Проверка на хипотези: t-тест и Z-тест

Нулева и алтернативна хипотеза, p-стойност и критична стойност; Z-тест за един и за два относителни дяла; Z-тест и t-тест за една средна, за две независими и за две свързани извадки

Упражнение 4 · Медицинска статистика · 2026/2027

гл. ас. д-р Костадин Костадинов, д.м.

Катедра по социална медицина и обществено здраве · Медицински университет – Пловдив

Какво ще можете след това упражнение

1. Да формулирате **нулева** и **алтернативна** хипотеза за популационни параметри.
2. Да вземате решение по **критична стойност** и по **p-стойност** и да формулирате извода правилно.
3. Да прилагате **Z-тест** за един и за два относителни дяла.
4. Да прилагате **Z-тест** и **t-тест** за една средна, **t-тест** за две независими и за две свързани извадки.
5. Да свързвате резултата от теста с **доверителния интервал на разликата**.

План на часа (90 мин)

1. Домашна 3: отговори — 5 мин
2. Логика на проверката — 15 мин
3. Z-тест за относителен дял — 16 мин
4. Тестове за средни — 24 мин
5. Задачи в час — 12 мин
6. Текуща проверка (1 т.) — 15 мин

Калкулатор и лист с формули. По учебника: двустранни тестове, $\alpha = 0{,}05$.

Домашна 3: отговори

1 · Наднормено тегло (84/350)

- $\hat{p} =$ **24,00 %**; $S_p =$ **2,28 п.п.**
- 95 % CI **[19,53; 28,47] %**
- ± 3 п.п.: **779** (при 24 %); **1 068** (при 50 %)
- ± 2 п.п.: **1 752** — d е на квадрат

2 · Систолно налягане

- А: **[128,84; 135,96]**; Б: **[134,24; 141,96]** mmHg — припокриват се
- разлика **–5,7** mmHg; $\mathrm{SE}_\Delta = 2{,}596$; $t = 2{,}021$
- 95 % CI **[–10,95; –0,45]** mmHg — **значима**

3 · Пушене I / VI курс

- I: 25,83 % **[20,30; 31,37]**; VI: 18,64 % **[13,49; 23,78]**
- разлика **7,20 п.п.**; $\mathrm{SE}_\Delta = 3{,}857$
- 95 % CI **[–0,36; 14,76]** п.п. — **не е значима**; не доказва липса на разлика

СЪДЪРЖАНИЕ

1. **Логика на проверката на хипотези**
2. Z-тест за относителен дял
3. Z-тест и t-тест за средни
4. Задачи в час, домашна работа и текуща проверка

Нулева и алтернативна хипотеза

Нулева хипотеза H_0 (работна)

Няма разлика или ефект **в популацията**:
$H_0 : \mu_1 = \mu_2$ или $H_0 : P_1 = P_2$.
Приема се временно за вярна, за да изчислим колко необичайни са данните.

Алтернативна хипотеза H_1

Разлика има: $H_1 : \mu_1 \neq \mu_2$ (двустранна).
В курса — само двустранни тестове.

Пример (учебник)

Две антихипертензивни терапии; средно систолно налягане 137 и 127 mmHg.
H_0: средното систолно налягане в популациите, лекувани с A и с B, е еднакво ($\mu_A = \mu_B$).
H_1: различно е ($\mu_A \neq \mu_B$).

Неправилно

„H_0: средното в **извадката** е равно на...“ — за извадката знаем стойностите; хипотезите са за **параметри**.

