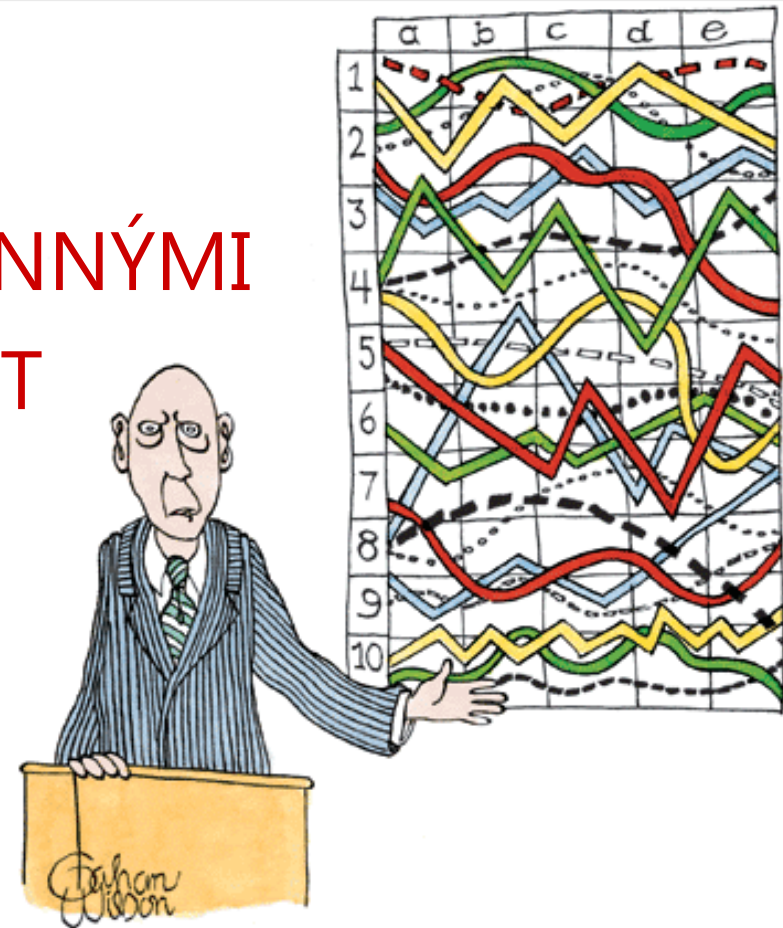

VZTAHY MEZI PROMĚNNÝMI

KORELAČNÍ KOEFICIENT



"I'll pause for a moment so you can let this information sink in."

Výzkumné otázky...

- Hypotézy o vzájemné souvislosti jevů:
 - Predikuje intelekt akademický úspěch?
 - Mají dobří češtináři i dobré známky z matematiky?
 - Existuje souvislost mezi mírou depresivní a anxiózní symptomatiky?
 - Jsou měsíční příjem a délka pracovní doby dobrými prediktory životní spokojenosti?
 - Jsou různá umělecká nadání specifická, nebo vycházejí ze stejného „všeobecného“ talentu?
-

Vztahy vs. kauzální vztahy

- Pozorujeme zvýšenou pravděpodobnost společného výskytu 2 jevů (hodnot)
 - Úsudek na příčinu je **problematický** – založen na teorii a výzkumném designu, který vyprodukoval data
 - Úsudek na příčinu je **potřebný**
 - tvorba teorie, úsudek o stabilitě v čase
 - rozlišování mezi náhodou a pravidelnostmi
-

