ТЕСТИРАЊЕ СТАТИСТИЧКИХ ХИПОТЕЗА

ТЕСТИРАЊЕ СТАТИСТИЧКЕ ХИПОТЕЗЕ ЗАСНОВАНО НА ЈЕДНОМ УЗОРКУ

АРИТМЕТИЧКА СРЕДИНА

1. **У извјештају једног производног предузећа од 3.000 радника, стоји да је просјечна плата у том предузећу 600 КМ са просјечним одступањем од 40 КМ. Случајно смо изабрали 80 радника тог предузећа и утврдили просјечну плату од 590 КМ. Испитати уз 5% ризика да ли наведену тврдњу можемо прихватити као тачну?**

РЈЕШЕЊЕ

$$N = 3000$$

$$\mu_0 = 600$$

$$n = 80$$

$$\sigma = 40$$

$$\bar{x} = 590$$

1. корак: формулисање нулте и алтернативне хипотезе;

$$H_0: \mu = 600$$

$$H_1: \mu \neq 600$$

2. корак: избор статистике теста;
Пошто је варијанса основног скупа позната, а узорак већи од 30 елемената, користи се статистика Z теста...

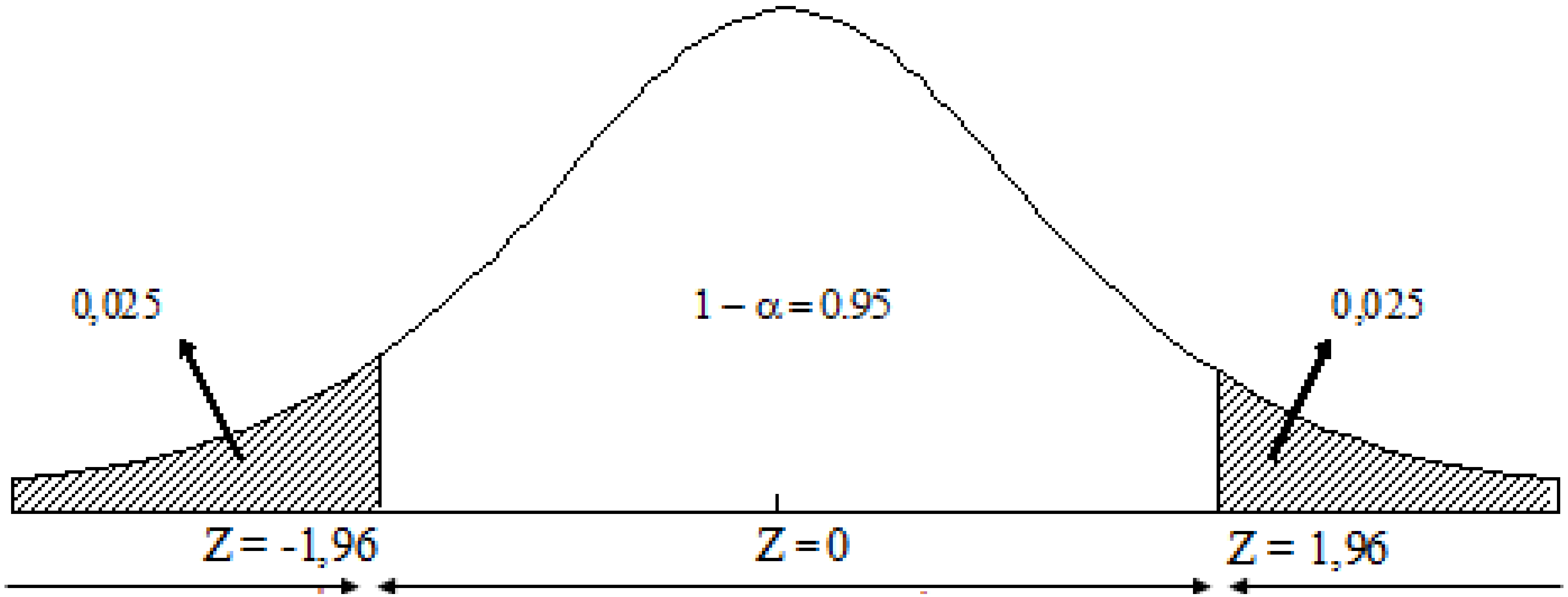
3. корак: избор нивоа значајности теста;
Закључак се доноси уз 5% ризика!

