Wstęp do uczenia maszynowego

Testowanie hipotez statystycznych

Ewa Szczurek + BW (modyfikacje)

bartek@mimuw.edu.pl
Instytut Informatyki
Uniwersytet Warszawski

4 marca 2024

Testowanie hipotez na przykładzie inwestowania giełdowego

- Odziedziczyłeś pieniądze po cioci z Ameryki
- Prosisz kolegę aby zainwestował te pieniądze za Ciebie
- Rok pózniej kolega twierdzi, że wszystkie Twoje pieniądze przegrał na inwestycjach giełdowych.

Jak sprawdzić hipotezy

- Kolega oszukuje (H_1)
- Kolega te pieniądze stracił uczciwie (H_0)?

Testowanie hipotez na przykładzie inwestowania giełdowego

- Przyjmijmy hipotezę zerową (kolega jest uczciwy). Oszacujmy, jakie są szanse, ze uczciwy inwestor straciłby wszystkie pieniądze w rok na giełdzie?
  - Prześledzmy, jakie były straty tegorocznych inwestorów operujących własnymi pieniędzmi (na pewno uczciwych)

Testowanie hipotez na przykładzie inwestowania giełdowego

Nasze wnioski będą zależeć od zachowania uczciwych inwestorów.

Rozważmy przypadki

1. Okazuje się, że sporo inwestorów miało duże straty.
   - trudno odrzucić H_0 (uczciwość kolegi).
   - nie dowodzi to uczciwości, tylko nie daje podstaw do jej odrzucenia.
2. Okazuje się, że traci tylko 1 na 1000 inwestorów.
   - uczciwość kolegi jest bardzo nieprawdopodobna
   - odrzucamy hipotezę zerową

Hipotezy statystyczne

Dane: przestrzeń statystyczna $(\Omega, F, \mathcal{P})$, gdzie $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ rodzina miar probabilistycznych, a także dwa komplementarne podzbiory przestrzeni parametrów, $\Theta_0 \cap \Theta_1 = \emptyset$ i $\Theta_0 \cup \Theta_1 = \Theta$.

Hipotezy statystyczne

Stwierdzenia postaci $\theta \in \Theta_0$ oraz $\theta \in \Theta_1$, nazywane odpowiednio hipotezą zerową i alternatywną.

Intuicja: mamy przypuszczenia co do rozkładu danej populacji (jego postaci funkcyjnej lub wartości parametrów). Prawdziwość przypuszczenia sprawdzamy na podstawie wyników próby losowej.

Hipotezy statystyczne: przykład

Eksperyment odpowiadający obserwacji liczby K_n sukcesów w n doświadczeniach w schemacie Bernoulliego, polegających na rzucie monetą. Mamy zadane

- $\Omega = \{0, 1, \dots n\}$
- F- rodzina wszystkich możliwych podzbiorów Ω
- $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, gdzie rozkład P_θ zmiennej losowej K_n przyjmującej wartości $k = 0, \dots, n$, jest dwumianowy
$$P_\theta(k) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$
dla nieznanego parametru $\theta \in (0, 1)$.

Możemy wykorzystać eksperyment w celu weryfikacji hipotezy o symetryczności monety, przyjmując
$$H_0 : \theta = 0.5, H_1 : \theta \neq 0.5$$
Tutaj $\Theta_0 = \{0.5\}$, a $\Theta_1 = (0, 0.5) \cup (0.5, 1)$.

Hipotezy statystyczne

Hipoteza prosta

Hipoteza, dla której odpowiadający jej zbiór Θ_i ($i = 0, 1$) składa się z tylko jednego elementu.

Hipoteza złożona

Hipoteza, dla której odpowiadający jej zbiór Θ_i ($i = 0, 1$) zawiera więcej niż jeden parametr.

Hipoteza parametryczna

Dotyczy wartości parametrów określonego rozkładu.

Hipoteza nieparametryczna

Nie dotyczy wartości parametrów rozkładu

Przykłady hipotez statystycznych

Weźmy skokową zmienną losową.

