

# Informatika a aplikovaná statistika

ZS 2015/2016  
FaF VFU Brno

Jiří Pazourek



# Motto:

*There are three kinds of lies: lies, damned lies, and statistics*

Když má hlavu v sauně a nohy v ledničce, hovoří statistik o příjemné průměrné teplotě.

*Franz Josef Strauß*

**NEW YORK** Nové laureáty humorných Nobelových cen ohlásili ve čtvrtek organizátoři jejich každoročního udělování na americké Harvardově univerzitě. O cenách od roku 1991 rozhoduje komise časopisu *Annals of Improbable Research* (Anály nepravděpodobného výzkumu), jejímiž členy jsou držitelé skutečných Nobelových cen. Ceny za směšný, výstřední, zbytečný nebo nevědecký výzkum udělují v deseti kategoriích.

Letos se vyznamenali zejména vědci z oboru biologie a lékařství, vyplývá z protokolu o oceněných laureátech. Zvláštní ohlas měl například objev, že prakticky všem savcům žijícím na Zemi trvá procedura močení zhruba 21 vteřin.

Mezi oceněnými se objevil i obětavý student americké Cornellovy univerzity, který se při studiu účinků včelích žihadel nechal dvěstěkrát bodnout na pětadvaceti místech svého těla. Zjistil přitom, že nejbolestivější je zásah žihadla do chrípí, horního rtu a pohlavního údu.

20. ZÁŘÍ 2015

# MOODLE

<https://amos.vfu.cz/moodle/>

The screenshot shows the website of the Faculty of Veterinary Medicine and Pharmacy, Brno University of Life Sciences. The header features the university's logo and name in Czech and English. A navigation bar contains links for 'INFORMACE O UNIVERZITĚ', 'ÚŘEDNÍ DESKA', 'VÝZKUM, VÝVOJ A ROZVOJ', 'STUDIUM, VZDĚLÁVÁNÍ', 'UCHAZEČI O STUDIUM', 'MÉDIA A VEŘEJNOST', 'LIDÉ NA VFU', 'RYCHLÉ ODKAZY', and 'AMBULANCE, POHOTOVOST, LABORATOŘE'. The 'RYCHLÉ ODKAZY' section is active, displaying a list of links under the heading 'Odkazy' and 'Servery VFU a odkazy na ně'. A red arrow points from the 'RYCHLÉ ODKAZY' menu item to the 'Veterinární toxikologická databáze' link in the list.

**VETERINÁRNÍ A FARMACEUTICKÁ UNIVERZITA BRNO**

**UNIVERSITAS VETERINARIA ET PHARMACEUTICA BRUNENSIS**

**Fakulta veterinárního lékařství**

**Fakulta veterinární hygieny a ekologie**

**Farmaceutická fakulta**

**Včelařství**

**INFORMACE O UNIVERZITĚ**   **ÚŘEDNÍ DESKA**   **VÝZKUM, VÝVOJ A ROZVOJ**   **STUDIUM, VZDĚLÁVÁNÍ**   **UCHAZEČI O STUDIUM**   **MÉDIA A VEŘEJNOST**   **LIDÉ NA VFU**   **RYCHLÉ ODKAZY**   **AMBULANCE, POHOTOVOST, LABORATOŘE**

VFU Brno > Rychlé odkazy

**RYCHLÉ ODKAZY**

**AKTUALITY**

**2012-09-10**  
VFU uzavřela spolupráci se SurGal Clinic za přítomnosti premiéra  
Rektor VFU Vladimír Večerek podepsal významnou smlouvu o spolupráci s nemocnicí v Brně SurGal Clinic. Smlouva byla slavnostně podepsána za přítomnosti předsedy české vlády Petra Nečase. Vysoká škola bude s klinikou spolupracovat v řadě oblastí. Oba partneri vzájemně dojednali společné plnění vědecko-výzkumných úkolů a spolupráci při vzdělávací činnosti.

