

MNBS081p;c Biostatistika Nutriční specialista – přednáška a cvičení (jaro 2018)



PŘEDNÁŠEJÍCÍ A CVIČÍCÍ:
JAN MUŽÍK
INSTITUT BIOSTATISTIKY A ANALÝZ
muzik@iba.muni.cz

Důležité informace



- Lze 2 absence, praktická cvičení je ale nutné doma dopracovat
- Ukončení:
 - Cvičení: aktivní účast
 - Přednáška: závěrečná teoretická a praktická písemka
- Materiály v IS
- Software: Microsoft Office - Excel 2010, Statistica 12
- Možnost pracovat na vlastním počítači (složky se z univerzitních počítačů po vypnutí počítače mažou)
- Výuka: 14:00-15:40, učebna A1/IBA, 6. patro COMPK6, Kamenice 3

Organizace výuky (Út 20. 2. až Út 29. 5. 2018)



- 01 Úvod, Příprava dat, Excel – základy, základní vzorce
- 02 Excel – vzorce, grafy, kontingenční tabulky
- 03 Excel – kontingenční tabulky, grafy; Dotazníky
- 04 Statistická analýza dat; Soft. Statistica – úvod
- 05 Testování hypotéz; Soft. Statistica – základní funkce, grafy, popisná statistika
- 06 Dokončení teorie testování; Zjišťování normality dat; Jednovýběrové testy; Dvouvýběrové parametrické a neparametrické testy
- 07 Párové parametrické a neparametrické testy; Kontingenční tabulky
- 08 Kontingenční tabulky – testy dobré shody – pokračování
- 09 ANOVA, korelace, základy regresní analýzy; Opakování
- 10 **Závěrečný test 22. 5. 2018 14:00-15:40**

Motivace



- 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 nalezena ve fázi přípravy dat se promítne do všech dalších kroků a může zapříčinit neplatnost výsledků a nutnost opakování analýzy.

Čas na otázky



I. Příprava dat, MS Excel



Datová tabulka Zásady správné tvorby dat Představení MS Excel

DATA – ukázka uspořádání datového souboru

Parametry (znaky)

Opakování

Pařet	Číslek	a1u cel.10 ⁹	aly% %	a2% %	a3u% %	aly% %	aly cel.10 ⁹	a2e cel.10 ⁹	a3u cel.10 ⁹	ay cel.10 ⁹	a1c %	a1sk m/s10 ³	a1Nas m/s10 ³	a1OY m/s10 ³	a1NaO m/s10 ³
3	1	4									33	72		32	
4	2	76	8	38	66	24	06	44	50	18	33	95	19	48	10
8	3	4	3	52	55	40	01	21	22	16	22	77	35	33	15
11	4	61	5	39	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	11	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	33	5
21	9	6	10	67	77	19	06	40	46	11	32	120	26	52	11
22	10	33	4	55	59	39	01	18	20	13	28	81	42	24	12
37	11	38	10	60	70	30	04	23	27	11	32	111	42	29	11
38	12	64	2	76	78	17	01	49	50	11	25	366	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	45	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	25	101	25	54	13
6	1	88	11	72	83	12	10	63	73	11	44	268	356	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	13	44	103	178	63	109
24	7	82	1	72	73	25	01	59	60	21	41	219	349	57	96
26	8	103	1	85	86	3	01	88	89	03	41	364	411	112	126
29	9	50	1	74	75	21	01	37	38	11	39	83	221	32	85
30	10	119	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	271	63	155
32	12	108	36	50	76	8	39	54	93	09	27	146	157	106	114
33	13	118	22	54	76	16	26	64	90	19	45	246	274	63	7
34	14	170	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

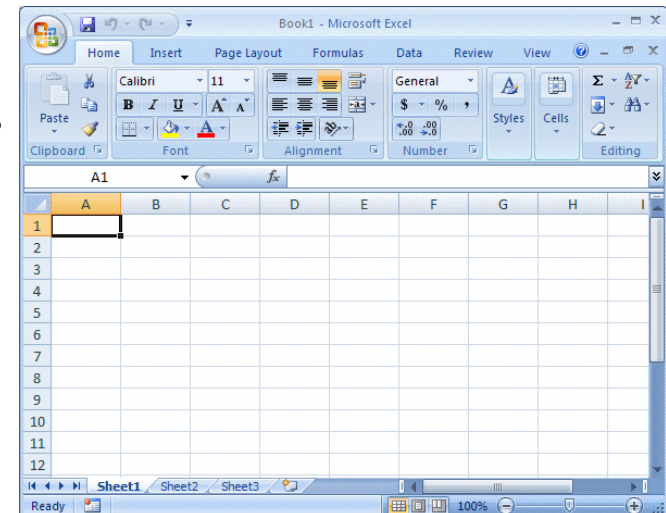


- Správné a přehledné uložení dat je základem jejich pozdější analýzy
- Je vhodné rozmyslet si předem jak budou data ukládána
- Pro počítačové zpracování dat je nezbytné ukládat data v tabulární formě
- Nejvhodnějším způsobem je uložení dat ve formě databázové tabulky
 - Každý **sloupec** obsahuje pouze **jediný typ dat**, identifikovaný hlavičkou sloupce;
 - Každý **řádek** obsahuje **minimální jednotku dat** (např. pacient, jedna návštěva pacienta apod.);
 - Je nepřipustné kombinovat v jednom sloupci číselné a textové hodnoty;
 - Komentáře jsou uloženy v samostatných sloupcích;
 - U textových dat je nezbytné kontrolovat překlepy v názvech kategorií;
 - Specifickým typem dat jsou data, u nichž je nezbytné kontrolovat, zda jsou 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

MS Excel



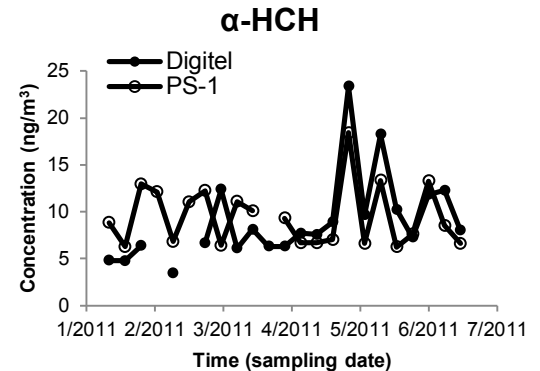
- Tabulkový procesor.
- Aktualizace každé 2 až 3 roky; nové funkce, rozšíření počtu řádků a sloupců, změna formátu.
- Starší formát: .xls, novější: .xlsx.
- Aktuální verze 2013 umožňuje ukládat tabulku až o 1 048 576 řádcích a 16 384 sloupcích.
- Maximální velikost buňky je 32 767 znaků.



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).



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

P. biní	2	Pohlaví			
Počet z	Délka				
Číslo	ryby2	Číslo	rv1	Váha	?
1	1	<input type="checkbox"/>	(Zobrazit vše)		
2	2	<input checked="" type="checkbox"/>	68		
26	26	<input checked="" type="checkbox"/>	99		
106	106	<input checked="" type="checkbox"/>	102		
121	121	<input checked="" type="checkbox"/>	109		
160	160	<input checked="" type="checkbox"/>	112		
34	34	<input checked="" type="checkbox"/>	120		
45	45	<input checked="" type="checkbox"/>	173		
70	70	<input checked="" type="checkbox"/>	28		
72	72	<input checked="" type="checkbox"/>	29		
87	87	<input checked="" type="checkbox"/>	29		
Celkový součet		OK	Storno		

II. Tvorba dat v MS Excel



Import a export dat

Struktura dat

Manuální zadávání nových dat

Tipy a triky jak se v datech pohybovat

Editace listů

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.

- **Export dat**

- Ukládáním souborů ve formátech podporovaných jinými SW, časté jsou textové soubory, dbf soubory nebo starší verze Excelu
- Přímé kopírování přes schránku Windows

Databázová struktura dat v Excelu



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

	A	B	C	D	E	F
1	CisloRyby	taxon	ParazitCelyNazev	ParazitZkratka	typParazita	PocetParazitu
2	97	Monogenea	Gyrodactylus sp.	Gsp	ektoparaz	1
3	97	Monogenea	Paradiplozoon homoion	PH	ektoparaz	23
4	97	acanthocephala	Acanthocephallus lucii (Müller, 1776)	Acl	endopara	1
5	98	Crustacea	Lernaea cyprinacea Linnaeus, 1758	LER	ektoparaz	9
6	98	Monogenea	Paradiplozoon homoion	PH	ektoparaz	7
7	99	Monogenea	Paradiplozoon homoion	PH	ektoparaz	3
8	100	Monogenea	D. rutili Gläser, 1965	Drut	ektoparaz	2
9	100	Monogenea	D. crucifer Wagener, 1857	Dcru	ektoparaz	2
10	100	Monogenea	Paradiplozoon homoion	PH	ektoparaz	1
11	101	Monogenea	D. crucifer Wagener, 1857	Dcru	ektoparaz	2
12	101	Monogenea	D. nanus Dogiel et Bychowsky, 1934	Dnan	ektoparaz	6
13	101	Monogenea	Paradiplozoon homoion	PH	ektoparaz	3
14	102	Monogenea	D. cabaleroi	Dcab	ektoparaz	2
15	102	Monogenea	D. crucifer Wagener, 1857	Dcru	ektoparaz	2
16	102	Monogenea	Paradiplozoon homoion	PH	ektoparaz	1
17	103	Monogenea	D. crucifer Wagener, 1857	Dcru	ektoparaz	2
18	103	Crustacea	Argulus foliaceus (Linnaeus, 1758)	AF	ektoparaz	1
19	103	Monogenea	Paradiplozoon homoion	PH	ektoparaz	2
20	104	Monogenea	D. nanus Dogiel et Bychowsky, 1934	Dnan	ektoparaz	2

