Komputerowa analiza danych doświadczalnych

Wykład 10

12.05.2022

dr inż. Łukasz Graczykowski

lukasz.graczykowski@pw.edu.pl

Semestr letni 2021/2022

Rozkłady t-Studenta i χ^2

Testy statystyczne

Rozkład χ2

Rozkład χ2

- Załóżmy (dla uproszczenia), że badamy populację opisaną standardowym rozkładem normalnym (Gaussa) o wartości oczekiwanej 0 i wariancji 1
- Pobieramy z niego próbę n-elementową i tworzymy sumę kwadratów:

$$\chi^2 = X_1^2 + X_2^2 + \ldots + X_n^2$$

- Zmienna χ^2 podlega **rozkładowi chi-kwadrat** o n stopniach swobody
- Funkcja gęstości prawdopodobieństwa:

$$f(\chi^2) = k \cdot (\chi^2)^{\lambda-1} e^{-1/2\chi^2} \qquad k = \frac{1}{\Gamma(\lambda) 2^{\lambda}}$$

- Dystrybuanta:

$$F(\chi^2) = \frac{1}{\Gamma(\lambda) 2^{\lambda}} \int_0^{\chi^2} u^{\lambda-1} e^{-1/2u} du$$

  – gdzie $\lambda = 1/2n$ a n to liczba stopni swobody

Rozkład χ2

- Funkcja charakterystyczna rozkładu χ^2 ma postać:

$$\phi_{\chi^2}(t)=(1-2it)^{-\lambda}$$

- Korzystając z jej własności funkcji charakterystycznej otrzymujemy natychmiast, że suma dwóch różnych χ^2 o n_1 i n_2 stopniach swobody daje również rozkład χ^2 o $n=n_1+n_2$ stopniach swobody

- Różniczkując funkcję charakterystyczną możemy wyznaczyć wartość oczekiwaną i wariancję rozkładu χ^2:

$$E\{X^2\}=-i\phi_{\chi^2}'(0)=2\lambda\equiv n$$

$$E\{(X^2)^2\}=-i\phi_{\chi^2}''(0)=4\lambda^2+4\lambda$$

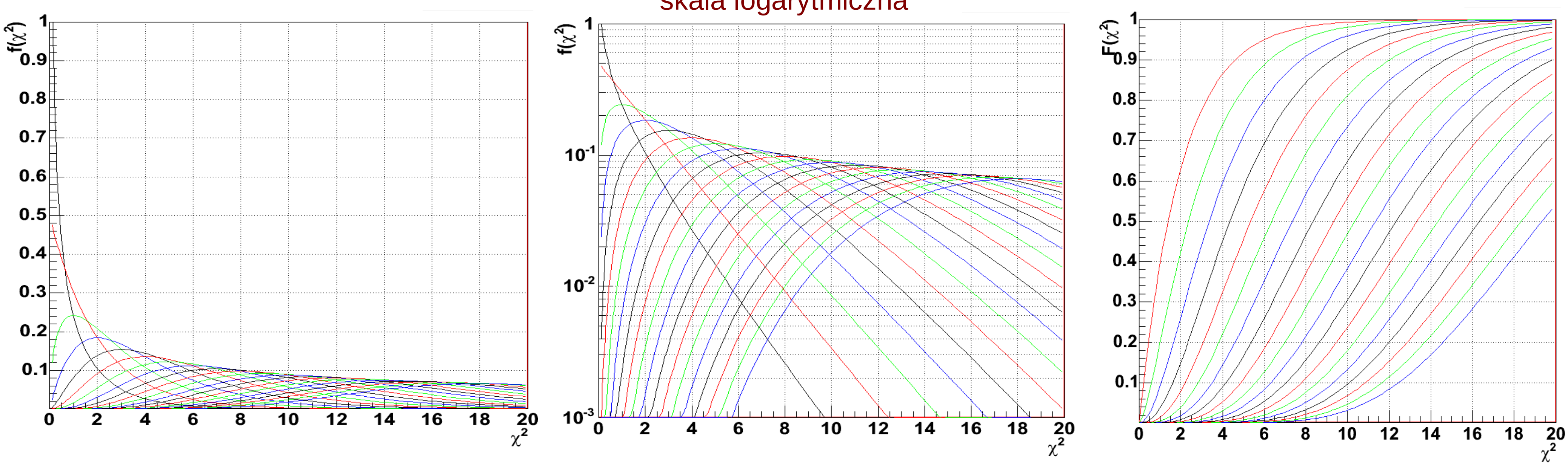
$$\sigma^2(X^2)=E\{(X^2)^2\}-(E\{X^2\})^2=4\lambda\equiv 2n$$

- Czyli wartość wartość oczekinawa rozkładu χ^2 wynosi n a wariancja $2n$

Rozkład χ2

- Wykresy rozkładu i dystrybuanty rozkładu χ^2 dla n od 1 do 20

skala logarytmiczna



- W rzeczywistych przypadkach mamy do czynienia z pełnym rozkładem Gaussa o dowolnym a i σ. Wprowadzamy wtedy odpowiednie przeskalowanie

$$\chi^2 = X^2 = \frac{(X_1 - a)^2 + (X_2 - a)^2 + \ldots + (X_n - a)^2}{\sigma^2}$$

- a w ogólnym przypadku wielowymiarowym, gdy zmienne są zależne:

$$\chi^2 = X^2 = (\boldsymbol{X} - \boldsymbol{a})^T B (\boldsymbol{X} - \boldsymbol{a})$$

Rozkład t-Studenta

Rozkład t-Studenta

- Załóżmy $X_1,\ldots, X_n$ są zmiennymi losowymi pochodzącymi z rozkładu normalnego $\mathcal{N}(\mu,\sigma^2)$
- Estymator wartości oczekiwanej (średnia z próby) to:

  $$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  a estymator wariancji to:

  $$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

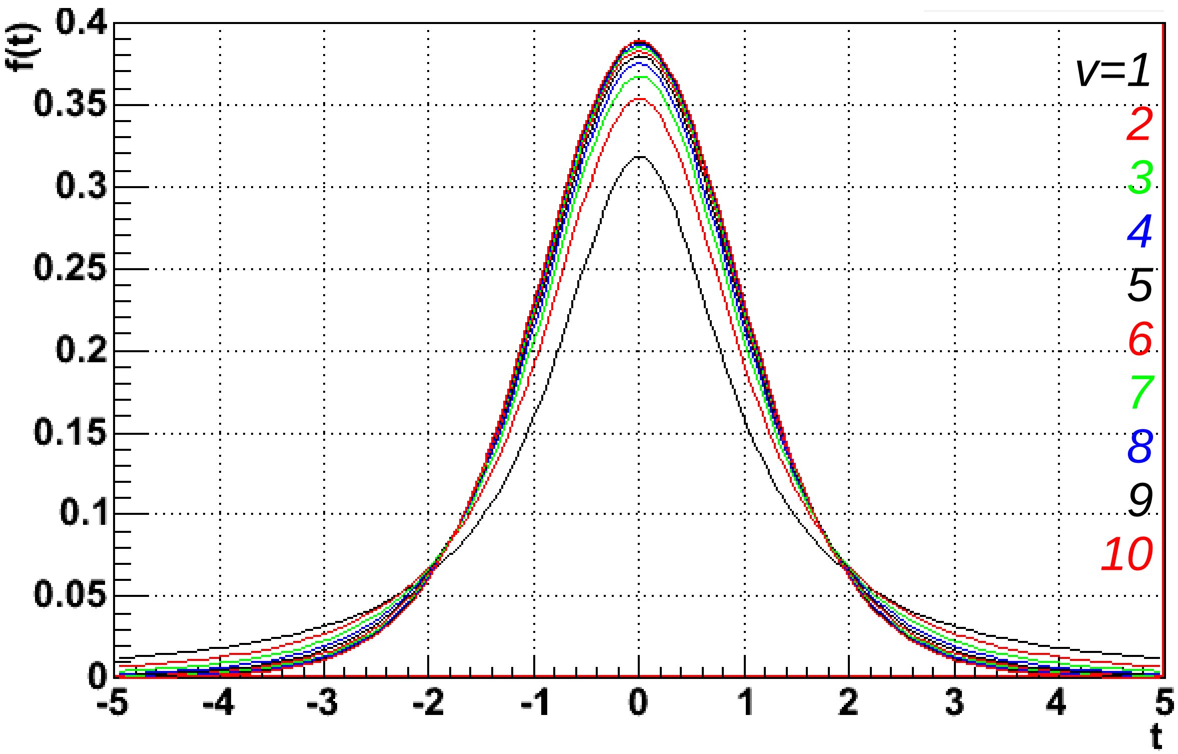
- Wtedy, zmienna $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ ma rozkład normalny standardowy (o średniej 1 i odchyleniu 0)
- A zmienna $\frac{\bar{X}-\mu}{S/\sqrt{n}}$, gdzie zastąpiliśmy odchylenie jego estymatorem ma **rozkład t-Studenta** o n-1 stopniach swobody

Rozkład t-Studenta

- Funkcja gęstości prawdopodobieństwa

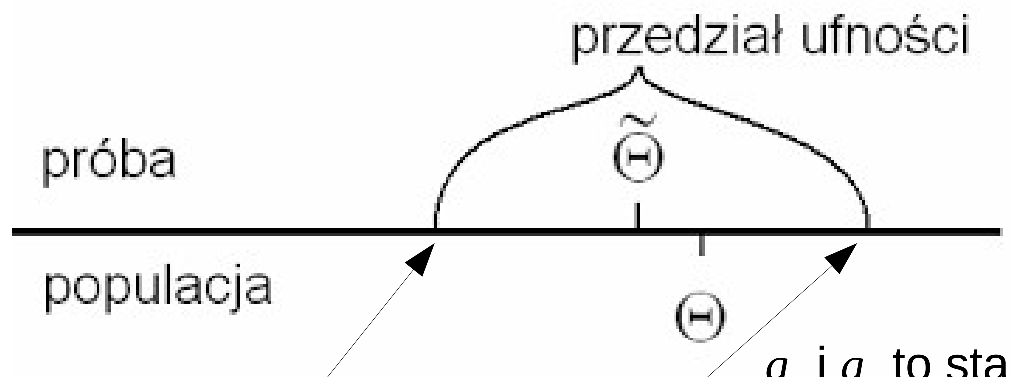
$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$

ν– liczba stopni swobody



- Jak widać, rozkład t-Studenta jest symetryczny względem 0
- Rozkład **dąży do rozkładu Gaussa gdy** $\nu \rightarrow \infty$
- Z symetrii względem 0 mamy związek (analogicznie jak dla Gaussa): $P(|t| \leq t) = 2F(|t|) - 1$

Przedział ufności



g_1 i g_2 to statystyki próby

Postać przedziału ufności jest następująca:

$$\{g_1 \le \Theta \le g_2\}, \qquad P = 1 - \alpha$$

Stwierdzamy, że z prawdopodobieństwem 1-α przedział ufności (g_1, g_2) zawiera szacowany parametr populacyjny Θ.

Wielkość α to **poziom istotności** (lub **ryzyko błędu**, że określony na podstawie próby przedział nie zawiera parametru Θ), zaś 1-α to **poziom ufności**.

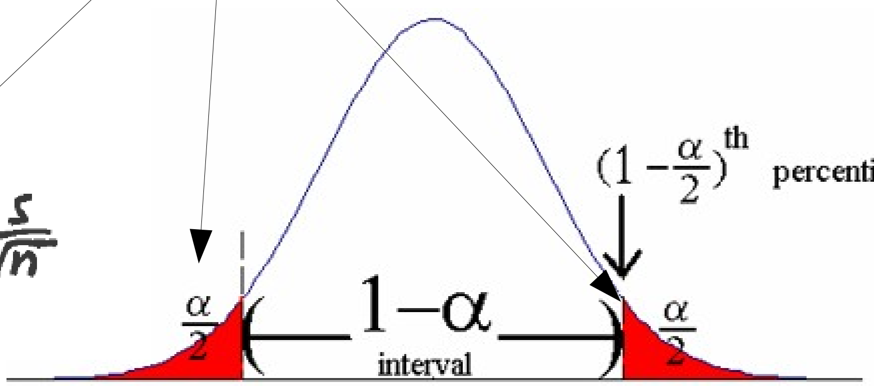
Poziom ufności

- Rozkłady χ^2 oraz t-Studenta pozwalają nam wprowadzić pojęcie **poziomu ufności** dla estymowanych wartości średnich oraz wariancji
- **Wartość średnia z próby**
  - pokazaliśmy, że rozkład t-Studenta ma wielkość $\dfrac{\bar{X}-\mu}{S/\sqrt{n}}$
  - możemy zatem zdefiniować prawdopodobieństwo wokół rzeczywistej wartości oczekiwanej populacji

