

MUNI
MED

BIOSTATISTIKA

Osnova

- Excel: opakování, příprava dat, základní vzorce
- Základy popisné statistiky
- Základní rozdělení pravděpodobnosti, testování hypotéz
- Parametrické testy
- Neparametrické testy
- Analýza kontingenčních tabulek
- Základy korelační analýzy a lineární regrese

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.

Příprava dat, MS Excel

Datová tabulka

Zásady správné tvorby dat

Možnosti MS Excel

Ukázka datového souboru

Parametry (znaky)

Základní jednotka dat!

| | A | AC | AD | AE | AF | AG | AH | AI | AJ |
|----|----|---------------|---------------|--------|---------------|--------------|---------------|---------------|----------------|
| 1 | ID | obvod_pasu_po | obvod_boku_po | WHR_po | WHR_riziko_po | syst_tlak_po | diast_tlak_po | hypertenze_po | cholesterol_po |
| 2 | 1 | 86,7 | 98,6 | 0,88 | 1 | 103 | 68 | 0 | 3,82 |
| 3 | 2 | 70,3 | 82,9 | 0,85 | 1 | 118 | 75 | 0 | 6,18 |
| 4 | 3 | 61,2 | 88,3 | 0,69 | 1 | 114 | 74 | 0 | 3,90 |
| 5 | 4 | 81,6 | 87,3 | 0,93 | 1 | 127 | 73 | 0 | 5,06 |
| 6 | 5 | 89,2 | 104,2 | 0,86 | 3 | 135 | 99 | 1 | 6,24 |
| 7 | 6 | 74,2 | 100,1 | 0,74 | 1 | 111 | 81 | 0 | 3,44 |
| 8 | 7 | 114,2 | 108,5 | 1,05 | 3 | 136 | 80 | 0 | 4,17 |
| 9 | 8 | 65,1 | 82,5 | 0,79 | 1 | 118 | 98 | 1 | 2,87 |
| 10 | 9 | 81,5 | 79,0 | 1,03 | 3 | 116 | 87 | 0 | 4,20 |
| 11 | 10 | 75,9 | 124,9 | 0,61 | 1 | 125 | 82 | 0 | 4,12 |
| 12 | 11 | 89,8 | 111,6 | 0,80 | 2 | 108 | 84 | 0 | 2,83 |
| 13 | 12 | 67,8 | 87,9 | 0,77 | 1 | 120 | 64 | 0 | 4,71 |
| 14 | 13 | 95,7 | 92,7 | 1,03 | 3 | 169 | 112 | 1 | 4,67 |
| 15 | 14 | 86,1 | 88,5 | 0,97 | 2 | 111 | 82 | 0 | 4,91 |
| 16 | 15 | 86,4 | 101,2 | 0,85 | 1 | 108 | 79 | 0 | 4,99 |
| 17 | 16 | 68,1 | 86,3 | 0,79 | 1 | 122 | 81 | 0 | 4,83 |
| 18 | 17 | 65,0 | 77,9 | 0,83 | 1 | 139 | 81 | 0 | 4,17 |
| 19 | 18 | 92,5 | 72,8 | 1,27 | 3 | 143 | 77 | 1 | 5,89 |
| 20 | 19 | 69,4 | 81,6 | 0,85 | 2 | 130 | 90 | 0 | 5,90 |
| 21 | 20 | 93,7 | 90,8 | 1,03 | 3 | 136 | 83 | 0 | 4,64 |
| 22 | 21 | 71,4 | 83,5 | 0,86 | 3 | 123 | 83 | 0 | 4,75 |
| 23 | 22 | 95,0 | 95,4 | 1,00 | 3 | 141 | 91 | 1 | 4,18 |

