

Data pro vzdělávání – jak data správně sesbírat a interpretovat

CORE042 – Výzkum v 21. století: data pohledem vaším a vašich kolegů
8. přednáška

Mgr. Ing. Lubomír Prokeš, Ph.D.
Katedra fyziky, chemie a odborné přípravy
Pedagogická fakulta

Vzdělání a kognitivní kapitál

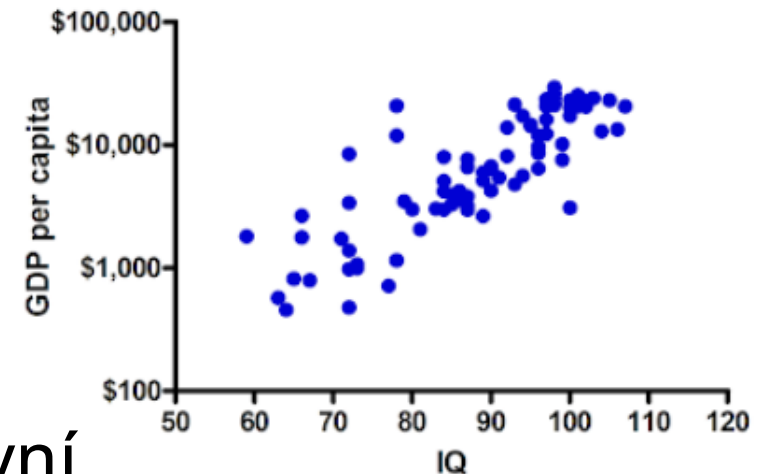
„Národy nečiní velikými především jejich velikáni, ale vyspělost
nesčíslných průměrných lidí.“
– Josef Jungmann

Vzdělávání = proces směřující primárně k rozvoji vědomostí, dovedností, návyků a schopností.

Vzdělanostní společnost = stav společnosti, kdy vědění vstupuje do tržního procesu zhodnocování jako prvek přinášející vyšší zisk než klasická průmyslová výroba.

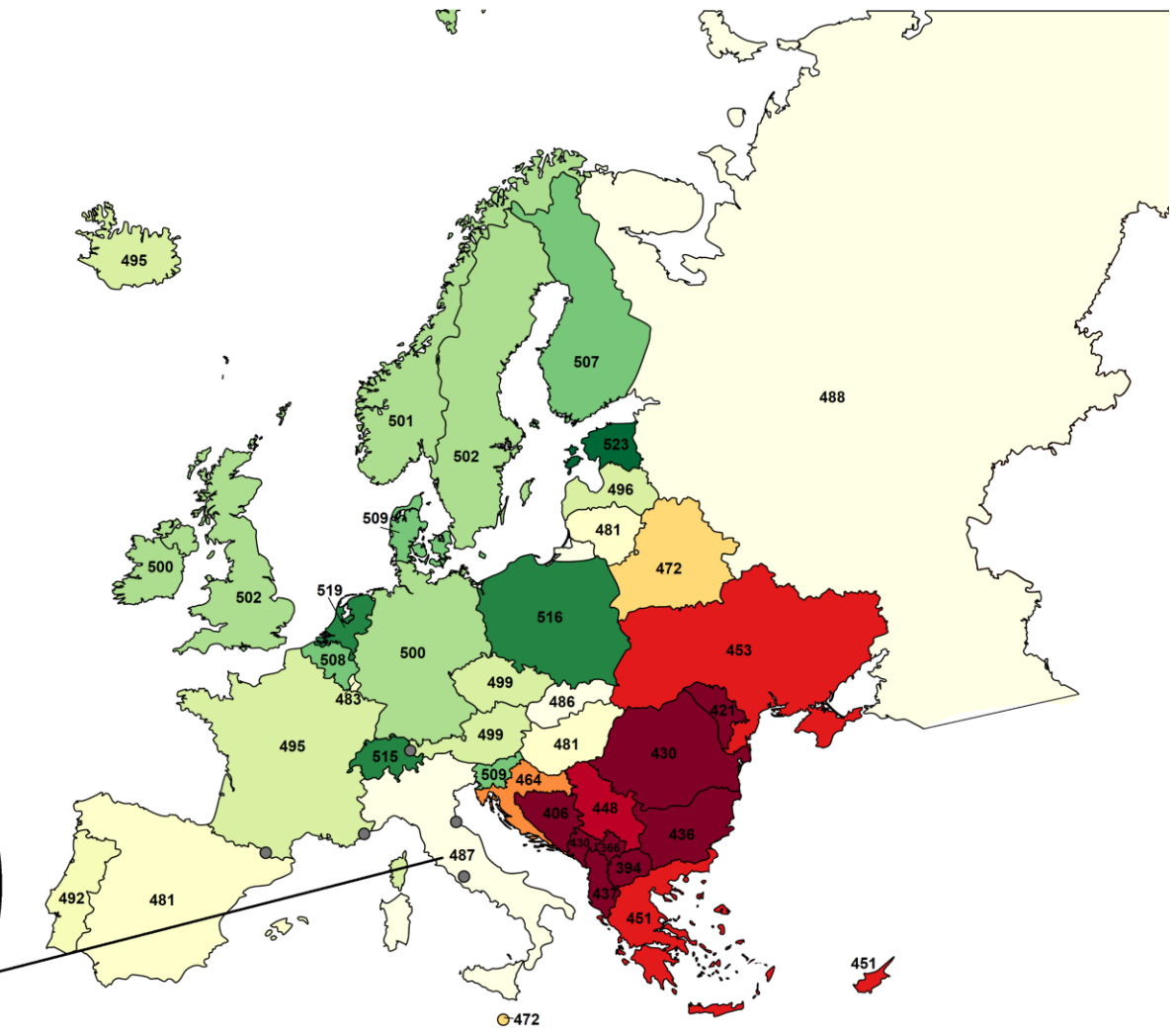
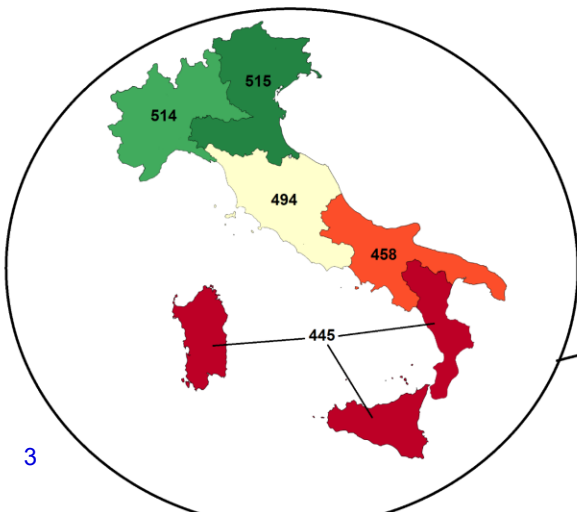
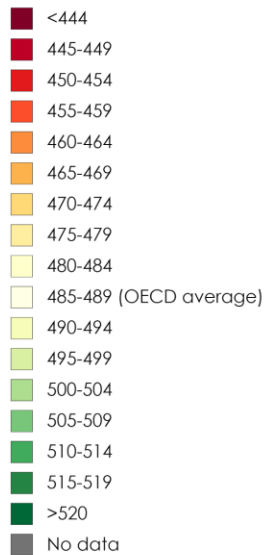
Kognitivní kapitál (Heiner Rindermann & James Thompson) – spočívá v lidských kognitivních schopnostech (měřených koeficientem IQ nebo výsledky testování úspěchů ve vzdělání).

Existuje pozitivní korelace mezi průměrnou kognitivní kompetencí a mírou blahobytu jednotlivých států.



PISA 2018

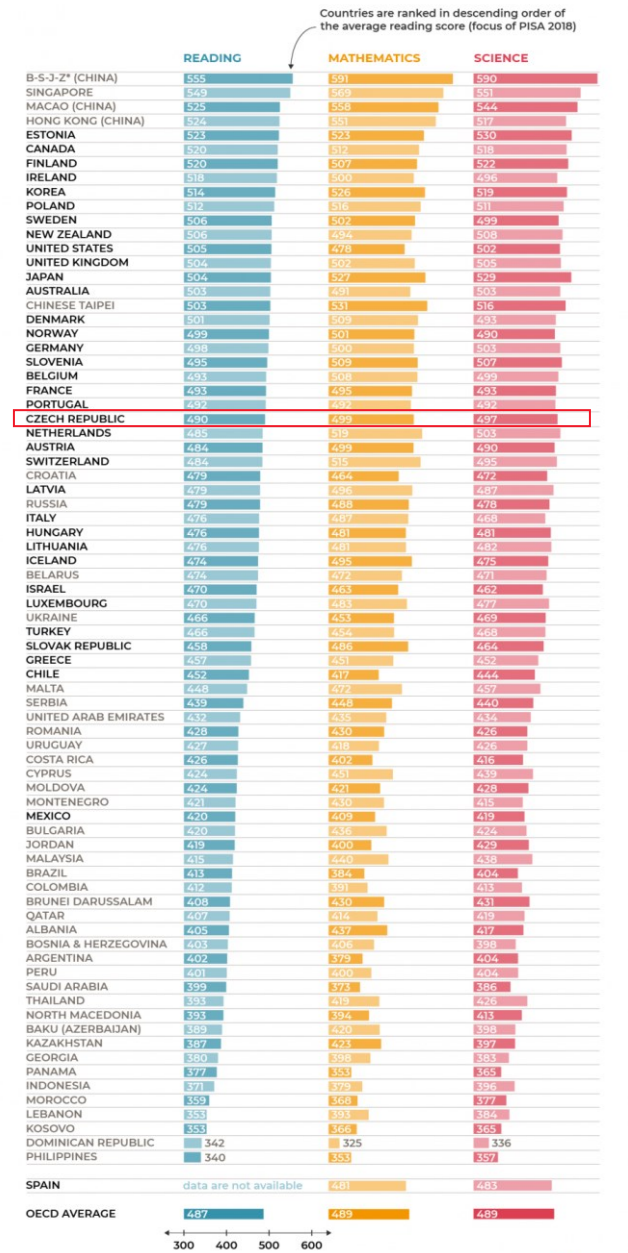
PISA 2018 mean score in mathematics literacy



SOURCES
 Italy: imgur.com/pk8NLBz.png
 Europe: oecd.org/pisa/PISA-results_ENGLISH.png

PISA 2018 results

Snapshot of students' performance in reading, mathematics and science



Source: OECD, PISA 2018 Database || *B-S-J-Z refers to Beijing, Shanghai, Jiangsu and Zhejiang

PISA worldwide ranking

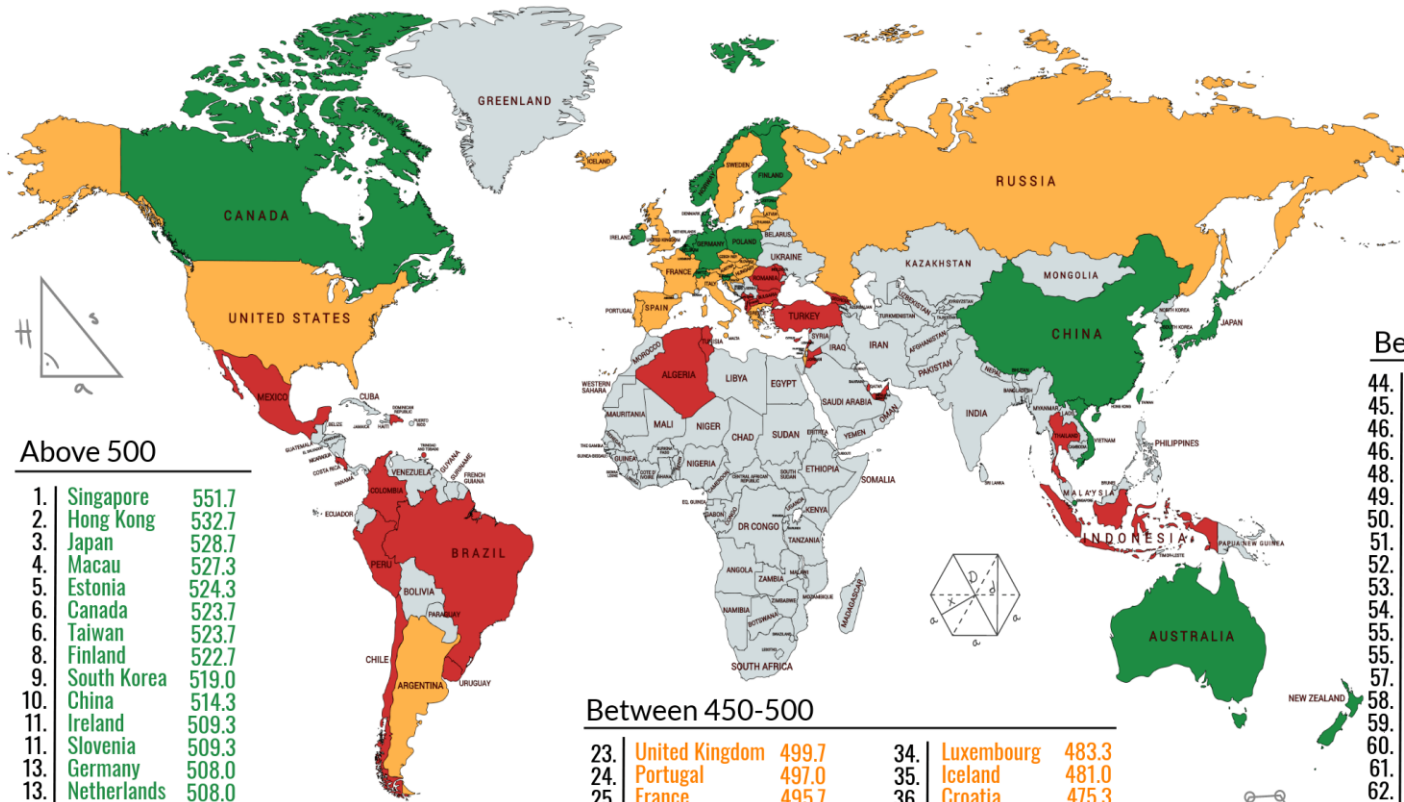
average score of math, science and reading

factsmaps.com

Source: OECD, 2015-2016

The Program for International Student Assessment (PISA) is a worldwide study by OECD in 70 nations of 15-year-old students' scholastic performance on mathematics, science and reading.

above 500 450-500 below 450



Above 500

| | | |
|-----|-------------|-------|
| 1. | Singapore | 551.7 |
| 2. | Hong Kong | 532.7 |
| 3. | Japan | 528.7 |
| 4. | Macau | 527.3 |
| 5. | Estonia | 524.3 |
| 6. | Canada | 523.7 |
| 6. | Taiwan | 523.7 |
| 8. | Finland | 522.7 |
| 9. | South Korea | 519.0 |
| 10. | China | 514.3 |
| 11. | Ireland | 509.3 |
| 11. | Slovenia | 509.3 |
| 13. | Germany | 508.0 |
| 13. | Netherlands | 508.0 |
| 15. | Switzerland | 506.3 |
| 16. | New Zealand | 505.7 |
| 17. | Denmark | 504.3 |
| 17. | Norway | 504.3 |
| 19. | Poland | 503.7 |
| 20. | Belgium | 502.7 |
| 21. | Australia | 502.3 |
| 21. | Vietnam | 502.3 |

Between 450-500

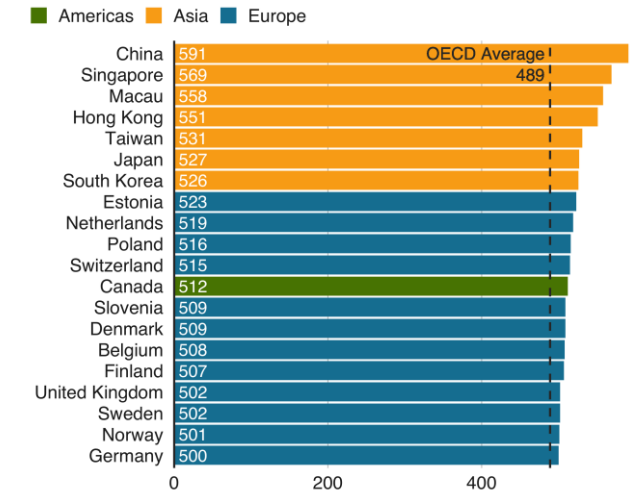
| | | | | | |
|-----|----------------|-------|-----|------------|-------|
| 23. | United Kingdom | 499.7 | 34. | Luxembourg | 483.3 |
| 24. | Portugal | 497.0 | 35. | Iceland | 481.0 |
| 25. | France | 495.7 | 36. | Croatia | 475.3 |
| 25. | Sweden | 495.7 | 37. | Lithuania | 475.0 |
| 27. | Austria | 492.3 | 38. | Hungary | 474.7 |
| 28. | Russia | 492.0 | 39. | Israel | 472.0 |
| 29. | Spain | 491.7 | 40. | Argentina | 468.7 |
| 30. | Czech Republic | 490.7 | 41. | Malta | 463.7 |
| 31. | United States | 487.7 | 42. | Slovakia | 463.0 |
| 32. | Latvia | 486.7 | 43. | Greece | 458.7 |
| 33. | Italy | 485.3 | | | |

Below 450

| | | |
|-----|-----------------|-------|
| 44. | Chile | 443.0 |
| 45. | Bulgaria | 439.7 |
| 46. | Cyprus | 437.7 |
| 46. | Romania | 437.7 |
| 48. | UA Emirates | 432.7 |
| 49. | Uruguay | 430.0 |
| 50. | Turkey | 424.3 |
| 51. | Trinidad & Tob. | 423.0 |
| 52. | Moldova | 421.3 |
| 53. | Montenegro | 418.7 |
| 54. | Costa Rica | 415.7 |
| 55. | Albania | 415.0 |
| 55. | Thailand | 415.0 |
| 57. | Mexico | 411.7 |
| 58. | Columbia | 410.3 |
| 59. | Qatar | 407.3 |
| 60. | Georgia | 405.3 |
| 61. | Jordan | 399.0 |
| 62. | Indonesia | 395.3 |
| 63. | Brazil | 395.0 |
| 64. | Peru | 394.0 |
| 65. | Lebanon | 376.3 |
| 66. | Tunisia | 371.3 |
| 67. | Macedonia, Rep. | 369.0 |
| 68. | Algeria | 362.0 |
| 69. | Kosovo | 355.7 |
| 70. | Dominican Rep. | 339.3 |

