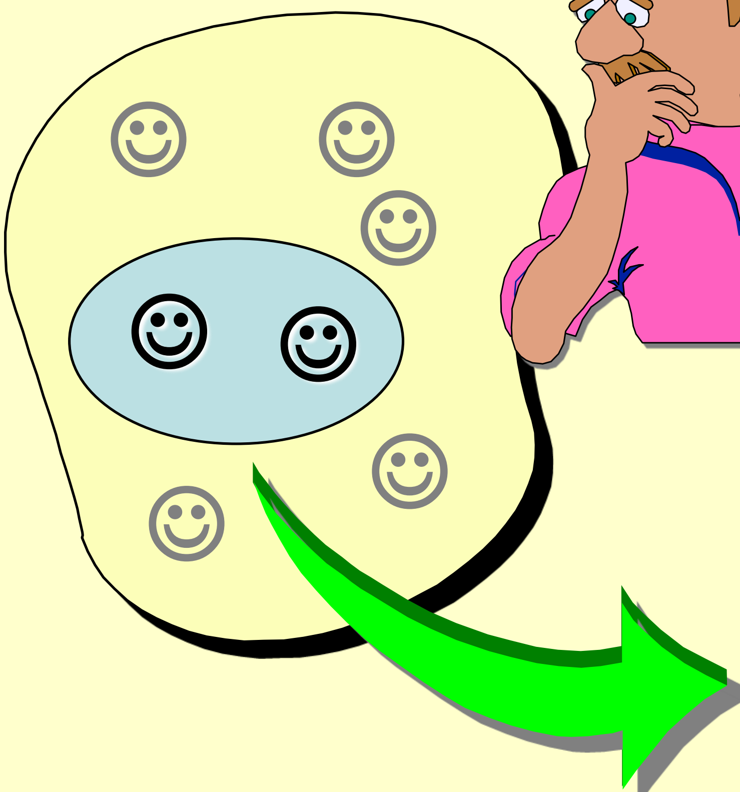
Тестване на хипотези

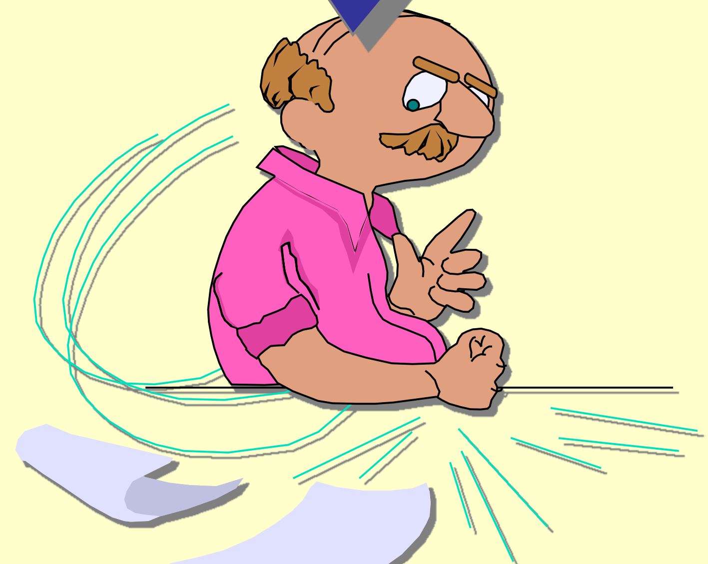
Използваме случайна извадка за да научим нещо за популацията

популация

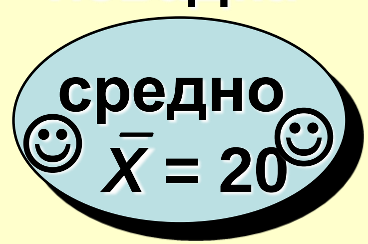


Вярвам, че средната възраст е 50 (хипотеза).

Не приемам хипотезата! Никак не си близо до истината.



Случайна извадка



Какво е хипотеза?

