

## СТАТИСТИЧЕСКА ПРОВЕРКА НА ХИПОТЕЗИ

### 1. Значение на сравняването на данни от извадки

В предходната глава е представено използването на извадковите статистики за оценка на параметрите на популацията. При някои неекспериментални стратегии, когато целта на изследванията е да се опишат характеристиките на специфични популации, такова оценяване на параметрите на популацията е достатъчно за анализа на данните.

Други изследователски стратегии включват събиране на данни за целите на проверка на хипотези. В медицинската практика често се провеждат изследвания върху ефективността на различни методи на лечение при болни с еднакви заболявания. Проучва се влиянието на определени фактори за възникването и развитието на съответни заболявания като се използват опитни и контролни групи или се сравняват резултатите в едни и същи групи преди и след провеждане на съответна интервенция. В експерименталните лаборатории се провеждат непрекъснато изследвания, при които се работи с една или повече опитни и контролни групи и т. н. Възниква въпросът доколко наблюдаваните различия в извадки могат да се приемат за достоверни, дължащи се на закономерни фактори или те са резултат на случайност.

Отговорът на тези въпроси се дава чрез сравняване на резултатите в наблюдаваните извадки с помощта на различни тестове за значимост, които позволяват на изследователя да установи дали данните подкрепят или опровергават формулираната от него изследователска хипотеза. Използваната статистическа процедура се нарича **проверка (тестване) на хипотези**. Това е значително по-



високо ниво на изследователска работа, отколкото описателната характеристика да данните.

За да формулира хипотези за различията между групите и за взаимовръзките между променливите величини, изследователят трябва да познава добре литературата по проблема, да извърши критичен анализ на публикувани данни, да избере подходящ дизайн на проучването и методи за събиране на достоверна информация.

Проверката на хипотези се опира на по-мощни статистически тестове, които позволяват да се определи дали различията между оценъчните индикатори от извадката и параметрите на популацията или различията между статистиките на различни извадки се дължат на случайност или са значими, закономерно обусловени от конкретни фактори и причини. В това се заключава същината и основната цел на сравняването на данни от извадки.

## 2. Основни понятия при сравняване на данни и проверка на хипотези

### 2.1. Същност и видове статистически хипотези

*Хипотезите* представляват *предположения за същността на факти или явления от заобикалящия ни свят, твърдения за характера на връзката между две или повече групи наблюдения.*

Един от основоположниците на статистическата теория за проверка на хипотези – Ю. Нейман – дава следното определение: „*статистическа хипотеза се нарича всяко предположение относно някаква функция на честотите на наблюдаеми случайни променливи*“. От тази гледна точка, *статистическите хипотези са предположения за параметрите или формата на разпределение на случайните променливи и за връзката между две или повече променливи.*

Формулирането и проверката на хипотези представлява съществена част от статистическите изводи и заключения. Хипотезите обикновено се позовават на някаква теория, за която или се



вярва, че е истина, или тя трябва да бъде използвана като аргумент, който предстои да се докаже. Подходящ пример за хипотеза е твърдението, че дадено ново лекарство е по-добро от използваното досега за лечение на същите симптоми.

При всеки разглеждан проблем, изучаваният въпрос се опростява в 2 конкуриращи се твърдения/хипотези, между които ще трябва да се направи избор: **нулевата хипотеза, означавана с  $H_0$ , срещу алтернативната хипотеза, означавана с  $H_1$** . Тези две хипотези не се третираат еднакво – много по-голямо внимание се отдава на  $H_0$ . Срещат се две чести ситуации:

1. Проведен е експеримент, който се опитва да опровергае или отхвърли определена хипотеза, най-често нулевата. Следователно, ние ѝ даваме приоритет, така че тя не може да бъде отхвърлена, ако няма достатъчно силни доказателства срещу нея. Например,  $H_0$ : няма различие във вкуса на обикновената и диетичната кока-кола срещу  $H_1$ : има различие.

2. Ако една от двете хипотези е „по-проста“, ние даваме приоритет на нея, така че „по-сложната“ не се приема, докато не се наберат достатъчно доказателства срещу по-простата. Например, „по-просто“ е да се твърди, че няма разлика във вкуса на обикновената и диетичната кока-кола, отколкото да се каже, че има разлика.

Хипотезите често са твърдения за параметрите на популацията като очаквана стойност – напр.,  $H_0$  може да гласи, че очакваният ръст на 10-годишните момчета в Плевенска област не е различен от този на 10-годишните момчета в страната. Друга хипотеза може да се отнася до формата на разпределение на дадена изучавана характеристика – например: ръстът на 10-годишните момчета в Плевенска област има нормално разпределение.

## Нулева хипотеза

Нулевата хипотеза ( $H_0$ ), предполага, че няма различие или ако такова видимо се наблюдава, то се дължи единствено на случайност. Например, в даден клиничен опит върху ново лекарство,  $H_0$  би могла да гласи: новото лекарство не е по-добро от използваното



понастоящем. Такава  $H_0$  може да се запише по-кратко: *няма разлика между двете лекарства*.

На  $H_0$  се отдава специално внимание. Това се дължи на факта, че тя се отнася до твърдението, което се проверява, докато  $H_1$  се отнася до твърдението, което трябва да бъде прието, ако/когато  $H_0$  се отхвърли.

Окончателното заключение след прилагане на теста за проверка на хипотези винаги се дава от гледна точка на  $H_0$  – „ $H_0$  се отхвърля в полза на  $H_1$ “ или „ $H_0$  не се отхвърля“. Никога не се прави заключение от типа „ $H_1$  се отхвърля“ или дори „ $H_1$  се приема“. Ако се направи заключение „ $H_0$  не се отхвърля“, то това не означава обезателно, че  $H_0$  е вярна, то само предполага, че няма достатъчно доказателства  $H_0$  да бъде отхвърлена в полза на  $H_1$ .

### Алтернативна хипотеза

Алтернативната хипотеза,  $H_1$ , представлява твърдение за това, което статистическият тест за проверка трябва да установи. Например, в даден клиничен опит с ново лекарство  $H_1$  би могла да бъде, че новото лекарство има различен ефект в сравнение с досега използваното. В този случай алтернативната хипотеза  $H_1$  е *двустранна (ненасочена)*, т.е. тя не определя дали новото лекарство е по-добро или по-лошо, а само постулира, че то има различен ефект от досегашното.

Алтернативната хипотеза  $H_1$  би могла също да гласи, че новото лекарство е по-добро от досегашното. В такъв случай тя е *едностранна (насочена)*, тъй като се конкретизира, че ефектът от новото лекарство е по-добър от досега използваното лекарство за същите симптоми.

Действително, традицията в статистиката е наложила да се формулира и проверява нулевата хипотеза  $H_0$ .

Има обаче и друго виждане по този проблем. Когато се формулира хипотеза, това означава, че изследователят вярва, че има различие или зависимост. Тогава по-разбираемо е да се формулира предположение за това какви различия или взаимовръзки се очак-



ват, отколкото да се формулира условие за нулева хипотеза, т.е. че няма различие.

Например, да предположим, че няма „значимо различие“ между повишаването на теглото при деца, кърмени от майките си и при изкуствено хранени деца. Ако действително нямаме достатъчно представа по този въпрос, ние не бихме формулирали хипотеза, а бихме задали въпрос: „Има ли разлика в повишаването на теглото между кърмените и изкуствено хранените деца?“. Ако имаме обосновка и научна информация по този въпрос, бихме формулирали следната хипотеза: „Кърмените деца наддават повече през първия месец, отколкото изкуствено хранените“. Така формулирана хипотезата звучи като  $H_1$ , а не като  $H_0$ .

## Прости и сложни хипотези

**Простата хипотеза** характеризира пълно разпределението на популацията и е еднозначно дефинирана. Например,  $H_0$ : разпределението на случайната величина  $X$  е нормално.

**Сложната хипотеза** не определя пълно разпределението на популацията и се дефинира нееднозначно. Например,  $H_1$ : разпределението на случайната величина  $X$  е или нормално, или асиметрично.

**Хипотези за една извадка и за повече от една извадка.** Хипотезите могат да се отнасят до стойността на един отделен параметър в популацията и те се проверяват чрез сравнение на хипотетичната стойност на параметъра със стойността на извадковата статистика.

Много по-честа е практиката да се сравняват две или повече извадки и тогава хипотезите се отнасят до повече от една извадка.

## 2.2. Грешки от I и II род

Грешките, които могат да бъдат допуснати при проверката на хипотези, се определят от гледна точка на  $H_0$ . След анализирането на данните, изследователят възприема  $H_0$  (ако няма значими раз-



личия) или отхвърля  $H_0$  (ако наистина има значими различия). Отхвърлянето на  $H_0$  означава, че са били открити значими различия, но тъй като нито едно проучване не е перфектно, **винаги има шанс за грешка**.

Могат да бъдат направени две потенциални грешки – **грешка от I род и грешка от II род**. За да разберем същността на тези грешки, нека представим възможностите за решения относно  $H_0$  чрез четирикратна таблица 2x2:

Табл. 11.1. Грешки от I-ви и II-ри род

Действителна ситуация	Взето решение	
	$H_0$ се отхвърля	$H_0$ не се отхвърля
$H_0$ е вярна	Грешка от I род	Правилно решение
$H_1$ е вярна	Правилно решение	Грешка от II род

Ако  $H_0$  е вярна и сме я възприели, значи сме взели правилно решение. Ако  $H_0$  е невярна и ние я отхвърлим, това също ще бъде правилен отговор.