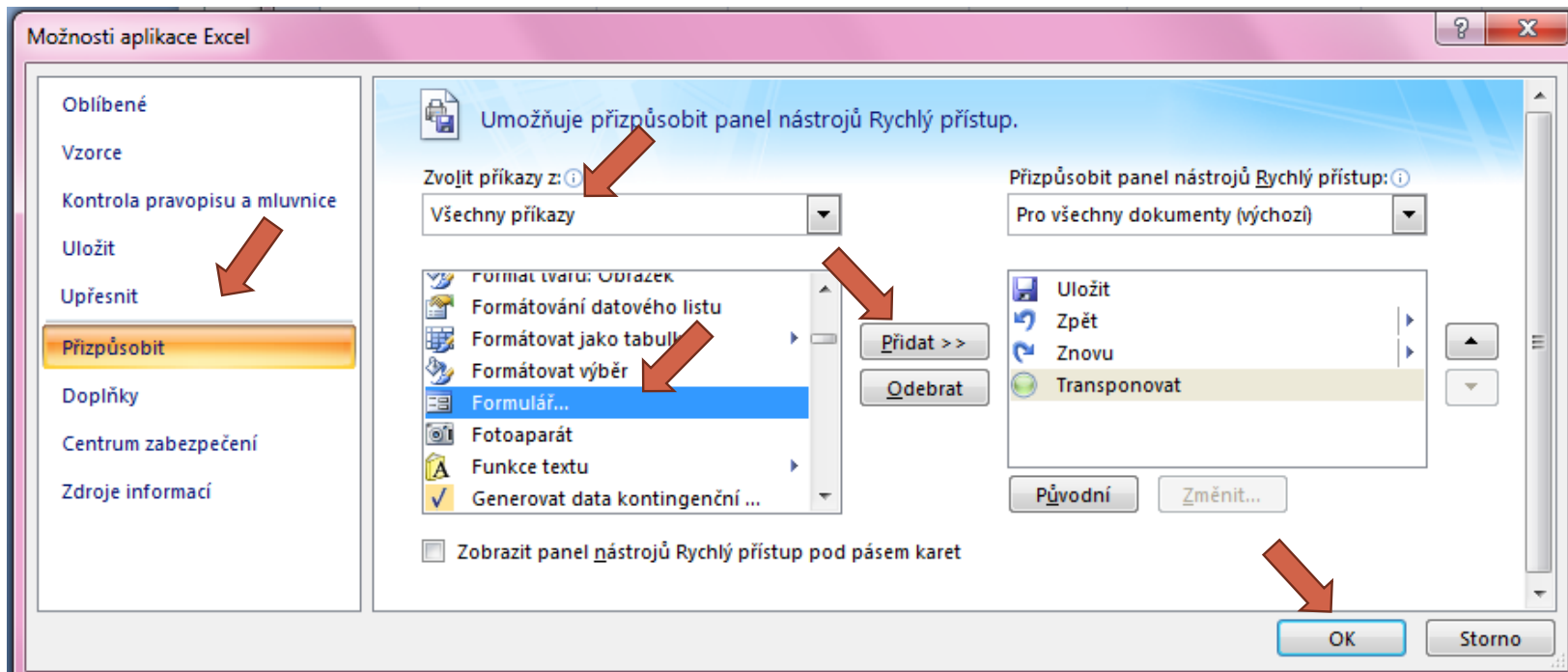
Řádky tabulky =
jednotlivé záznamy
(taxon, lokalita,
měření, pacient atd.)

Excel neumožňuje pojmenování řádků a sloupců vlastními názvy.

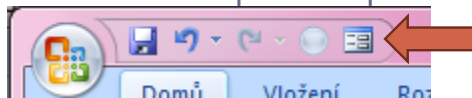
Automatický zadávací formulář I.



- Aplikaci automaticky zadávaného formuláře je nutné aktivovat
 - „Tlačítko Office“ → „Možnosti aplikace Excel“



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

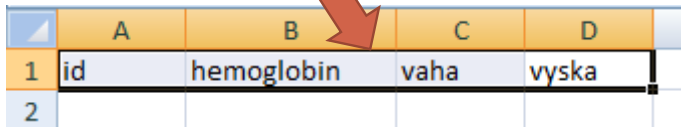


Automatický zadávací formulář II.

- Slouží k usnadnění zadávání dat do databázových tabulek
- Po označení načítá automaticky hlavičky sloupců jako zadávané položky

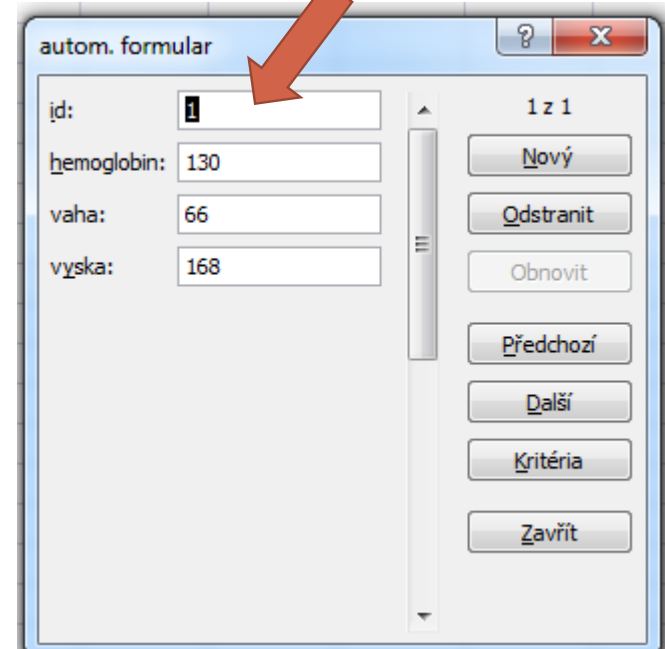
3. Vyplníme údaje pro hodnocený subjekt

1. Označíme názvy sloupců datové matice

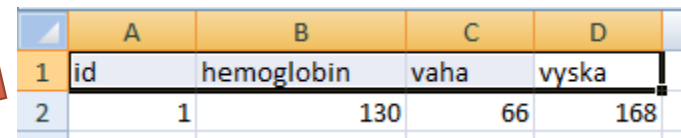


	A	B	C	D
1	id	hemoglobin	vaha	vyska
2				

2. Klikneme na novou ikonu Formulář v panelu nástrojů



4. Do datové tabulky se doplní zadané údaje



	A	B	C	D
1	id	hemoglobin	vaha	vyska
2	1	130	66	168

Vyzkoušej !!!

Automatické seznamy



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

Ověření dat

Nastavení Zpráva při zadávání Chybové hlášení

Ověřovací kritérium

Povolit: Seznam

Rozsah: je mezi

Zdroj: =\$A\$1:\$A\$4

Použít tyto změny u všech ostatních buněk se stejným nastavením

Vymazat vše OK Storno



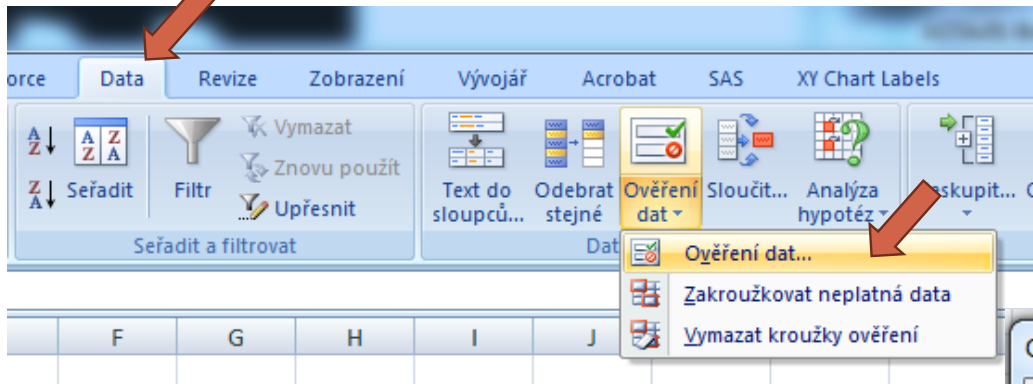
F	G	H
	Datova tabulka	
id	BMI	
1	podváha (BMI<18,5)	
2	obezita (BMI>29,9)	
3	nadváha (BMI: 25,0-29,9)	
	podváha (BMI<18,5)	
	norma (BMI: 18,5-24,9)	
	nadváha (BMI: 25,0-29,9)	
	obezita (BMI>29,9)	

Vyzkoušej !!!

