

Statystyka Matematyczna

Anna Janicka

wykład IX, 25.04.2016

TESTOWANIE HIPOTEZ STATYSTYCZNYCH

Plan na dzisiaj

1. Hipoteza statystyczna
2. Test statystyczny
3. Błędy I-go i II-go rodzaju
4. Poziom istotności, p -value
5. Schemat przeprowadzania testu statystycznego
6. Moc testu, rozmiar testu



Hipoteza statystyczna

ogólnie: pewna wypowiedź na temat rozkładu prawdopodobieństwa rządzącego interesującym nas zjawiskiem (obserwowaną zmienną losową)

cel: chcemy wnioskować o prawdziwości tej hipotezy na podstawie zaobserwowanych wartości zmiennej losowej



Przykłady hipotez statystycznych

- $X_1, X_2, \dots, X_n$ są próbą z rozkładu wykładniczego
- $X_1, X_2, \dots, X_n$ są próbą z rozkładu *normalnego* (to zakładamy) z param (5, 1)
- $EX_i = 7$ (wartość oczekiwana rozkładu to 7)
- $\text{Var } X_i > 1$ (wariancja rozkładu jest większa niż 1)
- $X_1, X_2, \dots, X_n$ są niezależne
- $EX_i = EY_j$ ($X_1, X_2, \dots, X_n$ oraz $Y_1, Y_2, \dots, Y_m$ mają takie same wartości oczekiwane)



Typy hipotez

☐ hipotezy

- parametryczne: dotyczą parametrów rozkładu
- nieparametryczne: dotyczą innych własności/postaci rozkładu

☐ hipotezy

- proste: wyznaczają dokładnie jeden rozkład
- złożone: wyznaczają rodzinę rozkładów



Hipoteza zerowa i alternatywna

Hipoteza zerowa: „podstawowa”, ozn. H_0

Hipoteza alternatywna: kontr-hipoteza – hipoteza, jaką przyjmujemy w przypadku odrzucenia hipotezy zerowej, ozn. H_1

np.:

- $H_0 : \lambda = 1, \quad H_1 : \lambda \neq 1$
- $H_0 : \lambda = 1, \quad H_1 : \lambda = 2$
- $H_0 : \lambda = 1, \quad H_1 : \lambda > 1$



Hipoteza zerowa i alternatywna – cd.

Hipotezy zerowa i alternatywna nie są równoprawne.

Hipoteza zerowa: stwierdzenie, wniosek z dotychczas obowiązującej teorii, przyjmowane za prawdziwe dopóki nie pojawią się obserwacje „bardzo trudne do pogodzenia” z tym przypuszczeniem. Albo „spekulacja”.

Hipoteza alternatywna: możliwość brana pod uwagę, jeśli zmuszeni będziemy do odrzucenia hipotezy zerowej



Test statystyczny

Procedura, która na podstawie konkretnych obserwacji (tj. dla każdej wartości obserwowanej zmiennej losowej) prowadzi do jednej z dwóch decyzji:

- odrzucić hipotezę zerową (na rzecz alternatywnej)
- nie odrzucać hipotezy zerowej

„odrzucamy H_0 ”

„nie ma podstaw do odrzucenia H_0 ”



Test statystyczny formalnie

Punkt wyjścia: model statystyczny

- $X = (X_1, X_2, \dots, X_n)$ – wektor obserwacji $\in \mathbf{X}$
- $X \sim P_\theta, \{P_\theta: \theta \in \Theta\}$ – rodzina rozkładów

Hipotezy H_0, H_1 :

- $H_0: \theta \in \Theta_0$
- $H_1: \theta \in \Theta_1$

t. że $\Theta_0 \cap \Theta_1 = \emptyset$

(hipotezy się wzajemnie wykluczają)



Test statystyczny formalnie – cd.

