného mínění". Jedná se o situace, kdy jsou prezentovány jako průzkumy veřejného mínění, ale ve skutečnosti nejsou provedeny metodou, která by zaručovala objektivní a reprezentativní vzorek populace.

### Příklad

**Online anketa na zpravodajském webu:** Zpravodajský server zveřejní otázku typu „Kdo by měl vyhrát nadcházející volby?“ a umožní čtenářům hlasovat. Tento průzkum však není reprezentativní, protože ho vyplní jen návštěvníci webu, kteří mají zájem hlasovat.

## Příklady nenáhodného výběru

**Quota sample** je typ výběru vzorku, kde jsou účastníci vybíráni na základě určených kvót, aby vzorek odpovídal určitým charakteristikám nebo proporcím ve srovnání s populací. Tento typ výběru není náhodný, ale je řízen tak, aby dosáhl určitých cílů týkajících se reprezentativnosti vzorku. Quota sampling může být efektivní při zajištění reprezentativnosti vzorku vzhledem k určitým charakteristikám, ačkoliv se jedná o nenáhodnou metodu výběru.

### Příklad

**Marketingový průzkum pro nový produkt:** Společnost vyvíjející nový kosmetický produkt chce provést průzkum mezi potenciálními zákazníky. Rozhodnou se, že vzorek musí zahrnovat 60 % žen a 40 % mužů, přičemž 40 % z nich musí být ve věku 18-34 let, 35 % ve věku 35-49 let a 25 % starších 50 let. Výběr respondentů je prováděn tak, aby tyto skupiny byly zastoupeny podle stanovených kvót.

# Typy chyby při práci se vzorkem

## Výběrová chyba (Sampling Error):

- Jedná se o rozdíl mezi výsledkem získaným z výběrového průzkumu a výsledkem, který by byl dosažen, pokud by celá populace byla zahrnuta do průzkumu.
- Výběrová chyba vzniká kvůli náhodným variacím ve výběru vzorku.
- Může být snížena zvětšením velikosti vzorku, ale nemůže být úplně eliminována.

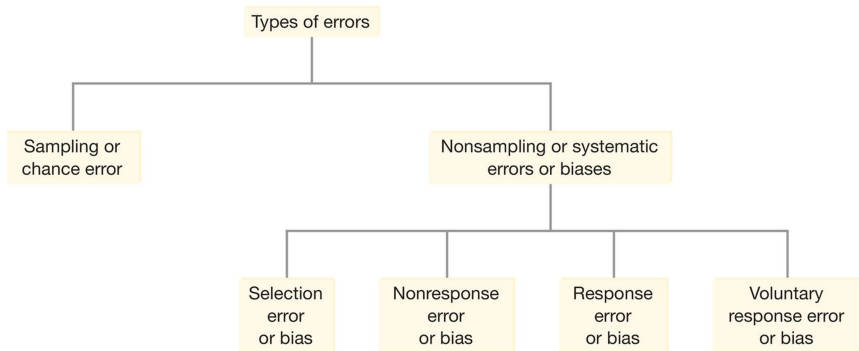
## Typy chyb při práci se vzorkem

### Chyby nezpůsobené výběrem (zkreslení) (Nonsampling errors or biases):

- Jedná se o chyby, které vznikají při sběru, záznamu a tabulaci dat a nejsou přímo spojeny s výběrem vzorku.
- Zkreslení může zahrnovat chyby v procesu měření, neúplnost nebo nepřesnost odpovědí, chyby při zaznamenávání dat, atd.
- Jsou to chyby v procesu sběru a zpracování dat, které mohou mít vliv na přesnost výsledků, ale nemají co do činění s náhodným výběrem vzorku.