kwantyle! $F(x_q)=P(X<x_q)=q$

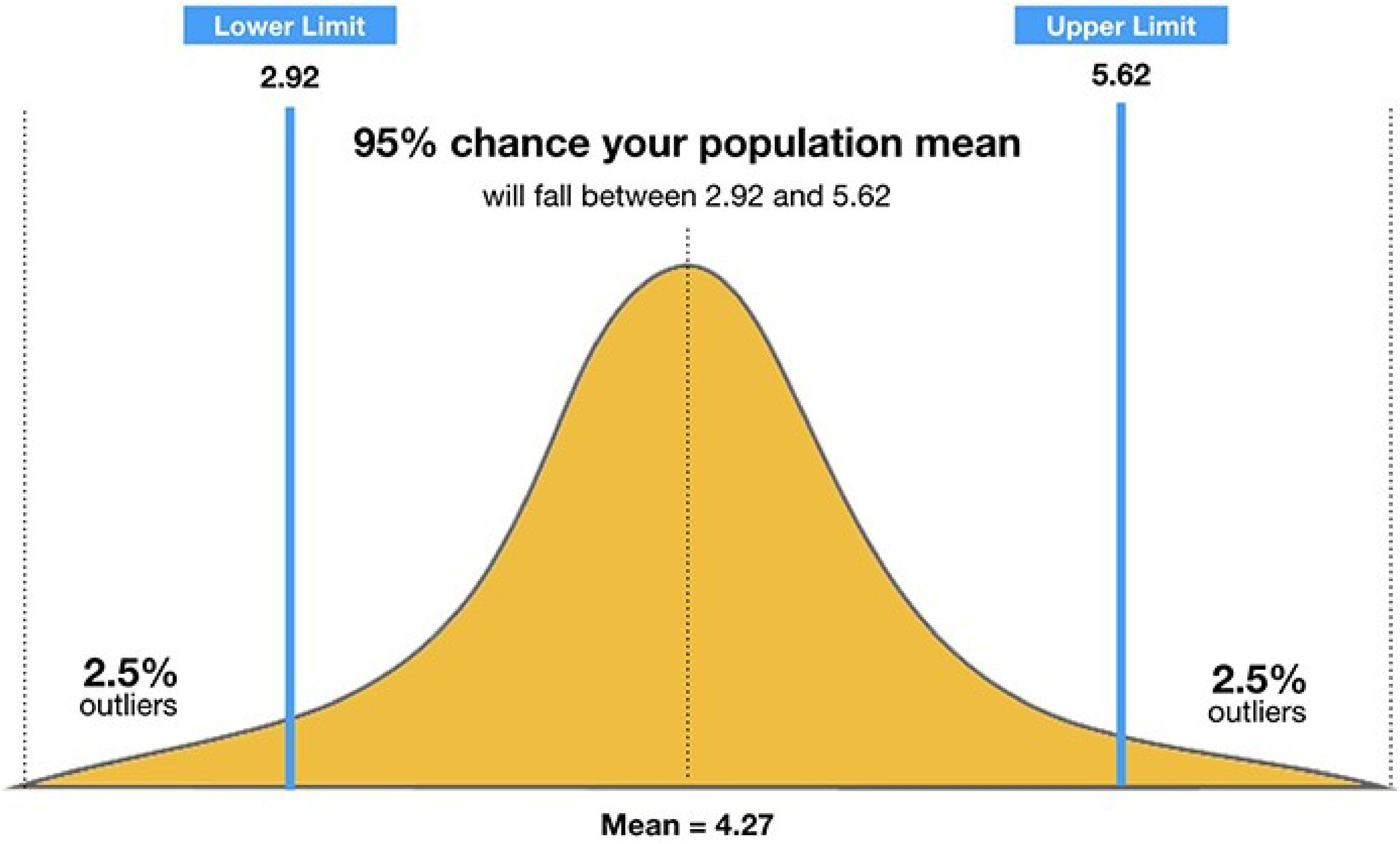
$$X \sim N(\mu,\sigma) \qquad \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

$$\Rightarrow \bar{x} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}$$

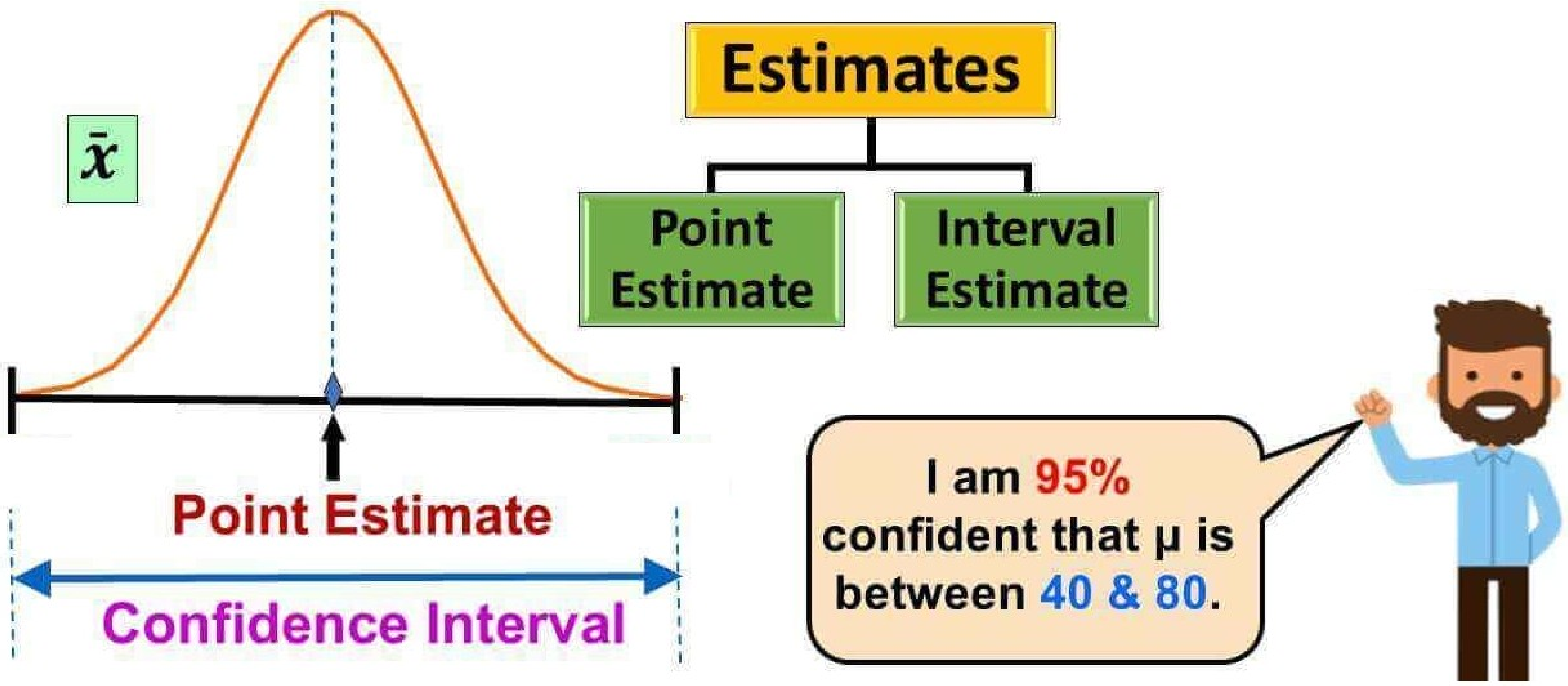


  - oczekujemy, że nasz wynik (estymowana wartość średnia z próby) mieści się w przedziale wyznaczonym przez poziom ufności
  - Dla **dużych wartości *n*** możemy to prawdopodobieństwo estymować rozkładem normalnym (**t-Student → rozkład normalny**)

Poziom ufności

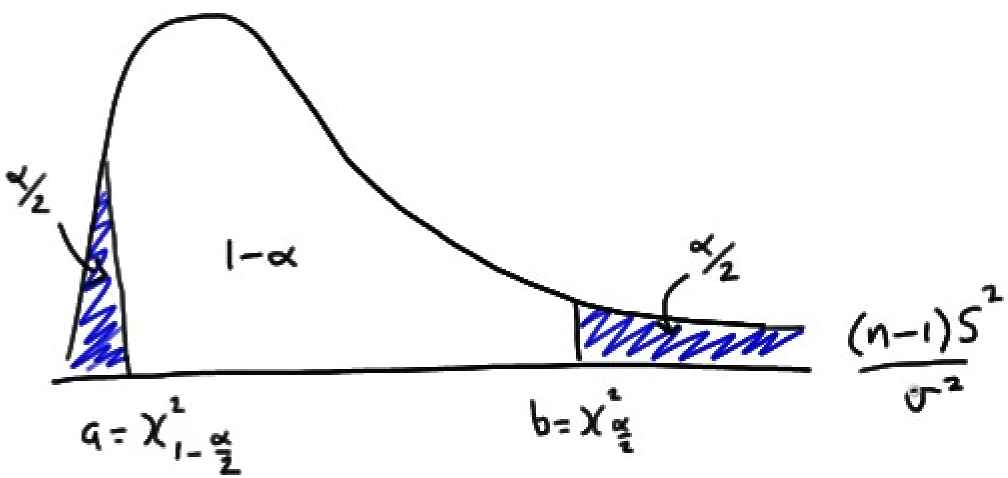


Poziom ufności



Poziom ufności

- Rozkłady χ^2 oraz t-Studenta pozwalają nam wprowadzić pojęcie **poziomu ufności** dla estymowanych wartości średnich oraz wariancji
- **Wariancja z próby**
  - Można, że rozkład χ^2 ma wielkość $\frac{n-1}{\sigma^2} S^2$ estymatyor wariancji
  - możemy zatem zdefiniować prawdopodobieństwo wokół wariancji



$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

  - oczekujemy, że prawdziwa wariancja populacji (na podstawie estymowania wariancji z próby) mieści się w przedziale wyznaczonym przez poziom ufności

https://en.wikipedia.org/wiki/Confidence_interval

Weryfikacja hipotez statystycznych

Weryfikacja hipotez statystycznych

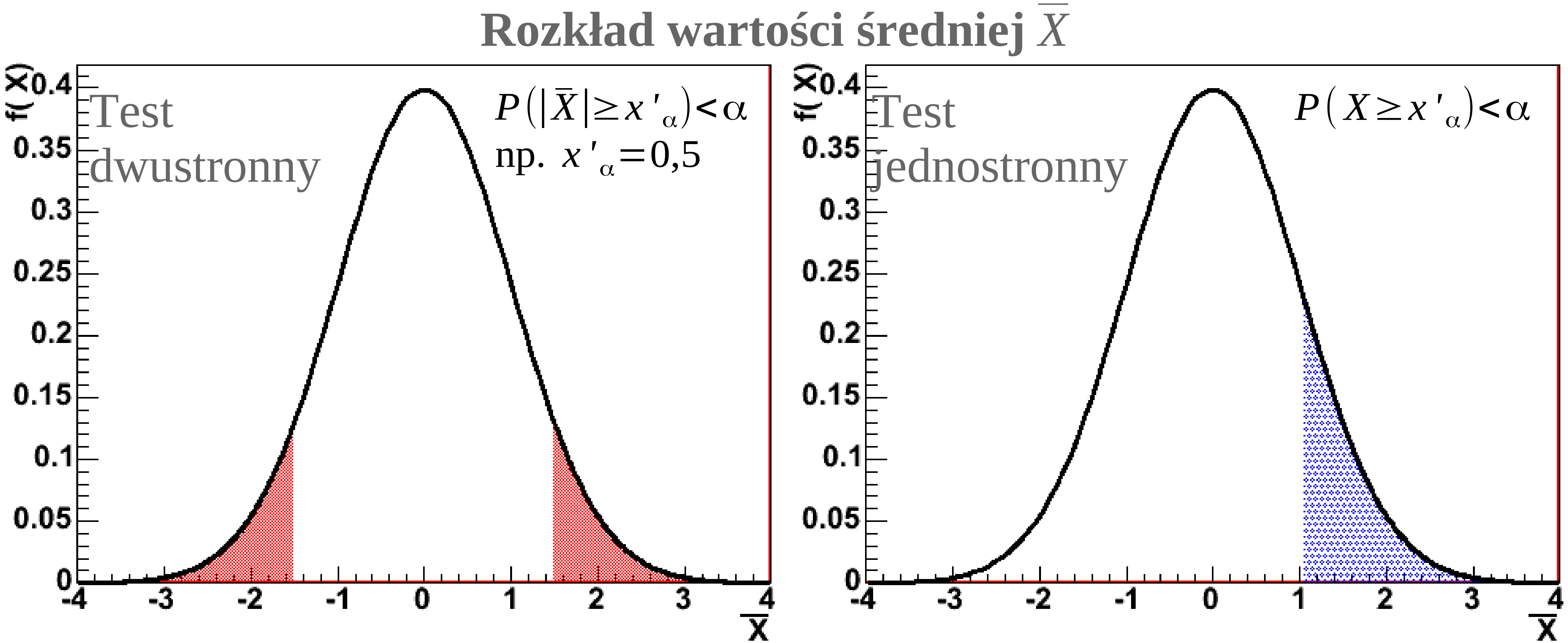
- **Przykład:** rozważamy zmienną losową X opisaną standardowym rozkładem Gaussa (średnia 0, odchylenie 1). Pobieramy 10-elementową próbę, uzyskaliśmy średnią arytmetyczną: $\bar{X}=0{,}5$
- Jak na podstawie tej jednej realizacji próby (np. wyniku eksperymentu) możemy stwierdzić, czy pochodzi ona z takiej populacji? Innymi słowy, naszą **hipotezą** jest: **próba losowa pochodzi z rozkładu Gaussa o średniej 0 i odchyleniu 1**
- Procedura weryfikacji hipotezy nazywana jest **testem statystycznym**
- **Jeżeli hipoteza jest słuszna (nasze założenie)** to <u>wartość średnia</u> (będąca również zmienną losową) $\bar{X}$ ma rozkład normalny ze średnią 0 i odchyleniem std. $1/\sqrt{10}$

$$\sigma^2(\bar{X})=\frac{1}{n}\sigma^2(X)=\frac{1}{10}\cdot 1\Rightarrow\sqrt{\sigma^2(\bar{X})}=\frac{1}{\sqrt{10}}$$

