



# STATISTICKÁ ANALÝZA DAT



# **Máš data? Tak z nich vytřískej to nejlepší!**

(Návod na vytvoření dechberoucí analýzy)

# CO TĚ V KURZU ČEKÁ

- Ahoj! V tomto ebooku a následně i kurzu, Ti chci poskytnout kompletní návod, který Ti pomůže co nejlépe zpracovat sehnaná data.

Po kurzu budeš ohromen/a tím, co všechno je možné z dat zjistit a jak s nimi pracovat. **Tak pojďme na to!**

## Co mám na účtě?

- Více než 100 **školení statistiky** pro vysokoškolské studenty
- **Vzdělávací platformu**, kterou prošlo již přes 1000 studentů  
- [www.statistickyneklasicky.cz](http://www.statistickyneklasicky.cz)
- **Vlastní knihu** - Neučebnice statistiky

**AUTORKA  
KURZU**

**ADRIANA  
ŘEHÁČKOVÁ**

## OBSAH KURZU

- **Kvantitativní výzkum**
- Cíl práce a statistické hypotézy
- Sběr dat a tvorba dotazníku
- Základní práce s daty
- Tvorba grafů v Excelu
- Práce s kvalitativními daty
- Práce s kvantitativními daty
- Regrese a korelace
- Časové řady

