и m_1 (се добиваа различни оптимални критични домени, кога $m_0 < m_1$ и кога $m_0 > m_1$). Тоа значи дека во овој случај не постои рамномерно оптимален критичен домен при тестирање на $H_0 : m = m_0$ против $H_1 : m \neq m_0$.

Но затоа, рамномерно оптималниот критичен домен при тестирање на $H_0 : m = m_0$ против $H_1^+ : m > m_0$ е $C_0 = \{x = (x_1, x_2, ..., x_n) : \overline{x} \geq c_0\}$, каде $c_0 = m_0 + \frac{1}{\sqrt{n}}\Phi_0^{-1}(\frac{1-2\alpha}{2})$. Слично, оптималниот критичен домен при тестирање на $H_0 : m = m_0$ против $H_1^- : m < m_0$ е $C_0 = \{x = (x_1, x_2, ..., x_n) : \overline{x} \leq c_0\}$, каде $c_0 = m_0 + \frac{1}{\sqrt{n}}\Phi_0^{-1}(\frac{2\alpha-1}{2})$.

5.2.3 Тестови со коефициент на подобност

Нека $\mathcal{P} = \{F(x, \theta) : \theta \in \Theta\}$ е фамилијата од допустливи распределби на обележјето X, каде θ е непознат параметар и Θ е просторот од параметри. При тестирање на простата хипотеза $H_0 : \theta = \theta_0$ против простата хипотеза $H_1 : \theta = \theta_1$ одговор на прашањето за најмоќен тест ни дава Лемата на Нејман-Пирсон. Во случај кога барем една од хипотезите, нултата или алтернативната, е сложена за наоѓање на критична област со големина α се користи **методот со коефициент на подобност** кој претставува еден вид на обопштување на тестот на Нејман-Пирсон.

При тестирање на нултата хипотеза $H_0 : \theta \in \Theta_0$ против алтернативната хипотеза $H_1 : \theta \in \Theta_1$ со методот со коефициент на подобност се користи тест статистиката

$$\lambda(X_1, ..., X_n) = \frac{\max_{\theta \in \Theta_0} L(X_1, ..., X_n; \theta)}{\max_{\theta \in \Theta} L(X_1, ..., X_n; \theta)} \tag{5.9}$$

која се нарекува **коефициент на подобност** (LR - likelihood ratio). Да забележиме дека ако $\widehat{\theta}$ е такво да $\max_{\theta\in\Theta} L(X_1, ..., X_n; \theta) = L(X_1, ..., X_n; \widehat{\theta})$, тогаш $\widehat{\theta}$ е ML оценувач за параметарот θ. Исто така, ако H_0 е проста хипотеза т.е. $\Theta_0 = \{\theta_0\}$, тогаш $\max_{\theta\in\Theta_0} L(X_1, ..., X_n; \theta) = L(X_1, ..., X_n; \theta_0)$.

За коефициентот на подобност важи

$$0 \leq \lambda(x_1, ..., x_n) \leq 1.$$

Веројатноста дека вредноста на $\lambda(x_1, ..., x_n)$ е голема, е поголема тогаш кога хипотезата H_0 е точна и обратно, таа е релативно мала кога H_0 не е точна. Затоа, разумно е критичната област C да се одбере така да ги содржи оние вредности на $x = (x_1, ..., x_n)$ за кои соодветниот коефициент на подобност не е поголем од даден број c $(0 < c \leq 1)$ т.е. има облик

$$C = \{x = (x_1, ..., x_n) : \lambda(x_1, ..., x_n) \leq c\}.$$

Постапката за наоѓање на критичен домен со големина α со метод со коефициент на подобност е да при дадени вредности на големината на примерокот

n и нивото на значајност на тестот α се наоѓа вредноста на константата c од условот да

$$\max_{\theta\in\Theta_0} P\{\lambda(X_1,...,X_n)\leq c\}=\alpha.$$

Кога нултата хипотеза H_0 е проста, овој услов преминува во

$$P\{\lambda(X_1,...,X_n)\leq c|\theta=\theta_0\}=P_0\{\lambda(X_1,...,X_n)\leq c\}=\alpha.$$

5.2.4 Тестови за параметрите на нормална распределба

Методот со коефициент на подобност овозможува да на едноставен начин се определат некои од тестовите за параметрите на нормална распределба $\mathcal{N}(m,\sigma^2)$. Претпоставуваме дека се зададени големината на примерокот n и нивото на значајност α.