Weryfikacja hipotez statystycznych

- Jak na podstawie **konkretnej realizacji próby** sprawdzić, czy założona hipoteza jest prawdziwa?
  - **I:** musimy ustalić pewną wartość prawdopodobieństwa α (zwanego **poziomem istotności**,z reguły mała wartość, np. 0,01, albo 0,03, czy 0,05)
  - **II:** pytamy, czy prawdopodobieństwo zaobserwowania określonych wartości próby jest mniejsze niż α: $P(|\bar{X}| \geq 0{,}5) < \alpha$
  - **nierówność <u>spełniona</u>** – jest mało prawdopodobne, aby próba pochodziła z rozkładu określonego przez testowaną hipotezę → **możemy ją odrzucić**
  - **prawdopodobieństwo zaobserwowania tego, że $|\bar{X}|$ jest duże, jest <u>bardzo małe</u>, ale takie nam się trafiło – więc prawdopodobnie (z prawdodobieństwem $1\text{-}\alpha$) nasza hipoteza nie jest słuszna**
  - **III:** jeśli prawdopodobieństwo jest mniejsze niż przyjęta wartość prawdopodobieństwa (poziom istotności) α, odrzucamy hipotezę na zadanym poziomie istotności

Weryfikacja hipotez statystycznych

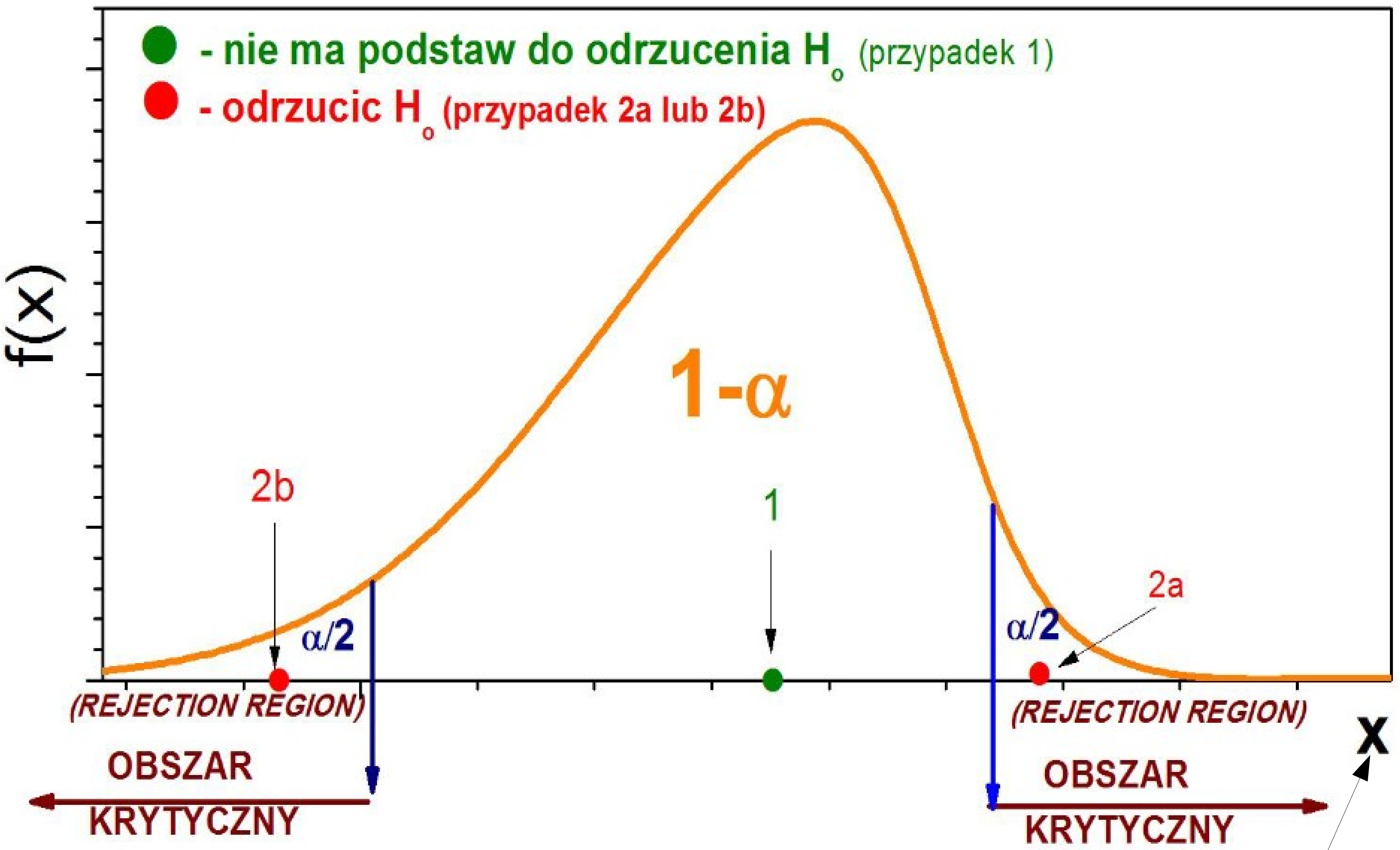


- Jeśli (w naszym przykładzie) wartość średnia znajduje się w zaznaczonym obszarze (nazywamy go **obszarem krytycznym**), to hipotezę odrzucamy
  - jeśli oczekujemy rozkładu normalnego o średniej 0 i małym odchyleniu (np. 10), a z próby losowej (konkretny eksperyment) mamy średnią 1000, to lądujemy w **“ogonie”** rozkładu średniej i na podstawie tej konkretnej próby **odrzucamy** hipotezę (**ale na podstawie innej próby moglibyśmy zaakceptować**)

Weryfikacja hipotez statystycznych

- W ogólnym przypadku używamy innych wielkości niż średnia:
  - definiujemy jakąś (wygodną dla nas) **statystykę testową** T (np. różnicę między wynikiem eksperymentu a krzywą teoretyczną)
  - ustalamy poziom istotności α
  - wyznaczamy taki zbiór U, który określa obszar zmienności statystyki testowej T, taki że prawdopodobieństwo znalezienia się w nim jest ograniczone wartością α: $P(T \in U) = \alpha$
  - z pobranej próby wyznaczamy konkretną wartość statystyki testowej T': jeżeli znajduje się ona **wewnątrz** obszaru krytycznego U, **odrzucamy hipotezę** (mówimy: krzywa teoretyczna nie opisuje wyniku eksperymentu), czyli odrzucamy hipotezę, jeżeli $T' \in U$

Weryfikacja hipotez statystycznych



rozkład statystyki testowej

Hipoteza zerowa i alternatywna

- Każde testowanie hipotez zaczynamy od założenia, że **nie ma związku** między zmiennymi lub nie ma różnic między porównywanymi grupami (np. nie ma różnicy między teorią a danymi)

Hipoteza zerowa (H_0) zawsze mówi o <u>braku związku</u> między zmiennymi lub o braku różnicy.

- **Hipoteza alternatywna (badawcza)**, H_A, stawiana jest przez nas i dotyczy sytuacji wytłumaczenia zjawiska, ***kiedy hipoteza zerowa jest fałszywa***
  - z metodologicznego punktu widzenia nie jest możliwe, aby w pełni udowodnić jej prawdziwość.
  - wystarczy, ze pojawi się <u>jeden</u> przypadek zaprzeczający hipotezie i staje się ona fałszywa (odnosi się to również do teorii).

Hipotezy w badaniach <u>falsyfikujemy</u> a nie potwierdzamy.

(odrzucamy hipotezę zerową na rzecz alternatywnej, ***lub***
nie stwierdzamy podstaw do odrzucenia hipotezy zerowej)

Hipoteza zerowa i alternatywna

- Hipoteza może być różna, przykłady hipotez:
  - średni wzrost Polaków to 175 cm
  - 2% dzieci w wieku szkolnym nie lubi czekolady
  - poziom szczęścia dwie minuty po zjedzeniu dużej porcji lodów jest wyższy niż przed zjedzeniem tejże dużej porcji lodów
- **Hipoteza zerowa** (H_0) to w uproszczeniu taka, gdy nie widzimy różnicy – np. czujemy się tak samo po zjedzeniu lodów jak przed
  - w przypadku pomiarów, np. wartość χ2 jest mała → teoria opisuje dane
- **Hipoteza alternatywa** (H_A) to przeciwieństwo hipotezy zerowej, którą <u>możemy zdefiniować na kilka sposobów</u>, np.:
  - np. jest różnica w szczęściu w zjedzeniu lodów (test dwustronny)
  - poziom szczęścia po zjedzeniu lodów jest mniejszy (test jednostronny)
  - poziom szczęścia po zjedzeniu lodów jest większy (test jednostronny)
  - w przypadku pomiarów χ2 jest duże → teoria nie opisuje danych (test jednostronny – rozkład χ2 jest niesymetryczny)
- **Statystyka testowa** - to funkcja próby, na podstawie której wnioskuje się o odrzuceniu lub nie hipotezy statystycznej – wielkość mająca swój rozkład prawd.

Test dobroci χ^2 dopasowania

Test χ^2 dobroci dopasowania

- Mamy N pomiarów g_i, $i=1, 2, \ldots, N$ oraz ich niepewności σ_i
- Wartości $f_i, \; i=1,2,\ldots,N$ określają nam prawdziwy rozkład danej wielkości mierzonej (**np. znaleziony poprzez estymację**)
- **Dla każdego pomiaru liczymy wielkość** u_i**:** $u_i = \dfrac{g_i - f_i}{\sigma_i}, \; i=1,2,\ldots,N$
- Jeśli nasza teoria (wartości f_i) jest prawdziwa, to rozkłady różnic u_i mają postać **standardowego** rozkładu normalnego **– nasza hipoteza**
- Jeśli tak, to rozkład χ^2 o N stopniach swobody będzie miała wielkość:

$$T = \sum_{i=1}^{N} u_i^2 = \sum_{i=1}^{N} \left(\frac{g_i - f_i}{\sigma_i} \right)^2$$

- **(Subiektywnie)** oczekujemy małej wartości wielkości T
- Gdy hipoteza jest **fałszywa**, wówczas poszczególne różnice u_i przyjmują duże wartości (wartość T jest duża)
- Jak określić granicę zmienności T? Można zauważyć, że granica ta jest określona **kwantylem** $\chi^2_{1-\alpha}$, czyli: $\boxed{P(T > \chi^2_{1-\alpha}) = \alpha}$

$$\boxed{\begin{aligned} F(x_q) = P(X \leq x_q) = q \\ P(X > x_q) = 1 - q \end{aligned}}$$

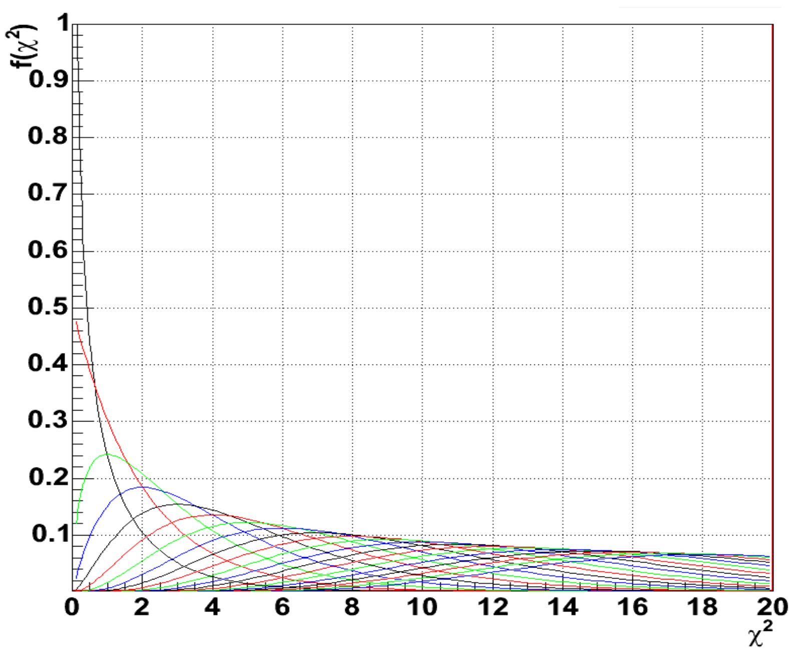
Test χ^2 dobroci dopasowania

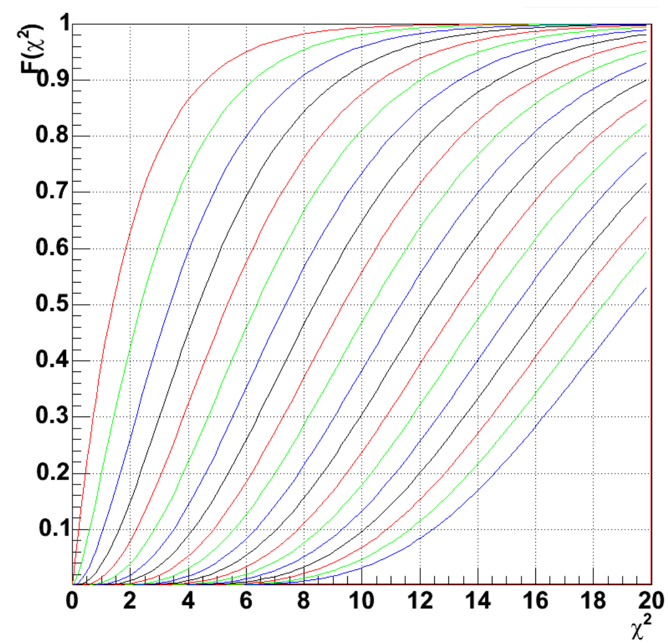
- **<u>Podsumowując</u>, w naszym przypadku musimy <u>dla danej realizacji próby (wyniku eksperymentu)</u> wyznaczyć wartość testową T i porównać ją z odpowiednim kwantylem rozkładu** χ^2 **o odpowiedniej liczbie stopni swobody:**

$$\boxed{T > \chi^2_{1-\alpha}}$$

- **Jeżeli ten warunek jest spełniony, to hipotezę odrzucamy** (punkty teoretyczne nie opisują danych eksperymentalnych na zadanym poziomie istotności)

- Skąd wziąć kwantyl? Z tablic lub z dystrybuanty:





Test χ^2 dobroci dopasowania

Wyobraźmy sobie, że rzucamy monetą 100 razy (N=100).

