

# Úvod a motivace

Statistika 1, Katedra sociologie Filozofické fakulty UK

Jaromír Mazák

# Statistika je když...

Efektivita vakcíny Pfizer–BioNTech je 73 %.

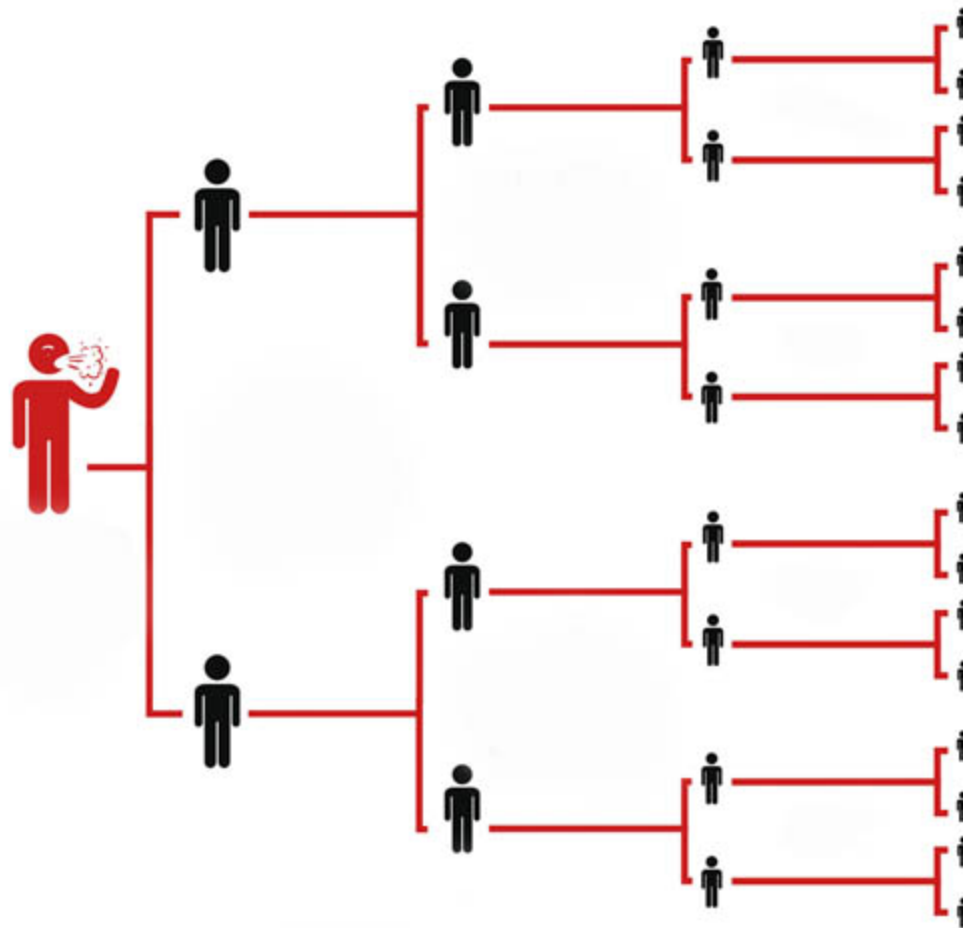
11,5 % lidí v Česku žije v chudobě.

