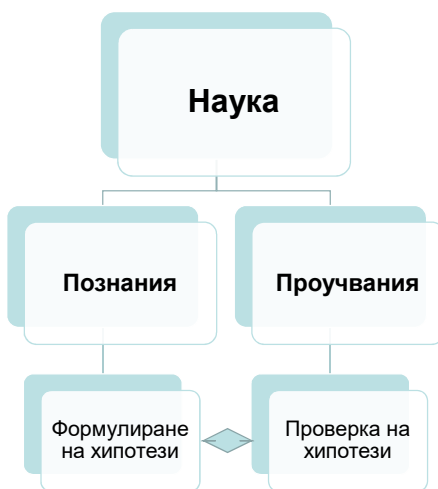
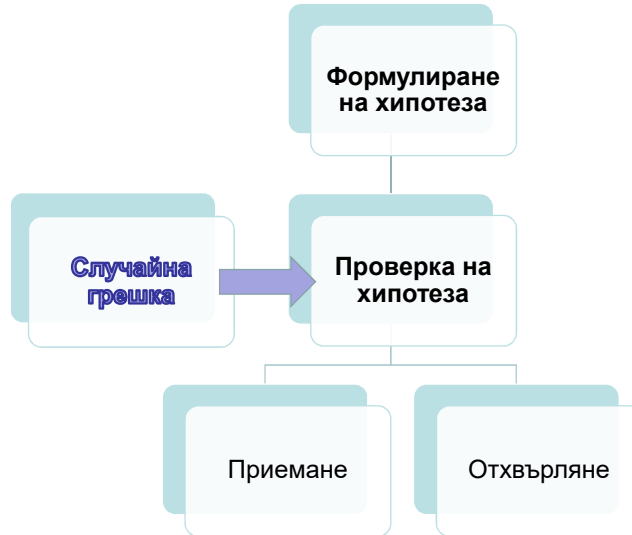


Основни статистически анализи

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.**

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

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

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)

Правило за р-стойността

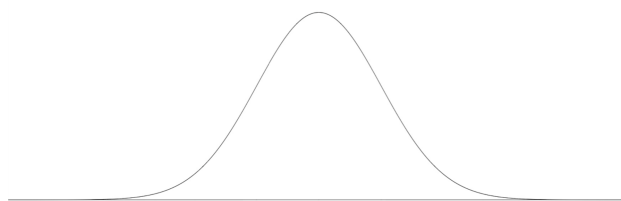
- Ако $p \geq 0.05$, H_0 се приема като правдоподобна. В заключение твърдим, че **няма статистически значима разлика** между сравняваните извадки.
- Ако $p < 0.05$, H_0 се отхвърля като неправдоподобна и приемаме H_1 . В заключение твърдим, че **има статистически значима разлика** между сравняваните 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

Параметрични и непараметрични тестове



Normal distribution



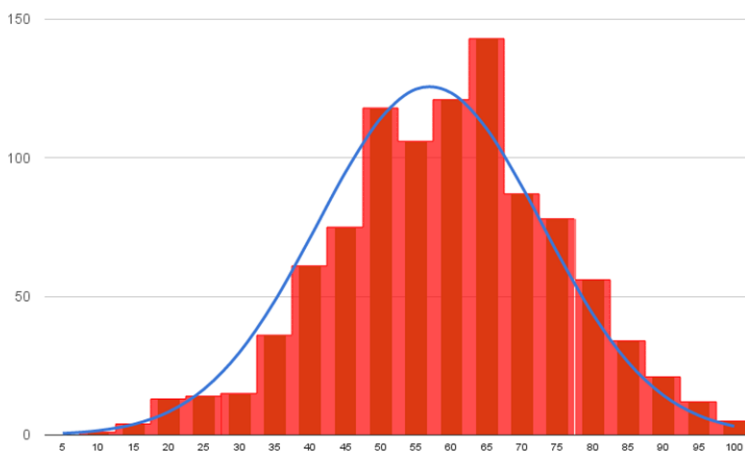
Paranormal distribution

Проверка за нормалност

- **Графичен метод:** графично сравняване на плътността на нормалното разпределение с хистограмата на относителните некумулирани честоти, получена на базата на направената извадка.
- Предимства – нагледност, недостатъци – ниска надеждност.

Проверка за нормалност

- **Графичен метод**



Проверка за нормалност

- **Чрез тест (критерий) за нормалност:**
 - K^2 тест на D'Agostino
 - Тест на Jarque–Bera
 - Тест на Anderson–Darling
 - Тест на Cramér–von Mises
 - Тест на Lilliefors
 - **Тест на Kolmogorov–Smirnov**
 - Тест на Shapiro–Wilk
- Най-прецизна оценка от научна гледна точка

Тест на Колмогоров-Смирнов

- Тестът на Колмогоров-Смирнов е един от най-мощните непараметрични критерии, особено при голям брой единици на наблюдение.
- Методът се базира на **сравняването на кумулативни редове** и има три разновидности:
 - Съпоставка и търсене на съответствие между извадка и генерална съвкупност
 - Между две извадки с еднакъв брой единици на наблюдение
 - Между две извадки с различен брой единици на наблюдение

Непараметрични тестове

- От статистическите хипотези, проверката на които изисква използването на непараметрични тестове, изследователят най-често се среща с **хипотеза, отнасяща се към самото статистическо разпределение**.
 - Нулевата хипотеза в този случай гласи, че **различиято между фактическите и теоретическите очаквани разпределения е обусловено само от причини от категорията на случайността**.
- Непараметричните критерии имат няколко **основни преимущества** в сравнение с параметричните:
 - независимост от формата на разпределение;
 - неограниченост от начина на отчитане;
 - по-малка трудоемкост.

