

KLASYCZNE NARZĘDZIA ZARZĄDZANIA JAKOŚCIĄ DIAGRAM KORELACJI

dr inż. Bartosz SOLIŃSKI

**Akademia Górniczo-Hutnicza
Wydział Zarządzania
Katedra Zarządzania Przedsiębiorstwem**

Diagram korelacji w literaturze przedmiotu występuje także pod innymi nazwami m.in.: wykres rozrzutu, wykres korelacji, wykres zmiennych. Służy on do graficznego przedstawienia relacji między dwoma zmiennymi (cechami). Stanowi więc zbiór punktów na płaszczyźnie, odpowiadający zbiorowi par liczb (x_i, y_i) , gdzie x_i oznacza i -tą obserwację zmiennej niezależnej X , y_i oznacza i -tą obserwację zmiennej zależnej Y .

Na istnienie korelacji zmiennych (cech) wskazuje ich wzajemny związek, który oznacza, że zmienne wpływają na siebie. Diagram korelacji nie bada natomiast związku przyczynowo-skutkowego zachodzącego między zmiennymi, a mówi jedynie o tym, czy istnieje pewien związek korelacyjny między dwiema zmiennymi.

Diagram korelacji używany jest do:

- stwierdzenia istnienia zależności pomiędzy zmiennymi,
- stwierdzenia kierunku związku,
- pokazania siły związku.

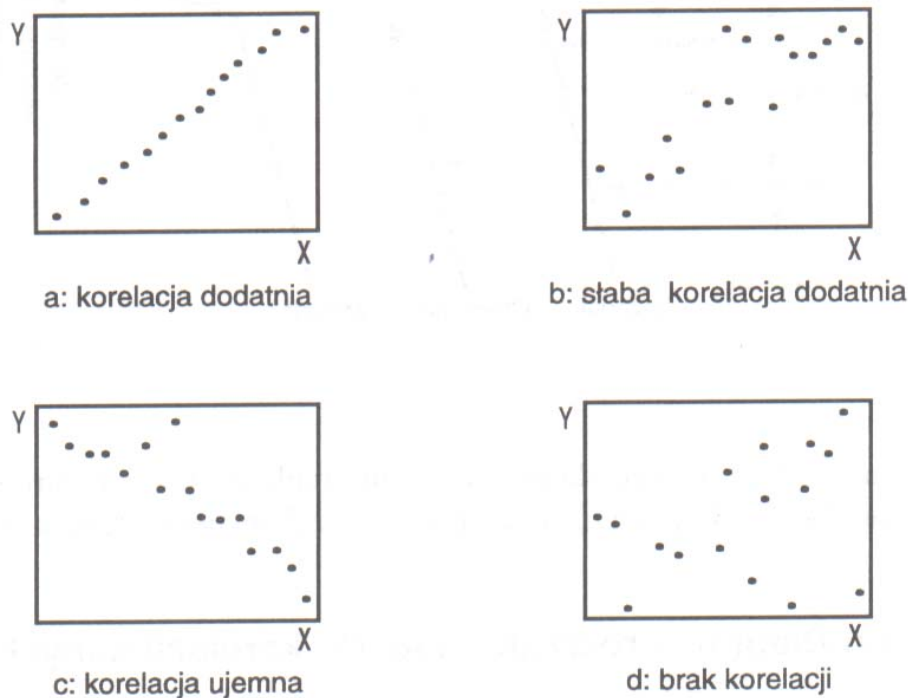
Procedura sporządzania diagramu korelacji jest następująca:

1. zebrać dwie pary danych, należących do dwóch zbiorów danych, których związek ma być analizowany,
2. na układ współrzędnych x i y nanieść pary danych,
3. zbadać kształt rozmieszczenia punktów w celu wykrycia rodzaju i siły zależności,
4. zbadać siłę zależności, obliczając współczynnik korelacji Pearsona (opcjonalnie),
5. zbadać istotność korelacji (opcjonalnie).

