

Logika testování hypotéz

David Lacko

Vědecké poznání

- Zdroje poznání (Charles Peirce)

- Metoda tradice
- Metoda authority
- Metoda a priori
- **Metoda vědy**

- **Vědecké poznání**

- Empirické a racionální poznání
- **Induktivní** a deduktivní myšlení
- *„Je předmětné, systematické, empirické, kritické, kontrolovatelné, reprodukovatelné a sociálně a kulturně podmíněné.“*

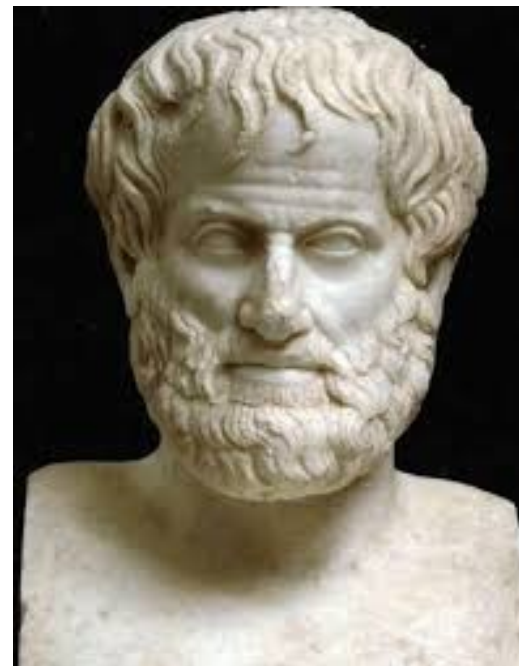


Dedukce vs. indukce

- Dedukce (top-down logic)
 - Matematika, logika, „jistota“
 - z obecného usuzuji na jednotlivé
- Indukce (bottom-up logic)
 - Hypotézy, empirie, „nejistota“ (pravděpodobnost)
 - od jednotlivého k obecnému (generalizace)
 - *„Proces odvozování obecných zákonů či principů z pozorování specifických instancí“*
- Zabývá se nimi epistemologie i statistika
- Indukce tvoří nové vědomosti, dedukce pouze přeuspořádává to, co už víme

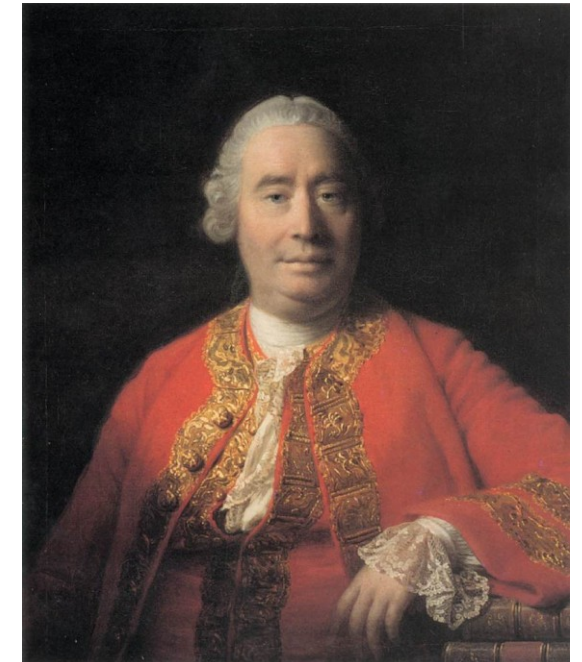
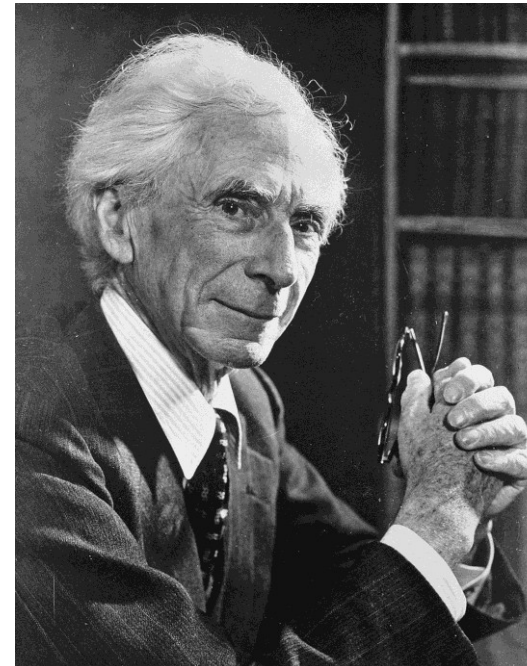
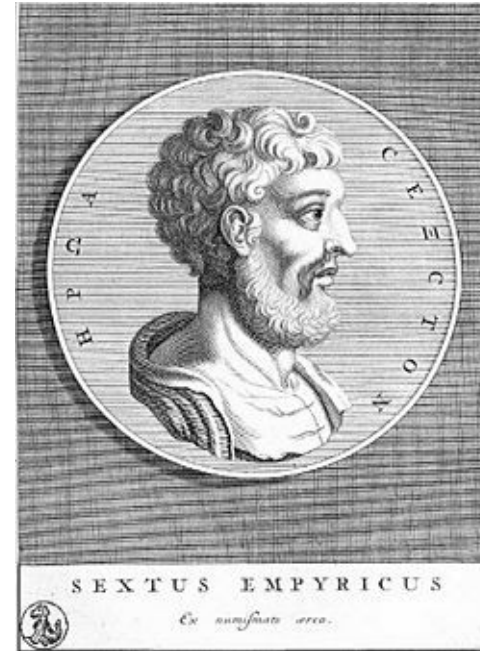
Induktivní přístup

