

# Úvod do statistických konzultací

Tipy a triky

Michal Kulich

KPMS MFF UK

S použitím práce následujících kolegů a kolegyň:  
Scott Emerson, Don Martin, Lianne Sheppard, Paul Sampson,  
Brian Joiner, Kathleen Kerr

# Organizace

- ▶ Informace pro klienty na stránce  
<http://www.karlin.mff.cuni.cz/~kpms/?vyber=konzultace>
- ▶ Informace pro studenty na stránce  
<http://www.karlin.mff.cuni.cz/~kulich/vyuka/consult>
- ▶ Každý týden 8 volných termínů konzultačních schůzek,  
zaplňujeme 2–3 z nich

# Organizace

## Postup ze strany klienta

- ▶ Najde volný termín v kalendáři
- ▶ Vyplní zadávací formulář
- ▶ Zašle na KPMS mailem
  - ▶ vybraný termín
  - ▶ vyplněný zadávací formulář
  - ▶ přílohy (ukázky datového souboru, formuláře pro sběr dat, výstupy z předchozích analýz, články vztahující se k tématu apod.)

# Organizace

## Rozvrhování schůzek

- ▶ Organizátor konzultací posoudí vhodnost tématu a
  - ▶ potvrdí termín
  - ▶ stanoví dohlízejícího pedagoga
  - ▶ přiřadí dvojici studentů (hlavní a pomocný konzultant)
  - ▶ najde místnost
- ▶ KPMS (paní Naxerová) oznámí termín a místo konzultace klientovi a studentům
- ▶ Dohlízející pedagog a studenti dostanou zadávací formulář a přílohy

# Organizace

## Jak se připravit na schůzku

- ▶ Prostudovat zadávací formulář
- ▶ Pokud možno si ujasnit o jaký statistický problém jde
- ▶ Najít si potřebné informace jak statistické tak z protilehlého oboru (poznámky z přednášek, učebnice/skripta, monografie, články, Wikipedie, Google, . . . )
- ▶ Cítíte-li potřebu se klienta ještě před schůzkou na něco důležitého vyptat nebo si vyžádat doplňující materiály, učiňte tak (máte kontakty)

# Organizace

## Průběh schůzky

Schůzku řídí hlavní konzultant (zpočátku možná dohlížející pedagog)

- ▶ Přivítejte klienta, představte se a seznámte se s ním
- ▶ Nastiňte, jak bude schůzka probíhat
- ▶ Dejte klientovi slovo, aby uvedl a vysvětlil svůj problém
- ▶ **Dělejte si poznámky (!!)**
- ▶ Vhodně zvolenými otázkami a diskusí s klientem se snažte zjistit podstatu problému a okolnosti potřebné pro jeho řešení
- ▶ Snažte se najít řešení nebo alespoň částečně užitečnou radu
- ▶ Je-li zřejmé, že by pomohlo sjednání další schůzky (ve stejném složení), sjednejte ji s klientem ihned.
- ▶ Shrňte výsledek schůzky a rozlučte se s klientem

# Organizace

## Po schůzce

- ▶ Hlavní konzultant připraví zprávu o schůzce (1–2 stránky)
- ▶ Termín vypracování zprávy: ideálně druhý den, nejpozději do 2–3 dnů
- ▶ Zprávu o schůzce hlavní konzultant pošle klientovi, organizátorovi konzultací a všem účastníkům schůzky
- ▶ Včas a uspojivě zpracovaná zpráva je zásadní

# Průběh konzultace

## Stanovení cíle

Nejdůležitější a nejtěžší je pochopit, **co vlastně klient chce**

- ▶ K čemu potřebuje řešení svého problému?
- ▶ Jaký význam má klientův problém v kontextu jeho oboru?
- ▶ Co přesně je cílem jeho výzkumu?
- ▶ K čemu budou použity jeho výsledky?



# Průběh konzultace

## Stanovení cíle

- ▶ Snažte se odlišit důležité informace od nepodstatných
- ▶ Pokud klient nemá jasno a zmiňuje více různých hledisek (obvykle nekompatibilních), snažte se rozpoznat to hlavní a určit sobě i jemu hlavní cíl a účel konzultace i jeho výzkumu
- ▶ Jde o iterativní proces
- ▶ Používejte otázky a obraty jako
  - ▶ Povězte nám, co vlastně děláte?
  - ▶ Proč na tom záleží?
  - ▶ Jestli tomu dobře rozumím, jde vám o to, že...
- ▶ Nebojte se předstírat naivitu a ptát se i na věci, které jsou vám (zdánlivě) jasné

# Průběh konzultace

## Stanovení cíle

- ▶ Prvotní je zjistit, zdali vůbec jde o statistický problém: Je možné klientův problém přeformulovat do pravděpodobnostního modelu a statistické úlohy nebo ne?
- ▶ Pokud je statistická formulace problému možná, o jaký typ problému jde?
  - ▶ odhadování parametrů
  - ▶ testování hypotéz
  - ▶ predikce
  - ▶ optimalizace
  - ▶ něco jiného?
- ▶ Nepouštějte se do žádných statistických úvah, dokud nechápete aspoň zhruba smysl, cíl, význam a důležitost klientova problému v celém jeho kontextu

# Průběh konzultace

## Znalost provedení experimentu

Pokud klient potřebuje pomoci s analýzou nebo interpretací hotových dat, je nezbytné zjistit všechny relevantní informace o experimentu, z něhož data pocházejí

- ▶ Jaké veličiny jsou k dispozici a jaký je jejich význam pro řešený problém? Která veličina je odezva, které jsou prediktory?
- ▶ Kolik pozorování je k dispozici?
- ▶ Jaká je struktura dat? Lze ji načrtnout na papír?
- ▶ Jde o nezávislá pozorování? Není v datech nějaká korelační struktura (časové řady, skupinově korelovaná pozorování, opakovaná měření na stejném jedinci apod.)
- ▶ Byla v experimentu provedena randomizace?

