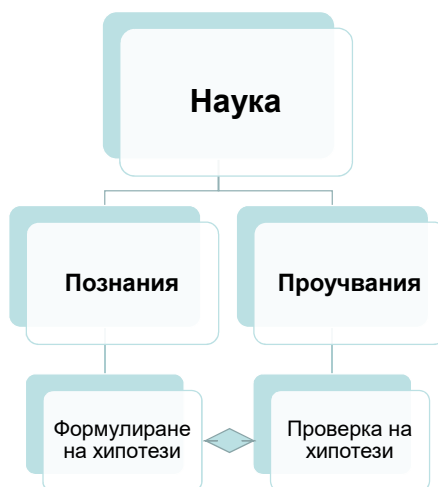
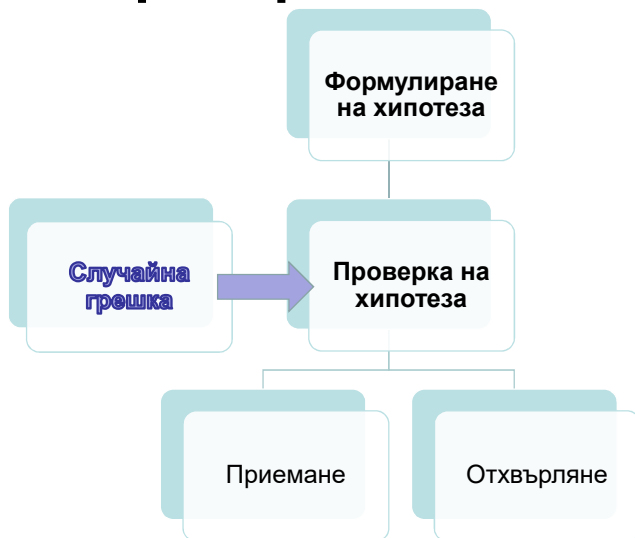


## Проверка на хипотеза. Параметрични тестове

### Проверка на хипотеза



## Проверка на хипотеза



## Хипотеза

- Хипотезата е **предположение за същността на даден факт**, респективно предположение за същността на определена закономерна поредица от факти.
- Хипотезата като предположение е **необходимо да бъде доказана**.
  - Доказателството ѝ предполага тя да изхожда от положения, които изглеждат правдоподобни, допустими. Освен това доказателството на хипотезата изисква съответно опитно, практическо потвърждаване на верността ѝ.
  - При всяко медицинско проучване изследователят поставя пред себе си хипотеза, която след завършване на наблюдението трябва да докаже или отхвърли с помощта на статистически методи.

## Значение на статистиката

- Захарен диабет
  - Експериментална група: Кръвна захар:  $\bar{X}_1 = 103$  mg/dl
  - Контролна група: Кръвна захар:  $\bar{X}_2 = 107$  mg/dl
- Рак на панкреас
  - Експериментална група 1-годишна преживяемост:  $p_1 = 23\%$
  - Контролна група: 1-годишна преживяемост:  $p_2 = 20\%$

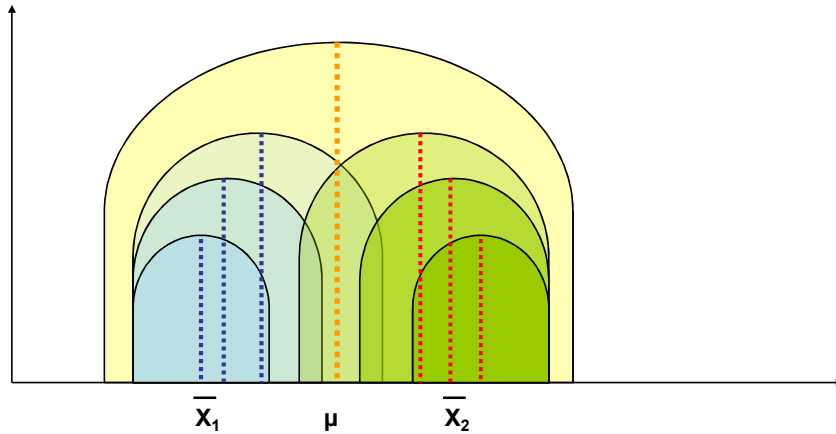
## Има ли разлика?

