

1FP416 - Počítačová podpora oceňování podniku

6. přednáška – Regresní analýza v excelu



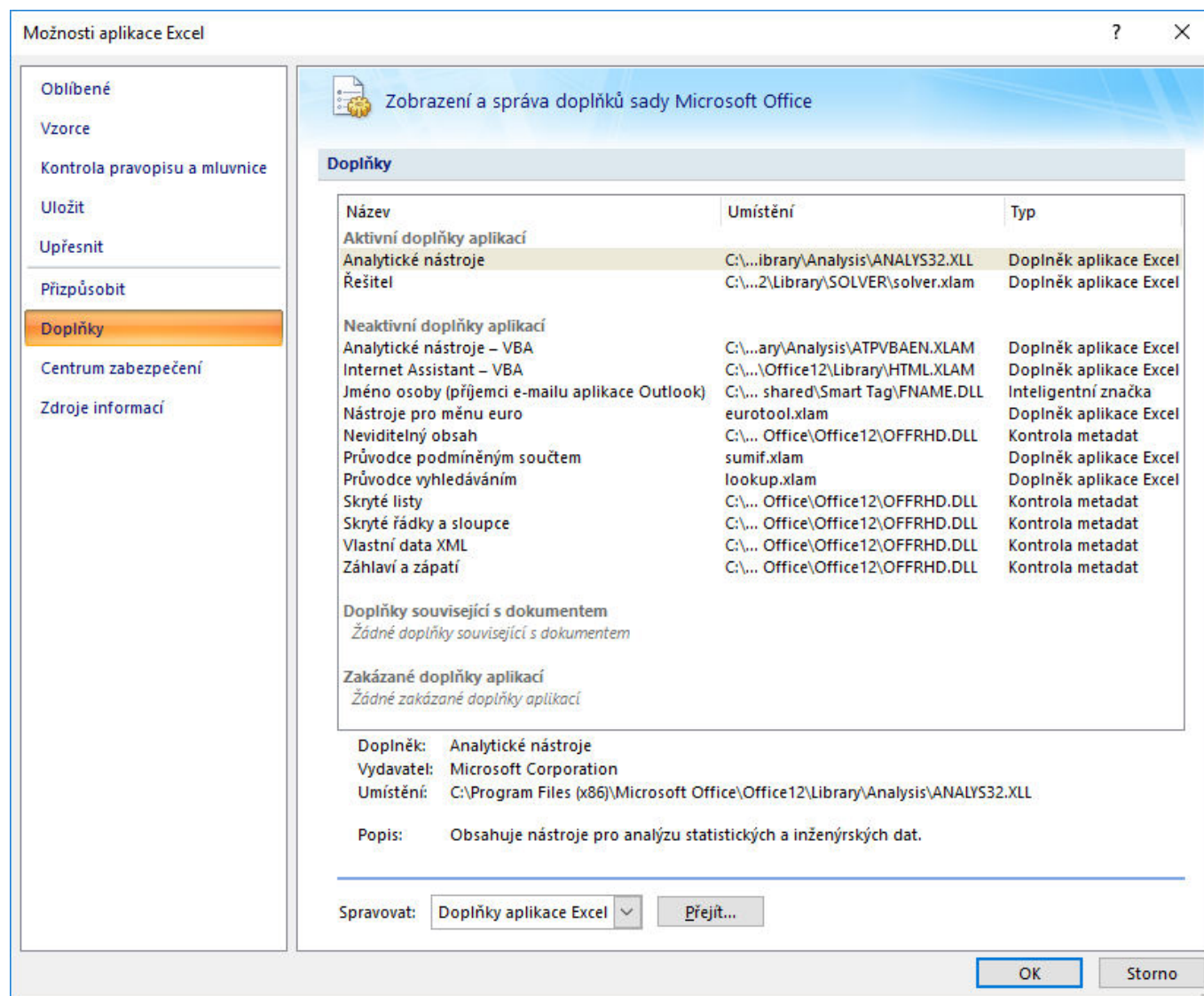
EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