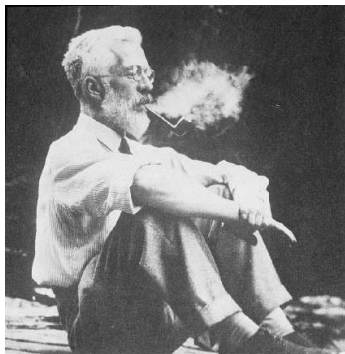
# Průběh konzultace

## Znalost provedení experimentu

- ▶ Jakým způsobem byli vybráni jedinci (experimentální jednotky), kteří se v datech nacházejí? Jak byl určen rozsah výběru? Jak byly určeny experimentální podmínky jednotlivých pozorování?
- ▶ Byl experiment proveden podle jasně daného písemného postupu? (Ano? Sem s ním!)
- ▶ Jak probíhal sběr a zpracování dat? Byly zavedeny nějaké kontrolní mechanismy zajišťující správnost dat?
- ▶ Mohou být zaznamenaná měření nějakým způsobem neúplná nebo zkreslená?
- ▶ Jsou v datech chybějící pozorování? Jak vznikla?

Nepouštějte se do žádných statistických úvah, dokud nevíte, co přesně data obsahují a jakým způsobem byla sesbírána!!

## Průběh konzultace



*“To call in the statistician after the experiment is done may be no more than asking him to perform a postmortem examination: he may be able to say what the experiment died of.”*

*R. A. Fisher (1890-1962)*

# Průběh konzultace

## Komunikace s klientem

- ▶ Většina klientů nerozumí ani základním statistickým pojmům a definicím
- ▶ Terminologii je třeba překládat do pojmů, které jsou pro klienta srozumitelné a intuitivní, byť na úkor ztráty přesnosti vyjadřování
- ▶ Zásadní je, aby konzultant rozuměl principům statistických metod a uměl je vysvětlit (konstanty vs. náhodné veličiny, parametr vs. jeho odhad, testování hypotéz, pojem *p-hodnota*, interval spolehlivosti)

# Průběh konzultace

## Komunikace s klientem

- ▶ Komunikace vyžaduje vzájemný respekt a vyváženost: ani arogance ani podřazenost
- ▶ Situace klienta a konzultanta je symetrická; ani jeden se nevyzná v oboru toho druhého
- ▶ Konzultant musí mít zdravou sebedůvěru; neexistuje statistik, který by znal všechno
- ▶ Nenechte se vmanipulovat do řešení, která se vám nelíbí

# Průběh konzultace

## Komunikace s klientem

- ▶ Složitý problém se snažte zjednodušit, rozložit na posloupnost dílčích snazších problémů
- ▶ Vždy hledejte co nejjednodušší přijatelnou cestu
- ▶ Řešení musí být v silách klienta — lepší je jednodušší snáze pochopitelná metoda než složité nevysvětlitelné řešení
- ▶ Buďte kreativní, nesnažte se napasovat problém na převybranou metodu nebo přesvědčit klienta, aby změnil cíle svého výzkumu



# Průběh konzultace

## Časté chytky

- ▶ Rozdíl mezi exploratorní a konfirmatorní analýzou
  - ▶ **Exploratorní:** Generování vědeckých hypotéz, hledání souvislostí a závislostí. Často stačí průzkumová analýza, netřeba klást důraz na formálně správné testování hypotéz.
  - ▶ **Konfirmatorní:** Hledání správné odpovědi na předem jasně formulovanou otázku. Nutno mít dobře provedený experiment se spolehlivými daty a čistě provedenou robustní analýzu.
- ▶ Problém mnohonásobných porovnávání
- ▶ Co je prediktor a co je odezva?
- ▶ Chci prokázat, že obě ošetření dávají stejné výsledky
- ▶ Správná interpretace p-hodnot
- ▶ Statistická versus praktická signifikance

# Průběh konzultace

*“The statistician must be instinctively and primarily a logician and a scientist in the broader sense, and only secondarily a user of the specialized statistical techniques.*

*The statistician who knows only statistics is a danger and a blight. It is less important to know how to use statistics, than it is to know how and when **not** to use it.”*

*Malcolm Rorty, 92nd Annual Meeting of ASA, 1930*

# Průběh konzultace

## Uzavření schůzky

- ▶ Shrňte klíčové body
  - ▶ Specifikace problému a řešených otázek
  - ▶ Důležité aspekty studie a dat
  - ▶ Učiněná rozhodnutí, navržená řešení
  - ▶ Další kroky
- ▶ Zjistěte, jestli klient souhlasí se specifikací problému a chápe doporučení
- ▶ Přesvědčte se, že nebyly přehlédnuty některé důležité aspekty

# Po konzultaci

## Zpráva o schůzce

- ▶ Forma: dopis klientovi
- ▶ Kdy schůzka proběhla, kdo se jí účastnil, kdo sepsal zprávu
- ▶ Struktura zprávy:
  - ▶ Popis problému, specifikace jeho cílů
  - ▶ Klíčové informace o klientově studii a datech
  - ▶ Otázky, které byly během schůzky diskutovány, navržené postupy a řešení
  - ▶ Závěrečné poznámky
- ▶ Zprávu pište formou, které klient bude rozumět
- ▶ Buďte diplomatičtí, vyhýbejte se zevšeobecnujícím negativním formulacím (data byla shledána nepoužitelnými, experiment byl proveden nekvalitně, položená otázka nedávala smysl)