Jiří Dvořák; Marie Snětinová

Simpsonův paradox

Rozhledy matematicko-fyzikální, Vol. 97 (2022), No. 1, 29–34

Persistent URL: http://dml.cz/dmlcz/151064

Simpsonův paradox

Jiří Dvořák, Marie Snětinová, MFF UK, Praha

Abstrakt. V tomto příspěvku představíme neintuitivní a na pohled paradoxní situaci, kdy například hráč může být v každém kole horší než jeho soupeř, ale na konci zápasu je stejně prohlášen vítězem. Ukazujeme, že jde čistě o volbu úhlu pohledu: má zvítězit ten, kdo je v každém kole lepší, nebo ten, kdo je celkově lepší? Tato volba může mít v praxi závažné důsledky a při formulaci závěrů je vždy žádoucí být pečlivý a opatrný. Co když změna úhlu pohledu otočí závěry naruby?

Zkouška je od slova „zkusit"

Představme si dva hypotetické kamarády, Aloise a Bedřicha, spolužáky v prvním ročníku vysoké školy. V zimním semestru Alois podstoupil osm zkoušek, ve dvou z nich uspěl (25% úspěšnost). Bedřich v zimním semestru podstoupil pět zkoušek, uspěl jen u jedné (20% úspěšnost). Alois si nemohl odpustit rýpnutí: „Podívej, Bedřichu, jsem lepší student!"

S odřenýma ušima oba studenti postoupili do letního semestru a v letním zkouškovém období Alois podstoupil dvě zkoušky, v obou uspěl (100% úspěšnost). Bedřich podstoupil pět zkoušek, ve čtyřech uspěl (80% úspěšnost). Po poslední zkoušce to Alois zhodnotil jednoznačně: „Vidíš, pořád jsem lepší." Bedřich se ale nedal: „Nene, podívej: každý jsme měli dohromady deset zkoušek, tys uspěl ve čtyřech, já v pěti. Já byl lepší!"

Ponechme stranou otázku, který student byl „ve skutečnosti" lepší. Oba předkládají korektní argumenty, oba svá tvrzení opírají o pravdivá data, viz tabulku 1.

	Zimní semestr		Letní semestr		Celkem	
Alois	2/8	**25 %**	2/2	**100 %**	4/10	40 %
Bedřich	1/5	20 %	4/5	80 %	5/10	**50 %**

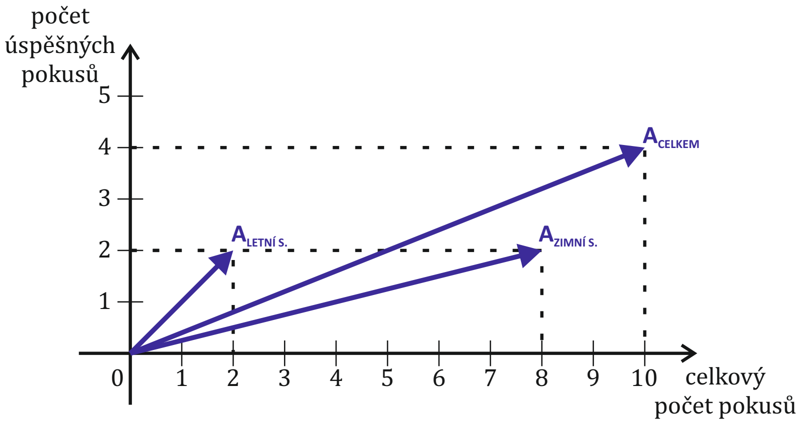
Tabulka 1: Poměr úspěšně složených zkoušek ku počtu všech skládaných zkoušek v jednotlivých semestrech i celkově

Jak je možné, že se jejich závěry tak liší? Procentuálně vzato, v každém ze semestrů byl úspěšnější Alois (25 % vs. 20 % v zimním semestru, 100 % vs. 80 % v letním semestru), celkově za akademický rok byl úspěšnější Bedřich (50 % vs. 40 %). To ale (možná překvapivě) není v rozporu. Pro svou neintuitivnost je tento fenomén označován jako „Simpsonův paradox“.

Užitečný způsob, jak se v daném problému vyznat, poskytuje vektorová reprezentace. Aloisův zimní semestr můžeme znázornit pomocí vektoru (8,2), neboli dva úspěchy z osmi pokusů, viz obr. 1. Sklon tohoto vektoru potom určuje Aloisovu úspěšnost: $2/8 = 0{,}25$. Aloisův letní semestr pak znázorníme pomocí vektoru (2,2). Aloisovo snažení za oba semestry se pak dá popsat součtem těchto dvou vektorů:

$$(8, 2) + (2, 2) = (8 + 2, 2 + 2) = (10, 4),$$

neboli čtyři úspěchy z deseti zkoušek, viz obr. 1.

