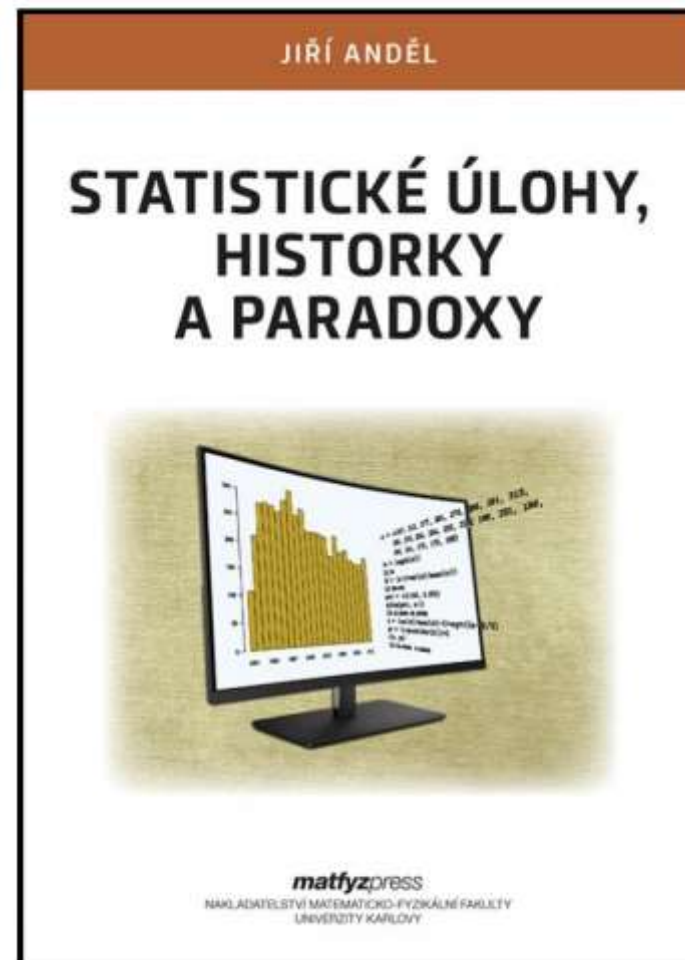
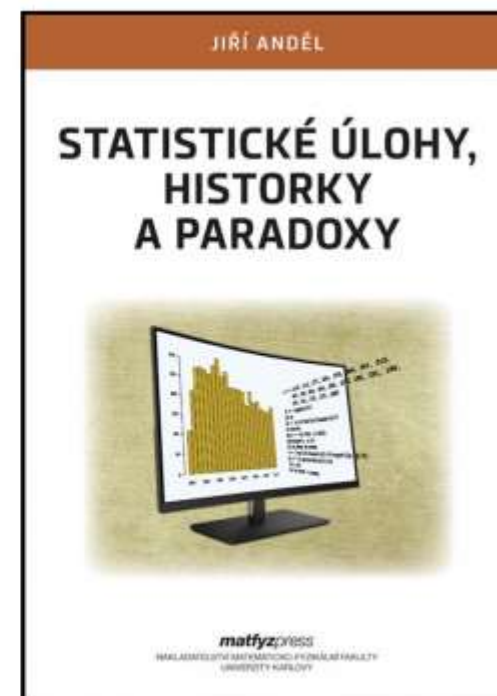
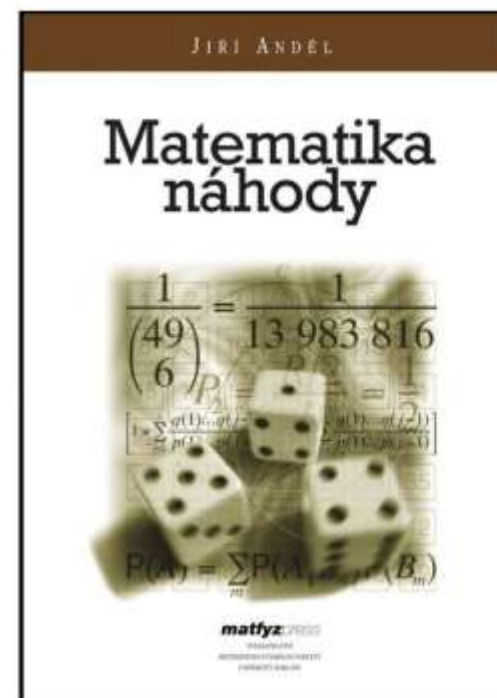
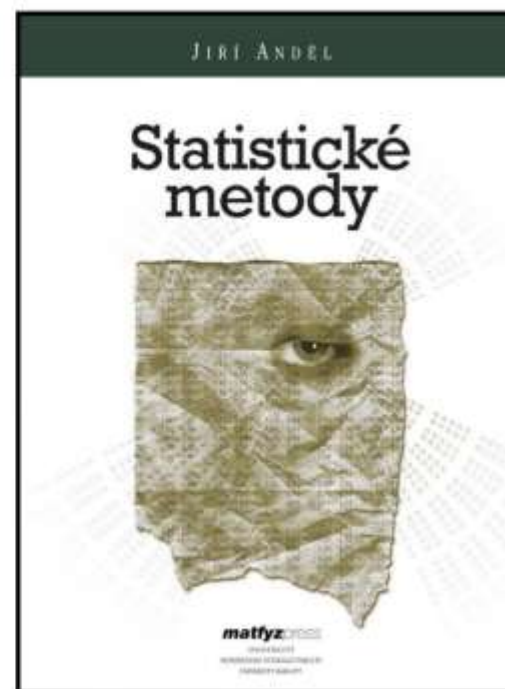
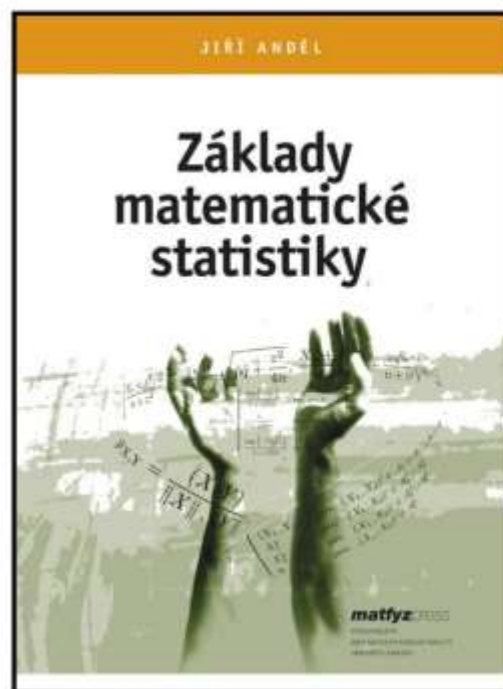