Test hipotezy H_0 przeciw H_1 :

statystyka $\delta: \mathbf{X} \rightarrow \{0, 1\}$

wartość 1 interpretujemy jako decyzję o odrzuceniu H_0 (na rzecz H_1) zaś 0 jako nieodrzućenie H_0

Obszar (zbiór) krytyczny testu:

$K = \{x \in \mathbf{X} : \delta(x) = 1\}$ – zbiór wyników, przy których odrzucamy H_0 ;

Obszar (zbiór) afirmacji testu:

$A = \{x \in \mathbf{X} : \delta(x) = 0\}$ – zbiór wyników, przy których nie odrzucamy H_0



$$K \cup A = \mathbf{X}, K \cap A = \emptyset$$

Test statystyczny formalnie – cd. (2)

Obszar krytyczny testu przeważnie ma postać

$$K = \{x \in \mathbf{X} : T(x) > c\}$$

dla pewnej statystyki T (tzw. **statystyki testowej**) oraz liczby c (tzw. **wartości krytycznej**), odpowiednio dobranych

Opisy testu statystycznego (równoważne):

- podanie T i c
- podanie K
- podanie δ



często obszarem krytycznym testu nazywa się przedział wartości statystyki, a nie prowadzący do niego zakres wartości obserwacji

Test statystyczny – przykład

Sprawdzamy, czy moneta jest symetryczna

Rzucamy tą monetą 400 razy

$$X \sim B(400, p)$$

□ $H_0 : p = 1/2, \quad H_1 : p \neq 1/2$

□ Jakie wyniki skłonią nas do odrzucenia H_0 ?

■ $|X - 200| < c$ – nie odrzucamy H_0 .

■ $|X - 200| \geq c$ – odrzucamy H_0 na rzecz H_1 .

tn. $T(x) = |x - 200|$

→ jakie powinno być c ?



Błędy I-go i II-go rodzaju

Z uwagi na losowość obserwacji, zawsze jest możliwość popełnienia błędu

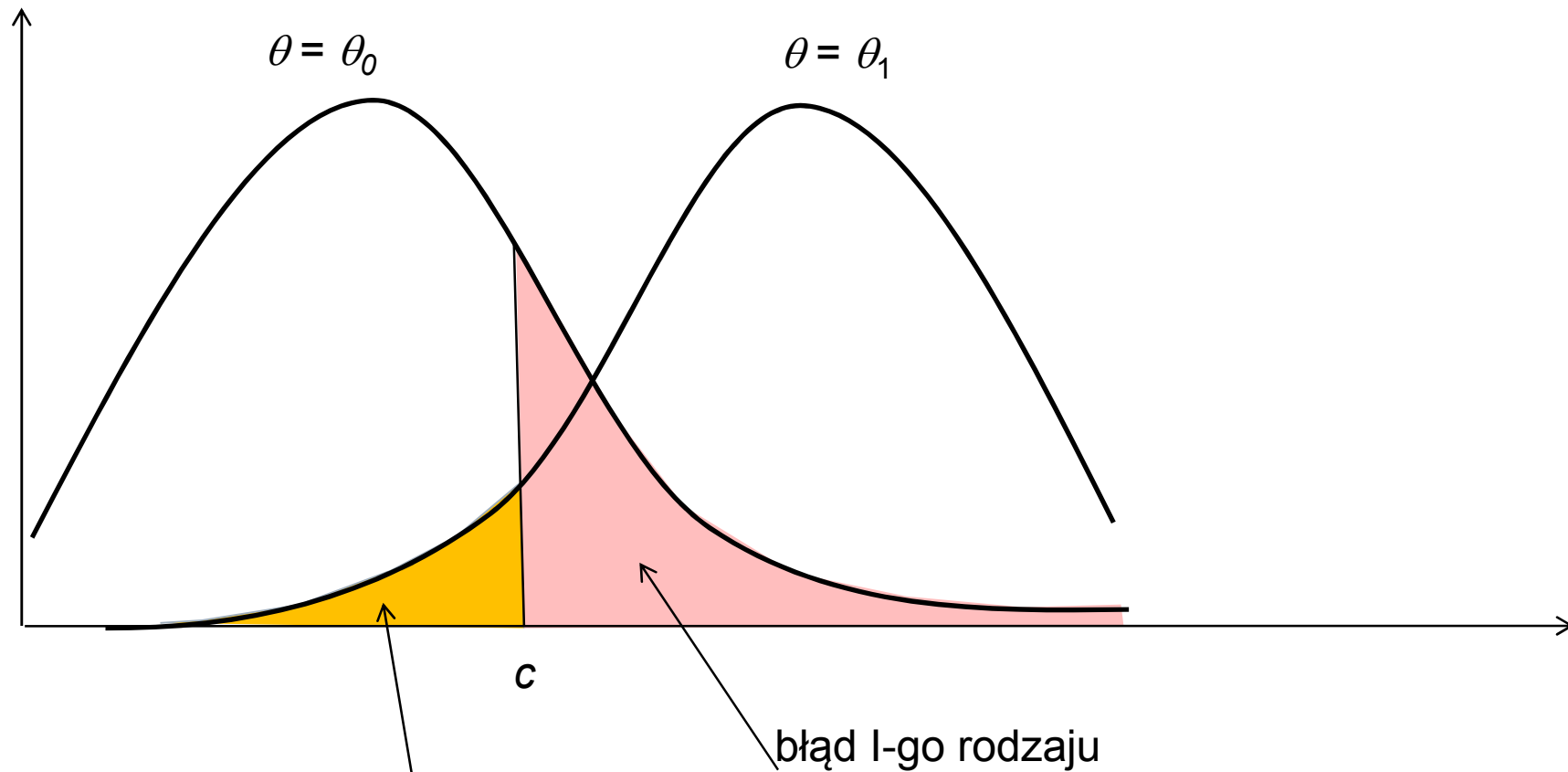
decyzja	Stan faktyczny	
	H_0 prawdziwa	H_0 fałszywa
odrzuć H_0	błąd I-go rodzaju	OK
nie odrzucać H_0	OK	błąd II-go rodzaju