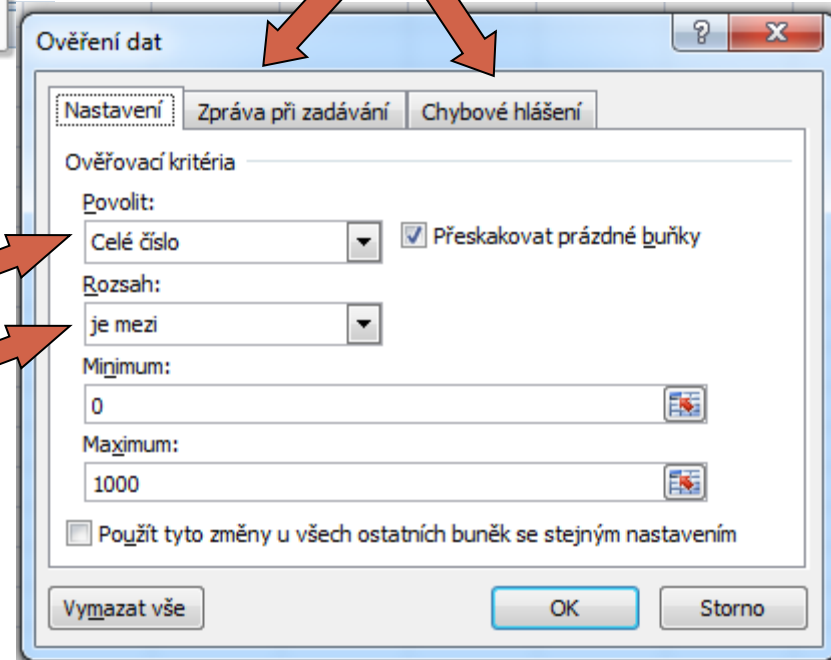
Automatická kontrola dat



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



komunikace s uživatelem



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

Rozsahy hodnot, načtení seznamů apod.

Vyzkoušej !!!

Seznamy I.

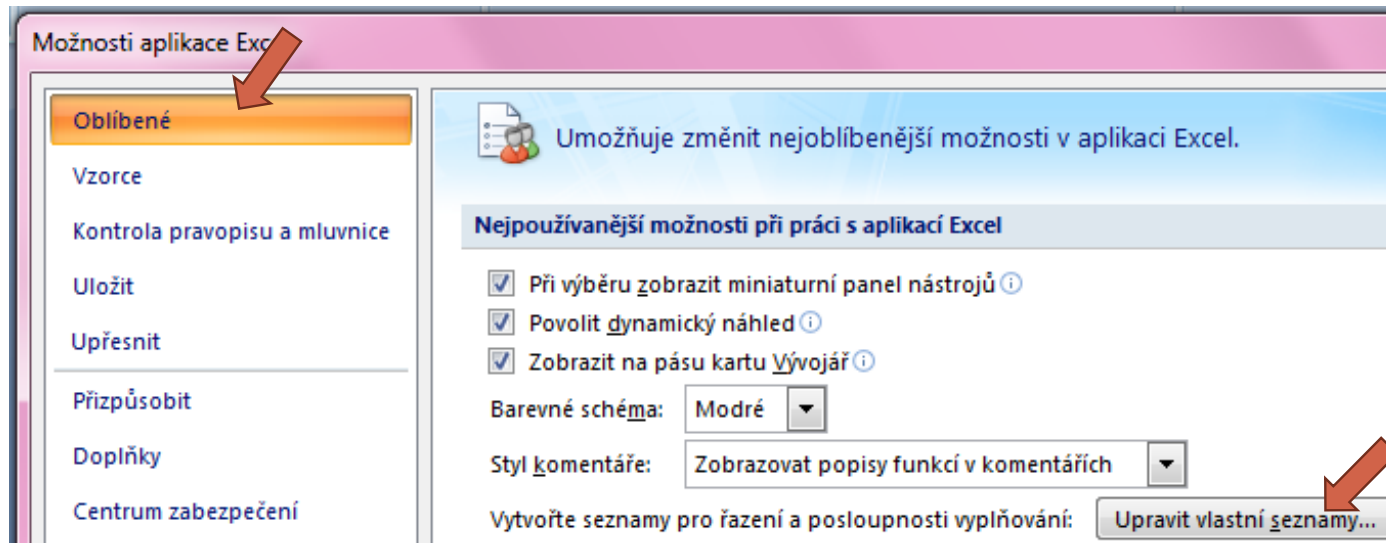


- Skupiny hodnot zachovávající logické pořadí, některé jsou zabudované (např. dny v týdnu, měsíce v roce), další je možné uživatelsky vytvořit, slouží pro účely řazení a automatického vyplňování dat

○ „Tlačítko Office“



→ „Možnosti aplikace Excel“

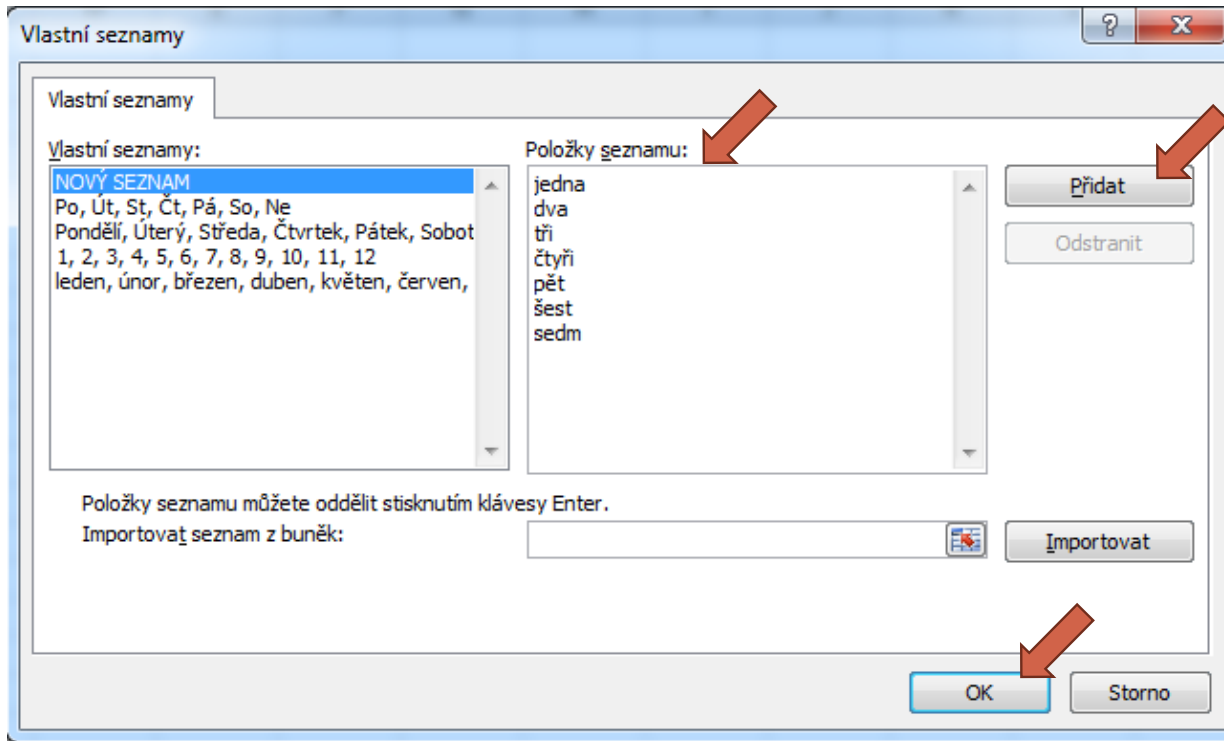


Seznamy II.