Во случај на две независни нормално распределени обележја X и Y со распределби $\mathcal{N}(m_1,\sigma_1^2)$ и $\mathcal{N}(m_2,\sigma_2^2)$ соодветно, претпоставуваме дека се зададени големините на примероците n_1 и n_2 соодветно и нивото на значајност α.

(1) Тестирање на $H_0: m=m_0$, против $H_1: m\neq m_0$, кога σ^2 е позната.

Кога $X\sim\mathcal{N}(m,\sigma^2)$ при σ^2 познато, функцијата на подобност е

$$L(x,m)=(2\pi\sigma^2)^{-n/2}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-m)^2\},\ x=(x_1,...,x_n)\in\mathbb{R}^n,\ m\in\mathbb{R}.$$

При σ^2 познато ML оценувач за m е $\overline{X}_n$, односно $\mathcal{L}(m)=L(x,m)$ достигнува максимум во $m=\overline{x}=\frac{1}{n}(x_1+...+x_n)$. И бидејќи нултата хипотеза H_0 е проста (т.е. $\Theta_0=\{m_0\}$), за коефициентот на подобност имаме

$$\lambda(x_1,...,x_n)=\frac{L(x,m_0)}{L(x,\overline{x})}=$$

$$=\exp\{-\frac{1}{2\sigma^2}(\sum_{i=1}^n(x_i-m_0)^2-\sum_{i=1}^n(x_i-\overline{x})^2)\}=\exp\{-\frac{n}{2\sigma^2}(\overline{x}-m_0)^2)\}.$$

Според методот со коефициент на подобност, критичната област C ги содржи сите оние $x=(x_1,...,x_n)\in\mathbb{R}^n$ за кои

$$\exp\{-\frac{n}{2\sigma^2}(\overline{x}-m_0)^2)\}\leq c,$$

каде $0<c\leq 1$. Последното неравенство е еквивалентно со

$$\left|\frac{\overline{x}-m_0}{\sigma}\sqrt{n}\right|\geq c_0,$$

каде $c_0 = \sqrt{-2\ln c} \geq 0$, односно критичната област може да се запише во обликот
$$C = \{x = (x_1, ..., x_n) \in \mathbb{R}^n : \Big|\frac{\overline{x} - m_0}{\sigma}\sqrt{n}\Big| \geq c_0\}.$$

Вредноста на бројот c_0 ја одредуваме од даденото ниво на значајност на тестот α, односно од условот
$$P_0\{\Big|\frac{\overline{X}_n - m_0}{\sigma}\sqrt{n}\Big| \geq c_0\} = \alpha.$$

Повторно користевме дека нултата хипотеза H_0 е проста. Сега, бидејќи статистиката
$$Z = \frac{\overline{X}_n - m_0}{\sigma}\sqrt{n} \sim \mathcal{N}(0,1),$$
имаме дека
$$\alpha = P_0\{|Z| \geq c_0\} = 1 - P_0\{-c_0 < Z < c_0\} = 1 - (\Phi_0(c_0) - \Phi_0(-c_0)) = 1 - 2\Phi_0(c_0),$$
каде $\Phi_0(c_0) = \int_0^{c_0} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$ е Лапласовиот интеграл. Значи, за c_0 имаме дека е бројот
$$c_0 = \Phi_0^{-1}(\frac{1-\alpha}{2}).$$