Prawdopodobieństwo uzyskania orła lub reszki wynosi 50%, a więc wartość oczekiwana wygląda następująco:

	Orzeł	Reszka
Wartość oczekiwana (N)	50	50

My jednak otrzymaliśmy: 52 orły i 48 reszek.

Czy nasze wyniki są zgodne z modelem teoretycznym (wartością oczekiwaną)?
W celu sprawdzenia przeprowadzamy test chi kwadrat

Test χ^2 dobroci dopasowania

PRZYKŁAD 1 – RZUT MONETĄ

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

- $\chi 2$ - wartość statystyki 'chi kwadrat'
- O - częstość obserwowana - (observed frequency)
- E - częstość oczekiwana - (expected frequency)

OBLICZANIE CHI KWADRAT - KROKI

1) Odejmujemy oczekiwane frekwencje od obserwowanych
2) Podnosimy do kwadratu
3) Dzielimy przez wartość oczekiwaną
4)Liczymy stopnie swobody (w tym wypadku liczba kategorii (orzeł/reszka) -1)

	Orzeł	Reszka	Suma
Wartość otrzymana (O)	52	48	100
Wartość oczekiwana (E)	50	50	50
Różnica (O-E)	2	-2	0
Kwadrat różnicy (O-E)^2	4	4	0
Iloraz kwadratu różnicy $\frac{(O-E)^2}{E}$	4/50=0,08	4/50=0,08	0,16

Chi kwadrat = 0,16

Df = 2-1=1

Test χ^2 dobroci dopasowania

Porównujemy nasz wynik z wartościami krytycznymi rozkładu chi kwadrat (odcinającymi obszar krytyczny oznaczany przez alfę, tj. obszary odrzucenia/nieodrzucenia) odpowiadającym liczbie stopni swobody i poziomowi prawdopodobieństwa z jakim chcemy wnioskować.

Tablica wartości krytycznych dla wartości chi kwadrat

n/α	0,995	0,99	0,98	0,975	0,97	0,96	0,95	0,9	0,1	0,05	0,04	0,03	0,025	0,02	0,01	0,005
1	0,0000	0,0002	0,0006	0,0010	0,0014	0,0025	0,0039	0,0158	2,7055	$\boxed{3,8415}$	4,2179	4,7093	5,0239	5,4119	6,6349	7,8794
2	0,0100	0,0201	0,0404	0,0506	0,0609	0,0816	0,1026	0,2107	4,6052	5,9915	6,4378	7,0131	7,3778	7,8240	9,2103	10,5966
3	0,0717	0,1148	0,1848	0,2158	0,2451	0,3002	0,3518	0,5844	6,2514	7,8147	8,3112	8,9473	9,3484	9,8374	11,3449	12,8382
4	0,2070	0,2971	0,4294	0,4844	0,5351	0,6271	0,7107	1,0636	7,7794	9,4877	10,0255	10,7119	11,1433	11,6678	13,2767	14,8603
5	0,4117	0,5543	0,7519	0,8312	0,9031	1,0313	1,1455	1,6103	9,2364	11,0705	11,6443	12,3746	12,8325	13,3882	15,0863	16,7496
6	0,6757	0,8721	1,1344	1,2373	1,3296	1,4924	1,6354	2,2041	10,6446	12,5916	13,1978	13,9676	14,4494	15,0332	16,8119	18,5476
7	0,9893	1,2390	1,5643	1,6899	1,8016	1,9971	2,1673	2,8331	12,0170	14,0671	14,7030	15,5091	16,0128	16,6224	18,4753	20,2777
8	1,3444	1,6465	2,0325	2,1797	2,3101	2,5366	2,7326	3,4895	13,3616	15,5073	16,1708	17,0105	17,5345	18,1682	20,0902	21,9550
9	1,7349	2,0879	2,5324	2,7004	2,8485	3,1047	3,3251	4,1682	14,6837	16,9190	17,6083	18,4796	19,0228	19,6790	21,6660	23,5894
10	2,1559	2,5582	3,0591	3,2470	3,4121	3,6965	3,9403	4,8652	15,9872	18,3070	19,0207	19,9219	20,4832	21,1608	23,2093	25,1882
11	2,6032	3,0535	3,6087	3,8157	3,9972	4,3087	4,5748	5,5778	17,2750	19,6751	20,4120	21,3416	21,9200	22,6179	24,7250	26,7568
12	3,0738	3,5706	4,1783	4,4038	4,6009	4,9385	5,2260	6,3038	18,5493	21,0261	21,7851	22,7418	23,3367	24,0540	26,2170	28,2995
13	3,5650	4,1069	4,7654	5,0088	5,2210	5,5838	5,8919	7,0415	19,8119	22,3620	23,1423	24,1249	24,7356	25,4715	27,6882	29,8195
14	4,0747	4,6604	5,3682	5,6287	5,8556	6,2426	6,5706	7,7895	21,0641	23,6848	24,4855	25,4931	26,1189	26,8728	29,1412	31,3193
15	4,6009	5,2293	5,9849	6,2621	6,5032	6,9137	7,2609	8,5468	22,3071	24,9958	25,8162	26,8479	27,4884	28,2595	30,5779	32,8013

(…)

0,16<3,8415 → nie ma podstaw do odrzucenia hipotezy zerowej

Test χ^2 i doświadczalny rozkład częstości

- Możemy również rozważać zmienną losową X, (opisaną rozkładem $f(x)$) którą dzielimy na r przedziałów (histogram): $\xi_1, \xi_2, \dots, \xi_i, \dots, \xi_r$
- Całkując $f(x)$ w przedziałach otrzymujemy prawdopodobieństwo p_i zaobserwowania zmiennej X w danym przedziale (binie):
$$p_i = P(x \in \xi_i) = \int_{\xi_i} f(x)\,dx \,; \quad \sum_{i=1}^{r} p_i = 1$$
- Z pobranej próby o liczebności n oznaczamy przez n_i elementy leżące w danym przedziale ξ_i
- Oczywiście zachodzi relacja: $n = \sum_{i=1}^{r} n_i$
suma wejść w poszczególnych binach równa jest liczebności próby
- **Oczekiwalibyśmy** (zakładając prawdziwość $f(x)$)**, że:** $n_i = n p_i$
- **Hipoteza:** zakładamy, że dla dużych wartości liczb n_i ich wariancja równa się n_i (patrz dyskusja o rozkładzie Poissona) i że rozkład wielkości u_i: $u_i^2 = \frac{(n_i - n\,p_i)^2}{n_i}, \;\; lub \;\; u_i^2 = \frac{(n_i - n\,p_i)^2}{n p_i}$ ma rozkład Gaussa

Test χ^2 i doświadczalny rozkład częstości

- Wtedy, suma kwadratów: $T=\sum_{i=1}^{r} u_i^2$
- Będzie miała (dla dużych n) rozkład χ^2
- **Jaka jest liczba stopni swobody?** Z defninicji histogramu mamy jedno **równanie więzów**: $n=\sum_{i=1}^{r} n_i$
- Zatem zmienne u_i **nie są niezależne**, więc liczba stopni swobody równa się r-1
- Oczywiście, jeżeli dodatkowo estymujemy p parametrów rozkładu na podstawie pomiarów (wprowadzamy p kolejnych wiezów uzależniających od siebie wielkości u_i), to liczba stopni swobody wynosi r-1-p
- Wartość T porównujemy, tak jak do tej pory, z kwantylami rozkładu χ^2 o określonej liczbie stopni swobody dla zadanego poziomu istotności α: $T>\chi^2_{1-\alpha}$
- **Jeśli nierówność jest spełniona – odrzucamy hipotezę**

Test χ^2 - przykład

Zadanie

Weryfikacja hipotez statystycznych (5 pkt.)

- Przeprowadzono eksperyment naświetlania wodorowej komory pęcherzykowej wiązką fotonów w celu badania oddziaływań fotonów z protonami. Fotony powodują powstawanie par elektron-pozyton, które mogą być wykorzystane do monitorowania wiązki fotonów. Częstość występowania zdjęć z 0,1,2,... parami elektron-pozyton powinna podlegać rozkładowi Poissona. Należy wczytać dane z pliku (w pierwszej kolumnie znajduje się liczba par elektronowych na zdjęciu k, a w drugiej liczba zdjęć zawierających k par elektronowych). Widzimy, że rozkład ten przypomina rozkład Poissona - próbujemy zatem obliczyć estymator największej wiarygodności dla parametry rozkładu Poissona (patrz Wykład 10 slajd 13) (1 pkt.)
- Narysować na jednym wykresie punkty pomiarowe i dopasowanie (metodą estymatora największej wiarygodności).
- Sprawdzić jakość dopasowania za pomocą testu χ2. W tym celu należy zaimplementować funkcję obliczającą statystykę testową χ2 zgodnie z wzorem $T = \sum_k \frac{(n_k - np_k)^2}{np_k}$

gdzie: nk - liczba obserwacji w k-tym binie, npk - przewidywana przez teorię liczba przypadków w k-tym binie

- Określić liczbę stopni swobody i obliczyć wartość statystyki testowej. (1 pkt.)
- Zaimplementować funkcję zwracającą wynik testu χ2 na zadanym poziomie istotności α

Wykorzystując zaimplementowaną funkcję zweryfikować hipotezę mówiacą, że dane pomiarowe podlegają rozkładowi Poissona. Dobrać odpowiednią wartość poziomu istotności. Uwaga! Kwanyl możemy odczytać z policzonej na ostatnich zajęciach dystrybuanty. (2 pkt.)

Test χ^2 - przykład

Zadanie

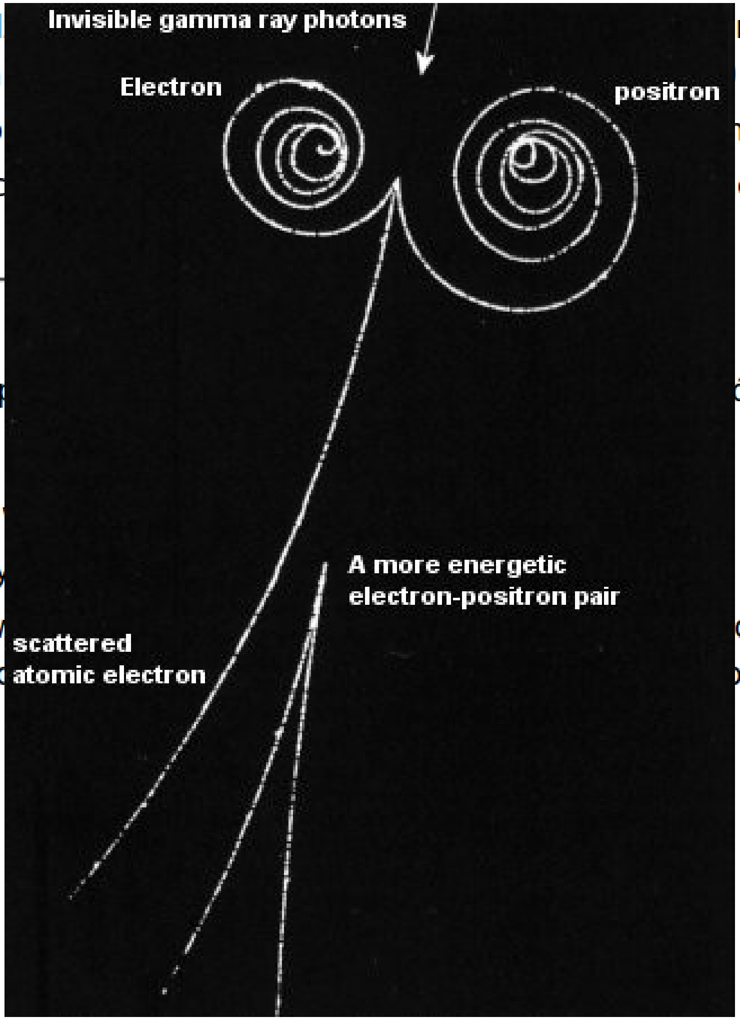
Weryfikacja hipotez statystycznych (5 pkt.)