**Грешка от I род** се получава, ако отхвърлим  $H_0$ , когато в действителност тя е вярна. Това би се случило, когато данните сочат статистически значим резултат, а фактически няма различие. Например, сравняваме две групи, обучавани по два различни метода. Резултатите в първата група са по-добри от втората и ние бихме отхвърлили  $H_0$ . Може да се окаже, обаче, че децата в първата група са просто по-надарени, а методът на обучение въобще не е имал значение.

**Вероятността за грешка от I род** се означава с  $\alpha$  (**алфа**) и може да бъде точно изчислена и намалена чрез промяна на нивото на значимост на хипотезата. Обикновено се работи с ниво на вероятност не по-малко от 95%, т.е.  $H_0$  се приема за вярна само ако нейната значимост е над 0.05. Ако приемем ниво на значимост 0.01 вместо 0.05 възможността за допускане на грешка от I род намалява 5 пъти. При това, обаче, ще стане по-трудно да се установи



значим резултат, т.е. ще намалее силата на статистическия тест и ще се увеличи рискът за грешка от II род.

Грешката от I род се счита за по-сериозна и по-важна за избягване, отколкото грешка от II род. Процедурата за проверка на хипотези, следователно, може да се приспособи така, че да гарантира „ниска“ вероятност за погрешно отхвърляне на  $H_0$ , но тази вероятност никога не е 0.

**Грешка от II род** се допуска, ако приемем  $H_0$ , т.е. според данните от проучването тя излиза вярна, а всъщност вярна е  $H_1$  – например приемаме, че даден метод не е ефективен, а всъщност вярно е тъкмо обратното – този метод е ефективен и съществен. В клиничен опит с ново лекарство грешка от II род би се допуснала, ако направим извод, че двете лекарства имат еднакъв ефект, а фактически те имат различен ефект, т.е.  $H_0$  не се отхвърля, а фактически тя е погрешна.

**Вероятността за грешка от II род** се означава с  $\beta$  (*бета*) и тя обикновено е неизвестна. Най-често се дължи на малки размери на извадките. За избягване на грешка от II род може да се намали нивото на значимост, или пък да се увеличи размера на извадката, или да се намалят източниците на странично вариране и да се увеличи т.нар. „**размер на ефекта**“ (влиянieto на независимата променлива).

**Вероятността за грешки от II род  $\beta$  е в обратна зависимост от мощността на теста, който се изразява като  $1-\beta$ .** Следователно, колкото по-малко е  $\beta$ , толкова по-мощен е критерият за проверка на хипотезите.

Двата вида грешки са тясно свързани: **намаляването на вероятността за грешка от I род увеличава риска за грешка от II род и обратно.** Въпросът се свежда до това, кой вид грешка е по-допустима за изследователя. Това зависи най-вече от проучването и от това какви биха били последиците за изследваните лица от допускане на грешка от I или от II род. При избора на критерий за проверка на хипотези най-добър се счита този тест, който намалява до минимум вероятността за грешки, както от I род, така и от II



род. Ако разполагаме с няколко критерия, аналогични по отношение на грешките от I род, ще трябва да изберем онзи критерий, при който вероятността за грешки от II род е най-малка. Именно такъв критерий се нарича „най-мощен критерий“. Като правило параметричните методи осигуряват по-малък риск за грешки от II род, отколкото непараметричните.

### 2.3. Статистическа значимост на хипотезите

При анализа на данни от извадки изследователят най-често се стреми да установи има ли статистически значима връзка между променливите или статистически значима разлика между групите.

В основата на понятието „статистическа значимост“ лежи идеята за вероятност. Ако вероятността, че даден резултат се проявява случайно е само 1 на 1000, то тя е много ниска и има висока вероятност този резултат да се дължи на съществен фактор или причина. Такъв резултат се нарича „статистически значим“.

**Нивото на значимост** представлява **вероятността, над която  $H_0$  се приема за вярна, и под която  $H_0$  се отхвърля**. Означава се с  $\alpha$ . Колкото по-малка е тази вероятност, толкова по-малка е възможността за грешка от I род. Обикновено се използва ниво на значимост 0.05. Това означава, че за да бъде един резултат статистически значим, не трябва да има повече от 5 на 100 вероятност този резултат да се дължи на случайност. В случаи на висока степен на отговорност на медицинските решения се приема още по-високо ниво на значимост – 0.01 или даже 0.001, което означава, че има съответно само 1 на 100 или 1 на 1000 вероятност за случаен резултат. Нивото на значимост се определя още в етапа на планиране на проучването в зависимост от целта и характера на изследването и от възможността за увеличаване на обема на извадката. Трябва да се отчитат и последиците от неправилното отхвърляне или приемане на  $H_0$ , например, ако приемането на невярна  $H_0$  има по-сериозни последици, отколкото отхвърлянето ѝ, на нивото на значимост  $\alpha$  може да присвои по-голяма стойност и обратно.





Практическото определяне на нивото на значимост на дадена хипотеза се опира на изчисляване на специални статистически критерии и сравняване на техните вероятностни стойности (означавани с  $p$ ) с таблични критични стойности, за да се провери дали вероятността е по-малка от 0.05, 0.01 и т.н. Резултатите се означават като  $p > 0.05$ ,  $p < 0.05$ ,  $p < 0.01$  и т.н. Компютърните програми дават още по-точни вероятности (напр.  $p < 0.057$  и т.н.).

$P$ -стойността на теста се сравнява с възприетото ниво на значимост и ако то е по-малко, резултатът е значим, т.е. ако нулевата хипотеза трябва да бъде отхвърлена на ниво  $\alpha = 0.05$ , това се записва като  $p < 0.05$ .  $P$ -стойността подчертава силата на доказателствата за отхвърляне на  $H_0$ . Колкото по-малко е  $p$ , толкова по-убедително е отхвърлянето на  $H_0$ .

### 3. Статистически тестове за проверка на хипотези

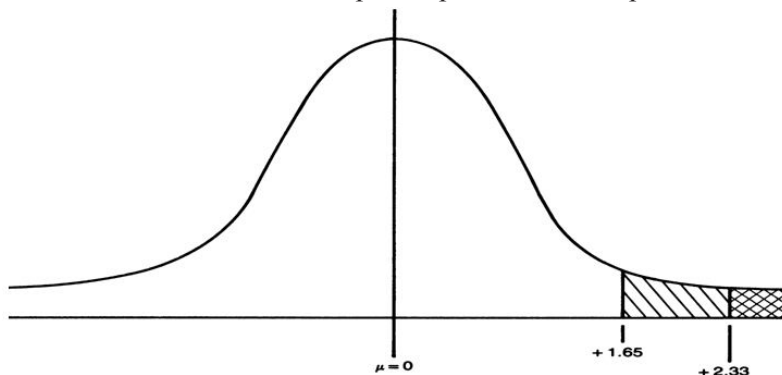
*Статистическият тест (критерий)* представлява величина, изчислена от данните в извадката, чиято стойност се използва, за да се реши дали  $H_0$  трябва да бъде отхвърлена или не.

#### 3.1. Едностранин и двустранин тест

„Страни“ или „опашки“ в разглеждания контекст се наричат краищата на кривата на разпределението на извадката или популацията.

*Едностранин тест е този, в който стойностите, при които можем да отхвърлим  $H_0$  са разположени изцяло в единия край на вероятностното разпределение (фиг. 11.1).* Прилага се при насочена алтернативна хипотеза, т.е. когато не само се установяват различия между групите, но се конкретизира и посоката на тези различия. С други думи, критичната област за едностранин тест представлява серия от стойности, които са по-малки или по-големи от критичната стойност на теста.

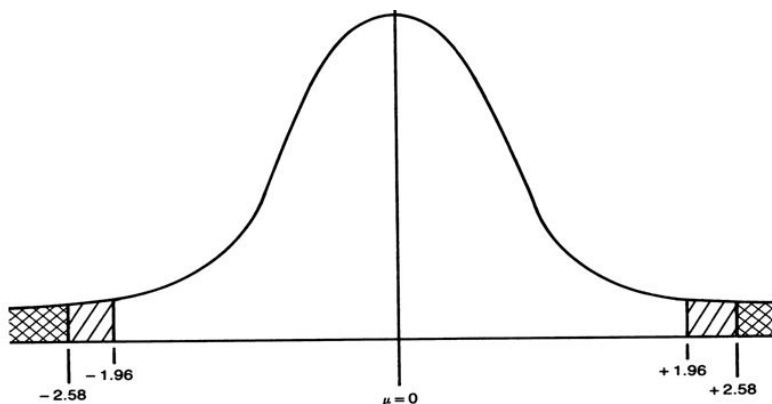
Двустранен тест е този, в който стойностите, при които можем да отхвърлим  $H_0$  са разположени в двата края на вероятностното разпределение (фиг. 11.2). Прилага се при ненасочена алтернативна хипотеза, когато не се конкретизира посоката на различията.



Фиг. 11.1. Едностраниен тест за значимост при нормална крива

Изборът между едностраниен или двустранен тест се определя от целта на изследването и се прави още при планиране на проучването. На практика, **едностранен тест се използва**, ако формулираната хипотеза указва посоката на очакваните различия, а **двустранен тест** – в останалите случаи.

Едностраниният тест е по-мошен, тъй като стойността на критерия не трябва да бъде толкова голяма, за да бъде значима на определено ниво. Например, значимост при двустранен тест 0.05 съответства на значимост при едностраниен тест 0.025. При двустранен тест и  $n > 120$  стойностите на  $z$  или  $t$  трябва да достигнат  $\pm 1.96$  за  $p=0.05$  и  $\pm 2.58$  при  $p=0.01$ , докато при едностраниен тест те са доста по-ниски – 1.65 и 2.33 за същите нива на значимост.



