Čemu se (ne)máme divit

Mirko Navara
Centrum strojového vnímání, katedra kybernetiky
elektrotechnická fakulta ČVUT, Praha
http://cmp.felk.cvut.cz/˜navara

27. 4. 2024

Motto

Kniha přírody je psána jazykem matematiky.

Galileo Galilei

Motto

Kniha přírody je psána jazykem matematiky.

Galileo Galilei

Ale dnes řekneme něco skoro bez vzorců.

Chaos v obálkách

Posílám 4 obálky, v každé je na začátku 10 lístků.
Kdykoli k vám dojdou, vyberte náhodně 2 obálky
a z jedné do druhé přesuňte 1 lístek (pokud tam je).

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)
Jak dlouho potrvá, než se rozhodne?

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)
Jak dlouho potrvá, než se rozhodne?

počet kol n	pravděpodobnost konce v kole n	pravděpodobnost konce do kola n
1	$\frac{2}{3}$	$\frac{2}{3} \doteq 0.67$

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)
Jak dlouho potrvá, než se rozhodne?

počet kol n	pravděpodobnost konce v kole n	pravděpodobnost konce do kola n
1	$\frac{2}{3}$	$\frac{2}{3} \doteq 0.67$
2	$\frac{1}{3} \cdot \frac{2}{3}$	$\frac{8}{9} \doteq 0.89$

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)
Jak dlouho potrvá, než se rozhodne?

počet kol n	pravděpodobnost konce v kole n	pravděpodobnost konce do kola n
1	$\frac{2}{3}$	$\frac{2}{3} \doteq 0.67$
2	$\frac{1}{3} \cdot \frac{2}{3}$	$\frac{8}{9} \doteq 0.89$
3	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{26}{27} \doteq 0.96$

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)
Jak dlouho potrvá, než se rozhodne?

počet kol n	pravděpodobnost konce v kole n	pravděpodobnost konce do kola n
1	$\frac{2}{3}$	$\frac{2}{3} \doteq 0.67$
2	$\frac{1}{3} \cdot \frac{2}{3}$	$\frac{8}{9} \doteq 0.89$
3	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{26}{27} \doteq 0.96$
4	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{80}{81} \doteq 0.99$
...	...	...

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)
Jak dlouho potrvá, než se rozhodne?

počet kol n	pravděpodobnost konce v kole n	pravděpodobnost konce do kola n
1	$\frac{2}{3}$	$\frac{2}{3} \doteq 0.67$
2	$\frac{1}{3} \cdot \frac{2}{3}$	$\frac{8}{9} \doteq 0.89$
3	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{26}{27} \doteq 0.96$
4	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{80}{81} \doteq 0.99$
...	...	...

Na 99 % budeme po 4 kolech hotovi,

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)
Jak dlouho potrvá, než se rozhodne?

počet kol n	pravděpodobnost konce v kole n	pravděpodobnost konce do kola n
1	$\frac{2}{3}$	$\frac{2}{3} \doteq 0.67$
2	$\frac{1}{3} \cdot \frac{2}{3}$	$\frac{8}{9} \doteq 0.89$
3	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{26}{27} \doteq 0.96$
4	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{80}{81} \doteq 0.99$
...	...	...

Na 99 % budeme po 4 kolech hotovi, ale kdy to bude určitě?

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)
Jak dlouho potrvá, než se rozhodne?

počet kol n	pravděpodobnost konce v kole n	pravděpodobnost konce do kola n
1	$\frac{2}{3}$	$\frac{2}{3} \doteq 0.67$
2	$\frac{1}{3} \cdot \frac{2}{3}$	$\frac{8}{9} \doteq 0.89$
3	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{26}{27} \doteq 0.96$
4	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{80}{81} \doteq 0.99$
...	...	...

Na 99 % budeme po 4 kolech hotovi, ale kdy to bude určitě?
Nikdy!

Chytřejší náhodné losování

Nápad: Změňme pravidla tak, aby bylo možno rozhodnout po konečném počtu kroků.

Chytřejší náhodné losování

Nápad: Změňme pravidla tak, aby bylo možno rozhodnout po konečném počtu kroků.
S volbou mezi třemi výsledky to pro dva hráče nejde, pro tři ano.

Chytřejší náhodné losování

Nápad: Změňme pravidla tak, aby bylo možno rozhodnout po konečném počtu kroků.
S volbou mezi třemi výsledky to pro dva hráče nejde, pro tři ano.
Ale dva hráči se mohou dohodnout např., že budou používat jen dva znaky (nůžky, papír);
jeden hráč vyhrává, dají-li stejné znaky, druhý při různých znacích.

Chytřejší náhodné losování

Nápad: Změňme pravidla tak, aby bylo možno rozhodnout po konečném počtu kroků.
S volbou mezi třemi výsledky to pro dva hráče nejde, pro tři ano.
Ale dva hráči se mohou dohodnout např., že budou používat jen dva znaky (nůžky, papír);
jeden hráč vyhrává, dají-li stejné znaky, druhý při různých znacích.
Je to spravedlivé a rozhodne se hned v prvním kole!

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub dvakrát po sobě**.
Je to spravedlivé?

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub dvakrát po sobě**.
Je to spravedlivé?
Ano. Líc a rub jsou rovnocenné.

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **líc a po něm rub**.
Je to spravedlivé?

