

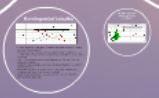
Zobrazení dvourozměrných dat, korelační koeficient

Dvourozměrná analýza
 - jaké vztahy mezi dvěma proměnnými
 - lze jako máv jedna proměnná ovlivňuje druhou proměnnou?
 - Proč jsou vztahy pozorovány (zajímavé)?
 - jaké jsou možnosti interpretace vztahů?
 - jaké proměnné ovlivňuje druhou?
 - jaké proměnné ovlivňuje druhou?
 - jaké proměnné ovlivňuje druhou?
 - jaké proměnné ovlivňuje druhou?
 - jaké proměnné ovlivňuje druhou?



Číselná a vizuální prezentace
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými

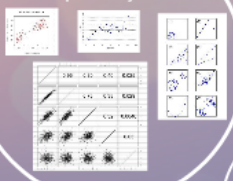
Kontingenční tabulka
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými



Bodový graf - scatterplot
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými



Různé podoby vztahu mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými



Korelace
 Standardizovaný sdílený rozptyl
 Pearsonův součinný, momentový koeficient korelace
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými

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

1. Měří sílu nelineárních korelačních součinných koeficientů na nejnižší (nejvyšší) vztah? a) 0,95 b) 0,99 c) -0,77 d) 0,1 e) 1,05
2. Při reprezentativní vzorku se ve věku 15, 33, 30, 41 a 54 let provalil občan na moře. Počet vyvolání pohotovosti je 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100.
3. Korelace mezi X a Y je 0,85. Koeficient mezi X a Y je 0,85. Koeficient mezi X a Y je 0,85. Koeficient mezi X a Y je 0,85. Koeficient mezi X a Y je 0,85.

Interpretace vztahů
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými
 - jaké vztahy mezi dvěma proměnnými



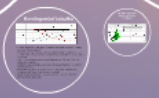
Zobrazení dvourozměrných dat, korelační koeficient

Dvourozměrná analýza
 - jaké vztahy mezi dvěma proměnnými
 - lze jako máv jítka považovat vztahy dvojích proměnných?
 - Proč jsou vztahy pozorovány (zároveň)?
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?



Číselná a vizuální prezentace
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?

Kontingenční tabulka
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?



Bodový graf - scatterplot
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?

Různé podoby vztahu mezi dvěma proměnnými
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?

Korelace
 Standardizovaný sdílený rozptyl

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

Pearsonův součinný, momentový koeficient korelace
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?

1. Měří sílu nelineárních korelačních vztahů
 2. Měří sílu nelineárních korelačních vztahů
 3. Měří sílu nelineárních korelačních vztahů

Lineární korelace
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?
 - jaké vztahy pozorovat (zároveň)?

Dvourozměrná analýza

- zkoumá vztahy mezi dvěma proměnnými
- **Do jaké míry jedna proměnná ovlivňuje druhou proměnnou?**

Př.

- Predikuje intelekt akademický úspěch?
- Mají dobří češtináři i dobré známky z matematiky?

