Fundusze Europejskie
dla Rozwoju Społecznego

Dofinansowane przez
Unię Europejską

# Metody analizy danych z programem Statistica

# STATYSTYKI OPISOWE

**Statystyka**, to nauka, której przedmiotem zainteresowań są metody badań i analizy zjawisk występujących w przyrodzie, technice itp.

Oparta jest na teorii rachunku prawdopodobieństwa.

Została zapoczątkowana w drugiej połowie XVII wieku przez francuskich matematyków: B. Pascala (1623 -1662) i P. Fermata (1601-1665).

Ogół badanych przypadków (zjawisk) określonego typu (zbiorowość) nazywamy **populacją generalną**.

Populacja musi być dobrze określona i nie musi dotyczyć zbiorowości ludzkiej.
Przykład: Wszyscy pracownicy UPP.

Ponieważ często nie mamy możliwości badania całej populacji, dlatego analizę przeprowadzamy na podstawie **próby** wylosowanej z całej populacji.

Próbą nazywamy podzbiór populacji generalnej.

Jednostka doświadczalna, to obiekt, dla którego wykonuje się analizę statystyczną.

Przykład: pracownik, pojedyncza roślina, poletko doświadczalne itd.

Wyniki pomiarów badanej cechy, uzyskane dla pobranej próby, niewiele mówią o samej próbie ani tym bardziej o populacji.

# WNIOSKOWANIE STATYSTYCZNE

Są dwie podstawowe metody wnioskowania statystycznego:

**Estymacja** jest to szacowanie wartości parametrów na podstawie próby.

**Testowanie (weryfikacja) hipotez** – sprawdzenie słuszności przypuszczeń dotyczących postaci rozkładu cechy lub wartości jego parametrów.

Do opisu badanego zjawiska wykorzystujemy pewne parametry takie jak: średnia, maksimum, minimum, odchylenie standardowe itd.

W statystyce takie parametry, uzyskane na podstawie próby, nazywa się ESTYMATORAMI. Parametry te często określa się jako: Statystyki opisowe.

**Definicja**

**Estymatorem** nazywamy funkcję, za pomocą której szacujemy prawdziwą wartość nieznanego parametru populacji.

# ESTYMACJA PARAMETRYCZNA

A. Estymacja punktowa

Polega na tym, że wartości estymatorów uzyskane na podstawie jednej próby są traktowane jako oszacowania parametrów populacji (np. $\mu \approx \bar{x}$ ).

B. Estymacja przedziałowa

Polega na wyznaczeniu takiego przedziału [L, U], którego końce i długość są zmiennymi losowymi i który z określonym prawdopodobieństwem 1 - $\alpha$ zawiera szacowany parametr populacji $\theta$.

Przedział taki nazywamy przedziałem ufności i zapisujemy go

$$P(L \leq \theta \leq U) \approx 1 - \alpha$$

Prawdopodobieństwo 1 - $\alpha$ jest nazywane współczynnikiem ufności.
W badaniach rolniczych 1 - $\alpha$ jest najczęściej równe 0,90; 0,95; 0,99.

Interpretacja współczynnika ufności

Współczynnik $1 - \alpha$ oznacza, że na 100 pobranych prób i wyliczonych na ich podstawie 100 przedziałów ufności, tylko $\alpha$ przedziałów nie zawiera parametru $\theta$.

# Liczbowe charakterystyki

## 1. Miary położenia

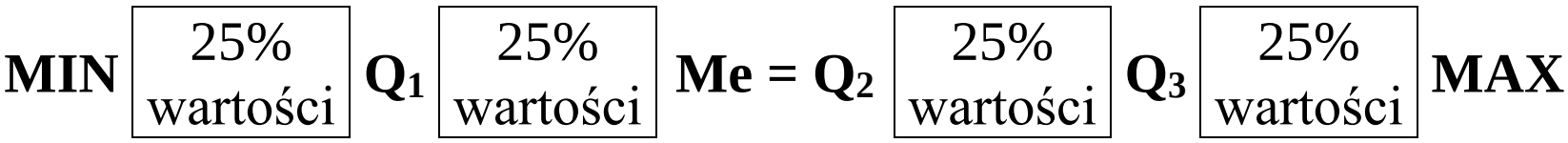
a) średnia arytmetyczna ($\bar{x}$)

b) mediana Me (kwartyl drugi, wartość środkowa, $Q_2$) - jej wartość zależy od n:

- dla n nieparzystego mediana znajduje się na miejscu {(n + 1)/2},
- dla n parzystego mediana równa się średniej liczb z miejsc {n/2} oraz {n/2+1};

c) kwartyl dolny $Q_1$ (kwartyl pierwszy);

d) kwartyl górny $Q_3$ (kwartyl trzeci);




e) wartość modalna (moda, dominanta) (Mo)

## 2. Miary zmienności (rozproszenia)

a) Rozstęp (R)
b) wariancja z próby ($s^2$)
c) odchylenie standardowe (s)
d) współczynnik zmienności (w)

**Przykład:**

Nauczyciel języka angielskiego przeprowadził wśród uczniów swojej klasy test rozumienia ze słuchu oraz test gramatyczny. W teście rozumienia ze słuchu można było uzyskać wynik od 0 do 40, w teście gramatycznym wynik o 0 do 100. Badana grupa uczniów uzyskała w teście rozumienia ze słuchu średni wynik 20pkt, odchylenie standardowe wyniosło 8pkt. W przypadku testu gramatycznego, średni wynik wyniósł 50, odchylenie standardowe wyniosło 12pkt.

