Глава 14

Проверка на параметрични хипотези

14.1 Основни етапи на проверката

Проверката на статистически параметрични хипотези с прилагане на критерий за значимост може да се извърши на следните етапи:

1. Формулират се основната (H_0) и алтернативната (H_1) хипотези;

2. Избира се ниво на значимост α;

3. Избира се статистика Z на критерия за проверка на хипотезата H_0;

4. Определя се разпределението на статистиката Z, при условие, че е вярна хипотезата H_0;

5. В зависимост от формулировката на алтернативната хипотеза H_1 се определя критичната област W или областта на допустимите стойности $\overline{W}$;

6. Получава се извадката от наблюдения $x_1, x_2, \ldots, x_n$ и се пресмята наблюдаваната стойност на критерия $Z_0 = Z(x_1, x_2, \ldots, x_n)$;

7. Взема се статистическо решение:

   ако $Z_0 \in W$, то хипотезата H_0 се отхвърля, понеже не се съгласува с резултатите от наблюдението;

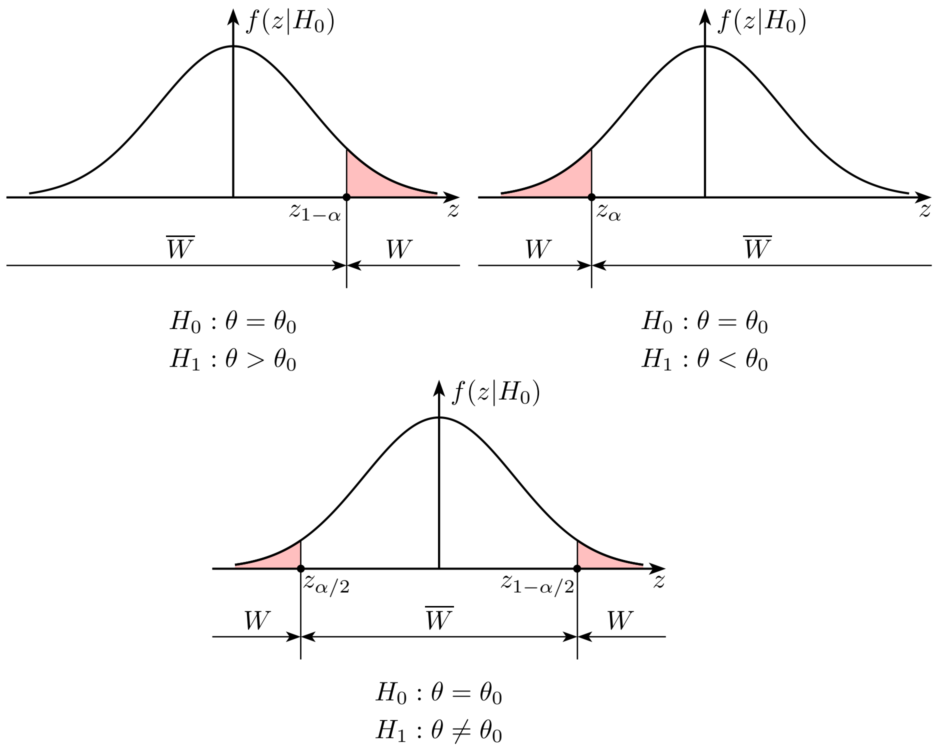
ако $Z_0 \notin W$ ($Z_0 \in \overline{W}$), то хипотезата H_0 се приема, т.е. счита се, че хипотезата H_0 не противоречи на резултатите от наблюдението.

Нека на статистическа проверка подлежи параметричната хипотеза $H_0 : \theta = \theta_0$, че даден параметър θ на генералната съвкупност е равен на хипотетично число θ_0.

Ще направим някои пояснения за етапите $4 - 7$ на проверката:

- Обикновено за статистика на критерия се избира случайна величина Z, чийто квантили z_p са зададени таблично: статистика с нормално разпределение $U = N(0, 1)$, статистика T с разпределение на Стюдънт, статистика с разпределение χ^2, статистика F с разпределение на Фишер и т.н.

- Нека Z е избраната статистика и z_p е нейният квантил от ред p: $P(Z < z_p) = p$.



Фиг. 14.1: Критични области

Тогава критичната област W се определя от неравенствата:

$$\begin{array}{ll} Z \geq z_{1-\alpha} & (\text{при алтернативна хипотеза } H_1^{(1)} : \theta > \theta_0); \\ Z \leq z_{\alpha} & (\text{при алтернативна хипотеза } H_1^{(2)} : \theta < \theta_0); \\ Z \leq z_{\alpha/2} \text{ или } Z \geq z_{1-\alpha/2} & (\text{при алтернативна хипотеза } H_1^{(3)} : \theta \neq \theta_0). \end{array}$$

На фиг. 14.1 е показано разположението на критичната област W при различните алтернативни хипотези. Тук $f(z|H_0)$ е плътността на разпределение на статистиката Z на критерия, при условие, че е вярна хипотезата H_0, $\overline{W}$ е областта на допустимите стойности на хипотезата, $P(Z \in \overline{W}|H_0) = 1 - \alpha$.

14.2 Хипотези за параметрите на една генерална съвкупност

Предполагаме, че генералната променлива X на дадена генерална съвкупност има нормално разпределение $N(m, \sigma)$ и е получена извадка от наблюдения на X:

$$x_1, x_2, \ldots, x_n. \tag{14.1}$$

Ще покажем как се извършва проверката на някои хипотези за параметрите m и σ на това разпределение.

Забележка 14.1. *Предположението, че X има нормално разпределение трябва да се основава на някакви теоретични факти или на предишни статистически изследвания на генералната съвкупност. Ако липсват такива обосновки, трябва (с използване на извадката (14.1)) да се провери хипотезата, че X има нормално разпределение (вж. раздел 14.2). Ако тази хипотеза не бъде отхвърлена, можем да извършим изложените по-долу проверки.*

14.2.1 Проверка на хипотезата, че средното m е равно на хипотетично число m_0

Проверява се хипотезата $H_0 : m = m_0$.

До проверка на такава хипотеза често се стига при качествения контрол на дадено серийно производство.

Например при производството на дадено лекарство количеството на основна съставка във всяка таблетка по стандарт трябва да е равно на m_0. На практика това количество е случайна величина X с нормално разпределение $N(m, \sigma)$. Може да се предположи, че генералното средно m на това количество в произведен партида е равно на стандартното количество m_0. Това предположение се проверява, като на базата на извадката от наблюдавани стойности (14.1) се провери параметричната хипотеза $H_0 : m = m_0$.

Проверката на хипотезата $H_0 : m = m_0$ се извършва в два случая:
- при известна дисперсия σ^2;
- при неизвестна дисперсия.

