$$n > \left(\frac{2{,}575.6}{1}\right)^2 \Rightarrow n > 238{,}85.$$

Следователно, обемът на извадката трябва да е около 240.♦

Общи задачи.

1. За изследването на нормално разпределена случайна величина X е получена извадката 36 34 33 36 38 39 40 32.
а) Да се намерят точковите оценки на математическото очакване EX и дисперсията DX.
б) Да се намерят доверителните интервали с доверителна вероятност $\gamma = 0{,}95$ за EX и DX.

2. Наблюдава се работата на автомат, който произвежда детайли по даден стандарт като за определяне на точността му са взети 30 детайла и е изчислена дисперсията на извадката $s_X^2 = 2{,}9$.
а) Коя от числените характеристики на извадката оценява точността на автомата?
б) Да се намери доверителният интервал за точността на автомата с доверителна вероятност $\gamma = 0{,}95$.

3. Автомат пълни бутилки като съдържанието им по стандарт е 300 мл. За определяне на точността на дозировката са проверени 10 бутилки и са получени данните 299, 276, 283, 301, 297, 281, 300, 291, 295, 291. а) Да се намери 90%-доверителен интервал за генералното средно квадратично отклонение на количеството течност в бутилките. б) Ако отклонението е по-голямо от 10мл, то автоматът трябва да се настрои. Има ли основание за пренастройка на апарата?

4. При измерване на физична величина X са получени резултатите -2, 0, -1, 1, 3. Да се определи с надеждност $\gamma = 0{,}95$ доверителният интервал за стойността на величината X, ако е известно, че точността на измерването е $\pm 1{,}9$.

5. В резултат на 5 измервания са получени резултатите 10, 9, 12, 15, 10. Да се намерят представителната грешка с доверителна вероятност $\gamma = 0{,}99$ на оценката за генералната средна на измерваната величина. Какъв трябва да е обемът на извадката, че представтелната грешка да се намали три пъти?

6. Ако се приеме, че са дадени извадки от нормално разпределени величини, да се намерят доверителни интервали за:
а) средното време , необходимо за почистване на дома, по извадката от зад. 1, **§22**;
б) средната стойност и средното отклонение от тази стойност на броя X на дефектните изделия по данните от зад. 3, **§22**;
в) средната стойност и средното отклонение от тази стойност на броя X на пътниците в един автобус по данните от зад. 5, **§21**.

7. От автомат за пакетиране на брашно са взети за контролно измерване 15 пакета и е изчислено $\bar{x} = 956g$.
а) Ако точността на дозиране е 20 g; да се оцени с надеждност $\gamma = 0{,}9$ средното тегло на пакет брашно;
б) Ако извадъчната дисперсия е $\tilde{s}_X = 20g$ да се оцени точността на автомата.

8. По проект дължината X на детайл трябва да бъде 50,0 см. За контрол на качеството е получена извадката 4,99 4,91 4,95 4,97 5,00 5,01 4,83 4,91 4,83 4,76. Да се оцени дължината на произвежданите детайли по дадената извадка. Какъв трябва да е обемът на извадката, че тази дължина да е оценена с грешка, по-малка или равна на 0,5 см.

ПРОВЕРКА НА ХИПОТЕЗИ

§29. Статистически хипотези. Нулева и конкурираща хипотеза. Грешка от първи и грешка от втори род. Статистически критерий.

Ще се спрем на друга основна задача на математическата статистика – проверка на предположение (хипотеза) относно разпределението или параметър на разпределението на случайна величина X. Задачи от такъв характер са често срещани в различни области на науката, например, при обработка на научни резултати, прогнозиране на икономическото и общественото развитие и др. Решаването им преминава през следните етапи:

1) събиране на данни;
2) изказване на някаква хипотеза;
3) приемане или отхвърляне на хипотезата по утвърдени правила.

Нулева и конкурираща хипотеза.

Нека се изучава случайната величина X (дискретна или непрекъсната), разпределението на която е неизвестно, или е неизвестен някой от параметрите на разпределението й.

Всяко предположение относно закона на разпределение или параметрите на разпределението на случайната величина X се нарича статистическа хипотеза.

Статистическите хипотези се разделят на:

Прости - ако еднозначно определят разпределението на величината X. В противен случай хипотезите се наричат сложни.

Например, нека X има показателно разпределение. Хипотезата $\lambda = 3$ е проста, тъй като тогава плътността на разпределение е $p_X(x) = 3e^{-3x},\ x > 0$. Хипотезата $0 < \lambda < 3$ е сложна хипотеза, тъй като законът на величината не е еднозначно определен.

Нека с H_0 е означена хипотезата, която ни интересува и която трябва да се провери (*нулева хипотеза*). Заедно с нея се разглежда и една от алтернативните й хипотези H_1, която се нарича *конкурираща*.

Конкуриращата хипотеза H_1 определя каква хипотеза би трябвало да приемем, ако отхвърлим хипотезата H_0.

Например, ако θ е неизвестен параметър на разпределението на величината X, то конкурираща хипотеза на хипотезата $H_0 : \{\theta = \theta_0\}$ може да е една от хипотезите

$$H_1^{(1)} : \{\theta \neq \theta_0\},\ H_1^{(2)} : \{\theta > \theta_0\} \text{ или } H_1^{(3)} : \{\theta < \theta_0\}.$$

Ще разглеждаме ***само прости нулеви хипотези.***

Грешки от първи и втори род . Ниво на значимост.

Грешките, които може да направим са:

- *да отхвърлим правилна хипотеза H_0 - <u>грешка от първи род.</u>*

Вероятността да се допусне грешка от първи род се нарича <u>ниво на значимост</u> и се означава с α;

- *да приемем неправилна хипотеза H_0 - <u>грешка от втори род</u>.*

Вероятността да се допусне грешка от втори род се означава с β.

Очевидно, трябва да се предполага, че α и β са малки числа.

Забележка 29.1. Дефинициите и смисъла на грешките от първи и втори род ще обясним със съпоставяне на решението за приемане на нулевата хипотеза с решението на съд за виновността на обвиняемия. Хипотезата H_0 е невинност на обвиняемия като обвинението представя доказателства за отхвърлянето й. Ако доказателствата са убедителни (значими), то хипотезата за невинност се отхвърля. Един от най-важните принципи, от които се ръководи съдът при взимане на решение, е да не бъде осъден невинен човек, (отхвърляне на вярна хипотеза H_0), т.е. да не се допусне грешка от първи род.

На този принцип са организирани и изложените по-долу тестове за проверка на хипотези – заедно с конкуриращата хипотеза се определя и каква е допустимата вероятност α за отхвърляне на вярна хипотеза H_0. Стойността на нивото на значимост α се избира от порядъка на 0,1, 0,05, 0,01 и т.н. По този начин се гарантира, че вероятността да се направи грешка е минимална.

