



Analytické nástroje v MS Excelu

Marie Franců

rozšířené možnosti programu

Obsah

Úvodem	5
Rozšířené možnosti výpočtů	6
Co jsou to Doplnky	6
Jak se doplňky zapnou	6
Jaké doplňky se v Excelu nachází	8
Analytické nástroje	9
Co vše se zde nachází	9
Přehled položek z dialogového okna:	9
Jak s nástroji pro analýzu dat pracovat	11
Anova.....	12
Anova: jeden faktor	12
Anova: dva faktory s opakováním	13
Anova: dva faktory bez opakování	15
Korelace	17
Kovariance	19
Popisná statistika	20
Exponenciální vyrovnání.....	21
Dvouvýběrový F-test pro rozptyl.....	23
Fourierova analýza	25
Histogram	26
Klouzavý průměr	29
Generátor pseudonáhodných čísel.....	31
Typy generování čísel	32
Rovnoměrné	32
Normální	33
Binomické	35
Poissonovo	36
Vzorky	36
Diskrétní.....	37
Pořadová statistika a percentily	39
Regrese	41
Vzorkování	46
Studentovy t-testy.....	48
Dvouvýběrový párový t-test na střední hodnotu.....	48
Dvouvýběrový t-test s rovností / nerovností rozptylů	51

Dvouvýběrový z-test na střední hodnotu	53
Řešitel – a co to vůbec je.....	55
Kde se Řešitel nachází	55
Co je to Řešitel.....	56
Jak tedy vše použít	56
Jak lze hledat řešení	62
Čím se začne	62
Co když MS Excel nemůže najít řešení?	64
Co dodat na závěr	65
Přílohy.....	66
S čím se můžete ve statistických výpočtech setkat	66

Úvodem

Možná víte, nebo jen tušíte, že MS Excel obsahuje mnoho funkcí, které dost lidí při běžné práci nevyužívá. Pokud se však o tento program začnete zajímat víc, může jednoduše zjistit, že některé možnosti mohou velmi usnadnit práci, hlavně při složitějších výpočtech nebo zpracování statistických dat. Přesto, že MS Excel není dokonalý program pro statistické výpočty jako některé speciální nástroje přímo k tomu určené, umí si v mnoha případech se statistickými výpočty velmi dobře poradit.

Vždycky jsem si říkala, že bych si v klidu měla projít a opravdu prozkoušet jednotlivé analytické nástroje, které MS Excel nabízí. Vždy jsem to ale nějak odkládala. Až nyní jsem si řekla, že to konečně udělám. A zároveň jsem to využila tak, že jsem vše začala psát do jednoduché příručky. Myslím, že tím vznikl jednoduchý návod také pro ostatní uživatele MS Excelu, kteří potřebují své údaje statisticky zpracovat, a přitom nechtějí investovat čas a ani peníze do speciálních statistických programů.

Navíc jsem ještě do příručky přidala možnost práce s nástrojem Řešitel, který umí například i řešit soustavu rovnic o více neznámých. A na závěr najdete návod na použití nástroje jednoduššího, a to je Hledání řešení.

Rozšířené možnosti výpočtů

Vedle standardních statistických funkcí, dostupných např. na kartě Vzorce ve skupině **Knihovna funkcí** pod ikonou **Statistické**, obsahuje MS Excel i sadu komplexních statistických nástrojů ve formě takzvaných doplňků, ve kterých se nachází například také **Analytické nástroje**.

Co jsou to Doplňky

Jak bylo řečeno před chvílí, tak mnoho nástrojů pro zpracování statistických dat jsou obsaženy v takzvaných **Doplňcích**. Po správném nastavení stačí s nimi jen začít pracovat.