- Zbiór hipotez dopuszczalnych zawiera wszystkie rozkłady skokowe.
- Hipoteza: że populacja ma rozkład Poissona
  - nieparametryczna oraz złożona (na Θ_i składają się wszystkie rozkłady Poissona, różniące się wartością λ).
- Hipoteza: że populacja ma rozkład Poissona z $\lambda = 0.5$
  - parametryczna, prosta.

Zasady weryfikacji hipotez

Test statystyczny

Reguła postępowania, która każdej próbie przyporządkowuje decyzję przyjęcia lub odrzucenia hipotezy H_0.

Zasady weryfikacji hipotez

$X_1, \ldots, X_n$ - próba losowa. Kolejne kroki:

- sformułowanie **hipotezy zerowej**, weryfikowanej
  - $H_0 : \theta \in \theta_0$ dla $\theta_0 \subset \Theta$
- sformułowanie **hipotezy alternatywnej**, będącej dopełnieniem H_0
  - $H_1 : \theta \in \theta_1$ dla $\theta_1 = \Theta \setminus \theta_0$
- Przyjęcie **statystyki testowej** $Z_n(X_1, \ldots, X_n)$
- Określenie obszarów: **zbioru krytycznego** (odrzuceń H_0) W i **zbioru przyjęć** W' (nieodrzuceń H_0), takich, że
  - $W \cap W' = \emptyset, W \cup W' = \mathbb{R}$
  - jeżeli $Z_n(X_1, \ldots, X_n) \in W$ to H_0 odrzucamy
  - jeżeli $Z_n(X_1, \ldots, X_n) \in W'$ to nie ma podstaw do odrzucenia H_0
- Ale, ale – jak dobrać Z_n i jak dobrać W?

Błędy

Weryfikujemy hipotezy na podstawie próby. Ponieważ w większości schematów może się zdarzyć (z małym prawdopodobieństwem) próba, która jest niereprezentatywna dla rozkładu, możemy popełnić dwa rodzaje błędów:

- **Błąd I rodzaju**:
  - odrzucimy H_0, gdy w istocie jest ona prawdziwa
  - na przykład: oskarżamy uczciwego kolegę o oszustwo
  - z prawdopodobieństwem $\alpha(W) = P(Z_n \in W|H_0)$ (mówimy na **poziomie istotności** α)
- **Błąd II rodzaju**:
  - przyjmiemy H_0, gdy jest ona fałszywa
  - na przykład: przyjmujemy, że kolega, który nas oszukał, jest uczciwy
  - z prawdopodobieństwem $\beta(W) = P(Z_n \in W'|H_1) = P(Z_n \in \mathbb{R} \setminus W|H_1)$.

Na ogół nie da się minimalizować $\alpha(W)$ i $\beta(W)$ naraz.

Moc

Moc testu

$$P(Z_n \in W | H_1) = 1 - \beta(W)$$

Test najmocniejszy

Test o tak dobranym obszarze krytycznym W, aby zminimalizować $\beta(W)$ przy ustalonym z góry poziomie istotności (prawdopodobieństwie błędu I rodzaju) $\alpha = \alpha(W)$.

Podział testów

Wyróżniamy testy

1. istotności
2. zgodności

Parametryczne testy istotności

Weryfikują hipotezy dotyczące wartości parametrów, na przykład

$$H_0 \quad : \quad \theta = \theta_0 \tag{1}$$
$$H_1 \quad : \quad \theta \neq \theta_0 \tag{2}$$

Test istotności dla wartości średniej (A)