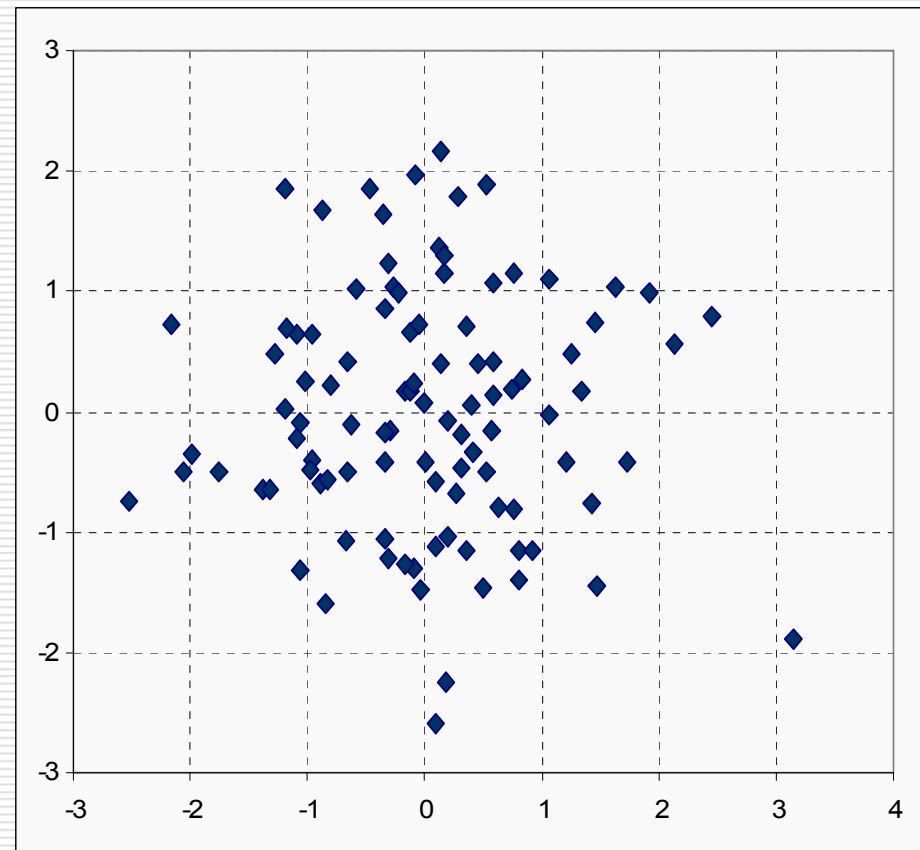
Statistické zachycení vztahu

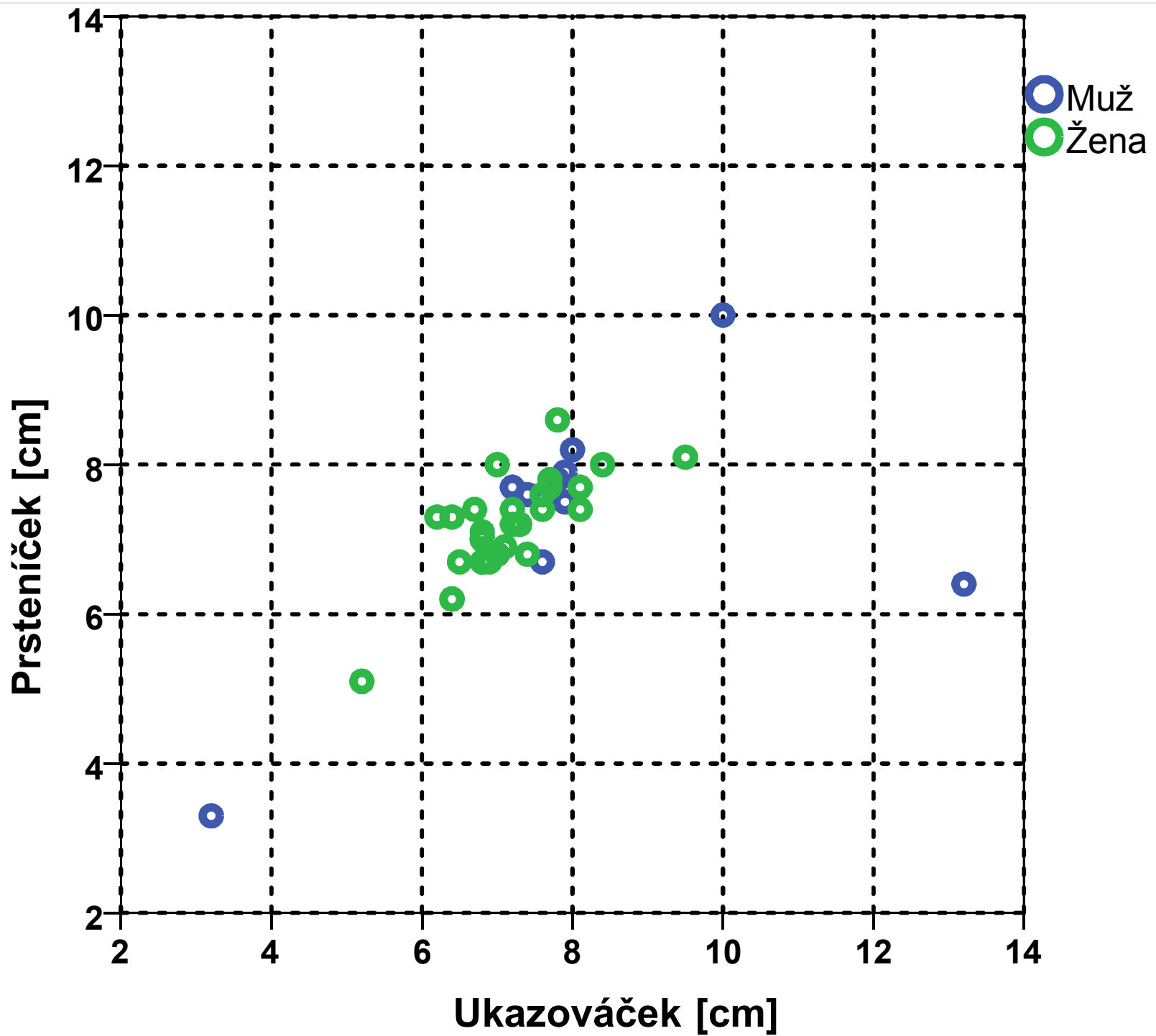
- ... závisí na úrovni měření vztahovaných proměnných – kategorické X metrické