- Przeprowadzono eksperyment naświetlania wodorowej komory pęcherzykowej wiązką fotonów w celu badania oddziaływań fotonów z protonami. Fotony powodują powstawanie par elektron-pozyton, które mogą być wykorzystane do monitorowania wiązki fotonów. Częstość występowania zdjęć z 0,1,2,... parami elektron-pozyton powinna podlegać rozkładowi Poissona. Należy wczytać dane z pliku (w pierwszej kolumnie znajduje się liczba par elektronowych na zdjęciu k, a w drugiej liczba zdjęć zawierających k par elektronowych). Wid na - próbujemy zatem obliczyć estymator największej wiarygodności dla slajd 13) (1 pkt.)
- Narysować na jednym wykresie punkty p największej wiarygodności).
- Sprawdzić jakość dopasowania za pomoc ć funkcję obliczającą statystykę testową χ2 zgodnie z wzorem $T = \sum_k \frac{(n_k$

gdzie: nk - liczba obserwacji w k-tym binie, np ów w k-tym binie

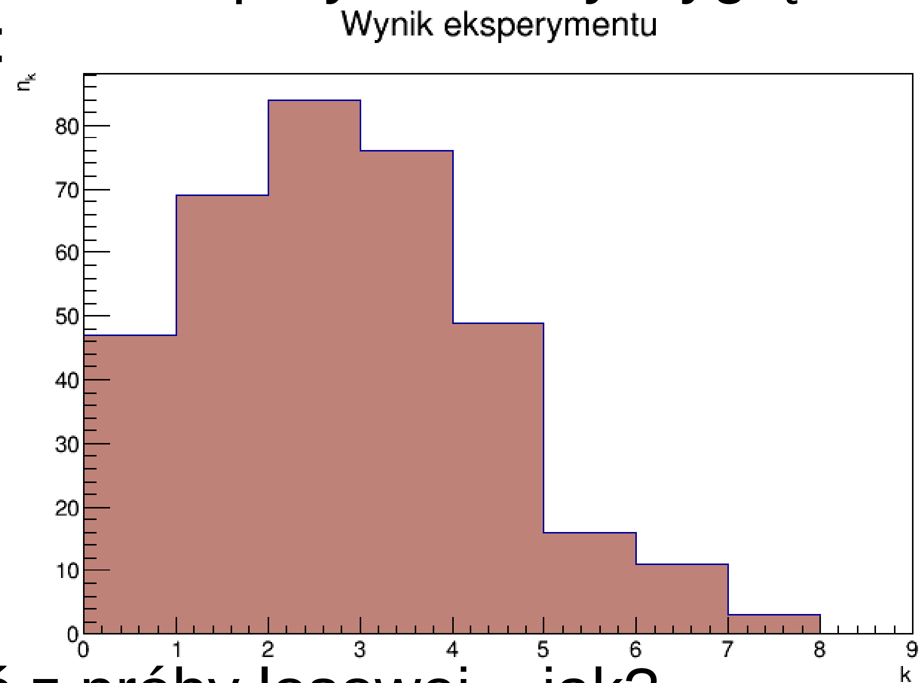
- Określić liczbę stopni swobody i obliczyć
- Zaimplementować funkcję zwracającą wy

Wykorzystując zaimplementowaną funkcję zw owe podlegają rozkładowi Poissona. Dobrać odpowiednią wartość poziomu istotn onej na ostatnich zajęciach dystrybuanty. (2 pkt.)



Test χ^2 - przykład

- Po wczytaniu danych z pliku histogram eksperymentalny wygląda następująco (nasza próba losowa):



- **Zakładamy hipotezę:** teoria mówi to jest rozkład Poissona („na oko" zresztą tak wygląda)

- Rozkład Poissona ma tylko jeden parametr (wartość średnią):

$$f(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

- Musimy go zatem jakoś wyznaczyć z próby losowej – jak? **Na przykład metodą największej wiarygodności** – szukamy estymatora nieobciążonego największej wiarygodności o minimalnej wariancji – wyprowadziliśmy go sobie na Wykładzie 10:

$$\frac{dl}{d\lambda} = l' = \sum_{j=1}^{N} \left\{ \frac{k^{(j)}}{\lambda} - 1 \right\} = \frac{1}{\lambda} \sum_{j=1}^{N} \left\{ k^{(j)} - \lambda \right\} = \boxed{\frac{N}{\lambda} (\bar{K} - \lambda)}$$

$$\boxed{\tilde{\lambda} = \bar{K}, \quad \sigma^2(\bar{K}) = \frac{N}{\lambda}}$$

Przypomnienie – definicja estymatora o min. wariancji:

$$\boxed{\begin{aligned} l' &= A(\lambda)(\tilde{\lambda} - \lambda) \\ \sigma^2(S) &= \frac{1}{|A(\lambda)|} \end{aligned}}$$

Test χ^2 - przykład

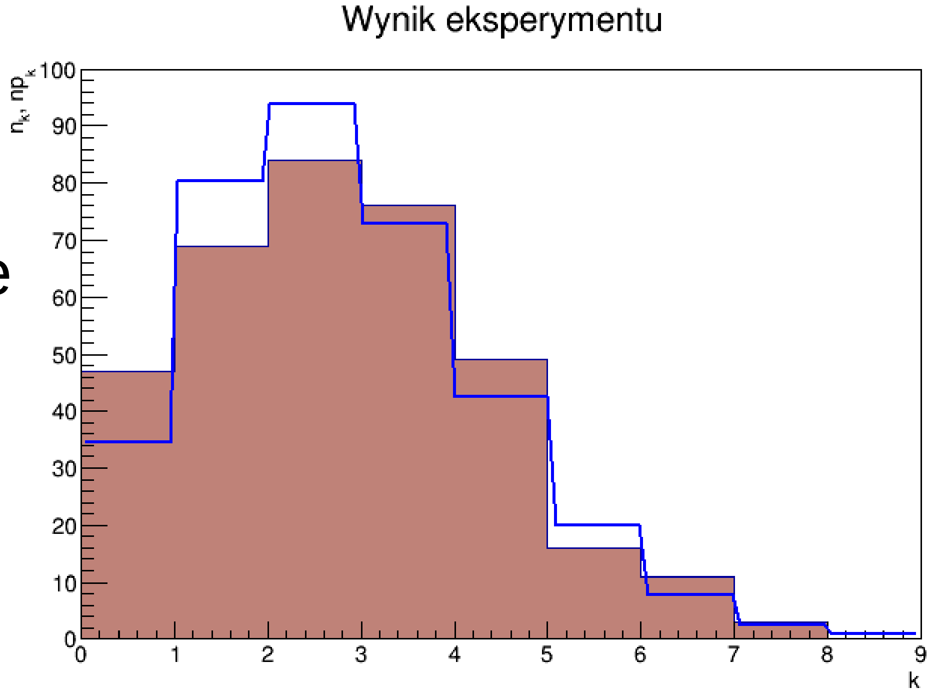
- Czyli estymatorem największej wiarygodności o minimalnej wariancji dla rozkładu Poissona jest średnia arytmetyczna z próby
- Oczywiście w naszym przypadku mamy histogram, który zawiera jakąś całkowitą liczbę wejść (całka z histogramu nie jest równa 1), wobec tego do średniej dodajemy wagi w postaci liczby wejść w danym binie i średnia staje się średnią ważoną:

$$\tilde{\lambda}=\frac{\sum_k k\cdot n_k}{\sum_k n_k}$$

- W naszym przypadku wartość ta wynosi mniej więcej: $\tilde{\lambda}\approx 2{,}33$
- Rysujemy więc funkcję:

$$n\cdot p_k=n\cdot f(k)=n\cdot\frac{\tilde{\lambda}^k}{k!}e^{-\tilde{\lambda}},\text{gdzie } n=\sum_k^7 n_k$$

- Jak teraz sprawdzić, czy faktycznie nasza hipoteza jest słuszna?
- **Testujemy dobroć dopasowania**



Test χ^2 - przykład

- Musimy zatem wyznaczyć wartość statystyki testowej T:

$$T=\sum_{k=0}^{7} u_i^2=\sum_{k=0}^{7} \frac{(n_k-n\,p_k)^2}{np_k}\approx 10{,}53$$

- Co dalej? Zakładamy poziom istotności, na przykład: $\alpha=0{,}01$
- Musimy jeszcze określić liczbę stopni swobody – ile ich jest?
  - liczba binów (8) minus 1 minus liczba parametrów (1)

$$r-1-p=8-1-1=6$$

- Teraz szukamy odpowiedniego kwantyla rozkładu χ^2 o 6 stopniach swobody: $\chi^2_{1-\alpha}=\chi^2_{0{,}99}\approx 16{,}81$
- Porównujemy statystykę z kwantylem: $T=10{,}51<\chi^2_{0{,}99}=16{,}81$
- **Warunek** $T>\chi^2_{1-\alpha}$ **nie jest spełniony, zatem na poziomie istotności** $\alpha=0{,}01$ **nie ma podstaw do odrzucenia hipotezy**

Test równości wariancji (test F-Fishera)

Test równości wariancji (F-Fishera)

- **Problem:** porównywanie wariancji populacji o jednakowych wartościach średnich
- **Przykład:** pomiar tej samej wielkości dwoma przyrządami pomiarowymi (zakładamy brak niepewności systematycznych – typu B)
- **Pytanie (hipoteza):** czy pomiary będą miały jednakowe wariancje (czy dokładność pomiaru jest jednakowa dla obu przyrządów)?
- Załóżmy, że rozważane populacje mają rozkład normalny
- Pobieramy próby o liczebności N_1 i N_2
- Dla każdej z pobranych prób wyznaczamy wariancję i liczymy iloraz

$$F = s_1^2 / s_2^2 \qquad s^2(X) = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2 \qquad s^2(\bar{X}) = \frac{1}{N \cdot (N-1)} \sum_{i=1}^{N} (X_i - \bar{X})^2 = s^2(X)/N$$

- Jeśli hipoteza o równości wariancji jest **prawdziwa**, to iloraz F powinien być bliski jedności $F \sim 1$

Test równości wariancji (F-Fishera)

- Można udowodnić, że tego typu wielkość ma rozkład F-Fishera

$$f(F)=\left(\frac{f_1}{f_2}\right)^{\frac{1}{2}f_1}\frac{\Gamma\left(\frac{1}{2}(f_1+f_2)\right)}{\Gamma\left(\frac{1}{2}f_1\right)\Gamma\left(\frac{1}{2}f_2\right)}F^{\frac{1}{2}f_1-1}\left(1+\frac{f_1}{f_2}F\right)^{-\frac{1}{2}(f_1+f_2)} \qquad f_1=N_1-1;\ \ f_2=N_2-1$$

- Szukamy zatem analogicznie wartości granicznej określającej obszar krytyczny, która jest odpowiednim kwantylem rozkładu F-Fischera:

$$P\left(\frac{s_1^2}{s_2^2}>F_{1-\alpha}\right)=\alpha$$

- Ostatecznie sprawdzamy zatem warunek: $\boxed{\frac{s_1^2}{s_2^2}>F_{1-\alpha}}$
- **To jest test jednostronny**, na ogół posługujemy jednak **testem dwustronnym**:

$$\boxed{\frac{s_g^2}{s_k^2}>F_{1-\alpha/2}(f_g,f_k),\text{gdzie f to liczby stopni swobody}}$$

Test równości wariancji (F-Fishera)

- Czyli w praktyce musimy zweryfikować hipotezę:

$$\frac{s_g^2}{s_k^2} > F_{1-\alpha/2}(f_g, f_k) = F''_{\alpha}(f_g, f_k)$$

- Indeksy g i k oznaczają większą i mniejszą wariancję z próby, czyli:

$$s_g^2 > s_k^2$$

- Jeżeli nierówność jest **spełniona**, to hipotezę o równości wariancji można **odrzucić**

Test równości wariancji – przykład

Numer pomiaru	Przyrząd 1	Przyrząd 2				
1	100	97	0,2	0,04	-2,86	8,16
2	101	101	1,2	1,44	1,14	1,31
3	102	102	2,2	4,84	2,14	4,59
4	100	99	0,2	0,04	-0,86	0,73
5	98	101	-1,8	3,24	1,14	1,31
6	97	98	-2,8	7,84	-1,86	3,45
7	100	101	0,2	0,04	1,14	1,31
8	101		1,2	1,44		
9	99		-0,8	0,64		
10	100		0,2	0,04		
Średnia	99,8	99,86				
Stopnie swobody	9	6				
S^2	19,6	20,86				
S^2/f	2,18	3,48				
F	1,6					

