

# I. Statistika ve vědecké praxi



Pozice statistické analýzy ve vědě a klinické praxi  
Význam statistických výstupů

# Anotace

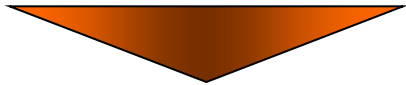


- Statistická analýza biologických dat je jedním z nástrojů, s jejichž pomocí se snažíme zjistit odpovědi na naše otázky týkající se pochopení živé přírody. Jako každý nástroj je i statistickou analýzu nezbytné na jedné straně korektně využívat a na druhou stranu nepřeceňovat její možnosti.
- Klíčovým faktem při statistické analýze dat je nahlížení na realitu prostřednictvím vzorku a přijmutí toho, že výsledky naší analýzy jsou jen tak dobré, jak dobrý je náš vzorek. Reprezentativnost a náhodnost vzorku spolu s jeho velikostí jsou důležité faktory ovlivňující věrohodnost našich závěrů.

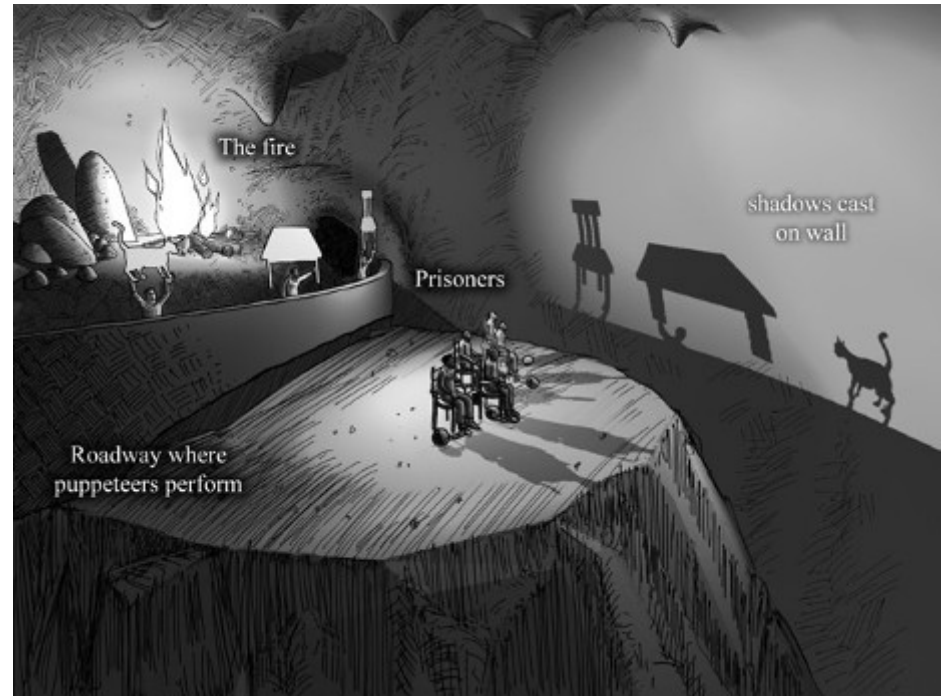
# Výzkum, realita, statistika



- Výzkum je naším způsobem porozumění realitě
- Ale jak přesné a pravdivé je naše porozumění?



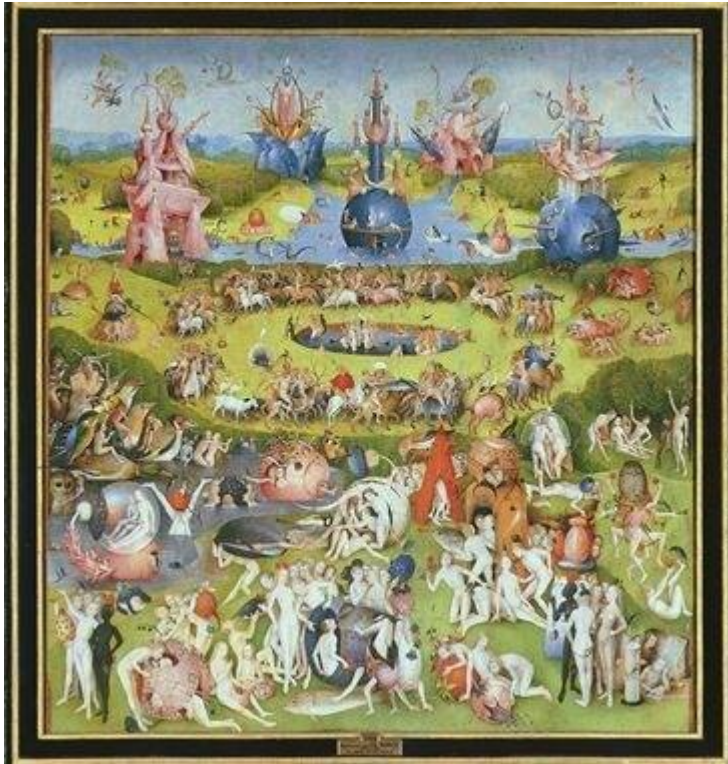
**Statistika** je  
jedním z nástrojů  
vnášejících do  
našich výsledků  
určitou spolehlivost.



# Význam variability

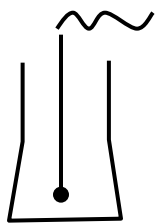


- Naše realita je variabilní a statistika je vědou zabývající se variabilitou
- Korektní analýza variability a její pochopení přináší užitečné informace o naší realitě
- V případě deterministického světa by statistická analýza nebyla potřebná



# Biostatistika - různé přístupy k variabilitě

## Variabilita opakovaných měření



Data

2,1  
2,8  
3,2  
1,2  
5,2  
2,9

chyba

## Variabilita znaku v populaci



165 cm



140 cm



182 cm



163 cm

rozptyl znaku, přirozená variabilita

## Variabilita modelovaných dat



chyba = nepřesnost modelu

## Variabilita časových řad



fluktuační, časová proměnlivost

## Variabilita ve skladbě biologických společenstev

DRUH 1	15
DRUH 2	30
DRUH 3	40
DRUH 4	14



biodiverzita

# Pojem VARIABILITA má mnoho významů .....



*... a ty určují přístup k jejímu  
hodnocení*

*Maskování a  
minimalizace  
vlivu*

*Respektování a  
odhadování vlivu*

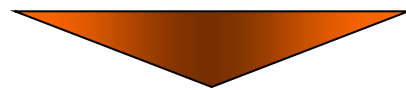
*Přímé využití k  
predikcím chování  
systému*

# Statistika – význam a definice



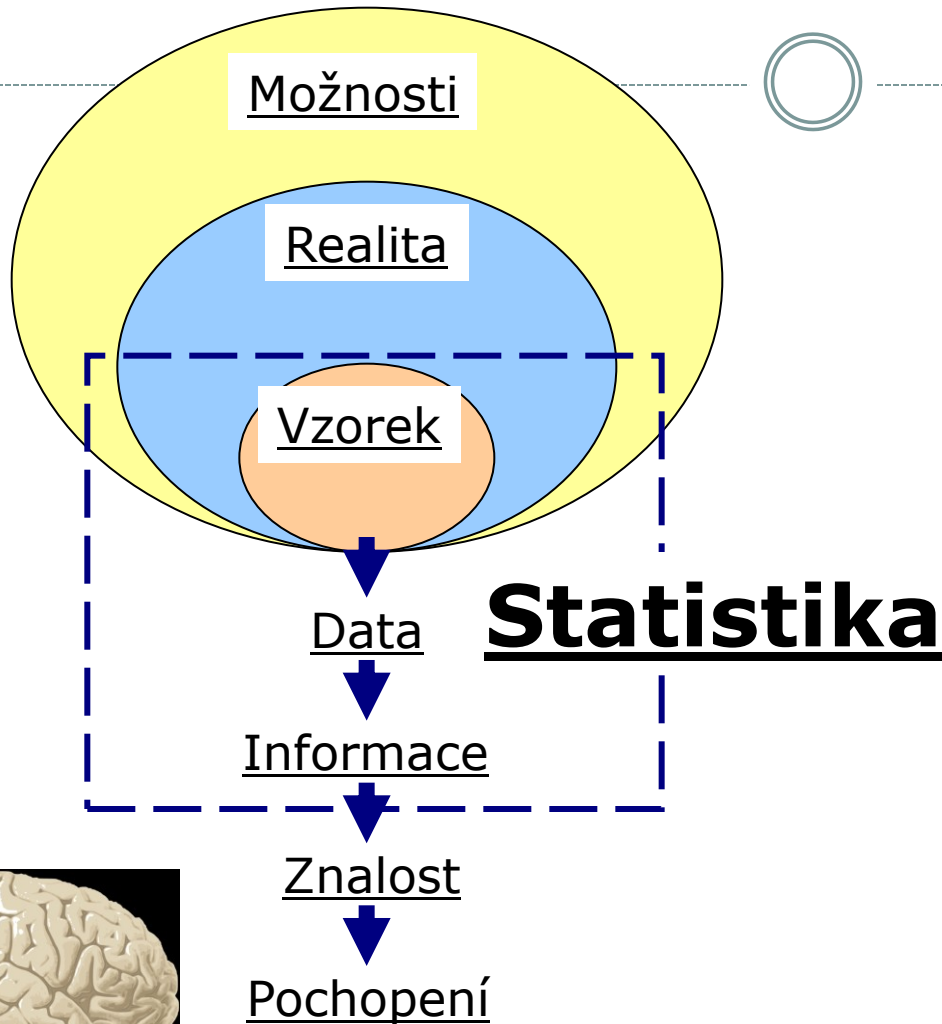
## **WWW.WIKIPEDIA.ORG:**

Statistika je matematickou vědou zabývající se shromážděním, analýzou, interpretací, vysvětlením a prezentací dat. Může být aplikována v širokém spektru vědeckých disciplín od přírodních až po sociální vědy. Statistika je využívána i jako podklad pro rozhodování, kdy nicméně může být záměrně i nevědomky zneužita.



Statistika využívá matematické modely reality k zobecnění výsledků experimentů a vzorkování. Statistika funguje korektně pouze pokud jsou splněny předpoklady jejích metod a modelů.

# Co může statistika říci o naší realitě?



Statistika není schopna činit závěry o jevech neobsažených v našem vzorku.

Statistika je nasazena v procesu získání informací z vzorkovaných dat a je podporou v získání naší znalosti a pochopení problému.

Statistika není náhradou naší inteligence !!!



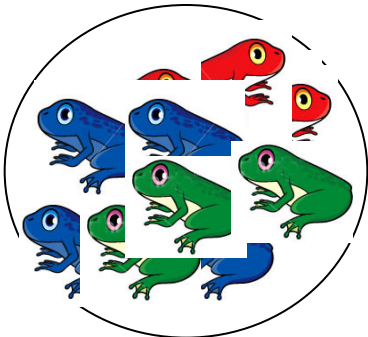


# Cílová populace

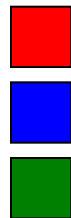


- **Cílová populace – klíčový pojem statistického zpracování**
  - Skupina objektů o nichž se chceme něco dozvědět (např. pacienti s danou diagnózou, všichni lidé nad 60 let, měření hemoglobinu v dané laboratoři)
  - Musí být definována ještě před zahájením sběru dat
  - Na cílové populaci probíhá vzorkování dat, které musí cílovou populaci dobře (reprezentativně) charakterizovat

**Cílová populace**



**Klíčové faktory  
cílové populace**



**Design  
experimentu a  
vzorkovací plán**



**Vzorkování a  
analýza dat**



# Statistika a zobecnění výsledků



***Neznámá  
cílová  
populace***

***Vzorek***

***Analýza***

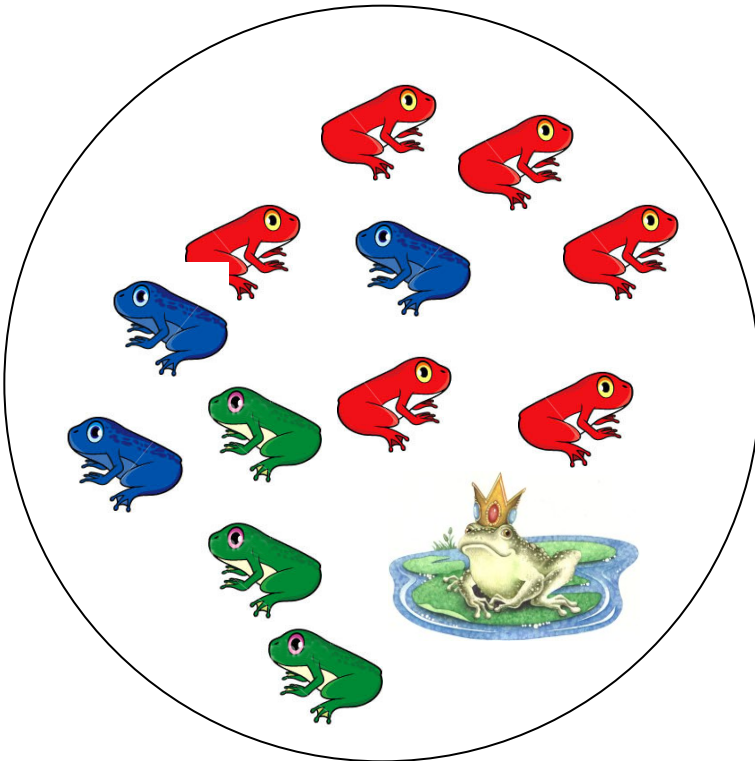
***Díky zobecnění  
výsledků známe  
vlastnosti cílové  
populace***

- Cílem analýzy není pouhý popis a analýza vzorku, ale zobecnění výsledků ze vzorku na jeho cílovou populaci
- Pokud vzorek nereprezentuje cílovou populaci, vede zobecnění k chybným závěrům

# Vzorkování a jeho význam ve statistice



- Statistika hovoří o realitě prostřednictvím vzorku!!!
  - Statistické předpoklady korektního vzorkování



**Representativnost:** struktura vzorku musí maximálně reflektovat realitu

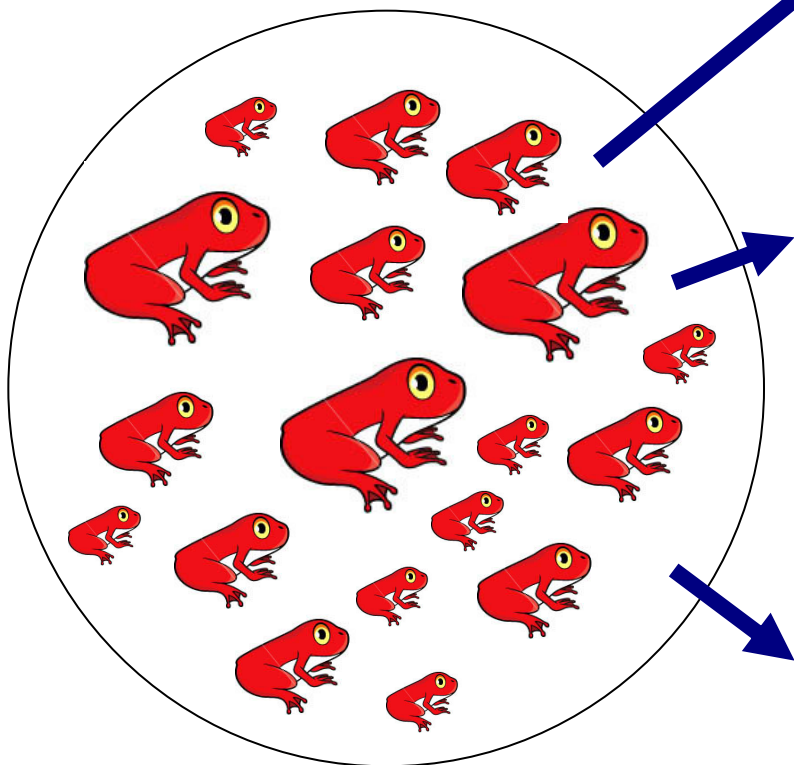


**Nezávislost:** několikanásobné vzorkování téhož objektu nepřináší ze statistického hlediska žádnou novou informaci



# Velikost vzorku a přesnost statistických výstupů


Existuje skutečné rozložení  
a skutečný průměr měřené  
proměnné



Z jednoho měření nezjistíme nic

Vzorek:  → ??????