"Jedna proměnná ovlivňuje druhou=

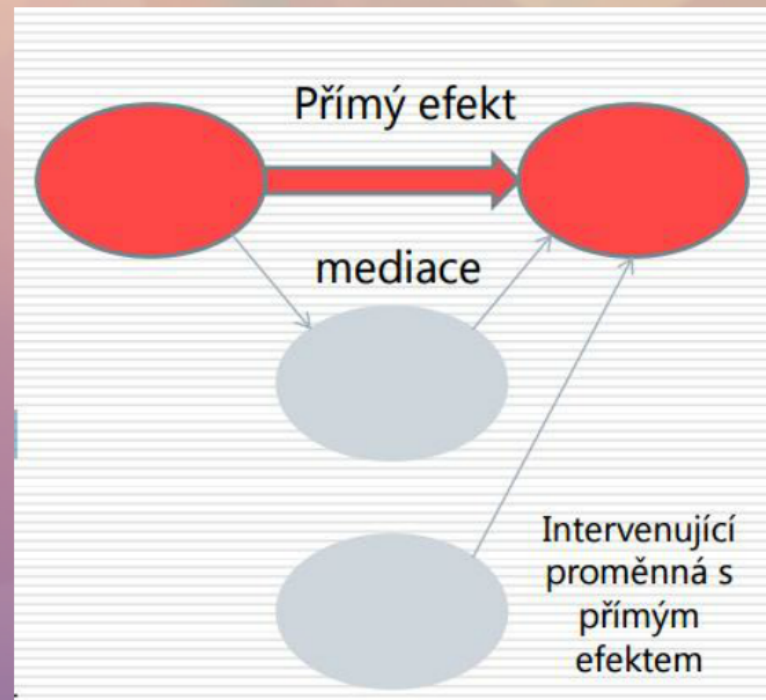
- mezi proměnnými existuje vztah, pokud rozložení hodnot jedné proměnné je asociováno s rozložením hodnot druhé proměnné

Statistická závislost

- hodnotě jedné veličiny (proměnné) odpovídá celké množství hodnot jiné veličiny
- př. výška žáků se s přibývajícím věkem zvětšuje (ale nelze tvrdit, že určitému věku přináleží určitá výška)

- cílem v
- prověřov
- v huma
ambiciózn

- cílem výzkumu je obvykle prověřovat kauzální vztahy
- v humanitních vědách velmi ambiciózní



Závislá a nezávislá proměnná

Nezávislá proměnná

- jejím chováním se vysvětluje chování závislé proměnné
- příčinná proměnná - v důsledku jejich změny se mění vysvětlovaná proměnná.

Závislá proměnná

- její chování se snažíme vysvětlit
- mění se v důsledku chování nezávislé proměnné

Intervenující proměnná

- zasahuje do vztahu mezi závislou a nezávislou proměnnou a ovlivňuje je
- obvykle není možné identifikovat všechny intervenující proměnné

Kontingenční tabulka

- způsob, jak popsat dvourozměrná rozdělení
- dá se použít pro všechny úrovně měření
- nejvhodnější pro nominální úroveň (nemá příliš mnoho hodnot)
- nevhodná, když máme mnoho hodnot - nepřehlednost

		známka z matematiky					celkem
		1	2	3	4	5	
známka z čj	1	82	40	8	1	0	131
	2	71	200	73	17	0	361
	3	4	75	109	25	0	213
	4	1	7	23	24	1	56
	5	0	0	2	1	2	5
celkem		158	322	215	68	3	766

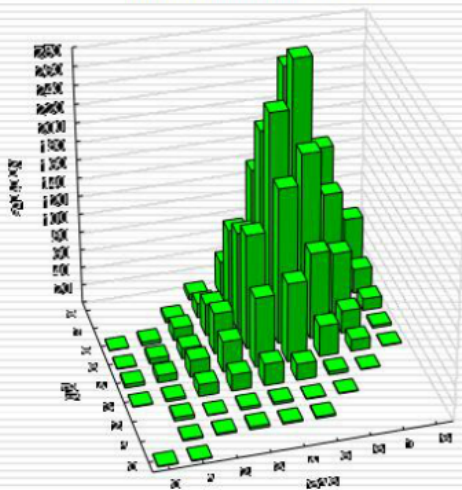
Kontingenční tabulka

		známka z matematiky					celkem
		1	2	3	4	5	
známka z čj	1	82	40	8	1	0	131
	2	71	200	73	17	0	361
	3	4	75	109	25	0	213
	4	1	7	23	24	1	56
	5	0	0	2	1	2	5
celkem		158	322	215	68	3	766