- Aristoteles - Sokratův pojem *epagogé* (Cicerův překlad – *inductio*)
- Pozitivismus, novotopozitivismus, induktivismus, empirismus
- Důraz na **verifikaci**
- Opakované dokazování je základem vědy
- A. Comte, R. Carnap (Vídeňský kruh), a další...



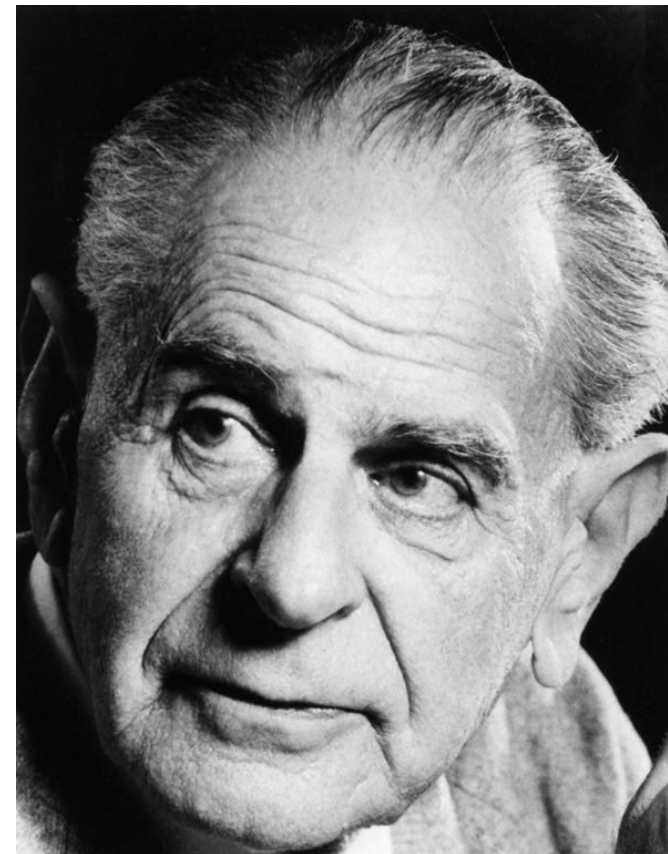
Problém indukce

- Sextus Empiricus
- Francis Bacon
 - Východ slunce (to, že zatím každý den vyšlo slunce neznamená, že vyjde i zítra)
- **David Hume – Problém indukce**
 - *A Treatise of Human Nature*
 - *An Enquiry concerning Human Understanding*
- Bertrand Russell
 - Krocán-vědec (to, že mě každý den krmí neznamená, že mě jednou nesní)
- -> Verifikace nestačí



