

# I. 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



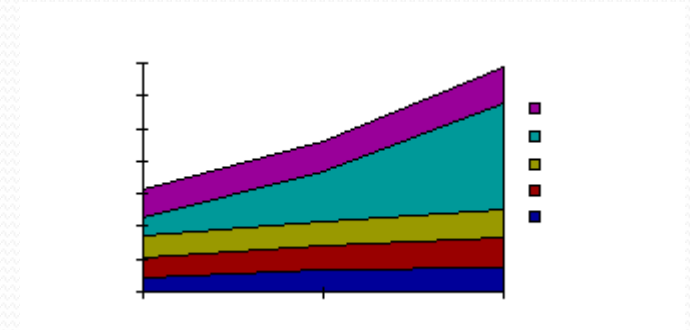
Pačet	Uvek	a1u cel.10%	aly% %	a2% %	a3u% %	aly% %	aly cel.10%	a2 cel.10%	a3u cel.10%	aly cel.10%	a4c %	a1sk m/s10 <sup>3</sup>	a1Nus m/s10 <sup>3</sup>	a1OZ m/s10 <sup>3</sup>	a1NaO m/s10 <sup>3</sup>
3	1	4									33	12		32	
4	2	7,6	8	58	66	24	06	44	50	1,8	33	95	19	48	10
8	3	4	3	52	55	40	01	21	22	1,6	22	77	35	33	15
11	4	61	5	59	64	35	03	36	39	21	33	103	26	49	13
12	5	69	3	85	88	9	02	59	61	06	37	81	13	45	7
14	6	59	15	55	70	19	09	33	41	1,1	32	137	33	61	15
16	7	8	18	75	93	7	14	60	74	06	34	151	20	59	8
20	8	96	3	72	75	23	03	69	72	22	40	77	11	38	5
21	9	6	10	67	77	19	06	40	46	1,1	32	120	26	52	11
22	10	33	4	55	59	39	01	18	20	1,3	28	81	42	24	12
37	11	38	10	60	70	30	04	23	27	1,1	32	111	42	29	11
38	12	64	2	76	78	17	01	49	50	1,1	25	356	73	115	23
39	13	68	1	57	58	39	01	39	39	27	20	234	59	71	18
49	14	85	7	67	74	26	06	57	63	22	30	156	25	108	17
51	15	93	7	57	64	35	07	53	60	33	35	129	21	23	4
52	16	22	10	55	66	34	02	12	15	07	33	46	30	12	8
55	17	99	3	78	81	10	03	77	80	01	30	189	24	140	18
56	18	5	2	80	82	13	01	40	41	07	26	101	25	54	13
6	1	88	11	72	83	12	10	63	73	1,1	44	268	366	145	199
9	2	92	2	66	68	28	02	61	63	26	42	168	269	76	122
13	3	100	7	83	90	8	07	83	90	08	54	181	201	81	9
15	4	96	1	75	76	23	01	72	73	22	45	343	47	124	169
17	5	60									45	40		21	
19	6	72	2	78	80	18	01	56	58	1,3	44	103	178	63	109
24	7	82	1	72	73	25	01	59	60	21	41	209	349	57	96
26	8	103	1	85	86	3	01	88	89	03	41	364	41,1	112	126
29	9	50	1	74	75	21	01	37	38	1,1	39	83	221	32	85
30	10	11,9	1	51	52	47	01	61	62	56	33	83	134	52	84
31	11	72	3	53	55	29	02	38	40	21	28	109	27,1	63	155
32	12	108	36	50	76	8	39	54	93	09	27	146	157	106	114
33	13	11,8	22	54	76	16	26	64	90	1,9	45	246	274	63	7
34	14	17,0	1	82	83	16	02	139	141	27	34	440	312	119	84
40	15	100	8	72	80	4	08	72	80	04	37	176	220	52	65

# 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 nepřípustné 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

# 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)



P. bini	2		
Počet z Délka	Pohlaví		
Číslo ryby2	Číslo rvt	Váha	?
1			
2			
3	26		
4	106		
5	121		
6	160		
7	34		
8	45		
9	70		
10	72		
11	87		
12			
13			
14			
15			
16			
17			

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



# 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+HOME – přesunutí na levý horní roh tabulky
- CTRL+END – přesunutí na pravý dolní roh tabulky
- CTRL+A – výběr celého listu
- CTRL + klepnutí myši do buňky – výběr jednotlivých buněk
- SHIFT + klepnutí myši 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

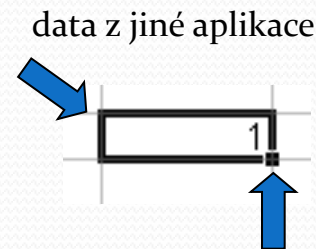
## • Kopírování a vkládání

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

## • Myš a okraje buňky

- Chycení myši 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).

