

Ing. Jan Kaiser
3. interní klinika 1. LF UK a VFN



1. LÉKAŘSKÁ
FAKULTA
Univerzita Karlova

PŘÍPRAVA DAT PRO STATISTICKÉ ZPRACOVÁNÍ



ING. JAN KAISER

Vzdělání

- Bc. Aplikovaná informatika (FIS VŠE)
- Ing. Kognitivní informatika (FIS VŠE)
- PhD. Biomedicínská informatika (1. LF UK) in progress

OBOR VÝZKUMU

- Telemedicína
- Obezitologie

KONTAKT

vyuka@jankaiser.cz

MATERIÁLY

vyuka.jankaiser.cz



Program přednášky

1. Typy a struktura dat
2. Záznam a reprezentace dat
3. Základy práce s tabulkovými procesory
4. Základy práce s programem Statistica



1.

TYPY A STRUKTURA DAT

1.1 Základní dělení podle statistiků

1. Kvalitativní data

- Nominální (*pohlaví, barva očí...*)
- Ordinální (*dosažené vzdělání...*)

2. Kvantitativní (numerická) data

- Diskrétní (*počitatelné*)
 - „*počet X v Y*“
- Kontinuální (*měřitelné*)
 - „*výška*“, „*váha*“, „*teplota*“...



1.2 Základní dělení podle programátorů

1. Číselná data („integers“)
2. Desetinná čísla („floats“)
3. Textová data („strings“)
4. Binární (Booleovské) hodnoty („boolean“)
5. Datum („date“)
6. Datum a čas („datetime“)
7. Pole („array“)
8. Objekt („object“)
9. Nil („null“)
10. Nedefinováno („undefined“)



1.3 „Normalizace“ dat

- Ve statistice např.:
 - *Min-max normalizace*
 - *Z-skóre standardizace*
 - *Logaritmická transformace*
- Pro naše účely ale potřebujeme především **mít strukturovaná data**
- Často se v datech setkáme s hodnotami „zakódovanými“ v textu, které potřebujeme extrahovat
 - Ideálně chceme získat číselné hodnoty, booleovské hodnoty, případně rozumně definované kategorické proměnné



1.4 „Normalizace“ dat - příklad

Záznam 1:

Pacient Jan Novák, 34 let, m. Stěžuje si na časté dýchací potíže, občasné záchvaty astmatu. Kuřák, denně 1 krabička. V posledním roce 2x v nemocnici s respiračními problémy. Hmotnost se pohybuje okolo 90 kg.

Záznam 2:

Eva Svobodová, 29, ž. Navštívila kvůli vysokému krevnímu tlaku. Naměřen TK 140/90. Nevykazuje žádné další rizikové faktory, nekouří, sportuje. Minulý rok bez hospitalizace. Hmotnost kolem 65 kg.

Záznam 3:

Tomáš Dvořák, věk 45, muž. Diagnóza: diabetes 2. typu, dekompenzovaný. Kuřák, ale snaží se přestat. Hospitalizován jednou minulý rok kvůli komplikacím s cukrovkou. Hmotnost 85.

Výsledek:

	Pac. 1	Pac. 2	Pac. 3
Věk	34	29	45
Pohlaví	muž	žena	muž
Kuřák	ano	ne	ano
Dg.	astma	zvýšený TK	DM2
Hospitalizace	2	0	1
Hmotnost	90	65	85

1.4 „Normalizace“ dat - příklad

Záznam 1:

Pacient Jan Novák, 34 let, m. Stěžuje si na časté dýchací potíže, občasné záchvaty astmatu. Kuřák, denně 1 krabička. V posledním roce 2x v nemocnici s respiračními problémy. Hmotnost se pohybuje okolo 90 kg.

Záznam 2:

Eva Svobodová, 29, ž. Navštívila kvůli vysokému krevnímu tlaku. Naměřen TK 140/90. Nevykazuje žádné další rizikové faktory, nekouří, sportuje. Minulý rok bez hospitalizace. Hmotnost kolem 65 kg.

Záznam 3:

Tomáš Dvořák, věk 45, muž. Diagnóza: diabetes 2. typu, dekompenzovaný. Kuřák, ale snaží se přestat. Hospitalizován jednou minulý rok kvůli komplikacím s cukrovkou. Hmotnost 85.

Výsledek:

	Pac. 1	Pac. 2	Pac. 3
Věk	34	29	45
Pohlaví	muž	žena	muž
Kuřák	1	0	1
Astma	1	0	0
DM2	0	0	1
TK sys.		140	
TK dias.		90	
Hospitalizace	2	0	1
Hmotnost	90	65	85

1.5 Selekce dat

- Potřebujeme všechna data z anamnézy?
- Můžeme je nějak agregovat bez ztráty vypovídající hodnoty?
- Co je naše výzkumná otázka, cíl, či hypotéza?
 - Např.:
 - Místo dg. jen informace zda pacient bere léky ovlivňující hmotnost
 - Vyřazení hodnot TK
 - Vyřazení počtu hospitalizací



1.6 Cíle, hypotézy, výzkumné otázky

1. Hypotéza

- „*Frekvence návštěv nutričního terapeuta pozitivně souvisí s úspěšností pacienta v redukčním programu.*“
- „*Náhrada osobních schůzek s terapeutem on-line komunikací nemá vliv na úspěšnost pacienta v redukčním programu.*“

2. Výzkumná otázka

- „*Existuje souvislost mezi frekvencí návštěv nutričního terapeuta a úspěšností pacienta v redukčním programu?*“
- „*Má náhrada osobních schůzek s terapeutem on-line komunikací vliv na úspěšnost pacienta v redukčním programu?*“

3. Cíl

- „*Určit vliv frekvence návštěv nutričního terapeuta na úspěšnost pacienta v redukčním programu.*“
- „*Stanovit dopad nahrazení osobních schůzek s terapeutem on-line komunikací na úspěšnost pacienta v redukčním programu.*“
- „*Identifikovat faktory, které mají významný vliv na úspěšnost pacientů v redukčním režimu.*“

1.5 Dokumentace

Ať už naše data strukturujeme a „normalizujeme“ jakkoliv, je potřeba tento proces zdokumentovat:

- 1. V metodice práce**, tak aby bylo jasné:
 - Jaká data jsme sbírali.
 - Jak jsme je transformovali.
 - Jakou podmnožinu dat jsme vybrali.
- 2. Pro statistika**, aby měl všechn důležitý kontext.