Vzorek určité velikosti poskytuje  
odhad reálné hodnoty s  
definovanou spolehlivostí

Vzorek:  → **Odhad  
průměru  
atd.**

Vzorkování všech existujících  
objektů poskytne skutečnou  
hodnotu dané popisné  
statistiky, nicméně tento přístup je  
ve většině případech nereálný.

# Různá role statistiky při různě velkém vzorku



**Malá data**

**Velká data**

**Obrovská data**



**Umění  
prodat**



**Umění  
pochopit**



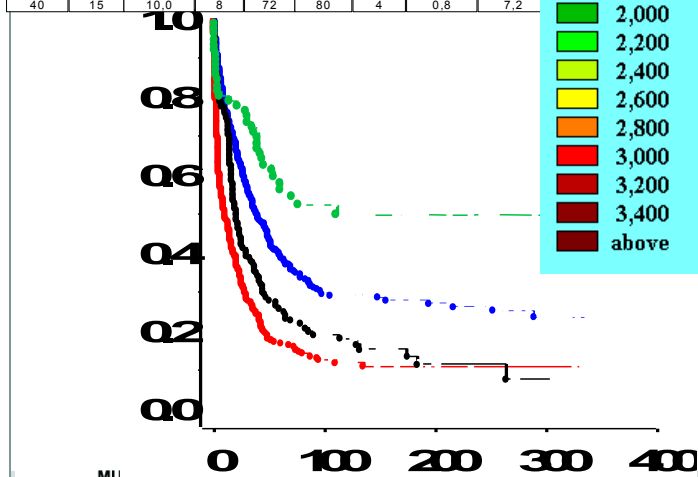
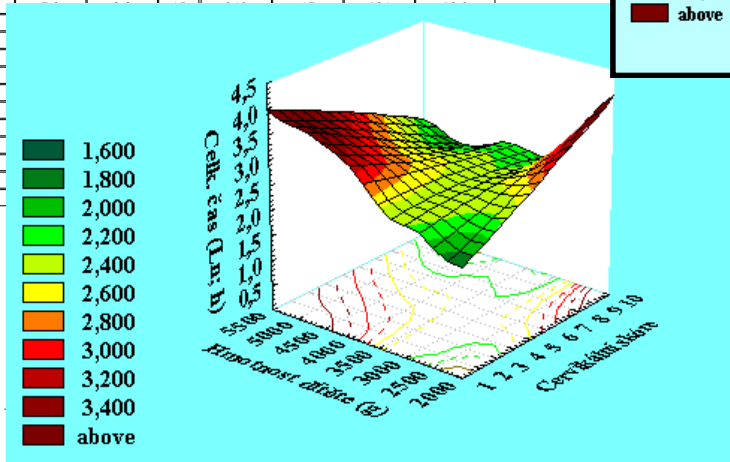
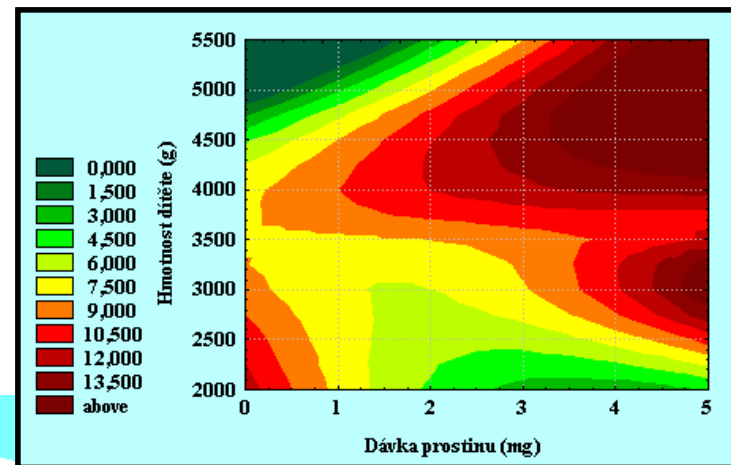
**Umění  
uchopit**



# Přístup biostatistiky

Pacient	Clovek	aLeu	aTy%	aSe%	aNeu%	aLy%	aTy	aSe	aNeu	aLy	aHtc	aCLsk	aCLNeus	aCLOZ	aCLNeuO
		cell.10 <sup>9</sup> / <sub>l</sub>	%	%	%	%	cell.10 <sup>9</sup> / <sub>l</sub>	cell.10 <sup>9</sup> / <sub>l</sub>	cell.10 <sup>9</sup> / <sub>l</sub>	cell.10 <sup>9</sup> / <sub>l</sub>	%	mV.s.10 <sup>3</sup>	mV.s.10 <sup>3</sup>	mV.s.10 <sup>3</sup>	mV.s.10 <sup>3</sup>
3	1	4									33	72		32	
4	2	7,6	8	58	66	24	0,6	4,4	5,0	1,8	33	95	19	48	10
8	3	4	3	52	55	40	0,1	2,1	2,2	1,6	22	77	35	33	15
11	4	6,1	5	59	64	35	0,3	3,6	3,9	2,1	33	103	26	49	13
12	5	6,9									37	81	13	45	7
14	6	5,9									32	137	33	61	15
16	7	8									34	151	20	59	8
20	8	9,6									40	77	11	38	5
21	9	6									32	120	26	52	11
22	10	3,3									28	81	42	24	12
37	11	3,8	10	60	70	30	0,4	2,3	2,7	1,1	32	111	42	29	11
38	12	6,4	2	76	78	17	0,1	4,9	5,0	1,1	25	366	73	115	23
39	13	6,8	1	57	58	39	0,1	3,9	3,9	2,7	20	234	59	71	18
49	14	8,5	7	67	74	26	0,6	5,7	6,3	2,2	30	156	25	108	17
51	15	9,3	7	57	64	35	0,7	5,3	6,0	3,3	35	129	21	23	4
52	16	2,2	10	56	66	34	0,2	1,2	1,5	0,7	33	46	30	12	8
55	17	9,9	3	78	81	10	0,3	7,7	8,0	0,1	30	189	24	140	18
56	18	5	2	80	82	13	0,1	4,0	4,1	0,7	26	101	25	54	13
6	1	8,8	11	72	83	12	1,0	6,3	7,3	1,1	44	268	36,6	145	19,9
9	2	9,2	2	66	68	28	0,2	6,1	6,3	2,6	42	168	26,9	76	12,2
13	3	10,0	7	83	90	8	0,7	8,3	9,0	0,8	54	181	20,1	81	9
15	4	9,6	1	75	76	23	0,1	7,2							
17	5	6,0													
19	6	7,2	2	78	80	18	0,1	5,6							
24	7	8,2	1	72	73	25	0,1	5,9							
26	8	10,3	1	85	86	3	0,1	8,8							
29	9	5,0	1	74	75	21	0,1	3,7							
30	10	11,9	1	51	52	47	0,1	6,1							
31	11	7,2	3	53	56	29	0,2	3,8							
32	12	10,8	36	50	76	8	3,9	5,4							
33	13	11,8	22	54	76	16	2,6	6,4							
34	14	17,0	1	82	83	16	0,2	13,9							
40	15	10,0	8	72	80	4	0,8	7,2							

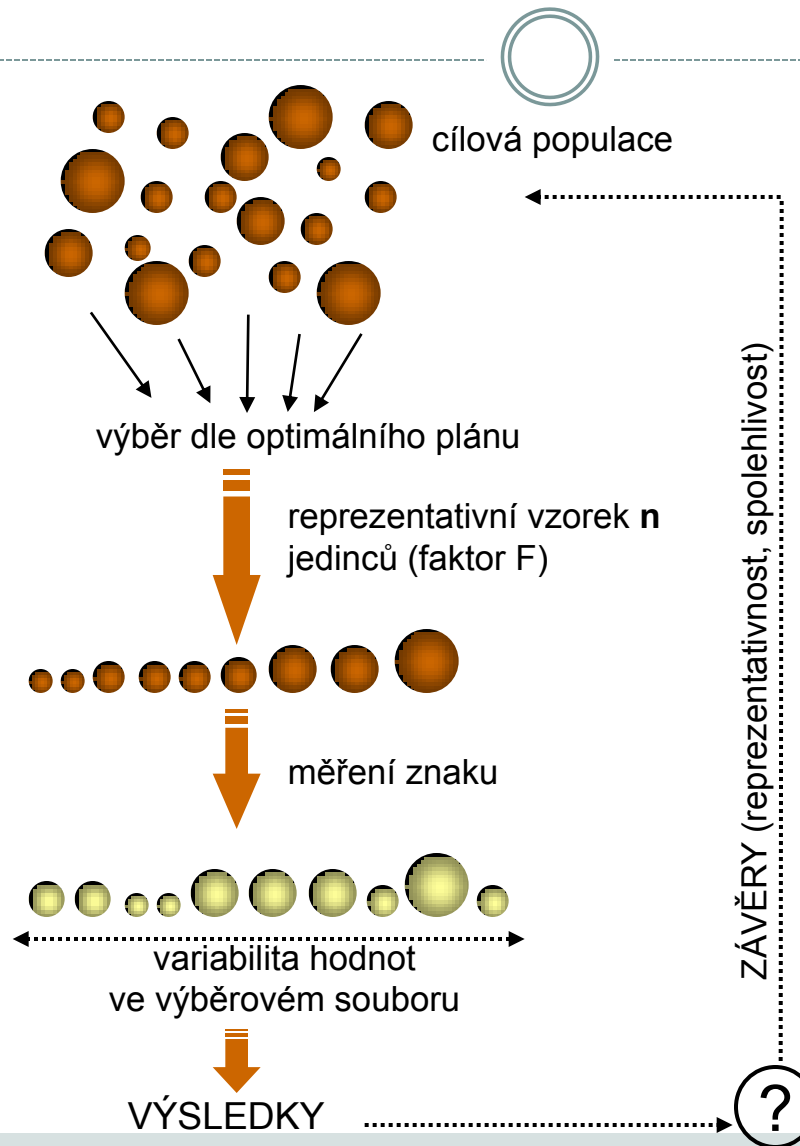
**Data**



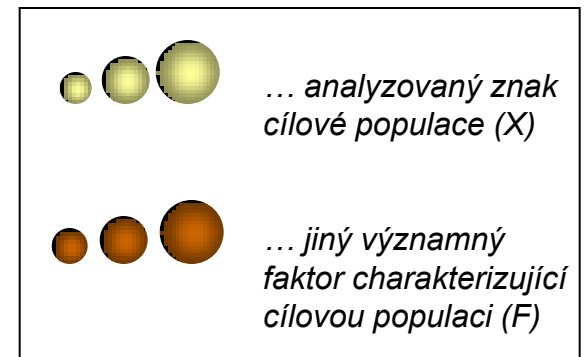
Schopnost: vidět data – komunikovat – interpretovat - prodávat

# Experimentální design: nezbytná výbava biologa

Účel analýzy:  
Popisný

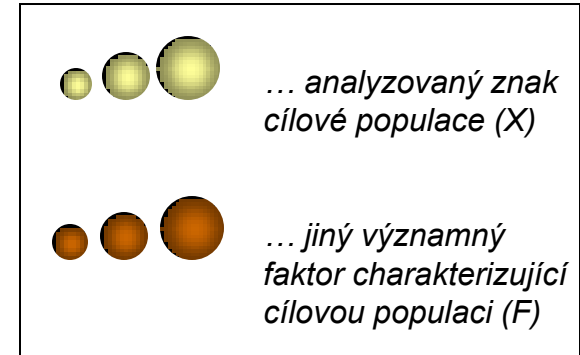
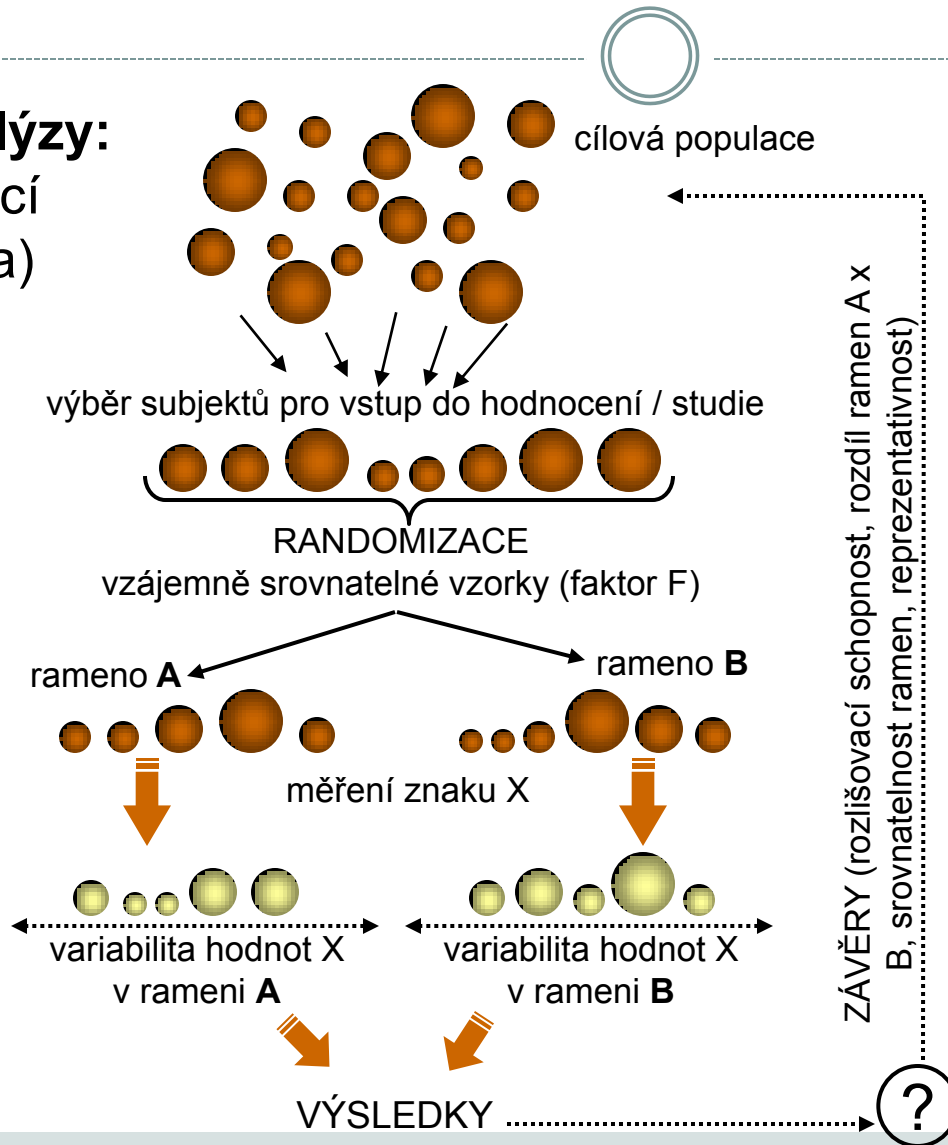