$P_\theta(K)$ dla $\theta \in \Theta_0$ – p-stwo błędu I-go rodzaju

$P_\theta(A)$ dla $\theta \in \Theta_1$ – p-stwo błędu II-go rodzaju

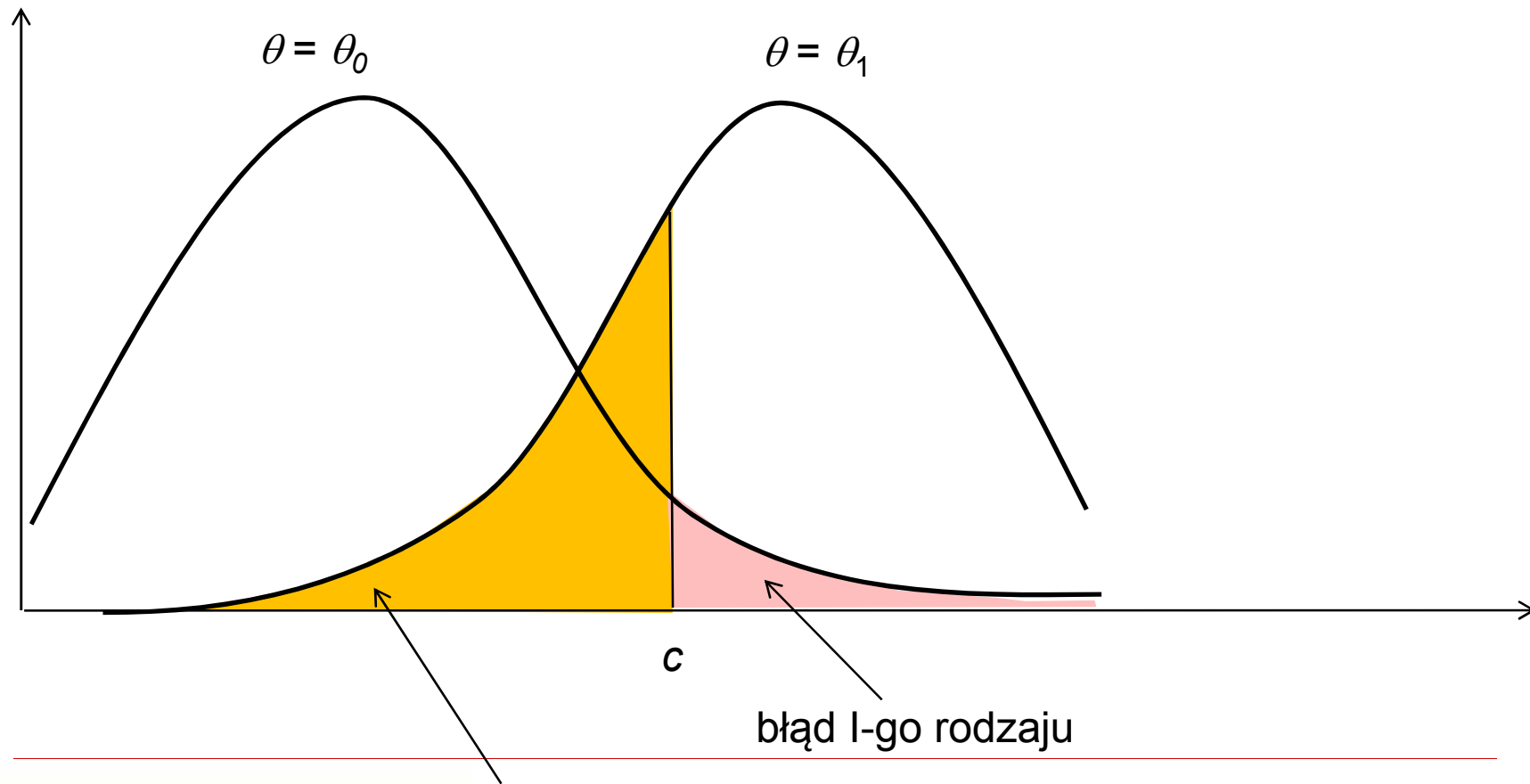
Błędy I-go i II-go rodzaju: interpretacja graficzna (1)