2.

ZÁZNAM A REPREZENTACE DAT

2.1 Tabulkové procesory

- Data běžně reprezentujeme v **tabulkách**
- Proto využíváme **tabulkové procesory**:
 - MS Excel (<https://uvt.cuni.cz/UVT-1011.html> & <https://uvt.cuni.cz/UVT-1086.html>)
 - Zdarma pod univerzitní licencí.
 - Google Sheets (<https://sheets.google.com/>)
 - Snadné sdílení, ale pozor co a s kým sdílíte!
 - Citlivá data v cloudu není dobrý nápad!

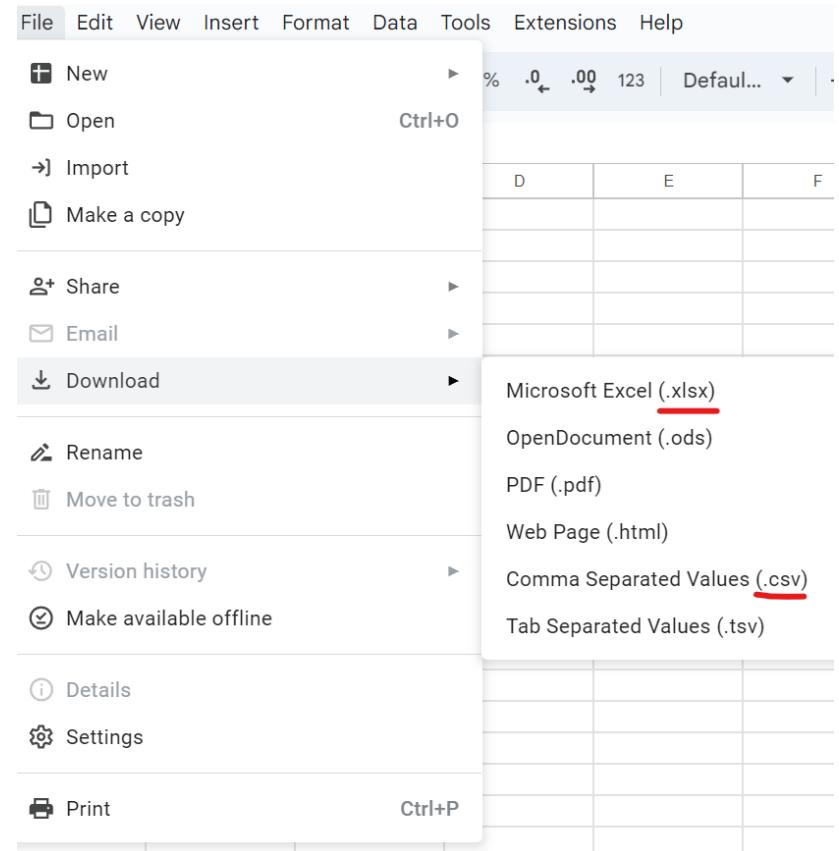


2.2 Formáty a kompatibilita

Ne vždy data zaznamenáváme sami
a ne vždy jsou od začátku v
tabulkovém procesoru:

- V zápisníku / sešitě (*raději rovnou do Excelu / Google Sheetu, prosím!*)
- Ve specializovaném software:
 - Export do **XLS(X), CSV**
 - **CTRL+C** a **CTRL+V**
 - Ruční přepis

A někdy chceme Google Sheet
exportovat do Excelu:



2.3 CSV

„Comma Separated Values“

- Velmi jednoduchý formát kompatibilní se všemi tabulkovými procesory.
- Pozor: data nejsou nutně oddělená jen čárkou, ale např. středníkem či tabulátorem.
- Pozor také na znakovou sadu!
 - Středoevropská: *Windows-1250*
 - Nejuniverzálnější: *UTF-8*

The screenshot shows the Microsoft Excel ribbon with the 'Data' tab selected. Below the ribbon, there are several buttons: 'Načíst data', 'Z Text/CSV', 'Z webu', 'Z tabulky nebo oblasti', 'Z obrázku', 'Poslední zdroje', 'Existující připojení', and 'Aktualizovat vše'. A red arrow points from the 'Načíst data' button to the 'Z Text/CSV' button. Another red arrow points from the 'Oddělovač' dropdown menu to the 'Středník' option. The main area displays a preview of a CSV file named 'zv_pcobc.csv'. The preview includes a header row and several data rows. The columns are labeled: NAZCOBCE, PSC, NAZPOST, KODOKRESU, NAZOKRESU, and NAZOBCE. The data rows show various place names and their corresponding codes and postcodes.

NAZCOBCE	PSC	NAZPOST	KODOKRESU	NAZOKRESU	NAZOBCE
Albertamy	36235	Albertamy	3403	Karlovy Vary	Albertamy
Adamov	54232	Úpice	3610	Trutnov	Trutnov
Adamov	67904	Adamov 1	3701	Blansko	Adamov
Adamov	28601	Čáslav	3205	Kutná Hora	Adamov
Adamov	37371	Rudolfov	3301	České Budějovice	Adamov
Adolfovice	79001	Jeseník 1	3811	Jeseník	Bělá pod Pradědem
Albeř	37833	Nova Bystřice	3303	Jindřichův Hradec	Nova Bystřice
Albeřice	36452	Žlutice	3403	Karlovy Vary	Verušičky
Alberovice	25765	Čechtice	3201	Benešov	Loket
Albrechtice	34201	Sušice 1	3404	Klatovy	Sušice
Albrechtice	38901	Vodňany	3307	Strakonice	Drahonice
Albrechtice	28401	Kutná Hora 1	3205	Kutná Hora	Malešov
Albrechtice	46345	Pěnčín u Liberce	3505	Liberec	Pěnčín
Albrechtice	59257	Rozsochy	3714	Žďár nad Sázavou	Rozsochy
Albrechtice	56301	Lnářkroun	3611	Ústí nad Orlicí	Albrechtice
Albrechtice	73543	Albrechtice u Českého Těšína	3803	Karviná	Albrechtice
Albrechtice nad Orlicí	51722	Albrechtice nad Orlicí	3607	Rychnov nad Kněžnou	Albrechtice nad Orlicí
Albrechtice nad Vltavou	39816	Albrechtice nad Vltavou	3305	Písek	Albrechtice nad Vltavou
Albrechtice u Frýdlantu	46331	Chrastava	3505	Liberec	Frýdlant
Albrechtice u Rýmařova	79351	Břidličná	3801	Bruntál	Břidličná