- Korzystamy z kwantyli funkcji F

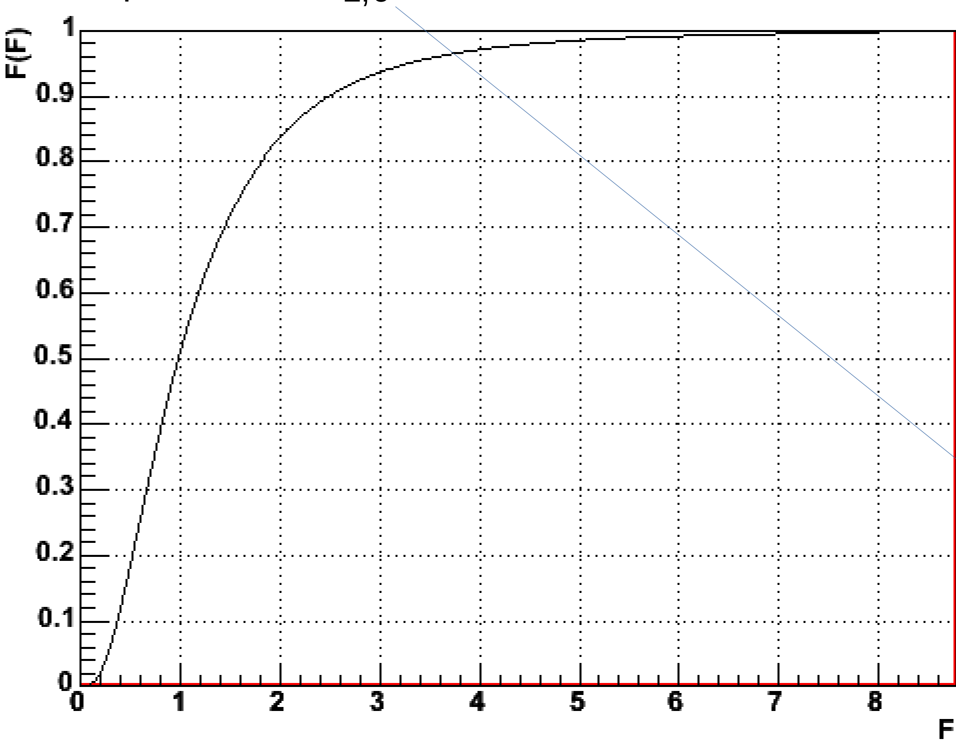
$$F''_{0,2}(6,9)=F_{0,9}(6,9)=2.51$$

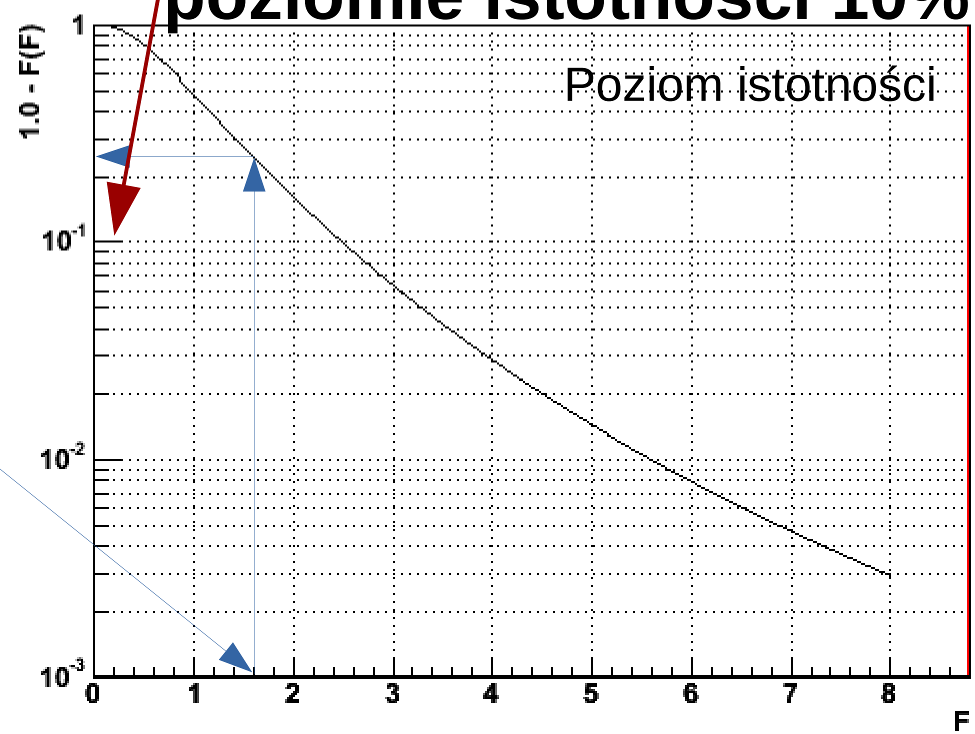
$$\boxed{F''_{0,1}(6,9)=F_{0,95}(6,9)=3.29}$$

$$F''_{0,02}(6,9)=F_{0,99}(6,9)=5.61$$

$$F''_{0,01}(6,9)=F_{0,995}(6,9)=6.89$$

- 1.6<3.29 – **nie ma podstaw do odrzucenia hipotezy na poziomie istotności 10%**





Porównanie wartości średnich (test t-Studenta)

Test równości średnich (t-Studenta)

- **Problem:** porównywanie wartości średnich dwóch prób losowych
- **Przykład:** badamy średni wzrost studentek 1 roku w Warszawie (populacja X) oraz w Nowym Jorku (populacja Y)
- **Pytanie (hipoteza):** czy wartości średnie obu populacji, na podstawie pobranych prób losowych, są jednakowe?
- Tak postawiona hipoteza cicho zakłada, że X i Y to te same populacje
- Powyższe rozważania możemy **uogólnić** na porównanie wartości średnich dwóch prób losowych z populacji X oraz Y o liczebnościach N_1 i N_2

Zastosowanie testu t-Studenta

- **Hipoteza:** równość wartości średnich z obu populacji: $\hat{x}=\hat{y}$
- Zakładamy (z centralnego twierdzenia granicznego), że wartości średnie z prób mają rozkład normalny z wariancjami średnich:

$$\sigma^2(\bar{X})=\sigma^2(X)/N_1, \quad \sigma^2(\bar{Y})=\sigma^2(Y)/N_2$$

- Wariancje średnich są estymowane przez estymatory:

$$s_{\bar{X}}^2=\frac{1}{N_1(N_1-1)}\sum_{j=1}^{N_1}(X-\bar{X})^2 \qquad s_{\bar{Y}}^2=\frac{1}{N_1(N_1-1)}\sum_{j=1}^{N_1}(Y-\bar{Y})^2$$

- Różnica wartości średnich z próby również ma rozkład zbliżony do normalnego: $\Delta=\bar{X}-\bar{Y}\Rightarrow\sigma^2(\Delta)=\sigma^2(\bar{X})+\sigma^2(\bar{Y})$
- Jeśli hipoteza jest prawdziwa, wówczas oczywiste jest, że $\hat{\Delta}=0$ oraz iloraz $\Delta/\sigma(\Delta)$ powinien podlegać rozkładowi Gaussa

Test różnic t-Studenta

- Skoro tak, to oczywiście $\sigma^2(X)=\sigma^2(Y)$, zatem można je estymować za pomocą jednego estymatora jako średnią ważoną z dwóch prób:

$$s^2=\frac{(N_1-1)s_X^2+(N_2-1)s_Y^2}{(N_1-1)+(N_2-1)}$$

- Wtedy możemy zdefiniować estymatory:

$$s_{\bar{X}}^2=s^2/N_{1,}\quad s_{\bar{Y}}^2=s^2/N_{2,}\quad s_\Delta^2=s_{\bar{X}}^2+s_{\bar{Y}}^2=\frac{N_1+N_2}{N_1\cdot N_2}s^2$$

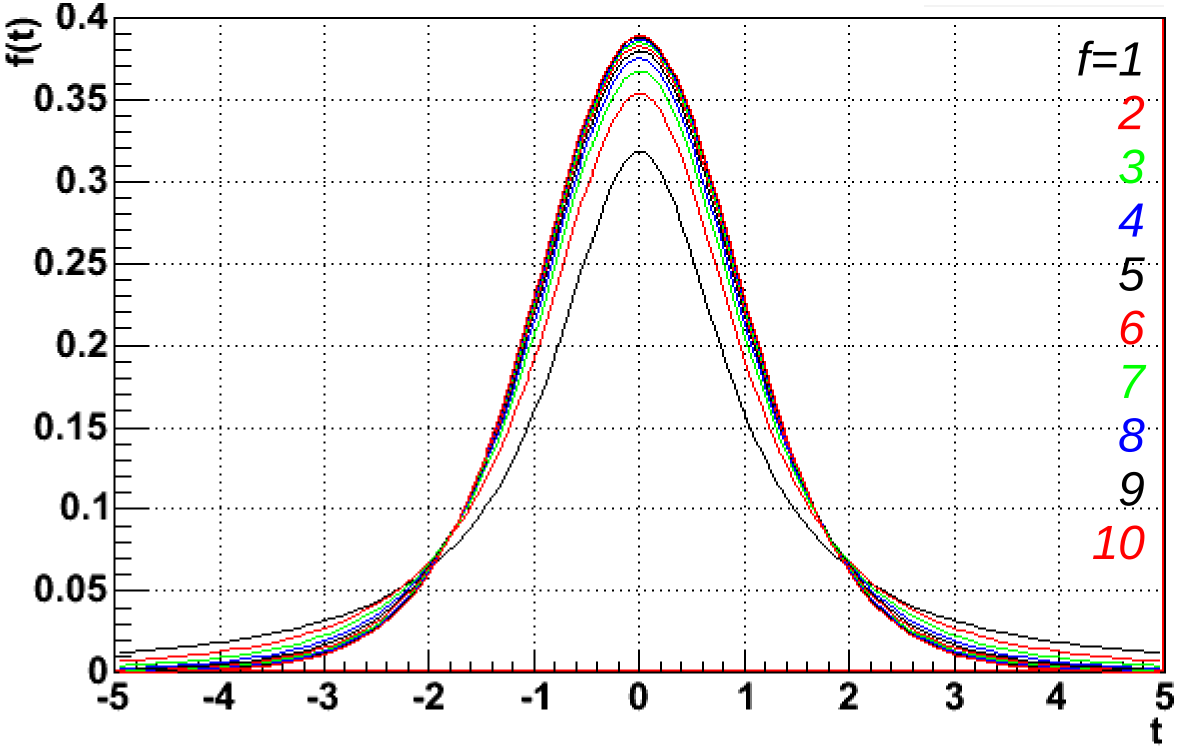
- Można udowodnić, że zmienna $\Delta/s(\Delta)$ podlega **rozkładowi t-Studenta** z liczbą stopni swobody $f=N_1+N_2-2$

- Równość wartości średnich można więc weryfikować posługując się **testem różnic Studenta**

- $\Delta/s(\Delta)$ obliczana jest na podstawie wyników dwóch prób. Jej wartość bezwzględną porównujemy z kwantylem rozkładu Studenta o liczbie stopni swobody f dla ustalonego poziomu istotności α. Sprawdzamy nierówność (**spełniona – odrzucamy hipotezę**):

$$|t|=\frac{|\Delta|}{s_\Delta}=\frac{|\bar{X}-\bar{Y}|}{S_\Delta}>t'_\alpha=t_{1-\frac{1}{2}\alpha}$$

Rozkład t-Studenta



- Jak widać, rozkład t-Studenta jest symetryczny względem 0
- Rozkład dąży do rozkładu Gaussa gdy $f \to \infty$
- Z symetrii względem 0 mamy związek (analogicznie jak dla Gaussa): $P(|t| \leq t) = 2F(|t|) - 1$

- Możemy wyznaczyć graniczne wartości $\pm t'_\alpha$ odpowiadające poziomowi istotności α poprzez całkę:

$$\int_0^{t'_\alpha} f(t)\,dt = \frac{1}{2}(1-\alpha), \quad gdzie \quad t'_\alpha = t_{1-\frac{1}{2}\alpha}$$

- Kwantyle $t'_\alpha = t_{1-\frac{1}{2}\alpha}$ są stablicowane dla różnych poziomów istotności α oraz liczby stopni swobody *f*

- Jest to dwustronny test t-Studenta (jednostronny analogicznie)

