Глава 11

СТАТИСТИЧЕСКА ПРОВЕРКА НА ХИПОТЕЗИ

Г. Грънчарова

В тази глава:

11.1. Значение и същност на сравняването на данни от извадки

При някои неекспериментални стратегии, когато целта на изследванията е да се опишат характеристиките на популацията на базата на данните от извадки, статистическото оценяване е достатъчно за анализа на данните.

Други изследователски стратегии имат за цел проверка и сравняване на хипотези. Например, 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. Естествено 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 (тестуване) на хипотези.*** Това е значително по-високо ниво на изследователска работа, отколкото описателната характеристика да данните.

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.

11.2. Основни понятия при проверка на хипотези

11.2.1. 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 третират 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 й 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., $\mathbf{H_0}$ 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 ($\mathbf{H_0}$), предполага, че няма различие или ако такова видимо се наблюдава, то се дължи единствено на случайност. Например, в даден клиничен опит върху ново лекарство, $\mathbf{H_0}$ би могла да гласи: новото лекарство не е по-добро от използваното понастоящем. Такава $\mathbf{H_0}$ може да се запише по-кратко: *няма разлика между двете лекарства.*

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

Окончателното заключение след прилагане на теста за проверка на хипотези винаги се дава от гледна точка на $\mathbf{H_0}$ - ***"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 $\mathbf{H_0}$ е вярна, то само предполага, че няма достатъчно доказателства $\mathbf{H_0}$ да бъде отхвърлена в полза на $\mathbf{H_1}$**.**

Алтернативна хипотеза

Алтернативната хипотеза, $\mathbf{H_1}$, представлява твърдение за това, което статистическият тест за проверка трябва да установи. Например, в даден клиничен опит с ново лекарство $\mathbf{H_1}$ би могла да бъде, че новото лекарство има различен ефект в сравнение с досега използваното. В този случай $\mathbf{H_1}$ 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. $\mathbf{H_1}$ би могла също да гласи, че новото лекарство е по-добро от досегашното. В такъв случай тя е ***еднострaнна (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 $\mathbf{H_0}$. Има 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 очакват, отколкото да се формулира условие за нулева хипотеза. 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_1: 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.

11.2.2. Грешки от I и II род

Грешките, които могат да бъдат допуснати при проверката на хипотези, се определят от гледна точка на $\mathbf{H_0}$. След анализирането на данните, изследователят възприема $\mathbf{H_0}$ (ако няма значими различия) или отхвърля $\mathbf{H_0}$ (ако наистина има значими различия). Отхвърлянето на $\mathbf{H_0}$ 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. Те се наричат ***грешка от I род и грешка от II род.*** За да разберем същността на тези грешки, нека представим възможностите за решения относно $\mathbf{H_0}$ чрез таблица 2х2:

Действителна ситуация	**Взето решение**	
	H_0 се отхвърля	**H_0 се приема**
H_0 е вярна	**Грешка от I род**	**Правилно решение**
H_1 е невярна	**Правилно решение**	**Грешка от II род**

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

Грешка от I род се получава, ако отхвърлим $\mathbf{H_0}$, 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 $\mathbf{H_0}$. 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 ***α (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. $\mathbf{H_0}$ 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 $\mathbf{H_0}$, но тази вероятност никога не е 0.

Грешка от II род се допуска, ако приемем $\mathbf{H_0}$, а в същност тя е невярна – напр., 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. $\mathbf{H_0}$ не се отхвърля, а фактически тя е погрешна.

Вероятността за грешка от II род се означава с ***β (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"*** (влиянието на независимата променлива).

Вероятността за грешки от II род β е в обратна зависимост от мощността на теста, който се изразява като 1-β. Следователно, колкото по-малко е β, 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.

11.2.3. 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 α. 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 $\mathbf{H_0}$. например, ако приемането на невярна $\mathbf{H_0}$ има по-сериозни последици, отколкото отхвърлянето й, на нивото на значимост α 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 **р)** 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.).

Р-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 $\boldsymbol{\alpha = 0.05}$, това се записва като $\mathbf{p < 0.05}$. Р-стойността подчертава силата на доказателствата за отхвърляне на $\mathbf{H_0}$. Колкото по-малко е р, толкова по-убедително е отхвърлянето на $\mathbf{H_0}$.

