

Zásady tvorby databáze, seznamy, organizace dat, třídění, funkce, výpočty a souhrny v Excelu

Po přihlášení se do sítě (viz login name + password v okně Login) budete mít přístupný síťový disk F:\, na kterém jsou uložena data pro praktická cvičení ze statistiky.

Otevření souboru s daty pro zpracování

- V aplikaci **Tento počítač** zvolte disk F:\
- Na disku F:\ otevřete složku „SOFTWARE \ biostatistika \ data“
- Otevřete excelovský sešit „data_1.xls“

Popis struktury excelovského sešitu s daty

V excelovském sešitu „data_1.xls“ jsou na jednotlivých listech uložena data z různých klinických studií.

1. List „data preventivní prohlídky“
na listu jsou data z preventivních prohlídek zaměstnanců. V rámci prohlídek byly u zaměstnanců sledovány rizikové faktory kardiovaskulárního onemocnění a na základě těchto údajů bylo zjištěno současné riziko kardiovaskulárního onemocnění a bylo odhadnuto riziko v 60 letech věku zaměstnance. (Použité zkratky: **STK** ... systolický tlak krve, **DTK** ... diastolický tlak krve, **BMI** ... body mass index, **LDL** ... low-density lipoprotein cholesterol, **HDL** ... high-density lipoprotein cholesterol).
2. List „screeningová studie“
na listu jsou uloženy základní údaje probandů screeningové studie zaměřené na časný záchyt karcinomu prostaty u mužů nad 40 let věku.
3. List „nádory nadledvin“
na listu jsou uloženy údaje z CT vyšetření pacientů s pozitivním nálezem na nadledvinách. V sloupcích **rozměr1**, ..., **rozměr3** je zaznamenána velikost nalezeného útvaru změřená ve třech směrech.
4. List „pacienti s leukémií“
na listu jsou data uložena ve formátu datum. U každého pacienta je uvedeno datum stanovení diagnózy leukémie, datum zahájení léčby lékem Interferon a datum zahájení léčby lékem Glivec (lék Glivec byl nasazen pro nedostatečnou léčebnou odpověď na Interferon).

Databázová struktura uložených dat

Data na všech listech jsou uložena v tzv. databázovém formátu.

V 1. řádku jsou uvedeny názvy měřených parametrů a na každém dalším řádku (počínaje řádkem č. 2) jsou uloženy údaje vztahující se ke konkrétnímu pacientovi.

Pozor! : Je důležité, aby údaje vztahující se k jednomu pacientovi, byly uloženy pouze na jednom řádku. První sloupec tabulky obsahuje většinou identifikaci pacienta.

Tento formát uložení dat umožňuje použít ke zpracování analytické nástroje programu MS Excel a je také základním formátem akceptovatelným většinou statistických programů.

ÚKOL č. 1 - Třídění tabulky s daty

Na listu „data preventivní prohlídky“ seřídíte celou tabulku s daty podle definovaného kritéria:

- Podle **věku** zaměstnanců vzestupně, tj. od nejmladšího až po nejstaršího zaměstnance. Zjistěte minimum a maximum věku.
- Podle hodnot celkového **cholesterolu** sestupně, tj. od nejvyšší hodnoty po nejnižší. Zjistěte maximální hodnotu cholesterolu.
- Podle hodnot **BMI** sestupně. Zjistěte maximální hodnotu BMI.

Na listu „data preventivní prohlídky“ seřídíte celou tabulku s daty podle dvou definovaných kritérií:

- Podle **pohlaví** vzestupně a současně podle **věku** sestupně.
Tj. po seřídění bude tabulka „rozdělena“ na dva bloky – muže a ženy a každý blok bude uspořádán podle věku od nejstaršího zaměstnance po nejmladšího. Zjistěte věk nejstaršího muže a věk nejstarší ženy.
- Podle **kouření** vzestupně a současně podle hodnot **STK** (systolického krevního tlaku) sestupně.
Tj. na začátek tabulky se přesunou údaje o kuřácích (tj. **ano** ve sloupci **kouření**) a teprve potom budou následovat údaje o nekuřácích (tj. **ne** ve sloupci **kouření**). Kuřáci i nekuřáci budou uspořádáni podle hodnot STK od nejvyšší hodnoty po nejnižší. Zjistěte nejvyšší hodnotu STK u kuřáků a nejvyšší hodnotu STK u nekuřáků.

NÁVOD NA ŘEŠENÍ:

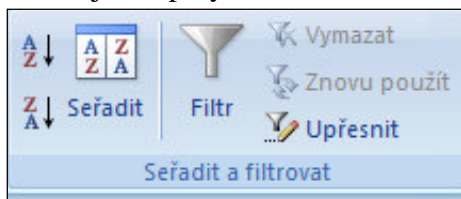
Třídění podle věku

- Na listu „data preventivní prohlídky“ klikněte do jedné z buněk sloupce **věk**.

Pozor! Při třídění tabulky nikdy neoznačujete sloupec, podle kterého chcete třídit. Pokud byste to udělali, seřídila by se pouze čísla v označeném sloupci a údaje ve zbývajících sloupcích by se třídění netýkalo - „zůstaly by stát na stejném místě“ (tj. u jednotlivých zaměstnanců by se potom objevil špatný údaj o věku).