**?**  
**Reprezentativnost**  
**Spolehlivost**  
**Přesnost**



# Experimentální design: nezbytná výbava biologa

**Účel analýzy:**  
Srovnávací  
(2 ramena)



**?**  
**Srovnatelnost**  
**Spolehlivost**  
**Přesnost**



# Praktická a statistická významnost



- Samotná statistická významnost nemá žádný reálný význam, je pouze měřítkem náhodnosti hodnoceného jevu
- Pro vyhodnocení reálné významnosti je nezbytné znát i reálně významné hodnoty

		<i>Praktická významnost</i>	
		<b>ANO</b>	<b>NE</b>
<i>Statistická významnost</i>	<b>ANO</b>	OK, praktická i statistická významnost je ve shodě, jednoznačný závěr	Významný výsledek je statistický artefakt velkého vzorku, prakticky nevyužitelné
	<b>NE</b>	Výsledek může být pouhá náhoda, neprůkazný výsledek	OK, praktická i statistická významnost je ve shodě, jednoznačný závěr

# Obecné schéma využití statistické analýzy



## **Experimentální design**

Jak velký vzorek je nezbytný pro statisticky relevantní výsledky?  
Klíčová stratifikační kritéria cílové populace.

## **Vzorkování**

Vzorkovací plán zabezpečující náhodnost a reprezentativnost vzorku.

## **Uložení a management dat**

Uložení dat ve vhodné formě a jejich vyčištění předcházející vlastní analýze je klíčovým krokem statistické analýzy.

## **Vizualizace dat**

Grafická inspekce dat je nezbytným krokem analýzy vzhledem ke schopnosti lidského mozku primárně akceptovat obrazová data. Poskytne vhled do dat, představu o jejich rozložení, vazbách proměnných apod.

## **Popisná analýza**

Popisná analýza umožňuje vyhodnotit srovnáním s existující literaturou realističnost naměřených rozsahů dat.

## **Testování hypotéz**

Testování vazeb mezi různými proměnnými s cílem navzájem vysvětlit jejich variabilitu a tím přispět k pochopení řešeného problému.

## **Modelování**

Možným vyvrcholením analýzy je využití získaných znalostí a pochopení problému k vytvoření prediktivních modelů.

# Stochastické modelování: predikce neurčitých jevů



Prospektivně – modelově - postihuje chování jevů při respektování variability

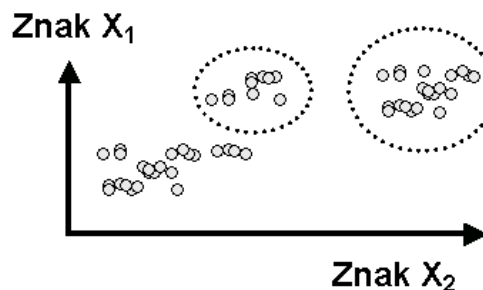
## Pravděpodobnostní vztahy

Anamnéza x Výsledek vyšetření pacienta

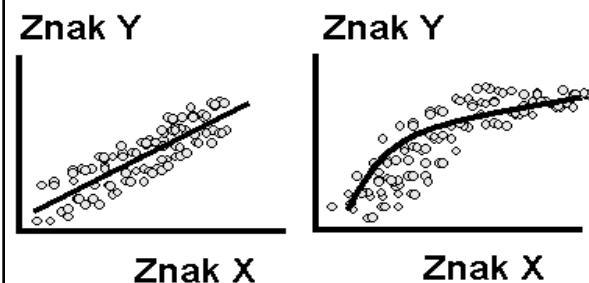
	Karcinom	Benigní léze	Benigní riziková	Zdravá	
Pozitivní anamnéza	2,22	34,44	0,00	63,33	100%
Negativní anamnéza	1,06	28,23	0,96	69,75	100%

$p < 0.05$

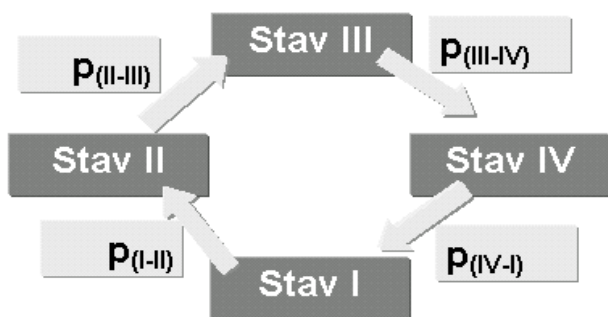
## Vícerozměrná diskriminace



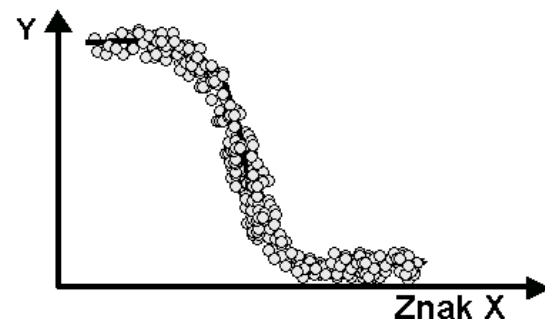
## Funkční vztahy znaků



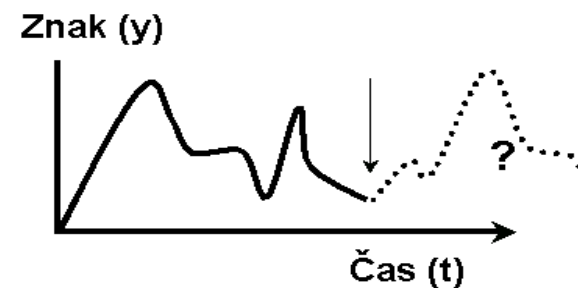
## Markovovy řetězce



## Logistické modely

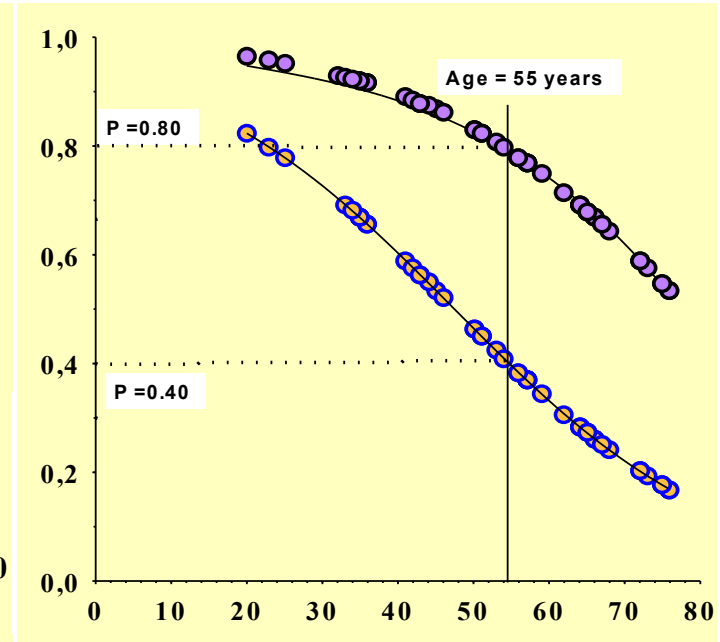
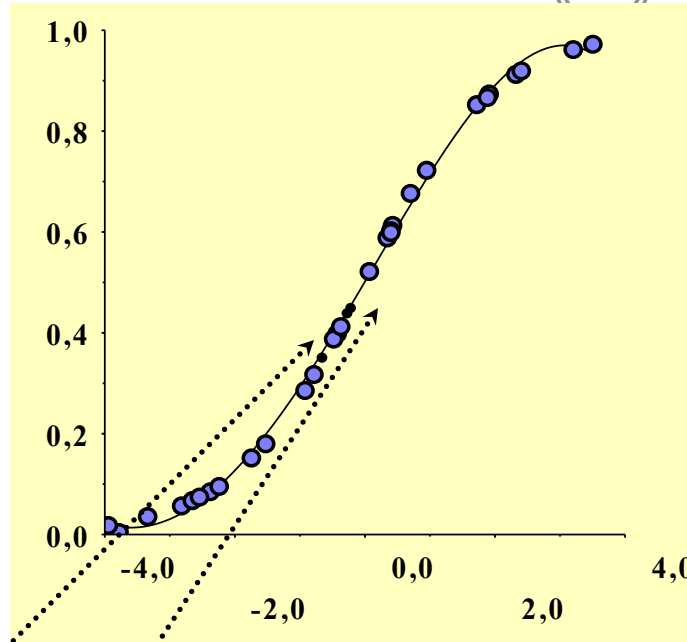


## Chování systému v čase



# Stochastické modelování: predikce neurčitých jevů

Osa Y  
Predikovaná  
pravděpodobnost



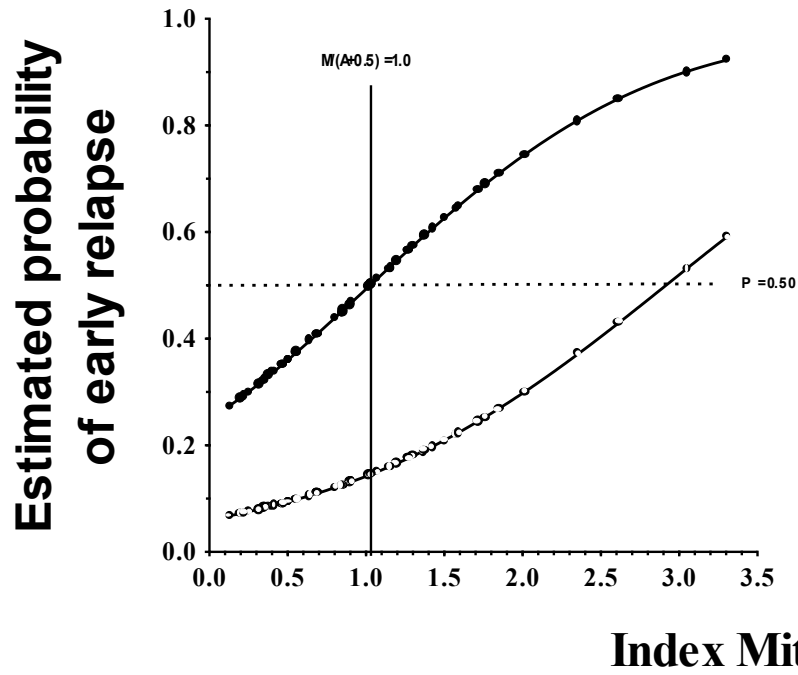
Osa X  
Parametr nebo kombinace parametrů

Data konkrétních pacientů (subjektů)  
k přímému hodnocení

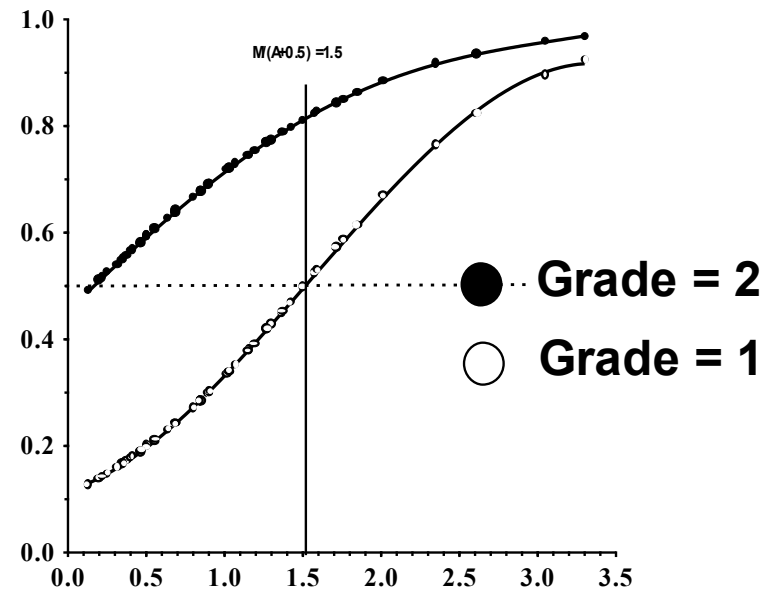
# Stochastické modelování: predikce neurčitých jevů

## Maligní lymfomy: Pravděpodobnost časného relapsu

### Stádium I - II



### Stádium III - IV



Schopnost: vytvářet prakticky využitelné nástroje

# II. Příprava dat



**Klíčový význam korektního uložení získaných dat**  
**Pravidla pro ukládání dat**  
**Čištění dat před analýzou**

# Anotace



- Současná statistická analýza se neobejde bez zpracování dat pomocí statistických software. Předpokladem úspěchu je správné uložení dat ve formě „databázové“ tabulky umožňující jejich zpracování v libovolné aplikaci.
- Neméně důležité je věnovat pozornost čištění dat předcházející vlastní analýze. Každá chyba, která vznikne nebo není nalezeno ve fázi přípravy dat se promítne do všech dalších kroků a může zapříčinit neplatnost výsledků a nutnost opakování analýzy.