Фиг. 11.2. Двустранен тест за значимост при нормална крива

### 3.2. Параметрични и непараметрични тестове (критерии)

Главното различие между двата вида тестове е предположението, което трябва да се направи за данните в популацията преди прилагане на съответния критерий. При параметричните тестове (напр.  $t$ -критерий, дисперсионен анализ и др.) се предполага, че проучваната променлива е нормално разпределена в популацията и дисперсията е еднаква при различни нива на променливата. Непараметричните тестове не зависят от вида на разпределението и могат да се прилагат при всички форми на разпределения.

До неотдавна се приемаше, че параметричните тестове трябва да се използват главно при интервални и пропорционални скали, а непараметричните – при номинални и ординални скали. В последните години се подчертава, че използването на параметрични техники при ординални данни не изопачава резултатите. Доста резултати при прилагане на параметрични методи са почти идентични с тези при непараметричните тестове.

Параметричните тестове са по-мощни и по-гъвкави, докато непараметричните са по-бързи и лесни за изчисляване. Последното днес вече не е особено предимство, тъй като компютърната обра-

ботка на данните прави всички изчислителни процедури достатъчно бързи и леки.

### Критична стойност и критична област на теста

**Критичната стойност ( $u$ )** на даден тест представлява *прагът, спрямо който се сравнява изчислената стойност на статистическия критерий, за да се определи дали да се отхвърли или не  $H_0$* . Тя зависи от нивото на значимост и от това дали тестът е едностранен или двустранен.

**Критичната област** представлява *серия от стойности на статистическия тест, при които  $H_0$  се отхвърля*; т.е. извадковото пространство за статистическия тест се разделя на две части; едната област (критичната област) води към отхвърляне на нулевата хипотеза, а другата не. Следователно, ако изчислената стойност на статистическия тест попада в критичната област,  $H_0$  се отхвърля и обратно.

### Мощност на теста

**Мощността (силата) на теста** определя вероятността да се приеме алтернативната хипотеза ( $H_1$ ), когато тя в действителност е вярна. С други думи, *силата на даден тест измерва способността на теста за правилно решение и вероятността за недопускане на грешка от II род*. По-мощен е този тест, който има по-голяма вероятност да отхвърли  $H_0$  и по такъв начин, да докаже статистически значим резултат. Силата на теста се изчислява чрез изваждане на вероятността за грешка от II род от единица и се изразява като  $1 - \beta$ . Максималната сила на даден тест може да бъде 1. В идеалния случай ние се стремим към мощност на теста, близка до 1.

### Степени на свобода

При изчисляване на статистическите критерии за значимост и в таблици за критичните им стойности се използва понятието „степенни на свобода“ ( $k$  или  $df$  – degrees of freedom). Както се посочва в предходната глава 12, степените на свобода са свързани с броя



на резултатите в серия от данни и с идеята за свобода на вариране. Например, три резултата (1, 5, 6) имат три степени на свобода, тъй като всяко от тези числа е независимо от другото. След като обаче изчислим средната, една от степените на свобода се губи, тъй като тези три резултата не са вече свободни да варират. Средната от тези резултати е 4. След като знаем средната и два от трите резултата, можем да изчислим колко ще бъде третият резултат – той не може да варира, а трябва да приеме такава стойност, че средната да бъде 4.

При работа с една извадка степента на свобода  $df = n - 1$ . При проверка на хипотези по-често се работи с две извадки и  $df = (n_1 - 1) + (n_2 - 2) = n_1 + n_2 - 2$ .

#### 4. Основни принципи и процедури за сравняване и проверка на статистически хипотези

Проверката на хипотези представлява процес на вземане на решение в статистически смисъл относно това дали находките в дадено изследване отразяват случайност или „действителен“ ефект при съответно ниво на вероятност.

Математическите процедури за проверка на хипотези се основават на прилагането на теорията на вероятностите и подбора на извадки. Поради вероятностния характер на процеса, грешките при вземане на решения не могат да бъдат напълно елиминирани. Независимо от това, прилаганите основни принципи и процедури позволяват да се определи нивото на вероятност, при което можем да твърдим, че получените данни в конкретно изследване подкрепят експерименталната хипотеза. Тази процедура е фундаментална за анализа на данните, както и за вземане на клинични решения.

##### 4.1. Основни принципи при проверка на хипотези

1. Сравняването се свежда до *проверка (тестуване, оценка) на значимостта на съответна хипотеза*. За работна хипотеза най-често се приема нулевата. Следователно, *сравняването на хипо-*



тези представлява доказване на правдоподобността на нулевата хипотеза.

2. Доказването на верността на дадена хипотеза е свързано с определяне на нейната статистическа вероятност (значимост). За вярна се приема  $H_0$ , която е подкрепена с вероятност по-голяма от 0.05, т. е.  $p > 0.05$ . И обратно – при вероятност  $p < 0.05$   $H_0$  не е вярна и се възприема  $H_1$ . Сумата от вероятностите на  $H_0$  и  $H_1$  винаги е равна на единица или 100%. Следователно, ако  $H_0$  има вероятност по-висока от 0.05 (5%), то  $H_1$  има вероятност съответно по-малка от 0.95 (95%) и обратно.

3. Сравняването се опира на доказване на значимостта на различията между фактическите резултати (честоти) и очакваните честоти (ако е вярна  $H_0$ , т.е. изучаваните фактори не оказват влияние).

4. За оценка на величината на различията се прибегва до изчисляване на различни параметрични и непараметрични критерии, чиито конкретни стойности се сравняват с критични таблични значения и се оценява нивото на вероятност на  $H_0$  и  $H_1$ .

5. При всяка проверка на хипотези изследователят е поставен пред задачата да вземе определено решение за потвърждаване или отхвърляне на проверяваната хипотеза. Стремехът във всички случаи е да се вземе правилно решение. Теорията за проверка на хипотези обаче допуска известни рискове за вземане на неправилно решение, но статистиката е разработила средства за измерване и контролиране на тези рискове.

6. В зависимост от постановката на проучването различаваме:

- \* *сравняване на резултати в независими извадки* (напр. между опитни и контролни групи, между няколко опитни групи, третирани по различен начин през един и същ времемен период);
- \* *сравняване на резултати в зависими извадки* (в една и съща извадка преди и след провеждане на дадено лечение или друга интервенция); такива извадки се наричат извад-



ки от типа „преди-после“ (before-after).

7. Според вида на изучаваните променливи най-често се сравняват:

- \* средни величини;
- \* коефициенти и пропорции;
- \* корелационни коефициенти и др.

8. Според методите за проверка на хипотези различаваме:

- \* параметрични методи за сравняване на две групи наблюдения;
- \* параметрични методи при повече от две групи наблюдения (дисперсионен анализ);
- \* непараметрични методи, основаващи се предимно на  $\chi^2$

#### 4.2. Логика на процедурата за проверка на хипотези

Независимо от използвания метод, *процедурата за проверка на хипотези преминава през следните общоприети стъпки:*

1. *Формулиране на  $H_1$* , която предсказва какво възнамеряваме да оценим – дали независимата променлива оказва значимо влияние върху зависимата променлива, т.е. има действително различие между групите. В тази стъпка се уточнява също дали се касае за еднопосочност или двупосочност.

2. *Формулиране на работната хипотеза  $H_0$* , която гласи, че всяко различие в данните се дължи просто на случайност: независимата променлива не оказва ефект върху зависимата променлива, т.е. наблюдаваното различие е случайно.

3. *Формулиране на нивото на значимост,  $\alpha$  (алфа)*. Според установената практика използваме вероятност  $\alpha = 0.05$  или  $\alpha = 0.01$ . Това означава, че ако вероятността на  $H_0$  е по-малка от 0.05 или 0.01, можем да отхвърлим  $H_0$ .



4. Изчисляване на цифровата стойност на статистическия тест за определяне вероятността на  $H_0$

5. Сравняване на изчислената стойност на статистическия тест с табличните стойности за съответни стандартни разпределения при конкретно вероятностно ниво на значимост;

6. Вземане на решение за приемане или отхвърляне на нулевата хипотеза според  $p$ -стойността на теста.

## 5. Параметрични методи за проверка на хипотези

### 5.1. Сравняване на две групи наблюдения чрез $t$ -критерия на Стюdent

#### *Сравняване на средни величини*

Това е един от най-често използваните тестове за значимост. Опира се на разликата между средните аритметични величини в извадките. Възможни са няколко ситуации:

- независими малки по обем извадки ( $n < 30$ ) с еднакви дисперсии и нормално или близко до нормалното разпределение;
- независими големи по обем извадки с равни или различни дисперсии;
- зависими (корелирани) извадки.

И при трите ситуации логиката на работа е еднаква. Има само незначителни различия във формулите за  $t$ -критерия.

Нека разгледаме втората ситуация на основата на следния **пример**: В ортопедично отделение се въвежда нова рехабилитационна програма, с която се цели да се намали следоперативния престой на пациентите. От наблюдение на 100 пациенти, изписани преди внедряване на новата програма е установена средна продължителност на престоя след операция  $\bar{x}_1 = 30$  дни и стандартно отклонение  $s_1 = 8$  дни. За група от 64 пациенти рехабилитирани по новата



програма средния престой е  $\bar{x}_2 = 24$  дни и  $s_2 = 6$  дни. **Доказват ли тези резултати, че новата рехабилитационна програма оказва влияние върху следоперативния средния престой на болните в сравнение с използваната преди това?**

В последната фраза ясно е посочено какво се стреми да докаже изследователят.