Jinak obecně se dá říci, že **doplňky jsou programy**, které do MS Excelu přidávají různé volitelné příkazy a funkce. Přičemž Excel může obsahovat několik typů doplňků:

- doplňky programu MS Excel,
- vlastní doplňky modelu COM (Component Object Model) – rozšiřuje možnosti MS Excelu, jsou vytvářeny v programovacím jazyce (Visual Basic, Visual C++ a Visual J++) a mají příponu .dll nebo .exe,
- rozšiřující balíky XML (mohou propojením sešitu s databázemi např. usnadnit vyplňování formulářů nebo sdílení dat),
- doplňky pro automatizaci – instalaci obvykle zajišťuje vývojář daného doplňku

Základní sada doplňku je hned k dispozici při instalaci a další jsou přístupné na serveru Microsoftu. Další rozšíření se dají najít i na jiných serverech v rámci internetu.

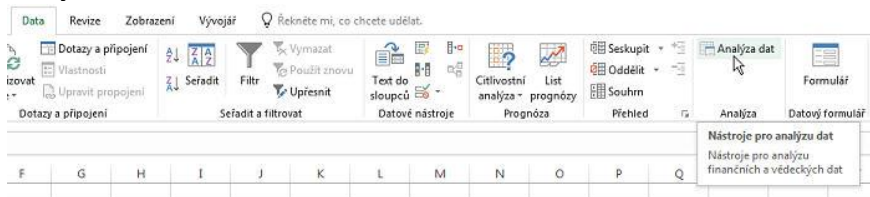
Jinak doplňky se na počítači standardně umísťují do složky AddIns (nebo do některé jejich podsložky) umístěné ve složce C:\Users\uživatel\AppData\Roaming\Microsoft.

Jak se doplňky zapnou

Doplňky, které jsou součástí standardní instalace Excelu je třeba před použitím aktivovat (zapnout). Proto v nabídce karty **Soubor** se klepne

Analytické nástroje

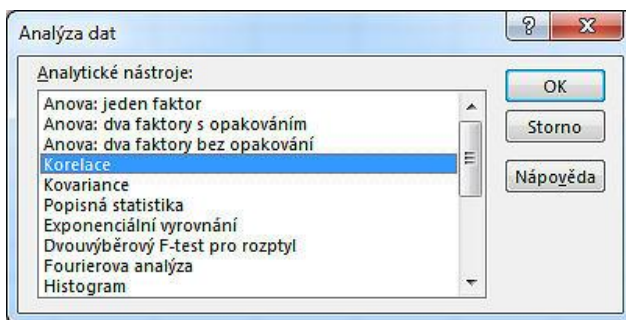
Zaškrtněte položku **Analytické nástroje** a poté klepněte na tlačítko **OK**. Tím se na konci karty **Data** zobrazí skupina **Analyzá** s ikonou **Analyzá dat**.



Obrázek 3: Kde se Analyzá dat najde

Co vše se zde nachází

Po klepnutí na ikonu **Analyzá dat** se zobrazí seznam dostupných nástrojů.



Obrázek 4: Analytické nástroje v Excelu

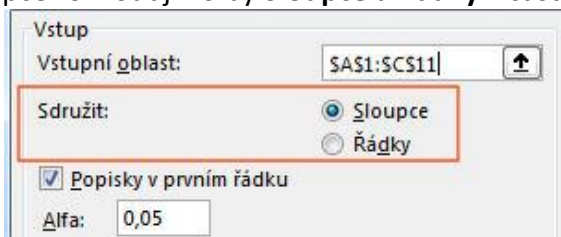
Přehled položek z dialogového okna:

- **Anova: jeden faktor** – ukazuje, zda má na hodnotu závislé veličiny statisticky významný vliv hodnota jedné sledované nezávislé veličiny.
- **Anova: dva faktory s opakováním** – zjišťuje, zda mají na hodnotu závislé veličiny statisticky významný vliv hodnoty dvou sledovaných nezávislých veličin. Přičemž každá kombinace nezávislých veličin je vyjádřena více hodnotami.
- **Anova: dva faktory bez opakování** – zjišťuje, zda mají na hodnotu závislé veličiny statisticky významný vliv hodnoty

- **Dvouvýběrový z-test na střední hodnotu** – zjišťuje, zda mají dva výběry stejnou nebo odlišnou střední hodnotu, zároveň ale je třeba znát rozptyly souborů, ze kterých výběry pocházejí.

