

# Analiza zależności cech statystycznych

## Współczynnik korelacji liniowej Pearsona

### oraz współczynnik korelacji rang Spearmana

**Literatura:** Ostasiewicz, Rusnak, Siedlecka. Statystyka elementy teorii i zadania. Wydawnictwo AE we Wrocławiu, Wrocław 1999 (2006 – wydanie 6), str. 274-297.

#### Przykład 1. Współczynnik korelacji Pearsona. Proste regresji (szereg rozdzielczy szczegółowy)

Przeprowadzono badanie zależności: X - ładowności posiadanych środków transportu [tony] oraz Y – osiąganą przez środki transportu średnią prędkości [km/h] w pewnej firmie transportowej. Dane przedstawia tabela:

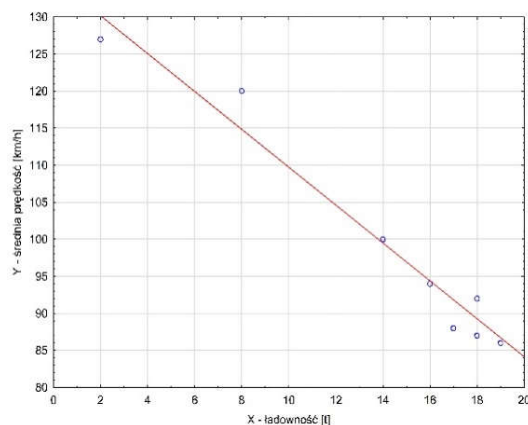
Ładowność $X_i$ [tony]	Średnia prędkość $Y_i$ [km/h]
14	100
18	92
19	86
2	127
8	120
2	127
8	120
16	94
17	88
18	87

Na podstawie zebranych danych statystycznych przeprowadzić następujące analizy:

- Zilustrować na wykresie rozrzutu wzajemną relację (zależność) badanych zmiennych
- Obliczyć współczynnik korelacji liniowej Pearsona oraz podać jego interpretację praktyczną
- Wyznaczyć proste regresji  $Y=f(X)$  i  $X=f(Y)$  oraz zinterpretować współczynniki tych prostych regresji

Ad a) Rysunek (Rys. 1) przedstawia zaobserwowaną zależność średniej prędkości samochodów od ich ładowności. Czerwoną linią zilustrowano prostą regresji  $Y=f(X)=a \cdot X+b$ .

Można zauważyć, że jest to zależność bardzo istotna ale przeciwna (duży ale ujemny współczynnik korelacji). Wraz ze wzrostem ładowności maleje średnia prędkość rozwijana przez te samochody.



**Rys. 1.** Wykres rozrzutu Y względem X.

Ad b)

W przypadku danych statystycznych podanych w postaci szeregu rozdzielczego szczegółowego współczynnik korelacji liniowej Pearsona (unormowaną miarę współzależności cech) wyznaczamy ze wzoru definicyjnego:

$$R_{xy} = \frac{C(X,Y)}{S_x * S_y} = \frac{\frac{1}{n} \sum_{i=1}^n [(x_i - \bar{x}) * (y_i - \bar{y})]}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} * \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\sum_{i=1}^n [(x_i - \bar{x}) * (y_i - \bar{y})]}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} * \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

gdzie:  $C(X,Y)$  – jest tzw. cowaiancją - miarą współzależności cech,  $S_x$  – jest odchyleniem standardowym cechy X, zaś  $S_y$  – jest odchyleniem standardowym cechy Y.