Недоказано твърдение за популационння параметър

Параметър е числова характеристика на популацията. Може да бъде:

- Популационна средна стойност,
- Популационна пропорция,
- Популационна дисперсия

• *Пример:* Верно ли е, че средната седмична сума, която студентите дават за алкохол е 10 лв?

Обща идея за тестване на хипотези

<u>Първо,</u> формулираме две хипотези

Нулева хипотеза

- “Убеждението” което подлежи на тестване
- Винаги съдържа знаците : $=$, $\leq$, или $\geq$
- Означение: H_0

Алтернативна хипотеза

- Противоположна на нулевата хипотеза
- Винаги съдържа знаците: $\neq$, $<$, or $>$
- Означение H_1

Пример:

H_0: $\mu \geq 3$ H1: $\mu<3$

Основни идеи за тестване

Първи начин:

Критична област → (Област на отхвърляне):

Идея: да намалим грешката от I род:

- **Грешка от I род**: Нулевата хипотеза се отхвърля, когато е верна.

Ниво на значимост **Означение α**

Избираме α = P(грешка от I род) =
P(H_0 да се отхвърли, ако H_0 е верна)
=> **Типични стойности** **0,01** **0,05** **0,1**

Построяваме област на отхвърляне (критична област), така че лицето над нея под кривата на плътността =α

Тестване на хипотези относно популационната средна стойност μ

Видове тестове

$H_0: \mu = константа$

Зависи от вида на алтернативната хипотеза

Едностранен тест

$H_1: \mu < константа$

Лявостранен тест

или

$H_1: \mu > константа$

дясностранен тест

Двустранен тест

$H_1: \mu \neq константа$

еквивалентно на

$H_1: \mu > константа$

$\mu < константа$

Два начина

Критична област

p-стойност

Алгоритъм за тестване на хипотези

критична област

1. Напишете нулевата H_0 и алтернативната H_1 хипотеза

2. Определете нивото на значимост α

3. Определете статистиката и извадковото разпр.

4. Получете критичната област

5. Направете извод

6. Интерпретация на извода

Тестване на хипотези за неизвестното популационно средно μ (σ е известно)

Предположения

– Популацията е нормално разпределена

– Ако не е нормална,то при голям обем на извадката (ЦГТ) можем да апроксимираме с нормална ($n \geq 30$)