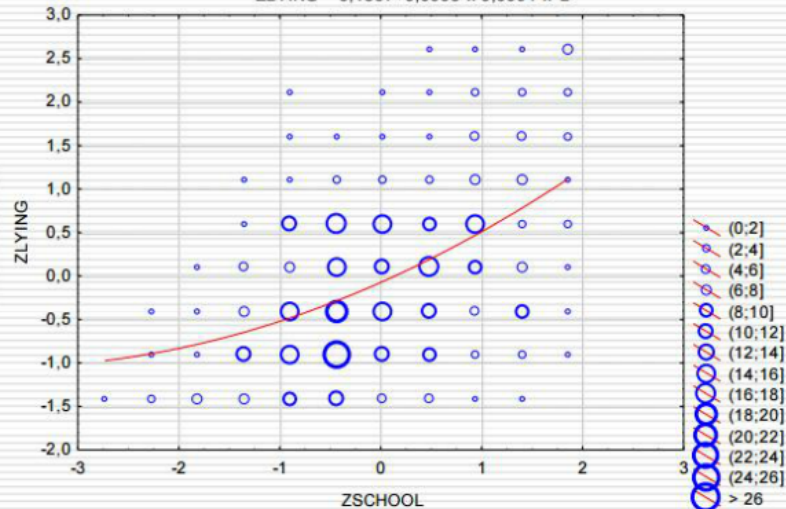
- v hlavní diagonále kontingenční tabulky více nakumulované hodnoty než jinde - lineární trend
- Hodnoty je třeba přehledně uspořádat (stejně jako u tabulky četností)
- Pro data všech úrovní měření, nejvhodnější pro diskrétní prom. s málo hodnotami
- Buňky mohou obsahovat absolutní četnosti, rel. četnosti (řádkové, sloupcové, celkové)
- Poslední sloupec/řádek obsahuje tzv. sloupcové/řádkové marginální (relativní) četnosti
- Je grafickou podobou je trojrozměrného sloupcový diagramu či histogramu (může tedy obsahovat i intervaly)
- Relativně vysoké četnosti v jedné z diagonál naznačují lineární provázanost proměnných

Grafická zobrazení dvourozměrného rozdělení

Bivariate Histogram of B15 against B16
b_test_akt.sta 149v*3080c
Include condition: v133 = 1

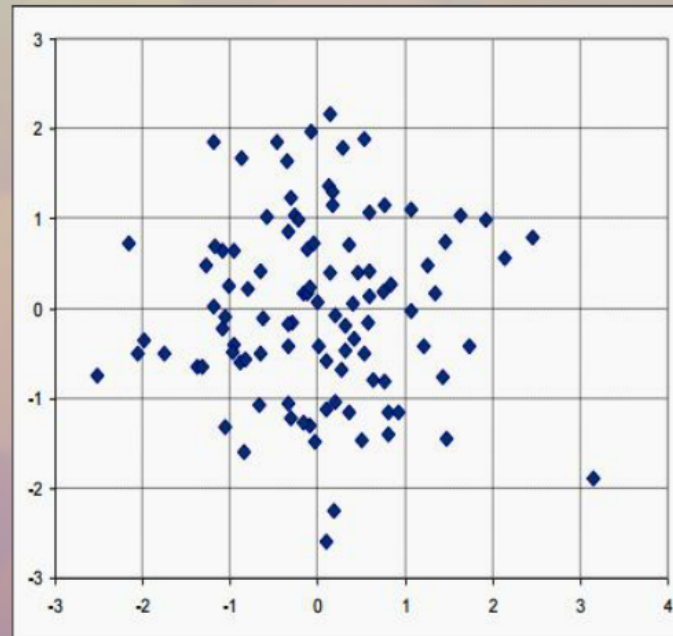


Scatterplot of ZLYING against ZSCHOOL
rudý říjen.sta 41v*481c
ZLYING = 0,1397+0,0903*x-0,0094*x^2



