ESF evropský sociální fond v ČR

EVROPSKÁ UNIE

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Střední průmyslová škola a Vyšší odborná škola technická Brno, Sokolská 1

Šablona: **Inovace a zkvalitnění výuky prostřednictvím ICT**

Název: **Databázové funkce tabulkového procesoru**

Téma: **MS Excel – závislost údajů - korelace**

Autor: **Ing. Kotásek Jaroslav**

Číslo: **VY_32_INOVACE_33–12**

Anotace: *Prezentace nás seznamuje pomocí základní teorie a objasňujícího příkladu s pojmem korelace. Žák se naučí počítat korelační koeficient u dvou sad údajů. Prezentace je určena pro žáky 2. ročníku technického lycea. Vytvořeno: prosinec 2012.*

Podpora digitalizace a využití ICT na SPŠ CZ.1.07/1.5.00/34.0632

Korelace

Ve statistice pracujeme často s dvojrozměrnými soubory, tj. se dvěma sadami údajů. Korelace pak analyzuje vzájemnou závislost těchto dvou sad údajů.

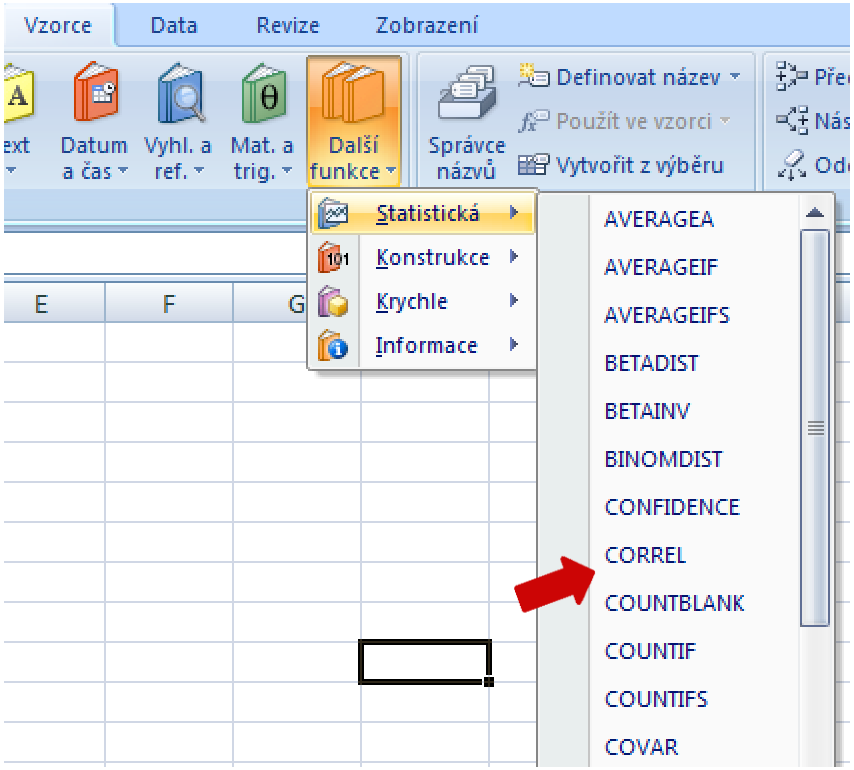
Danými sadami údajů může být například:

- údaj o hmotnosti a výšce osob (zde jistou závislost předpokládáme, vyšší člověk většinou i více váží)
- údaj o vzdělání a výši platu (zde můžeme také předpokládat přímo úměrnou závislost)
- údaj o věku a množství vykonané práce (zde je úměrná závislost pouze do určitého produktivního věku, s rostoucím věkem pak množství vykonané práce klesá)
- údaje ze školní klasifikace – závislost průměrného prospěchu na absenci, průměrného prospěchu na určité známce aj.

Z matematického hlediska se korelace definuje jako kovariance dělená součinem směrodatných odchylek obou statistických souborů.