Zásady pro ukládání dat

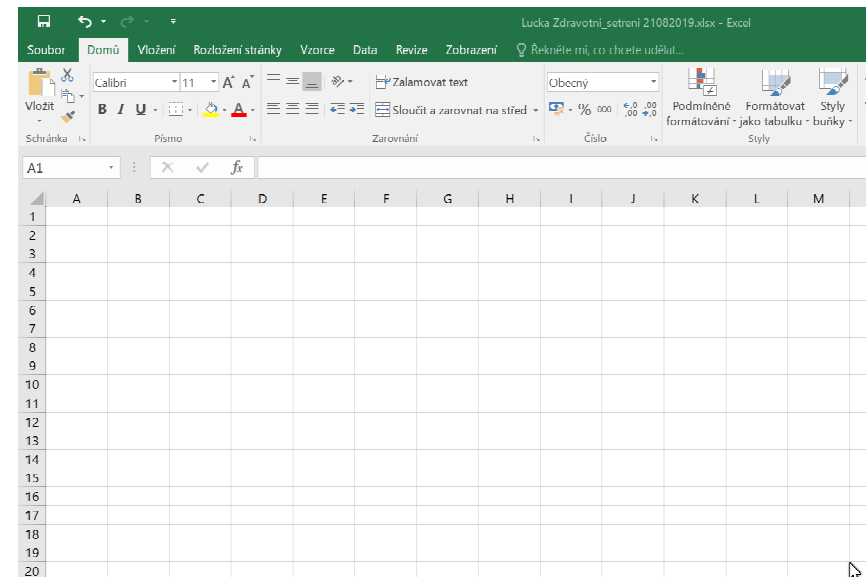
- Správné a přehledné uložení dat je základem jejich pozdější analýzy.
- Je vhodné rozmyslet si předem jak budou data ukládána.
- Pro počítačové zpracování dat je nezbytné ukládat data v tabulární formě.
- Nejvhodnějším způsobem je uložení dat ve formě databázové tabulky.
- Takto uspořádaná data je v tabulkových nebo databázových programech možné převést na libovolnou výstupní tabulku.
- Pro základní uložení a čištění dat menšího rozsahu je možné využít aplikací MS Excel.

Zásady pro ukládání dat

- 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 je nezbytné kontrolovat překlepy v názvech kategorií;
- Specifickým typem dat jsou data, u nichž je nezbytné kontrolovat, zda jsou uloženy v korektním formátu.

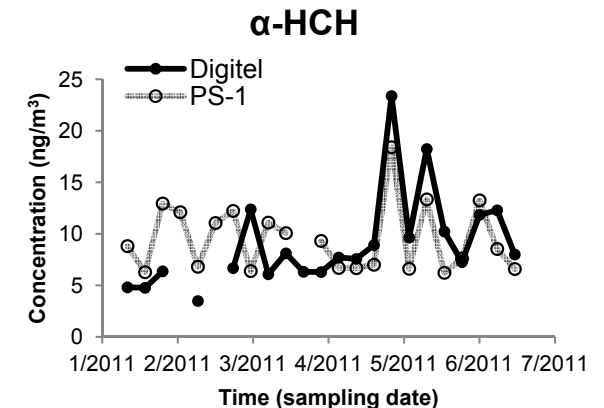
MS Excel

- Tabulkový procesor.
- Aktualizace každé 2 až 3 roky; nové funkce, rozšíření počtu řádků a sloupců, změna formátu.
- Starší formát: .xls, novější: .xlsx.
- Aktuální verze 2016 umožňuje ukládat tabulku o 1 048 576 řádcích a 16 384 sloupcích.



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.



| | | |
|----|--------------------|-----|
| 16 | | |
| 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 | | | |
|---------------|----------------|-----------|------|---|
| Počet z Délka | | Pohlaví | | |
| Číslo | ryby2 | Číslo rvt | Váha | ? |
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | 26 | | | |
| 8 | 106 | | | |
| 9 | 121 | | | |
| 10 | 160 | | | |
| 11 | 34 | | | |
| 12 | 45 | | | |
| 13 | 70 | | | |
| 14 | 72 | | | |
| 15 | 67 | | | |
| 16 | Celkový součet | | | |
| 17 | | | | |

(Zobrazit vše)
 68
 99
 102
 109
 112
 120
 173
 28
 29

Import a export dat

Import dat

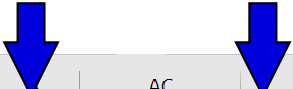
- Manuální zadávání
- Import – podpora importu ze starších verzí Excelu, textových souborů, databází apod.
- Kopírování přes schránku Windows – vkládání z nejrůznějších aplikací – MS Office, Statistica atd.