Bodový graf - scatterplot

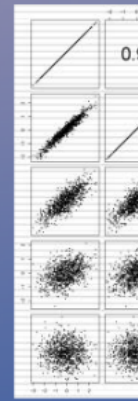
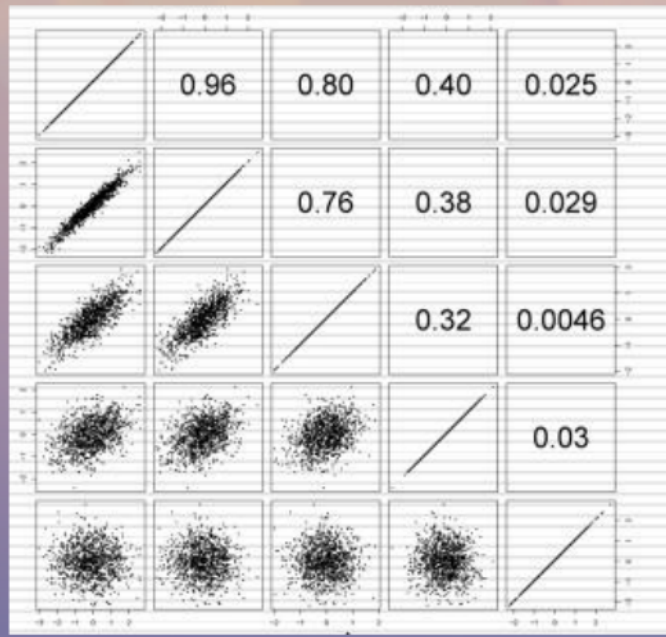
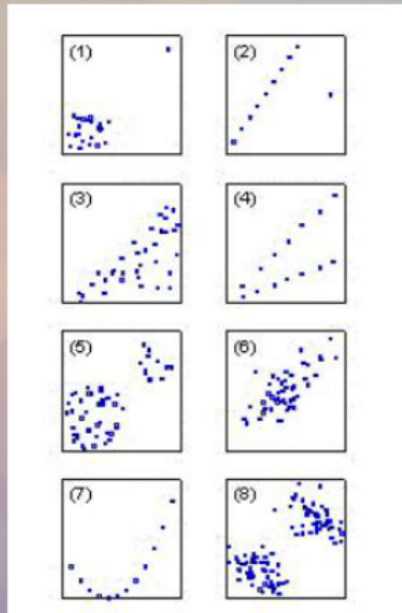
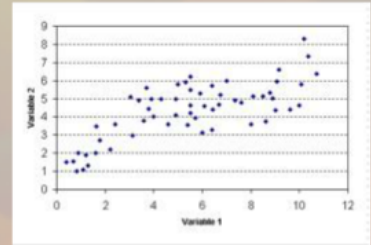
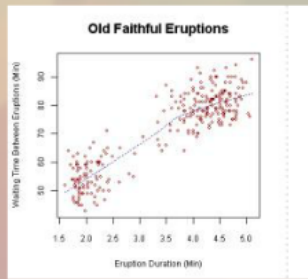
- Používá se na poměrové úrovni, zobrazuje přesné polohy odpovědi každého respondenta
- těsně související proměnné obvykle uspořádány do elipsy (čím užší a protáhlejší, tím těsnější vztah)
- Nahrazuje kontingenční tabulku, jsou-li obě proměnné spojité
- Pro proměnné s málo body měření nemá smysl
 - Každá 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