14.2.1.1 Дисперсията σ^2 на генералната променлива X е известна

Предполага се, че дисперсията σ^2 на генералната променлива X е известна, например от предишен опит, или е намерена теоретично, или е оценена по извадка с голям обем.

Оценка на генералното средно m е статистическото средно

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n).$$

При предположение, че хипотезата H_0 е вярна, $\overline{X}$ има нормално разпределение с математическо очакване $E\overline{X} = m_0$ и дисперсия $D\overline{X} = \frac{\sigma^2}{n}\cdot$ Тогава статистиката

$$Z = \frac{\overline{X} - m_0}{\sigma/\sqrt{n}}$$

има стандартно нормално разпределение $U = N(0, 1)$.

Затова критичната област W се определя от неравенствата:

$$\begin{array}{ll}
Z \geq u_{1-\alpha} & \text{(при алтернативна хипотеза } H_1^{(1)} : m > m_0); \\
Z \leq u_{\alpha} & \text{(при алтернативна хипотеза } H_1^{(2)} : m < m_0); \\
|Z| \geq u_{1-\alpha/2} & \text{(при алтернативна хипотеза } H_1^{(3)} : m \neq m_0).
\end{array}$$

Тук се отчита фактът, че $u_{\alpha/2} = -u_{1-\alpha/2}$ и че условието $|Z| \geq u_{1-\alpha/2}$ е изпълнено, когато $Z \leq -u_{1-\alpha/2}$ или $Z \geq u_{1-\alpha/2}$.

Ще отбележим, че при алтернативна хипотеза $H_1^{(3)} : m \neq m_0$ хипотезата $H_0 : m = m_0$ се приема, ако е изпълнено условието

$$-u_{1-\alpha/2} < \frac{\overline{x} - m_0}{\sigma/\sqrt{n}} < u_{1-\alpha/2}.$$

където $\overline{x}$ е наблюдаваната стойност на статистиката $\overline{X}$:

$$\overline{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n).$$

Като решим тези неравенства относно m_0, получаваме, че хипотезата H_0 (при алтернатива $H_1^{(3)} : m \neq m_0$) се приема, ако е изпълнено условието

$$\overline{x} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} < m_0 < \overline{x} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}.$$

Това означава, че хипотезата H_0 се приема, ако m_0 попада в двустранния доверителен интервал за параметъра m (вж. раздел 12.6).

14.2.1.2 Дисперсията на генералната променлива X е неизвестна

Ако дисперсията σ^2 на генералната променлива X е неизвестна или не може да се оцени достатъчно точно (например при малък обем на извадката), то при проверката на основната хипотеза $H_0 : m = m_0$ се използва статистиката

$$Z = \frac{\overline{X} - m_0}{\overline{S}/\sqrt{n}},$$

където $\overline{X}$ е статистическото средно, а $\overline{S}^2$ е неизместената статистическа дисперсия (вж. раздел 10.1):

$$\overline{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

При предположение, че хипотезата H_0 е вярна, статистиката Z има разпределение на Стюдънт $T(n-1)$ с $(n-1)$ степени на свобода.

Затова критичната област W се определя от неравенствата:

$$\begin{array}{ll}
Z \geq t_{1-\alpha}(n-1) & (\text{при алтернативна хипотеза } H_1^{(1)} : m > m_0); \\
Z \leq t_{\alpha}(n-1) & (\text{при алтернативна хипотеза } H_1^{(2)} : m < m_0); \\
|Z| \geq t_{1-\alpha/2}(n-1) & (\text{при алтернативна хипотеза } H_1^{(3)} : m \neq m_0).
\end{array}$$

Тук се отчита фактът, че $t_{\alpha/2}(n-1) = -t_{1-\alpha/2}(n-1)$ и че условието $|Z| \geq t_{1-\alpha/2}(n-1)$ е изпълнено, когато $Z \leq -t_{1-\alpha/2}(n-1)$ или $Z \geq t_{1-\alpha/2}(n-1)$.

Ще отбележим, че при алтернативна хипотеза $H_1^{(3)} : m \neq m_0$ хипотезата $H_0 : m = m_0$ се приема, ако е изпълнено условието

$$-t_{1-\alpha/2}(n-1) < \frac{\overline{x} - m_0}{\overline{s}/\sqrt{n}} < t_{1-\alpha/2}(n-1),$$

където $\overline{x}$ и $\overline{s}$ са наблюдаваните стойности на статистиките $\overline{X}$ и $\overline{S}$.

Като решим тези неравенства относно m_0, получаваме, че хипотезата H_0 (при алтернатива $H_1^{(3)} : m \neq m_0$) се приема, ако е изпълнено условието

$$\overline{x} - \frac{\overline{s}}{\sqrt{n}} u_{1-\alpha/2} < m_0 < \overline{x} + \frac{\overline{s}}{\sqrt{n}} u_{1-\alpha/2}.$$

Това означава, че хипотезата H_0 се приема, ако m_0 попада в двустранния доверителен интервал за параметъра m при неизвестна дисперсия (вж. раздел 12.6).

14.2.2 Проверка на хипотезата, че дисперсията σ^2 е равна на хипотетично число σ_0^2

Проверява се хипотезата $H_0 : \sigma^2 = \sigma_0^2$.

До проверка на такава хипотеза се стига, когато искаме да проверим точността на прибори или инструменти, точността на количествен метод за изследване, устойчивостта на технологичен процес.

Например при производството на дадено лекарство количеството на основна съставка във всяка таблетка по стандарт трябва да е равно на m_0 и да има дисперсия, не по-голяма от σ_0^2 . На практика това количество е случайна величина X с нормално разпределение $N(m, \sigma)$. Може да се предположи, че генералната дисперсия σ^2 на това количество в произведен партида е равно на σ_0^2. Това предположение се проверява, като на базата на извадката от наблюдавани стойности (14.1) се провери параметричната хипотеза $H_0 : \sigma^2 = \sigma_0^2$.

Проверката на хипотезата $H_0 : \sigma^2 = \sigma_0^2$ се извършва в два случая:

- при известно средно m_0;
- при неизвестно средно.

14.2.2.1 Средното m_0 е известно

Предполага се, че средното m_0 на статистическата променлива X е известно, например от предишен опит, или е намерено теоретично, или е оценено по извадка с голям обем.

В този случай, оценката $D_0 = S_0^2$ на генералната дисперсия σ^2 се изчислява по формулата

$$D_0 = S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m_0)^2.$$

При предположение, че хипотезата H_0 е вярна, статистиката

$$Z = \frac{nS_0^2}{\sigma_0^2}$$

има разпределение $\chi^2(n)$ с n степени на свобода.

