

Experimental Data Analysis

in ©MATLAB

Lecture 1:

Introduction to the Experimental Data Analysis,
Introduction to Data:

Motor Speech Disorders & Basal Ganglia Dysfunction

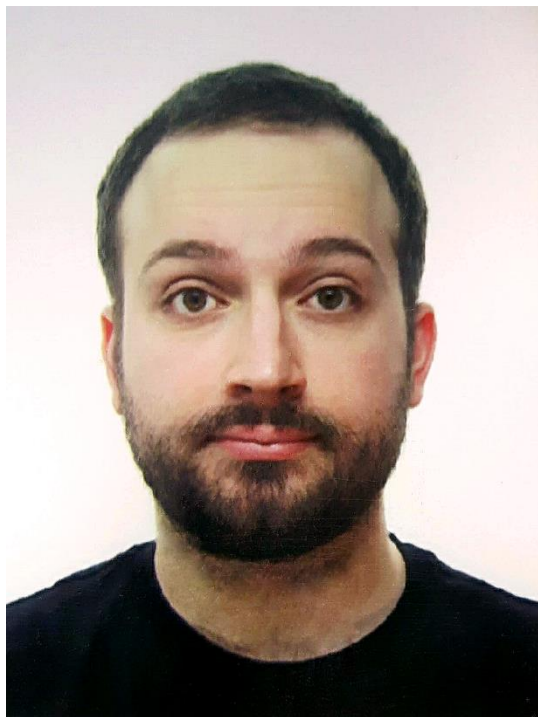
Jan Rusz

Czech Technical University in Prague



Praktická cvičení

Petr Krýže



Martin Šubert



Proč je důležitá „analýza experimentálních dat“?

VÝZKUM

- využití při řešení různých výzkumných a vývojových projektů ve všech odvětvích výzkumu
- statistika je jazykem pro kvantitativní interpretaci dat
- typické otázky jsou statistického charakteru:
 - „Dokážu rychlostí chůze definovat typ onemocnění?“
 - „Je tato nová léčebná metoda lepší než dosud užívaná?“
- smyslem pokroku současnosti je publikace nových pozorování, které se neobejdou bez interpretace dat
- na základě pozorování dochází k vývoji nových technologií, léčebných postupů atd.
- rozvoj výpočetní techniky umožnil kvantitativní zpracování dat s využitím různých statistických metod

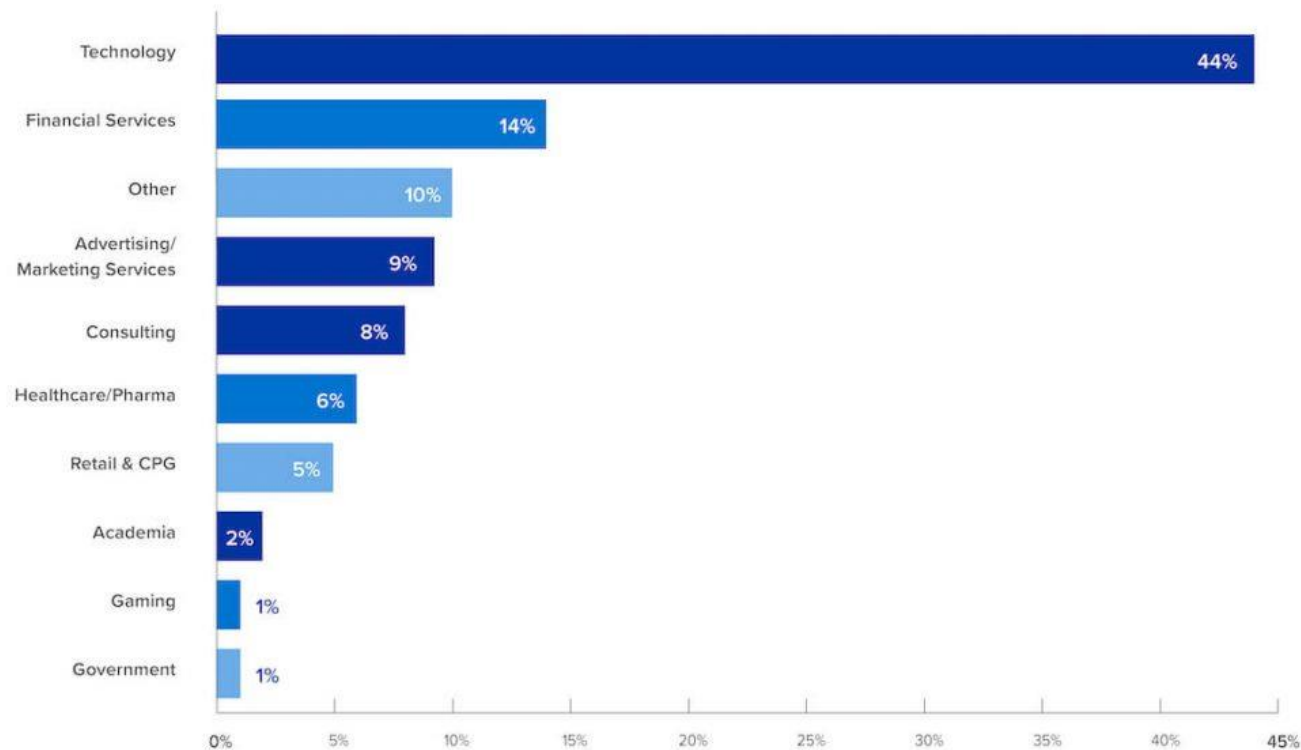
Proč je důležitá „analýza experimentálních dat“?