2.3 CSV – zdroják

1 NAZCOBCE;PSC;NAZPOST;KODOKRESU;NAZOKRESU;NAZOBCE
2 Abertamy;36235;Abertamy;3403;Karlov Vary;Abertamy
3 Adamov;54232;Úpice;3610;Trutnov;Trutnov
4 Adamov;67904;Adamov 1;3701;Blansko;Adamov
5 Adamov;28601;Čáslav;3205;Kutná Hora;Adamov
6 Adamov;37371;Rudolfov;3301;České Budějovice;Adamov
7 Adolfovice;79001;Jeseník 1;3811;Jeseník;Bělá pod Pradědem
8 Alberštejn;37833;Nová Bystřice;3303;Jindřichův Hradec;Nová Bystřice
9 Alberštejn;36452;Žlutice;3403;Karlov Vary;Verušičky
10 Alberovice;25765;Čechtice;3201;Benešov;Loket
11 Albrechtice;34201;Sušice 1;3404;Klatovy;Sušice
12 Albrechtice;38901;Vodňany;3307;Strakonice;Drahonice
13 Albrechtice;28401;Kutná Hora 1;3205;Kutná Hora;Malešov
14 Albrechtice;46345;Pěnčín u Liberce;3505;Liberec;Pěnčín
15 Albrechtice;59257;Rozsochy;3714;Žďár nad Sázavou;Rozsochy
16 Albrechtice;56301;Lanškroun;3611;Ústí nad Orlicí;Albrechtice
17 Albrechtice;73543;Albrechtice u Českého Těšína;3803;Karviná;Albrechtice
18 Albrechtice nad Orlicí;51722;Albrechtice nad Orlicí;3607;Rychnov nad Kněžnou;Albrechtice nad Orlicí
19 Albrechtice nad Vltavou;39816;Albrechtice nad Vltavou;3305;Písek;Albrechtice nad Vltavou
20 Albrechtice u Frýdlantu;46331;Chrastava;3505;Liberec;Frýdlant
21 Albrechtice u Rýmařova;79351;Břidličná;3801;Bruntál;Břidličná
22 Albrechtice v Jizerských horách;46843;Albrechtice v Jizerských horách;3504;Jablonec nad Nisou;Albrechtice v Jizerských horách
23 Albrechtičky;74255;Albrechtičky;3804;Nový Jičín;Albrechtičky
24 Albrechtovice;38301;Prachatice;3306;Prachatice;Záblatí
25 Aleje-Zátiší;35002;Cheb 2;3402;Cheb;Františkovy Lázně
26 Alenina Lhota;39133;Jistebnice;3308;Tábor;Jistebnice
27 Alexovice;66491;Ivančice;3703;Brno-venkov;Ivančice
28 Alojzov;79804;Určice;3709;Prostějov;Alojzov
29 Alšovice;46821;Bratříkov;3504;Jablonec nad Nisou;Pěnčín
30 Alžbětín;34004;Železná Ruda;3404;Klatovy;Železná Ruda
31 Amálín;79397;Slezské Rudoltice;3801;Bruntál;Slezské Rudoltice
32 Amerika;25901;Votice;3201;Benešov;Votice
33 Andělice;33901;Klatovy 1;3404;Klatovy;Dolany
34 Andělkov;46373;Habartice u Frýdlantu;3505;Liberec;Višňová
35 Andělská Hora;46331;Chrastava;3505;Liberec;Chrastava
36 Andělská Hora;36471;Bochoř;3403;Karlov Vary;Andělská Hora
37 Andělská Hora;79331;Světlá Hora;3801;Bruntál;Andělská Hora
38 Anděl Strážce;46342;Hodkovice nad Mohelkou;3504;Jablonec nad Nisou;Frýdštejn
39 Anenská Studánka;56301;Lanškroun;3611;Ústí nad Orlicí;Anenská Studánka
40 Anenská Ves;35709;Habartov;3409;Sokolov;Krajková
41 Annín;34201;Sušice 1;3404;Klatovy;Dlouhá Ves