# Výběrová chyba (Sampling Error) a zkreslení (Nonsampling errors or biases)



## Typy chyb nezpůsobené náhodným výběrem

Seznam členů cílové populace, který je používán k výběru vzorku, se nazývá výběrový rámec. Chyba, která vzniká

1. kvůli tomu, že výběrový rámec není pro populaci reprezentativní, se nazývá **chyba (volby) výběru**.
2. tím, že mnoho lidí zahrnutých do vzorku neodpovídá na průzkum, se nazývá **chyba nereprezentativní odpovědi**.
3. když lidé zahrnutí do průzkumu neposkytují správné odpovědi se nazývá **chyba nesprávné odpovědi**.
4. když průzkum není proveden na náhodně vybraném vzorku, ale na dotazníku zveřejněném v časopise nebo novinách a lidé jsou pozváni k odpovědi na tento dotazník se nazývá **chyba dobrovolných odpovědí**.

## Techniky náhodného výběru

1. **Jednoduchý náhodný výběr** tato technika výběru zaručuje, že každý vzorek o stejné velikosti má stejnou pravděpodobnost být vybrán.
2. **Systematický náhodný výběr** je metoda výběru vzorku, při níž se prvky populace vybírají s pravidelným intervalem nebo systémem. To znamená, že každý k-tý prvek populace je zahrnut do vzorku, kde „k“ představuje určitý krok nebo interval. Tato metoda může být efektivní a snadno implementovatelná, pokud je seznam populace dostupný a pravidelně uspořádaný.

## Techniky náhodného výběru

- 3. Stratifikovaný náhodný výběr** je metoda výběru vzorku, při níž je populace rozdělena do homogenních skupin nebo vrstev na základě určité charakteristiky (strat). Následně se v každé vrstvě provede náhodný výběr prvků, a tyto vzorky jsou následně kombinovány do jednoho celkového vzorku. Tato metoda pomáhá zaručit, že každá vrstva populace je zastoupena ve vzorku, což umožňuje přesnější odhady parametrů pro celou populaci.
- 4. Klastrovaný náhodný výběr** je metoda výběru vzorku, při níž je populace rozdělena do seskupení nebo klastrů. Náhodně se vyberou některé klastry a všechny jednotky v těchto vybraných klastru jsou zahrnuty do vzorku. Tato metoda usnadňuje výzkumníkům práci s velkými populacemi, kde není praktické nebo ekonomické vybrat náhodný vzorek ze všech jednotek, a umožňuje efektivní získání reprezentativního vzorku.

## Příklad

Představte si, že vláda chce zjistit průměrný příjem domácností v určitém městě, které má 10 000 domácností. Toto město je rozděleno do 100 čtvrtí (každá čtvrť má přibližně 100 domácností). Dále víme, že domácnosti jsou rozděleny do tří příjmových skupin:

- Nízkopříjmové domácnosti (4 000 domácností),
- Středněpříjmové domácnosti (4 000 domácností),
- Vysokopříjmové domácnosti (2 000 domácností).

Vláda chce provést průzkum na vzorku 500 domácností.

Průzkum proveďte pomocí:

1. Jednoduchého náhodného výběru,
2. Systematického náhodného výběru,
3. Stratifikovaného náhodného výběru,
4. Klastrovaného náhodného výběru.

## Příklad - Řešení A: Jednoduchý náhodný výběr

**Definice populace:** Populací je všech 10 000 domácností ve městě.

**Výběr vzorku:** Vláda rozhodne použít jednoduchý náhodný výběr. To znamená, že každé domácnosti přidělí číslo od 1 do 10 000 a poté pomocí náhodného generátoru čísel vybere 500 čísel. Domácnosti s těmito čísly budou zahrnuty do vzorku.

**Výsledky:** Poté, co jsou data shromážděna z vybraných domácností, může vláda odhadnout průměrný příjem všech domácností ve městě na základě tohoto reprezentativního vzorku.

Tento způsob zajišťuje, že každá domácnost má stejnou šanci být vybrána, což minimalizuje výběrovou zaujatost a zvyšuje spolehlivost výsledků.