Export dat

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

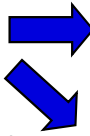
Databázová struktura dat v Excelu

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



| | AC | AD | AE | AF | AG | AH | AI | AJ | |
|----|----|---------------|---------------|--------|---------------|--------------|---------------|---------------|----------------|
| 1 | ID | obvod_pasu_po | obvod_boku_po | WHR_po | WHR_riziko_po | syst_tlak_po | diast_tlak_po | hypertenze_po | cholesterol_po |
| 2 | 1 | 86,7 | 98,6 | 0,88 | 1 | 103 | 68 | 0 | 3,82 |
| 3 | 2 | 70,3 | 82,9 | 0,85 | 1 | 118 | 75 | 0 | 6,18 |
| 4 | 3 | 61,2 | 88,3 | 0,69 | 1 | 114 | 74 | 0 | 3,90 |
| 5 | 4 | 81,6 | 87,3 | 0,93 | 1 | 127 | 73 | 0 | 5,06 |
| 6 | 5 | 89,2 | 104,2 | 0,86 | 3 | 135 | 99 | 1 | 6,24 |
| 7 | 6 | 74,2 | 100,1 | 0,74 | 1 | 111 | 81 | 0 | 3,44 |
| 8 | 7 | 114,2 | 108,5 | 1,05 | 3 | 136 | 80 | 0 | 4,17 |
| 9 | 8 | 65,1 | 82,5 | 0,79 | 1 | 118 | 98 | 1 | 2,87 |
| 10 | 9 | 81,5 | 79,0 | 1,03 | 3 | 116 | 87 | 0 | 4,20 |
| 11 | 10 | 75,9 | 124,9 | 0,61 | 1 | 125 | 82 | 0 | 4,12 |
| 12 | 11 | 89,8 | 111,6 | 0,80 | 2 | 108 | 84 | 0 | 2,83 |
| 13 | 12 | 67,8 | 87,9 | 0,77 | 1 | 120 | 64 | 0 | 4,71 |
| 14 | 13 | 95,7 | 92,7 | 1,03 | 3 | 169 | 112 | 1 | 4,67 |
| 15 | 14 | 86,1 | 88,5 | 0,97 | 2 | 111 | 82 | 0 | 4,91 |
| 16 | 15 | 86,4 | 101,2 | 0,85 | 1 | 108 | 79 | 0 | 4,99 |
| 17 | 16 | 68,1 | 86,3 | 0,79 | 1 | 122 | 81 | 0 | 4,83 |
| 18 | 17 | 65,0 | 77,9 | 0,83 | 1 | 139 | 81 | 0 | 4,17 |
| 19 | 18 | 92,5 | 72,8 | 1,27 | 3 | 143 | 77 | 1 | 5,89 |
| 20 | 19 | 69,4 | 81,6 | 0,85 | 2 | 130 | 90 | 0 | 5,90 |
| 21 | 20 | 93,7 | 90,8 | 1,03 | 3 | 136 | 83 | 0 | 4,64 |
| 22 | 21 | 71,4 | 83,5 | 0,86 | 3 | 123 | 83 | 0 | 4,75 |
| 23 | 22 | 95,0 | 95,4 | 1,00 | 3 | 141 | 91 | 1 | 4,88 |

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



Excel neumožňuje pojmenování řádků a sloupců vlastními názvy.

Typy 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 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

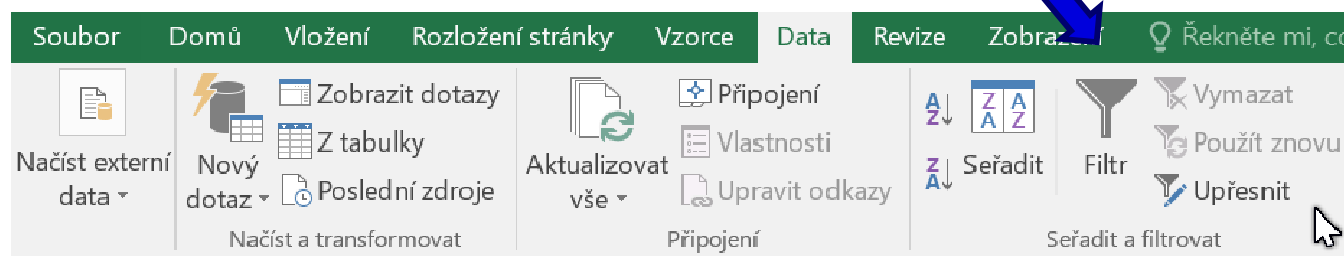
Myš a okraje buňky

- 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)