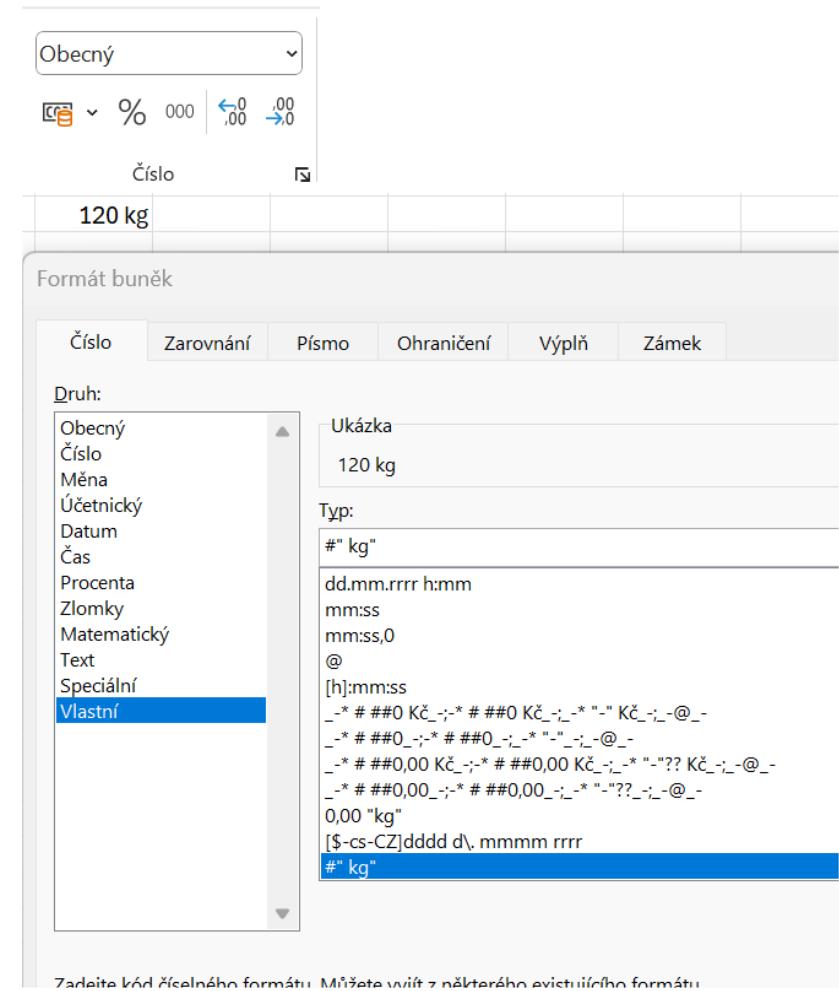
2.4 Na co si dát pozor

- **Nemixovat datové typy**

- Nemíchat např. numerické a textové hodnoty v jednom sloupci!
- Pokud hodnota chybí, nechám pole prázdné.
- Nepíšu „-“ nebo „nezapsáno“, pokud hodnota není nula, nepíšu „0“

- **Nepsat jednotky ručně**

- Tabulkové procesory mají funkci pro nastavení formátu (e.g. „kg“, „Kč“...), které doplní jednotku při zachování čistě číselné hodnoty.
- Buď přidejte jednotky takto formátováním, nebo je uveďte v záhlaví.



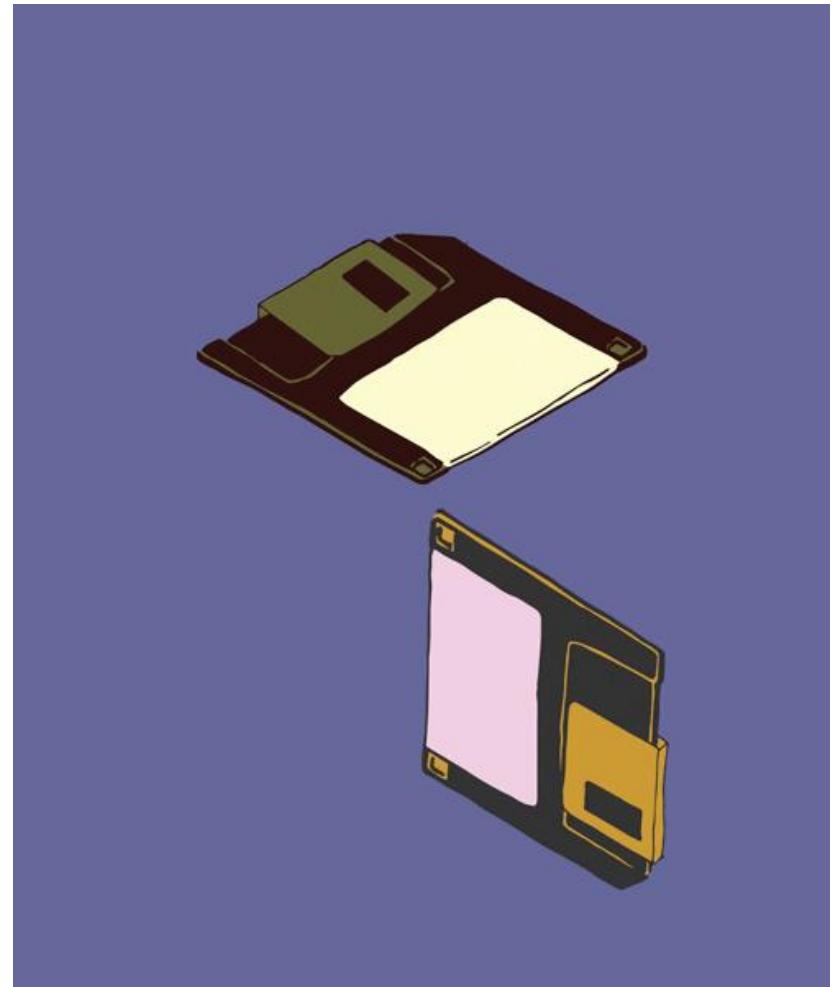
2.4 Na co si dát pozor

- **Automatická konverze na datum**
 - Excel i Google Sheets jsou notoricky známé tím, že cokoliv se jen trochu podobá datu převádí na datum!
- **Naformátujte si tabulku předem**
 - Než začnete vůbec zaznamenávat data, je dobré mít představu o tom jaká.
 - Nadepište si sloupce a nastavte buňkám správné datové typy – ušetří vám to bolest hlavy v budoucnu!



2.4 Na co si dát pozor

- **Kontrolujte, jestli se něco nerozbilo**
 - Speciálně pokud něco někam kopírujete!
 - Viz např. konverze na datum z předchozího slide
- **Průběžně ukládejte.**
- **Zálohujte!**



3.

ZÁKLADY PRÁCE S TABULKOVÝMI PROCESORY

3.1 Struktura

- Sloupce
 - A – ZZ
- Buňky
 - 1 – 9 999 999
- Listy
 - Hodnoty lze referencovat i mezi listy

Tip: Listy a sloupce lze také skrýt.