Wyliczając współczynniki zmienności, odpowiednio: (8/20)*100% = 40% oraz (12/50)*100% = 24% nauczyciel mógł stwierdzić, że w przypadku rozumienia ze słuchu wśród uczniów zachodzi większa zmienność, uczniowie są bardziej zróżnicowani pod względem tej umiejętności niż w przypadku testu gramatycznego. Dzięki tej informacji, nauczyciel może spróbować zmiejszyć różnice w tej umiejętności pomiędzy uczniami poprzez wprowadzenie odpowiednich ćwiczeń. Jako, że testy miały inną skalę, to posługując się współczynnikiem zmienności można było porównać dla tych dwóch testów zmienność w badanej grupie.

$$w = \frac{S}{\overline{x}} \cdot 100\%$$

**Współczynnik skośności (skośność) (miara asymetrii)** określa w jaki sposób rozmieszczone są dane względem średniej. Gdy jest ujemna to mówimy, że rozkład empiryczny badanej cechy jest lewostronnie skośny, czyli większość danych ma wartości większe od średniej arytmetycznej ($\bar{x} < m_e < m_o$). Gdy jest dodatnia to mówimy, że rozkład badanej cechy jest prawostronnie skośny, czyli większość danych ma wartości mniejsze od średniej arytmetycznej ($\bar{x} > m_e > m_o$). Gdy jest równy zero, to mówimy, że rozkład badanej cechy jest symetryczny względem średniej arytmetycznej ($\bar{x} = m_e = m_o$).

**Współczynnik spłaszczenia (kurtoza) (miara koncentracji)** bada „spiczastość" wykresu rozkładu badanej cechy względem wykresu rozkładu normalnego. Gdy jest ujemna to wykres rozkładu badanej cechy leży poniżej wykresu rozkładu normalnego, czyli rozkład jest bardziej spłaszczony od normalnego. Gdy jest dodatnia to wykres rozkładu badanej cechy leży powyżej wykresu rozkładu normalnego, czyli rozkład jest bardziej wysmukły niż normalny. Wartość kurtozy dla rozkładu normalnego wynosi zero.

# PRZEGLĄD TESTÓW DLA JEDNEJ I DWÓCH POPULACJI

# TESTOWANIE HIPOTEZ

- Hipoteza parametryczna – przypuszczenie dotyczące wartości parametrów populacji.
- Testowanie hipotezy – reguła postępowania prowadząca do zweryfikowania danej hipotezy.
- Weryfikacja hipotezy – na podstawie próby.
- W testowaniu hipotez, w każdej sytuacji, biorą udział dwie hipotezy, które się wykluczają. Oznaczone są jako:

| | |
|---|---|
| $H_0$ | *hipoteza zerowa* |
| $H_1$ | *hipoteza alternatywna* |

Zawsze jedna z nich jest prawdziwa, a druga fałszywa. NIE WIEMY, KTÓRA Z NICH !!!

NIE PODEJMUJE SIĘ NIGDY DECYZJI O PRZYJĘCIU HIPOTEZY ZEROWEJ $H_0$ !!

dlaczego?

Aby uchronić się przed popełnieniem jednego z dwóch błędów – I lub II rodzaju.

- Błąd I rodzaju – polega na odrzuceniu hipotezy $H_0$, gdy faktycznie jest ona prawdziwa.
- Błąd II rodzaju – polega na przyjęciu hipotezy $H_0$, gdy faktycznie jest ona fałszywa.
- W teście kontroluje się jedynie prawdopodobieństwo popełnienia błędu I rodzaju, które nazywane jest poziomem istotności i oznaczane jako $\alpha$ (alfa).

***Nie uwzględnia się konsekwencji popełnienia błędu II rodzaju (stąd nie przyjmujemy $H_0$).***

# Rozkład normalny (rozkład Gaussa)

(*Najważniejszy w przyrodzie rozkład)*

**Definicja**

Zmienna losowa *X* typu ciągłego ma rozkład normalny o parametrach $\mu$ oraz $\sigma$, jeżeli funkcja gęstości f(*x*) jest postaci:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \qquad x, \mu \in \Re \qquad \sigma \in \Re^+$$

Symbolicznie rozkład normalny oznaczamy:

