proLékaře.cz | 7.6.2026

Analýza dat v neurologii

LXXIV. Neparametrický Spearmanův koeficient korelace

L. Dušek, T. Pavlík,
J. Jarkovský, J. Koptíková

Institut biostatistiky a analýz,
LF MU, Brno

✉
prof. RNDr. Ladislav Dušek, Ph.D.
Institut biostatistiky a analýz,
LF MU, Brno
e-mail: dusek@iba.muni.cz

V minulých dílech seriálu jsme se věnovali výkladu Pearsonova korelačního koeficientu, který je také označován jako parametrická nebo lineární korelace. Jeho hodnocení je totiž smysluplné pouze při splnění předpokladu normálního rozložení hodnot u obou do korelace vstupujících proměnných *X* a *Y*. Na příkladech v předchozím díle jsme doložili, že asymetrie rozložení nebo výskyt odlehlých hodnot zásadně zkreslují odhad tohoto korelačního koeficientu a také výsledek jeho statistického hodnocení. Jak tedy postupovat v situacích, kdy rozložení hodnot korelovaných proměnných není normální? V takovém případě můžeme buď proměnné transformovat nějakou normalizující funkcí anebo použijeme tzv. neparametrickou korelaci, která nevyžaduje normalitu rozložení hodnot. Nejčastěji používanou neparametrickou mírou korelace je Spearmanův korelační koeficient (r_s), jehož výkladu budeme věnovat tento díl seriálu.

Připomeňme, že neparametrické statistiky jsou tzv. robustní, tedy více či méně necitlivé vůči odchylkám od normality analyzovaných proměnných. Neparametrické postupy typicky převádějí původní kvantitativní hodnoty proměnných na pořadí („rank") a tím se od vlivu odlehlých hodnot oprošťují. Z tohoto postupu vychází i vztah pro výpočet Spearmanova korelačního koeficientu, který přibližuje příklad 1. Původní hodnoty proměnných *X* a *Y* jsou nejprve převedeny na

Výpočet Spearmanova korelačního koeficientu (r_s) je založený na porovnání pořadí hodnot analyzovaných proměnných. Hodnoty proměnných *X* a *Y* jsou seřazeny dle velikosti (každá proměnná samostatně) a je jim přiděleno pořadí. Rozdíly v pořadí odpovídajících si hodnot následně vstupují do výpočtu korelace.

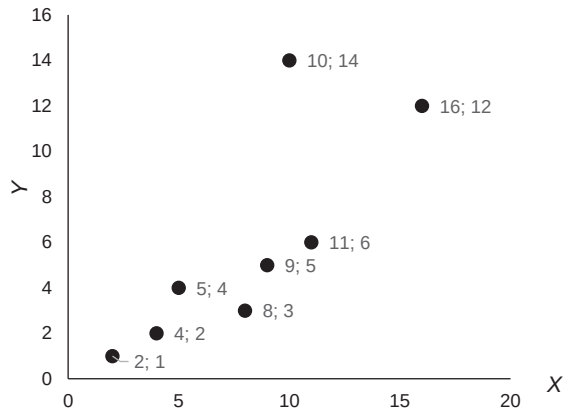
Příkladová data pro výpočet (N = 8)

X	*Y*	pořadí *X*	pořadí *Y*	d_i
2	1	1	1	0
4	2	2	2	0
5	4	3	4	–1
10	14	6	8	–2
11	6	7	6	1
9	5	5	5	0
8	3	4	3	1
16	12	8	7	1

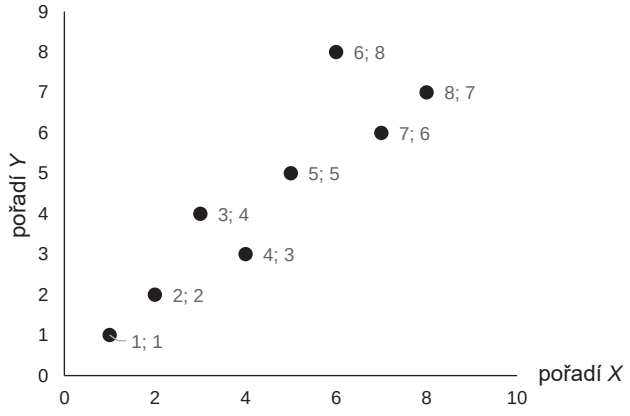
Vztah pro výpočet Spearmanova korelačního koeficientu, kde d_i je rozdíl v pořadí hodnot analyzovaných proměnných.