4. корак: формулисање правила на основу којих ће се вршити закључивање;
Критична област (област одбацивања нулте хипотезе) је распоређена симетрично на крајевима Z распореда, па доњу и горњу вриједност налазимо у Таблицама бр. 3

$$F(Z_{\alpha/2}) = 1 - \frac{\alpha}{2} = 1 - 0{,}025 = 0{,}975 \Rightarrow Z_{\alpha/2} = 1{,}96$$

Управо, на основу критичне вриједности формирају се правила одлучивања:

Ho не треба одбацити ако је $|Z| \leq 1{,}96$
Ho треба одбацити ако је $|Z| > 1.96$



5. корак: одређивање реализоване вриједности;

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

$$\frac{n}{N} = \frac{80}{3000} = 0{,}026 < 0{,}05 \Rightarrow$$ не користи се поправни фактор

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{590 - 600}{\frac{40}{\sqrt{80}}} = -\frac{10}{4{,}47} = -2{,}24$$

Закључак:

$$|Z| > Z_{\alpha/2} \Rightarrow$$

одбацујемо нулту хипотезу и (са 5% ризика) закључујемо да не можемо прихватити уводну претпоставку као истиниту, тј. просјечне плате се статистички значајно разликују од 600 КМ, како је наведено у извјештају!

2. Према декларацији произвођача сијалица просјечан вијек трајања његових производа износи више од 2000 часова, са просјечним одступањем од 70 часова. Случајно смо изабрали 30 сијалица и утврдили просјечан вијек трајања тих сијалица од 1988 часова.

Провјерити да ли можемо, уз 5% ризика, прихватити тврдњу овог произвођача као тачну (под претпоставком да распоред сијалица према вијеку трајања има карактеристике нормалног распореда).

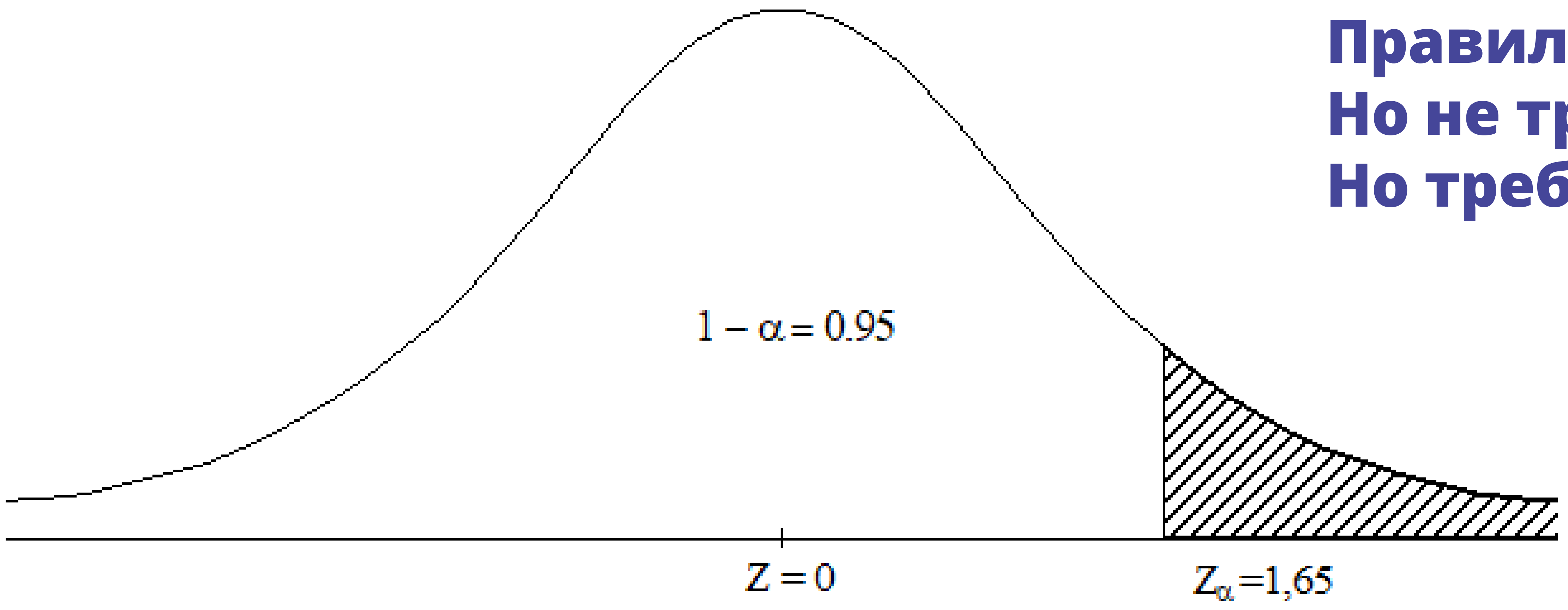
РЈЕШЕЊЕ:

$$\mu_0 = 2000$$
$$\sigma = 70$$
$$n = 30$$
$$\bar{x} = 1988$$
$$\alpha = 0{,}05$$

$$H_0\!: \mu \le 2000$$
$$H_1\!: \mu > 2000$$

Пошто је позната варијанса, а основни скуп нормално распоређен користи се статистика Z теста, док је ризик при закључивању 5%. На основу нулте хипотезе се може закључити да је критична област у потпуности са десне стране.

$$F(Z_\alpha) = 1 - \alpha = 1 - 0{,}05 = 0.95 \Rightarrow Z_\alpha = 1{,}65$$



Правила одлучивања гласе:
Ho не треба одбацити ако је Z≤1,65
Ho треба одбацити ако је Z >1,65

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{1988 - 2000}{\frac{70}{\sqrt{27}}} = -0{,}89$$

Закључак?

3. Један пољопривредни комбинат жели да утврди да ли у текућој години може очекивати са засијаних површина прошлогодишњи принос пшенице од 3 t/ha. Случајно одабраних 16 ha засијане површине дало је просјечан принос од 2,6 t/ha, са просјечним одступањем од 0,23 t. Утврдити уз 5% ризика да ли се може очекивати прошлогодишњи принос пшенице са засијаних површина, ако знамо да су засијане површине нормално распоређене према приносу.

Рјешење:

$$\mu_0 = 3,$$
$$n = 16,$$
$$\bar{x} = 2{,}6$$
$$s = 0{,}23$$

хипотезе:

$$H_0 : \mu = 3$$
$$H_1 : \mu \neq 3$$

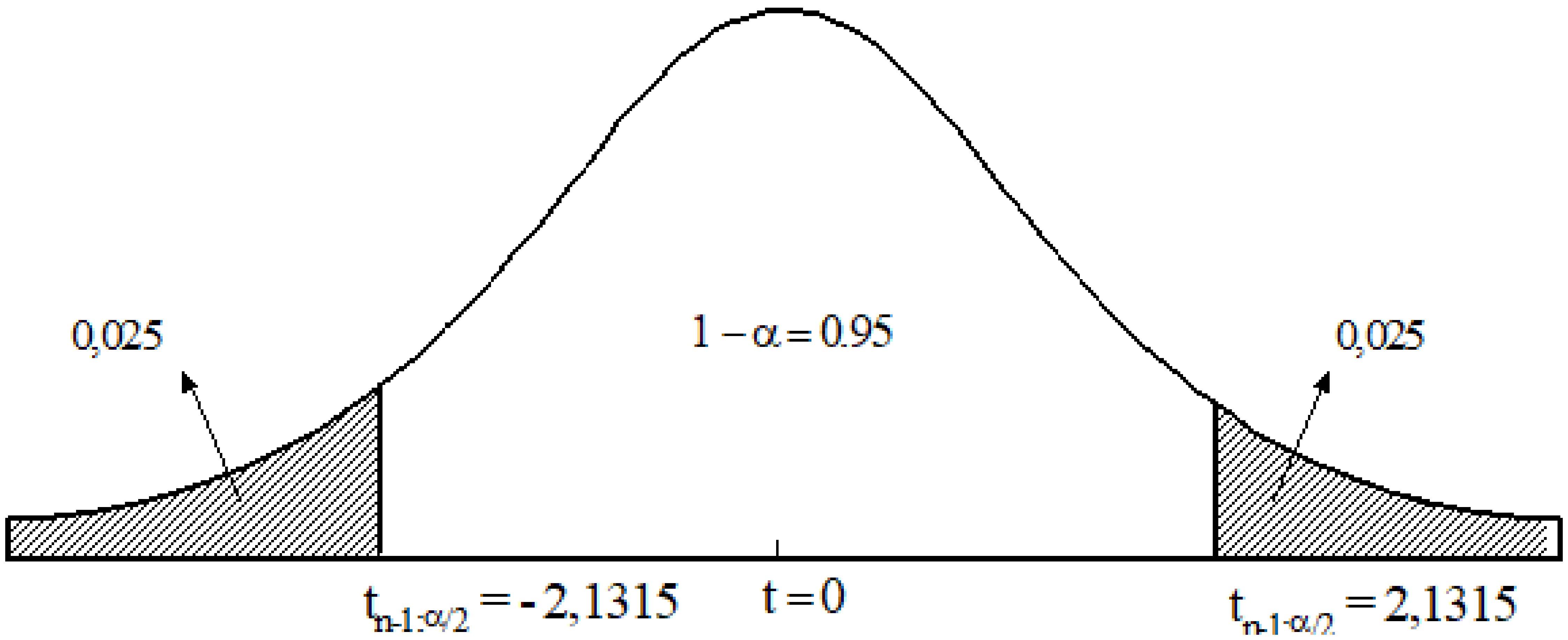
- **Пошто је непозната варијанса, а основни скуп нормално распоређен користи се статистика t теста, док је ризик при закључивању 5%.**

- **Критична област (област одбацивања нулте хипотезе) је распоређена симетрично на крајевима t распореда, па доњу и горњу вриједност налазимо у Таблицама…**

$t \rightarrow n-1=16-1=15$

$= \mathbf{2{,}1315}$

$t \rightarrow \alpha/2=0{,}05/2=0{,}025$



Правило одлучивања гласи:

Но неђемо одбацити ако је: $\left|t\right| \le t_{n-1;\alpha/2} (= 2{,}1315)$

Но ђемо одбацити ако је: $\left|t\right| > t_{n-1;\alpha/2} (= 2{,}1315)$

Реализована вриједност...

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{2{,}6 - 3}{\frac{0{,}23}{\sqrt{16}}} = \frac{-0{,}4}{0{,}0575} = -6{,}95$$

Закључак је...
Одбацујемо нулту хипотезу и уз 5% ризика закључујемо да се не може очекивати прошлогодишњи принос од 3 t/ha. Другим ријечима, овогодишњи принос пшенице се статистички значајно разликује од прошлогодишњег...

Б) Ако се постави питање (сасвим логично) да ли ђе принос пшенице бити већи од 2.5 t/ha, ствари се мијењају...