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

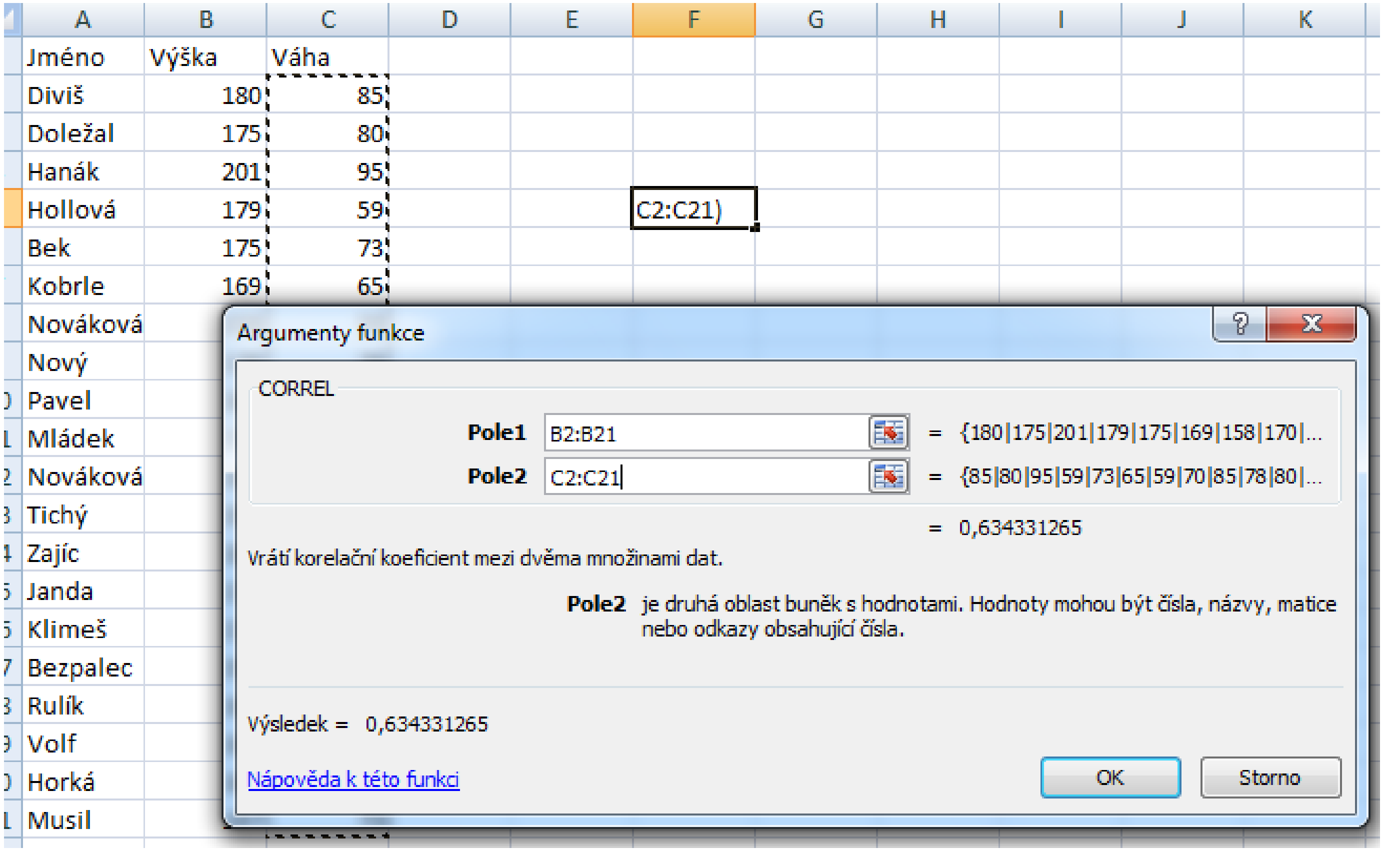
Výsledek se vždy pohybuje v intervalu [-1, 1], v případě zcela přímé závislosti platí, že existují koeficienty a, b takové, že y=ax+b a výsledek je 1, v případě zcela nepřímé závislosti existují a, b takové, že y=-ax+b a výsledek je -1.
Funkci na výpočet korelačního koeficientu CORREL najdeme mezi statistickými funkcemi.