# Statistika a covid

# Základní reprodukční číslo

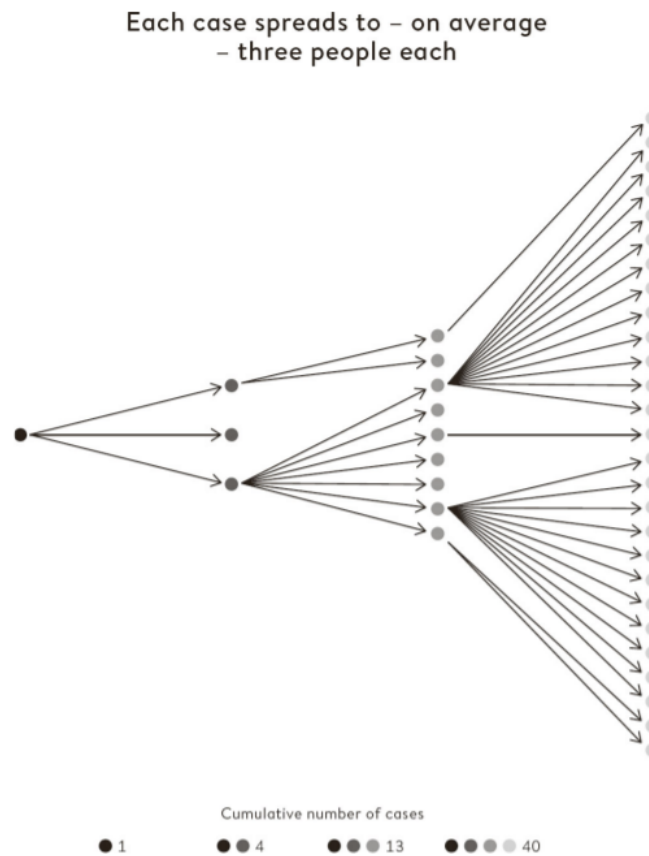
$$R_0$$

# Základní představa pro $R_0 = 2$



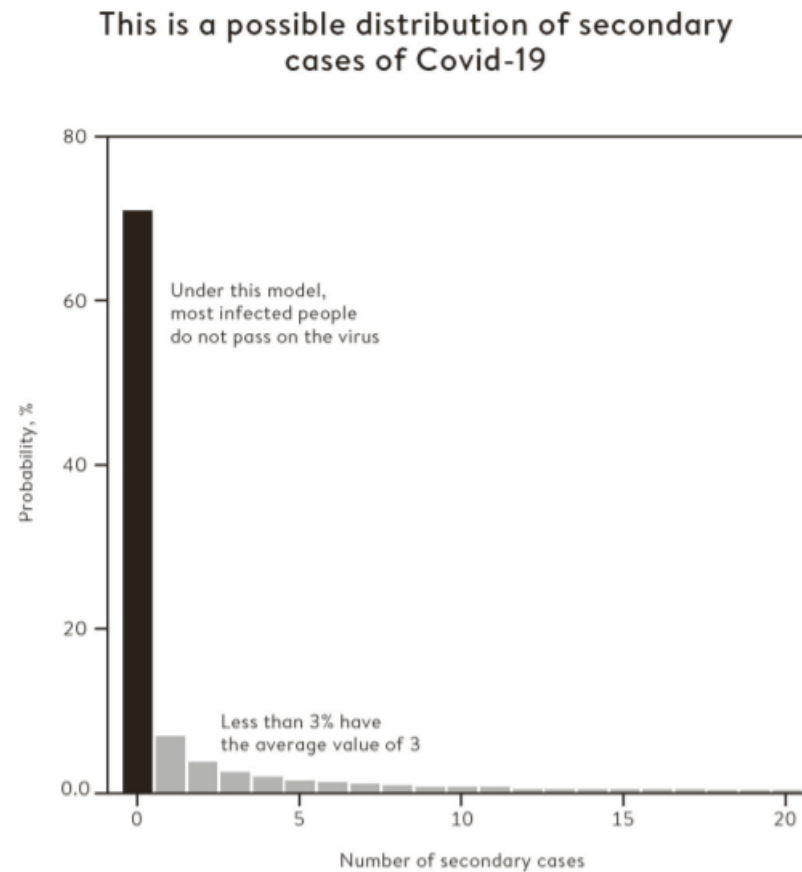
[Zdroj obrázku Michigan Health Lab](#)

# Blíže skutečnosti (pro $R_0 = 3$ )



Převzato z Spiegelhalter and Masters (2021)

# Odhad rozložení počtu nakažených



Source: Estimating the overdispersion in Covid-19 transmission using outbreak sizes outside China (Wellcome Open Research, 2020)

Převzato z Spiegelhalter and Masters (2021)

