§5. Проверка на статистически хипотези

1. Статистически хипотези. В този раздел ще изложим накратко основните идеи на теорията на ***проверка на статистически хипотези***, пренебрегвайки голяма част от важните за строго математическо изложение детайли, което по същество няма да попречи за овладяване основните рецепти за нейното използване.

Целта отново ще бъде оценка за параметрите на популацията според стойностите на наблюденията от дадена извадка. Статистическата хипотеза като правило се състои в някакво ***предположение относно вида на разпределението*** на наблюдаваната величина. За да разкрием механизма на основните идеи ще си послужим с IQ примера от **§4**. В този случай наблюдаваната величина X представлява резултати от изпълнението на IQ-тест върху извадка от $n = 64$ души, при което след обработка е получено $\bar{x} = 106.547$ и $s = 9.976$. Разполагаме и с норма на теста $\mu_0 = 100$. Целта на задачата е да отговорим на въпроса доколко наблюдаваната извадка може да се разглежда като представителна за популацията върху която е стандартизиран IQ-теста? Ако в крайна сметка приемем, че това е така, то различието между нормата $\mu_0 = 100$ и емпиричното средно $\bar{x} = 106.547$ трябва да се разглежда като продукт на случайността, която винаги в някаква степен съпровожда такива изследвания. В противен случай трябва да направим заключението, че ***наблюдаваният ефект*** на различие $\bar{x} = 106.547 > \mu_0 = 100$ има закономерен характер или още, че ефектът е ***статистически значим***, при някаква степен на сигурност на последния извод.

При анализа на този пример в **§4** посредством техниката на доверителните интервали, получихме достатъчно убедителен аргумент, че въпросният ефект наистина е статистически значим. Сега предстои да направим същия извод в контекста на теорията на проверка на статистически хипотези. Първата стъпка към това се състои във формулирането на статистическата хипотеза H_0, която се нарича още ***нулева хипотеза (null hypothesis)***, че средното на нормалната популация, за която се отнася нашата извадка, има стойност $\mu_0 = 100$, което се записва така

$$H_0 : \mu = \mu_0 .$$

Нулевите хипотези се наричат още ***хипотези за нулев резултат (нулев ефект)***, който термин достатъчно красноречиво говори за своето съдържание. В общия случай една хипотеза за нулев резултат гласи, че наблюдаваният резултат (наблюдаваният ефект) има случаен характер. Тук наблюдаваният резултат е различието между емпиричното средно и нормата на теста. Ако разполагаме с достатъчно статистически аргументи за отхвърлянето на нулевата хипотеза H_0, то можем да направим статистически обосновано заключение за наличието на съществен резултат, което в повечето случаи представлява главната цел на дадено практическо или научно изследване.

Отхвърлянето на нулевата хипотеза означава, че на практика приемаме някаква ***алтернативна хипотеза*** H_{alt}, която за нашия пример избираме да има максимално общия вид (***двустранна – ненасочена алтернатива***)

(5.1) $\quad H_{alt} : \mu \neq \mu_0 .$

Отхвърлянето (***rejecting***) на нулевата хипотеза H_0, което означава фактически приемането на алтернативата H_{alt}, или нейното ***приемане (accepting)*** представляват двете възможни решения от страна на изследователя, предприети въз основа на статистически съображения. При всяко от тези решения съществува известна вероятност за грешка. Отхвърлянето на нулевата хипотеза H_0, когато тя всъщност е валидна представлява грешка, която се нарича ***грешка от първи род***, а приемането на нулевата хипотеза H_0, когато тя всъщност не е валидна представлява грешка, която се нарича ***грешка от втори род***. И двата вида грешки са нежелани следствия от решението на изследователя. Тук обаче следва да се има предвид, че поради важни причини от предметно и техническо естество, грешката от първи род представлява в много по-висока степен нежелано последствие, като основните причини за такова състояние на нещата са две. Първата причина вече я споменахме – тя се състои в това, че в болшинството от случаите целта на изследователят е да докаже съществуването на някакъв ефект, което технически се състои в решението за отхвърляне на нулевата хипотеза H_0, а при такова действие ни заплашва единствено грешка от първи род. Втората и на практика по-важна причина се състои в това, че самата теория на проверка на статистически хипотези е технически пригодена да контролира по елементарен начин вероятността за грешка от първи род, докато контрола на грешката от втори род представлява сложна и трудно обозрима математическа цел.

