AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE

Podstawowe testy statystyczne

Statystyka

Dr inż. Janusz Majewski
Katedra Informatyki

Testy statystyczne

Zasadniczą domeną statystyki jest weryfikacja hipotez statystycznych, czyli pewnych przypuszczeń dotyczących rozkładu populacji.

Hipotezy:

- Parametryczne – dotyczą wartości parametrów rozkładu populacji.
- Nieparametryczne – dotyczą typu rozkładu populacji.

Proces weryfikacji hipotezy nosi nazwę testu statystycznego. Gdy weryfikacji podlega hipoteza parametryczna mówimy o teście parametrycznym.

Testy statystyczne

Hipoteza zerowa H_0 – podstawowa hipoteza statystyczna sprawdzana danym testem.

Hipoteza alternatywna H_1 – hipoteza konkurencyjna w stosunku do hipotezy zerowej (jeżeli odrzuca się hipotezę zerową, to przyjmuje się hipotezę alternatywną).

Błąd pierwszego rodzaju polega na odrzuceniu hipotezy podstawowej (zerowej), która jednak jest prawdziwa.

Błąd drugiego rodzaju polega na przyjęciu hipotezy podstawowej (zerowej), która jednak jest fałszywa.

Testy statystyczne

Błąd pierwszego rodzaju polega na odrzuceniu hipotezy podstawowej (zerowej), która jednak jest prawdziwa.

Błąd drugiego rodzaju polega na przyjęciu hipotezy podstawowej (zerowej), która jednak jest fałszywa.

Poziom istotności α – to prawdopodobieństwo popełnienia błędu pierwszego rodzaju (np. 0.05, 0.01, 0.1).

Moc testu $1-\beta$ - to prawdopodobieństwo niepopełnienia błędu drugiego rodzaju (β - oznacza prawdopodobieństwo popełnienia błędu drugiego rodzaju)

Testy statystyczne

Uwagi: testowanie hipotezy statystycznej nie jest tożsame z logicznym udowadnianiem jej prawdy lub fałszu. Hipoteza zerowa jest pewną naszą „teorią”. Jeśli „rzeczywistość” reprezentowana przez próbę nie zgadza się z teorią, to powinniśmy teorię (czyli hipotezę zerową) odrzucić. Czyniąc tak musimy wiedzieć, że bardzo rzadko, ale jednak czasem (z prawdopodobieństwem niewielkim α) hipoteza zerowa rzeczywiście jest słuszna, a tylko wynik próby przypadkowo odbiega od niej.

Testy statystyczne

H_0: **oskarżony jest niewinny**

H_1: **oskarżony jest winny**

Odbywa się przewód sądowy. Sędzia może:

1) uznać, że wina została udowodniona i wydać wyrok skazujący, tzn. przyjąć H_1 odrzucając H_0 (popełniając przy tym błąd z prawdopodobieństwem α),

2) uznać, że wina nie została udowodniona, co nie jest równoważne z przyjęciem, że oskarżony jest niewinny, tzn. może orzec, że nie ma podstaw do odrzucenia hipotezy H_0 (**co nie znaczy: przyjąć H_0**).

Testy statystyczne

Na ogół hipotezy zerowe nie są tymi, na których badaczowi szczególnie zależy. Formułowane są one zwykle w sposób „neutralny”, np. że nie ma różnic między dwiema populacjami, że nie ma związku między dwoma zjawiskami. Są one na ogół mało konstruktywne i nietwórcze dla badacza – eksperymentatora. Jemu właśnie zależy na wykazaniu istnienia różnic, związków i zależności wzajemnych. Badacz na ogół pragnie obalenia niechcianej neutralnej hipotezy zerowej i przyjęcia porządnej hipotezy alternatywnej. I do takiego właśnie postępowania znajduje w statystyce bardzo dobre narzędzie.

$\mu = \mu_0$

Testy statystyczne

Ogólny schemat przeprowadzenia parametrycznego testu statystycznego (test istotności) przedstawiono na schemacie blokowym z następnego slajdu.

Komentarze dotyczą testu istotności dla średniej:

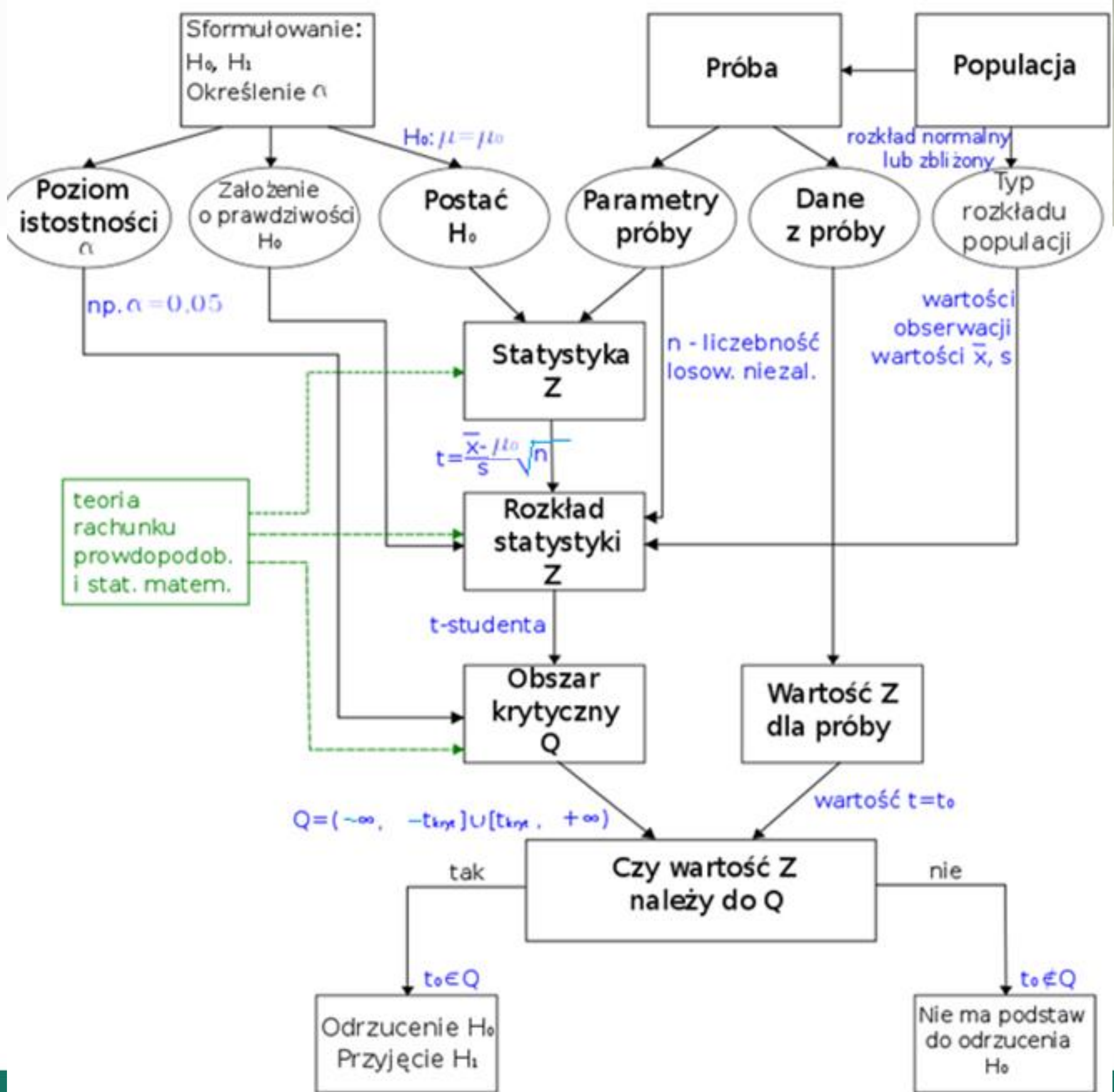
H_0: $\mu = \mu_0$ (μ_0 – wartość ustalona)