**Практическата работа преминава през следните етапи:**

**I етап – формулиране на  $H_0$ :** новата рехабилитационна програма не оказва влияние върху следоперативния престой на пациентите; доколкото такова различие се наблюдава, то се дължи на случайност.

**II етап – изчисляване на емпиричната стойност на  $t$ -критерия**

Напр., при независими големи извадки с различни дисперсии:

$$(11.1) \quad t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{|30 - 24|}{\sqrt{\frac{8^2}{100} + \frac{6^2}{64}}} = \frac{6}{1,1} = 5,45$$

където:

$|\bar{x}_1 - \bar{x}_2|$  – абсолютната разлика между средните в двете извадки

$s_1$  и  $s_2$  – дисперсиите в двете извадки

$n_1$  и  $n_2$  – брой случаи в двете извадки

**III етап – определяне степените на свобода  $df = n_1 + n_2 - 2 = 164 - 2 = 162 = \infty$**

**IV етап – определяне нивото на значимост на  $H_0$**  чрез сравняване на емпиричната стойност на  $t$ -критерия при съответната степен на свобода  $df$  в таблицата за критични стойности на  $t$  (**Приложение 1**). При  $t = 5.45$  и  $df = \infty$  последния ред на таблицата намираме  $p_{(t)} < 0.001$

**V етап – формулиране на заключението**

Получената стойност  $p_{(t)} < 0.001$  е доста отдалечена от критичната стойност за правдоподобност на  $H_0$  (0.05 и даже 0.01). Следо-



вателно,  $H_0$  не е вярна. Тя се отхвърля и се възприема  $H_1$  – има значимо различие в средния следоперативен престой в сравняваните групи. С други думи, новата рехабилитационна програма оказва съществено влияние върху срока на възстановяване след ортопедична операция, т.е средният престой на болните от втората група е значимо по-малък от този в първата група .

### Сравняване на коефициенти и пропорции

Ако изчислените статистики в наблюдаваните извадки са пропорции или интензивни показатели, то тогава t-критерий се изчислява по следния начин:

$$(11.2) \quad t = \frac{|p_1 - p_2|}{\sqrt{\frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}}}$$

където:

$p_1$  е изчислената статистика (пропорция, коефициент за честота) в първата извадка

$q_1$  – стойност допълваща  $p_1$  до 1, 100, 1 000 и т. н. ( $1 - p_1$ ,  $100 - p_1$ ,  $1000 - p_1$ )

$n_1$  – броят на случаите в първата група

$p_2, q_2, n_2$  – съответните показатели за втората извадка

След определяне стойността на t-критерия и степените на свобода  $df$  по таблицата се намира  $p_{(t)}$  и по същия начин се формулират изводите.

Ако  $H_0$  е вярна, т. е. нейната вероятност е  $p > 0.05$ , тя се приема и се прави заключение, че не съществува закономерно различие. Доколкото такова може да се наблюдава, то е резултат от влиянието на случайни фактори.

Ако  $H_0$  не е вярна, т. е. нейната вероятност е  $p < 0.05$ , тя се отхвърля и се възприема алтернативната хипотеза и се прави извод, че съществува значимо различие между статистиките в сравняваните групи, което се дължи на влиянието на съществени, закономерни фактори и причини.



**Определяне на необходим брой случаи (обем на извадките) при сравняване на две извадки.**

При планиране на дадено проучване изследователят се нуждае от реалистична оценка за това колко лица са необходими за проверка на формулираната хипотеза. Изходният момент при сравняването и проверката на хипотези е колко лица са необходими, за да се установи значимо различие.

Разработени са множество отделни таблици за определяне на размера на извадката при различни методи: при работа с  $t$ -критерий, при дисперсионен и регресионен анализ и т.н. Когато в постановката на проучването се предвижда да се работи с две или повече групи, обемът на извадките може да се определи чрез преобразуване на формулите за сравняване на средни величини или пропорции.

Възможно е да се наблюдават равен брой случаи в сравняваните групи или при предварително зададен брой случаи в едната група може да се определи необходимият брой за втората група. И в двата случая се изхожда от **принципната постановка за очаквано достоверно различие**, т. е. очаква се да получим стойност на  $t$ -критерия по-голяма от 2, на която съответства вероятност на нулевата хипотеза  $p < 0.05$ .

Предполагаемите стойности на показателите в сравняваните извадки отново се вземат от литературни данни, предишни проучвания или се установяват при пилотни проучвания (микропроучвания).

## **5.2. Сравняване на повече от две групи чрез дисперсионен анализ**

Доста често в изследователската практика се провеждат проучвания върху повече от две групи с цел да се определи дали съществуват достоверни различия между групите (например, групи болни лекувани по няколко различни начина; групи болни лекувани с различни дози на едно и също лекарство средство и т.н.). Ако решим да анализираме различията между групите чрез  $t$ -критерия

на Стюdent, ще трябва да извършим доста сравнения по двойки групи, но всяко от тях ще се отнася само за 2 групи. Вероятността за грешка от I род ще нараства с броя на извършваните сравнения.

Вместо извършване на серия от отделни сравнения, би могло да се приложи метод на статистически анализ, който разглежда всички групи едновременно. Този метод се нарича **дисперсионен анализ (Analysis of Variance – ANOVA)** и се основава на сравняване на вътрегруповите и междугруповите дисперсии.

**Условия за прилагане на дисперсионния анализ:**

- \* **независимата променлива (факторът)** се предпочита да е измерена **върху номинална скала** с две или повече нива (напр., полът е номинална променлива с две нива; семейното положение, социалният статус, религията и др. са номинални променливи с повече от 2 нива);
- \* **зависимата променлива (резултатът)** да бъде **непрекъсната** (продължителна), измерена върху интервална или пропорционална скала;
- \* **зависимата променлива (резултатът)** да има **нормално разпределение**;
- \* **сравняваните групи** да бъдат **взаимно изключващи се** (независими една от друга);
- \* **сравняваните групите** да имат **еднакви дисперсии** (изискване за хомогенност на дисперсиите).

Дисперсионният анализ се оказва доста силен статистически метод. Дори когато не са спазени строго горните условия, могат да се получат резултатите доста близки до истината – напр., изискването за номинална скала на независимата променлива почти не се съблюдава. Възможно е тя да бъде представена върху рангова, интервална или даже върху пропорционална скала.

Дисперсионният анализ се прилага по няколко **схеми**.

**А. В зависимост от броя на независимите променливи:**

- \* **еднофакторен дисперсионен анализ** (one-way ANOVA) – изучаване влиянието на една независима променлива величина, която може да има две или повече нива;

- \* **двуфакторен дисперсионен анализ** (two-way ANOVA) – изучаване влиянието на две независими променливи величини;
- \* **многофакторен дисперсионен анализ** (MANOVA) – изучаване влиянието на повече от две независими променливи величини;

### **Б. В зависимост от броя на случаите в сравняваните групи:**

- \* **равномерен комплекс** – еднакъв брой случаи в групите;
- \* **неравномерен комплекс** – нееднакви по размер групи.

Оттук произтичат названията: еднофакторен дисперсионен анализ, равномерен комплекс; двуфакторен дисперсионен анализ, неравномерен комплекс и т.н. Важно условие при подбора на проучваните фактори при двуфакторен или многофакторен дисперсионен анализ е пълната им независимост.

### **Логическа основа на дисперсионния анализ**

Статистическият въпрос при дисперсионния анализ се опира на нулевата хипотеза, т.е на допускането, че всички групи са равни, извлечени от една и съща популация и всякакво наблюдавано различие в резултатите в сравняваните групи се дължи на случайност.

Принципната разлика между проверката на хипотези с помощта на  $t$ -критерия на Стюdent и дисперсионния анализ е в това, че при първия подход се сравняват директно средните аритметични величини в групите, а вторият подход се опира на сравняване на дисперсиите.

Следователно, същността на проверката на  $H_0$  при дисперсионния анализ е свързана с обсъждане на причините за варирането на резултатите. В тази връзка следва да разграничим:

- \* **Вътрегрупово (вътрешно) вариране** – дължи се на индивидуалната вариабилност на наблюдаваните лица, т.е. на естествените, присъщи на отделните индивиди черти и свойства, по които те се отличават едни от други; тако-

ва вариране е налице винаги, дори и при подбор на лица, еднакви по възраст, пол и други характеристики. То не е свързано с действие на закономерни фактори и причини и има случаен характер.

- \* **Междугрупово (външно) вариране** – дължи се на различия в условията на провежданите експерименти и в нивата на изучаваните фактори (лечебни подходи, експозиции и др.). Следователно, то отразява влиянието на проучваните фактори или условия, т.е. дължи се на закономерни фактори и причини.
- \* **Общото вариране** между сравняваните групи може да се представи като **сума от вътрегруповото и междугруповото вариране**. Това позволява директното сравняване на двата вида вариране.

Нека означим междугруповата дисперсия с  $s_e^2$  ( $e$  в индекс – от external), вътрегруповата дисперсия с  $s_i^2$  ( $i$  в индекс – от internal) и общата дисперсия с  $s_t^2$  ( $t$  в индекс – от total). Какви ситуации са възможни?

1. Ако общото вариране е еднакво със средната на дисперсиите на отделните групи (т.е. с вътрегруповото вариране), то междугруповото вариране ще бъде нула, защото общото вариране е сума от вътрегруповото и междугруповото вариране ( $s_e^2 + s_i^2 = s_t^2$ , а  $s_i^2 = s_t^2$ , то  $s_e^2 = 0$ ). От това произтича, че средните аритметични в отделните групи не се различават.