	Kategorická	Metrická
Kategorická	Kontingenční tabulka Složený sloupcový diagram <i>Chí-kvadrát</i>	
Metrická	Složené podoby jednorozměrných zobrazení <i>Rozdíl popisných statistik</i>	Bodový diagram <i>Korelace</i>

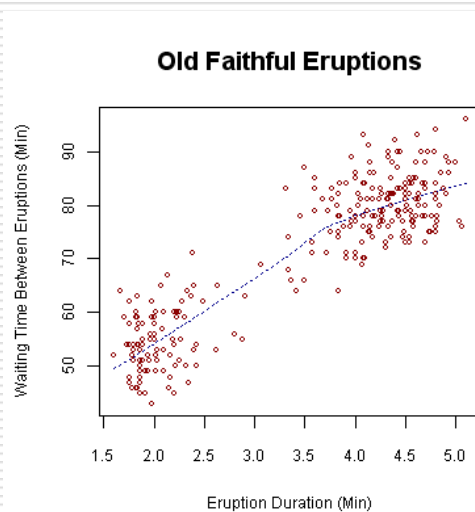
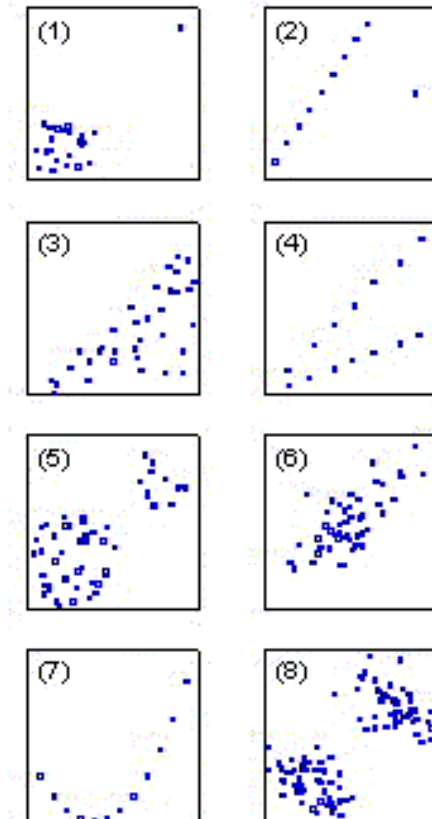
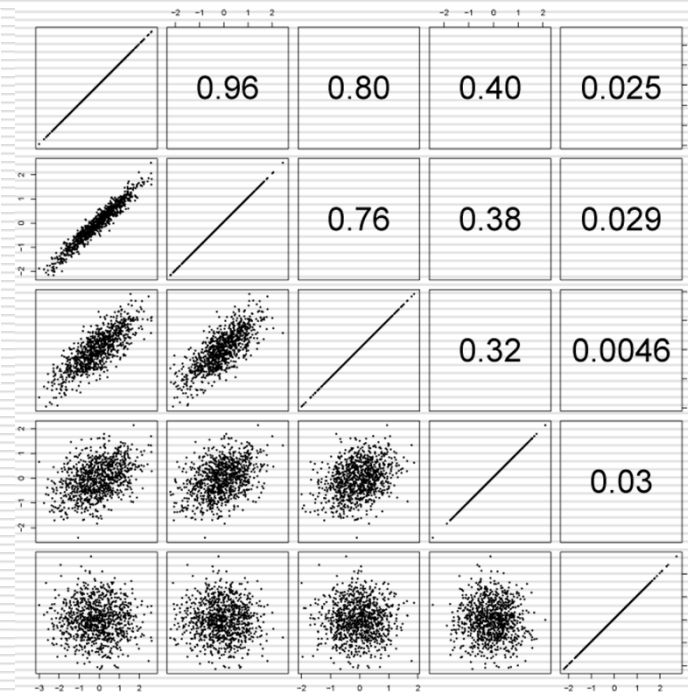
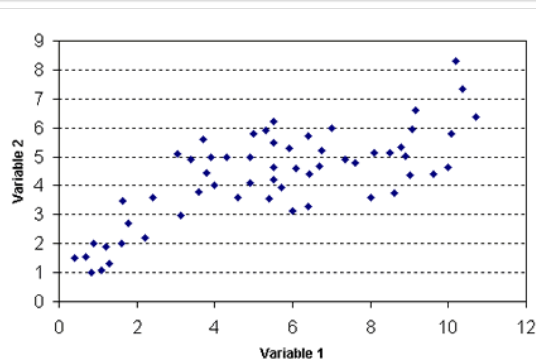
Bodový graf - scatterplot

- Bodový graf – scatterplot(scattergram)
- Nahrazuje kontingenční tabulku, jsou-li obě proměnné spojité; pro proměnné s málo body měření nemá smysl
- Každá osa reprezentuje jednu proměnnou, každý bod je jedna zkoumaná osoba (jednotka)
- Poskytuje tím lepší evidenci o vztahu dvou proměnných...
 - ...čím více měření jsme provedli
 - ...čím přesnější jednotlivá měření byla
- Počet stejných měření může reprezentovat např. velikost bodu
 - Frequency scatterplot