Jak s nástroji pro analýzu dat pracovat

Některé nástroje se mohou sice trochu lišit, ale obecně je důležité mít předem připravená data, protože většina z nich bude požadovat zadání tak zvané **Vstupní oblasti**. Tedy oblasti dat v listu, kde máte data připravená k analýze. Většinou se očekává, že data jsou ve sloupcích, případně řádcích. O tom, zda budou jako výběry chápány řádky či sloupce rozhodují volby **Sloupce** a **Řádky** v části **Sdružit**.



Obrázek 5: Zadávání vstupní oblasti

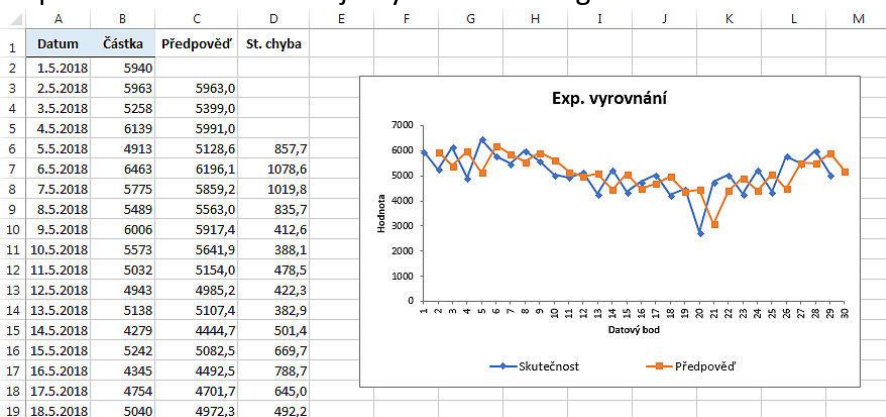
Jestliže jsou v záhlaví řádků nebo sloupců uvedeny jejich názvy a je třeba je zahrnout do vstupní oblasti, tak se musí zaškrtnout volba **Popisky v prvním řádku** (sloupci) případně volba **Popisky**. Jinak jestli se označí oblast i s textovými popisky a zapomenete tuto zatrhnout, tak při spuštění nástroje se zobrazí upozornění na to, že vstupní oblast obsahuje nečíselná data.

Důležitou položkou je i hladina významnosti **Alfa**. Standardně se počítá s hladinou významnosti 0,05 (lze ji však snadno změnit).

Další důležitou částí dialogových oken je nastavení **Možnosti výstupu**. Zde je třeba většinou určit, kde se mají zobrazit výsledky analýzy. Pokud se zvolí **Výstupní oblast**, pak se jen klepne do již existující buňky na listu. Při volbě **Nový list** dochází k vytvoření nového samostatného listu pro zobrazení výsledků. Jestliže do sousedního pole se napíše název listu, tak nový list bude hned i pojmenován. A poslední možnost **Nový sešit** vytvoří samostatný sešit.

Vstupní data se zadají jednoduše pomocí myši. Kromě vstupních dat je třeba ještě napsat takzvaný **koeficient útlumu**. Obvykle se používají hodnoty **0,2 až 0,3**, které říkají o kolik procent (**20 % až 30 %**) bude aktuální předpověď upravena podle chyby předchozí předpovědi. Větší hodnoty dovolují rychlejší reakci na chybu, ale mohou zase vést k chybným předpovědím, malé zase mohou způsobit příliš pomalou předpověď hodnot.

Ještě lze zatrhnout volby **Vytvořit graf** a **Standardní chyby** a u **Možnosti výstupu** si zadejte jednu buňku, od které se mají výsledná data zobrazovat (Obrázek 20 má zvolenou buňku C2). A po klepnutí na tlačítko **OK** se již výsledná data i graf zobrazí.