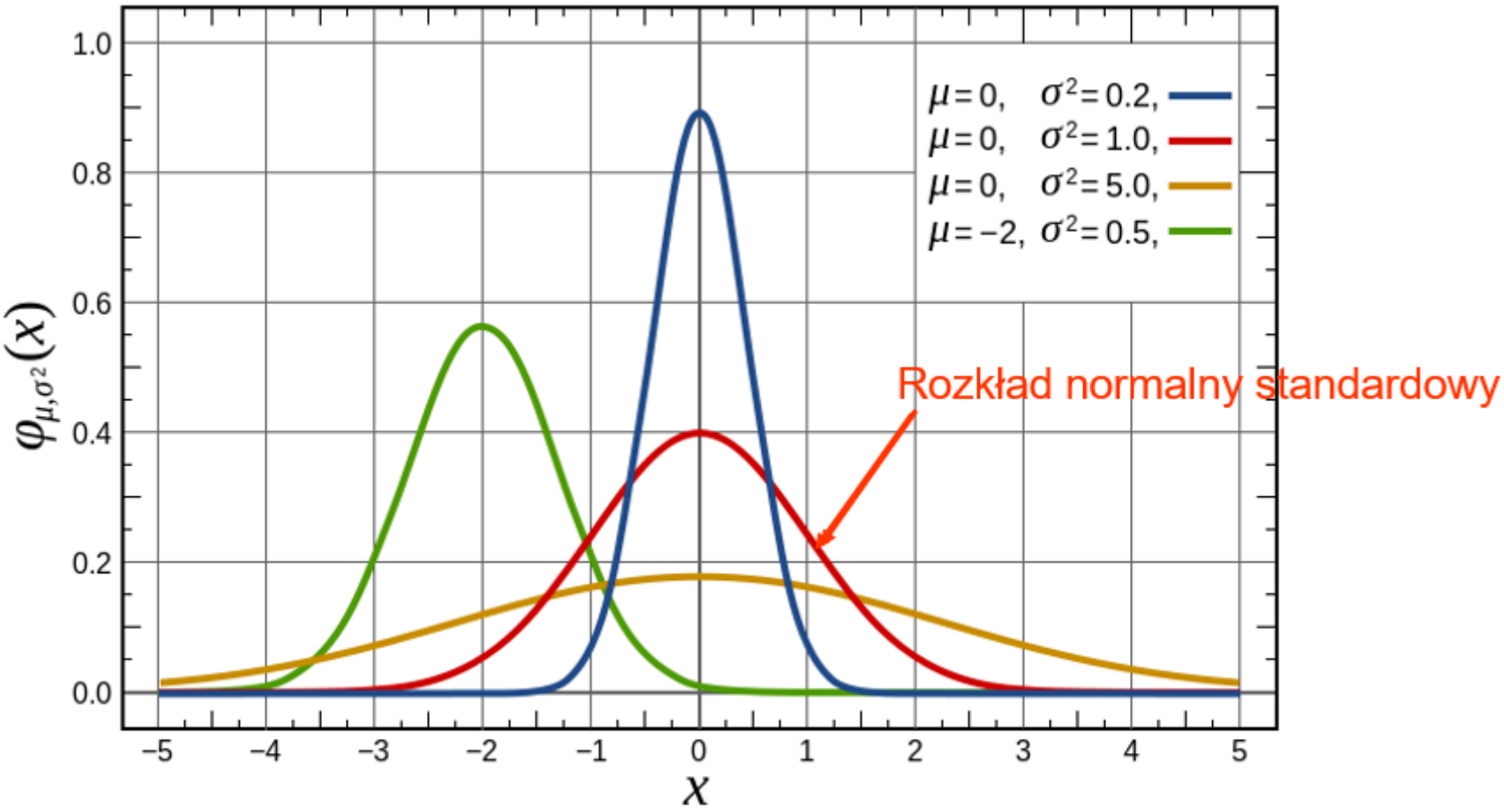
$$X \sim N(\mu, \sigma) \qquad \text{lub} \qquad X \sim N(\mu, \sigma^2)$$

**Parametry rozkładu:** $E(X) = \mu$ $D^2(X) = \sigma^2$ $D(X) = \sigma$

**Wykresem funkcji gęstości jest krzywa symetryczna względem prostej $x = \mu$ i jest nazywana**

**<u>krzywą normalną (krzywą Gaussa)</u>**

[F. Gauss XVIII / XIX]




Wartość μ decyduje o przesunięciu krzywej względem osi rzędnych.

Pole ograniczone krzywą i osią odciętych jest równe 1.

# Test hipotezy dla sredniej $\mu$, gdy wariancja populacji $\sigma^2$ jest nieznana

Założenia: a) populacja ma rozkład $N(\mu, \sigma)$;
b) parametry $\mu$ i $\sigma$ są nieznane.
c) Pobieramy próbę o liczebności n.

Wybór hipotezy zerowej i alternatywnej:

| A) | B) | C) |
|---|---|---|
| $H_0: \mu = \mu_0$<br>$H_1: \mu \neq \mu_0$ | $H_0: \mu \geq \mu_0$<br>$H_1: \mu < \mu_0$ | $H_0: \mu \leq \mu_0$<br>$H_1: \mu > \mu_0$ |
| $H_1$ - dwustronna | $H_1$ – lewostronna | $H_1$ – prawostronna |

## Porównanie dwóch populacji pod względem tej samej cechy X

Niech $x_{11}, x_{12}, \ldots x_{1n_1}$ będą obserwacjami cechy w pierwszej populacji ($X_1$), natomiast $x_{21}, x_{22}, \ldots x_{2n_2}$ będą obserwacjami tej cechy w drugiej populacji ($X_2$).