Сега ще опишем рецептата за отхвърляне или приемане на разискваната нулева хипотеза $H_0 : \mu = \mu_0 = 100$. Да изберем едно достатъчно малко ***ниво на значимост*** α, например стойността по подразбиране $\alpha = 0.05$, което α по смисъла на теорията за проверка на статистически хипотези позволява да се интерпретира като ***вероятност за грешка от първи род***. При валидна нулева хипотеза H_0, величината

$$t_{emp}(n-1) = \sqrt{n}\,\frac{\overline{x} - \mu_0}{s}$$

се подчинява на t-разпределение на Стюдънт с $n-1$ степени на свобода. Тук $t_{emp}(n-1)$ се нарича ***проверяваща статистика*** за нулевата хипотеза H_0. Математически се обосновава, че отхвърлянето на H_0 в този случай става когато за стойността на $t_{emp}(n-1)$ е изпълнено неравенството

(5.2) $$\left|t_{emp}(n-1)\right| \geq t_{1-\frac{\alpha}{2};n-1},$$

където $t_{1-\frac{\alpha}{2};n-1}$ е критичното значение на съответното t-разпределение. При този избор на $\alpha = 0.05$ имаме

$$t_{1-\frac{\alpha}{2};n-1} = -t_{\frac{\alpha}{2};n-1} = t_{0.975;63} = -t_{0.025;63} = 1.998,$$

а за $t_{emp}(63)$ имаме

$$t_{emp}(63) = \sqrt{64}\,\frac{106.547 - 100}{9.976} = 5.250,$$

които резултати в съответствие с (5.2) водят до отхвърляне на нулевата хипотеза H_0. Областта от стойности W за $t_{emp}(n-1)$, определена от (5.2), при които се отхвърля нулевата хипотеза H_0 се нарича ***критична област*** за H_0, която в този случай представлява обединение на два интервала

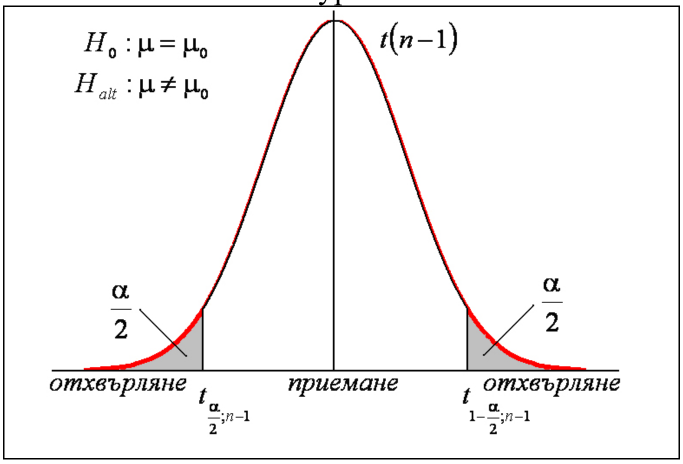
$$W = \left(-\infty, t_{\frac{\alpha}{2};n-1}\right] \cup \left[t_{1-\frac{\alpha}{2};n-1}, \infty\right).$$

Последното показва, че ***областта на приемане*** на H_0 представлява интервалът

$$\left(t_{\frac{\alpha}{2};n-1}, t_{1-\frac{\alpha}{2};n-1}\right).$$

Видът (5.2) на критичната област показва, че вероятността за грешка от първи род не може да надвишава отнапред фиксираното $\alpha = 0.05$ и именно в това се състои споменатият по-горе контрол на грешката от първи род. Намаляването на α обаче води до увеличаване възможностите за грешка от втори род. Областите на отхвърляне и приемане на H_0 са илюстрирани на следващата фигура.

Фигура 5.1.