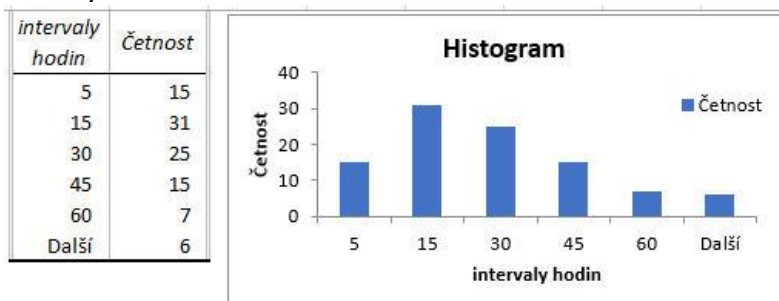
Obrázek 21: Výsledek Exponenciálního vyrovnání

V první buňce výstupní tabulky se zobrazí chyba, ve druhé odkaz na první hodnotu a od třetího řádku již probíhá samotné vyrovnávání. Pokud je **Koeficient útlumu 0,2**, tak vzorec vynásobí číslem 0,8 (tj. doplňkem do jedné) předchozí původní hodnotu a přičte k ní předchozí vyrovnanou hodnotu vynásobenou číslem 0,2. Do výpočtu dalších vyrovnávaných hodnot jsou takto zahrnuty vždy všechny předchozí hodnoty s tím, že časově vzdálenější hodnoty jsou brány s menší váhou.

Do pole **Vstupní oblast** se zadá oblast dat. Do pole **Hranice tříd** pak tabulku s připravenými hranicemi. Pokud tato tabulka obsahuje i záhlaví, tak se zaškrtně volba **Popisky**. Je obvyklé hranice seřadit od nejmenší k největší, ale není to nutné. Nástroj si umí hranice v případě potřeby seřadit sám.

V dolní části **Možnosti výstupu** se určuje to, kam se má vytvořit tabulka výstupní. Dále lze ještě zaškrtnout volby **Pareto** (tříděný histogram), **Kumulativní procentuální podíl** a **Vytvořit graf**.

Pokud žádná z uvedených voleb nebude zaškrta, tak se vytvoří jen tabulka, která bude obsahovat jen informaci o četnostech hodnot v jednotlivých třídách.

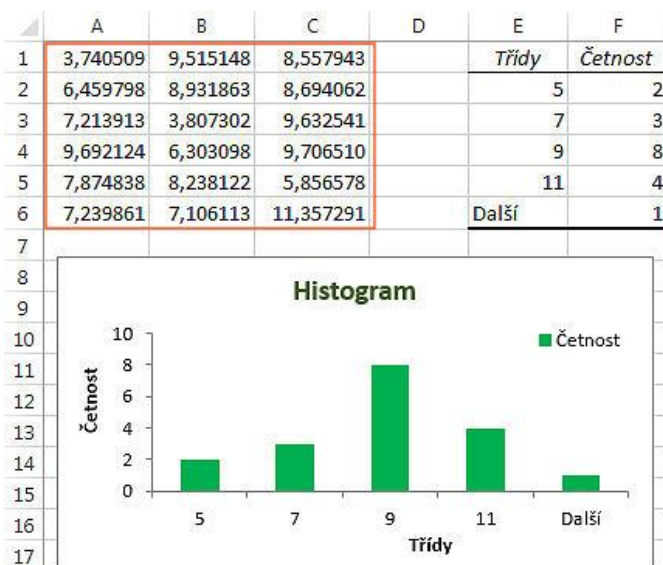


Obrázek 27: Výsledná tabulka četností a histogram

Třídy ve výsledné tabulce jsou označeny svou horní hranicí, takže např. řádek označený číslem 15 obsahuje počet hodnot z intervalu od 5 do 15, pozor, hodnota 5 patří ještě do předchozí třídy. Hodnoty, které jsou větší než poslední mez, jsou zařazeny do třídy s názvem Další.

Zaškrtnutím volby **Pareto (tříděný histogram)** a **Kumulativní procentuální podíl** se ve výstupní tabulce objeví další sloupce. Volba **Pareto (tříděný histogram)** zobrazí třídy a jejich četnosti seřazené sestupně podle četnosti, tím lze hned vidět třídy s největší četností. Volba **Kumulativní procentuální podíl** zobrazí četnosti přepočtené na procenta, a navíc ještě kumulativně sečtené.