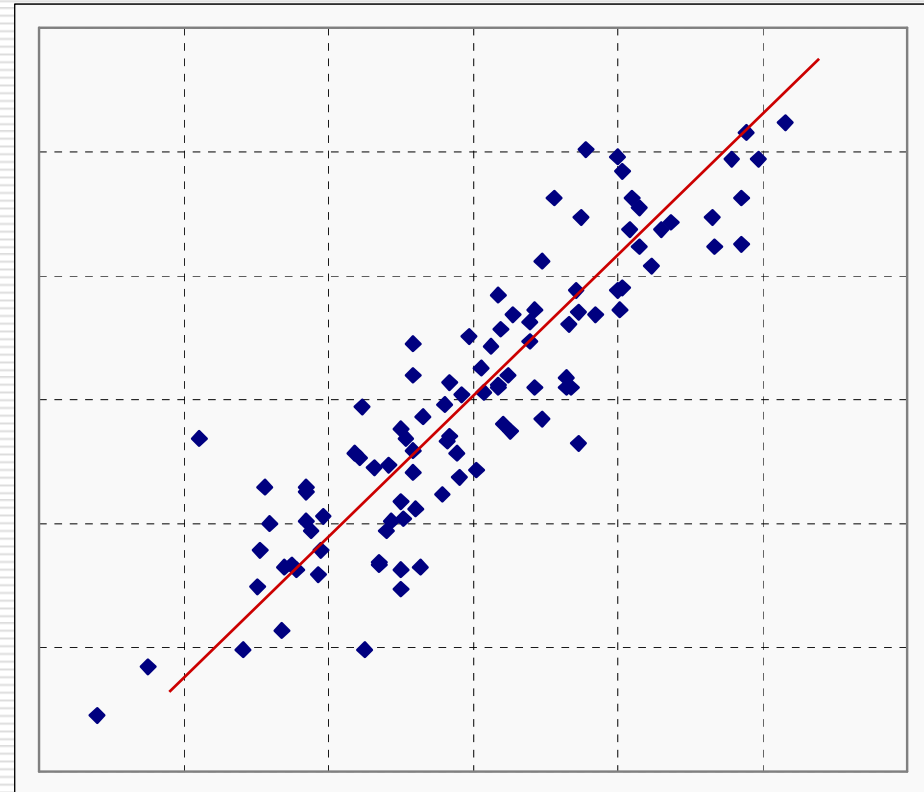
Různé podoby/druhy vztahu



Pouze takto vypadající scattery zobrazují vztah mezi 2 proměnnými, který je lineární a dobře (=smysluplně, výstižně) popsateľný pomocí Pearsonova korelačního koeficientu. U ostatních jde buď o vztahy nelineární, nebo je problém v heterogenitě, outlierech...

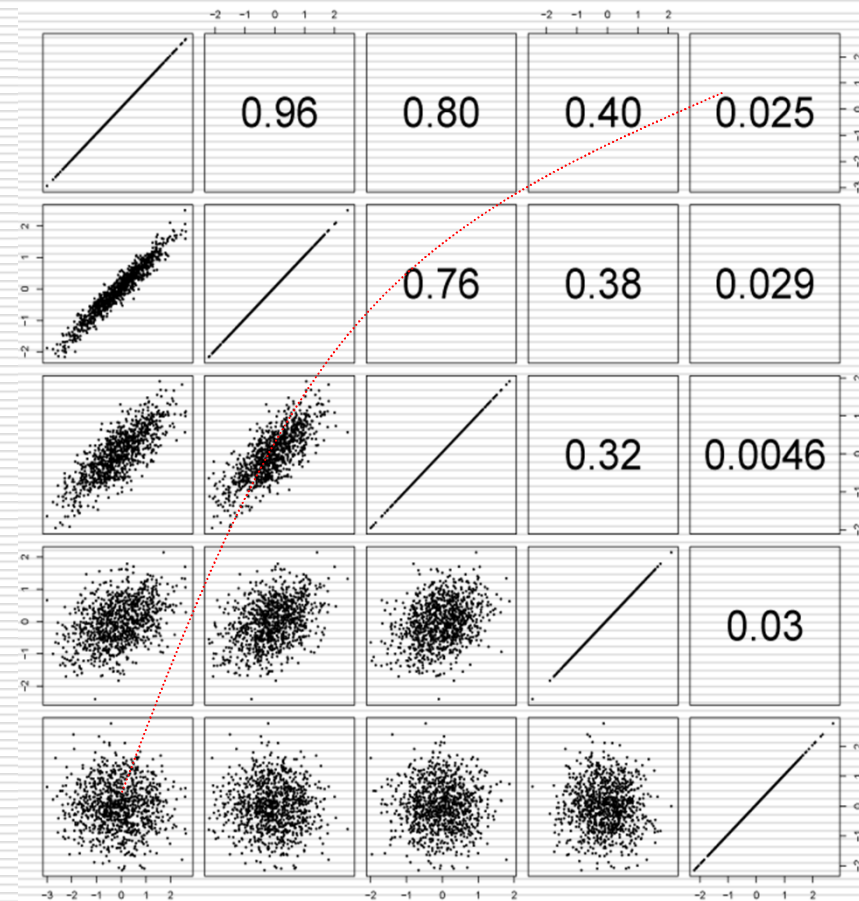
Lineární souvislost, vztah

- Lineární vztah je to, co se obvykle míní slovem korelace.
- Je to monotónní vztah, který se dá popsat slovy čím více X, tím více/méně Y.
- Projevuje se tak, že scatterplot se dá proložit „ideální“ přímkou
 - $y = ax + b$Tato funkce/přímka popisuje strmost vztahu.
Korelace popisuje **těsnost** vztahu.