# DATA – ukázka uspořádání datového souboru

## Parametry (znaky)



Opakování



Pacient	Clovek	aLeu cell.10 <sup>6</sup> /	aTy% %	aSe% %	aNeu% %	aLy% %	aTy cell.10 <sup>6</sup> /	aSe cell.10 <sup>6</sup> /	aNeu cell.10 <sup>6</sup> /	aLy cell.10 <sup>6</sup> /	aHtc %	aCLsk mV.s.10 <sup>3</sup>	aCLNeus mV.s.10 <sup>3</sup>	aCLOZ mV.s.10 <sup>3</sup>	aCLNeuO mV.s.10 <sup>3</sup>
3	1	4									33	72		32	
4	2	7,6	8	58	66	24	0,6	4,4	5,0	1,8	33	95	19	48	10
8	3	4	3	52	55	40	0,1	2,1	2,2	1,6	22	77	35	33	15
11	4	6,1	5	59	64	35	0,3	3,6	3,9	2,1	33	103	26	49	13
12	5	6,9	3	85	88	9	0,2	5,9	6,1	0,6	37	81	13	45	7
14	6	5,9	15	55	70	19	0,9	3,3	4,1	1,1	32	137	33	61	15
16	7	8	18	75	93	7	1,4	6,0	7,4	0,6	34	151	20	59	8
20	8	9,6	3	72	75	23	0,3	6,9	7,2	2,2	40	77	11	38	5
21	9	6	10	67	77	19	0,6	4,0	4,6	1,1	32	120	26	52	11
22	10	3,3	4	55	59	39	0,1	1,8	2,0	1,3	28	81	42	24	12
37	11	3,8	10	60	70	30	0,4	2,3	2,7	1,1	32	111	42	29	11
38	12	6,4	2	76	78	17	0,1	4,9	5,0	1,1	25	366	73	115	23
39	13	6,8	1	57	58	39	0,1	3,9	3,9	2,7	20	234	59	71	18
49	14	8,5	7	67	74	26	0,6	5,7	6,3	2,2	30	156	25	108	17
51	15	9,3	7	57	64	35	0,7	5,3	6,0	3,3	35	129	21	23	4
52	16	2,2	10	56	66	34	0,2	1,2	1,5	0,7	33	46	30	12	8
55	17	9,9	3	78	81	10	0,3	7,7	8,0	0,1	30	189	24	140	18
56	18	5	2	80	82	13	0,1	4,0	4,1	0,7	26	101	25	54	13
6	1	8,8	11	72	83	12	1,0	6,3	7,3	1,1	44	268	36,6	145	19,9
9	2	9,2	2	66	68	28	0,2	6,1	6,3	2,6	42	168	26,9	76	12,2
13	3	10,0	7	83	90	8	0,7	8,3	9,0	0,8	54	181	20,1	81	9
15	4	9,6	1	75	76	23	0,1	7,2	7,3	2,2	45	343	47	124	16,9
17	5	6,0									45	40		21	
19	6	7,2	2	78	80	18	0,1	5,6	5,8	1,3	44	103	17,8	63	10,9
24	7	8,2	1	72	73	25	0,1	5,9	6,0	2,1	41	209	34,9	57	9,6
26	8	10,3	1	85	86	3	0,1	8,8	8,9	0,3	41	364	41,1	112	12,6
29	9	5,0	1	74	75	21	0,1	3,7	3,8	1,1	39	83	22,1	32	8,5
30	10	11,9	1	51	52	47	0,1	6,1	6,2	5,6	33	83	13,4	52	8,4
31	11	7,2	3	53	56	29	0,2	3,8	4,0	2,1	28	109	27,1	63	15,5
32	12	10,8	36	50	76	8	3,9	5,4	9,3	0,9	27	146	15,7	106	11,4
33	13	11,8	22	54	76	16	2,6	6,4	9,0	1,9	45	246	27,4	63	7
34	14	17,0	1	82	83	16	0,2	13,9	14,1	2,7	34	440	31,2	119	8,4
40	15	10,0	8	72	80	4	0,8	7,2	8,0	0,4	37	176	22,0	52	6,5



# Zásady pro ukládání dat



- Správné a přehledné uložení dat je základem jejich pozdější analýzy
- Je vhodné rozmyslet si předem jak budou data ukládána
- Pro počítačové zpracování dat je nezbytné ukládat data v tabulární formě
- Nejvhodnějším způsobem je uložení dat ve formě databázové tabulky
  - Každý sloupec obsahuje pouze jediný typ dat, identifikovaný hlavičkou sloupce
  - Každý řádek obsahuje minimální jednotku dat (např. pacient, jedna návštěva pacienta apod.)
  - Je nepřipustné kombinovat v jednom sloupci číselné a textové hodnoty
  - Komentáře jsou uloženy v samostatných sloupcích
  - U textových dat nezbytné kontrolovat překlepy v názvech kategorií
  - Specifickým typem dat jsou datумы u nichž je nezbytné kontrolovat, zda jsou datумы uloženy v korektním formátu
- Takto uspořádaná data je v tabulkových nebo databázových programech možné převést na libovolnou výstupní tabulku
- Pro základní uložení a čištění dat menšího rozsahu je možné využít aplikací MS Office

# Ukládání dat v MS Office



- **MS Excel**

- 📄 Kontingenční tabulky – rychlá sumarizace rozsáhlých tabulek
- 📄 Možnost výpočtů a grafových výstupů přímo v aplikaci
- 📄 Visual Basic – složitější aplikace
- Omezení tabulky na 256x65536 buněk (do verze 2003)
- Omezená kontrola chyb při zadávání



- **MS Access**

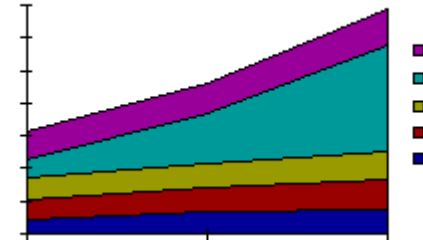
- 📄 Plnohodnotná databáze vhodná pro velké množství dat, řádky omezeny v podstatě jen dostupnou pamětí
- 📄 Kontrola typu dat
- 📄 Relace tabulek – omezení velikosti souboru
- 📄 Visual Basic a formuláře – složitější aplikace
- Omezení tabulky na 255 sloupců
- Výpočty a grafy jsou složitější než v Excelu



# Možnosti MS Excel



- Správa a práce s tabulárními daty
- Řazení dat, výběry z dat, přehledy dat
- Formátování a přehledné zobrazení dat
- Zobrazení dat ve formě grafů
- Různé druhy výpočtů pomocí zabudovaných funkcí
- Tvorba tiskových sestav
- Makra – zautomatizování častých činností
- Tvorba aplikací (Visual Basic for Applications)



Počet z	Délka	Pohlaví
1	2	
2		
3		
4		
5		
6		
7	26	
8	106	
9	121	
10	160	
11	34	
12	45	
13	70	
14	72	
15	87	
16	<b>Celkový součet</b>	
17		

10		
17	10	2
18	12	3
19	5	4
20	8	5
21	4	8
22	7	9
23	9	11
24	suma součtinů řádků	310
25		



# Import a export dat



- **Import dat**
  - Manuální zadávání
  - import – podpora importu ze starších verzí Excelu, textových souborů, databází apod.
  - kopírování přes schránku Windows – vkládání z nejrůznějších aplikací – MS Office, Statistica atd.
  - využití textových souborů jako kompatibilního formátu pro přenos dat mezi různými aplikacemi
- **Export dat**
  - Ukládáním souborů ve formátech podporovaných jinými SW, časté jsou textové soubory, dbf soubory nebo starší verze Excelu
  - Přímé kopírování přes schránku Windows

# Tipy a triky



- **Výběr buněk**

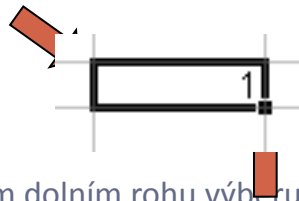
- CTRL+A – výběr celého listu
- CTRL + klepnutí myší do buňky – výběr jednotlivých buněk
- SHIFT + klepnutí myší na jinou buňku – výběr bloku buněk
- SHIFT + šipky – výběr sousedních buněk ve směru šipky
- SHIFT+CTRL+END (HOME) – výběr do konce (začátku) oblasti dat v listu
- SHIFT+CTRL+šipky – výběr souvislého řádku nebo sloupce buněk
- SHIFT + klepnutí na objekty – výběr více objektů

- **Kopírování a vkládání**

- CTRL+C – zkopírování označené oblasti buněk
- CTRL+V – vložení obsahu schránky – oblast buněk, objekt, data z jiné aplikace

- **Myš a okraje buňky**

- Chycení myší za okraj umožňuje přesun buňky nebo bloku buněk



- Při chycení čtverečku v pravém dolním rohu výběru je tažením možno vyplnit více buněk hodnotami původní buňky (ve vzorcích se mění relativní odkazy, je také možné vyplnění hodnotami ze seznamu – např. po sobě jdoucí názvy měsíců).

# Databázová struktura dat v Excelu



Sloupce tabulky = parametry záznamů, hlavička udává obsah sloupce – stejný údaj v celém sloupci

Jednotlivé záznamy  
(taxon, lokalita atd.)

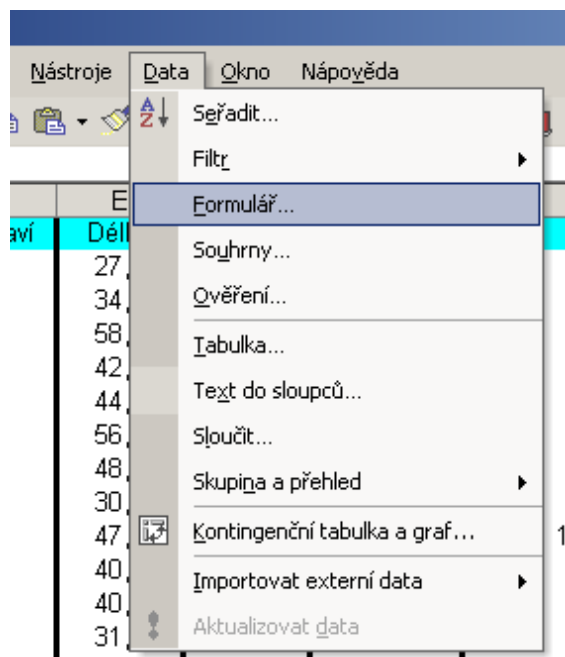


	A	B	C	D	E	F	G	H	I
1	Číslo	Značka	Společ	Pohlaví	Délka	Váha	P. anguillae	P. bini	
2	1	1	1	m	27,5	23,0	2	2	
3	2	2	2	f	34,0	62,5	0	2	
4	3	5	3	f	58,0	230,0	0	0	
5	4	6	4	f	42,0	155,0	0	0	
6	5	7	5	f	44,0	149,8	0	0	
7	6	8	6	f	56,0	323,0	0	1	
8	7	9	7	m	48,5	178,2	0	0	
9	8	10	8	f	30,5	47,7	4	6	
10	9	11	9	f	47,0	175,9	5	14	
11	10	12	10	f	40,0	85,1	5	9	
12	11	14	11	f	40,0	101,0	0	0	
13	12	15	12	f	31,0	84,0	15	9	
14	13	16	13	f?	22,0	9,0	0	0	
15	14	17	14	f	42,0	108,0	1	3	
16	15	18	15	f	44,0	130,0	0	0	
17	16	19	16	f	37,0	85,0	2	5	
18	17	20	17	f	50,0	212,0	1	8	

# Automatický zadávací formulář



- Slouží k usnadnění zadávání dat do databázových tabulek
- Načítá automaticky hlavičky sloupců jako zadávané položky



Nový záznam

Vyhledávání

Názvy sloupců

Obsah dané buňky - editovatelný

# Automatické seznamy



- Vytváří se z hodnot buněk v daném sloupci a umožňují vložit hodnotu výběrem ze seznamu již zadaných hodnot – usnadnění zadávání

Sloupec z něž je seznam vytvořen a pro který platí

Taxon	Abundance	Lokalita	etc.

Buňka, do níž se vloží vybraná hodnota

Glo

- Vymout
- Kopírovat
- Vložit
- Vložit jinak...
- Vložit buňky...
- Odstranit...
- Vymazat obsah
- Vložit komentář
- Formát buněk...
- Vybrat ze seznamu...**
- Přidat kukátko
- Hypertextový odkaz...

Caryophyllaeides fennica ( Schneider, 1902 )

**Piscicola geometra ( Linnaeus, 1761 )**

Acanthocephallus lucii ( Müller, 1776 )

Apophallus mühlungi Jägerskiöld, 1899

Argulus foliaceus ( Linnaeus, 1758 )

Caryophyllaeides fennica ( Schneider, 1902 )

D. cabaleroi

D. crucifer Wagener, 1857

D. fallax Wagener, 1857

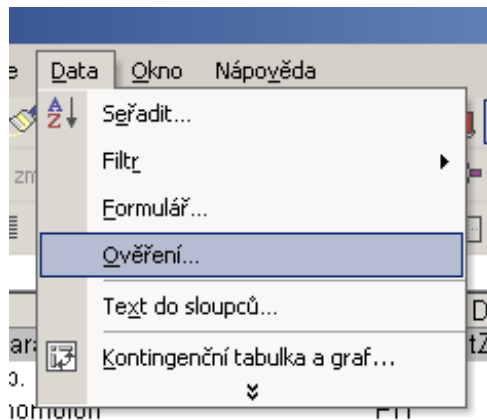
D. nanus Dogiel et Bychowsky, 1934



# Automatická kontrola dat



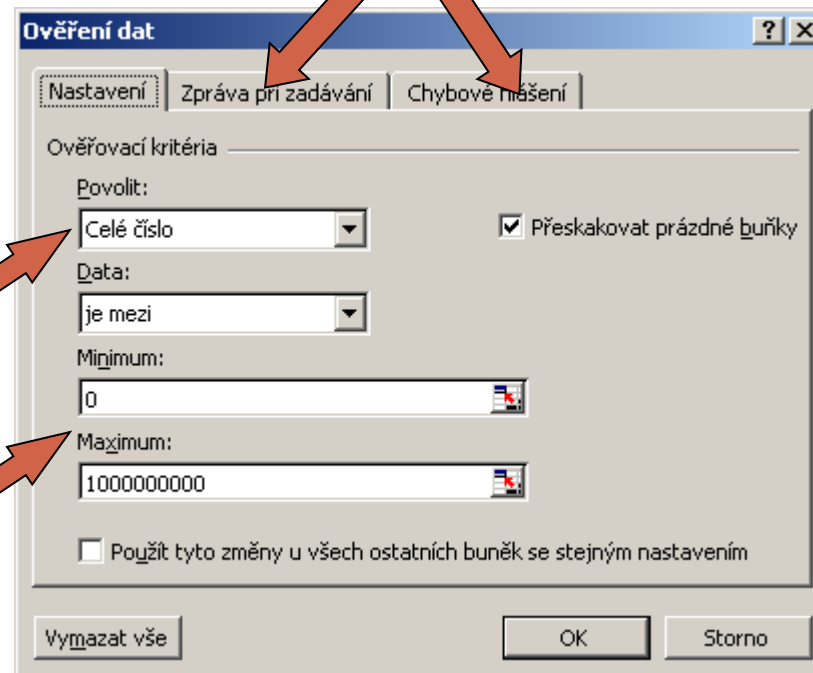
- Umožňuje ověřit typ, rozsah nebo povolit pouze určitý seznam hodnot zadávaných do sloupce databázové tabulky



Co je povoleno – definiční obory čísel, seznamy, vzorce atd.

