Základy korelační analýzy

Doposud jsme se z hlediska biostatistiky zabývali hodnocením spojitých a diskrétních náhodných veličin v jedné nebo více odlišitelných experimentálních skupinách. Tato kapitola představuje úvod do korelační analýzy, s jejíž pomocí hodnotíme vzájemný vztah dvou a více spojitých náhodných veličin. Nejjednoduššími nástroji pro kvantifikaci tohoto typu vztahu jsou tzv. korelační koeficienty, dva nejpoužívanější z nich zde uvádíme: Pearsonův a Spearmanův korelační koeficient.

Předpokládané výstupy z výuky:

1. Student zná reálné příklady hodnocení vztahu dvou spojitých veličin
2. Student umí definovat Pearsonův a Spearmanův korelační koeficient a zná rozdíly v jejich použití
3. Student rozumí principům výpočtu Pearsonova a Spearmanova koeficientu
4. Student je schopen identifikovat situace kdy je výpočet Pearsonova a Spearmanova koeficientu zavádějící
5. Student dokáže sestrojit interval spolehlivosti pro Pearsonův i Spearmanův koeficient
6. Student je schopen testovat hypotézu, že korelační koeficient je roven nule

---

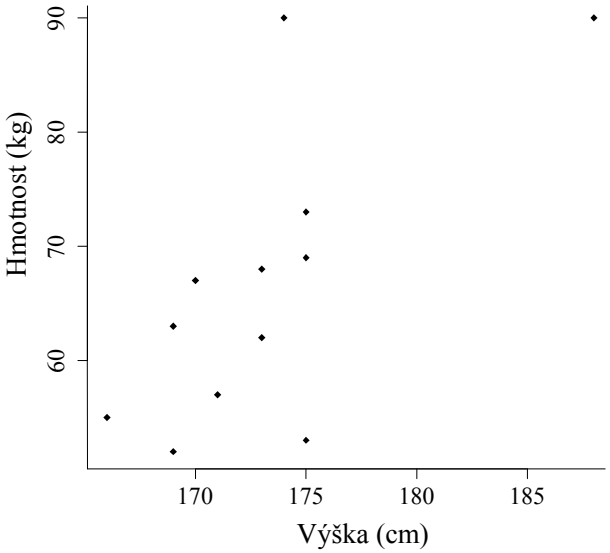
1 Úvod

Problematika hodnocení vztahů a souvislostí mezi dvěma a více spojitými veličinami, která je základem tzv. **korelační** a **regresní analýzy** (*correlation and regression analysis*) je velmi významnou oblastí biostatistiky [1, 2]. Úkoly, které můžeme řešit pomocí tohoto typu metod, jsou následující:

1. Zjistit, zda mezi sledovanými spojitými veličinami existuje potenciální vztah, např. zda vyšší hodnoty jedné náhodné veličiny souvisejí s nižšími hodnotami jiné náhodné veličiny. Můžeme se např. ptát, zda výše systolického krevního tlaku souvisí s konzumací sodíku, nebo zda vyšší hladina krevní glukózy souvisí s vyšší hladinou jiné látky v krevní plazmě.
2. Predikovat hodnoty jedné náhodné veličiny na základě znalosti hodnot jiných náhodných veličin. Naším cílem může být např. predikce hodnot koncentrací nějaké těžko měřitelné látky v prostředí na základě znalosti koncentrací látek příbuzných, které však těžko měřitelné nejsou.
3. Kvantifikovat vztah mezi dvěma spojitými náhodnými veličinami, např. pro použití jedné z nich na místo té druhé jako diagnostického testu. Můžeme si např. klást otázku, jak moc spolu souvisí hladiny dvou krevních bílkovin, když bychom měření jedné z nich chtěli nahradit druhou.

Nejjednodušším způsobem, jak zjistit, zda hodnoty dvou spojitých náhodných veličin spolu nějak souvisí, je vykreslení **bodového grafu** (*scatter plot*), který nám ukazuje, jak hodnoty jedné veličiny rostou nebo klesají v závislosti na druhé veličině. Příklad bodového

grafu je uveden na obrázku 1, kde je zobrazena výška a hmotnost studentů předmětu Biostatistika pro matematickou biologii v jarním semestru 2010. Výsledek je očekávaný, s vyšší výškou má tendenci růst i hmotnost, nicméně vzhledem k tomu, že zobrazené body neleží na přímce, nelze říci, že by mezi výškou a hmotností byl přesně lineární vztah.



