Statystyka i rachunek prawdopodobieństwa

Wykład 11: Hipotezy statystyczne

Mariusz Tarnopolski

Instytut Astronomii UMK

Oprócz estymacji parametrów rozkładu zmiennych losowych najważniejszym zadaniem statystyki jest weryfikacja hipotez statystycznych na podstawie zaobserwowanej próby losowej pobranej z populacji generalnej.

Oprócz estymacji parametrów rozkładu zmiennych losowych najważniejszym zadaniem statystyki jest weryfikacja hipotez statystycznych na podstawie zaobserwowanej próby losowej pobranej z populacji generalnej.

Testy statystyczne nie dają bezwzględnej pewności o słuszności hipotezy statystycznej, ale pozwalają ocenić/oszacować/skwantyfikować jej słuszność.

Przykład. Firma farmaceutyczna twierdzi, że jej nowy lek na chrapanie pomaga 90% pacjentów w zaledwie 2 tygodnie.

Przykład. Firma farmaceutyczna twierdzi, że jej nowy lek na chrapanie pomaga 90% pacjentów w zaledwie 2 tygodnie.
Lekarz przepisał ten lek 15 pacjentom i przebadał ich po 2 tygodniach.
Uzyskał wyniki:

1. nie chrapało 12 osób
2. nadal chrapały 3 osoby

Spodziewana liczba wyleczonych to $0.90 \cdot 15 = 13.5 \approx 14$ osób.

Czy firma farmaceutyczna kłamie? Czy próbka jest za mała? Wyniki są z pewnością stochastyczne—jakiego wyniku (przedziału ufności) można się spodziewać? Czy można uznać stanowisko firmy za zasadne?

Czy firma farmaceutyczna kłamie? Czy próbka jest za mała? Wyniki są z pewnością stochastyczne—jakiego wyniku (przedziału ufności) można się spodziewać? Czy można uznać stanowisko firmy za zasadne?

Testowanie hipotez statystycznych w 7 krokach:

1. Sformułuj hipotezę zerową H_0
2. Wybierz statystykę testową
3. Określ poziom istotności
4. Określ obszar krytyczny/odrzuceń
5. Oblicz wartość statystyki testowej z próby, zakładając prawdziwość hipotezy zerowej
6. Sprawdź, czy wartość statystyki wpada w obszar krytyczny/odrzuceń
7. Podejmij decyzję

$\boxed{1.}$ Postawienie hipotezy statystycznej H_0, że interesujące nas zjawisko podlega danemu rozkładowi prawdopodobieństwa (którego parametry lub postać postulujemy) oraz określenie hipotezy alternatywnej H_1, którą jesteśmy skłonni przyjąć w przypadku odrzucenia hipotezy H_0. W procesie testowania zakładamy prawdziwość H_0. Jeśli znajdziemy (w toku testowania) mocne dowody przeciwko niej, to ją odrzucamy na korzyść H_1. Zwykle H_1 jest pewnego rodzaju zaprzeczeniem H_0.

$\boxed{1.}$ Postawienie hipotezy statystycznej H_0, że interesujące nas zjawisko podlega danemu rozkładowi prawdopodobieństwa (którego parametry lub postać postulujemy) oraz określenie hipotezy alternatywnej H_1, którą jesteśmy skłonni przyjąć w przypadku odrzucenia hipotezy H_0. W procesie testowania zakładamy prawdziwość H_0. Jeśli znajdziemy (w toku testowania) mocne dowody przeciwko niej, to ją odrzucamy na korzyść H_1. Zwykle H_1 jest pewnego rodzaju zaprzeczeniem H_0.
Np. H_0 o przyjęciu przez parametr Q rozkładu prawd. wartości Q_0 zapiszemy jako

$$H_0: \, Q = Q_0$$

Hipoteza alternatywna to:

$$H_1: \, Q \neq Q_0 \quad \text{lub} \quad H_1: \, Q < Q_0 \quad \text{lub} \quad Q > Q_0 \quad \text{lub} \quad Q = Q_1$$