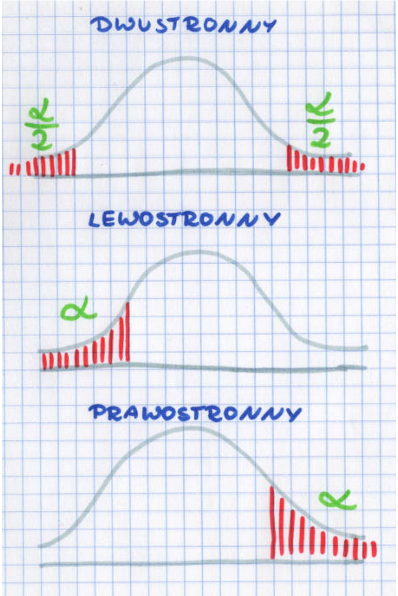
- Założenia: populacja ma rozkład normalny z **nieznanym** parametrem μ i **znanym** σ.
- $H_0 : \mu = \mu_0$
- przeciwko jednej z możliwych hipotez alternatywnych
  1. $H_1 : \mu < \mu_0$,
  2. $H_2 : \mu > \mu_0$, lub
  3. $H_3 : \mu \neq \mu_0$
- Fakt: Średnia z próby pochodzącej z rozkładu normalnego $N(\mu, \sigma)$ ma rozkład $N(\mu, \frac{\sigma}{\sqrt{n}})$
- Przy H_0 prawdziwej,
  - statystyka $\bar{X}$ ma rozkład $N(\mu_0, \frac{\sigma}{\sqrt{n}})$.
  - $\leftrightarrow$ statystyka $U = \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n}$ ma rozkład $N(0, 1)$.

Test istotności dla wartości średniej (A)

- .. przy H_0 prawdziwej, statystyka $U = \frac{\bar{X}-\mu_0}{\sigma}\sqrt{n}$ ma rozkład $N(0,1)$.
- Stąd, dla poziomu istotności α obszary krytyczne to
  1. $W = (-\infty, -q(1-\alpha))$ dla alternatywy $H_1 : \mu < \mu_0$,
  2. $W = (q(1-\alpha), \infty)$ dla alternatywy $H_2 : \mu > \mu_0$,
  3. $W = (-\infty, -q(1-\alpha/2)) \cup (q(1-\alpha/2), \infty)$ dla $H_3 : \mu \neq \mu_0$,

  gdzie $q(p)$ to kwantyl rzędu p rozkładu standardowego normalnego.

Obszary krytyczne dla różnych hipotez alternatywnych



- $H_0 : \mu = \mu_0$
- $H_1 : \mu \neq \mu_0$
- $W = (-\infty, -q(1-\alpha/2)) \cup (q(1-\alpha/2), \infty)$

- $H_0 : \mu \geq \mu_0$
- $H_1 : \mu < \mu_0$,
- $W = (-\infty, -q(1-\alpha))$

- $H_0 : \mu \leq \mu_0$
- $H_1 : \mu > \mu_0$,
- $W = (q(1-\alpha), \infty)$. [a]

Przykład pizzowy

Przykład pizzowy

Pizzeria zapewnia, że średnio dostarcza pizzę do klienta w 28min. By to sprawdzić zmierzono czas 100 losowo wybranych dostaw. Obliczono średni czas dostaw 31.5min. Znamy odchylenie standardowe: 5min.

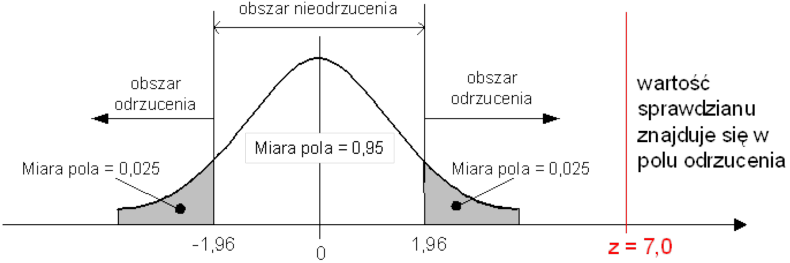
- Czy możemy uwierzyć w zapewnienia Pizzerii?

- $H_0 : \mu = 28$
- $H_1 : \mu \neq 28$
- Statystyka testowa

$$z = \frac{31.5 - 28}{5}\sqrt{100} = 7$$

- Poziom istotności 5%
- Obszar krytyczny $(-\infty; -1.96) \cup (1.96; +\infty)$

Przykład pizzowy



Test istotności dla wartości średniej (B)