H_1: $\mu \neq \mu_0$

$\alpha = \alpha_0$ (α_0 – wartość ustalona)

Populacja o rozkładzie normalnym lub zbliżonym.

Próba o liczebności *n*.


Sformułowanie:
H₀, H₁
Określenie α
Próba
Populacja
H₀: μ = μ₀
rozkład normalny lub zbliżony
Poziom istotności α
Założenie o prawdziwości H₀
Postać H₀
Parametry próby
Dane z próby
Typ rozkładu populacji
np. α = 0.05
wartości obserwacji wartości x̄, s
Statystyka Z
n - liczebność losow. niezal.
t = (x̄ - μ₀)/s √n
teoria rachunku prowdopodob. i stat. matem.
Rozkład statystyki Z
t-studenta
Obszar krytyczny Q
Wartość Z dla próby
wartość t=t₀
Q=(-∞, -t_kryt]∪[t_kryt, +∞)
Czy wartość Z należy do Q
tak
nie
t₀∈Q
t₀∉Q
Odrzucenie H₀ Przyjęcie H₁
Nie ma podstaw do odrzucenia H₀

Testy statystyczne – procedura postępowania

1) Sformułowanie hipotezy H_0, hipotezy H_1, określenie poziomu istotności α.
2) Ewentualne sprawdzenie typu rozkładu lub innych założeń.
3) Zbudowanie statystyki Z i określenie rozkładu tej statystyki przy założeniu prawdziwości H_0 oraz uwzględnienie parametrów próby (np. liczebności) i ewentualnie parametrów populacji.
4) Określenie w rozkładzie statystyki Z (przy uwzględnieniu wartości α) tzw. obszaru krytycznego Q, tzn. zbioru takich wartości Z, które są bardzo mało prawdopodobne (prawdopodobieństwo α) jeżeli są spełnione wszystkie założenia, w szczególności założenie prawdziwości hipotezy H_0.
$$P(Z \in Q) = \alpha$$
5) Obliczenie wartości statystyki Z na podstawie próby.
6) Sprawdzenie, czy wyznaczone Z należy do obszaru krytycznego. Jeśli tak – odrzucamy H_0 przyjmując H_1, jeśli nie – nie ma podstaw do odrzucenia H_0.

Test istotności dla średniej

- Populacja ma rozkład normalny o średniej μ i wariancji σ^2 (μ i σ^2 – nieznane).
- Próba o liczebności n jest losowana niezależnie.
- Poziom istotności α.

H$_0$: $\boldsymbol{\mu = \mu_0}$ (μ_0 – konkretna liczba)
H$_1$: $\boldsymbol{\mu \neq \mu_0}$ (test dwustronny)

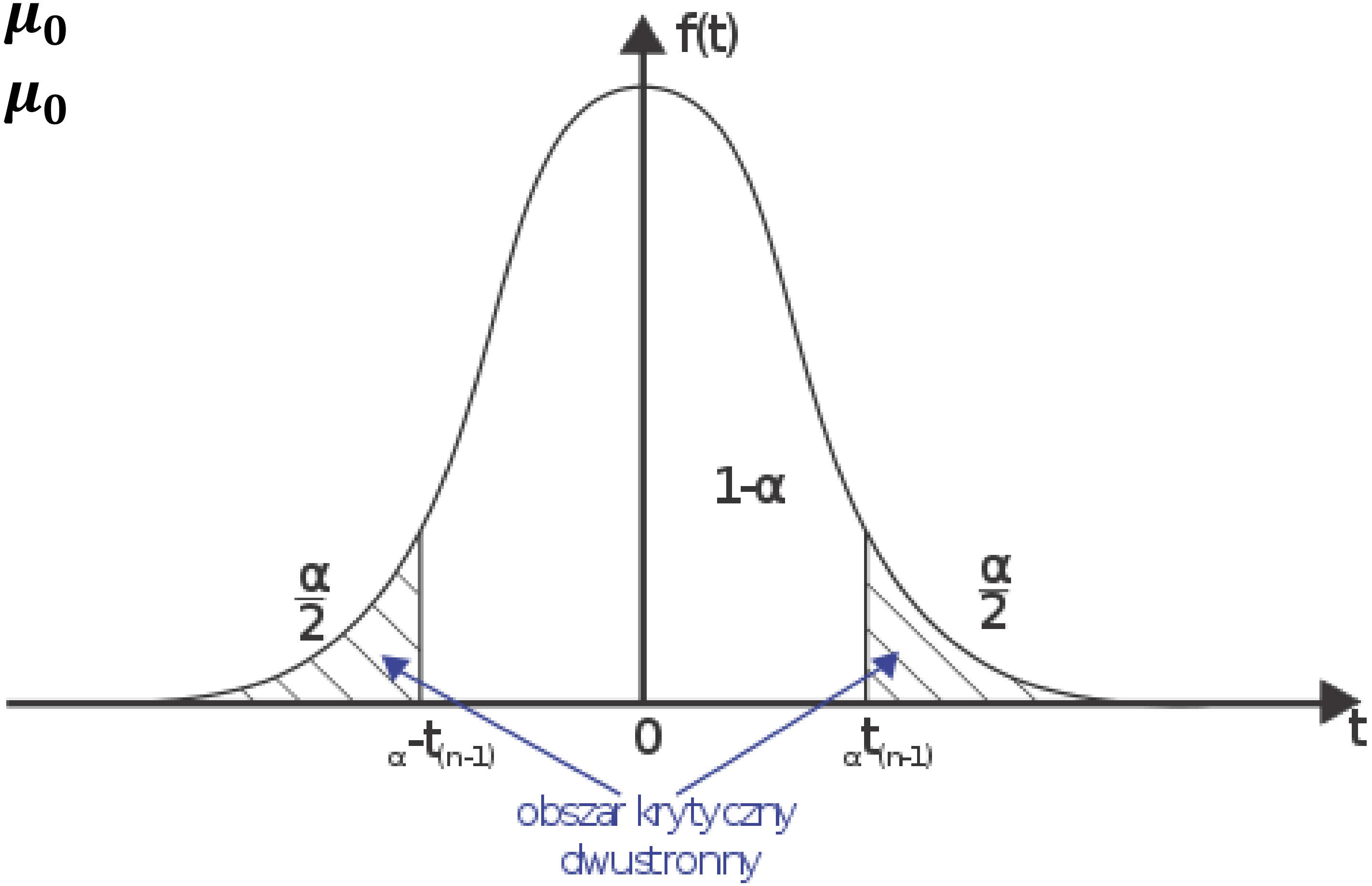
$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{\bar{x} - \mu_0}{s}\sqrt{n}$$