rozkłady statystyki testowej przy założeniu prawdziwości
hipotezy zerowej i alternatywnej



Błędy I-go i II-go rodzaju: interpretacja graficzna (2)

rozkłady statystyki testowej przy założeniu prawdziwości
hipotezy zerowej i alternatywnej



Poziom istotności

Test jest na **poziomie istotności** α , jeśli dla każdego $\theta \in \Theta_0$ mamy $P_\theta(K) \leq \alpha$.

Zwykle: szukamy testów o możliwie najmniejszym p-stwie popełnienia błędu II-go rodzaju dla ustalonego poziomu istotności α , zwykle = 0,1 lub 0,05 lub 0,01

Błąd I-go rodzaju zwykle ważniejszy – nie tylko konserwatyzm.

Test statystyczny – przykład cd.

Wyznaczanie obszaru krytycznego

Chcemy: poziom istotności $\alpha = 0,01$

Tzn. szukamy c t. że (przy założeniu $p = 1/2$)

$$P(|X - 200| > c) = 0,01$$

Z tw. de Moivre'a – Laplace'a mamy dla dużych n !

$$P(|X - 200| > c) \approx 2 \Phi(-c/10), \text{ żeby} \\ = 0,01 \text{ to } c \approx 25,8$$

Na poziomie istotności około 0,01 odrzucamy H_0 gdy liczba orłów mniejsza niż 175 lub większa niż 225



$$K = \{0, 1, \dots, 174\} \cup \{226, 227, \dots, 400\}$$

Test statystyczny – przykład cd. (2).

p-value

Nieco inne pytanie: co by było, gdyby liczba orłów była równa 220 ($T = 20$)?

Mamy:

$$P_{1/2}(|X - 200| > 20) \approx 0,05$$

p-value: prawdopodobieństwo błędu I-go rodzaju, gdyby przyjąć za wartość krytyczną uzyskaną wartość statystyki testowej

A zatem: p-value dla wartości statystyki testowej $T = 20$ wynosi ok. 0,05



p-value

p-value – prawdopodobieństwo pojawienia się wartości obserwacji „co najmniej tak samo ekstremalnych” jak zaobserwowane (przeczących hipotezie zerowej nie mniej niż te zaobserwowane)

decyzje:

- $p\text{-value} < \alpha$ – odrzucamy hipotezę zerową
- $p\text{-value} \geq \alpha$ – nie ma podstaw do odrzucenia hipotezy zerowej



Test statystyczny – przykład cd. (3)

Wpływ wyboru hipotezy alternatywnej

Dla innej hipotezy alternatywnej...