Asia holds top seven rankings in maths

Top 20 average national scores on PISA tests in 2018

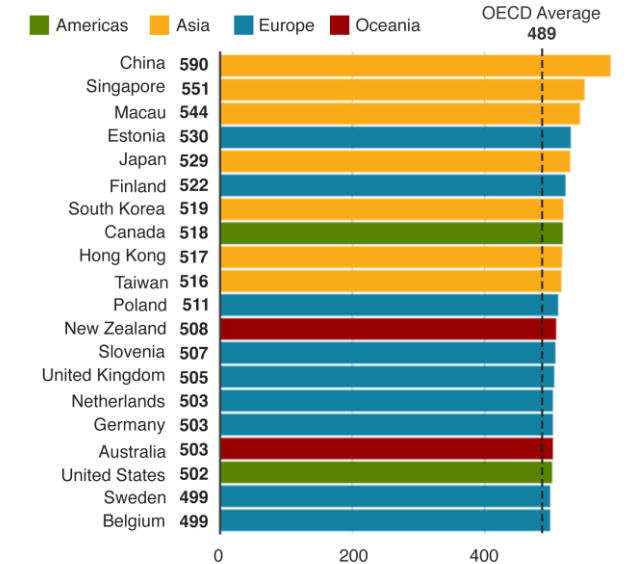


Source: OECD, 2015

BBC

Chinese students ahead in science

Top 20 in science on 2018 Pisa tests

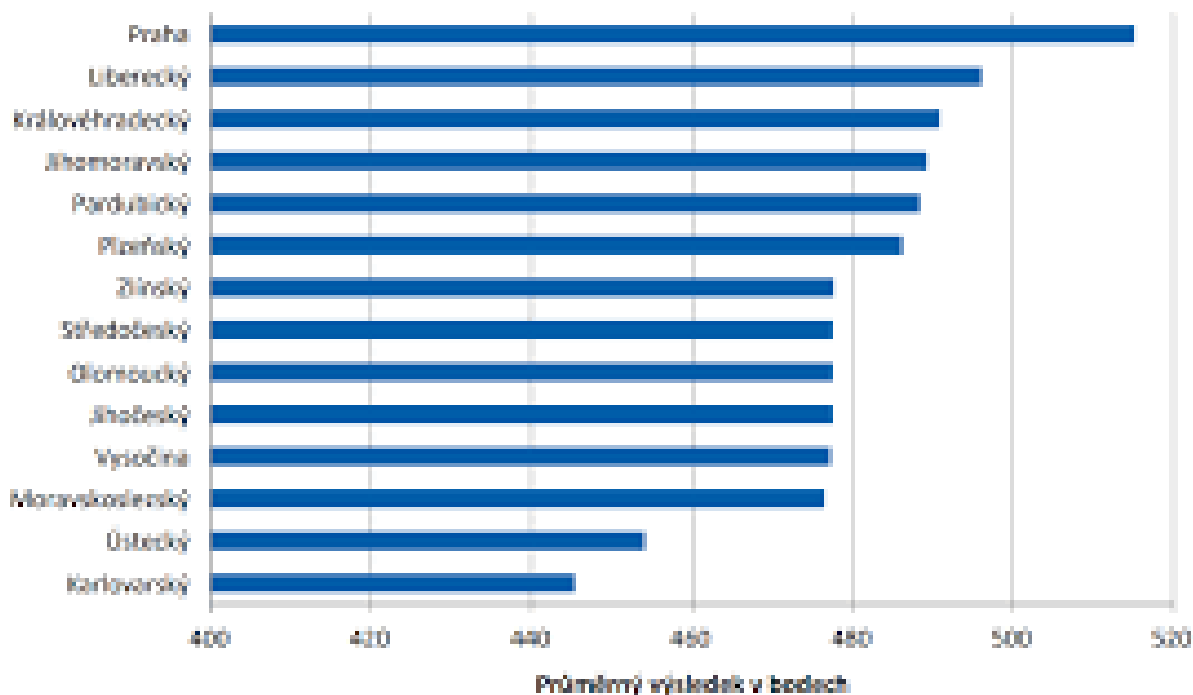


Source: OECD

BBC

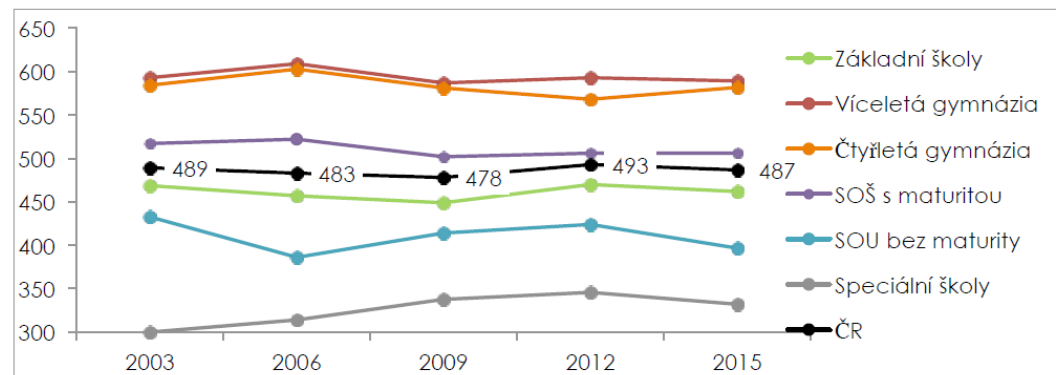
Analýza vzdělávání v České republice

Rešerše a dodatečná analýza vybraných problémů v českém vzdělávání.

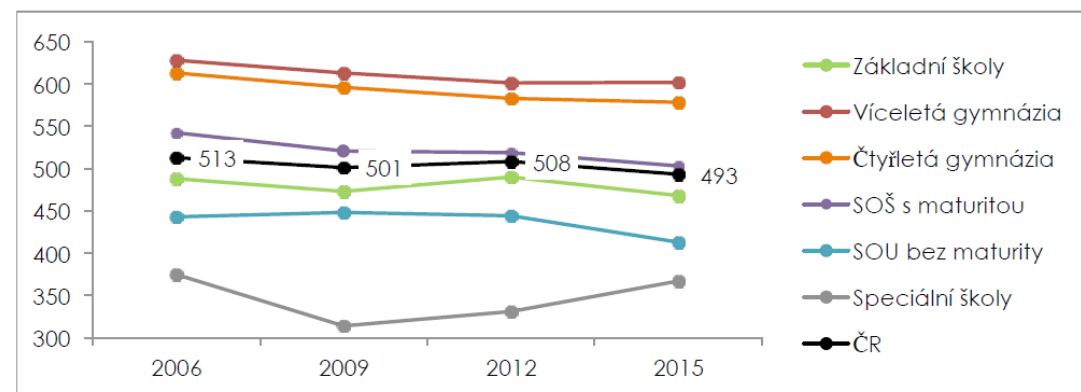


Přírodovědná gramotnost (PISA 2015) – srovnání krajů.

Čtenářská gramotnost (PISA 2003–2015)



Matematická gramotnost (PISA 2006–2015)



Tab. B1.5: Školství celkem – veřejné výdaje na školství v běžných cenách v letech 2006 až 2016

Včetně výdajů Ministerstva obrany.

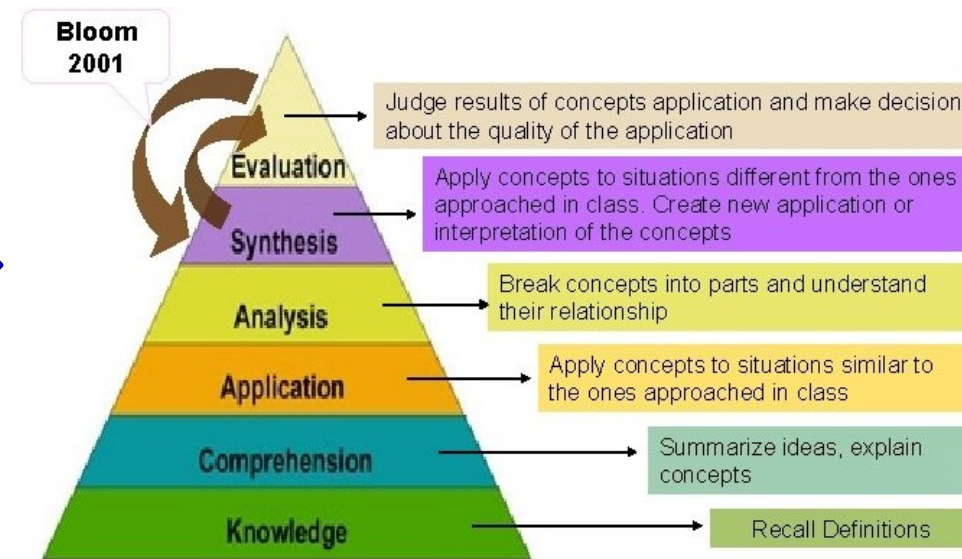
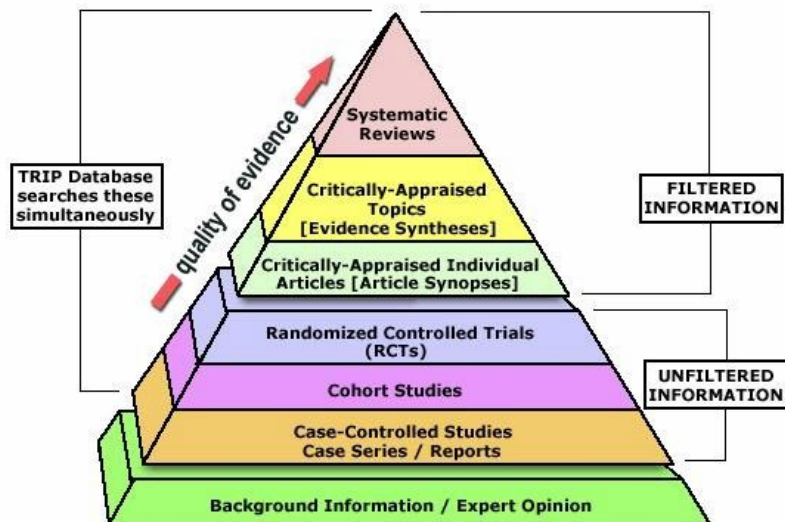
v mld. Kč

| | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|---------------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| | 1) | 1) | 1,2) | | | | | | | | |
| Hrubý domácí produkt v běžných cenách | 3 507,1 | 3 831,8 | 4 015,3 | 3 921,8 | 3 953,7 | 4 033,8 | 4 059,9 | 4 098,1 | 4 313,8 | 4 554,6 | 4 715,1 |
| Výdaje na školství celkem | 142,8 | 153,0 | 151,0 | 163,9 | 163,0 | 173,7 | 171,4 | 172,8 | 178,7 | 182,8 | 173,4 |
| Výdaje na školství v % HDP | 4,1% | 4,0% | 3,8% | 4,2% | 4,1% | 4,3% | 4,2% | 4,2% | 4,1% | 4,0% | 3,7% |
| Výdaje státního rozpočtu | 1 020,6 | 1 092,3 | 1 083,9 | 1 167,0 | 1 156,8 | 1 155,5 | 1 152,4 | 1 173,1 | 1 211,6 | 1 297,3 | 1 219,8 |

Pedagogický výzkum

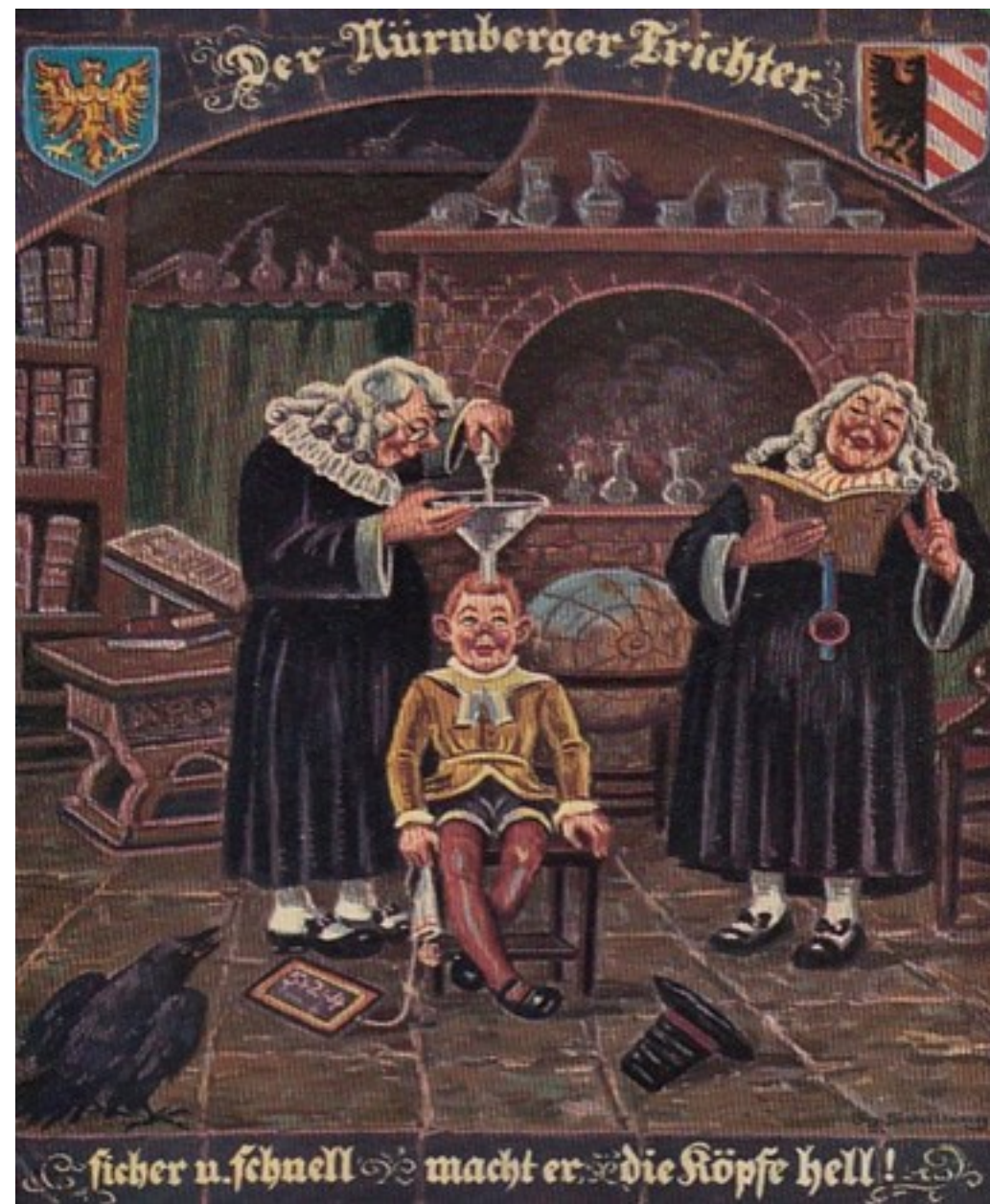
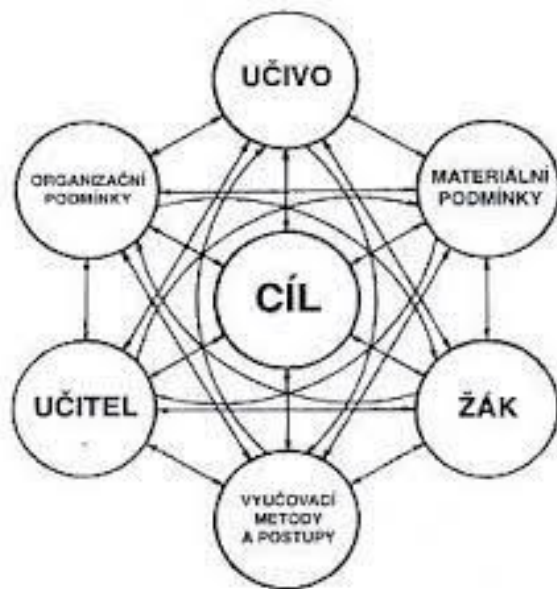
Pedagogický výzkum je plánované získávání údajů o edukační realitě za účelem zkvalitnění výchovy a vzdělávání.

Vzdělávání založené na důkaze (Evidence-based education, EBE) je snaha o systematické využívání poznatků získaných výzkumem v oblasti vzdělávání, provádění meta-analýz a snahu tyto prvky následně převést do praxe.



Cíle pedagogického výzkumu

Kvalitní data, jejich správná analýza a interpretace mohou být pro pedagogy i administrativní pracovníky základem kvalitního plánování a rozhodování.



Plynulý přenos dat:

je nezbytný pro efektivní rozhodování na základě datových výstupů.

Datově gramotní konzumenti: chápou význam dat a dokážou z nich vyvodit závěry.

Tvůrci schopní přenášet data plynule: předávají konzumentům informace a smysluplné datové výstupy s ohledem na priority, potřeby a perspektivu konzumenta dat.

Kultura plynulého přenosu dat: přístup organizace, vedení a zaměstnanců ke sběru a hodnocení dat a rozhodování na základě dat.

Ekosystém datového produktu: definuje standardy, nástroje a procesy a zprostředkovává plynulý přenos dat v organizaci.

Teorie duálního systému myšlení (Amos Tversky & Daniel Kahneman)

Intuitivní (emocionální, tzv. *rychlé*) – je velmi rychlé, málo náročné na energii. Je evolučně velmi staré. Hlavním úkolem je ochránit nás před nebezpečím a umožnit nám přežít. Výsledkem jsou myšlenkové zkratky (heuristiky, “rules of thumb”).

Záměrné (kognitivní, tzv. *pomalé*) - je značně energeticky náročné a evolučně velmi mladé, vyskytuje se pouze u vyšších primátů (člověk, šimpanz) a krkavcovitých ptáků (vrány). Výsledkem je myšlení logické a racionální.

Mysl vždy přednostně nabízí méně náročné intuitivní řešení (systém 1).

"Mozek je v podstatě líný kus masa. Nechce plýtvat energií."

Gregory Berns

MUNI

Kognitivní zkreslení (cognitive biases): mylné úsudky či nelogické a iracionální interpretace skutečnosti vycházející z nevhodného použití systému 1 nebo 2 (chyby v poznání a emocionální předsudky). Uplatňují se v závislosti na kontextu.



V některých situacích mohou dlouhé přemýšlení a vyčerpávající analýza faktů (systém 2) rozhodování zbrzdit či dokonce znemožnit. Přemýšlení tak může bránit přijetí intuitivního řešení (systém 1).

Vyřčení otázky, metodologie a plánování výzkumu

Identifikace a formulace problému a navazujících otázek.

Zpracování řešerše a vyhodnocení příslušné literatury.

Formulace hypotézy.

Plánování výzkumu.

Plány pedagogického výzkumu

Výběrové (statistické) šetření

Experiment

Pozorování (observační studie)

Případové studie

Metaanalýza publikovaných dat

Big data

Nástroje pedagogického výzkumu (metody sběru dat)

Dotazník

Interview

Didaktické testy

Pozorování



Sekundární data a dokumenty - metaanalýza

Analýza sociálních sítí

Text mining

} big data

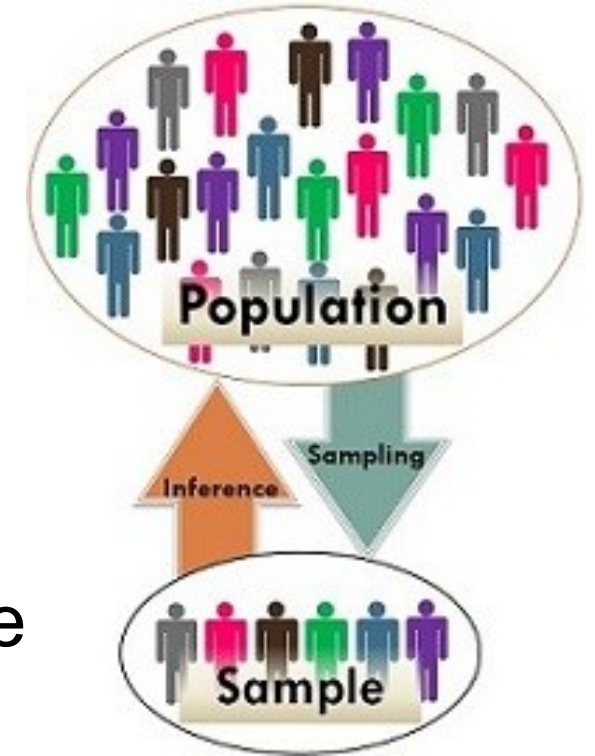
I. Výběrové šetření

= zaznamenání údajů o jedincích z téže skupiny (výběr, vzorek), reprezentující větší množinu jedinců (populace, základní soubor).

Prostý náhodný výběr = reprezentativní vzorek ze zvoleného základního souboru

(! v praxi může být obtížně realizovatelný).

Skupinový náhodný výběr - náhodný výběr se provádí pro celé skupiny (např. třídy, školy, ...)



Exhaustivní (totální) výběr: získáváme data od všech prvků zvoleného základního souboru (**census**).

Záměrný výběr - o výběru daného prvku nerozhoduje náhoda, ale úsudek výzkumníka či zkoumané osoby (samy se zařadí do výběru na základě výzvy výzkumníka).

Výběr na základě dobrovolnosti (anketa)

Výběr na základě dostupnosti

Kvótní výběr (určitý počet jedinců v dané kategorii)

Výběr „průměrných jednotek“ („typických“ případů)

Na záměrný výběr lze aplikovat pouze popisnou statistiku, nikoli statistické testování !!