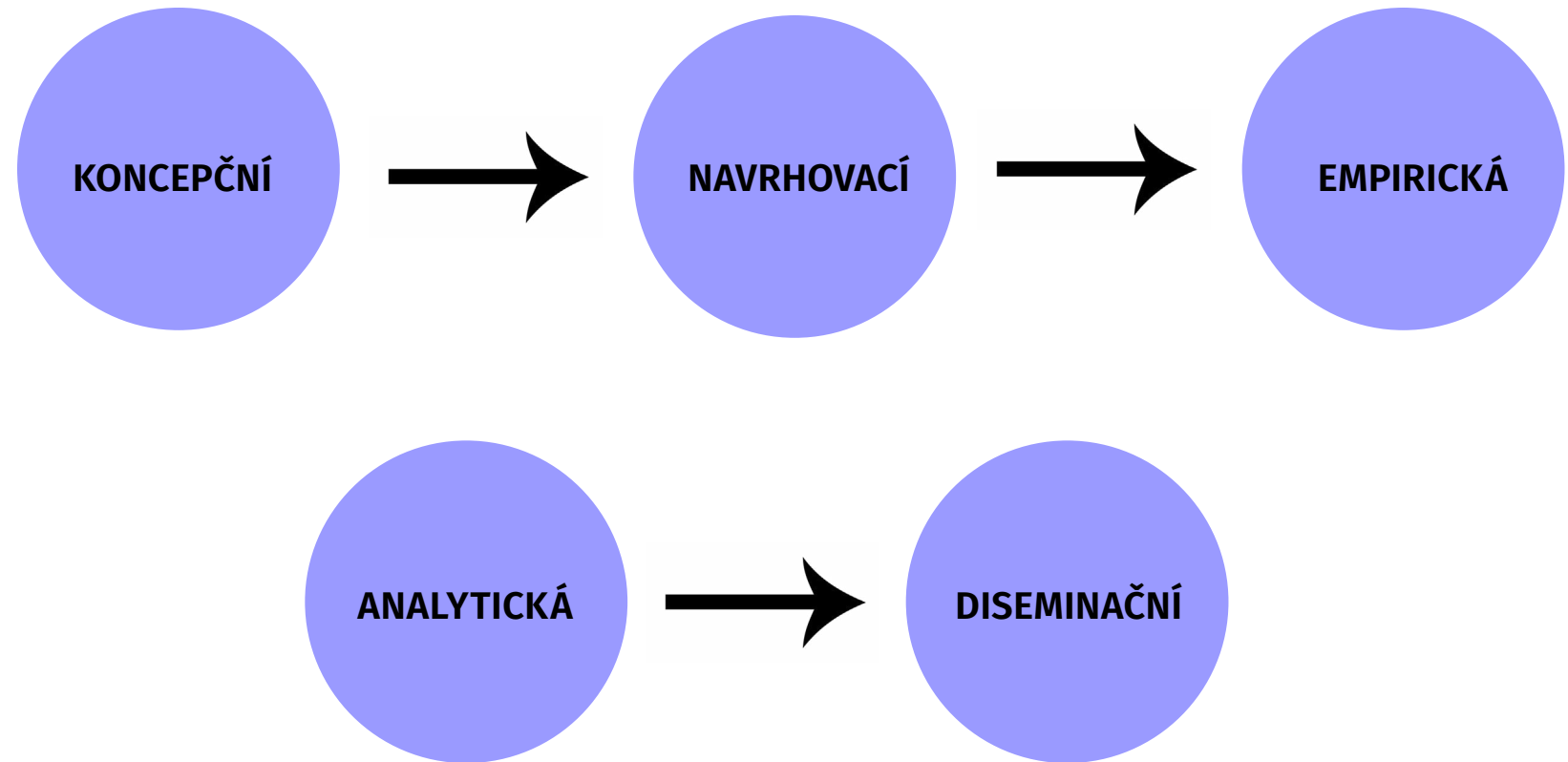
# Kvantitativní analýza

- Relativně velký vzorek respondentů
- Omezené množství informací o velkém počtu respondentů
- Data získávána pomocí dotazníků, sekundární analýza dat,...
- Snadná generalizace na populaci s měřitelnou validitou
- Statistické zpracování dat
- Testování hypotéz

# Kvalitativní analýza

- Malý počet respondentů
- Mnoho informací o malém počtu respondentů
- Data získaná na základě rozhovorů, nestandardizovaných pozorování,...
- Generalizace na populaci je obtížná, někdy i nemožná
- Nestatistické zpracování dat
- Vytváření nových hypotéz, teorií

# FÁZE KVANTITATIVNÍHO PROCESU



**KVANTITATIVNÍ  
VÝZKUM**

# KONCEPČNÍ FÁZE

- Jako první se zaměříme na oblast našeho výzkumu a uděláme **předběžnou teoretickou analýzu** - Zjistíme co nejvíce informací o dané problematice (literatura, názory odborníků)
- Další částí je **stanovení tématu výzkumu** (zpravidla téma bakalářské/diplomové práce) a následná **formulace problému** (Více v kapitole “Cíl práce a statistické hypotézy”).
- Pokud máme správně definovaný výzkumný problém, přistoupíme ke stanovení **cíle výzkumu**.
- Poslední fází je na základě cíle **stanovit hypotézy** (Více v kapitole “Cíl práce a statistické hypotézy”).

# NAVRHOVACÍ FÁZE

- Jako první si rozplánujeme co jak uděláme a kolik času na jakou část potřebujeme.
- Další částí je **charakteristika zkoumaného objektu**, zpravidla pracujeme s výběrovým souborem, který by měl být **reprezentativní** (měl by mít stejné charakteristiky jako základní soubor, obsahovat dostatečný počet pozorování).
- **Zvolit formu sběru dat** - dotazníkové šetření, pozorování, volně dostupná data,...  
PS: užitečné tipy a odkazy v kapitole “Sběr dat a tvorba dotazníku”
- Poslední fází je **finalizace a možný předvýzkum**. Tedy kontrola předešlých kroků a minimalizace možných chyb a následně tvorba předvýzkumu, který je zpravidla prováděn na menším vzorku dat.



# EMPIRICKÁ FÁZE

- **V této fázi výzkumu se získávají data o daném problému. Může se jednat o dohledání potřebných volně dostupných dat, dotazníková šetření či pozorování.**
- **Námi získaná data připravíme pro další analýzu - zkontrolujeme zda v datech nejsou chybějící/neúplné odpovědi, data přeneseme do vhodného programu, okódujeme a připravíme se na další část výzkumu.**

# **ANALYTICKÁ FÁZE**

- **V této fázi je provedena analýza získaných dat pomocí vhodně zvolených statistických metod.**
- **Tato fáze je zaměřená na výpočet základních popisných statistik a vytvoření tabulek četností, následně jsou data vhodně zobrazena pomocí grafů.**
- **Následuje statistická část, pomocí které například dokážeme identifikovat závislosti mezi proměnnými, predikovat vývoj dat a otestovat stat. hypotézy.**