- Z hlavního menu zvolte položku **Data**.
- Ve skupině **Seřadit a filtrovat** klikněte na ikonu , chcete-li třídit vzestupně (tj. od nejnižší hodnoty věku po nejvyšší), nebo klikněte na ikonu , (chcete-li třídit sestupně (tj. od nejvyššího věku po nejnižší). Zjistěte minimum a maximum věku.

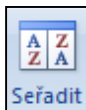
Nástroje skupiny **Seřadit a filtrovat**



- Stejným způsobem seřadíte celou tabulku s daty podle hodnot celkového cholesterolu (sloupec **cholesterol**) sestupně. Zjistíte maximální hodnotu cholesterolu.
- Seřadíte tabulku podle hodnot **BMI** sestupně. Zjistíte maximální hodnotu BMI.

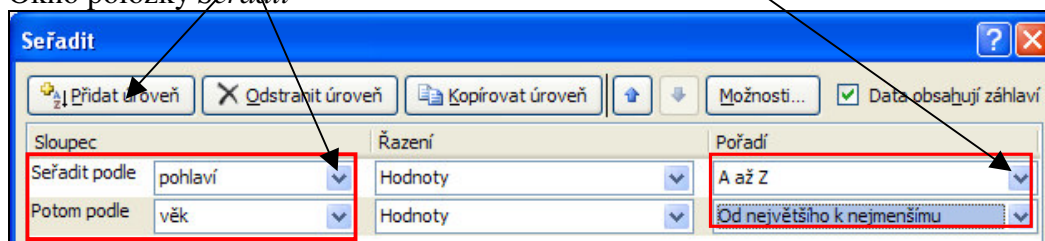
Třídění podle dvou znaků - **pohlaví** a **věk**

- Klikněte do libovolné neprázdné buňky v tabulce, kterou chcete třídit.
- Z hlavního menu zvolte položku **Data** a ve skupině **Seřadit a filtrovat** klikněte na ikonu **Seřadit**



- V poli **Seřadit podle** vyberte ze seznamu název 1. znaku, podle kterého chcete tabulku řadit – (vyberte **pohlaví**), a v poli **Pořadí** zvolte typ řazení **A až Z**, tj. vzestupně.
- Klikněte na tlačítko **Přidat úroveň** a vyberte 2. znak, podle kterého se budou data řadit ve 2. úrovni (vyberte **věk**), zvolte typ řazení **od největšího k nejmenšímu**, tj. sestupně. Potvrďte **OK**.

Okno položky **Seřadit**



Zjistíte věk nejstaršího muže a věk nejstarší ženy.

- Stejným způsobem seřadíte tabulku podle znaku **kouření** vzestupně a podle hodnot systolického krevního tlaku (sloupec **STK**) sestupně. Zjistíte nejvyšší hodnotu STK u kuřáků a nejvyšší hodnotu STK u nekuřáků.

ÚKOL č. 2 - Výpočet BMI

Na listu „*screeningová studie*“ vložte do tabulky s daty nový sloupec před sloupec *Vzdělání*. Pojmenujte tento sloupec **BMI** a vypočítejte pro každého muže hodnotu BMI.

NÁVOD NA ŘEŠENÍ:

- Klikněte na záložku listu „*screeningová studie*“.
- Vložte nový sloupec před sloupec **Vzdělání** – tj. klikněte na záhlaví sloupce E pravým tlačítkem myši a z místní nabídky zvolte **Vložit buňky**. Do buňky E1 napište **BMI**.
- K výpočtu BMI použijte údaje o tělesné výšce a tělesné hmotnosti mužů. Vzorec pro výpočet BMI:

$$BMI = \frac{hmotnost [kg]}{výška^2 [m^2]}$$

Poznámka: MS Excel nemá definovanou funkci pro výpočet BMI, vzorec pro výpočet proto musíte definovat sami.

Ve vzorci pro výpočet BMI se používá údaj o výšce v metrech. V tabulce s daty je výška zaznamenána v centimetrech a musí být převedena na metry (např. vydělena 100).

4) Klikněte do buňky E2 a napište jeden z následujících vzorců:

=D2/((C2/100)*(C2/100))

=D2/(C2/100)^2

=D2/POWER(C2/100;2)

=D2/(C2*C2/10000)

Poznámka:

Použijete-li ve vzorci místo konkrétních čísel adresy buněk s těmito čísly, můžete vzorec zkopírovat jednoduchým způsobem na zbývající řádky.

K převodu výšky na metry můžete místo C2/100 použít také zápis C2*.01

Pro zapsání znaku ^ (stříška) přepněte na anglickou klávesnici, např. levým **Alt+Shift**, znak ^ najdete na klávese s číslem 6.