$\boxed{1.}$ Postawienie hipotezy statystycznej H_0, że interesujące nas zjawisko podlega danemu rozkładowi prawdopodobieństwa (którego parametry lub postać postulujemy) oraz określenie hipotezy alternatywnej H_1, którą jesteśmy skłonni przyjąć w przypadku odrzucenia hipotezy H_0. W procesie testowania zakładamy prawdziwość H_0. Jeśli znajdziemy (w toku testowania) mocne dowody przeciwko niej, to ją odrzucamy na korzyść H_1. Zwykle H_1 jest pewnego rodzaju zaprzeczeniem H_0.
Np. H_0 o przyjęciu przez parametr Q rozkładu prawd. wartości Q_0 zapiszemy jako

$$H_0 : Q = Q_0$$

Hipoteza alternatywna to:

$$H_1 : Q \neq Q_0 \quad \text{lub} \quad H_1 : Q < Q_0 \quad \text{lub} \quad Q > Q_0 \quad \text{lub} \quad Q = Q_1$$

Testujemy hipotezę, że lek usuwa chrapanie u 90% pacjentów:

$$\begin{aligned} H_0 &: p = 0.9 \quad && \text{twierdzenie firmy} \\ H_1 &: p < 0.9 \quad && \text{firma kłamie} \end{aligned}$$

$\boxed{2.}$ Wybór odpowiedniej dla danego testu statystyki testowej (*test statistic*, *TS*), która ma znany rozkład i której wartość w próbie posłuży do weryfikacji H_0.

$\boxed{2.}$ Wybór odpowiedniej dla danego testu statystyki testowej (*test statistic, TS*), która ma znany rozkład i której wartość w próbie posłuży do weryfikacji H_0.

X to liczba osób, która przestała chrapać — liczba sukcesów. Przy założeniu H_0 mamy $p = 0.9$. Badamy prawdopodobieństwo sukcesu w próbie 15-elementowej.

Zatem X ma rozkład dwumianowy, $X \sim \mathcal{D}(15, 0.9)$. Jest to podstawa dla naszej TS.

$\boxed{3.}$ Określenie poziomu istotności α. Jest to taki poziom prawdopodobieństwa, przy którym wyniki uzyskane z próby uznajemy za tak mało prawdopodobne, że pozwalają odrzucić H_0.

$\boxed{3.}$ Określenie poziomu istotności α. Jest to taki poziom prawdopodobieństwa, przy którym wyniki uzyskane z próby uznajemy za tak mało prawdopodobne, że pozwalają odrzucić H_0.
Zwykle przyjmuje się $\alpha = 0.05$ lub $\alpha = 0.01$.
Ponieważ weryfikacja H_0 oparta jest na próbie losowej, możliwe jest popełnienie błędu przy decydowania o odrzuceniu (lub nie) H_0. Poziom istotności α jest też prawdopodobieństwem popełnienia tzw. błędu I rodzaju—odrzucenie H_0 gdy jest ona w rzeczywistości prawdziwa.

$\boxed{3.}$ Określenie poziomu istotności α. Jest to taki poziom prawdopodobieństwa, przy którym wyniki uzyskane z próby uznajemy za tak mało prawdopodobne, że pozwalają odrzucić H_0.
Zwykle przyjmuje się $\alpha = 0.05$ lub $\alpha = 0.01$.
Ponieważ weryfikacja H_0 oparta jest na próbie losowej, możliwe jest popełnienie błędu przy decydowania o odrzuceniu (lub nie) H_0. Poziom istotności α jest też prawdopodobieństwem popełnienia tzw. błędu I rodzaju—odrzucenie H_0 gdy jest ona w rzeczywistości prawdziwa.
Przyjmuje się też często poziom β błędu II rodzaju—przyjęcie hipotezy fałszywej. Podobnie jak α, powinien on być dość mały (0.05 lub 0.01).

$\boxed{3.}$ Określenie poziomu istotności α. Jest to taki poziom prawdopodobieństwa, przy którym wyniki uzyskane z próby uznajemy za tak mało prawdopodobne, że pozwalają odrzucić H_0.
Zwykle przyjmuje się $\alpha = 0.05$ lub $\alpha = 0.01$.
Ponieważ weryfikacja H_0 oparta jest na próbie losowej, możliwe jest popełnienie błędu przy decydowania o odrzuceniu (lub nie) H_0. Poziom istotności α jest też prawdopodobieństwem popełnienia tzw. błędu I rodzaju—odrzucenie H_0 gdy jest ona w rzeczywistości prawdziwa.
Przyjmuje się też często poziom β błędu II rodzaju—przyjęcie hipotezy fałszywej. Podobnie jak α, powinien on być dość mały (0.05 lub 0.01).