Статистическите методи са необходими  
за оценяване на различията, когато те са твърде малки,  
за да бъдат разпознати само въз основа на клиничен опит.

## Значение на статистиката

- Захарен диабет
  - Експериментална група: Кръвна захар:  $\bar{X}_1 = 103$  mg/dl
  - Контролна група: Кръвна захар:  $\bar{X}_2 = 107$  mg/dl
- Захарен диабет (по-голяма извадка)
  - Експериментална група: Кръвна захар:  $\bar{X}_1 = 104$  mg/dl
  - Контролна група: Кръвна захар:  $\bar{X}_2 = 105$  mg/dl

## Значение на статистиката

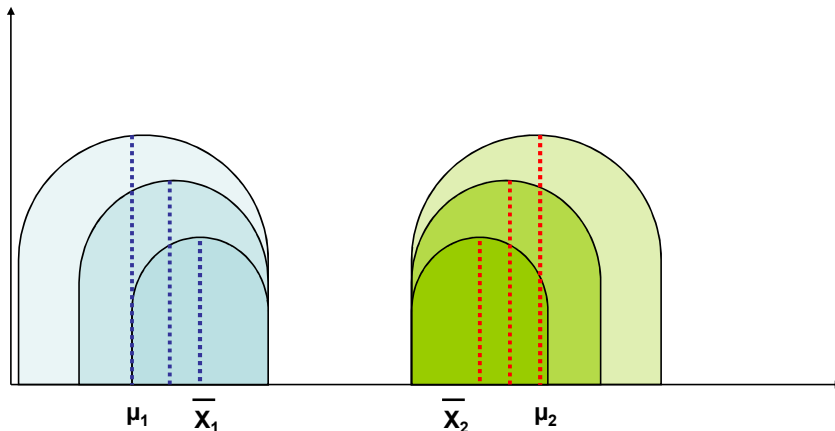


- Сравнение на две средни / относителни дялове:
  - Ако извадките са образувани от **една и съща популация**, то между техните средни **няма статистически значима разлика**.

## Значение на статистиката

- Захарен диабет
  - Експериментална група: Кръвна захар:  $\bar{X}_1 = 103$  mg/dl
  - Контролна група: Кръвна захар:  $\bar{X}_2 = 107$  mg/dl
- Захарен диабет (по-голяма извадка)
  - Експериментална група: Кръвна захар:  $\bar{X}_1 = 99$  mg/dl
  - Контролна група: Кръвна захар:  $\bar{X}_2 = 112$  mg/dl

## Значение на статистиката



- Сравнение на две средни / относителни дялове:
  - Ако извадките са образувани от **две различни популации**, то между техните средни **има статистически значима разлика** (защото произхождат от различни популации).

## Проверка на хипотеза

- Формулиране на предположение (работна хипотеза)
- Събиране на данни (доказателства)
- Потвърждаване или отхвърляне на предположението

## Проверка на хипотеза

### Съдебен процес

- Формулиране на предположение (работна хипотеза)
  - **Начална пледоария (презумция за невинност)**
- Събиране на данни
  - **Събиране на доказателства и показания**
- Потвърждаване или отхвърляне на предположението
  - **Оправдателна или осъдителна присъда**

## Проверка на хипотеза

- Два подхода за проверка на хипотеза:
  - Класически – **чрез определяне на критична област** (област на отхвърляне)
  - **Чрез използване на р-стойност** (предпочита се в престижните международни списания)
- **Двата подхода дават един и същ краен резултат!**

## Стъпки при проверка на хипотеза

- Класически подход (чрез определяне на критична област):
  - Дефиниране на нулева, алтернативна и работна хипотези
  - Определяне на ниво на значимост
  - Избор на статистически тест (критерий) и определяне на теоретична (критична, таблична) стойност
  - Изчисляване на емпирична стойност
  - Сравнение на теоретичната и емпирична стойности
  - Извод и интерпретация

## Нулева и алтернативна хипотези

- **Нулева хипотеза ( $H_0$ )** – различието между сравняваните показатели или наблюдавани явления е случайно, несъществено.
  - Нулевата хипотеза твърди, че няма статистически достоверна разлика в сравняваните статистически показатели. Въпреки, че в извадките може да се наблюдава известна разлика, тя е случайна и не може да бъде обобщена за генералните съвкупности.
- **Алтернативна хипотеза ( $H_1$ )** – различието между сравняваните показатели или явления е съществено, значимо.
  - Алтернативна хипотеза твърди, че констатираната разлика в емпиричните данни е статистически достоверна и може да бъде обобщена за генералните съвкупности.