## Příklad - Řešení B: Systematický náhodný výběr

**Definice populace:** Populací je všech 10 000 domácností ve městě.

**Stanovení intervalu:** Aby vláda získala vzorek 500 domácností z celkového počtu 10 000, určí se výběrový interval. Tento interval se vypočítá jako podíl celkové populace a požadovaného počtu vzorků, tedy

$$\frac{10\,000}{500} = 20.$$

To znamená, že každá 20. domácnost bude zahrnuta do vzorku.

**Výběr počátečního bodu:** Vláda nejprve náhodně vybere počáteční domácnost mezi prvními 20 domácnostmi (například domácnost číslo 12).

## Příklad - Řešení B: Systematický náhodný výběr

**Výběr vzorku:** Po výběru počáteční domácnosti číslo 12 se pak vybírají další domácnosti podle pevného intervalu, tedy každá 20. domácnost. To znamená, že do vzorku budou zahrnuty domácnosti s čísly 12, 32, 52, 72, a tak dále, až dokud nebude vybráno všech 500 domácností.

**Výsledky:** Stejně jako u jednoduchého náhodného výběru, shromáždí se data od vybraných domácností, a na základě těchto dat může vláda odhadnout průměrný příjem všech domácností ve městě.

Tento způsob je často jednodušší na provedení než jednoduchý náhodný výběr, zejména když je seznam všech prvků v populaci uspořádaný a přístupný. Systematický náhodný výběr zajišťuje, že vzorek je rovnoměrně rozložený po celé populaci.



## Příklad - Řešení C: Stratifikovaný náhodný výběr

**Rozdělení populace:** Populace 10 000 domácností je rozdělena do tří strat (nizkopříjmové, středněpříjmové, vysokopříjmové domácnosti).

**Určení velikosti vzorku v každé stratě:** Vzorek 500 domácností by měl reprezentovat podíl každé příjmové skupiny v populaci. Proto se určí, kolik domácností bude vybráno z každé straty:

- Nizkopříjmové domácnosti:  $500 \times \left(\frac{4000}{10000}\right) = 200$  domácností
- Středněpříjmové domácnosti:  $500 \times \left(\frac{4000}{10000}\right) = 200$  domácností
- Vysokopříjmové domácnosti:  $500 \times \left(\frac{2000}{10000}\right) = 100$  domácností

## Příklad - Řešení C: Stratifikovaný náhodný výběr

**Náhodný výběr z každé straty:** V každé z těchto skupin (strat) se náhodně vybere požadovaný počet domácností. Například z nízkopříjmových domácností se náhodně vybere 200 domácností, ze středněpříjmových 200 domácností a z vysokopříjmových 100 domácností.

**Výsledky:** Po shromáždění dat z vybraných domácností může vláda provést analýzu průměrného příjmu, přičemž vzorek zohledňuje různé příjmové skupiny ve městě. To zajišťuje, že vzorek je reprezentativní a zahrnuje všechny významné podskupiny populace.

Stratifikovaný náhodný výběr je zvláště užitečný, když existují významné rozdíly mezi podskupinami v populaci, které by měly být zohledněny při analýze výsledků. Tento přístup pomáhá získat přesnější odhady pro celou populaci.

## Příklad - Řešení D: Klastrovaný náhodný výběr

**Rozdělení populace na klastry:** Populace 10 000 domácností je rozdělena do 100 čtvrtí (klastrů), přičemž každá čtvrť představuje jeden klastr.

**Výběr klastrů:** Z těchto 100 čtvrtí vláda náhodně vybere 5 čtvrtí (klastrů), protože potřebuje vzorek 500 domácností a každá čtvrť má přibližně 100 domácností.

**Sběr dat:** V rámci vybraných 5 čtvrtí jsou pak zahrnuty všechny domácnosti, takže je zahrnuto všech 500 domácností z těchto čtvrtí do průzkumu.

**Výsledky:** Po shromáždění dat od všech domácností ve vybraných čtvrtích může vláda odhadnout průměrný příjem domácností ve městě.