Współczynnik korelacji możemy również policzyć z innego równoważnego wzoru wykorzystując zależności:

Kowariancję możemy inaczej zapisać w postaci:

$$\frac{1}{n} \sum_{i=1}^n [(x_i - \bar{x}) * (y_i - \bar{y})] = \frac{1}{n} \sum_{i=1}^n (x_i * y_i) - \bar{x} * \bar{y}$$

$$\text{Natomiast odchylenia standardowe możemy zapisać jako: } S_x = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2}, S_y = \sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2}$$

Wtedy współczynnik korelacji obliczamy ze wzoru (który często jest łatwiejszy do obliczeń):

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

Skorzystamy ze wzoru (2). Obliczenia przedstawiono w tabeli:

Ładowność $X_i$ [tony]	Średnia prędkość $Y_i$ [km/h]	$x_i * y_i$	$x_i^2$	$y_i^2$
14	100	1400	196	10000
18	92	1656	324	8464
19	86	1634	361	7396
2	127	254	4	16129
8	120	960	64	14400
2	127	254	4	16129
8	120	960	64	14400
16	94	1504	256	8836
17	88	1496	289	7744
18	87	1566	324	7569
$\Sigma = 122$	$\Sigma = 1041$	$\Sigma = 11684$	$\Sigma = 1886$	$\Sigma = 111067$

**Średnie dla statystyk:**

$\bar{x} = \frac{122}{10} = 12,2$ ;  $\bar{y} = \frac{1041}{10} = 104,1$ . Zatem średnia ładowność środków transportu wynosi ok. 12 (12.2) [tony], zaś średnią prędkość jaką rozwijały pojazdy floty transportowej firmy wynosiła ok. 104 (104.1) [km/h].

**Miary zmienności – odchylenia standardowe:**

Wariancja  $S_x^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = \frac{1886}{10} - 12,2^2 = 39,76$ . Zatem odchylenie  $S_x = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2} = \sqrt{39,76} = 6,31$ .

Wariancja  $S_y^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2 = \frac{111067}{10} - 104,1^2 = 269,89$ . Zatem odchylenie  $S_y = \sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2} = \sqrt{269,89} = 16,43$ .

Otrzymane miary zmienności obu statystyk oraz średnie wartości wskazują na pewną zmienność ich wartości. Oznacza to, że typowy przedział zmienności dla ładowności jest postaci: (5,89-18,5) [tony], zaś dla prędkości średniej tych pojazdów postaci: (87,67-120,53) [km/h].

Współczynnik korelacji liniowej Pearsona liczony ze wzoru (2) jest zatem następujący:

$$R_{xy} = \frac{\frac{1}{10} * 11684 - 12,2 * 104,1}{6,31 * 16,43} = -0,98$$

Interpretacja praktyczna: istnieje bardzo duża prawie perfekcyjna (idealna) ale przeciwna zależność liniowa (bardzo duża bliska -1 wartość współczynnika korelacji) pomiędzy prędkością maksymalną a ładownością. Wyrażają tę zależność tzw. proste regresji, które teraz wyznaczymy.

Ad c)

Można określić dwie postacie prostych regresji:

- pierwszą określając liniową zależność  $Y = f(X) = a_1 * X + b_1$
- drugą określając na odwrót liniową zależność  $X = f(Y) = a_2 * Y + b_2$

Aby określić równanie regresji w pierwszej postaci skorzystamy z zależności:  $Y - \bar{y} = R_{xy} * \frac{S_y}{S_x} (X - \bar{x})$ . Przekształcając tę zależność otrzymujemy:  $Y = R_{xy} * \frac{S_y}{S_x} * X + \left( \bar{y} - R_{xy} * \frac{S_y}{S_x} * \bar{x} \right) = a_1 * X + b_1$ ,

gdzie: współczynnik kierunkowy:  $a_1 = R_{xy} * \frac{S_y}{S_x}$ , zaś wyraz wolny:  $b_1 = \bar{y} - R_{xy} * \frac{S_y}{S_x} * \bar{x}$ .

Podobnie przekształcając zależność:  $X - \bar{x} = R_{xy} * \frac{S_x}{S_y} (Y - \bar{y})$  otrzymujemy równanie regresji:  $X = a_2 * Y + b_2$ , gdzie: współczynnik kierunkowy:  $a_2 = R_{xy} * \frac{S_x}{S_y}$ , zaś wyraz wolny:  $b_2 = \bar{x} - R_{xy} * \frac{S_x}{S_y} * \bar{y}$ .