**Odkazy**

**Servery VFU a odkazy na ně**

**Veterinární toxikologická databáze**

- ISKaM - Kolejní systém
- Moodle - E-learning
- OBDU Public
- EZproxy - přístup k elektronickým on-line zdrojům pro studenty a zaměstnance VFU Brno odkudkoli z počítačů mimo VFU Brno
- Webové rozhraní pro zapisování sportů na UTVS
- SafeQ - Kopírovací a tiskové služby (přístup jen z areálu VFU)

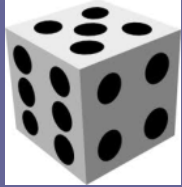
<http://new.euromise.org/czech/tajne/ucebnice/html/html/statist.html> (Zvárová)

<http://www.itl.nist.gov/div898/handbook/>

# 1.

Cíle:

- Chápat vliv „náhody“ a význam statistiky
- Chápat pojmy **náhodný jev (pokus)**, výběrový soubor, statistický znak
- Činnosti statistiky
- Umět sbírat statistická data = z pozorování sestavit tabulku (prostou, skupinovou, kombinační)
- Umět vytvořit histogram nebo jiný graf



# K čemu je statistika?



Statistika nám pomáhá odhalovat zákonitosti v našem **stochastickém\*** světě. Umožňuje nám **odfiltrovat** či alespoň **číselně popsat** vliv „náhody“ na naše experimentální data.

*Statistika* je v určitém smyslu jazykem pro shromažďování dat, manipulaci s nimi a jejich kvantitativní interpretaci.

\* náhodilý, náhodný, opak deterministický. Primitivní příklad stochastického jevu je házení hrací kostkou nebo mincí, protože výsledek se nedá jednotlivě předvídat, nicméně velké množství takových jevů se dá zkoumat **statisticky** co do pravděpodobnosti výsledků.

# Náhodný děj, náhodný pokus, náhodný jev

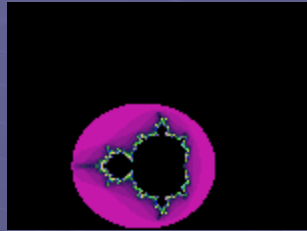
- Každý děj (proces) v přírodě probíhá za jistých podmínek...
- Zajímáme se pouze o děje, které se za **stejných, opakovaných** podmínek opakují = tzv. **HROMADNÉ DĚJE**

# děj náhodný, děj deterministický

- Jsou děje, které mají při zachování podmínek vždy stejný výsledek = **děje deterministické**.
- Jiné, ačkoliv jsou podmínky stále stejné, mají výsledek pokaždé jiný = **děje náhodné** = těmi se zabývá **počet pravděpodobnosti**

# "Náhoda" vs. předurčení

*přísný determinismus = při dokonalé znalosti počátečních podmínek lze, alespoň hypoteticky, přesně předpovědět budoucí chování jakéhokoli systému a světa jako celku.*



Laplaceova mechanistická představa vesmíru jako "**hodinového stroje**":

*"Rozum, který by v daný okamžik znal všechny síly řídící přírodu a vzájemné polohy objektů v ní a který by byl dostatečně výkonný, aby tato data mohl podrobit analýze, by mohl shrnout do jediného systému pohyb největších vesmírných těles i nejlehčích atomů: pro takový rozum by nebylo nic nejisté a budoucnost stejně jako minulost by byly přítomné před jeho zrakem" (Pierre Simon Laplace, 1812).*

Kvantová fyzika však ukazuje, že v mikroměřítcích prostoru a času je příroda skutečně a principiálně **řízena náhodou**. Např. to, zda se určitá elementární částice či radioaktivní jádro v daném okamžiku **rozpadne či nerozpadne**, je náhodná záležitost; mezi jádrem, které se má vzápětí rozpadnout a tím, které nikoli (nebo až za dlouhou dobu), nelze nalézt vůbec žádný rozdíl. Teprve v souboru velkého počtu částic či radioaktivních jader se objevují **(statistické) pravidelnosti**.