- Популационната дисперсия (станд.откл.) σ^2 е известно

Да припомним, че $\bar{X} \in N(\mu, \frac{\sigma^2}{n})$

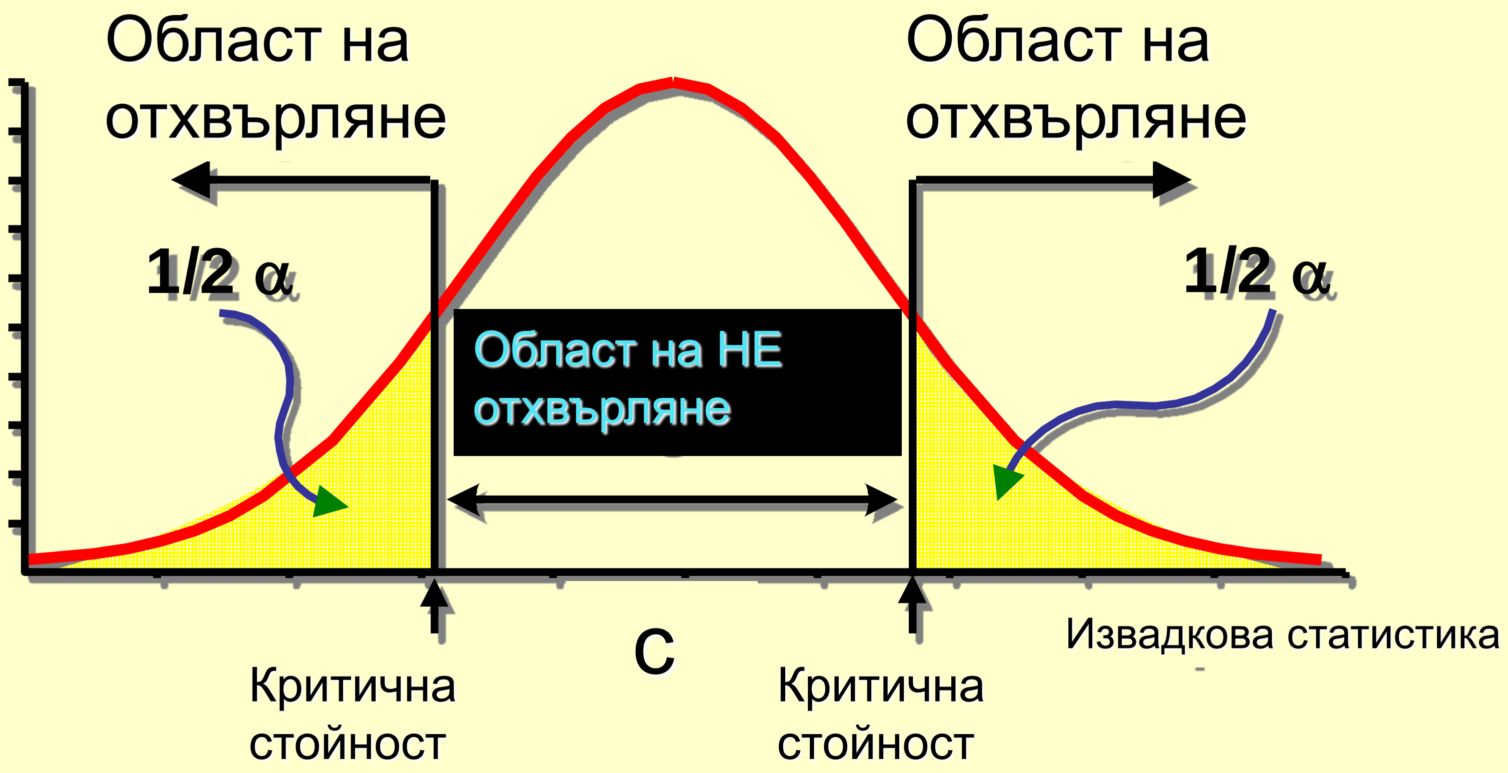
Използваме Z-статистиката

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Двустранен Z –тест за средното μ
(σ Известно)

$H_0\!: \mu = c$ $\qquad\qquad$ $H_1\!: \mu \neq c$ (означава < или >)

Избираме ниво на значимост α



Критична област при двустранен Z – тест за средното μ (σ Известно)

Използваме избраното от нас ниво на значимост α, за да намерим

$$z_\alpha : \qquad P(Z > z_\alpha) = \alpha$$

Критичната област се състои от два интервала (два клона) (- ∞, - $z_{1-\frac{\alpha}{2}}$) и ($z_{1-\frac{\alpha}{2}}$, +∞).

Как се взема решение

Ако статистиката $Z = \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n}$ за конкретната извадка

•има стойност в критичната област – то **отхвърляме** H0
•ако стойността й е извън КО – **няма основание да отхвърлим** H0

Едностранен тест за средната стойност μ (σ известно)