Statistické úlohy, historky a paradoxy

Z poslední knihy profesora Jiřího Anděla
čte Ondřej Vencálek

10. 6. 2021





Z obsahu

1	Několik statistických příběhů	13
1.1	Zápis studentů na MFF	13
1.2	Nečekané vlastnosti dat	13
1.2.1	Nadprůměrný počet nohou	13
1.2.2	Intelligence Angličanů a Skotů	13
1.2.3	Vážený průměr	15
1.2.4	Průměrný počet zapsaných studentů	15
1.2.5	Některé paradoxy	16
1.3	Chybná interpretace dat	21
1.3.1	Opilí řidiči	21
1.3.2	Nakládání zeleniny	21
1.4	Chyby v datech	21
1.5	Sčítání obyvatelstva	22
2	Základní pojmy	23
3	Prokládání přímky	47
3.1	Úvod	47
3.2	Metoda LAD	48
3.2.1	Popis metody	48
3.2.2	Laplaceova metoda	49
3.3	Metoda nejmenších čtverců	51
3.4	Předpověď pomocí regresní přímky	54
3.5	Historie metody nejmenších čtverců	57
3.6	Ortogonální regrese	59
4	Současně působící faktor	61
4.1	Korelace a kauzalita	61
4.2	Počet čápů a narozených dětí	61

Z obsahu

5	Udílání Oscarů	67
5.1	Základní informace	67
5.2	Histogramy věku herců a hereček	67
5.3	Závislost věku herců na roku udělení ceny	69
6	Problém dopravního policisty	77
7	Případ německých tanků	81
7.1	Odhad počtu prvků	81
7.2	Problém německých tanků	83
8	Kontingenční tabulky	85
8.1	Test nezávislosti	85
8.2	The lady tasting tea	88
8.3	Lindleyova varianta	90
8.4	Předpovědi v tabulkách	91
9	Horká ruka	95
10	Kdy je náskok dostatečný	107
11	Několik vítězných zápasů	109
12	Dílčí výsledky her	125
13	Žravé hry	129
14	Úloha von Neumanna	133
15	Děti a pravděpodobnosti	139
15.1	Úvod	139
15.2	Pravděpodobnost narození chlapce	139
15.3	Pohlaví sourozenců	141
16	Studenti a učitelé	143
16.1	Bernoullijské pokusy	143
16.2	Iterace úspěchů	143
16.3	Detekce zfalšovaných dat	144
16.4	Podvody u zkoušky	146

Průměrný počet zapsaných studentů (kap. 1.2.4)

Americké školy často inzerují průměrný počet studentů (average class sizes) zapsaných ve vyučovaných kurzech. Hemenway (1982) uvádí následující příklad. Na začátku akademického roku 1980/81 se na Harvard School of Public Health konalo 111 kurzů. V některých byl zapsán jen jeden student. Nejvíce studentů si zapsalo Epidemiologii, a to 229. Vedení školy oznámilo, že průměrný počet studentů v jednom kurzu je 14.5. Dotazem u studentů se však zjistilo, že každý z nich v průměru navštěvoval kurzy se 78 účastníky.