Př.: komplex podmínek ....normální atmosférický tlak,  
jsme u hladiny moře

děj .... ohříváme destilovanou vodu

Výsledek vždy stejný ... voda vře při 100 stupních

--> deterministický děj

Př.: komplex podmínek..... chov laboratorních myší, životní  
podmínky jsou sterilní a pořád stejné

děj .... naočkujeme jim nějaký virus

Výsledek náhodný ... některé myši zemřou, jiné ne

--> náhodný děj

někdy změna podmínek učiní z deterministického jevu jev náhodný - v prvním případě,  
když vezmeme libovolnou vodu (minerální, povrchovou, mořskou) nebo dokonce  
libovolnou kapalinu, její bod varu se stane náhodným

Studium „náhodných“ jevů, jak bylo řečeno, slouží k **zobecnění** – toto zobecnění může být začátkem poznání vztahu a nakonec dokonce kauzality tzv. přírodního zákona.

Např. mimozemšťan by po přiletu pozoroval např. mravence a lidí : jejich chování a zvyklosti: podle určitých kritérií by mohl považovat mravence za pokročilejší (organizovanější). Chování lidí bylo pro něj velmi „náhodné“ – bezdůvodné. Až dlouhým studiem by nakonec zjistil některé zákonitosti biologické, sociální, ...

**Vytrhování faktů bez souvislostí může vést k nesmyslným závěrům: Nejvíce lidí zemře v posteli. Z toho logicky vyplývá poučení: Chcete-li žít, neuléhejte do postele.**





# Náhodný jev – náhodný pokus

Takové pokusy, které jsou neomezeně mnohokrát opakovatelné, nazýváme **náhodné pokusy**. Jednoduchými příklady náhodných pokusů jsou: házení hracími kostkami nebo mincemi, tahání losů z osudí...

**Avšak teprve náhodné pokusy jiného druhu činí z počtu pravděpodobnosti důležitou vědu: téměř všechny experimenty prováděné v biologii, chemii, medicíně atd. mají svou náhodnou stránku.**

Ani měření např. délky tyče, hmotnosti tělesa, pH atd. nedává „jasnou“ odpověď – měření s tímž experimentátorem i tímž měřidlem přinese mírně odlišné údaje (zde máme na mysli reálnou přesnost měření, viz dále).

---

**Počtem pravděpodobnosti (součástí statistiky) se zabýváme proto, abychom tuto "náhodnou" stránku zachytili či vhodně vyjádřili.**

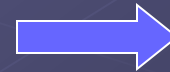
# Vlivy "náhody" na naše experimentální data

1. Chyba malých čísel, omezená velikost výběrového souboru (vzorku) - reprezentativnost vzorku!
2. Studovaný jev (například vliv léku na tělesnou váhu) může být překryt vlivy jiných faktorů, které jsme nezohlednili (tělesná teplota, genetická dispozice, metabolismus, střídání ročních období...)
3. Experimentální nástroje, které k získávání dat používáme, jsou technicky omezené - chyby měření (přesnost měření)

# ad 1) Populace a výběr (základní soubor a výběrový soubor)



populace



Výběr 43 osob

# ad 2) Vědecký pokus

Kontrolované faktory:  
Koncentrace, teplota, pH,  
objem, chemikálie....

NEkontrolované faktory:

Faktor 1  
(tlak)

Faktor 2  
(vlhkost)

Faktor 3  
(počet  
dioptrií)

Faktor 4  
(nemoc  
dětí)

Faktor 5  
(toaletní papír  
na WC)

# ad 3) Přesnost měření

Tabulka stanovení koncentrace (mg/l)  
téhož analytu ve vzorku  
v různých dnech / různými osobami / pracovišti:  
Výsledek - přístrojové možnosti

č.	c (mg/l)
1	26,4
2	26,0
3	26,1
4	27,4
5	26,7
6	28,1
7	24,6





# Činnost statistiky

## Statistické metody **explorační** (deskriptivní, induktivní)

sběr data pozorováním, popis, třídění, klasifikace, charakterizace.... **Popis** jevů, jaké **byly**.