Obr. 1 Bodový graf hodnot výšky a hmotnosti studentů matematické biologie.

2 Pearsonův korelační koeficient

Nevýhodou bodového grafu je samozřejmě absence kvantifikace funkčního vztahu sledovaných veličin. Kvantifikace obecného funkčního vztahu je obtížná, pro kvantifikaci lineárního vztahu náhodných veličin byl zaveden tzv. **Pearsonův korelační koeficient** (*Pearson correlation coefficient*). V teoretické podobě ho lze pro náhodné veličiny X a Y s nenulovým rozptylem vyjádřit následovně:

$$R(X,Y) = \frac{E((X-EX)(Y-EY))}{\sqrt{DX}\sqrt{DY}}. \qquad (11.1)$$

Je důležité zdůraznit, že Pearsonův korelační koeficient charakterizuje pouze lineární vztah, jinak řečeno odráží pouze variabilitu kolem lineárního trendu. Pro kvantifikaci nelineárních závislostí je naprosto nevhodný. Základní vlastností Pearsonova korelačního koeficientu je, že nabývá pouze hodnot z intervalu $\langle -1,1 \rangle$ s tím, že hodnota $R(X,Y)$ je kladná, když vyšší hodnoty náhodné veličiny X souvisí s vyššími hodnotami náhodné veličiny Y, a naopak je záporná, když nižší hodnoty X souvisí s vyššími hodnotami Y. Hodnoty 1, respektive -1, získáme pouze v případě, kdy body zobrazené v bodovém grafu leží na přímce s kladnou, respektive zápornou směrnicí.

2.1 Výpočet Pearsonova korelačního koeficientu

Teoretický výpočet $R(X,Y)$ je podmíněn znalostí konkrétního rozdělení pravděpodobnosti náhodného vektoru (X,Y), což se v praxi stává velmi zřídka. Lineární vztah náhodných veličin X a Y tak kvantifikujeme na základě výběrového souboru. Výběrový Pearsonův korelační

koeficient standardně značíme r a při jeho výpočtu vycházíme z realizace dvourozměrného náhodného vektoru o rozsahu n, tedy dvojic pozorovaných hodnot náhodných veličin X a Y pro první až n-tou experimentální jednotku:

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}, \ldots, \begin{pmatrix} x_n \\ y_n \end{pmatrix}. \tag{11.2}$$

Výpočet výběrového Pearsonova korelačního koeficientu je pak následující:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y}, \tag{11.3}$$

kde $\bar{x}$ a $\bar{y}$ jsou výběrové průměry, s_x a s_y jsou výběrové směrodatné odchylky. Na obrázku 2 jsou zobrazeny realizace náhodných veličin X a Y a k nim příslušné výběrové korelační koeficienty pro čtyři různé situace: graf vlevo nahoře odpovídá úplné lineární závislosti; graf vpravo nahoře ukazuje příklad relativně silné záporné korelace; vlevo dole pak vidíme slabě kladně korelované veličiny; vpravo dole jsou nakonec zobrazeny veličiny nekorelované.

Příklad 1. Vypočítejme výběrový Pearsonův korelační koeficient kvantifikující korelaci mezi výškou a hmotností studentů předmětu Biostatistika pro matematickou biologii v jarním semestru 2010. Pozorované hodnoty (realizace náhodného vektoru o rozsahu n = 13) jsou uvedeny v tabulce 1, navíc jsou předmětem obrázku 1.