Тест за приемане или отхвърляне на нулевата хипотеза. Нека са определени хипотезите H_0 и H_1 и нивото на значимост α като за проверка на хипотезата H_0 е взета извадка $(X_1,\ldots,X_n)$ с обем n. Тъй като разполагаме единствено с числени данни, то тестовете за проверка са организирани така, че по определено правило от извадката да се получи едно число (наричано наблюдавана стойност) и в зависимост от стойността му да се вземе решение за приемане или отхвърляне на нулевата хипотеза. За целта:

1. Избира се подходяща статистика $\mathcal{K}=K(X_1,\ldots,X_n)$, която да има предварително известно разпределение, т.е. разпределение, което не зависи от наблюдаваните стойности на признака X. Тогава изчислената по извадката стойност на $\mathcal{K}$ е една от нейните възможни стойности. Параметрите на разпределението на случайната величина $\mathcal{K}$ се наричат <u>*степени на свобода*</u> и се определят в предположение, че хипотезата H_0 е вярна.

Случайната величина $\mathcal{K}$, чрез която се проверява статистическа хипотеза се нарича <u>статистически критерий</u> или просто <u>критерий</u>. Формулата $\mathcal{K} = K(X_1,\ldots,X_n)$, по която критерият се изразява чрез наблюдавните стойности, се нарича <u>статистика на критерия</u> $\mathcal{K}$. Изчислената стойност на статистиката се означава с $K_{набл.}$ и се нарича <u>наблюдавана стойност на критерия</u>.

Пример 29.1 Нека X е нормално разпределена величина и от извадка с обем n сме изчислили извадъчната средна $\overline{x}$, която се явява наблюдавана стойност на величината $\overline{X}=\frac{1}{n}(X_1+\ldots+X_n)$ - също нормално разпределена с $E\overline{X}=EX$ и $\sigma\overline{X}=\frac{\sigma X}{\sqrt{n}}$ (виж **забележка 15.2**). Тогава величината $Z=\frac{\overline{X}-EX}{\sigma X/\sqrt{n}}$ има стандартно нормално разпределение. Следователно,

- величината $Z\sim N(0,1)$, може да бъде статистически критерий,
- формулата $Z=\frac{\overline{X}-EX}{\sigma X/\sqrt{n}}$ е статистика на критерия,
- наблюдаваната стойност е $Z_{набл.}=\frac{\overline{x}-EX}{\sigma X/\sqrt{n}}$.♦

2. В зависимост от нивото на значимост α и конкуриращата хипотеза H_1 се определя

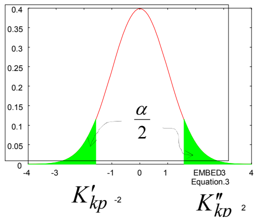
област D, за която ако $K_{набл.}\in D$ хипотезата се H_0 отхвърля. Тази област се нарича <u>критична област</u>.

Определянето на критичната област се основава на принципа, че малко вероятните събития са практически невъзможни. Тогава, ако хипотезата H_0 е вярна, то наблюдаваната стойност $K_{набл.}$ практически не попада в областите на малка вероятност на критерия $\mathcal{K}$. Следователно, имаме основание да обявим за критична такава област D, в която величината $\mathcal{K}$ приема стойности с малка вероятност, при това вероятността $P(\mathcal{K}\in D)$ да бъде равна на α. Очевидно, $K_{набл.}$ е една от възможните стойности на $\mathcal{K}$. В случай, че $K_{набл.}$ попадне в областта D при вярна хипотеза H_0, то ще направим грешка от първи род. Тъй като $P(K_{набл}\in D)=\alpha$, то вероятността за грешка от първи род е равна на α.

Областта D, определена от стойностите на $K_{набл.}$, за които имаме основание да отхвърлим нулевата хипотеза, се нарича <u>критична област</u>.

Ако отхвърлим хипотезата H_0 трябва да приемем хипотезата H_1. Затова освен от нивото на значимост, критичната област зависи и от конкуриращата хипотеза. Критичната област може да бъде :

<u>*Двустранна*</u> – ако $D=(-\infty,\ K'_{кр})\cup(K''_{кр},\infty))$.

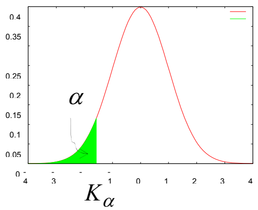


фиг.29.1.

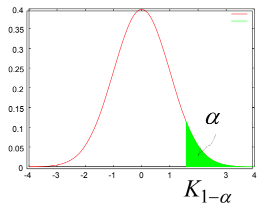
На фиг 29.1. е начертана графиката на плътността на критерия (при предположение, че хипотезата е вярна) като $K'_{кр.}=K_{\frac{\alpha}{2}}$ и $K''_{кр}=K_{1-\frac{\alpha}{2}}$ са квантили от ред $\frac{\alpha}{2}$ и $1-\frac{\alpha}{2}$. Лицата на двете защриховани области са равни на $\frac{\alpha}{2}$, а сумата им е равна на нивото на значимост α.

<u>*Едностранна*</u> –

- ако $D=(-\infty,\ K_{кр})$ - лявостранна (фиг 29.2а)
- ако $D=(K_{кр},\ \infty)$ - дясностранна (фиг 29.2б).



Фиг 29.2а.



Фиг 29.2б.

Тук критичните стойности са съответно квантилите K_α и $K_{1-\alpha}$ от ред α и $1-\alpha$, а лицата на защрихованите области са равни на α.

3. Взима се решение за приемане или отхвърляне на хипотезата.

- Ако $K_{набл.}\notin D$, то нулевата хипотеза H_0 се приема.
- Ако $K_{набл.}\in D$, то нулевата хипотеза H_0 се отхвърля.

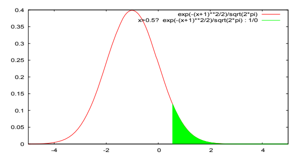
Връзка между вероятностите за грешка от първи и втори род. Мощност на критерия.

Нека са определени хипотезите H_0 и H_1. Има две възможности да вземем правилно решение – когато приемем вярна хипотеза H_0 и когато отхвърлим грешна хипотеза H_0, а всички случаи са дадени в таблицата

Решение / *Реалност*	H_0 *се приема*	H_0 *се отхвърля*
H_0 *е вярна*	Правилно решение Вероятност: $1-\alpha$	Грешка от I –ви род Вероятност: α (ниво на значимост)
H_0 *е грешна*	Грешка от II –ри род Вероятност: β	Правилно решение Вероятност: $1-\beta$ (мощност на критерия)

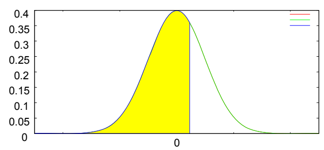
Вероятността да отхвърлим грешна хипотеза се нарича <u>мощност на критерия.</u>

Вероятностите α и β за грешки от първи и втори род са зависими една от друга. За да се види тази зависимост, да разгледаме графиките на плътностите на разпределението на критерия $\mathcal{K}$ съответно при условие, че H_0 е вярна (фиг. 29.3а) и при условие, че H_1 е вярна (29.3б). Защрихованите области определят областта на грешно решение.