Těsnost vztahu

- Čím těsnější (=intenzivnější, silnější) vztah 2 proměnných je, tím jsou body více nahuštěny okolo nějaké přímky
- Těsnost nesouvisí se sklonem té přímky, ale pouze s tím, jak moc se scatterplot podobá přímce.
- Těsnost se udává bezrozměrným číslem od 0 do 1, kde 0=žádný vztah(těsnost) a 1= maximální vztah (data na diagonále v obrázku napravo)
- Znaménko udává, zda jde o vztah čím víc, tím víc (+) nebo o vztah čím víc, tím míň (-)
- Rozsah je tedy od -1 do 1
- Těsnost -> kovariance



Korelace (=standardizovaný sdílený rozptyl)

- Chceme-li se zbavit obtížně interpretovatelných jednotek u kovariance, dosáhneme toho podobně jako při výrobě z-skórů – podělením deviačního skóru příslušnou směrodatnou odchylkou (=standardizace)

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{X_i - m_x}{s_x} \right) \left(\frac{Y_i - m_y}{s_y} \right) = \frac{c_{xy}}{s_x s_y}$$

- Zakroužkovanou část vzorce už ale známe – to je transformace na z-skór. Korelace jednodušeji je tedy:

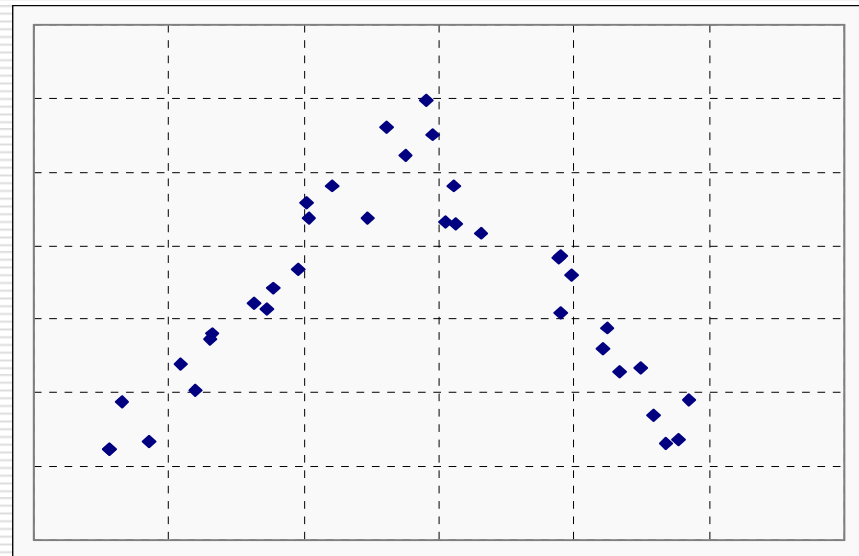
$$r_{xy} = \frac{\sum_{i=1}^n z_{X_i} z_{Y_i}}{n-1}$$

Vlastnosti popsaneho koeficientu korelace I.

- Jde o tzv. Pearsonův součinnový, momentový koeficient korelace
 - patří tedy do kategorie momentových ukazatelů (viz předchozí přednáška) a platí pro něj podobné věci:
 - nutná intervalová a vyšší úroveň měření
 - velký vliv odlehlých hodnot na výsledek
 - je vhodný pro popis normálně rozložených proměnných
 - vyjadřuje pouze sílu(těsnost) lineárního vztahu
- Nabývá hodnot v rozmezí -1 až 1
 - 0 = žádný vztah
 - 1(-1) = dokonalý kladný (záporný) vztah = identita proměnných = přímá úměra
- Korelace nepopisuje funkční vztah dvou proměnných, ale pouze jeho těsnost.