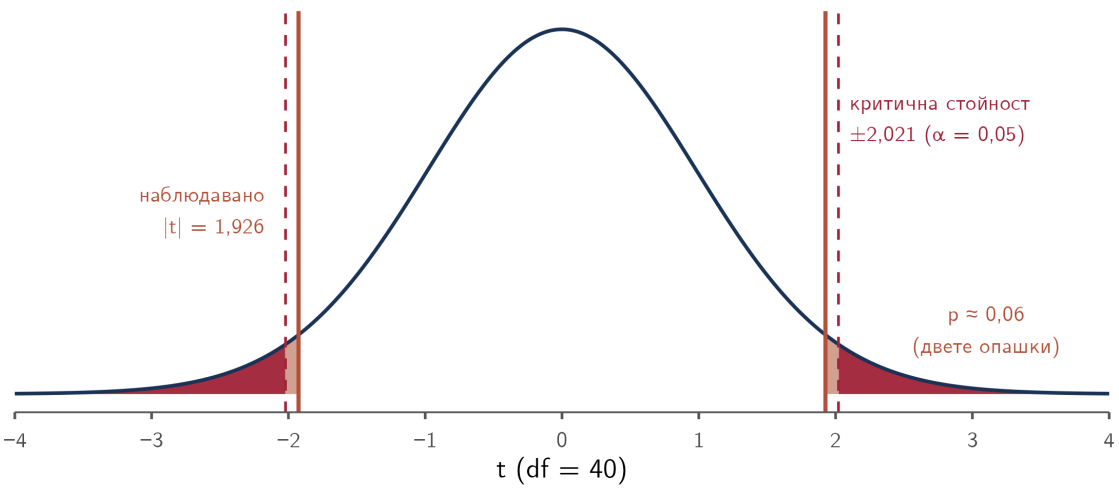
Пет стъпки на всеки тест

№	**Стъпка**	**Пример: две терапии, систолно налягане**
1	Хипотези H_0 и H_1	$H_0 : \mu_A = \mu_B$; $H_1 : \mu_A \neq \mu_B$
2	Ниво на значимост α	$\alpha = 0{,}05$ (двустранно)
3	Тестова статистика и степени на свобода	$t = 1{,}926$; $\mathrm{df} \approx \min(40; 40) = 40$
4	Критична стойност **или** p-стойност	$t_{\text{кр}} = 2{,}021$; p между 0,05 и 0,10
5	Решение и извод с думи, размер на ефекта	не отхвърляме H_0; няма **статистически значима** разлика (разлика 10 mmHg)

Критична стойност или p-стойност

|t| = 1,926 не стига до критичната стойност 2,021

Оранжево + тъмночервено: p-стойността (≈ 0,06). Тъмночервено: областта на отхвърляне (5 %).
p > α → не отхвърляме H_0



Класически подход

$|t_{\text{емп}}| > t_{\text{кр}} \longrightarrow$ отхвърляме H_0.
Тук $1{,}926 < 2{,}021 \longrightarrow$ не отхвърляме.

p-стойност от таблицата

Ред $\text{df} = 40$: 1,684 · 2,021 · 2,704 (колони 0,10 · 0,05 · 0,01).
$1{,}684 < 1{,}926 < 2{,}021 \longrightarrow$ **0,05 < p < 0,10**
(софтуер, $\text{df} = 40$: $p = 0{,}061$)

Колоните на t-таблицата в листа ($t_{0,05}$, $t_{0,025}$, $t_{0,005}$) са за двустранни $\alpha = 0{,}10$; 0,05; 0,01.

Правило за p-стойността

p	Решение	Извод
$p < \alpha$	отхвърляме H_0	има **статистически значима** разлика
$p \geq \alpha$	не отхвърляме H_0	няма **статистически значима** разлика

p-стойността е вероятността — **ако H_0 е вярна** и допусканията на теста са изпълнени — да се получи тестова статистика поне толкова крайна, колкото наблюдаваната (при двустранен тест — по абсолютна стойност). Тя не е вероятността H_0 да е вярна.

Неправилни изводи

- $p = 0{,}06 \longrightarrow$ „няма разлика“, „терапиите са еднакво ефективни“
- $p = 0{,}001 \longrightarrow$ „ефектът е голям“
- „приемаме H_0“

Правилно

„Разликата от 10 mmHg не е статистически значима ($p = 0{,}06$); данните не позволяват да отхвърлим H_0.“ Добавете размера на ефекта и 95 % CI.

Грешки от I и II род

Решение \ истина	H_0 **е вярна (няма ефект)**	H_0 **е невярна (има ефект)**
отхвърляме H_0	**грешка I род** (фалшиво положителен), вероятност α	вярно решение; **мощност** $1-\beta$
не отхвърляме H_0	вярно решение	**грешка II род** (фалшиво отрицателен), вероятност β

При фиксирана извадка по-малко α $\longrightarrow$ повече β. Мощността расте с по-голям ефект, по-голямо n и по-малко разсейване.