На сличен начин се добива дека критичните области при тестирање на

(а) $H_0 : m = m_0$, против $H_1 : m > m_0$, кога σ^2 е позната,

(б) $H_0 : m = m_0$, против $H_1 : m < m_0$, кога σ^2 е позната,

соодветно се дадени со

(а) $C = \{x = (x_1, ..., x_n) \in \mathbb{R}^n : \frac{\overline{x} - m_0}{\sigma}\sqrt{n} \geq c_0\},\ c_0 = \Phi_0^{-1}(\frac{1-2\alpha}{2}),$

(б) $C = \{x = (x_1, ..., x_n) \in \mathbb{R}^n : \frac{\overline{x} - m_0}{\sigma}\sqrt{n} \leq -c_0\},\ c_0 = \Phi_0^{-1}(\frac{1-2\alpha}{2}).$

(2) Тестирање на $H_0 : m = m_0$, против $H_1 : m \neq m_0$, кога σ^2 не е позната.

Кога дисперзијата σ^2 не е позната, нултата хипотеза може да се запише како $H_0 : (m, \sigma^2) = (m_0, \sigma^2)$, односно тогаш $\Theta_0 = \{(m_0, \sigma^2) : \sigma^2 > 0\}$, каде m_0 е фиксен реален број. Тогаш, задачата преминува во тестирање на сложена хипотеза $H_0 : t = (m, \sigma^2) \in \Theta_0$ против сложена хипотеза $H_1 : t = (m, \sigma^2) \in \Theta_1 = \Theta \setminus \Theta_0$, каде множеството допустливи вредности за параметрите на нормалната распределба е $\Theta = \{(m, \sigma^2) : m \in \mathbb{R},\ \sigma^2 > 0\}$.

Имајќи во предвид дека функцијата на подобност е

$$L(x, m, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2\},$$

за $x = (x_1, ..., x_n) \in \mathbb{R}^n,\ m \in \mathbb{R},\ \sigma^2 > 0$, и бидејќи ML оценувачи за m и σ^2 се $\overline{X}_n$ и $\overline{S}_n^2$ соодветно, имаме дека

$$\max_{t\in\Theta} L(x, t) = \max_{m\in\mathbb{R},\, \sigma^2>0} L(x, m, \sigma^2) = L(x, \overline{x}, s^2),$$

каде $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$, и дека

$$\max_{t\in\Theta_0} L(x, t) = \max_{\sigma^2>0} L(x, m_0, \sigma^2) = L(x, m_0, \sigma_0^2),$$

каде $\sigma_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m_0)^2$. Тогаш, коефициентот на подобност е

$$\lambda(x_1, ..., x_n) = \frac{L(x, m_0, \sigma_0^2)}{L(x, \overline{x}, s^2)} = \Big(\frac{\sum_{i=1}^{n}(x_i - m_0)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\Big)^{-n/2}.$$

Според методот со коефициент на подобност, критичната област C ги содржи сите $x = (x_1, ..., x_n) \in \mathbb{R}^n$ за кои важи

$$\Big(\frac{\sum_{i=1}^{n}(x_i - m_0)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\Big)^{-n/2} \leq c,$$

каде $0 < c \leq 1$. Последното неравенство е еквивалентно со

$$\frac{(\overline{x} - m_0)^2}{s^2} \geq c^{-2/n} - 1,$$

односно со

$$\Big|\frac{\overline{x} - m_0}{s}\sqrt{n-1}\Big| \geq c_0,$$

каде $c_0 = \sqrt{(n-1)(c^{-2/n} - 1)} \geq 0$, па критичната област е

$$C = \{x = (x_1, ..., x_n) \in \mathbb{R}^n : \Big|\frac{\overline{x} - m_0}{s}\sqrt{n-1}\Big| \geq c_0\},$$

каде c_0 се одредува од даденото ниво на значајност α. Познато ни е дека статистиката

$$T = \frac{\overline{X}_n - m_0}{\overline{S}_n}\sqrt{n-1} \sim t_{n-1}.$$

Затоа од

$$\alpha = P_0\{|T| \geq c_0\},$$

имаме дека

$$c_0 = t_{n-1,\alpha},$$

каде означуваме со $t_{n,\alpha}$ број (кој се чита од таблица) за кој важи $P\{|t_n| > t_{n,\alpha}\} = \alpha$, каде t_n е случајна променлива со студентова распределба со n степени на слобода.