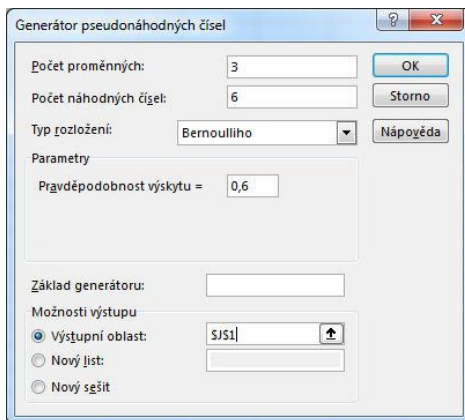
Při zaškrtnutí těchto posledních dvou voleb bude mít výsledný graf jiný vzhled.



Obrázek 36: Ukázka výsledných hodnot

Bernoulliho

U tohoto rozložení se používá pole **Pravděpodobnost výskytu**. Vyjadřuje se jako desetinné číslo z intervalu 0 až 1 (dá se říct, jako procento výskytu zapsané desetinným číslem, tedy 60 % odpovídá hodnotě 0,6).

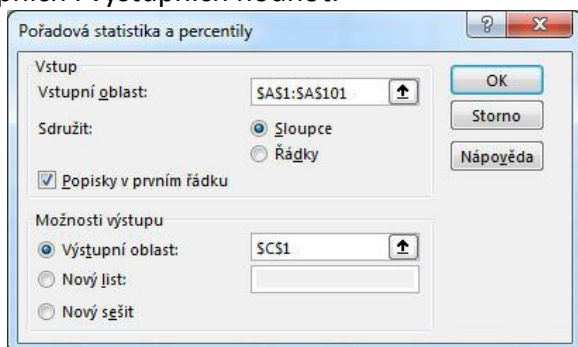


Obrázek 37: Vyplnění zadání

Pořadová statistika a percentily

Tento nástroj zjistí relativní pořadí hodnot ve skupině dat, kterou sledujete. Ukáže, která hodnota je první, druhá atd. Zároveň ukáže také procentní pořadí každé hodnoty skupiny dat pro snadnější analyzování relativního postavení hodnot v dané skupině.

Na listě MS Excelu existuje tabulka s počtem hodin strávených u počítače. Pokud se v dialogovém okně **Analýza dat** vybere položka Pořadová statistika a percentily, tak se zobrazí dialogové okno pro zadání vstupních i výstupních hodnot.



Obrázek 44: Dialogové okno pro pořadovou statistiku

Do pole **Vstupní oblast** se zadá oblast sledovaných dat, dále se řekne, zda jsou data ve sloupci či řádku a zda mají i popisky. Po označení výstupní oblasti se jen klepne na tlačítko **OK**.

	A	B	C	D	E	F
1	počet hodin u počítače		Bod	počet hodin u počítače	Pořadí	Percentil
2	42		71	75	1	100,00%
3	22		45	68	2	98,90%
4	19		65	67	3	96,90%
5	2		93	67	3	96,90%
6	8		78	65	5	95,90%
7	56		76	62	6	94,90%
8	0		34	60	7	93,90%
9	34		6	56	8	89,80%
10	34		16	56	8	89,80%
11	38		54	56	8	89,80%
12	24		84	56	8	89,80%
13	20		95	55	12	88,80%
14	28		27	48	13	87,80%
15	0		32	45	14	83,80%
16	33		33	45	14	83,80%

Obrázek 45: Výsledná tabulka z uvedených dat

Regrese

Pod položkou Regrese se skrývá analytický nástroj pro vyhodnocení lineárního vztahu mezi dvěma či více proměnnými. Pro lineární regresi je použita metoda nejmenších čtverců, která proloží přímkou sadou pozorování. Regrese tak umožňuje analyzovat, jakým způsobem ovlivňují hodnoty jedné či více **nezávisle proměnných** hodnotu jedné **závisle proměnné**.

