KORELACJA I REGRESJA WIELU ZMIENNYCH

- korelacja i regresja wieloraka
- korelacja i regresja cząstkowa

WSPÓŁCZYNNIK KORELACJI CZĄSTKOWEJ

$$r_{ij \cdot kl \ldots z} = \frac{-P_{ij}}{\sqrt{P_{ii} P_{jj}}}$$ subskrypty główne, następcze

P - macierz współczynników korelacji zmiennych objaśniających

P_{ij} - dopełnienie algebraiczne macierzy *P* powstałe w wyniku skreślenia *i-tego* wiersza i *j-tej* kolumny

$$P = \begin{bmatrix} 1 & r_{12} & r_{13} & \ldots & r_{1z} \\ r_{21} & 1 & r_{23} & \ldots & r_{2z} \\ r_{31} & r_{32} & 1 & \ldots & r_{3z} \\ r_{z1} & r_{z2} & r_{z3} & \ldots & 1 \end{bmatrix}$$

na głównej przekątnej $r_{ii}=1$
macierz symetryczna $r_{zk}=r_{kz}$

- Współczynnik korelacji cząstkowej informuje, jaka jest zależność między zmiennymi określonymi przez subskrypty główne z wyłączeniem oddziaływania wszystkich pozostałych zmiennych;

- Współczynnik korelacji cząstkowej może być mniejszy lub większy od współczynników korelacji prostej danej pary cech, może także wystąpić zmiana znaku;

- W przypadku analizy trzech zmiennych macierz *P* ma następującą postać:

$$P = \begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{bmatrix}$$

współczynniki korelacji cząstkowej określić można następująco:

$$r_{12\cdot 3} = \frac{-P_{12}}{\sqrt{P_{11}P_{22}}} = \frac{-\left[-\begin{vmatrix} r_{21} & r_{23} \\ r_{31} & 1 \end{vmatrix}\right]}{\sqrt{\begin{vmatrix} 1 & r_{23} \\ r_{32} & 1 \end{vmatrix} \cdot \begin{vmatrix} 1 & r_{13} \\ r_{31} & 1 \end{vmatrix}}} = \frac{r_{12} - r_{23}r_{13}}{\sqrt{(1-r_{23}^2)(1-r_{13}^2)}}$$

WSPÓŁCZYNNIK KORELACJI WIELORAKIEJ

- Zależność jednej zmiennej od zespołu innych cech

R_w lub $R_{1.234...k}$

pierwszy subskrypt oznacza zmienną objaśnianą, a pozostałe to zmienne objaśniające

$$R_{1\cdot 234...k} = \sqrt{1 - \frac{\det P}{\det R}}$$

gdzie: P - macierz współczynników korelacji zmiennych objaśniających i zmiennej objaśnianej

R - macierz współczynników korelacji pomiędzy zmiennymi objaśniającymi

$$P = \begin{bmatrix} 1 & r_{12} & r_{13} \cdots & r_{1k} & r_{1\,k+1} \\ r_{21} & 1 & r_{23} \cdots & r_{2k} & r_{2\,k+1} \\ r_{31} & r_{32} & 1 \cdots & r_{3k} & r_{3\,k+1} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{k+1\,1} & r_{k+1\,2} & r_{k+1\,3} \cdots & r_{k+1\,k} & 1 \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & r_{23} & \cdots & r_{2k} \\ r_{32} & 1 & \cdots & r_{3k} \\ \cdots & \cdots & \cdots & \cdots \\ r_{k2} & r_{k3} & \cdots & 1 \end{bmatrix}$$

- Współczynnik korelacji wielorakiej R_w przyjmuje wartości:
  $0 < R \leq 1$

- Współczynnik korelacji wielorakiej R_w określa jedynie natężenie korelacji liniowej ale nie wskazuje jej kierunku

- Wyznacznik macierzy R jest miara stopnia skorelowania zmiennych objaśniających $X_1, X_2, \ldots , X_k$

O silnej korelacji świadczy bliska zeru wartość det **R**.

W przypadku zmiennych ortogonalnych (niezależnych), wyznacznik ten osiąga swoją górną granicę równą jedności

$$\det R = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 \end{bmatrix} = 1$$

- R_w^2 - współczynnik determinacji - informuje w jakim stopniu (%) zmienność zmiennej zależnej Y jest wyjaśniona przez zmienność zespołu zmiennych objaśniających X_i

- Współczynnik korelacji wielorakiej może przyjmować wysoką wartość wywołaną nie tylko silna korelacja między zmiennymi Y i $X_1, X_2, \ldots , X_k$ ale także z powodu silnego skorelowania zmiennych objaśniających

- W analizie wielowymiarowej stosuje się często korektę współczynnika korelacji wielokrotnej, gdyż zwiększanie liczby zmiennych objaśniających nadmiernie zwiększa wartość R_w

$$_k R_w^2 = 1 - (1 - R_w^2)\left(\frac{n-1}{n-m}\right)$$

n – liczba obserwacji
m – liczba szacowanych parametrów łącznie z wyrazem wolnym

- **W przypadku trzech zmiennych *R_w*** obliczyć można według poniższego wzoru:

$$R_{1\cdot 23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}} \qquad R_{1\cdot 23} = \sqrt{1 - (1 - r_{12}^2)(1 - r_{13\cdot 2}^2)}$$