## Работна хипотеза

- **Работна хипотеза** – хипотезата, възприета от изследователя и проверявана по време на научното проучване, се нарича работна хипотеза.
- **В практиката нулевата хипотеза се приема за работна!**

## Избор на тест (критерий)

- Изборът на тест зависи от:
  - Дизайн на проучването:
    - обем на извадката
    - степени на свобода
    - брой извадки за сравнение
    - свързани или независими извадки
    - едностранен или двустранен тест
  - Променливи за сравнение
    - тип
    - скала
    - разпределение



## Избор на тест (критерий)

Дизайн	Непараметрични		Параметрични
	Номинална	Ординална	Ординална, интервална, пропорционална
1 извадка	$\chi^2$ тест за съответствие	Wilcoxon тест	t-test за 1 извадка
2 независими извадки	$\chi^2$ тест	Mann-Whitney U-тест	t-test за 2 извадки
2 свързани извадки	McNemar тест	Wilcoxon тест	t-test за 2 сдвоени извадки
K независими извадки	$\chi^2$ тест	Kruskal-Wallis H-тест	ANOVA
K свързани извадки		Friedman тест за сдвоени извадки	ANOVA с повторни измервания

## t-тест за 1 извадка



Има ли разлика между **извадковата статистика** и наличната референтна стойност за **популяционен параметър**?

## **t-тест за 2 независими извадки**



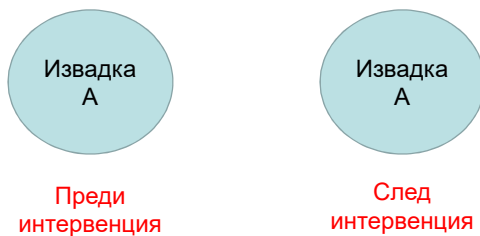
Има ли разлика между двете **извадкови статистики**?

## **Дисперсионен анализ (ANOVA)**



Има ли разлика между трите **извадкови статистики**?

## t-тест за 2 свързани извадки



Има ли разлика (**промяна**) в извадкова статистика преди и след интервенцията?

## Изчисляване на емпирична стойност

$$t = \frac{|\overline{X}_1 - \overline{X}_2|}{\sqrt{S_{x_1}^2 - S_{x_2}^2}}$$

t-test на Стюдънт  
(познат още като u-критерий)

## р-стойност

- **р-стойността представлява вероятността да се наблюдава констатираната или по-екстремна разлика в резултат на случайност** (т.е. ако нулевата хипотеза е вярна).
- р е между 0 и 1.
  - **Колкото по-голяма е р-стойността**, толкова по-вероятно е да наблюдаваме констатираната разлика в резултат на случайност. Т.е. различието се дължи най-вероятно на естествената вариабилност в рамките на изследваната популация, а не на обективно действащ фактор.
  - **Колкото по-малка е р-стойността**, толкова по-малко вероятно е да наблюдаваме констатираната разлика само и единствено в резултат на случайност. Т.е. различието се дължи най-вероятно на определен обективно действащ фактор (например по-ефективна терапия).
- **В практиката не ни трябва точната стойност на р, а само да знаем дали е по-голяма или по-малка от избраното ниво на значимост (най-често 0.05).**

## Ниво на значимост

- Нивото на значимост се означава с  $\alpha$  и се избира от изследвателя (т.е. зависи от дизайна на проучването).
- Представлява прагова вероятност, под която даден резултат (разлика, зависимост...) се приема за статистически значим.
  - Ако вероятността за появата на получената по извадката статистика е по-малка от приетото ниво на значимост, то нулевата хипотеза се отхвърля като несъответстваща на наблюдаваната ситуация.
  - Ако обаче тази вероятност е по-голяма от праговата стойност, нулевата хипотеза не се отхвърля, т.е. тя се приема за правдоподобна.
- Двете най-често използвани нива на значимост са 0.05 и 0.01.
- При класическия подход с критична област нивото на значимост служи за определяне на теоретичната стойност.