## Statistické metody **konfirmační** (prediktivní, deduktivní)

Tvorba a potvrzování hypotéz, modelů - jaké **budou projevy** dané jevu?

# Náhodný jev - Statistické pozorování - data

Výsledkem statistického šetření, pozorování či měření je většinou **nepřehledný** soubor dat. Proto se snažíme charakterizovat tato data menším počtem čísel, která jsou pro nás srozumitelnější a lze z nich vyčíst hledané obecné vztahy a zákonitosti.

K tomu **setřídíme** prvky souboru podle hodnot sledovaného znaku (proměnné). Vytváříme skupiny prvků souboru se stejnou hodnotou znaku a určujeme počet hodnot v dané skupině, tzv. četnost neboli frekvenci. Při třídění získáme **rozdělení četností, které udává kolikrát se jednotlivé hodnoty znaku (tj. varianty) v souboru opakuji.**

Toto rozdělení četností lze shrnout v tabulce. Celkem jednoduché je to v případě nespojitě proměnné, když proměnná nabývá jen několika variant hodnot (známka ve škole, zaokrouhlená výška podobně starých osob...).

# Druhy statistických znaků

statisticky sledovaná vlastnost  
náhodného jevu = statistický znak.

# Statistické znaky

Vlastnosti, sledované na prvcích (jednotkách statistického šetření) výběru či populace, nazýváme **znaky (veličiny)**. Znaky, které sledujeme, dělíme na kvalitativní a kvantitativní.

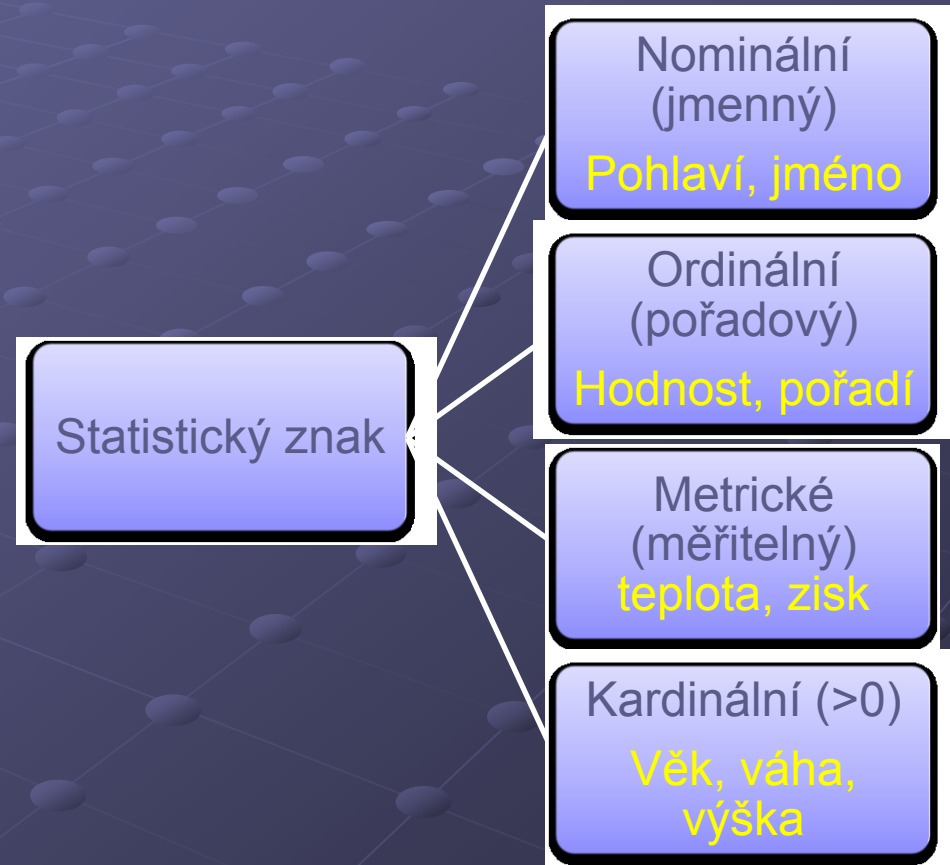
Jsou-li varianty zkoumané vlastnosti dány pouze slovním vyjádřením, jde o *znak kvalitativní*. Jsou-li varianty vyjádřeny **číslem**, jde o **znak kvantitativní**.