Decyzja	Hipoteza H_0	
	jest prawdziwa	jest fałszywa
przyjąć hipotezę H_0	decyzja poprawna	decyzja błędna **błąd drugiego rodzaju** $\boldsymbol{\beta}$
odrzucić hipotezę H_0	decyzja błędna **błąd pierwszego rodzaju** $\boldsymbol{\alpha}$	decyzja poprawna

$\boxed{3.}$ Określenie poziomu istotności α. Jest to taki poziom prawdopodobieństwa, przy którym wyniki uzyskane z próby uznajemy za tak mało prawdopodobne, że pozwalają odrzucić H_0.
Zwykle przyjmuje się $\alpha = 0.05$ lub $\alpha = 0.01$.
Ponieważ weryfikacja H_0 oparta jest na próbie losowej, możliwe jest popełnienie błędu przy decydowania o odrzuceniu (lub nie) H_0. Poziom istotności α jest też prawdopodobieństwem popełnienia tzw. błędu I rodzaju—odrzucenie H_0 gdy jest ona w rzeczywistości prawdziwa.
Przyjmuje się też często poziom β błędu II rodzaju—przyjęcie hipotezy fałszywej. Podobnie jak α, powinien on być dość mały (0.05 lub 0.01).

Decyzja	Hipoteza H_0	
	jest prawdziwa	jest fałszywa
przyjąć hipotezę H_0	decyzja poprawna	decyzja błędna **błąd drugiego rodzaju** $\boldsymbol{\beta}$
odrzucić hipotezę H_0	decyzja błędna **błąd pierwszego rodzaju** $\boldsymbol{\alpha}$	decyzja poprawna

Przyjmijmy $\alpha = 0.05$.

$\boxed{4.}$ Należy dokonać podziału wszystkich wartości TS na dwa dopełniające się obszary. Jeden z nich to **obszar krytyczny** K_α, tj. taki zakres TS, że przy założeniu słuszności H_0 prawdopodobieństwo przyjęcia przez statystykę wartości z tego obszaru równe jest poziomowi istotności α:

$$P\left(TS \in K_\alpha | H_0\right) = \alpha$$

Jest to **obszar odrzuceń** hipotezy H_0, tzn. jeśli wartość TS znajdzie się w tym obszarze to odrzucimy H_0.

$\boxed{4.}$ Należy dokonać podziału wszystkich wartości TS na dwa dopełniające się obszary. Jeden z nich to **obszar krytyczny** K_α, tj. taki zakres TS, że przy założeniu słuszności H_0 prawdopodobieństwo przyjęcia przez statystykę wartości z tego obszaru równe jest poziomowi istotności α:

$$P\left(TS \in K_\alpha | H_0\right) = \alpha$$

Jest to **obszar odrzuceń** hipotezy H_0, tzn. jeśli wartość TS znajdzie się w tym obszarze to odrzucimy H_0.
Dla $H_1 : Q > Q_0$ budujemy obszar krytyczny prawostronny;
dla $H_1 : Q < Q_0$ — lewostronny;
dla $H_1 : Q \neq Q_0$ — dwustronny.

Alternatywnie (acz równoważnie) można policzyć p-wartość: prawdopodobieństwo otrzymania wyniku takiego jak w próbie lub gorszego, przy zał. H_0. Jeśli p-wartość jest mniejsza niż α, to TS z próby wpada w obszar krytyczny—odrzucimy H_0. p-wartość jest też nazywana krytycznym poziomem istotności.

https://en.wikipedia.org/wiki/Misuse_of_p-values

Alternatywnie (acz równoważnie) można policzyć p-wartość: prawdopodobieństwo otrzymania wyniku takiego jak w próbie lub gorszego, przy zał. H_0. Jeśli p-wartość jest mniejsza niż α, to TS z próby wpada w obszar krytyczny—odrzucimy H_0. p-wartość jest też nazywana krytycznym poziomem istotności.