Jeśli cecha X jest cechą ciągłą to, zakładamy, że w obu populacjach analizowana cecha ma rozkład normalny:

$$X_1 \sim N(\mu_1, \sigma_1^2), \quad X_2 \sim N(\mu_2, \sigma_2^2).$$

**Uwaga:**

W celu porównania dwóch średnich stosujemy następujące testy statystyczne:

1. **Test t-Studenta** – wykorzystujemy go, gdy niezależne próby pochodzą z dwóch populacji o rozkładach normalnych i jednocześnie spełnione jest założenie jednorodności wariancji.
2. **Test t-Studenta z poprawką Cochrana-Coxa** – stosujemy go, gdy niezależne próby pochodzą z dwóch populacji o rozkładach normalnych, ale wariancje tych populacji nie są jednorodne.
3. **Test Mediany** – stosujemy go, gdy niezależne próby pochodzą z populacji, które nie mają rozkładu normalnego.

Wybór hipotezy zerowej i alternatywnej:

| A) | B) | C) |
|---|---|---|
| $H_0: \mu_1 = \mu_2$<br>$H_1: \mu_1 \neq \mu_2$ | $H_0: \mu_1 \geq \mu_2$<br>$H_1: \mu_1 < \mu_2$ | $H_0: \mu_1 \leq \mu_2$<br>$H_1: \mu_1 > \mu_2$ |
| $H_1$ - dwustronna | $H_1$ – lewostronna | $H_1$ – prawostronna |

# KORELACJA I REGRESJA LINIOWA

Przy badaniu populacji, równocześnie ze względu na dwie lub więcej cech mierzalnych, interesuje nas związek (zależność) jaki zachodzi między tymi cechami oraz charakter tego związku.

Przy badaniu zależności (lub współzależności) między cechami posługujemy się pojęciami regresji i korelacji, przy czym korelacja zajmuje się siłą zależności (współzależności), a regresja kształtem zależności.

# Współczynnik korelacji liniowej

1) Zakładamy, że interesują nas dwie cechy (zmienne) X i Y w populacji dwucechowej.

2) Gdy współzależność między zmiennymi ma charakter liniowy, to miarą współzależności (w populacji) jest współczynnik korelacji $\rho$ (ro):

$$\rho = \frac{\mathrm{Cov}(X,\ Y)}{\sigma_x \cdot \sigma_y}$$

gdzie:

Cov(X, Y) oznacza kowariancję (w populacji),

$\sigma_x$ - odchylenie standardowe zmiennej X,

$\sigma_y$ - odchylenie standardowe zmiennej Y.

Estymatorem współczynnika korelacji ρ (w populacji) jest współczynnik korelacji liniowej z próby, który oznaczamy literą r.

Obliczamy go na podstawie n par ($x_i$, $y_i$) wyników próby dwuwymiarowej

| | | | | |
|---|---|---|---|---|
| $x_1$ | $x_2$ | $x_3$ | … | $x_n$ |
| $y_1$ | $y_2$ | $y_3$ | … | $y_n$ |

według wzoru

$$r = \frac{S_{xy}}{S_x \cdot S_y} \qquad r \in [-1 ; 1]$$

gdzie

$S_x = \sqrt{S_x^2}$ i $S_y = \sqrt{S_y^2}$ - odchylenia standardowe

$S_{xy}$ - kowariancja z próby postaci

$$S_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$S_{xy} = \frac{1}{n-1}[\sum_{i=1}^{n} x_i y_i - n \cdot \bar{x} \cdot \bar{y}]$$

$$S_{xy} = \frac{1}{n-1}[\sum_{i=1}^{n} x_i y_i - \frac{1}{n}(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)]$$

Stawiamy hipotezy:

$H_0: \rho = 0,$

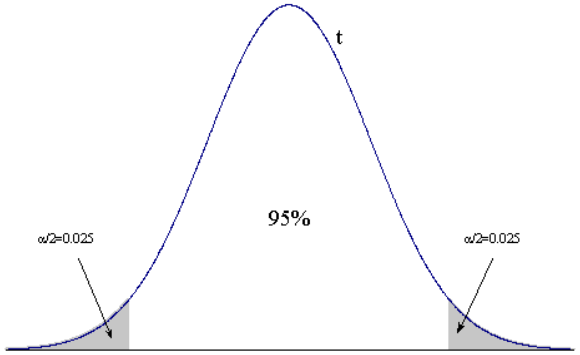
$H_1: \rho \neq 0,$

przyjmujemy poziom istotności $\alpha$ i weryfikujemy hipotezę $H_0$. Wyznaczamy wartość statystyki:

$$t = r \cdot \sqrt{\frac{n-2}{1-r^2}}$$

o rozkładzie t-studenta z $(n-2)$ stopniami swobody oraz wyznaczmy obszar krytyczny:

$$K = (-\infty,\ -t_{\alpha/2;n-2}) \cup (t_{\alpha/2;n-2},\ \infty).$$



## Właściwości współczynnika korelacji r

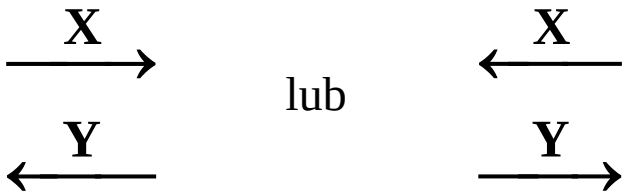
1) $-1 \leq r \leq 1$

2) $r > 0$ korelacja dodatnia;