$$r_s = 1 - \frac{6\sum_{i=1}^{N} d_i}{N(N^2-1)} = 1 - \frac{6\left(0^2+0^2+(-1)^2+(-2)^2+1^2+0^2+1^2+1^2\right)}{8(8^2-1)} = 0{,}9048$$

Graf s původními hodnotami *X* a *Y*



Graf s pořadími hodnot *X* a *Y*



Příklad 1. Výpočet Spearmanova korelačního koeficientu.

proLékaře.cz | 7.6.2026

Hodnota Spearmanova korelačního koeficientu může být testována na statistickou významnost, stejně jako tomu je u parametrické Pearsonovy korelace. Typická nulová hypotéza je $r_s = 0$, alternativní pak $r_s \neq 0$. Testová statistika je počítána obdobně jako v případě Pearsonova korelačního koeficientu a má Studentovo rozdělení (t) s $n - 2$ stupni volnosti. Další, často využívanou možností, je permutační přístup k testování statistické významnosti r_s. Příklady níže ukazují výsledky testu při různých hodnotách Spearmanova korelačního koeficientu.

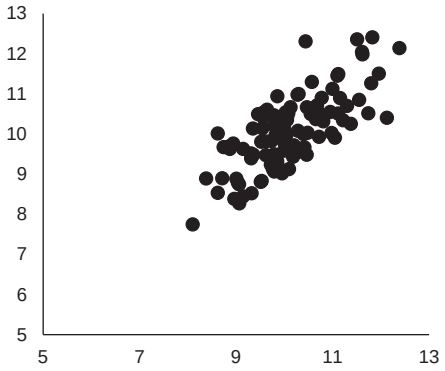
Vztah pro výpočet Spearmanova korelačního koeficientu

$$r_s = 1 - \frac{6\sum_{i=1}^{N} d_i}{N(N^2 - 1)}$$

Test statistické významnosti Spearmanova korelačního koeficientu

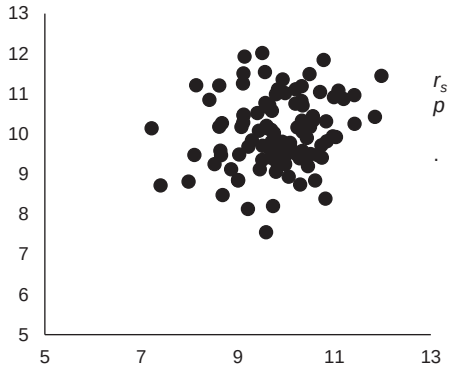
$$t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}}$$

2a. Statisticky významná kladná korelace



r_s = 0,707
p < 0,001

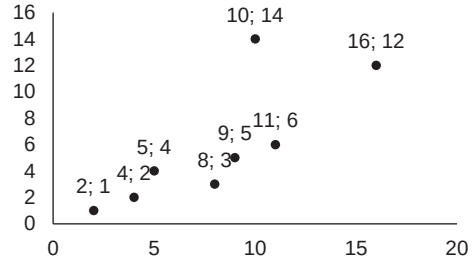
2b. Statisticky nevýznamná korelace



r_s = 0,148
p = 0,140

Příklad 2. Testování statistické významnosti Spearmanova korelačního koeficientu.