На сличен начин се добива дека критичните области при тестирање на

(а) $H_0 : m = m_0$, против $H_1 : m > m_0$, кога σ^2 не е позната,

(б) $H_0 : m = m_0$, против $H_1 : m < m_0$, кога σ^2 не е позната,

соодветно се дадени со

$$\text{(а) } C = \{x = (x_1, ..., x_n) \in \mathbb{R}^n : \frac{\overline{x} - m_0}{s}\sqrt{n-1} \geq c_0\},\ c_0 = t_{n-1,2\alpha},$$

$$\text{(б) } C = \{x = (x_1, ..., x_n) \in \mathbb{R}^n : \frac{\overline{x} - m_0}{s}\sqrt{n-1} \leq -c_0\},\ c_0 = t_{n-1,2\alpha}.$$

(3) Тестирање на $H_0 : \sigma^2 = \sigma_0^2$, против $H_1 : \sigma^2 \neq \sigma_0^2$.

Покрај постапката на одредување на критичната област C со помош на коефициенттот на подобност, при тестирање на хипотези за параметрите на нормална распределба, може да се примени и постапка со интервал на доверба. Тогаш, при тестирање на $H_0 : \theta = \theta_0$ против $H_1 : \theta \neq \theta_0$ со ниво на значајност α, хипотезата H_0 се прифаќа ако врз основа на податоците $x = (x_1, ..., x_n)$ вредноста θ_0 е обфатена со интервалот $(L(x), U(x))$, каде $(L(X_1, ..., X_n), U(X_1, ..., X_n))$ е $100(1-\alpha)\%$ интервал на доверба за θ, во спротивно H_0 се отфрла. Значи, критичната област има облик

$$C = \{x = (x_1, ..., x_n) \in \mathbb{R}^n : L(x) \geq \theta_0 \text{ или } U(x) \leq \theta_0\}.$$

При тоа, ако статистиките $L(X_1, ..., X_n)$ и $U(X_1, ..., X_n)$ го определуваат интервалот на доверба за θ со минимална должина, тогаш големината на вака дефинираната критична област е α, имено важи

$$P_0\{(X_1, ..., X_n) \in C\} = 1 - P_0\{L(X_1, ..., X_n) < \theta_0 < U(X_1, ..., X_n)\} = 1 - (1-\alpha) = \alpha.$$

Така, знаеме дека $100(1-\alpha)\%$ интервал на доверба за σ^2 е

$$(\frac{n\overline{S}_n^2}{b}, \frac{n\overline{S}_n^2}{a}),$$

каде $a = \chi_{n-1,1-\alpha/2}$ и $b = \chi^2_{n-1,\alpha/2}$, при што со $\chi^2_{n,\alpha}$ го означуваме бројот (кој се чита од таблица) за кој важи $P\{\chi^2_n > \chi^2_{n,\alpha}\} = \alpha$, каде χ^2_n е случајна променлива со χ^2 распределба со n степени на слобода. Тогаш, критичната област е

$$C = \{x = (x_1, ..., x_n) \in \mathbb{R}^n : \frac{ns^2}{\sigma_0^2} \geq b \text{ или } \frac{ns^2}{\sigma_0^2} \leq a\},$$

каде $a = \chi_{n-1,1-\alpha/2}$ и $b = \chi^2_{n-1,\alpha/2}$.