Dlatego równanie pierwszej prostej regresji - opisującej zależność średniej prędkości od ładowności jest postaci:

$$Y = a_1 * X + b_1 = -0,98 * \frac{16,43}{6,31} * X + \left( 104,1 + 0,98 * \frac{16,43}{6,31} * 12,2 \right) = -2,56 * X + 135,28.$$

Natomiast równanie drugiej prostej regresji - opisującej zależność ładowności od średniej prędkości jest następujące:

$$X = a_2 * Y + b_2 = -0,98 * \frac{6,31}{16,43} * Y + \left( 12,2 + 0,98 * \frac{6,31}{16,43} * 104,1 \right) = -0,38 * Y + 51,39.$$

#### Interpretacja praktyczna parametrów prostej regresji (1):

Współczynnik kierunkowy -2,56 interpretujemy jako wzrost (spadek) wartości zmiennej zależnej Y, gdy zmienna niezależna (w tym przypadku X) zmieni się o 1 jednostkę.

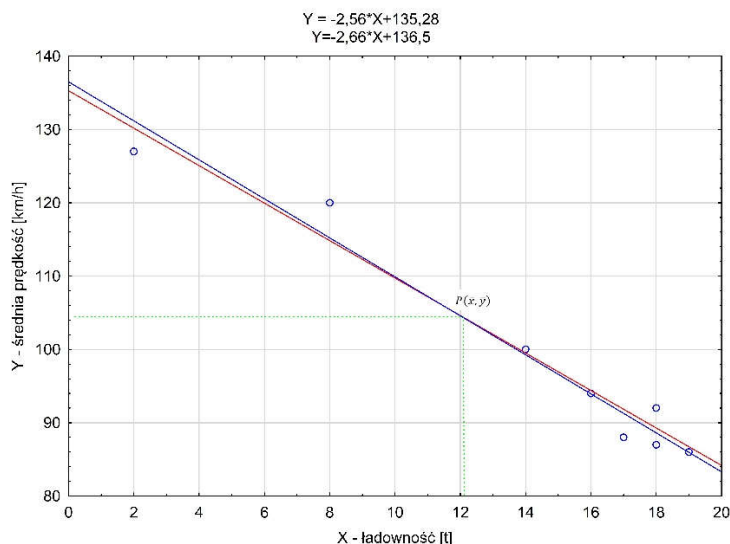
Oznacza to, że jeżeli ładowność zostanie zwiększona o 1 tonę to średnia prędkość zmaleje o 2,56 [km/h]. Jeżeli zwiększymy proporcjonalnie o 10 [ton], to ulegnie zmniejszeniu o ok. 25,6 [km/h].

Natomiast wyraz wolny 135,28 interpretujemy jako pewną stałą prędkość średnią niezależną od ładowności (gdyby ładowność była równa zero) i od niej jest zmniejszana (maleje) wartość prędkości liniowo względem wzrastającej ładowności pojazdu o wartość 2,56 [km/h] na każdy wzrost ładowności o 1 [tonę].

Jeżeli przekształcimy drugie równanie regresji (2), tak aby można było narysować obie te proste w jednym układzie współrzędnych (czyli wyznaczyć z niego zależność Y względem X), to otrzymamy równanie:

$$Y = -2,66 * X + 136,5.$$

Rysunek (rys. 2) pokazuje wzajemną zależność pomiędzy dwoma prostymi regresji. Przecinają się one w jednym punkcie wspólnym  $P(\bar{x}, \bar{y}) = (12,2, 104,1)$ .



Rys. 2. Interpretacja prostych regresji.