Analytický nástroj Regrese

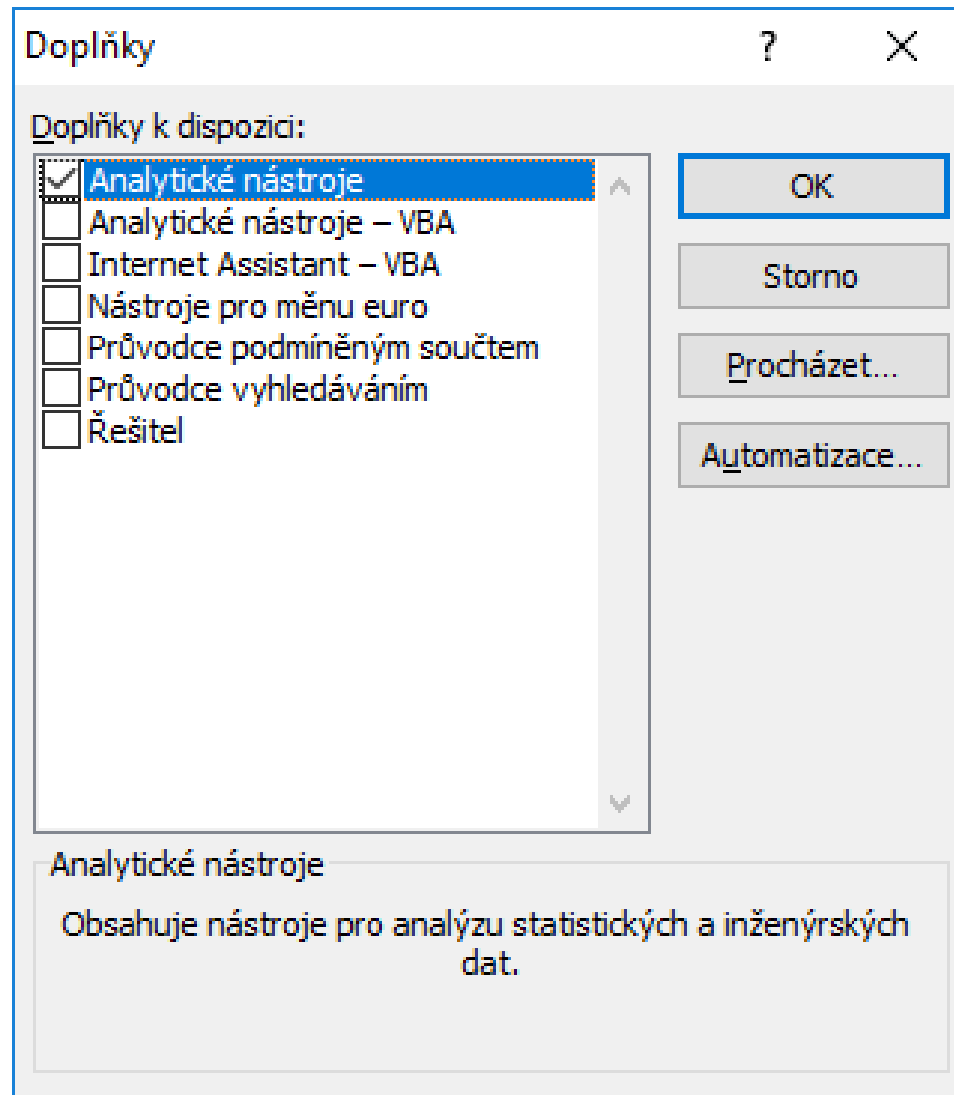
- Nejkomplexnějším nástrojem v MS Excel pro regresní analýzu je **analytický nástroj Regrese**, který je součástí doplňku **Analytické nástroje**. Kromě výše uvedených statistik navíc poskytuje informace o p-value regresních parametrů a regresního modelu, upravený koeficient determinace, hodnoty reziduí, graf regresní funkce apod.
- Nástroj Regrese je demonstrován na známé keynesiánské spotřební funkci, pomocí které zkoumáme závislost individuálních spotřebních výdajů na průměrných měsíčních příjmech v jednotlivých letech.