Затова критичната област W се определя от неравенствата:

$Z \geq \chi^2_{1-\alpha}(n)$ (при алтернативна хипотеза $H_1^{(1)} : \sigma^2 > \sigma_0^2$);
$Z \leq \chi^2_{\alpha}(n)$ (при алтернативна хипотеза $H_1^{(2)} : \sigma^2 < \sigma_0^2$);
$Z \leq \chi^2_{\alpha/2}(n)$ или $Z \geq \chi^2_{1-\alpha/2}(n)$
(при алтернативна хипотеза $H_1^{(3)} : \sigma^2 \neq \sigma_0^2$).

Ще отбележим, че при алтернативна хипотеза $H_1^{(3)} : \sigma^2 \neq \sigma_0^2$ хипотезата $H_0 : \sigma^2 = \sigma_0^2$ се приема, ако е изпълнено условието

$$\chi^2_{\alpha/2}(n) < \frac{ns_0^2}{\sigma_0^2} < \chi^2_{1-\alpha/2}(n),$$

където s_0^2 е наблюдаваната стойност на статистиката S_0^2:

$$s_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m_0)^2.$$

Като решим тези неравенства относно σ_0^2, получаваме, че хипотезата H_0 (при алтернатива $H_1^{(3)} : \sigma^2 \neq \sigma_0^2$) се приема, ако е изпълнено условието

$$\frac{ns_0^2}{\chi^2_{1-\alpha/2}(n)} < \sigma_0^2 < \frac{ns_0^2}{\chi^2_{\alpha/2}(n)}\,.$$

Това означава, че хипотезата H_0 се приема, ако σ_0^2 попада в двустранния доверителен интервал за параметъра σ^2 в случая, когато средното m е известно.

14.2.2.2 Средното m е неизвестно

Ако средното m на статистическата променлива X е неизвестно или не може да се оцени достатъчно точно (например при малък обем на извадката), то при проверката на основната хипотеза $H_0 : \sigma^2 = \sigma_0^2$ се използва статистика

$$Z = \frac{(n-1)\overline{S}^2}{\sigma_0^2},$$

където

$$\overline{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

е неизместената статистическа дисперсия и $\overline{X}$ е статистическото средно.

При предположение, че хипотезата H_0 е вярна, статистиката Z има разпределение $\chi^2(n-1)$ с $(n-1)$ степени на свобода.

Затова критичната област W се определя от неравенствата:

$Z \geq \chi^2_{1-\alpha}(n-1)$ (при алтернативна хипотеза $H_1^{(1)} : \sigma^2 > \sigma_0^2$);
$Z \leq \chi^2_{\alpha}(n-1)$ (при алтернативна хипотеза $H_1^{(2)} : \sigma^2 < \sigma_0^2$);
$Z \leq \chi^2_{\alpha/2}(n-1)$ или $Z \geq \chi^2_{1-\alpha/2}(n-1)$
(при алтернативна хипотеза $H_1^{(3)} : \sigma^2 \neq \sigma_0^2$).

Ще отбележим, че при алтернативна хипотеза $H_1^{(3)} : \sigma^2 \neq \sigma_0^2$ хипотезата $H_0 : \sigma^2 = \sigma_0^2$ се приема, ако е изпълнено условието

$$\chi^2_{\alpha/2}(n-1) < \frac{(n-1)\overline{s}^2}{\sigma_0^2} < \chi^2_{1-\alpha/2}(n-1),$$

където $\overline{s}^2$ е наблюдаваната стойност на статиката $\overline{S}^2$:

$$\overline{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Като решим тези неравенства относно σ_0^2, получаваме, че хипотезата H_0 (при алтернатива $H_1^{(3)} : \sigma^2 \neq \sigma_0^2$) се приема, ако е изпълнено условието

$$\frac{(n-1)\overline{s}^2}{\chi^2_{1-\alpha/2}(n-1)} < \sigma_0^2 < \frac{(n-1)\overline{s}^2}{\chi^2_{\alpha/2}(n-1)}.$$

Това означава, че хипотезата H_0 се приема, ако σ_0^2 попада в двустранния доверителен интервал за параметъра σ^2 в случая, когато средното m е неизвестно.

Пример 14.1. *От нормално разпределена съвкупност е направена извадка с обем $n = 17$ и по нея е пресметнатата неизместената статистическа дисперсия $\overline{s}^2 = 0.24$. С ниво на значимост $\alpha = 0.05$ да се провери хипотезата $H_0 : \sigma^2 = 0.18$ при алтернативна хипотеза $H_0 : \sigma^2 > 0.18$.*

Решение: Проверява се хипотеза от вида $H_0 : \sigma^2 = \sigma_0^2$ при неизвестно средно. Тогава се използва статистиката

$$Z = \frac{(n-1)\overline{S}^2}{\sigma_0^2}.$$

Имаме, че $n = 17$, $\sigma_0^2 = 0.18$, $\overline{s}^2 = 0.24$ и наблюдаваната стойност на статистиката е

$$Z_0 = \frac{(17-1)\cdot 0.24}{0.18} = 21.33.$$

При предположение, че хипотезата H_0 е вярна статистиката Z има разпределение $\chi^2(f)$ с $f = n-1 = 16$ степени на свобода. Понеже алтернативната хипотеза е от вида $H_0 : \sigma^2 > \sigma_0^2$, то критичната област се определя с неравенството $Z \geq \chi^2_{1-\alpha}(n-1)$. От Таблица 5 определяме квантила $\chi^2_{0.95}(16) = 26.3$. Понеже $Z_0 = 21.33 < 26.3$, то Z_0 не попада в критичната област и хипотезата H_0 не се отхвърля.

14.3 Хипотези за параметрите на две генерални съвкупности (независими извадки)

В практиката се налага да се сравняват параметрите на статистически променливи X и Y на две различни генерални съвкупности. Най-често такова сравнение се извършва в случая, когато променливите X и Y са непрекъснати, измерват се с едни и същи мерни единици и са нормално разпределени.

При предположение, че X и Y имат нормални разпределения ($X = N(m_1, \sigma_1)$ и $Y = N(m_2, \sigma_2)$), ще покажем как се извършват проверките на някои хипотези за сравнение на параметрите m_1, m_2 и σ_1, σ_2 на тези разпределения.

Сравнението се извършва на базата на **независимите извадки**

$$x_1, x_2, \ldots, x_{n_1}, \tag{14.2}$$

$$y_1, y_2, \ldots, y_{n_2}, \tag{14.3}$$

получени при наблюдения на променливите X и Y във всяка от разглежданите генерални съвкупности поотделно.