$\bar{x}$ – średnia z próby,
s - odchylenie standardowe z próby,
Rozkład t dla prawdziwej H$_0$.
Odrzucamy H$_0$ gdy $|t| \geq {}_{\alpha}\, t_{(n-1)}$
${}_{\alpha}t_{(n-1)}$ – wartość krytyczna rozkładu t dla danego α w teście dwustronnym

Test istotności dla średniej - dwustronny

H_0: $\boldsymbol{\mu = \mu_0}$

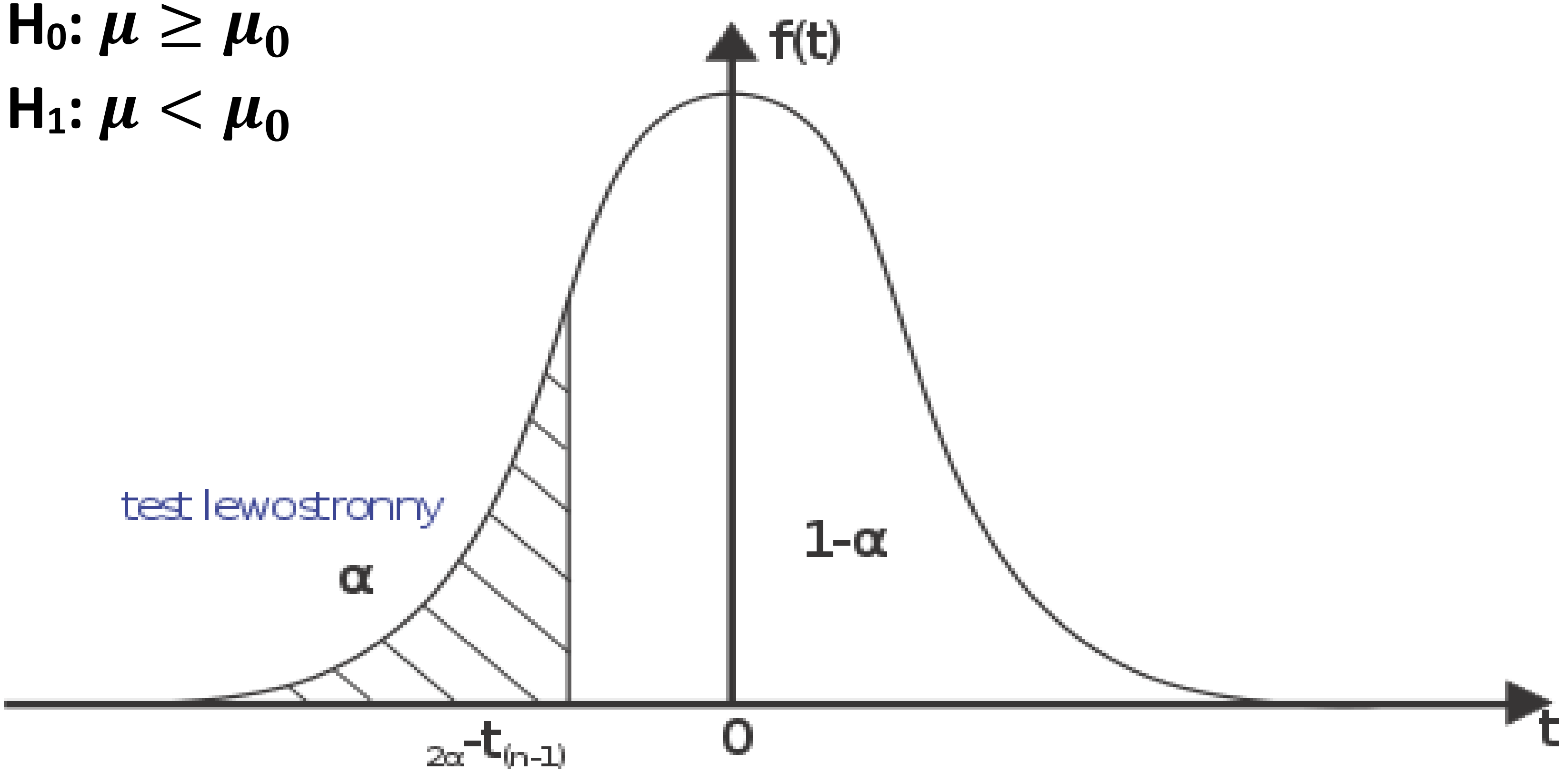
H_1: $\boldsymbol{\mu \neq \mu_0}$



Test istotności dla średniej - lewostronny

H$_0$: $\boldsymbol{\mu} \geq \boldsymbol{\mu_0}$

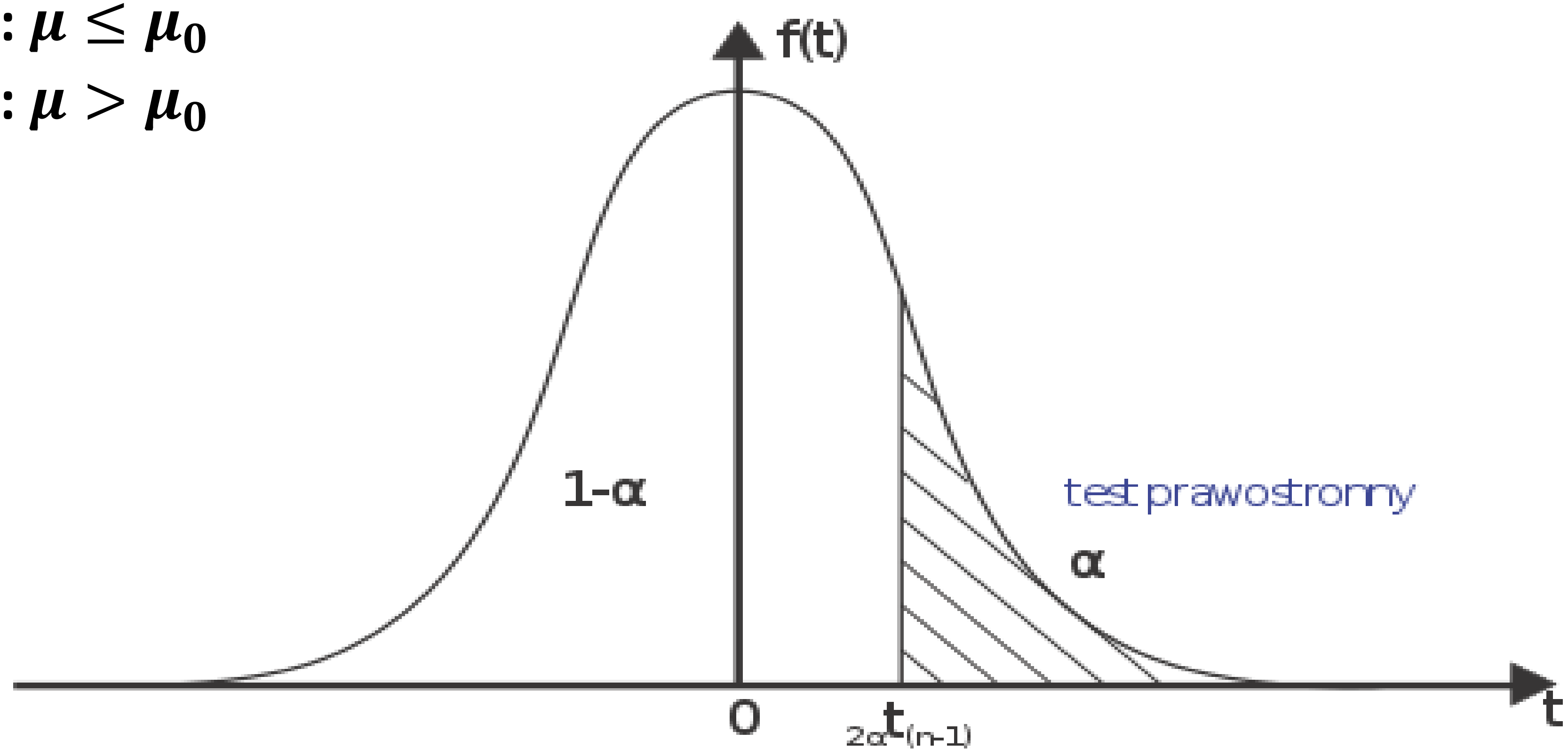
H$_1$: $\boldsymbol{\mu} < \boldsymbol{\mu_0}$



Test istotności dla średniej - prawostronny

H₀: $\mu \leq \mu_0$

H₁: $\mu > \mu_0$



Test istotności dla średniej

Przykład: W doświadczeniu biochemicznym bada się czas życia żywych komórek w toksycznym środowisku. Rozkład tego czasu można uznać za normalny. Dokonano 8 pomiarów i otrzymano następujące czasy życia komórek: 4,7; 4,0; 3,8; 6,2; 5,5; 4,5; 6,0 [godz.]. Dla $\alpha = 0{,}05$ sprawdzić hipotezę, że średni czas życia komórek wynosi 4 godziny:

$$\mu_0 = 4{,}0 \qquad \frac{s}{\sqrt{n}} = 0{,}315$$

$$\bar{x} = 5{,}0 \qquad t = 3{,}17$$

$$n = 8 \qquad {}_{0,05}t_{(7)} = 2{,}365$$

$$s^2 = 0{,}794 \qquad |t| \geq_{0,05} t_{(7)}$$

Średni czas życia jest istotnie różny od 4,0 godz.

Przedział ufności dla $1 - \alpha = 0{,}95$ wynosi $5{,}0 \pm 0{,}75$

