

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

Pacient	Clovek	aLeu cell.10 ⁶ /	aTy% %	aSe% %	aNeu% %	aLy% %	aTy cell.10 ⁶ /	aSe cell.10 ⁶ /	aNeu cell.10 ⁶ /	aLy cell.10 ⁶ /	aHtc %	aCLsk mV.s.10 ³	aCLNeus mV.s.10 ³	aCLOZ mV.s.10 ³	aCLNeuO mV.s.10 ³
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

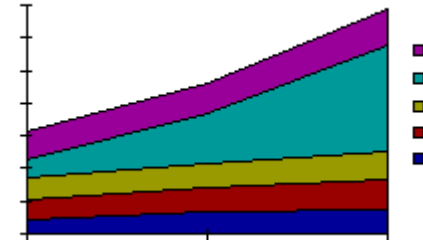


- Správné a přehledné uložení dat je základem jejich pozdější analýzy
- Je vhodné rozmyslet si předem jak budou data ukládána
- Pro počítačové zpracování dat je nezbytné ukládat data v tabulární formě
- Nejvhodnějším způsobem je uložení dat ve formě databázové tabulky
 - Každý sloupec obsahuje pouze 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

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	1	
6	2	
7	26	
8	106	
9	121	
10	160	
11	34	
12	45	
13	70	
14	72	
15	87	
16	Celkový součet	
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**

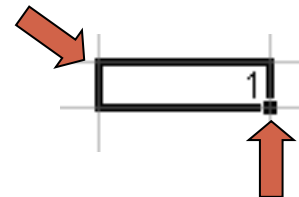
- CTRL+HOME – přesunutí na levý horní roh tabulky
- CTRL+END – přesunutí na pravý dolní roh tabulky
- CTRL+A – výběr celého listu
- CTRL + klepnutí myší do buňky – výběr jednotlivých buněk
- SHIFT + klepnutí myší na jinou buňku – výběr bloku buněk
- SHIFT + šipky – výběr sousedních buněk ve směru šipky
- 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).



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“

	F	G	H	
	poslední kontrola	pohlaví	nemocný	tíž
9	9.4.2010	muž		1
10	29.3.2010	muž		1

- 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.)



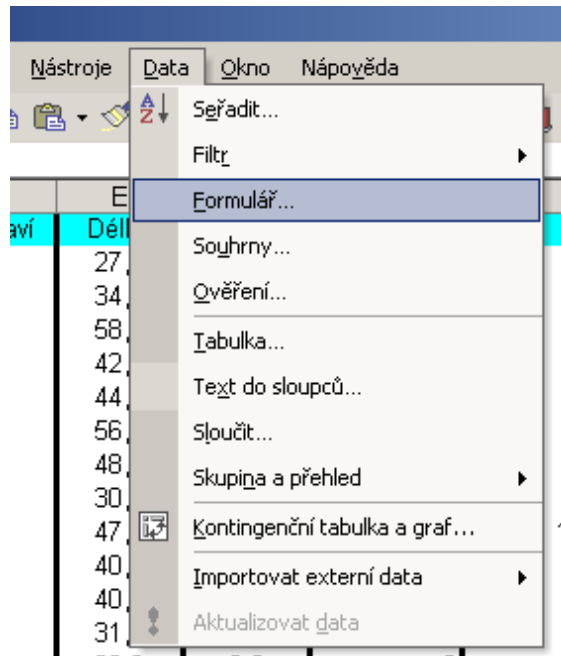
	A	B	C	D	E	F	G	H	I
1	Číslo	Značka	Společ	Pohlaví	Délka	Váha	P. anguillae	P. bini	
2	1	1	1	m	27,5	23,0	2	2	
3	2	2	2	f	34,0	62,5	0	2	
4	3	5	3	f	58,0	230,0	0	0	
5	4	6	4	f	42,0	155,0	0	0	
6	5	7	5	f	44,0	149,8	0	0	
7	6	8	6	f	56,0	323,0	0	1	
8	7	9	7	m	48,5	178,2	0	0	
9	8	10	8	f	30,5	47,7	4	6	
10	9	11	9	f	47,0	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