# Efektivní reprodukční číslo

$$R_t$$

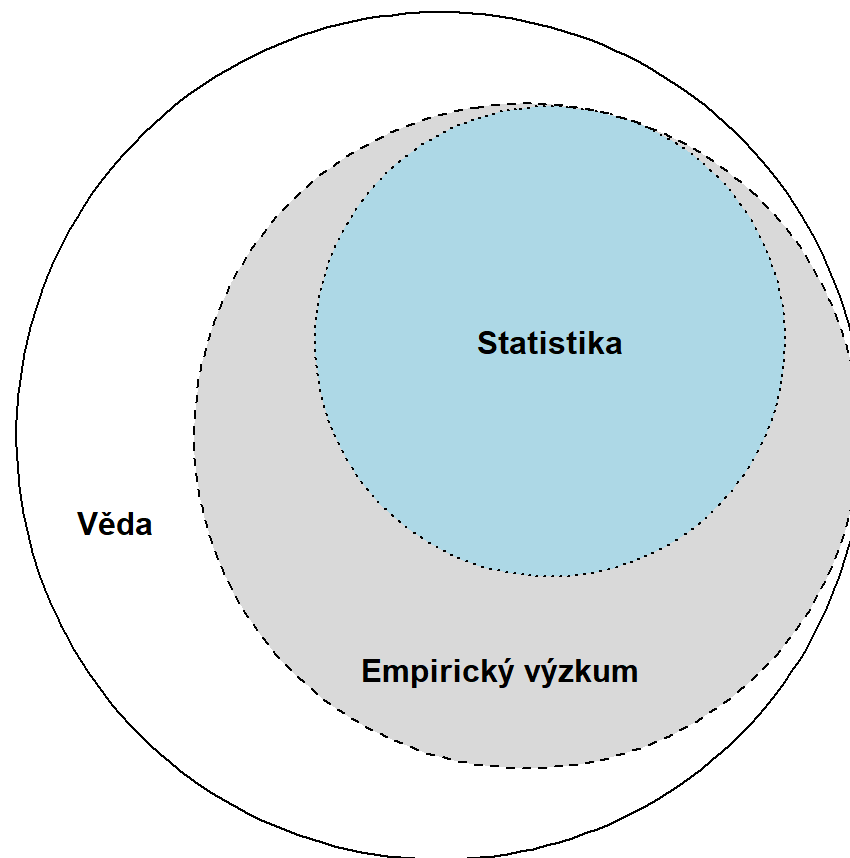


**Statistika jako investigativní proces**

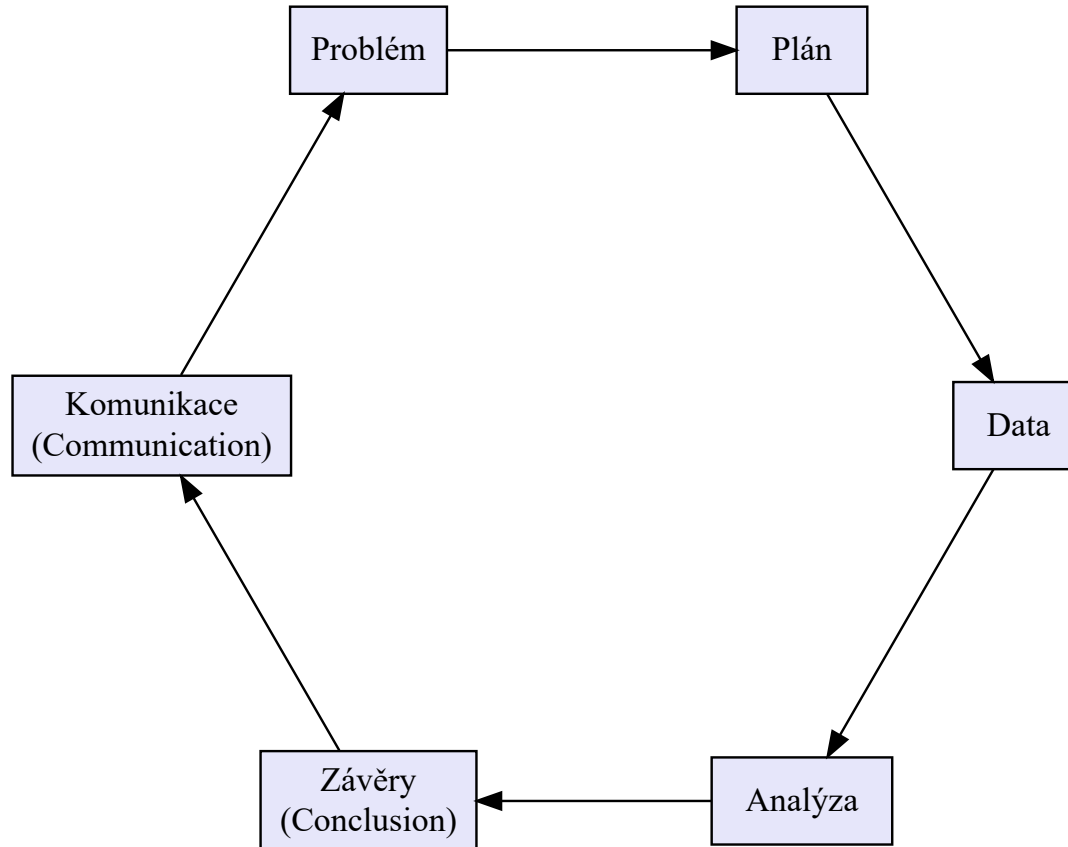
# Statistika je...