Rozsahy hodnot, načtení seznamů apod.

komunikace s uživatelem



# Seznamy



- Skupiny hodnot zachovávající logické pořadí, některé jsou zabudované (např. dny v týdnu, měsíce v roce), další je možné uživatelsky vytvořit, slouží pro účely řazení a automatického vyplňování dat

## Existující seznamy

Možnosti

Barva Mezinárodní Ukládání Kontrola chyb Pravopis Bezpečení

Zobrazení Výpočty Úpravy Obecné Převod Vlastní seznamy Graf

Vlastní seznamy: NOVÝ SEZNAM, Po, Út, St, Čt, Pá, So, Ne, Pondělí, Úterý, Středa, Čtvrtek, Pátek, Sobot, I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII, leden, únor, březen, duben, květen, červen, spring 97, autumn 97, spring 98

Položky seznamu: Přidat, Odstranit

Položky seznamu můžete oddělit stisknutím klávesy Enter.

Importovat seznam z buněk: Importovat

OK Storno

Výběr buněk pro nový seznam

Načtení nového seznamu

# Řazení dat

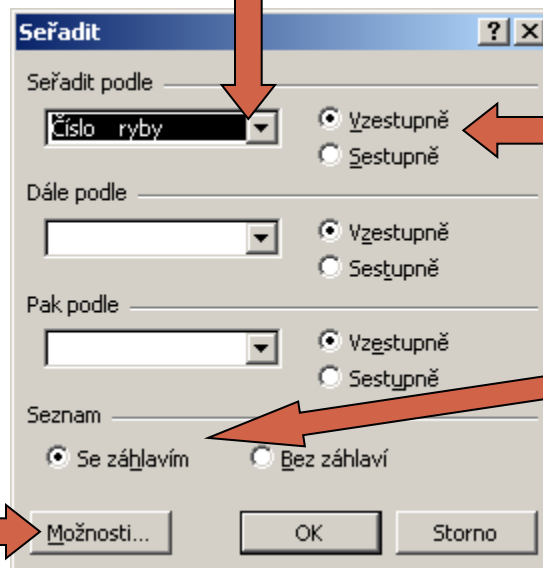
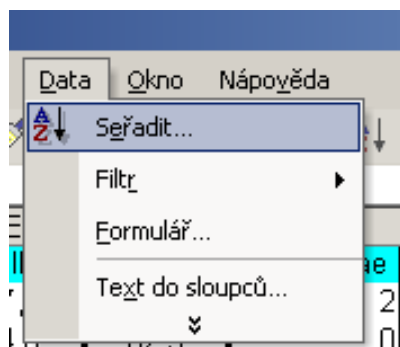


- Řazení dat je nejjednodušším způsobem jejich zpřehlednění, užitečným hlavně u menších/výsledkových tabulek



Zkontrolujte, zda seřazení nezničí vazby mezi buňkami = kontrola oblasti, kterou řadíte.

Podle čeho řadit



Směr řazení – vzestupně, sestupně

Využít první řádek oblasti jako záhlaví

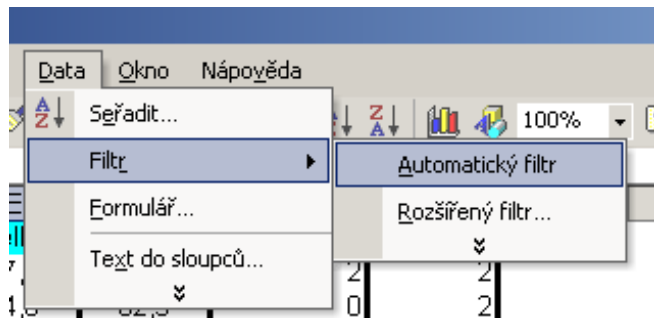
Další možnosti – řazení řádků, řazení podle seznamu

# Automatický filtr



- Pomocí automatického filtru je snadné vybírat úseky dat pro další zpracování na základě hodnot ve sloupcích databázové tabulky, výběr je možný i podle více sloupců (např. určitá skupina pacientů)
- Funkce automaticky rozezná hlavičky sloupců v souvislé oblasti buněk
- U sloupců použitých pro filtraci jsou rozbalovací seznamy zbarveny modře
- **Výhodné pro čištění dat (vyhledávání překlepů, kombinace textu a čísel)**

Výběr hodnot pro filtraci



Rozbalení seznamu hodnot nalezených ve sloupci

	A	B	C	D	E
1	Číslo	Značka	Společ	Pohlav	Délka
2	1	1	1	(Vše)	27,5
3	2	2	2	(Prvních 10...)	34,0
4	3	5	3	(Vlastní...)	58,0
5	4	6	4	f?	42,0
6	5	7	5	m	44,0
7	6	8	6	f	56,0
8	7	9	7	m	48,5

# III. Vizualizace dat



## Typy grafické vizualizace Rizika desinterpretace grafického zobrazení dat

# Anotace

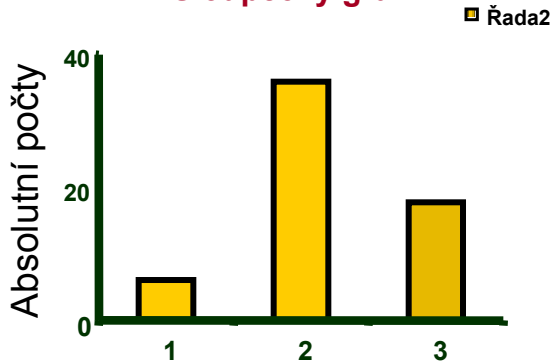


- Prvním krokem v analýze dat je jejich vizualizace. Různé typy dat nám umožňují získání představy o rozložení dat, zastoupení kategorií i vztazích proměnných navzájem. Prostřednictvím vizualizace získáváme vhled do dat a začínáme vytvářet hypotézy o zákonitostech panujících mezi proměnnými v hodnoceném souboru dat.

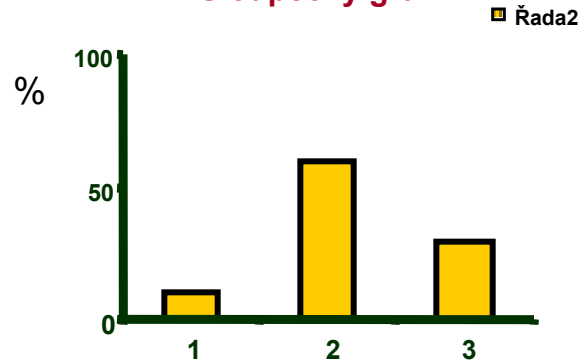
# Grafická prezentace dat - umění komunikace

## 1. Výskyt kategorií (1, 2, 3)

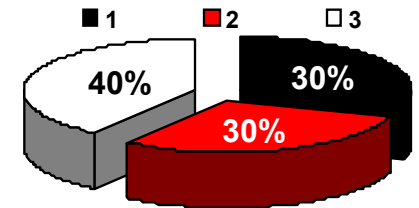
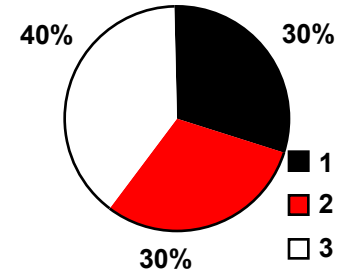
Sloupcový graf



Sloupcový graf

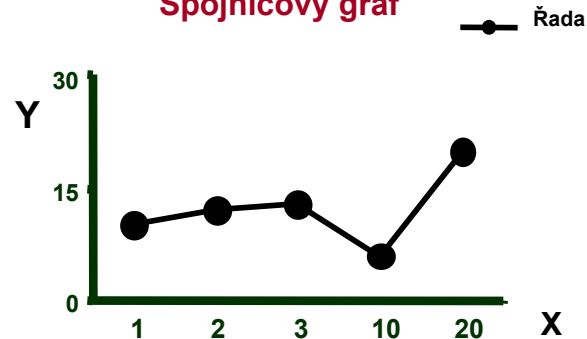


Koláčový (výsečový) graf

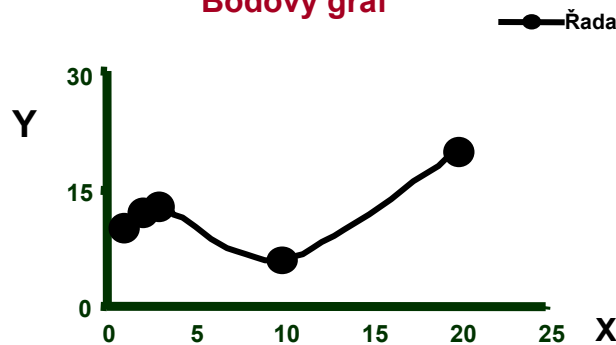


## 2. Vývoj hodnot (v čase) Y vs. X (t)

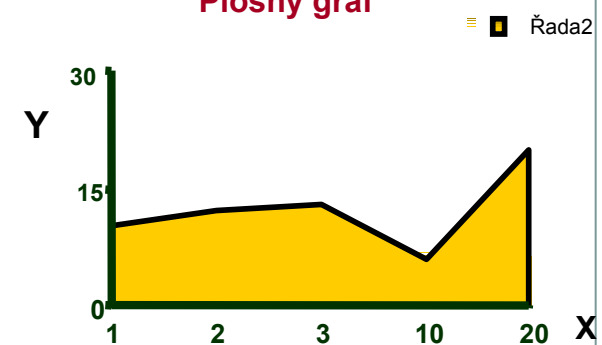
Spojnicový graf



Bodový graf



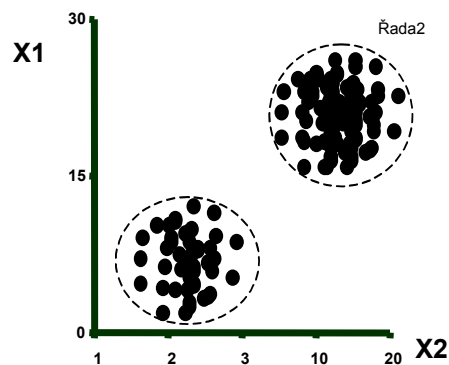
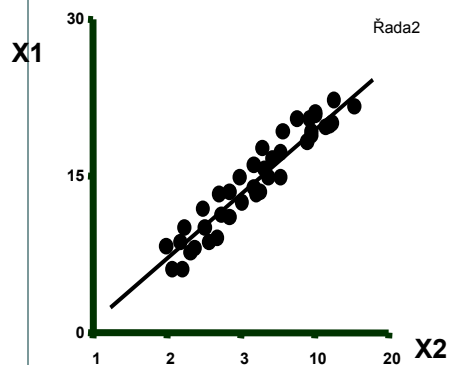
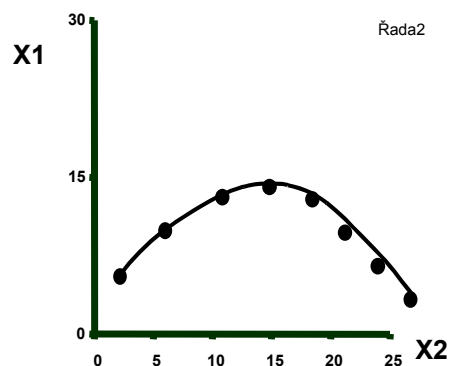
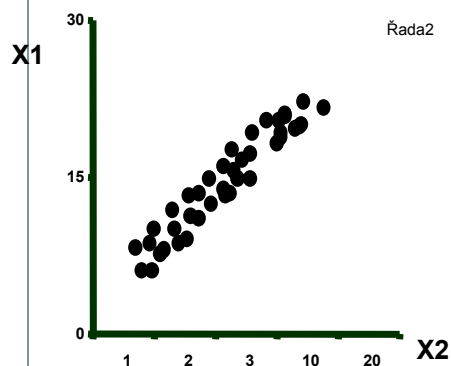
Plošný graf



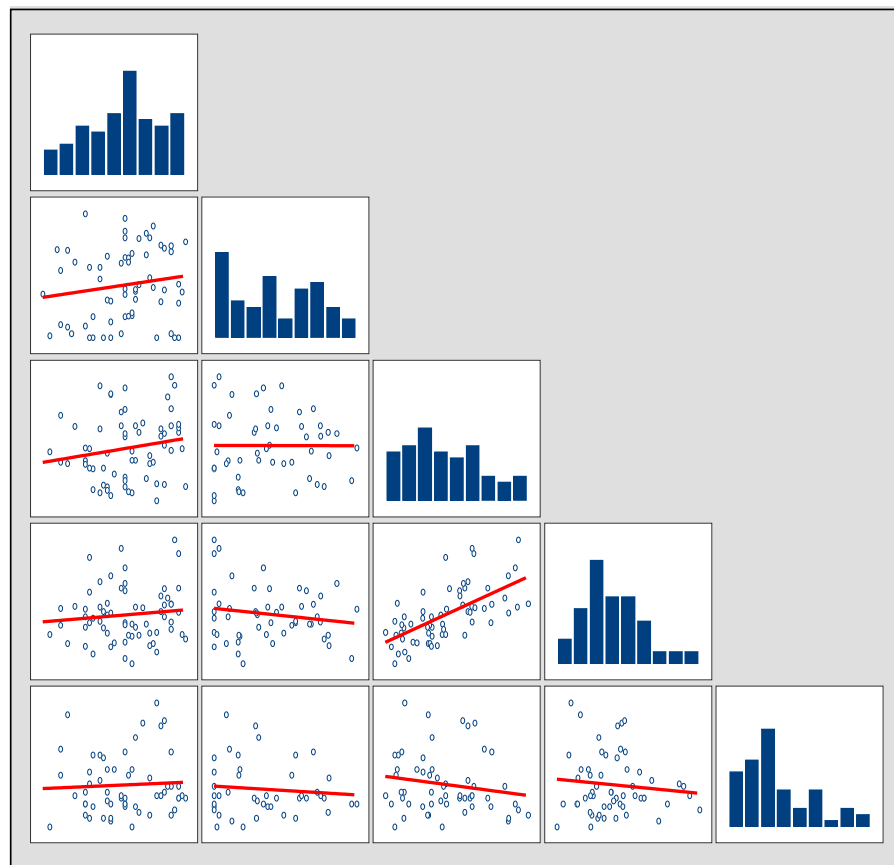
# Grafická prezentace dat - umění komunikace