В рамките на горния подход отхвърлянето на нулевата хипотеза H_0 при някакво достатъчно малко ниво на значимост α представлява решение с висока степен на валидност, понеже грешката, която ни заплашва в този случай има предварително фиксирана малка вероятност α. Приемането на нулевата хипотеза H_0 обаче ни заплашва с практически трудна за оценка грешка от втори род и по тази причина представлява решение с ниска степен на валидност. Тези две решения могат сполучливо да бъдат сравнение със следния съдебен казус. Отхвърлянето на нулевата хипотеза H_0 можем да сравним с произнасяне на присъда въз основа на убедителни аргументи, докато приемането на нулевата хипотеза H_0 представлява "освобождаване поради липса на достатъчно доказателства".

Отхвърлянето на нулевата хипотеза H_0 представлява на практика декларация за съществуването на ефект от закономерен характер, който ефект представлява определен интерес за изследователя или за онези, които ползват резултатите от неговото изследване.

Приемането на нулевата хипотеза H_0 най-правилно се формулира като "данните от опита не дават основания за отхвърлянето на нулевата хипотеза H_0".

2. Едностранна (насочена) алтернатива. Ако бяхме предварително сигурни, че не е реалистично да разглеждаме алтернативни възможности, при които $\mu < \mu_0 = 100$, то вместо двустранната (ненасочена) алтернатива (5.1) можем да предложим ***едностранна*** (***насочена***) алтернативна хипотеза

$$H_{alt} : \mu > \mu_0 .$$

В такъв случай критичната област за нулевата хипотеза H_0 при ниво на значимост α представлява интервалът

$$t_{emp}(n-1) \geq t_{1-\alpha;n-1} ,$$

който при $\alpha = 0.05$ за нашия пример има вида

$$t_{emp}(63) \geq 1.670 = t_{0.95;63} .$$

Разглеждането на едностранна алтернатива повишава съществено възможността за отхвърлянето на нулевата хипотеза H_0 в сравнение с двустранната алтернатива. Например, ако бяхме получили $t_{emp}(63)=1.683$, то при едностранна алтернатива трябваше да отхвърлим H_0, докато при двустранна алтернатива няма достатъчно основания за отхвърлянето. За разглеждания пример извадката беше формирана между учащи се от специализирано учебно заведение. Ако това заведение беше от елитарен тип, то притежаваме достатъчно основание да предположим, че средното за такава популация не може да бъде съществено по-ниско от нормата за средно и да предложим едностранна алтернатива $H_{alt}:\mu>\mu_0$. Ако обаче заведението беше ориентирано за работа с проблематични обучаеми, то притежаваме вече основание за предлагане на другата едностранна алтернатива

$$H_{alt}:\mu<\mu_0,$$

при която критичната област за отхвърляне на H_0 е интервалът

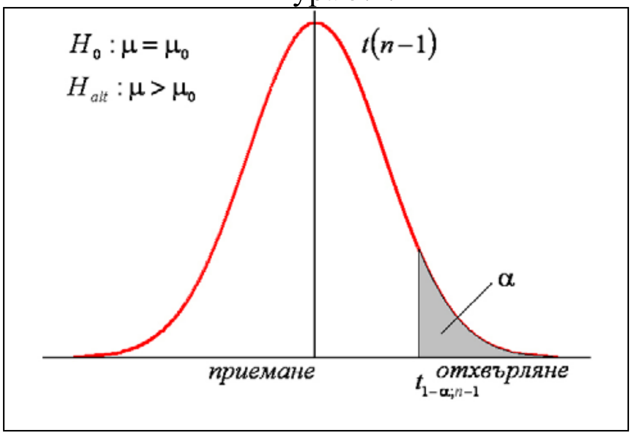
$$t_{emp}(n-1)\le t_{\alpha;n-1},$$

който при $\alpha=0.05$ за разглеждания пример има вида

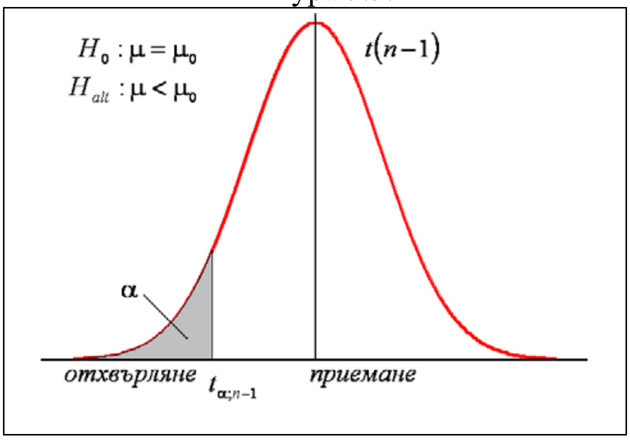
$$t_{emp}(63)\le -1.670=t_{0.05;63}.$$

Областите за отхвърляне и приемане на нулевата хипотеза H_0 при различните едностранни алтернативи са показани на фигури 5.2-3.

