

Jiří Anděl

Několik statistických příběhů

Učitel matematiky, Vol. 8 (2000), No. 2, 116–118

Persistent URL: <http://dml.cz/dmlcz/150935>

Terms of use:

© Jednota českých matematiků a fyziků, 2000

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://dml.cz>

NĚKOLIK STATISTICKÝCH PŘÍBĚHŮ

JIŘÍ ANDĚL

*Je to sice zvláštní a vypadá to
skoro jako nemožné,
ale v našich úvahách žádnou chybu nevidím.*

Užitečnost teorie pravděpodobnosti a matematické statistiky je dnes uznávána nejen ve světě, ale i u nás. Postupně i zde lidé zjišťují, že rozhodování za přítomnosti prvku náhody je třeba opřít o statistická data. Uvedu alespoň jeden příklad ze školského prostředí. Zkušenosti z posledních několika let ukazují, že se do 1. ročníku Matematicko-fyzikální fakulty Univerzity Karlovy (MFF UK) na obor matematika zapisuje jen asi 60% počtu přijatých studentů. Je to zřejmě proto, že se někteří hlásí na více fakult a MFF mají spíše v záloze. Pokud uspějí jinde, o studium na MFF již nemají zájem. Pokládá-li fakulta za vhodné, aby se na obor matematika do 1. ročníku z kapacitních důvodů zapsalo 120 posluchačů, pak jednoduchý výpočet ukazuje, že je jich třeba přijmout asi 200.

Zdalo by se, že není tedy příliš obtížné správně interpretovat statistická data. Přitom většinou pravý opak je pravdou. I tak jednoduchý ukazatel, jako je aritmetický průměr, může mít vlastnosti, které neškoleného (ale někdy i školeného) uživatele překvapí. Uvedu dvě historiky.

Historika č. 1 (*Teaching Statistics* 13, 1991, str. 23): Naprostá většina lidí má nadprůměrný počet noh.

Historika č. 2 (*Teaching Statistics* 14, 1992, str. 29): Je známo, že se Angličané a Skoti mezi sebou dobírají. V jedné hospodě v Edinburgu byl (prý) vyslechnut následující výrok: „Když se Skot přestěhuje do Anglie, zvýší se úroveň inteligence na obou místech.“ Ptáme se:

a) Je to možné?

Úvodní kapitola z knihy *Matematika náhody*, kterou autor připravuje pro nakladatelství MATFYZPRESS.

b) Co se asi říká v hospodě v Newcastlu, pokud jde o emigraci Skotů?

Odpověď na první otázku je kladná. Představme si pro jednoduchost, že existují jen dva Angličani a dva Skotové. Angličani se jmenují A_1 a A_2 , Skotové S_1 a S_2 . V Edinburgu si patrně představují, že s inteligencí je to takhle:

	A_1	A_2	S_1	S_2
IQ	100	120	140	160

Průměrná inteligence Angličanů je tedy 110 a Skotů 150. Přestěhuje-li se S_1 do Anglie (a Skotové věří, že to může udělat jen člověk s nižší inteligencí), bude tam pak průměr 120. Ve Skotsku zůstane jen S_2 , a tak tam bude průměr 160. V obou místech se tedy úroveň inteligence zvýšila. Je snadné tato data modifikovat na libovolně velký počet Angličanů a Skotů.

A co se říká v hospodě v Newcastlu? Prý se tam tvrdí, že se ke stěhování rozhodne chytrý Skot, a tak úroveň inteligence poklesne na obou místech. V Newcastlu se pochopitelně vychází z představy, že inteligence je takováto:

	S_1	S_2	A_1	A_2
IQ	100	120	140	160

Podle nich průměrná inteligence Skotů je 110 a Angličanů 150. Až se přestěhuje S_2 do Anglie, bude průměrná inteligence Skotů 100 a Angličanů 140. Klesne na obou místech.

Nejde-li jen o průměry, ale o mírně složitější ukazatele, může se situace zamotat ještě víc. Řekněme, že se ve dvou okresech (řekněme v okrese A a v okrese B) sleduje podíl žen v populaci. Obyvatelé jsou přitom rozděleni do dvou kategorií — na mladší a na starší. Představme si, že byly zjištěny následující údaje:

kategorie	okres A			okres B		
	ženy	muži	celkem	ženy	muži	celkem
mladší	5	6	11	3	4	7
starší	6	3	9	9	5	14
celkem	11	9	20	12	9	21

Kdyby se někomu tyto počty obyvatel zdály na okres příliš nízké, může si představit, že údaje jsou ve stotisících. Výsledky dalšího rozboru zůstanou stejné.

V mladší věkové kategorii je podíl žen na okrese A větší než na okrese B, protože

$$\frac{5}{11} = 45,5\% > \frac{3}{7} = 42,9\%.$$

Ve starší věkové kategorii je tomu také tak, protože

$$\frac{6}{9} = 66,7\% > \frac{9}{14} = 64,3\%.$$

Vezmeme-li však celková data, zjistíme, že na okrese A je podíl žen menší než na okrese B, protože

$$\frac{11}{20} = 55,0\% < \frac{12}{21} = 57,1\%.$$

Tato data ilustrují tzv. Simpsonův paradox. Ten se dá čistě matematicky zapsat následovně. Nechť a_i , b_i , c_i a d_i ($i = 1, 2$) jsou kladná čísla. Pak

$$\left\{ \frac{a_1}{b_1} < \frac{c_1}{d_1}, \frac{a_2}{b_2} < \frac{c_2}{d_2} \right\} \not\Rightarrow \left\{ \frac{a_1 + a_2}{b_1 + b_2} < \frac{c_1 + c_2}{d_1 + d_2} \right\}.$$

Nad tím se nejspíš nikdo nepozastaví. Zato numerické ilustrace bývají dosti působivé.

Uveďme ještě jednu statistickou historku, která varuje před zbrklou interpretací statistických dat. Statistiky ukazují, že 10% dopravních nehod způsobují opilí řidiči. Z toho plyne, že zbývajících 90% dopravních nehod je způsobeno střízlivými řidiči. Nemělo by se tedy střízlivým řidičům zakázat řízení vozidel? (*Teaching Statistics* 6, 1984, str. 52.)