I nie obejmuje wartości hipotetycznej 4,0 godz.

Test różnicy między dwoma średnimi

Z dwóch populacji o rozkładach normalnych oraz takich samych wariancjach wylosowano niezależnie próby o liczebnościach n_1 i n_2. Zweryfikować hipotezę mówiącą, że średnie w obu populacjach są równe.

H$_0$: $\boldsymbol{\mu_1 = \mu_2}$
H$_1$: $\boldsymbol{\mu_1 \neq \mu_2}$

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

gdzie:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

$\overline{x_1}$ – średnia z próby 1-szej populacji,
$\overline{x_2}$ – średnia z próby 2-giej populacji,
n_1 – liczebność próby 1-szej populacji,
n_2 – liczebność próby 2-giej populacji,
s_1^2 – wariancja z próby 1-szej populacji,
s_2^2 – wariancja z próby 2-giej populacji.

Statystyka *t* ma rozkład t-Studenta o $n_1 + n_2 - 2$ stopniach swobody.

H$_0$ odrzucamy, gdy $|t| \geq {}_{\alpha}\, t_{(n_1+n_2-2)}$

Test różnicy między dwoma średnimi

Gdy nie można założyć równości wariancji w obu populacjach stosujemy statystykę t:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

która ma rozkład t-Studenta o liczbie swobody ν danej wzorem:

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right)^2 \frac{1}{n_1 + 1} + \left(\frac{s_2^2}{n_2}\right)^2 \frac{1}{n_2 + 1}} - 2$$

H_0 odrzucamy, gdy $|t| \geq_\alpha t_{(\nu)}$.