Doplněk Analytické nástroje



zdroj: autor

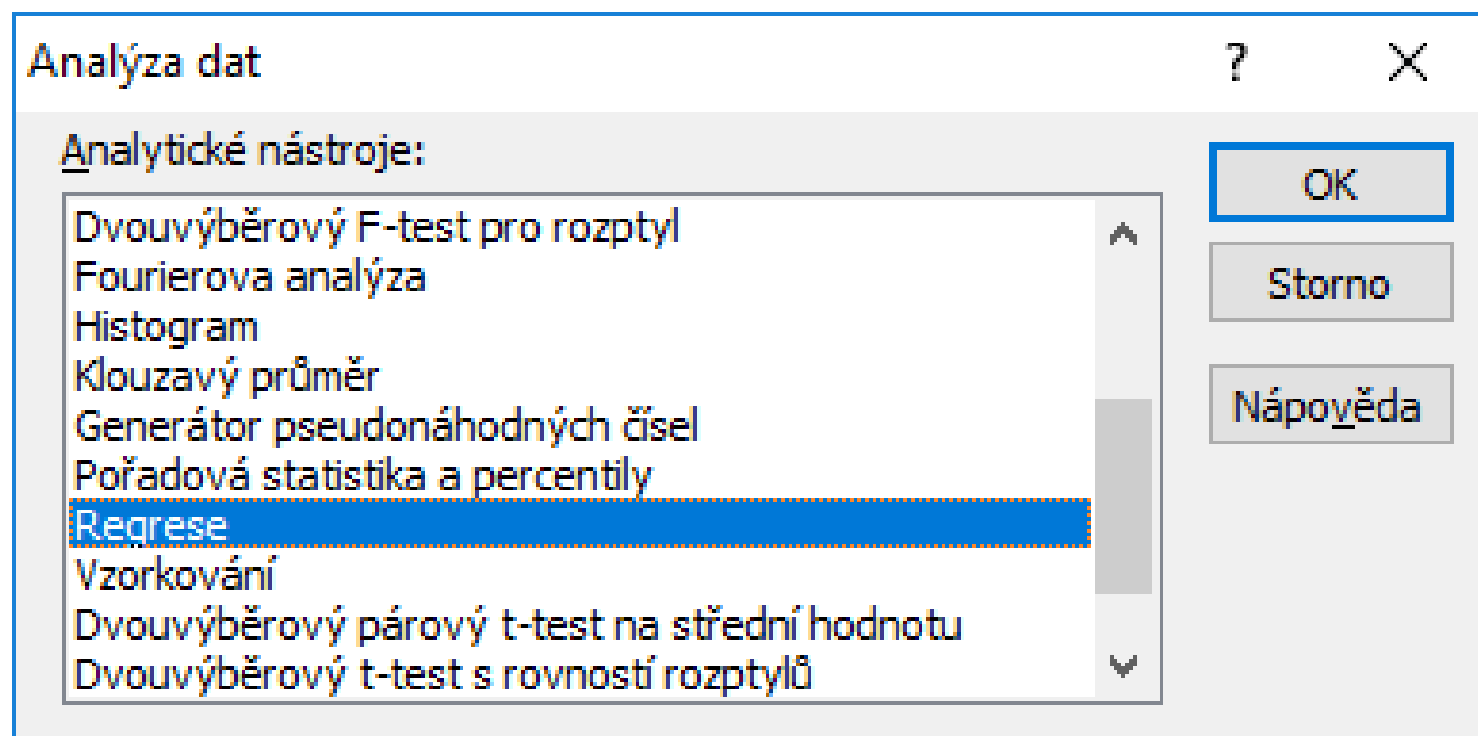
Doplněk Analytické nástroje



zdroj: autor

Možnosti nástroje Regrese

- Máte-li vyřešen předchozí krok, pokračujte stiskem tlačítka **Analýza dat** na kartě **Data** a vyberte nástroj **Regrese**.



zdroj: autor

Možnosti nástroje Regrese

- Zobrazí se okno, ve kterém vyplníte vstupní data a nastavíte možnosti výstupu. Nastavit můžete například hladinu spolehlivosti jinou než 95 %, model bez úrovně konstanty, graf regresní přímky, hodnoty reziduí apod.

zdroj: autor

Možnosti nástroje Regrese

- Základní a povinné vstupy jsou oblasti s hodnotami závisle proměnné y (Vstupní oblast Y), a hodnotami jedné nebo více nezávisle proměnných x (Vstupní oblast X). Výběr dalších vstupů závisí na požadavcích analytika a specifikaci modelu.