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **líc a po něm rub**.
Je to spravedlivé?
Ano. Až padne poprvé líc, následující hod rozhodne.

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub a po něm líc**.
Je to spravedlivé?

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub a po něm líc**.
Je to spravedlivé?
Ne!

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub a po něm líc**.
Je to spravedlivé?
Ne!
Padne-li napoprvé líc, Bob čeká, jestli padne ještě jednou.

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub a po něm líc**.
Je to spravedlivé?
Ne!
Padne-li napoprvé líc, Bob čeká, jestli padne ještě jednou.
Padne-li napoprvé rub, Alice už jen čeká na svou jistou výhru.

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub a po něm líc**.
Je to spravedlivé?
Ne!
Padne-li napoprvé líc, Bob čeká, jestli padne ještě jednou.
Padne-li napoprvé rub, Alice už jen čeká na svou jistou výhru.
A pokud padl líc a rub, Alice opět očekává jistou výhru.
Sázím 3:1 na Alici.

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub a po něm líc**.
Je to spravedlivé?
Ne!
Padne-li napoprvé líc, Bob čeká, jestli padne ještě jednou.
Padne-li napoprvé rub, Alice už jen čeká na svou jistou výhru.
A pokud padl líc a rub, Alice opět očekává jistou výhru.
Sázím 3:1 na Alici.

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub a po něm líc**.
Je to spravedlivé?
Ne!
Padne-li napoprvé líc, Bob čeká, jestli padne ještě jednou.
Padne-li napoprvé rub, Alice už jen čeká na svou jistou výhru.
A pokud padl líc a rub, Alice opět očekává jistou výhru.
Sázím 3:1 na Alici.

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob získává bod, padne-li **líc dvakrát po sobě** (a **hází se dál**).
Alice získává bod, padne-li **rub a po něm líc**.
Je to spravedlivé?

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob získává bod, padne-li **líc dvakrát po sobě** (a **hází se dál**).
Alice získává bod, padne-li **rub a po něm líc**.
Je to spravedlivé?
Bob může bodovat třeba pořád, Alice nanejvýš v každém druhém hodu.

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob získává bod, padne-li **líc dvakrát po sobě** (a **hází se dál**).
Alice získává bod, padne-li **rub a po něm líc**.
Je to spravedlivé?
Bob může bodovat třeba pořád, Alice nanejvýš v každém druhém hodu.
Ale je to spravedlivé! (Pokud hrají dlouho.)

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob získává bod, padne-li **líc dvakrát po sobě** (a **hází se dál**).
Alice získává bod, padne-li **rub a po něm líc**.
Je to spravedlivé?
Bob může bodovat třeba pořád, Alice nanejvýš v každém druhém hodu.
Ale je to spravedlivé! (Pokud hrají dlouho.)
Kdybychom hru sledovali pozpátku, po líci by v následujícím kroku jeden bodoval: líc Bob, rub Alice.

Házení mincí

Alice a Bob házejí mincí (**každý svou**).
Bob čeká, až padne **líc dvakrát po sobě**.
Alice čeká, až padne **líc a po něm rub**.
Vyhrává ten, kdo dosáhne cíle dříve (po menším počtu hodů).
Je to spravedlivé?

Házení mincí

Alice a Bob házejí mincí (**každý svou**).
Bob čeká, až padne **líc dvakrát po sobě**.
Alice čeká, až padne **líc a po něm rub**.
Vyhrává ten, kdo dosáhne cíle dříve (po menším počtu hodů).
Je to spravedlivé?
Ne!
Bob potřebuje v průměru 6 hodů, Alice 4.

Monty Hall problem

Troje dveře, za jedněmi je výhra.
Moderátor ví, za kterými.

Monty Hall problem

Troje dveře, za jedněmi je výhra.
Moderátor ví, za kterými.
Vybereme jedny dveře.
Moderátor nám ukáže jedny z dveří, které jsme nevybrali, a za nimi nic není.
Nyní smíme změnit svoji volbu; máme to udělat?

Monty Hall problem

Troje dveře, za jedněmi je výhra.
Moderátor ví, za kterými.
Vybereme jedny dveře.
Moderátor nám ukáže jedny z dveří, které jsme nevybrali, a za nimi nic není.
Nyní smíme změnit svoji volbu; máme to udělat?

Názor sedláka: Je to jedno, máme šanci $1/3$.

Monty Hall problem

Troje dveře, za jedněmi je výhra.
Moderátor ví, za kterými.
Vybereme jedny dveře.
Moderátor nám ukáže jedny z dveří, které jsme nevybrali, a za nimi nic není.
Nyní smíme změnit svoji volbu; máme to udělat?

Názor sedláka: Je to jedno, máme šanci $1/3$.

Názor praktika: Znovu náhodně vybereme.

Monty Hall problem

Troje dveře, za jedněmi je výhra.
Moderátor ví, za kterými.
Vybereme jedny dveře.
Moderátor nám ukáže jedny z dveří, které jsme nevybrali, a za nimi nic není.
Nyní smíme změnit svoji volbu; máme to udělat?

Názor sedláka: Je to jedno, máme šanci $1/3$.

Názor praktika: Znovu náhodně vybereme.
Teď totiž vybíráme jedny ze dvojích dveří, máme šanci $1/2$.