Test dla danych sparowanych

Czasem mamy do czynienia z dwoma próbami, które można traktować jako zbiory obserwacji dotyczących tych samych obiektów. Przykładowo niech x_i oraz y_i będą wartościami pewnej cechy oznaczonej u n pacjentów, odpowiednio przed i po kuracji. Wówczas zamiast liczyć średnie osobno z wyników przed oraz po kuracji i testować hipotezę H_0: $\mu_x = \mu_y$ lepiej jest najpierw dla każdego pacjenta (obiektu) obliczyć różnicę:

$$z_i = x_i - y_i$$

a następnie weryfikować hipotezę o tym, że wartość średnia różnic w całej populacji jest równa zeru.

$$H_0: \mu_z = 0$$

Test dla danych sparowanych

$$z_i = x_i - y_i$$

$$H_0: \mu_z = 0$$

Jest to tzw. test dla danych sparowanych. Obliczamy statystykę t:

$$t = \frac{\bar{z}}{s_z}\sqrt{n}$$

gdzie:

$\bar{z}$ – wartość średnia różnic z próby.

$$s_z = \sqrt{\frac{\sum_i (z_i - \bar{z})^2}{n-1}}$$

Statystyka t ma rozkład t-Studenta o n-1 stopniach swobody.

H_0 odrzucamy, gdy: $|t| \geq_\alpha t_{(n-1)}$.

Test dla danych sparowanych – przykład

Przykład: W próbie klinicznej nowego środka do leczenia anurezy każdy z 29 pacjentów przez 14 dni otrzymywał lek, a przez inne 14 – placebo. Kolejność przyjmowania tych środków była u każdego pacjenta ustalana losowo. Uzyskane wyniki obserwacji na następnym slajdzie.

Liczba „suchych” nocy (na 14 badanych) u pacjentów otrzymujących lek i placebo.

Lp.	Lek	Placebo	Różnica L-P	Lp.	Placebo	Lek	Różnica L-P
1	8	5	3	2	12	11	-1
3	14	10	4	5	5	8	2
4	8	0	8	8	13	9	-4
6	9	7	2	10	8	8	0
7	11	6	5	12	8	9	1
9	3	5	-2	14	4	8	4
11	6	0	6	15	8	14	6
13	0	0	0	17	2	4	2
16	13	12	1	20	8	13	5
18	10	2	8	23	9	7	-2
19	7	5	2	26	7	10	3
21	13	13	0	29	7	6	-1
22	8	10	-2				
24	7	7	0				
25	9	0	9				
27	10	6	4				
28	2	2	0				

Test dla danych sparowanych – przykład

$$n = 29$$

$$\bar{z} = 2{,}172$$

$$s_z^2 = 11{,}005$$

$$t = 3{,}53$$

$${}_{0{,}05}t_{(28)} = 2{,}048$$

$$t > t_{kryt}$$

Różnica między skutecznością leku a placebo jest istotna.

Test istotności dla wariancji

Pobrano próbę o liczebności n z populacji <u>o rozkładzie normalnym</u>.

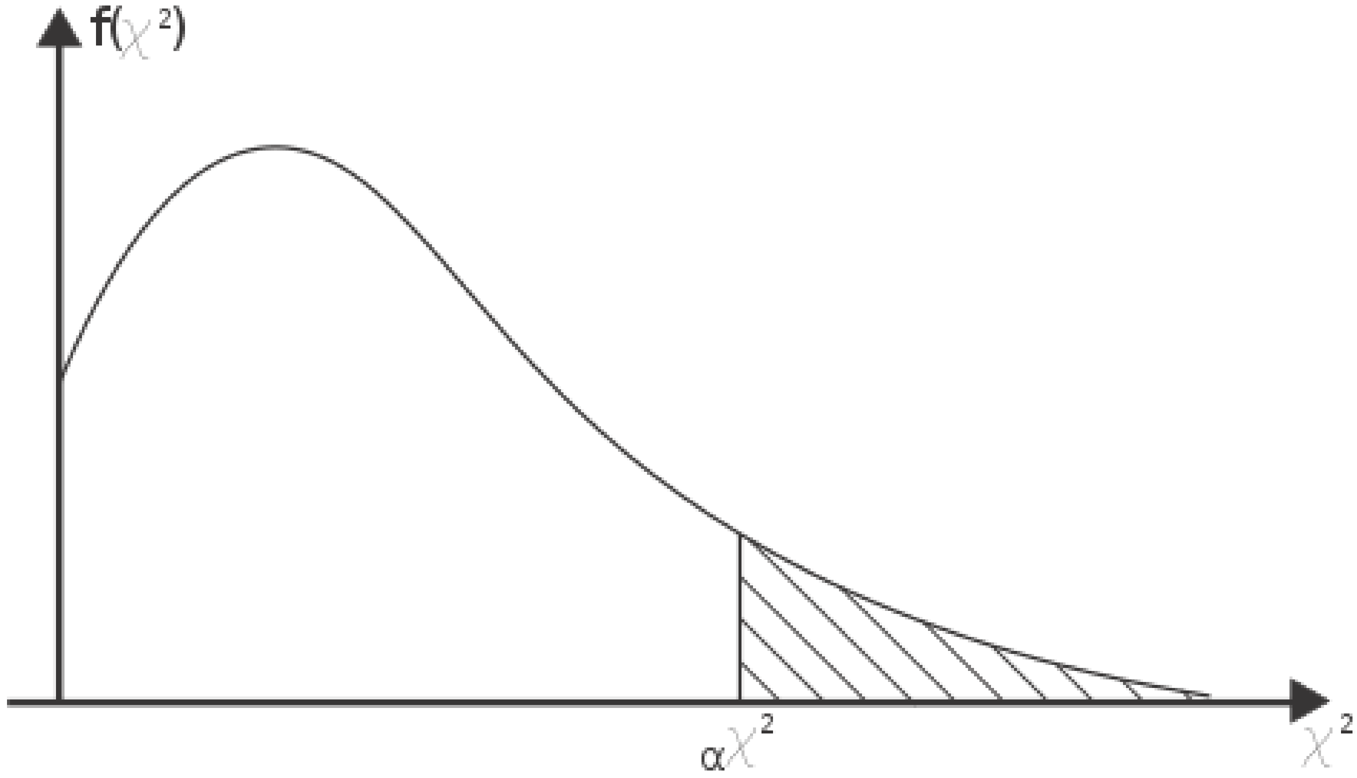
H$_0$: $\boldsymbol{\sigma^2 = \sigma_0^2}$
H$_1$: $\boldsymbol{\sigma^2 > \sigma_0^2}$

(test jednostronny, na ogół tylko wariancja większa od pewnego progu jest niekorzystna)