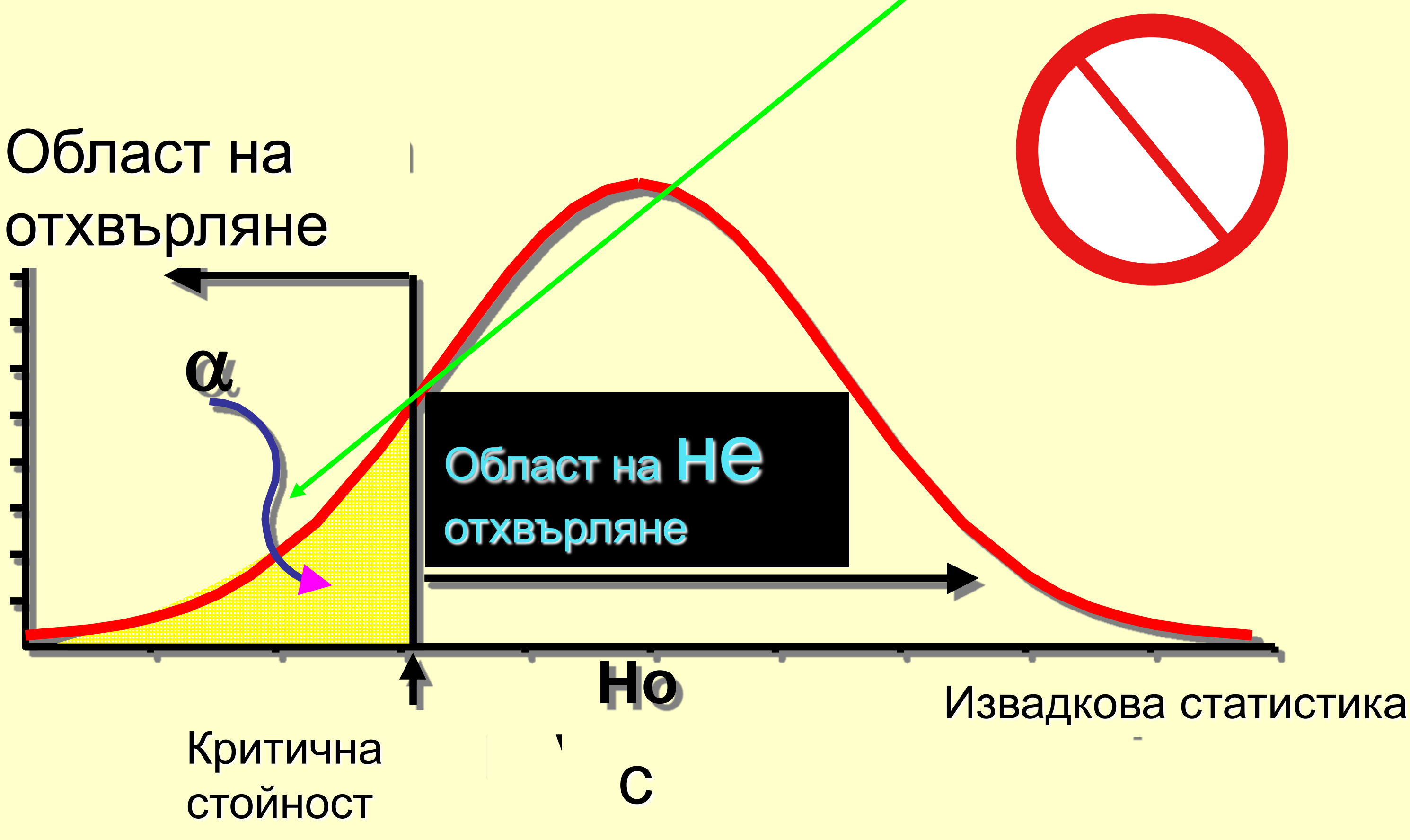
- **Предположения**
  - **Популацията е нормално разпределена**
  - **Ако не е нормална, можем да я апроксимираме с нормална ($n \geq 30$)**
- **Алтернативната хипотеза има знаците $<$ или $>$**
- **Z-статистика**