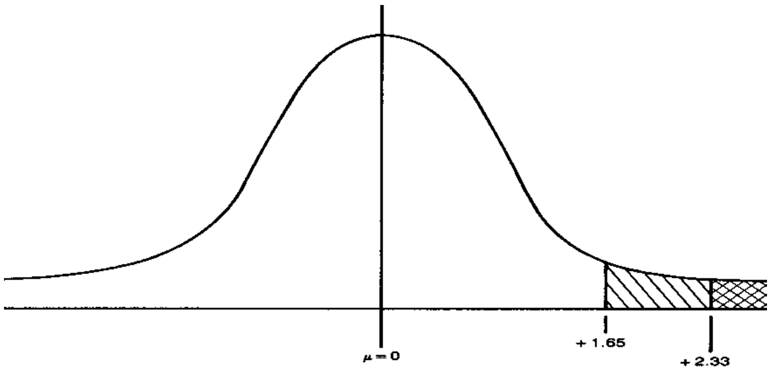
11.3. Статистически тестове за проверка на хипотези

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

Едностранен и двустранен тест

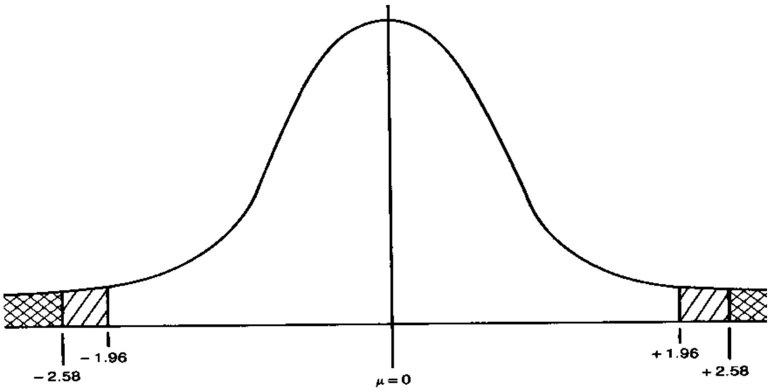
"Страни" или "опашки" в разглеждания контекст се наричат краищата на кривата на разпределението на извадката или популацията.

Едностранен 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 едностранния тест представлява серия от стойности, които са по-малки или по-големи от критичната стойност на теста.



Фиг. 11.1. Едностранен тест за значимост при нормална крива

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. Критичната област за двустранния тест представлява серия от стойности, които са по-малки от първата критична стойност на теста и серия от стойности, които са по-големи от втората критична стойност на теста.



Фиг. 11.2. Двустранен тест за значимост при нормална крива

Изборът между едностранен 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.

Едностранният тест е по-мощен, 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 едностранен тест 0.025. При двустранен тест и $n > 120$ стойностите на **z** или **t** трябва да достигнат **± 1.96** за p=0.05 и **± 2.58** при p=0.01, докато при едностранен тест те са доста по-ниски – 1.65 и 2.33 за същите нива на значимост.

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 обработка 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* $\boldsymbol{H_0}$. 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* $\boldsymbol{H_0}$ *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, $\mathbf{H_0}$ се отхвърля и обратно.

Мощност на теста

Мощността (силата) на теста определя вероятността да се приеме алтернативната хипотеза ($\mathbf{H_1}$), 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-β.** 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 сравняването им с таблиците за критичните им стойности се използва понятието "степени на свобода" (***k*** или ***df*** – degrees of freedom).

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

11.4. Основни принципи и процедури за проверка на хипотези

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

11.4.1. 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 хипотези 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 $\mathbf{H_0}$ и $\mathbf{H_1}$ винаги е равна на единица или 100%. Следователно, ако $\mathbf{H_0}$ има вероятност по-висока от 0.05 (5%), то $\mathbf{H_1}$ 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 $\mathbf{H_0}$, т.е. изучаваните фактори не оказват влияние).

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

5. При всяка проверка на хипотези изследователят е поставен пред задачата да вземе определено решение за потвърждаване или отхвърляне на проверяваната хипотеза. Стремежът 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 временен 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 извадки 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***, свързани с отношението на разликата между средни величини и относителни дялове към съответната стохастична грешка;
* ***параметрични методи при повече от две групи наблюдения***, основаващи се на отношението на дисперсии (дисперсионен анализ);
* ***непараметрични методи***, основаващи се предимно на χ^2 .

