Лекция 18

§18. Доверителни интервали. Проверка на хипотези

1. Извадкови разпределения. В този раздел ще разгледаме разпределенията на някои функции от случайните величини на дадена извадка, известни като основни извадкови разпределения.

В статистиката една от най-важните помощни задачи се явява при известна плътност на разпределение $f(x)$ и дадена вероятност α, $(0<\alpha<1)$, да се определи онова x, за което $F(x)=\alpha$, известна като задача за намиране ***квантили (quantile)*** на дадено разпределение. Например нека Z е разпределена нормално стандартно, $Z\in N(0,1)$. Следващата таблица съдържа няколко от най-често използваните квантили, които ще означаваме чрез z_α, $\Pr(Z<z_\alpha)=\alpha$.

Таблица 18.1.

α	z_α
0.005	-2.576
0.01	-2.326
0.025	-1.96
0.05	-1.645
0.95	1.645
0.975	1.96
0.99	2.236
0.995	2.576

Поради симетрията около нулата на нормалното стандартно разпределение е в сила формулата

$$z_\alpha=-z_{1-\alpha}.$$

Най-често използваното критично значение за нормалното стандартно разпределение е $z_{0.975}=1.96$.

χ^2***-разпределение*** (***chi-square distribution***). Казва се, че непрекъснатата величина X има χ^2-разпределение с r степени на свобода (***chi-square distribution with*** r ***degrees of freedom***) и се пише $X\in\chi^2(r)$, когато X има плътност

$$f(x)=\frac{x^{\frac{r}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{r}{2}}\Gamma\left(\frac{r}{2}\right)},\ x>0,$$ ($\Gamma(\bullet)$ – гама функция на Ойлер),

и $f(x)=0$ за $x\le 0$.

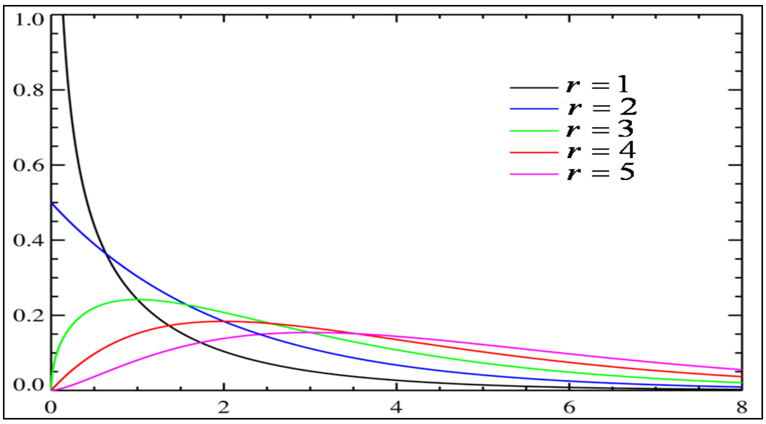


Рис. 18.1.

Една величина X има $\chi^2(r)$ разпределение, когато представлява сбор от квадратите на r независими и нормално стандартно разпределени величини. На диаграма 18.1 са приведени няколко графики на плътност за различни r.

$\chi^2(r)$ разпределението не е симетрично около средната си точка, както нормалното разпределение. Неговите квантили ще означаваме чрез $\chi^2_{\alpha;r}$, т.е ако $X \in \chi^2(r)$, то $\Pr\left(X < \chi^2_{\alpha;r}\right) = \alpha$. Да приведем например няколко такива стойности.

Таблица 18.2.

α	r	$\chi^2_{\alpha;r}$
0.05	2	$\chi^2_{0.05;2} = 0.106$
0.95	2	$\chi^2_{0.95;2} = 5.991$
0.025	2	$\chi^2_{0.025;2} = 0.051$
0.975	2	$\chi^2_{0.975;2} = 7.738$
0.005	2	$\chi^2_{0.005;2} = 0.010$
0.005	2	$\chi^2_{0.995;2} = 10.597$
0.05	10	$\chi^2_{0.05;10} = 3.940$
0.95	10	$\chi^2_{0.95;10} = 18.307$
0.025	10	$\chi^2_{0.025;10} = 3.247$
0.975	10	$\chi^2_{0.975;10} = 20.483$
0.005	10	$\chi^2_{0.005;10} = 2.156$
0.995	10	$\chi^2_{0.995;10} = 25.188$

t-разпределение на Стюдънт. Казва се, че непрекъснатата величина X има разпределение на Стюдънт с r степени на свобода (***Student's t-distribution with r degrees of freedom***) и се пише $X \in t(r)$, когато X има плътност

$$f(x) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi}\,\Gamma\left(\frac{r}{2}\right)\left(1+\frac{x^2}{r}\right)^{\frac{r+1}{2}}}.$$