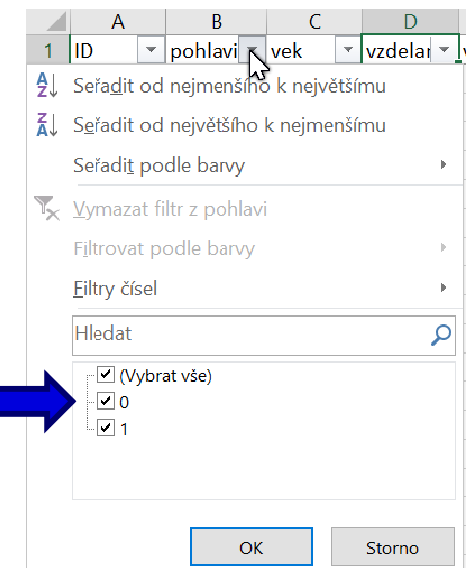
Automatický filtr

- 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**)



Výběr hodnot pro filtraci



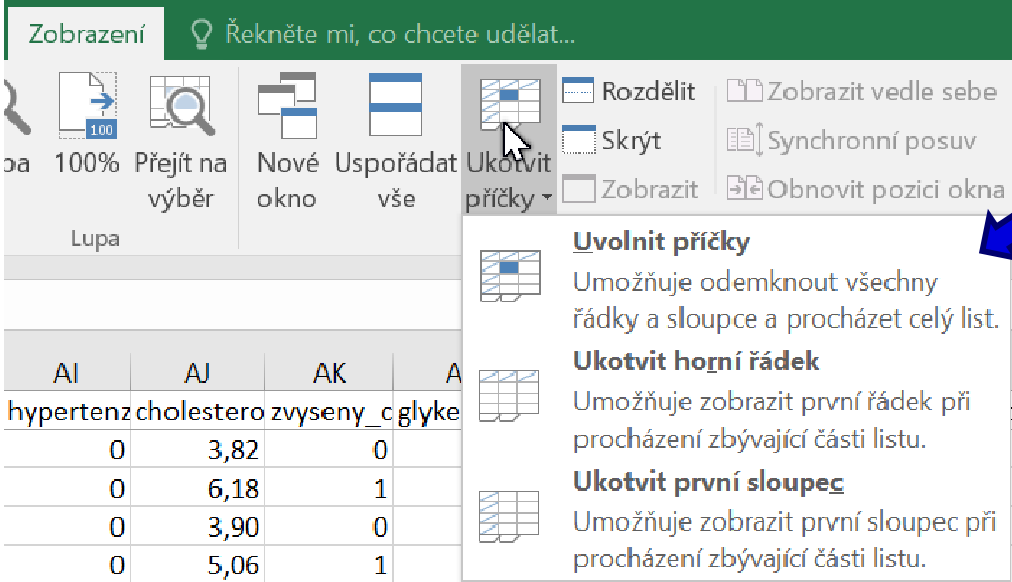
2. Objeví se rozbalovací šipka s výčtem všech unikátních hodnot v daném sloupci dat

The image shows an Excel table with columns A through E. A blue arrow points to the dropdown arrow in the 'pohlaví' column header. The table contains the following data:

| | A | B | C | D | E |
|---|----|---------|-----|---------|-------|
| 1 | ID | pohlaví | vek | vzdelai | vyska |
| 2 | 1 | 0 | 55 | 1 | 182 |
| 3 | 2 | 0 | 56 | 2 | 169 |
| 4 | 3 | 1 | 59 | 3 | 169 |

