

Weryfikacja hipotez statystycznych

Sformułowanie problemu

- **Weryfikacja hipotez statystycznych** jest drugą (po estymacji) metodą uogólniania wyników uzyskanych w próbie na całą populację.
- Weryfikacja hipotez statystycznych służy do odpowiadania na pytania o własności rozkładu zmiennej w populacji, na przykład: „czy średnia zmiennej X w populacji ma wartość m ?” albo „czy zmienne X i Y w populacji są stochastycznie niezależne?”.
- Opisana w dalszym ciągu wykładu metoda weryfikacji hipotez opiera się na **teorii podejmowania decyzji**.
- Jako autorów tego podejścia do problemu wymienia się zazwyczaj **Jerzego Neymana i Egona S. Pearsona**

Sformułowanie problemu

- **Weryfikacja hipotez statystycznych** jest drugą (po estymacji) metodą uogólniania wyników uzyskanych w próbie na całą populację.
- Weryfikacja hipotez statystycznych służy do odpowiadania na pytania o własności rozkładu zmiennej w populacji, na przykład: „czy średnia zmiennej X w populacji ma wartość m ?” albo „czy zmienne X i Y w populacji są stochastycznie niezależne?”.
- Opisana w dalszym ciągu wykładu metoda weryfikacji hipotez opiera się na **teorii podejmowania decyzji**.
- Jako autorów tego podejścia do problemu wymienia się zazwyczaj **Jerzego Neymana i Egona S. Pearsona**

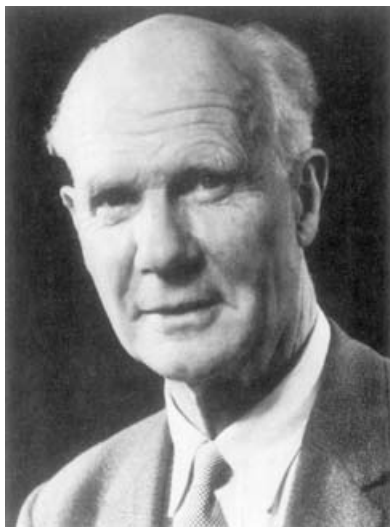
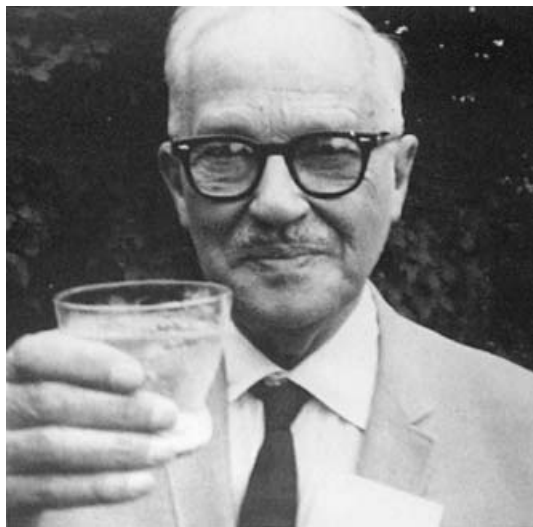
Sformułowanie problemu

- **Weryfikacja hipotez statystycznych** jest drugą (po estymacji) metodą uogólniania wyników uzyskanych w próbie na całą populację.
- Weryfikacja hipotez statystycznych służy do odpowiadania na pytania o własności rozkładu zmiennej w populacji, na przykład: „czy średnia zmiennej X w populacji ma wartość m ?” albo „czy zmienne X i Y w populacji są stochastycznie niezależne?”.
- Opisana w dalszym ciągu wykładu metoda weryfikacji hipotez opiera się na **teorii podejmowania decyzji**.
- Jako autorów tego podejścia do problemu wymienia się zazwyczaj Jerzego Neymana i Egona S. Pearsona

Sformułowanie problemu

- **Weryfikacja hipotez statystycznych** jest drugą (po estymacji) metodą uogólniania wyników uzyskanych w próbie na całą populację.
- Weryfikacja hipotez statystycznych służy do odpowiadania na pytania o własności rozkładu zmiennej w populacji, na przykład: „czy średnia zmiennej X w populacji ma wartość m ?” albo „czy zmienne X i Y w populacji są stochastycznie niezależne?”.
- Opisana w dalszym ciągu wykładu metoda weryfikacji hipotez opiera się na **teorii podejmowania decyzji**.
- Jako autorów tego podejścia do problemu wymienia się zazwyczaj **Jerzego Neymana i Egona S. Pearsona**

Jerzy Neyman i Egon S. Pearson



Sformułowanie problemu

Definicja

Hipotezę statystyczną nazywa się dowolne przypuszczenie dotyczące rozkładu jednej lub wielu zmiennych statystycznych w populacji (lub w kilku populacjach, jeśli badanie polega na porównywaniu populacji).

- Hipotezy dotyczące parametrów rozkładu, np. średniej, wariancji... nazywa się **hipotezami parametrycznymi**.
- Hipotezy dotyczące typu rozkładu zmiennej albo zależności bądź niezależności między zmiennymi nazywa się **hipotezami nieparametrycznymi**.
- Podstawą wnioskowania statystycznego jest wynik doświadczenia losowego (np. losowanie próby...).
- Hipoteza, która wyznacza jednoznacznie rozkład statystyki z próby to **hipoteza prosta**.
- Hipoteza, która nie określa w sposób jednoznaczny rozkładu wyników doświadczenia to **hipoteza złożona**.

Sformułowanie problemu

Definicja

Hipotezę statystyczną nazywa się dowolne przypuszczenie dotyczące rozkładu jednej lub wielu zmiennych statystycznych w populacji (lub w kilku populacjach, jeśli badanie polega na porównywaniu populacji).

- Hipotezy dotyczące parametrów rozkładu, np. średniej, wariancji... nazywa się **hipotezami parametrycznymi**.
- Hipotezy dotyczące typu rozkładu zmiennej albo zależności bądź niezależności między zmiennymi nazywa się **hipotezami nieparametrycznymi**.
- Podstawą wnioskowania statystycznego jest wynik doświadczenia losowego (np. losowanie próby...).
- Hipoteza, która wyznacza jednoznacznie rozkład statystyki z próby to **hipoteza prosta**.
- Hipoteza, która nie określa w sposób jednoznaczny rozkładu wyników doświadczenia to **hipoteza złożona**.

Sformułowanie problemu

Definicja

Hipotezę statystyczną nazywa się dowolne przypuszczenie dotyczące rozkładu jednej lub wielu zmiennych statystycznych w populacji (lub w kilku populacjach, jeśli badanie polega na porównywaniu populacji).

