



Bivariate Correlations – korelace

Korelace

Procedura *Bivariate Correlations* umožňuje vypočítat Pearsonův lineární korelační koeficient a dva typy neparametrických koeficientů korelace – Kendallovo tau-b a Spearmanovo ρ . K dispozici je rovněž test nulovosti korelačních koeficientů. Kromě toho lze zobrazit kovarianční matici vstupních proměnných a některé další statistiky.

Korelace charakterizuje vzájemný vztah dvou číselných nebo ordinálních proměnných. Tento vztah vyjadřujeme korelačním koeficientem. Sledujeme-li vztahy většího počtu proměnných současně, tabelujeme korelační koeficienty do tzv. korelační matice – čtvercového schématu zobrazujícího hodnoty koeficientů pro všechny dvojice vstupních proměnných. Vzhledem k tomu, že korelační koeficient je symetrickou mírou (nezáleží na pořadí proměnných), je rovněž korelační matice symetrická.

Pearsonův lineární korelační koeficient vyjadřuje míru lineární závislosti dvou číselných proměnných. Před jeho výpočtem je třeba ověřit, zda data neobsahují odlehlá pozorování, která by mohla získané závěry zkreslit. Tento typ koeficientu není vhodný tam, kde mezi proměnnými existuje jiný typ závislosti než lineární.

Pearsonův lineární korelační koeficient nabývá hodnot z intervalu $\langle -1, 1 \rangle$. Je-li jeho absolutní hodnota rovna jedné, data leží přesně na přímce. Korelační koeficient roven jedné charakterizuje přímou úměrnost (přímka je rostoucí), korelační koeficient roven mínus jedné odpovídá nepřímé úměrnosti (přímka je klesající). Při zkoumání reálných dat se však s těmito hraničními hodnotami korelačního koeficientu téměř nesetkáváme (data neleží přesně na přímce), ale zajímá nás, do jaké míry se přímce přibližují. Čím blíže jedné je absolutní hodnota koeficientu, tím lépe přímka data vystihuje a tím silnější lineární závislost mezi proměnnými existuje. Jestliže neexistuje lineární vztah mezi zkoumanými proměnnými, je jejich korelační koeficient roven nule.

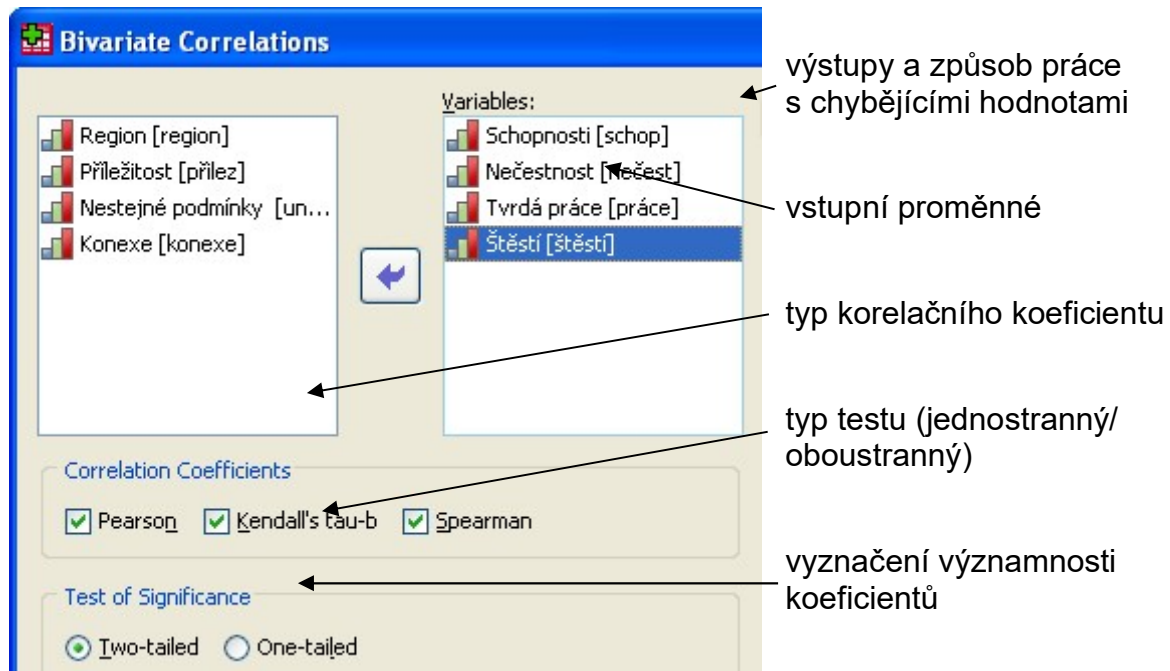
V některých případech je vhodnější užít místo Pearsonova lineárního korelačního koeficientu neparametrické korelační koeficienty – například tam, kde je rozložení některé z proměnných výrazně šikmé, data obsahují vzdálená pozorování nebo máme k dispozici pouze pořadí hodnot. SPSS nabízí dva typy neparametrických korelačních koeficientů: **Spearmanovo ρ** a **Kendallovo tau-b**. Tyto koeficienty neměří lineární závislost, ale vyjadřují, do jaké míry si odpovídají pořadí hodnot dvou proměnných. Z tohoto důvodu jsou také méně citlivé na odlehlá pozorování. Oba koeficienty nabývají hodnot z intervalu $\langle -1, 1 \rangle$. Čím je absolutní hodnota koeficientu větší, tím je popisovaný vztah mezi proměnnými silnější.