**Kvantitativní znak** je buď :

- spojitý, jestliže jednotlivé varianty znaku mohou nabývat jakékoliv hodnoty z určitého intervalu nebo rozmezí (*výška, hmotnost* apod.), či
- diskrétní (nespojité), jsou-li varianty znaku vyjádřeny oddělenými čísly. V případě pouze dvou možných hodnot je to znak alternativní (např. *pohlaví*), jinak množný (*počet onemocnění, počet dětí* apod.).



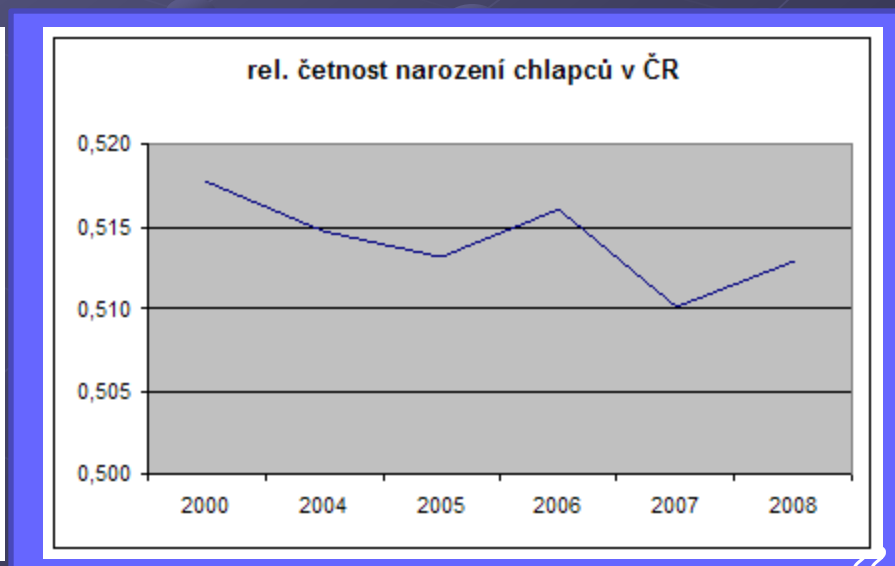
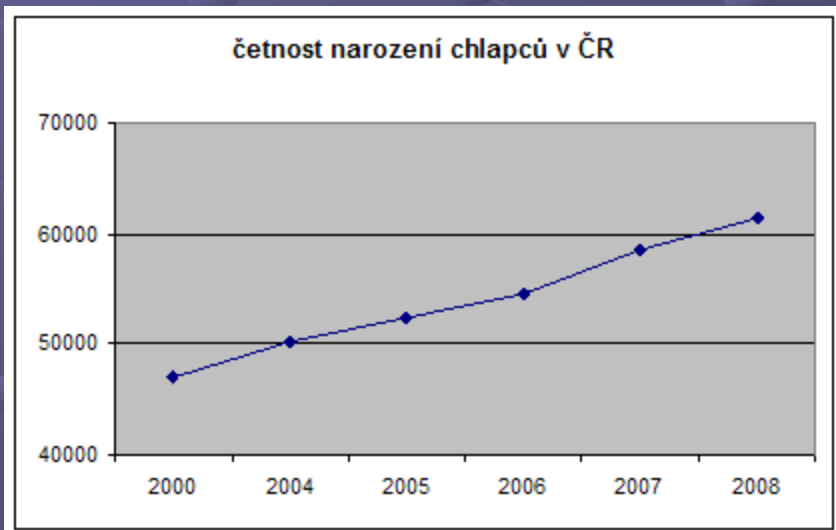
# Statistických znak (veličina, proměnná)



# Náhodný jev – četnost výskytu

**Náhodným jevem** označujeme výsledek náhodného pokusu, o kterém lze až po provedení tohoto pokusu rozhodnout, zda nastal nebo nenastal (např. náhodný jev  $A$  "narození chlapce").