2. Ако дисперсията за цялата група е доста по-голяма от средното вариране в рамките на отделните подгрупи, то съществува значима разлика между средните аритметични поне в две от подгрупите. В този случай, вътрегруповото вариране няма да е равно на общото вариране. Разликата между тях ще отразява междугруповото вариране.

Отговорът на въпроса „Има ли значимо (сигнификантно) различие между резултатите в групите?“ се свежда до сравняване на междугруповото и вътрегруповото вариране.



За да определим дали различията между групите са достатъчно големи и позволяват да се отхвърли нулевата хипотеза, трябва да изберем подходящ критерий за значимост. При три и повече групи подходящият тест е **F**-критерий (по името на Р. А. Фишер – един от създателите на дисперсионния анализ), който се изчислява като отношение на междугруповата към вътрегруповата дисперсия:

$$(11.3) \quad F = \frac{s_e^2}{s_i^2}$$

– Ако не съществува реално различие между сравняваните групи, т.е. различие в действието на изучаваните условия, то междугруповата и вътрегруповата дисперсии ще бъдат почти еднакви и тяхното съотношение ще бъде равно или близко до единица.

– Ако има реални, истински различия между групите, свързани с влиянието на проучваните фактори, то отношението **F** ще бъде по-голямо от единица.

Стойностите на критерия **F** се оценяват по специална таблица, изградена на същата принципна основа, както тази за стойностите на критерия **t** (*Приложение 1*), която позволява да се направи заключение за вероятността на нулевата хипотеза и да се формулира крайният извод за достоверността на различията между групите.

### Дисперсионен анализ при еднофакторен равномерен комплекс

**Пример:** Проведено е проучване в дом за стари хора с цел подобряване на ориентацията и самопомощта сред пациентите със заболяване на Алцхаймер. Изследвани са 3 групи по 10 болни с еднакъв стадий на заболяването. **В първата група** е прилагана стимулация в стаите на пациентите чрез музика, цветове, часовници и календари. Във **втората група** са провеждани периодични занимания за ориентация на пациентите, включващи информиране за време, дни на седмицата, сезон, посетители и др. **Третата група** е получавала рутинни грижи и е използвана за контрола. Ориентацията на пациентите е оценявана чрез паметов тест и решаване на проблем. Получени са следните резултати (**табл. 11.2**):



**Табл. 11.2. Ориентация и самопомощ при три групи пациенти с болестта на Алцхаймер в зависимост от прилаганата стимулация**

Занятия за ориентирание	Повишена стимулация	Контролна група
$\bar{x} = 64.0$	$\bar{x} = 53.0$	$\bar{x} = 42.0$
$s = 17.17$	$s = 14.12$	$s = 17.66$
Междугрупова дисперсия – $s_e^2 = 1210$		
Вътрегрупова дисперсия – $s_i^2 = 268.9$		

**Статистическият въпрос тук е:** Различават ли се групите по резултатите в ориентацията? Приложените интервенции променили ли са ориентацията на пациентите?

**Процедурата на сравняване** на групите преминава през следните етапи:

**I етап** –  $H_0$ : няма съществено различие в резултатите; доколкото такова се наблюдава, то е резултат на случайност.

**II етап** – изчисляване на емпиричната стойност на F-критерия по формулата:

$$(11.4) \quad F = \frac{s_e^2}{s_i^2} = \frac{1210}{268,9} = 4,5$$

Изчисляването на вътрегруповата и междугруповата дисперсия е доста трудна и дълга процедура и се извършва предимно с помощта на компютърни програми.

### **III етап – определяне степените на свобода $df$ ( $\kappa$ )**

При дисперсионния анализ се определят два вида степени на свобода:

$df_1$  ( $\kappa_1$ ) – за по-голямата междугрупова дисперсия; в случая  $df_1 =$  броя на групите  $r - 1 = 2$

$df_2$  ( $\kappa_2$ ) – за вътрегруповата дисперсия  $df_2 = (n_1 - 1) + (n_2 - 1) + (n_3 - 1) = (n - r) = 30 - 3 = 27$

**IV етап – определяне нивото на значимост на  $H_0$  чрез сравняване на емпиричната стойност на F-критерия с теоретичната**





при съответните степени на свобода и възприетото ниво на значимост на  $H_0$  (**Приложение 3**).

Разполагаме  $df_1$  върху най-горния ред на таблицата и  $df_2$  върху колоната от лявата страна на таблицата. Пресичането на съответния ред и колона ни дава критичните стойности на  $F$  при ниво на значимост на  $H_0$  0.05. Критичната стойност на  $F$  е 3.35. Това е стойността, която се изисква за отхвърляне на  $H_0$  при двустранен тест.

*Следователно, на базата на сравнение на емпиричната стойност на  $F$ -критерия с двете критични стойности, можем да заключим, че  $H_0$  се отхвърля при ниво на вероятност по-малко от 0.05 и по-голямо от 0.01.*

***У етап – формулиране на заключението:*** Отхвърляйки  $H_0$ , можем да направим заключение, че групите се различават помежду си, т.е. приложените интервенции представляват значим фактор, който е довел до различия в резултатите за ориентацията на пациентите в отделните групи.

Анализът може да продължи и да се определи източника на значимите разлики. При групи с еднакъв размер се използва метод на Тюки, а при различни по обем извадки – метод на Шефе.

### **Дисперсионен анализ при еднофакторен неравномерен комплекс**

В практиката на научните проучвания е за предпочитане изследваните групи да се оформят с равен брой случаи, но по различни причини (икономически, етични и др.) по-често се работи с неравен брой.

**Пример 1:** Под влияние на три различни дози серотонин (в мг/кг) е измервано теглото на тумор. Да се определи с помощта на дисперсионен анализ дали дозата серотонин е значим (закономерен) фактор, който оказва влияние върху теглото на тумора, т.е. да се проведе еднофакторен дисперсионен анализ. Изходните данни са представени в **табл. 11.3**:

**Табл. 11.3. Тегло на тумора в зависимост от дозата на серотонина**

Дози в мг/кг $- x_i$	Тегло на тумора (дг) – $y_i$ (резултат, зависима променлива)	$\bar{x} \pm S$	Брой тумори
2.5	2.0; 2.5; 2.7; 1.7; 1.9; 2.4;	$2,2 \pm 0,39$	6
5.0	1.4; 1.6; 1.7; 2.0; 1.8;	$1,7 \pm 0,22$	5
25.0	0.9; 1.0; 1.3; 1.6; 1.5; 1.4; 1.4;	$1,3 \pm 0,26$	7

Независимо, че броят на изследваните случаи не е голям ( $6 + 5 + 7 = 18$ ), за прилагане на дисперсионен анализ се препоръчва ползване на съвременна изчислителна техника с готови статистически програмни продукти.

При работа с приложна програма STATGRAPHICS (може и с SPSS или друга програма) резултатите от дисперсионния анализ се представят по следния начин (**табл. 11.4**):

**Табл. 11.4. Влияние на дозите серотонин върху средното тегло на тумора**

Източници на девиация	Девиация	Степени на свобода	Дисперсия	$F_{em}$	$P$
Обща девиация	3.98	17	1.40		
Изследван фактор (доза)	2.62	2	1.31	14.5	0.01
Неизследвани и/или случайни фактори	1.36	15	0.09		

В първата колона на таблицата са посочени източниците на вариране в зависимата променлива (Source of variation): междугрупово вариране (Between group) – от изучавания фактор и вътрегрупово вариране (Within groups) – от случайни фактори и фактори, невключени в анализа.

Във втората колона са дадени девиациите (Сума от квадратите на разликите между груповите средни и общата средна – Sum of Squares).



В третата колона са степените на свобода (к или df).

В четвъртата колона е дисперсията (Mean square).

В петата колона е изчислената стойност на  $F_{em}$  – ( $F$ -ratio).

В последната колона е оцененото ниво на значимост (Significance level). Не е необходимо изследователят да познава детайлно процедурите за изчисляване на  $F$ -тест. Интерес представлява нивото на значимост и всеки изследовател с лекота трябва да борави с него, да формулира нулева ( $H_0$ ) и алтернативна хипотеза ( $H_1$ ), както и да формулира изводи от получените резултати за нивото на значимост.

*Ако нивото на значимост е по-малко от зададеното (или от 0.05) –  $H_0$  се отхвърля и изводът е, че независимата променлива оказва значимо, закономерно влияние върху стойностите на зависимата променлива. Обратно, ако нивото на значимост е по-голямо от зададеното (или 0.05), изучаваният фактор не оказва съществено влияние.*

В разглеждания пример, според  $H_0$  дозата на серототина не оказва закономерно влияние върху теглото на тумора. Теоретичната стойност на  $F_t = 3,68$  при  $p = 0,05$  и степени на свобода (2,15). Тъй като  $F_{em} > F_t$  ( $14.5 > 3.68$ ), а  $p < 0.05$ , трябва да се отхвърли  $H_0$  и да се приеме  $H_1$ , т. е. **въздействието на серотонина върху теглото на тумора е закономерно**. Очевидно е, че при високите дози на серотонина средното тегло на тумора е по-малко в сравнение с ниските дози.

### **Дисперсионен анализ при многофакторен равномерен комплекс**

Почти няма разлика в процедурите за прилагането на многофакторния /Multifactor Analysis of Variance/ и еднофакторния дисперсионен анализ. Единственото ограничение е изучаваните фактори да са независими помежду си, т. е. да липсва връзка между тях. Ако има такава зависимост (а това се показва като резултат в таблицата от дисперсионния анализ – Interactions при ниво на значимост по-малко от 0.05), не бива да се използват резултатите за

научни изводи и заключения, т.е. дисперсионният анализ е неприложим.