Číslo ryby:	1	1 z 19
Značka ryby:	1	Nový
Společ číslo:	1	Odstranit
Pohlaví:	m	Obnovit
Délka:	27,5	Předchozí
Váha:	23	Další
P. anguillae:	2	Kritéria
P. bini:	2	Zavřít

Nový záznam

Vyhledávání

Názvy sloupců

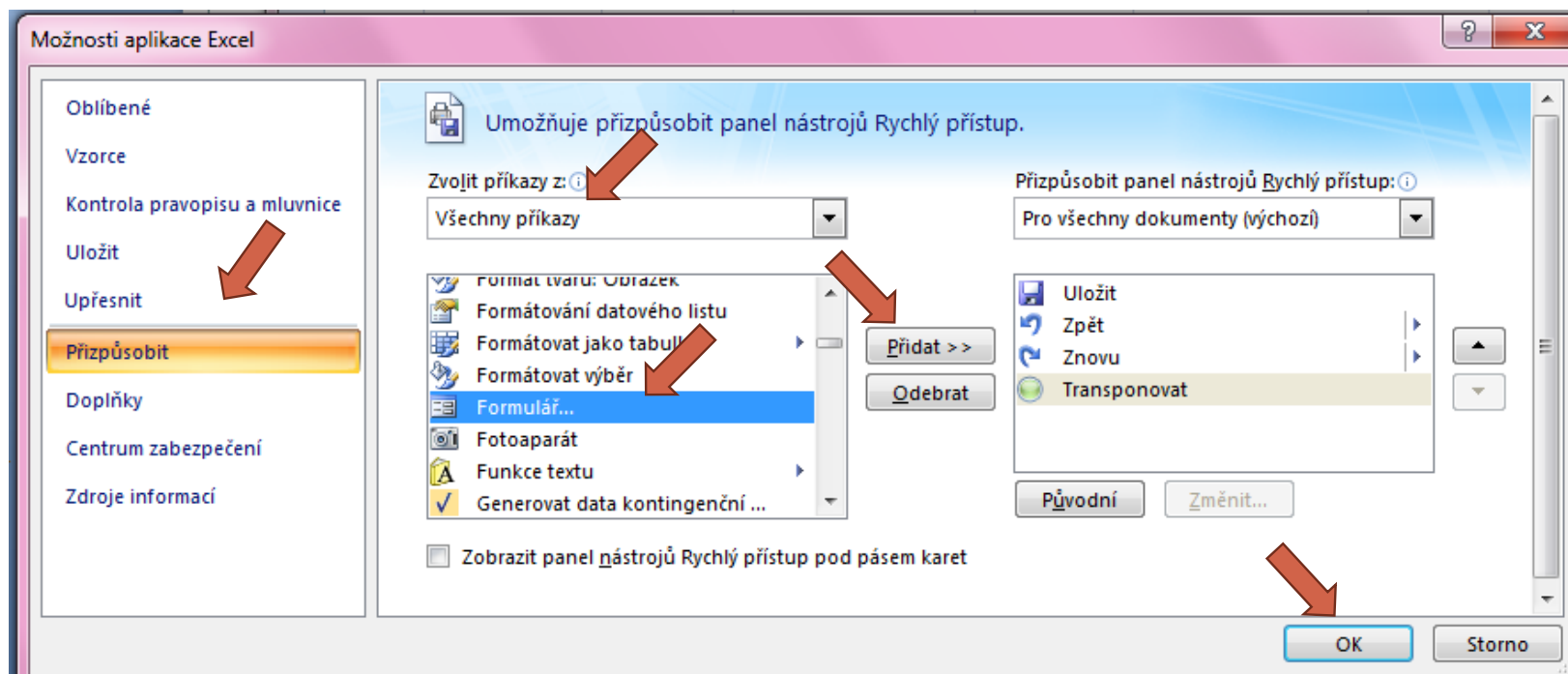
Obsah dané buňky - editovatelný

Automatický zadávací formulář II.