... základ pro kvantitativní empirický výzkum, zaměřená na porozumění světu, vytváření vědění a opory v rozhodování. Aby tohle zvládla, nestačí chroustat čísla, je potřeba skutečně rozumět tomu, co znamenají.

# Statistika v kontextu vědy



# PPDACC cyklus



PPDAC cyklus vymysleli MacKay and Oldford (2000) Explicitně na PPDACC ho rozšiřuje Spiegelhalter (2019)

# Problém

- Problém zpravidla definujeme skrze výzkumnou otázku.
- Objevila se nová infekce. Jak moc se šíří mezi lidmi?
- Vymyslete příklad relevantní výzkumné otázky pro sociologii.

# Plán

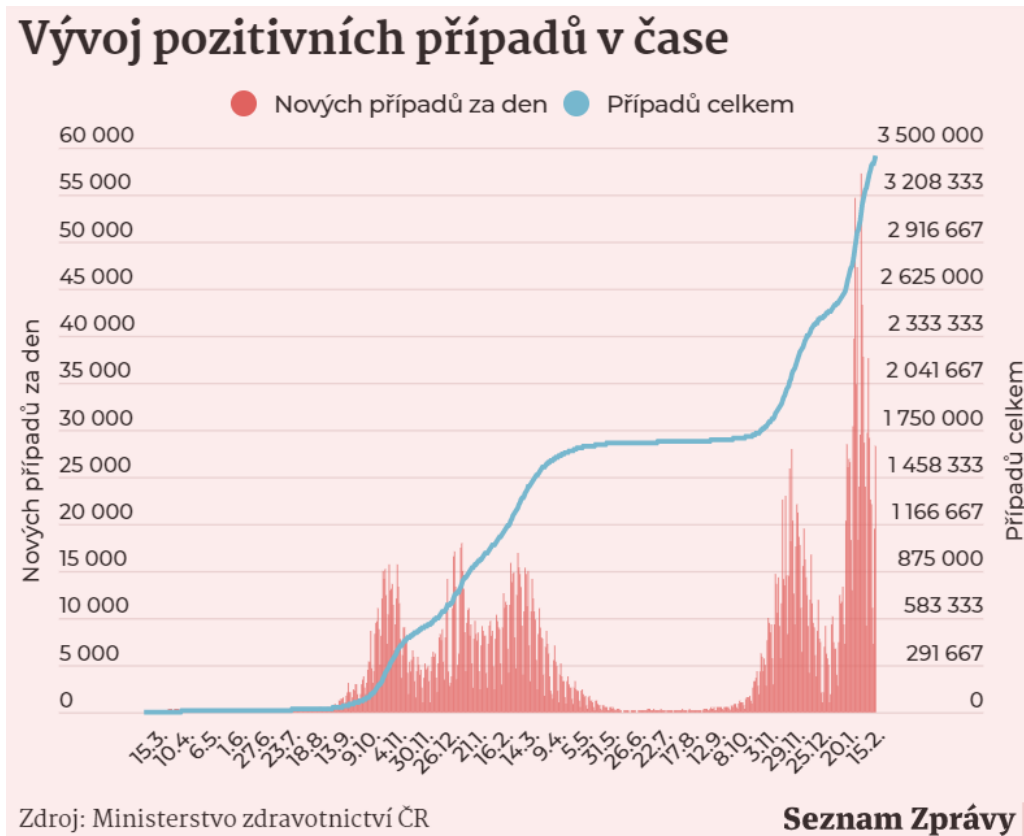
- Co měřit a jak? Jak data shromažďovat na jednom místě? Jaká relevantní data už existují? Jak se k nim dostat? Atd.
- Vytvořit datovou oporu: všechna relevantní pracoviště sdílejí informace o nových infekcích v přesně stanoveném formátu podle přesně stanovených definic.