# CÍL PRÁCE A STATISTICKÉ HYPOTÉZY

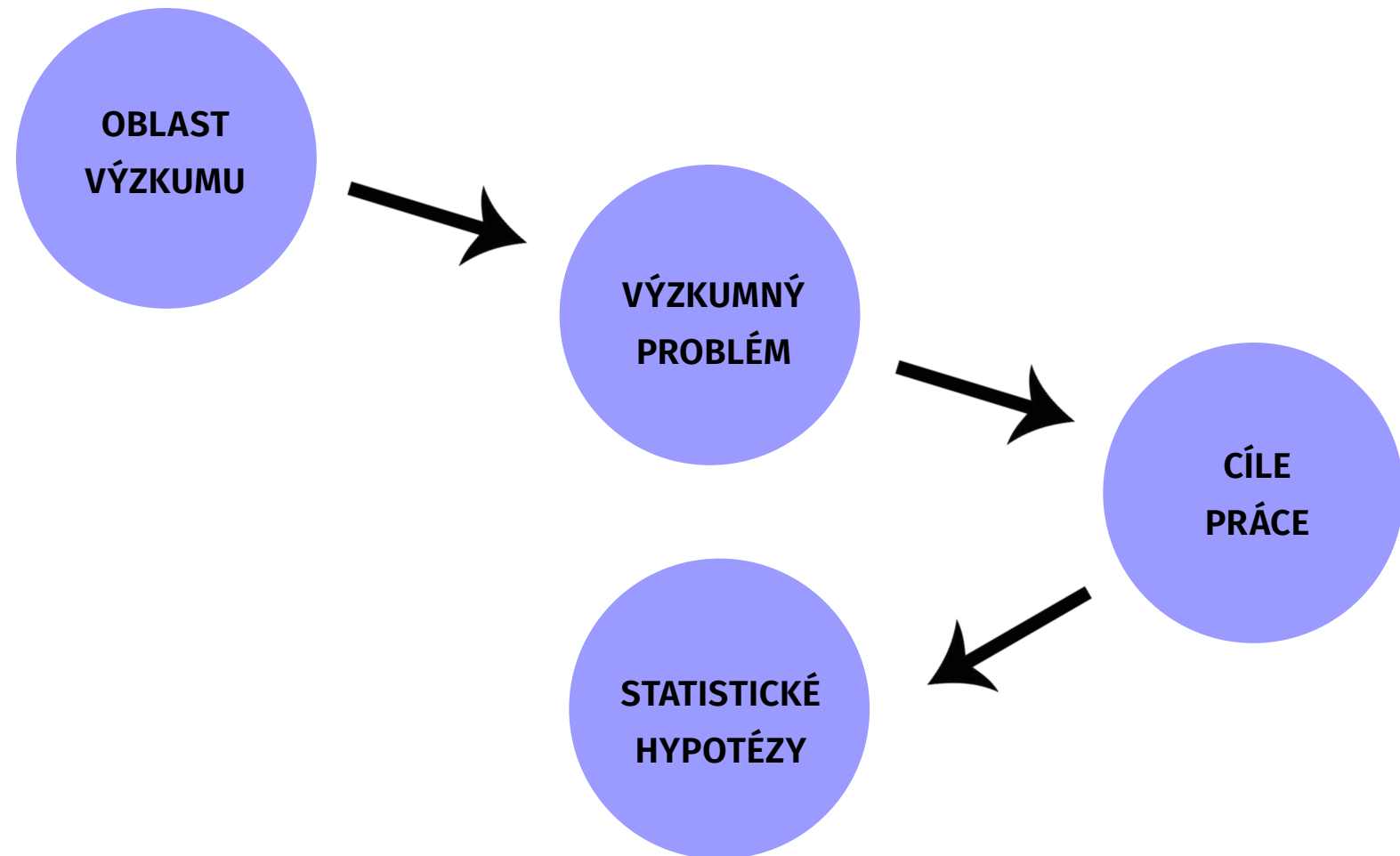
- Kvantitativní výzkum
- **Cíl práce a statistické hypotézy**
- Sběr dat a tvorba dotazníku
- Základní práce s daty
- Tvorba grafů v Excelu
- Práce s kvalitativními daty
- Práce s kvantitativními daty
- Regrese a korelace
- Časové řady

# CÍL PRÁCE A STATISTICKÉ HYPOTÉZY

- Na následujících stránkách si ukážeme jak stanovit výzkumný problém, cíl/e práce a jak správně formulovat statistické hypotézy.
- Správně stanovené cíle práce jsou jednou z nejdůležitějších částí celé výzkumné části.

**CÍL PRÁCE  
A STATISTICKÉ  
HYPOTÉZY**

# VÝZKUMNÝ PROBLÉM, CÍLE A HYPOTÉZY



**CÍL PRÁCE  
A STATISTICKÉ  
HYPOTÉZY**

# OBLAST VÝZKUMU

- Jedná se o oblast která nás zajímá a ve které budeme realizovat výzkum.
- **Příklady:** Matematická gramotnost u studentů středních škol, sebevražednost u drogově závislých, spotřeba potravin v českých domácnostech.