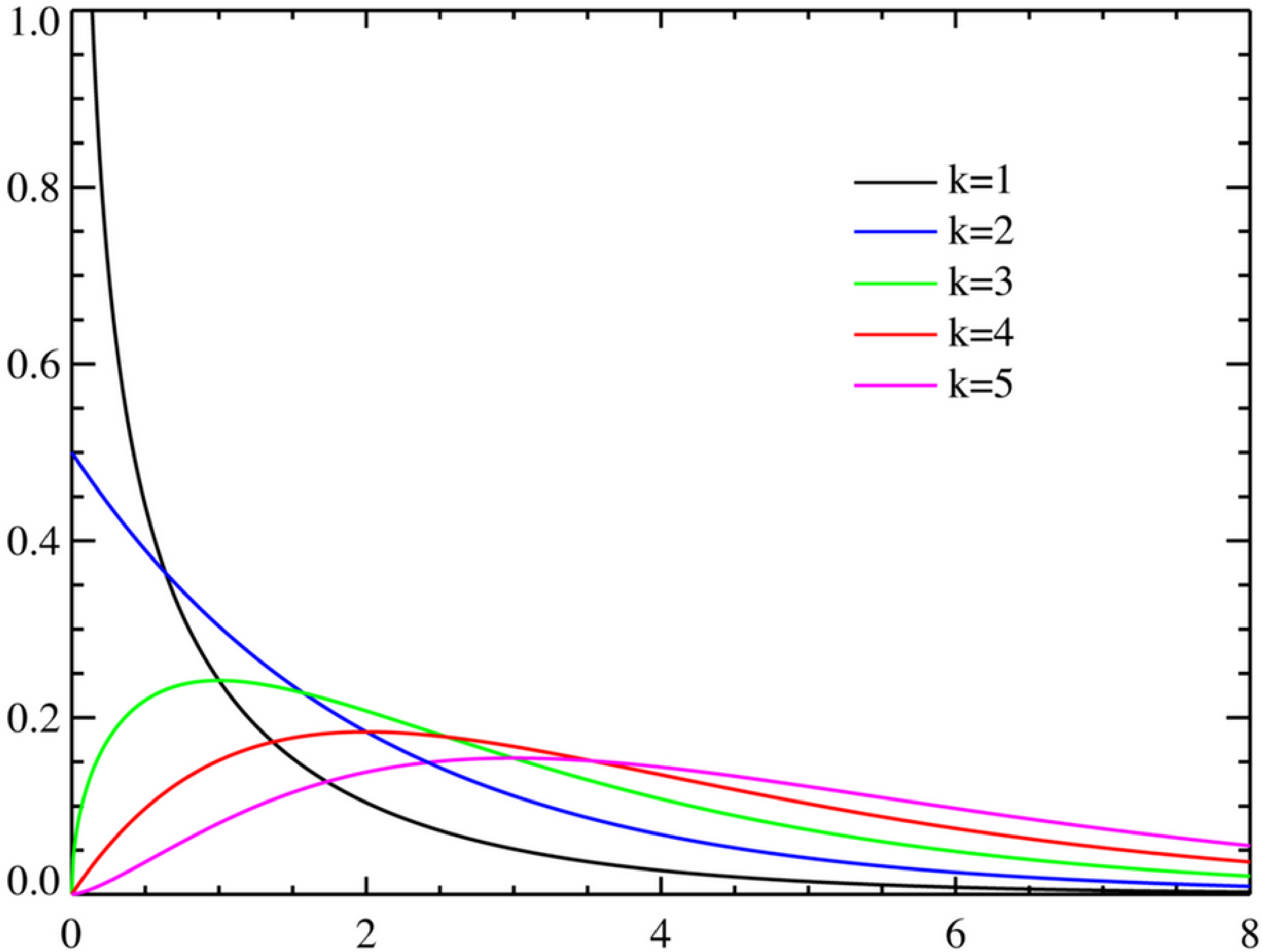
$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$ gdzie: s^2 – wariancja z próby

H$_0$ odrzucamy, gdy: $\chi^2 \geq {}_{\alpha}\chi^2_{(n-1)}$.

Test istotności dla wariancji



Rozkłady χ^2
źródło [Wikipedia]



Porównanie dwóch wariancji

Z dwóch populacji <u>o rozkładach normalnych</u> wylosowano próby o liczebnościach n_1 i n_2. Badamy:

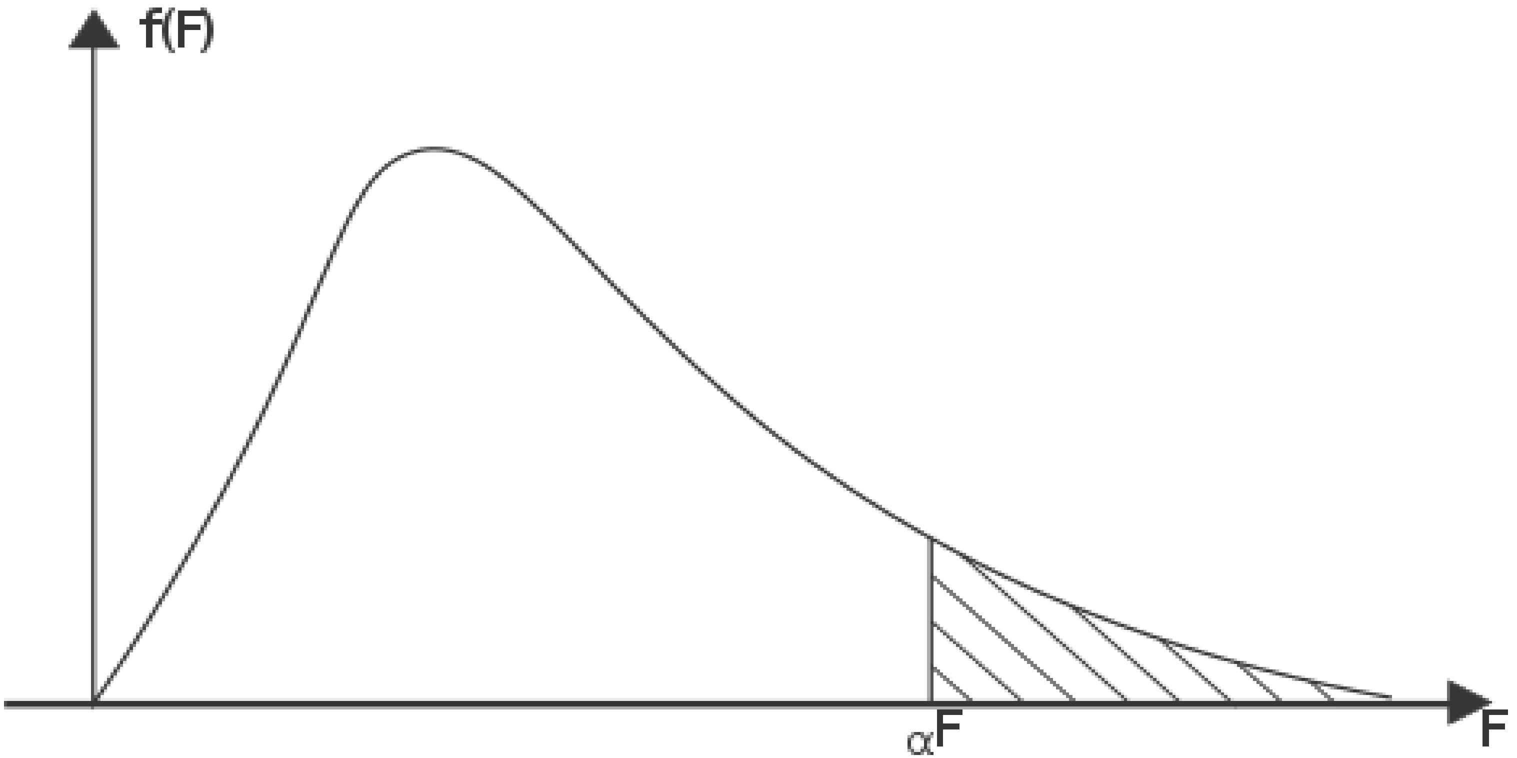
H$_0$: $\boldsymbol{\sigma_1^2 = \sigma_2^2}$
H$_1$: $\boldsymbol{\sigma_1^2 > \sigma_2^2}$
(test jednostronny)