**Пример 2** (по източник N8): Проведено е проучване за изучаване на факторите, влияещи върху скоростта на утаяване на еритроцитите (СУЕ). Изследван е серум на 72 пациенти – 36 мъже и 36 жени, разпределени по 12 според възрастта (18–20 г. – код 1; 21–40 г. – код 2 и 41–60 г. – код 3) и според време (на денонощието) на изследването (сутрин – код 1; обед – код 2 и вечер – код 3). Да се приложи многофакторен дисперсионен анализ, за да се установи дали полът, възрастта и времето на изследване са значими фактори за нивото на СУЕ. Да се сравнят средните, за да се установи, дали има сигнификантна разлика между тях.

Според  $H_0$  СУЕ не зависи от пола, възрастта и времето на изследване. Получените резултати са представени в **табл. 11.5**.

Компютърната програма кодира факторите с букви: А – пол; В – възраст и С – време на изследване.

**Първо**, в таблицата се проверява нивото на значимост на  $H_0$  относно взаимовръзката между факторите (Factor Interactions – между АВ (пол и възраст); АС (пол и време на изследване); ВС (възраст и време на изследване) и АВС (пол, възраст и време на изследване)). Според  $H_0$  няма значима, закономерна връзка между изучаваните фактори. Вижда се, че нивото на значимост на всички комбинации и общо е по-голямо от 0.05, т. е.  $H_0$  се приема – между изучаваните фактори няма взаимовръзка, което ни дава основание да приложим многофакторният дисперсионен анализ.

**Второ**, в таблицата се проверява нивото на значимост на  $H_0$ , според която полът, възрастта и времето на изследване не са значими фактори за СУЕ. Вижда се, че нивото на значимост на  $H_0$  за пола (0.0000) и за възрастта (0.0000) е по-малко от 0.05, а за времето на изследване е по-голямо от 0.05 (0.11).

Това показва, че полът и възрастта са значими, закономерни фактори, които оказват влияние върху скоростта на утаяване на еритроцитите, а времето на изследване не оказва закономерно влияние върху СУЕ.

**Табл. 11.5. Резултати от дисперсионен анализ чрез софтуерен пакет Statgraphics**

Analysis of Variance for sue - Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:pol	72,0	1	72,0	36,00	0,0000
B:agegr	108,0	2	54,0	27,00	0,0000
C:hor	9,0	2	4,5	2,25	0,1152
INTERACTIONS					
AB	12,0	2	6,0	3,00	0,0581
AC	3,0	2	1,5	0,75	0,4772
BC	5,0	4	1,25	0,63	0,6467
ABC	5,0	4	1,25	0,63	0,6467
RESIDUAL	108,0	54	2,0		
TOTAL (CORRECTED)	322,0	71			

All F-ratios are based on the residual mean square error.

The StatAdvisor

The ANOVA table decomposes the variability of sue into contributions due to various factors. Since Type III sums of squares (the default) have been chosen, the contribution of each factor is measured having removed the effects of all other factors. The P-values test the statistical significance of each of the factors. Since 2 P-values are less than 0,05, these factors have a statistically significant effect on sue at the 95,0% confidence level.

При работа със STATGRAPHICS се получават бързо и лесно резултатите за средните стойности на СУЕ в таблици (Means table) – *табл. 11.6*.

В първата колона (Level) на *табл. 11.6* са кодовете на различностите на факторните променливи – пол, възраст, време на изследване и комбинациите между тях.

Във втората колона (Count) е броят на случаите във всяка различност (група).

В третата колона (Mean) са средните стойности на всяка група.

В четвъртата колона (Std. Error) е специфичната стандартна грешка на средните в групите.

**Табл. 11.6. Резултати от дисперсионен анализ чрез Statgraphics**

Table of Least Squares Means for sue  
with 95,0 Percent Confidence Intervals

Level	Count	Mean	Std. Error	Lower Limit	Upper Limit
GRAND MEAN	72	5,0			
pol					
1	36	4,0	0,235702	3,52744	4,47256
2	36	6,0	0,235702	5,52744	6,47256
agegr					
1	24	3,5	0,288675	2,92124	4,07876
2	24	5,0	0,288675	4,42124	5,57876
3	24	6,5	0,288675	5,92124	7,07876
hor					
1	24	4,5	0,288675	3,92124	5,07876
2	24	5,25	0,288675	4,67124	5,82876
3	24	5,25	0,288675	4,67124	5,82876
pol by agegr					
1 1	12	3,0	0,408248	2,18151	3,81849
1 2	12	4,0	0,408248	3,18151	4,81849
1 3	12	5,0	0,408248	4,18151	5,81849
2 1	12	4,0	0,408248	3,18151	4,81849
2 2	12	6,0	0,408248	5,18151	6,81849
2 3	12	8,0	0,408248	7,18151	8,81849
pol by hor					
1 1	12	3,5	0,408248	2,68151	4,31849
1 2	12	4,5	0,408248	3,68151	5,31849
1 3	12	4,0	0,408248	3,18151	4,81849
2 1	12	5,5	0,408248	4,68151	6,31849
2 2	12	6,0	0,408248	5,18151	6,81849
2 3	12	6,5	0,408248	5,68151	7,31849
agegr by hor					
1 1	8	3,0	0,5	1,99756	4,00244
1 2	8	4,0	0,5	2,99756	5,00244
1 3	8	3,5	0,5	2,49756	4,50244
2 1	8	4,25	0,5	3,24756	5,25244
2 2	8	5,0	0,5	3,99756	6,00244
2 3	8	5,75	0,5	4,74756	6,75244
3 1	8	6,25	0,5	5,24756	7,25244
3 2	8	6,75	0,5	5,74756	7,75244
3 3	8	6,5	0,5	5,49756	7,50244

The StatAdvisor

This table shows the mean sue for each level of the factors. It also shows the standard error of each mean, which is a measure of its sampling variability. The rightmost two columns show 95,0% confidence intervals for each of the means. You can display these means and intervals by selecting Means Plot from the list of Graphical Options.

В последните две колони е долната (Upper limit) и горната граница (Lower limit) на доверителният интервал на средните при 95 % доверителна вероятност, формиран по един от следните методи:

- **Conf. Int.** – доверителен интервал, изграден чрез t-коэффициент и стандартната грешка на средните;
- **LSD** (Least Significant Differences) – доверителен интервал по метода на най-малките значими разлики;
- **Tukey method** (Tukey's Honest Significant Differences) – доверителен интервал по метода на Тюки;
- **Scheffe method** – доверителен интервал по метода на Шефе (при равни условия е най-консервативен);
- **Bonferroni method** – доверителен интервал по метода на Бонферони.

Когато чрез ANOVA е установено, че  $F$  отношението е значимо, за сравняване на комбинациите между средните на трите изследвани фактора се прилагат специални методи, наричани **post hoc тестове** за множествени сравнения на контрасти. Сравняването по двойки групи може да се извърши и с t-теста на Студент, но при използването му многократно, съществено се увеличава вероятността за грешка от I род.

**Post hoc тестовете** преодоляват вероятността за грешка тип I, т.е. вероятността за грешки на множествените сравнения, освен това и преодоляват вероятността за грешка на експеримента. При прилагането на post hoc тестовете се работи с предварително фиксирана (желана) вероятност за грешка от I род която остава еднаква при всички двойки сравнения. Затова при тези тестове се говори за вероятност за грешка на сравненията и вероятност за грешка на експеримента.

Post hoc тестовете, които се използват за множествени сравнения и формиране на доверителните интервали на разликите между груповите средни, достъпни в компютърната програма STATGRAPHICS в Multiple range tests, са: LSD (Least Significant Differences); Tukey (Tukey's Honest Significant Differences); Scheffe; Bonferroni; Newman-Kenls и Duncan.

Решението за това кой метод да се приложи зависи от броя на изследваните случаи в групите (равен или не) и вида на разпреде-



лението (нормално или асиметрично). Методът на Шефе се препоръчва при сложни сравнения и контрасти, при неравни извадки, но когато F-отношението на ANOVA се проверява при  $\alpha=0,05$ , post hoc теста трябва да се проверява при  $\alpha=0,10$ . Методите на Тюки, Нюман-Колс и Дънкан са статистически по-мощни при двойки сравнения (контрасти) с равен обем на извадките.

Както при еднофакторния, така и при многофакторния ANOVA, с post hoc анализа се установява между кои двойки групи разликата е сигнификантна. В табл. 11.7 такава разлика е отбелязана със звездичка при  $p=0,05$ . Ако изследователят желае, може да промени нивото на значимост.