Typické chyby v dotaznících pro výběrová šetření

Zkreslení v otázce

Úmyslné: „Souhlasíte, že ...“ – pokud respondent nemá silně opačný názor, má tendenci odpovědět kladně (tzv. **rámování**).

Neúmyslné: nepřesně/nejasně či složitě formulovaná otázka.

Pořadí otázek může ovlivnit výslednou odpověď (viz **efekt prvního dojmu**).

Důvěrnost otázek a anonymita: anonymní dotazník obvykle poskytne pravdivější informace.

Paradox volby (rozhodovací paralýza): čelí mu respondenti při rozhodování mezi mnoha alternativami. Větší složitost výběru zvyšuje úzkost a ta brání v rozhodování.

Chyby při sběru dat pro výběrová šetření

Nesprávná představa o populaci - může vyvolat zkreslení, zobecnění tak může být de facto prováděno pro jinou populaci (např. v důsledku klamu přeživších).

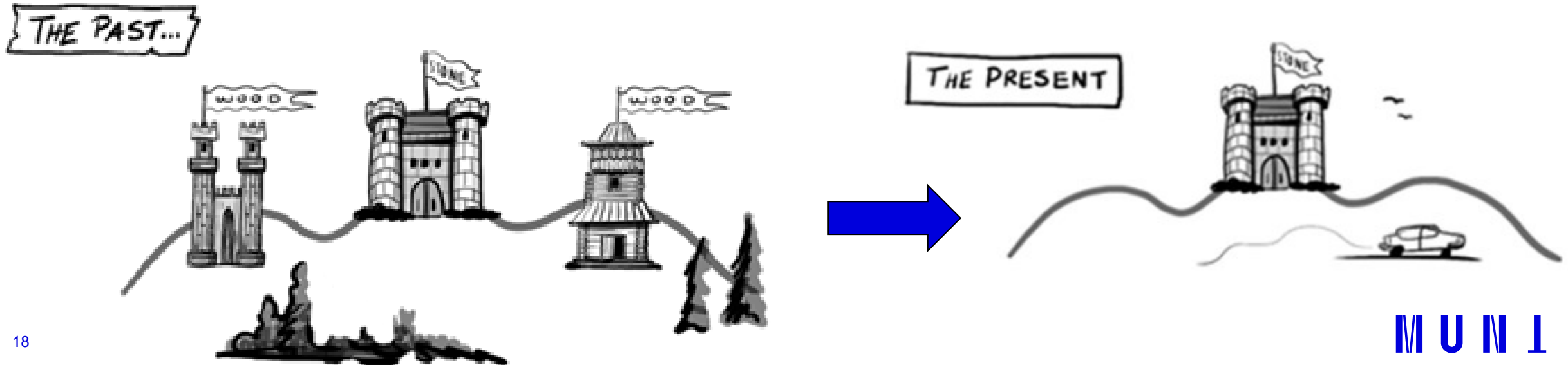
Nereprezentativní výběr - vede k systematické chybě, zobecnění tak může být de facto prováděno pro jinou populaci. Navíc se mohou objevit i **non-response** (NA) hodnoty.

Nevhodná metoda zjišťování zkoumaného jevu a experimentální chyby

Nevhodný rozsah výběru - ovlivňuje bodové i intervalové odhady a testování hypotéz

Klam přeživších (survivorship bias)

= logický klam, založený na vyšší viditelnosti těch, kteří "přežili" určitý proces a neviditelností „nepřeživších“ jedinců. Jeho důsledkem mohou být příliš optimistické závěry.



Výsledky testování a monitorování (interregionální a mezinárodní srovnání)

United Nations Educational, Scientific and Cultural Organization (UNESCO; od 1958)

International Association for the Evaluation of Educational Achievement (IEA; od 1967):

TIMSS (přírodověda)

PIRLS (čtenářská gramotnost)

ICCS (občanská výchova)

Organisation for Economic Co-operation and Development (OECD; od 2000)

PISA (matematická, čtenářská a přírodovědná gramotnost)

TALIS (postoje a názory učitelů)

PIAAC (funkční gramotnost dospělých)

II. Observační (pozorovací) studie

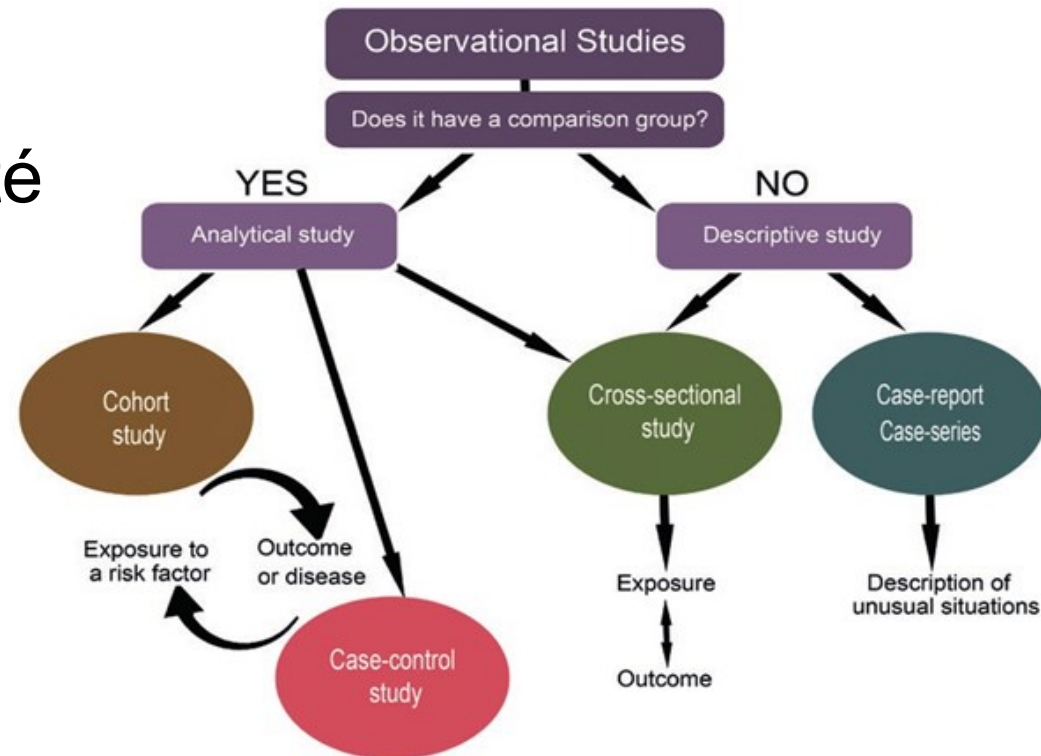
= sledování jedinců a měření proměnných, bez záměrného ovlivňování kterékoli z proměnných.

Cílem pozorování je získat popis určité skupiny jedinců nebo situace.

Případová studie = získání podrobné, hloubkové znalosti o jednom nebo několika málo případech.

Standardizované pozorování = snaha o objektivizaci pozorování,

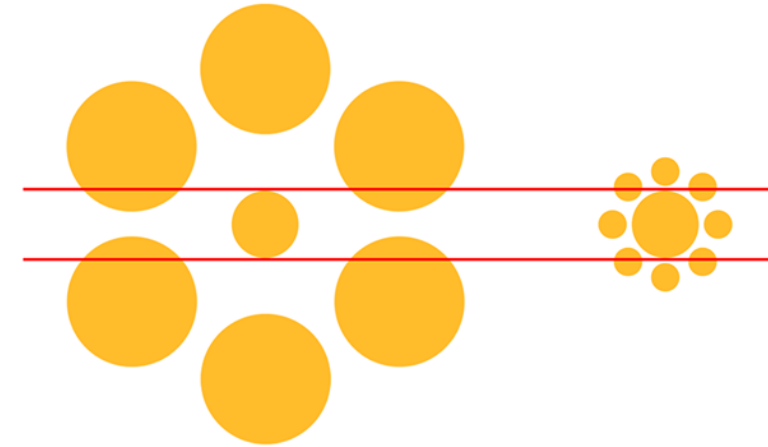
využívá různé druhy posuzovacích škál.



Kognitivní zkreslení při sběru dat



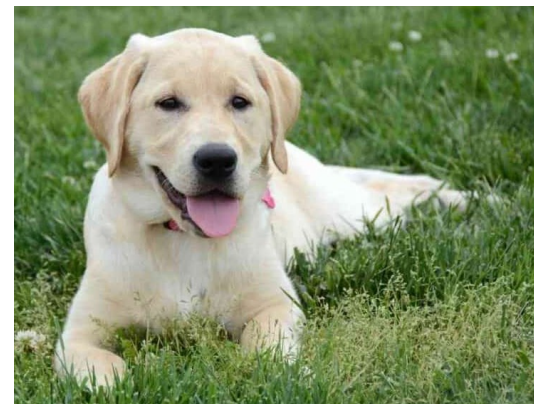
Efekt asymetrické dominance (efekt volavky, decoy effect, efekt návnady) = důsledek toho, že nehodnotíme věci na základě absolutních hodnot, ale prostřednictvím srovnání s dalšími možnostmi. Něco posuzujeme jako hezčí, dražší, větší, atd., když máme současně srovnání s něčím nevzhledným, laciným, malým, atd.



Efekt toho, co je přítomné (feature-positive effect): je obtížnější všimnout si, že něco chybí, než že je něco přítomné. Chybění některých atributů je méně nápadné, než přítomnost atributů jiných.

Haló efekt (efekt svatozáře) = převládnutí prvního nebo nejsilnějšího dojmu, který následně potlačí všechny ostatní vjemy, čímž brání jejich správnému rozlišení a vyhodnocení, sklon předpokládat, že se pozitivní vlastnosti seskupují.

Předpoklad, že eseje napsané krásným rukopisem jsou lepší než ty naškrábané.



Efekt salience: nějaký nápadný znak (vlastnost, atribut) si získá větší pozornost, než by mu měla náležet. Že je něco nápadné, nemusí znamenat, že je to důležité, nicméně to může zastínit skutečně důležité příznaky, které mohou být méně nápadné. Tento znak se následně může stát tzv. „kotvou“.

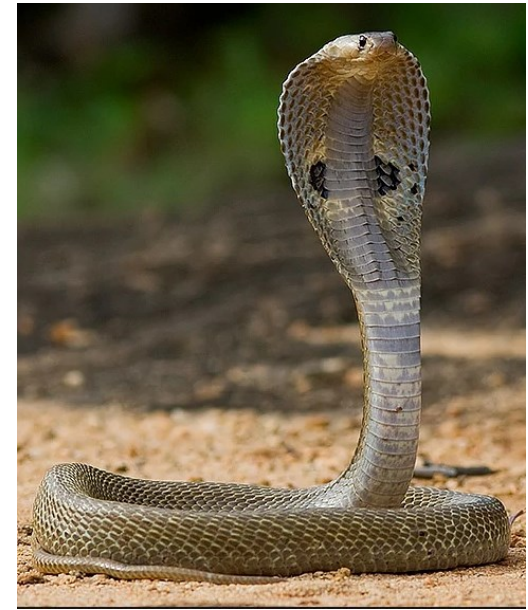
Goodhartovo pravidlo: "Jestliže se určitý ukazatel stane cílem, potom přestane být dobrým ukazatelem." Jde vlastně o záměrné ovlivnění hodnoty příslušného ukazatele, ten pak ztrácí hodnotu pro observační studie.

Např. když se známky stanou cílem, ztratí svou hodnotu ukazatele stavu vzdělání.

Campbellovo pravidlo: „Čím více se jakýkoli kvantitativní ukazatel používá k rozhodování, tím více bude předmětem vnějšího ovlivňování a tím více se naruší procesy, které má sledovat.“

Např. jestliže bude mít měření výsledků (plošné testování) vliv na hodnocení jednotlivých žáků či celé školy, zcela jistě to bude mít přímý dopad na výuku, která se bude zaměřovat především na to, co se bude testovat.

Goodhartovo a Campbelllovo pravidlo lze považovat za varianty tzv. **kobřího efektu (cobra effect)**: lidé reagují na každé nové pravidlo, regulaci a příkaz, a jejich reakce vedou k výsledkům, které mohou být zcela odlišné od zamýšlených výsledků (pokus o řešení problému často původní stav ještě zhorší).

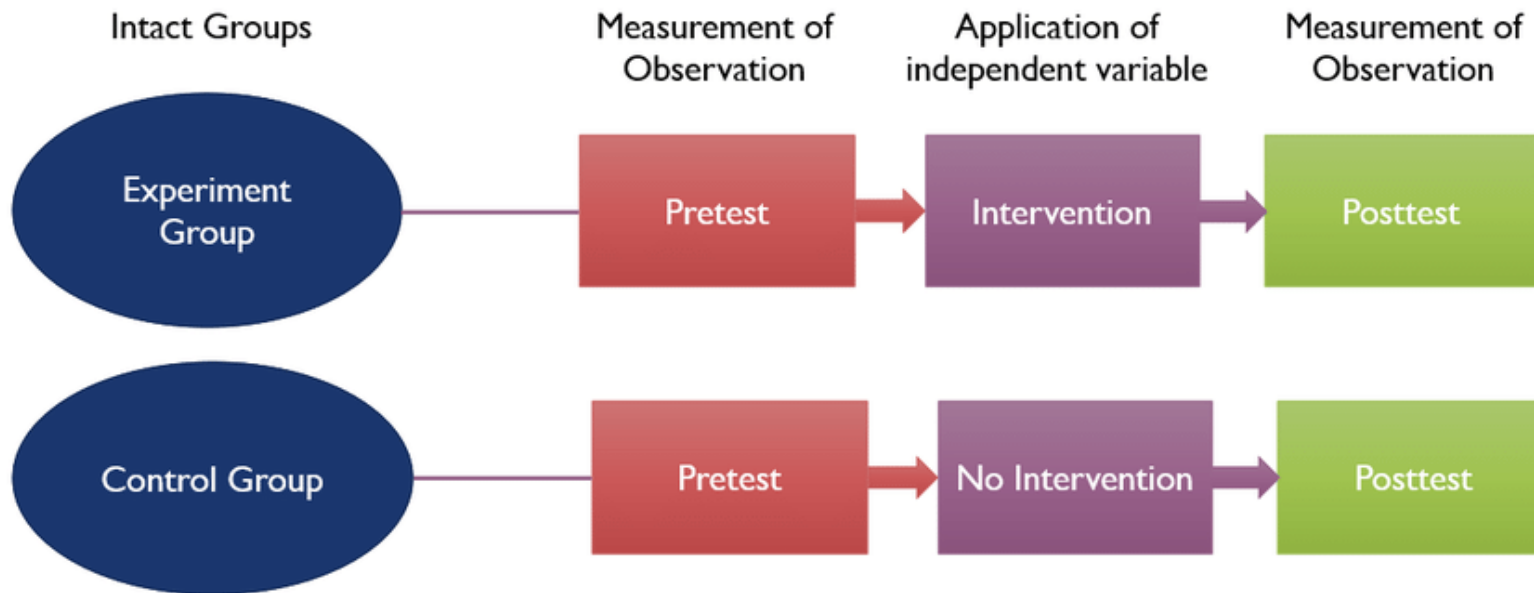


Tendence k průměru: sklon přisuzovat jevům spíše střední intenzitu, než intenzitu vysokou či nízkou, např. vlivem nezkušenosti, opatrností či shovívavostí pozorovatele.

Hawthornský efekt (efekt pozorovatele): chování subjektů je ovlivňováno samotným výzkumem (pozorováním), účastníci reagují na to, že jsou středem pozornosti.

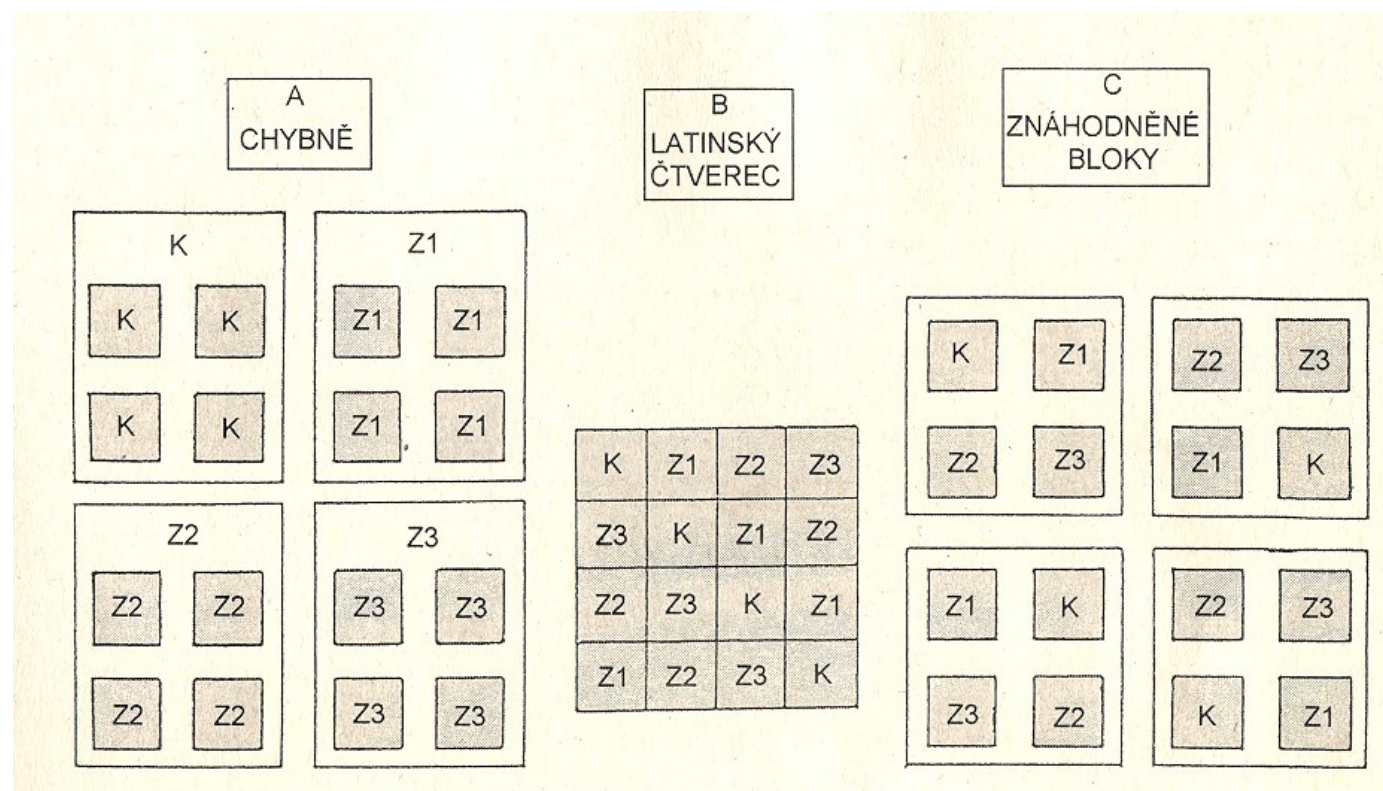
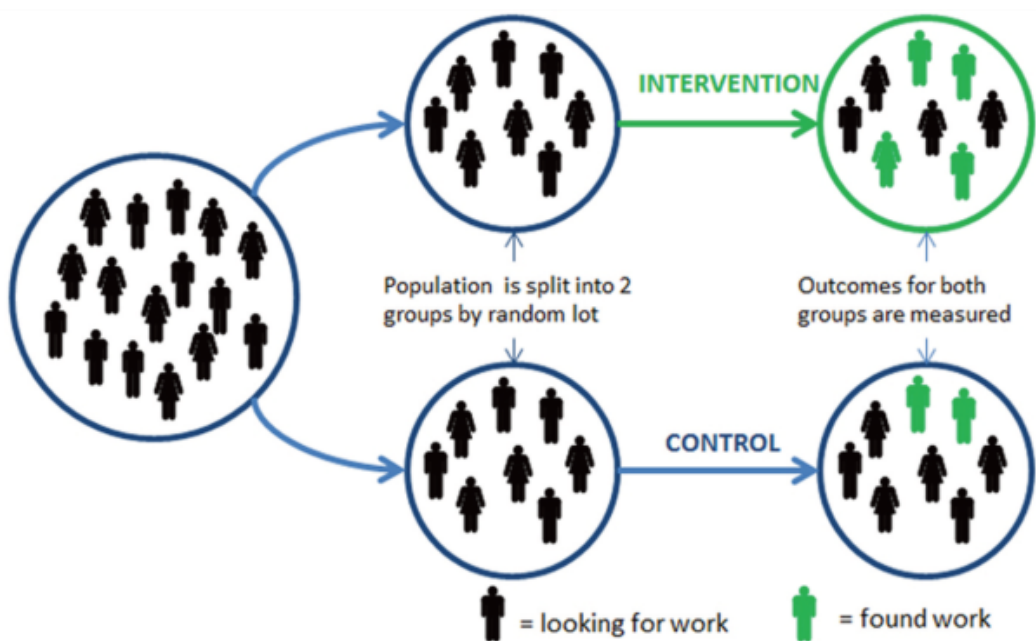
III. Experiment

zkoumá změny, které nastaly u jedné či více skupin jedinců (pokusných jednotek) po aplikaci nějakého ošetření (sledování příčin a následků). V průběhu experimentu je jedna skupina ošetřena způsobem A a druhá skupina způsobem B (nebo zůstane neošetřena).