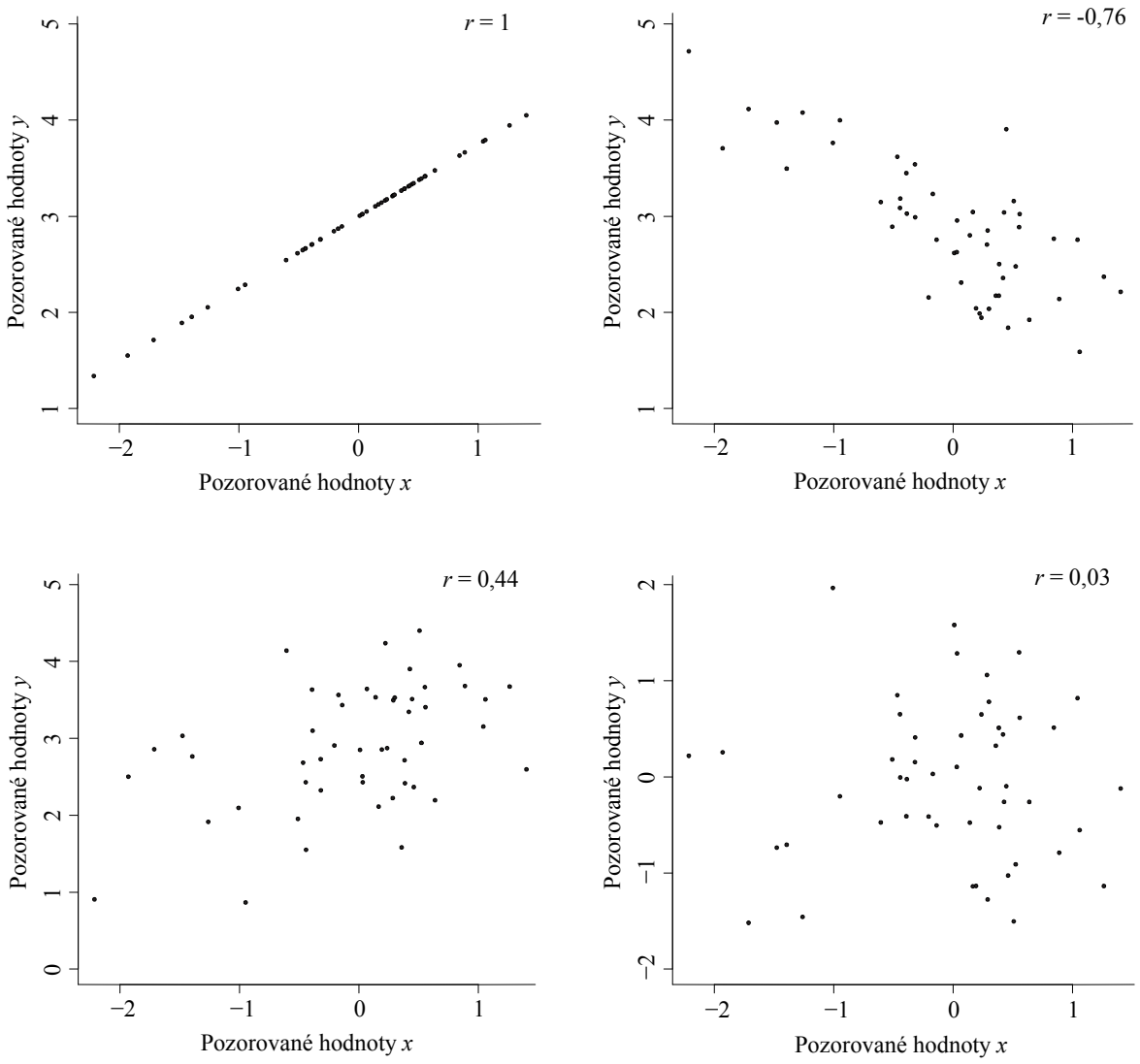
Tabulka 1 *Pozorované hodnoty výšky a hmotnosti 13 studentů.*

175	166	170	169	188	175	176	171	173	175	173	174	169
69	55	67	52	90	53	57	57	68	73	62	90	63

Výpočet výběrových statistik pro jednoduchost vynecháme (laskavý čtenář si je může jednoduše dopočítat na základě dat v tabulce 1), dosazením do vztahu (11.3) získáme následující hodnotu výběrového Pearsonova korelačního koeficientu:

$$r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y} = \frac{148\,929 - 148\,417{,}2}{(13-1) * 5{,}3 * 12{,}5} = 0{,}64\,. \tag{11.4}$$

Hodnota r =0,64 ukazuje na silnou korelaci, kdy s vyšší výškou roste i hmotnost, což odpovídá očekávání, nicméně je třeba si uvědomit malou velikost výběrového souboru a dvě odlehlé hodnoty na obrázku 1 odpovídající hmotnosti 90 kg, které úplně nekorespondují se zbytkem souboru. Obě tyto skutečnosti ovlivňují výslednou hodnotu r.