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Лявостранен тест

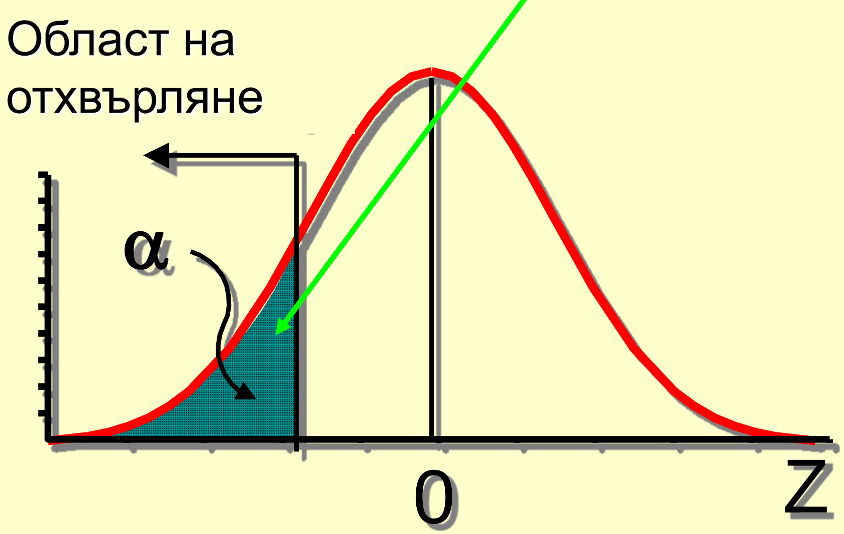
$$H_0\colon \mu \geq c$$

$$H_1\colon \mu < c$$

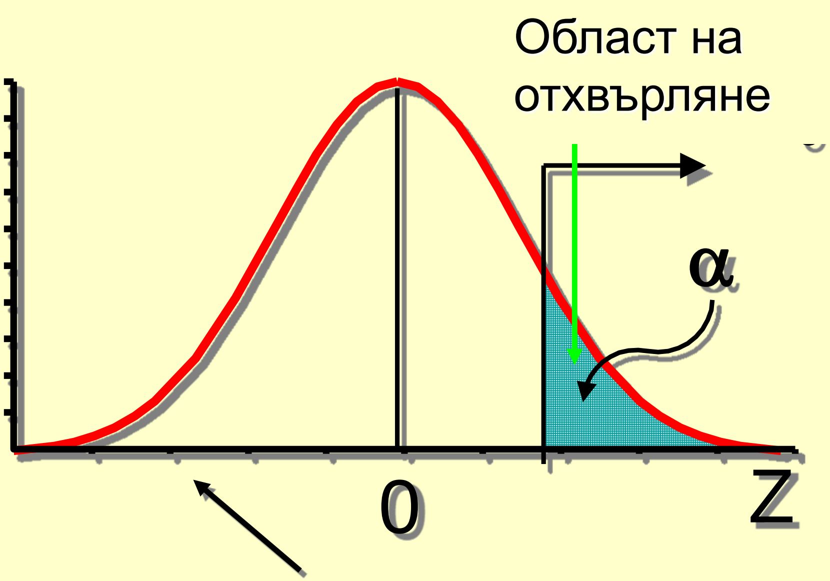


Едностранен Z тест за средна стойност

$H_0: \mu = 0 \quad H_1: \mu < 0$



$H_0: \mu = 0 \quad H_1: \mu > 0$



Малки стойности не противоречат на H_0 - **не** се отхвърля!

Проверка на хипотези за средната стойност μ, когато популационната дисперсия е неизвестна (σ неизвестно)

Предположения

- Популацията е нормално разпределена
- Малък обем (n≤30)

■Използваме t -разпределение и следната статистика

$$t = \frac{\bar{X} - \mu}{\dfrac{S}{\sqrt{n}}}$$

t-тест за неизвестното популационно средно

Тестване на хипотези относно популационната средна стойност μ

Два начина

Критична област

p-стойност

Изчисляване на р-стойност за проверка на хипотези

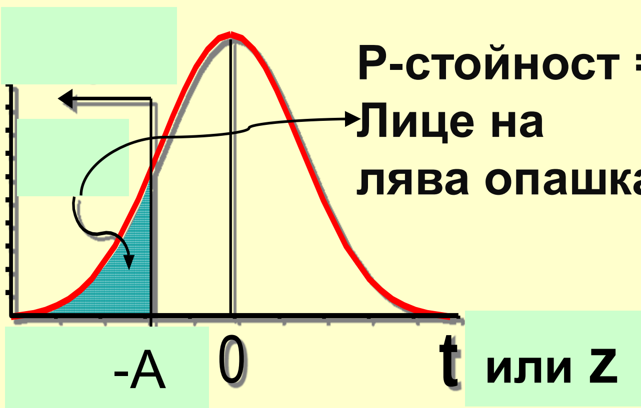
- За дадена извадка се изчислява статистиката (например, извадково средно) и получената число се замества в $Z = \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n}$