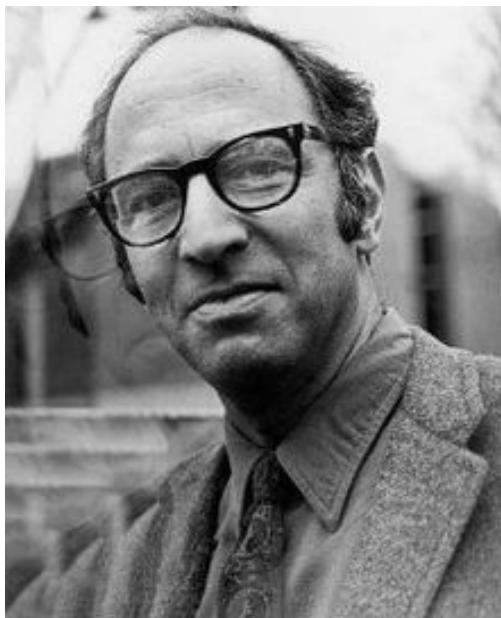
Snaha o „vyřešení“ problému indukce

- Karl Popper – Kritický racionalismus (hypoteticko-deduktivní metoda)
 - Černá labuť (můžeme pozorovat nekonečně mnoho bílých labutí, neznamená to ale, že existují jenom bílé labutě. Naopak, stačí pozorovat jednu černou labuť a víme jistě, že není pravda, že by existovali jenom bílé labutě.
- *Logik der Forschung*:
 - 1. Indukce neexistuje.
 - 2. Hypotézy lze jen **falsifikovat**, nikoli verifikovat - verifikace neexistuje.
 - 3. Demarkační kritérium - vědecká je jen falsifikovatelná teorie.
 - 4. Smyslem vědy je falsifikace.



Následovatelé

- Thomas S. Kuhn - Struktura vědeckých revolucí
- Imre Lakatos – Kritický racionalismus s principem indukce
- Paul Feyerabend - Metodologický anarchismus
- Post-modernismus ... (viz Sokalova aféra)



Alternativní „řešení“

- Immanuel Kant – Racionalismus
- David Papineau – Reliabilismus
- Thomas Bayes – Bayesovská inference
- A mnoho dalších



Statistický pohled

- Ronald Fisher
- Jerzy Neyman and Egon Pearson
- *„Induction is the glory of science and the scandal of philosophy“* – C. D. Broad
- *„Popper supplied the philosophy, and Fisher, Neyman and colleagues supplied the statistics“*
– Queen, Quinn & Keough



Testování hypotéz

- Co je to hypotéza?
 - Z řeckého *hypóthesis* – předpoklad, domněnka.
 - Má podobu tvrzení o vztahu či souvislosti dvou či více proměnných.
 - Je psána výrokovou formou, je velice konkrétní.
 - Jednoduchost, verifikovatelnost, falsifikovatelnost, srozumitelnost.
 - X výzkumná otázka
- Co je to testování hypotéz?
 - Proces rozhodování o tom, jestli přijmeme, nebo zamítneme hypotézu.
 - Podobně jako indukce, snaha o generalizaci výsledku získaného ze vzorku na populaci.