### Przykład 2. Współczynnik korelacji Pearsona. Proste regresji (tabela korelacyjna)

Przeprowadzono badanie zależności: X – staż pracy [lata] oraz Y – płacy brutto pracowników [tys. zł] w pewnej firmie logistycznej w 1995 roku. Wyniki badania dla n=100 pracowników w postaci dwudzielczej tabeli korelacyjnej licznosci z opracowanymi przedziałami zmienności dla płacy brutto oraz stażu pracy przedstawia tabela:

X – Staż pracy [lata]	Y – Płaca brutto [tys. zł]			
	[0-1)	[1-2)	[2-3)	[3-4)
[0-10)	25	4	3	1
[10-20)	18	4	1	0
[20-30)	20	10	0	0
[30-40)	12	2	0	0

Na podstawie otrzymanych danych statystycznych wyznaczyć:

- współczynnik korelacji liniowej Pearsona oraz podać jego interpretację praktyczną
- proste regresji  $Y=f(X)$  i  $X=f(Y)$  oraz zinterpretować współczynniki tych prostych regresji

### Rozwiązanie:

Oznaczmy  $n_{ij}$  – licznosci rozkładu łącznego wektora (X,Y) badanych cech statystycznych. Jest to taka liczba wyników w tabeli korelacyjnej, gdzie zmienna X przyjmuje wartości z „i” – tego swojego przedziału zmienności, zaś statystyka Y z „j-tego” (np.  $n_{32} = 10$  – było 10 pracowników, którzy mieli staż pracy w przedziale od 20-30 lat, zaś zarobki w przedziale od 1-2 tys. zł.)

Podobnie możemy określić tzw. licznosci rozkładów brzegowych (tylko dla pojedynczych cech).

$n_{i*} = \sum_{j=1}^k n_{ij}$  – licznosci w wierszach (dla rozkładu tylko statystyki X), „k”-liczba kolumn tabeli korelacyjnej

$n_{*j} = \sum_{i=1}^w n_{ij}$  – licznosci w kolumnach (dla rozkładu tylko statystyki Y), „w”-liczba wierszy tabeli korelacyjnej

X - Staż pracy [lata]	Y – Płaca brutto [tys. zł]				Suma $n_{i*}$	$n_{i*} * \bar{x}_i$	$n_{*j} * \bar{y}_j$	$n_{i*} * \bar{x}_i^2$	$n_{*j} * \bar{y}_j^2$
	[0-1)	[1-2)	[2-3)	[3-4)					
[0-10)	25	4	3	1	33	$33*5=165$	$75*0,5=37,5$	$33*5^2=825$	$75*0,5^2=18,75$
[10-20)	18	4	1	0	23	$23*15=345$	$20*1,5=30$	$23*15^2=5175$	$20*1,5^2=45$
[20-30)	20	10	0	0	30	$30*25=750$	$4*2,5=10$	$30*25^2=18750$	$4*2,5^2=25$
[30-40)	12	2	0	0	14	$14*35=490$	$1*3,5=3,5$	$14*35^2=17150$	$1*3,5^2=12,25$
Suma $n_{*j}$	75	20	4	1	100	1750	81	41900	101

W badaniu objętych było zatem np. 30 pracowników, którzy mieli staż pracy od 20-30 lat nieważne jaką posiadali płacę brutto. Podobnie było np. 75 pracowników mających płacę brutto do 1 tys. zł, nieważne jaki mieli staż (bez rozbijania tej grupy ze względu na staż).

Ad a)

W przypadku danych statystycznych zgrupowanych w postaci tabeli korelacyjnej (dwudzielczej) współczynnik korelacji liniowej Pearsona możemy obliczyć ze wzoru:

$$R_{xy} = \frac{\frac{1}{n} \sum_{i=1}^w \sum_{j=1}^k (n_{ij} * \hat{x}_i * \hat{y}_j) - \frac{1}{n} \sum_{i=1}^w (n_{i*} * \hat{x}_i) * \frac{1}{n} \sum_{j=1}^k (n_{*j} * \hat{y}_j)}{\sqrt{\frac{1}{n} \sum_{i=1}^w n_{i*} * \hat{x}_i^2 - \left(\frac{1}{n} \sum_{i=1}^w (n_{i*} * \hat{x}_i)\right)^2} * \sqrt{\frac{1}{n} \sum_{j=1}^k n_{*j} * \hat{y}_j^2 - \left(\frac{1}{n} \sum_{j=1}^k (n_{*j} * \hat{y}_j)\right)^2}} \quad (3)$$

gdzie:  $\hat{x}_i$  – oznacza środek i-tego przedziału zmienności dla statystyki X, zaś  $\hat{y}_j$  – oznacza środek j-tego przedziału zmienności dla statystyki Y.