p-wartość jest często interpretacyjnie nadużywana: nie określa prawdopodobieństwa prawdziwości H_0 (bez sensu—albo jest, albo nie jest prawdziwa), różne wartości p-wartości są nieinformatywne, tzn. nie mówią, która hipoteza jest bardziej prawdopodobna.
p-wartość mówi: prawdopodobieństwo uzyskania wartości TS takiej jak z próby lub większej, przy założeniu rozkładu zmiennej TS jak w H_0.

https://en.wikipedia.org/wiki/Misuse_of_p-values

Drugi obszar, K'_α, będący dopełnieniem obszaru krytycznego, to obszar odrzucenia H_0, czyli obszar przyjęcia hipotezy H_1:

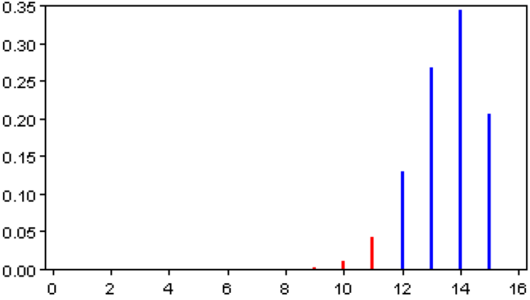
$$P\left(TS \in K'_\alpha | H_1\right) = \beta$$

Jeśli wartość TS znajdzie się w tym obszarze to odrzucimy hipotezę H_1, czyli przyjmiemy hipotezę H_0.
Ponieważ K_α i K'_α się dopełniają, oraz H_0 i H_1 się wzajemnie wykluczają, wystarczy posługiwać się jednym obszarem krytycznym—zazwyczaj K_α.

Drugi obszar, K'_α, będący dopełnieniem obszaru krytycznego, to obszar odrzucenia H_0, czyli obszar przyjęcia hipotezy H_1:

$$P\left(TS \in K'_\alpha | H_1\right) = \beta$$

Jeśli wartość TS znajdzie się w tym obszarze to odrzucimy hipotezę H_1, czyli przyjmiemy hipotezę H_0.
Ponieważ K_α i K'_α się dopełniają, oraz H_0 i H_1 się wzajemnie wykluczają, wystarczy posługiwać się jednym obszarem krytycznym—zazwyczaj K_α.
Dla leku na chrapanie, konstruujemy lewostronny obszar krytyczny, który (przyjmując $\alpha = 0.05$) rozciąga się do kwantyla $\mathcal{D}(15, 0.9)_{0.05} \approx 11$:



Number of Successes	Number of Failures	Exact Probability	Cumulative Probability
0	15	0.000%	0.000%
1	14	0.000%	0.000%
2	13	0.000%	0.000%
3	12	0.000%	0.000%
4	11	0.000%	0.000%
5	10	0.000%	0.000%
6	9	0.000%	0.000%
7	8	0.003%	0.003%
8	7	0.028%	0.031%
9	6	0.194%	0.225%
10	5	1.047%	1.272%
11	4	4.284%	5.556%
12	3	12.851%	18.406%
13	2	26.690%	45.096%
14	1	34.315%	79.411%
15	0	20.589%	100.000%

$\boxed{5.}$ Statystyka testowa przyjęła wartość $X = 12$. Wyznaczamy p-wartość jako $P(X \leqslant 12)$; jeśli będzie ona mniejsza niż α to H_0 odrzucimy. Otrzymujemy

$$P(X \leqslant 12) = 0.18406$$

$\boxed{5.}$ Statystyka testowa przyjęła wartość $X = 12$. Wyznaczamy p-wartość jako $P(X \leqslant 12)$; jeśli będzie ona mniejsza niż α to H_0 odrzucimy. Otrzymujemy
$$P(X \leqslant 12) = 0.18406$$

$\boxed{6.}$ Statystyka z próbki wynosi $X = 12$ i nie wpada w obszar krytyczny K_α. Mamy równoważnie, że p-wartość $> \alpha$, zatem ostatecznie