Podobnosti skupin se dosahuje **randomizací**: jedince do obou skupin se z populace vybírají náhodným výběrem.

Randomizace je náhodné rozdělení jedinců do skupin. Zvyšuje pravděpodobnost, že si skupiny budou podobné a výsledky nebudou ovlivněny (eliminace vlivu nežádoucích proměnných).



Typy proměnných z hlediska experimentu

Nezávisle proměnná – proměnná, jejíž hodnotu můžeme účelově ovlivnit a měnit. Experimenty se musí vždy navrhnout tak, aby jednotlivé proměnné na sebe nepůsobily navzájem rušivě.

Závisle proměnná – proměnná, u které očekáváme změny v důsledku změn hodnot nezávisle proměnné.

Intervenující/nežádoucí proměnná – rušivý element, který do výzkumu vstupuje nechtěně a může zkreslit jeho výsledek, vliv těchto proměnných se snažíme eliminovat randomizací.

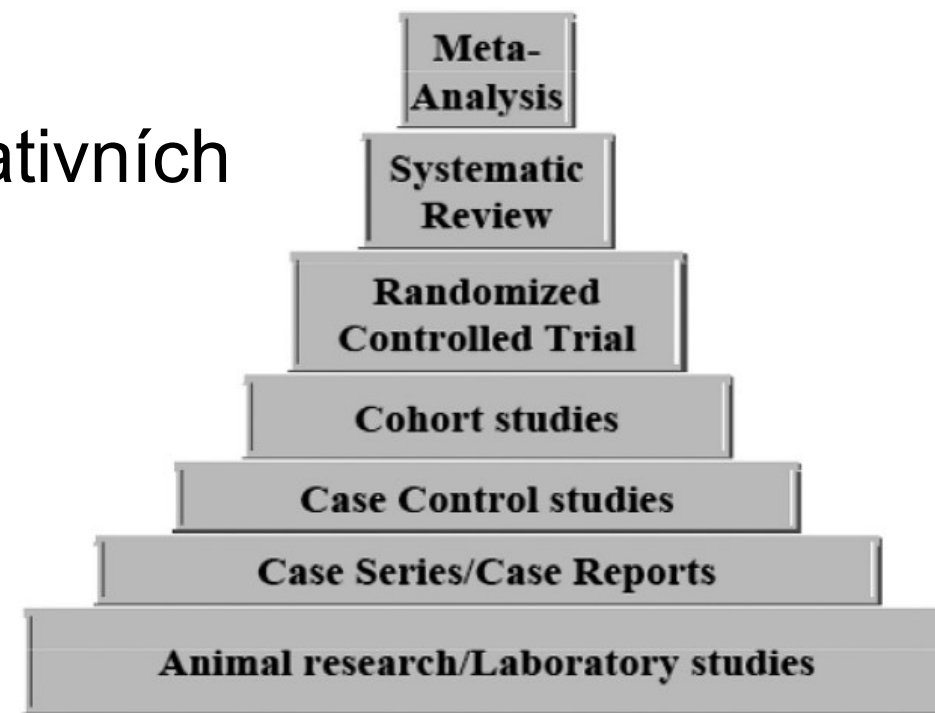
Skrytá proměnná = prostorový/časový trend, systematické chyby, ...

IV. Metaanalýza dat

Metaanalýza souhrnně **analyzuje data z více na sobě nezávislých studií**. Cílem je identifikace a kvantifikace převažujících trendů nebo zjištění příčin rozdílných závěrů prací.

Metaanalýza dokáže odhalit ignorování negativních výsledků (**publikační zkreslení**)

Metaanalýza dokáže odhalit **opakované publikování** pozitivních výsledků.



Sekundární data

- archivovaná nebo publikovaná **data**

UNESCO (<http://uis.unesco.org/>)

IEA (<https://www.iea.nl/>)

World Bank (<https://data.worldbank.org/topic/education>)

MŠMT (<https://www.msmt.cz/vzdelavani/skolstvi-v-cr/statistika-skolstvi/mezinarodni-statistiky>)

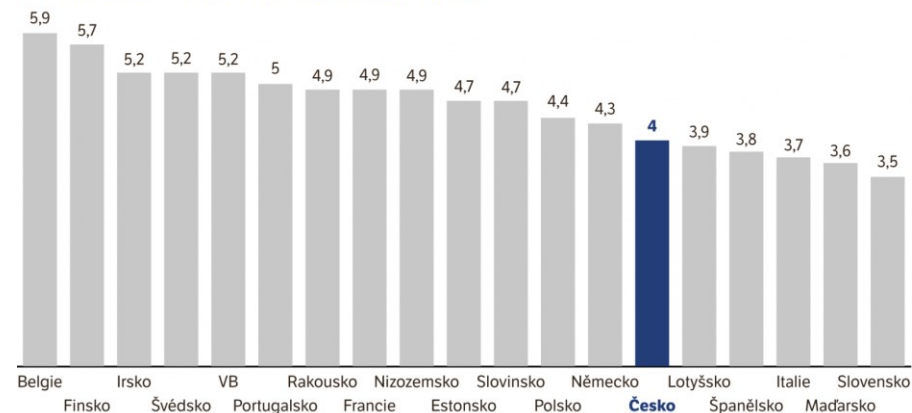
MŠMT (<https://www.msmt.cz/vzdelavani/skolstvi-v-cr/statistika-skolstvi/otevrena-data>)

ČŠI (<https://csicr.cz/cz/>)

CERMAT (<https://data.ceremat.cz/>)

Otevřená data v ČR (<https://data.gov.cz/datov%C3%A9-sady>)

VÝDAJE NA ŠKOLSTVÍ V PROCENTECH HDP (2012)



29- archivované nebo publikované **dokumenty** (osobní, oficiální)

V. Big data

= soubory dat, jejichž velikost (**objem**) je mimo schopnosti zachycovat, spravovat a zpracovávat data běžně používanými softwarovými prostředky (používají se cloudy) v rozumném čase (**rychlost**). Data nemusí mít přesně definovanou strukturu (**nestrukturovanost**) a mohou být přítomny v různých formátech (**nehomogenita** a **nekonzistence**).

Virtuální výukové prostředí dokáže vytěžit velké množství dat o každém žákovi. Např. u online kurzů (*Duolingo, Khan Academy, Coursera, aj.*), které big daty disponují, jich lze efektivně využít jako zpětné vazby k inovaci a zkvalitnění výuky.

Jak vhodně zmanipulovat výzkum

Testování na ideálním (resp. dokonale abnormálním) vzorku: nízká prediktivní schopnost (overfitting) a generalizovatelnost.

Testování oproti „braku“: výsledky vypadají efektivněji (kontrast v důsledku efektu asymetrické dominance).

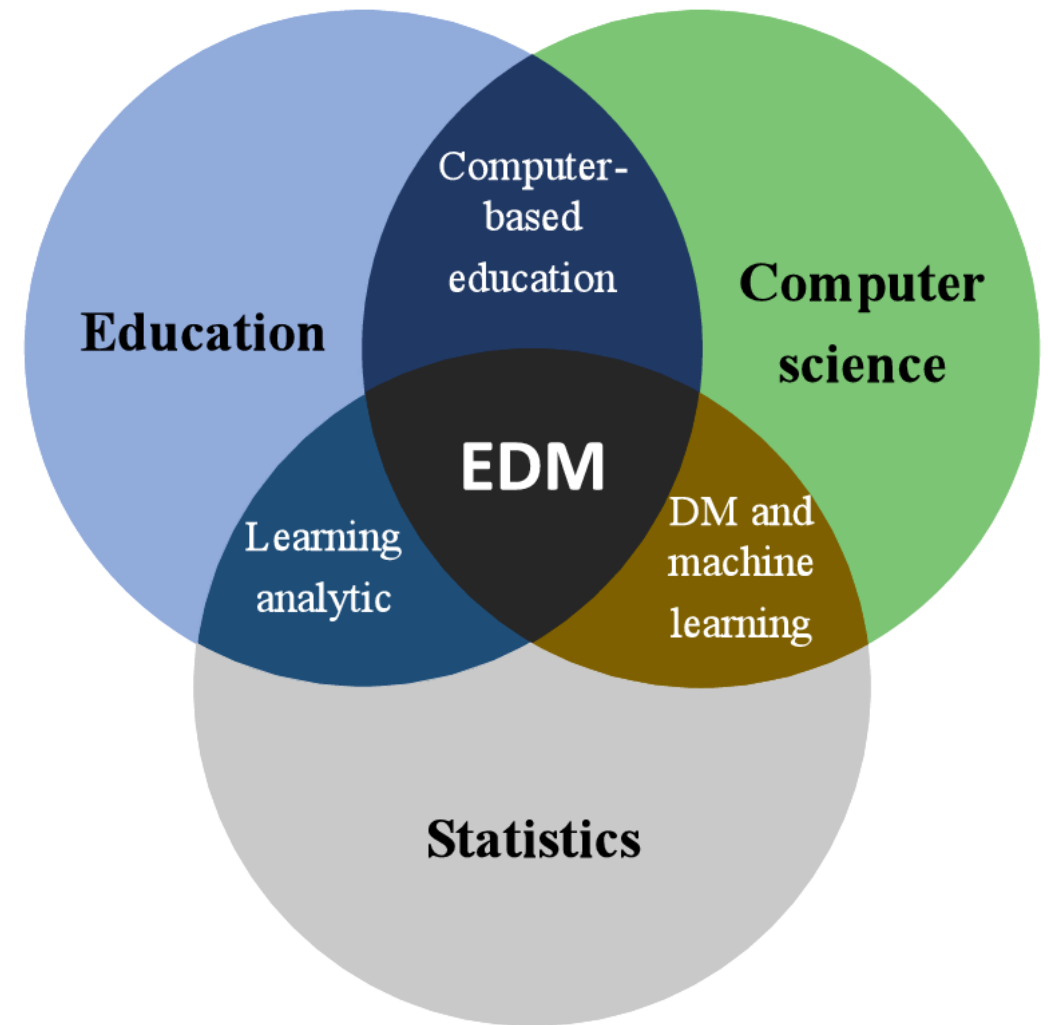
Ukončení výzkumu dostatečně brzy („dokud to vychází“), resp. **dostatečně pozdě** („dokud nedostanu vhodný výsledek“) zvyšuje pravděpodobnost příznivých výsledků.

Vylučování subjektů z výběru (intention to treat): odstranění extrémních hodnot z výsledků, resp. vylučování subjektů během výzkumu.

³Analýza **vhodně vybraných podskupin.**

Datová věda (data science)

- sjednocuje statistiku, analýzu dat, data mining, strojové učení (machine learning) a související metody s cílem pochopit a analyzovat skutečné jevy na základě dat. Využívá techniky a teorie čerpané z mnoha oblastí matematiky, statistiky, informatiky a matematické informatiky

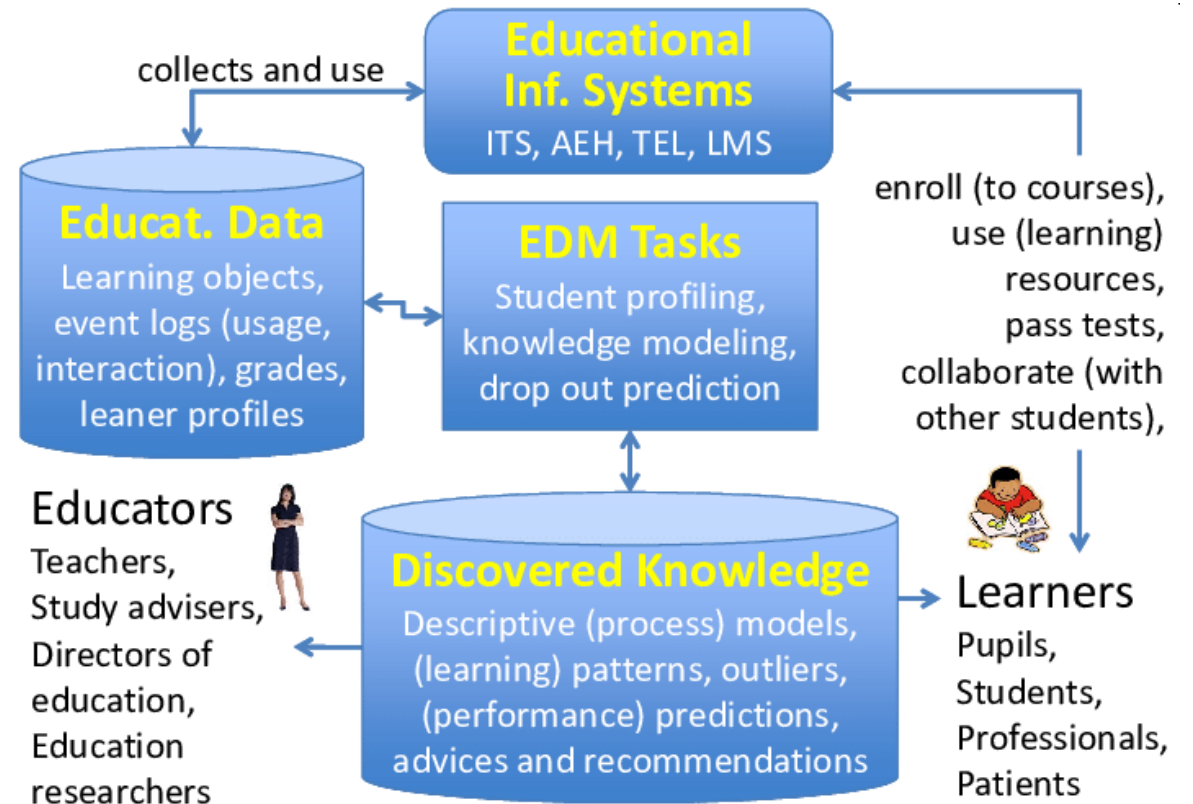


Data mining je proces hledání vzorů v datech. Používá se pro zjišťování nových poznatků a hypotéz v protikladu ke klasické **statistické analýze**, která má za cíl je potvrdit nebo vyvrátit.

Learning analytics

Analýza kurzů, práce tříd, resp. škol (predikce a hledání konfigurací).

Analýza online učení – zpracování dat v reálném čase (spokojenost s kursem, možná dezorientace studenta a snaha o nápravu)



Statistika

Zahrnuje organizaci dat a jejich popis užitím grafů a numerických souhrnů.

Popisná (deskriptivní) statistika se snaží několika čísly (průměr, medián, rozptyl, koeficient korelace, aj.), případně obrázky, stručně vystihnout podstatné informace o daných datech.

Exploratorní (průzkumová) analýza dat využívá k analýze dat grafické nástroje a vizualizaci dat. Jejím úkolem je zejm. ověřit předpoklady, usnadnit výběr vhodných statistických nástrojů a technik, případně též navrhnout hypotézy o příčinách pozorovaných jevů.

Statistická indukce (statistická inference): zobecňování úsudků o vlastnostech populace založený na informacích z náhodných výběrů.

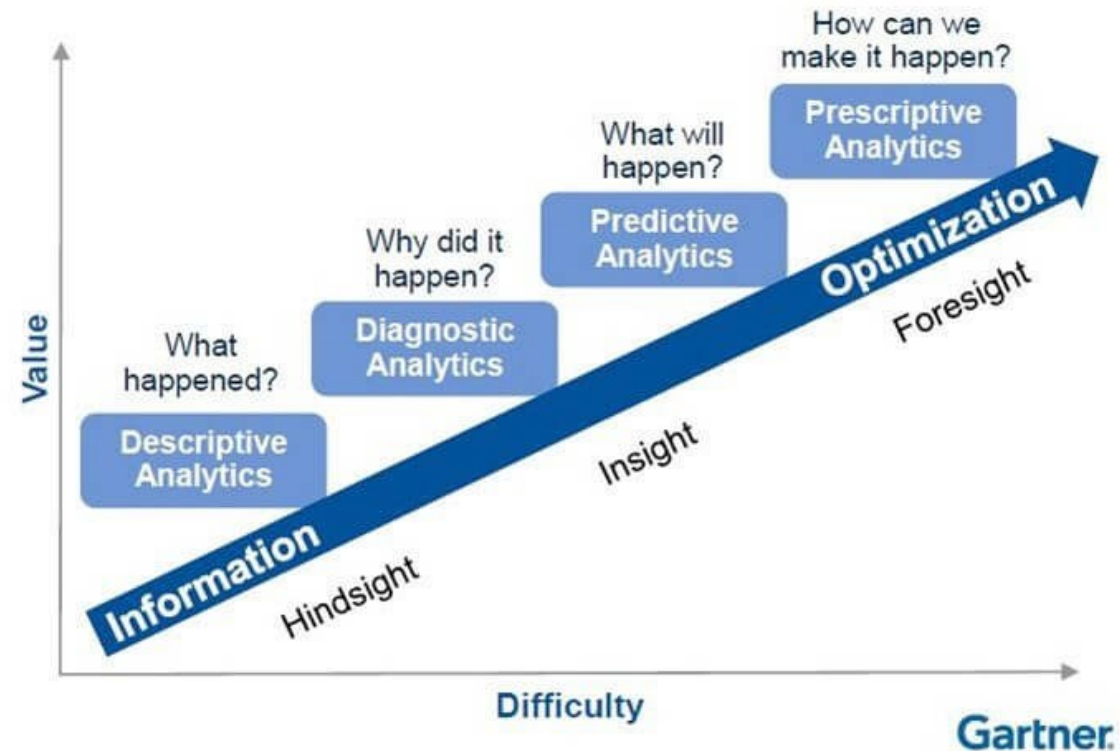
Strategie používání dat

Popisná (deskriptivní) analytika:
Co se stalo?