Monty Hall problem

Troje dveře, za jedněmi je výhra.
Moderátor ví, za kterými.
Vybereme jedny dveře.
Moderátor nám ukáže jedny z dveří, které jsme nevybrali, a za nimi nic není.
Nyní smíme změnit svoji volbu; máme to udělat?

Názor sedláka: Je to jedno, máme šanci $1/3$.

Názor praktika: Znovu náhodně vybereme.
Teď totiž vybíráme jedny ze dvojích dveří, máme šanci $1/2$.

Názor matematika: Vždy měníme.

Monty Hall problem

Troje dveře, za jedněmi je výhra.
Moderátor ví, za kterými.
Vybereme jedny dveře.
Moderátor nám ukáže jedny z dveří, které jsme nevybrali, a za nimi nic není.
Nyní smíme změnit svoji volbu; máme to udělat?

Názor sedláka: Je to jedno, máme šanci $1/3$.

Názor praktika: Znovu náhodně vybereme.
Teď totiž vybíráme jedny ze dvojích dveří, máme šanci $1/2$.

Názor matematika: Vždy měníme. Máme šanci $2/3$!

Nejlepší hráč golfu

Občas se někomu podaří z odpalu přímo trefit jamku.

Nejlepší hráč golfu

Občas se někomu podaří z odpalu přímo trefit jamku.
Ale nedaří se mu to zopakovat.

Nejlepší hráč golfu

Občas se někomu podaří z odpalu přímo trefit jamku.
Ale nedaří se mu to zopakovat.
Jednomu se to podařilo.
Je to nejlepší hráč golfu?

Nejlepší hráč golfu

Občas se někomu podaří z odpalu přímo trefit jamku.
Ale nedaří se mu to zopakovat.
Jednomu se to podařilo.
Je to nejlepší hráč golfu? Není.
Ne proto, že to nedokázal zopakovat.

Nejlepší hráč golfu

Občas se někomu podaří z odpalu přímo trefit jamku.
Ale nedaří se mu to zopakovat.
Jednomu se to podařilo.
Je to nejlepší hráč golfu? Není.
Ne proto, že to nedokázal zopakovat.
Podrobnější analýza říká, že za rok je na celém světě asi 100 000 000 odpalů, takže pravděpodobnost, že se to někomu z tolika hráčů podaří, není zas tak malá.

Úmrtí po očkování

Jeden člověk v ČR po očkování proti covidu zemřel ještě dřív, než odešel z očkovacího centra.
Máme se bát?

Úmrtí po očkování

Jeden člověk v ČR po očkování proti covidu zemřel ještě dřív, než odešel z očkovacího centra.
Máme se bát? Ne!
Z 10 000 000 lidí zemře asi 400 za den, tj. v průměru 1 za 4 minuty.

Úmrtí po očkování

Jeden člověk v ČR po očkování proti covidu zemřel ještě dřív, než odešel z očkovacího centra.
Máme se bát? Ne!
Z 10 000 000 lidí zemře asi 400 za den, tj. v průměru 1 za 4 minuty.
Stejně nebezpečná tedy vyjde návštěva jakéhokoli jiného místa, např. restaurace, muzea, úřadu…

Úmrtí po očkování

Jeden člověk v ČR po očkování proti covidu zemřel ještě dřív, než odešel z očkovacího centra.
Máme se bát? Ne!
Z 10 000 000 lidí zemře asi 400 za den, tj. v průměru 1 za 4 minuty.
Stejně nebezpečná tedy vyjde návštěva jakéhokoli jiného místa, např. restaurace, muzea, úřadu...

Nejnebezpečnější činnost je ležet v posteli. Tam umřelo nejvíc lidí.

První statistické průzkumy

Astrologové, zejména Johannes Kepler.

První statistické průzkumy

Astrologové, zejména Johannes Kepler.

- Položili počátky statistiky i astrologii.

První statistické průzkumy

Astrologové, zejména Johannes Kepler.

- Položili počátky statistiky i astrologii.
- Astrologická znamení mj. už dávno nejsou, co bývala.

První statistické průzkumy

Astrologové, zejména Johannes Kepler.

- Položili počátky statistiky i astrologii.
- Astrologická znamení mj. už dávno nejsou, co bývala.
- Přesto na datu narození závisí např. úspěšnost ve sportu (fotbal, šach).

První statistické průzkumy

Astrologové, zejména Johannes Kepler.

- Položili počátky statistiky i astrologii.
- Astrologická znamení mj. už dávno nejsou, co bývala.
- Přesto na datu narození závisí např. úspěšnost ve sportu (fotbal, šach).
- Jenže kvůli věkovým rozdílům při srovnávání.

První statistické průzkumy

Astrologové, zejména Johannes Kepler.

- Položili počátky statistiky i astrologii.
- Astrologická znamení mj. už dávno nejsou, co bývala.
- Přesto na datu narození závisí např. úspěšnost ve sportu (fotbal, šach).
- Jenže kvůli věkovým rozdílům při srovnávání.
- Nicméně byly zjištěny drobné odchylky mezi stejně starými dětmi narozenými v různých ročních obdobích, vysvětlují se sezónními rozdíly ve výživě matek.

První statistické průzkumy

Astrologové, zejména Johannes Kepler.