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

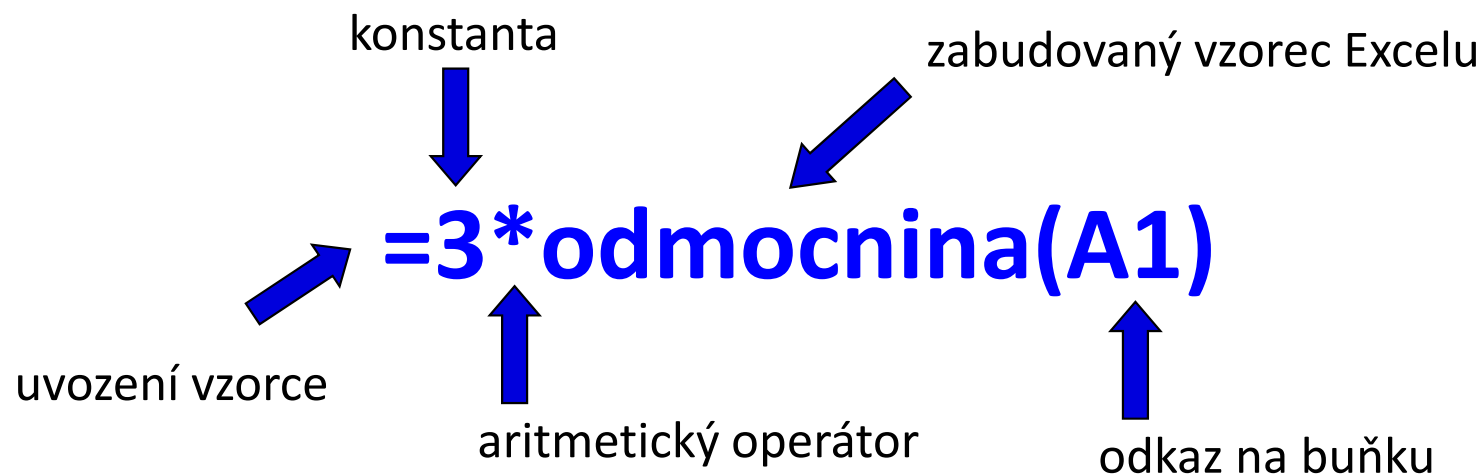


Ukotví řádky nad označenou buňkou a sloupce vlevo od označené buňky

| AI | AJ | AK | A |
|-----------|-----------|---------|-------|
| hypertenz | cholester | zvyseny | glyke |
| 0 | 3,82 | 0 | |
| 0 | 6,18 | 1 | |
| 0 | 3,90 | 0 | |
| 0 | 5,06 | 1 | |

Vzorce

- vpisují se do buněk sešitu
- vzorce jsou vždy uvozeny = (lze též + -)
- aritmetické operátory + zabudované funkce Excelu
- pro „sčítání“ nečíselných položek se používá &
- 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

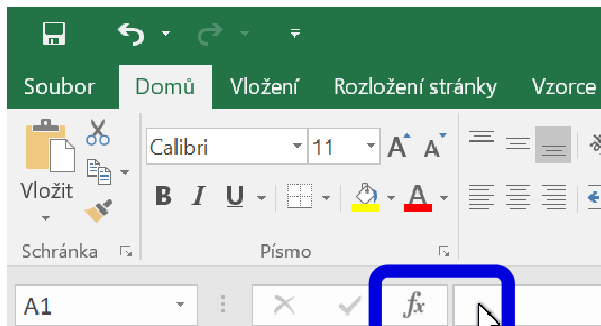
Absolutní odkaz

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

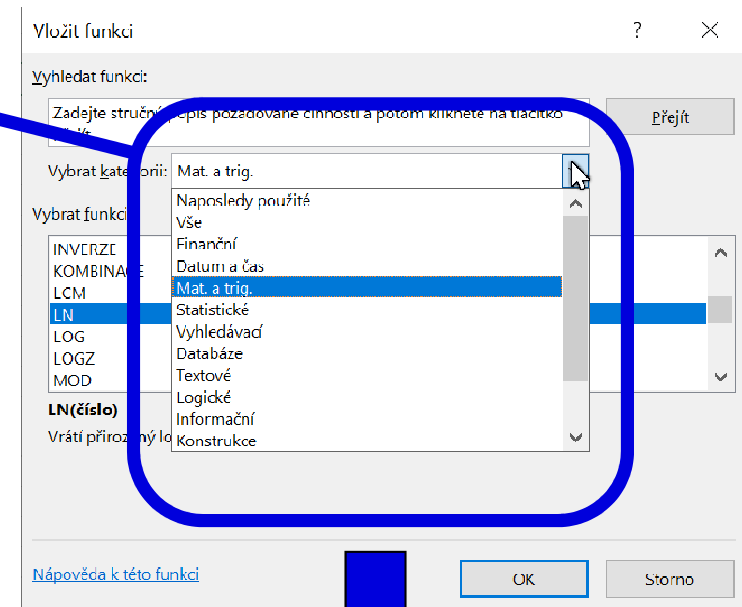
uzamčení řádku → **\$A\$1** ← uzamčení sloupce