*Gdy wzrasta (lub maleje) wartość jednej cechy i jednocześnie wzrasta (lub maleje) wartość drugiej cechy.*

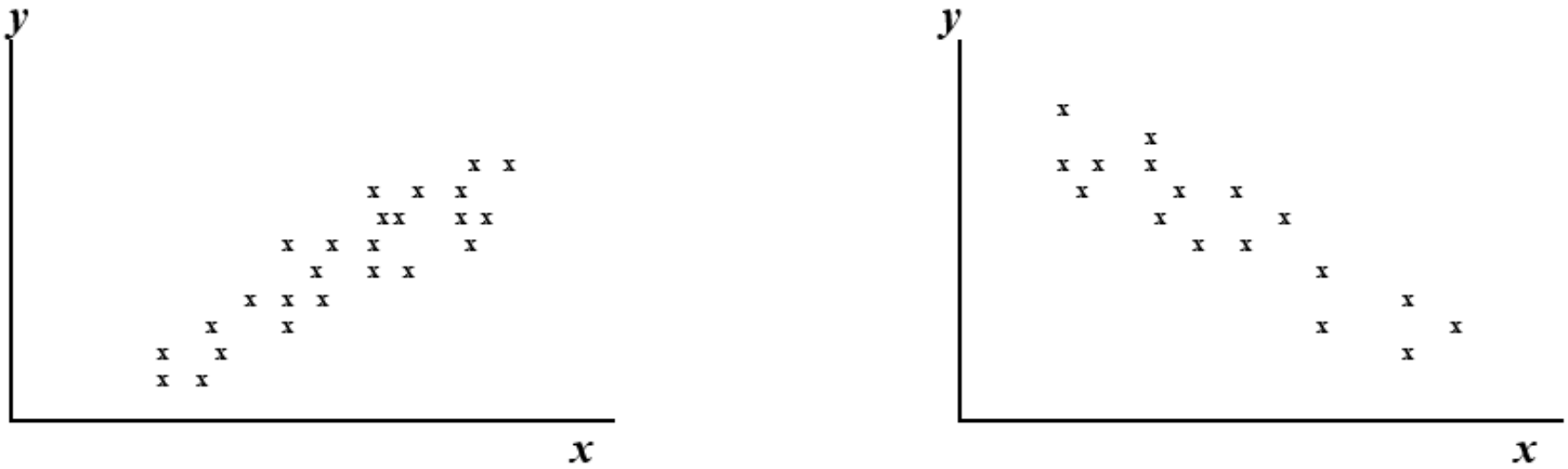
3) $r < 0$, korelacja ujemna;

*Gdy wzrostowi wartości jednej zmiennej towarzyszy zmniejszanie się wartości drugiej zmiennej.*

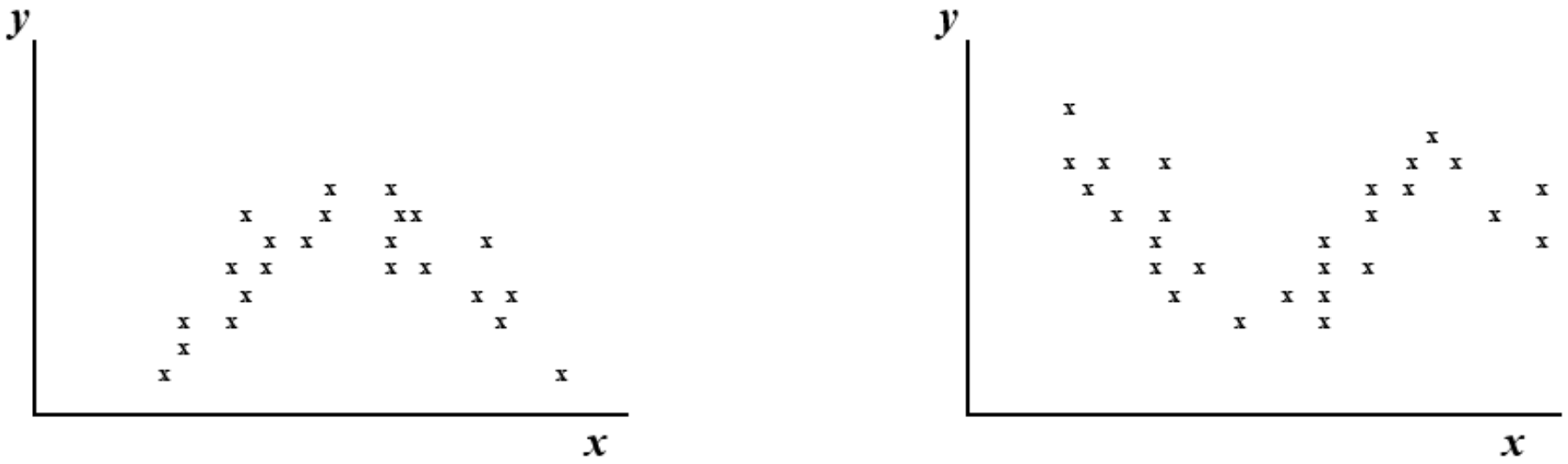



4) Gdy $r = -1$ lub $r = 1$, to wtedy między zmiennymi X i Y istnieje ścisła zależność liniowa.
5) Gdy $r = 0$, to jest brak korelacji.