Testování hypotéz

- Nulová hypotéza (H_0)
 - Ta, kterou se snažíme vyvrátit (falsifikovat)
 - Snaha vyvrátit nulovou a potvrdit tak alternativní
 - Je opakem výzkumné hypotézy
 - Většinou předpokládá neexistenci jevu, fenoménu či efektu, nepřepokládá děj, souvislosti
 - Např. *neexistuje statisticky významný rozdíl mezi skupinami*
 - Occamova bříta
 - V textu se většinou neuvádí, ale autor si jí musí být vědom

Testování hypotéz

- Alternativní hypotéza ($H_1, H_2, H_{2.1}, \dots H_n$)
 - Naše výzkumná hypotéza
 - Zdůrazňuje vztah, efekt, jev a jejich existenci.
 - Oboustranná
 - Např. *existuje statisticky významný rozdíl váhy mezi muži a ženami*
 - $H_0: M = \check{Z}$, $H_1: M \neq \check{Z}$
 - Jednostranná
 - Např. *Ženy mají statisticky významně nižší průměrnou váhu než muži*
 - $H_0: M \leq \check{Z}$, $H_1: M > \check{Z}$

Časté chyby v tvorbě hypotéz

- 1) Chybí vztah mezi dvěma a více jevy
 - Např. *Muži jsou inteligentní.*
- 2) Není ověřitelným výrokem
 - Např. *Pokud se objeví šikana, žáci ji nebudou umět zvládat.*
- 3) Dopředu odhaduje výsledek
 - Např. *s drogami se setkalo 90 % studentů VŠ.*
- 4) Vysvětluje výsledky výzkumu
 - Např. *Inteligentnější žáci užívají častěji drogy, protože si je umí snáze sehnat.*
- 5) Samozřejmá tvrzení
 - Např. *Bohatší rodiny si mohou dovolit dražší vybavení.*
- 6) Zbožná přání, apely na morálku, neoznamovací formát
 - Např. *Všichni lidé by měly odmítnout vraždu.*

Fisherův test signifikance

- 1) Výběr vhodného testu
- 2) Nastavení nulové (H_0 , null ne nutně nil) hypotézy
- 3) Spočítání teoretické pravděpodobnosti výsledku pod H_0
- 4) Posouzení statistické významnosti výsledku
 - P-hodnota $>$ nebo $<$ než stanovená hladina významnosti
 - α – a posteriori
- 5) Interpretace statistické významnosti výsledku

Fisherův test signifikance

- Pravděpodobnost výsledků pod nulovou hypotézou
 - Nesignifikantní výsledky mohou být ignorovány
 - P-hodnota se vypisuje v přesné podobě (např. $p = 0.049$ nebo $p = 0.051$)
- + Je flexibilní (a posteriori), vhodný pro explorativní výzkumy
- chybí power analysis i alternativní hypotéza