PRŮMYSL/PRAXE

- analýza dat nám pomůže najít skryté známky, neznámé souvislosti, marketingové trendy atd.
- průmysl začíná být závislý na zpracování velkých objemů dat
- zpracování dat běžného života (mobilní telefony)
- rozhodovací a predikční systémy

Proč je důležitá „analýza experimentálních dat“?

What Industries Do Data Scientists Work In?



„s rychlostí současného rozvoje technologií se s vysokou pravděpodobností stane v budoucích letech analýza a zpracování dat základní dovedností“

Co Vás naučí „analýza experimentálních dat“?

- praktická aplikace základních statistických metod na různé typy dat
 - základní statistické metody (1/3 předmětu)
 - pokročilé statistické metody (2/3 předmětu)
 - základy strojového učení (3/3 předmětu)
- získání základní dovednosti pro samostatné řešení výzkumného projektu
- seznámení s problematikou zpracování dat v neurovědách

Návrh a plánování projektu

- nelze studovat celou populaci, která nás zajímá
- cílem výzkumu/vývoje je zobecňování poznatků z daného výběru na celou populaci
- definice cílů, účelů a typu studie
 - **cíl 1**: vytipovat osoby v riziku rozvoje neurologického onemocnění (např. Parkinsonova nemoc)
 - **účel 1**: provést preventivní opatření
 - **typ 1**: skupinová srovnání, tzv „cross-sectional“
 - **cíl 2**: nalézt vedlejší účinky léčby
 - **účel 2**: optimalizace terapie
 - **typ 2**: longitudinální sledování
- je potřeba znát dobře zkoumanou problematiku („state-of-the-art“)
=> riziko vzniku nadbytečného výzkumu/vývoje
- před zahájením je důležité formulovat hypotézy, které mají být kriticky prověřeny
 - **hypotéza 1**: porucha řeči bude brzkým ukazatelem neurologické nemoci
 - **hypotéza 2**: vysoké dávky dopaminergní medikace povedou k rozvoji koktavosti
- zásadní chybou je formulace hypotéz až podle nashromážděných dat