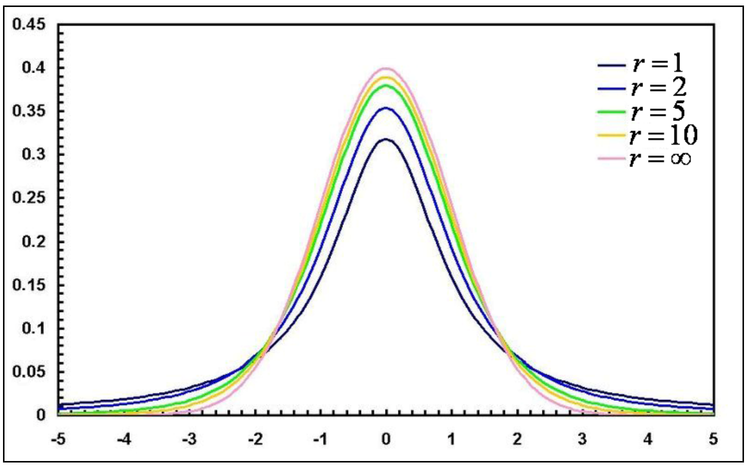


Рис. 18.2.

Степените на свобода (***degrees of freedom – df***) представляват параметър на това разпределение. Разпределението на Стюдънт е симетрично около нулата и неговият външен вид наподобява твърде много нормалното стандартно разпределение. При достатъчно голямо r, например $r > 120$, се приема, че разпределението на Стюдънт е на практика идентично с нормалното стандартно разпределение. На следващата диаграма са приведени няколко графики на плътност за различни r.

Една величина X има $t(r)$ разпределение, когато представлява частно

$$X = \frac{Z}{\sqrt{\chi^2(r)/r}},$$

където Z и $\chi^2(r)$ са две независими величини, разпределени съответно $N(0,1)$ и $\chi^2(r)$.

Квантилите на $t(r)$ разпределението ще означаваме чрез $t_{\alpha;r}$, т.е. ако $X \in t(r)$, то $\Pr(X < t_{\alpha;r}) = \alpha$. Да приведем например няколко такива стойности.

Таблица 18.3.

α	r	$t_{\alpha;r}$
0.95	10	$t_{0.95;10} = 1.812$
0.95	20	$t_{0.95;20} = 1.725$
0.95	30	$t_{0.95;30} = 1.697$
0.975	10	$t_{0.975;10} = 2.228$
0.975	20	$t_{0.975;20} = 2.086$
0.975	30	$t_{0.975;30} = 2.042$
0.99	10	$t_{0.99;10} = 2.764$
0.99	20	$t_{0.99;20} = 2.528$
0.99	30	$t_{0.99;30} = 2.457$

И тук поради симетрията на $t(r)$ разпределението около нулата е в сила формулата

$$t_{\alpha;r} = -t_{1-\alpha;r},$$

от която например въз основа на таблица 18.3 можем да пресметнем

$$t_{0.025;10} = t_{1-0.025;10} = t_{0.975;10} = -2.228.$$

F*-разпределение** (F*-distribution**). Казва се, че непрекъснатата величина X има F-разпределение на Фишер със степени на свобода m и n (***Fisher's F-distribution with degrees of freedom*** m ***and*** n) и се пише $X \in F(m,n)$, когато X има плътност

$$f(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)\sqrt{\frac{(mx)^m n^n}{(mx+n)^{m+n}}}}{x\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}, \quad x > 0,$$

и $f(x) = 0$ за $x \le 0$. Една величина X има $F(m,n)$ разпределение, когато представлява частно

$$X = \frac{\chi^2(m)/m}{\chi^2(n)/n},$$

където $\chi^2(m)$ и $\chi^2(n)$ са две независими χ^2-разпределени величини със съответните степени на свобода. F -разпределението не е симетрично относно средната си точка и неговите форми наподобяват формите на χ^2-разпределението. Квантилите на $F(m,n)$ разпределението ще означаваме чрез $F_{\alpha;m,n}$, т.е. ако $X \in F(m,n)$, то $\Pr(X < F_{\alpha;m,n}) = \alpha$.