- Položili počátky statistiky i astrologii.
- Astrologická znamení mj. už dávno nejsou, co bývala.
- Přesto na datu narození závisí např. úspěšnost ve sportu (fotbal, šach).
- Jenže kvůli věkovým rozdílům při srovnávání.
- Nicméně byly zjištěny drobné odchylky mezi stejně starými dětmi narozenými v různých ročních obdobích, vysvětlují se sezónními rozdíly ve výživě matek.
- Rozhodně je mnoho podstatnějších vlivů, např. výživa dětí.

Čínské horoskopy

- Podle roku narození předpovídají možný rok úmrtí.

Čínské horoskopy

- Podle roku narození předpovídají možný rok úmrtí.
- Zjištěna závislost u populace Číňanů v USA.

Čínské horoskopy

- Podle roku narození předpovídají možný rok úmrtí.
- Zjištěna závislost u populace Číňanů v USA.
- Jenže funguje jen u těch, kteří tomu věří.

Čínské horoskopy

- Podle roku narození předpovídají možný rok úmrtí.
- Zjištěna závislost u populace Číňanů v USA.
- Jenže funguje jen u těch, kteří tomu věří.
- A ti umírají v průměru o 2 roky dříve!

Výhoda astrologie

Výhoda astrologie

- I kdybych zamíchal horoskopy jako karty, nehnal by mě nikdo k zodpovědnosti.

Výhoda astrologie

- I kdybych zamíchal horoskopy jako karty, nehnal by mě nikdo k zodpovědnosti.
- Pokud někdo chce poradit, jak zhodnotit své úspory nebo vyhrát soud s podvodníkem, potřebuje solidnější rady, i když ani ty nebudou zaručené.

Každý může být někomu dobrým astrologem (i finančním poradcem)

Obešleme 128 lidí.
Polovině předpovíme růst akcií, polovině pokles.

Každý může být někomu dobrým astrologem (i finančním poradcem)

Obešleme 128 lidí.
Polovině předpovíme růst akcií, polovině pokles.
V každé skupině polovině předpovíme růst jiných akcií, polovině pokles.

Každý může být někomu dobrým astrologem (i finančním poradcem)

Obešleme 128 lidí.
Polovině předpovíme růst akcií, polovině pokles.
V každé skupině polovině předpovíme růst jiných akcií, polovině pokles.
...
Po 7 kolech zbyde jeden, jemuž jsme vše předpověděli správně.
Ten si řekne: „*To nemůže být náhoda! Jinak by se na víc než* 99 % *někdy zmýlil. Tomu věřím!*“

Typické otázky

- Kde se déle žije, víc pije, víc vydělává?
- Co nám škodí/prospívá?
- Jak vyhodnotit úspěšnost předpovědí počasí?
- Proč mohou selhat volební předpovědi?

Zbohatlíkov [Swoboda]

Průměrný příjem rodiny je 82 320.
Příjem více než poloviny rodin je $\geq$ 29 000.

Zbohatlíkov [Swoboda]

Průměrný příjem rodiny je 82 320.
Příjem více než poloviny rodin je $\geq$ 29 000.
Nejčetnější příjem rodiny je 18 000.
Většina lidí nemá ani 7 500.
88 % lidí má méně než 25 000.

Zbohatlíkov [Swoboda]

Příjmy 25 rodin (v závorce počet členů rodiny):

1 200 000	(3)	60 000	(1)	45000	(2)
150 000	(5)	51 000	(3)	42 000	(2)
86 000	(4)	49 000	(4)	38 000	(4)
37 000	(3)	20 000	(7)	14 000	(1)
35 000	(5)	18 000	(3)	13 000	(4)
32 000	(3)	18 000	(8)	11 000	(1)
29 000	(3)	18 000	(4)	10 000	(2)
26 000	(4)	16 000	(3)		
24 000	(4)	16 000	(2)		

Zbohatlíkov [Swoboda]

Příjmy 25 rodin (v závorce počet členů rodiny):

1 200 000	(3)	60 000	(1)	45000	(2)
150 000	(5)	51 000	(3)	42 000	(2)
86 000	(4)	49 000	(4)	38 000	(4)
37 000	(3)	20 000	(7)	14 000	(1)
35 000	(5)	18 000	(3)	13 000	(4)
32 000	(3)	18 000	(8)	11 000	(1)
29 000	(3)	18 000	(4)	10 000	(2)
26 000	(4)	16 000	(3)		
24 000	(4)	16 000	(2)		

„Pan milionář si žije skvěle, obklopen několika tucty chudých otroků.“

Průměr nebo medián?

[Gonick, Smith] V r. 1984 University of Virginia oznámila, že průměrný nástupní plat absolventů katedry rétoriky a komunikací byl (tehdy impozantních) 55 000 \$.

Průměr nebo medián?

[Gonick, Smith] V r. 1984 University of Virginia oznámila, že průměrný nástupní plat absolventů katedry rétoriky a komunikací byl (tehdy impozantních) 55 000 $. Příčinou nebyly tam vyučované dovednosti, ale jeden absolvent – basketbalový profesionál Ralph Sampson.

Průměr nebo medián?

[Gonick, Smith] V r. 1984 University of Virginia oznámila, že průměrný nástupní plat absolventů katedry rétoriky a komunikací byl (tehdy impozantních) 55 000 $. Příčinou nebyly tam vyučované dovednosti, ale jeden absolvent – basketbalový profesionál Ralph Sampson.