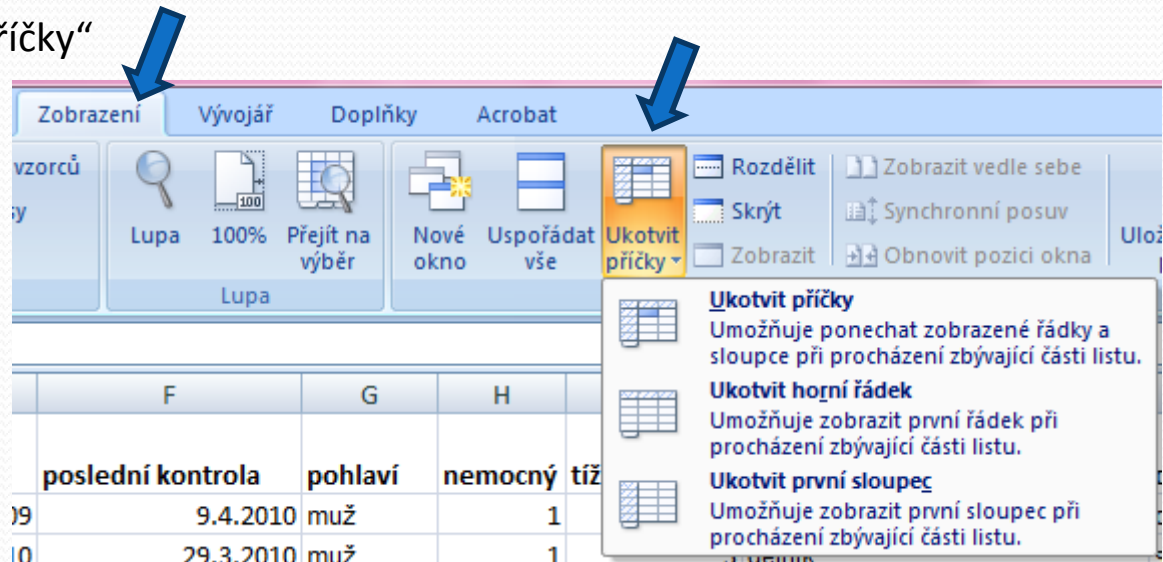


# Ukotvení příček

- Umožňuje ukotvení libovolných řádků a sloupců pro pohodlné vkládání a prohlížení dat v tabulce
- Umožňuje číst řádky/sloupce ze začátku tabulky i po přesunutí se dále
- Záložka „Zobrazení“ → „Ukotvit příčky“

- Nabízené možnosti:

- Ukotvit příčky – ukotví řádky nad označenou buňkou a sloupce vlevo od označené buňky
- Ukotvit horní řádek
- Ukotvit první sloupec
- ! Ukotvení zrušíme opětovným odkliknutím možnosti ukotvení příček





# Databázová struktura dat v Excelu

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

Jednotlivé záznamy  
(taxon, lokalita,  
pacient 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 I.

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

Microsoft Office 2003 a starší

The image shows a screenshot of the Microsoft Office 2003 interface. On the left, the 'Data' menu is open, and the 'Formulář...' option is highlighted. A blue arrow points from this menu item to the 'data' dialog box on the right. The dialog box contains several input fields for data entry, with blue arrows pointing to specific elements:

- Nový záznam**: Points to the 'Nový' button.
- Názvy sloupců**: Points to the input field for 'Číslo ryby:'.
- Obsah dané buňky - editovatelný**: Points to the input field for 'Značka ryby:'.
- Vyhledávání**: Points to the 'Kritéria' button.

The dialog box also includes buttons for 'Odstranit', 'Obnovit', 'Předchozí', 'Další', and 'Zavřít'. The 'data' dialog box is titled 'data' and shows '1 z 19' records.