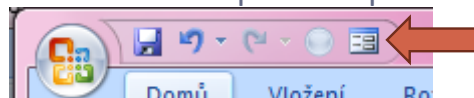
Microsoft Office 2007

- Aplikaci automaticky zadávaného formuláře je nutné aktivovat

- „Tlačítko Office“ → „Možnosti aplikace Excel“



- Automatický zadávací formulář spustíme pomocí nové ikonky na panelu nástrojů Rychlý přístup; dále stejné



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

Context menu for the selected cell:

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

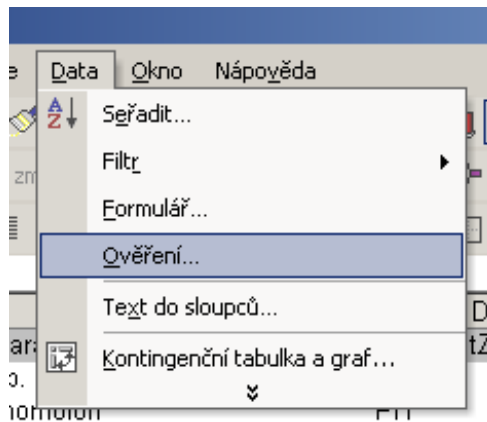
Dropdown list of taxonomic names:

- 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



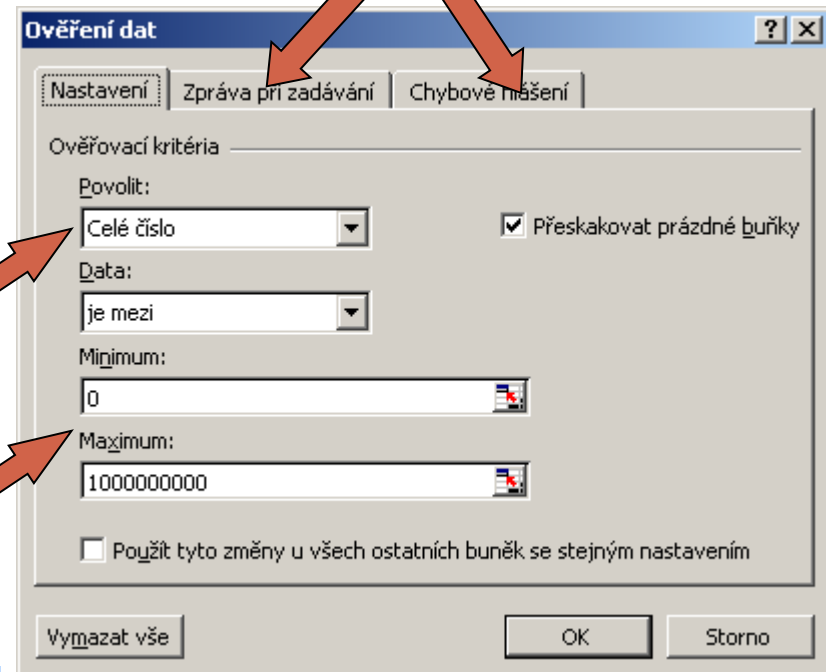
- Umožňuje ověřit typ, rozsah nebo povolit pouze určitý seznam hodnot zadávaných do sloupce databázové tabulky



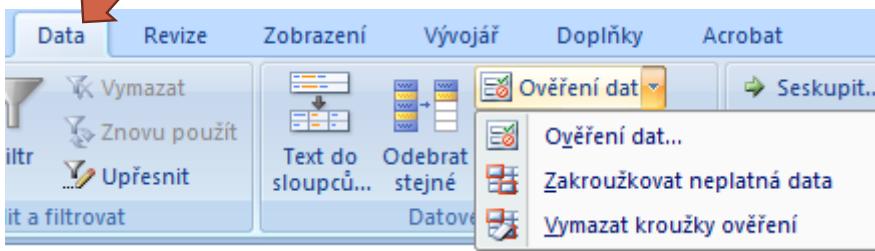
Co je povoleno – definiční obory čísel, seznamy, vzorce atd.