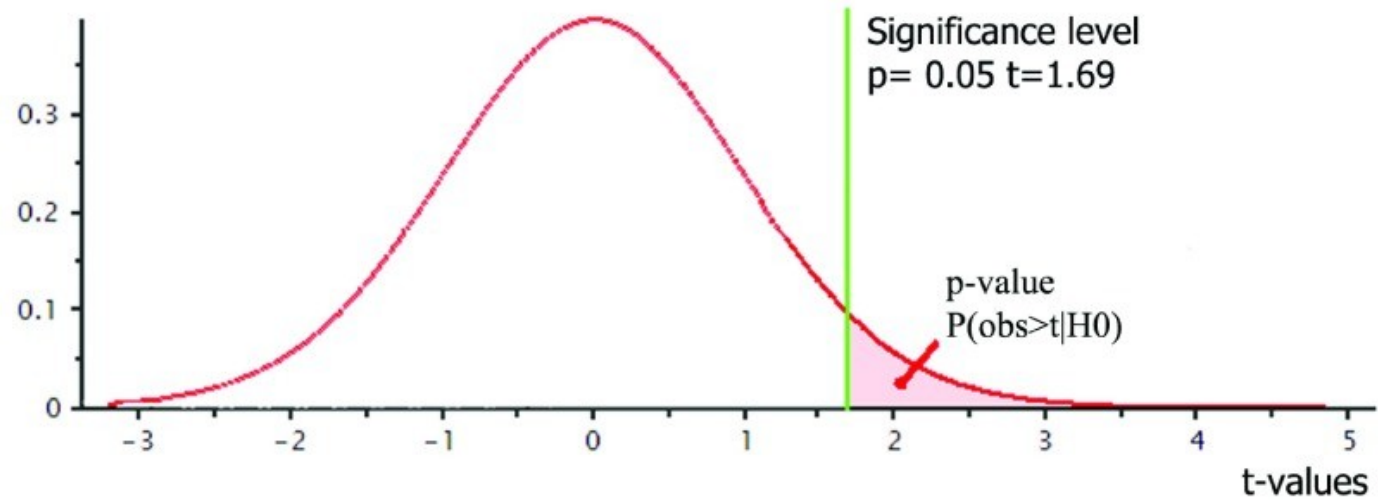
Neyman-Pearsonův test akceptace

- 1) Nastavení očekávané velikosti účinku v populaci (power analysis)
- 2) Výběr vhodného (nejsilnějšího) testu
- 3) Nastavení hlavní hypotézy (H_M) – de facto H_0
 - Chyba I. řádu, α – a priori
- 4) Nastavení alternativní hypotézy (H_A)
 - Chyba II. Řádu, β
- 5) Spočítání velikosti vzorku (N) vyžadovaného pro dostatečnou sílu testu ($1-\beta$)
- 6) Spočítání kritické hodnoty testu
- 7) Spočítání testové hodnoty pro výzkum (likelihood-ratio test)
- 8) Rozhodnutí ve prospěch hlavní nebo alternativní hypotézy

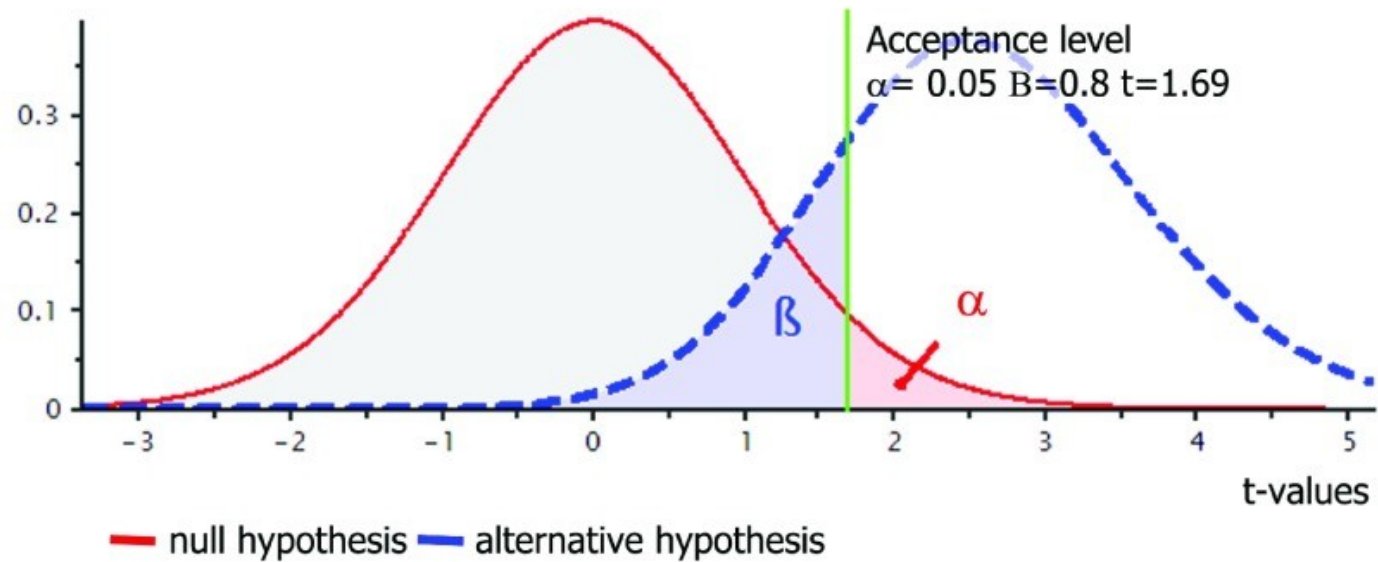
Neyman-Pearsonův test akceptace

- Pokud je test signifikantní -> zamítá se hlavní hypotéza a akceptuje se alternativní
 - Pokud je test nesignifikantní, ale má vysokou sílu testu -> přijímá se hlavní hypotéza
 - Pokud je test nesignifikantní a má malou sílu testu -> data jsou neprůkazná
 - Pokud se ignoruje ES a β , jedná se o Fisherův přístup
- + Je silnější, vhodný pro opakovaná měření, deduktivní
- inklinuje k mechanickému používání (a priori), méně flexibilní