## Příklad - Řešení D: Klastrovaný náhodný výběr

Klastrovaný náhodný výběr je užitečný, když je populace geograficky rozsáhlá nebo jinak obtížně přístupná, protože výběr celých klastrů snižuje náklady a zjednodušuje logistiku. Přestože může být tento způsob méně přesný než stratifikovaný výběr, je často praktičtější v situacích, kdy není možné nebo vhodné náhodně vybírat jednotlivé prvky z celé populace.

# Obsah přednášky

Statistika a typy statistik

Základní termíny

Typy proměnných

Typy datových sad

Populace vs. Vzorek

Šetření

# Šetření

Jedná se o statistickou metodu, která se zabývá plánováním, provedením a analýzou vybraného problému s cílem získat relevantní a spolehlivé informace o vlivu různých faktorů na zkoumaný proces nebo systém.

Tato metoda pomáhá identifikovat klíčové faktory, optimalizovat podmínky a minimalizovat vliv variability na výsledky šetření.

## Šetření - design

**Ošetření (treatment)** je podmínka (nebo soubor podmínek), kterou výzkumník ukládá na skupinu prvků.

Skupina prvků, která obdrží ošetření, se nazývá **ošetřovaná skupina - treatment group**, a skupina prvků, která nedostane ošetření, se nazývá **kontrolní skupina - control group**.

**Randomizace** postup, při kterém jsou prvky náhodně přiřazeny k různým skupinám.

## Pozorovací studie vs. Řízený experiment

**Pozorovací studie (Observační studie)** je typ šetření, kdy je přiřazování prvků k různým ošetřením dobrovolné a experimentátor pouze pozoruje výsledky studie.

**Řízený experiment** je typ šetření, kdy experimentátor řídí (náhodné) přiřazování prvků k různým ošetřením (treatmentům).

**Dvojitě zaslepený experiment** je typ šetření, kde jak účastníci, tak experimentátor sbírající data nevědí, která skupina obdržela aktivní ošetření a která kontrolní podmínky. Šetření je navrženo s cílem eliminovat možné zkreslení v hodnocení výsledků.



## Příklad

Představte si, že farmaceutická společnost vyvinula nový lék k léčbě nemoci XYZ. Aby zjistila, zda je tento lék při léčbě nemoci XYZ účinný, musí být testován na skupině lidí.

Předpokládejme, že existuje 100 osob s nemocí XYZ; 50 z nich se dobrovolně rozhodne užívat tento lék, zatímco zbývajících 50 se rozhodne jej nebrat. Výzkumník pak porovnává míry uzdravení obou skupin pacientů.

Je toto příklad navrženého experimentu nebo pozorovací studie?

## Příklad - řešení

Toto je příklad **pozorovací studie**, protože 50 pacientů se dobrovolně připojilo k ošetřovací skupině; nebyli náhodně vybráni.

V tomto případě mohou výsledky studie být neplatné, protože účinky léku mohou být zkresleny jinými proměnnými. Všichni pacienti, kteří se rozhodli vzít lék, nemusí být podobní těm, kteří se rozhodli jej nebrat.

Je možné, že osoby, které se rozhodly vzít lék, jsou ve stadiích pokročilého onemocnění. Následně nemusí mnoho ztratit tím, že se dobrovolně rozhodnou zvolit být ve skupině na kterou je aplikováno ošetření. Pacienti v obou skupinách se mohou také lišit v dalších faktorech, jako jsou věk, pohlaví a podobně.

## Shrnutí přednášky:

Statistika a typy statistik

Základní termíny

Typy proměnných

Typy datových sad

Populace vs. Vzorek

Šetření

## Co si nastudovat na následující týden?

**Příprava na cvičení:** Leaflet 02  
Koncepty a procedury 02

**Povinná literatura:** Mann (2016), Kapitola 2

# Děkuji za pozornost!

## DATA: BY THE NUMBERS