Vstup	Popis
Popisky	Obsahují-li zadané oblasti proměnných názvy sloupců.
Konstatnta je nula	Má-li být model bez úrovně konstanty, tj. $b_0 = 0$.
Hladina spolehlivosti	Změna hladiny spolehlivosti na požadovanou úroveň. Defaultně nastaveno na 95 %.

zdroj: autor

Možnosti nástroje Regrese

- **Hladina spolehlivosti** určuje s jakou pravděpodobností se pokoušíme prokázat statistickou významnost regresních parametrů (t-test) a regresního modelu (F-test). Hladina spolehlivosti se také používá pro intervalové odhady regresních parametrů.

Individuální spotřební funkce

- V následující tabulce jsou vstupní data pro odhad jednoduché spotřební funkce. Spotřební výdaje spotřebitele jsou reprezentovány závisle proměnnou y a disponibilní příjem nezávisle proměnnou x . Spotřební výdaje a disponibilní příjem jsou uvedeny jako průměrné roční hodnoty v tisících Kč.

Individuální spotřební funkce

	A	B	C	D
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				

Spotřební výdaje a příjmy v tis. Kč

Rok	y	x
2010	38,0	36,1
2011	43,7	49,2
2012	45,2	51,7
2013	46,0	53,9
2014	41,3	43,0
2015	48,1	59,0
2016	49,8	62,1
2017	51,8	67,0
Σ	363,9	421,9

www.financevpraxi.cz

Individuální spotřební funkce

- Výběrová spotřební funkce má lineární tvar $Y = b_0 + b_1x$. Jde o jednofaktorovou regresní funkci. Zajímají nás odhady parametrů b_0 a b_1 , jejich kvalita a kvalita modelu jako celku. Statisticky průkazný model může být použit pro předpověď budoucích hodnot spotřebních výdajů na základě očekávaného disponibilní příjmu v následujících letech.
- V našem konkrétním případě odhadujeme model s úrovnovou konstantou, vybrané oblasti dat neobsahují popisky a volíme hladinu spolehlivosti 99 %.

Individuální spotřební funkce

Regrese

Vstup

Vstupní oblast Y:

Vstupní oblast X:

Popisky Konstanta je nula

Hladina spolehlivosti %

Možnosti výstupu

Výstupní oblast:

Nový list:

Nový sešit

Rezidua

Rezidua Graf s rezidui

Standardní rezidua Graf regresní přímky

Normální pravděpodobnost

Graf pravděpodobnosti

OK

Storno

Nápověda

zdroj: autor

Výstup nástroje Regrese

- Výstup regresní analýzy obsahuje 3 základní tabulky, které se týkají kvality regresního modelu, rozptylu ANOVA, odhadů a kvality regresních parametrů. Detailněji je tato problematika rozebrána v článku [Testování regresního modelu](#).
- První tabulka s názvem Regresní statistika obsahuje vypočtené statistiky, které vypovídají o shodě regresního modelu s empirickými daty. Excel však používá zcela odlišnou terminologii než je zvykem, proto jsem k jednotlivým názvům dopsal standardní názvy.

Výstup nástroje Regrese

<i>Regresní statistika</i>		VYSVĚTLIVKY
Násobné R	0,999483758	= koeficient korelace
Hodnota spolehlivosti R	0,998967783	= koeficient determinace
Nastavená hodnota spolehlivosti R	0,998795747	= upravený koeficient determinace
Chyba stř. hodnoty	0,156869921	= standardní chyba odhadu Y
Pozorování	8	= počet pozorování

zdroj: autor

Výstup nástroje Regrese

- Druhá tabulka obsahuje Analýzu rozptylu ANOVA. ANOVA rozkládá rozptyl empirických hodnot závisle proměnné *Celkem* (TSS) na rozptyl vyrovnaných hodnot *Regrese* (ESS) a rozptyl reziduální složky modelu *Rezidua* (RSS).

	<i>Rozdíl</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Významnost F</i>
Regrese	1	142,8930383	142,8930383	5806,731064	3,4382E-10
Rezidua	6	0,147649034	0,024608172		
Celkem	7	143,0406873			

zdroj: autor

Výstup nástroje Regrese

- Další důležitou statistikou v rámci tabulky ANOVA je *F-test*, kterým se testuje statistická významnost regresního modelu. Hodnota vypočteného *F-testu* je 5 806,73. *Významnost F* představuje hladinu významnosti, na které ještě můžeme zamítnout nulovou hypotézu o statistické nevýznamnosti modelu. Tato hodnota se standardně označuje jako *F-value*. Hodnota *F-value* svědčí o vysoké statistické průkaznosti modelu.
- Třetí základní tabulka Odhady koeficientů obsahuje odhady regresních parametrů b_j , standardní chyby odhadů $s(b_j)$, testovací statistiku *t-test*, minimální hladinu významnosti *p-value*, na které ještě lze zamítnout nulovou hypotézu a intervaly spolehlivosti 95 % a 99 %.