Графика на критерия, при условие, че хипотезата H_0 е вярна

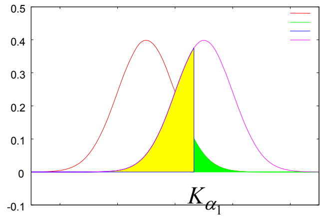
Фиг. 29.3а



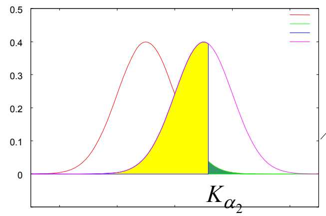
Графика на критерия при условие че хипотезата H_1 е вярна

фиг.29.3б

На фиг. 29.4. относно обща координатна система са начертани областите на грешни решения при нива на значимост $\alpha=0{,}1$ и $\alpha=0{,}05$. Забелязваме, че с намаляване на вероятността за грешка от първи род от 0,1 на 0,05 (критичната точка се мести надясно), се увеличава вероятността за грешка от втори род.



Области на грешки от първи и втори род при ниво на значимост $\alpha_1 = 0{,}1$



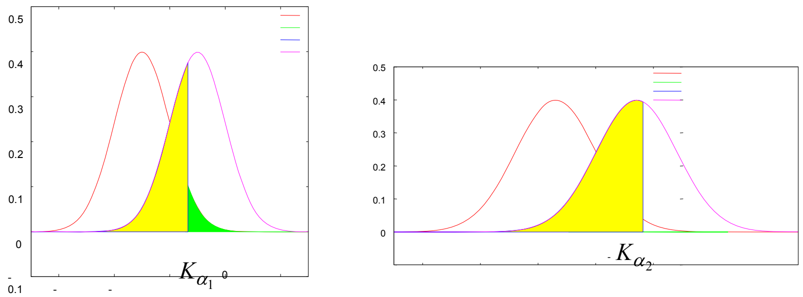
Области на грешки от първи и втори род при ниво на значимост $\alpha_2 = 0{,}05$

Фиг.29.4.

На фиг. 29.5. са съпоставени графиките на плътностите на критерия и областите на грешки от първи и втори род при различни стойности на обема на извадката.

– лявата – при по-голям обем на извадката,
– дясната – при по-малък.

Както се вижда, при нарастване на обема на извадката областите на грешки от първи и втори род се стесняват.



Фиг.29.5.

Следователно,

- *При намаляваме на вероятността α на грешка от първи род се увеличава вероятността β за грешка от втори род и обратно.*
- *При увеличаване на обема на извадката вероятността за грешки от първи и втори ред намалява.*

Накрая, ще обобщим резултатите от тази глава:

- *За всяка хипотеза H_0 е определен критерий $\mathcal{K}$, и подходяща статистика, чрез която от извадката се пресмята наблюдаваната стойност $K_{набл.}$ на критерия.*
- *Параметрите на критерия $\mathcal{K}$ се наричат степени на свобода и в повечето случаи зависят от обема на извадката.*
- *Параметрите на статистиката на критерия се определят в предположение, че хипотезата H_0 е вярна.*
- *Критичната област (областта на отхвърляне на H_0) зависи от конкуриращата хипотеза H_1 и от нивото на значимост α.*
- ***Проверката на хипотезта H_0 при конкурираща хипотеза H_1 и ниво на значимост α се извършва по следния начин:***

1) От извадката се пресмята $K_{набл.}$.
2) В съответствие с H_1, нивото на значимост α и броя на степените на свобода се определя критичната област D.
3) Взима се решение: ако $K_{набл.} \notin D$ - приема се H_0,
ако $K_{набл.} \in D$ - отхвърля се H_0.

§30. Непараметрични тестове - χ^2 - критерий на Пирсън за вида на разпределението.

Хипотезите и тестовете за приемането им може да разделим на две групи :

<u>Параметрични</u> - хипотези, които се отнасят за някои количествени параметри на разпределението на изучавания признак X,
<u>Непараметрични</u> - хипотези относно други характеристики на признака X.

Към непараметричните хипотези спадат хипотезите относно вида на разпределението, например, хипотезата $H_0 = X \sim B(N, p)$. Проверката за верността на хипотезата H_0 се извършва на базата на извадка с обем n:

$$\begin{array}{c|cccc} X & x_1 & x_2 & \dots & x_k \\ \hline m_i & m_1 & m_2 & \dots & m_k \end{array}, \qquad n = \sum_{i=1}^{k} m_i \ .$$

Преди проверката, от извадката понякога се намират оценки на необходимите параметри на разпределението (за биномното разпределение това са параметрите N и p).

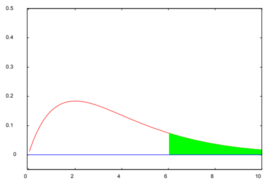
Нека броят на определените от извадката параметри на разпределението е l.

След като е определен какъв е видът на разпределението (например, биномно разпределение) и какви са параметрите на това разпределение (параметрите N и p), се пристъпва към сравняване на резултатите от наблюденията с теоретичните резултати, изчислени съгласно издигнатата хипотеза. Разглеждат се разликите $m_i - m_i'$ между набюдаваните честоти m_i и теоретичните честоти m_i' – честотите, които би трябвало да имат вариантите, ако хипотезата H_0 е вярна. Очевидно, $\sum (m_i' - m_i)^2$ е оценка за различията им, но се използва статистиката

$$\chi^2_{\text{набл}} = \sum_{i=1}^{k} \frac{(m_i - m_i')^2}{m_i'},$$

която има χ^2-разпределение със степени на свобода $s = k - l - 1$.

Критичната област е дясностранна (защрихованата област на фиг. 30.1), определена от критичната стойност $\chi_\alpha^{2(kp)}(s)$ от ред α на случайната величина $\chi^2(s)$ (равна на квантила $\chi^2_{1-\alpha}(s)$ от ред $1-\alpha$).



Фиг. 30.1

Границите за приемане или отхвърляне се определят от избраното ниво на значимост. По този начин:

Проверката на хипотезата H_0 за вида на разпределението при ниво на значимост α се провежда по следния начин:

1) намират се теоретичните честоти m_i' и наблюдаваната стойност

$$\chi^2_{\text{набл}} = \sum_{i=1}^{k} \frac{(m_i - m_i')^2}{m_i'}$$

2) изчисляват се степените на свобода $s = k - l - 1$, където:

- *k е броят на различните варианти в извадката,*
- *l е броят на параметрите на разпределението, оценките на които са определени от извадката*

3) от таблицата се взима $\chi^2_{\text{кр.}} = \chi^2_{1-\alpha}(s)$, където $\chi^2_{1-\alpha}(s)$ е квантилът от ред 1-α на χ^2-разпределението със степени на свобода s.

4) Взима се решение:

Ако $\chi^2_{\text{набл.}} < \chi^2_{\text{кр.}}$, то хипотезата H_0 се приема.