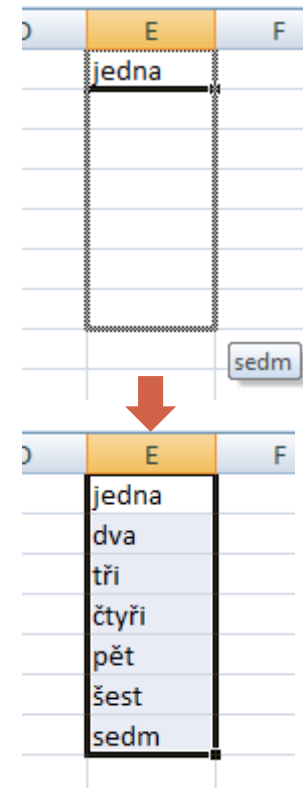
1. Definice seznamu

- Upravit vlastní seznamy →



2. Využití při tvorbě dat

Zápis jedné hodnoty ze seznamu a protažení do dalších buněk



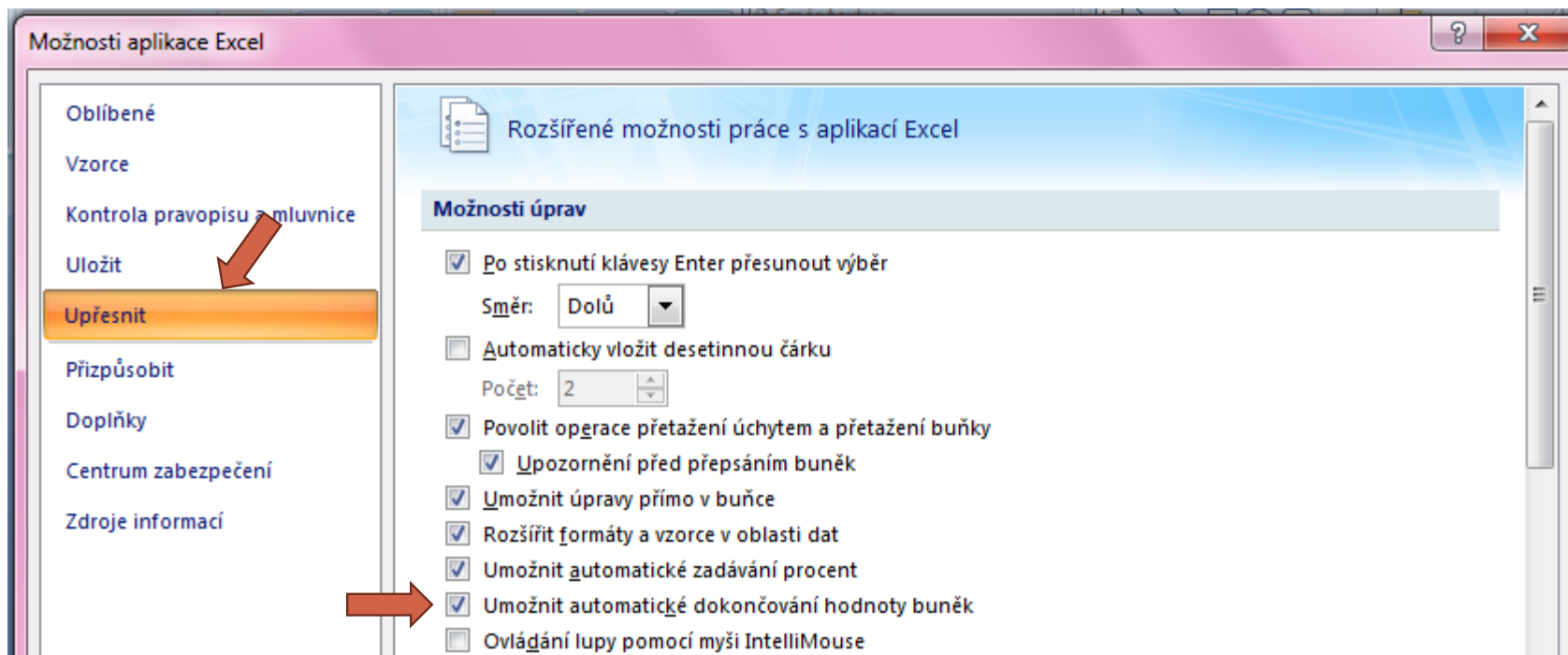
Automaticky byly doplněny následující složky seznamu

Vyzkoušej !!!

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“



Tipy a triky jak se v datech pohybovat



• Výběr buněk

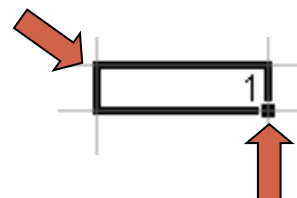
- CTRL+HOME – přesunutí na levý horní roh tabulky
- CTRL+END – přesunutí na pravý dolní roh tabulky
- CTRL+A – výběr celého listu
- CTRL + klepnutí myši do buňky – výběr jednotlivých buněk
- SHIFT + klepnutí myši na jinou buňku – výběr bloku buněk
- SHIFT + šipky – výběr sousedních buněk ve směru šipky
- 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, data z jiné aplikace

• Myš a okraje buňky

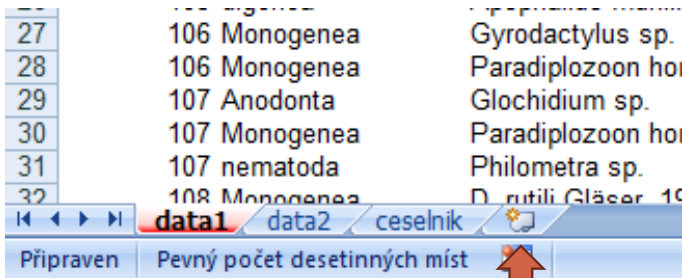
- Chycení myši za okraj umožňuje přesun buňky nebo bloku buněk
- Při chycení čtverečku v pravém dolním rohu výběru je tažením možno vyplnit více buněk hodnotami původní buňky (ve vzorcích se mění relativní odkazy, je také možné vyplnění hodnotami ze seznamu – např. po sobě jdoucí názvy měsíců).



Editace listů

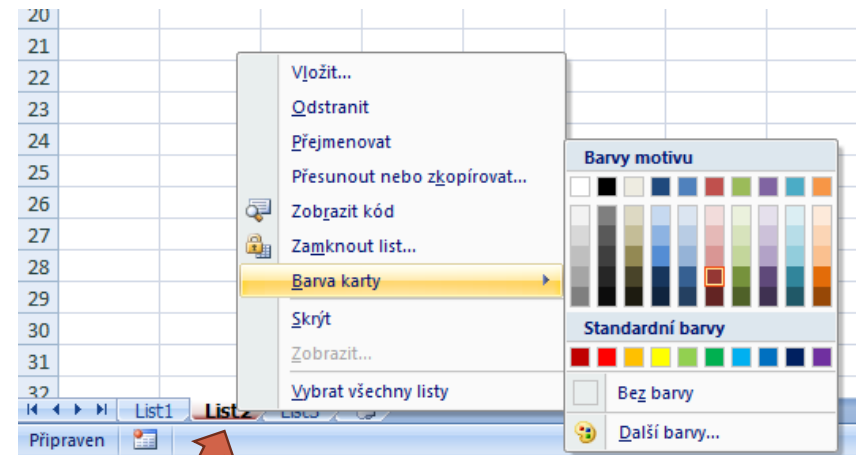


- Excelovský soubor (sešit) se skládá z listu(ů) (List1, List2, ...), které je možné libovolně pojmenovat, obarvit, kopírovat, přesouvat jejich pořadí na listě atd.
- Ve vzorcích lze odkazovat na jiné listy než ve kterém se nacházíme.
- V jednotlivých listech lze ukládat např. různé datové tabulky, číselníky, seznamy atd.



- **Vložení listu:**

- Záložkou na spodní listě
- Klávesovou zkratkou: **Shift + F11**



Klik pravým tlačítkem myši na záložku listu

III. Práce s daty v MS Excel aneb jak se v datech vyznat



Formáty buněk
Seřazení dat
Filtrování
Ukotvení příček
Podmíněné formátování

Formáty buněk

- **Obecný** (bez formátu)
- **Číslo** (desetinná místa, oddělení 1000)
- **Měna** (desetinná místa, jednotky - symbol)
- **Datum** (různé formátování – 24.9.2015, 24. září 2015, 24-9-15, ...)
- **Čas**
- **Procenta**
- **Text**
- **Další** (vč. nastavení vlastního formátu)...

01_Excel_Data [Režim kompatibility] - I

Domů Vložení Rozložení stránky Vzorce Data Revize Zobrazení Vývojář PDF Acrobat

Arial 10

Obecný

Číslo

Formát buněk

Číslo Zarovnání Písmo Ohraničení Výplň Zámek

Druh:

- Obecný
- Číslo
- Měna
- Účetnický
- Datum
- Čas
- Procenta
- Zlomky
- Matematický
- Text
- Speciální
- Vlastní

Ukázka
CisloRyby

Desetinná místa: 0

Oddělovat 1000 (,)

Záporná čísla:

- 1234
- 1234
- 1234
- 1234

Číslo se používá pro obecné zobrazování čísel. Formát měny a účetnický formát poskytují speciální formáty pro peněžní hodnoty.