Obr. 2 Ukázky realizací náhodných veličin X a Y a vypočtené výběrové korelační koeficienty.

2.2 Interval spolehlivosti pro Pearsonův korelační koeficient

Jako každou výběrovou statistiku je i výběrový Pearsonův korelační koeficient r vhodné doplnit $100(1 - \alpha)\%$ intervalem spolehlivosti, který nám dá informaci o variabilitě tohoto odhadu. Na rozdíl od výpočtu bodového odhadu, který lze vypočítat na datech z různých rozdělení, je však v případě, že chceme rozhodovat o vlastnostech Pearsonova korelačního koeficientu (např. konstruovat interval spolehlivosti pro r nebo testovat hypotézy o r), nutné učinit předpoklad o normalitě náhodných veličin X a Y. Jinými slovy, při výpočtu r předpokládáme realizaci dvourozměrného náhodného vektoru z dvourozměrného normálního rozdělení o rozsahu n. Dalším problémem při konstrukci intervalu spolehlivosti pro r je fakt, že výběrové rozdělení výběrového korelačního koeficientu není normální. Abychom byli schopni interval spolehlivosti zkonstruovat, je třeba použít transformaci na náhodnou veličinu W, přičemž transformace je následující:

$$W = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right). \tag{11.5}$$

Lze ukázat, že náhodná veličina W má normální rozdělení s rozptylem přibližně $D(W) = 1/(n-3)$, kde n je velikost výběrového souboru. Vzhledem k normalitě veličiny W má $100(1-\alpha)\%$ interval spolehlivosti pro její střední hodnotu tvar

$$(d^*, h^*) = w \pm z_{1-\alpha/2}\frac{1}{\sqrt{n-3}}, \tag{11.6}$$

kde $z_{1-\alpha/2}$ je příslušný kvantil standardizovaného normálního rozdělení. Výsledný $100(1-\alpha)\%$ interval spolehlivosti pro r pak dostaneme zpětnou transformací ve tvaru

$$(d, h) = \left(\frac{\exp(2d^*)-1}{\exp(2d^*)+1}; \frac{\exp(2h^*)-1}{\exp(2h^*)+1}\right), \tag{11.7}$$

Příklad 2. Navážeme na příklad 1, kde byl vypočítán výběrový korelační koeficient pro vztah výšky a hmotnosti studentů biostatistiky. Nyní pro $r = 0{,}64$ zkonstruujeme 95% interval spolehlivosti. Realizace transformované náhodné veličiny je následující:

$$w = \frac{1}{2}\ln\frac{1+0{,}64}{1-0{,}64} = 0{,}758, \tag{11.8}$$

Interval spolehlivosti pro střední hodnotu náhodné veličiny W s $\alpha = 0{,}05$ má tvar

$$(d^*, h^*) = 0{,}758 \pm 1{,}96/\sqrt{13-3} = (0{,}138; 1{,}377), \tag{11.9}$$

z čehož plyne výsledný 95% interval spolehlivosti pro výběrový korelační koeficient vztahu výšky a hmotnosti studentů biostatistiky

$$(d, h) = \left(\frac{\exp(2d^*)-1}{\exp(2d^*)+1}; \frac{\exp(2h^*)-1}{\exp(2h^*)+1}\right) = (0{,}14; 0{,}88). \tag{11.10}$$

Z výsledku vidíme, že 95% interval spolehlivosti je velmi široký, neboť připouští jak hodnoty odpovídající silné korelaci ($r = 0{,}88$), tak hodnoty odpovídající velmi slabé, nebo spíše žádné korelaci ($r = 0{,}14$). Zde je na vině zejména malý rozsah výběrového souboru, neboť je zřejmé, že na základě $n = 13$ pozorování je velmi obtížné dělat zásadní závěry ohledně vztahu dvou náhodných veličin.