Tip na navigaci mezi buňkami:

- Ctrl + ↓
- Ctrl + ↑
- Ctrl + →
- Ctrl + ←

A1	:	X ✓ fx	NAZCOBCE		
A	B	C	NAZCOBCE	PSC	NAZPOST
1			NAZCOBCE		
2	Abertamy	36235	Abertamy		
3	Adamov	54232	Úpice		
4	Adamov	67904	Adamov 1		
5	Adamov	28601	Čáslav		
6	Adamov	37371	Rudolfov		
7	Adolfovice	79001	Jeseník 1		
8	Alber	37833	Nová Bystřice		
9	Alberice	36452	Žlutice		
10	Alberovice	25765	Čechtice		
11	Albrechtice	34201	Sušice 1		
12	Albrechtice	38901	Vodňany		
13	Albrechtice	28401	Kutná Hora 1		
14	Albrechtice	46345	Pěnčín u Liberce		
15	Albrechtice	59257	Rozsochy		
16	Albrechtice	56301	Lanškroun		
17	Albrechtice	73543	Albrechtice u Českého		
18	Albrechtice nad Orlicí	51722	Albrechtice nad Orlicí		
19	Albrechtice nad Vltavou	39816	Albrechtice nad Vltavou		
20	Albrechtice u Frýdlantu	46331	Chrastava		
21	Albrechtice u Rýmařova	79351	Břidličná		
22	Albrechtice v Jizerských horách	46843	Albrechtice v Jizerských		
23	Albrechičky	74255	Albrechičky		
24	Albrechtovice	38301	Prachatice		
25	Aleje-Zátiší	35002	Cheb 2		
26	Alojzína Lhotka	20122	Hrádečno		

3.2 Formátování a automatické doplňování

- Číselné formátování viz předchozí kapitola.
 - Nezapomeňte číselný formát nastavit předem.
- Desetinná místa: 
 - Pozor: jedná se jen o formát, program stále počítá se všemi desetinnými místy!
- Automatické doplňování:
 - Pokud máte logickou řadu, at' už číselnou nebo časovou, výběrem a roztažením buněk vám Excel doplní následující hodnoty.
 - Podobně funguje i u vzorců, viz dále...

	A	B	C
1	JAN		
2	FEB		
3	MAR		
4	APR		
5	MAY		
6	JUN		
7	JUL		
8	AUG		
9	SEP		
10	OCT		
11	NOV		
12	DEC		

3.3 Vzorce

- $=\text{SUMA}(A1:D1)$
- $=\text{PRŮMĚR}(A1:D1)$
- $=\text{MIN}(A1:D1)$
- $=\text{MAX}(A1:D1)$
- $=\text{SMODCH.P}(A1:D1)$
- $=\text{COUNTIF}(A1:D1, \text{"muž"})$
- ... a další (viz např.
<https://office.lasakovi.com/excel/funkce/>)

Tip: Excel má několik módů vložení, pokud kopírujete buňku se vzorcem, můžete ji vložit např. jen jako hodnotu nebo naopak jen vzorec.

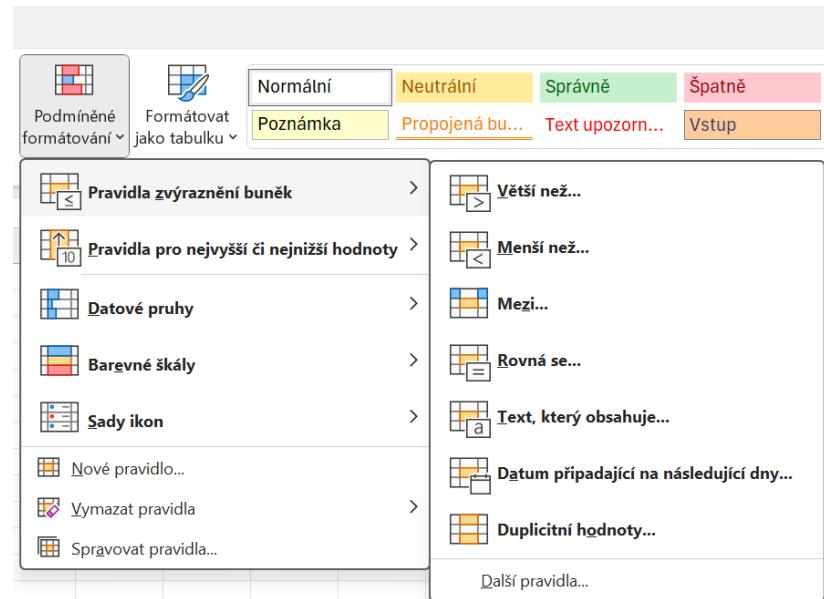
- A co znamená
 - $=\text{SUMA}(\$A\$1:\$D\$1)$
 - $=\text{SUMA}(\$A1:\$D1)$
 - $=\text{SUMA}(A\$1:D\$1)?$

→ Zkuste v buňce se vzorcem stisknout klávesu **F4**:

- Znak „\$“ definuje tzv. fixaci odkazu. Zkuste si vzorec roztahnout na další buňky s a bez dolaru.
- Pokud není přítomen, odkazy se automaticky inkrementují.

3.4 Podmíněné formátování

- Pokud potřebujete zvýraznit určité hodnoty, bude se hodit.
- Lze podbarvit např.
 - Duplicítní hodnoty
 - Buňky, které obsahují určitý text
 - Čísla větší / menší než X
 - Čísla pod / nad průměrem vzorku
- Lze také definovat vlastní pravidla pomocí vzorce
- Dobré např. jako pomůcka při čištění dat (*do samotné práce ale barvičky nepatří!*)