# Automatický zadávací formulář II.

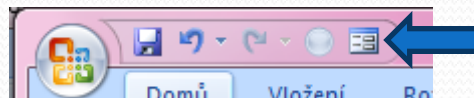
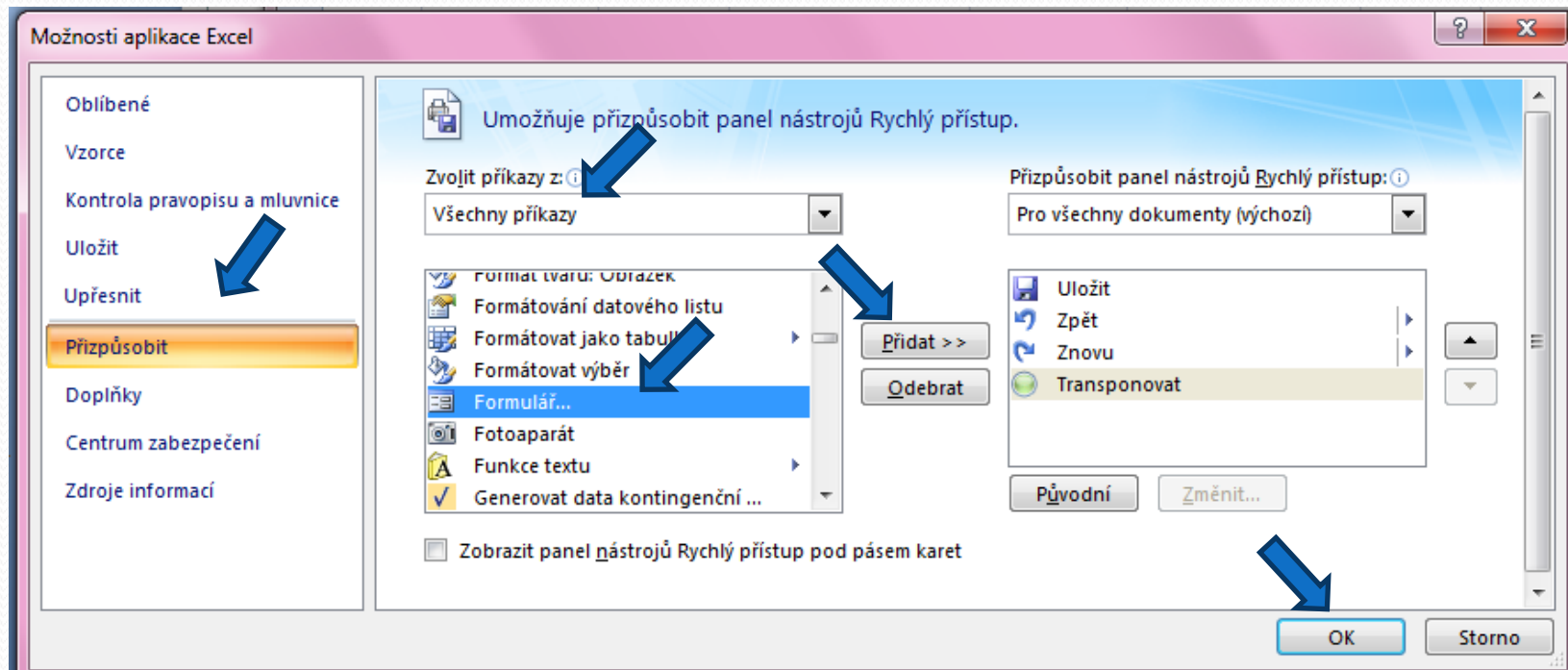
Microsoft Office 2007

• Aplikaci automaticky zadávaného formuláře je nutné aktivovat

• „Tlačítko Office“



→ „Možnosti aplikace Excel“



Vytvořil Institut biostatistiky a analýz,  
Masarykova univerzita

# 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

laon	urbae	Locata	ec

Buňka, do níž se vloží vybraná hodnota

1909  
Linnaeus, 1758

1857  
aceum ( Rudolp  
jicii ( Müller, 177  
5  
ion  
Bychowsky, 193  
nnica ( Schneid  
i, 1937  
1857  
ion  
nnica ( Schneid  
Linnaeus, 176