$\boxed{5.}$ Statystyka testowa przyjęła wartość $X = 12$. Wyznaczamy p-wartość jako $P(X \leqslant 12)$; jeśli będzie ona mniejsza niż α to H_0 odrzucimy. Otrzymujemy

$$P(X \leqslant 12) = 0.18406$$

$\boxed{6.}$ Statystyka z próbki wynosi $X = 12$ i nie wpada w obszar krytyczny K_α. Mamy równoważnie, że p-wartość $> \alpha$, zatem ostatecznie

$\boxed{7.}$ nie ma podstaw do odrzucenia H_0, zatem uzyskana wartość $X = 12$ jest zgodna statystycznie z twierdzeniem firmy o 90% skuteczności leku.

A co gdyby próbka była większa niż 15 osób...?

A co gdyby próbka była większa niż 15 osób...?

Zbadano 100 osób przyjmujących lek na chrapanie przez 2 tygodnie.

Uzyskano wyniki:

1. nie chrapało 80 osób
2. nadal chrapało 20 osób

Spodziewana liczba wyleczonych to $0.90 \cdot 100 = 90$ osób.

A co gdyby próbka była większa niż 15 osób...?

Zbadano 100 osób przyjmujących lek na chrapanie przez 2 tygodnie.

Uzyskano wyniki:

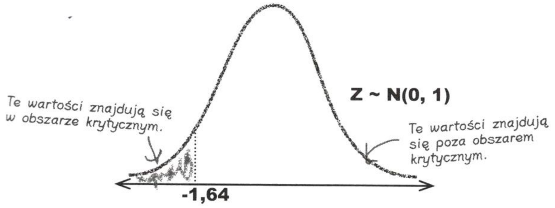
1. nie chrapało 80 osób
2. nadal chrapało 20 osób

Spodziewana liczba wyleczonych to $0.90 \cdot 100 = 90$ osób.

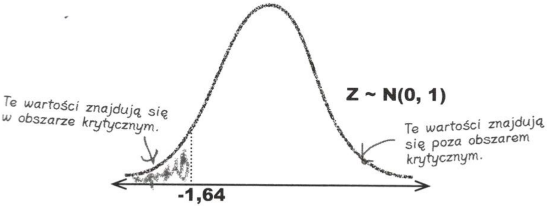
Czy 80 osób to podejrzanie mała liczba?

1.
$$H_0 : p = 0.9 \quad \text{twierdzenie firmy}$$
$$H_1 : p < 0.9 \quad \text{firma kłamie}$$

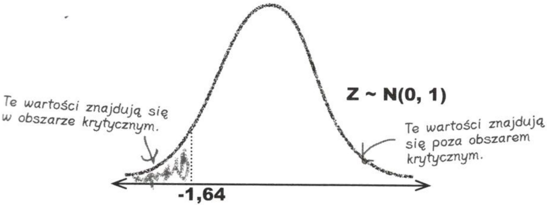
2. Zmienna losowa o rozkładzie dwumianowym $X \sim D(100, 0.9)$ ma $\mathbb{E}(X) = np = 90$, $V(X) = npq = 9$. Dla dużej probki ($n = 100 > 50$) standaryzujemy: $Z = \dfrac{X - 90}{\sqrt{9}}$ oraz przybliżamy (CLT) standardowym rozkładem normalnym, $Z \sim \mathcal{N}(0, 1)$.
3. $\alpha = 0.05$
4. Kwantyl $u_{0.05} = -1.64$
5. $TS = \dfrac{80 - 90}{3} = -3.33$



1. $$H_0 : p = 0.9 \quad \text{twierdzenie firmy}$$
$$H_1 : p < 0.9 \quad \text{firma kłamie}$$
2. Zmienna losowa o rozkładzie dwumianowym $X \sim D(100, 0.9)$ ma $\mathbb{E}(X) = np = 90$, $V(X) = npq = 9$. Dla dużej probki ($n = 100 > 50$) standaryzujemy: $Z = \dfrac{X - 90}{\sqrt{9}}$ oraz przybliżamy (CLT) standardowym rozkładem normalnym, $Z \sim \mathcal{N}(0, 1)$.
3. $\alpha = 0.05$
4. Kwantyl $u_{0.05} = -1.64$
5. $TS = \dfrac{80 - 90}{3} = -3.33$