11.4.2. Основни стъпки на процедурата за проверка на хипотези

Независимо от това какъв метод се използва, ***процедурата за проверка на хипотези преминава през следните общоприети стъпки:***

1. Формулиране на H_1 - какво възнамеряваме да оценим: че резултатите са "действителни" или "значими", т.е. независимата променлива оказва влияние върху зависимата променлива или има действително различие между групите. Уточнява се също вида на хипотезата – едно- или двупосочна.

2. Формулиране на работната хипотеза H_0 , която е логично противоположна на **H_1.** Нулевата хипотеза гласи, че всяко различие в данните се дължи просто на случайност, т.е. независимата променлива не оказва ефект върху зависимата променлива; различието се дължи на случайност.

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

4. Изчисляване на цифровата стойност на статистическия тест за определяне вероятността на H_0 от наблюдаваните данни.

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.

11.4.3. Избор на статистически тестове за значимост

Изборът на подходящи статистически критерии за анализ на данните в конкретни изследвания се определя от следните ***основни съображения:***

1. От скалата за измерване на променливите величини в извадките (номинална, ординална, интервална, пропорционална).

2. От броя на групите (извадките) в дадено изследване (1, 2 или повече).

3. От вида на сравняваните извадки: ***независими*** или ***зависими извадки*** (повторни измервания на едни и същи лица).

Посочените съображения могат да бъдат обобщени в следната таблица:

Табл. 11.1. Избор на статистически тестове за значимост

Скала	**Две извадки**		**Три и повече извадки**	
	Независими	**Зависими**	**Независими**	**Зависими**
Номинална	Хи-квадрат (също и за една група)	МакНемар	Хи-квадрат	Тест на Кохран
Ординална	Медианен тест, U критерий на Ман-Уйтни	Критерии на Уилкоксон Знаков тест	Медианен тест Тест на Крускал-Уолис	Тест на Фридман
Интервална/ пропорционална	z или t-критерий (независим)	z или t-критерий (корелиран) (paired)	Дисперсионен анализ – F-критерий на-Фишер	Дисперсионен анализ – F-критерий на Фишер

В зависимост от дизайна на конкретното изследване, още преди събирането на данните трябва да се избере подходящ тест за анализ на данните.

Пример: Изследовател иска да оцени ефективността на ново лечение в сравнение с конвенционално лечение. Резултатът се измерва чрез 5-степенна ординална скала. Всяко изследвано лице се отнася към една от двете групи. Какъв тест би използвал изследователят за проверка на H_0?

В този случай измерителната скала е ординална и се наблюдават 2 независими извадки. Изследователят може да избере медианен тест или тест на Ман-Уйтни.

По подобен начин може да бъде направен всеки друг избор. Голямото разнообразие на тестовете не позволява детайлното им разглеждане. За това съществува огромна статистическа литература, от която при необходимост може да обърне всеки изследовател.

11.5. Параметрични методи за проверка на хипотези

11.5.1. Сравняване на две групи наблюдения

Сравняване на средни величини чрез t-критерия на Стюдент

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 изчисляване на 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 $\mathbf{s_1} = 8$ 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?

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:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{|30 - 24|}{\sqrt{\frac{8^2}{100} + \frac{6^2}{64}}} = 6/1.1 = 5.45$$

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 - степени 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 критичните 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). Следователно, **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:

$$t = \frac{|p_1 - p_2|}{\sqrt{\dfrac{p_1 \times q_1}{n_1} + \dfrac{p_2 \times q_2}{n_2}}}$$

където:

$\mathbf{p_1}$ – изчислената статистика (пропорция, коефициент за честота) в първата извадка

$\mathbf{q_1}$ – стойност допълваща p_1до 1, 100, 1 000 и т. н. ($1-p_1$, $100-p_1$, $1000-p_1$)

$\mathbf{n_1}$ - броят на случаите в първата група

$\mathbf{p_2}$, $\mathbf{q_2}$, $\mathbf{n_2}$ - съответните показатели за втората извадка

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

Ако $\mathbf{H_0}$ 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 $\mathbf{H_0}$ 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.