3.5 Čištění dat

- Někdy data chybí
- Nebo obsahují *outliery*
- Případně překlepy či chyby

Outliery = Hodnoty, které se výrazně liší od ostatních pozorovaných hodnot

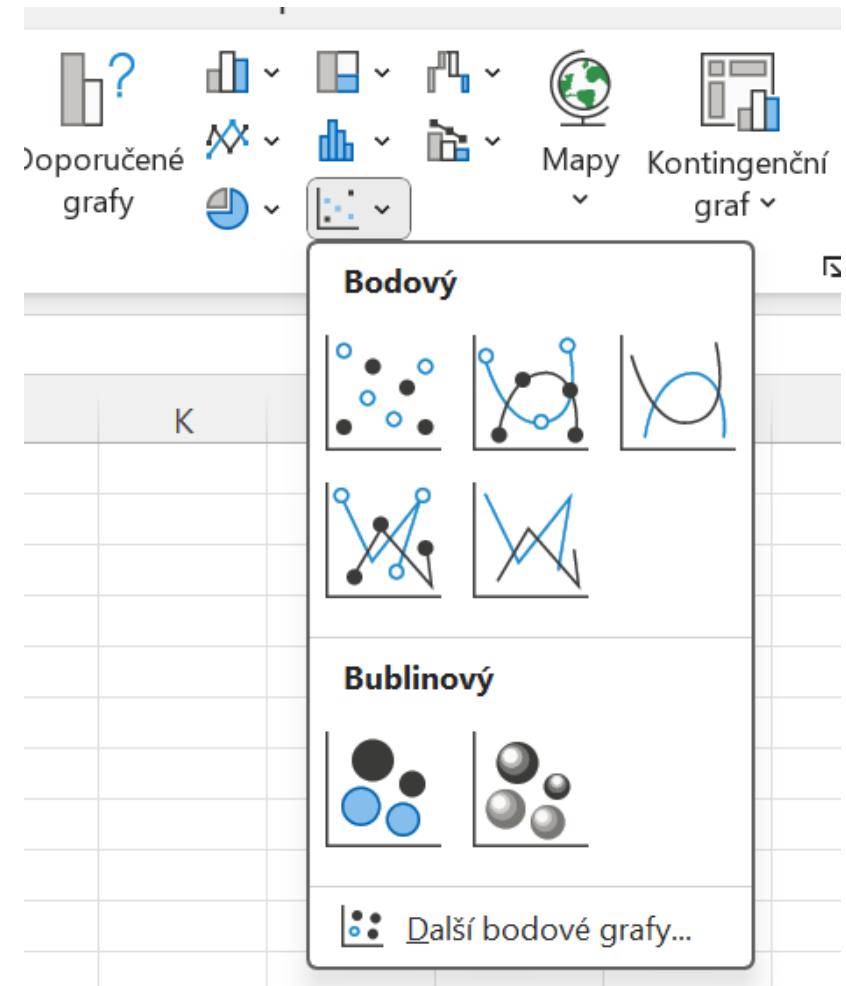
Jak je identifikovat?

- Vizualizací v grafu
- Podmíněným formátováním

Jak se jich zbavit?

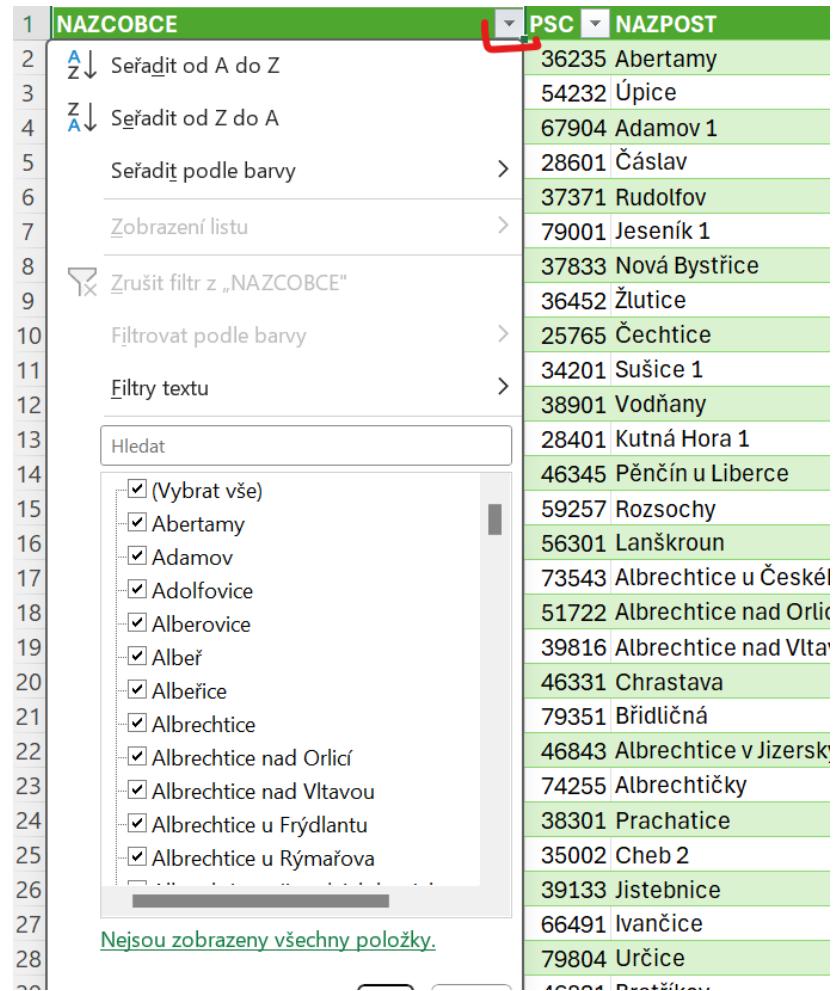
- Data → Odebrat duplicity
- *Najít a nahradit (CTRL + H)*
- Ručně opravit

Prevence: *Data → Ověření dat*



3.6 Práce s tabulkami – řazení a filtrování

- Vybrat data → pravé myšítko → Seřadit / Filtr
 - Pozor! Musíte vybrat všechna data, pokud vyberete jen jeden sloupec, ztratíte návaznost napříč řádky!
- Lepší cesta:
 - Vybrat data → Domů → Formátovat jako tabulku
 - Kliknout na šipečku u nadpisu sloupce a mohu řadit či filtrovat



The screenshot shows a context menu for a table column titled "NAZCOBCE". The menu includes options for sorting (A-Z, Z-A, by color), viewing the list, unfiltering, filtering by color, filtering text, and searching. A checked checkbox indicates "Select all". The main table lists various locations with their codes and names.