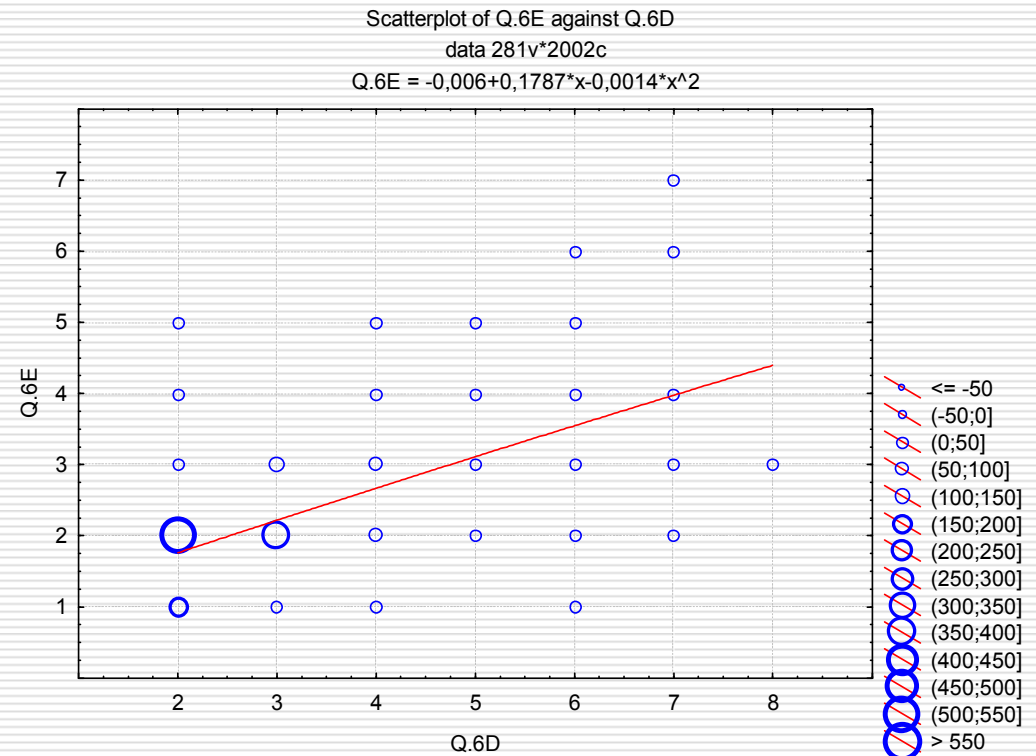
Vlastnosti Pearsonova koeficientu korelace II.

- r^2 = koeficient determinace (někdy D , R^2)
 - = proporce sdíleného rozptylu
 - V důsledku toho:
 $0,3-0,1 \neq 0,7-0,5$
- $r = 0$ neznamená, že mezi rozděleními proměnných není žádná souvislost, znamená pouze, že mezi nimi není *lineární* vztah.



Vlastnosti Pearsonova koeficientu korelace III.

- Kdy nemá korelace smysl?
 - V1: Kolik hodin denně sledujete televizi?
 - V2: Kolik hodin denně sledujete televizní zpravodajství?
 - Proč? 😊
- Korelace proměnných se společnou příčinnou:
 - Swoboda: platy kněží a ceny vodky v průběhu doby korelují!
 - IQ dětí a velikost a jejich výška prý také...
 - ... kovariance proměnných se společnou příčinnou je základem dalších metod analýzy dat v psychologii: analýzy reliability a faktorové analýzy.



Korelační koeficienty pro pořadová data

(podrobněji přednáška 7)

- vhodné nejen pro pořadová data, ale i pro intervalová, která mají rozložení výrazně odlišné od normálního
- zachycují i nelineární monotónní vztahy (viz Hendl, s260)
- ukazatele toho, nakolik jsou pořadí podle korelovaných dvou proměnných stejná
- Spearmanův koeficient ρ – r_s
 - založený na velikosti rozdílů v pořadí
 - ekvivalentem Pearsonova koeficientu na pořadových datech
 - lze interpretovat r^2
- Kendallův koeficient tau – τ (s variantami „b“ nebo „c“)
 - založený na počtu hodnot (prvků výběrového souboru) mimo pořadí
 - vyjadřuje spíše pravděpodobnost, že se prvky výběrového souboru uspořádají podle obou proměnných do stejného pořadí