2.3 Test hypotézy o nulové korelaci dvou náhodných veličin

I v případě malého výběrového souboru, jaký byl použit např. v příkladech 1 a 2, je logické klást si otázku, zda je či není korelace dvou sledovaných veličin nulová. Tato situace vede na testování následujících hypotéz:

$$H_0 : r = 0, \qquad H_1 : r \neq 0. \tag{11.11}$$

Pro testování je nezbytný předpoklad realizace dvourozměrného náhodného vektoru o rozsahu n z normálního rozdělení, což znamená, že máme k dispozici náhodný vektor

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}, \ldots, \begin{pmatrix} x_n \\ y_n \end{pmatrix}, \qquad \begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \end{pmatrix} \right). \tag{11.12}$$

Za platnosti nulové hypotézy pak má statistika

$$T = r\sqrt{\frac{n-2}{1-r^2}} \tag{11.13}$$

Studentovo t rozdělení pravděpodobnosti s $n - 2$ stupni volnosti. Pro oboustrannou alternativu zamítáme nulovou hypotézu na hladině významnosti $\alpha = 0{,}05$, když hodnota testové statistiky přesáhne v absolutní hodnotě kvantil $t_{1-\alpha/2}^{(n-2)}$. Je třeba poznamenat, že testovou statistiku T nelze použít pro testování obecné hypotézy $H_0 : r = r_0 \neq 0$, neboť pro r různé od nuly nemá testová statistika Studentovo t rozdělení. Postup pro testování hypotézy $H_0 : r = r_0 \neq 0$ lze najít např. v [3].

Příklad 3. Provedení testu o nulové korelaci dvou náhodných veličin opět demonstrujeme na datech výšky a hmotnosti studentů biostatistiky. Realizace testové statistiky dané vztahem (11.13) je následující

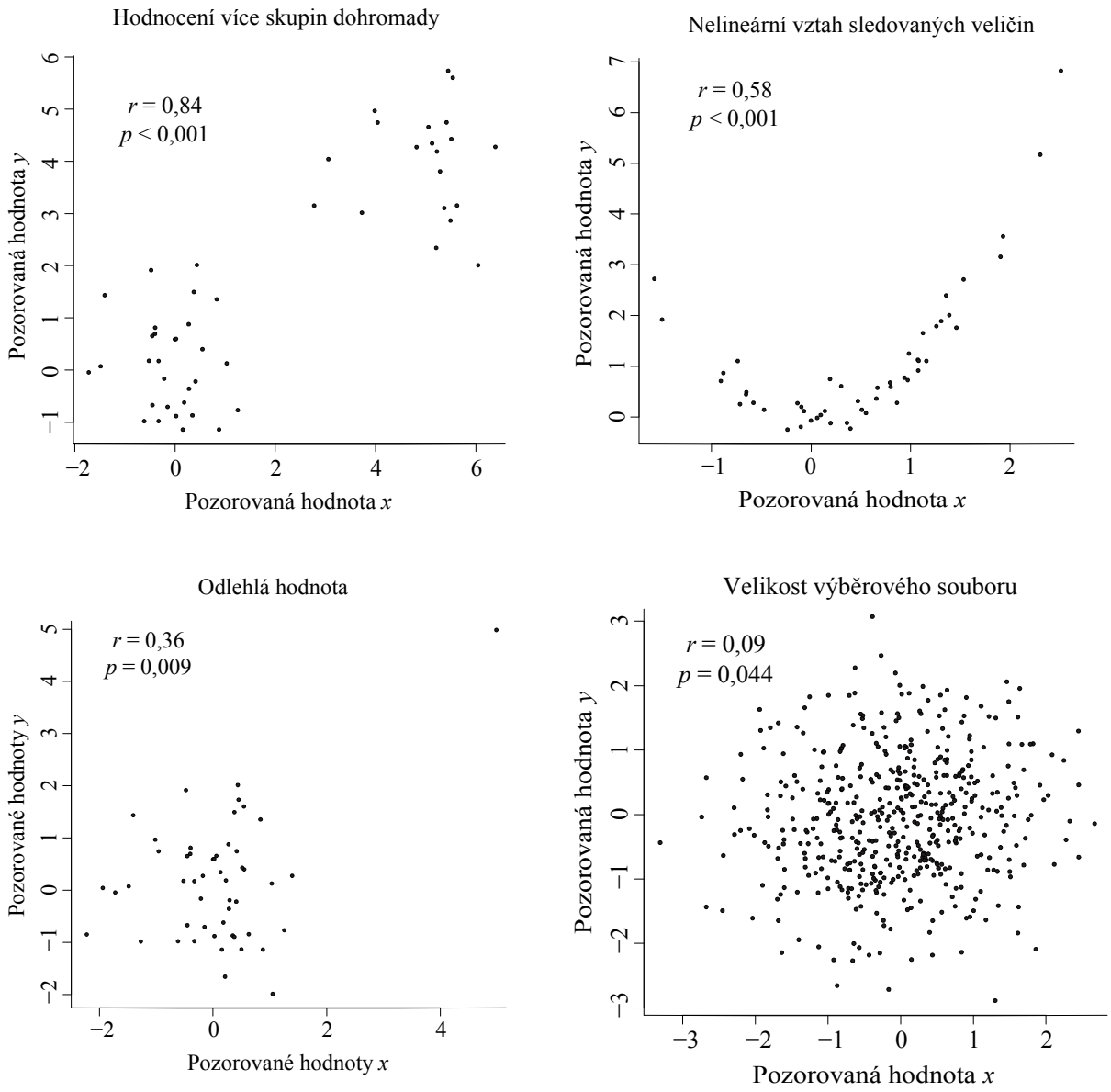
$$t = r\sqrt{\frac{n-2}{1-r^2}} = 0{,}64\sqrt{\frac{13-2}{1-0{,}64^2}} = 2{,}76. \tag{11.14}$$