Výstup nástroje Regrese

	<i>Koeficienty</i>	<i>Chyba stř. hodnoty</i>	<i>t Stat</i>	<i>Hodnota P</i>	<i>Dolní 95%</i>	<i>Horní 95%</i>	<i>Dolní 99,0%</i>	<i>Horní 99,0%</i>
Hranice	21,951	0,314	69,940	5,74827E-10	21,183	22,719	20,787	23,114
Soubor X 1	0,446	0,006	76,202	3,4382E-10	0,432	0,461	0,425	0,468
	b_j	$s(b_j)$	t-test	p-value	intervaly spolehlivosti 95 %		intervaly spolehlivosti 99 %	

zdroj: autor

Výstup nástroje Regrese

- Na základě odhadů b_0 a b_1 nyní můžeme sestavit výběrovou regresní funkci Y v indexním tvaru

$$Y_i = 21,951 + 0,466x_i$$

zdroj: autor

- Pomocí funkce můžeme pro konkrétní hodnotu x_i dopočítat vyrovnané hodnoty regresní přímky

$$Y_{2010} = 21,951 + 0,446 \cdot 36,1 = 38,0$$

$$Y_{2011} = 21,951 + 0,446 \cdot 49,2 = 43,9$$

$$Y_{2012} = 21,951 + 0,446 \cdot 51,7 = 45,0$$

⋮

$$Y_{2017} = 21,951 + 0,446 \cdot 67,0 = 51,9$$

zdroj: autor

Výstup nástroje Regrese

- Nad rámec základního výstupu nás ještě zajímají hodnoty reziduální složky. Ty můžeme využít pro testování existence autokorelace v podobě grafického vývoje v čase a výpočet [Durbin-Watsonova testu](#).

	<i>Pozorování</i>	<i>Očekávaná Y</i>	<i>Rezidua</i>
	1	38,0449726	-0,091068632
	2	43,9287466	-0,211674644
	3	45,0378468	0,113225216
	4	46,0239774	0,010886601
	5	41,1242095	0,219374472
	6	48,2636545	-0,1415425
	7	49,64508	0,115095991
	8	51,8554965	-0,014296504

zdroj: autor

Regresní analýza v Excelu (jeden z podkladů pro prognózu trhu) Excel 2007+

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
<i>1</i>	Rok	HDP	Inflace	Trh
<i>2</i>	2014	10 000	104	200
<i>3</i>	2015	11 000	102	202
<i>4</i>	2016	11 200	102	250
<i>5</i>	2017	12 000	101	270
<i>6</i>	2018	12 300	101	
<i>7</i>	2019	12 500	102	

zdroj: autor

Regresní analýza v Excelu (jeden z podkladů pro prognózu trhu) Excel 2007+

- **a) Časový trend**
- Oblast dat: D1:D5
- Vložení – Graf – spojnicový
- Klepnout na čáru funkce v grafu pravým tlačítkem – vyvolá se nabídka
- Příkaz: **Přidat spojnici trendu**
- Možnosti spojnice trendu: – lze zadat různé funkční vztahy (*zvolit*)
 - Zobrazit rovnici regrese (*zatrhnout*)
 - Zobrazit hodnotu spolehlivosti R (= koef. determinace) (*zatrhnout*) $\text{Korelační koeficient} = \sqrt{R}$
 - Odhad dopředu: x period *Zavřít*
- Pokud jsou regresní koeficienty v grafu vidět v matematickém formátu, lze klepnout pravým tlačítkem myši na box s rovnicí a vybrat jiný formát čísla (Příkaz Formát popisku spojnice trendu z místní nabídky).
- **Prognóza:** $25,8x + 166$ přepsat do buněk tabulky (x je budoucí rok 5, 6 atd.)
- Stejný postup můžeme použít i v případě jednoduché regrese, kdy je jedna nezávislá proměnná, ale místo spojnicového grafu použijeme graf XY (X = nezávislá, Y = závislá)