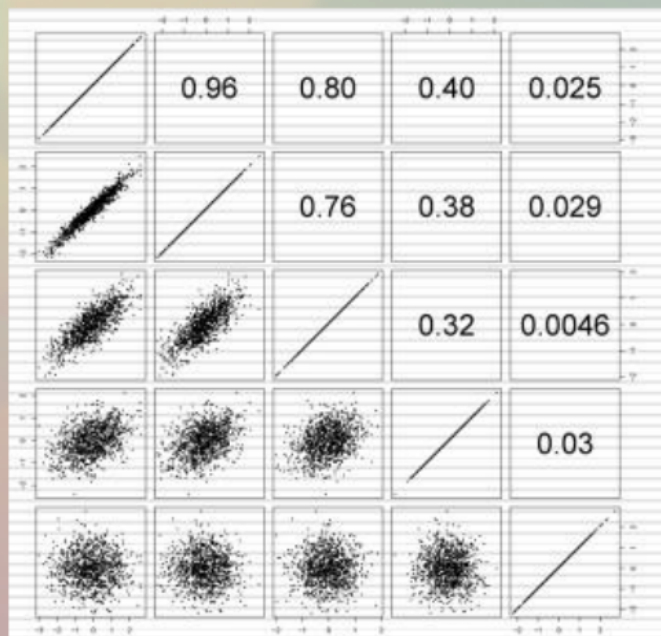




Korelac
vztah dv
pouze j

Různé podoby vztahu mezi dvěma proměnnými





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

- Lineární vztah, korelace.
- Je to monotónní, čím více X, tím více Y.
- Projevuje se takto „ideální“ přímkou.

$$y = ax + b$$

- Tato funkce/přímka popisuje strmost vztahu.
- Korelace popisuje těsnost vztahu.

Těsnost

- Čím těsnější (=intenzivnější) vztah 2 proměnných, tím jsou body v scatterplotu nahuštěny okolo přímkou.
- Těsnost nesouvisí s sklonem té přímkou, pouze s tím, jak nahuštěný je scatterplot podél přímkou.

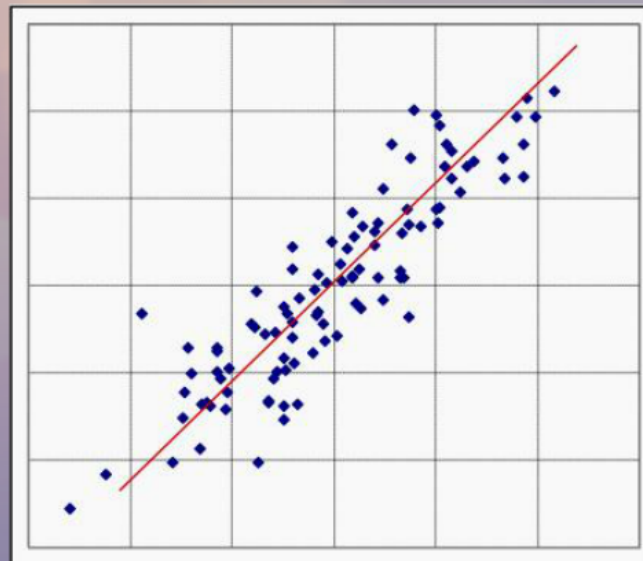
- Těsnost se udává od -1 do 1, kde 0=žádný vztah (data na diagonále).
- Znaménko udává směr vztahu (+) nebo o vzájemnosti (-).
- Rozsah je te...

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

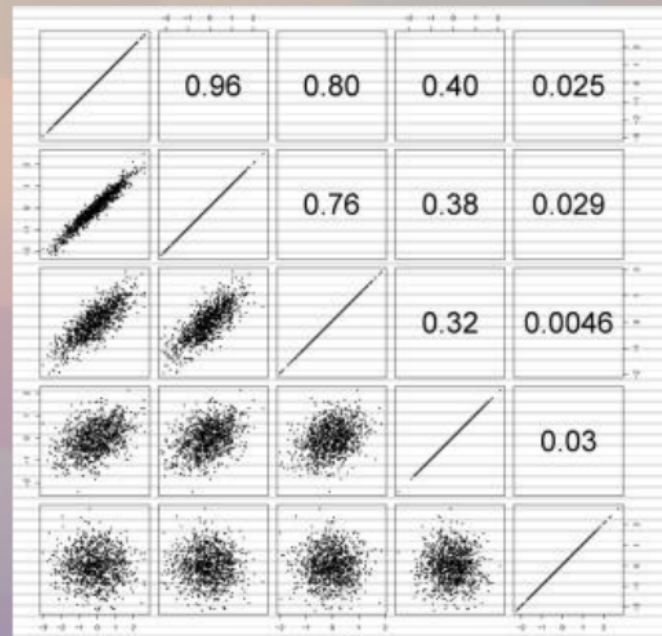
$$y = ax + b$$

- 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

Korelace

Standardizovaný sdílený rozptyl

$$r_{xy} = \frac{\sum z_x z_y}{n-1}$$

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

Pearsonův součinný, momentový koeficient korelace

- nutná intervalová a vyšší úroveň měření
- velký vliv odlehlých hodnot na výsledek
- je vhodný pro popis normálně rozložených proměnných (alespoň unimodální)
- vyjadřuje pouze sílu(těsnost) lineárního vztahu