- Hipotezy dotyczące parametrów rozkładu, np. średniej, wariancji... nazywa się **hipotezami parametrycznymi**.
- Hipotezy dotyczące typu rozkładu zmiennej albo zależności bądź niezależności między zmiennymi nazywa się **hipotezami nieparametrycznymi**.
- Podstawą wnioskowania statystycznego jest wynik doświadczenia losowego (np. losowanie próby...).
- Hipoteza, która wyznacza jednoznacznie rozkład statystyki z próby to **hipoteza prosta**.
- Hipoteza, która nie określa w sposób jednoznaczny rozkładu wyników doświadczenia to **hipoteza złożona**.

Sformułowanie problemu

Definicja

Hipotezę statystyczną nazywa się dowolne przypuszczenie dotyczące rozkładu jednej lub wielu zmiennych statystycznych w populacji (lub w kilku populacjach, jeśli badanie polega na porównywaniu populacji).

- Hipotezy dotyczące parametrów rozkładu, np. średniej, wariancji... nazywa się **hipotezami parametrycznymi**.
- Hipotezy dotyczące typu rozkładu zmiennej albo zależności bądź niezależności między zmiennymi nazywa się **hipotezami nieparametrycznymi**.
- Podstawą wnioskowania statystycznego jest wynik doświadczenia losowego (np. losowanie próby...).
- Hipoteza, która wyznacza jednoznacznie rozkład statystyki z próby to **hipoteza prosta**.
- Hipoteza, która nie określa w sposób jednoznaczny rozkładu wyników doświadczenia to **hipoteza złożona**.

Sformułowanie problemu

Definicja

Hipotezę statystyczną nazywa się dowolne przypuszczenie dotyczące rozkładu jednej lub wielu zmiennych statystycznych w populacji (lub w kilku populacjach, jeśli badanie polega na porównywaniu populacji).

- Hipotezy dotyczące parametrów rozkładu, np. średniej, wariancji... nazywa się **hipotezami parametrycznymi**.
- Hipotezy dotyczące typu rozkładu zmiennej albo zależności bądź niezależności między zmiennymi nazywa się **hipotezami nieparametrycznymi**.
- Podstawą wnioskowania statystycznego jest wynik doświadczenia losowego (np. losowanie próby...).
- Hipoteza, która wyznacza jednoznacznie rozkład statystyki z próby to **hipoteza prosta**.
- Hipoteza, która nie określa w sposób jednoznaczny rozkładu wyników doświadczenia to **hipoteza złożona**

Sformułowanie problemu

Definicja

Hipotezę statystyczną nazywa się dowolne przypuszczenie dotyczące rozkładu jednej lub wielu zmiennych statystycznych w populacji (lub w kilku populacjach, jeśli badanie polega na porównywaniu populacji).

- Hipotezy dotyczące parametrów rozkładu, np. średniej, wariancji... nazywa się **hipotezami parametrycznymi**.
- Hipotezy dotyczące typu rozkładu zmiennej albo zależności bądź niezależności między zmiennymi nazywa się **hipotezami nieparametrycznymi**.
- Podstawą wnioskowania statystycznego jest wynik doświadczenia losowego (np. losowanie próby...).
- Hipoteza, która wyznacza jednoznacznie rozkład statystyki z próby to **hipoteza prosta**.
- Hipoteza, która nie określa w sposób jednoznaczny rozkładu wyników doświadczenia to **hipoteza złożona**

Sformułowanie problemu

- Hipotezę statystyczną, którą chcemy zweryfikować nazywamy **hipotezą zerową** i oznaczamy zazwyczaj h_0 .
- Hipotezę zerową weryfikuje się zawsze w stosunku do innej hipotezy, która nazywana jest **hipotezą konkurencyjną** i oznaczana h_1 .
- Hipoteza zerowa musi być (ze względów technicznych) hipotezą prostą, hipoteza konkurencyjna może być złożona.
- Proces weryfikacji ma na celu podjęcie decyzji o rodzaju podejmowanego działania. Każdej z hipotez h_0 i h_1 odpowiada działanie właściwe w przypadku, gdy ta hipoteza jest prawdziwa: a_0 dla h_0 i a_1 dla h_1 .

Sformułowanie problemu

- Hipotezę statystyczną, którą chcemy zweryfikować nazywamy **hipotezą zerową** i oznaczamy zazwyczaj h_0 .
- Hipotezę zerową weryfikuje się zawsze w stosunku do innej hipotezy, która nazywana jest **hipotezą konkurencyjną** i oznaczana h_1 .
- Hipoteza zerowa musi być (ze względów technicznych) hipotezą prostą, hipoteza konkurencyjna może być złożona.
- Proces weryfikacji ma na celu podjęcie decyzji o rodzaju podejmowanego działania. Każdej z hipotez h_0 i h_1 odpowiada działanie właściwe w przypadku, gdy ta hipoteza jest prawdziwa: a_0 dla h_0 i a_1 dla h_1 .

Sformułowanie problemu

- Hipotezę statystyczną, którą chcemy zweryfikować nazywamy **hipotezą zerową** i oznaczamy zazwyczaj h_0 .
- Hipotezę zerową weryfikuje się zawsze w stosunku do innej hipotezy, która nazywana jest **hipotezą konkurencyjną** i oznaczana h_1 .
- Hipoteza zerowa musi być (ze względów technicznych) hipotezą prostą, hipoteza konkurencyjna może być złożona.
- Proces weryfikacji ma na celu podjęcie decyzji o rodzaju podejmowanego działania. Każdej z hipotez h_0 i h_1 odpowiada działanie właściwe w przypadku, gdy ta hipoteza jest prawdziwa: a_0 dla h_0 i a_1 dla h_1 .

Sformułowanie problemu

- Hipotezę statystyczną, którą chcemy zweryfikować nazywamy **hipotezą zerową** i oznaczamy zazwyczaj h_0 .
- Hipotezę zerową weryfikuje się zawsze w stosunku do innej hipotezy, która nazywana jest **hipotezą konkurencyjną** i oznaczana h_1 .
- Hipoteza zerowa musi być (ze względów technicznych) hipotezą prostą, hipoteza konkurencyjna może być złożona.
- Proces weryfikacji ma na celu podjęcie decyzji o rodzaju podejmowanego działania. Każdej z hipotez h_0 i h_1 odpowiada działanie właściwe w przypadku, gdy ta hipoteza jest prawdziwa: a_0 dla h_0 i a_1 dla h_1 .

Sformułowanie problemu

- Procedura weryfikacji hipotez nie jest bezbłędna. Możemy popełnić dwa typy błędów.
- **Błąd pierwszego rodzaju** to odrzucenie hipotezy h_0 w przypadku, kiedy jest ona prawdziwa, tzn. podjęcie działania a_1 w sytuacji gdy właściwe było a_0 .
- **Błąd drugiego rodzaju** to nie odrzucenie hipotezy h_0 w sytuacji kiedy jest ona fałszywa, tzn. podjęcie działania a_0 podczas kiedy właściwe było a_1 .