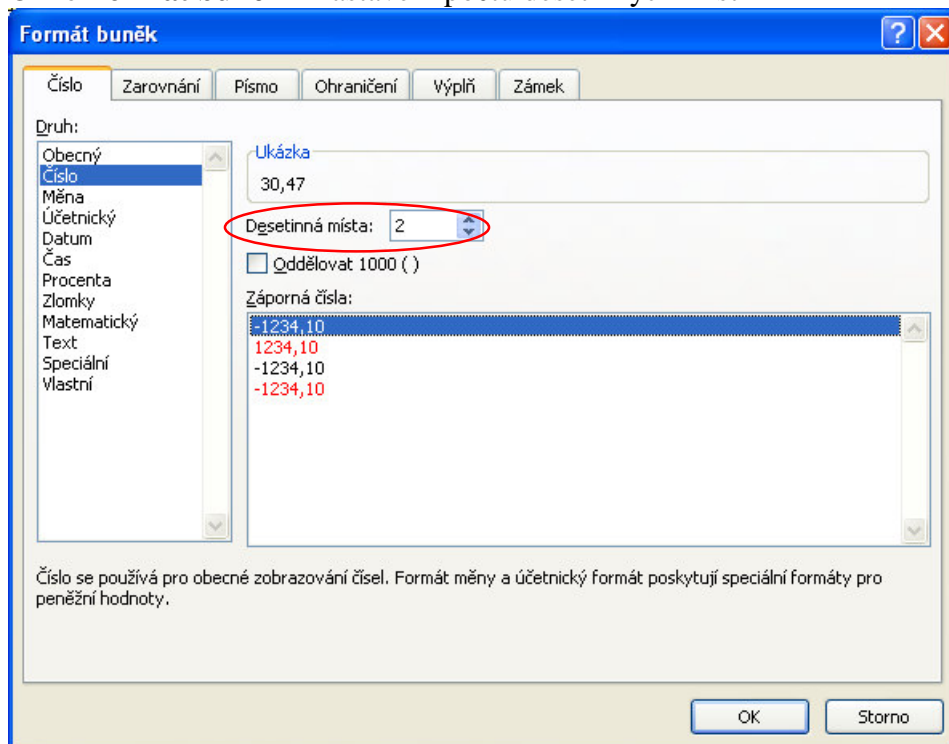
5) Výslednou hodnotu BMI zaokrouhlete na 2 desetinná místa.

6) Klikněte pravým tlačítkem myši na buňku s výsledkem a z místní nabídky zvolte **Formát buněk**,

7) Na kartě **Číslo** zvolte **Druh: Číslo** a v poli **Desetinná místa** nastavte 2.

8) Potvrďte **OK**.

Okno **Formát buněk** – nastavení počtu desetinných míst



9) Zkopírujte vzorec s výpočtem BMI na zbývající řádky.

- 10) Klikněte na buňku E2 s výpočtem a zatažením za výplňový úchyt zkopírujte výpočet i na ostatní řádky tabulky (nebo dvakrát klikněte na výplňový úchyt a tabulka se automaticky doplní).

Vložení vzorce pro výpočet BMI

fx =D2/((C2/100)*(C2/100))				
C	D	E	F	G
výška	Hmotnost	BMI	Vzdělání	Bydliště
168	86	30,47	2	4
172	69		3	1
178	72		4	3

st	BMI	Vzd
	30,47	

výplňový úchyt buňky

zatahněte dolů,
nebo 2krát klikněte

ÚKOL č. 3 - Výpočet maximální a průměrné hodnoty

Na listu „*nádory nadledvin*“ doplňte pro každého pacienta:

- maximální rozměr (z čísel ve sloupcích **rozměr1**, **rozměr2** a **rozměr3**)
- průměrný rozměr (z čísel ve sloupcích **rozměr1**, **rozměr2** a **rozměr3**).

NÁVOD NA ŘEŠENÍ:

- Klikněte na záložku listu „*nádory nadledvin*“.
- K výpočtu maximální hodnoty z čísel ve sloupci **rozměr1**, **rozměr2** a **rozměr3** použijte funkci **MAX()**
 - Klikněte kurzorem do buňky G2 a klikněte na ikonu **Vložit funkci**  na začátku řádku vzorců.

A1					
	A	B	C	D	E
1					

- V seznamu **Vybrat kategorií** zvolte **Statistické** a v seznamu **Vybrat funkci**: vyberte funkci **MAX**, potvrďte **OK**.
 - Do pole **Číslo1** napište adresu oblasti buněk D2:F2 (nebo klikněte na tlačítko na konci pole **Číslo1** a v excelovské tabulce označte kurzorem oblast buněk pro výpočet a klikněte opět na tlačítko na konci řádku s adresou). Potvrďte **OK**.
- Průměrný rozměr vypočítejte obdobným způsobem pomocí funkce **PRŮMĚR()**.

Funkci **MAX()**, resp. **PRŮMĚR()** můžete do buňky napsat také přímo z klávesnice ve tvaru

=MAX(D2:F2), resp. **=PRŮMĚR(D2:F2)**.

Poznámka: Názvy funkcí, podobně jako adresy buněk, můžete psát malými nebo velkými písmeny. Do kulatých závorek vždy napište adresu oblasti buněk, v nichž se nachází čísla pro výpočet.