Fisher significance testing



Neyman-Pearson acceptance testing



Null Hypothesis Significant testing (NHST)

- Poprvé Everett F. Lindquist
- Mnoho autorů a úprav
 - 1940-1960 nespočet tzv. „cookbooks“
 - Často ne-statistici
- Sami autoři obou přístupů nepovažovali své teorie za propojitelné
- Kombinace obou předchozích přístupů
 - N-P procedurální základ, F filozofický základ
- Když se mluví o testování hypotéz, zpravidla se jedná o NHST



NHST jako hybrid je zdrojem nedorozumění

- *„The confusion of many users (...) is probably due to mixing up two distinct approaches to testing hypotheses - Fisher's significance test on one side and Neyman-Pearson's theory of statistical decision on the other - into an "inconsistent hybrid that every decent statistician would reject" (Gigerenzer, 1993)*
- Mnoho miskoncepcí a dezinterpretací
- NHST je často kritizováno
- Ačkoliv je to dominantní přístup, je stále vysoce kontroverzní

Možné výsledky NHST

- Pracuje s **nulovou hypotézou** (typicky status quo) a **alternativní jednostrannou nebo oboustrannou hypotézou**
- $p < \alpha$ = **zamítnutí nulovou hypotézu** (reject)
 - Což se většinou interpretuje jako důkaz pro alternativní hypotézu
- $p > \alpha$ = **nedochází k zamítnutí nulové hypotézy** (fail to reject)
 - Data jsou tedy neprůkazná
- Častá chyba nejen v DP = nesignifikantní výsledek značí podporu nulové hypotézy
 - Omyl, absence důkazu není důkazem absence!
 - (testování ekvivalence)

Kritika NHST

- *“What used to be called judgment is now called prejudice and what used to be called prejudice is now called a null hypothesis. In the social sciences, particularly, it is dangerous nonsense (dressed up as the “scientific method”) and will cause much trouble before it is widely appreciated as such.” (A. W. F. Edwards)*
- *„[Statistical significance testing] is based upon a fundamental misunderstanding of the nature of rational inference, and is seldom if ever appropriate to the aims of scientific research.“ (W. W. Rozeboom)*

4 nejčastější výtky NHST v sociálních vědách

- 1) Senzitivita na velikost vzorku
 - Malý vzorek – silné výsledky nesignifikantní
 - Velký vzorek – triviální výsledky signifikantní
- 2) Nulová (nil-null) hypotéza je téměř vždy fake (a nemá potom smysl ji vůbec vyvracet – strawman argumentační faul)
- 3) Nepříjemně vysoký výskyt chyb. II. typu (a tedy slabá síla testu)
- 4) Nedorozumění a zneužití

Alternativy

- Nahrazení p-hodnoty jinými statistickými indikátory
- Opuštění spojení „statistická významnost“, snížení hladiny významnosti a opuštění jejího mechanického stanovování
- **„Nová statistika“** (Geoff Cumming) - odhady hodnot parametrů, konfidenčních intervalů a velikostí efektů
 - Adekvátní doplnění tradičního NHST
- Bayesovská statistika
- Tukeyho explorativní datová analýza
- A jiné ...



I CAN'T BELIEVE SCHOOLS
ARE STILL TEACHING KIDS
ABOUT THE NULL HYPOTHESIS.

I REMEMBER READING A BIG
STUDY THAT CONCLUSIVELY
DISPROVED IT *YEARS* AGO.