Забележка 14.2. *Предположението, че X и Y имат нормално разпределение трябва да се основава на някакви теоретични факти или на предишни статистически изследвания на генералните съвкупности. Ако липсват такива обосновки, трябва с използване на извадките* (14.2) *и* (14.3) *да се проверят поотделно хипотезите, че X и Y имат нормално разпределение* (*вж. раздел* 14.2). *Ако тези хипотези не бъдат отхвърлени, можем да извършим изложените по-долу проверки.*

14.3.1 Сравнение на дисперсиите σ_1^2 и σ_2^2

Проверява се основната хипотеза $H_0 : \sigma_1^2 = \sigma_2^2$.

До проверка на такава хипотеза се стига, когато се прави сравнение на точността на два измерителни прибора, сравнява се точността на два метода за измерване и др. Очевидно, за предпочитане е този измерителен прибор, който осигурява по-малка дисперсия на резултатите от измерванията.

Ще отбележим, че условието $\sigma_1^2 = \sigma_2^2$ е задължително при извършването на някои проверки на хипотези и затова често се налага да се проверява дали това условие е изпълнено.

Проверката на хипотезата $H_0 : \sigma_1^2 = \sigma_2^2$ се извършва в два случая:

- при известни средни m_1 и m_2;
- при неизвестни средни m_1 и m_2.

14.3.1.1 Средните m_1 и m_2 са известни

Предполага се, че средните m_1 и m_2 на статистическите променливи X и Y са известни, например от предишен опит, или са намерени теоретично, или са оценени по извадки с голям обем.

В този случай, оценките $D_{01} = S_{01}^2$ и $D_{02} = S_{02}^2$ на генералните дисперсии σ_1^2 и σ_2^2 се изчисляват по формулите

$$S_{01}^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(X_i - m_1)^2, \quad S_{02}^2 = \frac{1}{n_2}\sum_{i=1}^{n_2}(Y_i - m_2)^2.$$

Нека s_{01}^2 и s_{02}^2 са наблюдаваните стойности на статистиките S_{01}^2 и S_{02}^2 и $s_{01}^2 > s_{02}^2$.

При предположение, че хипотезата H_0 е вярна, статистиката

$$Z = \frac{S_{01}^2}{S_{02}^2}$$

има разпределение на Фишер $F(n_1, n_2)$ с n_1 и n_2 степени на свобода.

Затова критичната област W се определя от неравенствата:

$$\begin{array}{ll} Z \geq F_{1-\alpha}(n_1, n_2) & \text{(при алтернативна хипотеза } H_1^{(1)} : \sigma_1^2 > \sigma_2^2\text{);} \\ Z \geq F_{1-\alpha/2}(n_1, n_2) & \text{(при алтернативна хипотеза } H_1^{(3)} : \sigma_1^2 \neq \sigma_2^2\text{).} \end{array}$$

Ще отбележим, че при алтернативна хипотеза $H_1^{(3)} : \sigma_1^2 \neq \sigma_2^2$ хипотезата $H_0 : \sigma_1^2 = \sigma_2^2$ се приема, ако е изпълнено условието

$$\frac{s_{01}^2}{s_{02}^2} < F_{1-\alpha/2}(n_1, n_2).$$

14.3.1.2 Средните m_1 и m_2 са неизвестни

Предполага се, че средните m_1 и m_2 на статистическите променливи X и Y са неизвестни.

В този случай, оценките $\overline{S}_1^2$ и $\overline{S}_2^2$ на генералните дисперсии σ_1^2 и σ_2^2 се изчисляват по формулите

$$\overline{S}_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \overline{X})^2, \quad \overline{S}_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \overline{Y})^2. \tag{14.4}$$

Нека $\overline{s}_1^2$ и $\overline{s}_2^2$ са наблюдаваните стойности на статистиките $\overline{S}_1^2$ и $\overline{S}_2^2$ и $\overline{s}_1^2 > \overline{s}_2^2$.

При предположение, че хипотезата H_0 е вярна, статистиката

$$Z = \frac{\overline{S}_1^2}{\overline{S}_2^2}$$

има разпределение на Фишер $F(n_1-1, n_2-1)$ с n_1-1 и n_2-1 степени на свобода.

Затова критичната област W се определя от неравенствата:

$$Z \geq F_{1-\alpha}(n_1-1, n_2-1) \quad (\text{при алтернативна хипотеза } H_1^{(1)}: \sigma_1^2 > \sigma_2^2);$$
$$Z \geq F_{1-\alpha/2}(n_1-1, n_2-1)(\text{при алтернативна хипотеза } H_1^{(3)}: \sigma_1^2 \neq \sigma_2^2).$$

Ще отбележим, че при алтернативна хипотеза $H_1^{(3)}: \sigma_1^2 \neq \sigma_2^2$ хипотезата $H_0: \sigma_1^2 = \sigma_2^2$ се приема, ако е изпълнено условието

$$\frac{\overline{s}_1^2}{\overline{s}_2^2} < F_{1-\alpha/2}(n_1-1, n_2-1).$$

Пример 14.2. *По две независими извадки с обеми $n_1 = 11$ и $n_2 = 14$, извлечени от две нормално разпределени генерални съвкупности X и Y, са пресметнати неизместените статистически дисперсии: $\overline{s}_X^2 = \overline{s}_1^2 = 0.76$, $\overline{s}_Y^2 = \overline{s}_2^2 = 0.38$. При ниво на значимост $\alpha = 0.05$ да се провери хипотезата $H_0: \sigma_1^2 = \sigma_2^2$ за равенство на генералните дисперсии при алтернативна хипотеза $H_1: \sigma_1^2 > \sigma_2^2$.*

Решение: Понеже средните на X и Y са неизвестни, то ще използваме статистиката

$$Z = \frac{\overline{S}_1^2}{\overline{S}_2^2}.$$

Намираме отношението на по-голямата към по-малката неизместени статистически дисперсии

$$Z_0 = \frac{\overline{s}_1^2}{\overline{s}_1^2} = \frac{0.76}{0.38} = 2.$$

Според вида на алтернативната хипотеза критичната област се задава с неравенството $Z \geq F_{1-\alpha}(n_1-1, n_2-1) = c_r$. От Таблица 8 с $\alpha = 0.05$ и степени на свобода $k_1 = n_1 - 1 = 11 - 1 = 10$, $k_2 = n_2 - 1 = 14 - 1 = 13$ определяме дясната критична точка

$$c_r = F_{0.95}(10, 13) = 2.67.$$

Понеже $Z_0 = 2 < 2.67 = c_r$, то Z_0 не попада в критичната област и няма основание за отхвърляне на основната хипотеза H_0.

14.3.2 Сравнение на средните m_1 и m_2

Проверява се основната хипотеза $H_0 : m_1 = m_2$.