Ако $\chi^2_{\text{набл.}} \geq \chi^2_{\text{кр.}}$, то H_0 се отхвърля.

Изчисляване на теоретичните честоти. Нека $p_X(x)$ и $F_X(x)$ са съответно плътността и функцията на разпределение на величината X съгласно издигната хипотеза H_0. Теоретичните честоти изчисляваме по формулата

$$m_i' = np_i' \quad (i = 1, \ldots, k),$$

където вероятностите p_i' пресмятаме в зависимост от разпределението на извадката по следния начин:

- *Ако статистическото разпределение е неинтервално, т.е. извадката е дадена във вида* $\begin{array}{c|cccc} X & x_1 & x_2 & \ldots & x_k \\ \hline m_i & m_1 & m_2 & \ldots & m_k \end{array}$, *то*

$$p_i' = P(X = x_i) = p_X(x_i) \quad (i = 1, \ldots, k).$$

- *Ако статистическото разпределение е интервално, т.е. извадката е дадена примерно във вида*

$$\begin{array}{c|cccc} [x_{i-1}, x_i) & [x_0, x_1) & [x_1, x_2) & \ldots & [x_{k-1}, x_k) \\ \hline m_i & m_1 & m_2 & \ldots & m_k \end{array},$$

то $\quad p_i' = P(x_{i-1} < X < x_i) = F_X(x_i) - F_X(x_{i-1}) \quad (i = 1, \ldots, k).$

Пример 30.1. Проверява се хипотезата H_0, че 41% от хората имат кръвна група А, 9% - кръвна група В, 46% - нулева кръвна група и 4 % - кръвна група АВ. Да се провери тази хипотеза с ниво на значимост $\alpha = 0{,}02$, ако са получени следните данни: $\begin{array}{c|cccc} \text{кръвна група} & A & B & O & AB \\ \hline m_i & 74 & 25 & 86 & 15 \end{array}$.

Решение. 1) Тъй като извадката има обем 200, то съгласно хипотезата,

$$m_1' = np_1' = 0{,}41.200 = 82, \quad m_2' = 0{,}09.200 = 18,$$

$$m_3' = 0{,}46.200 = 92, \quad m_4' = 0{,}04.200 = 8,$$

откъдето

$$\chi^2_{\text{набл}} = \sum_{i=1}^{4} \frac{(m_i - m_i')^2}{m_i'} = \frac{(74-82)^2}{82} + \frac{(25-18)^2}{18} + \frac{(86-96)^2}{96} + \frac{(15-8)^2}{8} = 10{,}02$$

2) От извадката не сме изчислявали параметри на разпределението, следователно, броят на степените на свобода е $s = 4 - 0 - 1 = 3$

3) Определяме $\chi^2_{\text{кр.}} = \chi^2_{1-\alpha}(s) = \chi^2_{1-0{,}05}(3) = \chi^2_{0{,}95}(3) = 7{,}81$.

4) Тъй като $\chi^2_{\text{набл.}} > \chi^2_{\text{кр.}}$, то хипотезата се отхвърля.♦

Забележка 30.1. Ще обърнем внимание, че в този пример беше разгледана не количествена, а качествена величина, която приема 4 различни "стойности" A, B, O и AB.

Пример 30.2. При ниво на значимост $\alpha = 0{,}1$ да се провери хипотезата, че величината X, за която е получена извадката $\begin{array}{c|cccc} (x_{i-1}, x_i) & (0,2) & (2,4) & (4,6) & (6,8) \\ \hline m_i & 6 & 5 & 8 & 5 \end{array}$, има равномерно разпределение.

Решение.

1) *Определяне на параметрите на разпределението.* В пример **26.1** получихме оценки $a^*=0{,}18$ и $b^*=7{,}82$ за интервала $[a,b]$ от стойности на величината X.

2) *Пресмятане на* $\chi^2_{набл}$. Плътността на разпределението на равномерно разпределена величина в интервала $[0{,}18;7{,}82]$ е

$p_X(x)=\dfrac{1}{7{,}82-0{,}18}=0{,}131$ за $x\in[0{,}18;7{,}82]$ и $p_X(x)=0$ при $x\notin[0{,}18;7{,}82]$,

а функцията на разпределение е

$$F_X(x)=\begin{cases}0 & x<0{,}18\\ 0{,}131x-0{,}024 & 0{,}18\le x\le 7{,}82\\ 1 & x>1\end{cases}.$$

Изчисляваме теоретичните вероятности:

$p'_1=P(0<X\le 2)=F_X(2)-F_X(0)=0{,}238$,

$p'_2=P(2<X\le 4)=F_X(4)-F_X(2)=0{,}500-0{,}238=0{,}262$,

$p'_3=P(4<X\le 6)=F_X(6)-F_X(4)=0{,}262$, $p'_4=P(6<X\le 8)=F_X(8)-F_X(4)=0{,}238$.

Умножавайки по обема $n=24$, получаваме теоретичните честоти. Пресмятанията подреждаме в таблица:

$(x_{i-1},x_i]$	m_i	$m'_i=np'_i$	$\frac{(m_i-m'_i)^2}{m'_i}$
$(0,2]$	6	5,71	0,014
$(2,4]$	5	6,28	0,261
$(4,6]$	8	6,28	0,828
$(6,8]$	5	5,71	0.089
		$\sum=$	1,192

. Следователно, $\chi^2_{набл.}=1{,}192$.

3. *Изчисляване на* $\chi^2_{кр}$. При определянето на степените на свобода имаме предвид, че интервалите на извадката са 4 и че от тази извадка са изчислени два параметъра на разпределението – границите интервала [0,18; 7,82], т.е. $s=k-l-1=4-2-1=1$, Следователно, при $\alpha=0{,}1$

$$\chi^2_{кр}=\chi^2_{1-0{,}1}(1)=\chi^2_{0{,}9}(1)=2{,}706\,.$$

4. *Решение за приемане или отхвърляне на хипотезата*. Тъй като $\chi^2_{\text{набл.}}<\chi^2_{\text{кр.}}$, то няма основание за отхвърляне на хипотезата за равномерно разпределение на величината X.

Пример 30.3. Контролира се размерът X на детайлите, изработвани на струг (виж таблицата по-долу). С ниво на значимост $\alpha=0{,}1$ да се провери хипотезата за нормално разпределение на контролирания размер.

Решение. **1)** С помощта на таблицата от извадката изчисляваме оценката $\bar{x}$ на $a=EX$ и оценката $\tilde{s}_x$ на $\sigma=\sigma X$:

(x_{i-1},x_i)	m_i	$x^*=(x_i+x_{i-1})/2$	$m_i x_i$	$m_i x_i^2$
(3, 4)	5	3,5	17,5	61,25
(4, 5)	15	4,5	67,5	303,75
(5, 6)	23	5,5	126,5	695,75
(6, 7)	19	6,5	123,5	802,75
(8, 9)	6	7,5	45	337,5
$\sum=$	68	27,5	380	2201

Получаваме: $\bar{x}=5{,}5882$, $\overline{x^2}=32{,}3676$, $\tilde{s}_x^2=1{,}1563$, $\tilde{s}_x=1{,}0753$.