Np. przegramy, jeśli na monecie będzie
wypadał orzeł *za często*.

□ $H_0 : p = \frac{1}{2}, \quad H_1 : p > \frac{1}{2}$

H_0 mogłoby
brzmieć $p \leq \frac{1}{2}$

□ Jakie wyniki skłonią nas do odrzucenia H_0 ?

■ $X - 200 \leq c$ – nie odrzucamy H_0 .

■ $X - 200 > c$ – odrzucamy H_0 na rzecz H_1 .

tzn. $T(x) = x - 200$



Test statystyczny – przykład cd. (4)

Wpływ wyboru hipotezy alternatywnej

Również z tw. de Moivre'a – Laplace'a:

$$P_{1/2}(X - 200 > c) \approx 0,01 \text{ dla } c \approx 23,3,$$

a zatem na poziomie istotności 0,01

odrzucaamy $H_0 : p = 1/2$ na rzecz $H_1 : p > 1/2$

gdy liczba orłów jest równa co najmniej 224

A co gdy wypadnie 220 orłów?

p-value wynosi ok. 0,025; nie odrzucaamy H_0



Schemat przeprowadzania testu statystycznego

1. Określenie modelu statystycznego
2. Postawienie hipotezy zerowej H_0 i alternatywnej H_1
3. Wybór poziomu istotności α
4. Wybór statystyki testowej T / zdefiniowanie obszaru krytycznego K
5. Decyzja: zależna od tego, czy wartość statystyki testowej „wpada” do obszaru krytycznego (ew. z porównania p-value i α)



Moc testu (przy hipotezie alternatywnej)

$P_{\theta}(K)$ dla $\theta \in \Theta_1$ – moc testu (przy hipotezie alternatywnej)

Funkcja mocy testu:

$$\beta : \Theta_1 \rightarrow [0, 1] \text{ t. że } \beta(\theta) = P_{\theta}(K)$$

Zwykle: szukamy testów na zadanym poziomie istotności o jak największej mocy.



Test statystyczny – przykład cd. (5)

Moc testu

- Testujemy $H_0 : p = 1/2$ przeciw $H_1 : p = 3/4$
testem: $T(x) = X - 200$, $K = \{T(x) > 23,3\}$
(tj. na poziomie istotności $\alpha = 0,01$)

Moc testu dla hipotezy alternatywnej:

$$\begin{aligned}\beta(3/4) &= P(T(x) > 23,3 \mid p = 3/4) = P_{3/4}(X > 223,3) \\ &\approx 1 - \Phi((223,3 - 300)/5\sqrt{3}) \approx \Phi(8,85) \approx 1\end{aligned}$$