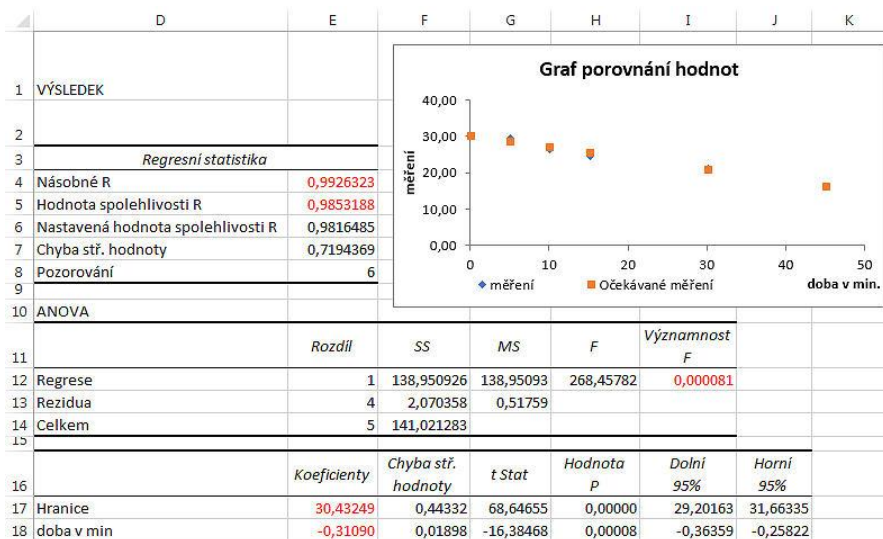
Například v tabulce laboratorního měření je třeba ověřit to, jak doba mezi odběrem vzorku a jeho analýzou může ovlivnit výsledek. To znamená, že čas je nezávislá veličina a výsledek měření je závislá.

	A	B
	výsledek měření na čase čekání	
1	do zpracování vzorku	
2	doba v min.	měření
3	0	30,44
4	5	29,86
5	10	26,87
6	15	24,86
7	30	21,34
8	45	16,58

Obrázek 46: Tabulka měření

V dialogovém okně **Analýza dat** se nyní vybere položka **Regrese**. Po klepnutí na tlačítko **OK** se zobrazí její dialogové okno, do kterého se zadávají hodnoty, s kterými se má počítat.

Pozor, první pole se nazývá **Vstupní oblast Y**, do které je třeba zadat závislé hodnoty. Teprve druhé pole je **Vstupní oblast X**, kam se zadávají hodnoty nezávislé. Excel čeká, že to budou sloupce se stejným počtem hodnot. Také je důležité označit to, zda sloupce mají či nemají popisky. To znamená nezapomenout zatrhnout volbu **Popisky**, pokud jsou označeny i nadpisy sloupců.



Obrázek 48: Výstupní tabulka s grafem

Druhá tabulka s názvem **Anova** lze zkontrolovat statistickou významnost analýzy. Když je hodnota **Významnost F** menší než **0,05**, tak výsledek se dá považovat za statisticky průkazný.

Při zaškrtnutí volby **Graf regresní přímky** se bude na vodorovné ose nacházet **nezávislá veličina X** a ose svislé **závislá veličina Y**. Přičemž první datová řada představuje původní hodnoty ze vstupní tabulky a druhá řada ukazuje hodnoty očekávané, které jsou vypočteny z rovnice regresní přímky. Pohledem lze tak zkontrolovat, jestli mají původní hodnoty skutečně lineární průběh a porovnat je s průběhem regresní přímky.

Na základě třetí tabulky se pak dá sestavit rovnice regresní přímky vyjadřující vztah mezi veličinami. Obecný tvar rovnice je **y=ax+b**, kde pak **koeficient a** se nachází ve sloupci **Koeficienty** na řádce veličiny **X** (doba v min) a **koeficient b** na řádce **Hranice**. V daném případě by byl tvar **y=-0,3109*x+ 30,4325**.

Tato rovnice není obvykle ideální, proto je užitečné si všimnout ještě sloupců s názvem **Dolní 95 %** a **Horní 95 %**. Pak by se do grafu dal udělat pás hodnot ohraničený právě těmito dvěma přímkami, které by vyjadřovaly určitý pás spolehlivosti.

