

Kolumna $G(x)$ ukazuje stopień przyrastania ogólnej liczby nabywców, gdy przesuwamy się do następnej kategorii wielkości zakupu, natomiast kolumna $G'_x(x)$ ukazuje stopień przyrastania ogólnej liczby sprzedanych opakowań tym nabywcom. Stopień koncentracji określa współczynnik Giniego, który przyjmuje wartość:

$$K(x) = 1 - 0,4394 = 0,5606.$$

Można stwierdzić, że w tej grupie nabywców występuje umiarkowana tendencja do nabywania znacznej liczby opakowań produktu przez nieznaną liczbę konsumentów.

7.2.3. Miary współwystępowania

Sposób pomiaru współwystępowania zjawisk charakteryzujących rynek zależy od rodzaju danych, które te zjawiska opisują. Zestaw najpopularniejszych podejść stosowanych w analizie współwystępowania przedstawia tabela 7.10.

Tabela 7.10. Najpopularniejsze metody pomiaru współwystępowania

Cecha zależna	Cecha niezależna		
	dane nominalne	dane porządkowe	dane przedziałowe lub ilorazowe
Dane nominalne	współczynnik V Cramera	współczynnik V Cramera	
Dane porządkowe	współczynnik V Cramera	współczynniki korelacji rang	
Dane przedziałowe lub ilorazowe	współczynnik eta	współczynnik eta	współczynnik korelacji Pearsona

7.2.3.1. Współczynnik V Cramera

Podstawą pomiaru siły współwystępowania dwóch cech opisywanych przez dane nominalne jest tabela kontyngencji, która zestawia w wierszach i kolumnach kategorie jednej i drugiej cechy nominalnej. Na przecięciu każdego wiersza i każdej kolumny znajduje się wielkość określająca liczbę jednostek, które w zbiorze danych posiadają jednocześnie obie kategorie (wyróżnioną w wierszu i w kolumnie). Schemat tablicy kontyngencji dla trzech kategorii cechy zależnej oraz dwóch kategorii cechy niezależnej przyjmuje następującą postać.

Tabela 7.11. Schemat tablicy kontyngencji

Kategorie cechy zależnej	Kategorie cechy niezależnej		Ogółem
	A	B	
Q	n_{QA}	n_{QB}	$n_{QA} + n_{QB} = n_Q$
P	n_{PA}	n_{PB}	$n_{PA} + n_{PB} = n_P$
R	n_{RA}	n_{RB}	$n_{RA} + n_{RB} = n_R$
Ogółem	$n_{QA} + n_{PA} + n_{RA} = n_A$	$n_{QB} + n_{PB} + n_{RB} = n_B$	n

Gdyby cechy były niezależne, wówczas liczebności wewnątrz tablicy powinny przyjmować wartości $\hat{n}_{ij}$ wyznaczone zgodnie ze wzorem:

$$\hat{n}_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}.$$

Im większa jest rozbieżność między liczebnościami obserwowanymi empirycznie i liczebnościami teoretycznymi (określającymi brak zależności), tym silniejsza jest zależność obu cech. Miarą rozbieżności jest statystyka χ^2 , którą wyznacza wzór:

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}},$$

gdzie: i, j – numery kolejnych wierszy i kolumn w tablicy kontyngencji, przy czym $i = 1, 2, \dots, k$, zaś $j = 1, 2, \dots, l$.

Wartość statystyki χ^2 jest podstawą konstrukcji współczynnika V Cramera, według wzoru:

$$C = \sqrt{\frac{\chi^2}{n(g-1)}},$$

gdzie: $g = \min\{k, l\}$.

Współczynnik ten przyjmuje wartości z przedziału $[0, 1]$ i im jest większy, tym silniejsza jest zależność między cechami nominalnymi. Może on być również miarą zależności między cechami dowolnego typu, jeśli tylko da się je zestawić w tablicę kontyngencji.

Przykład 7.15

Dla zbiorowości firm, których łączny rozkład według typu dostawcy oraz liczby reklamacji przedstawia tabela 7.12, rozkład teoretyczny, czyli taki, jaki powinien wystąpić, gdyby cechy te były niezależne, przyjmuje postać zawartą w tablicy 7.13.