11.5.2. 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 лекарствено средство и т.н.). Ако решим да анализираме различията между групите чрез t-критерия на Стюдент, 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 ***основни 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*** – при еднакъв брой случаи в групите;

* ***неравномерен комплекс*** – при 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 Стюдент и дисперсионния анализ е в това, че при първия подход се сравняват директно средните аритметични величини в групите, а дисперсионният анализ 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; такова 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 $s_t^2 = s_i^2$, 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 критерий за значимост. При три и повече групи подходящият тест е **F**-критерий (критерий на Фишер по името на един от създателите на този метод).

$$F = \frac{s_e^2}{s_i^2}$$

- 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 таблица, която позволява да се направи заключение за вероятността на нулевата хипотеза и да се формулира крайният извод за достоверността на различията между групите.

11.6. Непараметрични методи за проверка на хипотези

11.6.1. Същност на непараметричните методи

Понятието “***непараметрични методи***” обединява голяма група статистически ***критерии*** за проверка на хипотези.

Първо предимство - *използват се независимо от формата на разпределение,* т. е. при всички форми на разпределение на случаите.

Второ предимство - *прилагат се както при качествени, така и при количествени променливи*. При променливи величини върху ординални скали, се използват само непараметрични методи.

Трето предимство - *те са по-лесно приложими 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.

11.6.2. Критерий хи-квадрат (χ^2)

Логическа основа на критерия χ^2

Критерият χ^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_t).***

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

Теоретични честоти (f_t) са тези, които биха се получили, ако е вярна нулевата хипотеза. Наричат се още ***очаквани честоти***.

Логически следва изводът, че колкото по-голяма е разликата между f и f_t, толкова по-голяма е вероятността различието да бъде съществено, т. е. да се отхвърли $\mathbf{H_0}$ и обратно - колкото по-малка е разликата между f и f_t, толкова по-голяма е вероятността за потвърждаване на $\mathbf{H_0}$.

Оценката на научни хипотези с помощта на критерият χ^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 нива (разновидности). Наричат се още таблици 2 х 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 χ^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 (очакваните) честоти за всяка клетка от таблицата при условие, че H_0 е вярна.

Намирането на очакваните честоти става по формулата:

$$f_t = \frac{\textbf{Брой случай в сумарния ред x брой случаи в сумарната колона}}{\textbf{общ брой наблюдавани случаи}}$$

При четирикратна таблица е достатъчно да се определи теоретичната честота за една от клетките на таблицата, а останалите три теоретични честоти допълват 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 - брой на нивата на независимата (факторната) променлива

R - брой на разновидностите на зависимата променлива (резултата)

При четирикратна таблица df винаги е единица - df = (2-1)(2-1) = 1.

При многократни таблици df винаги е по-голяма от единица, тъй като поне единият признак има повече от две разновидности.

Ако многократната таблица включва групировка по повече от две променливи величини, то:

S е равно на броя на разновидностите на първата променлива;

R е произведение от броя на разновидностите на останалите променливи.

Понятието ***"степени на свобода"*** означава ***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 и общия брой наблюдавани случаи.

IV етап – изчисляване на стойността на χ^2 по формулата

$$\chi^2 = \Sigma \frac{(f - f_t)^2}{f_t}$$

За четирикратна таблица стойността на χ^2 може да се изчисли и чрез следната формула:

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

където:
a, b, c, d са съответните фактически честоти в таблицата
(a+b), (c+d), (a+c), (b+d) са 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 χ^2 съгласно получената емпирична стойност на χ^2 и съответната степен на свобода.

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

VII етап - формулиране на крайния извод

Ако $p>0.05$, $\mathbf{H_0}$ се приема и обратно - при $p<0.05$ $\mathbf{H_0}$ се отхвърля и се приема алтернативната хипотеза $\mathbf{H_1}$.

Нивото на значимост на $\mathbf{H_0}$ (при определено значение на df) е в обратна зависимост от величината на χ^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 $\mathbf{H_0}$.

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, в която тютюнопушенето е независима променлива (проучваният фактор), а ХОБЗ е резултатът (зависима променлива).