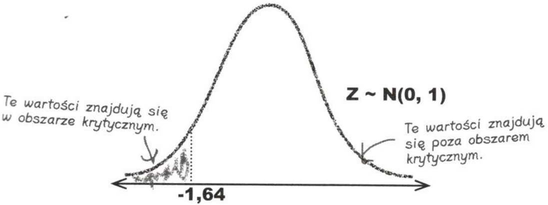


1. $$H_0 : p = 0.9 \quad \text{twierdzenie firmy}$$
$$H_1 : p < 0.9 \quad \text{firma kłamie}$$
2. Zmienna losowa o rozkładzie dwumianowym $X \sim D(100, 0.9)$ ma $\mathbb{E}(X) = np = 90$, $V(X) = npq = 9$. Dla dużej probki $(n = 100 > 50)$ standaryzujemy: $Z = \dfrac{X - 90}{\sqrt{9}}$ oraz przybliżamy (CLT) standardowym rozkładem normalnym, $Z \sim \mathcal{N}(0, 1)$.
3. $\alpha = 0.05$
4. Kwantyl $u_{0.05} = -1.64$
5. $TS = \dfrac{80 - 90}{3} = -3.33$

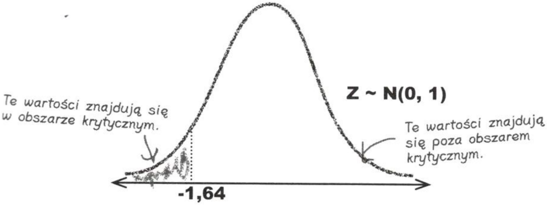


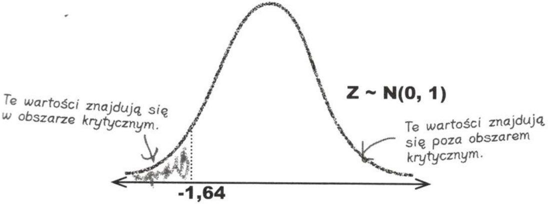
1.
$$H_0 : p = 0.9 \quad \text{twierdzenie firmy}$$
$$H_1 : p < 0.9 \quad \text{firma kłamie}$$

2. Zmienna losowa o rozkładzie dwumianowym $X \sim D(100, 0.9)$ ma $\mathbb{E}(X) = np = 90$, $V(X) = npq = 9$. Dla dużej probki ($n = 100 > 50$) standaryzujemy: $Z = \dfrac{X - 90}{\sqrt{9}}$ oraz przybliżamy (CLT) standardowym rozkładem normalnym, $Z \sim \mathcal{N}(0, 1)$.
3. $\alpha = 0.05$
4. Kwantyl $u_{0.05} = -1.64$
5. $TS = \dfrac{80 - 90}{3} = -3.33$

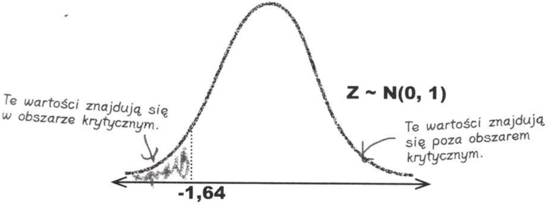


1. $$H_0 : p = 0.9 \quad \text{twierdzenie firmy}$$ $$H_1 : p < 0.9 \quad \text{firma kłamie}$$
2. Zmienna losowa o rozkładzie dwumianowym $X \sim D(100, 0.9)$ ma $\mathbb{E}(X) = np = 90$, $V(X) = npq = 9$. Dla dużej probki ($n = 100 > 50$) standaryzujemy: $Z = \dfrac{X - 90}{\sqrt{9}}$ oraz przybliżamy (CLT) standardowym rozkładem normalnym, $Z \sim \mathcal{N}(0, 1)$.
3. $\alpha = 0.05$
4. Kwantyl $u_{0.05} = -1.64$
5. $TS = \dfrac{80 - 90}{3} = -3.33$