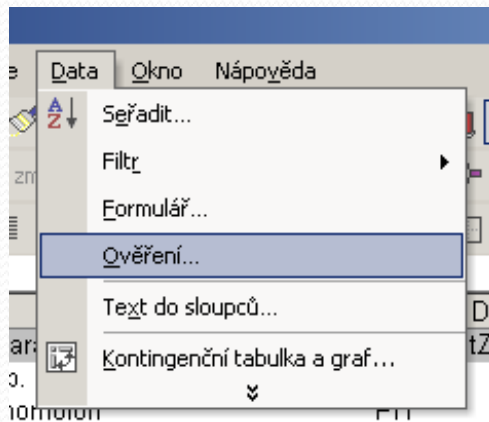
Glo

- Vyjmout
- 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

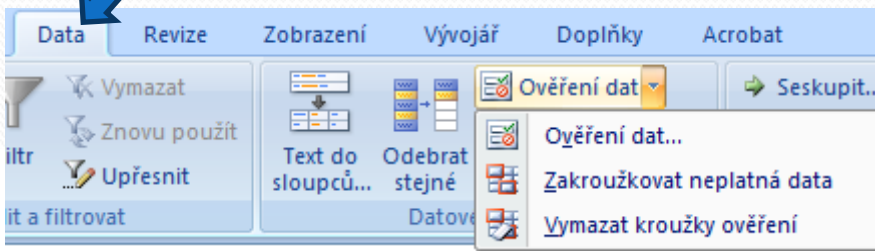
- Umožňuje ověřit typ, rozsah nebo povolit pouze určitý seznam hodnot zadávaných do sloupce databázové tabulky



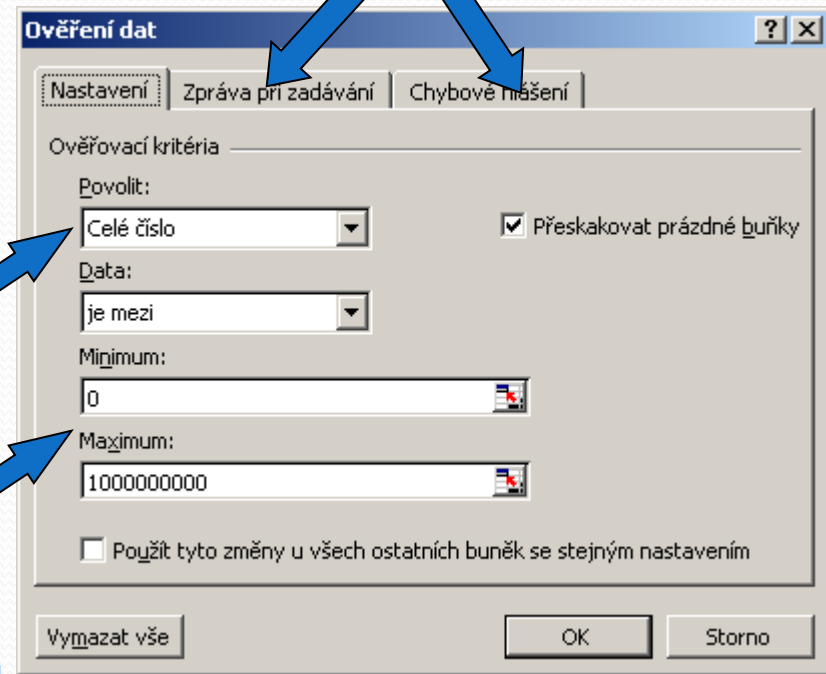
Co je povoleno – definiční obory čísel, seznamy, vzorce atd.

Microsoft Office 2007

Rozsahy hodnot, načtení seznamů apod.