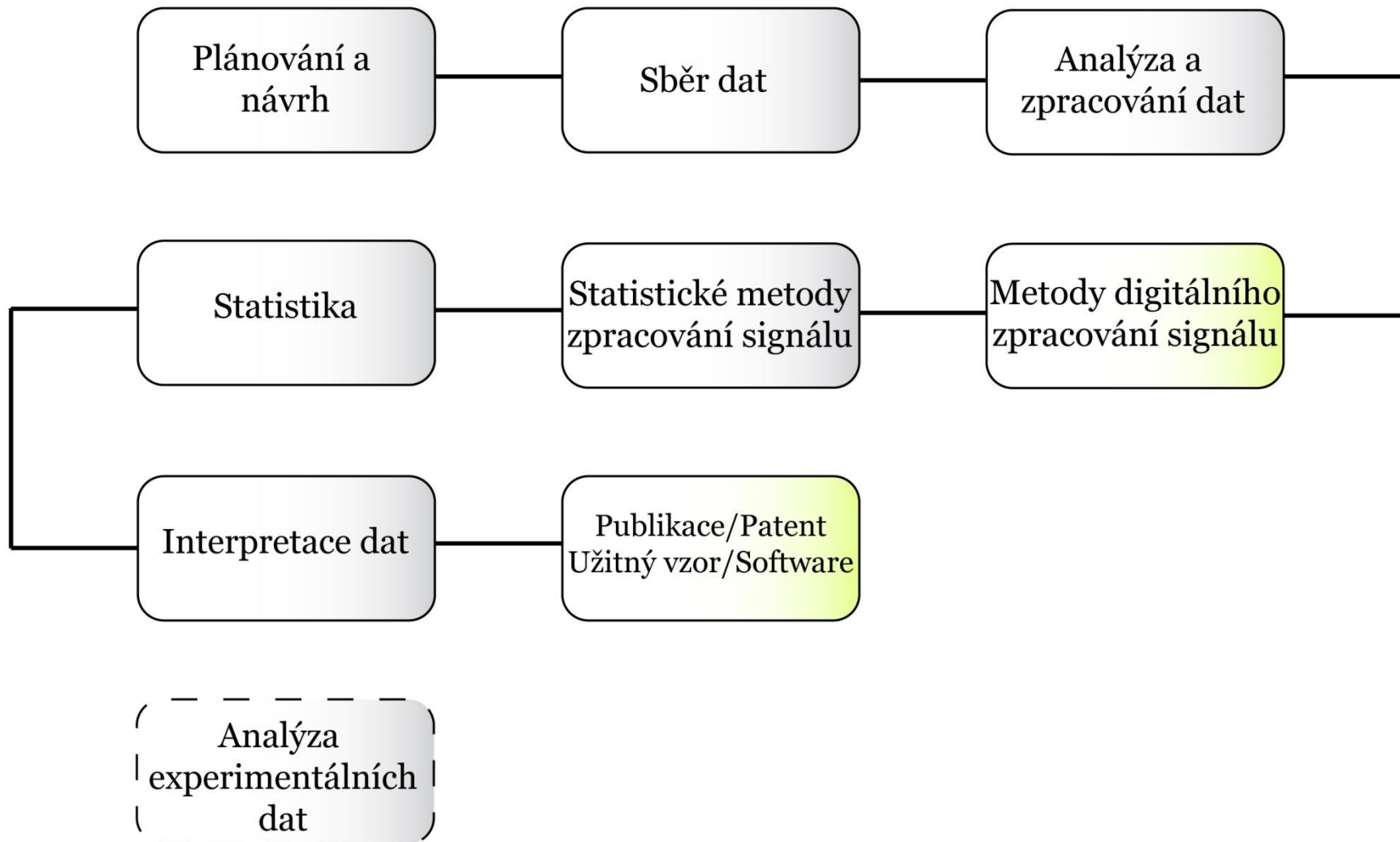
Sběr dat

- definujeme populaci, které se výzkum týká
- výhodné je naplánovat tzv. pilotní studii na malém vzorku studované populace
- ověření navržených metod pozorování, měření a zpracování údajů
- definování „inclusion/exclusion“ kritérií
- ověření kvality získaných údajů
 - vyloučení formálních chyb
 - náhodné chyby (chyby z nepozornosti)
- systematické chyby (špatné otázky, špatné nastavení přístroje)
- možnost statistického odhadu počtu vzorků nezbytných pro nalezení statistické významnosti

Analýza a zpracování dat

- předzpracování signálu, filtrace: odstranění rušivých složek signálu, odstranění stejnosměrné složky
- segmentace signálu pomocí DSP metod
 - analýza signálu v časové oblasti (prahování, obálka signálu, počet průchodů nulou...)
 - analýza signálu ve frekvenční oblasti (spektrum, autokorelace)
- segmentace signálu pomocí statistických metod (střední hodnota, medián, směrodatná odchylka, klastrování...)
- vývoj ukazatelů pro vyhodnocení měřeného jevu (momenty, délka trvání segmentu, šikmost spektra...)