Działanie	h_0 prawdziwa	h_1 prawdziwa
a_0	✓	błąd II rodzaju
a_1	błąd I rodzaju	✓

Sformułowanie problemu

- Procedura weryfikacji hipotez nie jest bezbłędna. Możemy popełnić dwa typy błędów.
- **Błąd pierwszego rodzaju** to odrzucenie hipotezy h_0 w przypadku, kiedy jest ona prawdziwa, tzn. podjęcie działania a_1 w sytuacji gdy właściwe było a_0 .
- **Błąd drugiego rodzaju** to nie odrzucenie hipotezy h_0 w sytuacji kiedy jest ona fałszywa, tzn. podjęcie działania a_0 podczas kiedy właściwe było a_1 .

Działanie	h_0 prawdziwa	h_1 prawdziwa
a_0	✓	błąd II rodzaju
a_1	błąd I rodzaju	✓

Sformułowanie problemu

- Procedura weryfikacji hipotez nie jest bezbłędna. Możemy popełnić dwa typy błędów.
- **Błąd pierwszego rodzaju** to odrzucenie hipotezy h_0 w przypadku, kiedy jest ona prawdziwa, tzn. podjęcie działania a_1 w sytuacji gdy właściwe było a_0 .
- **Błąd drugiego rodzaju** to nie odrzucenie hipotezy h_0 w sytuacji kiedy jest ona fałszywa, tzn. podjęcie działania a_0 podczas kiedy właściwe było a_1 .

Działanie	h_0 prawdziwa	h_1 prawdziwa
a_0	✓	błąd II rodzaju
a_1	błąd I rodzaju	✓

Sformułowanie problemu

- Procedura weryfikacji hipotez nie jest bezbłędna. Możemy popełnić dwa typy błędów.
- **Błąd pierwszego rodzaju** to odrzucenie hipotezy h_0 w przypadku, kiedy jest ona prawdziwa, tzn. podjęcie działania a_1 w sytuacji gdy właściwe było a_0 .
- **Błąd drugiego rodzaju** to nie odrzucenie hipotezy h_0 w sytuacji kiedy jest ona fałszywa, tzn. podjęcie działania a_0 podczas kiedy właściwe było a_1 .

Działanie	h_0 prawdziwa	h_1 prawdziwa
a_0	✓	błąd II rodzaju
a_1	błąd I rodzaju	✓

Przykład: hipoteza o studencie

Przykład

Wykładowca przeprowadza egzamin ze statystyki. Egzamin polega na zadawaniu 3 spośród 9 pytań obejmujących całość materiału. Wykładowca uważa, że do zaliczenia wystarczy, że student zna $\frac{2}{3}$ materiału, ale zdecydowanie nie wystarczy gdy zna $\frac{1}{3}$ materiału. Wykładowca losuje trzy pytania i na podstawie tego, jak student odpowiada, musi podjąć decyzję dotyczącą zaliczenia

Przykład: hipoteza o studencie

- Wykładowca formułuje dwie hipotezy:
 - h_0 = „student umie $\frac{2}{3}$ materiału”
 - h_1 = „student umie $\frac{1}{3}$ materiału”
- Hipotezom tym odpowiadają dwa działania
 - a_0 = „zaliczyć”
 - a_1 = „nie zaliczyć”
- Doświadczenie losowe polega na wylosowaniu 3 spośród 9 pytań i zadaniu ich studentowi. Wynik doświadczenia to 0, 1, 2 albo 3 poprawne odpowiedzi.

Przykład: hipoteza o studencie

Rozkład prawdopodobieństwa wyników doświadczenia przy założeniu prawdziwości h_0 i h_1 :

Gdy h_0 jest prawdziwa

x_i	$P(X = x_i h_0)$
0	1/84
1	18/84
2	45/84
3	20/84
Σ	1

Gdy h_1 jest prawdziwa

x_i	$P(X = x_i h_1)$
0	20/84
1	45/84
2	18/84
3	1/84
Σ	1

	0	1	2	3
d_1	a_0	a_0	a_0	a_0
d_2	a_1	a_0	a_0	a_0
d_3	a_0	a_1	a_0	a_0
d_4	a_0	a_0	a_1	a_0
d_5	a_0	a_0	a_0	a_1
d_6	a_1	a_1	a_0	a_0
d_7	a_1	a_0	a_1	a_0
d_8	a_1	a_0	a_0	a_1
d_9	a_0	a_1	a_1	a_0
d_{10}	a_0	a_1	a_0	a_1
d_{11}	a_0	a_0	a_1	a_1
d_{12}	a_1	a_1	a_1	a_0
d_{13}	a_1	a_1	a_0	a_1
d_{14}	a_1	a_0	a_1	a_1
d_{15}	a_0	a_1	a_1	a_1
d_{16}	a_1	a_1	a_1	a_1

Przykład: hipoteza o studencie

Definicja

Przyporządkowanie wynikom doświadczenia elementów ze zbioru działań nazywa się funkcją decyzyjną.

- Przykładem funkcji decyzyjnej jest

	0	1	2	3
d_2	a_1	a_0	a_0	a_0
- Zbiór wyników doświadczenia, którym odpowiada a_1 nazywa się **obszarem krytycznym** K

	0	1	2	3
d_2	a_1	a_0	a_0	a_0

, $K = \{0\}$
- Zbiór wyników doświadczenia, którym odpowiada a_0 nazywa się **obszarem przyjęcia**

	0	1	2	3
d_2	a_1	a_0	a_0	a_0

, $\{1, 2, 3\}$

Przykład: hipoteza o studencie

Problem

Jak racjonalnie wybrać najlepszą funkcję decyzyjną?

- Zakładamy maksymalne dopuszczalne **prawdopodobieństwo popełnienia błędu I-go rodzaju**.
- Prawdopodobieństwo to oznacza się α i nazywa **poziom istotności**.
- Spośród wszystkich funkcji decyzyjnych wybieramy te, dla których prawdopodobieństwo popełnienia błędu I-go rodzaju nie przekracza α .
- typowe poziomy istotności: 0, 1, 0,05, 0,01...

Przykład: hipoteza o studencie

Problem

Jak racjonalnie wybrać najlepszą funkcję decyzyjną?

- Zakładamy maksymalne dopuszczalne **prawdopodobieństwo popełnienia błędu I-go rodzaju**.
- Prawdopodobieństwo to oznacza się α i nazywa **poziom istotności**.
- Spośród wszystkich funkcji decyzyjnych wybieramy te, dla których prawdopodobieństwo popełnienia błędu I-go rodzaju nie przekracza α .
- typowe poziomy istotności: 0, 1, 0,05, 0,01...

Przykład: hipoteza o studencie

Problem

Jak racjonalnie wybrać najlepszą funkcję decyzyjną?

- Zakładamy maksymalne dopuszczalne **prawdopodobieństwo popełnienia błędu I-go rodzaju**.
- Prawdopodobieństwo to oznacza się α i nazywa **poziom istotności**.
- Spośród wszystkich funkcji decyzyjnych wybieramy te, dla których prawdopodobieństwo popełnienia błędu I-go rodzaju nie przekracza α .
- typowe poziomy istotności: 0, 1, 0,05, 0,01...