**Табл. 11.7. Резултати от сравняване на средните по двойки чрез LSD post hoc тест**

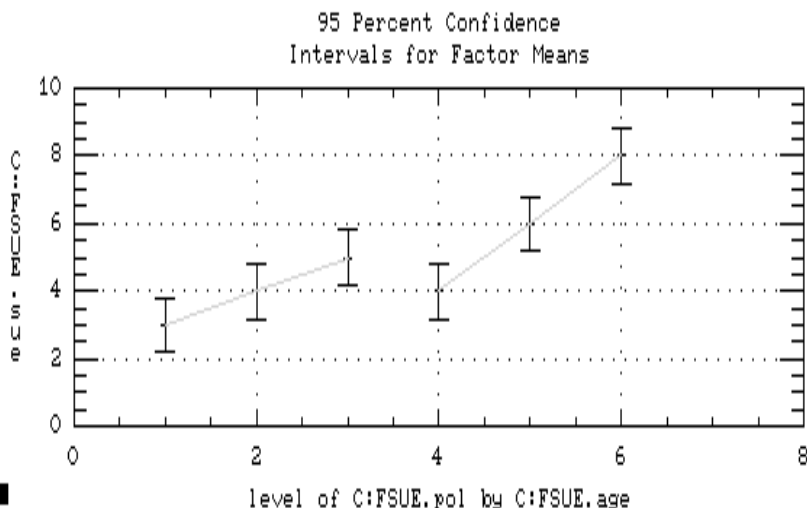
Multiple Range Tests for sue by agegr			
-----			
Method: 95,0 percent LSD			
agegr	Count	LS Mean	Homogeneous Groups
-----			
1	24	3,5	X
2	24	5,0	X
3	24	6,5	X
-----			
Contrast		Difference	+/- Limits
-----			
1 - 2		*-1,5	0,818177
1 - 3		*-3,0	0,818177
2 - 3		*-1,5	0,818177
-----			
* denotes a statistically significant difference.			
The StatAdvisor			
-----			
This table applies a multiple comparison procedure to determine which means are significantly different from which others. The bottom half of the output shows the estimated difference between each pair of means. An asterisk has been placed next to 3 pairs, indicating that these pairs show statistically significant differences at the 95,0% confidence level. At the top of the page, 3 homogenous groups are identified using columns of X's. Within each column, the levels containing X's form a group of means within which there are no statistically significant differences. The method currently being used to discriminate among the means is Fisher's least significant difference (LSD) procedure. With this method, there is a 5,0% risk of calling each pair of means significantly different when the actual difference equals 0.			



В табл. 11.7 са представени резултатите от сравняването на средните стойности по двойки чрез LSD post hoc теста. Със звездички са отбелязани при трите двойки разлики, т.е. изводът е, че съществува сигнификантна разлика на средните стойности на СУЕ между трите възрастови групи – 18–20 г. с 21–40 г; 18–20 г. с 41–60 г; и 21–40 г. с 41–60 г.

Резултатите от сравняванията, проведени с шестте post hoc теста са идентични, тъй като моделът на многофакторния дисперсионен анализ е равномерен комплекс.

При обсъждането е добре да се ползват и графични образи на резултатите. В компютърната програма се създава графика (Means plot) на груповите средни с доверителни интервали. Както се вижда от **фиг. 11.3**, че има сигнификантна разлика между средните стойности на СУЕ в двете групи по пол; както при мъжете, така и при жените с увеличаване на възрастта се увеличават стойностите на СУЕ.



**Фиг. 11.3.** Средни стойности на СУЕ и доверителни интервали при мъже и жени според възрастта

Накрая е необходимо да се разгледа накратко въпроса за повишаване мощността на статистическите тестове чрез сравняване



на двойки по-сложни контрасти. Ако броят на разновидностите на някой фактор е по-голям, за да се избегне припокриване на информацията, при сравняването по двойки на една и съща средна (процедурата намалява мощността на статистическия тест) групите може да се обединяват.

Например, ако е необходимо да се сравнят по двойки средните стойности на четири разновидности на фактора, трябва да се проверят шест нулеви хипотези:

$$1. H_0: \bar{x}_1 = \bar{x}_2$$

$$2. H_0: \bar{x}_1 = \bar{x}_3$$

$$3. H_0: \bar{x}_1 = \bar{x}_4$$

$$4. H_0: \bar{x}_2 = \bar{x}_3$$

$$5. H_0: \bar{x}_2 = \bar{x}_4$$

$$6. H_0: \bar{x}_3 = \bar{x}_4$$

Вижда се, че една и съща информация за всяка средна се припокрива при три двойки сравнявания. В този случай се препоръчва средните стойности да се подредят и номерират във възходящ ред ( $\bar{x}_1 < \bar{x}_2 < \bar{x}_3 < \bar{x}_4$ ).

При сравняването по двойки е възможно да се установи, че при всички двойки  $p > 0.05$ , т. е. приема се  $H_0$  – не съществува сигнификантна разлика между средните. В тези случаи изследователят може да обсъди нови хипотези и да комбинира средните по подходящ за съответната област на проучването начин. Например, една от възможните комбинации е проверка на три вместо на шест хипотези и резултатите от проверката са следните:

$$H_0: \bar{x}_1 = \bar{x}_4; p > 0,05$$

$$H_0: \bar{x}_2 = \bar{x}_3; p > 0,05$$

$$H_0: \bar{x}_1 + \bar{x}_4 = \bar{x}_2 + \bar{x}_3; p < 0,05,$$



т.е. няма сигнификантна разлика между първата и четвъртата средна и между втората и третата, но има сигнификантна разлика между групите с ниски и високи стойности в сравнение с групите със средни стойности (възможни са и други комбинации). При това сравняване с ANOVA се повишава мощността на статистическия тест.

## 6. Непараметрични методи за проверка на хипотези

### 6.1. Същност на непараметричните методи

С понятието „*непараметрични методи*“ се обединяват голяма група статистически критерии за проверка на хипотези. Тук се отнасят също и някои подходи за изчисляване на коефициенти на корелация, за изработване на нормативи, за изучаване на циклични промени и др., но основното предназначение на непараметричните методи е за доказване на научни хипотези.

Както отбелязахме по-горе, параметричните методи имат важно ограничително условие, заключаващо се в това, че те са приложими само при количествени променливи, които имат нормално или близко до нормалното разпределение. В действителност твърде често се срещаме с разпределения, различни от нормалното, или не разполагаме с точна информация за характера на разпределението. В такива случаи методично по-правилно е да се използват непараметрични критерии за доказване на хипотези.

Следователно: *първото и най-съществено предимство на непараметричните методи е в това, че те се използват независимо от формата на разпределение, т. е. при всички форми на разпределение на случаите.*

*Второто предимство се заключава в това, че те се прилагат както при качествени, така и при количествени променливи.*

При сравняване на извадки, в които променливите величини са измерени върху ординални (рангови) скали, се използват само непараметрични методи.



**Третото предимство** е свързано с това, че те са по-лесно приложими, елементарни като на изчислителната техника и по-икономични по отношение на разход на труд и средства.

**Недостатък** на непараметричните методи е **по-малката мощност** на използваните при тях статистически критерии.

## 6.2. Критерий хи-квадрат ( $\chi^2$ ) за оценка на научни хипотези

### Логическа основа на критерия $\chi^2$

**Критерият  $\chi^2$**  е най-често използваният непараметричен тест за оценка на хипотези. Създаден е от английския статистик К. Пирсон и се нарича още **критерий на Пирсон или критерий на съгласието** (съответствието).

Принципната основа на неговото приложение се състои в **сравняване на фактическите честоти ( $f$ ) с теоретичните честоти ( $f_i$ )**.

**Фактически честоти ( $f$ ) са тези, които се наблюдават в конкретното проучване.**

**Теоретични честоти ( $f_i$ ) – които биха се получили, ако е вярна нулевата хипотеза.** Наричат се още **оачквани честоти**.

Логически следва изводът, че колкото по-голяма е разликата между  $f$  и  $f_p$ , толкова по-голяма е вероятността различието да бъде съществено, т. е. да се отхвърли  $H_0$  и обратно – колкото по-малка е разликата между  $f$  и  $f_i$ , толкова по-голяма е вероятността за потвърждаване на  $H_0$ .

Оценката на научни хипотези с помощта на критерият  $\chi^2$  може да се извършва при различни схеми – при една променлива с две или повече разновидности, при две променливи с по две или повече разновидности и при повече от две променливи. Данните се представят в четирикратни и многократни таблици.



**Четирикратни таблици** са тези, в които независимата и зависимата променливи величини имат само по две разновидности. Наричат се още таблици  $2 \times 2$  или таблици на контингенция.

**Многократни таблици** се тези, в които поне една от променливите величини има повече от 2 нива.

### Методика на прилагане на критерия $\chi^2$

Последователността на работа и при двата начина на представяне на изходните данни (четирикратни и многократни таблици) е еднаква и преминава през следните стъпки:

**I етап – формулиране на  $H_0$ :** няма различие в сравняваните групи.

**II етап – изчисляване на теоретичните честоти за всяка клетка от таблицата при условие, че  $H_0$  е вярна.**

Ако работим с четирикратна таблица, достатъчно е да се определи теоретичната честота само за една от клетките на таблицата, а останалите три теоретични честоти ще се получат като допълващи до цифрите в сумарния ред и сумарната колона.

### III етап – определяне на степените на свобода ( $df$ или $k$ )

Степените на свобода се определят като  $df = (S - 1)(R - 1)$ , където  $S$  е брой нива на независимата (факторната) променлива, а  $R$  – брой на разновидностите на зависимата променлива

**При четирикратна таблица  $df$  винаги е единица.**

**При многократни таблици  $df$  винаги е по-голяма от единица**, тъй като поне единият признак има повече от две разновидности. Степента на свобода в този случай се изчислява като:

$df = (S - 1)(R - 1)$ , където  $S$  е брой разновидности на първата променлива;  $R$  е произведение от броя на разновидностите на останалите променливи.

Както бе отбелязано в предходните раздели, тълкуването на понятието „**степени на свобода**“ е доста трудно. Най-общо, **степените на свобода представляват брой свободни, истински независими значения, които могат да приемат променливите в дадено проучване, така че да не се промени крайният резултат.**



Напр., в четирикратна таблица само една от клетките може да приеме свободно значение. Числата в останалите клетки ще я допълват, за да не се променят резултатите в сумарния ред и сумарната колона на таблицата и общия брой наблюдавани случаи.