Diagnostická analytika: *Proč se to stalo?*

Prediktivní analytika: *Co se stane?*

Preskriptivní analytika: *Co musíme udělat, aby se to stalo?*

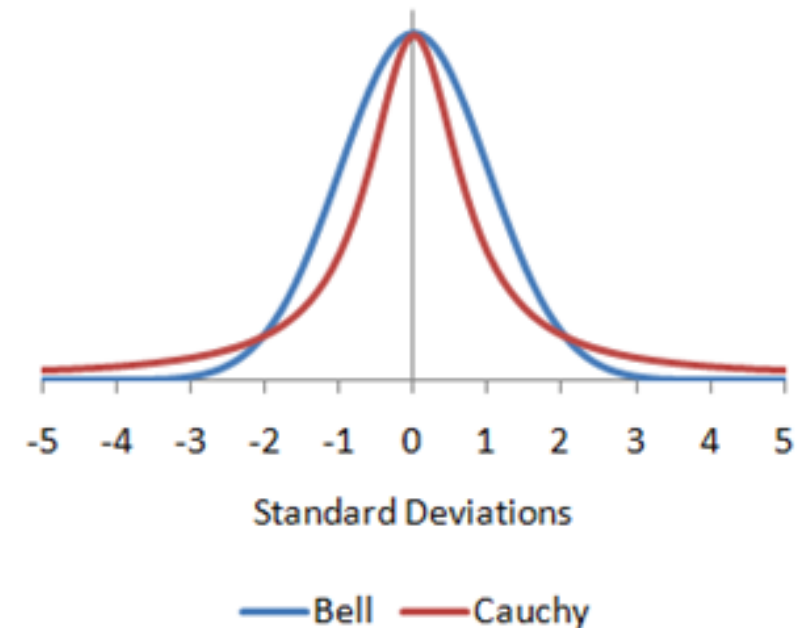
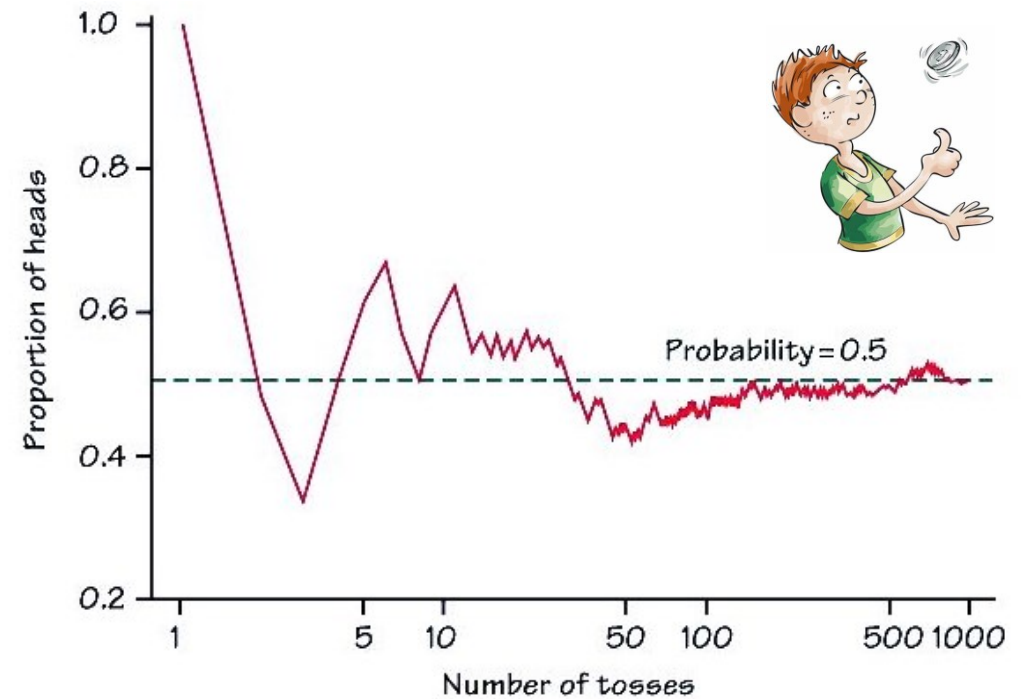


Zákon velkých čísel

Bernoulliho věta: relativní četnost sledovaného jevu stochasticky konverguje k jeho pravděpodobnosti.

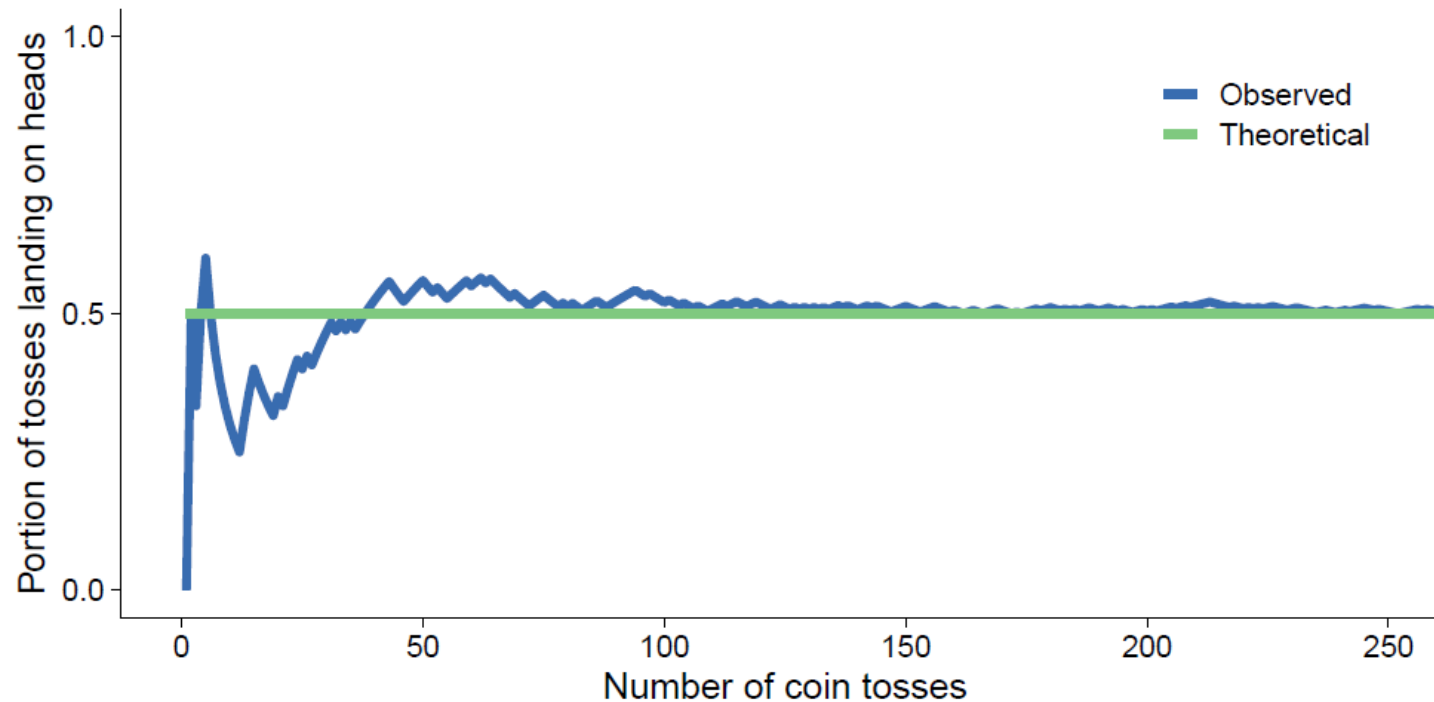
Čebyševova věta: pokud budeme opakovat stejný experiment (měření) dostatečně často, bude se průměrný výsledek blížit teoretické střední hodnotě rozdělení pravděpodobnosti.

Důsledek: u souborů s malým počtem jedinců je větší pravděpodobnost výskytu extrémních hodnot (čím menší soubor, tím větší rozptyl).



Zákon malých čísel

= nadhodnocení reprezentativnosti malých vzorků. Lidé mají tendenci věřit, že relativně malý počet pozorování je přesným a spolehlivým odrazem celkové populace.



Průměrný výkon u škol s malým počtem studentů obvykle vykazuje mnohem větší variabilitu než u škol s velkým počtem studentů.

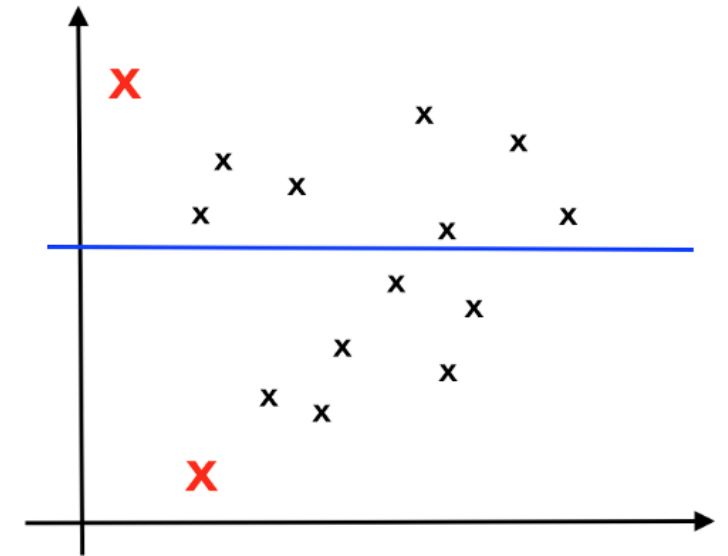
Regrese k průměru

Byly-li při prvním měření (pretest) získány extrémní hodnoty, při druhém měření (posttest) budou změřeny hodnoty blíže střední hodnotě populace. Objevuje se v každém systému, kde se kombinuje dovednost („talent“) a náhoda („štěstí“).

D. Kahneman:

úspěch = talent + štěstí

velký úspěch = talent + hodně štěstí



Studenti s horším výsledkem testu dosáhnou po absolvování doučovacího kurzu v obdobném testu v průměru lepší hodnoty. Tato změna se zdůvodňuje působením kurzu.

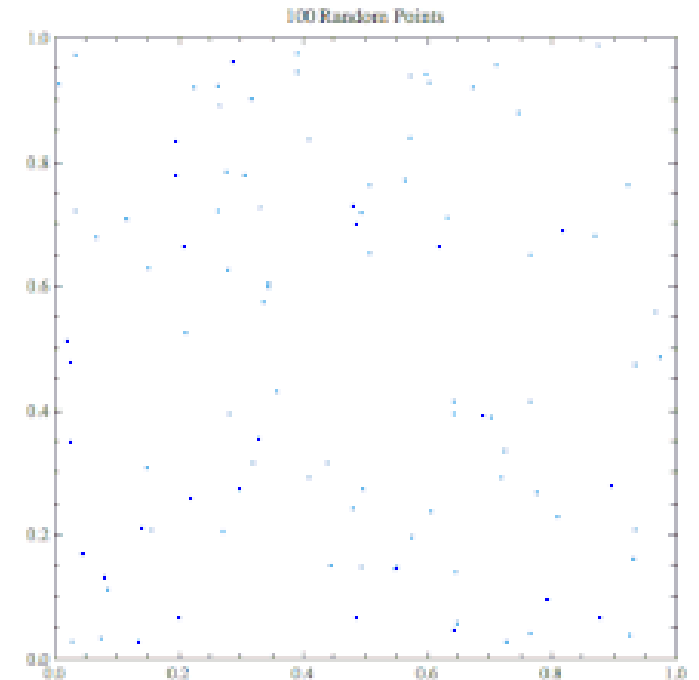
Iluze shlukování (iluze podobnosti, past hledání předloh, clustering illusion, hot hand bias) = snaha dát náhodnému jevu nějakou příčinu, tendence lidské psychiky vnímat náhodné shluky (klastry) v malém souboru náhodně rozmístěných dat jako nenáhodné.

Očekáváme velkou variabilitu, skutečně náhodné sekvence nám tak připadají seskupené a uspořádané. V případě řady vytvořené náhodnými hody mincí

OXXXOXXXOXXOOOXOOXXOO

je pro 11 prvních znaků je patrný zdánlivý trend.

Ověření náhodnosti např. Wald-Wolfowitzovým testem



Gamblerova iluze (gambler's fallacy, law of averages)

= chybná víra, že pokud je výskyt určitého jevu v dané chvíli mnohem častější než je běžné, je menší pravděpodobnost, že se tento jev vyskytne v budoucnu (nebo vice versa). Jde vlastně o snahu vidět změnu trendu v malém souboru dat.

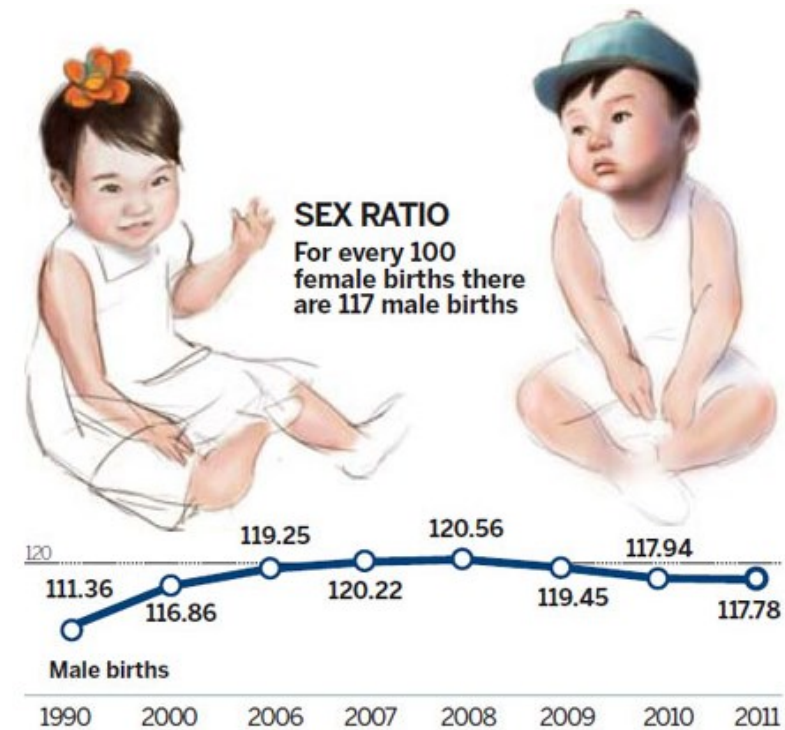


Někteří rodiče věří, že mají-li v řadě za sebou několik synů, resp. dcer, roste pravděpodobnost, že každý následující potomek bude pohlaví opačného.

Pravděpodobnost, že příští dítě bude syn je stejná u prvorodičů i u rodičů, kteří již mají 3 syny. Úvaha, že po narození 3 synů je vyšší pravděpodobnost narození dcery, je projevem **gamblerovy iluze**.

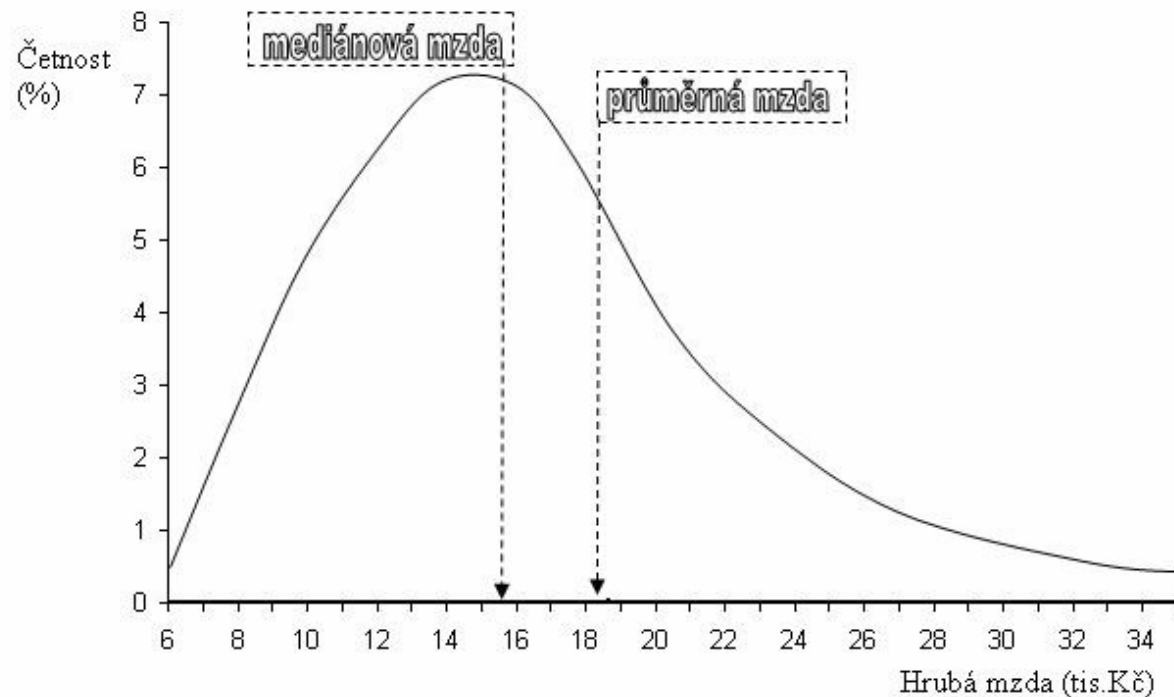


Turpinova rodina: 10 dcer a 3 synové



„Prokletí průměru“

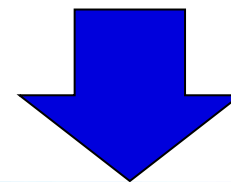
Průměrná mzda je příkladem nevhodného použití aritmetického průměru (vychýlený odhad), pod průměrem jsou téměř dvě třetiny zaměstnanců.



Průměr z extrémních hodnot

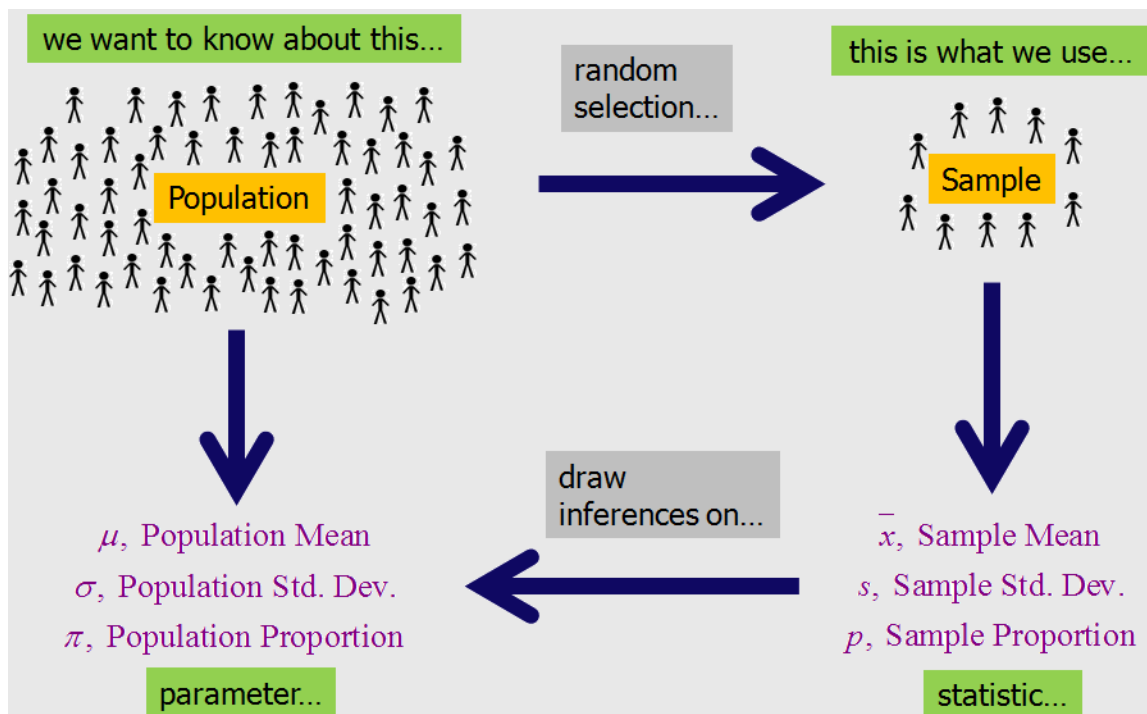
"z jedné strany hoří jako uhel vzňatý, z druhé strany mrzne v kámen ledovatý;"

K. J. Erben: Kytice (Záhořovo lože)



Statistické usuzování (statistická inference, statistická indukce)

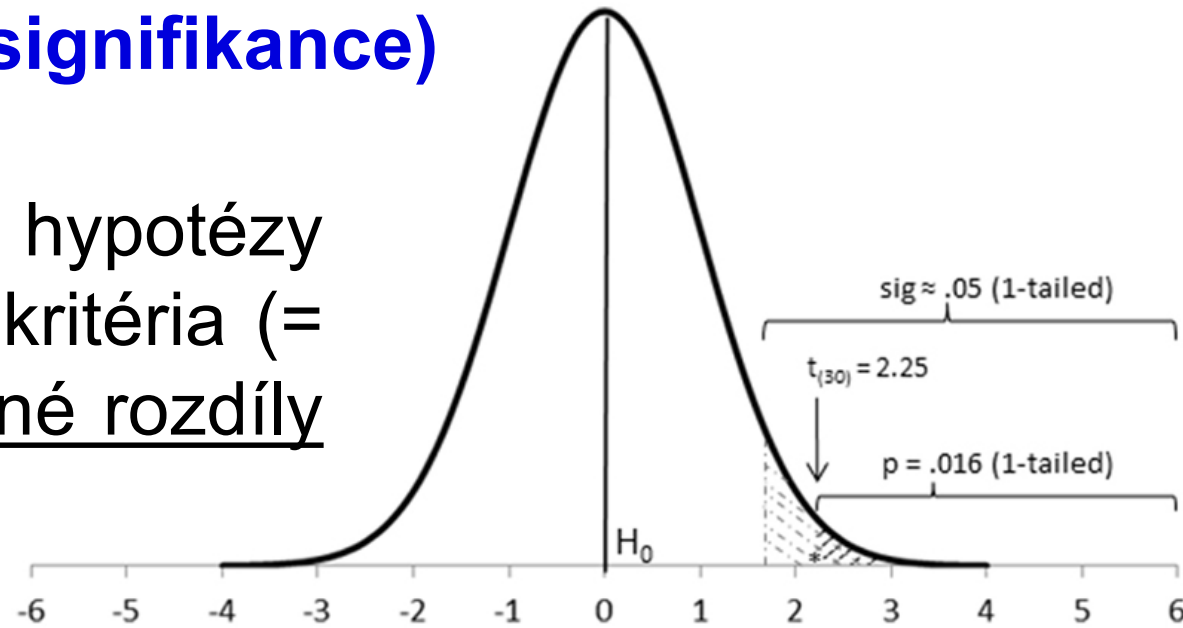
- usiluje o získání závěrů o základní populaci na základě analýzy a zobecnění výběrových dat, včetně zhodnocení spolehlivosti těchto závěrů, k čemuž využívá pravděpodobnostní pojmy (statistická inference, statistická indukce).