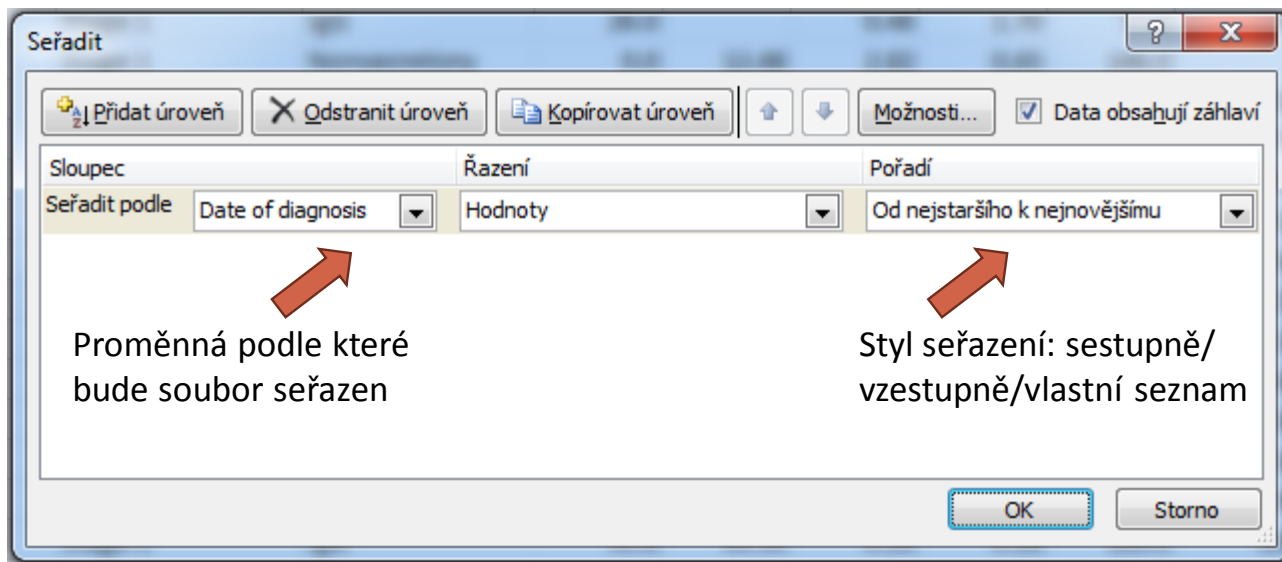
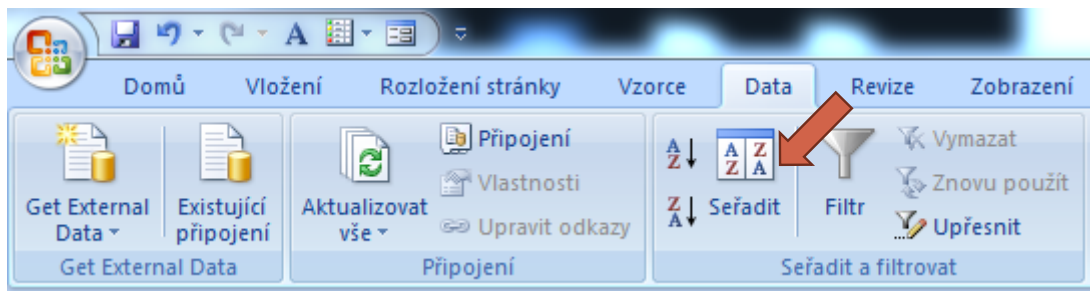
OK Storno

	A	B
1	CisloRyby	taxo
2	97	Monogenea
3	97	Monogenea
4	97	acanthocep
5	98	Crustacea
6	98	Monogenea
7	99	Monogenea
8	100	Monogenea
9	100	Monogenea
10	100	Monogenea
11	101	Monogenea
12	101	Monogenea
13	101	Monogenea
14	102	Monogenea
15	102	Monogenea
16	102	Monogenea
17	103	Monogenea
18	103	Crustacea
19	103	Monogenea
20	104	Monogenea
21	104	Monogenea
22	105	Monogenea
23	105	Monogenea
24	105	Monogenea
25	105	Monogenea
26	105	digenea
27	106	Monogenea
28	106	Monogenea

Řazení dat



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



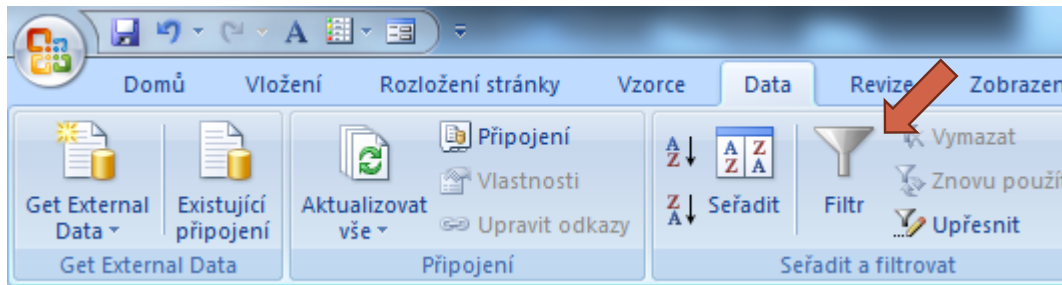
Zkontrolujte, zda seřazení nezničí vazby mezi buňkami = kontrola oblasti, kterou řadíte.

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
- **Výhodné pro čištění dat (vyhledávání překlepů, kombinace textu a čísel)**

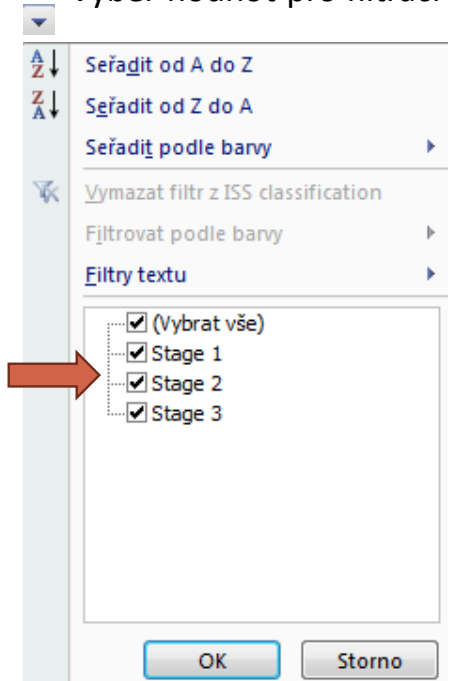
1. Zapnutí filtru (alternativa klávesová zkratka **Ctrl+Shift+L**)



2. Objeví se rozbalovací šipka s výčtem všech unikátních hodnot v daném sloupci dat

	A	B	C	D	E
1	ID	Date of birth	Date of diagnosis	Sex	ISS classification
2	189	26.12.1959	29.6.1994	F	Stage 1
3	5	9.8.1945	3.3.1997	M	Stage 1
4	192	19.7.1935	2.10.1997	M	Stage 2

Výběr hodnot pro filtraci



Rozšířený filtr



- Funguje podobně jako automatický filtr, ale seznam povolených hodnot není nutné vybírat ručně – je uveden v oblasti jinde na listu (nebo i na jiném listu).
- Podmínkou jsou shodná záhlaví filtrované oblasti a oblasti povolených hodnot.
- Prázdné buňky odpovídají prázdné podmínce – tj. je-li v oblasti povolených hodnot nějaká buňka prázdná, splní podmínku libovolná buňka filtrované oblasti.
- Čísla řádků filtrované oblasti jsou zobrazena modře.

Tlačítko Upřesnit na kartě Data

Revize Zobrazení Vývojář Doplněk Acrobat

Seřadit a filtrovat

Seřadit Filtr **Upřesnit**

	D
	ParazitZkratka
	Gsp
	DH

Upřesnit
Chcete-li omezit záznamy, které budou zahrnuty ve výsledné sadě dotazu, zadejte složitá kritéria.

Rozšířený filtr

Akce

Přímě v seznamu

Kopírovat jinam

Oblast seznamu:

Oblast kritérií:

Kopírovat do:

Bez duplicitních záznamů

OK Storno

Výběr oblasti cílových hodnot (přefiltrovaných)

Původní seznam včetně záhlaví

Oblast kritérií včetně záhlaví

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:

Ukotví řádky nad označenou buňkou a sloupce vlevo od označené buňky

	F	G	H	
	poslední kontrola	pohlaví	nemocný	tíž
19	9.4.2010	muž		1
.0	29.3.2010	muž		1

- Odstranění ukotvení: Po ukotvení příček se automaticky možnost „Ukotvit příčky“ změní na „Uvolnit příčky“

Podmíněné formátování



- Záložka „Domů“ → „Podmíněné formátování“.
- Barevné označení buněk nebo výplň buňky symbolem podle námi zadaných kritérií, např.:
 - numerická hodnota větší/menší než průměr
 - datum z konkrétního období
 - podobná slova
 - duplicitní údaje

- Co s barevnými buňkami?
- Použijeme filtr!