Srovnáme-li výslednou hodnotu testové statistiky t s kvantilem Studentova t rozdělení příslušným hladině významnosti $\alpha = 0{,}05$, tedy provedeme-li srovnání

$$t = 2{,}76 > 2{,}20 = t_{0{,}975}^{(11)} = t_{1-\alpha/2}^{(n-2)}, \tag{11.15}$$

zamítáme H_0 o tom, že mezi výškou a hmotností studentů biostatistiky je nulová korelace.

Jak bylo uvedeno výše, Pearsonův korelační koeficient kvantifikuje míru lineárního vztahu mezi náhodnými veličinami X a Y. Jeho výpočet je tedy naprosto nevhodný v situacích, kdy se o lineární vztah mezi X a Y nejedná. Obrázek 3 ukazuje čtyři situace, kdy výpočet výběrového Pearsonova korelačního koeficientu nemá smysl, respektive kdy může být jeho výpočet z hlediska interpretace zavádějící. Graf vlevo nahoře znázorňuje situaci, kdy výběrový soubor obsahuje dvě skupiny subjektů s odlišnými hodnotami náhodných veličin X i Y. Ve chvíli, kdy si tohoto nejsme vědomi, výpočet výběrového Pearsonova korelačního koeficientu indikuje silnou korelaci X a Y ($r = 0{,}84$), která je dokonce na daném souboru vysoce statisticky významná ($p < 0{,}001$). Tento výsledek je však statistický artefakt a ve skutečnosti není relevantní. Ideální by v tomto případě bylo soubor rozdělit a kvantifikovat korelaci v obou podsouborech zvlášť (podle obrázku je korelace X a Y v podsouborech naopak velmi malá). Graf vpravo nahoře ukazuje situaci, kdy je mezi veličinami X a Y nelineární vztah. Také zde je výsledný korelační koeficient ($r = 0{,}58$) relativně vysoký, statisticky významný a zároveň neodpovídá skutečnosti. Vlevo dole pak vidíme, jaký vliv má odlehlá hodnota v případě dvou nezávislých (a tedy i nekorelovaných) veličin X a Y. Vzhledem k nezávislosti bychom čekali realizaci r kolem 0, nicméně zde vidíme výsledné r rovno 0,36, opět statisticky významné ($p = 0{,}009$). Konečně, graf vpravo dole ukazuje vliv velikosti výběrového souboru na statistickou významnost korelačního koeficientu. V tomto případě je korelace mezi veličinami X a Y velmi slabá až žádná ($r = 0{,}09$), nicméně velikost výběrového souboru je tak velká ($n = 500$), že statistický test indikuje statisticky významný rozdíl r od hodnoty 0. Toto je klasický příklad rozporu mezi statistickou a praktickou významností výsledku, kdy je nezbytné kromě statistiky do výsledné interpretace zapojit i znalost dané problematiky. Všechny čtyři problematické případy lze velmi dobře odhalit s použitím bodového grafu, který by měl být jedním z prvních kroků při hodnocení vzájemného vztahu dvou spojitých náhodných veličin.