Тест на Пирсън (χ^2 -тест)

- Критерият на съгласие на Пирсън е **основен метод за оценка на хипотези, отнасящи се към самото статистическо разпределение**.
 - Такива хипотези, ако те се интерпретират в духа на нулевата хипотеза, се заключават в предположението, че **разликата между фактическите и очакваните разпределения е обусловена само от случайности**.
- Критерият на Пирсън се използва за статистическа оценка на резултатите от изследванията в случаи, в които не е необходимо да се знае абсолютната величина на самия признак и размерът на връзката, а се изисква **да се потвърди съществено ли е влиянието на изучавания фактор или е случайно**.

Тест на Пирсън (χ^2 -тест)

Пол	Ваксина против сезонен грип					Общо
	Не	По-скоро не	Не знам	По-скоро да	Да	
Мъже	130	180	100	50	40	500
Жени	90	70	150	120	70	500
Общо	220	250	250	170	110	1000

Наблюдавани честоти (ФАКТИ)

Полът на анкетираните **оказва ли влияние** върху нагласите за поставяне на ваксина против сезонен грип?

Дали полът **е фактор** за нагласите за поставяне на ваксина против сезонен грип?

Тест на Пирсън (χ^2 -тест)

$$\chi^2 = \sum \frac{(f_0 - f_T)^2}{f_T}$$

Таблица 3. χ^2 -разпределение

df	0.10	0.05	0.01
1	2.706	3.841	6.635
2	4.605	5.991	9.210
3	6.251	7.815	11.345
4	7.779	9.488	13.277
5	9.236	11.070	15.086
6	10.645	12.592	16.812
7	12.017	14.067	18.475
8	13.362	15.507	20.090
9	14.684	16.919	21.666
10	15.987	18.307	23.209

$$df = (R - 1) \times (C - 1)$$

Тест на Пирсън (χ^2 -тест)

Пол	Ваксина против сезонен грип					Общо
	Не	По-скоро не	Не знам	По-скоро да	Да	
Мъже	130	180	100	50	40	500
Жени	90	70	150	120	70	500
Общо	220	250	250	170	110	1000

Наблюдавани честоти (ФАКТИ)

Пол	Ваксина против сезонен грип					Общо
	Не	По-скоро не	Не знам	По-скоро да	Да	
Мъже	110	125	125	85	55	500
Жени	110	125	125	85	55	500
Общо	220	250	250	170	110	1000

Очаквани честоти (ОЧАКВАНИЯ)

$$\chi^2 = \sum \frac{(f_0 - f_T)^2}{f_T} = 11$$

Тест на Пирсън (χ^2 -тест)

- **Условия за прилагане:**
 - Голяма по обем представителна извадка (над 50 единици на наблюдение)
 - Всички очаквани (теоретични) честоти са поне 1 и поне 80% от тях са по-големи от 5
- **Заклученията зависят понякога от начина на групировка** на случаите според разновидностите на наблюдаваните признаци.

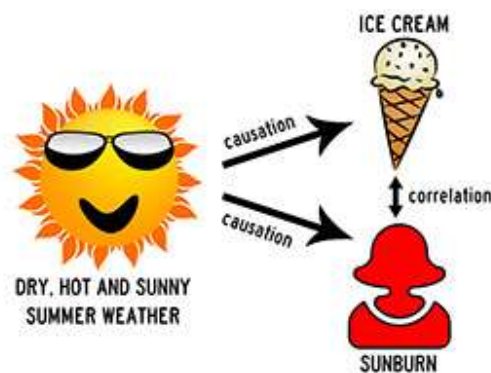
Тест на Пирсън (χ^2 -тест)

- Метод при изучаване на качествени алтернативни признаци

$$\chi^2 = \frac{n \times (a \times d - b \times c)^2}{(a + b) \times (c + d) \times (a + c) \times (b + d)}$$

Вид хранене	Здравно състояние		Всичко
	Заболели (+)	Здрави (-)	
Изкуствено (+)	43 a	50 b	93 a + b
Естествено (-)	5 c	87 d	92 c + d
Всичко	48 a + c	137 b + d	185 n

Зависимост и причинно-следствена връзка



Доказването на статистическа зависимост (статистическа връзка) не означава доказване на причинно-следствена връзка!

Причинно-следствена връзка в медицината се доказва посредством епидемиологични методи (критерии на Брадфорд-Хил).