Rozsahy hodnot, načtení seznamů apod.

komunikace s uživatelem



Microsoft Office 2007

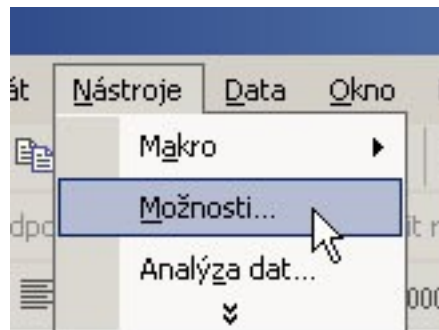


Seznamy I.

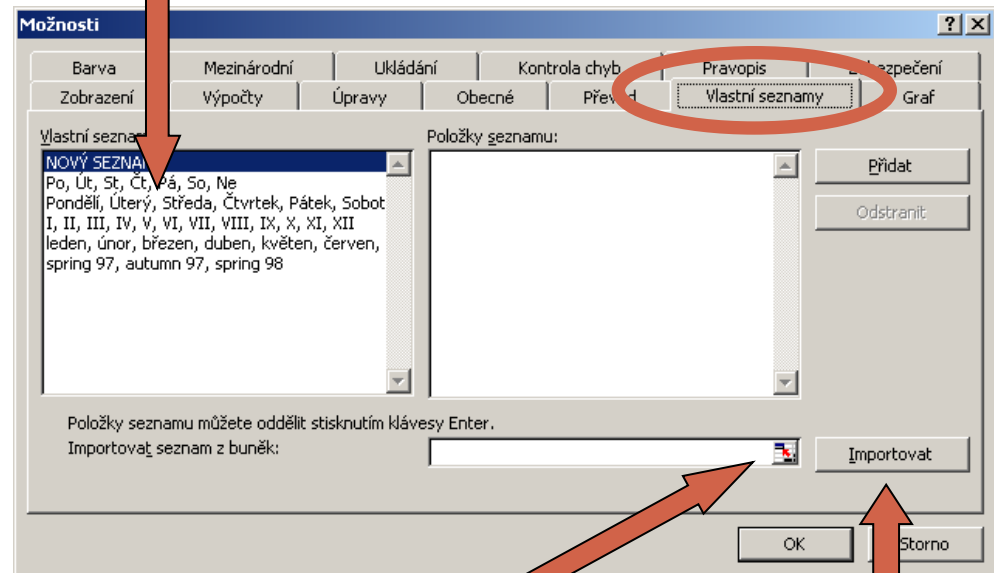


- Skupiny hodnot zachovávající logické pořadí, některé jsou zabudované (např. dny v týdnu, měsíce v roce), další je možné uživatelsky vytvořit, slouží pro účely řazení a automatického vyplňování dat