	123.0	320	2.35	41.1
0.45	129.0	218	2.36	48.5
	96.0	191	2.37	45.2

IV. Vzorce v Excelu

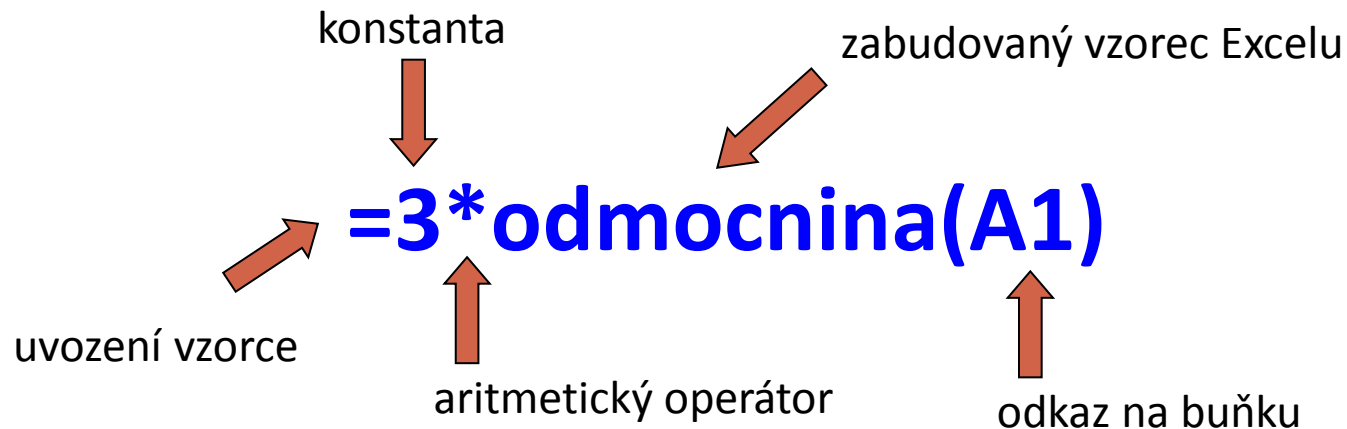


Tvorba vzorců
Absolutní a relativní odkazy
Kopírování vzorců

Vzorce



- vpisují se do buněk sešitu
- vzorce jsou vždy uvozeny = (lze též + -)
- aritmetické operátory + zabudované funkce Excelu
- pro „sčítání“ nečíselných položek se používá & (klávesová zkratka: **alt gr + c**)
- výpočet je založen buď na číselných konstantách nebo odkazech na buňky



Vzorce – odkaz na buňku



- **Relativní odkazy**

- **A1** = buňka 1. řádku sloupci A
- **A1:B6** = blok buněk – levý horní roh je v 1. řádku, sloupec A, pravý dolní na řádku 6, sloupec B
- relativní odkaz se při automatickém vyplnění buněk vzorcem posune
- mění se s kopírováním, při vložení a odstranění řádku nebo sloupce

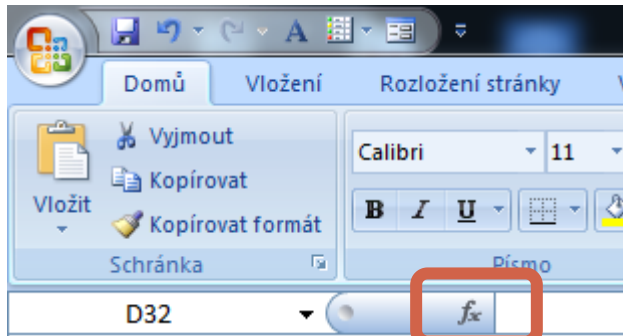
- **Absolutní odkaz**

- odkaz na buňku je pevně dán, při kopírování nebo automatickém vyplnění se nemění
- lze uzamknout jak řádky, tak sloupce samostatně

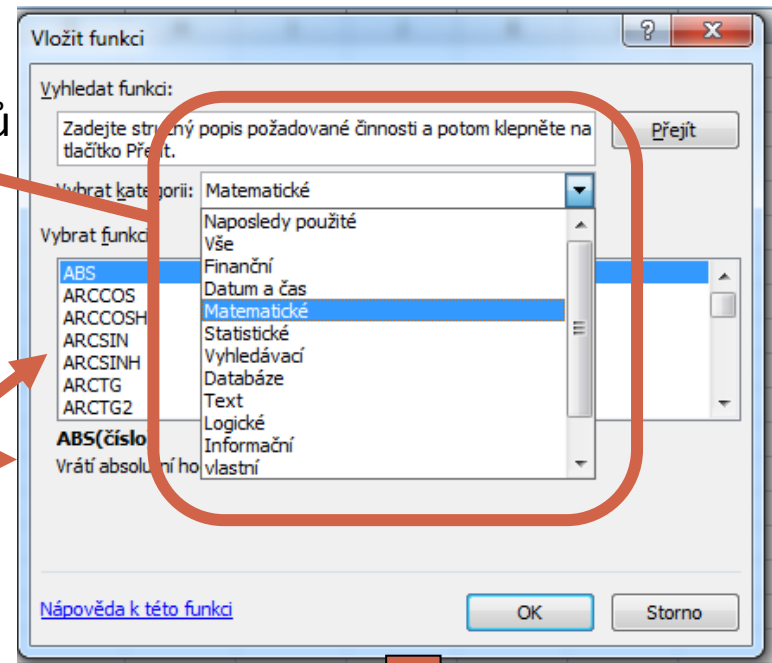
uzamčení řádku → **\$A\$1** ← uzamčení sloupce

Pamatuj: Adresu upevníme pomocí znaku **\$** (klávesová zkratka: **altr gr + ů** nebo **F4**)

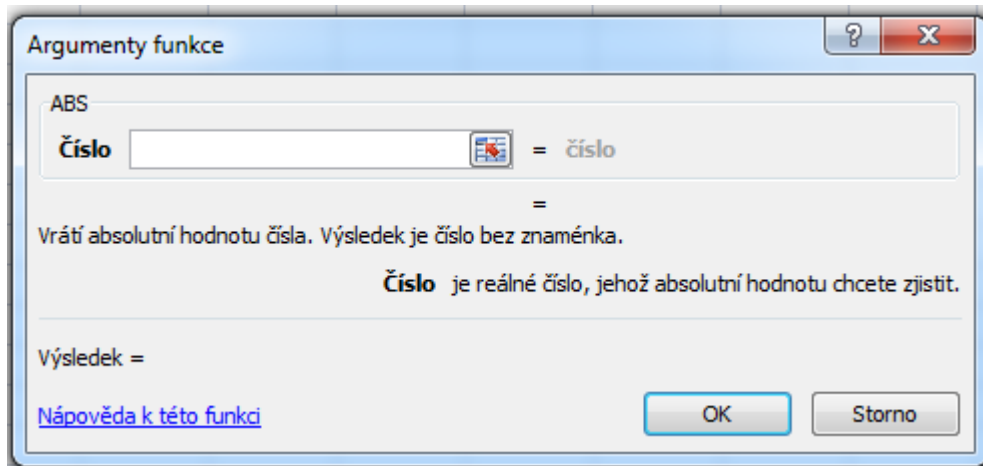
Vzorce – využití seznamu vzorců



Kategorie vzorců

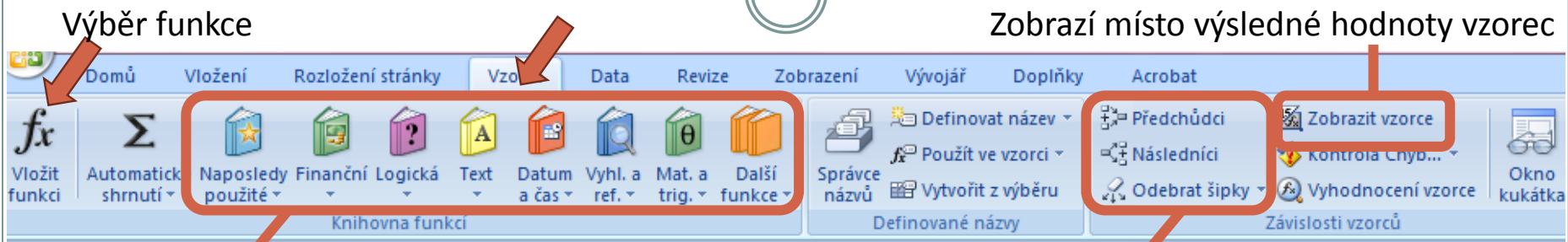


Funkce a její stručný popis



průvodce funkcí

Tvorba a závislosti vzorců



Výběr funkce z konkrétních knihoven

Označení a odznačení předchozích a následných vzorců

	A	B	C	D
1	Hemoglobin level (g/l)			Průměr
2	77.0			114.1
3	146.0			
4	87.0			
5	106.0			
6	125.0			
7	133.0			
8	117.0			
9	130.0			
10	86.4			
11	122.0			
12	123.0			
13	129.0			
14	96.0			
15	120.0			
16				

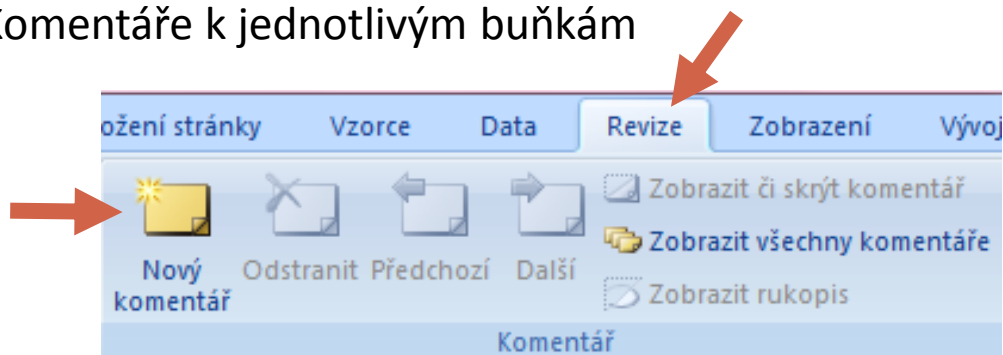
Zpřehlednění vzorců (zalomení řádku):

ALT+ENTER

Komentáře, sledování změn

- karta „Revize“

Komentáře k jednotlivým buňkám



možnost zamknout list či sešit dvojicí hesel

sledování změn a jejich schvalování nebo zamítní

Vzorce – užitečné funkce



Celkem 408 funkcí ve verzi MS Excel 2010, ve verzi 2013 přidáno 50 nových funkcí

- **SUMA** – součet číselných hodnot oblasti;
- **SUMIF** – podmíněný součet (podmínky v doplňkové oblasti);
- **PRŮMĚR** – aritmetický průměr číselných hodnot oblasti;
- **GEOMEAN** – geometrický průměr číselných hodnot oblasti;
- **COUNTIF** – počet hodnot oblasti splňujících zadanou podmínku;
- **KDYŽ** – logická podmínka (if);
- **MAX, MIN** – maximum/minimum číselných hodnot oblasti;
- **MEDIAN** – výpočet mediánu;
- **PERCENTILE** – výpočet percentilů;
- **DATUM (ROK, MĚSÍC, DEN)** – práce s kalendářními daty;
- **ABS** – absolutní hodnota;
- **SVYHLEDAT** – spojování tabulek podle identifikátoru - řádku.