# Data

- Sběr dat, uchovávání, získání, čištění.
- Reálná data jsou téměř vždy zatížena chybami a zkresleními - potřebujeme reflektovat a zaznamenat. Musíme co nejpřesněji rozumět tomu, co data znamenají, v čem se na ně můžeme spolehnout a v čem nikoliv.
- Jaké jsou vlastnosti testů, které ověřují, že je někdo nakažený? Může záležet na typu testu, jeho administraci atp. Zajímá nás míra falešné positivity a falešné negativity. Navíc zpočátku třeba testy nebyly k dispozici a diagnóza probíhala na základě odborného posouzení bez nich.
- S postupem času je potřeba zaznamenávat nové typy dat, například reinfekce. Nebo jestli byl infikovaný člověk očkovaný.

# Analýza

- Zpracovávání dat v softwaru, tabulky, grafy, testování hypotéz,...



[Zdroj Seznam zprávy](#)



# Závěry/Komunikace

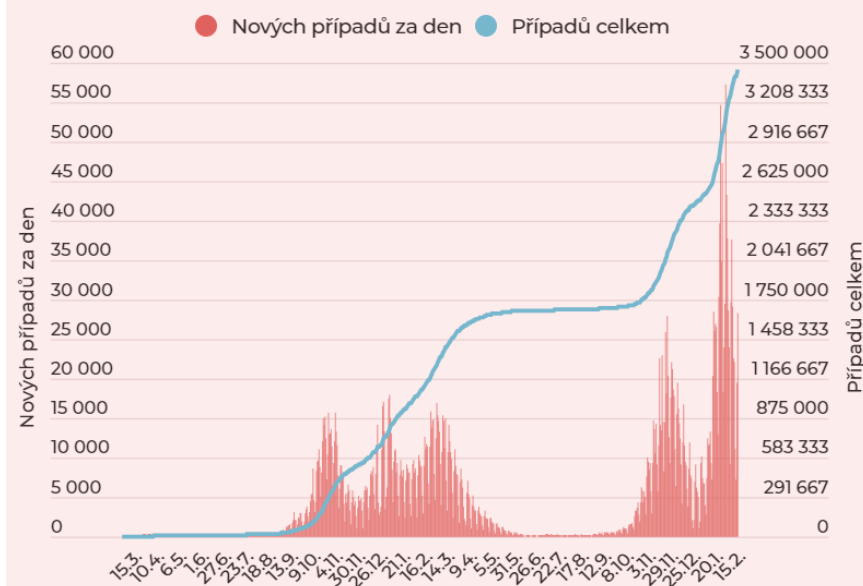
Čísla nemají možnost mluvit sama za sebe. My mluvíme za ně. My jim dáváme smysl.

—Nate Silver, Signál a šum (2015)

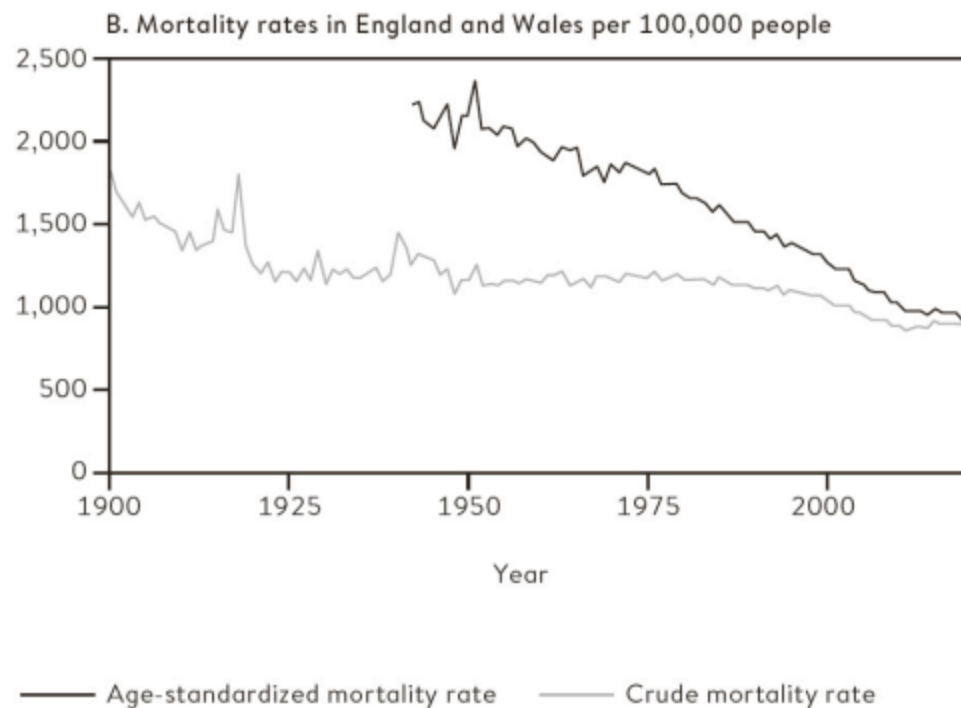
- Interpretace, generování nových nápadů, komunikace různým skupinám.