- 4) Zaokrouhlete výpočet průměrné hodnoty na 1 desetinné místo.
- 5) Klikněte na buňku pravým tlačítkem myši a z místní nabídky zvolte **Formát buněk**.
- 6) Zkopírujte oba výpočty pomocí výplňového úchytu na zbývající řádky – můžete kopírovat buď každý výpočet zvlášť, nebo po označení obou buněk můžete kopírovat oba výpočty najednou.

	D	E	F	G	H	
	rozměr1 (mm)	rozměr2 (mm)	rozměr3 (mm)	maximální rozměr	průměrný rozměr	
	69	58	80	80	69,0	meta
	30	18	37			meta
	105	93	100			

zatáhněte za výplňový úchyt,
nebo na úchyt poklepejte

- 7) Čísla v buňkách s výpočty zarovnejte na střed buněk.
- 8) Označte celý sloupec G a H, z hlavního menu zvolte položku **Domů** a na kartě **Zarovnání** klikněte na ikonu **Zarovnat na střed**.

ÚKOL č. 4 - Doplnění časových údajů, počítání s údaji typu datum

Na listu „*pacienti s leukémií*“ doplňte na konec tabulky tři nové sloupce:

	A	B	C	D	E	F	G
	pac. č.	datum stanovení diagnózy	datum nasazení Interferonu	datum nasazení Glivecu	doba od Dg. do nasazení Interferonu	doba od Dg. do nasazení Glivecu	délka léčby Interferonem
1							
2	1	15.02.01	01.01.02	14.12.02			
3	2	20.12.01	15.02.02	31.01.03			

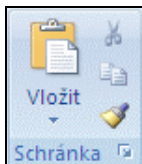
Zformátujte text v buňkách. Pro každého pacienta vypočítejte tyto časové údaje ve dnech.

NÁVOD NA ŘEŠENÍ:

- 1) Klikněte na záložku listu „*pacienti s leukémií*“
- 2) Do buňky E1 napište text „**doba od Dg. do nasazení Interferonu**“
- 3) Do buňky F1 napište text „**doba od Dg. do nasazení Glivecu**“ a
- 4) Do buňky G1 napište text „**délka léčby Interferonem**“.
- 5) Upravte formát buněk E1, F1 a G1 (nastavte tučné písmo, zalomení textu, barevnou výplň a ohraničení buněk).
 - a) Označte buňky E1, F1 a G1.
 - b) Klikněte do bloku pravým tlačítkem myši a z místní nabídky zvolte **Formát buněk**.
 - c) Z karty **Zarovnání** vyberte **Zarovnání textu vodorovně: – na střed**.
 - d) Zaklikněte volbu **Zalomit text**.

- e) Z karty **Písmo** zvolte **Řez písma: Tučné** a **Velikost: 12**.
 f) Z karty **Výplň** zvolte barvu výplně buněk.

Poznámka: K formátování můžete použít také formát konkrétní buňky. Klikněte na buňku D1 a zkopírujte její formát – tj. na kartě **Domů** klikněte na tlačítko  ve skupině **Schránka**. Potom klikněte na buňku, nebo označte oblast buněk, na kterou chcete tento formát použít.



- 6) Do buňky E2 napište vzorec **=C2-B2** (výsledkem bude počet dní od stanovení diagnózy do začátku léčby Interferonem)
- 7) Do buňky F2 napište vzorec **=D2-B2** (výsledkem bude počet dní od stanovení diagnózy do začátku léčby Glivecem)
- 8) Do buňky G2 napište vzorec **=D2-C2** (výsledkem bude počet dnů léčby Interferonem).
- 9) Označte buňky E2:G2 a zatažením za výplňový úchyt (nebo poklikáním na výplňový úchyt) zkopírujte vzorec i na ostatní řádky tabulky.

Pozor! Nezaměňujte označování buněk s kopírováním pomocí výplňového úchytu. Při označování buněk vždy klikněte dovnitř buňky.

	A	B	C	D	E	F	G
		datum stanovení diagnózy	datum nasazení Interferonu	datum nasazení Glivecu	doba od Dg. do nasazení Interferonu	doba od Dg. do nasazení Glivecu	délka léčby Interferonem
1	pac. č.						
2	1	15.02.01	01.01.02	14.12.02	320	667	347
3	2	20.12.01	15.02.02	31.01.03			
4	3	03.12.00	20.06.00	13.11.01			

zatahněte za výplňový úchyt

ÚKOL č. 5 - Výpočet základních popisných statistik pomocí funkcí Excelu

Vložte do sešitu nový list s názvem „*popisné statistiky*“, zkopírujte na tento list všechny kvantitativní znaky z listu „*data preventivní prohlídky*“.

Vypočítejte pro tyto znaky základní popisné statistiky, tj. minimum, maximum, medián, průměr, směrodatnou odchylku, počet hodnot. K výpočtu použijte definované funkce MS Excel.

TEORIE

Kvantitativní znaky popisujeme pomocí základních charakteristik polohy (průměr, medián) a charakteristik variability (směrodatná odchylka).

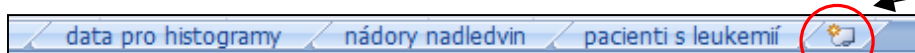
Medián je prostřední hodnota v řadě naměřených hodnot uspořádaných podle velikosti, rozděluje řadu s uspořádanými hodnotami na dvě poloviny – hodnoty v první polovině jsou menší než medián, hodnoty ve druhé polovině jsou větší než medián. Máme-li sudý počet hodnot, medián se potom vypočte jako průměr dvou prostředních hodnot řady.