Regresní analýza v Excelu (jeden z podkladů pro prognózu trhu) Excel 2007+

- **b) Vícenásobná regrese**
- Karta Data → příkaz Analýza dat
- Pokud příkaz není v nabídce, pak tlačítko Office (2007) nebo karta Soubor (2010 a vyšší) → tlačítko Možnosti aplikace Excel → příkaz Doplnky → tlačítko Přejít (při nastavené volbě Doplnky aplikace Excel) → zatrhnout Analytické nástroje → OK
- Karta Data → Analýza dat → Regrese
 - Vstupní oblast Y: D1 : D5 (Trh)
 - Vstupní oblast X: B1 : C5 (HDP a inflace)
 - Popisky – *zatrhnout OK*
- Výsledek v novém listu:
 - Násobné R = korelační koeficient 0,954
 - Hodnota spolehlivosti R = koef. determinace ($R = \text{Korelační koeficient}^2$)
 - Koeficienty
- Do buněk pro prognózu trhu zapíšeme rovnici pomocí regresních koeficientů. V daném případě zapíšeme např. do D6 (prognóza trhu pro rok 2010) rovnici:
 $-7645,47 + 0,13*B6 + 62,88*C6$
- Příkaz Analýza dat – Regrese pracuje pouze s lineární regresí. Většinu funkcí lze ale na lineární tvar převést – popis postupu viz kniha Mařík, M. a kol.: Metody oceňování podniku 2007, 2011 nebo 2018, Příloha B.

Základní funkce v Excelu

- sum
- sumif
- sumifs
- countif
- countifs
- iferror
- vlookup
- hlookup
- index&match
- concatenate
- pivot table

SUMIF, SUMIFS

SUMIF

- Sečte buňky ve vybrané oblasti podle zadaného kritéria.

=SUMIF(oblast;kritéria;součet)

- Popis argumentů
 - oblast - vyhodnocovaná oblast buněk
 - kritéria- podmínka, která definuje sčítané buňky - může být "1", 1, "<1", ">1", "slovo"
 - součet- oblast ve které se sčítají buňky, pokud vyhovují kritériu v oblasti

SUMIF, SUMIFS

SUMIFS

Sečte buňky ve vybrané oblasti podle zadaných kritérií.

=SUMIFS(oblast_součtu, oblast_kritérií1, kritérium1 [oblast_kritérií2, kritérium2]...)

- Popis argumentů
 - **oblast_součtu** - jedna nebo více buněk pro výpočet součtu, včetně čísel nebo názvů - *Povinný argument*
 - **oblast_kritérií1** - První vyhodnocovaná oblast pro přidružené kritérium1 *Povinný argument*
 - **kritérium1** Podmínka pro první oblast (vyjádřená číslem, výrazem, ...) *Povinný argument*
 - **oblast_kritérií2** - Druhá vyhodnocovaná oblast pro přidružené kritérium2 *Nepovinný argument*
 - **kritérium2** - Podmínka pro druhou oblast_kritérií2, (vyjádřená číslem, výrazem, ...) *Nepovinný argument*

COUNTIF, COUNTIFS

Obě funkce spadají do kategorie **statistických funkcí**, které vrací počet buněk splňující jedno kritérium nebo množinu kritérií.

Funkce COUNTIF() vrací počet buněk, které splňují pouze jedno kritérium.

=COUNTIF(oblast; kritérium)

COUNTIF, COUNTIFS

Funkce COUNTIFS() je nadstavbou jednokriteriální funkce. Funkce vrátí počet buněk, které splňují více kritérií. Funkci je však možné použít i s jednou podmínkou. Argumenty funkce v hranatých závorkách nejsou povinné.

**=COUNTIFS(oblast_kritérií1; kritérium1; [oblast_kritérií2; kritérium2]
...)**

IFERROR

Funkci IFERROR můžete použít k zachycení a zpracování chyb ve vzorci. IFERROR vrátí hodnotu, kterou určíte, pokud je výsledkem vzorce chyba. jinak vrátí výsledek vzorce.

=IFERROR(hodnota, hodnota_při_chybě)

Syntaxe funkce IFERROR má následující argumenty:

- **Hodnota** Povinný argument. Jedná se o argument, ve kterém se testuje výskyt chyby.
- **value_if_error** Povinný argument. Hodnota, která se vrátí, pokud je výsledkem vzorce chyba. Jsou vyhodnoceny následující typy chyb: #N/A, #VALUE!, #REF!, #DIV/0!, #NUM!, #NAME? nebo #NULL!.