# Závěry/Komunikace - ilustrace 1

## Vývoj pozitivních případů v čase

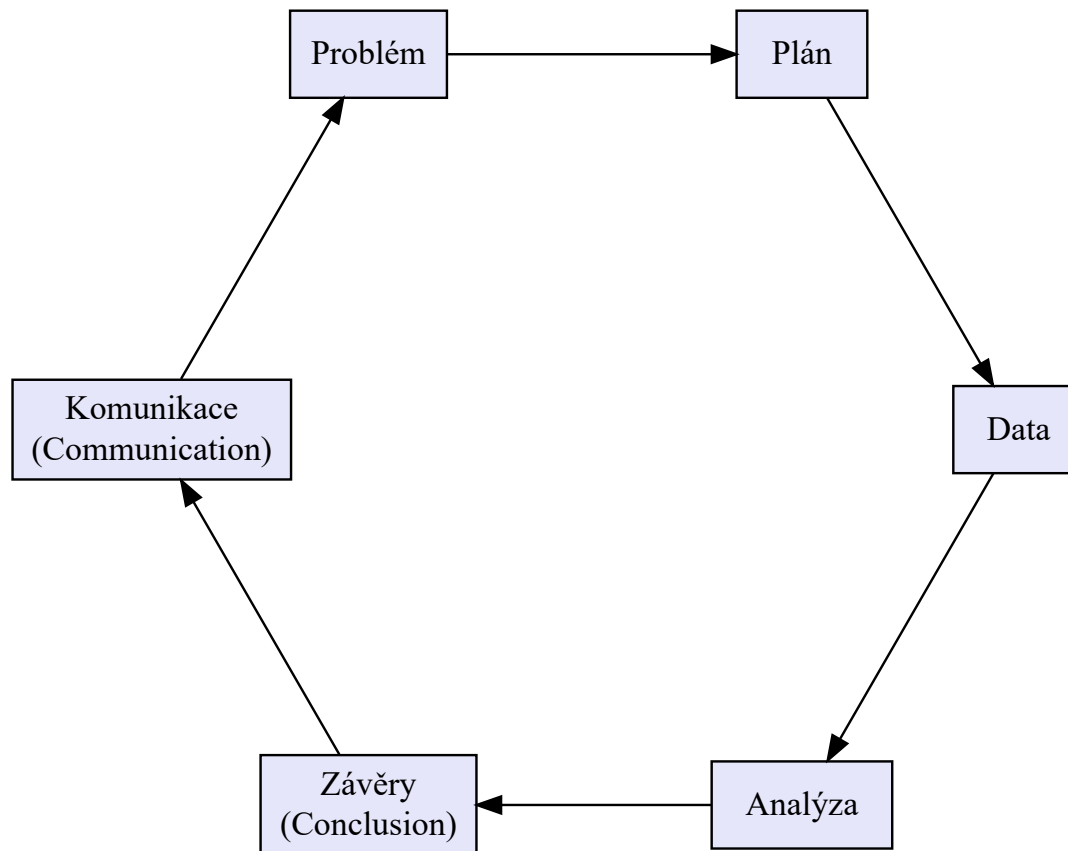


# Závěry/Komunikace - ilustrace 2



Obrázek z Spiegelhalter and Masters (2021) s odvoláním na ONS, končí rokem 2020.