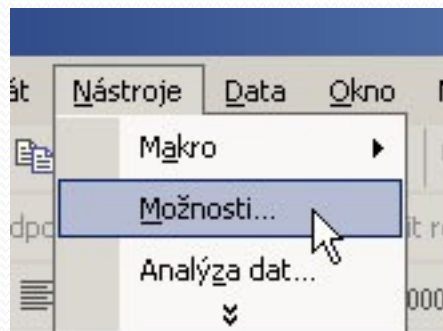
komunikace s uživatelem



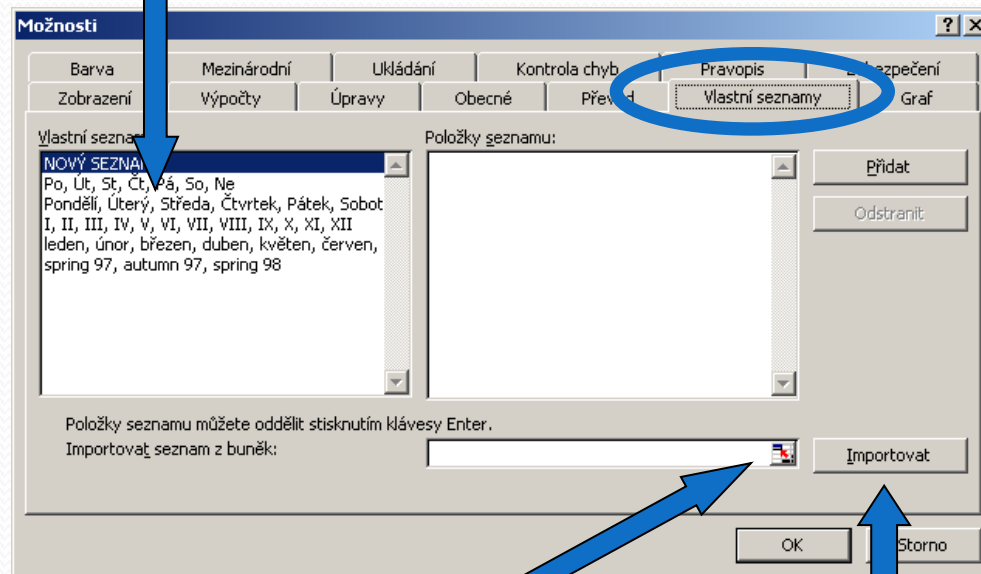
# Seznamy I.

- 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

Microsoft Office 2003 a starší



Existující seznamy



Výběr buněk pro nový seznam

Načtení nového seznamu

Vytvořil Institut biostatistiky a analýz,  
Masarykova univerzita

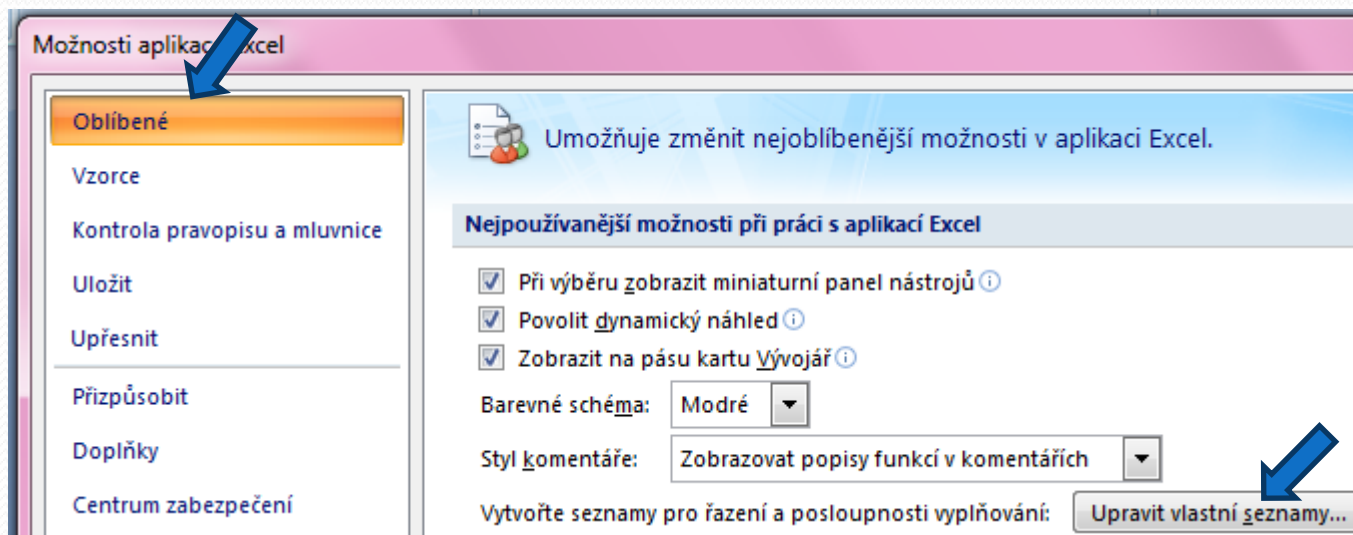
# Seznamy II.