Microsoft Office 2003 a starší



Existující seznamy



Výběr buněk pro nový seznam

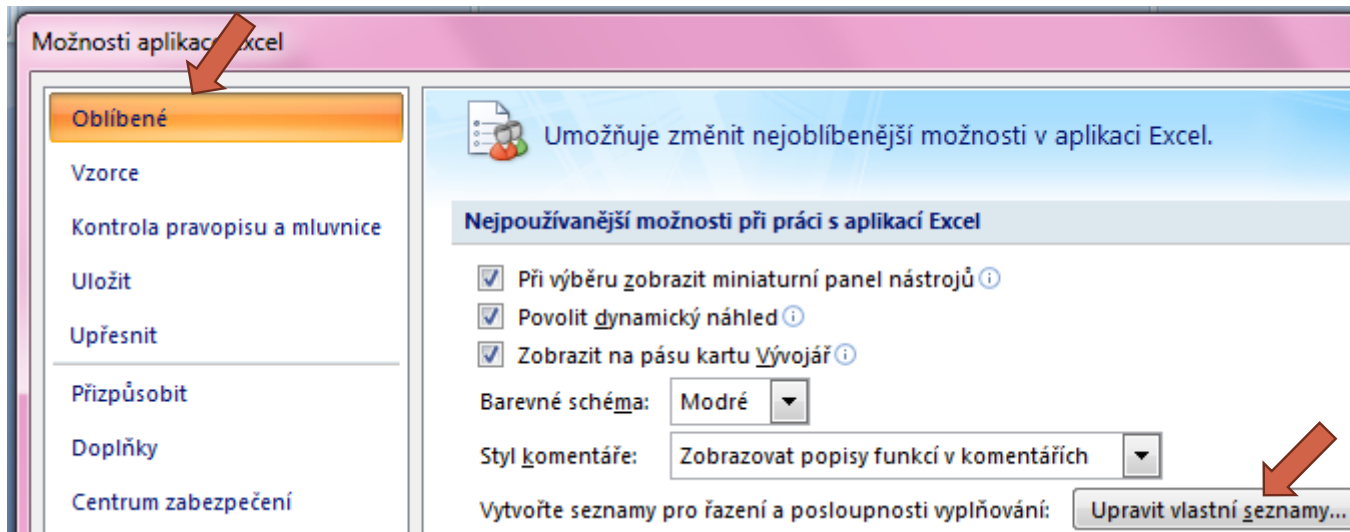
Načtení nového seznamu

Seznamy II.



Microsoft Office 2007

- „Tlačítko Office“ → „Možnosti aplikace Excel“



- Vlastní seznamy dále stejné (viz předchozí slide)

Řazení dat

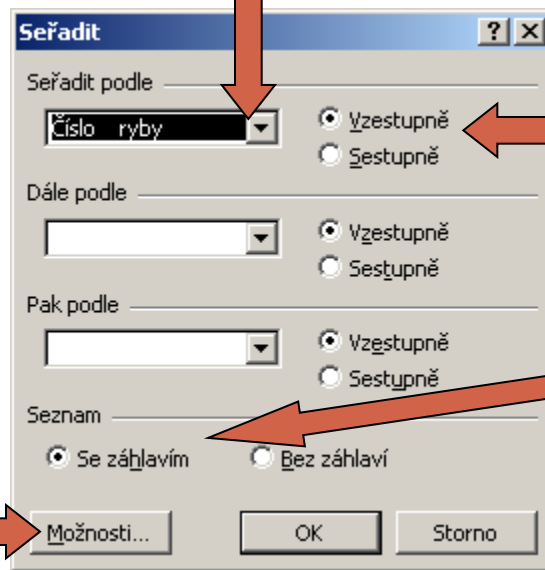
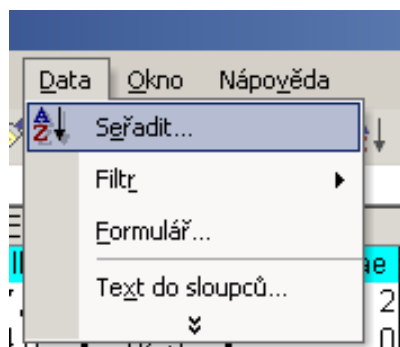


- Řazení dat je nejjednodušším způsobem jejich zpřehlednění, užitečným hlavně u menších/výsledkových tabulek



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

Podle čeho řadit



Směr řazení – vzestupně, sestupně

Využít první řádek oblasti jako záhlaví

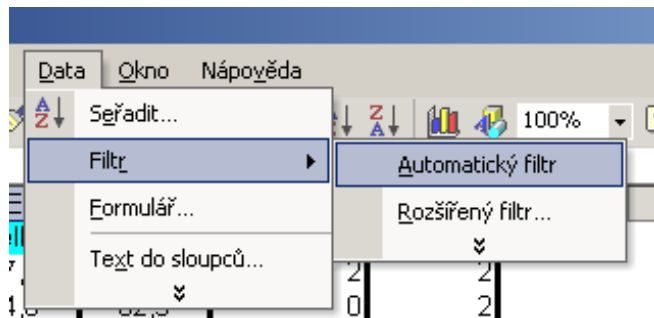
Další možnosti – řazení řádků, řazení podle seznamu

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

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“



→ „Možnosti aplikace Excel“