Nabývá hodnot v rozmezí -1 až 1

0 = žádný vztah

1(-1) = dokonalý kladný (záporný) vztah



Korelace nepopisuje funkční vztah dvou proměnných, ale pouze jeho směr a těsnost.

1. Který z následujících korelačních koeficientů ukazuje na nejtěsnější (nejsilnější) vztah?

a) 0,55 b) 0,09 c) -0,77 d) 0,1 e) 1,05

2. Pěti reprezentativním vzorkům lidí ve věku 15, 20, 30, 45 a 60 let jsme dali dotazník na měření politické konzervativnosti. Těmto 5 vzorkům v uvedeném pořadí vyšly následující průměrné hodnoty konzervativnosti: 60, 85, 80, 70, 65. Korelace mezi věkem a politickou konzervativností je

a) 1.0 b) -1.0 c) lineární d) nelineární

3. Korelace mezi X a Y je 0,60; korelace mezi X a W je -0,80. Má X těsnější lineární vztah s Y nebo s W?

Zobrazení dvourozměrných dat, korelační koeficient

Dvourozměrná analýza

Objeví vztah mezi dvěma proměnnými
 -> Jaké má jedna proměnná vliv na druhou?

1. Jaké jsou závislosti (záporná, kladná, žádná)?
 -> Jaké jsou závislosti (záporná, kladná, žádná)?

2. Jakou proměnnou považujeme za závislou?
 -> Jakou proměnnou považujeme za závislou?

3. Jakou proměnnou považujeme za nezávislou?
 -> Jakou proměnnou považujeme za nezávislou?



Základní statistika

1. Základní statistika
 -> Základní statistika

2. Základní statistika
 -> Základní statistika

3. Základní statistika
 -> Základní statistika

Kontingenční tabulka

1. Kontingenční tabulka
 -> Kontingenční tabulka

2. Kontingenční tabulka
 -> Kontingenční tabulka

3. Kontingenční tabulka
 -> Kontingenční tabulka



Bodový graf - scatterplot

1. Bodový graf - scatterplot
 -> Bodový graf - scatterplot

2. Bodový graf - scatterplot
 -> Bodový graf - scatterplot

3. Bodový graf - scatterplot
 -> Bodový graf - scatterplot

Různé podoby vztahu mezi dvěma proměnnými

1. Různé podoby vztahu mezi dvěma proměnnými
 -> Různé podoby vztahu mezi dvěma proměnnými

2. Různé podoby vztahu mezi dvěma proměnnými
 -> Různé podoby vztahu mezi dvěma proměnnými

3. Různé podoby vztahu mezi dvěma proměnnými
 -> Různé podoby vztahu mezi dvěma proměnnými

Korelace

Standardizovaný sdílený rozptyl

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

Pearsonův součinný, momentový koeficient korelace

- 1. Pearsonův součinný, momentový koeficient korelace
- 2. Pearsonův součinný, momentový koeficient korelace
- 3. Pearsonův součinný, momentový koeficient korelace

Nabývá hodnot v mezích -1 až 1
 0 = žádný vztah
 1(-1) = dokonalý (kladný/záporný) vztah

! Korelace nepropojuje funkční vztah dvou proměnných, ale pouze jeho směr a sílu.

1. Korelace mezi X a Y je 0,85 (kladná)
 2. Korelace mezi X a Y je 0,85 (kladná)
 3. Korelace mezi X a Y je 0,85 (kladná)

Průběh vztahu

1. Průběh vztahu
 -> Průběh vztahu

2. Průběh vztahu
 -> Průběh vztahu

3. Průběh vztahu
 -> Průběh vztahu