Příkladová data



Výpočet intervalu spolehlivosti Spearmanova korelačního koeficientu využívá stejně jako v případě Pearsonova korelačního koeficientu Fisherovy transformace; další možností je využití permutačního přístupu k odhadu intervalu spolehlivosti.

$$z = 0{,}5 \times ln\left(\frac{1+r}{1-r}\right)$$

$$směr.odch. = \sqrt{1/(n-3)}$$

$$(d^*, h^*) = z \pm 1{,}96 \times směr.odch.$$

$$r_s = 1 - \frac{6\sum_{i=1}^{N} d_i}{N(N^2-1)} = 0{,}9048$$

$$z = 0{,}5 \times ln\left(\frac{1+r}{1-r}\right) = 0{,}5 \times ln\left(\frac{1+0{,}9048}{1-0{,}9048}\right) = 1{,}498$$ Výpočet Fisherovy transformace

$$směr.odch. = \sqrt{1/(n-3)} = \sqrt{1/(8-3)} = 0{,}447$$

$$(d^*, h^*) = z \pm 1{,}96 \times směr.odch. = 1{,}498 \pm 1{,}96 \times 0{,}447 = (0{,}622;\ 2{,}375)$$

$$(d, h) = \frac{\exp(2 \times d^*) - 1}{\exp(2 \times d^*) + 1}; \frac{\exp(2 \times h^*) - 1}{\exp(2 \times h^*) + 1} =$$
$$= \frac{\exp(2 \times (0{,}622)) - 1}{\exp(2 \times (0{,}622)) + 1}; \frac{\exp(2 \times 2{,}375) - 1}{\exp(2 \times 2{,}375) + 1} =$$

Zpětná Fisherova transformace

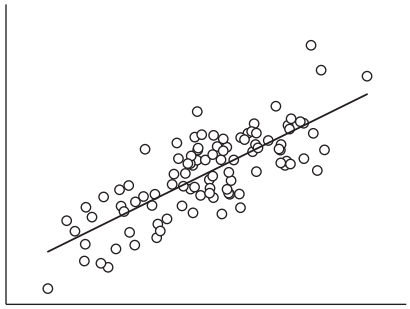
$$= \mathbf{(0{,}552;\ 0{,}983)}$$ 95% interval spolehlivosti Spearmanova korelačního koeficientu.

Příklad 3. Výpočet 95% intervalu spolehlivosti Spearmanova korelačního koeficientu (data z příkladu 1).

proLékaře.cz | 7.6.2026

4a.
Normálně rozložená data s kladným lineárním vztahem – velký vzorek dat

V této situaci poskytuje výpočet Pearsonova a Spearmanova korelačního koeficientu stejný výsledek, a to jak v numerické hodnotě, tak v statistickém testu.

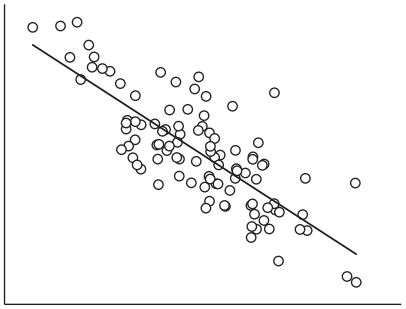


$N = 100$
$r = 0{,}777$
$p < 0{,}001$

$r_s = 0{,}733$
$p < 0{,}001$

4b.
Normálně rozložená data se záporným lineárním vztahem – velký vzorek dat

V této situaci poskytuje výpočet Pearsonova a Spearmanova korelačního koeficientu stejný výsledek, a to jak v numerické hodnotě, tak v statistickém testu.

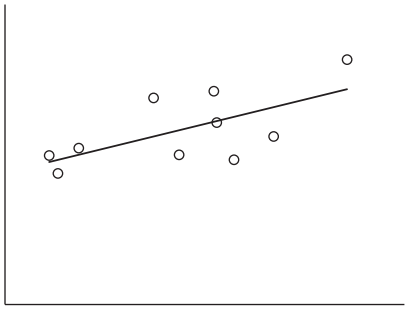


$N = 100$
$r = -0{,}788$
$p < 0{,}001$

$r_s = -0{,}753$
$p < 0{,}001$

4c.
Normálně rozložená data s lineárním vztahem – malý vzorek dat

Data vedou k obdobné numerické hodnotě Pearsonova a Spearmanova korelačního koeficientu, avšak liší se výsledek statistického testování v důsledku malé velikosti vzorku. Vzhledem k tzv. nižší síle neparametrického testu je Spearmanův korelační koeficient statisticky nevýznamný.