Bývalý premiér: „Co je to za stát, kde 2/3 obyvatel má podprůměrný příjem?“

Průměr nebo medián?

[Gonick, Smith] V r. 1984 University of Virginia oznámila, že průměrný nástupní plat absolventů katedry rétoriky a komunikací byl (tehdy impozantních) 55 000 $. Příčinou nebyly tam vyučované dovednosti, ale jeden absolvent – basketbalový profesionál Ralph Sampson.

Bývalý premiér: „Co je to za stát, kde 2/3 obyvatel má podprůměrný příjem?“
A co má být? Co navrhujete?

Průměr nebo medián?

[Gonick, Smith] V r. 1984 University of Virginia oznámila, že průměrný nástupní plat absolventů katedry rétoriky a komunikací byl (tehdy impozantních) 55 000 $. Příčinou nebyly tam vyučované dovednosti, ale jeden absolvent – basketbalový profesionál Ralph Sampson.

Bývalý premiér: „Co je to za stát, kde 2/3 obyvatel má podprůměrný příjem?“
A co má být? Co navrhujete?

Medián pravděpodobnostního rozdělení je taková hodnota, že
s pravděpodobností 1/2 dostáváme hodnoty **větší**,
s pravděpodobností 1/2 dostáváme hodnoty **menší**.

„Statistika je věda o přesném součtu nepřesných čísel“

„Statistika je věda o přesném součtu nepřesných čísel“

Není!

„Statistika je věda o přesném součtu nepřesných čísel“

Není!

- **Absolutní** chyba součtu roste, jen **relativní** chyba klesá (a zůstává nenulová).

„Statistika je věda o přesném součtu nepřesných čísel“

Není!

- **Absolutní** chyba součtu roste, jen **relativní** chyba klesá (a zůstává nenulová).
- Lépe by bylo hovořit o „přesném průměru nepřesných čísel“.

„Statistika je věda o přesném součtu nepřesných čísel“

Není!

- **Absolutní** chyba součtu roste, jen **relativní** chyba klesá (a zůstává nenulová).
- Lépe by bylo hovořit o „přesném průměru nepřesných čísel“.
- To vše však jen za určitých předpokladů (např. nezávislost).

Role statistiky ve vědě

[Janko: Jak vytváří statistika obrazy života a světa]

Není potřeba, pokud jde o zjevné zákonitosti, ale odhaluje nám řadu jemnějších vztahů.

Příklady závislostí:

- Jedovatost muchomůrky hlíznaté – nesporná

Příklady závislostí:

- Jedovatost muchomůrky hlíznaté – nesporná
- Jedovatost zmije – nezpochybňovaná

Příklady závislostí:

- Jedovatost muchomůrky hlíznaté – nesporná
- Jedovatost zmije – nezpochybňovaná
- Škodlivost dalších vlivů – diskutovaná
  - Hliníkové nádobí
  - Chronický únavový syndrom
  - Syndrom války v Zálivu
  - Letní čas

Příklady závislostí:

- Jedovatost muchomůrky hlíznaté – nesporná
- Jedovatost zmije – nezpochybňovaná
- Škodlivost dalších vlivů – diskutovaná
    - Hliníkové nádobí
    - Chronický únavový syndrom
    - Syndrom války v Zálivu
    - Letní čas
- Placebo efekt apod.

Příklady závislostí:

- Jedovatost muchomůrky hlíznaté – nesporná
- Jedovatost zmije – nezpochybňovaná
- Škodlivost dalších vlivů – diskutovaná
  - Hliníkové nádobí
  - Chronický únavový syndrom
  - Syndrom války v Zálivu
  - Letní čas
- Placebo efekt apod.
- Nepodložená tvrzení
  - Geopatogenní zóny
  - Účinky homeopatik

Chceme-li si udělat jasno, **neobejdeme se bez statistiky**.

K čemu je tedy statistika?

Statistika dovoluje rozpoznat **společné vlastnosti velkých souborů**.

K čemu je tedy statistika?

Statistika dovoluje rozpoznat **společné vlastnosti velkých souborů**.

Nabízí pouze problematický kompromis, nic víc ani očekávat nelze:

- Závěry nejsou zaručené, pouze hodně pravděpodobné.

K čemu je tedy statistika?

Statistika dovoluje rozpoznat **společné vlastnosti velkých souborů**.

Nabízí pouze problematický kompromis, nic víc ani očekávat nelze:

- Závěry nejsou zaručené, pouze hodně pravděpodobné.
- Nelze je přímo uplatnit na jednotlivce.

K čemu je tedy statistika?

Statistika dovoluje rozpoznat **společné vlastnosti velkých souborů**.

Nabízí pouze problematický kompromis, nic víc ani očekávat nelze:

- Závěry nejsou zaručené, pouze hodně pravděpodobné.
- Nelze je přímo uplatnit na jednotlivce.
- Hypotézy lze pouze **zamítat** jako nepravděpodobné, **nikoli potvrdit**, jeví-li se pravděpodobné.

Dvojitě zaslepený experiment

Jak statistik očkuje dvojčata.

Dvojitě zaslepený experiment

Jak statistik očkuje dvojčata.

Ideál: Experimentátor ani účastník experimentu nemá vědět, zda je ve výběrovém nebo kontrolním souboru.

Dvojitě zaslepený experiment

Jak statistik očkuje dvojčata.