Obecné schéma výzkumného/vývojového projektu



Anotace

V rámci předmětu „Analýza experimentálních dat“ se studenti naučí aplikovat základní metody statistických analýz a strojového učení pro vyhodnocení a interpretaci dat. V rámci cvičení budou studenti zpracovávat a vyhodnocovat dílčí úlohy na reálných datech z oblasti zpracování signálů v neurovědách. V rámci semestrální práce budou studenti řešit komplexní úlohu a na závěr prezentovat výsledky jejich práce. Cílem předmětu je studenty seznámit s praktickým využitím základních statistických metod a také naučit je kriticky myslet a získat dovednosti při samostatném řešení praktických úkolů.

Syllabus přednášek

1. Úvod do analýzy experimentálních dat, seznámení s daty
2. Základní statistické parametry, pravděpodobnostní distribuce, vykreslení statistických dat
3. Testování hypotéz, skupinové rozdíly, párové testy, velikost účinku
4. Korelace, testování normality dat, parametrické vs. neparametrické testy
5. Analýza variance, post-hoc testy
6. Chyby I. a II. typu, mnohočetná srovnání, odhad velikosti vzorku
7. Vícerozměrná analýza variance
8. Úvod do modelování, regresní analýza
9. Klasifikace s učitelem
10. Validace modelu
11. Klasifikace bez učitele
12. Rezerva, konzultace semestrálních prací
13. Redukce dimenzionality, interpretace dat
14. Prezentace výsledků semestrálních prací

- Anglické slidy (většinou grafické znázornění principů)
- Analýza dat je obrovský vědní obor a je zde nespočetně mnoho možností jak analyzovat data
- Poskytnou úvod do dané problematiky, který se následně ověří praktickou úlohou na cvičení
- Nejsou povinné

- Nejsou povinné, nemá je však smysl absolvovat bez přednášky
- Na cvičení se již nebude opakovat základní teorie z přednášek
- **Max 40 bodů ... cvičení v semestru (10 cvičení x 4 body = 40 bodů)**
- Jednotlivé úlohy jsou vytvořeny na bázi výsledků výzkumu skupiny SAMI
- Každý odevzdá cvičení do konce týdne (neděle 24:00 bez případných stržení bodů) přes moodle vypracované zadání, které bude bodově ohodnoceno (lze stihnout i v rámci cvičení)
- **Odevzdání 10 praktických úloh je povinné pro získání zápočtu**
- Navíc lze získat bonus za termínované bonusové úlohy bodované dle obtížnosti (dohromady až 5 bodů)

Semestrální práce

- Komplexní semestrální práce, řešeno ve dvojicích
- Na výběr bude několik typů semestrální práce
- **Odevzdání semestrální úlohy je povinné pro získání zápočtu**
- Prezentace výsledků je povinná a proběhne v posledním výukovém týdnu
- Hodnocení se zúčastní motivačně všichni studenti
- Max 35 bodů ... semestrální práce:
 - 30 bodů - kvalita zpracování a obsah odevzdané zprávy
 - 5 bodů - prezentace výsledků

Testy a zkouška

- Max 15 bodů ... 2 × kontrolní test (po 1. části a před vánoci)
- Max +-10 bodů ... zkouška
- Bonus max. 5 bodů ... posudek na semestrální práci zadaného kolegy

Celkově lze získat až 110 bodů (100 bodů + 10 bodů bonus)

- 90-100 bodů: A
- 80-99 bodů: B
- 70-89 bodů: C
- 60-79 bodů: D
- 50-69 bodů: E
- 0-49 bodů: F