Пресмятане на квантилите. Пресмятането на квантилите за споменатите, а също така и за други разпределения, представлява сложна изчислителна задача, която се решава автоматично от програмните среди за статистическа обработка. Преди време, когато такава възможност не е била достъпна, за решаването на тази задача са били използвани обширни таблици, които обикновено съпровождат по-старите учебници по статистика. По-нататък ще се убедим, че тази промяна има не само технически характер, но и внася съществено нов елемент в процедурите за проверка на статистически хипотези.

2. Доверителни интервали. Основната идея за доверителните интервали ще покажем върху сравнително елементарна ситуация, свързана с анализ на една метрична променлива, което обаче добре показва главните характеристики на подхода.

При детерминирането на типичния основен статистически модел за една метрична величина X, който се състои в предположението за нормалност на величината X, приведохме точкови оценки за популационното средно μ и популационната дисперсия σ^2, посредством извадковото средно $\bar{x}$ и извадковата дисперсия s^2. Техниката на доверителните интервали позволява да се направи нещо повече. Да разгледаме отначало задачата за доверителен интервал за популационното средно μ. Може да се докаже, че ако $X \in N(\mu, \sigma^2)$, то величината

(18.1) $$t_{emp} = \sqrt{n}\,\frac{\bar{X}-\mu}{s}$$

се подчинява на разпределение на Стюдънт с $n-1$ степени на свобода, $t_{emp} \in t(n-1)$. Да изберем едно достатъчно малка вероятност α, $(0 < \alpha < 1)$, обикновено $\alpha < 0.10$, която ще наречем ***ниво на значимост*** (***significance level***). Типичните стойности са $\alpha = 0.10$, $\alpha = 0.05$, $\alpha = 0.01$ или $\alpha = 0.001$. Стойността по ***подразбиране*** във всичките програмни среди за статистическа обработка е $\alpha = 0.05\ (5\%)$. В такъв случай вероятността $\gamma = 1-\alpha$ се нарича ***ниво на доверие (confidence level)***. При такъв избор на α стойността на γ се получава обикновено по-голяма от $0.9\ (90\%)$. Тогава съгласно определенията и (18.1) имаме

$$\gamma = 1-\alpha = P\left(t_{\frac{\alpha}{2};n-1} < \sqrt{n}\,\frac{\bar{X}-\mu}{s} < t_{1-\frac{\alpha}{2};n-1} \right),$$

от което след елементарно преобразуване получаваме

$$\gamma = P\left(\bar{X} - \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2};n-1} < \mu < \bar{X} + \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2};n-1} \right).$$

По този начин намерихме един ***случаен интервал***

(18.2) $$\left[\bar{X} - \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2};n-1}, \bar{X} + \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2};n-1} \right],$$

който с вероятност γ съдържа популационното средно μ. При записа на (18.2) е отчетен фактът, че $t_{\frac{\alpha}{2};n-1} = -t_{1-\frac{\alpha}{2};n-1}$. Последното трябва да се разбира така. Ако провеждаме повторни избори на извадки с обем n, то при около $\gamma.100\%$ от случаите популационното средно μ ще се съдържа в интервала (18.2), който може да се запише като

$$\bar{x} \pm \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2};n-1} .$$

Доверителният интервал (18.2) определя ***интервална оценка*** за параметъра μ, в съпоставка с понятието точкова оценка, която се определя от средното $\bar{x}$.

При голям брой степени на свобода имаме $t_{\alpha;n} \approx z_{\alpha}$ при всяко α, което дава възможност в този случай да опростим формулата (18.2) до

$$\left[\bar{x} - \frac{s}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{x} + \frac{s}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right],$$

което дава така наречения ***асимптотичен*** доверителен интервал.