2) За изчисляване на теоретичните честоти и $\chi^2_{набл}$ използваме таблицата

(x_{i-1},x_i)	m_i	$p_i=F\left(\frac{x_i-\bar{x}}{\tilde{s}_x}\right)$	$p_{i-1}=F\left(\frac{x_{i-1}-\bar{x}}{\tilde{s}_x}\right)$	$m'_i=n(p_i-p_{i-1})$	$\frac{(m_i-m'_i)^2}{m'_i}$
(3, 4)	5	0,069836	0,008042	4,201982	0,151555
(4, 5)	15	0,292175	0,069836	15,11904	0,000937
(5, 6)	23	0,649114	0,292175	24,27182	0,066642
(6, 7)	19	0,905391	0,649114	17,42689	0,142004
(8, 9)	6	0,987547	0,905391	5,586606	0,03059
				$\sum=$	0,391729

Следователно, $\chi^2_{набл}=0{,}3917$.

3) Изчисляваме $\chi^2_{кр}=\chi^2_{1-0{,}1}(5-2-1)=\chi^2_{0{,}9}(2)=4{,}61$.

4) Тъй като $\chi^2_{\text{набл.}}<\chi^2_{\text{кр.}}$, то приемаме хипотезата за нормално разпределение.

Упражнение.

1. По извадката $\begin{array}{c|ccccc} x_i & 1 & 2 & 3 & 4 & 5\\ \hline m_i & 110 & 130 & 70 & 90 & 100\end{array}$, може ли да се твърди с ниво на значимост 0,01, че дискретният признак X, които приема възможни стойности 1,2,3,4 или 5, е равномерно разпределен?

§31. Параметрични тестове - тестове за математическото очакване и дисперсията на нормално разпределена генерална съвкупност.

Параметричните тестове проверяват хипотези относно стойността на параметър θ на разпределението на изследван количествен признак X на генералната съвкупност. Както вече отбелязахме, предполагаме, че хипотезата H_0 е проста , т.е. $H_0=\{\theta=\theta_0\}$. Като конкурираща може да се приеме една от хипотезите $H_1^{(1)}:\{\theta\ne\theta_0\}$, $H_1^{(2)}:\{\theta>\theta_0\}$ или $H_1^{(3)}:\{\theta<\theta_0\}$, от която зависи вида на критичната област.

Пример 31.1. Да разгледаме нормално разпределена случайна величина X с неизвестно математическо очакване и известно средно квадратично отклонение $\sigma=1$. Нека от извадката $(X_1,\ldots,X_9)$ с обем n=9 е изчислена извадъчната средна $\bar{x}=2{,}5$. Ще проверим хипотезата $H_0=\{EX=3\}$ при ниво на значимост $\alpha=0{,}1$ и при конкурираща хипотеза:

а) $H_1=\{EX\neq 3\}$; б) $H_1=\{EX>3\}$; в) $H_1=\{EX<3\}$.

Решение. Ако допуснем, че хипотезата H_0 е вярна, то величината X и величините X_1, …, X_9 (резултатите от наблюденията на X) са разпределени по закона $N(3,1)$, т.е. те са нормално разпределени величини с математическо очакване $EX=EX_i=3$ и $\sigma X=\sigma X_i=1$ ($i=1,\ldots,9$). Тогава $\bar{x}=2{,}5$ е стойност на величината $\overline{X}=\frac{1}{9}(X_1+\ldots+X_9)$, която също е нормално разпределена.

Съгласно свойствата на математическото очакване и дисперсията

$$E\overline{X}=\frac{1}{9}(EX+\ldots+EX)=\frac{1}{9}9.3=3\,,\quad D\overline{X}=\frac{1}{9^2}(DX+\ldots+DX)=\frac{1}{81}.9.1=9.1=\frac{1}{9}\,,$$

$$\sigma\overline{X}=\sqrt{D\overline{X}}=\sqrt{\frac{1}{9}}=\frac{1}{3}\,.$$

Следователно, ако H_0 е вярна, то $\overline{X}\sim N\left(3,\frac{1}{3}\right)$ и $\bar{x}=2{,}5$ е една от нейните възможни стойности. По тази стойност ние трябва да приемем или отхвърлим H_0 при ниво на значимост α и при конкурираща хипотеза H_1.

Тъй като $\overline{X}$ е неотместена оценка на EX, то е малко вероятно наблюдаваната стойност $\bar{x}$ на $\overline{X}$ да се различава съществено от EX. Следователно, може да приемем, че хипотезата H_0 не е вярна само, ако $\bar{x}$ попадне в област с малка вероятност, т.е. в краищата на реалната права $(-\infty,\infty)$. По този начин ще бъде малко вероятно да отхвърлим вярна хипотеза H_0 и да приемем невярната хипотеза H_1. Следователно, критичната област D ще определим в зависимост от нивото на значимост α и от конкуриращата хипотеза H_1 по следния начин:

а) $H_1=\{EX\neq 3\}$, т.е. $H_1=\{3<EX \text{ или } 3>EX\}$ — критичната област съдържа възможни стойности както по-големи, така и по-малки от 3, т.е.

стойности на $\bar{x}$, за които $|\bar{x}-3|$ е голямо число. Областта ще бъде определена, ако намерим такова число δ, че вероятността за събитието $\{|\bar{x}-3|>\delta\}$ да бъде равна на нивото на значимост $\alpha=0{,}1$, т.е. P(да приемем хипотезата $H_1)=P(|\bar{x}-EX|>\delta)=0{,}1$.

Но $P(|\bar{x}-EX|>\delta)=1-P(|\bar{x}-EX|\leq\delta)\Rightarrow P(|\bar{x}-EX|\leq\delta)=1-0{,}1=0{,}9$.

Като използваме формула (13.3) за величината $\overline{X}\sim N\left(3,\frac{1}{3}\right)$, за δ получаваме $P(|X-EX|\leq\delta)=2F\left(\frac{\delta}{\sigma}\right)-1=0{,}9\Rightarrow 2F(3\delta)=1{,}9\Rightarrow F(3\delta)=0{,}95$.

От таблицата за функцията $F(x)$ определяме, че $3\delta=1{,}645\Rightarrow\delta=0{,}548$.

Тогава, областта за приемане на нулевата хипотеза е интервалът $[2{,}452;3{,}548]$, а областта на отхвърляне е двустранната област $D=(-\infty,2{,}452)\bigcup(3{,}548,\infty)$. На фиг. 29.1 са представени двете области като $K'_{кр.}=2{,}452$ и $K''_{кр.}=3{,}548$. Лицето на защрихованите области е равно на нивото на значимост $\alpha=0{,}1$. Тъй като $\bar{x}=2{,}5\notin D$, то нямаме основание да отхвърлим хипотезата H_0.