Przykład: hipoteza o studencie

Problem

Jak racjonalnie wybrać najlepszą funkcję decyzyjną?

- Zakładamy maksymalne dopuszczalne **prawdopodobieństwo popełnienia błędu I-go rodzaju**.
- Prawdopodobieństwo to oznacza się α i nazywa **poziom istotności**.
- Spośród wszystkich funkcji decyzyjnych wybieramy te, dla których prawdopodobieństwo popełnienia błędu I-go rodzaju nie przekracza α .
- typowe poziomy istotności: 0,1, 0,05, 0,01...

Przykład: hipoteza o studencie

Problem

Jak racjonalnie wybrać najlepszą funkcję decyzyjną?

- Zakładamy maksymalne dopuszczalne **prawdopodobieństwo popełnienia błędu I-go rodzaju**.
- Prawdopodobieństwo to oznacza się α i nazywa **poziom istotności**.
- Spośród wszystkich funkcji decyzyjnych wybieramy te, dla których prawdopodobieństwo popełnienia błędu I-go rodzaju nie przekracza α .
- typowe poziomy istotności: 0, 1, 0,05, 0,01...

x_i	$P(X=x_i h_0)$	
0	1/84	0,012
1	18/84	0,214
2	45/84	0,536
3	20/84	0,238
Σ	1	

	0	1	2	3	$P(I)$
d_1	a_0	a_0	a_0	a_0	0,000
d_2	a_1	a_0	a_0	a_0	0,012
d_3	a_0	a_1	a_0	a_0	0,214
d_4	a_0	a_0	a_1	a_0	0,536
d_5	a_0	a_0	a_0	a_1	0,238
d_6	a_1	a_1	a_0	a_0	0,226
d_7	a_1	a_0	a_1	a_0	0,548
d_8	a_1	a_0	a_0	a_1	0,250
d_9	a_0	a_1	a_1	a_0	0,750
d_{10}	a_0	a_1	a_0	a_1	0,452
d_{11}	a_0	a_0	a_1	a_1	0,774
d_{12}	a_1	a_1	a_1	a_0	0,762
d_{13}	a_1	a_1	a_0	a_1	0,464
d_{14}	a_1	a_0	a_1	a_1	0,786
d_{15}	a_0	a_1	a_1	a_1	0,988
d_{16}	a_1	a_1	a_1	a_1	0,012

$$\alpha = 0,3$$

	0	1	2	3	$P(I)$
d_1	a_0	a_0	a_0	a_0	0,000
d_2	a_1	a_0	a_0	a_0	0,012
d_3	a_0	a_1	a_0	a_0	0,214
d_4	a_0	a_0	a_1	a_0	0,536
d_5	a_0	a_0	a_0	a_1	0,238
d_6	a_1	a_1	a_0	a_0	0,226
d_7	a_1	a_0	a_1	a_0	0,548
d_8	a_1	a_0	a_0	a_1	0,250
d_9	a_0	a_1	a_1	a_0	0,750
d_{10}	a_0	a_1	a_0	a_1	0,452
d_{11}	a_0	a_0	a_1	a_1	0,774
d_{12}	a_1	a_1	a_1	a_0	0,762
d_{13}	a_1	a_1	a_0	a_1	0,464
d_{14}	a_1	a_0	a_1	a_1	0,786
d_{15}	a_0	a_1	a_1	a_1	0,988
d_{16}	a_1	a_1	a_1	a_1	1,000

Przykład: hipoteza o studencie

- Zakładamy maksymalne dopuszczalne **prawdopodobieństwo popełnienia błędu I-go rodzaju** $\alpha = 0,3$.
- Spośród wszystkich funkcji decyzyjnych wybieramy te, dla których prawdopodobieństwo popełnienia błędu I-go rodzaju nie przekracza α .
- Spośród wybranych w poprzednim punkcie funkcji wybieramy tę dla której prawdopodobieństwo popełnienia błędu drugiego rodzaju jest najmniejsze.
- Błąd drugiego rodzaju jest to nie odrzucenie h_0 gdy jest fałszywa, czyli podjęcie działania a_0 , gdy właściwe jest a_1 . Prawdopodobieństwo tego błędu oznaczmy β .

Przykład: hipoteza o studencie

- Zakładamy maksymalne dopuszczalne **prawdopodobieństwo popełnienia błędu I-go rodzaju** $\alpha = 0,3$.
- Spośród wszystkich funkcji decyzyjnych wybieramy te, dla których **prawdopodobieństwo popełnienia błędu I-go rodzaju nie przekracza α** .
- Spośród wybranych w poprzednim punkcie funkcji wybieramy tę dla której **prawdopodobieństwo popełnienia błędu drugiego rodzaju jest najmniejsze**.
- Błąd drugiego rodzaju jest to nie odrzucenie h_0 gdy jest fałszywa, czyli podjęcie działania a_0 , gdy właściwe jest a_1 . **Prawdopodobieństwo tego błędu oznaczmy β** .

Przykład: hipoteza o studencie

- Zakładamy maksymalne dopuszczalne **prawdopodobieństwo popełnienia błędu I-go rodzaju** $\alpha = 0,3$.
- Spośród wszystkich funkcji decyzyjnych wybieramy te, dla których prawdopodobieństwo popełnienia błędu I-go rodzaju nie przekracza α .
- Spośród wybranych w poprzednim punkcie funkcji wybieramy tę dla której prawdopodobieństwo popełnienia błędu drugiego rodzaju jest najmniejsze.
- Błąd drugiego rodzaju jest to nie odrzucenie h_0 gdy jest fałszywa, czyli podjęcie działania a_0 , gdy właściwe jest a_1 . Prawdopodobieństwo tego błędu oznaczmy β .

Przykład: hipoteza o studencie

- Zakładamy maksymalne dopuszczalne **prawdopodobieństwo popełnienia błędu I-go rodzaju** $\alpha = 0,3$.
- Spośród wszystkich funkcji decyzyjnych wybieramy te, dla których prawdopodobieństwo popełnienia błędu I-go rodzaju nie przekracza α .
- Spośród wybranych w poprzednim punkcie funkcji wybieramy tę dla której prawdopodobieństwo popełnienia błędu drugiego rodzaju jest najmniejsze.
- Błąd drugiego rodzaju jest to nie odrzucenie h_0 gdy jest fałszywa, czyli podjęcie działania a_0 , gdy właściwe jest a_1 . Prawdopodobieństwo tego błędu oznaczmy β .