На сличен начин се добива дека критичните области при тестирање на

$$\text{(а) } H_0 : \sigma^2 = \sigma_0^2, \text{ против } H_1 : \sigma^2 > \sigma_0^2,$$

$$\text{(б) } H_0 : \sigma^2 = \sigma_0^2, \text{ против } H_1 : \sigma^2 < \sigma_0^2,$$

соодветно се дадени со

$$\text{(а) } C = \{x = (x_1, ..., x_n) \in \mathbb{R}^n : \frac{ns^2}{\sigma_0^2} \geq c\},\ c = \chi^2_{n-1,\alpha},$$

$$\text{(б) } C = \{x = (x_1, ..., x_n) \in \mathbb{R}^n : \frac{ns^2}{\sigma_0^2} \leq c\},\ c = \chi^2_{n-1,1-\alpha}.$$

(4) Тестирање на $H_0 : m_1 = m_2$, против $H_1 : m_1 \neq m_2$, кога σ_1^2 и σ_2^2 се познати.

Повторно користиме метод со интервал на доверба, овој пат за разликата $m_1 - m_2$. Прво, за централна статистика ја земаме статистиката

$$Z = \frac{(\overline{X}_{n_1} - m_1) - (\overline{Y}_{n_2} - m_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1),$$

(покажи!). Заради симетричноста на распределбата на Z, значи дека бараме таков број c така што

$$1 - \alpha = P\{|Z| < c|H_0\} = P_0\{\left|\frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right| < c\},$$

од каде следи дека

$$c = \Phi_0^{-1}(\frac{1-\alpha}{2})$$

и критичната област е

$$C = \{(x, y) \in \mathbb{R}^{n_1+n_2} : \left|\frac{\overline{x} - \overline{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right| \geq c\}.$$

Ирена Стојковска

На сличен начин се добива дека критичните области при тестирање на

(а) $H_0 : m_1 = m_2$, против $H_1 : m_1 > m_2$, кога σ_1^2 и σ_2^2 се познати,

(б) $H_0 : m_1 = m_2$, против $H_1 : m_1 < m_2$, кога σ_1^2 и σ_2^2 се познати,

соодветно се дадени со

$$(а)\ C = \{(x, y) \in \mathbb{R}^{n_1+n_2} : \frac{\overline{x} - \overline{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \geq c\},\ c = \Phi_0^{-1}(\frac{1 - 2\alpha}{2}),$$

$$(б)\ C = \{(x, y) \in \mathbb{R}^{n_1+n_2} : \frac{\overline{x} - \overline{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq -c\},\ c = \Phi_0^{-1}(\frac{1 - 2\alpha}{2}).$$

(5) Тестирање на $H_0 : m_1 = m_2$, против $H_1 : m_1 \neq m_2$, кога $\sigma_1^2 = \sigma_2^2 = \sigma^2$ не се познати.

Овој пат за одредување на интервал на доверба за $m_1 - m_2$ за централна статистика се користи статистиката

$$T = \frac{(\overline{X}_{n_1} - m_1) - (\overline{Y}_{n_2} - m_2)}{\sqrt{n_1 \overline{S}_x^2 + n_2 \overline{S}_y^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}(n_1 + n_2 - 2)} \sim t_{n_1+n_2-2},$$

(види Теорема 3.5). Па, заради симетричноста на распределбата на T при одредување на интервал на доверба со минимална должина бараме број c така што

$$1 - \alpha = P\{|T| < c|H_0\} = P_0\{\Big|\frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sqrt{n_1 \overline{S}_x^2 + n_2 \overline{S}_y^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}(n_1 + n_2 - 2)}\Big| < c\},$$

од каде следи дека

$$c = t_{n_1+n_2-2,\, \alpha}$$

и критичната област е

$$C = \{(x, y) \in \mathbb{R}^{n_1+n_2} : \Big|\frac{\overline{x} - \overline{y}}{\sqrt{n_1 s_x^2 + n_2 s_y^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}(n_1 + n_2 - 2)}\Big| \geq c\}.$$

На сличен начин се добива дека критичните области при тестирање на

(а) $H_0 : m_1 = m_2$, против $H_1 : m_1 > m_2$, кога $\sigma_1^2 = \sigma_2^2 = \sigma^2$ не се познати,