С помощта на горните резултати можем да правим истински съдържателни статистически анализи.

Пример 18.1. Извършено е изследване за нивото на интелигентност посредством IQ-тест за интелигентност на група от $n = 64$ обучаеми в отговаряща на теста възраст, случайно подбрани от специализирано учебно заведение, при което са получени следните резултати.

118, 116, 109, 117, 116, 105, 104, 105, 90, 106, 110, 112, 126, 104, 116, 118, 120, 96, 111, 111, 96, 99, 122, 107, 113, 110, 101, 102, 114, 97, 118, 116, 98, 105, 94, 81, 121, 102, 113, 111, 109, 96, 97, 104, 94, 95, 127, 109, 93, 108, 95, 92, 105, 120, 110, 117, 94, 103, 105, 94, 92, 120, 109, 101

При IQ-тестовете за интелигентност е известно, че нормата за среден резултат е $\mu_0 = 100$, което естествено приемаме за средно за онази популация, към която е стандартизиран тестът. След обработка на резултатите от извадката са получени стойности $\bar{x} = 106.547$ и $s = 9.976$. В този случай имаме $\bar{x} = 106.547 > 100 = \mu_0$, т.е. средното от извадката надвишава нормата за средно, но е възможно този резултат да се дължи на известна игра на случайността. На базата на този факт можем да вземем решение, че средните постижения на учениците от специализираното учебно заведение са по-високи от нормата, което решение обаче включва съществен елемент на неопределеност. Тук техниката на доверителните интервали предлага по-добър критерий. Да изберем $\alpha = 0.05$ и да пресметнем $(1-\alpha)100\% = 95\%$ доверителен интервал по формулата (18.2). Получаваме интервала

$$\left[106.547 - \frac{9.976}{\sqrt{64}}1.998, 106.547 + \frac{9.976}{\sqrt{64}}1.998\right] = [104.055, 109.039], \ (t_{0.975;63} = 1.998)$$

което показва, че нормата (популационното средно) $\mu_0 = 100$ не се съдържа в този интервал, а по-точно лежи вляво от него. Последният факт вече представлява убедително доказателство, че средните постижения в тази група надвишават съществено нормата за средно. При $\alpha = 0.01$ имаме $t_{0.975;63} = 2.656$, откъдето получаваме следния $(1-\alpha)100\% = 99\%$ доверителен интервал

$$\left[106.547 - \frac{9.976}{\sqrt{64}}2.656, 106.547 + \frac{9.976}{\sqrt{64}}2.656\right] = [103.235, 109.859],$$

което потвърждава направеният извод с по-висока сигурност.

По-нататък ще дадем друго решение на разискваната проблем чрез техниката на проверка на статистически хипотези, което решение в основните си характеристики е идентично с приведеното.

Доверителни интервали за дисперсията се пресмятат въз основа на факта, че ако $X \in N\left(\mu, \sigma^2\right)$, то величината

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

се подчинява на χ^2-разпределение с $n-1$ степени на свобода, $\chi^2_{emp} \in \chi^2(n-1)$. Следвайки специфични статистически съображение, последният факт позволява да пресметнем $(1-\alpha)100\%$ доверителен интервал по формулата

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

За примера 18.1 с IQ-теста, при $\alpha = 0.05$ имаме $\chi^2_{0.025;63} = 42.950$ и $\chi^2_{0.975;63} = 86.830$, което дава следният 95% доверителен интервал

$$\left[\frac{(64-1)(9.976)^2}{86.830}, \frac{(64-1)(9.976)^2}{42.950}\right] = [72.207, 145.979],$$

за популационната дисперсия.

По-нататък в настоящия лекционен курс ще приведем начини за пресмятане на доверителни интервали в по-сложни ситуации.