$$\alpha = 0,3$$

	0	1	2	3	$P(I)$	$P(II)$
d_1	a_0	a_0	a_0	a_0	0,000	1,000
d_2	a_1	a_0	a_0	a_0	0,012	0,762
d_3	a_0	a_1	a_0	a_0	0,214	0,464
d_4	a_0	a_0	a_1	a_0	0,536	0,786
d_5	a_0	a_0	a_0	a_1	0,238	0,988
d_6	a_1	a_1	a_0	a_0	0,226	0,226
d_7	a_1	a_0	a_1	a_0	0,548	0,548
d_8	a_1	a_0	a_0	a_1	0,250	0,750
d_9	a_0	a_1	a_1	a_0	0,750	0,250
d_{10}	a_0	a_1	a_0	a_1	0,452	0,452
d_{11}	a_0	a_0	a_1	a_1	0,774	0,774
d_{12}	a_1	a_1	a_1	a_0	0,762	0,012
d_{13}	a_1	a_1	a_0	a_1	0,464	0,214
d_{14}	a_1	a_0	a_1	a_1	0,786	0,536
d_{15}	a_0	a_1	a_1	a_1	0,988	0,238
d_{16}	a_1	a_1	a_1	a_1	1,000	0,000

Przykład: hipoteza o studencie

Podsumowanie

Na poziomie istotności 0,3 optymalną funkcją decyzyjną jest funkcja d_6 , według której należy zaliczyć studentowi, który odpowiedział na dwa lub trzy pytania, a nie zaliczyć temu, który nie odpowiedział na żadne lub odpowiedział na jedno pytanie.

$$\begin{array}{c|cccc} & 0 & 1 & 2 & 3 \\ \hline d_6 & a_1 & a_1 & a_0 & a_0 \end{array} \quad K = \{0, 1\}$$

Prawdopodobieństwo popełnienia błędu I-go rodzaju dla tej funkcji jest równe $0,226 < \alpha$, prawdopodobieństwo popełnienia błędu II-go rodzaju $\beta = 0,226$.

$$\alpha = 0,1$$

	0	1	2	3	$P(I)$	$P(II)$
d_1	a_0	a_0	a_0	a_0	0,000	1,000
d_2	a_1	a_0	a_0	a_0	0,012	0,762
d_3	a_0	a_1	a_0	a_0	0,214	0,464
d_4	a_0	a_0	a_1	a_0	0,536	0,786
d_5	a_0	a_0	a_0	a_1	0,238	0,988
d_6	a_1	a_1	a_0	a_0	0,226	0,226
d_7	a_1	a_0	a_1	a_0	0,548	0,548
d_8	a_1	a_0	a_0	a_1	0,250	0,750
d_9	a_0	a_1	a_1	a_0	0,750	0,250
d_{10}	a_0	a_1	a_0	a_1	0,452	0,452
d_{11}	a_0	a_0	a_1	a_1	0,774	0,774
d_{12}	a_1	a_1	a_1	a_0	0,762	0,012
d_{13}	a_1	a_1	a_0	a_1	0,464	0,214
d_{14}	a_1	a_0	a_1	a_1	0,786	0,536
d_{15}	a_0	a_1	a_1	a_1	0,988	0,238
d_{16}	a_1	a_1	a_1	a_1	1,000	0,000

$$\alpha = 0,5$$

	0	1	2	3	$P(I)$	$P(II)$
d_1	a_0	a_0	a_0	a_0	0,000	1,000
d_2	a_1	a_0	a_0	a_0	0,012	0,762
d_3	a_0	a_1	a_0	a_0	0,214	0,464
d_4	a_0	a_0	a_1	a_0	0,536	0,786
d_5	a_0	a_0	a_0	a_1	0,238	0,988
d_6	a_1	a_1	a_0	a_0	0,226	0,226
d_7	a_1	a_0	a_1	a_0	0,548	0,548
d_8	a_1	a_0	a_0	a_1	0,250	0,750
d_9	a_0	a_1	a_1	a_0	0,750	0,250
d_{10}	a_0	a_1	a_0	a_1	0,452	0,452
d_{11}	a_0	a_0	a_1	a_1	0,774	0,774
d_{12}	a_1	a_1	a_1	a_0	0,762	0,012
d_{13}	a_1	a_1	a_0	a_1	0,464	0,214
d_{14}	a_1	a_0	a_1	a_1	0,786	0,536
d_{15}	a_0	a_1	a_1	a_1	0,988	0,238
d_{16}	a_1	a_1	a_1	a_1	1,000	0,000

Weryfikacja hipotezy o średniej.

Podobne rozumowanie można przeprowadzić w przypadku hipotezy statystycznej dotyczącej wartości średniej zmiennej X w populacji. Doświadczeniem losowym jest tutaj losowanie próby, a wynik tego doświadczenia to **wartość statystyki „średnia z próby”** w wylosowanej próbie.

- Ponieważ hipoteza zerowa musi być prosta może ona mieć jedynie następującą postać:
 - h_0 = „Średnia zmiennej X w populacji jest równa m_0 ”
- Przy założeniu, że wariancja $D^2(X)$ w populacji jest znana (lub oszacowana przy pomocy procedury estymacji) oraz próba jest wystarczająco liczna h_0 jednoznacznie określa rozkład statystyki „średnia z próby” jako $N(m_0, D(X)/\sqrt{n})$.
- Hipoteza konkurencyjna może mieć jedną z trzech postaci:
 - h_1 = „Średnia zmiennej X w populacji jest równa m_1 ”
 - h_1 = „Średnia zmiennej X w populacji jest większa (mniejsza) niż m_0 ”
 - h_1 = „Średnia zmiennej X w populacji jest różna od m_0 ”

Weryfikacja hipotezy o średniej.

Podobne rozumowanie można przeprowadzić w przypadku hipotezy statystycznej dotyczącej wartości średniej zmiennej X w populacji. Doświadczeniem losowym jest tutaj losowanie próby, a wynik tego doświadczenia to **wartość statystyki „średnia z próby”** w wylosowanej próbie.

- Ponieważ hipoteza zerowa musi być prosta może ona mieć jedynie następującą postać:
 - h_0 = „Średnia zmiennej X w populacji jest równa m_0 ”
- Przy założeniu, że wariancja $D^2(X)$ w populacji jest znana (lub oszacowana przy pomocy procedury estymacji) oraz próba jest wystarczająco liczna h_0 jednoznacznie określa rozkład statystyki „średnia z próby” jako $N(m_0, D(X)/\sqrt{n})$.
- Hipoteza konkurencyjna może mieć jedną z trzech postaci:
 - h_1 = „Średnia zmiennej X w populacji jest równa m_1 ”
 - h_1 = „Średnia zmiennej X w populacji jest większa (mniejsza) niż m_0 ”
 - h_1 = „Średnia zmiennej X w populacji jest różna od m_0 ”

Weryfikacja hipotezy o średniej.

Podobne rozumowanie można przeprowadzić w przypadku hipotezy statystycznej dotyczącej wartości średniej zmiennej X w populacji. Doświadczeniem losowym jest tutaj losowanie próby, a wynik tego doświadczenia to **wartość statystyki „średnia z próby”** w wylosowanej próbie.