Obr. 3 Problematické situace pro výpočet Pearsonova korelačního koeficientu.

3 Spearmanův korelační koeficient

Zatímco první situaci na obrázku 3 lze řešit rozdělením souboru na dva a následným výpočtem korelačního koeficientu v obou podsouborech, v situaci odpovídající grafu vpravo nahoře nemá smysl Pearsonův korelační koeficient počítat vůbec, neboť ten odráží pouze lineární závislost. Rozšíření směrem k hodnocení určitých forem nelineární závislosti představuje tzv. **Spearmanův korelační koeficient** (*Spearman rank-correlation coefficient*). Jedná se o neparametrický korelační koeficient, který je robustní vůči odlehlým hodnotám a obecně odchylkám od normality, neboť stejně jako řada dalších neparametrických metod pracuje pouze s pořadími pozorovaných hodnot. Na rozdíl od Pearsonova koeficientu korelace, který popisuje lineární vztah veličin X a Y, Spearmanův koeficient korelace popisuje, jak dobře vztah veličin X a Y odpovídá monotónní funkci, která může být samozřejmě nelineární.

Při výpočtu opět vycházíme z realizace dvourozměrného náhodného vektoru o rozsahu n, tedy dvojic pozorovaných hodnot náhodných veličin X a Y pro n subjektů. Dále definujme číslo x_{ri} jako pořadí hodnoty x_i v rámci vzestupně uspořádaných hodnot $x_1, ..., x_n$, číslo y_{ri} jako pořadí hodnoty y_i v rámci vzestupně uspořádaných hodnot $y_1, ..., y_n$, čísla $\bar{x}_r$ a $\bar{y}_r$ jako průměry hodnot x_{ri}, respektive y_{ri} (tedy jako průměrná pořadí), a čísla s_{x_r} a s_{y_r} jako odpovídající směrodatné odchylky. Spearmanův korelační koeficient, označme ho r_s, pak vypočítáme pomocí vzorce

$$r_s = \frac{\sum_{i=1}^{n} x_{ri} y_{ri} - n\bar{x}_r \bar{y}_r}{(n-1)s_{x_r} s_{y_r}}, \tag{11.16}$$

což není nic jiného než vzorec pro výběrový Pearsonův korelační koeficient počítaný na pořadích pozorovaných hodnot. Hodnoty r_s se pohybují stejně jako v případě koeficientu r v rozmezí od -1 do 1. Hodnot kolem nuly nabývá Spearmanův korelační koeficient v případě, že pořadí hodnot x_i a y_i jsou náhodně zpřeházená a mezi sledovanými veličinami není žádný vztah. Naopak hodnot -1 a 1 nabývá Spearmanův korelační koeficient v případě, že jedna z veličin je monotónní funkcí druhé veličiny.

Výpočetní alternativou ke vzorci (11.16) je výpočet založený na diferencích pořadí pozorovaných hodnot, které definujeme jako $d_i = x_{ri} - y_{ri}$. Hodnotu Spearmanova korelační koeficient pak odhadneme pomocí vztahu

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \tag{11.17}$$

Tento výpočet r_s platí přesně pouze pro neopakovaná pozorování, což znamená, že je citlivý na opakující se hodnoty, které vedou k průměrování pořadí. Vyskytuje-li se mezi hodnotami $x_1, ..., x_n$, respektive $y_1, ..., y_n$, množství shodných hodnot, je vhodnější použít k výpočtu Spearmanova korelačního koeficientu definiční vztah (11.16).

Příklad 4. Pro srovnání s hodnotou $r = 0{,}64$ vypočtenou v příkladu 1 odhadneme korelaci výšky a hmotnosti studentů biostatistiky také pomocí Spearmanova koeficientu korelace. Hodnoty potřebné k výpočtu jsou uvedeny v tabulce 2. Vzhledem k přítomnosti opakovaných hodnot u výšky i hmotnosti vypočteme nejprve Spearmanův korelační koeficient s použitím vzorce (11.16):

$$r_s = \frac{\sum_{i=1}^{n} x_{ri} y_{ri} - n\bar{x}_r \bar{y}_r}{(n-1)s_{x_r} s_{y_r}} = \frac{721{,}5 - 637}{(13-1) * 3{,}86 * 3{,}88} = 0{,}47. \tag{11.18}$$