В статистиката често се налага да се сравнят генералните средни m_1 и m_2 на две нормално разпределени статистически променливи X и Y и да се определи дали те са равни или се различават. Това може да се направи, като се сравнят статистическите средни $\overline{x}$ и $\overline{y}$ на тези променливи, пресметнати от извадките (14.2) и (14.3). Но понеже, като правило, средните $\overline{x}$ и $\overline{y}$ се оказват различни, то възниква въпросът: значително или незначително е това различие? На този въпрос може да се отговори, като се провери основната хипотеза $H_0 : m_1 = m_2$.

Ако след проверката основната хипотеза H_0 се приеме, то разликата между статистическите средни е незначителна и се обяснява със случайни причини, в частност, със случайния подбор на елементите в извадките. Например, ако физическите величини A и B имат равни истински размери, а средните аритметични $\overline{x}$ и $\overline{y}$ на резултатите от измерванията на тези величини са различни, то това различие е незначително.

Ако след проверката основната хипотеза H_0 се отхвърли, то разликата между статистическите средни е значителна и не може да бъде обяснена със случайни причини, а се обяснява с това, че генералните средни m_1 и m_2 са различни. Например, ако средното аритметично на резултатите от измерванията на физичната величина A значително се различават от средното аритметично на резултатите от измерванията на физичната величина B, то истинските размери на тези величини са различни.

Проверката на хипотезата $H_0 : m_1 = m_2$ се извършва по различни начини в разгледаните по-долу случаи:

14.3.2.1 Дисперсиите σ_1^2 и σ_2^2 са известни

Предполага се, че дисперсиите σ_1^2 и σ_2^2 на статистическите променливи X и Y са известни.

При предположение, че основната хипотеза $H_0 : m_1 = m_2$ е вярна, статистиката

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

има стандартно нормално разпределение $U = N(0, 1)$.

Затова критичната област W се определя от неравенствата:

$Z \geq u_{1-\alpha}$ (при алтернативна хипотеза $H_1^{(1)} : m_1 > m_2$);
$Z \leq u_{\alpha}$ (при алтернативна хипотеза $H_1^{(2)} : m_1 < m_2$);
$|Z| \geq u_{1-\alpha/2}$ (при алтернативна хипотеза $H_1^{(3)} : m_1 \neq m_2$).

Ще отбележим, че при алтернативна хипотеза $H_1^{(3)} : m_1 \neq m_2$ хипотезата $H_0 : m_1 = m_2$ се приема, ако е изпълнено условието

$$-u_{1-\alpha/2} < \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} < u_{1-\alpha/2}.$$

Забележка 14.3 (Големи независими извадки). *Нека статистическите променливи X и Y са непрекъснати, измерват се едни и същи мерни единици, но не са нормално разпределени. Тогава не можем да приложим изложения по-горе начин на проверка на хипотезата $H_0 : m_1 = m_2$.*

Обаче, ако независимите извадки (14.2) и (14.3) са „големи", т.е. $n_1 \geq 30$ и $n_2 \geq 30$, то статистическите средни $\overline{X}$ и $\overline{Y}$ приближено са нормално разпределени, а неизместените статистически дисперсии $\overline{S}_1^2$ и $\overline{S}_2^2$ (формули (14.4)) са достатъчно добри оценки на генералните дисперсии σ_1^2 и σ_2^2. Затова можем да считаме, че те са известни приближено. Тогава статистиката

$$Z^* = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\overline{S}_1^2}{n_1} + \dfrac{\overline{S}_2^2}{n_2}}}$$

приближено има стандартно нормално разпределение $U = N(0, 1)$ и може да проверяваме хипотезата $H_0 : m_1 = m_2$ така, както е изложено по-горе, като заменим статистиката Z със статистиката Z^*.

Понеже разгледаният критерий за проверка е приближен, то към получените с неговото прилагане изводи трябва да се отнасяме внимателно.

14.3.2.2 Дисперсиите σ_1^2 и σ_2^2 са неизвестни, но са равни

Предполага се, че статистическите променливи X и Y са разпределени нормално, но техните дисперсии σ_1^2 и σ_2^2 са неизвестни. Например извадките (14.2) и (14.3) са „малки" ($n_1 < 30$ или $n_2 < 30$) и не може да се получат добри оценки за генералните дисперсии σ_1^2 и σ_2^2. Поради това, методът за сравнение на средните, изложен в предната подточка, не може да се приложи.

Обаче, ако допълнително се предположи, че неизвестните генерални дисперсии σ_1^2 и σ_2^2 са равни, то може да се формулира критерий за сравнение на средните m_1 и m_2. Например, ако се сравняват средните на количествата на лечебните съставки на две партиди от лекарство, произведени на една и съща поточна линия, то естествено е да се допусне, че дисперсиите на контролираните количества са равни.

В случая, когато липсват подобни съображения, трябва да се подложи на проверка хипотезата $H_0 : \sigma_1^2 = \sigma_2^2$ (вж. точка 14.3.1). Ако тази хипотеза се приеме, може да се пристъпи към проверката на основната хипотеза $H_0 : m_1 = m_2$ по изложения по-долу начин.

В критерия за проверка се използва статистиката

$$Z = \frac{\overline{X} - \overline{Y}}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}},$$

където $\overline{X}$ и $\overline{Y}$ са статистическите средни,

$$S^2 = \frac{(n_1 - 1)\overline{S}_1^2 + (n_2 - 1)\overline{S}_2^2}{n_1 + n_2 - 2}$$

е „общата дисперсия”, а неизместените статистически дисперсии $\overline{S}_1^2$ и $\overline{S}_2^2$, пресметнати по формули (14.4), са оценки на дисперсиите σ_1^2 и σ_2^2.

При предположение, че основната хипотеза $H_0 : m_1 = m_2$ е вярна, статистиката Z има разпределение на Стюдънт $T(f)$ с $f = n_1 + n_2 - 2$ степени на свобода.

Тогава критичната област W се определя от неравенствата:

$Z \geq t_{1-\alpha}(f)$ (при алтернативна хипотеза $H_1^{(1)} : m_1 > m_2$);
$Z \leq t_{\alpha}(f)$ (при алтернативна хипотеза $H_1^{(2)} : m_1 < m_2$);
$|Z| \geq t_{1-\alpha/2}(f)$ (при алтернативна хипотеза $H_1^{(3)} : m_1 \neq m_2$).