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

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

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 (150.500):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. Степени на свобода. При четирикратна таблица df = 1.

4. Изчисляване на стойността на χ^2 чрез заместване във формулата.

$$\chi^2 = \frac{(100-75)^2}{75} + \frac{(400-425)^2}{425} + \frac{(50-75)^2}{75} + \frac{(450-425)^2}{425} = 19.6$$

Същата стойност се получава и ако заместим в опростената формула за χ^2

$$\chi^2 = \frac{N\,(ad - bc)^2}{(a+b).(c+d).(a+c).(b+d)} = \frac{1000.(100.450 - 50.4000)2}{500.500.150.850} = 19.6$$

5. Корекция на Йетс за непрекъснатост при работа с таблици 2х2 тук не се налага, тъй като очакваните честоти са изразени с дискретни числа, както фактическите честоти.

6. Определяне на вероятността на H_0. В таблицата за критичните стойности на χ^2 в реда, съответстващ на df =1, намираме че изчисленият χ^2= 19.6 би се разположил надясно от последната колона. Това означава, че възприетата $\mathbf{H_0}$ има вероятност p<0.001. Следователно, $\mathbf{H_0}$ е невярна. Тя се отхвърля и се възприема алтернативната хипотеза $\mathbf{H_1}$.

*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 ограничителни условия за използване на χ^2

Недостатък на непараметричните критерии е тяхната по-малка мощност. В конкретния случай с χ^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 е наложило използването на χ^2 като филтър за подбор на значими фактори, след което е целесъобразно да се приложат други критерии за силата на зависимостите - напр., коефициент на корелация.

Съществуват някои ***ограничителни условия за използване на χ^2:***

* χ^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 χ^2 при к=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 χ^2 при к>1, ако повече от 20% от очакваните честоти са по-малки или равни на 5.
* приложим е само при категорийни данни.

11.6.3. 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. Може да се използва не само когато данните в извадките са представени като преки измервания, но и когато те са във вид на рангове. За оценка на вероятността на H_0 се използват специални таблици за U при еднакви и нееднакви по размер извадки и с отчитане на посоката на теста (едно- или двустранен).

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 χ^2, след което се сравнява с таблицата за критичните стойности на χ^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. Знаковия тест е по-слаба алтернатива на знаково-ранговия 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-мощен от знаковия тест.

Критерий на Колмогоров-Смирнов – използва 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.

11.7. 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.

11.8. ВЪПРОСИ ЗА САМОПОДГОТОВКА

1. С каква цел се сравняват данни от репрезентативни проучвания?

А. за да се установи съществено ли е различието между тях

Б. за да се уеднакви структурата на средата от която са изчислени

В. за да се обобщят данните за популацията

2. Двустранна (ненасочена) хипотеза е тази, при която не се определя насоката на различието между сравняваните резултати, а само се установява съществуването на различие.

А. вярно Б. невярно

3. Едностранна (насочена) хипотеза е тази, при която се определя насоката на различието между сравняваните резултати.

А. вярно Б. невярно

4. Нивото на значимост на нулевата хипотеза представлява:

А. вероятността, че наблюдаваното различие се дължи на случайност

Б. величината на размера на извадката

В. нито едно от двете

5. Нивото на значимост на алтернативната хипотеза представлява:

А. вероятността, че наблюдаваното различие се дължи на случайност

Б. вероятността, че наблюдаваното различие се дължи на закономерни причини

В. и двете са верни

6. Хипотезата, че средната продължителност на предстоящия живот при мъжете и жените е различна, без да се посочва насоката на това различие, представлява пример за:

А. еднопосочна хипотеза

Б. двупосочна хипотеза

7. Каква е функцията на критичната таблична стойност на даден критерий?

А. тя е равна на изчислената стойност от наблюдаваните данни

Б. тя е границата (стойността), спрямо която се взема решение за приемане или отхвърляне на H_0

В. тя е център на разпределението на стойностите на $\boldsymbol{x}$

8. Решението за използване на едностранен или двустранен тест при проверка на хипотези обикновено се взема след анализиране на данните.

А. вярно Б. невярно

9. Вероятност на нулевата хипотеза $p = 0.001$ означава, че тя със сигурност е невярна.