б) $H_1=\{EX>3\}$. В този случай областта на отхвърляне на H_0 се определя от равенството $P(\bar{x}-3>\delta)=0{,}1$, т.е. $P(\bar{x}>\delta+3)=0{,}1$ Но

$$P(\bar{x}>\delta+3)=1-P(\bar{x}<\delta+3)=1-F\left(\frac{\delta+3-3}{1/3}\right)=1-F(3\delta)$$

Следователно, $F(3\delta)=0{,}9\Rightarrow 3\delta=1{,}285\Rightarrow\delta=0{,}43$.

От тук определяме критичната област $D=(\delta+3,\ \infty)=(3{,}43,\ \infty)$ и областта на приемане $(-\infty,\ 3{,}43]$ (фиг. 29.2б).

Тъй като $\bar{x}=2{,}5\in(-\infty,\ 3{,}48)$, то и в този случай приемаме хипотезата H_0.

в) $H_1=\{EX<3\}$. Областта на отхвърляне на H_0 определяме от равенството

$$P(\bar{x}-3<\delta)=0{,}1\Rightarrow P(\bar{x}<\delta+3)=F\left(\frac{\delta+3-3}{1/3}\right)=F(3\delta)=0{,}1\,.$$

Тъй като $0{,}1<0{,}5$, то $3\delta<0$ и за да намерим δ, ще използваме, че

$$F(3\delta)=1-F(-3\delta)\text{, т.е. } F(-3\delta)=1-0{,}1=0{,}9\,.$$

От таблицата получаваме $-3\delta=1{,}28$. Следователно, $\delta=-0{,}43$, а критичната област е $D=(-\infty,\ \delta+3)=(-\infty,\ 2{,}57)$ (фиг. 29.2а).

Тъй като $\bar{x}=2{,}5\in(-\infty,\ 2{,}57)$, то нямаме основание да приемем хипотезата $H_0:\{EX=3\}$ при конкурираща хипотеза $H_1=\{EX<3\}$. ♦

Както се вижда от примера, *при едни и същи резултати от наблюденията една хипотеза може да бъде приета или отхвърлена в зависимост от конкуриращата хипотеза.*

В разгледания пример, за проверка на хипотезата $H_0:\{EX=3\}$ използвахме, че ако хипотезата е вярна, то $\overline{X}\sim N(3,1/3)$. Очевидно, знаем също разпределението на статистиката $Z=\dfrac{\overline{X}-3}{1/3}\sim N(0,1)$. Намерените критични стойности са квантилите съответно от ред $1-\alpha/2=0{,}95$, $1-\alpha=0{,}9$ и $\alpha=0{,}1$ на стандартното нормално разпределение. Следователно, за хипотезата $H_0=\{EX=a\}$, подходяща за статистически критерий е величината $Z\sim N(0,1)$.

Ако средно квадратичното отклонение не е известно, то от извадката пресмятаме неговата оценка – поправеното средно квадратично отклонение $\tilde{s}_x$. Тогава като статистика на критерия за проверка на хипотезата $H_0=\{EX=a\}$ се използва величината $\dfrac{\overline{X}-a}{\tilde{s}_x/\sqrt{n}}$.

Доказва се, че тази величина има разпределение $t(n-1)$ на Стюдънт (*t*-разпределение с $n-1$ степени на свобода. Така обобщаваме:

Тестове за проверка на хипотезата $H_0=\{EX=a\}$ *относно математическото очакване* EX *на нормално разпределена величина с ниво на значимост* α:

- *<u>При известно</u>* σX *: Статистика:* $Z_{набл.}=\dfrac{\overline{x}-a}{\sigma X/\sqrt{n}}\sim N(0,1)$. (31.1)

Критична област: $|Z_{набл.}-a|>Z_{1-\frac{\alpha}{2}}$, *ако* $H_1=\{EX\neq a\}$;

$Z_{набл.}>Z_{1-\alpha}$, *ако* $H_1=\{EX>a\}$;

$Z_{набл.}<Z_{\alpha}$,. *ако* $H_1=\{EX<a\}$.

- *<u>При неизвестно</u>* σX *: Статистика:* $t_{набл.}=\dfrac{\overline{x}-a}{\tilde{s}_x/\sqrt{n}}\sim t(n-1)$. (31.2)

Критична област: $|t_{набл.}-a|>t_{\frac{\alpha}{2}}$, *ако* $H_1=\{EX\neq a\}$;

$t_{набл.}>t_{1-\alpha}$, *ако* $H_1=\{EX>a\}$;

$t_{набл.}<t_{\alpha}$, *ако* $H_1=\{EX<a\}$.

По подобен начин са организирани и тестовете за проверка на хипотезите, касещи дисперсията. Тук като статистически критерий се използва χ^2-разпределението с n или n-1 степени на свобода:

Тестове за проверка на хипотезата $H_0=\{DX=c^2\}$ *относно дисперсията* DX *на нормално разпределена величина с ниво на значимост* α:

- *<u>При известно</u>* $EX=\mu$ *: Статистика:*

$$\chi^2_{набл.}=\frac{ns_0^2}{c^2}\sim\chi^2(n), \quad \text{където } s_0^2=\frac{\sum(x_i-\mu)^2 m_i}{n}.$$

Критична област: $\chi^2_{набл}<\chi^2_{\frac{\alpha}{2}}(n)$, $\chi^2_{набл}>\chi^2_{1-\frac{\alpha}{2}}(n)$, *ако* $H_1=\{DX\neq c^2\}$;

$\chi^2_{набл.}>\chi^2_{1-\alpha}(n)$, *ако* $H_1=\{DX>c^2\}$;

$\chi^2_{набл.}<\chi^2_{\alpha}(n)$, *ако* $H_1=\{DX<c^2\}$.

- *<u>При неизвестно</u>* EX *: Статистика:* $\chi^2_{набл.}=\dfrac{(n-1)\tilde{s}_x^2}{c^2}\sim\chi^2(n-1)$.

Критична област: $\chi^2_{набл}<\chi^2_{\frac{\alpha}{2}}(n-1)$, $\chi^2_{набл}>\chi^2_{1-\frac{\alpha}{2}}(n-1)$ *ако* $H_1=\{DX\neq c^2\}$;

$\chi^2_{набл.}>\chi^2_{1-\alpha}(n-1)$, *ако* $H_1=\{DX>c^2\}$;

$\chi^2_{набл.}<\chi^2_{\alpha}(n-1)$, *ако* $H_1=\{DX<c^2\}$.

Тук с $\chi^2_{\alpha}(s)$ и $t_{\alpha}(s)$ са означени квантилите от ред α на χ^2-разпределението и *t*-разпределението s степени на свобода, които се взимат от таблиците на стр. 170 и 169.

Забележка 31.1. Освен с определяне на критичната област, проверката на хипотезата $H_0=\{\theta=\theta_0\}$ може да се извърши и чрез така наречената *p-стойност (p-value)*. Основните етапи тук са:

1. От извадката се изчислява наблюдаваната стойност $K_{набл.}$.