SVYHLEDAT (VLOOKUP) - VVYHLEDAT (HLOOKUP) - Funkce vyhledávací

Funkce SVYHLEDAT (VLOOKUP) vyhledá v prvním sloupci tabulky požadovanou/hledanou hodnotu a vrátí hodnotu, která je ve stejném řádku (jako hledaná hodnota), ale nachází se v jiném sloupci zdrojové tabulky.

=SVYHLEDAT(hledat;tabulka;sloupec[;typ])

- **hledat** - hodnota (číslo, logická hodnota, odkaz), hledaná v prvním sloupci tabulky.
- **tabulka** - prohledávaná tabulka u které je první sloupec prohledáván. Argument tabulka může být odkaz na oblast nebo název oblasti.
- **sloupec** - je celé číslo sloupce v argumentu **tabulka**, ze kterého se bude přiřazovat. Lze zadávat od 1 až max. hodnoty počtu sloupců v tabulce (tabulka). První sloupec má hodnotu 1.
- **typ** *nepovinný parametr* - nabývá logických hodnot (0 - NEPRAVDA - FALSE, 1 -PRAVDA - TRUE).
 - PRAVDA - (nezadáno/prázdné) - přesná shoda nebo nejvyšší hodnotu, která je menší než hledaná.
 - NEPRAVDA - vrátí hodnotu pokud najde přesnou shodu.

SVYHLEDAT (VLOOKUP) - VVYHLEDAT (HLOOKUP) - Funkce vyhledávací

Funkce VVYHLEDAT (HLOOKUP) vyhledá v prvním řádku tabulky požadovanou/hledanou hodnotu a vrátí hodnotu, která je ve stejném sloupci (jako hledaná hodnota), ale nachází se v jiném řádku zdrojové tabulky.

= VVYHLEDAT(hledat;tabulka;řádek[;typ])

- **Popis atributu funkce:**
- **hledat** - hodnota (číslo, logická hodnota, odkaz), hledaná v prvním řádku tabulky.
- **tabulka** - prohledávaná tabulka u které je první řádek prohledáván. Argument tabulka může být odkaz na oblast nebo název oblasti.
- **řádek** - je celé číslo řádku v argumentu **tabulka**, ze kterého se bude přiřazovat. Lze zadávat od 1 až max. hodnoty počtu řádků v tabulce (tabulka). První řádek má hodnotu 1.
- **typ** *nepovinný parametr* - nabývá logických hodnot (0 - NEPRAVDA - FALSE, 1 -PRAVDA - TRUE).
 - PRAVDA - (nezadáno/prázdné) - přesná shoda nebo nejvyšší hodnotu, která je menší než hledaná.
 - NEPRAVDA - vrátí hodnotu pokud najde přesnou shodu.

INDEX (INDEX)

Vrátí hodnotu nebo odkaz (index řádku, sloupce) na hodnotu z tabulky/oblasti.

= INDEX(pole;řádek [;sloupec])

=INDEX(odkaz; řádek; [sloupec]; [oblast])

Popis atributu funkce:

- **Odkaz** - na oblastí (jednu či více)
- **Řádek** - určuje řádek průsečíku
- **Sloupec**- určuje sloupec průsečíku
- **Oblast** - určuje oblast, ve které má ležet průsečík - v argumentu odkaz musí být patřičný počet oblastí

Match (POZVYHLEDAT)

Vyhledá zadanou položku v oblasti buněk a vrátí relativní pozici dané položky v oblasti.

= MATCH(co, prohledat, [shoda])

Popis atributu funkce:

- **co** - hodnotu, kterou chcete nalézt
- **prohledat** - kde se bude hledat, oblast buněk
- **shoda** - (-1, 0, 1)

CONCATENATE

Funkce **CONCATENATE**, která patří mezi [textové funkce](#), se používá ke spojení dvou nebo víc textových řetězců do jednoho řetězce.

= CONCATENATE(text1, [text2], ...)

text1 (povinný argument)

První položka určená ke spojení. Položka může být textová hodnota, číslo nebo odkaz na buňku.

text2;... (volitelné)

Další textové položky určené ke spojení. Maximální počet položek je 255, celkem 8192 znaků.

Děkuji za pozornost