- Ale gdy np. $H_1 : p = 0,51$

$$\beta(0,51) = P(T(x) > 23,3 \mid p = 0,51) \approx \Phi(1,93) \approx 0,973$$

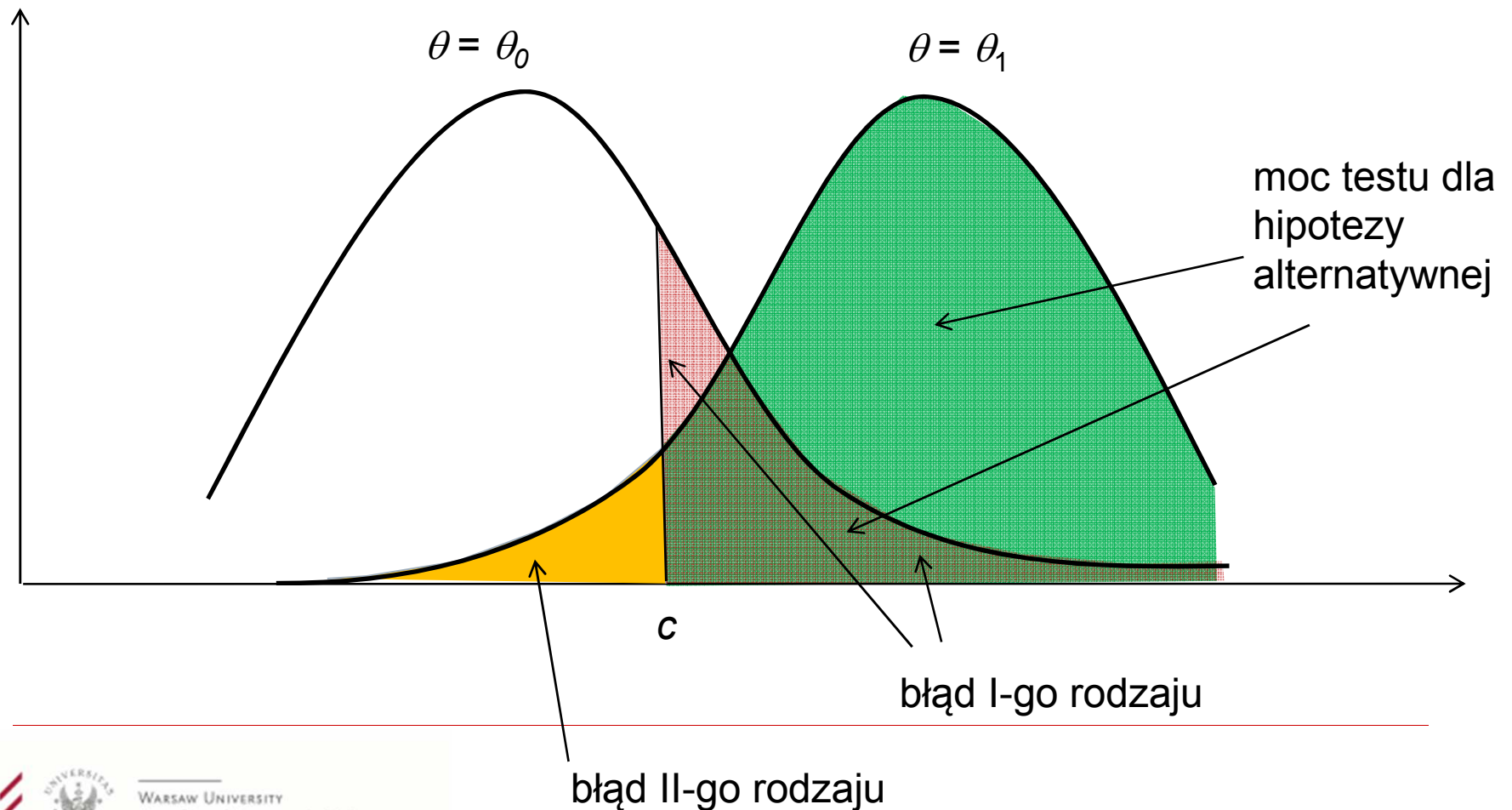
- A gdyby np. $H_1 : p = 1/4$ to dla statystyki testowej T

~~$$\beta(1/4) = P(T(x) > 23,3 \mid p = 1/4) \approx 1 - \Phi(14,23) \approx 0$$~~