- Ponieważ hipoteza zerowa musi być prosta może ona mieć jedynie następującą postać:
 - h_0 = „Średnia zmiennej X w populacji jest równa m_0 ”
- Przy założeniu, że wariancja $D^2(X)$ w populacji jest znana (lub oszacowana przy pomocy procedury estymacji) oraz próba jest wystarczająco liczna h_0 jednoznacznie określa rozkład statystyki „średnia z próby” jako $N(m_0, D(X)/\sqrt{n})$.
- Hipoteza konkurencyjna może mieć jedną z trzech postaci:
 - h_1 = „Średnia zmiennej X w populacji jest równa m_1 ”
 - h_1 = „Średnia zmiennej X w populacji jest większa (mniejsza) niż m_0 ”
 - h_1 = „Średnia zmiennej X w populacji jest różna od m_0 ”

Weryfikacja hipotezy o średniej.

Podobne rozumowanie można przeprowadzić w przypadku hipotezy statystycznej dotyczącej wartości średniej zmiennej X w populacji. Doświadczeniem losowym jest tutaj losowanie próby, a wynik tego doświadczenia to **wartość statystyki „średnia z próby”** w wylosowanej próbie.

- Ponieważ hipoteza zerowa musi być prosta może ona mieć jedynie następującą postać:
 - h_0 = „Średnia zmiennej X w populacji jest równa m_0 ”
- Przy założeniu, że wariancja $D^2(X)$ w populacji jest znana (lub oszacowana przy pomocy procedury estymacji) oraz próba jest wystarczająco liczna h_0 jednoznacznie określa rozkład statystyki „średnia z próby” jako $N(m_0, D(X)/\sqrt{n})$.
- Hipoteza konkurencyjna może mieć jedną z trzech postaci:
 - h_1 = „Średnia zmiennej X w populacji jest równa m_1 ”
 - h_1 = „Średnia zmiennej X w populacji jest większa (mniejsza) niż m_0 ”
 - h_1 = „Średnia zmiennej X w populacji jest różna od m_0 ”

Weryfikacja hipotezy o średniej.

Podobne rozumowanie można przeprowadzić w przypadku hipotezy statystycznej dotyczącej wartości średniej zmiennej X w populacji. Doświadczeniem losowym jest tutaj losowanie próby, a wynik tego doświadczenia to **wartość statystyki „średnia z próby”** w wylosowanej próbie.

- Ponieważ hipoteza zerowa musi być prosta może ona mieć jedynie następującą postać:
 - h_0 = „Średnia zmiennej X w populacji jest równa m_0 ”
- Przy założeniu, że wariancja $D^2(X)$ w populacji jest znana (lub oszacowana przy pomocy procedury estymacji) oraz próba jest wystarczająco liczna h_0 jednoznacznie określa rozkład statystyki „średnia z próby” jako $N(m_0, D(X)/\sqrt{n})$.
- Hipoteza konkurencyjna może mieć jedną z trzech postaci:
 - h_1 = „Średnia zmiennej X w populacji jest równa m_1 ”
 - h_1 = „Średnia zmiennej X w populacji jest większa (mniejsza) niż m_0 ”
 - h_1 = „Średnia zmiennej X w populacji jest różna od m_0 ”

Weryfikacja hipotezy o średniej.

Podobne rozumowanie można przeprowadzić w przypadku hipotezy statystycznej dotyczącej wartości średniej zmiennej X w populacji. Doświadczeniem losowym jest tutaj losowanie próby, a wynik tego doświadczenia to **wartość statystyki „średnia z próby”** w wylosowanej próbie.

- Ponieważ hipoteza zerowa musi być prosta może ona mieć jedynie następującą postać:
 - h_0 = „Średnia zmiennej X w populacji jest równa m_0 ”
- Przy założeniu, że wariancja $D^2(X)$ w populacji jest znana (lub oszacowana przy pomocy procedury estymacji) oraz próba jest wystarczająco liczna h_0 jednoznacznie określa rozkład statystyki „średnia z próby” jako $N(m_0, D(X)/\sqrt{n})$.
- Hipoteza konkurencyjna może mieć jedną z trzech postaci:
 - h_1 = „Średnia zmiennej X w populacji jest równa m_1 ”
 - h_1 = „Średnia zmiennej X w populacji jest większa (mniejsza) niż m_0 ”
 - h_1 = „Średnia zmiennej X w populacji jest różna od m_0 ”

Weryfikacja hipotezy o średniej.

Podobne rozumowanie można przeprowadzić w przypadku hipotezy statystycznej dotyczącej wartości średniej zmiennej X w populacji. Doświadczeniem losowym jest tutaj losowanie próby, a wynik tego doświadczenia to **wartość statystyki „średnia z próby”** w wylosowanej próbie.

- Ponieważ hipoteza zerowa musi być prosta może ona mieć jedynie następującą postać:
 - h_0 = „Średnia zmiennej X w populacji jest równa m_0 ”
- Przy założeniu, że wariancja $D^2(X)$ w populacji jest znana (lub oszacowana przy pomocy procedury estymacji) oraz próba jest wystarczająco liczna h_0 jednoznacznie określa rozkład statystyki „średnia z próby” jako $N(m_0, D(X)/\sqrt{n})$.
- Hipoteza konkurencyjna może mieć jedną z trzech postaci:
 - h_1 = „Średnia zmiennej X w populacji jest równa m_1 ”
 - h_1 = „Średnia zmiennej X w populacji jest większa (mniejsza) niż m_0 ”
 - h_1 = „Średnia zmiennej X w populacji jest różna od m_0 ”

Weryfikacja hipotezy o średniej.

Podobne rozumowanie można przeprowadzić w przypadku hipotezy statystycznej dotyczącej wartości średniej zmiennej X w populacji. Doświadczeniem losowym jest tutaj losowanie próby, a wynik tego doświadczenia to **wartość statystyki „średnia z próby”** w wylosowanej próbie.

- Ponieważ hipoteza zerowa musi być prosta może ona mieć jedynie następującą postać:
 - h_0 = „Średnia zmiennej X w populacji jest równa m_0 ”
- Przy założeniu, że wariancja $D^2(X)$ w populacji jest znana (lub oszacowana przy pomocy procedury estymacji) oraz próba jest wystarczająco liczna h_0 jednoznacznie określa rozkład statystyki „średnia z próby” jako $N(m_0, D(X)/\sqrt{n})$.
- Hipoteza konkurencyjna może mieć jedną z trzech postaci:
 - h_1 = „Średnia zmiennej X w populacji jest równa m_1 ”
 - h_1 = „Średnia zmiennej X w populacji jest większa (mniejsza) niż m_0 ”
 - h_1 = „Średnia zmiennej X w populacji jest różna od m_0 ”

Weryfikacja hipotezy o średniej.

