3.3 Емпириска функција на распределба

Основен проблем при статистичките истражувања е одредување на распределбата на обележјето X врз основа на даден примерок. Во математичката статистика постои теорема (позната како централна теорема на математичката статистика) која дава потврден одговор на прашањето дали простиот случаен примерок може да даде комплетна информација за распределбата на обележјето X. При тоа, точното одредување на распределбата на обележјето X бара големината n на примерокот неограничено да расте. Бидејќи во пракса може да се работи само со конечна големина на примерокот, распределбата на обележјето X може да се одреди само приближно, и колку е n поголемо толку е апроксимацијата поточна.

Начинот на кој примерокот ја одредува распределбата на обележјето X е даден преку неговата емпириска функција на распределба.

Дефиниција 3.1. Нека $(X_1, X_2, ..., X_n)$ е примерок кој одговара на обележје со функција на распределба $F(x)$. Функцијата $F_n : \mathbb{R} \to \chi$, каде χ е множеството од случајни променливи дефинирана со

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\{X_i < x\},$$

каде $I\{\cdot\}$ е индикатор на настан, се нарекува **емпириска функција на распределба** на примерокот $(X_1, X_2, ..., X_n)$.

Бидејќи сумата на n независни и еднакво распределени случајни променлви со Бернулиеви 0,1 распределби со параметар p, $0 < p < 1$ има $\mathcal{B}(n,p)$ распределба, заклучуваме дека случајната променлива $nF_n(x)$ има $\mathcal{B}(n,p)$ распределба каде $p = P\{X_i < x\} = F(x)$. Од тука следува дека распределбата на веројатност на емпириската функција на распределба е

$$P\{F_n(x) = \frac{k}{n}\} = \binom{n}{k} F^k(x)(1-F(x))^{n-k},\ k = 0, 1, ..., n.$$

Нека $(x_1, x_2, ..., x_n)$ е реализација на примерокот $(X_1, X_2, ..., X_n)$. Ги подредуваме броевите $x_1, x_2, ..., x_n$ во растечки редослед, и добиената низа ја означуваме со $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$. Тогаш, **реализација на емпириската функција на распределба** на примерокот $(X_1, X_2, ..., X_n)$ која одговара на реализацијата на примерокот $(x_1, x_2, ..., x_n)$ е дадена со

$$F_n(x) = \begin{cases} 0 & , x \leq x_{(1)} \\ \frac{k}{n} & , x_{(k)} < x \leq x_{(k+1)},\ k = 1, 2, ..., n-1 \\ 1 & , x > x_{(n)} \end{cases}$$

Ирена Стојковска

Се согледува дека реализацијата на емпириската функција на распределба се поклопува со функцијата на кумулативните релативни честоти на статистичките податоци $x_1, x_2, ..., x_n$.

Значењето на емпириската функција на распределба F_n се состои во тоа што за големи вредности на n таа претставува добра апроксимација на функцијата на распределба F на обележјето X, што е покажано со следните две тврдења.

Теорема 3.1. *Нека F е функцијата на распределба на обележјето X и F_n е емпириската функција на распределба на примерокот $(X_1, X_2, ..., X_n)$ кој одговара на обележјето X. Тогаш, за секој реален број x важи*

$$P\{\lim_{n\to\infty} F_n(x) = F(x)\} = 1.$$

Доказ. Нека $x \in \mathbb{R}$ е произволен реален број. Дефинираме случајни променливи $Y_i = I\{X_i < x\}$, $i = 1, 2, ..., n$, кои се независни и еднакво распределени (од $X_1, ..., X_n$ независни и еднакво распределени) со конечни математички очекувања $E(Y_i) = P\{X_i < x\} = F(x) < +\infty$. Следи дека за низата $\{Y_i\}$ важи силниот закон на големите броеви, т.е.

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \stackrel{a.s.}{\to} \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = F(x),$$

и имајќи ја во предвид дефиницијата на емпириската функција на распределба на промерокот $(X_1, X_2, ..., X_n)$, $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\{X_i < x\} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, следи тврдењето на теоремата т.е. $P\{\lim_{n\to\infty} F_n(x) = F(x)\} = 1$. ■