**IV етап – изчисляване на стойността на  $\chi^2$**

$$(11.5) \quad \chi^2 = \sum \frac{(f - f_t)^2}{f_t}$$

За четирикратна таблица изчислителната процедура за  $\chi^2$  може да бъде опростена и без да се определят очакваните честоти да се замести с фактическите честоти в следната формула:

$$(11.6) \quad \chi^2 = \frac{N(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

където:

$a, b, c, d$  са съответните фактически честоти в таблицата  $(a+b), (c+d), (a+c), (b+d)$  – числата в сумарния ред и сумарната колона

**V етап – корекция на Йетс**, ако таблицата е четирикратна (т.е.  $df = 1$ ) и ако очакваните честоти не са изразени с цели числа.

**VI етап – определяне на нивото на значимост на  $H_0$  в таблиците за критичните стойности на  $\chi^2$**  съгласно получената емпирична стойност на  $\chi^2$  и съответната степен на свобода.

Въз основа на стойността на  $\chi^2$  и степента на свобода, в таблицата за критичните стойности на  $\chi^2$  (**Приложение 2**) се търси вероятността  $p$ , с която се подкрепя  $H_0$ .

**VII етап – формулиране на крайния извод.**

Ако  $p > 0.05$ ,  $H_0$  се приема и обратно – при  $p < 0.05$   $H_0$  се отхвърля и се приема алтернативната хипотеза  $H_1$ .

Нивото на значимост на  $H_0$  (при определено значение на  $df$ ) е в обратна зависимост от величината на  $\chi^2$ . За практическата работа няма значение дали то е 0.01 или 0.001 и по-малко. Важно е само дали  $p$  е по-малко или по-голямо от 0.05, което е гранично ниво за отхвърляне или приемане на  $H_0$ .



**Пример:** При 10-годишно лонгитудинално наблюдение на две групи мъже (пушачи и непушачи) е проследено възникването на хронични обструктивни белодробни заболявания (ХОБЗ). Данните са представени в четирикратна таблица (табл. 11.8), в която тютюнопушенето е независима променлива (проучваният фактор), а ХОБЗ е резултатът (зависима променлива).

**Табл. 11.8. Честота на ХОБЗ и тютюнопушене**

	Със ХОБЗ	Без ХОБЗ	ОБЩО
<b>Пушачи</b>	100 (75) <i>a</i>	400 (425) <i>b</i>	<b>500 – <i>a</i> + <i>b</i></b>
<b>Непушачи</b>	50 (75) <i>c</i>	450 (425) <i>d</i>	<b>500 – <i>c</i> + <i>d</i></b>
<b>ОБЩО</b>	<b>150 <i>a</i> + <i>c</i></b>	<b>850 <i>b</i> + <i>d</i></b>	<b>1000</b>

Въз основа на получените данни трябва да отговорим на въпроса: „Тютюнопушенето оказва ли съществено влияние върху заболяемостта от ХОБЗ?“

**Последователността на работа е следната:**

**1. Формулиране на  $H_0$ :** Тютюнопушенето не оказва влияние върху заболяемостта от ХОБЗ. Различията в заболяемостта при пушачите и непушачите са несъществени; доколкото такива се наблюдават, то те се дължат на случайни фактори.

**2. Определяне на теоретичните честоти във всяка клетка на таблицата.** Първата теоретична стойност ще бъде равна на  $(150 \cdot 500) : 1000 = 75$ . За удобство нанасяме тази теоретична стойност в скоби редом със съответната фактическа честота. Останалите три теоретични честоти допълват резултатите в сумарния ред и сумарната колона и ги нанасяме в скоби в съответните клетки на таблицата.

**3. Определяне на степените на свобода.** При четирикратна таблица  $df = 1$ .

**4. Изчисляване на  $\chi^2$**  чрез заместване във формулата.

**5. Корекция на Йетс за непрекъснатост** при работа с таблици  $2 \times 2$  тук не се налага, тъй като очакваните честоти са изразени с дискретни числа, както фактическите честоти.

**6. Определяне на вероятността на  $H_0$**  По таблицата за критичните стойности на  $\chi^2$  в реда, съответстващ на  $df = 1$ , намираме че  $\chi^2 = 19.6$  се разполага надясно от последната колона, т.е. възприетата  $H_0$  има  $p < 0.001$ . Следователно,  $H_0$  е невярна. Тя се отхвърля и се възприема алтернативната хипотеза  $H_1$ .

**7. Крайният извод се формулира по следния начин:** Има съществено различие в заболяемостта от ХОБЗ при пушачи и непушачи. Тютюнопушенето оказва съществено влияние върху заболяемостта от ХОБЗ.

### **Оценка и условия за използване на $\chi^2$**

Недостатък на непараметричните критерии е тяхната по-малка мощност. В конкретния случай с  $\chi^2$ , това означава, че не може да се направи преценка за силата на влияние на тютюнопушенето, а се стига само до извода, че има такова влияние. Това обстоятелство е наложило широкото използване на  $\chi^2$  като филтър при подбор на значими фактори сред множество изследвани фактори, след което е целесъобразно да се приложат други критерии за силата на зависимостите – напр., коефициент на корелация.

### **Ограничителни условия за използване на $\chi^2$ :**

- \*  $\chi^2$  трябва да се изчислява само от абсолютни числа;
- \* изчислението трябва да се опира на достатъчно голям брой случаи; заключенията зависят от възприетия начин на групировка на признаците;
- \* не се прилага  $\chi^2$  при  $k = 1$ , ако има очаквани честоти по-малки или равни на 5 или в четирикратна таблица общият брой случаи е по-малък от 20.
- \* не може да се изчислява  $\chi^2$  при  $k > 1$ , ако повече от 20% от очакваните честоти са по-малки или равни на 5.
- \* приложим е само при категорийни данни.

### **Други непараметрични критерии**

***U-критерий на Ман-Уитни*** – един от най-мощните непараметрични тестове за сравняване на две популации на основата на





две независими извадки. Използва се за проверка на  $H_0$ , че двете популации имат идентично разпределение. ***U-критерия*** е аналог на *t*-критерия при сравняване на средни величини от независими извадки и се прилага, когато разпределението е неизвестно или се различава от нормалното. Изчисляването на емпиричната стойност на *U*-критерия се опира на прегрупиране на данните в извадките и подреждането им по рангове. За оценка на вероятността на  $H_0$  се използват специални таблици за *U* при еднакви и нееднакви по размер извадки и при едно– или двустранен тест.

***Медианен тест*** – използва се при две независими извадки и се опира на сравняване на медианите. Данните се представят в рангова скала. За всяка извадка се определя медианата и броя случаи, които попадат под и над нея. След това се построява четирикратна таблица и се изчислява  $\chi^2$ , след което се сравнява с таблицата за критичните стойности на  $\chi^2$ . Медианният тест може да се прилага и при повече от две независими извадки с ординално измерване на зависимата променлива.

***Критерий на Крускал-Уолис*** – използва се при три и повече извадки. Той е аналог на *F*-критерия при еднофакторен дисперсионен анализ и е подходящ, когато видът на разпределението е неизвестен или то се отличава от нормалното и резултатът (зависимата променлива) е върху ординална скала. Явява се логично разширение на теста на Ман-Уитни.

***Знаков тест*** – предназначен е за проверка на хипотези за медианата в популацията и често включва използване на независими извадки от типа „преди-после“. Нулевата хипотеза твърди, че разликата в медианите е нула. Не изисква нормално разпределение. Знаковият тест е по-слаба алтернатива на знаково-ранговият тест на Уилкоксон. Използва се при рангова скала.

***Знаково-рангов тест на Уилкоксон*** – използва се при зависими извадки (най-често от вида „преди-после“) и ординална скала на измерване на зависимата променлива. Опира се на сравняване на медианите. Той е по-мошен от знаковият тест.

***Критерий на Колмогоров-Смирнов*** – използва се при една и при две извадки за проверка на съответствието между емпиричното и

теоретичното разпределение. Използва се при непрекъснати променливи величини, но е възможен и при дискретни разпределения. Не изисква нормално разпределение.

### **Интерпретиране на статистическите тестове**

В процеса на анализа на данните изследователят избира подходящ тест, изчислява стойността на критерия и го сравнява с таблични стойности, за да установи дали резултатите са **статистически значими**. За медицинските проучвания, в допълнение към статистическата значимост на резултатите, е необходимо да се разглежда и тяхната **практическа и медицинска значимост**, която включва:

- **оценка на размера на ефекта**, т.е. колко големи са наблюдаваните зависимости или различия в данните;
- **социалната и клинична значимост** на такива фактори като ценова ефективност, качество на живот и др.

Размерът на ефекта в дадено проучване се отнася до действителните размери на наблюдаваните различия между групите или до силата на зависимостите между променливите. Важно е да се признае, че макар и да е достигната статистическа значимост, размерът на ефекта може да бъде клинично незначим или с толкова минимален ефект, че да не представлява интерес.

Внимателно трябва да се интерпретират и нулевите резултати, показващи отсъствие на влияние и ефект. Възможно е изследователят да пропусне даден ефект поради неговия малък размер и/или недостатъчно случаи в анализа. Статистическата сила на даден анализ измерва вероятността за точно откриване на действителен ефект от определен размер. Следователно, даден нулев резултат може да бъде функция на ниска статистическа сила, а не на липса на действителен ефект.

Съществуват критерии, извън статистическата значимост, които трябва да се отчитат преди вземане на решения за клиничната приложимост на изследванията. Те се повлияват в голяма степен от ценностите и икономическите ограничения в управлението на здравните грижи в дадена общност.