СЪДЪРЖАНИЕ

1. Логика на проверката на хипотези
2. **Z-тест за относителен дял**
3. Z-тест и t-тест за средни
4. Задачи в час, домашна работа и текуща проверка

Z-тест за един дял

$$Z = \frac{\hat{p} - P_0}{\sqrt{\frac{P_0(100-P_0)}{n}}}$$

P_0 — стойността по H_0 (референтна, теоретична). Грешката се смята с P_0, защото тестът предполага, че H_0 е вярна. Формулата не е в листа с формули.

Условие

Очакваните броеве $nP_0/100 \geq 10$ и $n(100 - P_0)/100 \geq 10$. Тук 160 и 40.

Пример (по учебника)

Нова терапия за рак на белия дроб: **146 от 200** пациенти (73 %) преживяват поне година; референтна ефективност 80 %.

1. $H_0 : P = 80\%$; $H_1 : P \neq 80\%$; $\alpha = 0{,}05$
2. $S = \sqrt{80 \times 20/200} = 2{,}828$
3. $Z = (73 - 80)/2{,}828 = -2{,}47$
4. $|Z| = 2{,}47 > 1{,}96$ (и $< 2{,}576$) $\rightarrow$ $0{,}01 < p < 0{,}05$ (софтуер: 0,013)

Отхвърляме H_0: едногодишната преживяемост е статистически значимо по-ниска от 80 %.

Z-тест за два дяла: безглутенова диета

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(100-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \quad \hat{p} = \frac{m_1 + m_2}{n_1 + n_2} \times 100$$

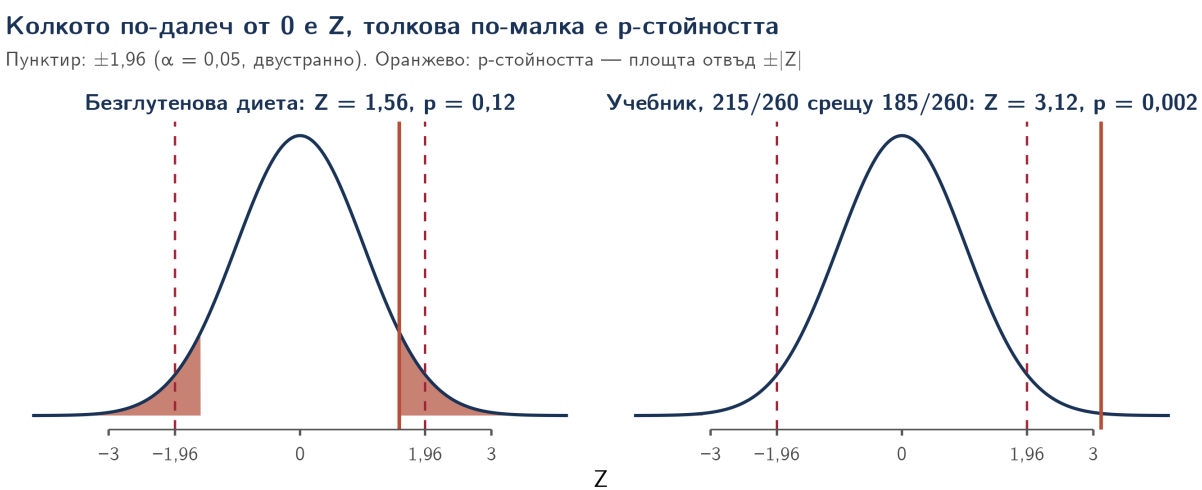
В листа с формули — с дроби: $p(1-p)$. $\hat{p}$ — **обединеният** дял.

Диета	С подобрение	n	$\hat{p}$
безглутенова	56	100	56 %
контролна	45	100	45 %

1. $H_0 : P_1 = P_2$; $H_1 : P_1 \neq P_2$; $\alpha = 0{,}05$
2. $\hat{p} = 101/200 \times 100 = 50{,}5\%$
3. $Z = 11/\sqrt{50{,}5 \times 49{,}5 \times (1/100 + 1/100)} = 11/7{,}071 = 1{,}56$
4. $1{,}56 < 1{,}645 \longrightarrow$ **p > 0,10** (софтуер: 0,12)

Не отхвърляме H_0: няма статистически значима разлика (разлика 11 п.п.; 95 % CI [–2,77; 24,77] п.п.).