Hladina významnosti (α) říká, s jakou pravděpodobností jsme neučinili správný závěr. Pokud je tedy hladina významnosti 5 %, pak pouze v 1 z 20 případů bychom se dopustili chybného závěru (tzv. chyby prvního druhu).

Statistická významnost (p-hodnota, signifikance)

= pravděpodobnost zamítnutí nulové hypotézy určená na základě hodnoty testového kritéria (= odhad pravděpodobnosti, že pozorované rozdíly jsou jen důsledkem náhody).



Pokud je hodnota **p** menší než **hladina významnosti α** , zamítneme nulovou hypotézu. Test nulové hypotézy tedy spočívá v tom, zda hodnota p překročí, nebo nepřekročí, zvolenou hodnotu hladiny významnosti α (nejčastěji 0.05, tj. 5 %).

Při interpretaci výsledků testů je třeba mít představu o věcné významnosti a o minimálním věcně významném rozdílu.

| | H_0 je pravdivá (pozitivní skutečnost) | H_0 je nepravdivá (negativní skutečnost) |
|---|---|---|
| H_0 se nezamítá (pozitivní výsledek testu) | Správné rozhodnutí (pravdivě pozitivní rozhodnutí) $P = 1 - \alpha$ Spolehlivost testu | Chyba II. druhu (falešně pozitivní rozhodnutí) $P = \beta$ Hladina významnosti |
| H_0 se zamítá (negativní výsledek testu) | Chyba I. druhu (falešně negativní rozhodnutí) $P = \alpha$ Hladina významnosti | Správné rozhodnutí (pravdivě negativní rozhodnutí) $P = 1 - \beta$ Síla testu |

Spíše než „statisticky významný“ by se hodilo spíše „statisticky detekovatelný“.

Test významnosti je detektiv, nikoliv soudce !!

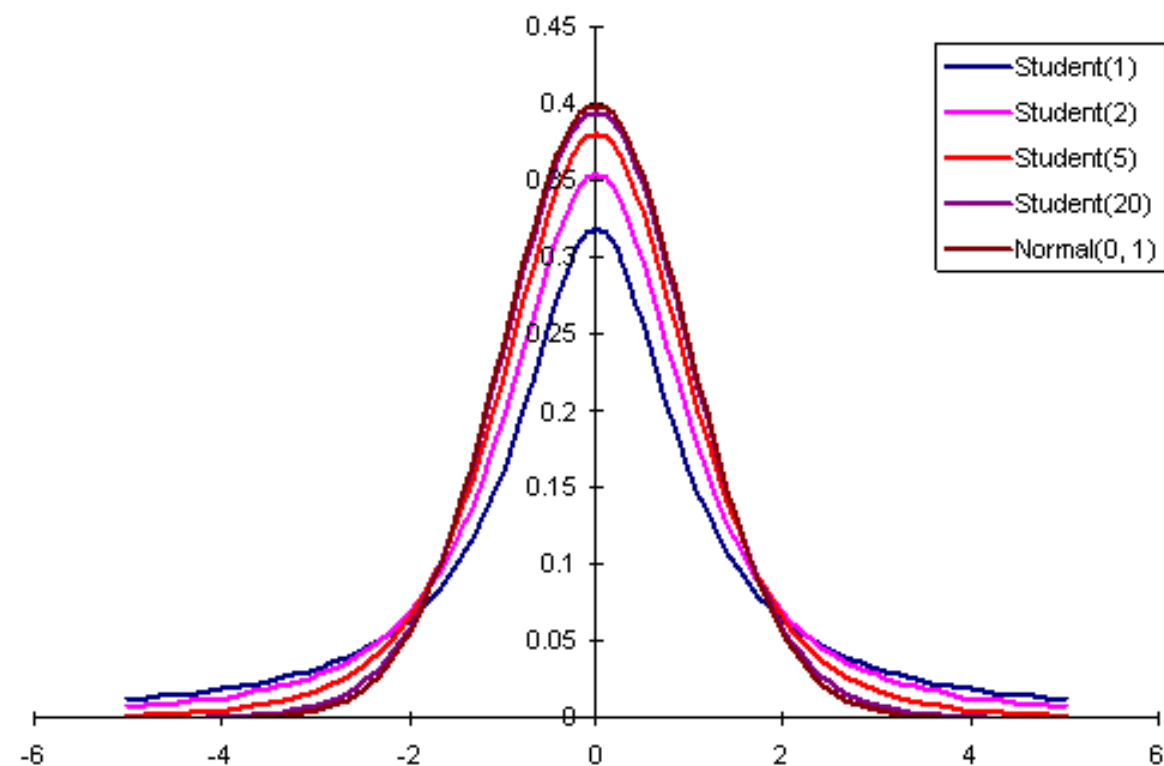
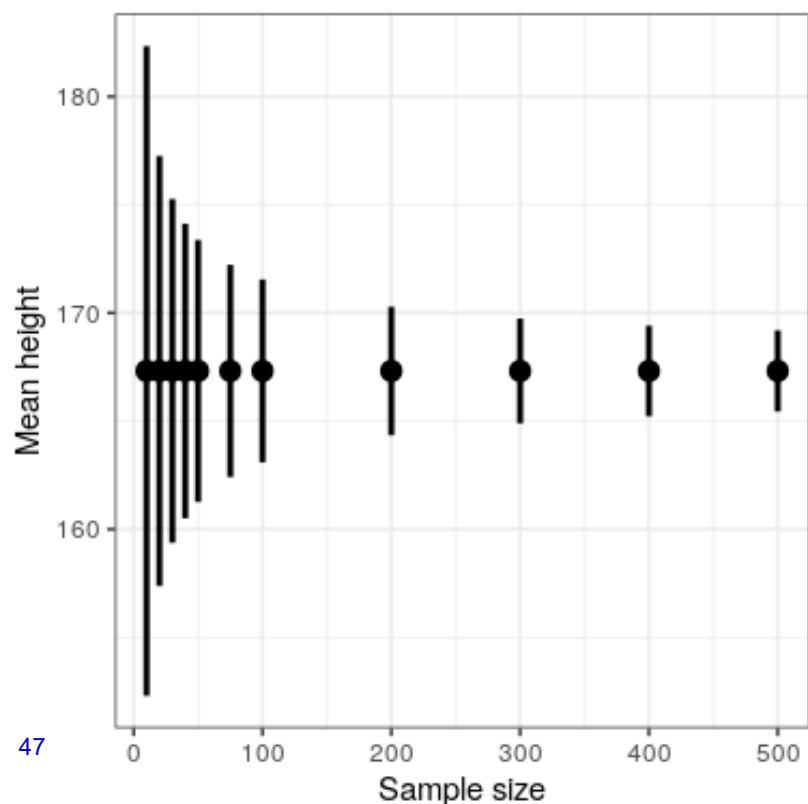
Chyba 1. druhu (klam „pozitivních“ výsledků) = odhalení statisticky významného rozdílu mezi soubory, i když ve skutečnosti mezi nimi neexistuje reálný rozdíl. Typickou příčinou je příliš velké datové soubory (a tím pádem velmi úzké intervaly spolehlivosti). Náhodný rozdíl mezi velmi početnými soubory se tak často jeví jako statisticky signifikantní.

Takové výsledky se uveřejňují spíše než výsledky negativní (tzv. **publication bias**).

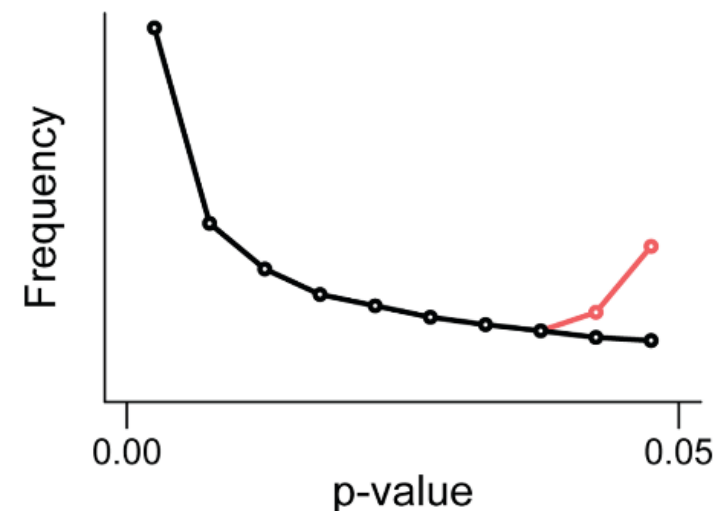
Badatelé, kteří nezjistí očekávaný „pozitivní“ efekt mohou někdy sami sebe přesvědčit, že důvodem neúspěchu je **chyba 2. druhu**.

Efekt prokletí vítěze – původní statisticky významný výsledek nelze spolehlivě reprodukovat ani mnohonásobným opakováním pokusů.

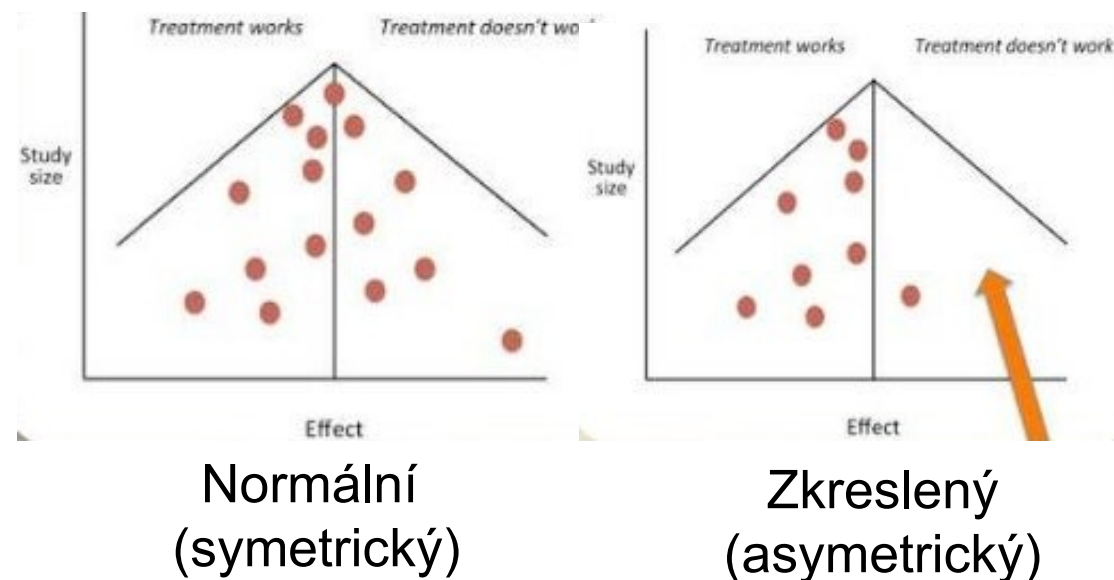
Chyba 2. druhu (klam nevýznamného významu) = mylné domněnání, že mezi skupinami neexistuje rozdíl. K této chybě dochází např. vlivem malé velikosti souboru (a tím pádem širokých intervalů spolehlivosti). Reálný rozdíl mezi málo početnými soubory se tak často nepodaří prokázat, protože není statisticky signifikantní.



p-hacking (data dredging) = cílené hledání statisticky významných výsledků, obvykle vyřazením „podezřelých“ hodnot ze souboru (**manipulace s výsledky**).



Publikační zkreslení (publication bias, „efekt šuplíku“) - výsledek experimentu nebo výzkumné studie ovlivňuje rozhodnutí, zda ji publikovat (předpojatost ve prospěch pozitivních, resp. statisticky významných výsledků). Jde o projev **klamu přeživších**.



| | Drogy přítomny | Drogy nepřítomny |
|--------------------------------|---|--|
| Pes upozorní psovoda na pach | Správné upozornění (těm věnujeme pozornost) | Falešný nález (o nich neslyšíme) |
| Pes neupozorní psovoda na pach | Falešné negativum (o tom se nikdo nedozví) | Správné neupozornění (nikdo si na ně nevzpomene) |

Bayesovská statistika je obor moderní statistiky, která pracuje s podmíněnými pravděpodobnostmi a dovoluje zpřesňovat pravděpodobnost výchozí hypotézy, jak se objevují další relevantní skutečnosti. Jádrem jejího matematického aparátu je Bayesova věta. Má rozsáhlé využití všude tam, kde se pracuje s nejistými znalostmi

Bayesova věta

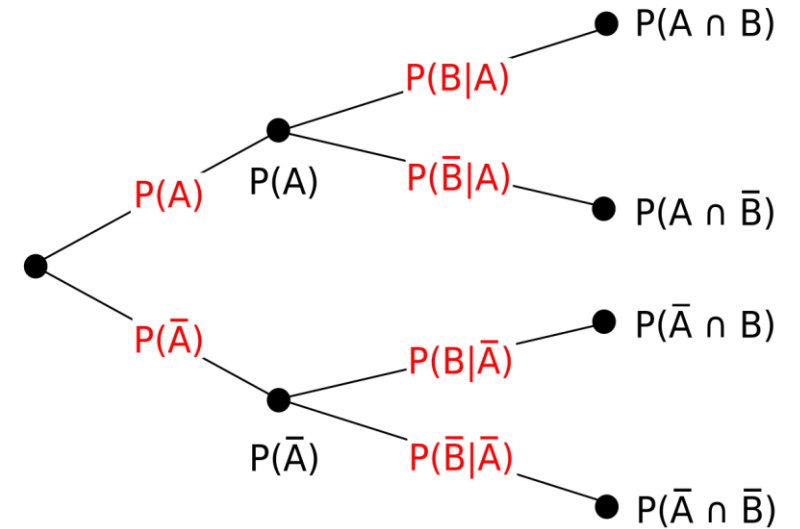
Bayesova věta popisuje pravděpodobnost výskytu události související s nějakou podmínkou. Udává, jak podmíněná pravděpodobnost nějakého jevu souvisí s opačnou podmíněnou pravděpodobností.

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)}$$

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B | A) = \frac{P(A \cap B)}{P(A)}$$

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B | A) \cdot P(A) + P(B | \neg A) \cdot P(\neg A)}$$



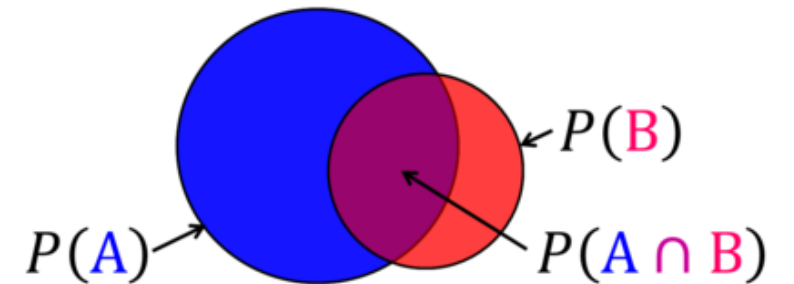
probability a hypothesis is true given the evidence

probability a hypothesis is true (before any evidence is present)

probability of seeing the evidence if the hypothesis is true

probability of observing the evidence

$$P(H/E) = \frac{P(H) P(E/H)}{P(E)}$$



V případě vraždy byl získán vzorek vrahovy DNA, přičemž je 0.1% šance že vzorek může být mylně přiřazen jinému člověku. Jaká je pravděpodobnost, že je vrahem muž, u něhož byla zjištěna shoda se vzorkem?

Chybná interpretace: je 0.1% šance že muž je nevinný a 99.9% šance, že je vrah.

Bayesovská interpretace: kromě pravděpodobnosti získané z výsledků analýzy DNA bere v úvahu i pravděpodobnost jevu, že zcela náhodná osoba může být vrahem (tato pravděpodobnost je asi 0.01%).

Máme tedy 2 tvrzení:

A = muž je vinen

B = mužova DNA se shoduje se vzorkem

Odtud

$$P(A) = 0.01\%$$

$$P(B) = 0.1\%$$

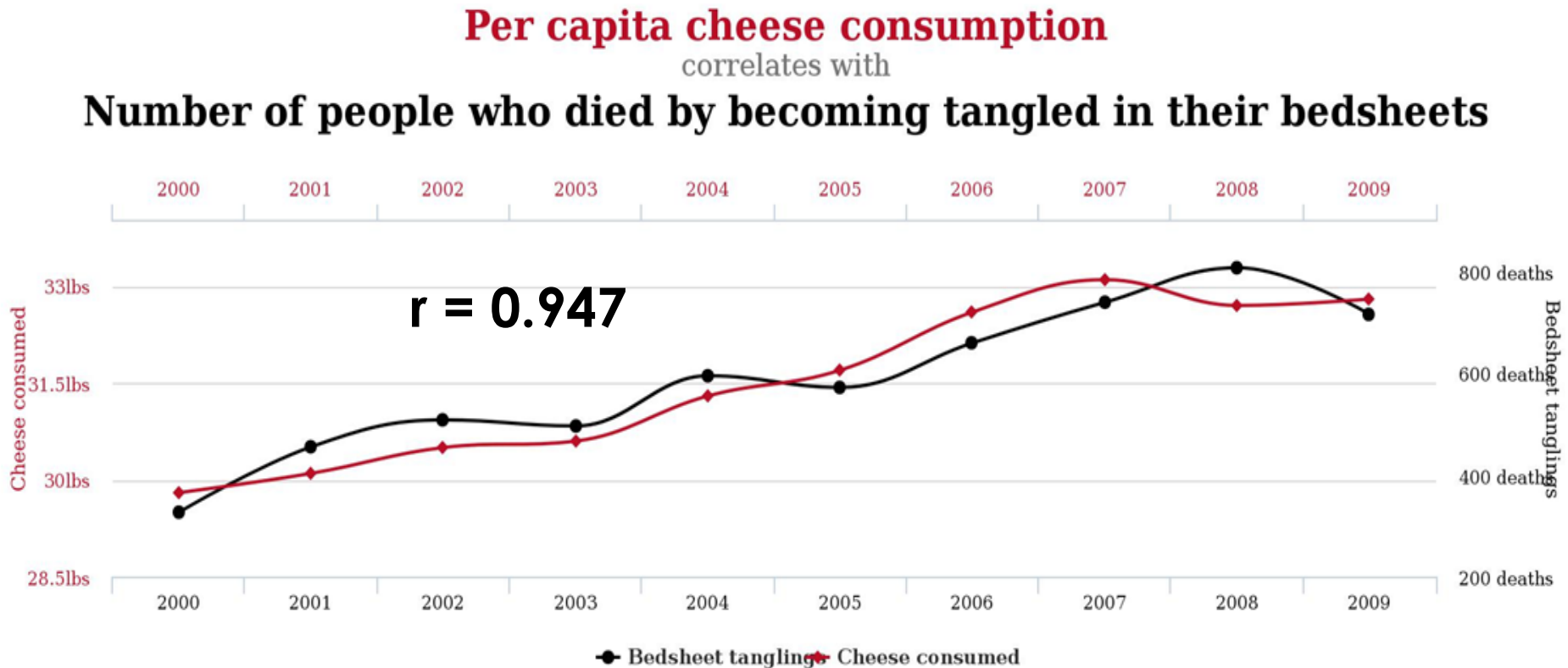
Pravděpodobnost, která nás zajímá je $P(A|B)$: pravděpodobnost, že muž je vinen, za předpokladu, že byla nalezena shoda jeho DNA s vrahovou.