Теорема 3.2. (Централна теорема на математичката статистика на Гливенко - Кантели) *Нека F е функцијата на распределба на обележјето X и F_n е емпириската функција на распределба на примерокот $(X_1, X_2, ..., X_n)$ кој одговара на обележјето X. Тогаш, важи*

$$P\{\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0\} = 1.$$

Доказ. Нека F е непрекината функција на распределба. Нека $\varepsilon > 0$. Тогаш, постои $m \in \mathbb{N}$ така што $\frac{1}{m} \leq \varepsilon$. Нека $-\infty = t_0 < t_1 < ... < t_{m-1} < t_m = +\infty$ се такви да $F(t_k) = \frac{k}{m}$, $k = 0, 1, 2, ..., m$ (постојат затоа што $F(x)$ е неопаѓачка). Тогаш, за $x \in (t_k, t_{k+1}]$ важи

$$F_n(x) - F(x) \leq F_n(t_{k+1}) - F(t_k) \leq F_n(t_{k+1}) - F(t_{k+1}) + \varepsilon, \tag{3.1}$$

$$F_n(x) - F(x) \geq F_n(t_k) - F(t_{k+1}) \geq F_n(t_k) - F(t_k) - \varepsilon, \tag{3.2}$$

затоа што $F(t_{k+1}) = \frac{k+1}{m} = \frac{k}{m} + \frac{1}{m} \leq F(t_k) + \varepsilon$.

Нека $A_k = \{w \in \Omega : \lim_{n\to\infty} F_n(t_k) = F(t_k)\}$, $k = 0, 1, 2, ..., m$ и $A = \cap_{k=0}^{m} A_k$. Тогаш, од Теорема 3.1 следи дека $P(A_k) = 1$, $k = 0, 1, 2, ..., m$, од каде и $P(A) = 1$. Тогаш, $\forall w \in A$, $\exists n_0 = n_0(w)$, така што $\forall n \geq n_0(w)$,

$$|F_n(t_k) - F(t_k)| \leq \varepsilon,\ k = 0, 1, 2, ..., m. \tag{3.3}$$

Од (3.1)-(3.3) следи дека $\forall w \in A$, $\exists n_0 = n_0(w)$, така што $\forall n \geq n_0(w)$,

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \leq 2\varepsilon, \tag{3.4}$$

од каде добиваме дека важи тврдењето на теоремата кога F е непрекината функција.

Нека сега, F е произволна функција на распределба (F е неопаѓачка и непрекината од лево). Нека $\varepsilon > 0$. Тогаш, постои $m \in \mathbb{N}$ и низа $-\infty = t_0 < t_1 < ... < t_{m-1} < t_m = +\infty$ така што

$$F(t_{k+1}) - F(t_k + 0) \leq \varepsilon. \tag{3.5}$$

Б.Г.О нека $\{t_1, t_2, ..., t_{m-1}\}$ ги содржи сите точки во кои F прави скок не помал од $\varepsilon/2$. Тогаш, за $x \in (t_k, t_{k+1}]$ важи

$$F_n(x) - F(x) \leq F_n(t_{k+1}) - F(t_k + 0) \leq F_n(t_{k+1}) - F(t_{k+1}) + \varepsilon, \tag{3.6}$$

$$F_n(x) - F(x) \geq F_n(t_k + 0) - F(t_{k+1}) \geq F_n(t_k + 0) - F(t_k + 0) - \varepsilon. \tag{3.7}$$

Нека $A_k = \{w \in \Omega : \lim_{n\to\infty} F_n(t_k + 0) = F(t_k + 0)\}$, $B_k = \{w \in \Omega : \lim_{n\to\infty} F_n(t_k) = F(t_k)\}$, $k = 0, 1, 2, ..., m$ и $A = \cap_{k=0}^{m} A_k B_k$. Од Теорема 3.1 следи дека $P(A_k) = P(B_k) = 1$, $k = 0, 1, 2, ..., m$, од каде и $P(A) = 1$. Значи, $\forall w \in A$, $\exists n_1 = n_1(w)$, така што $\forall n \geq n_1(w)$,

$$F_n(t_k + 0) - F(t_k + 0) \geq -\varepsilon,\ k = 0, 1, 2, ..., m, \tag{3.8}$$

$$F_n(t_k) - F(t_k) \leq \varepsilon,\ k = 0, 1, 2, ..., m. \tag{3.9}$$

Од (3.5)-(3.9) следи (3.4), од каде се добива дека важи тврдењето на теоремата и кога F е произволна функција на распределба. ∎

3.4 Статистики

Решавањето на проблемот на наоѓање на непознатата распределба на обележјето X, подразбира и одредување на некои карактеристики на обележјето. За таа цел се служиме со случајни променливи кои се функции од примерокот.