Ще отбележим, че при алтернативна хипотеза $H_1^{(3)} : m_1 \neq m_2$ хипотезата $H_0 : m_1 = m_2$ се приема, ако е изпълнено условието

$$-t_{1-\alpha/2}(f) < \frac{\overline{x} - \overline{y}}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} < t_{1-\alpha/2}(f),$$

където

$$s^2 = \frac{(n_1-1)\overline{s}_1^2 + (n_2-1)\overline{s}_2^2}{n_1+n_2-2}$$

е наблюдаваната стойност на общата дисперсия S^2, а $\overline{s}_1^2$ и $\overline{s}_2^2$ са наблюдаваните стойности на $\overline{S}_1^2$ и $\overline{S}_2^2$.

Пример 14.3. *По две независими извадки с обеми $n = 10$ и $m = 8$, извлечени от две нормално разпределени генерални съвкупности X и Y, са пресметнати статистическите средни: $\overline{x} = 142.3$, $\overline{y} = 145.3$ и неизместените статистически дисперсии: $\overline{s}_X^2 = 2.7$, $\overline{s}_Y^2 = 3.2$. При ниво на значимост $\alpha = 0.01$ да се провери хипотезата $H_0 : EX = EY$ за равенство на генералните средни при алтернативна хипотеза $H_1 : EX \neq EY$.*

Решение: Понеже дисперсиите на X и Y са неизвестни и извадките са „малки", то ще процедираме както в точка 14.3.2.2. За целта първо проверяваме хипотезата $H_0 : \sigma_X^2 = \sigma_Y^2$ за равенство на двете дисперсии (вж. точка 14.3.1.2). Определяме отношението на по-голямата към по-малката неизместени статистически дисперсии

$$Z_0 = \frac{\overline{s}_Y^2}{\overline{s}_X^2} = \frac{3.2}{2.7} \approx 1.185.$$

От Таблица 9 на разпределението на Фишер с $\alpha = 0.01$ и $k_1 = n_1 - 1 = 8 - 1 = 7$ и $k_2 = n_2 - 1 = 10 - 1 = 9$ степени на свобода определяме $F_{1-\alpha}(k_1, k_2) = F_{0.99}(7, 9) = 5.61$. Понеже $Z_0 = 1.185 < 5.61$, то хипотезата за равенството на двете дисперсии не се отхвърля.

Сега можем да пристъпим към проверката на основната хипотеза $H_0 : EX = EY$. За целта пресмятаме наблюдаваната стойност на общата дисперсия

$$s^2 = \frac{(n_1-1)\overline{s}_1^2 + (n_2-1)\overline{s}_2^2}{n_1+n_2-2} = \frac{7 \cdot 3.2 + 9 \cdot 2.7}{7+9} = 2.91875,$$

намираме $s = \sqrt{2.91875} \approx 1.71$ и определяме наблюдаваната стойност на използваната статистика

$$Z_0 = \frac{\overline{x} - \overline{y}}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \frac{142.3 - 145.3}{1.71\sqrt{1/8 + 1/10}} \approx -3.7.$$

От Таблица 6 на разпределението на Стюдънт $T(f)$ с ниво на значимост $\alpha = 0.01$ и $f = n_1 + n_2 - 2 = 8 + 10 - 2 = 16$ степени на свобода определяме $t_{1-\alpha/2}(f) = t_{0.995}(16) = 2.921$. Понеже $|Z_0| = 3.7 > 2.921$, то Z_0 попада в критичната област. Хипотезата за равенството на двете средни се отхвърля.

14.4 Сравнение на средните m_1 и m_2 на две зависими извадки

В предните параграфи се предполагаше, че наблюденията (14.2) и (14.3) са независими. Сега ще разгледаме случая, когато се сравняват две наблюдения

$$x_1, x_2, \ldots, x_n; \quad y_1, y_2, \ldots, y_n \tag{14.5}$$

на една статистическа променлива, които са извършени над елементите на една и съща извадка на дадена генерална съвкупност.

Например от партида с произведени детайли са подбрани по случаен начин n детайла, на които са направени две измервания на определен контролиран размер и получените наблюдения са: $x_1, x_2, \ldots, x_n$ – резултатите от измерванията с първия контролен прибор; $y_1, y_2, \ldots, y_n$ – резултатите от измерванията на същите детайли, направени в същия ред с втория контролен прибор. Тогава резултатите $x_i, y_i,\ i = 1, 2, \ldots, n$, които се отнасят за един и същ детайл са зависими и в този смисъл самите извадки са зависими. Понеже, като правило $x_i \neq y_i,\ i = 1, 2, \ldots, n$ и $\overline{x} \neq \overline{y}$, то възниква необходимостта да се провери значими или незначими са тези различия.

До подобно сравнение на резултатите от две измервания (наблюдения) на една и съща извадка от елементи на генерална съвкупност се стига, когато се изследва влиянието на някакъв фактор върху наблюдаваната статистическа променлива X. Нека Y е същата статистическа променлива, но след въздействието на фактора. Тогава $x_1, x_2, \ldots, x_n$ ще бъдат наблюденията преди въздействието на фактора, а $y_1, y_2, \ldots, y_n$ – наблюденията след въздействието.

Нека $EX = m_1$ и $EY = m_2$. За да се определи дали факторът е имал въздействие върху променливата X, се прави проверка на хипотезата $H_0 : m_1 = m_2$ при алтернативна хипотеза $H_1 : m_1 \neq m_2$.

Условията, при които се извършва такава проверка, са следните:

- Изучаваните елементи на извадката, за които се отчита изменението на измерваната статистическа променлива преди и след въздействието върху извадката, да са избрани случайно от генералната съвкупност;

- На всеки елемент от извадката да съответстват две измервания: x_i – до въздействието и y_i – след въздействието;

- Разликата $R = X - Y$ има нормално разпределение.

Нека $ER = r$. Тогава основната H_0 и алтернативната H_1 хипотези може да се запишат още във вида:

$$H_0 : r = 0, \qquad H_1 : r \neq 0.$$

Проверката на хипотезата $H_0 : r = 0$ се извършва в следния ред:

- Пресмятаме разликите $r_i = x_i - y_i,\ i = 1, 2, \dots, n$;

- Ако се налага, подлагаме на проверка хипотезата, че R има нормално разпределение и ако тази хипотеза не бъде отхвърлена, продължаваме със следващите стъпки;

- Пресмятаме наблюдаваните стойности $\overline{r}$ и $\overline{s}^2$ на статистическото средно
$$\overline{R} = \frac{1}{n}\sum_{i=1}^{n} R_i$$
и на неизместената статистическа дисперсия
$$\overline{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(R_i - \overline{R})^2;$$

- Пресмятаме наблюдаваната стойност Z_0 на статистиката
$$Z = \frac{\overline{R}\sqrt{n}}{\overline{S}},$$
която има разпределение на Стюдънт $T(f)$ с $f = n - 1$ степени на свобода;

- Задаваме ниво на значимост α и от Таблица 6 определяме квантила $t_{1-\alpha/2}(f)$;
- Вземаме решение:

  Ако $|Z_0| < t_{1-\alpha/2}(f)$, хипотезата H_0 се приема. Тогава може да се приеме, че факторът е оказал незначително въздействие върху променливата X;

  Ако $|Z_0| \geq t_{1-\alpha/2}(f)$, хипотезата H_0 се отхвърля. Това означава, че факторът е оказал значително въздействие върху променливата X.