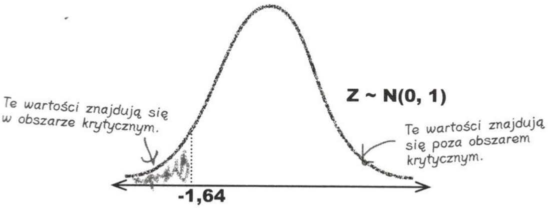




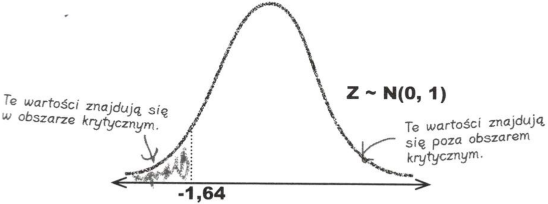
4. Kwantyl $u_{0.05} = -1.64$
5. $TS = \dfrac{80-90}{3} = -3.33$
6. Wartość TS wpada w obszar krytyczny K_α.
7. Odrzucamy na poziomie istotności 5% hipotezę H_0 o 90% skuteczności leku.
8. Inna droga: p-wartość $= P(Z < -3.33) = 0.0004$ (z tablic statystycznych).
   Jest to wartość mniejsza niż α, zatem odrzucamy H_0.



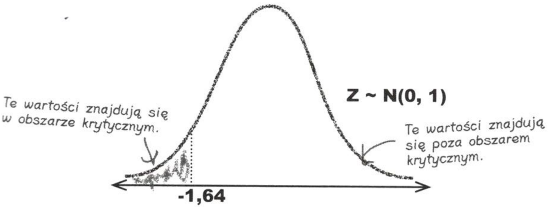
4. Kwantyl $u_{0.05} = -1.64$
5. $TS = \dfrac{80 - 90}{3} = -3.33$
6. Wartość TS wpada w obszar krytyczny K_α.
7. Odrzucamy na poziomie istotności 5% hipotezę H_0 o 90% skuteczności leku.
8. Inna droga: p-wartość $= P(Z < -3.33) = 0.0004$ (z tablic statystycznych).
   Jest to wartość mniejsza niż α, zatem odrzucamy H_0.



4. Kwantyl $u_{0.05} = -1.64$
5. $TS = \dfrac{80 - 90}{3} = -3.33$
6. Wartość TS wpada w obszar krytyczny K_α.
7. Odrzucamy na poziomie istotności 5% hipotezę H_0 o 90% skuteczności leku.
8. Inna droga: p-wartość $= P(Z < -3.33) = 0.0004$ (z tablic statystycznych).
   Jest to wartość mniejsza niż α, zatem odrzucamy H_0.



4. Kwantyl $u_{0.05} = -1.64$
5. $TS = \dfrac{80 - 90}{3} = -3.33$
6. Wartość TS wpada w obszar krytyczny K_α.
7. Odrzucamy na poziomie istotności 5% hipotezę H_0 o 90% skuteczności leku.
8. Inna droga: p-wartość $= P(Z < -3.33) = 0.0004$ (z tablic statystycznych).
   Jest to wartość mniejsza niż α, zatem odrzucamy H_0.



4. Kwantyl $u_{0.05} = -1.64$
5. $TS = \dfrac{80-90}{3} = -3.33$
6. Wartość TS wpada w obszar krytyczny K_α.
7. Odrzucamy na poziomie istotności 5% hipotezę H_0 o 90% skuteczności leku.
8. Inna droga: p-wartość $= P(Z < -3.33) = 0.0004$ (z tablic statystycznych).
   Jest to wartość mniejsza niż α, zatem odrzucamy H_0.

Przed użyciem zapoznaj się z treścią ulotki dołączonej do opakowania bądź skonsultuj się z lekarzem lub farmaceutą, gdyż każdy lek niewłaściwie stosowany zagraża Twojemu życiu lub zdrowiu.

Przed użyciem zapoznaj się z treścią ulotki dołączonej do opakowania bądź skonsultuj się z lekarzem lub farmaceutą, gdyż każdy lek niewłaściwie stosowany zagraża Twojemu życiu lub zdrowiu. Zwłaszcza, jeśli lek jest halucynacją AI.