Velmi často provádíme také test nulovosti korelačního koeficientu. Nulová hypotéza testu je formulována tak, že korelační koeficient je na celém základním souboru roven nule. Na základě získané significance potom rozhodujeme o zamítnutí/nezamítnutí této hypotézy.

Volání procedury v SPSS

Analyze → Correlate → Bivariate

Nastavení dialogu



- Do pole *Variables* přeneseme nejméně dvě proměnné, jejichž vzájemné vztahy sledujeme. Zadáme-li proměnných více, získáme korelační matici s hodnotami koeficientů pro všechny dvojice proměnných.
- V části *Correlation Coefficients* označíme požadované typy korelačního koeficientu (Pearsonův lineární korelační koeficient, Kendallovo tau-b, Spearmanovo ρ).
- Procedura rovněž automaticky provádí test nulovosti korelačních koeficientů. V části *Test of Significance* volíme mezi oboustrannou a jednostrannou alternativní hypotézou.
- Zaškrtnuté políčko *Flag significant correlations* určuje, zda mají být v korelační matici hvězdičkami vyznačeny koeficienty, které jsou na základě předchozího testu významně různé od nuly.

Tlačítko *Options*



Tlačítkem *Options* volíme další požadované výstupy a způsob práce s chybějícími hodnotami.

Statistics (Statistiky)

- *Means and standard deviations* – průměry a směrodatné odchylky vstupních proměnných.
- *Cross-product deviations and covariances* – kovarianční matice vstupních proměnných a součiny odchylek od průměrů (tj. hodnoty součtů, které tvoří číselník vzorce pro výpočet Pearsonova korelačního koeficientu).

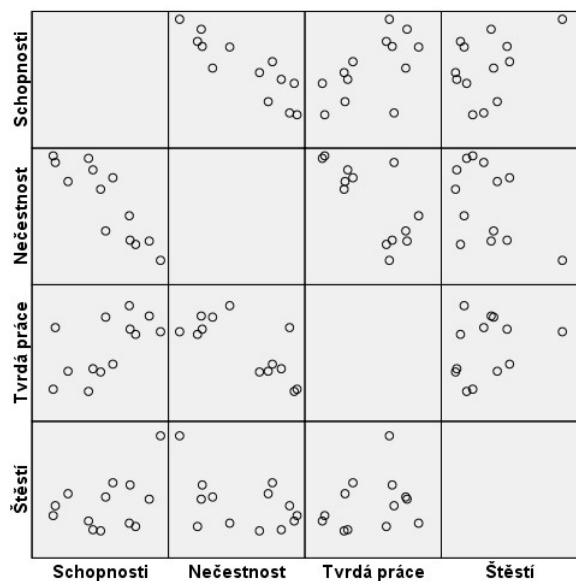
Missing Values (Chybějící hodnoty)

- *Exclude cases pairwise* – případy s chybějícími hodnotami u některé ze vstupních proměnných jsou vynechány z výpočtu korelačních koeficientů pouze tam, kde je to nezbytně nutné.
- *Exclude cases listwise* – případy s chybějícími hodnotami u některé ze vstupních proměnných jsou vyloučeny z výpočtu všech korelačních koeficientů.

Výstupy

Datový soubor obsahuje vybrané informace z mezinárodního výzkumného projektu *Sociální spravedlnost*, realizovaného v roce 1991. Ze třinácti zemí, které se výzkumu účastnily, máme k dispozici souhrnné údaje – průměrná hodnocení důležitosti vybraných položek pro získání bohatství. Naším cílem je podrobněji prozkoumat vzájemný vztah těchto proměnných.