Statistické funkce v MS Excel



Od verze 2007 obsahuje MS Excel některé pokročilé statistické funkce

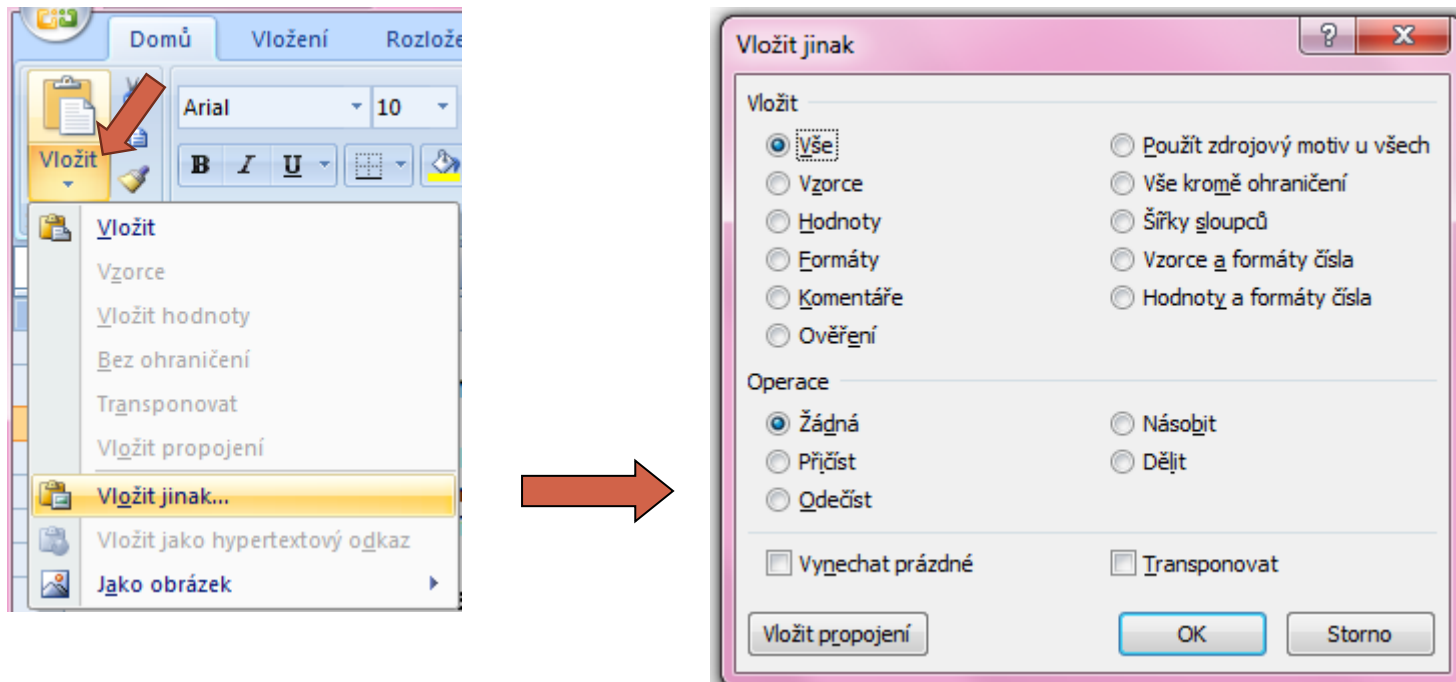
- **ZLEVA, ZPRAVA, ČÁST** – funkce pro ořezávání textových řetězců;
- **STANDARDNÍ MATEMATICKÉ FUNKCE (SIN, COS, LOG, LOGZ, EXP)** – a mnoho dalších.
- **CONFIDENCE** – výpočet intervalu spolehlivosti (při normálním rozdělení);
- **CORREL, PEARSON** – výpočet Pearsonova korelačního koeficientu;
- **COVAR** – výpočet kovariance dvou množin dat;
- **COUNTIF** – počet hodnot oblasti splňujících zadanou podmínku;
- **DEVSQ** – součet čtverců odchylek od výběrového průměru;
- **FDIST, GAMMADIST, CHIDIST, TDIST, NORMDIST** aj. – různá rozdělení pravděpodobnosti;
- **PRŮMODCHYLKA** – průměrná hodnota absolutních odchylek;
- **SLOPE** – směrnice lineárního modelu;
- **TTEST, ZTEST, CHITEST** – statistické testy shodnosti;

ŘADU DALŠÍCH FUNKCÍ VŠAK EXCEL POSTRÁDÁ A JE TŘEBA VYUŽÍT SILNĚJŠÍHO NÁSTROJE.

Kopírování / Vkládání



- Kopírování vzorců, textů, celých sloupců (zkopírování pomocí Ctrl+C; dále „Vložit jinak...“)



- Kopírování grafů z Excelu do Wordu:
Vložit jinak → Typ: Obrázek (rozšířený metasoubor)

V. Samostatné cvičení



Úkoly – část 1



- I. Vytvořte kopii listu zadání a nazvěte ji výsledky. (**nápověda**: přesunout nebo zkopírovat list → vytvořit kopii)
- II. Pomocí roztahování buněk vyplňte proměnnou „ID“ čísla od 1 do 89.
- III. Ukotvěte ID pacientů a názvy proměnných ve sloupcích. (**nápověda**: funkce ukotvení příček)
- IV. Spojte „Jméno“ a „Příjmení“ do jednoho sloupce (např. Zdeněk Novák..) (**nápověda**: vzoreček tažením roztáhněte na celý sloupec datového souboru)
- V. Spočítejte délku hospitalizace z „první kontrola“ a „poslední kontrola“. Je získaná hodnota všude reálná? Pokud ne, tak u kterých pacientů? (**nápověda**: zkontrolujte pomocí filtru)
- VI. Pomocí vzorce vyberte pouze „Rok poslední kontroly“ ze sloupce „poslední kontrola“. Seřadte datový soubor podle této nové proměnné. (**nápověda**: vyberte funkci z Knihovny funkcí – Datum a čas)
- VII. Sloupec „nemocný“ překódujte pomocí funkce „když“ následovně: 1-nemocný, 0 –zdravý.
- VIII. Převedte „výšku“ na metry.
- IX. Vypočítejte „BMI“. (**nápověda**: vzorec pro index tělesné hmotnosti najdete na internetu)
- X. Spočítejte k jaké změně „tepu před“ a „tepu po“ došlo (např. léčbě nebo podání léku) (**nápověda**: pozor na chybějící hodnoty)

Úkoly – část 2



- XI. Spočítej „**Počet oblíbených činností**“ (sloupec U-Y).
- XII. Spočítej minimální, maximální a průměrnou hodnotu leukocytů (proměnná „**Leukocyty**“).
- XIII. Pomocí podmíněného formátování nalezněte **duplicitní záznamy** dle jména pacienta. Jsou všechny Vámi označené záznamy skutečně duplicitní? Duplicitní údaj smažte. (**nápověda**: označ sloupec → karta „Domů“ → podmíněné formátování → zvýraznit pravidla buněk → duplicitní hodnoty → filtrovat podle barvy; pozor na proměnnou, podle které hledáte duplicitní záznamy).

VI. Kontingenční tabulky v Excelu



Kontingenční tabulka



- Frekvenční sumarizace dvou kategoriálních proměnných (binárních, nominálních nebo ordinálních proměnných).
- Obecně: **R x C kontingenční tabulka** (R – počet kategorií jedné proměnné, C – počet kategorií druhé proměnné).
- Speciální případ: 2 x 2 tabulka = čtyřpolní tabulka.
- Kontingenční tabulky: **absolutních četností, celkových procent, řádkových/sloupcových četností**
- Př.: Sumarizace vyšetřených osob podle pohlaví a výsledku diagnostického testu.

Pohlaví	Výsledek vyšetření		Celkem
	Nemocný	Zdravý	
Muž	45	11	56
Žena	25	6	31
Celkem	70	17	87



Jsou více nemocní muži nebo ženy?