Studentovy t-testy

Dvouvýběrový párový t-test na střední hodnotu

Studentův t-test pro dva výběry umožňuje ověřit, zda dva výběry mají v rámci stanovené hladiny významnosti stejnou střední hodnotu, nebo zda se jejich střední hodnoty liší o předem stanovenou hodnotu. Při tomto typu t-testu se nepředpokládá, že se rozptýly obou souborů rovnají. Párový test se často používá tehdy, když jsou pozorování či měření dat přirozeným způsobem spárována, například při dvojím testování skupiny (např. před nějakým pokusem a po něm).

	A	B
1	měření množství	
2	1. metoda	2. metoda
3	0,518	0,598
4	0,047	0,127
5	0,914	0,994
6	0,626	0,706
7	0,041	0,121
8	0,893	0,973
9	0,133	0,213
10	0,852	0,932
11	0,240	0,320
12	0,965	1,045
13	0,662	0,742
14	0,829	0,909
15	0,683	0,763
16	0,480	0,560
17	0,977	1,057

Obrázek 53: Ukázka vstupních dat

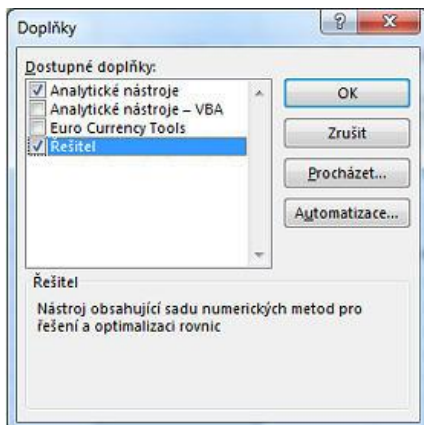
Po klepnutí na položku Dvouvýběrový párový t-test na střední hodnotu se zobrazí dialogové okno, do kterého je nyní třeba zadat vstupní údaje.

Řešitel – a co to vůbec je

Když jsem před několika lety poprvé učila tento doplněk v rámci kurzu MS Excelu, nazvali ho posluchači malou Sibylou. Proč asi? O co se vlastně jedná? Nevím, zda to pomůže objasnit oficiální definice, která říká, že Řešitel umí vypočítat řešení takzvaných citlivostních scénářů na základě upravitelných buněk a buněk s omezujícími podmínkami. Nevím, jestli se jí dá takto porozumět. Ale představte si, že je nutné rozhodnout o nějaké výrobě, prodeji či jiné službě, která je v plánu. Ale je potřeba zvážit, zda do toho jít, či ne. Chybí ale nějaké vyhodnocení, přitom víte, že jsou nějaké vstupní podmínky, například materiál, lidé, kteří mají určitou znalost, cena služeb související s prodejem,... Pokud je možné toto vše převést na nějaké matematické vyjádření pomocí soustavy linerárních rovnic o více neznámých, pak lze Řešitele použít, aby s rozhodnutím pomohl.

Kde se Řešitel nachází

Standardně tento nástroj v prostředí MS Excelu vidět není. Nachází se totiž mezi **Doplňky**, které lze kdykoliv podle potřeby zapnout či vypnout podobně jako **Analytické nástroje**.



Obrázek 61: Okno Doplňky s jednotlivými nástroji

To znamená, že v nabídce karty **Soubor** se musí klepnout na příkaz **Možnosti**. Poté v zobrazeném dialogovém okně se přejde na položku

Jak lze hledat řešení

Nadpis zní možná trochu zvláštně, ale MS Excel má opravdu v sobě nástroj, který se nazývá **Hledání řešení** (na kartě **Data** ve skupině **Prognóza** pod ikonou **Citlivostní analýza**).



Obrázek 69: Kde se Hledání řešení nachází

Jsou různé situace v životě, práci či podnikání, kdy se něco plánuje a bylo by užitečné, jak vše vyjde. Přičemž není potřeba hledat nějaká složitá řešení, ale jen nějaká konkrétní hodnota.

Například existuje plán uspořádat nějaký seminář nebo kurz, kde se ví, kolik by stál pronájem místnosti, lektor, reklama, případně další věci k tomu potřebné. Nyní je ale potřeba zjistit, kolik lidí by mělo přijít, aby se seminář zaplatil, aby nezůstal úplně ve ztrátě.