Grafické znázornění vztahů



Před výpočtem korelačního koeficientu je užitečné nejprve zobrazit vztahy mezi proměnnými graficky. V případě, že je vstupních proměnných více můžeme využít maticový bodový graf (*Graphs, Legacy Dialogs, Scatter/Dot*), který se skládá z bodových grafů pro všechny dvojice zadaných proměnných.

Z grafu ověříme, zda se v datech nevyskytují odlehlá pozorování nebo jiný typ problému, který by mohl další výsledky zkreslit. Zároveň si vytvoříme orientační představu o vztazích mezi proměnnými. Zde je například patrný lineární vztah (protiběžnost) mezi proměnnými *Schopnosti* a *Nečestnost*, zatímco proměnná *Štěstí* se nezdá být s ostatními příliš svázaná.

Vzhledem k tomu, že v datech nemáme žádná výrazně odlehlá pozorování, můžeme užít Pearsonův lineární korelační koeficient.

Následující výstupy byly získány pomocí nabídky *Bivariate Correlations*.

Popisné statistiky

Descriptive Statistics

	Mean	Std. Deviation	N
Schopnosti	3.58	.20	13
Nečestnost	3.59	.48	13
Tvrdá práce	3.28	.38	13
Štěstí	3.11	.24	13

Tabulka *Descriptive Statistics* poskytuje přehled o průměrech, směrodatných odchylkách a počtech platných pozorování všech vstupních proměnných.

Zde se například jenom nepatrně liší průměrné hodnoty proměnných *Schopnosti* a *Nečestnost*, zatímco mezi jejich směrodatnými odchylkami je výraznější rozdíl – názory na vliv nečestného chování pro získání bohatství se v jednotlivých zemích více diferencují než názory na vliv schopností, ačkoliv se jejich hodnocení v průměru pohybuje okolo stejné hodnoty.

Korelační matice – Pearsonův lineární korelační koeficient

Correlations

		Schopnosti	Nečestnost	Tvrdá práce	Štěstí
Schopnosti	Pearson Correlation	1	-.879**	.591*	.414
	Sig. (2-tailed)		.000	.033	.160
	N	13	13	13	13
Nečestnost	Pearson Correlation	-.879**	1	-.737**	-.476
	Sig. (2-tailed)	.000		.004	.100
	N	13	13	13	13
Tvrdá práce	Pearson Correlation	.591*	-.737**	1	.268
	Sig. (2-tailed)	.033	.004		.375
	N	13	13	13	13
Štěstí	Pearson Correlation	.414	-.476	.268	1
	Sig. (2-tailed)	.160	.100	.375	
	N	13	13	13	13

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Korelační matice obsahuje pro každou dvojici vstupních proměnných hodnotu Pearsonova lineárního korelačního koeficientu (*Pearson Correlation*) a significance testu nulovosti tohoto koeficientu (*Sig. (2-tailed)*). V případě že byla pro práci s chybějícími hodnotami použita metoda *pairwise*, je v každé buňce tabulky uveden také počet případů, ze kterých byl korelační koeficient spočítán. Při použití metody *listwise* je počet případů zapsán pod tabulkou.

Korelační koeficienty významně odlišné od nuly jsou v tabulce označeny hvězdičkami (jedna hvězdička odpovídá nenulovosti na hladině spolehlivosti 95 %, dvě hvězdičky hladině spolehlivosti 99 %).

Listy procedur IBM SPSS Statistics

Pro větší přehlednost upravíme tabulku pivotací tak, aby v horní vrstvě byly zobrazeny pouze hodnoty korelačních koeficientů. Takto získáme čtvercovou symetrickou matici s jedničkami na diagonále.

Tabulku můžeme ještě zpřehlednit skriptem *Obarvení tabulky*, který je volně k dispozici na stránkách www.acrea.cz. Skript podbarví jednotlivé buňky podle hodnot korelačního koeficientu. Záporným koeficientům odpovídá modrá barva, kladným červená. Barva je tím sytější, čím je absolutní hodnota koeficientu vyšší.

Correlations

Pearson Correlation

	Schopnosti	Nečestnost	Tvrdá práce	Štěstí
Schopnosti	1	-.879**	.591*	.414
Nečestnost	-.879**	1	-.737**	-.476
Tvrdá práce	.591*	-.737**	1	.268
Štěstí	.414	-.476	.268	1

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Z korelační matice vyplývá, že například mezi dvojicí proměnných *Schopnosti* a *Nečestnost* je silná negativní závislost. To znamená, že v zemích, kde mají schopnosti větší vliv na úspěch, má nečestné jednání vliv menší. Naopak proměnné *Schopnosti* a *Tvrdá práce* jsou korelovány kladně – v zemích, kde jsou pro získání bohatství důležité schopnosti je důležitější i tvrdá práce. Korelační koeficienty proměnné *Štěstí* s ostatními proměnnými jsou na 95% hladině spolehlivosti nevýznamné, lze tedy s 5% rizikem tvrdit, že mezi vlivem štěstí a ostatních uvedených faktorů není lineární závislost.