Ideál: Experimentátor ani účastník experimentu nemá vědět, zda je ve výběrovém nebo kontrolním souboru.
Někdy je těžké to dodržet.

Časté omyly v chápání statistiky

Proč máme statistiku respektovat i s těmito nedostatky:

- Kdy máme očekávat stoletou vodu?
- Kdy máme očekávat výhru v loterii?
- Kdo je na titulní stránce, s tím to půjde z kopce.

Časté omyly v chápání statistiky

Proč máme statistiku respektovat i s těmito nedostatky:

- Kdy máme očekávat stoletou vodu?
- Kdy máme očekávat výhru v loterii?
- Kdo je na titulní stránce, s tím to půjde z kopce.

Stejný princip platí i pro akcie a investiční fondy!

Kde všude je/má být statistika

$\Longrightarrow$ **je třeba vyučovat statistiku a pravděpodobnost už od základní školy, na odpovídající úrovni (kvalitativní, nikoli kvantitativní).**

Newcombův-Benfordův zákon

Vezměte

- poslední 3 nenulové cifry vašeho (nebo jiného) telefonního čísla,
- číslo domu,
- číslo bot,

Newcombův-Benfordův zákon

Vezměte

- poslední 3 nenulové cifry vašeho (nebo jiného) telefonního čísla,
- číslo domu,
- číslo bot,

vše vynásobte.

Newcombův-Benfordův zákon

Vezměte

- poslední 3 nenulové cifry vašeho (nebo jiného) telefonního čísla,
- číslo domu,
- číslo bot,

vše vynásobte.
Komu výsledek začíná číslicí 1?

Newcombův-Benfordův zákon

Vezměte

- poslední 3 nenulové cifry vašeho (nebo jiného) telefonního čísla,
- číslo domu,
- číslo bot,

vše vynásobte.
Komu výsledek začíná číslicí 1?
Na podobném principu se vyhledávají podvodná data u finančních transakcí, daňových dokladů apod.

Statistické odhady

Jsou zatíženy chybou; chceme, aby nebyla systematická:

Statistické odhady

Jsou zatíženy chybou; chceme, aby nebyla systematická:
Nestranný odhad je takový, že jeho střední hodnota je stejná jako střední hodnota odhadovaného rozdělení.

Statistické odhady

Jsou zatíženy chybou; chceme, aby nebyla systematická:
Nestranný odhad je takový, že jeho střední hodnota je stejná jako střední hodnota odhadovaného rozdělení.
Ve stavebnictví se zjevně nepoužívá. ☹

Statistické odhady

Jsou zatíženy chybou; chceme, aby nebyla systematická:
Nestranný odhad je takový, že jeho střední hodnota je stejná jako střední hodnota odhadovaného rozdělení.
Ve stavebnictví se zjevně nepoužívá. ☹
Někdy jsou jiná kritéria pro volbu odhadu, např. minimalizace rizik. (Odhad věku, IQ apod.)

Co lze říci o přesnosti volebních preferencí?

Ve vzorku 1000 z 10 000 000 voličů se 200 respondentů vyjádřilo, že budou volit určitou stranu.
Odhad preferencí je $200/1000 = 20\,\%$.
Jaká je garantovaná přesnost tohoto odhadu?

Co lze říci o přesnosti volebních preferencí?

Ve vzorku 1000 z 10 000 000 voličů se 200 respondentů vyjádřilo, že budou volit určitou stranu.
Odhad preferencí je $200/1000 = 20\,\%$.
Jaká je garantovaná přesnost tohoto odhadu?

S **jistotou** lze říci pouze, že preference jsou 0.000 02 **až** 0.999 92, protože zaručíme jen to, že aspoň 200 voličů ji bude volit a aspoň 800 ji volit nebude (a to ještě nesmí změnit názor).

Co lze říci o přesnosti volebních preferencí?

Ve vzorku 1000 z 10 000 000 voličů se 200 respondentů vyjádřilo, že budou volit určitou stranu.
Odhad preferencí je $200/1000 = 20\,\%$.
Jaká je garantovaná přesnost tohoto odhadu?

S **jistotou** lze říci pouze, že preference jsou 0.000 02 **až** 0.999 92, protože zaručíme jen to, že aspoň 200 voličů ji bude volit a aspoň 800 ji volit nebude (a to ještě nesmí změnit názor).

Více nám řekne informace, že preference jsou 16.7 **až** 23.3 % **s pravděpodobností** 99 %. (K tomu je potřebný správný výběr respondentů.)

Co lze říci o přesnosti volebních preferencí?

Ve vzorku 1000 z 10 000 000 voličů se 200 respondentů vyjádřilo, že budou volit určitou stranu.
Odhad preferencí je $200/1000 = 20\,\%$.
Jaká je garantovaná přesnost tohoto odhadu?

S **jistotou** lze říci pouze, že preference jsou 0.000 02 **až** 0.999 92, protože zaručíme jen to, že aspoň 200 voličů ji bude volit a aspoň 800 ji volit nebude (a to ještě nesmí změnit názor).

Více nám řekne informace, že preference jsou 16.7 **až** 23.3 % **s pravděpodobností** 99 %. (K tomu je potřebný správný výběr respondentů.)

Podobně: Kdy mám vyjet, abych byl na místě určitě včas?

Robustní statistika

- Klasická statistika pracovala především s chybami, které měly přibližně normální rozdělení, vznikaly **součem mnoha malých nezávislých chyb**.