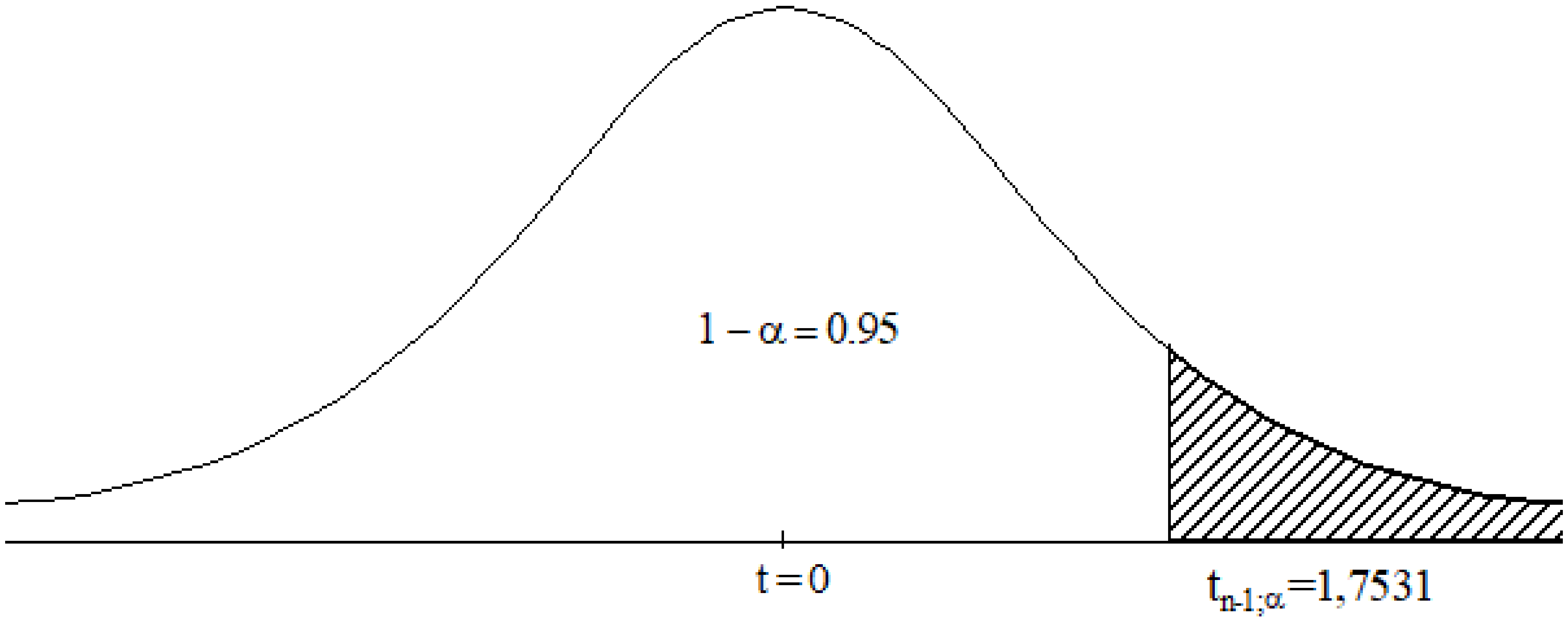
- Хипотезе ће гласити:

$$H_0: \mu \leq 2{,}5$$

$$H_1: \mu > 2{,}5$$

- Пошто је деснострани тест критична област је с једне стране, па имамо сљедећу ситуацију...

$$t \begin{cases} n-1=16-1=15 \\ \alpha=0{,}05 \end{cases} = \mathbf{1{,}7531}$$



Правила:

Но неђемо одбацити ако је: $t \le t_{n-1;\alpha}$

Но ђемо одбацити ако је: $t > t_{n-1;\alpha}$

Реализована вриједност:

$$t = \frac{2,6 - 2,5}{\frac{0,23}{\sqrt{16}}} = 1,7391$$

Закључак:

Прихвата се нулта хипотеза...

ТЕСТИРАЊЕ СТАТИСТИЧКЕ ХИПОТЕЗЕ ЗАСНОВАНО НА ЈЕДНОМ УЗОРКУ

ПРОПОРЦИЈА

1. На једној фудбалској утакмици случајно је изабрано 500 гледалаца међу којима је било 375 мушкараца. Испитати, уз 5% ризика, претпоставку да фудбалској утакмици присуствује:
   а) 70% мушкараца,
   б) мање од 30% гледалаца женског пола.