Z Bayesovy věty plyne $P(A|B) = P(B|A) \cdot P(A) / P(B)$

$P(B|A)$ je pravděpodobnost, že se mužova DNA shoduje s vrahovou: $P(B|A) = 1$. Odtud po dosazení $P(A|B) = 10\%$.

Falešná korelace (spurious correlation)

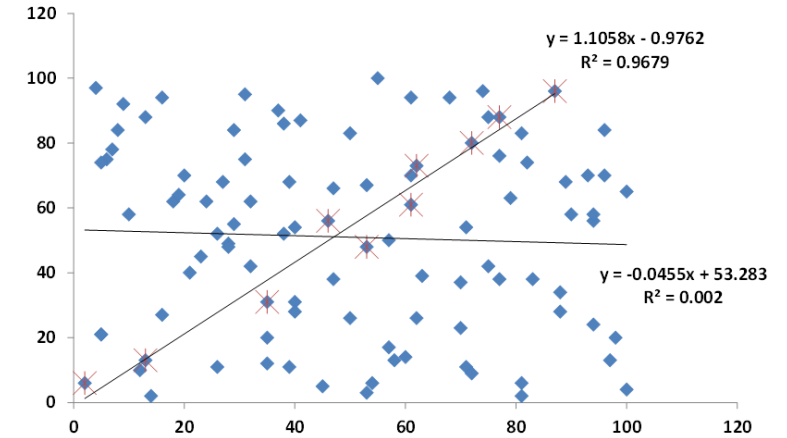
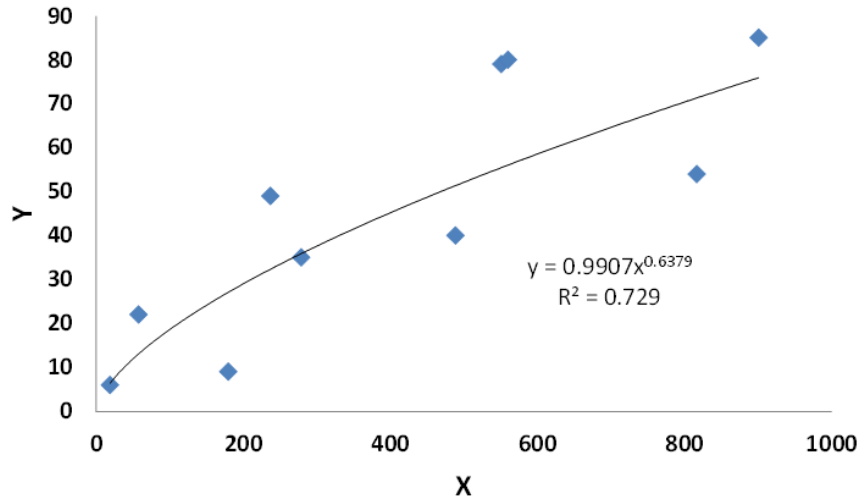
= pozorované korelace mezi proměnnými jsou pouze náhodné nebo jsou obě proměnné ovlivněny nějakou neviditelnou třetí proměnnou.



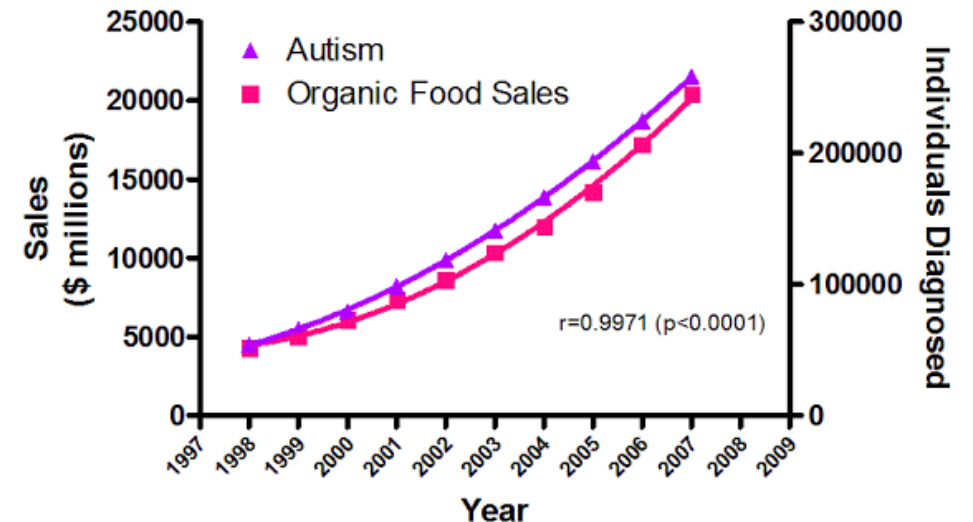
tylervigen.com

MUNI

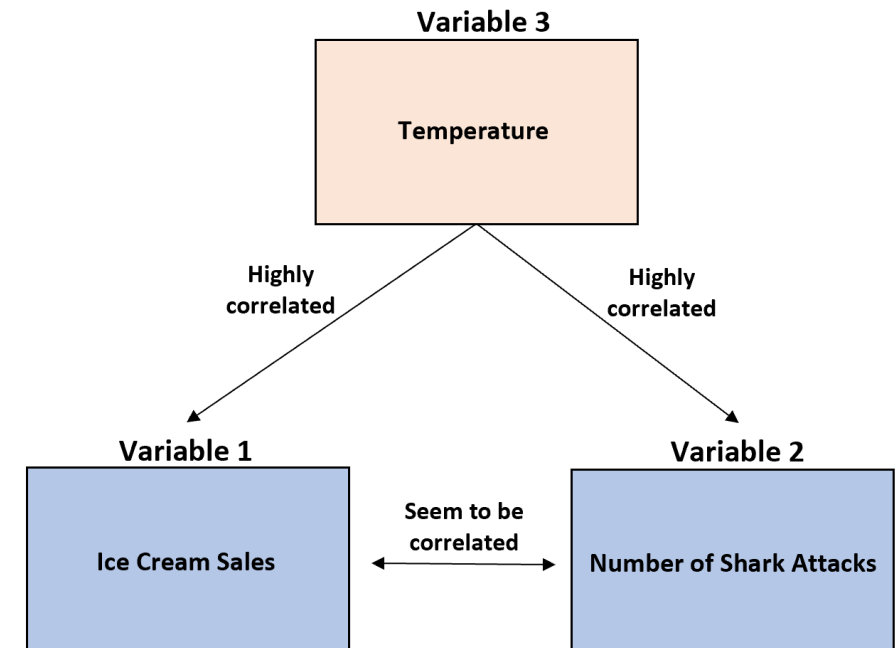
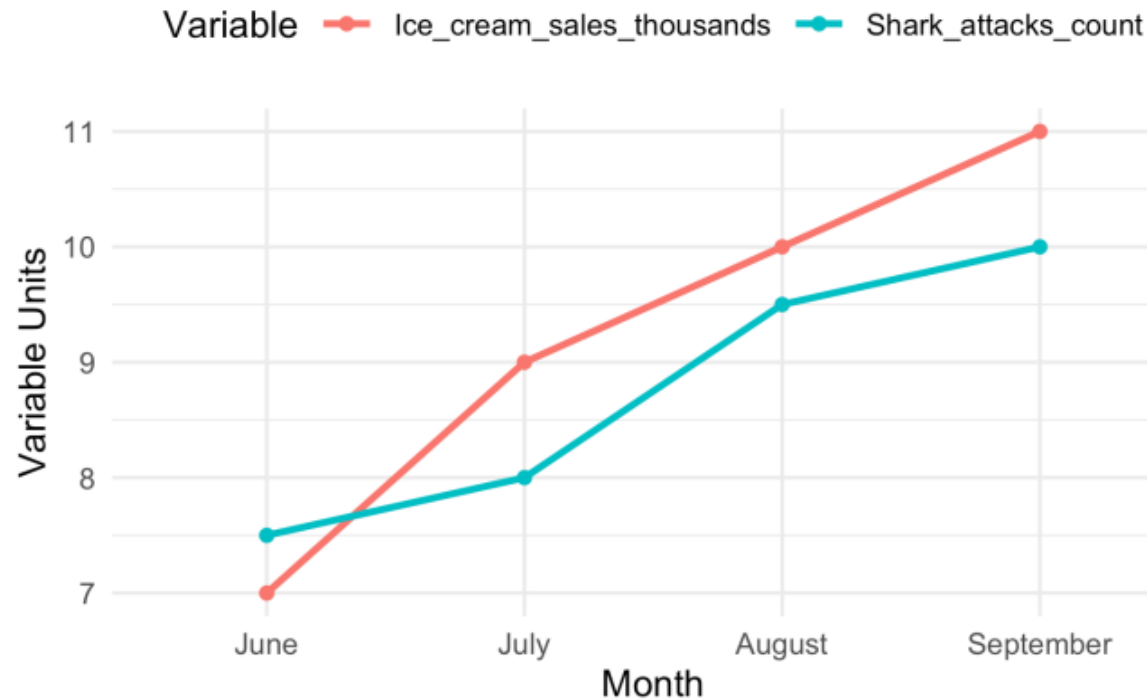
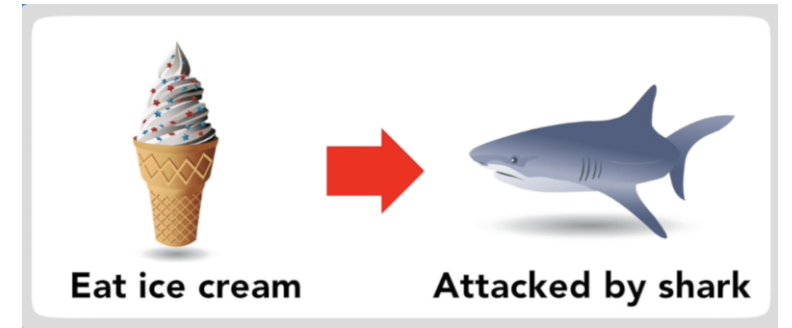
Zákon malých čísel: vnímání řady dat jako nenahodilé je dáno chybným úsudkem o variabilitě, která by se měla objevit v menším vzorku dat.



Nepříčinný časový vztah: jakékoli dvě nezávislé proměnné, které se lineárně mění s časem vykazují dokonalou korelaci (oba procesy se dějí zároveň).



Skrytá proměnná = sledované proměnné mohou být obě ovlivněny nějakou další, neviditelnou, proměnnou („problém třetí proměnné“).



Korelace vs. kauzalita

Kauzalita: jakákoli změna hodnoty jedné proměnné způsobí změnu hodnoty jiné proměnné



příčina

"The Usual"



Reverse Causality

důsledek



Simultaneity

souběžnost

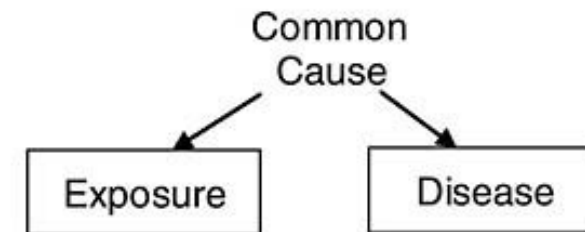


koincidence = náhodná asociace

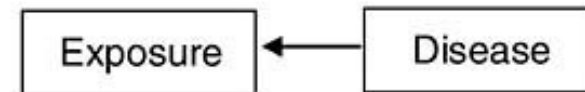
Causal Association



Spurious Association



Reverse Causality

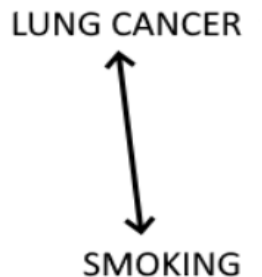


Korelace (asociace) neimplikuje kauzalitu: pokud nějaké dvě proměnné spolu korelují, nelze z toho vyvozovat, že jedna je příčinou a druhá kauzálním následkem. Kauzalita sice není vyloučena, ale nelze ji pouhou korelací prokázat. Lidé běžně za příčinu považují ten faktor, který nejvíce koreluje s následkem.

Jednocestný blud: pokud je X asociováno s Y v tom, že je vždy předchází, pak je vždy možnost, že Y je příčinou a ne důsledkem X (viz **efekt plavecké figury**).

Nezbytná a postačující příčina: i když je asociace mezi X a Y příčinná, nevyplývá z toho, že každé X bude následováno Y. Nezbytná příčina tedy není vždy postačující.

Všichni kuřáci nezemřou na rakovinu plic a ne všichni lidé, kteří na rakovinu plic umírají, jsou kuřáci. Kouření tak není nezbytnou ani postačující příčinou.



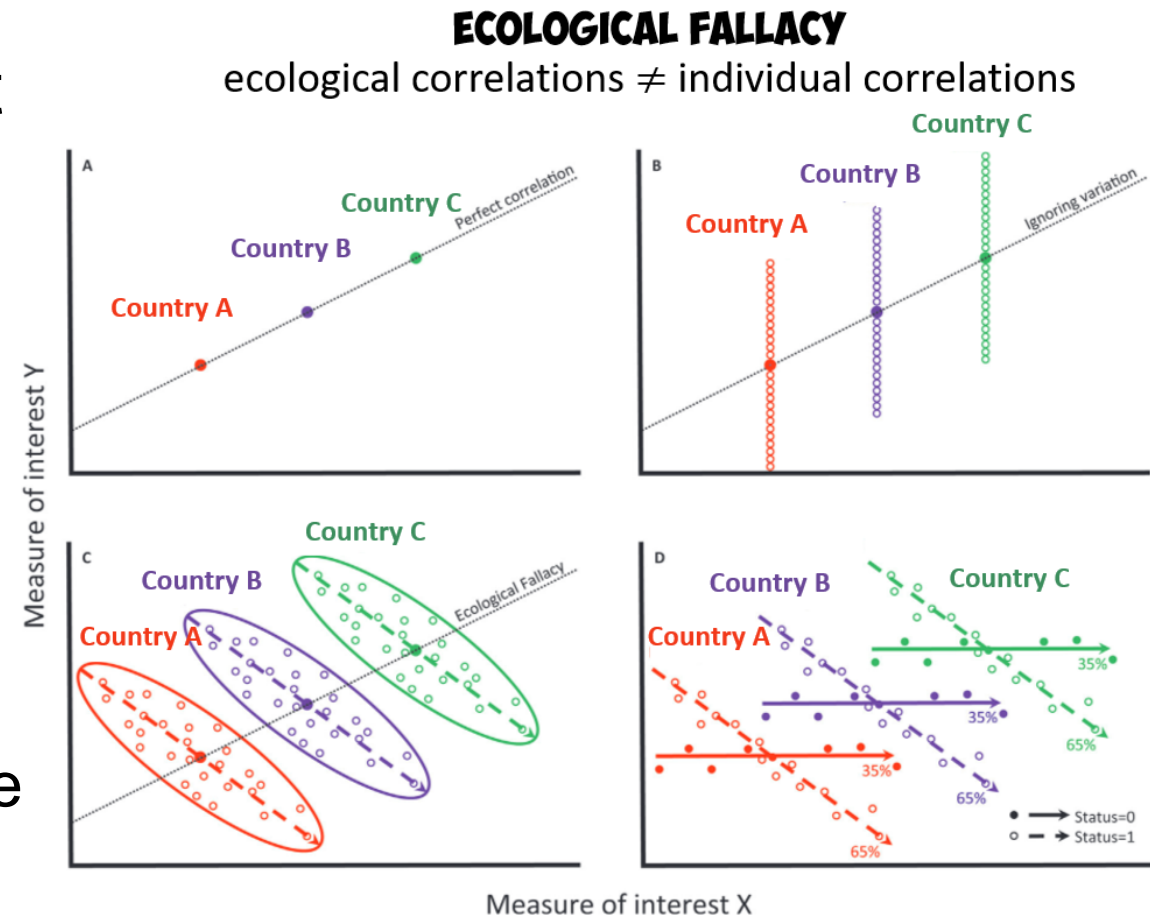
Ekologický klam (ecological fallacy)

= zkreslení, jež může nastat při extrapolaci výsledků agregovaných statistických dat pro daný soubor na jednotlivce z tohoto souboru.

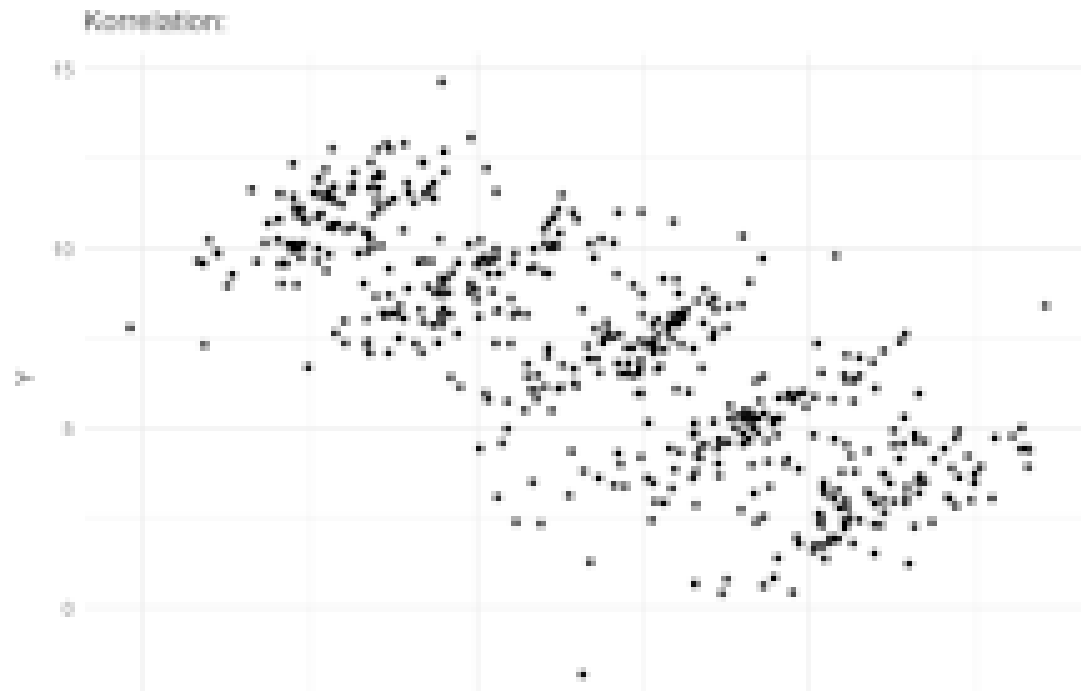
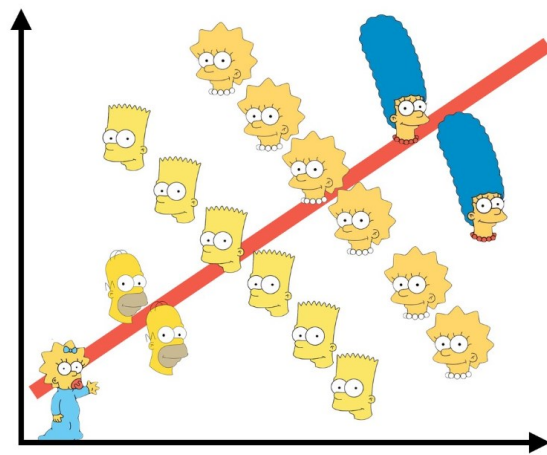
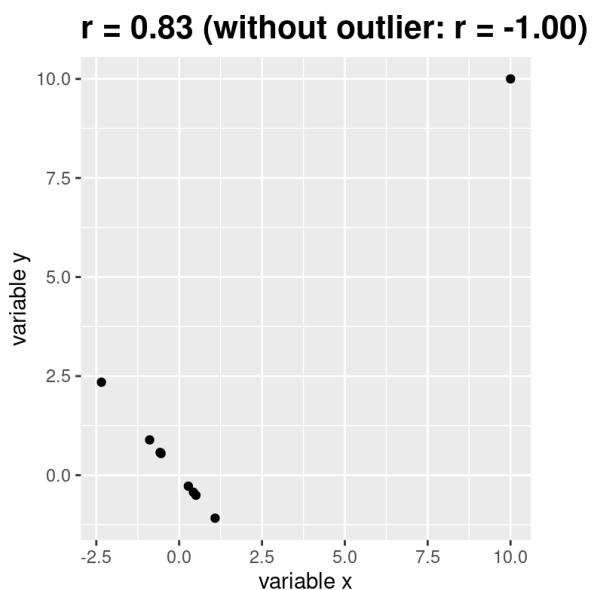
Lidé tíhnou k tomu, přisuzovat jednotlivým členům skupiny průměrné charakteristiky dané skupiny.