Pamatuj: Adresu upevníme pomocí znaku **\$** (klávesa **F4**)

Vzorce – využití seznamu vzorců

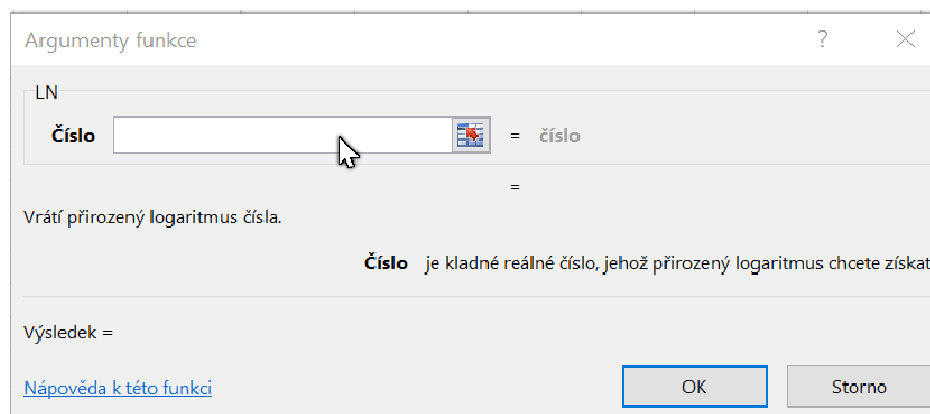


Kategorie vzorců



Funkce a její stručný popis

Průvodce funkcí



Vzorce – užitečné funkce

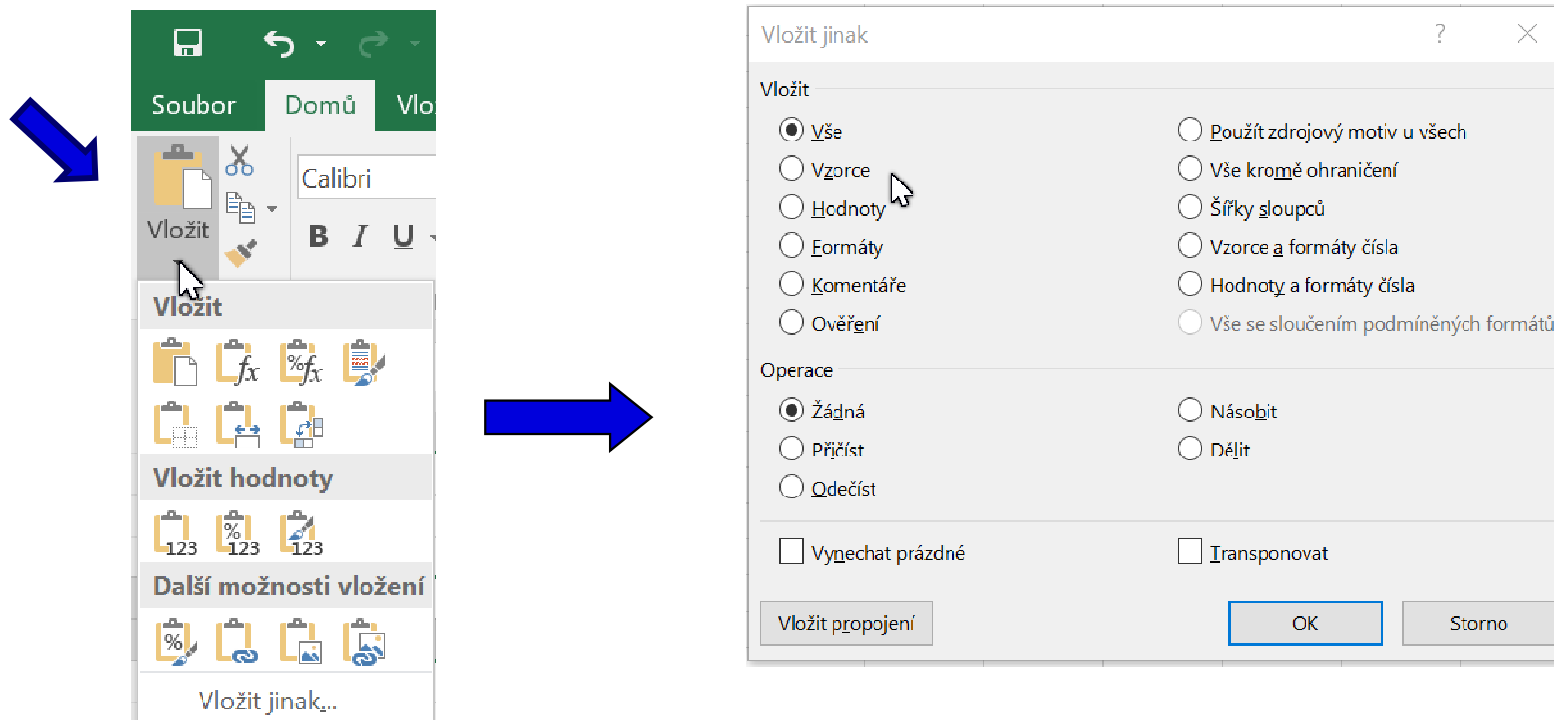
- **SUMA** – součet číselných hodnot oblasti;
- **SUMIF** – podmíněný součet (podmínky v doplňkové oblasti);
- **PRŮMĚR** – aritmetický průměr číselných hodnot oblasti;
- **GEOMEAN** – geometrický průměr číselných hodnot oblasti;
- **COUNTIF** – počet hodnot oblasti splňujících zadanou podmínku;
- **KDYŽ**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