$$n = 500$$

$$f = 375$$

$$\pi_0 = 0{,}70$$

$$p = \frac{f}{n} = \frac{375}{500} = 0{,}75$$

Хипотезе

$$H_0: \pi = 0{,}70$$

$$H_1: \pi \neq 0{,}70$$

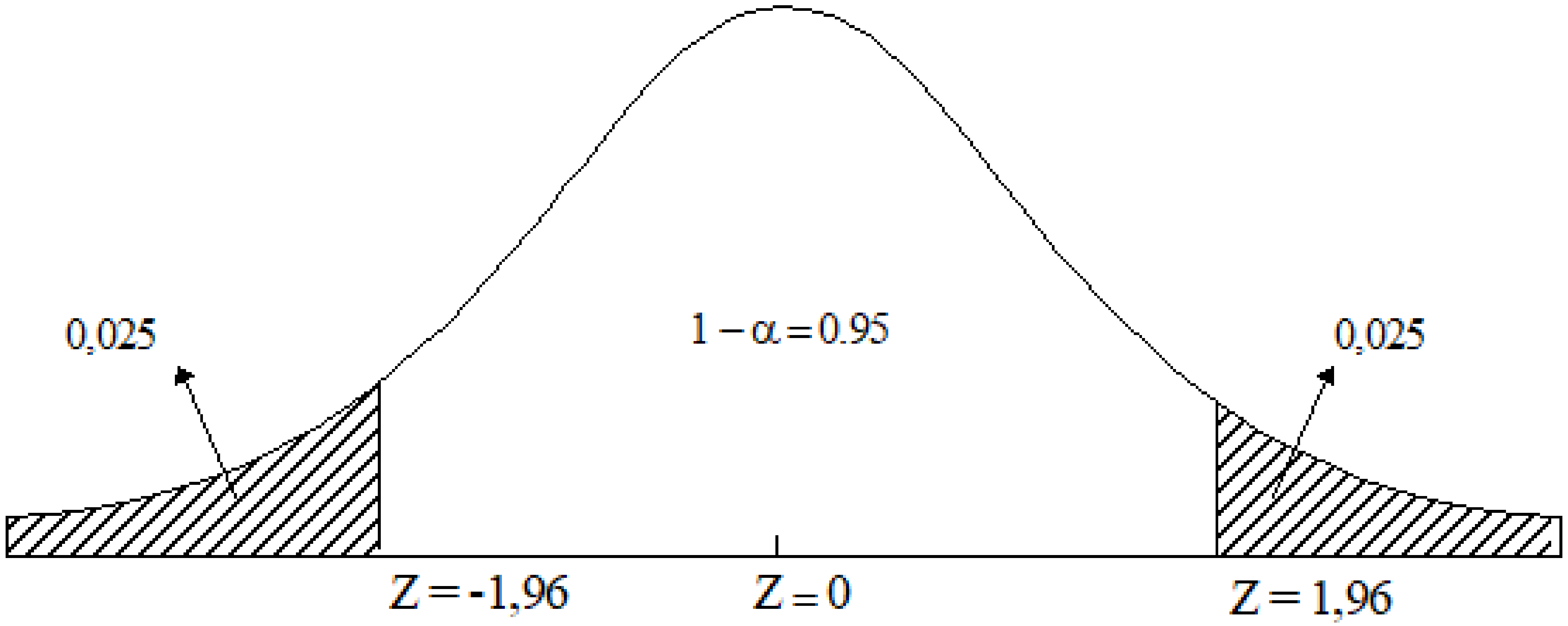
Увијек се користи статистика Z теста, ако важе сљедећи услови:

$$n \cdot \pi_0 > 5 \Rightarrow 500 \cdot 0{,}70 = 350 > 5$$

$$n \cdot (1 - \pi_0) > 5 \Rightarrow 500 \cdot (1 - 0{,}70) = 150 > 5$$

$$n \geq 30$$

Ризик је 5%, а критичне вриједности су распоређене симетрично, на крајевима нормалног распреда.



Правило одлучивања гласи:
Ho нећемо одбацити ако је: $|Z| \leq Z_{\alpha/2} (= 1{,}96)$
Ho ћемо одбацити ако је: $|Z| > Z_{\alpha/2} (= 1{,}96)$

Реализована вриједност износи:

$$Z = \frac{p - \pi_0}{s_p} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 \cdot (1 - \pi_0)}{n}}} = \frac{0{,}75 - 0{,}70}{\sqrt{\frac{0{,}70 \cdot (1 - 0{,}70)}{500}}} = \frac{0{,}05}{0{,}0205144} = 2{,}44$$