Směrodatná odchylka se počítá podle vzorce $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$, kde $i=1, \dots, n$

(x_1 je 1. naměřená hodnota, x_2 je 2. naměřená hodnota, atd.), n je počet naměřených hodnot, $\bar{x}$ je průměrná hodnota. Směrodatná odchylka je vlastně střední kvadratická odchylka naměřených dat od průměrné hodnoty, vyjadřuje tedy variabilitu naměřených hodnot kolem průměru.

NÁVOD NA ŘEŠENÍ:

- 1) Vložte do sešitu nový list a pojmenujte ho „popisná statistika“ - klikněte na ikonu **Vložit list** ve spodní části obrazovky.



Klikněte pravým tlačítkem myši na název „List1“ nově vloženého listu, z místní nabídky zvolte **Přejmenovat** a napište text **popisná statistika**.

(Nový list můžete také vložit, kliknete-li pravým tlačítkem myši na název listu, před který chcete nový list vkládat, a z místní nabídky zvolíte **Vložit ...**).

- 2) Z listu „data preventivní prohlídky“ okopírujte všechny kvantitativní znaky.
- 3) Označte celý sloupec pro kopírování kliknutím na záhlaví sloupce (tj. písmeno, které daný sloupec označuje). Zkopírujte sloupce **věk**, **STK**, **DTK**, **BMI**, **cholesterol**, **LDL**, **HDL** a **Triglyceridy**.
- 4) Zkopírované sloupce vložte na list „popisné statistiky“.
- 5) Na volné části listu, vpravo od tabulky s daty, vytvořte tabulku pro výpočet popisných statistik (viz obrázek dole).
- 6) Názvy popisovaných znaků zkopírujte z tabulky s daty, do řádků tabulky zadejte názvy počítaných statistik.

H	I	J	K	L	M	N	O	P	Q	R
Triglyceridy		věk	STK	DTK	BMI	cholesterol	LDL	HDL	Triglyceridy	
1,26		minimum								
0,56		maximum								
1,01		průměr								
1,30		medián								
1,60		směrodatná odchylka								
0,55		počet hodnot								
0,81										

- 7) Vypočítejte nejdříve všechny statistiky pro znak **věk**, pro ostatní znaky vzorce zkopírujte, např. použitím výplňového úchytu buněk.
- 8) Výpočet jednotlivých statistik můžete realizovat buď přes vkládání funkcí pomocí ikony  nebo můžete funkci napsat do buňky přímo:

=MIN(A:A)

=MAX(A:A)

=PRŮMĚR(A:A)

=MEDIÁN(A:A)

=SMODCH.VÝBĚR(A:A)

=POČET(A:A)

Poznámka: A:A je adresa celého sloupce A (s hodnotami věku). Místo adresy A:A můžete také použít adresu A2:A601, což je adresa oblasti buněk, ve kterých jsou uloženy údaje o věku.

- 9) Zaokrouhlete čísla s vysokým počtem desetinných míst a upravte formát tabulky (viz postup v předchozích návodech).

Upravená tabulka s výsledky

	J	K	L	M	N	O	P	Q	R
		věk	STK	DTK	BMI	cholesterol	LDL	HDL	Triglyceridy
minimum		19	90	50	17	2,5	0,59	0,57	0,35
maximum		72	180	110	43,5	10,05	7,52	2,63	20,46
průměr		35,6	125,9	78,8	24,7	4,95	2,88	1,42	1,46
medián		32,5	120	80	24,3	4,87	2,79	1,39	1,13
směrodatná odchylka		12,0	14,2	9,2	3,9	1,03	0,90	0,35	1,31
počet hodnot		600	600	600	599	600	583	590	599

ÚKOL č. 6 - Výpočet základních popisných statistik pomocí nástroje Analýza dat

Vypočítejte základní statistické charakteristiky znaků **věk**, **STK**, **DTK**, **BMI**, **cholesterol**, **LDL**, **HDL** a **Triglyceridy** z listu „*popisné statistiky*“. K výpočtu použijte analytický modul **Analýza dat**. Výslednou tabulku upravte.

NÁVOD NA ŘEŠENÍ:

Doplňek **Analytické nástroje** je program aplikace Microsoft Office Excel, který je k dispozici po instalaci sady Microsoft Office nebo aplikace Excel. Pokud jej však chcete v aplikaci Excel používat, je nutné jej nejprve zavést.