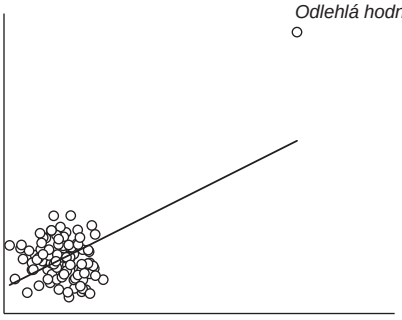


$N = 10$
$r = 0{,}652$
$p = 0{,}041$

$r_s = 0{,}515$
$p = 0{,}133$

4d.
Data nesplňující předpoklad normálního rozložení – velký vzorek dat

V této situaci se hodnoty Pearsonova a Spearmanova korelačního koeficientu výrazně liší. Velikost a statistická významnost Pearsonova korelačního koeficientu jsou zkresleny přítomností odlehlé hodnoty. Výpočet Spearmanova korelačního koeficientu tímto ovlivněn není a jeho výsledek je správný.



$N = 100$
$r = 0{,}533$
$p < 0{,}001$

$r_s = -0{,}135$
$p = 0{,}179$

Příklad 4. Srovnání hodnot Pearsonova a Spearmanova korelačního koeficientu.

pořadí (samostatně každá proměnná zvlášť) a následně je kalkulována hodnota korelace, která pracuje s diferencemi pořadí X a Y u jednotlivých objektů, kterých je N. Diference pořadí u i-tého řádku vstupní matice hodnot se označuje d_i. Výsledný vztah pro výpočet r_s je následující:

$$r_s = 1 - \frac{6\sum_{i=1}^{N} d_i}{N(N^2 - 1)}$$

Jsou-li pořadí hodnot X a Y naprosto shodná, pak součet všech hodnot d_i je roven nule a korelační koeficient dosahuje své maximální kladné hodnoty, tedy 1. Obdobně, pokud by hodnoty X a Y byly řazeny přesně opačně, pak by výsledná hodnota r_s byla –1 (záporná korelace). Je patrné, že výpočet r_s je velmi jednoduchý a lze ho provést i bez zvláštního softwarového vybavení.

Ze vztahu pro výpočet je zřejmé, proč je Spearmanův koeficient v literatuře často označován jako pořadová korelace („rank correlation"). Z tohoto označení také vyplývá interpretace neparametrické korelace, která je rozdílná od korelace Pearsonovy. Pearsonův korelační koeficient pracuje s původními kvantitativními hodnotami X a Y a na jeho hodnotu mají vliv nejen stejný směr v hodnotách obou proměnných, ale také kvantitativní rozdíly hodnot X a Y od jejich průměru. Jde o korelaci, která dosahuje maxima, pokud je mezi X a Y čistý přímkový vztah. Hodnoty Spearmanova korelačního koeficientu odrážejí pouze stejný směr hodnot X a Y (koeficient je počítán z pořadí, nikoli z původních hodnot), a tedy nijak ne-

proLékaře.cz | 7.6.2026

souvisí s tvarem vztahu obou proměnných. Hodnota r_s se tudíž může blížit maximu, i když mezi hodnotami *X* a *Y* není lineární vztah. Pro maximální neparametrickou korelaci stačí pouze, aby hodnoty obou proměnných rostly nebo klesaly ve stejném pořadí. Neparametrická korelace neodráží kvantitu, tedy „o kolik" se mění hodnota *X* v závislosti na hodnotě *Y*.

Výše zmíněné rozdíly mezi parametrickou a neparametrickou korelací jsou především interpretační. Pokud jde o dosažitelné hodnoty koeficientů, není mezi oběma metodickými postupy žádný rozdíl. Spearmanův korelační koeficient může stejně jako Pearsonova korelace nabývat hodnot od –1 do +1. Hodnoty r_s blízké nebo rovny nule ukazují na situaci, kdy jsou pořadí hodnot *X* a *Y* náhodně zpřeházená a mezi oběma veličinami není žádný vztah.

Rovněž odhad intervalu spolehlivosti pro neparametrickou korelaci a test její statistické významnosti (testujeme nulovou hypotézu $r_s = 0$) jsou prakticky totožné s výpočty pro Pearsonův korelační koeficient. Konkrétní postupy dokládají příklady 2 a 3.