Решен пример от учебника: 215/260 срещу 185/260



Едногодишна преживяемост без прогресия: А **82,69 %**, Б **71,15 %**.

- $\hat{p} = 400/520 = 76{,}92\%$
- $Z = 11{,}54/\sqrt{76{,}92 \times 23{,}08 \times 2/260} = 11{,}54/3{,}695 = 3{,}12$
- $3{,}12 > 2{,}576 \rightarrow$ **p < 0,01** (0,002)

Отхвърляме H_0: разликата от 11,5 п.п. е статистически значима.

95 % CI на разликата: [4,36; 18,71] п.п. — не съдържа 0.

Z-тест и интервал на разликата: една логика, две стандартни грешки

	Z-тест	**95 % CI на разликата**
въпрос	съвместими ли са данните с $P_1 = P_2$?	в какви граници е $P_1 - P_2$?
стандартна грешка	**обединен** дял: $\sqrt{\hat{p}(100-\hat{p})(1/n_1 + 1/n_2)}$	**отделни** дялове: $\sqrt{\hat{p}_1(100-\hat{p}_1)/n_1 + \hat{p}_2(100-\hat{p}_2)/n_2}$
диета	$Z = 1{,}56$; $p = 0{,}12$	[–2,77; 24,77] — съдържа 0
учебник	$Z = 3{,}12$; $p = 0{,}002$	[4,36; 18,71] — не съдържа 0

Добрата практика: съобщавайте и p-стойността, и разликата с 95 % CI.

СЪДЪРЖАНИЕ

1. Логика на проверката на хипотези
2. Z-тест за относителен дял
3. **Z-тест и t-тест за средни**
4. Задачи в час, домашна работа и текуща проверка

Z-тест за една средна: известно σ

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$

Използва се, когато популационното σ е **известно** (напр. от национален регистър). При неизвестно σ — t-тест; при много голямо n двата съвпадат (ред ∞ в таблицата).

Пример

Средното систолно налягане на 100 работници в нощни смени е 128,6 mmHg. В общата популация на същата възраст $\mu_0 = 125$ mmHg. За целите на задачата приемаме, че стандартното отклонение и при тези работници е известно: $\sigma = 15$ mmHg.

1. $H_0 : \mu = 125$; $H_1 : \mu \neq 125$
2. $Z = (128{,}6 - 125) / \left(15/\sqrt{100}\right) = 3{,}6 / 1{,}5 = 2{,}40$
3. $1{,}96 < 2{,}40 < 2{,}576 \rightarrow 0{,}01 < p < 0{,}05$ (0,016)

Отхвърляме H_0: средното налягане е статистически значимо по-високо (с 3,6 mmHg).

t-тест за една извадка

$$t = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}, \quad \mathrm{df} = n - 1$$

Сравнява средната на една извадка с референтна стойност. Условия: независими наблюдения; при малко n — приблизително нормално разпределение без силно отклоняващи се стойности. Формулата не е в листа, но знаменателят е същият като за CI: $S_{\overline{X}} = S/\sqrt{n}$.

Пример: пулс на 10 студенти (упр. 2)

$\overline{X} = 80{,}8$; $S = 6{,}529$ уд./мин. Различава ли се средният пулс от 72 уд./мин?

1. $H_0 : \mu = 72$; $H_1 : \mu \neq 72$
2. $S_{\overline{X}} = 6{,}529/\sqrt{10} = 2{,}065$
3. $t = 8{,}8/2{,}065 = 4{,}26$; $\mathrm{df} = 9$
4. ред 9: $4{,}26 > 3{,}250 \rightarrow$ **p < 0,01** (0,002)

Отхвърляме H_0: средният пулс е статистически значимо по-висок от 72 уд./мин; 95 % CI [76,13; 85,47].

t-тест за две независими извадки (лист с формули)

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}, \quad \mathrm{df} \approx \min(n_1 - 1, n_2 - 1)$$

Вариант с **различни** дисперсии (Уелч); приложим и при равни. $\mathrm{df} = \min$ е консервативно приближение; софтуерът изчислява точните df на Уелч.