## 3. Vztahy mezi proměnnými - korelace

Bodový - korelační diagram



Bodový - korelační diagram

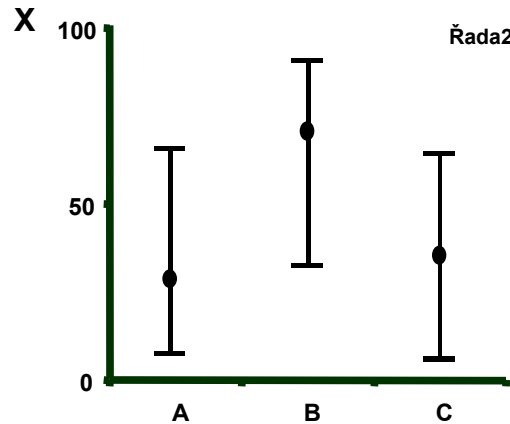
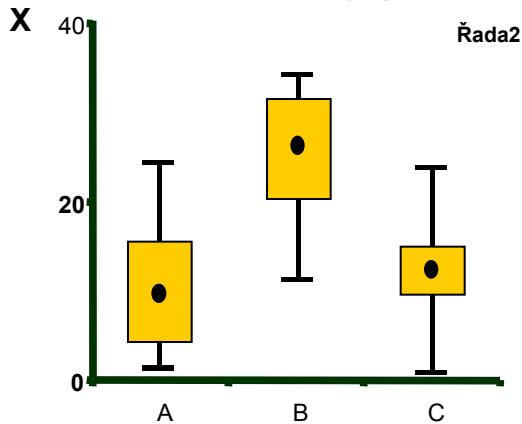




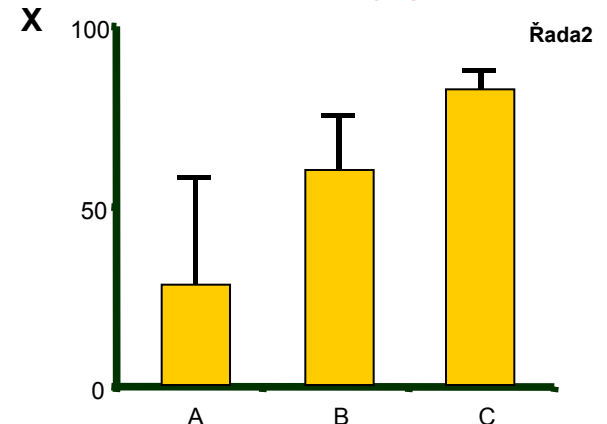
# Grafická prezentace dat - umění komunikace

## 4. Kvantitativní hodnoty parametru(ů) - $X$ - v rámci kategorií A, B, C

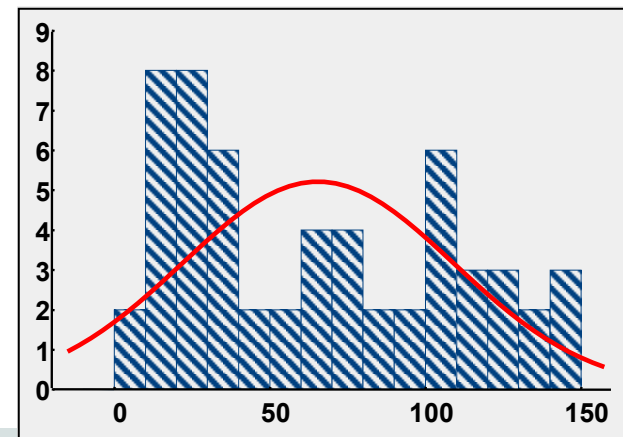
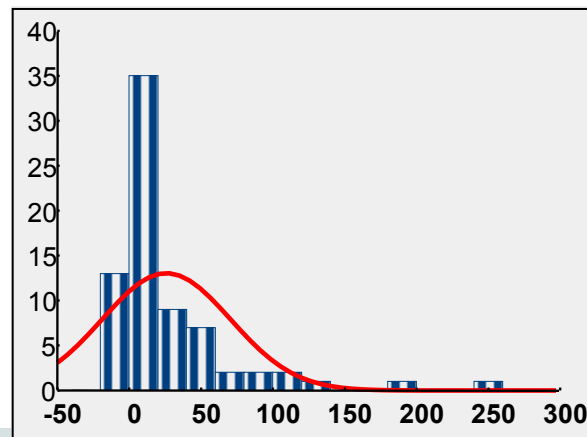
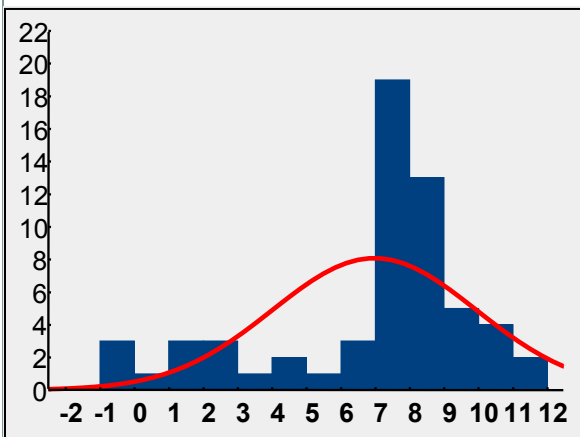
Krabicový graf



Sloupcový graf

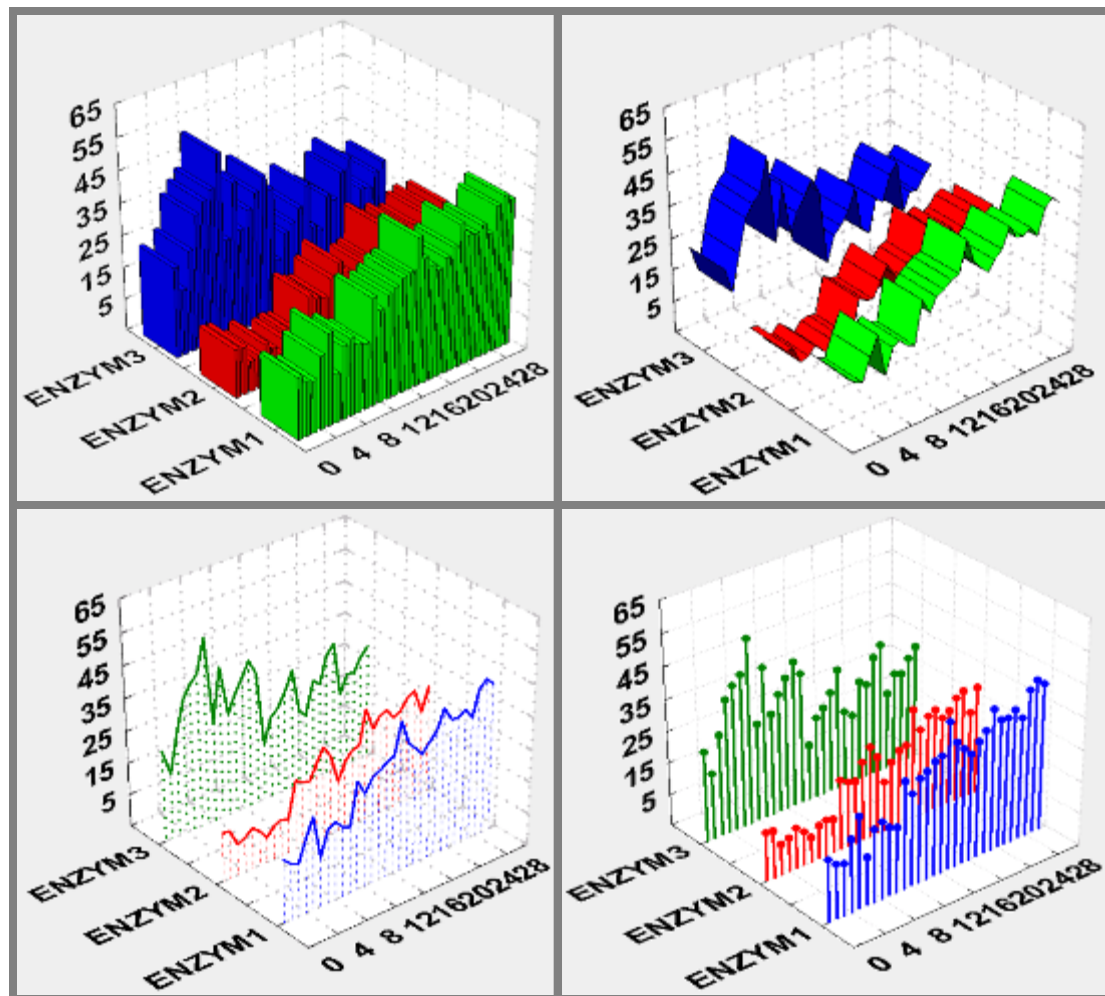
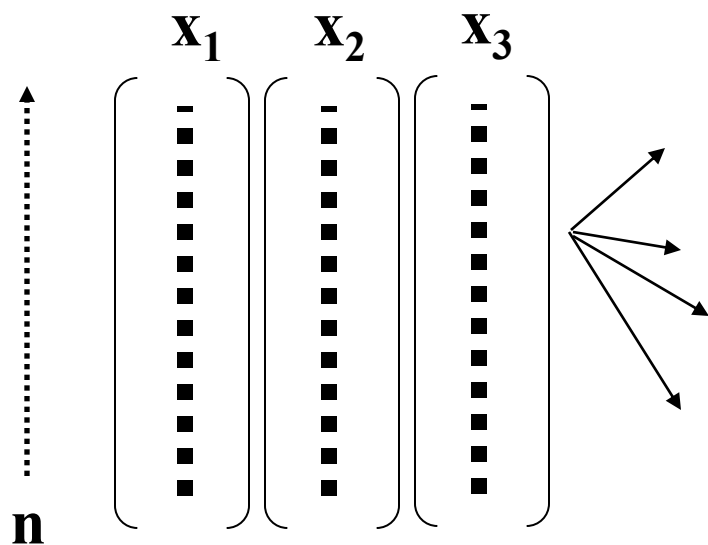


## 5. Histogram



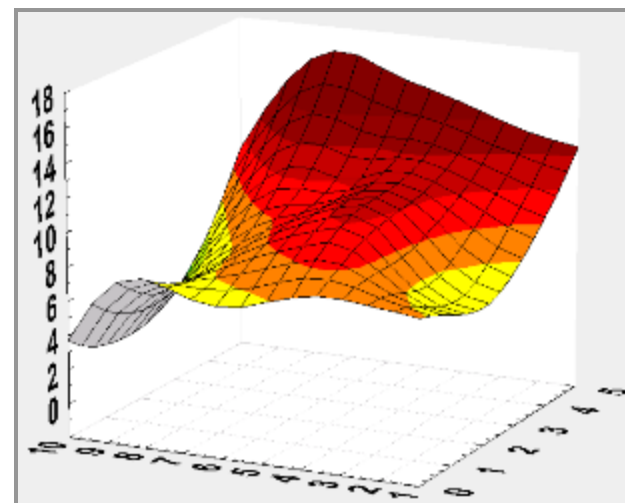
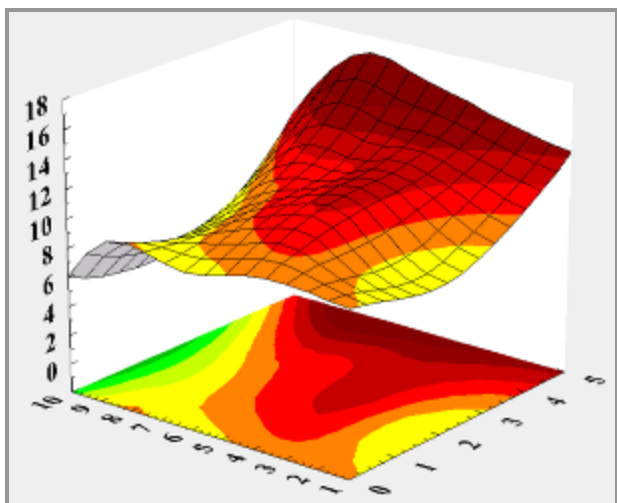
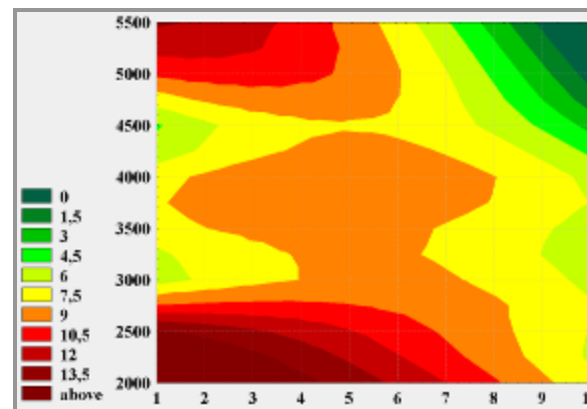
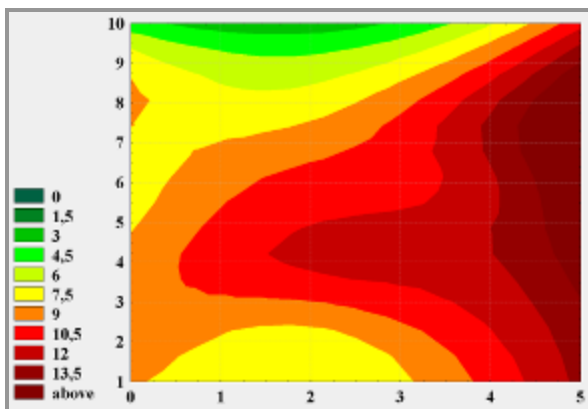
# Grafická prezentace dat - umění komunikace