Statistické funkce v MS Excel

- **CONFIDENCE.NORM** – výpočet intervalu spolehlivosti (při normálním rozdělení);
- **CORREL, PEARSON** – výpočet Pearsonova korelačního koeficientu;
- **COVARIANCE.S** – výpočet kovariance dvou množin dat;
- **COUNTIF** – počet hodnot oblasti splňujících zadanou podmínku;
- **DEVSQ** – součet čtverců odchylek od výběrového průměru;
- **F.DIST, GAMMA.DIST, T.DIST, NORM.DIST** aj. – různá rozdělení pravděpodobnosti;
- **PRŮMODCHYLKA** – průměrná hodnota absolutních odchylek;
- **SLOPE** – směrnice lineárního modelu;
- **T.TEST, Z.TEST, CHISQ.TEST** – statistické testy shodnosti;
- ŘADU DALŠÍCH FUNKCÍ VŠAK EXCEL POSTRÁDÁ A JE TŘEBA VYUŽÍT SILNĚJŠÍHO NÁSTROJE.

Kopírování a vkládání

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



MUNI
MED

Praktické cvičení



Datový soubor

Rehabilitace po mozkovém infarktu

| | A | B | C | D | E | F | G | H | |
|----|----|---------|-----|---------------------|----------------|-------------------------------|----------------------------|-------------------------------------|---------------|
| 1 | ID | Pohlaví | Věk | Etiologie | Lokalizace | Terapie | Komorbidity_k omplikace | Barthel_index_pr ed rehabilitaci | Katego red |
| 2 | 1 | muž | 82 | okluze nebo stenóza | mozkové tepny | jiná farmakologická terapie | 0 | 25 | vysoce |
| 3 | 2 | žena | 81 | embolie | mozkové tepny | jiná farmakologická terapie | 2 | 20 | vysoce |
| 4 | 3 | muž | 55 | okluze nebo stenóza | mozkové tepny | jiná farmakologická terapie | 0 | 35 | vysoce |
| 5 | 4 | žena | 46 | embolie | mozkové tepny | intravenózní trombolýza rt-PA | 0 | 20 | vysoce |
| 6 | 5 | muž | 76 | okluze nebo stenóza | mozkové tepny | jiná farmakologická terapie | 0 | 45 | částečn |
| 7 | 6 | muž | 72 | okluze nebo stenóza | mozkové tepny | jiná farmakologická terapie | 0 | 25 | vysoce |
| 8 | 7 | muž | 62 | trombóza | mozkové tepny | jiná farmakologická terapie | 0 | 40 | vysoce |
| 9 | 8 | muž | 64 | trombóza | přívodní tepny | jiná farmakologická terapie | 0 | 15 | vysoce |
| 10 | 9 | žena | 82 | okluze nebo stenóza | mozkové tepny | jiná farmakologická terapie | 0 | 10 | vysoce |
| 11 | 10 | muž | 58 | trombóza | mozkové tepny | jiná farmakologická terapie | 0 | 25 | vysoce |
| 12 | 11 | muž | 84 | okluze nebo stenóza | mozkové tepny | jiná farmakologická terapie | 0 | 40 | vysoce |
| 13 | 12 | žena | 92 | okluze nebo stenóza | mozkové tepny | jiná farmakologická terapie | 0 | 30 | vysoce |
| 14 | 13 | žena | 79 | embolie | mozkové tepny | jiná farmakologická terapie | 1 | 40 | vysoce |
| 15 | 14 | muž | 69 | trombóza | mozkové tepny | jiná farmakologická terapie | 3 | 45 | částečn |
| 16 | 15 | muž | 67 | okluze nebo stenóza | mozkové tepny | mechanická trombektomie | 0 | 25 | vysoce |
| 17 | 16 | žena | 70 | trombóza | přívodní tepny | mechanická trombektomie | 0 | 40 | vysoce |
| 18 | 17 | žena | 59 | trombóza | mozkové tepny | jiná farmakologická terapie | 0 | 25 | vysoce |
| 19 | 18 | žena | 63 | okluze nebo stenóza | přívodní tepny | jiná farmakologická terapie | 0 | 40 | vysoce |