Терапия	n	$\overline{X}$, mmHg	S
A	41	137	31
B	41	127	12

1. $H_0 : \mu_A = \mu_B$; $\alpha = 0{,}05$
2. $\mathrm{SE} = \sqrt{31^2/41 + 12^2/41} = \sqrt{26{,}95} = 5{,}191$
3. $t = 10/5{,}191 = 1{,}926$; $\mathrm{df} = 40$
4. $1{,}684 < 1{,}926 < 2{,}021 \rightarrow$ **0,05 < p < 0,10** (0,06)

Не отхвърляме H_0: разликата от 10 mmHg не е статистически значима.

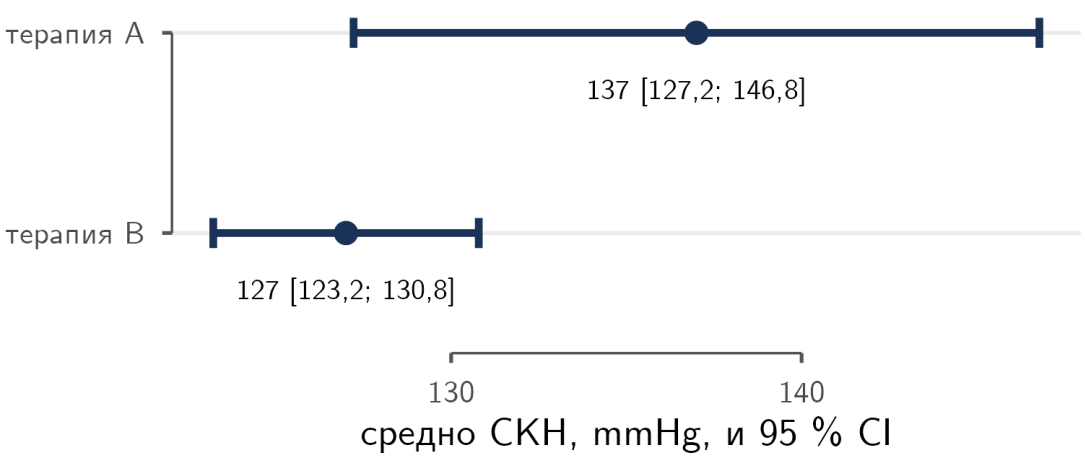
95 % CI на разликата: [–0,49; 20,49] mmHg — съдържа 0, но допуска и голям ефект.

Тестът и интервалът: едно решение

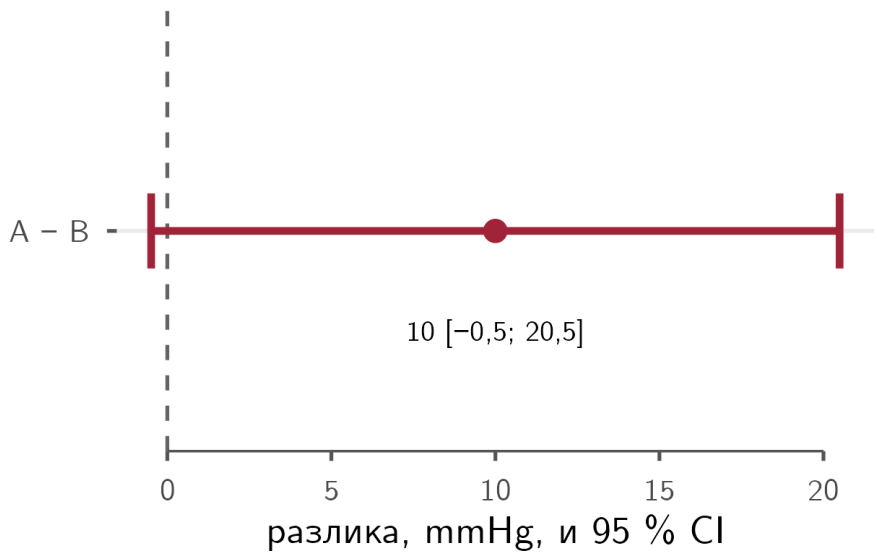
t-тестът и интервалът на разликата казват едно и също

t = 1,926 < 2,021 (p ≈ 0,06) ⇔ 95 % CI на разликата съдържа 0. Разлика от 10 mmHg обаче не е „сходен резултат“

Групите



Разликата



Интервалът показва повече от p-стойността: данните са съвместими както с липса на разлика, така и с разлика от около 20 mmHg. Учебникът (с. 108) заключава, че терапиите са „свързани със сходни клинични резултати“ — $p > 0{,}05$ не стига за това. Извод за клинична еквивалентност изисква предварително определени граници на приемлива разлика и специален анализ.

t-тест с равни дисперсии (обединено стандартно отклонение)

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_p^2(1/n_1 + 1/n_2)}}, \quad \mathrm{df} = n_1 + n_2 - 2$$

Не е в листа с формули. Предполага **равни популационни дисперсии**; тогава е малко по-мощен. Незначим тест на Левин не доказва равенство — затова в задачите използваме варианта на Уелч от листа.