Omówimy dokładnie przypadek weryfikacji pary hipotez

- h_0 = „Średnia zmiennej X w populacji jest równa m_0 ”
- h_1 = „Średnia zmiennej X w populacji jest równa m_1 ” ($m_1 > m_0$)

Działania odpowiadające tym hipotezom są oczywiście zależne od sytuacji, dlatego sformułujemy je bardzo ogólnie:

- a_0 = „postępujemy tak, jakby h_0 była prawdziwa”
- a_1 = „postępujemy tak, jakby h_1 była prawdziwa”

Weryfikacja hipotezy o średniej.

Omówimy dokładnie przypadek weryfikacji pary hipotez

- h_0 = „Średnia zmiennej X w populacji jest równa m_0 ”
- h_1 = „Średnia zmiennej X w populacji jest równa m_1 ” ($m_1 > m_0$)

Działania odpowiadające tym hipotezom są oczywiście zależne od sytuacji, dlatego sformułujemy je bardzo ogólnie:

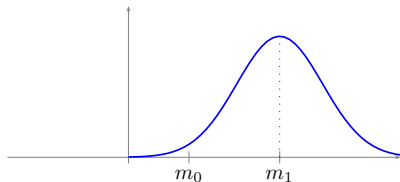
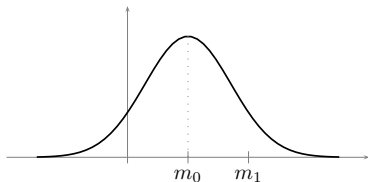
- a_0 = „postępujemy tak, jakby h_0 była prawdziwa”
- a_1 = „postępujemy tak, jakby h_1 była prawdziwa”

Weryfikacja hipotezy o średniej.

Przy założeniu, że $D^2(X)$ jest znane mamy następujące rozkłady warunkowe wyników doświadczenia losowego:

$$P(\bar{X}|h_0) = N(m_0, D(X)/\sqrt{n}),$$

$$P(\bar{X}|h_1) = N(m_1, D(X)/\sqrt{n})$$



Weryfikacja hipotezy o średniej.

Przyjmujemy poziom istotności α . Zdefiniowanie funkcji decyzyjnej jest równoznaczne z podaniem obszaru krytycznego K , spełniającego warunek:

$$P(\bar{X} \in K \mid h_0) \leq \alpha$$

czyli warunkowe prawdopodobieństwo, że statystyka „średnia z próby” przyjmie wartości leżące w K przy warunku, że h_0 jest prawdziwa nie przekracza α .

Jeśli do podejmowania decyzji używamy funkcji decyzyjnej d_K , której odpowiada obszar krytyczny K , to jeśli wartość $\bar{X}$ w wylosowanej próbie jest elementem K podejmujemy działanie a_1 , a jeśli nie jest elementem K podejmujemy działanie a_0 .

Weryfikacja hipotezy o średniej.

Przyjmujemy poziom istotności α . Zdefiniowanie funkcji decyzyjnej jest równoznaczne z podaniem obszaru krytycznego K , spełniającego warunek:

$$P(\bar{X} \in K \mid h_0) \leq \alpha$$

czyli warunkowe prawdopodobieństwo, że statystyka „średnia z próby” przyjmie wartości leżące w K przy warunku, że h_0 jest prawdziwa nie przekracza α .

Jeśli do podejmowania decyzji używamy funkcji decyzyjnej d_K , której odpowiada obszar krytyczny K , to jeśli wartość $\bar{X}$ w wylosowanej próbie jest elementem K podejmujemy działanie a_1 , a jeśli nie jest elementem K podejmujemy działanie a_0 .

Weryfikacja hipotezy o średniej.

Przyjmujemy poziom istotności α . Zdefiniowanie funkcji decyzyjnej jest równoznaczne z podaniem obszaru krytycznego K , spełniającego warunek:

$$P(\bar{X} \in K \mid h_0) \leq \alpha$$

czyli warunkowe prawdopodobieństwo, że statystyka „średnia z próby” przyjmie wartości leżące w K przy warunku, że h_0 jest prawdziwa nie przekracza α .

Jeśli do podejmowania decyzji używamy funkcji decyzyjnej d_K , której odpowiada obszar krytyczny K , to jeśli wartość $\bar{X}$ w wylosowanej próbie jest elementem K podejmujemy działanie a_1 , a jeśli nie jest elementem K podejmujemy działanie a_0 .

Weryfikacja hipotezy o średniej.

Przyjmujemy poziom istotności α . Zdefiniowanie funkcji decyzyjnej jest równoznaczne z podaniem obszaru krytycznego K , spełniającego warunek:

$$P(\bar{X} \in K \mid h_0) \leq \alpha$$

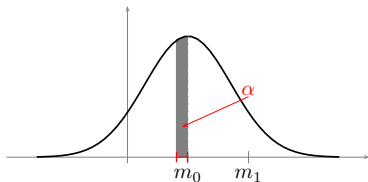
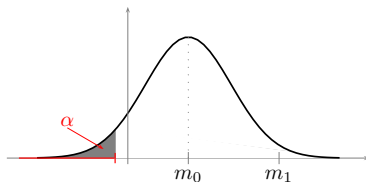
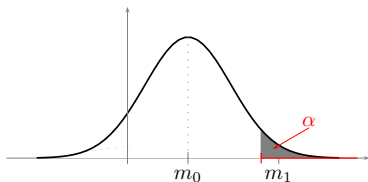
czyli warunkowe prawdopodobieństwo, że statystyka „średnia z próby” przyjmie wartości leżące w K przy warunku, że h_0 jest prawdziwa nie przekracza α .

Jeśli do podejmowania decyzji używamy funkcji decyzyjnej d_K , której odpowiada obszar krytyczny K , to jeśli wartość $\bar{X}$ w wylosowanej próbie jest elementem K podejmujemy działanie a_1 , a jeśli nie jest elementem K podejmujemy działanie a_0 .

Weryfikacja hipotezy o średniej.

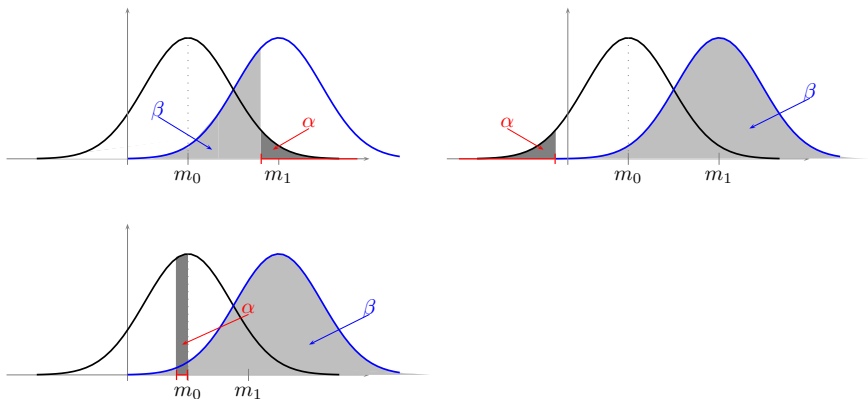
Przykłady obszarów krytycznych spełniających warunek

$$P(\bar{X} \in K \mid h_0) \leq \alpha:$$

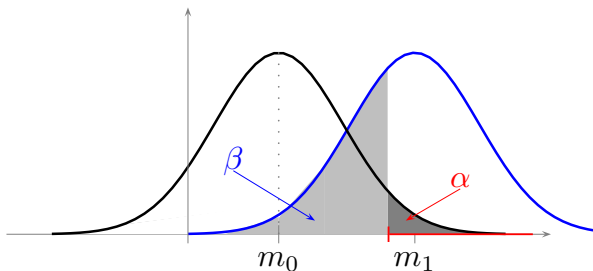


Weryfikacja hipotezy o średniej.