Фигура 5.2.



Фигура 5.3.

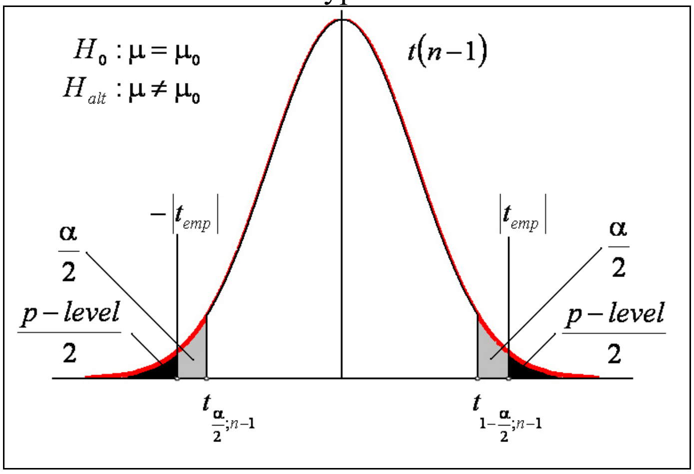


3. Оценено ниво на значимост. Изложеното дотук в настоящата лекция имаше за цел главно въвеждането в базовите идеи и терминология. В програмните среди за статистическа обработка решението за отхвърляне или приемане на H_0 става въз основа на една възможност за точно пресмятане на ***оцененото ниво на значимост***, което се означава с $p-level$, $p-value$ или ***само с*** p.

Идеята за използване на оцененото ниво на значимост p се вижда добре от фигура 5.4, която илюстрира случая на двустранна алтернатива. В този случай оцененото ниво на значимост p представлява вероятността, зададена от лицето под кривата на плътност за $t(n-1)$-разпределението, което се намира вдясно от $|t_{emp}(n-1)|$ и вляво от $-|t_{emp}(n-1)|$. Нека в резултат от наблюденията сме взели решение за отхвърлянето на нулевата хипотеза H_0, както при обсъждания по-горе пример. Това означава, че е налице $|t_{emp}(n-1)|\ge t_{1-\frac{\alpha}{2};n-1}$, което гарантира вероятност за грешка от първи род по-малка от зададеното ниво на значимост α. В тази ситуация стойността на p

сигурно няма да надвишава стойността на α, като в ***типичния случай*** на отхвърляне на H_0 стойността на p се явява ***съществено по малка*** от тази на α.

Фигура 5.4.



Ако бяхме "рискували" отгатвайки по някакъв начин предварително $\alpha = p$ (знаейки ориентировъчно какво ще се случи), който риск всъщност се състои в увеличаване възможността за грешка от втори род, то отново щяхме да отхвърлим нулевата хипотеза H_0, само че този път при по-малка вероятност за грешка (от първи род), която вероятност се явява точно стойността на p. Тук "рискът" е оправдан, понеже при отхвърлянето на нулевата хипотеза H_0 ни заплашва единствено грешка от първи род.

Описаният подход предлага една ***изключително елементарна*** в техническо отношение препоръка за отхвърляне или приемане на нулевата хипотеза H_0, понеже фактически известната стойност на p дава прецизна оценка на вероятността за грешка от първи род. Ако предварително сме задали ниво на значимост α, то отхвърлянето на H_0 става при условие, че $p \leq \alpha$. При работа с оцененото ниво на значимост обаче не се налага предварително да изискваме стойност на α. Единственото което остава да преценим е дали получената стойност на p ни удовлетворява в качеството на вероятност за грешка от първи род. ***Колкото е по-малка стойността на p, толкова е по-сигурно решението за отхвърляне на нулевата хипотеза H_0.*** В съответствие с подразбиращото се общоприето ниво на значимост $\alpha = 0.05$, отхвърлянето на H_0 става при $p \leq 0.05$. При стойности на p от 0.01 до 0.05 понякога се говори за маргинална значимост на наблюдавания ефект. При стойности $p < 0.001$ се говори за силна значимост.