	NAZCOBCE	PSC	NAZPOST
2	A↓ Seřadit od A do Z		36235 Abertamy
3	Z↓ Seřadit od Z do A		54232 Úpice
4	Seřadit podle barvy	>	67904 Adamov 1
5	Zobrazení listu	>	28601 Čáslav
6	Zrušit filtr z „NAZCOBCE“	>	37371 Rudolfov
7	Filtrovat podle barvy	>	79001 Jeseník 1
8	Filtry textu	>	37833 Nová Bystřice
9	Hledat		36452 Žlutice
10	<input checked="" type="checkbox"/> (Vybrat vše)		25765 Čechtice
11	<input checked="" type="checkbox"/> Abertamy		34201 Sušice 1
12	<input checked="" type="checkbox"/> Adamov		38901 Vodňany
13	<input checked="" type="checkbox"/> Adolfovice		28401 Kutná Hora 1
14	<input checked="" type="checkbox"/> Alberovice		46345 Pěnčín u Liberce
15	<input checked="" type="checkbox"/> Alber		59257 Rozsochy
16	<input checked="" type="checkbox"/> Alberice		56301 Lanškroun
17	<input checked="" type="checkbox"/> Albrechtice		73543 Albrechtice u Českého
18	<input checked="" type="checkbox"/> Albrechtice nad Orlicí		51722 Albrechtice nad Orlicí
19	<input checked="" type="checkbox"/> Albrechtice nad Vltavou		39816 Albrechtice nad Vltavou
20	<input checked="" type="checkbox"/> Albrechtice u Frýdlantu		46331 Chrastava
21	<input checked="" type="checkbox"/> Albrechtice u Rýmařova		79351 Břidličná
22	...		46843 Albrechtice v Jizerských
23	...		74255 Albrechtičky
24	...		38301 Prachatice
25	...		35002 Cheb
26	...		39133 Jistebnice
27	...		66491 Ivančice
28	...		79804 Určice
			40001 Drahonice

3.7 Makra & VBA Skripty

- V Excelu lze programovat!
- V menu *Automatizace* máte možnost spouštět předdefinované skripty, případně nahradit vlastní sekvenci operací – tzv. *Makro*.
- Po uložení můžete makro spouštět znova a znova, případně ho navázat na tlačítko.
- Makra lze také psát či upravovat přímo ve VBA.
- Více info pro zvědavé např. na <https://office.lasakovi.com/excel/vba/co-je-VBA/>



The screenshot shows a code editor window titled "Skript" (Script). The code is written in VBA and performs the following tasks:

- It defines a function named "main" that takes a parameter "workbook" of type "ExcelScript.Workbook".
- Inside the function, it gets the active worksheet from the workbook.
- It then pastes the range B5 from the selected sheet into the active worksheet, starting at cell E3.
- After pasting, it sets the formula for cell B5 to "=SUM(Tabulka3[@])".
- It retrieves the table "Tabulka3" from the workbook.
- It pastes the range D4 from the table "Tabulka3" into the active worksheet, starting at cell Sloupec3 in row 1.
- Finally, it moves the range D4 from the table "Tabulka3" to the active worksheet.

```
function main(workbook: ExcelScript.Workbook) {
    let selectedSheet = workbook.getActiveWorksheet();
    // Paste to range B5 on selectedSheet
    // from range E3 on selectedSheet
    selectedSheet.getRange("B5").copyFrom(
        selectedSheet.getRange("E3"),
        ExcelScript.RangeCopyType.all,
        false, false);
    // Set range B5 on selectedSheet
    selectedSheet.getRange("B5").
        setFormulaLocal("=SUM(Tabulka3[@])");
    ;
    let tabulka3 = workbook.getTable("Tabulka3");
    // Paste to range D4 on selectedSheet
    // from table tabulka3 cell in row 1
    // on column Sloupec3
    tabulka3.getColumn("Sloupec3").
        getRangeBetweenHeaderAndTotal().
        getRow(1).moveTo(selectedSheet.
        getRange("D4"));
}
```

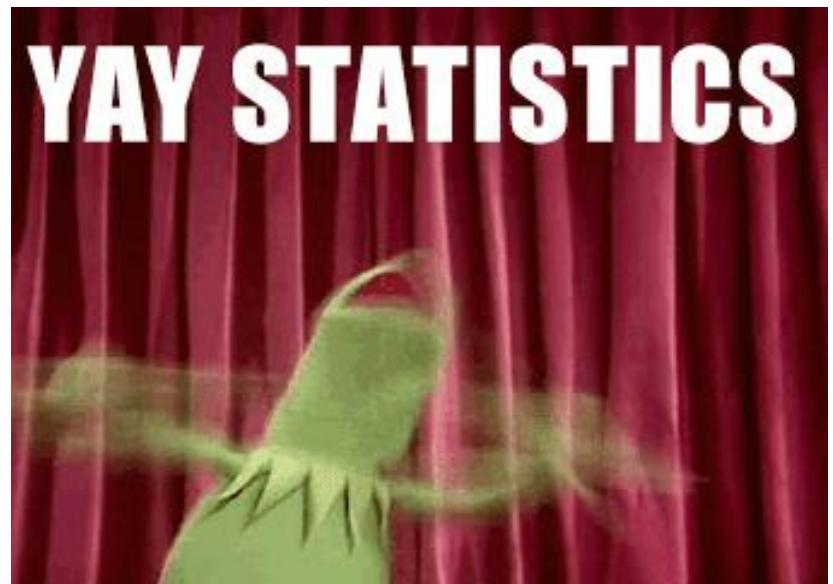
4.

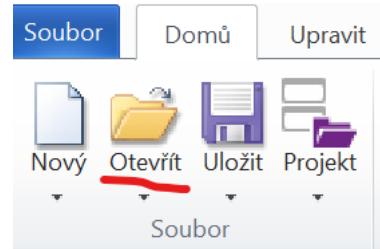
ZÁKLADY PRÁCE S PROGRAMEM STATISTICA

4.1 Program Statistica

TIBCO Statistica - statistický software
v češtině pro pokročilé i méně
pokročilé statistické analýzy

- <https://multilicence.lf1.cuni.cz/>
- <https://statistica.pro/>
- S instalací ideálně počkejte až vám přijde *Serial Number* a *Product Key* (může chvíli trvat).
- Instalační soubor spouštějte jako správce.
- Ve staženém zipu je i čeština, soubory ze složky *statistica_cestina_14* musíte zkopírovat do *C:\Program Files\TIBCO\Statistica*
 - <https://statistica.pro/knowledgebase-cz/installace-ceskeho-prostrediv-aplikaci-tibco-statistica-14/>





4.2 Import dat

- Statistica má integraci s MS Office.
- „Otevřít“ → Vyberete váš excelovský soubor (nebo CSV)
- → „Importovat vybraný list do tabulky“
- Pozn.: Statistica data validuje, takže pokud máte diskrepance v datových typech, upozorní vás na to!
- Po úspěšném importu se zobrazí okno s datovou tabulkou.