2. В зависимост от конкуриращата хипотеза::

- *При* $H_1:\{\theta>\theta_0\}$ *изчисляваме* $p=P(K>K_{набл.})$
- *При* $H_1:\{\theta<\theta_0\}$ *изчисляваме* $p=P(K<K_{набл.})$
- *При* $H_1:\{\theta\neq\theta_0\}$ *изчисляваме* $p=P(K<K_{набл.})+P(K>K_{набл.})$

3. ***Взимане на решение***.

- *Ако* $p<\alpha$, *то* $K_{набл.}$ *се намира в критичната област и хипотезата* H_0 *се отхвърля.*
- *Ако* $p\geq\alpha$, *то нямаме основание за отхвърляне на хипотезата* H_0.

Този начин за проверка на хипотезата е удобен при възможност за пресмятане на стойностите на функцията $F_K(x)$ на разпределение на статистическия критерий или при наличието на подробни таблици за разпределението на статистическия критерий.

Пример 31.2. От голяма партида резистори са избрани 36 резистора. Средната стойност на съпротивлението X на резисторите е $9{,}5k\Omega$. При ниво на значимост $\alpha=0{,}1$ да се провери хипотезата, че извадката е взета от партида, за която съпротивлението на резисторите е $10k\Omega$, при двустранна конкурираща хипотеза, ако: а) дисперсията на X е известна и равна на $4k\Omega^2$; б) дисперсията на X е неизвестна, а изборната поправена дисперсия е равна на $2{,}25k\Omega^2$;

Решение. По извадка, за която $\bar{x}=9{,}5$, проверяваме хипотезата $H_0=\{EX=10\}$ при конкурираща хипотеза $H_1=\{EX\neq10\}$.

а) Средно квадратичното отклонение $\sigma X=\sigma=\sqrt{4}=2$ е известно. Тогава статистически критерий е $Z\sim N(0{,}1)$, а статистиката на критерия е $\dfrac{\overline{X}-a}{\sigma/\sqrt{n}}$.

1) Параметрите на статистиката се определят от обема $n=36$ и от нулевата хипотеза, според която a=10, $\sigma=2$, а наблюдаваната стойност на $\overline{X}$ е 9,5. Следователно, наблюдаваната стойност на критерия е

$$Z_{набл}=\frac{9{,}5-10}{2/\sqrt{36}}=\frac{-0{,}5}{1/3}=-1{,}5\,.$$

2) При ниво на значимост $\alpha=0{,}1$ определяме критичната стойност

$$Z\;=Z_{1-\frac{\alpha}{2}}=Z_{0{,}95}=1{,}65\,.$$

3) Тъй като $|-1{,}5|<1{,}65$, то хипотезата H_0 се приема.

б) Тук оценка на средно квадратичното отклонение е $\tilde{s}_x=\sqrt{2{,}25}=1{,}5$. Като критерий използваме t-разпределението с $s=36-1=35$ степени на свобода, $t_{набл}=\dfrac{9{,}5-10}{1{,}5/\sqrt{36}}=\dfrac{-3}{1{,}5}=-2$, критичната стойност е $t_{1-\frac{\alpha}{2}}(35)=t_{0{,}95}(35)\approx1{,}69$. Тъй като $|-2|>1{,}69$, то хипотезата H_0 се отхвърля.

Пример 31.3. Да се намери мощността на критерия и вероятността за грешка от втори род β в пример **31.2а)**, ако е вярна конкуриращата хипотеза и действителната средна стойност на съпротивлението на резисторите е равно на $9\,k\Omega$.

Решение. В пример **31.2а** получихме, че критичната област при ниво на значимост 0,1 и областта на приемане са съответно $D=\{Z<-1{,}65\bigcup Z>1{,}65\}$ и $\overline{D}=\{Z\in[-1{,}65;\ 1{,}65]\}$. Статистиката на критерия

$$Z=\frac{\overline{X}-10}{2/\sqrt{36}} \tag{31.1}$$

изчислихме при предположение, че хипотезата $H_0=\{EX=10\}$ е вярна.

От (31.1) намираме съответната област на приемане за стойностите на величината $\overline{X}$ като решаваме относно $\overline{X}$ неравенството $-1{,}65\le\dfrac{\overline{X}-10}{1/3}\le1{,}65$, от което получаваме, че $\overline{X}\in[9{,}45;\ 10{,}55]$. Наблюдаваната стойност $\bar{x}=9{,}5$ е число в този интервал и затова нямаме основание да отхвърлим хипотезата H_0.

Но в действителност е вярна конкуриращата хипотеза, по-конкретно, хипотезата $H_1=\{EX=9\}$, т.е. $\overline{X}\sim N(9;1/3)$. Тогава намерената област $\overline{X}\in[9{,}45;\ 10{,}55]$ е област на грешка от втори род.

Първо ще намерим вероятността β за грешка от втори род $\beta=P(9{,}45\le\overline{X}\le10{,}55)$ като приложим формула (13.2):

$$\beta=F\left(\frac{10{,}55-9}{0{,}3333}\right)-F\left(\frac{9{,}45-9}{0{,}3333}\right)=F(4{,}65)-F(1{,}35)=0{,}9999-0{,}9115=0{,}0884\,.$$

Следователно, мощността на критерия е $1-\beta=1-0{,}0884\approx0.91$, т.е. при вярна хипотеза $H_1=\{EX=9\}$ и обем на извадката $n=36$ мощността на критерия е 0,91.

Пример 31.4. Точността на автомат, произвеждащ детайли, се характеризира с дисперсията на дължината X на детайлите. Ако тази величина е по-голяма от $4mm^2$, то автоматът трябва да се пренастрои. Извадъчната дисперсия на 15 случайно избрани детайла е равна на $6{,}8mm^2$. Необходимо ли е да се настройва автоматът, ако приемем ниво на значимост а) 0,01; б) 0,1?

Решение. Приемаме, че дължината X на детайлите има нормално разпределение. Автоматът трябва да се настройва, ако DX>4. Тъй като нулевата хипотеза трябва да бъде проста, то избираме $H_0=\{DX=4\}$ и $H_1=\{DX>4\}$ като хипотезата H_0 ще проверим при неизвестно математическо очакване. Според правилото за проверка на нулевата хипотеза изчисляваме $\chi^2_{набл}=\dfrac{(n-1)\tilde{s}_x^2}{c^2}=\dfrac{14\left(\frac{15}{14}.6{,}8\right)}{4}=25{,}5$.

Критичната стойност и решението за настройване на автомата определяме в зависимост от нивото на значимост:

а) $\chi^2_{кр}=\chi^2_{1-\alpha}(n-1)=\chi^2_{0{,}99}(14)=29{,}14$. Тъй като $\chi^2_{набл}<\chi^2_{кр}$, то приемаме нулевата хипотеза, т.е. не е необходимо настройване на автомата.

б) $\chi^2_{кр}=\chi^2_{1-\alpha}(n-1)=\chi^2_{0{,}9}(14)=21{,}06$. Тъй като $\chi^2_{набл}>\chi^2_{кр}$, то нулевата хипотеза се отхвърля и в този случай е необходима настройка на автомата.