Пример 14.4. *На две аналитични везни, в един и същи ред, са претеглени* 10 *проби на химически вещества и са получени следните резултати:*

x_i, mg	25	30	28	50	20	40	32	36	42	38
y_i, mg	28	31	26	52	24	36	33	35	45	40

При ниво на значимост $\alpha = 0.01$ да се определи значимо или незначимо се различават резултатите от претеглянията, при предположение, че те са разпределени нормално.

Решение: Проверява се хипотезата $H_0 : m_1 = m_2$ за равенство на средните на две зависими извадки на нормално разпределени случайни величини при алтернативна хипотеза $H_1 : m_1 \neq m_2$. За целта първо намираме разликите $r_i = x_i - y_i$:

$$r_i,\ mg| -3,\ -1,\ 2,\ -2,\ -4,\ 4,\ -1,\ 1,\ -3,\ -2$$

Пресмятаме наблюдаваните стойности $\overline{r}$ и $\overline{s}^2$ на статистическото средно $\overline{R}$ и на неизместената статистическа дисперсия $\overline{S}^2$:

$$S_r = -3 - 1 + 2 - 2 - 4 + 4 - 1 + 1 - 3 - 3 = -9,\ \overline{r} = \frac{S_r}{n} = \frac{-9}{10} = -0.9,$$
$$S_{rr} = 9 + 1 + 4 + 4 + 16 + 16 + 1 + 1 + 9 + 4 = 65,$$
$$\overline{s}^2 = \frac{nS_{rr} - S_r^2}{(n-1)n} = \frac{650 - 81}{90} \approx 6.32,\ \overline{s} = \sqrt{6.32} \approx 2.51.$$

Определяме наблюдаваната стойност Z_0 на статистиката $Z = \dfrac{\overline{R}\sqrt{n}}{\overline{S}}$:

$$Z_0 = \frac{\overline{r}\sqrt{n}}{\overline{s}} = \frac{-0.9\sqrt{10}}{2.51} \approx -1.13.$$

От Таблица 6 на разпределението на Стюдънт $T(f)$ с ниво на значимост $\alpha = 0.01$ и $f = n-1 = 10-1 = 9$ степени на свобода определяме $t_{1-\alpha/2}(f) = t_{0.995}(9) = 3.25$. Понеже $|Z_0| = 1.13 < 3.25$, то Z_0 не попада в критичната област. Хипотезата за равенството на двете средни се приема. Резултатите от претеглянията на двете аналитични везни се различават незначимо.

14.5 Проверка на хипотезата за стойността на неизвестна вероятност

Нека са направени n независими опита, при всеки от които дадено събитие A се сбъдва с неизвестна вероятност $p = P(A)$. Точкова оценка на тази вероятност е относителната честота $w = \dfrac{\overline{x}}{n}$ на настъпването на събитието A, където n е броят на опитите, а $\overline{x}$ е броят на успешните опити, при които се е сбъднало събитието A.

Основавайки се на тази оценка, може издигнем и проверим хипотезата $H_0 : p = p_0$, че неизвестната вероятност p е равна на хипотетично число p_0, което е близко до w.

Ако са изпълнени условията

$$n > 50,\ nw > 5,\ n(1-w) > 5 \tag{14.6}$$

и хипотезата H_0 е вярна, то статистиката

$$Z = \frac{w - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

има разпределение, близко до нормалното разпределение $U = N(0, 1)$.

Затова критичната област W се определя от неравенствата:

$Z \geq u_{1-\alpha}$ (при алтернативна хипотеза $H_1^{(1)} : p > p_0$);
$Z \leq u_{\alpha}$ (при алтернативна хипотеза $H_1^{(2)} : p < p_0$);
$|Z| \geq u_{1-\alpha/2}$ (при алтернативна хипотеза $H_1^{(3)} : p \neq p_0$).

За проверка на хипотезата $H_0 : p = p_0$ може да се използват също доверителните интервали за вероятността p (вж. раздел 12.7). При това хипотезата H_0 се приема при ниво на значимост α, ако съответния едностранен или двустранен доверителен интервал съдържа стойността p_0; в противен случай хипотезата H_0 се отхвърля.

14.6 Проверка на хипотезата за равенство на две неизвестни вероятности

За проверката на хипотезата $H_0 : p_1 = p_2$ за равенство на две неизвестни вероятности p_1 и p_2 се правят две серии от независими опити. Нека в серия от n_{1*} опита дадено събитие A се е появило n_{11} пъти, а в серия от n_{2*} опита същото събитие се е появило n_{21} пъти. Резултатите от двете серии от опити са представени в долната таблица.

Серия	Събития		Сума
	A	$\overline{A}$	
1	n_{11}	n_{12}	n_{1*}
2	n_{21}	n_{22}	n_{2*}
Сума	n_{*1}	n_{*2}	n

Проверява се хипотезата H_0 за равенство на вероятностите за поява на събитието A в двете серии от опити. Въвеждаме следните означения:

$$w_1 = \frac{n_{11}}{n_{1*}}, \quad w_2 = \frac{n_{21}}{n_{2*}}, \quad w = \frac{n_{11} + n_{21}}{n_{1*} + n_{2*}} = \frac{n_{*1}}{n} \cdot$$

Нека са изпълнени условията

$$\frac{n_{1*}n_{*1}}{n} > 5, \quad \frac{n_{1*}n_{*2}}{n} > 5, \quad \frac{n_{2*}n_{*1}}{n} > 5, \quad \frac{n_{2*}n_{*2}}{n} > 5. \tag{14.7}$$

Тогава за проверка на хипотезата $H_0 : p_1 = p_2$ се използва статистиката

$$Z = \frac{w_1 - w_2}{\sqrt{\overline{S}^2_{w_1 - w_2}}}\,, \tag{14.8}$$

където $\overline{S}^2_{w_1 - w_2}$ е оценката на дисперсията на разликата на случайните величини w_1 и w_2 и се пресмята по формулата

$$\overline{S}^2_{w_1 - w_2} = w(1 - w)\left(\frac{1}{n_{1*}} + \frac{1}{n_{2*}}\right). \tag{14.9}$$

Ако хипотезата H_0 е вярна, статистиката Z има разпределение, близко до стандартното нормално разпределение $U = N(0, 1)$.