Ирена Стојковска

Дефиниција 3.2. Нека $(X_1, X_2, ..., X_n)$ е примерок кој одговара на обележје X со функција на распределба $F(x)$ и нека $f : \mathbb{R}^n \to \mathbb{R}$, е Борелова функција. Тогаш, случајната променлива $f(X_1, X_2, ..., X_n)$ се нарекува **статистика**.

Пример 3.8. Нека $(X_1, X_2, ..., X_n)$ е примерок кој одговара на обележје X со функција на распределба $F(x)$.

а) Статистиката $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ се нарекува **средина на примерокот**.

б) Статистиката $\overline{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2$ се нарекува **дисперзија на примерокот**, додека $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{n}{n-1}\overline{S}_n^2$ е позната како **корегирана дисперзија на примерокот**.

в) Статистиката $Z_{n,k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ се нарекува **k-ти момент на примерокот**, додека $M_{n,k} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^k$ се нарекува **k-ти централен момент на примерокот**. Лесно се воочува дека $Z_{n,1} = \overline{X}_n$ и $M_{n,2} = \overline{S}_n^2$

г) За примерокот $(X_1, X_2, ..., X_n)$ ја форимираме соодветната **варијациона низа** $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ за која важи дека за секоја реализација $(x_1, x_2, ..., x_n)$ на примерокот $(X_1, X_2, ..., X_n)$ имаме

$$X_{(1)} = x_{(1)}, X_{(2)} = x_{(2)}, ..., X_{(n)} = x_{(n)},$$

каде $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ е реализацијата $(x_1, x_2, ..., x_n)$ подредена по големина. Тогаш, секој елемент од варијационата низа $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ се нарекува **подредена статистика**. Најмалата и најголемата подредена статистика може да се изразат како функции од примерокот на следниот начин

$$X_{(1)} = \min\{X_1, X_2, ..., X_n\}, \quad X_{(n)} = \max\{X_1, X_2, ..., X_n\}.$$

Пример 3.9. Кога разгледуваме дводимензионално обележје (X, Y), со непознати распределби F_X и F_Y за секое од обележјата X и Y соодветно, тогаш испитувањата ги вршиме со помош на **дводимензионален примерок**

$$(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n).$$

Пример за статистика (функција од дводимензионалниот примерок) е

$$R_{X,Y} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \overline{X}_n\,\overline{Y}_n}{\sqrt{\overline{S}^2_{Xn}\,\overline{S}^2_{Yn}}},$$

каде

$$\overline{S}^2_{Xn} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2,\ \overline{S}^2_{Yn} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2,$$

и се нарекува **коефициент на корелација на примерокот**.

3.5 Карактеристики на некои статистики

Бидејќи статистиките се користат при статистичкото заклучување, потребно е да се знаат нивните бројни карактеристики, точните распределби кога n е конечно (за примерок со мал обем или **мал примерок**) и асимптотското однесување на распределбите при $n \to \infty$ (за примерок со голем обем или **голем примерок**).

Нека X е обележје со непозната функција на распределба F. Да ги означиме бројните карактеристики на обележјето X со

$$EX = m,\ DX = \sigma^2,\ EX^k = m_k,\ E(X-m)^k = \mu_k.$$

Да забележиме дека $m_1 = m$, $\mu_1 = 0$ и $\mu_2 = \sigma^2$. Нека сега $(X_1, X_2, ..., X_n)$ е примерок кој одговара на обележјето X. Тогаш важи следното својство за бројните карактеристики на некои статистики изразени преку бројните карактеристики на обележјето X.

Својство 3.1. *Нека $(X_1, X_2, ..., X_n)$ е примерок кој одговара на обележјето X со $EX = m$, $DX = \sigma^2$, $EX^k = m_k$, $E(X-m)^k = \mu_k$. Тогаш,*

а) $E(\overline{X}_n) = m$, $D(\overline{X}_n) = \frac{\sigma^2}{n}$,

б) $E(\overline{S}^2_n) = \frac{n-1}{n}\sigma^2$, $D(\overline{S}^2_n) = \frac{(n-1)^2}{n^3}\left(\mu_4 - \frac{n-3}{n-1}\mu_2^2\right)$,

в) $E(Z_{n,k}) = m_k$, $D(Z_{n,k}) = \frac{1}{n}(m_{2k} - m_k^2)$.

Доказ. а) $E(\overline{X}_n) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\cdot nm = m$,
$D(\overline{X}_n) = D(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$.