Стойността на p се явява ***първично понятие*** и просто трябва да се остави да говори само за себе си, без да има особено важна необходимост да бъде сравнявано в детайли с традиционните нива на значимост. Неговото пресмятане се извършва автоматично от програмните среди за статистическа обработка. Стойността на p ***зависи*** разбира се от вида на алтернативната хипотеза H_{alt}. При избор на едностранна алтернатива неговата стойност е точно два пъти по-малка отколкото при съответната двустранна алтернатива и освен това никога на теория не може да достигне стойност нула, макар че в много практически ситуации неговата стойност се получава пренебрежимо малка. Стойността на p обикновено се закръглява до третия знак след десетичната точка, при което $p = 0.000$ означава $p < 0.001$.

За обсъждания пример при двустранна алтернатива имаме $p = 0.000002$, а при едностранната алтернатива имаме $p = 0.000001$. Тези пресмятания показват, че решението за отхвърляне на H_0 въз основа на данните от опита съдържа вероятност за грешка от порядъка на 0.000002, което е много-по малко от нивото на предварителния контрол по уговорка $\alpha = 0.05$.

Резултатът от статистическата обработка на данните от IQ-теста в разглеждания пример може да се опише с едно единствено изречение, в което е спомената цялата основна статистическа информация за опита и неговия резултат.

"Средният резултат 106.547 от провеждането на IQ-теста е съществено по-висок от нормата 100, [$t(63) = 5.250$; $p < 0.001$]."

От това изречение се подразбира, че нулевата хипотеза се проверява срещу двустранна алтернатива, понеже не е указано изрично нещо друго (от записа може да се пресметне даже и емпиричното стандартно отклонение). Подобен опростен израз е често срещан в текстовете и очевидно неговото правилно разчитане изисква определена статистическа грамотност, на която всъщност е посветен настоящият лекционен курс.

Програмните среди за статистическа обработка са ориентирани към описания току що подход за пресмятане оцененото ниво на значимост. От потребителя се иска преди всичко да познава добре съдържанието на специфичните нулеви хипотези, както и условията за приложимост на конкретния статистически метод.

От техническа гледна точка, отхвърлянето на нулевата хипотеза $H_0 : \mu = \mu_0$ срещу двустранната алтернатива $H_{alt} : \mu \neq \mu_0$ при ниво на значимост α е еквивалентно на това, хипотетичното средно μ_0 да лежи извън $(1-\alpha)100\%$ доверителен интервал за средното, респективно нейното приемане е еквивалентно на това, μ_0 да лежи вътре във въпросния интервал.

4. Хипотези за дисперсията. Дотук успоредно с въвеждането в апарата на проверка на статистически хипотези, изложихме подробно и различните видове хипотези, свързани със средното на една популация. Сега ще се запознаем с хипотезите за дисперсията. Нулевата хипотеза е

$$H_0 : \sigma^2 = \sigma_0^2,$$

което означава предположение за конкретна стойност σ_0^2 за дисперсията на популацията, към която принадлежи независимата нормална извадка с обем n. Двустранната алтернатива има вида

$$H_{alt} : \sigma^2 \neq \sigma_0^2.$$

Проверяващата статистика

$$\chi^2_{emp}(n-1) = \frac{(n-1)s^2}{\sigma_0^2}$$

на H_0 срещу H_{alt}, при валидна H_0, има χ^2-разпределение с $n-1$ степени на свобода. Областта за отхвърляне на H_0 (критичната област) при ниво на значимост α се състои от двата интервала

$$\left(0, \chi^2_{\frac{\alpha}{2};n-1}\right] \cup \left[\chi^2_{1-\frac{\alpha}{2};n-1}, \infty\right),$$

т.е. H_0 се отхвърля ако $\chi^2_{emp}(n-1) \leq \chi^2_{\frac{\alpha}{2};n-1}$ или $\chi^2_{emp}(n-1) \geq \chi^2_{1-\frac{\alpha}{2};n-1}$. Областта за приемане на H_0 представлява интервалът

$\left(\chi^2_{\frac{\alpha}{2};n-1}, \chi^2_{1-\frac{\alpha}{2};n-1}\right)$.