Obrázek 1: Aktivace funkce CORREL.

	A	B	C
1	Jméno	Výška	Váha
2	Diviš	180	85
3	Doležal	175	80
4	Hanák	201	95
5	Hollová	179	59
6	Bek	175	73
7	Kobrle	169	65
8	Nováková	158	59
9	Nový	170	70
10	Pavel	182	85
11	Mládek	174	78
12	Nováková	186	80
13	Tichý	190	92
14	Zajíc	180	73
15	Janda	176	68
16	Klimeš	170	80
17	Bezpalec	190	80
18	Rulík	183	72
19	Volf	183	75
20	Horká	180	59
21	Musil	187	73

Obrázek 2: Tabulka s vhodnými daty pro výpočet korelačního koeficientu.



Obrázek 3: Realizace funkce CORREL.

U tohoto příkladu je vidět poměrně výrazná závislost mezi údaji Výška a Váha. Korelační koeficient je 0,63 – veličiny jsou přímo úměrné.

	A	B	C	D	E	F	G	H	I	J
1		Průměr	MAT	VYT	Absence					
2		3,25	3	3	61					
3		2,25	3	1	60		Korelace průměr a absence			0,117759
4		2,08	3	2	57		Korelace Průměr a MAT			0,771648
5		1,5	2	1	0		Korelace Průměr a VYT			0,836706
6		2,58	4	3	34		Korelace MAT a VYT			0,675994
7		3,33	4	4	52		Korelace VYT a absence			0,181862
8		1,92	2	1	33					
9		2,33	2	3	90					
10		2	2	1	36					
11		1,83	2	1	24					
12		1,83	2	2	8					
13		2,17	2	2	82					
14		2,17	3	2	8					
15		2,08	2	2	108					
16		1,75	2	1	45					
17		2,5	3	2	33					
18		1,17	1	1	30					
19		2,33	3	2	34					
20		2,42	3	2	35					
21		3,83	4	4	24					
22		2,08	1	1	15					
23		2,25	3	2	68					
24		3,08	4	2	7					
25		3,33	3	3	64					
26		2,08	3	1	28					
27		2,67	2	2	34					
28		2,08	2	1	69					
29		1,75	2	1	40					
30		2,5	3	2	69					
31		1,42	1	1	22					
32		2,75	3	3	8					

Obrázek 4: Vyžití funkce CORREL pro více vztahů vždy mezi dvěma veličinami. Jedná se o výsledky pololetní klasifikace žáků jedné třídy.

U tohoto příkladu vidíme, že nejzávislejší vztah je mezi průměrnou známkou a známkou z výpočetní techniky, naopak velmi slabá přímá úměrná závislost je mezi průměrem a absencí.

	A	B	C	D	E	F
1	Jméno	Výška	Váha			
2	Diviš	180	85			
3	Doležal	170	80			
4	Hanák	200	95			
5	Hollová	128	59			1
6	Bek	156	73			
7	Kobrle	140	65			
8	Nováková	128	59			
9	Nový	150	70			
10	Pavel	180	85			
11	Mládek	166	78			
12	Nováková	170	80			
13	Tichý	194	92			
14	Zajíc	156	73			
15	Janda	146	68			
16	Klimeš	170	80			
17	Bezpalec	170	80			
18	Rulík	154	72			
19	Volf	160	75			
20	Horká	128	59			
21	Musil	156	73			

Obrázek 4: Vyžití funkce CORREL při úplné závislosti.

B2 | fx | =C2*2+10

	A	B	C	D	E
1	Jméno	Výška	Váha		
2	Diviš	180	85		
3	Doležal	170	80		
4	Hanák	200	95		
5	Hollová	128	59		
6	Bek	156	73		
7	Kobrle	140	65		

Obrázek 5: V buňkách ve sloupci B je vždy stejný vzorec s relativním posunutím.

Zde je úplná závislost výšky na váze. Stačí se podívat, jakým způsobem se z pole Váha počítá pole Hmotnost a vše je jasné.

Doplňující příklad:

Pokuste se vymyslet dvě sady datových údajů, které jsou pokud možno co nejvíce nepřímo závislé.