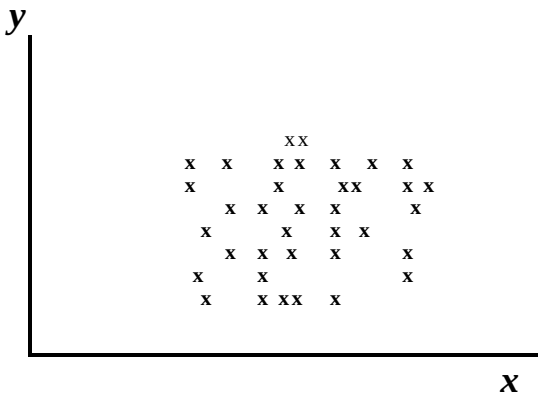
Najprostszym sprawdzianem zależności jest wykres rozrzutu (diagram korelacyjny). Jest to zbiór punktów ($x_i$ , $y_i$) płaszczyzny.




**r > 0** ***zależność liniowa*** **r < 0**

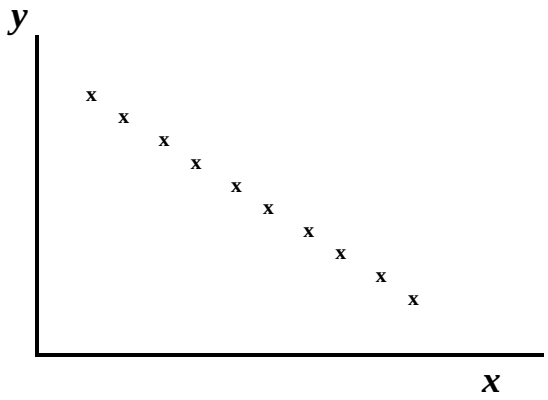



***zależność krzywoliniowa***




**r = 0**

***brak zależności***




**r = –1**

***zależność ścisła***

# Regresja liniowa

Na kształtowanie się, zmiennej zależnej Y ma wpływ nie tylko zmienna niezależna X ale również czynniki zakłócające, nazywane składnikiem losowym. Zależność statystyczną zapisujemy wówczas jako

$$y = f(x) + \varepsilon,$$

gdzie ε - składnik losowy.

Równanie regresji (model regresji) - równanie opisujące związek między cechami uwzględniające obecność składnika losowego.

X - niezależna - zmienna (cecha) objaśniająca
Y - zależna - zmienna (cecha) objaśniana

Wyznaczamy wartość oczekiwaną:

$$E(y) = E[f(x) + \varepsilon].$$

Ponieważ zmienna x jest niezależna od ε, to

$$E(y) = E[f(x)] + E(\varepsilon).$$

Ponadto wartość oczekiwana składnika losowego wynosi zero.

Stąd

$$E(y) = \beta_0 + \beta_1 x$$

$\beta_0$ - wyraz wolny regresji,
$\beta_1$ - współczynnik regresji.

Parametry $\beta_0$ i $\beta_1$ wyznaczamy tak, aby był spełniony warunek

$$\sigma^2 = E[Y - (\beta_0 + \beta_1 X)]^2 = \min$$

gdzie $\sigma^2$ wyraża wariancję błędu, który nie może być wyjaśniony regresją cechy Y względem X.

W praktyce, zależność między X i Y polega na oszacowaniu na podstawie próby nieznanych współczynników regresji i nieznanej wariancji błędu.
Parametry regresji najczęściej szacuje się za pomocą metody najmniejszych kwadratów.
Metoda najmniejszych kwadratów polega na minimalizacji sumy kwadratów odchyleń postaci:

$$SS = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 .$$

Metoda ta prowadzi do układu równań normalnych, z których uzyskujemy estymatory współczynników regresji $\beta_0$ i $\beta_1$ postaci:

$$b_1 = \frac{s_{xy}}{s_x^2}; \qquad b_0 = \bar{y} - b_1\bar{x}$$

Wtedy równanie regresji liniowej z próby jest

$$y = b_0 + b_1 x.$$

Interpretacja przyrodnicza współczynnika regresji $b_1$

Współczynnik regresji $b_1$ określa o ile przeciętnie wzrośnie (jeśli jest dodatni) lub zmaleje (jeśli jest ujemny) wartość cechy Y jeśli wartość cechy X wzrośnie o jednostkę.