$$F = \frac{s_1^2}{s_2^2}$$ gdzie: s_1^2, s_2^2 – wariancje z prób.

(Zawsze umieszczamy większą z wariancji w liczniku i oznaczamy jako s_1^2).

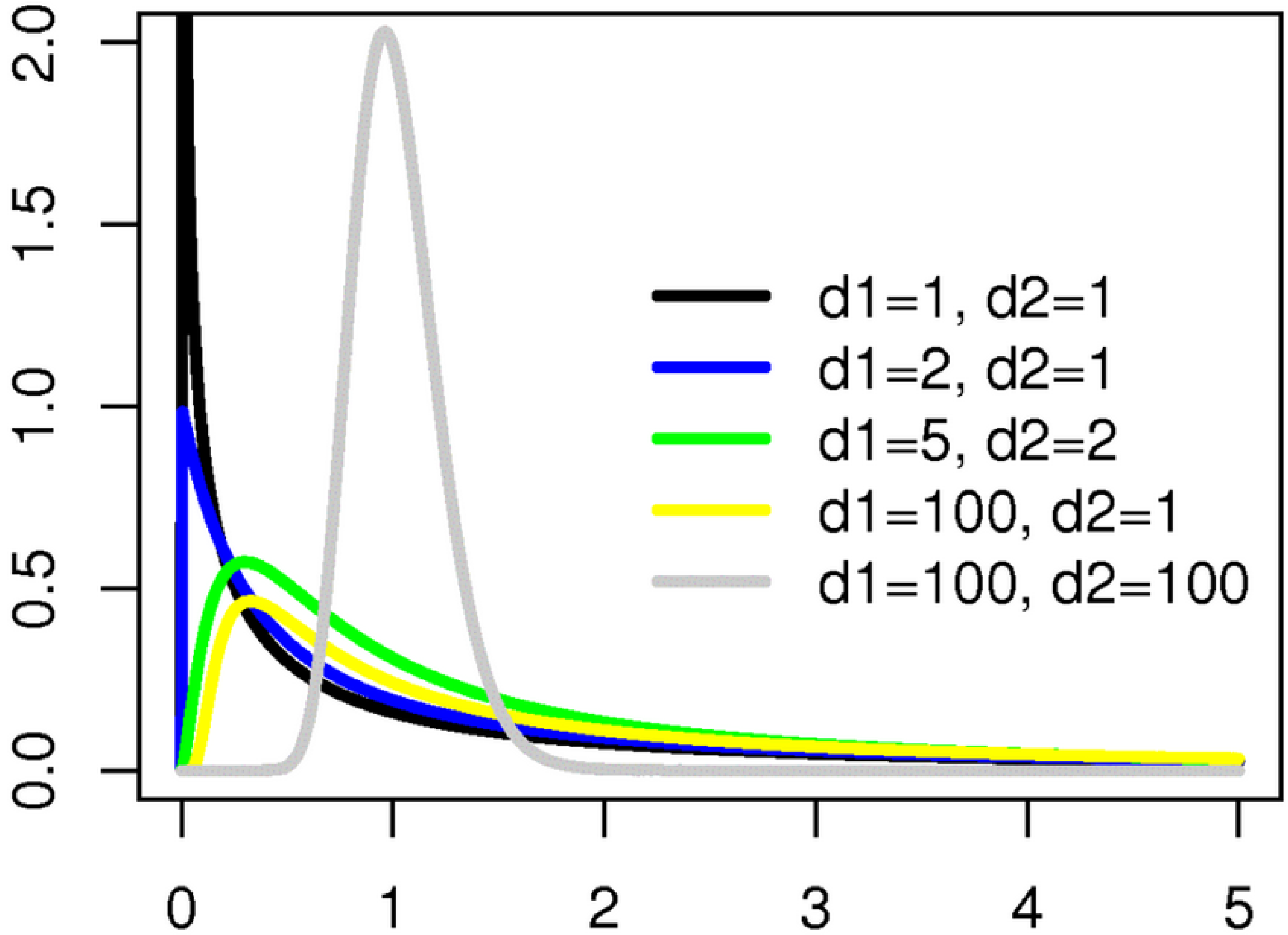
H$_0$ odrzucamy, gdy: $F \geq {}_{\alpha}F_{(n_2-1)}^{(n_1-1)}$

Porównanie dwóch wariancji



Rozkłady F Snedecora

źródło [Wikipedia]



Porównanie dwóch wariancji i dwóch średnich (przykład)

Przykład: Zmierzono w dwóch ulach średnicę komórek plastra zbudowanego przez pszczoły. Dla 7 wylosowanych komórek z plastra z pierwszego ula otrzymano wyniki (w mm): 5,36; 5,20; 5,28; 5,16; 5,30; 5,08; 5,23; analogicznie dla drugiego ula otrzymano: 5,15; 5,04; 5,30; 5,22; 5,19; 5,24; 5,12. Na poziomie istotności $\alpha = 0{,}05$ zweryfikować hipotezę, że średnie długości średnic komórek na plastrach pochodzących z dwóch różnych uli są równe.