Toto je ideální situace, kde se dá právě nástroj Hledání řešení využít. On totiž slouží pro hledání odpovídajícího řešení pro různé situace na základě postupných kroků (tzv. iterací).

Čím se začne

Důležité je celou situaci zapsat do přehledné tabulky, kdy by v jednotlivých buňkách byly náklady rozepsány. Určitě je třeba pronajmout místnost, zajistit reklamu, stravu, ubytování a samozřejmě i odměnu lektorovi.

Toto všechno jsou náklady, na základě, kterých by bylo užitečné vyzkoušet to, kolik lidí a v jaké hodnotě vstupenek by celou akci zaplatilo. Později lze výpočet také zopakovat pro různé ceny

Přílohy

S čím se můžete ve statistických výpočtech setkat

Statistický soubor nebo statistický výběr

Jaký je rozdíl? Spousta lidí tyto pojmy zaměňuje. Představte si, že potřebujete otestovat nějakou vlastnost u výrobku, nebo chování lidí za určité situace. Není ale prakticky možné otestovat všechny výrobky, nebo zjišťovat chování všech lidí v daném městě či státě. Proto se v praxi testuje jen jejich náhodný vzorek. Co to znamená? Všechny výrobky by představovali statistický soubor a vybraný vzorek je pak statistický výběr. Stejně tak u lidí. Lidé z celého města či státu tvoří statistický soubor, ale náhodně vybraná skupina už je statistický výběr. Charakteristiky zjištěné u statistického výběru lze pak s jistou mírou nepřesnosti zobecnit na celý soubor.

Hypotéza

Většina analytických nástrojů v MS Excelu pracuje s testováním statistických hypotéz. Jak takové testování prakticky probíhá?

Nejprve se vždy stanovuje takzvaná nulová hypotéza (například hodnota X je nezávislá na hodnotě Y).

Poté se stanoví hypotéza opačná, které se říká alternativní hypotéza (například hodnota X je závislá na hodnotě Y).

Dále se stanoví hladina významnosti (Alfa).

Nakonec se právě testováním může nulová hypotéza potvrdit či zamítnout. Jestliže je nulová hypotéza zamítnutá, pak platí hypotéza alternativní.

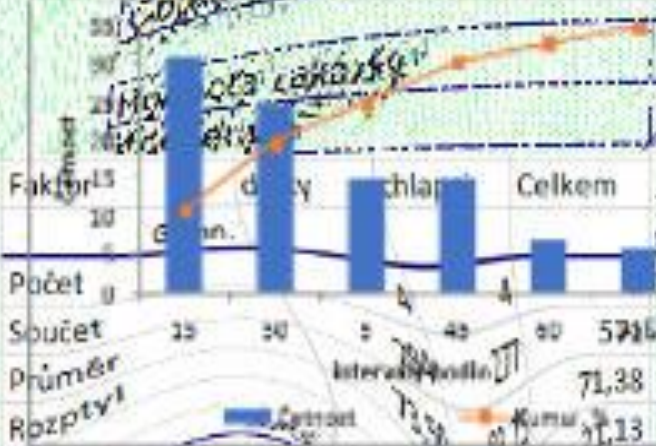
Hladina významnosti

Hladina významnosti by měla ukazovat pravděpodobnost, s jakou se neoprávněně zamítne nulovou hypotézu, i když ve skutečnosti platí. Standardně se volí výchozí hodnota hladiny významnosti na 0,05, tedy 5 %. Pokud se nulová hypotéza potvrdí, tak ji lze považovat za potvrzenou s 95 % spolehlivostí. Není-li k tomu nějaký speciální důvod, tak se většinou ponechává nastavená výchozí hladina významnosti.

Histogram



00000
 25.12.2023
 Novák



2 350 000 Kč
 46 000 Kč

Počet	Součet	Průměr	Rozptyl
4	157	49,25	45,58
4	180	45,00	3,33
8	377	47,13	26,13

Počet	Součet	Průměr	Rozptyl
8	509	63,63	9,13
4	293	63,00	15,25
4	257	64,25	27,50
8	377	47,13	26,13