- 1) Klikněte na tlačítko sady Microsoft Office (vlevo nahoře) a potom klikněte na tlačítko **Možnosti aplikace Excel**.
- 2) Klikněte na položku **Doplňky** a v rozevíracím seznamu **Spravovat** vyberte položku **Doplňky aplikace Excel**.
- 3) Klikněte na tlačítko **Přejít** a v seznamu **Doplňky k dispozici** zaškrtněte **Analytické nástroje**. Klikněte na tlačítko **OK**.
- 4) Na listu „*popisné statistiky*“ zvolte z hlavního menu položku **Data**, klikněte na **Analýza dat** a zvolte **Popisná statistika**.
- 5) V okně **Popisná statistika** doplňte adresu **Vstupní oblast** (na listu s daty označte celé sloupce A:H – tedy včetně názvů znaků v 1. řádku).
- 6) Zvolte **Popisky v prvním řádku** a **Celkový přehled**.
- 7) Definujte adresu **Výstupní oblast** - chcete-li výsledky na listu s daty, klikněte do pole pro definování výstupní oblasti a napište adresu buňky, do které se umístí levý horní roh výstupní tabulky. (Adresu můžete také vložit přímo, tj. kliknutím na buňku).
- 8) Potvrďte **OK**.

Popisná statistika

Vstup

Vstupní oblast:

Sdružit: ☒ Sloupce ☐ Řádky

☒ Popisky v prvním řádku

Možnosti výstupu

☒ Výstupní oblast:

☐ Nový list:

☐ Nový sešit:

☒ Celkový přehled

☐ Hladina spolehlivosti pro stř. hodnotu: %

☐ K-té největší:

☐ K-té nejmenší:

OK Storno nápověda

Výsledná tabulka:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	věk		STK		DTK		BMI		cholesterol		LDL		HDL		Triglyceridy
2															
3	Stř. hodno	35,56833	Stř. hodno	125,9417	Stř. hodno	78,8	Stř. hodno	24,72053	Stř. hodno	4,946483	Stř. hodno	2,880823	Stř. hodno	1,419712	Stř. hodno
4	Chyba stř.	0,488696	Chyba stř.	0,578176	Chyba stř.	0,376464	Chyba stř.	0,158676	Chyba stř.	0,041887	Chyba stř.	0,037116	Chyba stř.	0,014365	Chyba stř.
5	Medián	32,5	Medián	120	Medián	80	Medián	24,3	Medián	4,87	Medián	2,79	Medián	1,39	Medián
6	Modus	26	Modus	120	Modus	80	Modus	24	Modus	5,02	Modus	2,3	Modus	1,46	Modus
7	Směr. odc	11,97055	Směr. odc	14,16236	Směr. odc	9,221446	Směr. odc	3,883517	Směr. odc	1,026011	Směr. odc	0,896177	Směr. odc	0,348918	Směr. odc
8	Rozptyl vý	143,2942	Rozptyl vý	200,5726	Rozptyl vý	85,03506	Rozptyl vý	15,0817	Rozptyl vý	1,052699	Rozptyl vý	0,803133	Rozptyl vý	0,121744	Rozptyl vý
9	Špičatost	-0,64995	Špičatost	1,328986	Špičatost	0,890217	Špičatost	1,408469	Špičatost	1,352615	Špičatost	1,609665	Špičatost	0,157935	Špičatost
10	Šikmost	0,613101	Šikmost	0,837917	Šikmost	0,230501	Šikmost	0,868872	Šikmost	0,710912	Šikmost	0,79082	Šikmost	0,515566	Šikmost
11	#REF!	53	#REF!	90	#REF!	60	#REF!	26,5	#REF!	7,55	#REF!	6,93	#REF!	2,06	#REF!
12	Minimum	19	Minimum	90	Minimum	50	Minimum	17	Minimum	2,5	Minimum	0,59	Minimum	0,57	Minimum
13	Maximum	72	Maximum	180	Maximum	110	Maximum	43,5	Maximum	10,05	Maximum	7,52	Maximum	2,63	Maximum
14	Součet	21341	Součet	75565	Součet	47280	Součet	14807,6	Součet	2967,89	Součet	1679,52	Součet	837,63	Součet
15	Počet	600	Počet	600	Počet	600	Počet	599	Počet	600	Počet	583	Počet	590	Počet
16															

Program **Analytické nástroje** počítá, kromě základních popisných charakteristik polohy a variability dat, také tzv. charakteristiky rozdělení (distribuce) dat.

Šikmost vyjadřuje, jak jsou hodnoty symetricky či asymetricky rozloženy kolem střední hodnoty, zda v souboru převládají spíše nízké hodnoty (podprůměrné) nebo vysoké hodnoty (nadprůměrné). Je-li hodnota šikmosti kladná, převládají nízké hodnoty, je-li záporná, převládají vysoké hodnoty. Hodnoty blízké nule znamenají, že jsou hodnoty rozloženy rovnoměrně.

Špičatost vyjadřuje, jak jsou hodnoty koncentrovány kolem střední hodnoty, zda převládají spíše hodnoty blízké střední hodnotě nebo hodnoty odlišné od střední hodnoty. Je-li hodnota špičatosti kladná, jsou hodnoty koncentrovány kolem středu – tj. průměru nebo mediánu. Je-li hodnota záporná, hodnoty nejsou koncentrovány kolem středu. Je-li hodnota blízká nule, jsou hodnoty rozloženy rovnoměrně.

Statistika **rozptyl** je charakteristikou variability dat (je to druhá mocnina směrodatné odchylky).