Porównanie dwóch wariancji i dwóch średnich (przykład)

1) Sprawdzamy, czy można uznać, że **wariancje** średnic w obu plastrach są równe

$$n_1 = n_2 = 7$$

$$\overline{x_1} = 5{,}23$$

$$\overline{x_2} = 5{,}18$$

$$s_1^2 = 0{,}008767$$

$$s_2^2 = 0{,}0073$$

$$F = \frac{0{,}008767}{0{,}0073} = 1{,}201$$

Zawsze w liczniku większe oszacowanie wariancji!

$$_{0{,}05}F_{(6)}^{(6)} = 4{,}28$$

$$F < F_{kryt}.$$

Nie można odrzucić hipotezy o równości wariancji średnic komórek.

Porównanie dwóch wariancji i dwóch średnich (przykład)

2) Zakładamy więc, że wariancje są równe i obliczamy:

$$s^2 = \frac{6 \cdot 0{,}008767 + 6 \cdot 0{,}0073}{12} = 0{,}00803$$

$$t = \frac{5{,}23 - 5{,}18}{\sqrt{0{,}00803 \cdot \left(\frac{1}{7} + \frac{1}{7}\right)}} = 2{,}088$$

$${}_{0{,}05}t_{(12)} = 2{,}179$$

$$|t| < t_{kryt}\,.$$

Nie można odrzucić hipotezy o równości średnich średnic komórek.

Test istotności dla frakcji

Populacja ma rozkład dwupunktowy z nieznanym prawdopodobieństwem sukcesu π. Na podstawie dużej ($n > 100$) próby losowanej niezależnie weryfikuje się hipotezę:

H$_0$: $\boldsymbol{\pi = \pi_0}$
H$_1$: $\boldsymbol{\pi \neq \pi_0}$

$$p = \frac{r}{n}$$

r – liczba sukcesów,
n – liczba doświadczeń
p – frakcja z próby.

Stosujemy statystykę *u*:

$$u = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}}$$

u ma w przybliżeniu rozkład normalny standaryzowany, H$_0$ odrzucamy gdy: $|u| \geq_\alpha u$.

Porównanie dwóch frakcji

Z dwóch populacji o rozkładzie dwupunktowym i nieznanych prawdopodobieństwach sukcesu π_1 i π_2 pobrano dwie duże próby o liczebnościach n_1 i n_2 ($n_1 > 100, n_2 > 100$). W próbie o liczebności n_1 pochodzącej z populacji pierwszej stwierdzono r_1 sukcesów, w próbie drugiej r_2 sukcesów.

H$_0$: $\boldsymbol{\pi_1 = \pi_2}$
H$_1$: $\boldsymbol{\pi_1 \neq \pi_2}$

$$p_1 = \frac{r_1}{n_1}$$

$$p_2 = \frac{r_2}{n_2}$$

$$p = \frac{r_1 + r_2}{n_1 + n_2}$$

$$n = \frac{n_1 \cdot n_2}{n_1 + n_2}$$

$$u = \frac{p_1 - p_2}{\sqrt{\frac{p(1-p)}{n}}}$$

Statystyka *u* ma rozkład zbliżony do standaryzowanego rozkładu normalnego. H$_0$ odrzucamy gdy: $|u| \geq_{\alpha} u$.

Test McNemara

Czasami dysponujemy obserwacjami tworzącymi N par. Każda para obserwacji związana jest z tym samym obiektem. Para obserwacji to wynik dwu doświadczeń, z których każde może zakończyć się sukcesem.

		Próba 2		
		Sukces	Porażka	
Próba 1	Sukces	k	r	$k+r$
	Porażka	s	m	$s+m$
		$k+s$	$r+m$	N

Test McNemara

		Próba 2		
		Sukces	Porażka	
Próba 1	Sukces	k	r	$k+r$
	Porażka	s	m	$s+m$
		$k+s$	$r+m$	N

Frakcja sukcesu:

w próbie 1: $\frac{r+k}{N}$ w próbie 2: $\frac{k+s}{N}$

Badamy różnicę $\frac{k+r}{N} - \frac{k+s}{N} = \frac{r-s}{N}$

H$_0$: $\boldsymbol{\pi_1 - \pi_2 = 0}$

H$_1$: $\boldsymbol{\pi_1 - \pi_2 \neq 0}$

$$u = \frac{r-s}{\sqrt{r+s}}$$

Statystyka *u* ma rozkład zbliżony do normalnego standaryzowanego. H$_0$ odrzucamy gdy $|u| \geq_{\alpha} u$. Przedział ufności dla różnicy frakcji:

$$\frac{r-s}{N} \pm_{\alpha} u \frac{\sqrt{r+s}}{N}$$

Test McNemara

<u>Przykład:</u> Sto próbek plwociny posiano na dwóch podłożach A i B. Zadanie polega na porównaniu zdolności obu podłoży do wykrywania prątków gruźlicy. Wyniki przedstawiono w tabeli:

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="3">Podłoże B</td></tr>
<tr><td>+</td><td>-</td><td>Razem</td></tr>
<tr><td rowspan="3">Podłoże A</td><td>+</td><td>40
k</td><td>24
r</td><td>64
k+r</td></tr>
<tr><td>-</td><td>10
s</td><td>26
m</td><td>36
s+m</td></tr>
<tr><td>Razem</td><td>50
k+s</td><td>50
r+m</td><td>100
N</td></tr>
</table>

Test McNemara

		Podłoże B		
		+	-	Razem
Podłoże A	+	40 k	24 r	64 k+r
	-	10 s	26 m	36 s+m
	Razem	50 k+s	50 r+m	100 N

$$r = 24$$

$$s = 10$$

$$N = 100$$

$$\frac{r-s}{N} = \frac{24-10}{100} = 0{,}14$$

$$u = \frac{24-10}{\sqrt{24+10}} = 2{,}401$$

$$_{0{,}05}u = 1{,}960$$

$$|u| > u_{kryt}$$

Hipotezę o jednakowej przydatności obu podłoży należy odrzucić dla α=0,05. 95% przedział ufności dla różnicy częstości $\frac{r-s}{N}$ wynosi:

$$\frac{24-10}{100} \pm \frac{1{,}96\sqrt{24+10}}{100}$$ czyli $\mathbf{0,14 \pm 0,12}$.

$$_{0{,}01}u = 2{,}576$$

Dla $\alpha = 0{,}01$ hipotezy zerowej nie da się odrzucić.