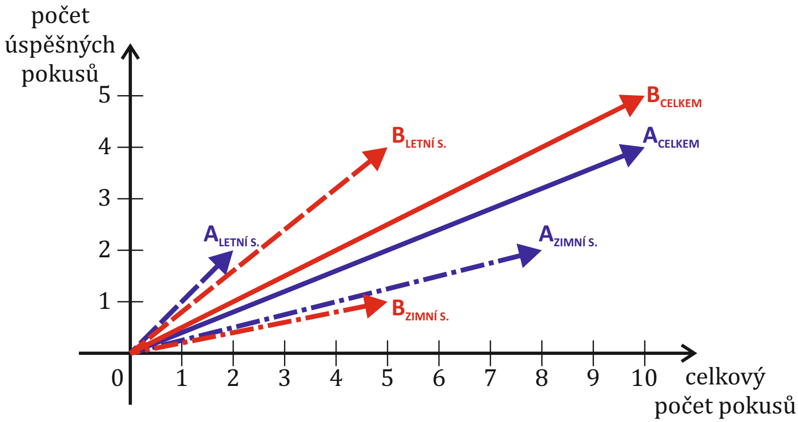


Obr. 1: Aloisova úspěšnost u zkoušek

Podobně pro Bedřichovy zkoušky použijeme vektory (5,1) a (5,4). Za oba semestry dohromady potom:

$$(5, 1) + (5, 4) = (5 + 5, 1 + 4) = (10, 5).$$

Obr. 2 ukazuje, že v každém semestru zvlášť je Alois úspěšnější (jeho dílčí vektory mají větší sklon), ale za oba semestry dohromady je Bedřich úspěšnější (součet Bedřichových vektorů má větší sklon než součet Aloisových vektorů). To přesně odpovídá „paradoxnímu" pozorování v tabulce výše, přestože o paradox ve skutečnosti nejde. Jednoduše řečeno, některé čtveřice vektorů tuto zajímavou vlastnost mají, přestože je poměrně neobvyklá.



Obr. 2: Porovnání úspěšnosti Bedřicha a Aloise

Poznamenejme, že taková paradoxní situace by nenastala, kdyby počty zkoušek v jednotlivých semestrech byly vyvážené, například kdyby každý ze studentů podstoupil v každém semestru pět zkoušek. Řekněme, že Alois uspěl v zimním semestru u a_1 z těchto pěti zkoušek, v letním semestru u a_2 zkoušek, a podobně Bedřich uspěl u b_1, respektive b_2 zkoušek. Pokud byl Alois v každém semestru úspěšnější, znamená to $a_1 > b_1, a_2 > b_2$. Celkově pak Alois uspěl u $a_1 + a_2$ z 10 zkoušek, Bedřich u $b_1 + b_2$ z 10 zkoušek. Z předchozího plyne, že $a_1 + a_2 > b_1 + b_2$ a Alois by tak byl celkově úspěšnější i při pohledu na celý akademický rok.

Uvedený příklad Simpsonova paradoxu je převzatý (a doplněný) ze stránky [1], respektive z knížky [2].

Tady jde o zdraví...

Ze stránky [1] přebíráme i následující příklad, tentokrát ze skutečné lékařské praxe, kde nesprávně vyvozené závěry mohou mít velmi závažné následky. Poznamenejme, že anglická stránka [3] uvádí k tomuto příkladu více podrobností.

V experimentální studii podstoupilo 350 pacientů operaci ledvinových kamenů klasickým postupem a 350 pacientů speciálním postupem (tzv. perkutánní nefrolitotomie). Zákrok byl hodnocen jako úspěšný, pokud po třech měsících po zákroku nebyly u daného pacienta pozorovány ledvinové kameny. V tabulce 2 vidíme, že v kategorii pacientů s malými kameny (do velikosti 2 cm) byl úspěšnější klasický operační postup, stejně jako v kategorii pacientů s velkými kameny (velikost nad 2 cm).

	Malé kameny		Velké kameny		Celkem	
Klasický postup	$\frac{81}{87}$	**93 %**	$\frac{192}{263}$	**73 %**	$\frac{273}{350}$	78 %
Speciální postup	$\frac{234}{270}$	87 %	$\frac{55}{80}$	69 %	$\frac{289}{350}$	**83 %**

Tabulka 2: Poměr úspěšných operací ku počtu provedených operací v jednotlivých kategoriích (malé vs. velké kameny) i celkově

Když zapomeneme na velikost kamenů a podíváme se na celkové výsledky, vychází jako úspěšnější naopak nový, speciální postup. Zdálo by se žádoucí tento typ zákroku nasadit do rutinní praxe, protože přece pacientům pomáhá s větší pravděpodobností než klasický postup. To by ale byla chyba a zbytečný hazard se zdravím pacientů! Když totiž přichází pacient do nemocnice, má buď malé kameny (a výhodnější je pro něj klasický postup), nebo velké kameny (a také je pro něj výhodnější klasický postup). Pro žádného jednotlivého pacienta není nový, speciální postup výhodnější.