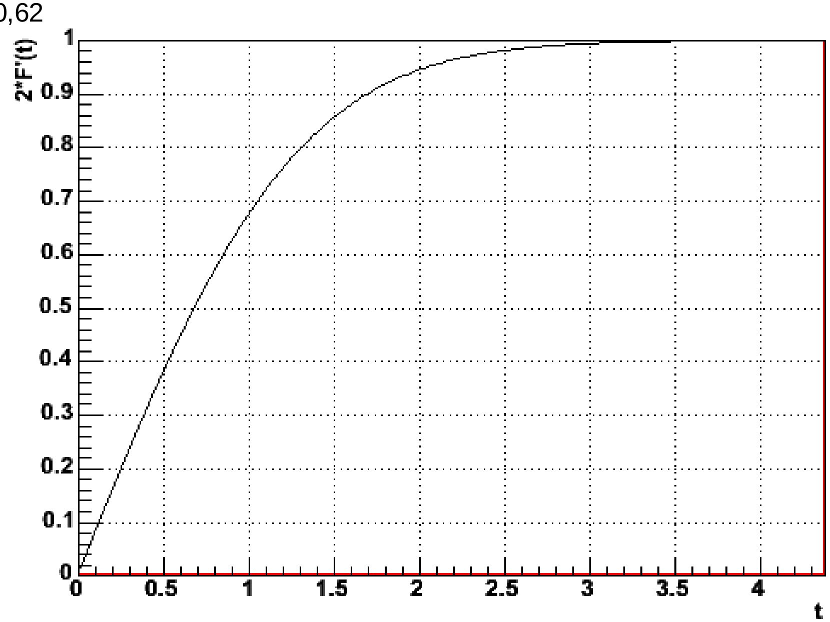
Test różnic t-Studenta - przykład

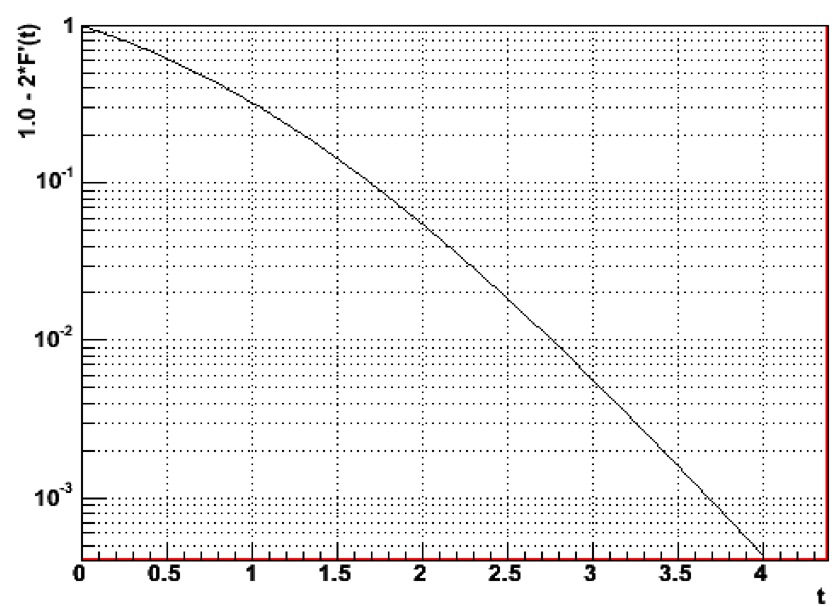
Numer pomiaru	Przyrząd 1	Przyrząd 2				
1	**100**	**97**	-0,21	0,04	-3,8	14,44
2	**101**	**101**	0,79	0,62	0,2	0,04
3	**102**	**102**	1,79	3,2	1,2	1,44
4	**100**	**99**	-0,21	0,04	-1,8	3,24
5	**98**	**101**	-2,21	4,89	0,2	0,04
6	**97**	**108**	-3,21	10,31	7,2	51,84
7	**100**	**101**	-0,21	0,04	0,2	0,04
8	**101**	**102**	0,79	0,62	1,2	1,44
9	**99**	**96**	-1,21	1,47	-4,8	23,04
10	**100**	**101**	-0,21	0,04	0,2	0,04
11	**98**		-2,21	4,89		
12	**101**		0,79	0,62		
13	**100**		-0,21	0,04		
14	**102**		1,79	3,2		
15	**103**		2,79	7,78		
16	**101**		0,79	0,62		
17	**99**		-1,21	1,47		
18	**100**		-0,21	0,04		
19	**102**		1,79	3,2		
Ilość pomiarów	19	10				
Średnia	**100,21**	**100,8**	**-0,59**			
Stopnie swobody	18	9				
S^2	43,16	95,6				
S^2/f	2,4	10,62				
S^2	49,1					
S^2 Delta	8,18					

- Mamy kwantyle:

$$t'_{0,2}(27)=t_{0,9}(27)=1.71$$
$$t'_{0,1}(27)=t_{0,95}(27)=2,05$$
$$t'_{0,02}(27)=t_{0,99}(27)=2,77$$
$$t'_{0,01}(27)=t_{0,995}(27)=3,05$$
$$t'_{0,004}(27)=t_{0,998}(27)=3,43$$
$$t'_{0,002}(27)=t_{0,999}(27)=3,69$$

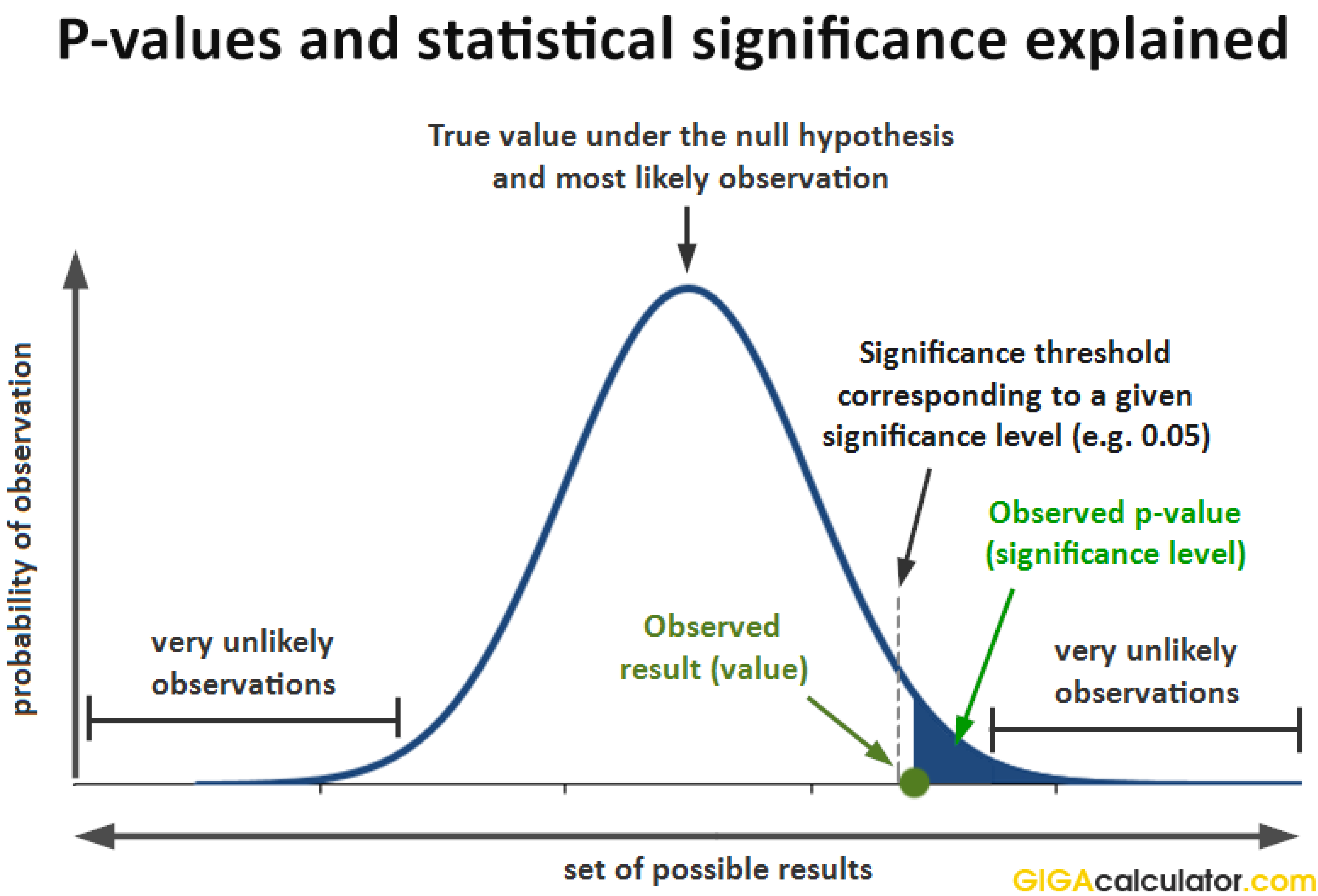
- **Hipotezy nie można odrzucić**





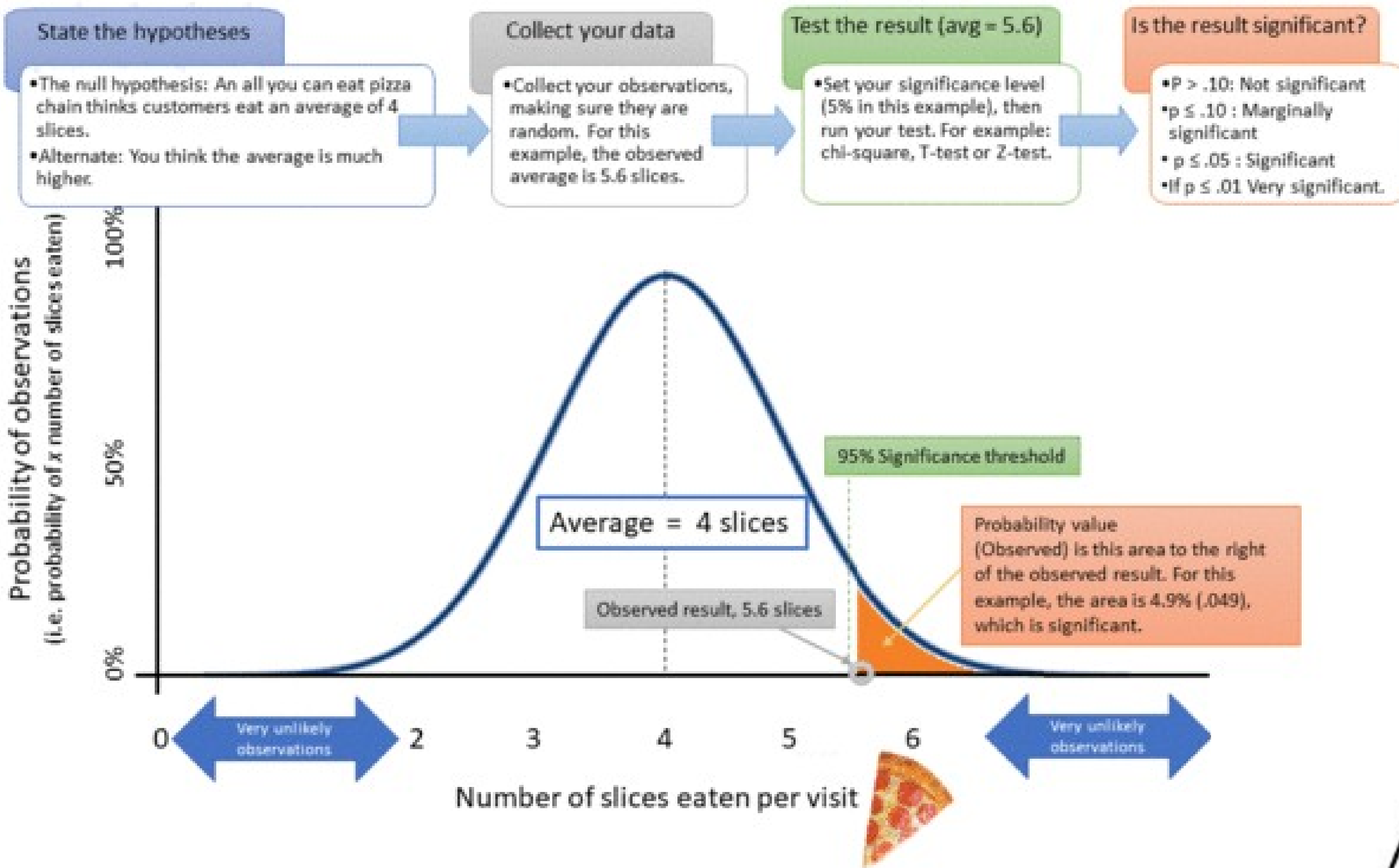
Wartość p
p-value

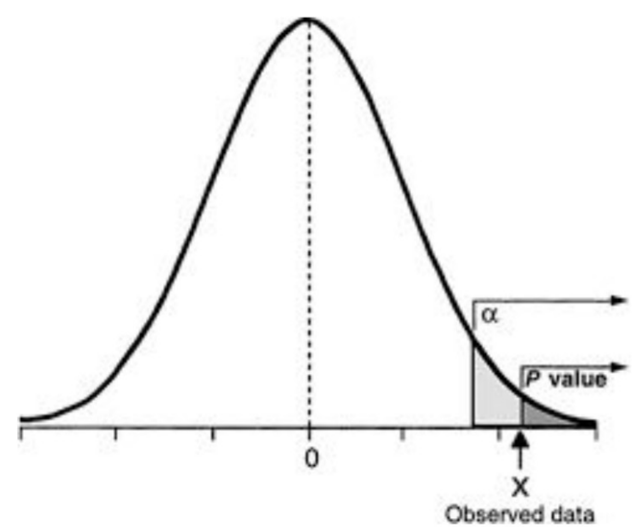
Wartość p



Wynik jest s**tatystycznie istotny (mała wartość p)**, zatem odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej (badawczej)

Wartość p


State the hypotheses
•The null hypothesis: An all you can eat pizza chain thinks customers eat an average of 4 slices.
•Alternate: You think the average is much higher.
Collect your data
•Collect your observations, making sure they are random. For this example, the observed average is 5.6 slices.
Test the result (avg = 5.6)
•Set your significance level (5% in this example), then run your test. For example: chi-square, T-test or Z-test.
Is the result significant?
•P > .10: Not significant
•p ≤ .10 : Marginally significant
• p ≤ .05 : Significant
•If p ≤ .01 Very significant.
Probability of observations
(i.e. probability of x number of slices eaten)
100%
50%
0%
95% Significance threshold
Average = 4 slices
Observed result, 5.6 slices
Probability value (Observed) is this area to the right of the observed result. For this example, the area is 4.9% (.049), which is significant.
0
Very unlikely observations
2
3
4
5
6
Very unlikely observations
Number of slices eaten per visit