(б) $H_0 : m_1 = m_2$, против $H_1 : m_1 < m_2$, кога $\sigma_1^2 = \sigma_2^2 = \sigma^2$ не се познати, соодветно се дадени со

$$\text{(а) } C = \{(x,y) \in \mathbb{R}^{n_1+n_2} : \frac{\overline{x}-\overline{y}}{\sqrt{n_1 s_x^2 + n_2 s_y^2}} \sqrt{\frac{n_1 n_2}{n_1+n_2}(n_1+n_2-2)} \geq c\},$$

$$\text{каде } c = t_{n_1+n_2-2,\, 2\alpha},$$

$$\text{(б) } C = \{(x,y) \in \mathbb{R}^{n_1+n_2} : \frac{\overline{x}-\overline{y}}{\sqrt{n_1 s_x^2 + n_2 s_y^2}} \sqrt{\frac{n_1 n_2}{n_1+n_2}(n_1+n_2-2)} \leq -c\},$$

$$\text{каде } c = t_{n_1+n_2-2,\, 2\alpha}.$$

(6) Тестирање на $H_0 : \sigma_1^2 = \sigma_2^2$, против $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Како централни статистики при одредување на интервал на доверба за σ_1^2/σ_2^2 се користат статистиките

$$F_1 = \frac{n_1(n_2-1)\sigma_2^2 \overline{S}_x^2}{n_2(n_1-1)\sigma_1^2 \overline{S}_y^2} \sim F_{n_1-1,\, n_2-1}, \text{ и } F_2 = \frac{n_2(n_1-1)\sigma_1^2 \overline{S}_y^2}{n_1(n_2-1)\sigma_2^2 \overline{S}_x^2} \sim F_{n_2-1,\, n_1-1},$$

(види Теорема 3.6). Имено, ако $s_x^2 > s_y^2$ тогаш критичната област е

$$C = \{(x,y) \in \mathbb{R}^{n_1+n_2} : \frac{n_1(n_2-1)s_x^2}{n_2(n_1-1)s_y^2} \geq c\}, \ c = F_{n_1-1,\, n_2-1,\, \alpha/2},$$

и ако $s_x^2 < s_y^2$ тогаш критичната област е

$$C = \{(x,y) \in \mathbb{R}^{n_1+n_2} : \frac{n_2(n_1-1)s_y^2}{n_1(n_2-1)s_x^2} \geq c\}, \ c = F_{n_2-1,\, n_1-1,\, \alpha/2}.$$

При тоа со $F_{n_1,\, n_2,\, \alpha}$ се означува бројот (кој се чита од таблица) за која важи $P\{F_{n_1,\, n_2} > F_{n_1,\, n_2,\, \alpha}\} = \alpha$, каде $F_{n_1,\, n_2}$ е случајна променлива која има Фишерова рспределба со n_1, n_2 степени на слобода.

На сличен начин се добива дека критичните области при тестирање на

$$\text{(а) } H_0 : \sigma_1^2 = \sigma_2^2, \text{ против } H_1 : \sigma_1^2 > \sigma_2^2,$$

$$\text{(б) } H_0 : \sigma_1^2 = \sigma_2^2, \text{ против } H_1 : \sigma_1^2 < \sigma_2^2,$$

соодветно се дадени со

$$\text{(а) } C = \{(x,y) \in \mathbb{R}^{n_1+n_2} : \frac{n_1(n_2-1)s_x^2}{n_2(n_1-1)s_y^2} \geq c\}, \ c = F_{n_1-1,\, n_2-1,\, \alpha},$$

$$\text{(б) } C = \{(x,y) \in \mathbb{R}^{n_1+n_2} : \frac{n_2(n_1-1)s_y^2}{n_1(n_2-1)s_x^2} \geq c\}, \ c = F_{n_2-1,\, n_1-1,\, \alpha}.$$