А. вярно Б. невярно

10. Ако стойността на p (H_0) за даден статистически критерий е $p>0.25$, то:

А. нулевата хипотеза трябва да се приеме

Б. нулевата хипотеза се отхвърля

В. нито едно от двете

11. Хипотезата, че средната продължителност на предстоящия живот при мъжете пушачи е по-ниска от тази при непушачите, е пример за:

А. двупосочна хипотеза

Б. еднопосочна хипотеза

12. Резултатите от даден експеримент са статистически значими, когато:

А. те са важни за статистиците, независимо от тяхната значимост за изследователите

Б. и изследователите, и статистиците ги считат за важни

В. резултатите са важни за изследователите, независимо от важността им за статистиците

Г. наблюдаваният ефект е твърде голям и не може да се обясни със случайност

13. Анализирани са проби за кръвна захар при 140 пациенти по два различни метода. При изчислен $t=2.63$, какъв извод трябва да направим?

А. няма значимо различие в двата метода

Б. има значимо различие в двата метода

14. Ако даден критерий е значим при $p<0.05$, то той е значим и при $p<0.01$.

А. вярно Б. невярно

15. Грешка от I род се допуска винаги, когато:

А. нулевата хипотеза не се отхвърля, когато тя е невярна

Б. алтернативната хипотеза се отхвърля, когато тя е вярна

В. нулевата хипотеза се отхвърля, а тя всъщност е вярна

17. Рискът за допускане на грешка от II род не зависи от допускането на грешка от I род.

А. вярно Б. невярно

18. Въпреки че говорим за два типа грешки, при проверката на дадена хипотеза можем на направим едновременно само една грешка.

А. вярно Б. невярно

19. Грешка от II род се отнася до:

А. отхвърляне на нулевата хипотеза, когато алтернативната е вярна

Б. избиране на грешно решение

В. приемане на нулевата хипотеза, когато тя е невярна

20. При извършване на тест за значимост на H_0 срещу H_1, грешка от II род представлява:

А. вероятността H_1 да е вярна

Б. вероятността за отхвърляне на H_0, ако H_1 е вярна

В. вероятността за приемане на H_0, ако H_1 е вярна

21. Коя хипотеза се приема най-често за работна при научните проучвания?

А. алтернативната

Б. нулевата

В. и двете се използват еднакво често

22. Сумата от вероятностите на нулевата и алтернативната хипотеза:

А. е в интервала от 0.0 до 1.0

Б. може да надхвърли 1.0

В. винаги е равна на 1.0

23. При какво ниво на значимост нулевата хипотеза се приема за вярна?

А. $p<0.05$ Б. $p>0.05$ В. $p<0.01$

24. Ако нулевата хипотеза е вярна, това означава, че:

А. има съществено различие между сравняваните показатели

Б. различието се дължи на закономерни причини

В. различието се дължи на случайни фактори

25. С нарастване на вероятността на нулевата хипотеза намалява вероятността на алтернативната хипотеза и обратно.

А. вярно Б. невярно

26. Направете извод за влиянието на тютюнопушенето върху заболеваемостта от рак на белите дробове, ако нивото на значимост на нулевата хипотеза е $p\ (\chi^2) < 0.05$.
А. няма съществено влияние
Б. има съществено влияние
В. не може да се направи извод от тези данни

27. Колкото е по-голямо различието между фактическите (f) и очакваните (ft) честоти при изчисляване на хи-квадрат:
А. толкова по-вероятно е да бъде отхвърлена нулевата хипотеза
Б. толкова по-вероятно е резултатите да бъдат незначими
В. нито едно от двете

28. От 60 болни мъже на възраст 50-54 г. с еднаква степен на хипертония 30 са лекувани по схема А и 30 - по схема Б. След три месеца състоянието на всички лекувани е оценено като подобрено, без промяна или влошено. Данните са представени в таблица с абсолютни числа. Изберете най-подходящия статистически тест за сравняване на двете лечебни схеми.
А. t - критерий на Стюдент
Б. коефициент на вариация
В. хи-квадрат

29. При сравняване на хи-квадрат и коефициента на корелация (r) кое от посочените твърдения не е вярно?
А. хи-квадрат е с по-малка мощност от r
Б. хи-квадрат се отнася към непараметричните критерии
В. хи-квадрат е с по-голяма мощност от r