- Założenia: populacja ma rozkład normalny, μ i σ **nieznane**.
- $H_0 : \mu = \mu_0$, przeciwko jednej z hipotez alternatywnych
  1. $H_1 : \mu < \mu_0$,
  2. $H_2 : \mu > \mu_0$, lub
  3. $H_3 : \mu \neq \mu_0$
- Dla próby z rozkładu normalnego o średniej μ, $T = \frac{\bar{X}_n - \mu}{S_n}\sqrt{n-1}$, gdzie $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$, ma rozkład t-Studenta o $n-1$ stopniach swobody.
- Zatem przy H_0 prawdziwej, statystyka $T = \frac{\bar{X} - \mu_0}{S_n}\sqrt{n-1}$ ma rozkład $t(n-1)$.
- Stąd, dla poziomu istotności α obszary krytyczne to
  1. $W = (-\infty, -t(1-\alpha, n-1))$ dla alternatywy H_1,
  2. $W = (t(1-\alpha, n-1), \infty)$ dla alternatywy H_2,
  3. $W = (-\infty, -t(1-\alpha/2, n-1) \cup (t(1-\alpha/2, n-1), \infty)$ dla H_3,

  gdzie $t(p, \nu)$ to kwantyl rzędu p rozkładu t-Studenta o ν stopniach swobody.

Test istotności dla dwóch średnich

- Założenia: dwie próby o licznościach n_1 i n_2, z populacji o rozkładach $N(\mu_1, \sigma_1)$, $N(\mu_2, \sigma_2)$. Znamy wariancje σ_1^2 i σ_2^2.
- $H_0 : \mu_1 = \mu_2$ (średnie są takie same)
- Fakt: Różnica średnich z prób $\bar{X}_1 - \bar{X}_2$ ma rozkład $N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$
- Jeśli H_0 prawdziwa to
  - zmienna $(\bar{X}_1 - \bar{X}_2)$ ma rozkład $N\left(0, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$
  - czyli statystyka $U = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$, ma rozkład $N(0, 1)$.
- Stąd, dla poziomu istotności α obszary krytyczne dla U to
  1. $W = (-\infty, -q(1-\alpha))$ dla alternatywy $H_1 : \mu_1 < \mu_2$,
  2. $W = (q(1-\alpha), \infty)$ dla alternatywy $H_2 : \mu_1 > \mu_2$,
  3. $W = (-\infty, -q(1-\alpha/2)) \cup (q(1-\alpha/2), \infty)$ dla $H_3 : \mu_1 \neq \mu_2$,

gdzie $q(p)$ to kwantyl rzędu p rozkładu standardowego normalnego.

Testowanie hipotez nieparametrycznych

Założenia:

- wszystkie obserwacje są niezależne statystycznie
- wartości z próby są porządkowalne

Czego nie zakładamy:

- konkretngo rozkładu populacji

Wiele testów nieparameterycznych opiera się o rangi obserwacji (a nie o ich wartości).

- Niewrażliwe na monotoniczne przekształcenia danych
- Niewrażliwe na wartości odstające

Test Manna-Whitneya/Wilcoxona

Założenia: X, Y - próby liczności m i n, z dwóch różnych populacji, o równych wariancjach.

- H_0: $P(X > Y) = P(Y > X)$.
- H_1: $P(X > Y) > P(Y > X)$ lub H_2: $P(X > Y) < P(Y > X)$ lub H_3: $P(X > Y) \neq P(Y > X)$.

Koncepcja:

- Pomieszajmy wartości z obu prób
- Przypiszmy każdej z wartości rangę
- Statystyka testowa R: suma rang wartości z próby Y
- Przy hipotezie zerowej, każde przypisanie rang wartościom z próby Y z $\binom{m+n}{n}$ możliwych jest jednakowo prawdopodobne
- Dla każdego z jednako prawdopodobnych przypisań możemy policzyć sumę rang i wyznaczyć rozkład statystyki testowej

Przykład: Porównujemy wyniki Treatment vs Control

Treatment	Control
1 (1)	6 (4)
3 (2)	4 (3)

- Przypisujemy wynikom dla czterech pacjentów ich rangi
- Suma rang w grupie Control wynosi $R = 7$
- Suma rang w grupie Treatment wynosi 3
- Czy taki wynik wskazuje na istotną różnicę pomiędzy Treatment a Control?