## 6. Zviditelnění primárních dat



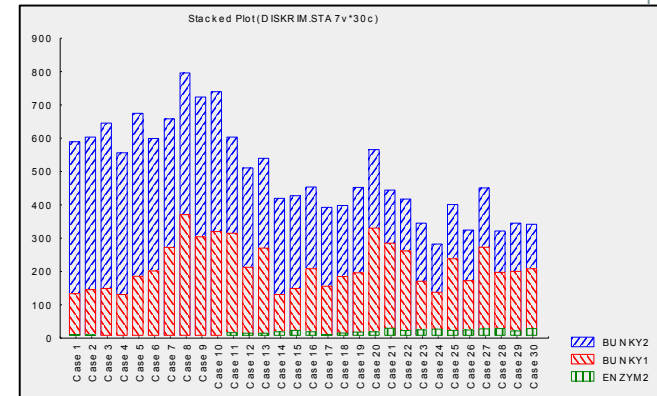
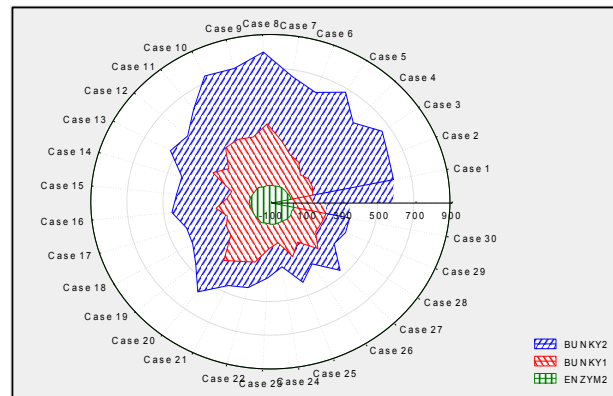
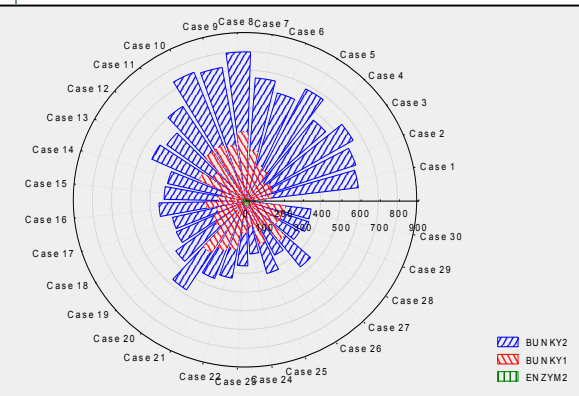
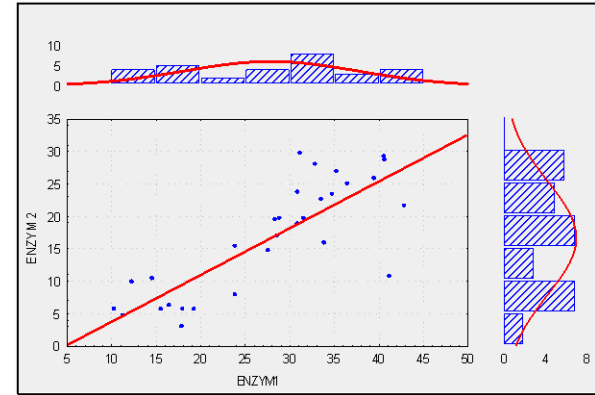
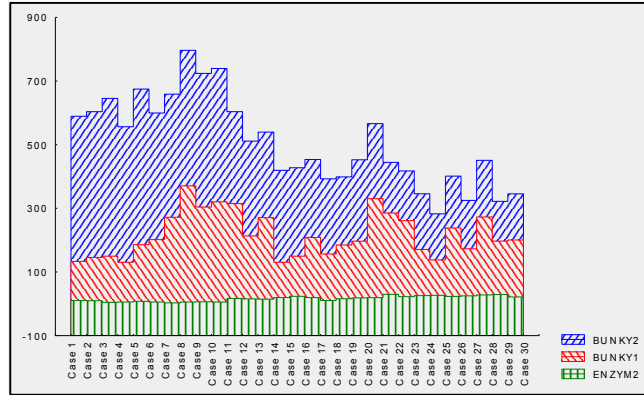
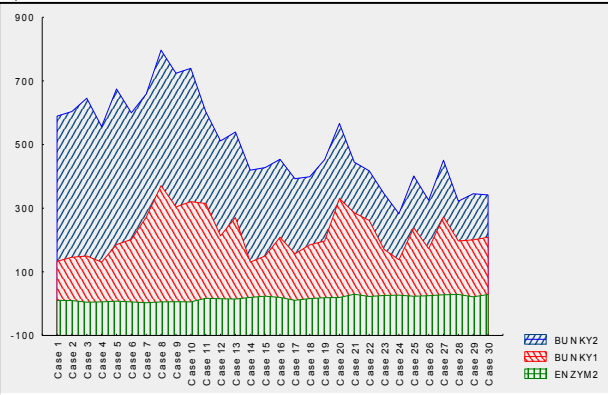
# Grafická prezentace dat - umění komunikace

## 7. Vztahy mezi proměnnými - interakce dvou parametrů, reakční plochy



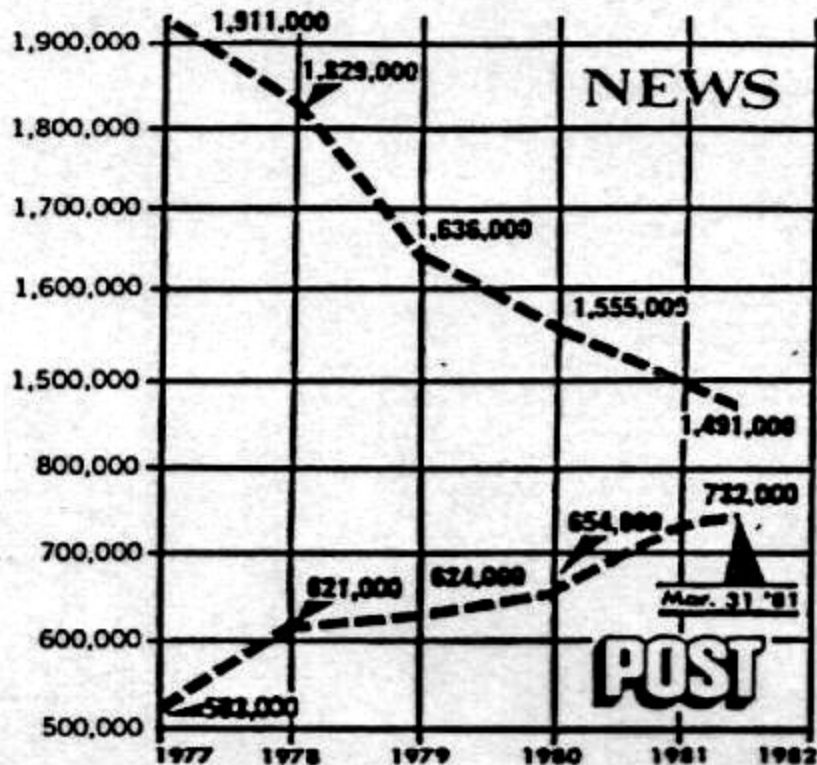
# Grafická prezentace dat - umění komunikace

## 8. Grafické zviditelnění má nekonečně mnoho možností

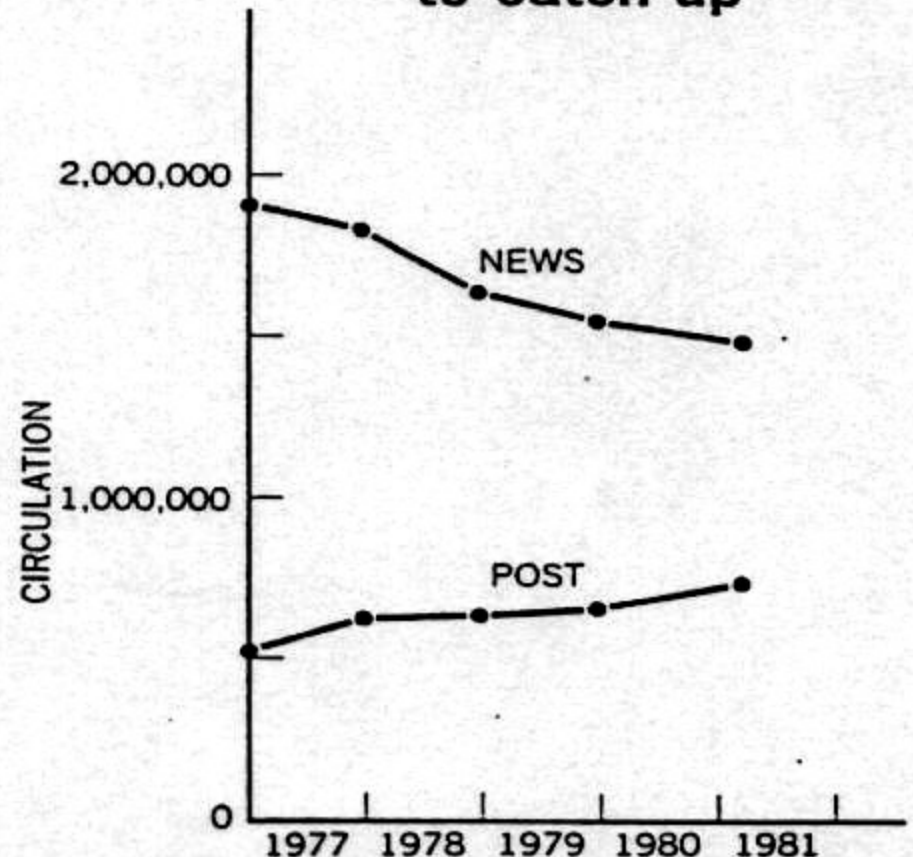


# Nesprávné užití grafů: problém rozsahu číselné osy

## The soaraway Post — the daily paper New Yorkers trust

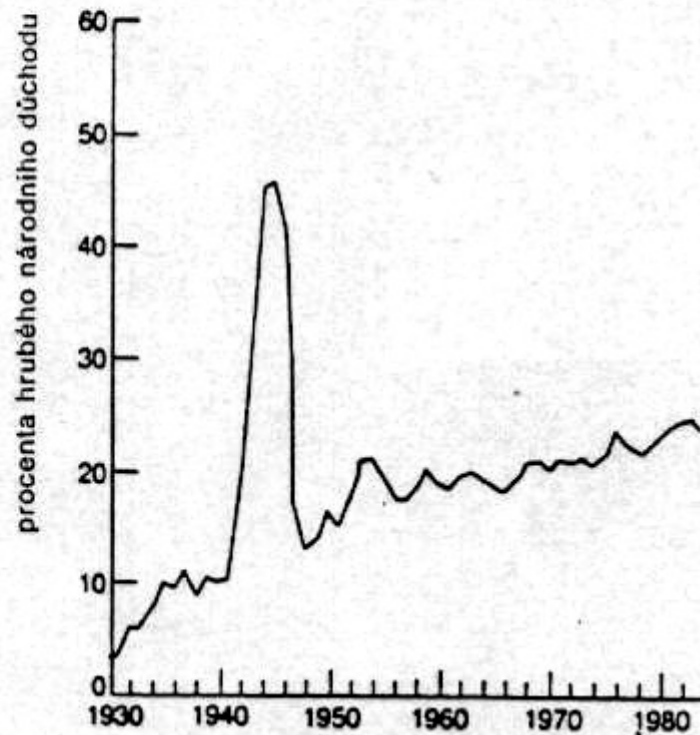
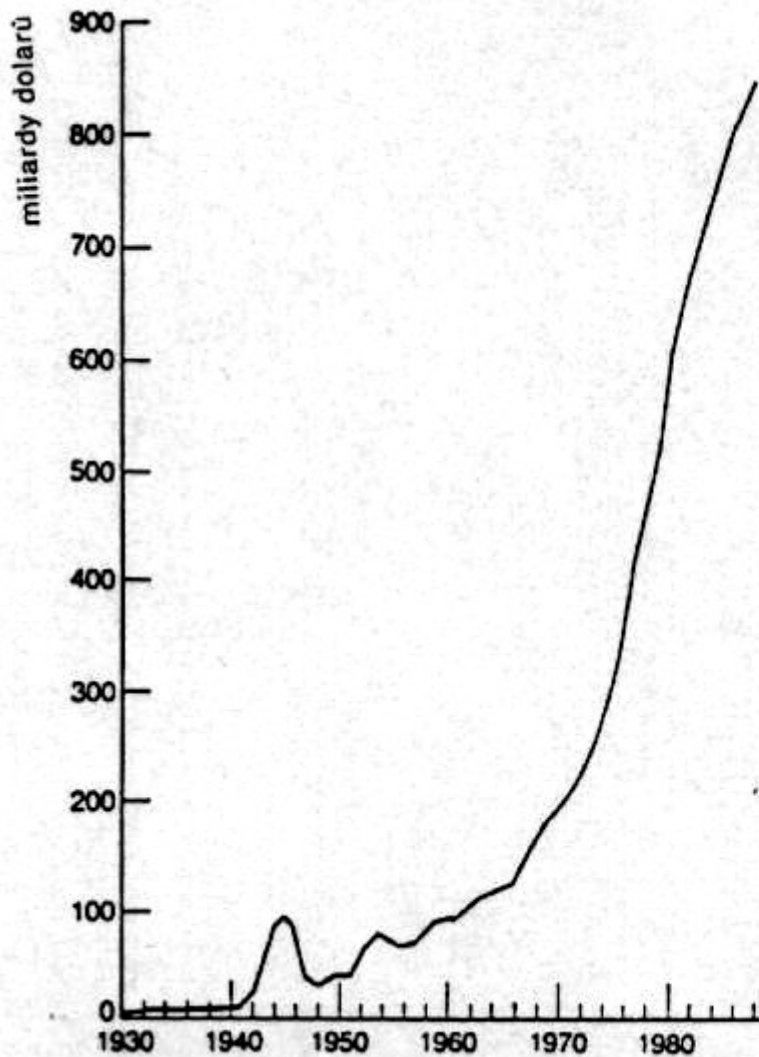