Tablica 7.12. Rozkład liczby reklamacji według typu dostawcy

Liczba reklamacji	Typ dostawcy			Ogółem
	krajowy producent z sektora publicznego	krajowy producent z sektora prywatnego	producent zagraniczny	
0	10	40	10	60
1	40	20	20	80
2	30	10	20	60
3	20	10	10	40
4	10	10	10	30
Ogółem	110	90	70	270

Tablica 7.13. Rozkład liczby reklamacji według typu dostawcy przy niezależności cech

Liczba reklamacji	Typ dostawcy		
	krajowy producent z sektora publicznego	krajowy producent z sektora prywatnego	producent zagraniczny
0	24,4	20,0	15,5
1	32,6	26,7	20,7
2	24,4	20,0	15,5
3	16,3	13,3	10,4
4	12,2	10,0	7,8

Na przykład: $\hat{n}_{11} = \frac{n_{1.} \cdot n_{.1}}{n} = \frac{110 \cdot 60}{270} = 24,4$, $\hat{n}_{12} = \frac{\hat{n}_{1.} \cdot n_{.2}}{n} = \frac{90 \cdot 60}{270} = 20$.

Rozbieżność między rozkładem faktycznym i teoretycznym, mierzona wartością statystyki χ^2 , wynosi zatem:

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} = \frac{(10 - 24,4)^2}{24,4} + \frac{(40 - 20)^2}{20} + \frac{(10 - 15,5)^2}{15,5} + \frac{(40 - 32,6)^2}{32,6} + \frac{(20 - 26,7)^2}{26,7} + \frac{(20 - 20,7)^2}{20,7} + \frac{(30 - 24,4)^2}{24,4} + \frac{(10 - 20)^2}{20} + \frac{(20 - 15,5)^2}{15,5} + \frac{(20 - 16,3)^2}{16,3} + \frac{(10 - 13,3)^2}{13,3} + \frac{(10 - 10,4)^2}{10,4} + \frac{(10 - 12,2)^2}{12,2} + \frac{(10 - 10)^2}{10} + \frac{(10 - 7,8)^2}{7,8} \approx 44,1177.$$

Współczynnik V Cramera przyjmuje wartość:

$$C = \sqrt{\frac{\chi^2}{n(g-1)}} = \sqrt{\frac{44,1177}{270 \cdot (3-1)}} \approx 0,28,$$

co oznacza, że między liczbą reklamacji i typem dostawcy występuje słaba zależność statystyczna.

7.2.3.2. Współczynniki korelacji rang

Współczynnik korelacji rang pozwala określić zbieżność lub rozbieżność danych o charakterze porządkowym, np. rang nadawanych produktom lub markom. Współczynnik korelacji rang Spearmana umożliwia porównanie rang nadanych jedynie dwóm cechom i wyznacza się go według wzoru:

$$r_d = 1 - 6 \cdot \frac{\sum (y_i - x_i)^2}{n(n^2 - 1)},$$

gdzie: x_i – rangi nadane pierwszej cesze, dla $i = 1, 2, \dots, n$; y_i – rangi nadane drugiej cesze; n – liczba obserwacji.

Współczynnik ten przyjmuje wartości z przedziału $[-1, 1]$, przy czym jego znak informuje o zbieżności (dodatni) lub rozbieżności (ujemny) ocen wyróżnionych cech, a wartość liczbowa określa siłę współwystępowania tych ocen.

Przykład 7.16

Tablica 7.14 przedstawia oceny wystawione przez dział marketingu dotyczące lojalności odbiorców oraz wizerunku marek konkurujących na pięciu rynkach, na których firma prowadzi swoją działalność.

Tablica 7.14. Rachunek współczynnika korelacji rang

Rynek	Ocena lojalności odbiorców w skali od 1 (ocena najniższa) do 5 (ocena najwyższa)	Ocena wizerunku marek konkurencyjnych w skali od 1 (ocena najniższa) do 5 (ocena najwyższa)
A	1	4
B	5	1
C	3	3
D	2	1
E	4	2

Wartość współczynnika korelacji rang Spearmana wynosi zatem:

$$r_d = 1 - 6 \cdot \frac{\sum (y_i - x_i)^2}{n(n^2 - 1)} = 1 - 6 \cdot \frac{(1-4)^2 + (5-1)^2 + (3-3)^2 + (2-1)^2 + (4-2)^2}{5(5^2 - 1)} = 1 - 1,5 = -0,5$$

Wynik oznacza, że oceny działu marketingu dotyczące lojalności odbiorców oraz wizerunku marek konkurencyjnych na poszczególnych rynkach wykazują widoczną rozbieżność o umiarkowanej sile, wyrażającą się współwystępowaniem opinii o spadku stopnia lojalności odbiorców wraz z poprawą wizerunku marek konkurencyjnych.