Każdej funkcji decyzyjnej odpowiada prawdopodobieństwo popełnienia błędu drugiego rodzaju β . Optymalna funkcja decyzyjna to ta, dla której β jest najmniejsze:



Weryfikacja hipotezy o średniej.



Obszar krytyczny ma postać $K = \langle k, \infty \rangle$. Jak znaleźć k ?

Weryfikacja hipotezy o średniej.

Granica obszaru krytycznego k spełnia warunek

$$P(\bar{X} > k | h_0) = \alpha$$

Przekształcamy:

$$\begin{aligned}\alpha = P(\bar{X} > k | h_0) &= P\left(\frac{\bar{X} - m_0}{D(X)/\sqrt{n}} > \frac{k - m_0}{D(X)/\sqrt{n}} \mid h_0\right) = \\ &= P(U > \frac{k - m_0}{D(X)/\sqrt{n}})\end{aligned}$$

Jeśli więc λ_α jest taką liczbą rzeczywistą, że

$$F_U(\lambda_\alpha) = 1 - \alpha$$

to

$$\frac{k - m_0}{D(X)/\sqrt{n}} = \lambda_\alpha \quad \text{zatem} \quad k = m_0 + \lambda_\alpha \frac{D(X)}{\sqrt{n}}$$

Weryfikacja hipotezy o średniej.

Granica obszaru krytycznego k spełnia warunek

$$P(\bar{X} > k | h_0) = \alpha$$

Przekształcamy:

$$\begin{aligned}\alpha = P(\bar{X} > k | h_0) &= P\left(\frac{\bar{X} - m_0}{D(X)/\sqrt{n}} > \frac{k - m_0}{D(X)/\sqrt{n}} \mid h_0\right) = \\ &= P(U > \frac{k - m_0}{D(X)/\sqrt{n}})\end{aligned}$$

Jeśli więc λ_α jest taką liczbą rzeczywistą, że

$$F_U(\lambda_\alpha) = 1 - \alpha$$

to

$$\frac{k - m_0}{D(X)/\sqrt{n}} = \lambda_\alpha \quad \text{zatem} \quad k = m_0 + \lambda_\alpha \frac{D(X)}{\sqrt{n}}$$

Weryfikacja hipotezy o średniej.

Granica obszaru krytycznego k spełnia warunek

$$P(\bar{X} > k | h_0) = \alpha$$

Przekształcamy:

$$\begin{aligned}\alpha &= P(\bar{X} > k | h_0) = P\left(\frac{\bar{X} - m_0}{D(X)/\sqrt{n}} > \frac{k - m_0}{D(X)/\sqrt{n}} \mid h_0\right) = \\ &= P(U > \frac{k - m_0}{D(X)/\sqrt{n}})\end{aligned}$$

Jeśli więc λ_α jest taką liczbą rzeczywistą, że

$$F_U(\lambda_\alpha) = 1 - \alpha$$

to

$$\frac{k - m_0}{D(X)/\sqrt{n}} = \lambda_\alpha \quad \text{zatem} \quad k = m_0 + \lambda_\alpha \frac{D(X)}{\sqrt{n}}$$

Weryfikacja hipotezy o średniej.

Granica obszaru krytycznego k spełnia warunek

$$P(\bar{X} > k | h_0) = \alpha$$

Przekształcamy:

$$\begin{aligned}\alpha = P(\bar{X} > k | h_0) &= P\left(\frac{\bar{X} - m_0}{D(X)/\sqrt{n}} > \frac{k - m_0}{D(X)/\sqrt{n}} \mid h_0\right) = \\ &= P(U > \frac{k - m_0}{D(X)/\sqrt{n}})\end{aligned}$$

Jeśli więc λ_α jest taką liczbą rzeczywistą, że

$$F_U(\lambda_\alpha) = 1 - \alpha$$

to

$$\frac{k - m_0}{D(X)/\sqrt{n}} = \lambda_\alpha \quad \text{zatem} \quad k = m_0 + \lambda_\alpha \frac{D(X)}{\sqrt{n}}$$

Weryfikacja hipotezy o średniej.

Granica obszaru krytycznego k spełnia warunek

$$P(\bar{X} > k | h_0) = \alpha$$

Przekształcamy:

$$\begin{aligned}\alpha = P(\bar{X} > k | h_0) &= P\left(\frac{\bar{X} - m_0}{D(X)/\sqrt{n}} > \frac{k - m_0}{D(X)/\sqrt{n}} \mid h_0\right) = \\ &= P(U > \frac{k - m_0}{D(X)/\sqrt{n}})\end{aligned}$$

Jeśli więc λ_α jest taką liczbą rzeczywistą, że

$$F_U(\lambda_\alpha) = 1 - \alpha$$

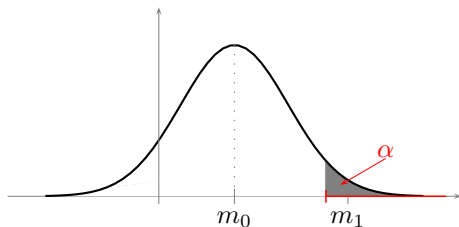
to

$$\frac{k - m_0}{D(X)/\sqrt{n}} = \lambda_\alpha \quad \text{zatem} \quad k = m_0 + \lambda_\alpha \frac{D(X)}{\sqrt{n}}$$

Weryfikacja hipotezy o średniej.

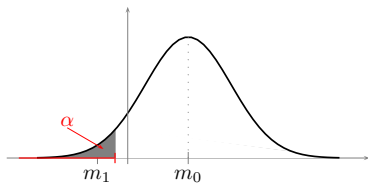
Obszary krytyczne są różne w zależności od postaci hipotezy konkurencyjnej:

- $h_1 = „E(X) = m_1 > m_0”$ lub $h_1 = „E(X) > m_0”$



Weryfikacja hipotezy o średniej.

- $h_1 = „E(X) = m_1 < m_0”$ lub $h_1 = „E(X) < m_0”$



Weryfikacja hipotezy o średniej.

- $h_1 = „E(X) \neq m_0”$