Rehabilitace po mozkovém infarktu

- Cvičný datový soubor obsahuje záznamy o **celkem 407 pacientech hospitalizovaných pro mozkový infarkt** na neurologickém oddělení akutní péče, kde jim byla poskytnuta terapie pro obnovu krevního oběhu v postižené části mozku.
- Po zvládnutí akutní fáze byl u pacientů vyhodnocen stupeň soběstačnosti v základních denních aktivitách (ADL) pomocí tzv. **indexu Barthelové (BI)** a byli přeloženi na **rehabilitační oddělení**.
- Po dvou týdnech byl opět dle BI vyhodnocen stupeň soběstačnosti a pacienti byli buď propuštěni do ambulantní péče, nebo přeloženi na oddělení následné péče.

Rehabilitace po mozkovém infarktu

Sbírané informace:

- základní demografické údaje (**pohlaví a věk**),
- informace o samotné diagnóze mozkové příhody (**etiologie a lokalizace uzávěru cévy**),
- informace o léčbě (typ indikované **terapie a výskyt komplikací**)
- informace o **způsobu ukončení rehabilitace**.
- Stupeň soběstačnosti před rehabilitací byl dodatečně zjištěn z neurologie a na konci rehabilitace byl vyplněn nový dotazník pro určení výsledného **indexu Barthelové**.

Úkol č. 1 – kontrola a příprava dat

1. Do všech řádků tabulky vyplňte do sloupce *Barthel_index_reference* hodnotu 64,4.
2. **Ukotvěte** ID pacientů a názvy proměnných ve sloupcích. (**nápověda**: vyber buňku pro levý horní roh → karta „Zobrazení“ → funkce Ukotvit příčky).
3. Zapněte **automatický filtr** nad celou datovou tabulkou a **zkontrolujte přítomnost chybných hodnot** ve sloupcích Pohlaví, Věk, Etiologie, Lokalizace, Terapie. Chybné hodnoty opravte. (**nápověda**: označ všechny sloupce → karta „Data“ → funkce Filtr).

Úkol č. 1 – kontrola a příprava dat

4. Pomocí **podmíněného formátování** nalezněte **duplicitní záznamy** ID pacientů. 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).
5. Spočítejte hodnoty ve sloupci *Barthel_index_po_rehabilitaci* jako celkový **součet** dosažených bodů v jednotlivých otázkách Barthelové testu po rehabilitaci. (**nápověda**: prostý součet jednotlivých buněk nebo funkce SUMA(...)).

Úkol č. 1 – kontrola a příprava dat

6. Spočítejte hodnoty ve sloupci *Barthel_index_zmena* jako **rozdíl** Barthelové indexu před a po rehabilitaci (**nápověda**: prostý vzorec pro rozdíl).
7. Sloupce *Barthel_index_pred_rehabilitaci* a *Barthel_index_po_rehabilitaci* **překódujte** do sloupců *Kategorie_zavislosti_pred_rehabilitaci* a *Kategorie_zavislosti_po_rehabilitaci* následovně: 0 až 40 = vysoce závislý, 45 až 100 = částečně soběstačný. (**nápověda**: pomocí funkce `KDYŽ(...)`).