Obliczmy najpierw średnie dla obu statystyk:

$$\text{Średnią dla statystyki X obliczamy ze wzoru: } \frac{1}{n} \sum_{i=1}^w (n_{i*} * \hat{x}_i) = \frac{1750}{100} = 17,5.$$

$$\text{Średnią dla statystyki Y obliczamy ze wzoru: } \frac{1}{n} \sum_{j=1}^k (n_{*j} * \hat{y}_j) = \frac{81}{100} = 0,81.$$

Zatem średni staz pracy pracowników wynosił 17,5 lat, zaś średnia płaca brutto ich zarobków na poziomie 0,81 tys. zł (ok. 800 zł).

Podobnie wyznaczmy potrzebne do wzoru (3) wartości odchyleń standardowych obu statystyk

$$\text{Wariancję dla statystyki X obliczamy ze wzoru: } \frac{1}{n} \sum_{i=1}^w n_{i*} * \hat{x}_i^2 - \left(\frac{1}{n} \sum_{i=1}^w (n_{i*} * \hat{x}_i)\right)^2 = \frac{41900}{100} - 17,5^2 = 112,75.$$

$$\text{Zatem odchylenie standardowe dla stazu pracy wynosi: } S_x = \sqrt{112,75} = 10,62 \text{ [lat].}$$

$$\text{Wariancję dla statystyki Y obliczamy ze wzoru: } \frac{1}{n} \sum_{j=1}^k n_{*j} * \hat{y}_j^2 - \left(\frac{1}{n} \sum_{j=1}^k (n_{*j} * \hat{y}_j)\right)^2 = \frac{101}{100} - 0,81^2 = 0,35.$$

$$\text{Zatem odchylenie standardowe dla zarobków wynosi: } S_y = \sqrt{0,35} = 0,59 \text{ [tys. zł].}$$

Aby wyznaczyć składową wzoru (3):  $\frac{1}{n} \sum_{i=1}^w \sum_{j=1}^k (n_{ij} * \hat{x}_i * \hat{y}_j)$  możemy skorzystać obliczeń w tabeli

- $n_{ij} * \hat{x}_i * \hat{y}_j$

X - Staż pracy [lata]	Y – Płaca brutto [tys. zł]				
	[0-1)	[1-2)	[2-3)	[3-4)	
[0-10)	25*5*0,5=62,5	4*5*1,5=30	3*5*2,5=37,5	1*5*3,5=17,5	
[10-20)	18*15*0,5=135	4*15*1,5=90	1*15*2,5=37,5	0*15*3,5=0	
[20-30)	20*25*0,5=250	10*25*1,5=375	0*25*2,5=0	0*25*3,5=0	
[30-40)	12*35*0,5=210	2*35*1,5=1-5	0*35*2,5=0	0*35*3,5=0	<b>Suma</b>
<b>Suma</b>	<b>657,5</b>	<b>600</b>	<b>75</b>	<b>17,5</b>	<b>1350</b>

$$\frac{1}{n} \sum_{i=1}^w \sum_{j=1}^k (n_{ij} * \hat{x}_i * \hat{y}_j) = \frac{1350}{100} = 13,5$$

$$\text{Zatem } R_{xy} = \frac{13,5 - 17,5 * 0,81}{10,62 * 0,59} = -0,11.$$

Istnieje dość mała (niewielka) zależność liniowa przeciwna pomiędzy płacą brutto pracowników a ich stażem pracy. Wraz ze wzrostem stażu pracy maleje płaca brutto pracowników. Więcej zarabiają pracownicy młodzi (z mniejszym stażem pracy)