Microsoft Office 2007

- „Tlačítko Office“



→ „Možnosti aplikace Excel“



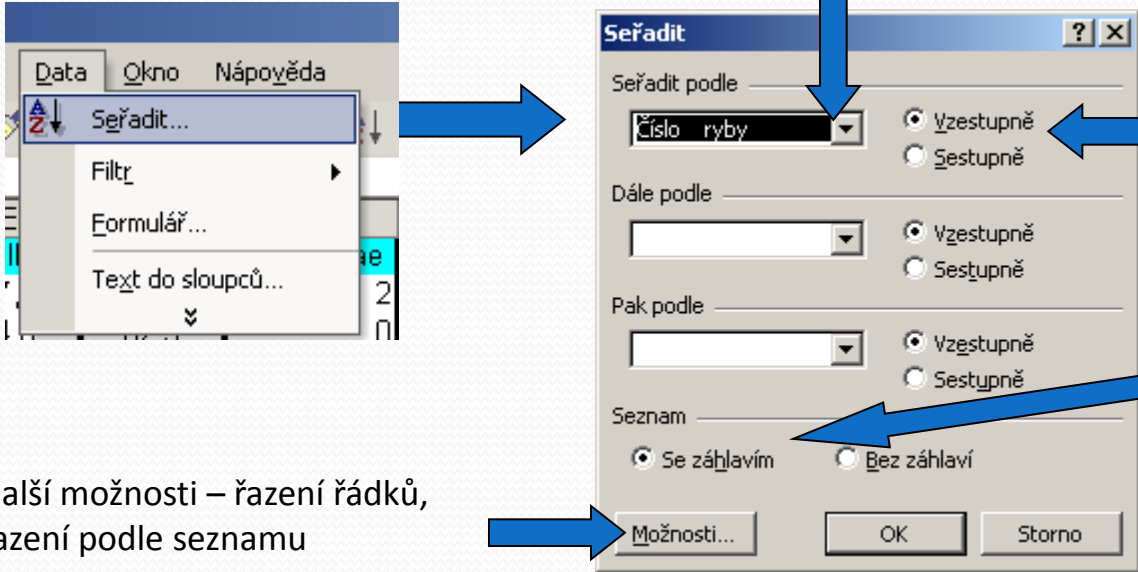
- Vlastní seznamy dále stejné (viz předchozí slide)

# 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

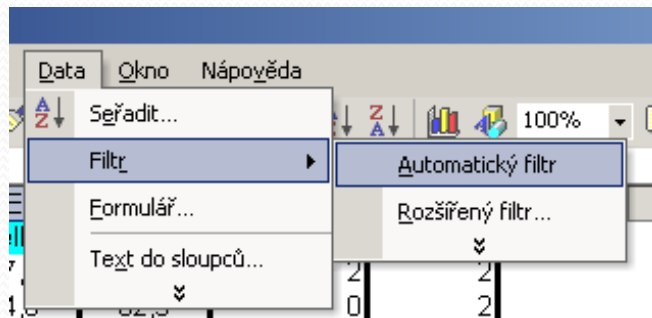
Vytvořil Institut biostatistiky a analýz,  
Masarykova univerzita



# 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

The image shows a Microsoft Excel spreadsheet titled 'Microsoft Excel - ryby.xls'. The spreadsheet has columns A through E. Columns A, B, and C have dropdown arrows, indicating they are filtered. The data in the table is as follows:

	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

A blue arrow points from the 'Automatický filtr' option in the previous image to the 'Společ' dropdown menu in this table, which is currently open, showing a list of values: (Vše), (Prvních 10...), (Vlastní...), f?, and m.

Vytvořil Institut biostatistiky a analýz,  
Masarykova univerzita

# Automatické dokončování hodnot buněk

- Vhodné pro textová pole; následně není nutné vypisovat celé slovo či slovní spojení, ale jen zvolit nabízené, již dříve použité slovo či slovní spojení
- Automatické dokončování hodnot buněk je nutné nastavit
  - „Tlačítko Office“ → „Možnosti aplikace Excel“