# VÝZKUMNÝ PROBLÉM

- Oproti výzkumné oblasti je nutné, aby byl výzkumný problém jasně definovaný.
- Výzkumný problém by měl prohloubit danou problematiku a přinést něco nového.
- **Příklady:**  
Příčiny nízké matematické gramotnosti u studentů středních škol.  
Vztah mezi drogovou závislostí a sebevraždností.  
Faktory, které ovlivňují spotřebu potravin v českých domácnostech.

# CÍLE PRÁCE

- Prvně si položíme otázky “Co budeme zkoumat?” a “Jak to budeme zkoumat?”.
- Cíle práce jsou **klíčovým bodem** celého našeho snažení.
- V případě kvantitativního výzkumu pak cíl práce zpravidla definujeme pomocí sloves: **analyzovat, prokázat, najít, zjistit,...**
- **Příklady:**  
„Popsat demografické a osobnostní charakteristiky studentů s nejvyšší a nejnižší matematickou gramotností. Najít příčiny nízké matematické gramotnosti. Analyzovat faktory mající vliv na nadměrnou spotřebu v některých českých domácnostech....“
- Cíle práce by měly být formulovány v **přítomném čase**, být **jednoznačné a srozumitelné**.



# STATISTICKÉ HYPOTÉZY

- Správně stanovená hypotéza by měla vycházet z teoretické části.
- Základní typy hypotéz jsou hypotézy o **souvislosti, následku a o rozdílu**.
- Hypotézy můžeme definovat buď **oboustranně** nebo **jednostranně**.
- Hypotéza je **oznamovací věta**. Vždy máme jak alternativní, tak nulovou hypotézu. Nulovou hypotézu není nutné v práci uvádět (Případně konzultujte s vedoucím).
- Při testování hypotéz používáme předem danou **hladinu významnosti** (zpravidla 5 %).
- **Prvně stanovíme hypotézy** a teprve potom sháníme/sbíráme data.
- Hypotéza musí být **testovatelná** (lze ji přijmout, či nepřijmout).

# STATISTICKÉ HYPOTÉZY - PŘÍKLADY

- V hypotéze musí být vždy zmíněná **statistická významnost**.

## **Příklady:**

„Mezi matematickou a čtenářskou gramotností existuje **statisticky významná závislost**.“

“S rostoucím skórem sociokenomického statusu, **signifikantně roste skóre matematické gramotnosti**.“

“Míra sebevraždy je **signifikantně vyšší u drogově závislých, než u osob, které drogy neužívají**.“

“Existuje **statisticky významná závislost mezi pohlavím a mírou drogové závislosti**.“

“Domácnosti se čtyřmi a více členy mají **stat. významně vyšší spotřebu než domácnosti s méně než čtyřmi členy**.“

# SBĚR DAT A TVORBA DOTAZNÍKU

- Kvantitativní výzkum
- Stanovení statistických hypotéz
- **Sběr dat a tvorba dotazníku**
- Základní práce s daty
- Tvorba grafů v Excelu
- Práce s kvalitativními daty
- Práce s kvantitativními daty
- Regrese a korelace
- Časové řady

# VOLNĚ DOSTUPNÁ DATA

Níže můžeš vidět pár užitečných odkazů, kde je možné najít data vhodná ke statistickým analýzám.

- [Český statistický úřad](#)
- [Eurostat](#)
- [Informační systém o průměrném výdělku](#)
- [Statistiky ministerstva práce a soc. věcí](#)
- [Data od Světové banky](#)
- [Statistický portál OSN](#)
- [Data ze všech možných odvětví \(data.world\)](#)
- [Mezinárodní stat. portál NationMaster](#)

SBĚR DAT  
A TVORBA  
DOTAZNÍKU

# TVORBA DOTAZNÍKU

- Pokud chceme vypracovat kvalitní výzkum, potřebujeme v první řadě sestavit kvalitní dotazník, díky kterému získáme hodnotná data. A právě to si teď ukážeme!
- Jak již bylo zmíněno, prvně je nutné definovat si výzkumný problém a výzkumné otázky (ideálně i stat. hypotézy) a teprve potom se pustit do tvorby dotazníku.

# ZÁKLADNÍ PRAVIDLA

- Motivujte respondenta k vyplnění dotazníku, objasněte mu smysl vašeho dotazníku a případně i kontakt, kde vás respondent může kontaktovat.
- Dotazník by měl být anonymní a mělo by to v něm být uvedeno.
- Otázky pokládejte srozumitelně a jednoduše.
- Složitě nebo neznámé pojmy definujte.
- Jednoduché a pokud možno i zajímavé otázky dejte na začátek, složitější na konec.