- Policzmy rozkład R przy założeniu hipotezy zerowej (przypisanie do Control i Treatment jest losowe)
  - Każde z $\binom{4}{2} = 6$ przypisań rang do pacjentów z grupy Control jest jednakowo możliwe

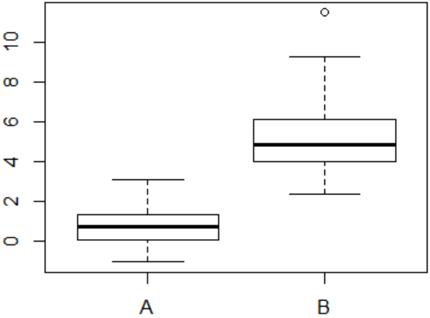
Ranks	R
{1,2}	3
{1,3}	4
{1,4}	5
{2,3}	5
{2,4}	6
{3,4}	7

Z tej tabeli możemy wywnioskować jaki jest rozkład R przy hipotezie zerowej

r	3	4	5	6	7
$P(R = r)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

Test Manna-Whitneya/Wilcoxona

Test potrafi wykryć przesunięcia rozkładów pomiędzy populacjami.



Testy zgodności

Sprawdza zgodność empirycznego rozkładu z próby z rozkładem hipotetycznym, lub zgodność dwóch rozkładów empirycznych.

Test zgodności

Niech $X_1, \ldots, X_n$ próba prosta pobrana z rozkładu o dystrybuancie F. Test zgodności dotyczy hipotez postaci

- $H_0 : F = F_0$
- $H_1 : F \neq F_0$

Test zgodności χ^2-Pearsona

$X_1, \ldots, X_n$- próba.

- Grupujemy wyniki w k rozłącznych typów wartości w próbie
- n_i -liczebność w klasie i, $i = 1, \ldots, k$, $\sum_{i=1}^{k} n_i = n$
- Zakładając prawdziwą H_0
  - obliczamy p_i, prawdopodobieństwo przyjęcia wartości z klasy i (znając F_0)
  - Oczekiwane liczebności klas powinny są równe np_i
  - Statystyka
  $$\chi^2 = \sum_i^k \frac{(n_i - np_i)^2}{np_i}$$
  ma asymptotyczny rozkład $\chi^2(k - l - 1)$, gdzie k-liczba klas, l-liczba nieznanych parametrów rozkładu (estymowanych korzystając z ML)
  - Duże wartości statystyki przemawiają za odrzuceniem H_0.
- Zbiór krytyczny na poziomie istotności α ma postać
$$W = (\chi^2(1 - \alpha, k - l - 1), \infty).$$

Test zgodności χ^2-Pearsona - przykład

Użyto komputera aby wylosować $n = 100$ liczb z rozkładu dwumianowego $Binom(5, 0.5)$. Oto pogrupowanie wyników losowania

Liczba sukcesów	0	1	2	3	4	5
Liczba obserwacji	3	16	36	32	11	2

Zweryfikuj na poziomie istotności $\alpha = 0.05$ hipotezę, że przedstawione wyniki rzeczywiście pochodzą z rozkładu $Binom(5, 0.5)$.

Test zgodności χ^2-Pearsona - przykład

- Dla tego rozkładu (czyli dla H_0 mamy

$$p_i = P(X = i) = \binom{5}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{5-i} = \frac{5!}{i!(5-i)!} \left(\frac{1}{2}\right)^5 .$$

- Stąd, teoretyczne liczności klas np_i dla $i = 0, 1, \dots, 5$ to $3.12, 15.62, 31.25, 31.25, 15.62, 3.12.$
- Wartość statystyki

$$\chi^2 = \sum_{i=0}^{5} \frac{(n_i - np_i)^2}{np_i} = 2.53$$

- Kwantyl rzędu $1 - \alpha$ rozkładu χ^2 o $k - 1 = 5$ stopniach swobody $\chi^2(0.95, 5) = 11.1$. Mamy $2.53 < 11.1$, a zatem nie ma podstaw do odrzucenia hipotezy, że komputer generuje liczby z rozkładu $Binom(5, 0.5)$.