Import textových souborů s oddělovači

Oddělovač proměnných: Středník

Možnosti importu:

- Vzít jména proměnných z prvního řádku
- Počet případů k přeskročení: 0
- Maximální počet případů pro import: 1001
- Znak pro desetinný oddělovač: ,
- Text. řetězec chybějících dat: .
- Přeskročit prázdné řádky

Možnosti prom. (klikněte níže pro povolení):

- Jméno
- Číslo: 1
- Hodnota chyb. dat: -999999998,00
- Typ dat
- Délka textu: 6
- Vstup. formát
- Odstranění nečíselných znaků
- Vynech.

Zastavit Rozšířené

Textový soubor (zv_psc_adr.csv)

PSC	TYP_PSC
10000	1
10001	1
10002	10
10003	1
10004	1
10005	1
10006	10
10007	1
10008	10
10009	1
10010	4
10011	1
10012	10
10013	10
10014	10
10015	10
10016	10
10017	3
10018	10
10019	10

Pozn.: Kliknutím na OK způsobi, že datový soubor bude importován jako tabulka za použití vybraných možností. Velikost cache tabulky je množství místa použité pro buffer náhledu pro zjištění typů proměnných.

Autoobn.

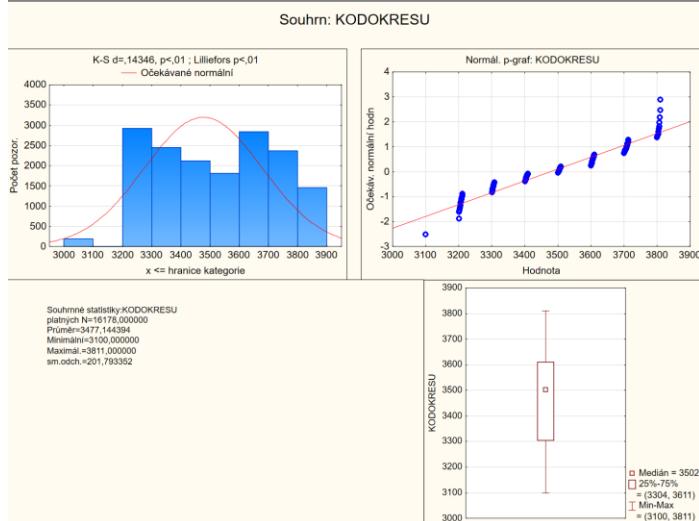
4.3 Popisné statistiky

- *Statistiky → Základní statistiky → Popisné statistiky*

Příklad souhrn:

Proměnná	Popisné statistiky (zv_pcobc)				
	platných N	Průměr	Minimální	Maximál.	sm.odch.
PSC	16178	46490,83	10000,00	79862,00	16434,05
KODOKRESU	16178	3477,14	3100,00	3811,00	201,79

Příklad Grafy 1:



Soubor Domů Upravit Zobrazit Vložit Formát Statistiky

Základní Vícenásobná ANOVA Neparametrické Prokládání Rozdělení a simulační regrese statistiky rozdělení

Popisné statistiky: zv_pcobc

Proměnné: PSC KODOKRESU

Rychlé Detailní výsledky Robustnost Normalita Pravd. & bod. grafy Kateg. grafy Možnosti

Souhrn: Statistiky Grafy 1 Grafy 2

Tabulky četností Histogramy

Krabicový graf všech proměnných

Zobrazení grafického srovnávacího souhrnu

Tip: Grafy i tabulky se dají kopírovat!

4.4 Korelace

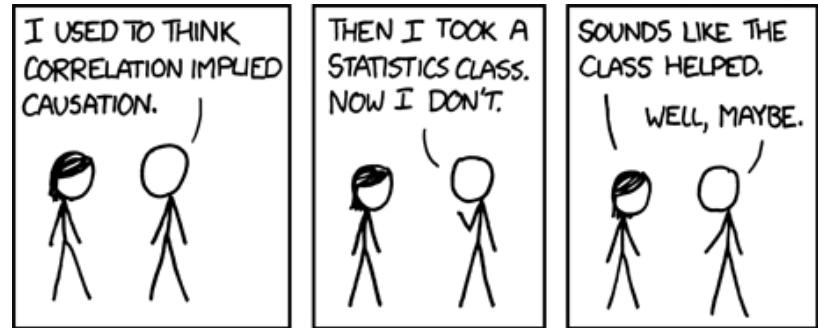
- *Statistiky → Základní statistiky → Korelační matice*
- Vyberete proměnné
- Statistica vyplivne tabulku s korelacemi a červeně označí statisticky významné korelace mezi proměnnými

Tip: Korelace ≠ kauzalita!

Korelace (zv_pcobc)		
Označ. korelace jsou významné na hlad. p < ,05000		
N=16178 (Celé případy vynechány u ChD)		
Proměnná	PSC	KODOKRESU
PSC	1,000000	0,949108
KODOKRESU	0,949108	1,000000

4.5 Další info ke statistice

- <https://statistica.pro/knowledgebase/>
- <https://www.wikiskripta.eu/w/Fórum:Testy/Pearsonův korelační koeficie nt>
- <https://www.wikiskripta.eu/w/Normální rozdělení>
- <https://www.wikiskripta.eu/w/Popisná statistika>



ZDROJ: XKCD (<https://xkcd.com/552/>)

Tip: nebojte se využít vedoucího, kamarády...



Díky za pozornost!

Dotazy?

KONTAKT

vyuka@jankaiser.cz

MATERIÁLY

vyuka.jankaiser.cz