Dále vypočteme hodnotu r_s i pomocí vztahu (11.17). V tomto případě dosadíme hodnoty z tabulky 2 následovně:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6*191}{13(13^2-1)} = 0{,}48\,, \tag{11.19}$$

Je vidět, že v tomto případě dávají oba výpočty koeficientu r_s velmi podobné výsledky, které odpovídají střední korelaci mezi výškou a hmotností. Oba výsledky se však liší od původně vypočtené hodnoty r = 0,64. Důvodem jsou dvě pozorování odpovídající hmotnosti 90 kg, které úplně nekorespondují se zbytkem souboru (viz obrázek 1). V tomto případě, kdy máme velmi limitovanou velikost výběrového souboru, je tedy lepší dát přednost neparametrické variantě, tedy hodnotě Spearmanova koeficientu korelace.

Tabulka 2 *Hodnoty pro výpočet Spearmanova koeficientu korelace výšky a hmotnosti studentů.*

Student	Výška: x_i	Pořadí výšky	Hmotnost: y_i	Pořadí hmotnosti	Rozdíl d_i	d_i^2
1	175	10	69	10	0	0
2	166	1	55	3	-2	4
3	170	4	67	8	-4	16
4	169	2,5	52	1	1,5	2,25
5	188	13	90	12,5	0,5	0,25
6	175	10	53	2	8	64
7	176	12	57	4,5	7,5	56,25
8	171	5	57	4,5	0,5	0,25
9	173	6,5	68	9	-2,5	6,25
10	175	10	73	11	-1	1
11	173	6,5	62	6	0,5	0,25
12	174	8	90	12,5	-4,5	20,25
13	169	2,5	63	7	-4,5	20,25

Konstrukce 100(1 – α)% intervalu spolehlivosti i test nulové hypotézy H_0: $r_s = 0$ probíhá pro Spearmanův korelační koeficient stejně jako pro koeficient Pearsonův. Co se týče konstrukce intervalu spolehlivosti, výběrové rozdělení r_s je pro výběry o velikosti alespoň 10 stejné jako výběrové rozdělení r. Pro větší vzorky, kdy je velikost souboru alespoň 30, je pak možné použít pro ověření nulové hypotézy $r_s = 0$ stejnou testovou statistiku jako v případě r danou vztahem (11.13). Pro zamítnutí H_0: $r_s = 0$ pak platí také stejná pravidla jako pro koeficient r.

Příklady k řešení:

1. U 10 pacientů byla změřena charakteristika *X*, a to dvakrát v rozmezí dvou dnů, před a po chirurgickém výkonu. Výsledky jsou dány tabulkou. Zajímá nás, jak spolu obě měření korelují. Vypočítejte Pearsonův koeficient korelace *r*.

Osoba	Hodnota před výkonem	Hodnota po výkonu
1	6	8
2	3	0,5
3	8	7
4	7	4
5	5	6
6	7	10
7	3	5
8	3,5	4
9	5,5	5
10	2,5	6
Průměr	5,05	5,55
SD	1,96	2,57

   [Výsledek: Pearsonův koeficient korelace $r = 0{,}545$]

2. Na datech z předchozího příkladu testujte nulovou hypotézu, že Pearsonův korelační koeficient *r* je roven nule, tedy, že mezi hodnotami před a po zákroku není lineární vztah.

   [Výsledek: $T = 1{,}84$ a příslušný kvantil Studentova t rozdělení $t_{(1-\alpha/2)}(n-2) = 2{,}31$, nulovou hypotézu nezamítáme]

Použitá literatura:

1. Sokal RR, Rohlf FJ. Biometry, The principles and practice of statistics in biological research. 3rd edition, W. H. Freeman and company, New York, 1995.
2. Andersen PK, Skovgaard LT. Regression with Linear Predictors. Springer, New York, 2010.
3. Zar JH. Biostatistical Analysis. 5th edition, Pearson Prentice-Hall, New Jersey, 2010.

Doporučená literatura:

1. Zvára K. Biostatistika. Nakladatelství Karolinum, Praha, 2006.
2. Zvárová J. Základy statistiky pro biomedicínské obory. Nakladatelství Karolinum, Praha, 2004.