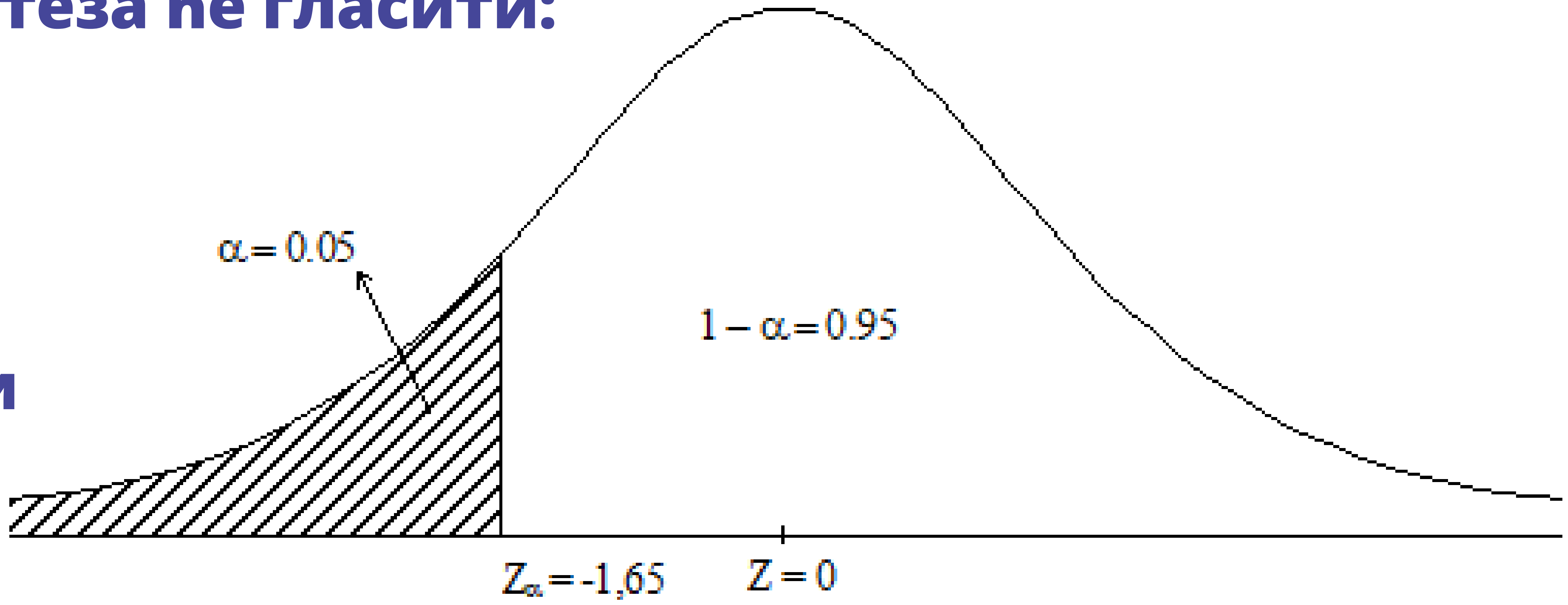
Закључак:
Одбацујемо нулту хипотезу и уз 5% ризика закључујемо да се учешће гледалаца мушког пола статистички значајно разликује од 70%.

б) нулта и алтернативна хипотеза ће гласити:

$$H_0 : \pi \geq 0{,}30$$

$$H_1 : \pi < 0{,}30$$

Овде се јасно види да се ради о једностраном тесту (лијево је критична област...)



- Са слике се јасно види да ћемо нулту хипотезу одбацити ако је реализована вриједност мања од -1,65 и обрнуто...

- Реализована вриједност:

$$Z = \frac{p - \pi_0}{s_p} = \frac{0{,}25 - 0{,}30}{0{,}0205134} = -2{,}44$$

- Закључак:

Одбацујемо нулту хипотезу и уз 5% ризика констатујемо да је учешће гледалаца женског пола мање од 30%.

2. Проценат масноће у једној литри млијека на једној фарми треба да износи најмање 3.5%. Случајно смо изабрали 250 литара млијека и утврдили да проценат масноће износи 3%. Испитати уз који највећи степен ризика можемо прихватити ову хипотезу.

ХВАЛА НА ПАЖЊИ!