Ad b)

Proste regresji:

- równanie pierwszej prostej regresji (opisującej zależność płacy brutto od stazu pracy) jest postaci:

$$Y = a_1 * X + b_1 = -0,11 * \frac{0,59}{10,62} * X + \left(0,81 + 0,11 * \frac{0,59}{10,62} * 17,5\right) = -0,006 * X + 0,92.$$

- równanie drugiej prostej regresji (opisującej zależność stazu pracy od płacy brutto) jest następujące:

$$X = a_2 * Y + b_2 = -0,11 * \frac{10,62}{0,59} * Y + \left(17,5 + 0,11 * \frac{10,62}{0,59} * 0,81\right) = -1,98 * Y + 19,1.$$

Z interpretacji praktycznej współczynników prostej regresji  $Y=f(X)$  wynika, że wzrost stażu pracy np. o 10 lat powoduje zmniejszenie się płacy brutto tych pracowników o ok. -0,06 (tys. zł – 60 zł). Natomiast stała płaca brutto niezależna od stażu pracy to 0,92 tys. zł (920 zł).

### Przykład 3. Współczynnik korelacji rang Spearmana

W tabeli podano dane statystyczne o wartościach wskaźników statystycznych w województwach Polski w 2016 roku: X – liczby motorów zarejestrowanych w tych województwach w przeliczeniu na 1 tys. osób, Y – liczby wypadków śmiertelnych w tych województwach w przeliczeniu na 100 tys. osób.

Dokonać rangowania wartości tych wskaźników, a następnie wyznaczyć nieparametryczny współczynnik korelacji rang Spearmana dla tych statystyk. Wyniki zinterpretować praktycznie.

Województwo	X – liczba motorów / 1000 osób	Y – liczba wypadków śmiertelnych / 100 000 osób
Dolnośląskie	14	13
Kujawsko-pomorskie	24	16
Lubelskie	24	15
Lubuskie	26	17
Łódzkie	21	17
Małopolskie	18	10
Mazowieckie	16	17
Opolskie	14	14
Podkarpackie	30	14
Podlaskie	21	18
Pomorskie	16	12
Śląskie	14	9
Świętokrzyskie	23	18
Warmińsko-mazurskie	24	20
Wielkopolskie	24	15
Zachodniopomorskie	24	14

Oparty o rangi współczynnik korelacyjny Spearmana jest alternatywą dla parametrycznego współczynnika korelacji liniowej Pearsona. Wyznaczamy go z następującego wzoru:

$$r_s = 1 - \frac{6 * \sum_{i=1}^n (d_i)^2}{n * (n^2 - 1)} = 1 - \frac{6 * \sum_{i=1}^n (r_i^{(y)} - r_i^{(x)})^2}{n * (n^2 - 1)} \quad (4)$$

gdzie:  $r_i^{(y)}$  – przypisane rangi dla i-tej uporządkowanej wartości cechy Y,  $r_i^{(x)}$  – przypisane rangi dla odpowiadającej i-tej wartości cechy X.

Dane statystyczne porządkujemy (sortujemy) np. od największej do najmniejszej według wartości cechy Y. Kolejna tabela przedstawia dane uporządkowane. Następnie przypisujemy rangi robocze poszczególnym wartościom zmiennych Y i X. Przyjmujemy, że ranga określa które miejsce w rankingu zajmuje dane województwo ze względu na badaną zmienną. Im niższa ranga tym wyższe miejsce w rankingu (np. pierwsze  $r_i = 1$ ). Porządkujemy wartości malejąco (od największych) zatem będziemy traktować województwo warmińsko-mazurskie jako zajmujące pierwsze miejsce w niekorzystnym rankingu największej liczby wskaźnika wypadków śmiertelnych w Polsce. Natomiast województwo podkarpackie jest pierwsze w rankingu ze względu na wartości wskaźnika liczby motorów.