- tečkované čáry odrážejí korelace prostřednictvím agregovaných dat
- barevné čáry odrážejí korelace jednotlivých úrovní.

Speciálním druhem ekologického efektu je tzv. Simpsonův paradox.



Simpsonův paradox je důsledkem nehomogenity dat. Soubor tvořen více skupinami s různými hodnotami průměrů a/nebo kovariancí (= nehomogenní soubor), může pro kontinuální data vykazovat falešnou korelaci (spurious correlation).

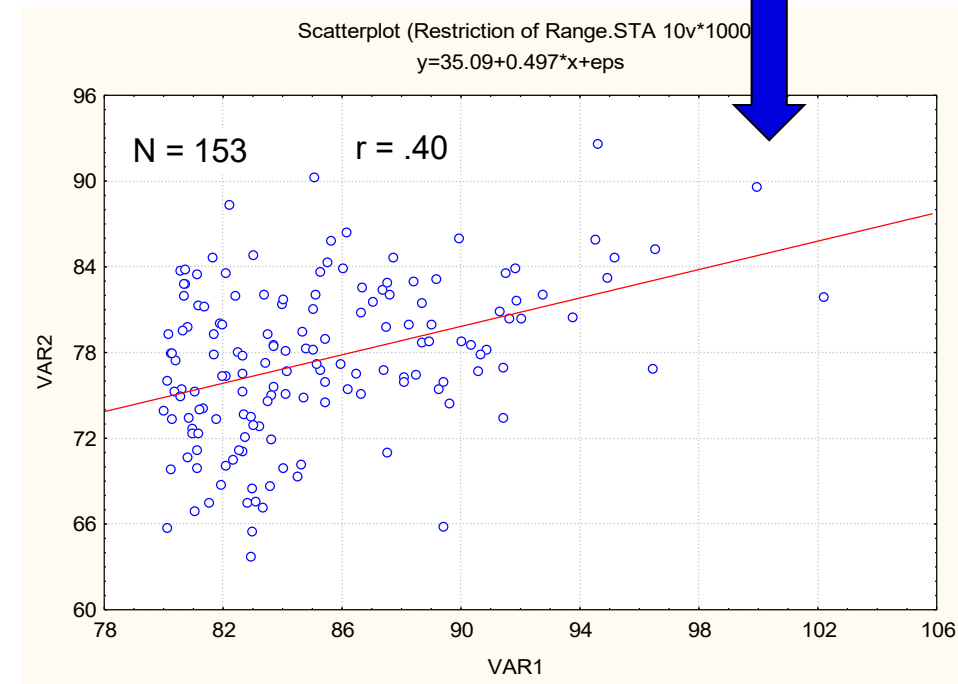
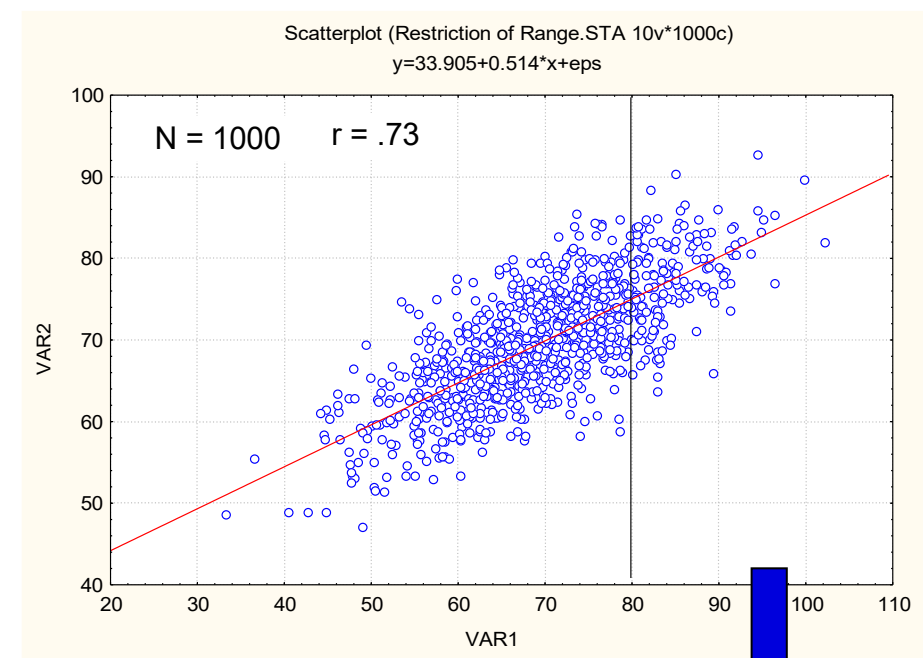


Ve statistikách dochází ke zkreslení, pokud statistický model **vynechá jednu nebo více relevantních proměnných**. To vede k modelu, který připisuje účinek chybějících proměnných těm, které byly zahrnuty. **M U N I**

Berksonův paradox

Dvě proměnné, které jsou za normalních podmínek nezávislé, se za jistých okolností mohou jevit závislými (nebo se směr závislosti může obrátit).

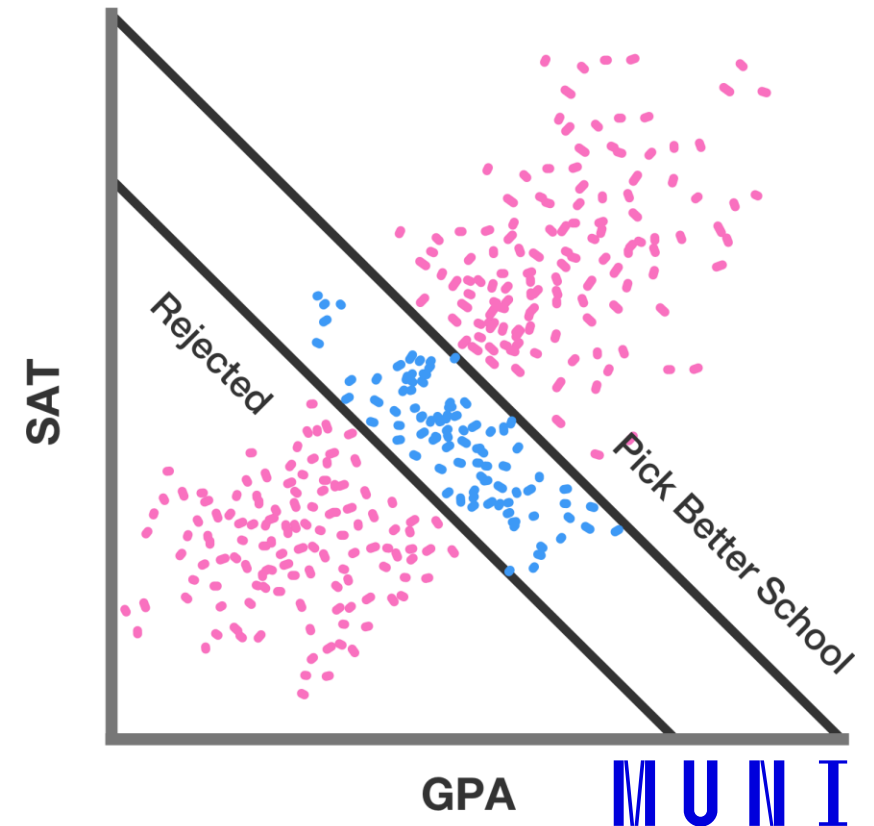
Dostupné hodnoty jsou v tomto případě ořezanou verzí celého souboru dat (např. **klam přeživších**). Korelace (a jiné popisné statistiky) takto omezeného souboru může být podhodnocena (nebo nadhodnocena).



Berksonův paradox

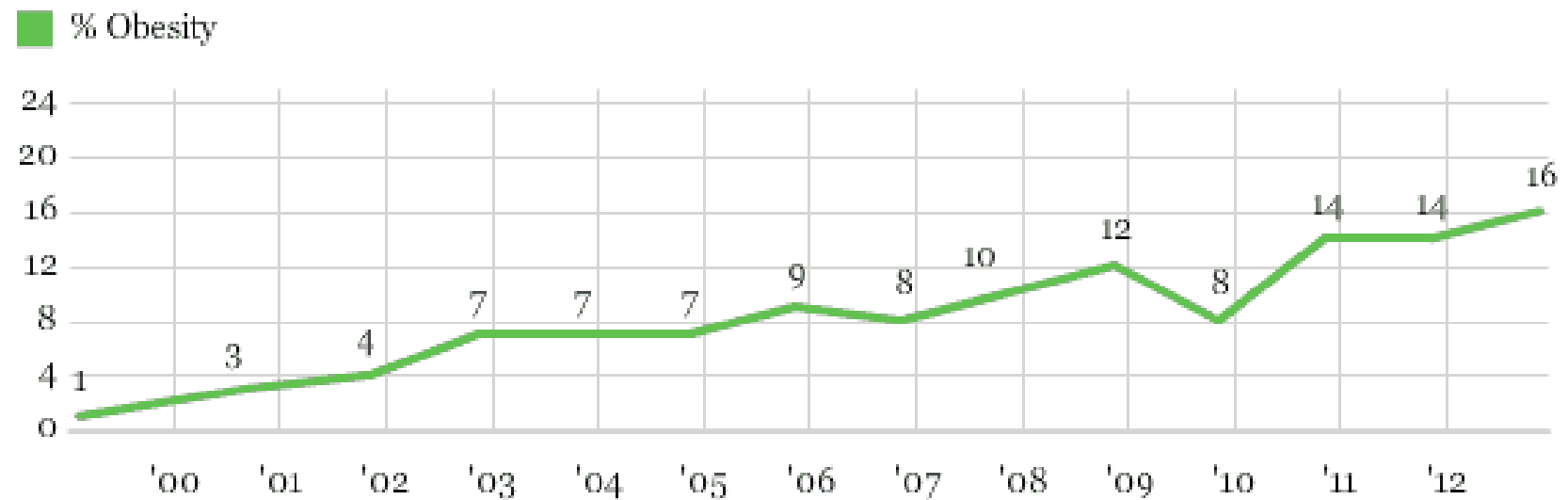
Univerzity v USA si často vybírají studenty na základě průměrného prospěchu na střední škole (GPA) a výsledky testů SAT - ty spolu pozitivně korelují. Přijímání jsou studenti s dostatečně vysokým GPA, nebo dostatečně vysoké skóre SAT, nebo nějakou kombinací obou.

Z grafu je patrné, že nikdo pod spodní čarou nebyl přijat a nikdo nad horní čarou se nerozhodl navštěvovat školu (uchazeči, kteří mají jak vysoký průměr GPA, tak vysoké skóre SAT, se pravděpodobně dostanou na prestižnější školu a proto nenastoupí). Studenti, kteří na školu skutečně nastoupí (jsou v grafu vyznačeny modrými tečkami) vykazují klesající trend, přestože celková populace (červené a modré tečky) vykazuje trend rostoucí .



Lineární funkci intuitivně rozumíme, modelované jevy by měly mít alespoň přibližně lineární průběh (mezi proměnnými x a y je lineární vztah).

Exponenciální funkce resp. **procentuální růst** nejsou intuitivně pochopitelné.



Počet obézních jedinců v populaci (v %) může dosáhnout maximálně 100 %, tj. v budoucnu tedy dojde k „ohnutí“ závislosti (nic co narůstá procentuálně neroste věčně).

Kargokultická statistika

= mechanické, rituální používání statistických metod bez pochopení jejich předpokladů, omezení a interpretace. Statistické termíny a postupy jsou používány jako zaklínadla, bez porozumění předpokladům či významu výpočtů, resp. terminologie.

Zavádění statistického softwaru se řídí především snadností použití a rozsahem statistických postupů, které software implementuje. Čím je software "výkonnější" a "uživatelsky přívětivější", tím více vybízí ke **kargokultické statistice**.



Nepoučený uživatel statistického software může často založit zásadní rozhodnutí na základě:

- 1) volby nesprávné metody analýzy dat, která poskytne nesmyslné výsledky
- 2) nesprávné interpretace správných výsledků analýzy dat

„Kargokultická“ věda

Fyzik Richard Feynman označil tímto pojmem obory, které sice vědu napodobovaly (měly časopisy, katedry, konference a další vnější atributy vědy), ale jejich výstupy neměly žádnou vědeckou hodnotu.

Označení vycházelo z tzv. „kargo kultu“ popsaného antropology u obyvatel Nové Guineje.



Kvalita interpretace výsledků

Na kvalitu interpretace výsledků (rozhodování) má vliv úplnost (neúplnost) informací a náhoda („štěstí“). Interpretace výsledků může vycházet z:

výsledků (intuitivní, „zkreslení výsledkem“)

procesu rozhodování

Samotná kvalitní data (výsledky) nezaručují jejich správnou interpretaci.

Lucasova kritika: osvědčené modely časem ztrácejí svoji platnost: co platilo dříve, již platit nemusí. I přesto, že tyto modely mohly být původně správné, jejich parametry se časem změnily tak, až se z nich staly modely nepoužitelné. Lucasovu kritiku lze považovat za variantu **kobřího efektu**.

Desinterpretace výsledků

Potvrzovací zkreslení (confirmation bias, myside bias, disconfirming evidence) = tendence zformulovat hypotézu a pak hledat cesty k jejímu potvrzení (tzv. „paradox texaského střelce“). Přehlížení faktů popírající naše přesvědčení (názory), výklad nejednoznačné informace tak, aby s nim byla v souladu.

Klam authority souvisí s tím, že dané tvrzení zastává někdo významný, objevuje v učebnicích či médiích, případně je obecně přijímáno v komunitě. Viz **haló efekt** a **zkušenostní zkreslení**.

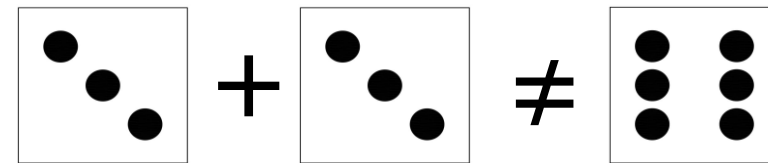
Past jediného důvodu (atribuční chyba, fallacy of the single cause): předpoklad jedné významné příčiny, namísto současného a vzájemného působení řady různých vlivů.

Slepota vůči alternativám (tunelové vidění): jsme-li postaveni před volbu A a zachování dosavadního stavu, máme tendenci porovnávat jen obě možnosti pouze mezi sebou a již neuvažovat případné další možnosti (B, C, ...).

Kotvení (anchoring effect): člověk se při rozhodování soustředí na jednu věc (informaci nebo rys), tzv. „kotvu“, a nevnímá širší souvislosti či další proměnné, které by v rozhodování mohly hrát klíčovou roli.

Klam konjunkce (conjunction fallacy, klamná shoda okolností): při rozhodování se opíráme o omezený počet faktorů, o nichž si myslíme, že danou situaci dostatečně charakterizují. Lidé mají tendenci tíhnout k příběhům, které působí „věrohodně“.

Klam Svatoplukových prutů je projevem víry, že více slabých důkazů či indicií je ekvivalentní spolehlivému důkazu.



Past indukce: sklon k vyvozování obecných závěrů z jednotlivostí.

Storytelling (past dobrého příběhu): vytváření souvislého příběhu z nesouvisejících informací (příběh, který chceme slyšet).

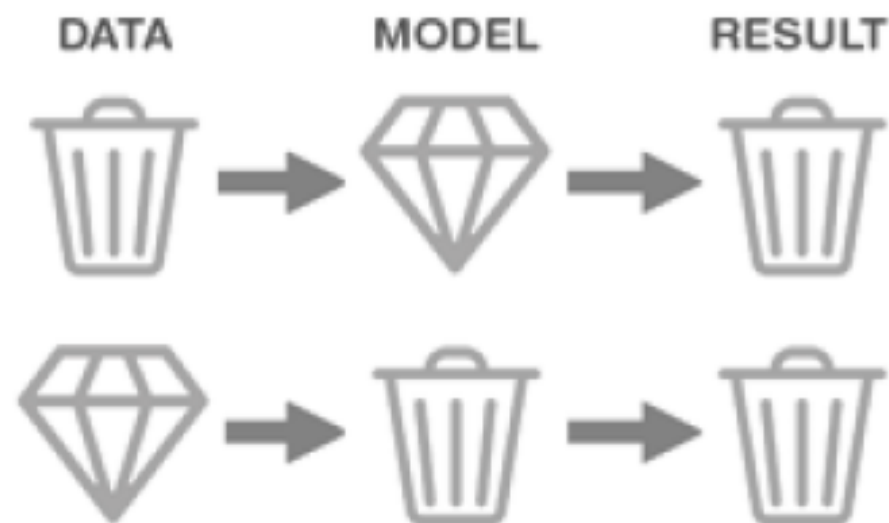
Efekt plavecké figury: záměna příčin za jejich důsledky.

GIGO (Garbage In, Garbage Out)

Chybná nebo nesmyslná vstupní data produkují chybný, nesmyslný a bezcenný výstup.

Nevhodná či nesprávná metoda analýzy dat, poskytne nesmyslné výsledky

Nesprávná interpretace správných výsledků analýzy dat vede k chybným závěrům.



Jak se vyhnout nesprávné interpretaci dat a kognitivnímu zkreslení?

Informace (data) nestačí jen sbírat a intuitivně vyhodnocovat. Je potřeba se nad nimi zamyslet a správně pochopit. Můžeme si položit například tyto otázky:

- Co nám data ve skutečnosti říkají?
- Co výsledky znamenají?
- Co nám chybí?
- Co nevíme?
- Zamýšlíme se nad nimi správně?
- Jak se na ně můžeme dívat nebo je chápat jinak?



„Učitel nemůže nikdy skutečně učit, pokud se stále neučí sám. Lampa nemůže nikdy rozsvítit jinou lampu, pokud nebude nadále hořet svým vlastním plamenem. Učitel, který studentům svou lekci pouze opakuje, může zatížit jejich mysl, ale nemůže je povzbudit.“

Rabíndranáth Thákur

Děkuji za pozornost !

prokes@chemi.muni.cz