- Upravte výslednou tabulku - ze všech vypočítaných statistik můžete nechat jen ty nejdůležitější a můžete také odstranit sloupce, ve kterých se duplikuje název

počítaných statistik. Upravte formát číselných hodnot, tj. nastavte formát s menším počtem desetinných míst.

Úpravy tabulky můžete začít tím, že před buňku s textem **věk** vložíte jednu prázdnou buňku – tím dosáhnete toho, že se názvy znaků přesunou nad sloupce s číselnými údaji.

10) Klikněte pravým tlačítkem myši na buňku s textem **věk**.

11) Z místní nabídky zvolte **Vložit buňky** a potvrďte **Posunout buňky vpravo**.

12) Klikněte na **OK**.

Názvy počítaných statistik nechejte pouze v prvním sloupci tabulky, z ostatních sloupců je můžete odstranit.

13) Označte všechny buňky, které chcete odstranit, klikněte do označené oblasti pravým tlačítkem myši a zvolte **Odstranit**.

14) Potvrďte **Posunout buňky vlevo**, klikněte na **OK**.

15) Zaokrouhlete čísla s vysokým počtem desetinných míst.

Upravená tabulka s výsledky by mohla vypadat takto:

	věk	STK	DTK	BMI	cholesterol	LDL	HDL	Triglyceridy
Stř. hodnota	35,57	125,94	78,80	24,72	4,95	2,88	1,42	1,46
Chyba stř. hodnoty	0,489	0,578	0,376	0,159	0,042	0,037	0,014	0,053
Medián	32,5	120	80	24,3	4,87	2,79	1,39	1,13
Modus	26	120	80	24	5,02	2,3	1,46	1,03
Směr. odchylka	11,97	14,16	9,22	3,88	1,03	0,90	0,35	1,31
Rozptyl výběru	143,29	200,57	85,04	15,08	1,05	0,80	0,12	1,71
Špičatost	-0,65	1,33	0,89	1,41	1,35	1,61	0,16	82,77
Šikmost	0,61	0,84	0,23	0,87	0,71	0,79	0,52	6,97
Minimum	19	90	50	17	2,5	0,59	0,57	0,35
Maximum	72	180	110	43,5	10,05	7,52	2,63	20,46
Počet	600	600	600	599	600	583	590	599

ÚKOL č. 7 - Konstrukce jednoduchých četnostních tabulek

Z dat na listu „data preventivní prohlídky“ vytvořte jednoduchou četnostní tabulku pro znak **kouření**, **pohlaví**, **věková kategorie**, **BMI hodnocení** a **cholesterol hodnocení** – tj. zjistěte počet a procento zaměstnanců v jednotlivých kategoriích daného znaku (tj. počet a procento kuřáků a nekuřáků, počet a procento mužů a žen, atd.).

TEORIE:

Základním nástrojem popisu kvalitativních (kategoriálních) znaků jsou četnostní tabulky, které popisují rozdělení četností naměřených hodnot v jednotlivých kategoriích znaku. V programu MS Excel lze tyto tabulky vytvořit pomocí nástroje **Kontingenční tabulka**.

NÁVOD NA ŘEŠENÍ:

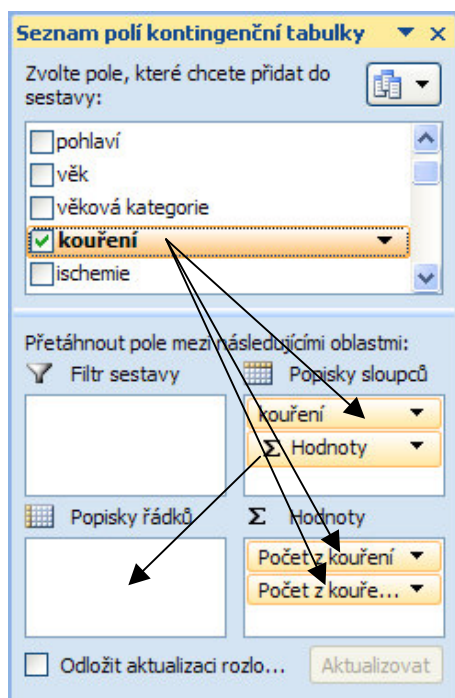
Zjištění počtu a procenta kuřáků a nekuřáků v souboru.

- 1) Klikněte na libovolnou buňku s daty v tabulce na listu "*data preventivní prohlídky*".
- 2) Z hlavního menu vyberte položku **Vložení** a klikněte na tlačítko **Kontingenční tabulka**.
Otevře se okno **Vytvořit kontingenční tabulku**.
- 3) V okně **Vytvořit kontingenční tabulku** se definuje adresa oblasti s daty a volí se umístění kontingenční tabulky (na nový nebo existující list).
V poli **Tabulka/oblast** nechejte adresu celé tabulky, která se do pole doplnila automaticky. Novou tabulku umístěte na **Nový list**.
- 4) Klikněte na **OK**.
- 5) Do sešitu se vloží nový list, který obsahuje schéma prázdné kontingenční tabulky a okno **Seznam polí kontingenční tabulky** se seznamem všech znaků ze specifikované oblasti dat.
Vytvořte kontingenční tabulku pro znak **kouření**.
- 6) Pole **kouření** přetáhněte z okna **Seznam polí kontingenční tabulky** do oblasti **Popisy sloupců**.