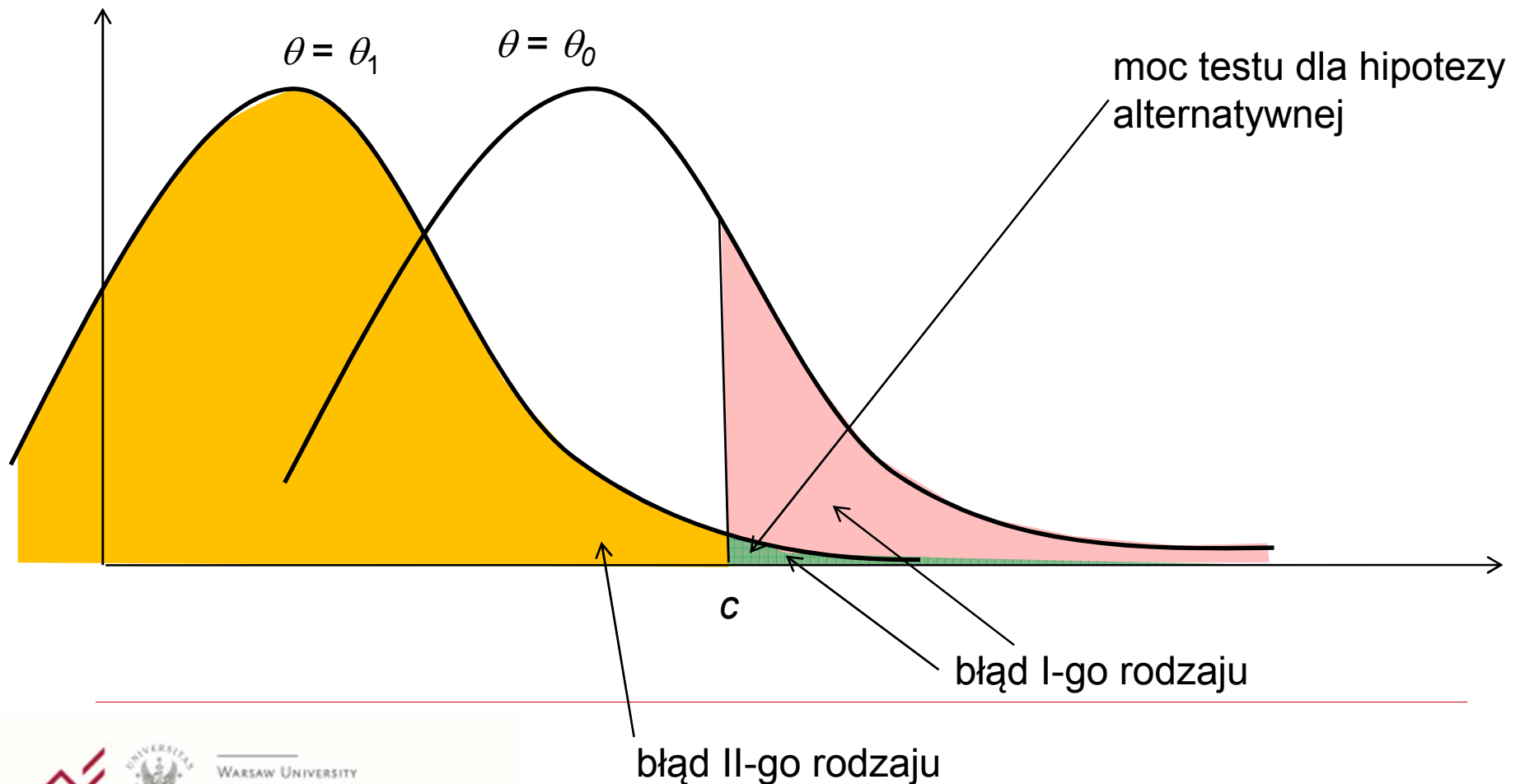
Moc testu: interpretacja graficzna (1)

rozkłady statystyki testowej przy założeniu prawdziwości
hipotezy zerowej i alternatywnej



Moc testu: interpretacja graficzna (2)

rozkłady statystyki testowej przy założeniu prawdziwości
hipotezy zerowej i alternatywnej



Czułość i swoistość

Swoistość – odsetek wyników *prawdziwie ujemnych* (gdy fałszywa H_0)

Czułość – odsetek wyników *prawdziwie dodatnich* (gdy prawdziwa H_0)

zwł. w badaniach medycznych (H_0 to choroba)



Rozmiar testu

czasem mówi się również o **rozmiarze testu**:

$$\sup_{\theta \in \Theta_0} P_{\theta}(K)$$

wówczas:

poziom istotności = α jeśli rozmiar testu nie przekracza α .

Czasem poszukuje się tzw. nieobciążonych testów: moc testu musi być co najmniej tak duża jak rozmiar testu.



WARSAW UNIVERSITY
Faculty of Economic Sciences