Пример 31.5. По извадка с обем n и при дадено ниво на значимост α да се проверят следните хипотези за нормално разпределената величина X като се използват p-стойности.

а) $H_0=\{EX=10\}$, $H_1=\{EX\neq 10\}$, $n=36$, $\alpha=0{,}1$, ако $\bar{x}=9{,}5$ и генералната дисперсия е $\sigma^2=4$; б) $H_0=\{EX=10\}$, $H_1=\{EX<10\}$, $n=36$, $\alpha=0{,}1$, ако $\bar{x}=9{,}3$ и изборната поправена дисперсия е $6{,}25k\Omega^2$ (виж пример **31.2а**);

в) $H_0=\{DX=4\}$, $H_1=\{DX>4\}$, $n=15$, $\alpha=0{,}1$, ако EX е неизвестно.

Решение. а) При използване p-стойности извършваме следното:

1) Изчисляваме $Z_{набл}=\dfrac{9{,}5-10}{2/\sqrt{36}}=\dfrac{-0{,}5}{1/3}=-1{,}5$.

2) За двустранна критична област изчисляваме

$$p=P(|Z|>|-1{,}5|)=1-P(|Z|\le 1{,}5)=1-[2F(1{,}5)-1]=2-2.0{,}9332=0{,}1336 .$$

3) Тъй като $p>\alpha$, то наблюдаваната стойност се намира в областта на приемане, т.е. нямаме основание за отхвърляне на хипотезата H_0

б) Изчисляваме: 1) $t_{набл}=\dfrac{9{,}3-10}{2{,}5/\sqrt{36}}=\dfrac{-4{,}2}{2{,}5}=-1{,}68$.

2) хипотезата $H_1=\{EX<10\}$ е лявостранна, следователно, за величината $T\sim t(n-1)$ (t-разпределение с $n-1=36-1=35$ степени на свобода трябва да изчислим вероятността $P(T<-1{,}68)$. Тъй като плътността на t-разпределението е четна функция, то

$$P(T<-1{,}68)=P(T>1{,}68)=1-P(T<1{,}68) .$$

В таблицата за квантилите на t-разпределението не са дадени данни за разпределението с 35, нито има точно стойността 1,68, но може да използваме, че големи стойности на n t-разпределението е близко до $N(0,1)$, затова може да приемем, че $P(T<1{,}68)<0{,}95$. Тогава $p=P(T<-1{,}68)=1-P(T<1{,}68)>1-0{,}95=0{,}05$.

3) Следователно, $p\approx 0{,}05<\alpha$ и наблюдаваната стойност се намира в критичната област, т.е. нулевата хипотеза трябва да се отхвърли. (Ако разполагаме с таблица от критичните стойности на t-разпределението, то $P(T<-1{,}68)=P(T>1{,}68)=0{,}050932$.)

в) 1) $\chi^2_{набл}=\dfrac{(n-1)\widetilde{s}_x}{c^2}=\dfrac{14\left(\frac{15}{14}.6{,}8\right)}{4}=25{,}5$,

2) Тъй като $H_1=\{DX>4\}$, то трябва да изчислим $p=P(\chi^2(14)>25{,}5)$, която получаваме като изчислим стойността $1-F_{\chi^2}(25{,}5;14)=0{,}02994$ (за 14 степени на свобода). Така определяме, че $p=0{,}02938>0{,}01$, т.е нямаме основание да отхвърлим нулевата хипотеза.

Забележка. Ако използваме таблицата на стр. 170, то $p=1-P(\chi^2(14)<25{,}5)$. Тъй като на реда за 14 степени числото 25,5 се намира между квантилите $\chi^2_{0{,}95}(14)=23{,}68$ и $\chi^2_{0{,}975}(14)=26{,}12$, то $0{,}95<P(\chi^2_5(14)<25{,}5)<0{,}975$ и $p>1-0{,}975\approx 0{,}025$, откъдето отново заключаваме, че няма основание за отхвърлянето на H_0.♦

Общи задачи (§§30, 31).

1. Дадена е извадката 17, 15, 5, 9, 13, 42, 8, 24, 34, 38, 29, 6. Групирайте данните в интервали (0,10), (10,20), (20,30), (30,40) и (40,50). а) да се намерят извадъчните средна и средно квадратично отклонение на извадката; б) да се начертае хистограмата на относителните честоти; в) да се провери хипотезата H_0, че величината X е разпределена равномерно в интервала [0, 50] (с ниво на значимост $\alpha=0{,}05$).

2. При изпълнение на определени условия на експлоатация е установено, че средната продължителност на безотказна работа на голяма партида уреди е 1000 със средно квадратично отклонение 100 часа. Изборната средна на случайна извадка с обем 25 се е оказала 970 часа като средно квадратичното отклонение на извадката съвпада с това на цялата партида. Може ли да твърдим, че условията за експлоатация на цялата партида са спазени, ако приемем ниво на значимост:
а) 0,10; б) 0,01?

3. Да се реши задача 2 при условие, че средно квадратичното отклонение на партидата е 120.

4. Твърди се, че изработваните в завод детайли имат диаметър 10 мм. Използвайки едностранен критерий да се провери с ниво на значимост 0,05 това твърдение, ако за извадка с обем n=16 се е оказало, че средния диаметър е равен на 10,3, считайки, че а) дисперсията е известна и е равна на 1мм^2; б) поправената дисперсия на извадката е 1,21мм^2.

5. Изпробва се нов метод за измерване на разстояния. Направени са 10 измервания на един и същ еталон и е изчислено, че поправеното средно квадратично отклонение от еталона е 10 мм. Съгласува ли се този резултат с хипотезата, че дисперсията на грешките от измерванията по този метод не е по-голяма от 50 мм^2 (приемаме ниво на значимост 0,05)

6. От генерална съвкупност е извлечена извадка с обем 150:

x_i	6	9	12	15	18	21	24	27	30	33
m_i	1	4	13	23	22	29	29	16	11	2

Да се провери с 10%-но ниво на значимост хипотезата за нормално разпределение на признака X на генералната съвкупност.

7. При изпитание на апаратура се отчита броят X на повредите. Резултатите от 60 изпитания са представени в таблицата:

x_i	0	1	2	3
m_i	42	11	4	3

. Да се провери хипотезата $X\sim Po(\lambda)$ с ниво на значимост а) α=0,001; б) α=0,01;

8. Дадена е извадката (X - брой на децата в 25 семейства) 3, 2, 0, 1, 4, 5, 3, 2, 4, 1, 0, 2, 5, 5, 2, 1, 2, 1, 0, 0, 3, 6, 2,1, 0. а) да се начертае полигонът на относителната честота. Изказано е предположение, че 20% от семействата в даден район нямат деца, 30% от семействата имат едно, 30% - две деца, а останалите 20% - повече от две деца. Да се провери това предположение на базата на дадената извадка с ниво на значимост 0,05.