Jádro problému je zde v nevyváženém rozdělení pacientů mezi oba operační postupy, s ohledem na velikost kamenů, kterými trpí. Speciální postup byl ve studii použitý zejména u pacientů s malými kameny, kde jsou ale oba operační postupy velmi úspěšné. Celková úspěšnost speciálního postupu je tedy nejvíce ovlivněna jeho výsledky na skupině malých kamenů. Naopak klasický postup byl použitý zejména u pacientů s velkými kameny, kde jsou oba postupy méně úspěšné. Celková úspěšnost klasického postupu je tedy nejvíce ovlivněna jeho výsledky na skupině velkých kamenů. Kdyby byli pacienti s velkými/malými kameny rozděleni mezi obě metody spravedlivě, k paradoxním závěrům bychom nemohli dojít, jak už jsme diskutovali výše. Studie tedy nebyla dobře navržena a teprve během jejího vyhodnocování se zjistilo, že velikost kamenů má zásadní roli pro úspěch zákroku, dokonce větší, než typ použitého zákroku.

Zkouška dospělosti

Pro zajímavost doplňme ještě reálný příklad z českého prostředí. Tabulka 3 ukazuje úspěšnost studentů u státní maturity z matematiky (jarní termín zkoušky, řádný termín) v letech 2015 a 2016. Jde o podíly studentů, kteří zkoušku úspěšně složili. Studenti jsou rozděleni do dvou skupin: gymnázia a všechny ostatní typy středních škol. Data je možné ověřit na stránce [4].

V tabulce si můžeme všimnout, že celková úspěšnost se v roce 2016 mírně zvýšila oproti roku 2015. To Cermat interpretoval jako mírné zlepšení maturantů v matematice. Na druhou stranu, podíly úspěšných studentů mezi roky 2015 a 2016 v každé ze skupin „gymnázia" a „ostatní" mírně klesly. Tento úhel pohledu zase podávala některá média jako mírné zhoršení maturantů v matematice.

	Gymnázia		Ostatní		Celkem	
2015	$\frac{6\,846}{7\,095}$	**96,5 %**	$\frac{7\,843}{12\,209}$	**64,2 %**	$\frac{14\,689}{19\,304}$	76,1 %
2016	$\frac{6\,662}{6\,916}$	96,3 %	$\frac{6\,386}{10\,072}$	63,4 %	$\frac{13\,048}{16\,988}$	**76,8 %**

Tabulka 3: Poměr počtu úspěšných maturantů ku počtu všech maturantů z matematiky v jednotlivých skupinách podle typu školy (gymnázia vs. ostatní) i celkově

V tomto případě je paradoxní situace způsobena tím, že ve skupině „ostatní" klesl mezi roky 2015 a 2016 počet maturantů podstupujících zkoušku z matematiky (ať už úspěšně nebo neúspěšně) mnohem výrazněji než ve skupině „gymnázia". Přestože v každé ze skupin došlo k mírnému snížení úspěšnosti, v celkovém podílu za daný rok hrála kategorie „ostatní" v roce 2016 menší roli (10 072 z celkových 16 988 studentů, tedy cca 59 %) než v roce 2015 (12 209 z celkových 19 304 studentů, tedy cca 63 %). v roce 2016 je tedy větší vliv skupiny „gymnázia", která celkovou úspěšnost vytahuje nahoru oproti roku 2015.

Když paradox není paradox

Na závěr zdůrazněme, že v tomto příspěvku jsme se vás snažili přesvědčit, že v uvedených příkladech vlastně o žádný paradox nejde. Tradiční označení „paradox" se odkazuje k neintuitivní možnosti dvojí, protikladné interpretace, která závisí na úhlu pohledu. Pouze kontext daného problému určuje, který úhel pohledu a potažmo který způsob interpretace je „správný“ ve smyslu užitečný pro zodpovězení položených otázek. V příkladu z lékařského prostředí jsme viděli, že volba vhodného či nevhodného úhlu pohledu může mít závažné důsledky. Dobrou zprávou je, že výskyt dat s touto vlastností je poměrně neobvyklý a patří zejména do oblasti statistických anekdot a folkloru. Přesto bychom měli být pokaždé při interpretaci dat opatrní a přemýšlet, zda při změně úhlu pohledu neumožňují jinou interpretaci.

Literatura

[1] `https://cs.wikipedia.org/wiki/Simpsonův_paradox` Dostupné online, citováno 30. 9. 2021.

[2] Anděl, J. *Statistické úlohy, historky a paradoxy.* Matfyzpress, Praha, 2018.

[3] `https://en.wikipedia.org/wiki/Simpson%27s_paradox` Dostupné online, citováno 30. 9. 2021.

[4] `https://vysledky.cermat.cz/data/Default.aspx` Dostupné online, citováno 30. 9. 2021.

* * * * *

NASYCENÉ PÁRY

Když teplota klesne pod bod
rosný,
vodní páry pocítí hlad
mocný.
Na kondenzačních jádrech se
zachytí
a za chvíli úplně se
nasytí.
Já bych ale nebyl
příliš rád,
kdybych se měl tím způsobem
stravovat!

Emil Calda: Úvod do obecné teorie prostoru, Karolinum, Praha, 2003