Poznámka: Přetáhnete-li pole **kouření** do oblasti **Popisy řádků**, objeví se popis kategorií "ne=nekuřák" a "ano=kuřák" v řádcích výsledné tabulky.

- 7) Pole **kouření** přetáhněte 2 krát do oblasti Σ **Hodnoty**.

Poznámka: Výsledná tabulka by měla obsahovat dva údaje: 1. absolutní četnosti (tj. počet kuřáků a nekuřáků) a také 2. relativní četnosti (tj. procento kuřáků a nekuřáků v celém souboru) – proto se do oblasti Σ **Hodnoty** přetahuje pole 2 krát.



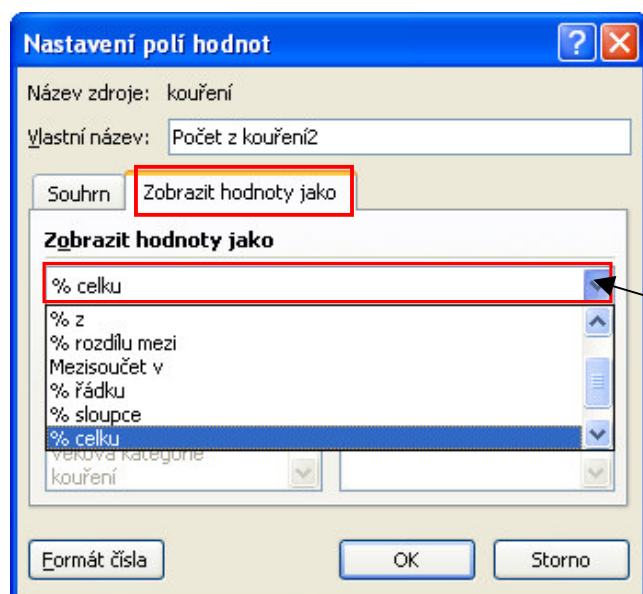
přetáhněte myší

- 8) Tlačítko Σ **Hodnoty** přetáhněte z oblasti **Popisky sloupců** do oblasti **Popisky řádků** (viz obrázek na předcházející straně).
- 9) Nastavte v tabulce výpočet procent.
- 10) Klikněte pravým tlačítkem myši na jedno z čísel v tabulce a z místní nabídky zvolte **Nastavení polí hodnot**.

	A	B	C	D
1				
2				
3		Popisky sloupců		
4	Hodnoty	ne	ano	Celkový součet
5	Počet z kouření	408	192	600
6	Počet z kouření2	408	192	600
7				

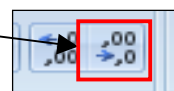
- 11) Klikněte na záložku **Zobrazit hodnoty jako** a v poli **Zobrazit hodnoty jako** vyberte **% celku**. Klikněte na **OK**.

Nastavení výpočtu procent



ze seznamu vyberte
% celku

- 12) Zaokrouhlete procenta na jedno desetinné číslo.
- 13) Označte buňky s procenty, klikněte na položku **Domů** v hlavním menu a ze skupiny **Číslo** klikněte na tlačítko **Odebrat desetinné místo**



- 14) Nastavte klasické rozložení kontingenční tabulky (v tabulce se místo textu **Popisky sloupců** zobrazí název pole **kouření**).
- 15) Klikněte pravým tlačítkem do tabulky a z místního menu zvolte **Možnosti kontingenční tabulky**.
- 16) Klikněte na **Zobrazit** a zaklikněte **Klasické rozložení kontingenční tabulky**.
- 17) Klikněte na **OK**.

18) Výslednou tabulku můžete dále upravit:

Místo textu **Počet z kouření** napište **počet**, místo **Počet z kouření2** napište **procento**.

Místo textu **ne** napište **nekuřáci** a místo **ano** napište **kuřáci**.

Upravená kontingenční tabulka pro znak **kouření**

kouření ▼			
Hodnoty	ne	ano	Celkový součet
Počet	408	192	600
Procento	68,0%	32,0%	100,0%

Závěr: Z tabulky lze vyčíst, že v souboru 600 zaměstnanců je 192 kuřáků (tj. 32,0 %) a 408 nekuřáků (tj. 68,0 %).

19) Klikněte opět na libovolnou buňku s daty na listu "*data preventivní prohlídky*" a stejným postupem vytvořte kontingenční tabulku pro znak **pohlaví**. Tabulku umístěte na list s první kontingenční tabulkou.

Poznámka: Chcete-li výslednou kontingenční tabulku umístit na existující list, zvolte umístění **Existující list** a potom klikněte na záložku se jménem listu, do kterého chcete tabulku umístit. V tomto listu klikněte do prázdné buňky, kam se umístí levý horní roh výsledné tabulky.

20) Vytvořte kontingenční tabulku pro znak **věková kategorie**.

21) Vytvořte kontingenční tabulku pro znak **BMI hodnocení**.

Poznámka: Při tvorbě nové kontingenční tabulky vždy nejdříve otevřete list, na kterém se nachází data, z kterých chcete kontingenční tabulku vytvořit.