Nyní ukážeme, že

$$\frac{1}{n} \sum_{i=1}^k n_i^2 \geq \frac{n}{k},$$

přičemž rovnost nastane právě tehdy, když $n_1 = \dots = n_k$. Uvažujme náhodnou veličinu Y , která může se stejnou pravděpodobností $1/k$ nabývat hodnot $n_1, \dots, n_k$. Jde tedy o počet účastníků náhodně vybraného kurzu. Její střední hodnota je rovna

$$\mathbb{E}Y = \frac{1}{k} \sum_{i=1}^k n_i = \frac{n}{k}$$

a její rozptyl je

$$\text{var } Y = \mathbb{E}Y^2 - (\mathbb{E}Y)^2 = \frac{1}{k} \sum_{i=1}^k n_i^2 - \frac{n^2}{k^2} \geq 0.$$

Vynásobením této nerovnosti konstantou k/n dostaneme žádanou nerovnost.¹

S podobnými problémy se setkáváme i v jiných situacích. Předpokládejme, že by se měl zjistit průměrný počet dětí v rodině. Správný postup by byl vytvořit seznam všech rodin a z nich pak určitý počet náhodně vybrat. Kdyby se naopak náhodně vybral určitý počet dětí a u každého z nich se zjistilo, kolik je dětí v jeho rodině, dostal by se zkreslený výsledek. Rodiny s větším počtem dětí by měly větší pravděpodobnost, že budou vybrány. Ale ani tento závěr by nemusel být obecně správný. Kdyby se děti vybíraly mezi studenty Harvardu, došlo by k jinému zkreslení. Je známo, že rodiny s velkým počtem dětí bývají sociálně slabší, a tak zřídka své děti posílají studovat na Harvard.

Některé paradoxy (kap. 1.2.5)

Muži nad 30

	Lektvar	Bez lektvaru
Vítězství	6	8
Prohra	6	9

$$(6 \times 9 = 54) > (6 \times 8 = 48)$$

Muži pod 30

	Lektvar	Bez lektvaru
Vítězství	6	22
Prohra	3	12

$$(6 \times 12 = 72) > (3 \times 22 = 66)$$

Ženy nad 30

	Lektvar	Bez lektvaru
Vítězství	4	17
Prohra	6	27

$$(4 \times 27 = 108) > (6 \times 17 = 102)$$

Ženy pod 30

	Lektvar	Bez lektvaru
Vítězství	4	43
Prohra	3	33

$$(4 \times 33 = 132) > (3 \times 43 = 129)$$

Všichni muži

	Lektvar	Bez lektvaru
Vítězství	12	30
Prohra	9	21

$$(12 \times 21 = 252) < (9 \times 30 = 270)$$

Všechny ženy

	Lektvar	Bez lektvaru
Vítězství	8	60
Prohra	9	60

$$(8 \times 60 = 480) < (9 \times 60 = 540)$$

Všichni hráči

	Lektvar	Bez lektvaru
Vítězství	20	90
Prohra	18	81

$$(20 \times 81 = 1620) = (18 \times 90 = 1620)$$

Nakládaná zelenina (kap. 1.4.2)

1. Skoro všichni tlustí lidé jedli nakládanou zeleninu. Její efekt se zřejmě kumuluje.
2. 99 % lidí, kteří zemřeli na rakovinu, jedlo nakládanou zeleninu.
3. 95 % lidí, kteří sympatizují s teroristy, jedlo nakládanou zeleninu.
4. 99 % lidí, kteří byli zapleteni do leteckých či automobilových havárií, jedlo nakládanou zeleninu během jednoho měsíce před neštěstím.
5. 93 % mladistvých delikventů pochází z rodin, kde se často podává nakládaná zelenina.
6. Všichni lidé, kteří se narodili mezi léty 1920 a 1930 a kteří jedli nakládanou zeleninu, mají svraštělou kůži, ztratili většinu svých zubů, mají křehké kosti a špatně vidí.
7. Přesvědčivá je zpráva jednoho významného týmu lékařských specialistů. Krysy krmené 10 kg nakládané zeleniny denně měly po 30 dnech nadmutá břicha. Jejich chuť ke zdravé výživě byla zcela podlomena.

Zhoubnému vlivu nakládané zeleniny se lze vyhnout změnou jídelního lístku. Doporučuje se jíst polévku z orchidejí. Prakticky nikdo z požívání této polévky neměl žádné problémy.

Medaile JČMF (22. 11. 2019)

prof. RNDr. Jiří Anděl, DrSc., MFF UK v Praze
za celoživotní přínos české a světové matematice a
obětavou práci při výchově nastupujících generací