# PPDACC cyklus - shrnutí



# Studium statistiky na Katedře sociologie FF UK

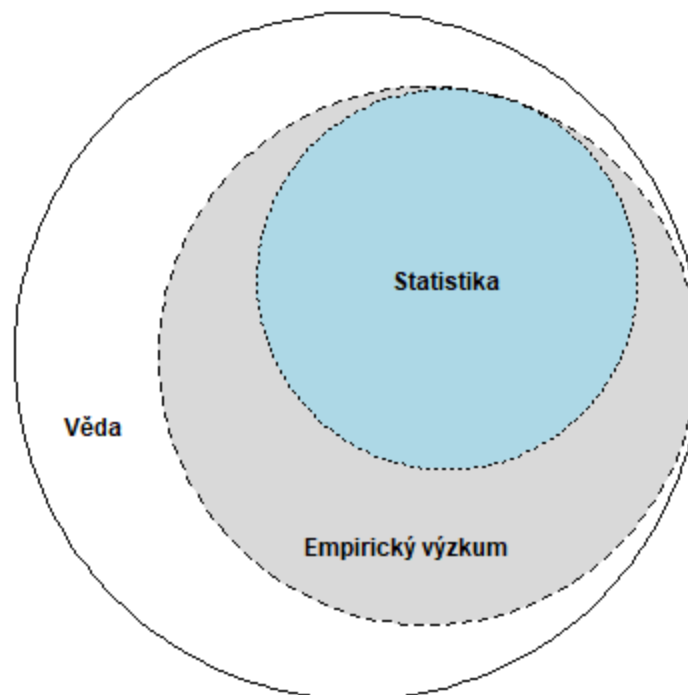
# Statistika a empirický výzkum u nás

## Statistika

- Statistika 1 + 2 (Bc.)
- Zpracování scg. dat (Bc.)
- Analýza dat v SPSS 1 + 2 (Mgr.)
- Úvod do vícerozměrné an. dat (Mgr.)
- Pokročilé stat. met. (Mgr.)
- Volitelné kurzy (hlavně v R)

## Empirický výzkum

- Scg. výzkum 1, 2, 3, 4 (Bc.)



**Statistika je proces, ale má různé cíle...**

# Zmátení pojmů: cíle vědy a cíle statistiky

## Cíle vědy (hierarchie)

- Deskripce - popis světa a jeho pravidelností
- Vysvětlení (explanace) - kauzální mechanismus, kauzální inference (umožňuje predikci a kontrafaktuální uvažování, tj. kreativitu)

Rozdíl mezi deskripcí a explanací ilustruje rozdíl mezi fakty a teorií.

## Cíle statistiky

- Deskripce dat (deskriptivní statistika)
- Inference z dat (vzorku) na populaci (inferenční, resp. induktivní statistika)

Z hlediska cílů vědy plní deskriptivní statistika a induktivní statistika stejnou funkci: deskripci. Vysvětlení nikdy nevychází jen ze statistiky, vždy potřebuje teorii.



# Teorie a fakta (data)

“Některé druhy gazel, když spatří predátora, vyskakují do výšky.”

“Chování gazel vysvětluje tzv. signalizační teorie, podle které predátorovi ukazuje, že má spoustu energie a bude obtížné ji chytit (vyhne se tím výdeji při nutnosti útěku).”



Autor foto: [Yathin sk - I photographed this springbok in Etosha National Park,](#)

CC BY-SA 3.0

A variable

**Proměnná: základní stavební kámen  
statistického uvažování**

# Proměnná

Proměnná: formální reprezentace určité vlastnosti či charakteristiky entit (věcí, osob, ...).

Statistika: zkoumání vlastností proměnných a vztahů mezi nimi.

# Příklad standardního zápisu proměnných v sociologii

ID	pohlaví	věk	postoj k placení školného
1	žena	22	rozhodně souhlasí
2	žena	48	spíše nesouhlasí
3	muž	35	neví/nechce odpovědět

# Vidět svět jako data

Prvním krokem statistického vyšetřování je překlopit svět na data.

- Kolik stromů je na světě?
- Jaká je úmrtnost novorozenců?

# Aplikovaná statistika potřebuje definice pojmů

Představte si, že se díváte na dvě ovce na poli. Kolik ovcí tam je? Dvě, samozřejmě. Až na to, že jedna z ovcí není ovce, je to jehně. A ta druhá ovce je těžce březí – vlastně rodí, každou chvíli porodí. Kolik že to bylo ovcí? Jedna? Dvě? Dvě a půl?