# DÉLKA DOTAZNÍKU A POČET RESPONDENTŮ

- Ideální délka dotazníku by měla být kolem 7 minut (20-25 otázek). Maximum je potom ideálně 15 minut.
- Počet respondentu? Ideálně platí, že čím více, tím lépe. Naprosto minimální (z mého pohledu však nedostatečná) hranice je 30 respondentů.
- Počet respondentů samozřejmě závisí na více faktorech, ze zkušenosti mohu říci, že pokud děláte běžné dotazníkové šetření některé cílové skupiny, snažte se získat minimálně 150-300 respondentů. Kolem 300 respondentů je již soubor, na který lze aplikovat různé metody a spíše se tak vyhnout nesplnění podmínek testů. Ideální je však mít kolem 500-1000 respondentů.

# SPRÁVNÝ TYP OTÁZEK

- Primárně pracujeme se 2 typy otázek, **otevřené a uzavřené**.
- **Otevřené otázky** jsou užitečné, ale doporučuji jejich počet minimalizovat. Dávají prostor respondentovi se vyjádřit a nám mohou poskytnout další zajímavé postřehy, zároveň je však velmi obtížné tyto otázky dále zpracovávat.
- **Uzavřené otázky** jsou pro náš výzkum klíčové a je možné s nimi dále pracovat a aplikovat na ně příslušné metody. Jednoduše se vyplňují. Nevýhodou může být snadné překliknutí (chyby) a nemusí vždy zcela vystihnout názor respondenta.



# UZAVŘENÉ OTÁZKY

- Uzavřené otázky můžeme dále rozdělit dle následujících typů:
- **Dichotomické otázky** - otázky, které mají 2 možnosti odpovědí (ano/ne, muž/žena)
- **Výběrové otázky** - Dávají respondentovi možnost vybrat jednu z nabízených alternativ. [V případě, že se např. jedná o kategorie věku, tak by měly být všechny intervaly stejně dlouhé (méně než 18, 18-27, 28-37, 38-47,...)]
- **Výčtové otázky** - Zde má respondent možnost vybrat několik nabízených alternativ zároveň
- **Škálové otázky** - Jsou skvělé v případě, že chceme měřit názory a postoje respondentů na některé ze zvolených škál.
- **Výběrové otázky se zpracují zpravidla lépe, než otázky výčtové.**

# ŠKÁLOVÉ OTÁZKY

- Pro mě, jako statistika, jsou škálové otázky jedny z nejoblíbenějších, dají se totiž ze všech výše zmíněných variant jednoduše kvantifikovat a statisticky analyzovat.
- **Hodnotící otázky** - Respondent vyjádří svůj postoj na stupnici, zpravidla podle Likertovy škály (např. silně nespokojen, mírně nespokojen, neutrální, mírně spokojen, velmi spokojen).
- **Škála pořadí** - Zde má respondent za úkol seřadit jednotlivé možnosti dle svých preferencí.
- **Škála konstantní sumy** - Zde respondent rozděljuje předem stanovené množství bodů (např. 100) mezi jednotlivé možnosti, opět dle svých preferencí.

# KDE DOTAZNÍK VYTVOŘIT

- Doporučuji tvořit dotazník online, jelikož pro další práci je nezbytné mít data v elektronické podobě.
- Za mě osobně zcela jednoznačně doporučuji **Google dotazníky**, které jsou **ZDARMA** a výsledky lze snadno exportovat do Excelu, samotná tvorba dotazníku je jednoduchá a intuitivní.
- Další možností je například **Survio**, které má velmi pěkné návrhy šablon, jednoduché ovládání a rovněž nabízí export dat. Musíte však počítat se zaplacením poplatku, v případě, že chcete data exportovat a dál s nimi pracovat.

# Práce s daty může úspěšně začít!

Skvělé videolekce, tutoriály a tvorbu kompletních analýz najdeš v kurzu **STATISTICKÁ ANALÝZA DAT** na webu [statistickyneklasicky.cz](http://statistickyneklasicky.cz)

Těším se na Tebe!

Adriana Řeháčková

+420 702 042 455

[info@statistickyneklasicky.cz](mailto:info@statistickyneklasicky.cz)

[www.statistickyneklasicky.cz](http://www.statistickyneklasicky.cz)