Četnost, s jakou nastává náhodný jev  $A$ , můžeme charakterizovat podílem  $m/n$ , kde  $n$  je délka posloupnosti (rozsah výběru=počet rodiček) a  $m$  je počet narozených chlapců. Číslo  $m$  nazýváme **absolutní četnost** a podíl  $m/n$  ( $0-1$ ) **relativní četnost** výskytu náhodného jevu  $A$  ve výběru o rozsahu  $n$ . Grafické znázornění četnosti „narození chlapce“ v ČR:





# Relativní četnost a pravděpodobnost

## Statistická definice (von Misessova)

**Nejpodstatnější! Je nutná pro náš další výklad statistiky.**

$$P(A) = \lim_{n \rightarrow \infty} \frac{m(A)}{n}$$

Podle této definice je relativní četnost jevu, se kterou jsme již seznámili v popisné statistice, odhadem pravděpodobnosti. Empiricky určená relativní četnost kolísá kolem hodnoty pravděpodobnosti a přibližuje se k ní s rostoucím počtem pokusů.

# Sběr a prezentace dat

Tabulka prostá  
(netříděná data)

**Tabulka 1.1:** Tělesná výška 27 dětí ve věku 9,5-10 let

Tělesná výška v cm								
130	140	136	141	139	133	149	151	139
136	138	142	127	147	139	135	141	143
132	146	151	146	141	141	131	142	141

Tabulka skupinová  
(data tříděná dle jednoho znaku)

**Tabulka 1.2:** Rozdělení dětí ve věku 9,5 - 10 let podle tělesné výšky (délka třídního intervalu 5 cm)

Střed třídního intervalu	Počet dětí
125	1
130	3
135	4
140	12
145	4
150	3
Celkem	27

Tabulka kombinační  
(data tříděná dle více znaků)

**Tabulka 1.3:** Rozdělení dětí ve věku 9,5 - 10 let podle pohlaví a tělesné výšky (délka třídního intervalu 5 cm)

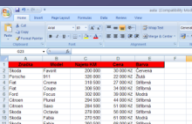
Střed třídního intervalu	Počet chlapců	Počet dívek	Počet dětí
125	1	-	1
130	1	2	3
135	3	1	4
140	7	5	12
145	1	3	4
150	2	1	3
Celkem	15	12	27

**Tabulka 1.4:** Vztah mezi věkem manželky a věkem manžela

Věk manžela	Věk manželky							Celkem
	15-24	25-34	35-44	45-54	55-64	65-74	75 a více	
15-24	193	50	1	-	-	-	-	244
25-34	231	1162	108	4	-	-	-	1505
35-44	12	408	977	92	4	-	-	1493
45-54	1	36	320	652	66	3	-	1078
55-64	-	5	37	211	358	34	1	646
65-74	-	1	6	24	105	133	10	279
75 a více	-	-	1	4	10	32	25	72
Celkem	437	1664	1450	987	543	202	36	5317



Tabulka korelační  
(závislost dvou kvant. znaků)