Затова критичната област W се определя от неравенствата:

$Z \geq u_{1-\alpha}$ (при алтернативна хипотеза $H_1^{(1)} : p_1 > p_2$);
$Z \leq u_{\alpha}$ (при алтернативна хипотеза $H_1^{(2)} : p_1 < p_2$);
$|Z| \geq u_{1-\alpha/2}$ (при алтернативна хипотеза $H_1^{(3)} : p_1 \neq p_2$).

Ако някое от условията (14.7) не е изпълнено, за проверката на хипотезата H_0 трябва да се използва критерият χ^2 (вж. раздел 14.5).

Пример 14.5. *За изучаване на ефективността на профилактично лекарство против алергия са изследвани две групи от хора, предразположени към това заболяване. Получени са следните резултати:*

Приемали лекарство		*Неприемали лекарство*	
Заболели	*Незаболели*	*Заболели*	*Незаболели*
3	172	32	168

Означават ли тези резултати, че лекарството е ефективно, ако нивото на значимост е $\alpha = 0.05$?

Решение: Означаваме с p_1 вероятността за заболяване, ако се приема лекарството, а с p_2 – ако не се приема. Проверява се хипотезата $H_0 : p_1 = p_2$ при алтернативна хипотеза $H_1 : p_1 < p_2$. Получените резултати оформяме в следната таблица

	Заболели	*Незаболели*	*Сума*
Приемали лекарство	3	172	$n_{1*} = 175$
Неприемали лекарство	32	168	$n_{2*} = 200$
Сума	$n_{*1} = 35$	$n_{*2} = 340$	$n = 375$

Определяме относителните честоти:

$$w_1 = \frac{3}{175} \approx 0.017,\ w_2 = \frac{32}{200} = 0.160,\ w = \frac{35}{375} \approx 0.093.$$

Проверяваме, че наистина са изпълнени условията (14.7):

$$\frac{n_{1*}n_{*1}}{n} \approx 16.33,\ \frac{n_{1*}n_{*2}}{n} \approx 158.67,\ \frac{n_{2*}n_{*1}}{n} \approx 18.67,\ \frac{n_{2*}n_{*2}}{n} \approx 181.33.$$

По формула (14.9) пресмятаме наблюдаваната стойност на дисперсията

$$\overline{s}^2_{w_1-w_2} = 0.093 \cdot 0.907 \left(\frac{1}{175} + \frac{1}{200} = \right) \approx 0.0009,$$

след което от (14.8) намираме наблюдаваната стойност на критерия:

$$Z_0 = \frac{0.017 - 0.160}{\sqrt{0.0009}} \approx -4.77.$$

Според вида на алтернативната хипотеза критичната област е лява едностранна и се определя от неравенството $Z \leq u_\alpha$. От Таблица 4 определяме квантила $u_\alpha = u_{0.05} = -u_{0.95} = -1.96$. Понеже $Z_0 = -4.77 < -1.96$, то Z_0 попада в критичната област и хипотезата H_0 се отхвърля. Може да приемем, че вероятността p_1 е значимо по-малка от вероятността p_2, т.е. лекарството е ефективно.

14.7 Проверка на хипотези с използване на доверителен интервал

Проверката на статистически параметрични хипотези с използване на критични области може да се извърши и чрез построяването на доверителен интервал. За всички параметрични хипотези, разгледани в раздел 14.2 , областта $\overline{W}$ на приемане на основната хипотеза $H_0 : \theta = \theta_0$ при ниво на значимост α съвпада с доверителния интервал за параметъра θ при доверителна вероятност $p = 1 - \alpha$. При това на едностранните критични области съответстват едностранните доверителни интервали, а на двустранната критична област съответства двустранният доверителен интервал. Основната хипотеза H_0 се приема, ако стойността θ_0 попада в съответния доверителен интервал; в противен случай хипотезата H_0 се отхвърля.

Ако се проверява основната хипотеза $H_0 : \theta_1 = \theta_2$, то се търси съответният доверителен интервал за разликата $\theta = \theta_1 - \theta_2$ и хипотезата H_0 се приема, ако нулата попада в този интервал и се отхвърля, ако не попада в него. Изключение имаме при проверката на хипотезата за равенство на две дисперсии $H_0 : \sigma_1^2 = \sigma_2^2$. Тогава се търси доверителният интервал за отношението $\theta = \sigma_1^2/\sigma_2^2$ на двете дисперсии и хипотезата H_0 се приема, ако единицата попада в този интервал и се отхвърля, ако не попада в него.

Пример 14.6. *По извадка с обем* $n = 16$, *извлечена от нормално разпределена генерална съвкупност* $X = N(m, \sigma)$, *са пресметнати статистическото средно* $\overline{x} = 118.2$ *и неизместеното статистическо стандарт-*

но отклонение $\overline{s} = 3.6$. *При ниво на значимост* $\alpha = 0.05$ *да се провери хипотезата* $H_0 : m = 120$ *при алтернативна хипотеза* $H_1 : m \neq 120$.

Решение: Понеже дисперсията σ^2 *е неизвестна, ще процедираме както в точка* 14.2.1.2*. Определяме наблюдаваната стойност на критерия:*

$$Z_0 = \frac{\overline{x} - m_0}{\overline{s}/\sqrt{n}} = \frac{(118.2 - 120)\sqrt{16}}{3.6} = \frac{-1.8 \cdot 4}{3.6} = -2.$$

Според алтернативната хипотеза критичната област е двустранна: $|Z| \geq t_{1-\alpha/2}(n-1)$. *От Таблица* 6 *определяме* $t_{1-\alpha/2}(n-1) = t_{0.975}(15) = 2.131$. *Понеже* $|Z_0| = 2 < 2.131$, *то* Z_0 *не попада в критичната област и хипотезата* H_0 *не се отхвърля.*

Да проверим хипотезата H_0 *с използване на доверителен интервал.*

Понеже дисперсията σ^2 *е неизвестна, ще определим доверителния интервал за средното* m, *както в раздел* 12.6*: От Таблица* 6 *определяме квантила* $t_{1-\alpha/2}(n-1) = t_{0.975}(15) = 2.131$*; Пресмятаме точността*

$$\Delta = \frac{\overline{s}}{\sqrt{n}} t_{1-\alpha/2}(n-1) = \frac{3.6 \cdot 2.131}{4} \approx 1.92$$

и намираме доверителния интервал

$I_{0.95} = (118.2 - 1.92, 118.2 + 1.92) = (116.28, 120.12)$.

Понеже $m_0 = 120 \in I_{0.95}$, *то хипотезата* H_0 *не се отхвърля.*