## Правило за р-стойността

- Ако  $p \geq 0.05$ ,  $H_0$  се приема като правдоподобна. В заключение твърдим, че **няма статистически значима разлика** между сравняваните извадки.
- Ако  $p < 0.05$ ,  $H_0$  се отхвърля като неправдоподобна и приемаме  $H_1$ . В заключение твърдим, че **има статистически значима разлика** между сравняваните извадки.
- **Колкото по-голяма емпирична стойност на критерия, толкова по-малка р-стойност.**

## Вземане на решение

Таблица 2. t-разпределение на Стюдънт

df	0.10	0.05	0.01
1	6.314	12.71	63.66
2	2.920	4.303	9.925
3	2.353	3.182	5.841
4	2.132	2.776	4.604
...	...	...	...
27	1.703	2.052	2.771
28	1.701	2.048	2.763
29	1.699	2.045	2.756
≥30	1.645	1.960	2.576

## Сравнение на теоретичната и емпирична стойности

- Ако изчислената емпирична стойност на критерия е **по-малка или равна** на теоретичната, то  $p \geq 0.05$ . Следователно,  $H_0$  **се приема** като правдоподобна. В заключение твърдим, че **няма статистически значима разлика** между сравняваните извадки.
- Ако изчислената емпирична стойност на критерия е **по-голяма** от теоретичната, то  $p < 0.05$ . Следователно,  $H_0$  **се отхвърля** като неправдоподобна и **приемаме  $H_1$** . В заключение твърдим, че **има статистически значима разлика** между сравняваните извадки.
- **Колкото по-голяма емпирична стойност на критерия, толкова по-малка p-стойност.**

## Чести грешки

- В практиката не трябва да проверяваме само дали даден резултат е статистически значим. Трябва да се проверяват също размера на оценявания резултат и точността му (ширината на интервала на доверителност).
- Статистически значим резултат не означава клинично значим резултат.
- Липсата на статистически значим резултат е също резултат сам по себе си.
- Ако повторим даден експеримент, p-стойността най-вероятно ще бъде различна.

## Чести грешки

- В статистиката не доказваме хипотези, а твърдим дали имаме основание да отхвърлим или не дадено твърдение.
- Нивото на значимост  $\alpha$  и  $p$ -стойността са различни понятия. Нивото на значимост се определя от изследователя, а  $p$ -стойността се изчислява въз основа на събраните данни.
- $P$ -стойността не показва вероятността за верност на работната хипотеза, а колко вероятно е наблюдаваната разлика да бъде резултат на случайност.

## Грешки при проверка на хипотеза

	Извод: Отхвърляме нулевата хипотеза	Извод: Приемаме нулевата хипотеза
Нулевата хипотеза е вярна	Грешка от I род ( <b>фалшиво-положителен резултат</b> )	OK
Нулевата хипотеза е невярна	OK	Грешка от II род ( <b>фалшиво-отрицателен резултат</b> )

## Грешки при проверка на хипотеза

- **Отхвърлянето на вярна нулева хипотеза се нарича грешка от I род.**
- Вероятността да бъде направена грешка от I род съвпада с нивото на значимост  $\alpha$ .
  - Ниско ниво на значимост  $\alpha$  намалява вероятността за грешка от I род, но увеличава вероятността за грешка от II род.
- **Приемането на невярна нулева хипотеза се нарича грешка от II род.**
- Вероятността да бъде направена грешка от II род се означава с  $\beta$ .

## Грешки при проверка на хипотеза

- **Мощност** – вероятността да се отхвърли нулевата хипотеза, когато тя е невярна.
  - С други думи, вероятност да потвърдим статистически значим резултат, когато такъв резултат в действителност е на лице.
- Мощността е равна на  $1 - \beta$ .
- Мощността зависи от обема на извадката, разсейването вътре в нея, нивото на значимост и размера на оценявания ефект.
  - Нуждаем се от по-голяма извадка, когато търсим по-малък ефект и когато разсейването е по-голямо.



## **Грешки при проверка на хипотеза**

- **Двата рода грешки имат сложна взаимовръзка и не е възможно техният размер да бъде минимизиран едновременно.**
  - Допускането на коя от двете грешки може да има по-тежки последствия зависи от контекста на решавания проблем и решение се взема на експертно равнище.
- В практиката обикновено се фиксира нивото на значимост и се избира такъв тест, който при определени условия ще доведе до най-малка вероятност за грешка от втори род (т.е. тест с възможно най-голяма мощност).