## Współczynnik determinacji $R^2$

Oblicza się ze wzoru

$$R^2 = r^2 \ (100\%)$$

## Interpretacja współczynnika $R^2$

- określa stopień dopasowania modelu regresji do danych,
- podaje w ilu procentach zmienność cechy Y została wyjaśniona zmiennością cechy X poprzez przyjęte równanie regresji.

## Prognoza

Na podstawie uzyskanego równania regresji możemy prognozować oczekiwaną wartość cechy Y dla danej wartości cechy X.

**Uwaga**

Nie wolno wyznaczać prognozy dla Y poza obszarem zmienności cechy X, dla której równanie zostało wyznaczone, $x \in [x_{min}, x_{max}]$.

Wartość prognozowaną (regresyjną) obliczamy:

$$\hat{y}_i = b_0 + b_1 x_i$$

# KORELACJA CECH JAKOŚCIOWYCH

# Test niezależności chi-kwadrat

Test niezależności cech niemierzalnych (jakościowych) przeprowadza się w następujący sposób:

**Formułujemy hipotezy:**

$H_0$ : cechy X i Y są niezależne

$H_1$ : cechy X i Y są zależne

Rozkład wartości dwóch cech podaje się za pomocą tabeli zwanej *tablicą wielodzielczą* lub *korelacyjną*.

*Tablica wielodzielcza (korelacyjna)*

| $X \setminus Y$ | $Y = y_1$ | $Y = y_2$ | $Y = y_3$ | ... | $Y = y_k$ |
|---|---|---|---|---|---|
| $X = x_1$ | $n_{11}$ | $n_{12}$ | $n_{13}$ | ... | $n_{1k}$ |
| $X = x_2$ | $n_{21}$ | $n_{22}$ | $n_{23}$ | ... | $n_{2k}$ |
| $X = x_3$ | $n_{31}$ | $n_{32}$ | $n_{33}$ | ... | $n_{3k}$ |
| ... | ... | ... | ... | ... | ... |
| $X = x_w$ | $n_{w1}$ | $n_{w2}$ | $n_{w3}$ | ... | $n_{wk}$ |

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

$$\hat{n}_{ij} = \frac{suma\ i - tego\ wiersza \times suma\ j - tej\ kolumny}{n}$$

gdzie: n - liczba par obserwacji,

w - liczba wierszy w tabeli wielodzielczej,

k - liczba kolumn w tablicy wielodzielczej,

$\chi^2$ - statystyka chi - kwadrat

Q = $(\chi^2_{(w-1)\times(k-1);\,1-\alpha}; +\infty)$

## Miary zależności pomiędzy cechami niemierzalnymi (jakościowymi)

a) współczynnik kontyngencji Pearsona $C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \in \left\langle 0, \frac{\sqrt{\frac{w-1}{w}} + \sqrt{\frac{k-1}{k}}}{2} \right\rangle$

b) współczynnik Cramera $V = \sqrt{\frac{\chi^2}{n \times \min(w-1;\ k-1)}} \in \langle 0,\ 1 \rangle$

c) współczynnik Czuprowa $T = \sqrt{\frac{\chi^2}{n \times \sqrt{(w-1) \times (k-1)}}} \in \langle 0,\ 1 \rangle$

d) współczynnik Yule'a $\Phi = \sqrt{\frac{\chi^2}{n}} \in \langle 0,\ 1 \rangle$ =

*Opis*: Między *cecha X* a *cecha Y* zachodzi skorelowanie o sile $\{C, V, T_{xy}, \Phi\}$%.

# ANALIZA WARIANCJI DOŚWIADCZEŃ JEDNOCZYNNIKOWYCH

# *Podstawowe pojęcia*

- **Materiał badawczy** - Substancja lub jakikolwiek materiał, organiczny czy nieorganiczny, w którym zachodzi proces lub zjawisko będące przedmiotem badań.

- **Jednostka badawcza** - Naturalnie lub umownie wyróżniona część materiału badawczego, która w całości jest jednakowo traktowana w trakcie procesu i względem której prowadzone są indywidualne obserwacje w czasie trwania i przy zakończeniu procesu produkcji.

- **Czynnik badawczy** - Czynnik, którego działaniu poddany jest odpowiedni materiał badawczy w celu obserwowania wpływu tego czynnika na zjawisko lub proces będący przedmiotem badania.

## ANOVA dla klasyfikacji pojedynczej

## 1. UKŁAD CALKOWICIE LOSOWY

W klasyfikacji pojedynczej bada się wpływ tylko jednego czynnika (kontrolowanego na *a* poziomach lub wariantach) na wyniki doświadczenia (klasyfikacja jednoczynnikowa).

Poziomy (lub warianty) czynnika - obiekty doświadczalne.

Powtórzenia obiektów – $n_1, n_2, ..., n_a$ - replikacje (mogą być różne).

$n_1 + n_2 + ... + n_a = n$ liczba obserwacji w doświadczeniu.

Najprostszym układem doświadczalnym jest układ całkowicie losowy.

Założenie: jednostki doświadczalne – jednorodne.