Защо точно $\alpha = 0.05\ (5\%)$***?*** В последния пример избрахме да пресметнем 95% доверителен интервал, въз основа на което направихме заключението за съществено различие между средното за постиженията от групата и нормата за средно. Ако бяхме избрали по-голяма стойност за α, то доверителният интервал щеше да се получи по-къс и следователно щяха да се увеличат шансовете да направим същия извод, но от друга страна изводът щеше да се получи по-малко достоверен. В този случай централният въпрос е къде е компромисната бариерна стойност, която разделя между естественото желание на даден експериментатор да потвърди съществуването на някакъв ефект и достоверността на извода, че такъв ефект съществува. ***Общоприетата*** такава бариерна стойност е $\alpha = 0.05$, което няма значение за математиката и представлява въпрос на договореност в средите на изследователите в областта на различните науки. В следващия раздел, посветен на проверка на статистически хипотези, стойността на нивото на значимост α ще придобие по-ясни очертания, запазвайки първоначалния си смисъл.

3. Проверка на статистически хипотези. В този раздел ще изложим накратко основните идеи на теорията на ***проверка на статистически хипотези***, пренебрегвайки част от важните за строго математическо изложение детайли, което по същество няма да попречи за овладяване основните подходи.

Целта отново ще бъде насочена към анализ 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 18.1. 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 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 ефект***, който термин достатъчно красноречиво говори за своето съдържание. В общия случай една хипотеза за нулев ефект гласи, че наблюдаваният (някакъв) ефект има случаен характер. Тук наблюдаваният ефект 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 ефект, 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***)

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

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 $\alpha = 0.05$. При валидна нулева хипотеза H_0, величината

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

следва t-разпределение на Стюдънт с $n-1$ степени на свобода. Тук $t_{emp}(n-1)$ се нарича ***проверяваща статистика*** за нулевата хипотеза H_0.

Може да се обоснове, че отхвърлянето на H_0 в този случай е целесъобразно да стане когато за стойността на $t_{emp}(n-1)$ е изпълнено неравенството

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

където $t_{1-\frac{\alpha}{2};n-1}$ е съответният квантил $t(n-1)$ разпределението. 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 $t_{sample}(63)$ получаваме стойността

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

които резултати в съответствие с (18.4) водят до отхвърляне на нулевата хипотеза H_0. Областта от стойности W за $t_{emp}(n-1)$, определена от (18.4), при които се отхвърля нулевата хипотеза H_0 се нарича ***критична област*** за H_0, която в този случай представлява обединение на два интервала

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

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

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

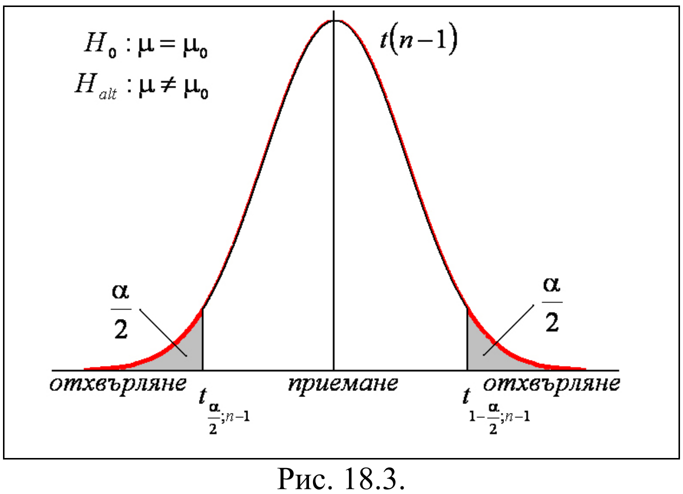


Рис. 18.3.

Видът (18.4) 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 вероятността за грешка от втори род. Областите на отхвърляне и приемане на H_0 са илюстрирани на следващата диаграма.

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. Приемането на нулевата хипотеза H_0 обаче ни заплашва с практически непредвидима грешка от втори род и по тази причина представлява решение с ниска степен на валидност. Тези две решения могат сполучливо да бъдат сравнение със 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".

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 (18.3) 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 стойност $t_{emp}(63) = 1.683$, 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

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

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 рис. 18.4-5.