При едностранна алтернатива $H_{alt}:\sigma^2 > \sigma_0^2$, областта за отхвърляне на H_0 представлява интервала $\left[\chi^2_{1-\alpha;n-1}, \infty\right)$, т.е. H_0 се отхвърля ако $\chi^2_{emp}(n-1) \geq \chi^2_{1-\alpha;n-1}$, а областта за приемане на H_0 представлява съответно интервала $\left(0, \chi^2_{1-\alpha;n-1}\right)$.

При едностранна алтернатива $H_{alt}:\sigma^2 < \sigma_0^2$, областта за отхвърляне на H_0 представлява интервала $\left(0, \chi^2_{\alpha;n-1}\right]$, т.е. H_0 се отхвърля ако $\chi^2_{emp}(n-1) \leq \chi^2_{\alpha;n-1}$, а областта за приемане на H_0 представлява съответно интервала $\left(\chi^2_{\alpha;n-1}, \infty\right)$.

Да предположим например, че освен с норма за средно $\mu_0 = 100$, в документацията на обсъжданият IQ-тест има и норма за стандартно отклонение $\sigma_0 = 12$. Наблюдаваното стандартно отклонение $s = 9.976$ е видимо по-малко от нормата. За да установим дали наблюдаваният ефект е статистически значим при ниво на значимост $\alpha = 0.05$, ще проверим нулевата хипотеза $H_0:\sigma = \sigma_0 = 12$ срещу двустранната алтернатива $H_{alt}:\sigma \neq \sigma_0$. В този случай

$$\chi^2_{emp}(63) = \frac{(64-1)(9.976)^2}{12^2} = 43.540 .$$

За съответните критични стойности на $\chi^2(63)$ разпределението имаме $\chi^2_{0.025;63} = 42.950$ и $\chi^2_{0.975;63} = 86.830$. В този случай вземаме решение да приемем H_0, понеже

$$\chi^2_{0.975;63} = 86.830 > 43.540 = \chi^2_{emp}(63) = 43.540 > 42.950 = \chi^2_{0.025;63} .$$

Стойността $\chi^2_{emp} = 43.540$ определя оценено ниво на значимост при двустранна алтернатива $p = 0.05846$. Анализирайки внимателно ситуацията на базата на точната стойност на p, можем да стигнем до решение за отхвърлянето на нулевата хипотеза H_0, понеже при такова решение ни заплашва грешка с вероятност по-малка от 6%. Такова решение очевидно е по-доброто от двете. Ако отначало бяхме избрали едностранна алтернатива $H_{alt}:\sigma^2 < \sigma_0^2$, то щяхме да получим $p = 0.029230 < 0.05 = \alpha$, което се явява напълно достатъчно основание да отхвърлим H_0, следвайки класическия подход. Изборът на едностранна алтернатива може да бъде обоснован от факта, че наблюдаваната група обучаеми е от специализирано заведение, което предполага по-голяма хомогенност на индивидите и респективно по-малка дисперсия на постиженията в сравнение с нормата $\sigma_0 = 12$. Предпочитанието на едностранна алтернатива пред двустранната трябва винаги да бъде добре обосновано.

Резултатите от направения анализ може да се запишат посредством изречението:

"Стандартното отклонение 9.976 в наблюдаваната група от провеждането на IQ-теста е съществено по-ниско от нормата 12, [$\chi^2(63) = 43.540$; $p = 0.058$]."

Последният пример насочва вниманието към важния за практиката случай, какво да се прави, когато p е по-голямо но близко до 0.05. Тук както и във всички случаи решението дали наблюдаваният ефект ще се разглежда като закономерен или случаен зависи изцяло от изследователя. Ако обаче се получи $p > 0.20$, то едва ли тази стойност би послужила като аргумент, чрез който въпросният изследовател би убедил потребителите на неговите резултати, че наблюдаваният ефект има наистина закономерен – неслучаен характер. В литературата практически не се срещат ситуации, при които ефект с $p > 0.10$ да се разглежда като неслучаен.

5. Тестове за нормалност. To be continued.