Test niezależności: χ^2-Pearsona

Założenia:

- Obserwujemy n par wartości zmiennych (X, Y) skokowych.
- $X \in (x_1, \ldots, x_r)$, $Y \in (y_1, \ldots, y_s)$
- Brzegowe rozkłady: $P(X = x_i) = p_{i.}$, $P(Y = y_j) = p_{.j}$
- Łączny rozkład zmiennych $P(X = x_i, Y = y_j) = p_{ij}$

Hipotezy:

- $H_0 : p_{ij} = p_{i.}p_{.j}$ dla każdego i, j (niezależność)
- $H_1 : p_{ij} \neq p_{i.}p_{.j}$

Test niezależności: χ^2 Pearsona

Statystyka testowa: na podstawie tablicy kontyngencji (zliczeń)

y/x	x_1	x_2	...	x_s
y_1	n_{11}	n_{12}	...	n_{1s}
...				
y_r	n_{r1}	n_{r2}	...	n_{rs}

Przyjmując, że H_0 prawdziwa

- Dla tablicy zliczeń:

$$T = \sum_{i}^{r} \sum_{j}^{s} \frac{(n_{ij} - n_{i.}n_{.j}/n)^2}{n_{i.}n_{.j}/n}$$

- T ma rozkład $\chi^2((r-1)(s-1))$
- Obszar krytyczny (χ_α, ∞), dla poziomu istotności α i $P(T > \chi_\alpha) = \alpha$.

Test niezależności: dokładny test Fishera

- Założenia: tak jak w teście χ^2 dla zmiennych skokowych, ale o dwóch możliwych wartościach.
- Hipoteza H_0 : zmienne niezależne.
- Oparty o tablicę zliczeń 2x2

	X= 0	**X=1**	**Row total**
Y = 0	*a*	*b*	*a+b*
Y = 1	*c*	*d*	*c+d*
Column total	*a+c*	*b+d*	*a+b+c+d = n*

Test niezależności: dokładny test Fishera

- Przy ustalonych zliczeniach brzegowych, jedna wartość w tabelce (np. a) determinuje pozostałe
- Przy założeniu niezależności, rozkład zliczeń a odpowiada prawdopodobieństwu a sukcesów w $a+c$ losowaniach bez zwracania z populacji, w której łącznie jest $a+b$ sukcesów i $c+d$ porażek.

Test niezależności: dokładny test Fishera

Zatem dla H_0 spełnionej, wartości w tabelce mają rozkład hipergeometryczny.

$$p(a) = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}} = \frac{(a+b)!\ (c+d)!\ (a+c)!\ (b+d)!}{a!\ \ b!\ \ c!\ \ d!\ \ n!},$$

gdzie $p = P(Y_{11} = a \mid \texttt{wartości brzegowe})$, gdzie `wartości brzegowe` są zadane przez

$$Y_{11}+Y_{12} = a+b,\ Y_{21}+Y_{22} = c+d,\ Y_{11}+Y_{21} = a+c,\ Y_{12}+Y_{22} = b+d.$$

Pani Bristol i mleko

Ronald Fisher użył w swojej książce takiego przykładu, opartego na ponoć faktycznym eksperymencie przeprowadzonym na Muriel Bristol, która twierdziła, że potrafi rozpoznać, czy do filiżanki najpierw wlano herbatę, czy najpierw mleko.

Jedyny przypadek, dla którego wynik testu wskazywałby na zdolność rozpoznania z poziomem istotności 0.05, to ten gdyby Pani Bristol zgadła kolejność wlania płynow we wszystkich ośmiu filiżankach.

The Design of Experiments (1935)

Podsumowanie

- Hipoteza statystyczna, prosta złożona, parametryczna, nieparametryczna
- Błędy I i II rodzaju
- Moc testu
- Test istotnosci dla wartosci sredniej dla próby normalnej ze znanym σ
- Test istotnosci dla wartosci sredniej dla próby normalnej z nieznanym σ
- Testy istotnosci dla dwóch srednich
- Test Manna-Whitneya/Wilcoxona
- Testy zgodności
- Test zgodności Pearsona
- Test niezależności Pearsona
- Dokładny test Fishera