Рис. 18.4

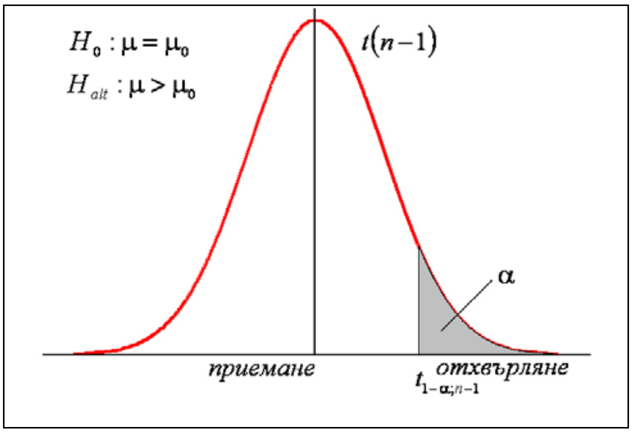
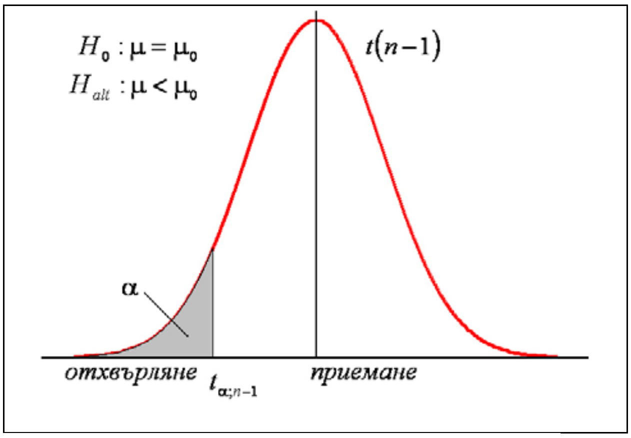


Рис. 18.5.



Оценено ниво на значимост. Изложеното дотук в предишните раздели 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.

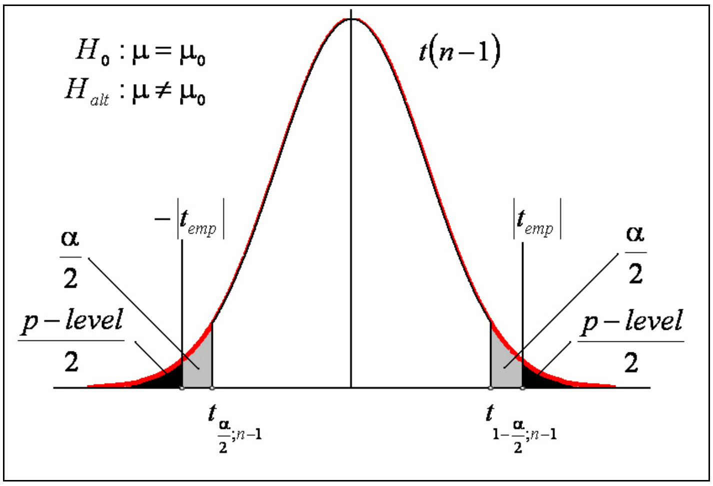


Рис. 18.6.

Идеята за използване на оцененото ниво на значимост p се вижда добре от диаграма 18.6, 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. Ако бяхме "рискували" да заложим предварително $\alpha = p$ (знаейки какво ще се случи), който риск всъщност се състои в увеличаване шанса за грешка от втори род, то отново щяхме да отхвърлим нулевата хипотеза H_0, само че този път при по-малка вероятност за грешка, 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 един ***изключително елементарен*** в техническо отношение начин за отхвърляне или приемане на нулевата хипотеза H_0, понеже фактически известната стойност на p дава ***точна оценка*** на вероятността за грешка от първи род. Сега вече е очевидно, че 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 задаваме 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 $p \leq 0.05$.

Стойността на p се явява ***първично понятие*** и просто трябва да се остави да говори само за себе си, без да има особена 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 едностранна алтернатива неговата стойност е точно два пъти по-малка от тази при 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. Стойности отбелязани с $p = 0.000$ означават, че $p < 0.001$.

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 $\alpha = 0.05$.

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, [$t(63) = 5.250$; $p = 0.000$]."

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 текст.

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.

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

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

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

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

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 $H_{alt} : \sigma > \sigma_0$, 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 $H_{alt} : \sigma < \sigma_0$, 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 пример 18.1 за 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 $\chi^2_{0.975;63} = 86.830$. Сега 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 $H_{alt}: \sigma < \sigma_0$, то щяхме да получим $p = 0.02923 < 0.05 = \alpha$, 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.