Ponieważ niektóre wartości statystyk mogą się powtarzać, to ostateczne rangi są wyliczane jako średnia z rang roboczych dla powtarzających się wyników. Ze względu na wypadkowość np. podlaskie i świętokrzyskie mają te same wartości 18, ale

przypisaliśmy im kolejne różne rangi 2 i 3. Wyliczamy zatem średnią równą:  $\frac{2+3}{2} = 2,5$  i ta wspólną średnią rangę (taką samą) przypisujemy obu województwom. Podobnie postępujemy w innych takich sytuacjach. Obliczenia podaje tabela.

Województwo	X	Y	Rangi				$d_i = r_i^{(y)} - r_i^{(x)}$	$d_i^2$
			$r_i^{(y)}$ rangi robocze	$r_i^{(y)}$ rangi właściwe	$r_i^{(x)}$ rangi robocze	$r_i^{(x)}$ rangi właściwe		
Warmińsko-mazurskie	24	20	1	1	3	5	1-5=-4	16
Podlaskie	21	18	2	2,5	9	9,5	2,5-9,5=-7	49
Świętokrzyskie	23	18	3	2,5	8	8	2,5-8=-5,5	30,25
Lubuskie	26	17	4	5	2	2	5-2=3	9
Łódzkie	21	17	5	5	10	9,5	5-9,5=-4,5	20,25
Mazowieckie	16	17	6	5	13	12,5	5-12,5=-7,5	56,25
Kujawsko-pomorskie	24	16	7	7	4	5	7-5=2	4
Lubelskie	24	15	8	8,5	5	5	8,5-5=3,5	12,25
Wielkopolskie	24	15	9	8,5	6	5	8,5-5=3,5	12,25
Opolskie	14	14	10	11	14	15	11-15=-4	16
Podkarpackie	30	14	11	11	1	1	11-1=10	100
Zachodniopomorskie	24	14	12	11	7	5	11-5=6	36
Dolnośląskie	14	13	13	13	15	15	13-15=-2	4
Pomorskie	16	12	14	14	12	12,5	14-12,5=1,5	2,25
Małopolskie	18	10	15	15	11	11	15-11=4	16
Śląskie	14	9	16	16	16	15	16-15=1	1
Suma:								384,5

Na podstawie wyników cząstkowych obliczeń w tabeli otrzymujemy wartość współczynnika korelacji rang Spearmana:

$$r_S = 1 - \frac{6 \cdot 384,5}{16 \cdot (16^2 - 1)} = 0,44.$$

**Interpretacja współczynnika:** Współczynnik Pearsona przyjmuje wartości podobnie jak współczynnik korelacji liniowej Pearsona z przedziału  $[-1,1]$ . Ponieważ jest on dodatni to istnieje średnia zależność zgodna w uporządkowaniu (przypisaniu) rang dla obu badanych statystyk. Raczej niskim rangom dla statystyki Y odpowiadają niskie rangi dla statystyki X (ale zgodność ta jest tylko na poziomie ok. 44%). Tym samym można wnioskować, że województwa, które posiadały wysoki wskaźnik X dla liczby motorów jeżdżących po ich drogach (wysokie miejsca w rankingu, niskie wartości rang) zajmowały także wysokie miejsca w niechlubnym rankingu wypadkowości (miały niższe wartości rang dla Y). Można powiedzieć że obserwuje się średnią zależność Y od X w województwach Polski w badanym roku.

- **Zadania do samodzielnego rozwiązania (z książki: Ostasiewicz, Rusnak, Siedlecka. Statystyka elementy teorii i zadania. Wydawnictwo AE we Wrocławiu, Wrocław 1999 (2006), str. 274-297).**

Zad. 11.1, Zad. 11.2, Zad. 11.5, Zad. 11.10, Zad. 11.15.