Ukázka kontingenční tabulky

	Nemocný	Zdravý	Celkem
Muž	45	11	56
Žena	25	6	31
Celkem	70	17	87

Kontingenční tabulka
absolutních četností

Větší počet nemocných mužů, který je dán pouze vyšším zastoupením mužů v celkovém vzorku (56 z 87)

	Nemocný	Zdravý	Celkem
Muž	80,4 %	19,6 %	100,0 %
Žena	80,6 %	19,4 %	100,0 %

Kontingenční tabulka
řádkových procent

Po výpočtu relativních četností vidíme, že se muži a ženy neliší ve výskytu onemocnění



**Jsou více nemocní
muži nebo ženy?**

Kontingenční tabulky v Excelu: zdroj dat a příprava dat



Kontingenční tabulka se dá vytvořit:

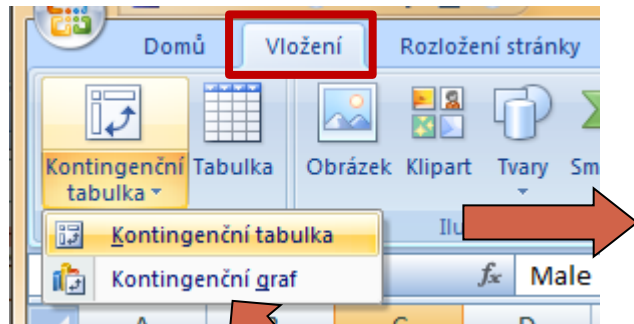
1. z tabulky v daném sešitě
2. z dat z jiného sešitu Excelu
3. z externích dat (např. MS Access)
4. ze sloučených dat z více oblastí - z různých listů nebo různých sešitů
5. z jiné kontingenční tabulky

Data musí být uspořádána formou standardního databázového seznamu:

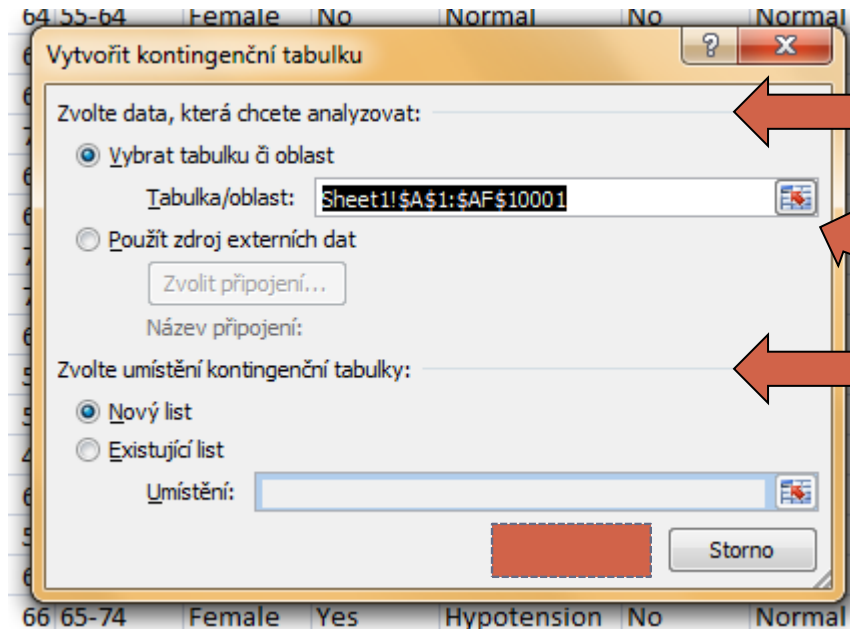
- V prvním řádku: názvy polí
- Další řádky: data

Vzhled tabulky: karta **Domů** → **Formátovat jako tabulku**

Vytvoření kontingenční tabulky v Excelu



Graf nebo tabulka



Zdroj dat (kromě Excelu i např. externí databáze)

Zdrojová oblast dat

Umístění tabulky

Kontingenční tabulky – rozvržení

Seznam polí kontingenční tabulky

Zvolte pole, které chcete přidat do sestavy:

- age
- agecat
- gender
- diabetes
- bp
- smoker
- choles
- active
- obesity
- angina
- mi
- nitro
- antictot

Přetáhnout pole mezi následujícími oblastmi:

- Filtr sestavy
- Popisky sloupců
- Popisky řádků
- Σ Hodnoty

Odložit aktualizaci rozlo... Aktualizovat

parametry, které je možné zobrazit v kontingenční tabulce

Seznam polí kontingenční tabulky

Zvolte pole, které chcete přidat do sestavy:

- age
- gender
- diabetes
- bp
- choles

Přetáhnout pole mezi následujícími oblastmi:

- Filtr sestavy
- Popisky sloupců
- Popisky řádků
- Σ Hodnoty

smoker

agecat

Počet z agecat

filtr

parametry ve sloupcích

parametry na řádcích

parametry dat

Kontingenční tabulky – nastavení II.

Kontingenční tabulka

Počet z agecat	Popisky sloupců		
Popisky řádků	No	Yes	Celkový součet
45-54	1694	501	2195
55-64	3015	863	3878
65-74	2200	661	2861
75+	816	250	1066
Celkový součet	7725	2275	10000

Seznam polí kontingenční tabulky

Zvolte pole, které chcete přidat do sestavy:

- age
- agecat
- gender
- diabetes
- bp
- smoker
- choles

Přetáhnout pole mezi nás

Filtr sestavy

Popisky řádků

agecat

- Přesunout nahoru
- Přesunout dolů
- Přesunout na začátek
- Přesunout na konec
- Přejít k filtru sestavy
- Přejít k popiskům řádků
- Přejít k popiskům sloupců
- Přejít k hodnotám
- Odstranit pole
- Nastavení polí hodnot...**

Počet z agecat

Nastavení polí hodnot

Název zdroje: agecat

Vlastní název: Počet z agecat

Souhrn

Zobrazit hodnoty jako

Kritéria shrnutí pole hodnoty

Zvolte typ kalkulačky, který chcete použít pro shrnutí dat z vybraného pole:

- Součet
- Počet**
- Průměr
- Maximum
- Minimum
- Součin

Způsob
sumarizace
položky

Formát čísla

OK

Storno

Aktualizace dat v kontingenční tabulce

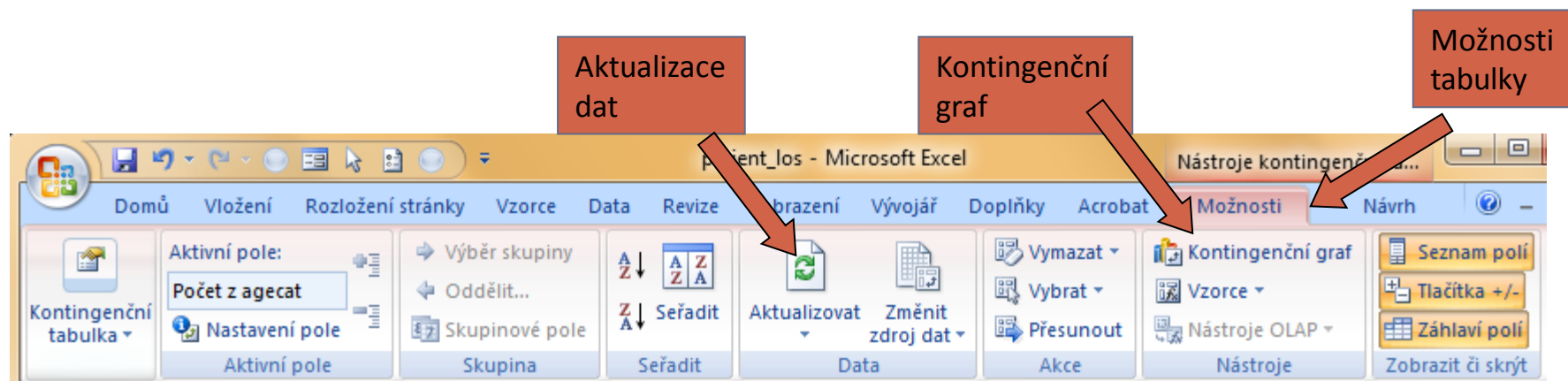


Při změně dat v tabulce se zdrojovými daty **nedojde** automaticky k aktualizaci dat v kontingenční tabulce.

Musíte provést aktualizaci dat.

1. Stůjíte kdekoliv v kontingenční tabulce
2. Na kartě **Možnosti** ve skupině **Data** klikněte na **Aktualizovat** (Alt+F5), nebo na **Aktualizovat vše** (Ctrl+Alt+F5)

Data z kontingenční tabulky lze vizualizovat pomocí **kontingenčního grafu**



Rozložení kontingenční tabulky



Po vytvoření se kontingenční tabulka zobrazí v tzv. **kompaktním formátu**. Lze ji zobrazit ale i ve formě **tabulky**, nebo ve formě **osnovy**.

1. Stůjte kdekoliv v kontingenční tabulce
2. Na kartě **Návrh** vyberte tlačítko **Rozložení sestavy** a volbu **Zobrazit ve formě osnovy nebo zobrazit ve formě tabulky**

Kompaktní formát - uspořádání tabulky aby zabírala co nejméně místa

Forma osnovy - řádková pole nižší úrovně je od vyšších úrovní odsazena, řádky nejsou odděleny čarami

Forma tabulky - klasická forma tabulky, pole nižší úrovně jsou v dalším sloupci

Vyzkoušej!