α
P value
0
X
Observed data

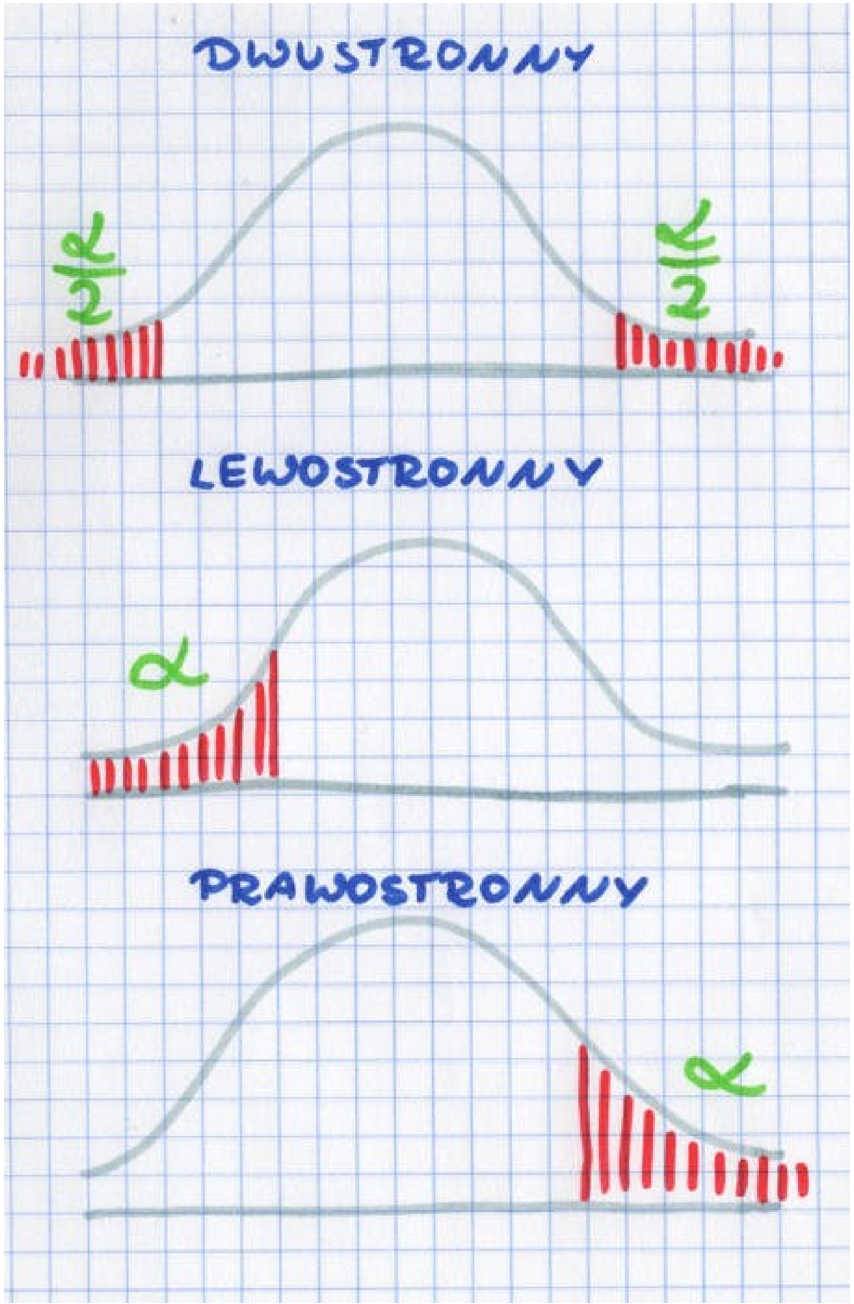
Hipotezy zerowa i alternatywna

Błędy I i II rodzaju

Na podstawie:
https://www.statystyczny.pl/hipotezy-statystyczne/

Statystyka testowa

- **Statystyka testowa** - to funkcja próby, na podstawie której wnioskuje się o odrzuceniu lub nie hipotezy statystycznej – wielkość mająca swój rozkład prawdopodobieństwa
- Z naszej próby losowej (eksperymentu) dostajemy jedną wartość – ona znajduje się gdzieś w tym rozkładzie
- Obszar krytyczny (obszar odrzuceń) jest zawsze na końcu rozkładu
  - jeśli hipoteza mówi, że coś jest różne – dwustronny
  - mniejsze lub większe – jednostronny
- **Statystyka testowa ma swój (różny) rozkład zarówno dla H_0 jak i H_A!!!**



Błąd I rodzaju

- **Błąd I rodzaju** to taki, gdzie odrzucamy hipotezę zerową a była ona prawdziwa
  - H_0: poziom szczęścia (<u>średnio w populacji</u>) po zjedzeniu dużej porcji lodów jest taki sam jak przed
  - H_A: poziom szczęścia (<u>średnio w populacji</u>) się zmienia
  - my na podstawie doświadczenia (<u>próby losowej</u>) zjedliśmy lody i np. przez te okropne wyrzuty sumienia, że znowu za dużo kalorii :) odrzucamy hipotezę zerową na rzecz alternatywnej
  - jeżeli w wyniku wielu prób losowych wynika, że jednak lądujemy w ogonie rozkładu owego szczęścia, to popełnimy właśnie błąd pierwszego rodzaju – bo odrzuciliśmy hipotezę zerową, która w rzeczywistości była prawdziwa
- Prawdopodobieństwo popełnienia błędu I rodzaju określa **poziom istotności** α, stąd oczekujemy, by był jak najmniejszy

Błąd II rodzaju

- **Błąd II rodzaju** to taki, gdy nie odrzucimy hipotezy w rzeczywistości fałszywej
  - ma on miejsce w sytuacji, kiedy jednak ten poziom szczęścia przed zjedzeniem lodów i po zjedzeniu w rzeczywistości się różni. Jeśli (**średnio w społeczeństwie, w populacji**) te wyrzuty sumienia z powodu zjedzenia dobrych i smacznych lodów powodują, że poziom szczęścia zdecydowanie spada po zjedzeniu lodów, a my stwierdzimy (**na podstawie próby losowej**), że nie ma żadnej różnicy, to wtedy popełniamy błąd II rodzaju. **Nie odrzucamy hipotezy zerowej, mimo że jest ona fałszywa.**
- **Prawdopodobieństwo** popełnienia błędu II rodzaju określamy jako β

Przykład: błędy I i II rodzaju

Z fabryki wykałaczek ma wyjechać partia produktów, która musi spełniać ustalone kryteria jakości (np. ostrość zakończenia x). Kontrola jakości polega na losowaniu pewnej liczby paczek, która jest sprawdzana pod kątem kryterium.

W kontroli jakości testujemy hipotezę: *cała partia spełnia wymagania jakości*

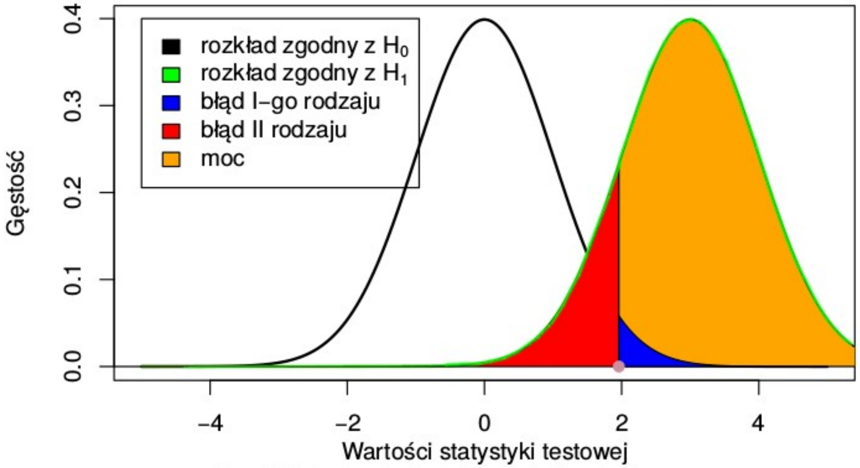
Wyobraźmy sobie, że w kontrolowanej partii znalazło się kilka paczek wadliwych.

Wariant A: Partia nie wyjeżdża z fabryki mimo, że duża jej część mogła być ok – błąd I rodzaju (odrzuca hipotezę zerową, ryzyko producenta).

Wariant B: Partia wyjedzie z fabryki mimo, a producent popełnia błąd II rodzaju (przyjmuje błędną hipotezę zerową) i konsument może trafić na nieodpowiednie wykałaczki (ryzyko konsumenta).

Moc testu

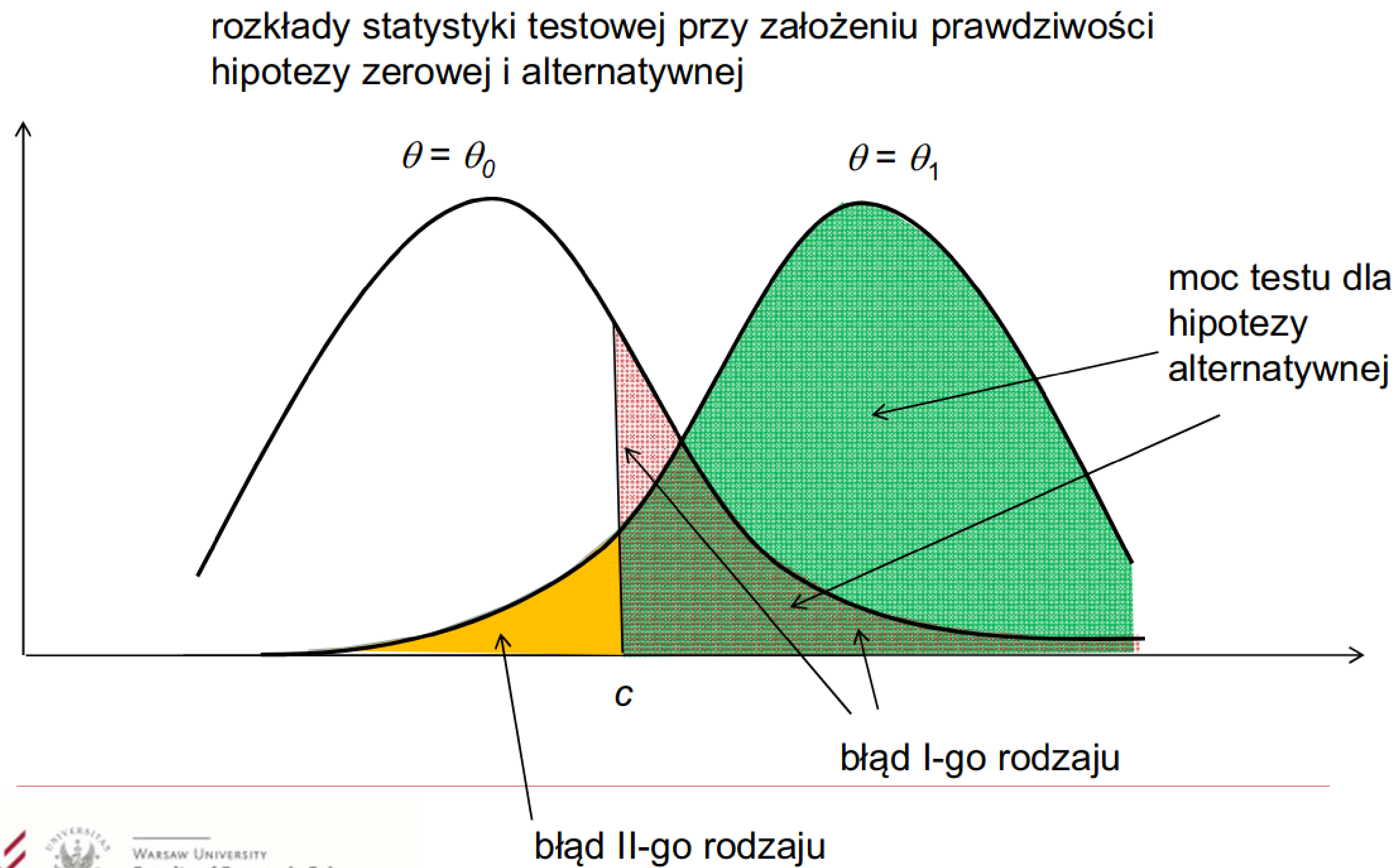
- Moc testu (prawdopodobieństwo, że prawidłowo odrzucimy hipotezę zerową) to 1-β. Inaczej mówiąc jest to prawdopodobieństwo niepopełnienia błędu II rodzaju.
- Moc testu zależy od kilku czynników:
  - Wielkości próby użytej w badaniu (im większa próba, tym większa moc testu).
  - Rzeczywistej wielkości efektu na tle losowej zmienności w populacji.
  - Przyjętego poziomu istotności α (między błędem I i II rodzaju jest taka zależność, że jeżeli zwiększamy prawdopodobieństwo popełnienia danego błędu, jednocześnie zmniejszamy je dla drugiego).



Rys.2 Relacje między błędem I, II rodzaju oraz mocą

Moc testu

Moc testu:
interpretacja graficzna (1)



Warsaw University
Faculty of Economic Sciences

Testy statystyczne

KONIEC