30. Двустранен тест за сравняване на хипотези се прилага при:
А. ненасочена хипотеза
Б. насочена хипотеза
В. не зависи от вида на хипотезата

31. Има ли закономерно различие между средното тегло при градски и селски новородени момчета, ако $t = 2.73$ и $df = \infty$ (безкрайност):
А. има закономерно различие
Б. няма закономерно различие
В. различието се дължи на случайни причини

32. Установете има ли закономерно различие между средния ръст на новородените момчета в два района, ако степента на свобода df = 200 и изчислена величина на t = 1.28:

А. има закономерно различие

Б. различието е съществено

В. различието е случайно

33. Установете има ли закономерно различие между средното тегло на новородени момчета (3400г) и новородени момичета (3250г), ако t = 2.85 и df = ∞ (безкрайност):

А. има закономерно различие

Б. няма закономерно различие

В. различието се дължи на случайни причини

34. Проведено е проучване, в което изследваните лица са разпределени в три групи – контролна, опитна А и опитна Б. След лечението се сравняват резултатите за трите групи. Кой е най-подходящият статистически метод за сравняване:

А. коефициент на корелация

Б. t-критерий

В. дисперсионен анализ

35. Ако трябва да изследваме разликата между 5 средни аритметични, коя е най-подходящата статистическа процедура?

А. t-критерий на Стюдент

Б. F-критерий на Фишер

В. хи-квадрат

36. Едностранен тест за сравняване на хипотези се прилага при:

А. насочена хипотеза

Б. ненасочена хипотеза

В. не зависи от вида на хипотезата

37. При обикновен дисперсионен анализ коя от следните мерки представлява оценка на варирането на индивидуалните измервания?

А. междугруповата дисперсия

Б. вътрегруповата дисперсия

В. общата дисперсия

38. За сравняване на пропорции в две извадки най-подходящ е:
А. t-критерият на Стюдент
Б. критерият на Фишер
В. методът на най-малките квадрати

39. За сравняване на резултати в две независими извадки, когато резултативната променлива е количествена, най-подходящ е:
А. методът на най-малките квадрати
Б. t-критерият на Стюдент
В. хи-квадрат

40. Непараметричните методи за проверка на хипотези се използват:
А. само при нормално разпределение
Б. само при алтернативно разпределение
В. при всички форми на разпределение

41. Кой от посочените методи се отнася към непараметричните?
А. хи-квадрат
Б. методът на Мартин
В. методът на най-малките квадрати

42. Кой е най-често използваният критерий при непараметричният анализ?
А. критерий хи-квадрат
Б. t-критерий на Стюдент
В. критерий на Фишер

43. Кой от посочените критерии не се отнася към непараметричните?
А. t-критерий на Стюдент
Б. критерий на Фишер
В. и двата не се отнасят към непараметричните

44. χ^2 (хи-квадрат) се изчислява само на основата на:
А. предварително изчислени средни величини
Б. абсолютни числа
В. предварително изчислени проценти

45. Степента на свобода (df) при хи-квадрат се определя:
А. df = n - 1
Б. df = r.c - 1 (r - редове, c - колони на таблицата)
В. df = (r - 1) (c - 1)

46. Изследовател иска да оцени ефективността на ново лечение в сравнение с конвенционално използвано лечение. Резултатът (зависимата променлива) ще се измерва чрез интервална скала. Всяко изследвано лице ще бъде отнесено към една от двете сравнявани групи. Броят на лицата във всяка от групите е по-малък от 30. Какъв тест би използвал изследователят за анализ на данните?

А. t - критерий на Стюдент

Б. хи-квадрат

В. коефициент на вариация

Отговори на въпросите от глава 11:

1А; 2А; 3А; 4А; 5Б; 6Б; 7Б; 8Б; 9А; 10А; 11Б; 12Г; 13Б; 14А; 15В; 16Б; 17Б; 18А; 19Б; 20В; 21Б; 22В; 23Б; 24В; 25А; 26Б; 27А; 28В; 29В; 30А; 31А; 32В; 33А; 34В; 35Б; 36А; 37Б; 38А; 39Б; 40В; 41А; 42А; 43В; 44Б; 45В; 46А