## 2. UKŁAD BLOKÓW LOSOWANYCH KOMPLETNYCH

Układ ten stosujemy, gdy podejrzewamy, że może występować systematyczna zmienność między powtórzeniami np.:

- zmienność glebowa w przypadku doświadczeń polowych,
- zmienność spowodowana wykonywaniem pomiarów lub analiz przez więcej niż jedną osobę lub więcej niż jedno urządzenia,
- zmienność między danymi pochodzącymi z dwóch lub więcej regionów.

Przykładowe schematy doświadczenia jednoczynnikowego z czterema poziomam czynnika (4 obiekty: A1, A2, A3, A4) oraz trzema powtórzeniami

Układ całkowicie losowy

| | | |
|---|---|---|
| A1 | A2 | A1 |
| A2 | A3 | A4 |
| A3 | A1 | A3 |
| A4 | A2 | A4 |

Obiekty rozlosowane na obszarze całego doświadczenia w sposób losowy

Układ losowanych bloków

| | | |
|---|---|---|
| A1 | A2 | A3 |
| A3 | A4 | A2 |
| A4 | A1 | A4 |
| A2 | A3 | A1 |

**blok 1** **blok 2** **blok 3**

**Kierunek zmienności glebowej**

Obiekty rozlosowane w obrębie bloków – w jednym bloku tylko raz występuje każdy obiekt

## Podstawowe założenia analizy wariancji

a) Analizowana cecha X jest mierzalna.

b) Rozważanych jest *a* niezależnych populacji (obiektów), które mają rozkłady normalne $N(\mu_i, \sigma_i)$, gdzie $i = 1, 2, \ldots, a$.

c) Rozkłady te mają jednakowe wariancje, $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_a^2$ (założenie jednorodności lub homogeniczności).

d) Obiekty są przydzielone w sposób losowy do poszczególnych jednostek eksperymentalnych (*zasada randomizacji*).

# *Badania jednoczynnikowe*

Model obserwacji w badaniach jednoczynnikowych
o układzie całkowicie losowanych

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

Efekt czynnika A

Błąd doświadczalny

Model obserwacji w badaniach jednoczynnikowych
o układzie bloków losowanych kompletnych

$$y_{ij} = \mu + \gamma_j + \alpha_i + \varepsilon_{ij}$$

i=1,2,…,a
j=1,2,…,r

Efekt bloku

$\varepsilon_{ij} \sim N(0;\sigma)$

# *Układ całkowicie losowy*

$H_0 : \mu_1 = \mu_2 = ... = \mu_v$

$H_1 :\sim H_0$

**Tabela analizy wariancji dla doświadczenia o układzie całkowicie losowy**

| Źródła zmienności | Stopnie swobody | Sumy kwadratów SS | Średnie kwadraty MS | F | p |
|---|---|---|---|---|---|
| Obiekty | v-1 | $SS_O$ | $MS_O$ | $F_0$ | $p_0$ |
| Błąd | n-v | $SS_E$ | $MS_E$ | | |
| Całkowita | n-1 | $SS_C$ | | | |

$F_{\alpha;v-1;n-v}$

# *Układ bloków losowanych kompletnych*

$H_0 : \mu_1 = \mu_2 = ... = \mu_v$

$H_1 :\sim H_0$

**Tabela analizy wariancji dla doświadczenia o układzie bloków losowanych kompletnych**

| Źródła zmienności | Stopnie swobody | Sumy kwadratów SS | Średnie kwadraty MS | F | p |
|---|---|---|---|---|---|
| Obiekty | v-1 | $SS_O$ | $MS_O$ | $F_0$ | $p_0$ |
| Bloki | r-1 | $SS_R$ | $MS_R$ | | |
| Błąd | (r-1)(v-1) | $SS_E$ | $MS_E$ | | |
| Całkowita | n-1 | $SS_C$ | | | |

$F_{\alpha;v-1;(v-1)(r-1)}$

# *Układ całkowicie losowy*

**Hipotezy szczegółowe - test dla różnicy średnich :**

| test Tukeya | | |
|---|---|---|
| $H_0: \mu_i - \mu_{i'} = 0$<br>$H_1: \sim H_0$ | gdy $\lvert\bar{x}_i - \bar{x}_{i'}\rvert > NIR^T$<br>$H_0$ odrzucamy | Najmniejsza istotna różnica<br>$NIR^T = \sqrt{\frac{MS_E}{r}} \cdot q_{\alpha,v,v_E}$ |

# *Układ bloków losowanych kompletnych*

| | | |
|---|---|---|
| $H_0: \mu_i - \mu_{i'} = 0$<br>$H_1: \sim H_0$ | gdy $\lvert\bar{x}_i - \bar{x}_{i'}\rvert > NIR^T$<br>$H_0$ odrzucamy | Najmniejsza istotna różnica<br>$NIR^T = \sqrt{\frac{MS_E}{r}} \cdot q_{\alpha,v,v_E}$ |

## DZIĘKUJĘ ZA UWAGĘ! 😊

Materiał powstał w ramach projektu „Najlepsi z natury! Podnoszenie kompetencji osób dorosłych przez UPP”, współfinansowanego ze środków Europejskiego Funduszu Społecznego Plus, w ramach programu Fundusze Europejskie dla Rozwoju Społecznego 2021-2027, na podstawie umowy nr FERS.01.05-IP.08-0469/23