## The Post struggles to catch up

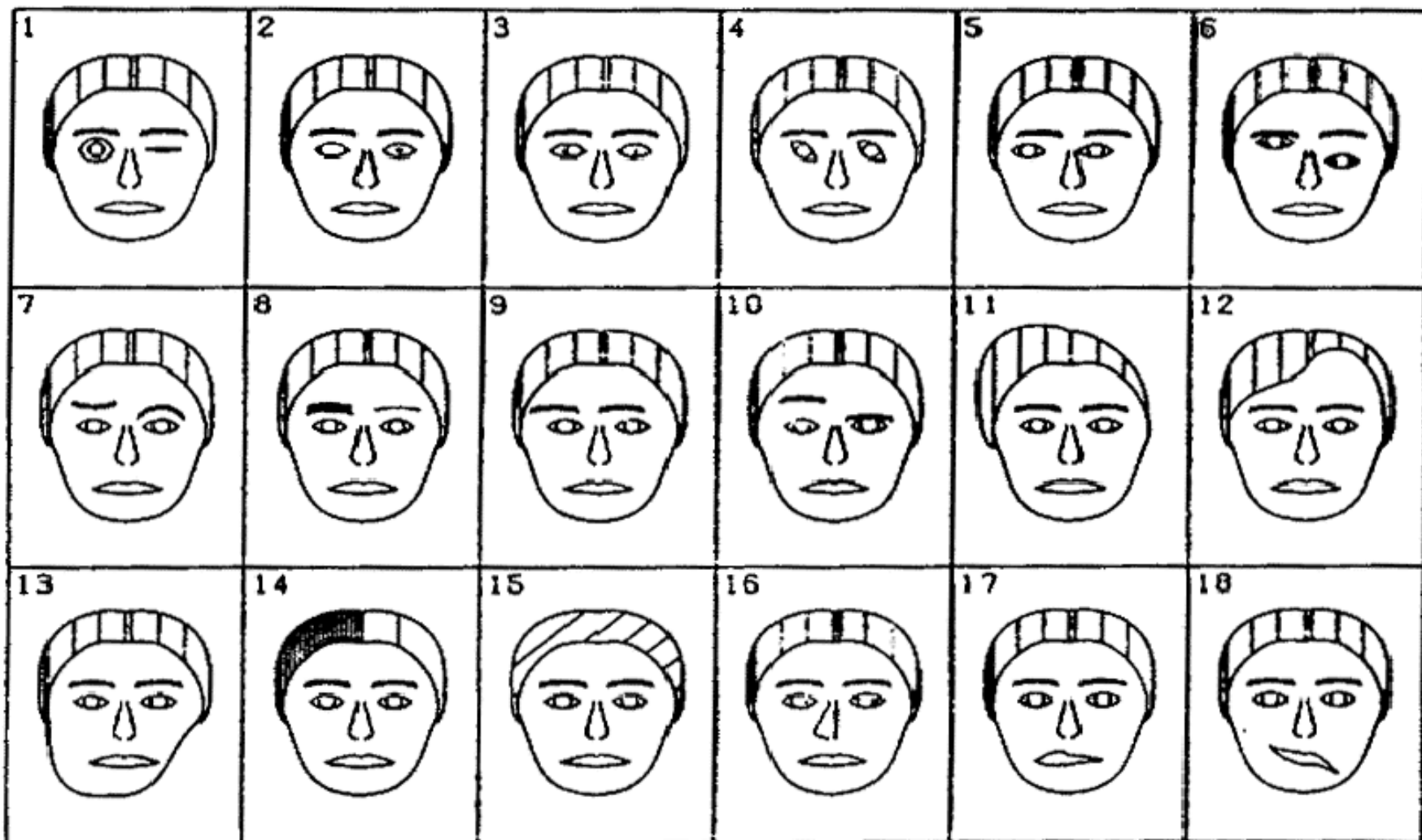




# Nesprávné užití grafů: problém standardizace hodnot



# Grafy zaměřené na vícerozměrné soubory dokáží zviditelnit i veliké soubory dat



# IV. Teoretické pozadí statistické analýzy



Jak vznikají informace  
Rozložení dat



# Anotace



- Základním principem statistiky je pravděpodobnost výskytu nějaké události. Prostřednictvím vzorkování se snažíme odhadnout skutečnou pravděpodobnost událostí. Klíčovou otázkou je velikost vzorku, čím větší vzorek, tím větší šance na projevení se skutečné pravděpodobnosti výskytu jevu.

# JAK vznikají informace ? základní pojmy

Skutečnost

**Náhoda**

(vybere jednu z možností pokusu)

**Jev**

*podmnožina všech možných výsledků pokusu/děje, o které lze říct, zda nastala nebo ne*

Pozorovatel

Rozliší, co nastalo

- a) **podle možností**
- b) **podle toho, jak potřebuje**

**Jevové pole**

*třída všech jevů, které jsme se rozhodli nebo jsme schopni sledovat*

**Skutečnost + Jevové pole = Měřitelný prostor**

**Experimentální jednotka** - *objekt, na kterém se provádí šetření*

**Populace** - *soubor experimentálních jednotek*    **Znak** - *vlastnost sledovaná na objektu*

**Sledovaná veličina** - *číselná hodnota vyjadřující výsledek náhodného experimentu*

Znak se stává náhodnou veličinou, pokud se jeho hodnota zjišťuje vylosováním objektu ze základního souboru

Výběr - výběrová populace - cílová populace

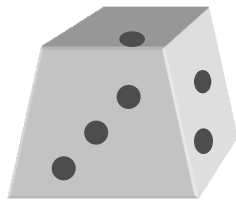
**Náhodný výběr**

**Reprezentativnost**

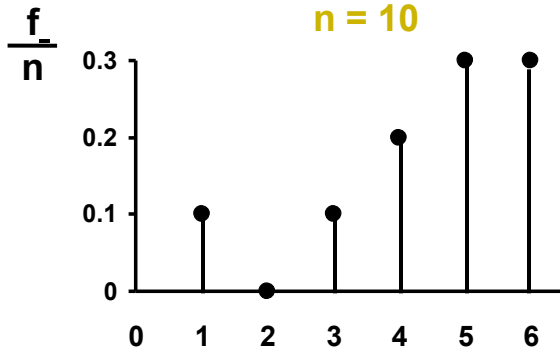
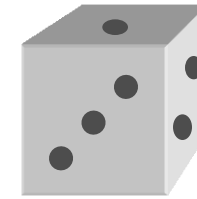
# JAK vznikají informace ?

„Empirical approach“

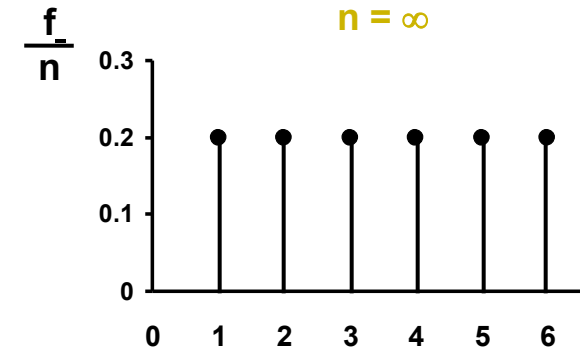
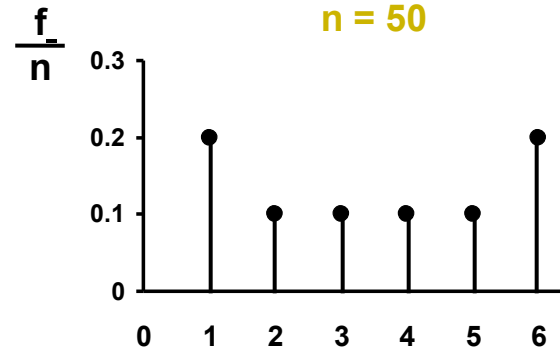
„Classical approach“



Empirický postup



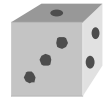
možné jevy: čísla 1 – 6



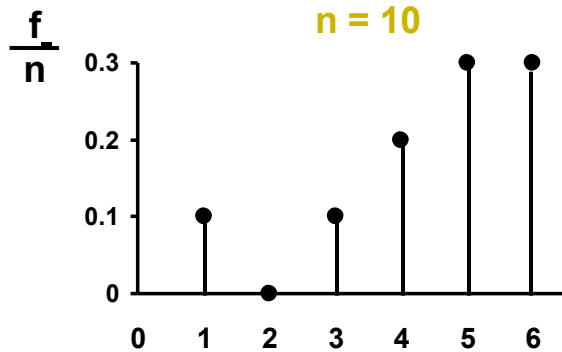
$n$  – počet hodů (opakování)

**U složitých stochastických systémů se pravda získá až po odvedení značného množství experimentální práce: musíme dát systému šanci se projevit**

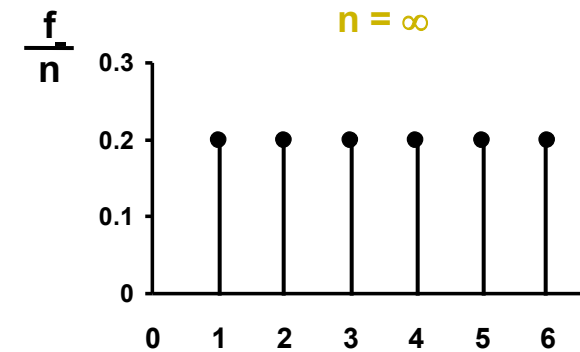
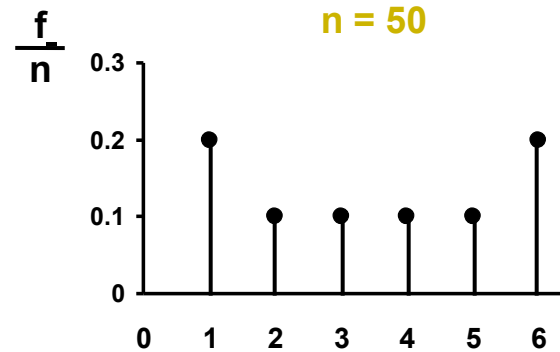
# JAK vznikají informace ?



Empirický postup



možné jevy: čísla 1 – 6



$n$  – počet hodů (opakování)



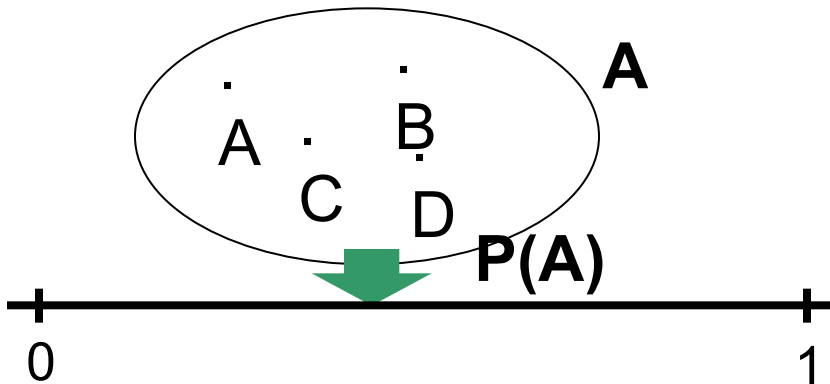
Při realizaci náhodného experimentu roste se zvyšujícím se počtem opakování pravdivá znalost systému (výsledky se stávají stabilnější) .... diskutabilní je ale ovšem míra zobecnění konkrétního experimentu

# Empirický zákon velkých čísel



Při opětovné nezávislé realizaci téhož náhodného experimentu se podíl výskytů sledovaného jevu mezi všemi dosud provedenými realizacemi zpravidla ustaluje kolem konstanty.

Pravděpodobnost je libovolná reálná funkce definovaná na jevovém poli  $A$ , která každému jevu  $A$  přiřadí nezáporné reálné číslo  $P(A)$  z intervalu  $0 - 1$ .



Z praktického hlediska je  
pravděpodobnost  
**idealizovaná relativní četnost**

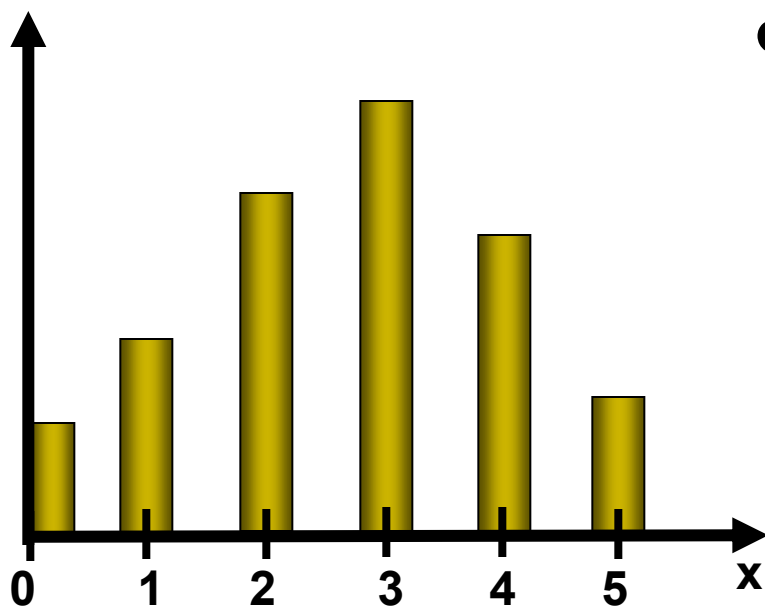
- $P(A) = 1$  ..... jev jistý
- $P(A) = 0$  ..... jev nemožný
- $P(A \cap B) = P(A) \cdot P(B)$  ..... nezávislé jevy
- $P(A \cap B) = P(A) \cdot P(B/A)$  ..... závislé jevy
- $P(A/B) = P(A \cap B) / P(B)$  ..... podmíněná pravděpodobnost

# Pravděpodobnost výskytu jevu – rozložení dat



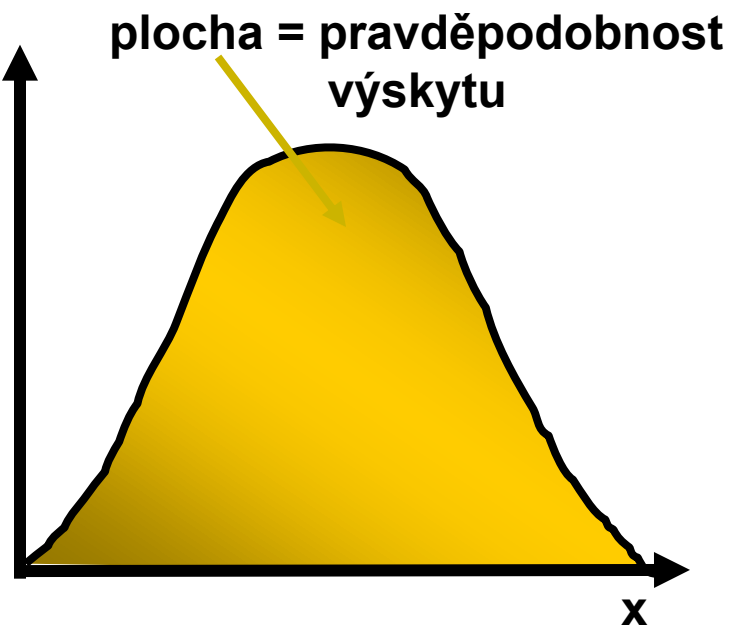
- ✦ existuje pravděpodobnost výskytu jevů (nedeterministické závěry)
- ✦ „vše je možné“: pouze jev s pravděpodobností 0 nikdy nenastane
- ✦ pravděpodobnost lze zkoumat retrospektivně i prospektivně

pravděpodobnost  
výskytu



počet chlapců v rodině s X dětmi

$\varphi(x)$



výška postavy