Същият пример

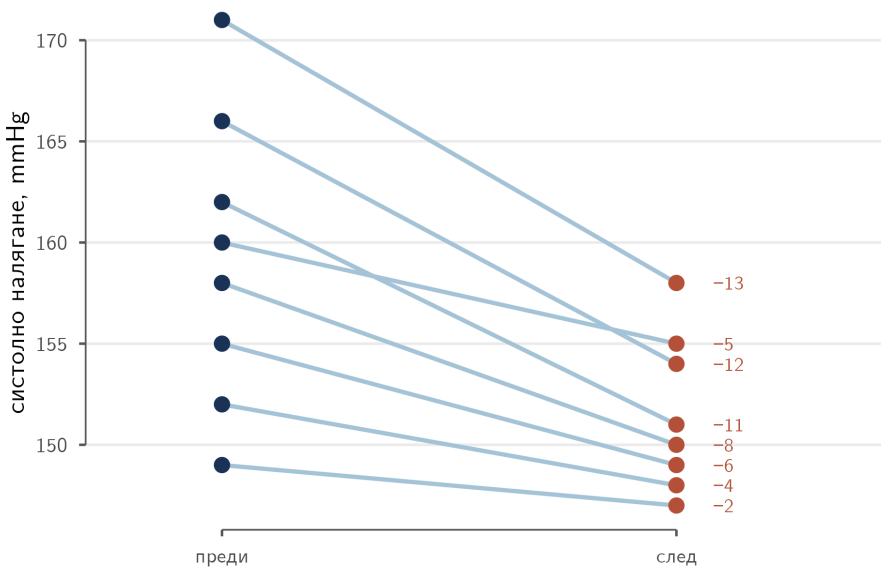
$S_p = \sqrt{(40 \times 31^2 + 40 \times 12^2)/80} = 23{,}51$
$t = 10/\left(23{,}51 \times \sqrt{2/41}\right) = 1{,}926$, $\mathrm{df} = 80 \rightarrow$ ред 50: $0{,}05 < p < 0{,}10$ (софтуер: 0,058)

При $n_1 = n_2$ двата варианта дават едно и също t. Тук стандартните отклонения явно се различават (31 срещу 12 mmHg; дисперсии 961 и 144) $\rightarrow$ вариантът на Уелч от листа.

t-тест за две свързани извадки

Всеки пациент е своя контрола

Анализира се разликата „след − преди“: средно −7,6 mmHg; t = −5,35, df = 7



$$t = \frac{\overline{d}}{S_d/\sqrt{n}}, \quad \mathrm{df} = n - 1$$

d — разликата „след – преди“ за всеки пациент. Условие: разликите да са приблизително нормално разпределени.

$H_0 : \mu_d = 0$; $H_1 : \mu_d \neq 0$ $(d = \text{след} - \text{преди})$

Разлики: –8, –11, –2, –13, –6, –12, –5, –4 mmHg

1. $\overline{d} = -61/8 = -7{,}625$ mmHg; $S_d = 4{,}033$
2. $t = -7{,}625 / \left(4{,}033/\sqrt{8}\right) = -5{,}35$; $\mathrm{df} = 7$
3. $5{,}35 > 3{,}499 \longrightarrow$ **p < 0,01** (0,001)

Систолното налягане спада статистически значимо: средно с 7,6 mmHg (95 % CI 4,3–11,0).