W przypadku gdy celem pomiaru jest identyfikacja porządku występującego wśród badanych obiektów ze względu na badaną własność tych obiektów, rangi określają pozycję obiektu w uszeregowanym zbiorze. Może się zdarzyć, że kilka obiektów zajmuje równorzędną pozycję, co uniemożliwia precyzyjne ich uszeregowanie. W takich sytuacjach każdemu z tych obiektów nadaje się taką samą rangę – wartość średnią z tych rang, które powinny być nadane tej grupie obiektów. Takie obiekty nazywa się wiązanymi lub połączonymi. Występowanie obiektów powiązanych zmniejsza wartość współczynnika korelacji, a uzyskane uporządkowanie określa się mianem słabego uporządkowania³. Do pomiaru stopnia zbieżności uszeregowania obiektów ze względu na ich dwie własności, co utożsamia się z siłą skorelowania tych cech, można zastosować współczynnik korelacji rang τ_b Kendalla, który wyznacza się według wzoru:

$$\tau_b = \frac{2R}{\sqrt{\frac{n(n-1)}{2} - T_x} \cdot \sqrt{\frac{n(n-1)}{2} - T_y}} = 1,$$

gdzie: $T_x = \frac{1}{2} \sum_{i=1}^k t_{xi} (t_{xi} - 1)$ oraz $T_y = \frac{1}{2} \sum_{i=1}^k t_{yi} (t_{yi} - 1)$; t_{xi} , t_{yi} – liczba rang powiązanych w każdej i -tej podgrupie rang w uszeregowaniu odpowiednio według cechy X i Y ; k – liczba podgrup z rangami powiązanymi w zbudowanym uporządkowaniu; R – liczba par tych rang, które po uporządkowaniu według pierwszej cechy, czyli:

dla $j > i$ zachodzi $x_j \geq x_i$,

zachowują relację:

³ Por. J. Steczkowski, A. Zeliaś, *Metody statystyczne w badaniach jakościowych*, Akademia Ekonomiczna w Krakowie, Kraków 1997, s. 190.

dla $j > i$ zachodzi $y_j > y_i$

dla cechy drugiej, kolejno dla każdego $j = 1, 2, \dots, n$; n – liczba jednostek objętych badaniem.

Współczynnik ten przyjmuje wartości z przedziału $[-1, 1]$ i interpretowany jest analogicznie jak współczynnik korelacji rang Spearmana.

Przykład 7.17

Tablica 7.15 przedstawia uporządkowanie, dokonane przez dział marketingu, pięciu rynków, na których firma prowadzi swą działalność, ze względu na lojalność odbiorców oraz wizerunek marek konkurujących na tych rynkach.

Tablica 7.15. Rachunek współczynnika korelacji rang τ_b Kendalla

Rynek	Pozycja rynku ze względu na lojalność odbiorców	Pozycja rynku ze względu na wizerunek marek konkurencyjnych
A	1,5	5
B	4,5	2
C	1,5	4
D	3,0	2
E	4,5	2

Ocena korelacji obu cech wymaga ustalenia:

– dla pierwszego uszeregowania według lojalności odbiorców (cecha X), że:

$$k = 2, t_{x1} = 2, \text{ oraz } t_{x2} = 2, \text{ zatem } T_x = \frac{1}{2} \sum_{i=1}^k t_{xi} (t_{xi} - 1) = \frac{1}{2} [2(2-1) + 2(2-1)] = 2,$$

– dla drugiego uszeregowania według wizerunku marek konkurencyjnych (cecha Y), że:

$$k = 1, t_{y1} = 3, \text{ zatem } T_y = \frac{1}{2} \sum_{i=1}^k t_{yi} (t_{yi} - 1) = \frac{1}{2} 3(3-1) = 3.$$

Uporządkowanie rynków według rosnących rang nadanych pierwszej cesze przedstawia tablica 7.16.

Tablica 7.16. Uporządkowanie rynków według rang nadanych lojalności konsumentów

Rynek	Nr kolejny	Pozycja rynku ze względu na lojalność odbiorców	Pozycja rynku ze względu na wizerunek marek konkurencyjnych
A	1	1,5	5
C	2	1,5	4
D	3	3,0	2
B	4	4,5	2
E	5	4,5	2

Zatem $R = 0 + 0 + 0 + 0 = 0$, gdyż w drugiej kolumnie, czyli ze względu na cechę Y , nie występują wartości, które spełniają warunek: dla $j > 1$, $y_j > 5$, nie występują wartości, które spełniają warunek: dla $j > 2$, $y_j > 4$, nie występują wartości, które spełniają warunek: dla $j > 3$, $y_j > 2$, oraz nie występują wartości, które spełniają warunek: dla $j > 4$, $y_j > 2$.