# Histogram četností

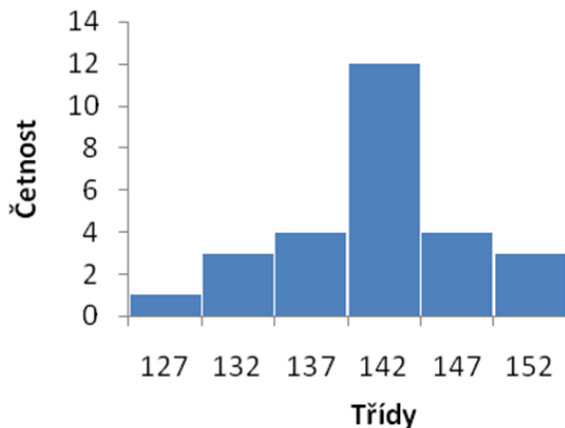
**Tabulka 1.1:** Tělesná výška 27 dětí ve věku 9,5-10 let

Tělesná výška v cm								
130	140	136	141	139	133	149	151	139
136	138	142	127	147	139	135	141	143
132	146	151	146	141	141	131	142	141



127	1	3,70%
128	0	
129	0	
130	1	3,70%
131	1	3,70%
132	1	3,70%
133	1	3,70%
134	0	
135	1	3,70%
136	2	7,41%
137	0	
138	1	3,70%
139	3	11,11%
140	1	3,70%
141	5	18,52%
142	2	7,41%
143	1	3,70%
144	0	
145	0	
146	2	7,41%
147	1	3,70%
148	0	
149	1	3,70%
150	0	
151	2	7,41%
	27	100%

**Histogram**



interval	Třidy	Četnost
123-127	127	1
128-132	132	3
133-137	137	4
138-142	142	12
143-147	147	4
148-152	152	3



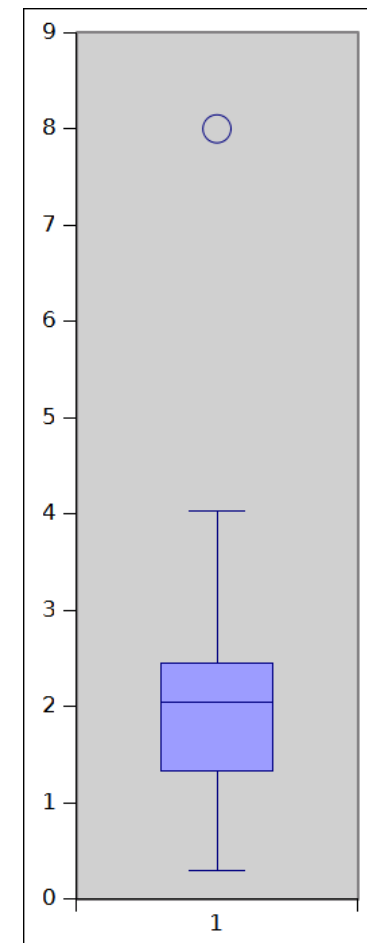
intervalové  
rozdělení

# Příklad z popisné statistiky:

Rybáři vylovili malý rybník v městském parku.

Byl to první výlov po 10ti letech, byly v něm ryby různých velikostí, celkem 30 kusů. Jejich váhy jsou tyto:

0,3 1,9 2,1 0,48 2,1 3,1 0,8 2,3 3,2 1  
1,6 2,2  
2,5 1,4 2,2 4 1,1 2 3,1  
1,95 2 1,5 2,56 8 1,01 2,15 4,3  
0,9 2,2 1,3



Vypočítejte:

Variační rozpětí  $R$ ... tj. maximální rozpětí jejich vah.

Průměrnou váhu ulovených ryb..

Rozhodněte pomocí mediánu, jestli je víc těžších ryb než průměr..

Mezikvartilové rozpětí  $R_q$ ,

tj. rozmezí vah, ve kterém je 50% ulovených kusů

Vypočtete směrodatnou odchylku a variační koeficient .. míry rozptýlenosti..

Rozdělte data do 6ti tříd a sestrojte histogram...

Nakonec na přiložený papír nakreslete krabicový diagram...

Jsou nějaké odlehlé hodnoty ... tj. takové, které jsou od krabice vzdálené o víc než  $1,5 R_q$ ?



<http://www.czso.cz/csu/redakce.nsf/i/statistiky>