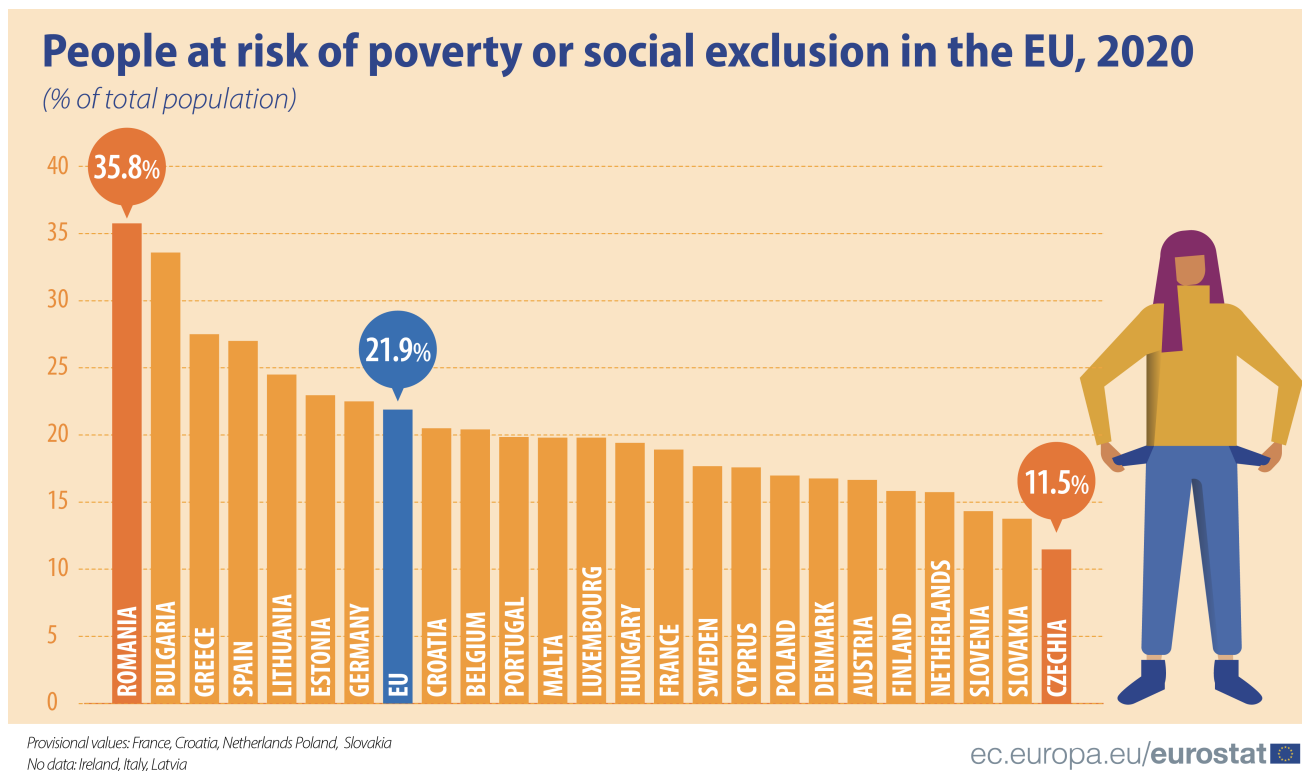
—[Tim Harford, How to make the World add up](#)

Tim Harford: “Avoid premature enumeration”

# Sociologický příklad: Ohrožení chudobou

Otázka: Máme v ČR ve srovnání se zbytkem EU hodně, nebo málo lidí ohrožených chudobou?

# Ohrožení chudobou: data



[Zdroj Eurostat](#)

- Jak byste nízkou míru ohrožení chudobou u nás vysvětlili?



# Co je ohrožení chudobou a sociálním vyloučením?

At risk of poverty or social exclusion, abbreviated as AROPE, corresponds to the sum of persons who are either at risk of poverty, or severely materially and socially deprived or living in a household with a very low work intensity.

[Eurostat](#)

- Ale co to znamená?

# Psychologický příklad: Násilné videohry

Otázka: Způsobuje hraní násilných počítačových her násilné chování u dětí?

# Operacionalizace

Překlopení pojmů z jazyka teorie do observačního jazyka.

# Reference v prezentaci

Následuje seznam referencí, které v prezentaci neměly aktivní proklik na zdroj.

MacKay, R. J., and R. W. Oldford. 2000. "Scientific Method, Statistical Method and the Speed of Light." *Statistical Science* 15 (3): 254–78. <https://doi.org/10.1214/ss/1009212817>.

Silver, Nate. 2015. *The Signal and the Noise: Why so Many Predictions Fail—but Some Don't*. Illustrated edition. New York, NY: Penguin Books.

Spiegelhalter, David. 2019. *The Art of Statistics: Learning from Data*. UK USA Canada Ireland Australia India New Zealand South Africa: Pelican.

Spiegelhalter, David, and Anthony Masters. 2021. *Covid by Numbers: Making Sense of the Pandemic with Data*. Pelican.