Wartość współczynnika korelacji rang τ_b Kendalla wynosi:

$$\tau_b = \frac{2R}{\sqrt{\frac{n(n-1)}{2} - T_x} \cdot \sqrt{\frac{n(n-1)}{2} - T_y}} - 1 = \frac{2 \cdot 0}{\sqrt{\frac{5(5-1)}{2} - 2} \cdot \sqrt{\frac{5(5-1)}{2} - 6}} - 1 = 0 - 1 = -1.$$

Wynik wskazuje na pełną rozbieżność lojalności konsumentów i wizerunku marek konkurencyjnych na badanych rynkach.

Jeżeli nie występują rangi powiązane, czyli uporządkowanie obiektów jest mocne, to wzór na współczynnik korelacji rang Kendalla ogranicza się do postaci zwanej często współczynnikiem korelacji rang τ_a Kendalla:

$$\tau_a = \frac{2R}{\frac{n(n-1)}{2}} - 1.$$

Współczynnik konkordancji Kendalla i Babingtona-Smitha umożliwia porównanie zgodności rang nadanych dowolnej liczbie cech. Wyznacza się go według wzoru:

$$r_k = \frac{12 \cdot S_r}{m^2 \cdot n(n^2 - 1)}.$$

gdzie: S_r – suma kwadratów różnic między sumą rang każdej cechy i ich średnią wartością; m – liczba obiektów ocenianych; n – liczba cech.

Współczynnik ten przyjmuje wartości z przedziału $[0, 1]$ i im wyższą ma wartość, tym silniejsza jest zgodność opinii wyrażanych za pomocą rang.

Przykład 7.18

Tablica 7.17 przedstawia ocenę wyrażoną za pomocą rang podanych przez dział marketingu, dotyczącą lojalności odbiorców, wizerunku marek konkurencyjnych, jakości oraz poziomu obsługi klienta na pięciu rynkach, na których firma prowadzi swą działalność.

Tablica 7.17. Rangi nadane poszczególnym rynkom ze względu na wyróżnione cechy

Rynek	Ocena lojalności odbiorców w skali od 1 (ocena najniższa) do 5 (ocena najwyższa)	Ocena wizerunku marek konkurencyjnych w skali od 1 (ocena najniższa) do 5 (ocena najwyższa)	Ocena jakości produktów w skali od 1 (ocena najniższa) do 5 (ocena najwyższa)	Ocena poziomu obsługi klienta w skali od 1 (ocena najniższa) do 5 (ocena najwyższa)	Razem
A	1	4	4	4	13
B	5	1	3	2	11
C	3	3	5	3	14
D	2	5	2	5	14
E	4	2	1	1	8

Średnia wartość wszystkich rang ze względu na cztery cechy wynosi:

$$\bar{x} = \frac{1}{2} m(n+1) = \frac{1}{2} 5(4+1) = 12,5.$$

Suma kwadratów różnic między sumą rang każdej cechy i ich średnią wartością wynosi:

$$S_r = (13-12,5)^2 + (11-12,5)^2 + (14-12,5)^2 + (14-12,5)^2 + (8-12,5)^2 = 27,25.$$

Współczynnik konkordancji przyjmuje zatem wartość:

$$r_k = \frac{12 \cdot S_r}{m^2 \cdot n(n^2 - 1)} = \frac{12 \cdot 27,25}{5^2 \cdot 4 \cdot (4^2 - 1)} \approx 0,218.$$

Wynik oznacza względnie niską zgodność ocen poszczególnych rynków ze względu na wyróżnione cechy.

7.2.3.3. Współczynnik eta

Współczynnik *eta* mierzy zależność występującą między cechami opisywanymi przez dane przedziałowe lub ilorazowe (cecha zależna) i cechami opisywanymi przez dane nominalne lub porządkowe (cecha niezależna), wśród których można wyróżnić k kategorii. Definiuje go wzór:

$$\eta = \sqrt{\frac{\sum_{i=1}^k (\bar{x}_i - \bar{x})^2 \cdot n_i}{\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2}},$$

gdzie: x_{ij} – wartość cechy zależnej dla j -tej jednostki w i -tej grupie wyodrębnionej według wartości cechy niezależnej; $\bar{x}_i$ – wartość średnia cechy zależnej w i -tej grupie wyodrębnionej według wartości cechy niezależnej; n_i – liczba jednostek należących do i -tej grupy wyodrębnionej według cechy niezależnej; $\bar{x}$ – wartość średnia cechy zależnej.

Współczynnik ten przyjmuje wartości z przedziału $[0, 1]$ i im wyższe ma wartości, tym silniejsza jest zależność cechy przedziałowej lub ilorazowej od nominalnej lub porządkowej.