Кой тест? (извадка от таблицата в листа с формули)

Признак	Групи	Условие	Тест
количествен	1	нормално разпределение	t-тест за една извадка (Z при известно σ)
количествен	2 независими	нормално разпределение	t-тест за две независими извадки
количествен	2 свързани	нормални разлики	t-тест за две свързани извадки
количествен	2 независими	не е нормално	U-тест на Mann–Whitney
относителен дял	1	$nP_0/100 \geq 10$ и $n(100 - P_0)/100 \geq 10$	Z-тест за един дял
относителен дял	2 независими	поне 10 със и 10 без събитието във всяка група	Z-тест за два дяла (или χ^2)

Чести грешки

Грешка	**Правилно**
хипотези за извадката	хипотези за **популационните** параметри
$p = 0{,}06 \longrightarrow$ „няма разлика“	„няма **статистически значима** разлика“
„приемаме H_0“	„не отхвърляме H_0“
S вместо $S/\sqrt{n}$ в знаменателя	в знаменателя е **стандартната грешка**
Z-тест за два дяла с отделните дялове в SE	за два дяла — **обединеният** дял; за един — P_0
свързани данни като независими групи	t-тест за **разликите**
само p-стойност в извода	разлика + 95 % CI + p

СЪДЪРЖАНИЕ

1. Логика на проверката на хипотези
2. Z-тест за относителен дял
3. Z-тест и t-тест за средни
4. **Задачи в час, домашна работа и текуща проверка**

Задачи в час

А · Множествена склероза

400 пациенти: **25 от 200** на експериментална терапия и **45 от 200** на конвенционална рецидивират до 6 месеца след терапията.

Има ли статистически значима разлика в дела на рецидивиралите? Посочете и 95 % CI на разликата.

Б · Холестерол

Нов продукт за понижаване на холестерола, по 30 пациенти в група. След 3 месеца ($\overline{X} \pm S_{\overline{X}}$): експериментална **213 ± 6** mg/dL, контролна **228 ± 7** mg/dL.

Изчислете t и определете интервала на p-стойността.

Подсказка: $S^2/n = S^2_{\overline{X}}$.

Решение

А. $\hat{p}_1 = 12{,}5\%$, $\hat{p}_2 = 22{,}5\%$; $\hat{p} = 70/400 = 17{,}5\%$

$Z = -10/\sqrt{17{,}5 \times 82{,}5 \times 2/200} = -10/3{,}800 = -2{,}63$

$|Z| = 2{,}63 > 2{,}576 \rightarrow$ **p < 0,01** (0,0085)

Отхвърляме H_0: рецидивите са с 10 п.п. по-редки при експерименталната терапия; 95 % CI [–17,38; –2,62] п.п.

Б. $t = (213 - 228)/\sqrt{6^2 + 7^2} = -15/9{,}220 = -1{,}63$

$\mathrm{df} = \min(29; 29) = 29$: $1{,}63 < 1{,}699 \rightarrow$ **p > 0,10** (0,11)

Не отхвърляме H_0: разликата от 15 mg/dL не е статистически значима.

Тук $n = 30$ в група. Ако обемите (или df) не са дадени, t може да се изчисли, но p-стойността не може да се определи от таблицата.

Обобщение и домашна работа

Да запомним

1. Хипотезите са за популацията; тестваме H_0.
2. $p < \alpha \longrightarrow$ отхвърляме H_0; $p \geq \alpha \longrightarrow$ няма **статистически значима** разлика.
3. Дялове: Z-тест (един дял — SE с P_0; два дяла — обединен дял); средни: t-тест (df!), Z само при известно σ.
4. Свързани данни $\longrightarrow$ t-тест за разликите.
5. Изводът: разлика + 95 % CI + p.

Домашна работа

Четири задачи — на сайта, „Медицинска статистика“: **kostadinoff.github.io**

- Z-тест за два дяла и за един дял
- t-тест за две независими и за две свързани извадки

Следващия път: **χ^2-тест**. В края — текуща проверка (1 т.) по **t-тест и Z-тест**.

Текуща проверка (1 точка): CI за дял и сравнение по CI

Правила

- **15 минути**, самостоятелно
- един казус на човек; съседите имат **различни** варианти
- позволени: калкулатор и лист с формули
- напишете име, факултетен номер и група

Какво се оценява

1. дял и 95 % CI в първата група
2. дял и 95 % CI във втората група
3. припокриват ли се интервалите
4. разлика, SE_{Δ} и 95 % CI на разликата
5. **заключение** в процентни пунктове

Благодаря за вниманието!