или в $t = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$

Тогава **р-стойността** се пресмята като лице на:

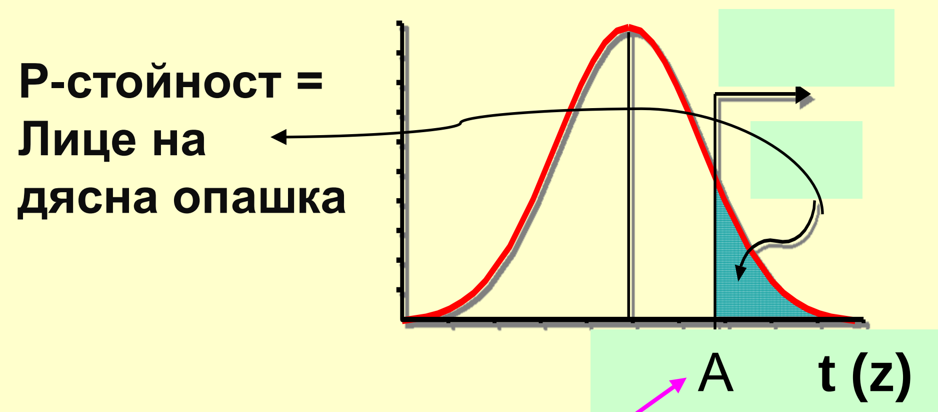
Определяне за р-стойност= Лице

Едностранен, ляв тест



Стойност на
статистиката

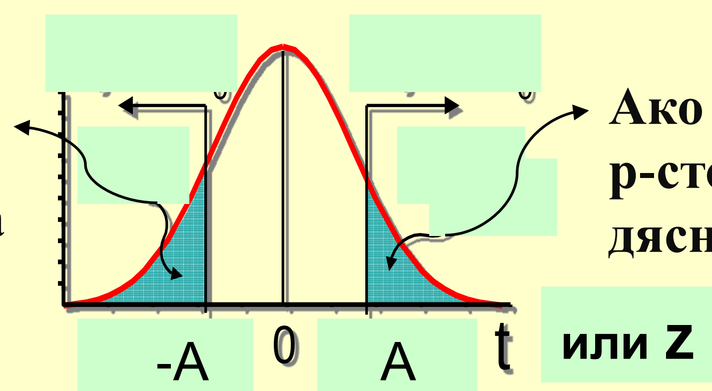
Едностранен, десен тест



Стойност на
статистиката

Двустранен тест

Ако *A* е отрицателно, р-стойност = 2* лицето на лявата опашка



Ако *A* е положително, р-стойност= 2* лицето на дясната опашка

Ст.ойност на статистика

Използване на p-стойност за проверка на хипотези

- За дадена извадка, **p-стойността** дава вероятността да се наблюдават и други екстремни извадки, ако нулевата хипотеза беше вярна.
- p-стойността е вероятност, т.е. това е число между 0 и 1.
- Близко до 0 означава "невъзможно"
- И така, ако p-стойността е "малка" (типично, по-малко от 0,05), тогава отхвърляме нулевата хипотеза.

Отхвърляме H_0 ако стойността е малка

***P*-стойност**	**Интерпретация**
По-малко от 0,01	**Голяма статистическа значимост. Много строги докателства срещу нулевата хипотеза**
От 0,01 до 0,1	**Статистически достатъчни доказателства срещу нулевата хипотеза.**
По-голямо от 0,1	**Недостатъчно основание за отхвърляне на нулевата хипотеза**

Алгоритъм за тестване на хипотези

(p-стойност)

1. Напишете нулевата и алтернативната хипотеза

2. Определете статистиката и извадковото разпределение

3. Пресметнете P-стойността за статистиката на теста

4. Направете извод

5. Дайте интерпретация на вашия извод