Przykład 7.19

W firmie oferującej na rynki lokalne soki owocowe w celu zwiększenia wielkości sprzedaży podjęto próbę ustalenia, czy kolor opakowania soku ma wpływ na wielkość sprzedaży. Zaobserwowane ilości sprzedanych opakowań według kolorów kształtowały się tak, jak w tablicy 7.18.

Tablica 7.18. Wielkość sprzedaży według koloru opakowania produktu

Kolor opakowania	Wielkość sprzedaży w poszczególnych dniach					
Zielony	18	22	22	23	22	23
Niebieski	20	18	19	21	20	20
Biały	10	11	12	11	11	10

Cechą zależną jest liczba sprzedanych opakowań, cechą niezależną zaś kolor opakowania. Wybrano 3 kolory, zatem występują 3 grupy produktów, których sprzedaż obserwowano. Oto niezbędne obliczenia w celu wyznaczenia współczynnika η :

$$\bar{x}_{\text{zielony}} = \frac{18 + 22 + 22 + 23 + 22 + 23 + 19}{7} \approx 21,28,$$

$$\bar{x}_{\text{niebieski}} = \frac{20 + 18 + 19 + 21 + 20 + 20 + 18}{7} \approx 19,43,$$

$$\bar{x}_{\text{biały}} = \frac{10 + 11 + 12 + 11 + 11 + 11 + 10}{7} \approx 10,86,$$

$$\bar{x} = \frac{18 + 22 + 22 + 23 + 22 + 23 + 19 + 20 + 18 + 19 + 21 + 20 + 20 + 18 + 10 + 11 + 12 + 11 + 11 + 11 + 10}{21} \approx 17,19,$$

$$\sum_{i=1}^k (\bar{x}_i - \bar{x})^2 \cdot n_i = (21,28 - 17,19)^2 \cdot 7 + (19,43 - 17,19)^2 \cdot 7 + (10,86 - 17,19)^2 \cdot 7 \approx 433,2425,$$

$$\begin{aligned} \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 &= (18 - 17,19)^2 + (22 - 17,19)^2 + (22 - 17,19)^2 + \\ &+ (23 - 17,19)^2 + (22 - 17,19)^2 + (23 - 17,19)^2 + (19 - 17,19)^2 + \\ &+ (20 - 17,19)^2 + (18 - 17,19)^2 + (19 - 17,19)^2 + (21 - 17,19)^2 + \\ &+ (20 - 17,19)^2 + (20 - 17,19)^2 + (18 - 17,19)^2 + (10 - 17,19)^2 + \\ &+ (11 - 17,19)^2 + (12 - 17,19)^2 + (11 - 17,19)^2 + (11 - 17,19)^2 + \\ &+ (11 - 17,19)^2 + (10 - 17,19)^2 = 467,2381. \end{aligned}$$

Wartość współczynnika η wynosi zatem:

$$\eta = \sqrt{\frac{\sum_{i=1}^k (\bar{x}_i - \bar{x})^2 \cdot n_i}{\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2}} = \sqrt{\frac{433,2425}{467,2381}} \approx 0,96.$$

Wynik wskazuje na bardzo silną zależność wielkości sprzedaży od koloru opakowania.

7.2.3.4. Współczynnik korelacji Pearsona

Współczynnik korelacji Pearsona mierzy zależność między dwiema cechami opisywanymi przez dane przedziałowe lub ilorazowe. Definiuje go wzór:

$$r(xy) = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}}{S(x) \cdot S(y)} = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}}{\sqrt{\left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2\right)} \sqrt{\left(\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2\right)}},$$

gdzie: x_i, y_i – wartości cech; i – numer obserwacji; n – liczba wszystkich obserwacji; $\bar{x}, \bar{y}$ – wartości średnie obu cech; $S(x), S(y)$ – wartości odchyłeń standardowych obu cech.

Współczynnik ten przyjmuje wartości z przedziału $[-1, 0, 1, 0]$, przy czym jego znak informuje o kierunku zależności, a wartość liczbową o jej sile. Współczynnik ten jest symetryczny, co oznacza, że jego wartość informuje o sile i kierunku zależności cechy X od Y oraz Y od X .

Przykład 7.20

Dla 10 punktów sprzedaży należących do badanej branży odnotowano następujące dane dotyczące powierzchni ekspozycyjnej (cecha X w m^2), liczby zatrudnionych sprzedawców (cecha Y) oraz miesięcznego przychodu ze sprzedaży (cecha Z w jednostkach pieniężnych).