Kovarianční matice a součiny odchylek od průměrů

Correlations

		Schopnosti	Nečestnost	Tvrdá práce	Štěstí
Schopnosti	Pearson Correlation	1	-.879**	.591*	.414
	Sig. (2-tailed)		.000	.033	.160
	Sum of Squares and Cross-products	.461	-.998	.524	.235
	Covariance	.038	-.083	.044	.020
	N	13	13	13	13
Nečestnost	Pearson Correlation	-.879**	1	-.737**	-.476
	Sig. (2-tailed)	.000		.004	.100
	Sum of Squares and Cross-products	-.998	2.794	-1.608	-.665
	Covariance	-.083	.233	-.134	-.055
	N	13	13	13	13
Tvrdá práce	Pearson Correlation	.591*	-.737**	1	.268
	Sig. (2-tailed)	.033	.004		.375
	Sum of Squares and Cross-products	.524	-1.608	1.703	.293
	Covariance	.044	-.134	.142	.024
	N	13	13	13	13
Štěstí	Pearson Correlation	.414	-.476	.268	1
	Sig. (2-tailed)	.160	.100	.375	
	Sum of Squares and Cross-products	.235	-.665	.293	.697
	Covariance	.020	-.055	.024	.058
	N	13	13	13	13

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Listy procedur IBM SPSS Statistics

Tabulku můžeme dále doplnit o kovarianční matici (*Covariance*) a součiny odchylek od průměrů (*Sum of Squares and Cross-products*), tj. hodnoty součtů z čitatele vzorce pro výpočet Pearsonova korelačního koeficientu.

Neparametrické korelace

Correlations			Schopnosti	Nečestnost	Tvrdá práce	Štěstí
Kendall's tau_b	Schopnosti	Correlation Coefficient	1,000	-,795**	,410	,179
		Sig. (2-tailed)	.	,000	,051	,393
		N	13	13	13	13
	Nečestnost	Correlation Coefficient	-,795**	1,000	-,359	-,179
		Sig. (2-tailed)	,000	.	,088	,393
		N	13	13	13	13
	Tvrdá práce	Correlation Coefficient	,410	-,359	1,000	,154
		Sig. (2-tailed)	,051	,088	.	,464
		N	13	13	13	13
	Štěstí	Correlation Coefficient	,179	-,179	,154	1,000
		Sig. (2-tailed)	,393	,393	,464	.
		N	13	13	13	13
Spearman's rho	Schopnosti	Correlation Coefficient	1,000	-,912**	,555*	,286
		Sig. (2-tailed)	.	,000	,049	,344
		N	13	13	13	13
	Nečestnost	Correlation Coefficient	-,912**	1,000	-,577*	-,280
		Sig. (2-tailed)	,000	.	,039	,354
		N	13	13	13	13
	Tvrdá práce	Correlation Coefficient	,555*	-,577*	1,000	,253
		Sig. (2-tailed)	,049	,039	.	,405
		N	13	13	13	13
	Štěstí	Correlation Coefficient	,286	-,280	,253	1,000
		Sig. (2-tailed)	,344	,354	,405	.
		N	13	13	13	13

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Matice neparametrických korelačních koeficientů mají obdobnou strukturu jako matice Pearsonova lineárního korelačního koeficientu, oba dva typy koeficientů jsou však uvedeny v jedné tabulce.

Hodnoty různých typů korelačních koeficientů není vhodné navzájem srovnávat, můžeme však porovnat jejich významnost. Zde vychází Kendallův koeficient pro proměnné *Tvrdá práce* a *Schopnosti* nebo *Nečestnost* na rozdíl od Pearsonova koeficientu nevýznamný. Spearmanův koeficient pro proměnné *Nečestnost* a *Tvrdá práce* je významný na 5% hladině významnosti, ale Pearsonův koeficient dokonce na 1% hladině. Tyto výsledky jsou v souladu s obecněji platným tvrzením, že při splnění předpokladů je vhodnější užít parametrickou variantu koeficientu (Pearsonův korelační koeficient), neboť s testováním nulovosti u neparametrických koeficientů je spojena větší pravděpodobnost chyby druhého druhu (nulovou hypotézu zamítáme méně často).