Robustní statistika

- Klasická statistika pracovala především s chybami, které měly přibližně normální rozdělení, vznikaly **součem mnoha malých nezávislých chyb**.
- S digitalizací dat dochází častěji k **hrubým chybám** (chyby přenosu dat, špatně umístěná desetinná čárka apod.); postupy je potřeba jim přizpůsobit.

Robustní statistika

- Klasická statistika pracovala především s chybami, které měly přibližně normální rozdělení, vznikaly **součem mnoha malých nezávislých chyb**.
- S digitalizací dat dochází častěji k **hrubým chybám** (chyby přenosu dat, špatně umístěná desetinná čárka apod.); postupy je potřeba jim přizpůsobit.
- V klasickém modelu by takové chyby byly považovány za vysoce nepravděpodobné.

Paradox hromady

- „Odborníci Fondu OSN pro otázky populace tvrdí, že se roku 2050 nemusíme bát. Svět podle nich začne být lidstvu skutečně těsný až ve chvíli, kdy počet obyvatel planety dosáhne čtrnácti miliard.“[6]

[6] Sandra Krištofová: Rok 2050: bude nás devět miliard. Planeta začne být lidstvu těsná. MF Dnes, 22. 3. 2001

Paradox hromady

- „Odborníci Fondu OSN pro otázky populace tvrdí, že se roku 2050 nemusíme bát. Svět podle nich začne být lidstvu skutečně těsný až ve chvíli, kdy počet obyvatel planety dosáhne čtrnácti miliard.“[6]
- … a televize bude při tom, až se narodí 14 000 000 000-tý občan …

[6] Sandra Krištofová: Rok 2050: bude nás devět miliard. Planeta začne být lidstvu těsná. MF Dnes, 22. 3. 2001

Paradox hromady

- „Odborníci Fondu OSN pro otázky populace tvrdí, že se roku 2050 nemusíme bát. Svět podle nich začne být lidstvu skutečně těsný až ve chvíli, kdy počet obyvatel planety dosáhne čtrnácti miliard.“[6]
- … a televize bude při tom, až se narodí 14 000 000 000-tý občan …
- Řešení paradoxu hromady ve fuzzy logice: S každým narozeným člověkem roste pravdivost výroku „Země je přelidněná“.

[6] Sandra Krištofová: Rok 2050: bude nás devět miliard. Planeta začne být lidstvu těsná. MF Dnes, 22. 3. 2001

Obvyklé aplikace statistiky

- V reklamě: Průzkumy trhu, zaměření reklamy na cílovou skupinu.

Obvyklé aplikace statistiky

- V reklamě: Průzkumy trhu, zaměření reklamy na cílovou skupinu.
- V politice: Totéž.

Obvyklé aplikace statistiky

- V reklamě: Průzkumy trhu, zaměření reklamy na cílovou skupinu.
- V politice: Totéž.
- V podnikání: Průzkumy trhu, odhad adekvátního rozsahu výroby a odbytu.

Obvyklé aplikace statistiky

- V reklamě: Průzkumy trhu, zaměření reklamy na cílovou skupinu.
- V politice: Totéž.
- V podnikání: Průzkumy trhu, odhad adekvátního rozsahu výroby a odbytu.

Obvyklé aplikace statistiky

Umí naznačit, co by spolu mohlo souviset.

Obvyklé aplikace statistiky

Umí naznačit, co by spolu mohlo souviset.
Problém je určit, zda je příčinná souvislost a kterým směrem.

Obvyklé aplikace statistiky

Umí naznačit, co by spolu mohlo souviset.
Problém je určit, zda je příčinná souvislost a kterým směrem.
Bohatší lidé se dožívají vyššího věku.
Ženy matky mají vyšší IQ než bezdětné.
...

Neobvyklé aplikace statistiky a pravděpodobnosti

- V bankovnictví a kriminalistice: Vytipování podezřelých finančních operací.

Neobvyklé aplikace statistiky a pravděpodobnosti

- V bankovnictví a kriminalistice: Vytipování podezřelých finančních operací.
- V daňovém systému: Vytipování zfalšovaných dat.

Neobvyklé aplikace statistiky a pravděpodobnosti

- V bankovnictví a kriminalistice: Vytipování podezřelých finančních operací.
- V daňovém systému: Vytipování zfalšovaných dat.
- Ve strategických hrách (šachy, go).

Neobvyklé aplikace statistiky a pravděpodobnosti

- V bankovnictví a kriminalistice: Vytipování podezřelých finančních operací.
- V daňovém systému: Vytipování zfalšovaných dat.
- Ve strategických hrách (šachy, go).
- V komunikaci (ethernet): Pokud si dva počítače „skáčí do řeči“ (=začaly vysílat oba současně), odmlčí se oba na náhodně zvolenou dobu, než pokus o vysílání zopakují.

Podezřelé losování

- V r. 2006 v ČR proběhlo výběrové řízení: Z 16 firem bylo pro 4 zakázky vylosováno do užšího výběru vždy 5 firem, jejichž nabídky byly dále posuzovány. Jedna z firem se zúčastnila prostřednictvím čtyř filiálek, tj. podala 4 z 16 přihlášek.
- S pravděpodobností $> 99\,\%$ by firma uspěla už v některém ze 3 kol.

