

Bonusový úkol 3: Korelace

Stáhněte si (nebo najděte z hodiny) dataset “gapminder.xlsx” ([stáhnutí zde](#)). Budeme zkoumat vztah mezi GDP na hlavu a očekávanou délkou dožití, a to ve dvou časových bodech, v roce 1957 a 2007.

1. Nejdříve příslušné proměnné graficky zobrazte (graf musí mít veškeré formální náležitosti, tj. název, popisky os, zřejmé jednotky). Pro zobrazení takového vztahu je nejlépe použít bodový graf (scatter plot), který jste viděli na přednášce. Excel vám ho pravděpodobně nabídne jako jeden z doporučených. Extra: nastavte popisky bodů tak, aby zobrazovaly zemi (přes Popisky dat -> Další možnosti).
2. Vypočítejte Pearsonův a Spearmanův korelační koeficient (viz návod níže).
3. Výsledky (grafy a koeficienty za roky 1957 a 2007) přehledně představte a určete, ve kterém roce byl vztah silnější. Zkuste určit, který korelační koeficient je v této situaci vhodnější.

K práci přiložte i **používaný soubor v Excelu**, kde budou zadané funkce a vytvořené grafy. Komentáře je možné psát přímo do Excelu (přes “Vložit textové pole”), nebo odevzdat jako soubor typu .doc nebo .pdf. Do názvu dokumentu zahrňte prosím své příjmení.

Počítání koeficientů v Excelu:

Pearsonův koeficient je jednoduchý, existují prakticky identické funkce =CORREL() nebo =PEARSON().

Spearmanův koeficient je složitější, bohužel neexistuje přímá funkce na jeho výpočet (v základní verzi Excelu). To ale nevadí, protože výpočet můžeme provést na základě znalosti vztahu mezi oběma koeficienty. Připomeňme si z přednášky, že Spearmanův koeficient je korelační koeficient aplikovaný na pořadí jednotlivých pozorování. To už se v Excelu udělá velice snadno. Na určení pořadí existuje funkce =RANK.AVG(), do které je třeba zadat tři parametry - porovnávanou hodnotu (tu, pro niž počítáte pořadí), celkový soubor hodnot (v rámci kterého chceme určit pořadí, pozor na zafixování oblasti všech dat pomocí znaku \$) a zda má jít sestupně (0) nebo vzestupně (1).

Zadaná funkce na ukázkových datech:

DOSADIT						fx	=RANK.AVG(A1;\$A\$1:\$A\$6;1)					
	A		B		C		D		E			
1	5		4		=RANK.AVG(A1;\$A\$1:\$A\$6;1)							
2	6		7									
3	2		6									
4	7		2									
5	9		8									
6	3		4									

Na základě sloupečků pořadí již pak spočítáte snadno Pearsonův koeficient. Podrobný návod v angličtině najdete například zde:

<https://www.real-statistics.com/correlation/spearmans-rank-correlation/>

Alternativním, i když ne nutně doporučeným způsobem, je přefiltrovat obě proměnné nezávisle na sobě podle pořadí vzestupně a pak vedle nich vytvořit sloupec 1, 2, 3, ... Spočítáním Pearsonova koeficientu na těchto hodnotách získáte Spearmanův pro původní proměnné.