Analizę diagramu rozpoczyna się od badania charakteru zaobserwowanej zależności. Korelacja pomiędzy zmiennymi może mieć różnorodny charakter. O pozytywnej korelacji pomiędzy zmiennymi mówimy wówczas, gdy równocześnie wzrost zmiennej X powoduje wzrost zmiennej Y . Negatywna korelacja występuje wówczas, gdy wzrost jednej zmiennej X

powoduje jednoczesny spadek drugiej zmiennej Y , a jeżeli wzrost jednej zmiennej X nie powoduje zmian drugiej zmiennej Y , wówczas mamy do czynienia z brakiem korelacji pomiędzy zmiennymi.

Przykłady różnego charakteru zaobserwowanych korelacji przedstawiono na rysunku 1. W przypadku (a) można zaobserwować silną korelację, można więc przypuszczać że zmienna X jest głównym czynnikiem wpływającym na dany problem (zmienną Y). Taką samą tendencję wzrostową można zauważyć na rysunku (b), jednak w tym przypadku obserwuje się dość silny rozrzut wartości, może świadczyć to, że X nie jest głównym czynnikiem, ale wartości zmiennej Y zależą w dużym stopniu od tej zmiennej. Na kolejnych rysunkach (c) przedstawiono korelację silną ujemną, a na (d) brak korelacji. W tym ostatnim przypadku zmienne nie wywierają na siebie żadnego wpływu i pozostają bez związku.



Rys. 1. Przykładowe diagramy korelacji

W celu zbadania tej zależności można posłużyć się także współczynnikiem korelacji populacji ρ :

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

w którym:

ρ – współczynnik korelacji w populacji,

$\text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$ – kowariancja dwóch zmiennych X i Y .

Wartość korelacji w populacji często nie jest znana więc szacuje się ją na podstawie zmienności losowej par wyników i obserwacji zmiennych X i Y , otrzymując współczynnik korelacji z próby, oznaczany przez r – nazywany współczynnikiem korelacji *Pearsona*.

Ten współczynnik często można używać jako punktową ocenę współczynnika korelacji populacji i najczęściej jest to wystarczające założenie przy analizie wpływu na siebie dwóch zmiennych. Współczynnik r przyjmuje wartości z zakresu $\langle -1,1 \rangle$, a o sile korelacji świadczy $|r|$, która może przyjąć wartość $\langle 0,1 \rangle$. Wyznaczenie współczynnika korelacji r , pozwala na interpretację siły zależności pomiędzy zmiennymi (tabela 1).

Tabela 1. Siła zależności pomiędzy zmiennymi

| Wartość współczynnika r | Znaczenie |
|-------------------------------|--------------------------|
| $\langle 0,7 ; 1,0 \rangle$ | silna korelacja dodatnia |
| $\langle 0,3 ; 0,7 \rangle$ | słaba korelacja dodatnia |
| $\langle -0,3 ; 0,3 \rangle$ | brak korelacji |
| $\langle -0,7 ; -0,3 \rangle$ | słaba korelacja ujemna |
| $\langle -0,7 ; -1,0 \rangle$ | silna korelacja ujemna |

Dodatkowo można także dokonać sprawdzenia istotności współczynnika korelacji ρ dla populacji na podstawie współczynnika r . W tym celu przeprowadza się test sprawdzający, czy zmienne są ze sobą skorelowane, stawiając hipotezy:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

i wykorzystując jako sprawdzian wartość statystyki t według wzoru:

$$t_{(n-2)} = \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}}$$

w którym:

n – liczność par zmiennych X i Y .

Następnie porównuje się otrzymaną wartość z wartością krytyczną odczytaną z tablic testu t -Studenta dla $n-2$ stopni swobody i przyjętego poziomu istotności α (Uwaga: odczytując z tablic t -Studenta, wartość testu dla $\alpha/2$). W przypadku gdy:

$$t_{obliczeniowe} > t_{tablicowe}$$

stwierdza się istotność współczynnika korelacji liniowej, na poziomie istotności α .