Jistou slabinou výpočtu Spearmanova korelačního koeficientu je práce s pořadími hodnot, neboť transformace původních hodnot proměnných do pořadí zásadně zužuje numerický rozsah hodnot. To se projeví zejména při práci s malými soubory dat, kdy říkáme, že neparametrické testy mají tzv. nižší sílu než testy parametrické. Tím je myšleno, že mají při stejné velikosti vzorku nižší schopnost rozpoznat neplatnost nulové hypotézy. Problémem také může být výskyt stejných hodnot, které pak v rámci proměnných *X* a *Y* dostávají stejná pořadí a ta se musí průměrovat. V takovém případě je v literatuře doporučován jiný vztah pro výpočet r_s:

$$r_s = \frac{\sum_{i=1}^{N} x_{ri} y_{ri} - n\bar{x}_r\bar{y}_r}{(N-1)\, s_{xr} s_{yr}}$$

Tento vztah je v podstatě vztahem pro výpočet Pearsonova korelačního koeficientu, avšak počítaného na pořadích vstupujících hodnot *X* a *Y*. Hodnota x_{ri} značí pořadí hodnoty x_i v rámci vzestupně uspořádaných hodnot *X*. Obdobně jsou takto převedeny hodnoty proměnné *Y*. Označení pomocí indexu *r* značí „rank", tedy pořadí. Hodnoty $\bar{x}_r$ a $\bar{y}_r$ jsou potom průměrnými pořadími v rámci hodnot *X* a Y, hodnota s_{xr} s_{yr} je součinem směrodatných odchylek rovněž počítaných na pořadích hodnot obou proměnných.

Čtenáře nyní jistě napadne otázka, kdy je tedy v praxi lepší použít neparametrickou korelaci místo parametrické. Obecné pravidlo vyplývá již z výše uvedeného výkladu. Spearmanova korelace by měla být jednoznačně preferována u dat, kde vstupující proměnné nesplňují podmínky normálního rozdělení, zejména pokud se v nich vyskytují odlehlé hodnoty. Není-li z nějakého důvodu smyslem korelace prokázat přímkový vztah *X* a *Y*, je neparametrický korelační koeficient dobrou volbou. Při analýze konkrétních dat lze ovšem vždy použít současně obě korelace a srovnat jejich výsledky. Významné rozdíly mezi neparametrickou a parametrickou korelací by pak měly být varováním a signálem, že je třeba věnovat pozornost rozložení hodnot a možným zkreslením. Tyto situace se snaží přiblížit ukázky uvedené na příkladu 4.

Česká neurologická společnost ČLS JEP

Česká neurologická společnost (ČNS) je součástí České lékařské společnosti Jana Evangelisty Purkyně (www.cls.cz).
Členem společnosti může stát lékař, farmaceut, případně jiný pracovník ve zdravotnictví a příbuzném oboru, který souhlasí s posláním a cíli ČLS JEP a zaváže se přispívat k jejich plnění. Každý může být členem více odborných společností.

Jak se stát členem ČNS?

- vyplňte přihlášku na webových stránkách ČNS www.czech-neuro.cz, registrovat se zároveň můžete také do jednotlivých sekcí ČNS
- po odeslání registrace získáte na e-mail potvrzení o úspěšném odeslání Vaší přihlášky
- schvalování žádostí o členství probíhá vždy na nejbližší výborové schůzi ČNS, o přijetí Vás bude informovat sekretariát ČNS (sekretariat@czech-neuro.cz)

Co vám členství v ČNS přinese?

- předplatné časopisu Česká a slovenská neurologie a neurochirurgie
- pravidelný elektronický zpravodaj s novinkami
- zvýhodněné podmínky účastí na pravidelném neurologickém sjezdu a jiných akcích
- možnost zúčastnit se soutěže o nejlepší neurologické publikace

Změny údajů

V případě změny Vašich údajů (jména, adresy, telefonu, e-mailu apod.) ji, prosím, nahlaste členské evidenci sekretariátu ČNS sekretariat@czech-neuro.cz. Změna bude nahlášena automaticky také vydavateli časopisu Česká a slovenská neurologie a neurochirurgie a Centrální evidenci členů ČLS JEP.