[7] Informace laskavě poskytl Prof. RNDr. J. Štěpán, DrSc., který se spolupracovníky vypracoval znalecký posudek a referoval o něm v televizi.

Podezřelé losování

- V r. 2006 v ČR proběhlo výběrové řízení: Z 16 firem bylo pro 4 zakázky vylosováno do užšího výběru vždy 5 firem, jejichž nabídky byly dále posuzovány. Jedna z firem se zúčastnila prostřednictvím čtyř filiálek, tj. podala 4 z 16 přihlášek.
- S pravděpodobností $> 99\,\%$ by firma uspěla už v některém ze 3 kol.
- Nebyla vybrána v žádném ze 4 kol.
  Firma zažalovala zadavatele, neboť považuje losování za zmanipulované.[7]

[7] Informace laskavě poskytl Prof. RNDr. J. Štěpán, DrSc., který se spolupracovníky vypracoval znalecký posudek a referoval o něm v televizi.

Podezřelé losování

- V r. 2006 v ČR proběhlo výběrové řízení: Z 16 firem bylo pro 4 zakázky vylosováno do užšího výběru vždy 5 firem, jejichž nabídky byly dále posuzovány. Jedna z firem se zúčastnila prostřednictvím čtyř filiálek, tj. podala 4 z 16 přihlášek.
- S pravděpodobností $> 99\,\%$ by firma uspěla už v některém ze 3 kol.
- Nebyla vybrána v žádném ze 4 kol.
  Firma zažalovala zadavatele, neboť považuje losování za zmanipulované.[7]
- Stejně neúspěšných bylo celkem 9 firem z 16. Kdyby se tak losovalo každý rok, tak nepravděpodobný výsledek by vyšel **v průměru jednou za 333 000 let**.

[7] Informace laskavě poskytl Prof. RNDr. J. Štěpán, DrSc., který se spolupracovníky vypracoval znalecký posudek a referoval o něm v televizi.

Neobvyklé aplikace statistiky a pravděpodobnosti[8]

[8]Mahdian, B.: Blind Digital Image Authentication. PhD. Thesis, FJFI ČVUT, Praha, 2008

Neobvyklé aplikace statistiky a pravděpodobnosti

Figure 1.7: An example of famous image manipulations of recent years.

Figure 1.8: An example of famous image manipulations of recent years.

Neobvyklé aplikace statistiky a pravděpodobnosti

- V kódování: Generování klíče pro šifrování obvykle začíná: „vezmeme dvě velká prvočísla p, q a vypočteme $N = p\,q$“.[9]

[9] Demlová, M.: Střípky z diskrétní matematiky. Inspirativní setkání s matematikou, FEL ČVUT, 8. března 2013

Neobvyklé aplikace statistiky a pravděpodobnosti

- V kódování: Generování klíče pro šifrování obvykle začíná: „vezmeme dvě velká prvočísla p, q a vypočteme $N = p\,q$“.[9]
- Ve skutečnosti se vezmou dvě náhodně vybraná čísla a provede se řada **nezávislých** testů, zda nejsou něčím dělitelná.

[9] Demlová, M.: Střípky z diskrétní matematiky. Inspirativní setkání s matematikou, FEL ČVUT, 8. března 2013

Neobvyklé aplikace statistiky a pravděpodobnosti

- V kódování: Generování klíče pro šifrování obvykle začíná: „vezmeme dvě velká prvočísla p, q a vypočteme $N = p\,q$“.[9]
- Ve skutečnosti se vezmou dvě náhodně vybraná čísla a provede se řada **nezávislých** testů, zda nejsou něčím dělitelná.
- Připouštíme, že to nemusí být prvočísla a vzniklá šifra RSA bude vadná, pokud riziko je $\leq 2^{-100} \doteq 7.9 \cdot 10^{-31}$.

[9] Demlová, M.: Střípky z diskrétní matematiky. Inspirativní setkání s matematikou, FEL ČVUT, 8. března 2013

Neobvyklé aplikace statistiky a pravděpodobnosti

- Buňky v mozku ptáků rozpoznávající směr (*head direction cells*) mají přesnost cca 45 stupňů. Průměruje se jich $> 200\,000$. Tím se dosahuje vysoká přesnost.[10]

---

[10] Pavel Němec: Neurální substrát magnetické kompasové orientace ptáků, přednáška, CTS, Praha 20.10.2011

[11] Matas, J.: Strojové učení a jak poznat ve fotografiích kluky od holek a jak z mnoha slabých studentů vytvořit jednoho experta. Inspirativní setkání s matematikou, FEL ČVUT, 13. prosince 2013

Neobvyklé aplikace statistiky a pravděpodobnosti

- Buňky v mozku ptáků rozpoznávající směr (*head direction cells*) mají přesnost cca 45 stupňů. Průměruje se jich $> 200\,000$. Tím se dosahuje vysoká přesnost.[10]
- Statistické rozpoznávání na základě „hlasování“ mnoha nedokonalých klasifikátorů.[11]

[10] Pavel Němec: Neurální substrát magnetické kompasové orientace ptáků, přednáška, CTS, Praha 20.10.2011

[11] Matas, J.: Strojové učení a jak poznat ve fotografiích kluky od holek a jak z mnoha slabých studentů vytvořit jednoho experta. Inspirativní setkání s matematikou, FEL ČVUT, 13. prosince 2013