б) Најнапред, $E(X_i^2) = DX_i + (EX_i)^2 = \sigma^2 + m^2$.
$E(\overline{S}^2_n) = E(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - E(\overline{X}^2_n) =$
$\frac{1}{n}\sum_{i=1}^{n}(\sigma^2 + m^2) - (D(\overline{X}_n) + (E(\overline{X}_n))^2) = (\sigma^2 + m^2) - (\frac{\sigma^2}{n} + m^2) = \frac{n-1}{n}\sigma^2$.

в) $E(Z_{n,k}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i^k) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^k) = \frac{1}{n}\sum_{i=1}^{n} m_k = m_k$,
$D(Z_{n,k}) = D(\frac{1}{n}\sum_{i=1}^{n} X_i^k) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i^k) = \frac{1}{n^2}\sum_{i=1}^{n}(EX_i^{2k} - (EX_i^k)^2) =$
$\frac{1}{n^2}\sum_{i=1}^{n}(m_{2k} - (m_k)^2) = \frac{1}{n}(m_{2k} - (m_k)^2)$. ■

Нормалната распределба има важна улога при проучување на случајните појави. Затоа, важно е да се знаат точните распределби на некои статистики, меѓу кои и основните статистики $\overline{X}_n$ и $\overline{S}_n^2$, во случај кога обележјето X има нормална распределба.

Теорема 3.3. *Нека $(X_1, X_2, ..., X_n)$ е примерок кој одговара на обележјето X кое има $\mathcal{N}(m, \sigma^2)$ распределба. Тогаш,*

а) Случајната променлива $\overline{X}_n$ има $\mathcal{N}(m, \frac{\sigma^2}{n})$ распределба, што значи дека случајната променлива $\frac{\overline{X}_n - m}{\sigma}\sqrt{n}$ има $\mathcal{N}(0,1)$ распределба,

б) Случајната променлива $\frac{n\overline{S}_n^2}{\sigma^2}$ има χ^2_{n-1} распределба,

в) Случајните променливи $\overline{X}_n$ и $\overline{S}_n^2$ се независни.

Доказ. а) Случајните променливи X_i, $i = 1, 2, ..., n$ се независни и еднакво распределени т.е. $X_i \sim \mathcal{N}(m, \sigma^2)$, $i = 1, 2, ..., n$, и ако во Својство 1.6 земеме $a_1 = ... = a_n = 1/n$ добиваме дека

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(m, \frac{\sigma^2}{n}). \tag{3.10}$$

Потоа, според Својство 1.4 имаме

$$\frac{\overline{X}_n - m}{\sigma}\sqrt{n} = \frac{\sum_{i=1}^{n}(X_i - m)}{\sqrt{n}\cdot\sigma} \sim \mathcal{N}(0,1).$$

б) Претходно покажавме дека $\overline{X}_n \sim \mathcal{N}(m, \frac{\sigma^2}{n})$, види (3.10), и затоа $\frac{X_i - \overline{X}_n}{\sigma} \sim \mathcal{N}(0,1)$, според Својство 1.6. Сега, бидејќи меѓу случајните променливи $\frac{X_i - \overline{X}_n}{\sigma}$, $i = 1, 2, ..., n$ постои една линеарна врска

$$\sum_{i=1}^{n}\frac{X_i - \overline{X}_n}{\sigma} = \frac{1}{\sigma}(\sum_{i=1}^{n} X_i - n\overline{X}_n) = 0,$$

заклучуваме дека (види ја Забелешка 1.1)

$$\frac{n\overline{S}_n^2}{\sigma^2} = \sum_{i=1}^{n}(\frac{X_i - \overline{X}_n}{\sigma})^2 \sim \chi^2_{n-1}. \tag{3.11}$$

в) Дефинираме $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$ со

$$Y_k = X_k - \overline{X}_n,\ k = 1, 2, ..., n-1,\ Y_n = \overline{X}_n.$$

Тогаш, постои несингуларна матрица M така што

$$\mathbf{Y} = \mathrm{M}\mathbf{X},$$

каде $\mathbf{X} = (X_1, X_2, ..., X_n)$ е случаен вектор од независни и еднакво распределени случајни променливи со нормални распределби. Следи дека и $\mathbf{Y}$ е случаен вектор кој има n-димензионална нормална распределба.

Сега, за $k = 1, 2, ..., n-1$ имаме

$$\begin{aligned} cov(Y_k, Y_n) &= E(Y_kY_n) - E(Y_k)E(Y_n) = E(Y_k\overline{X}_n) - E(X_k - \overline{X}_n)E(\overline{X}_n) = E(Y_k\overline{X}_n) = \\ &= E((X_k - \overline{X}_n)\overline{X}_n) = E(X_k\overline{X}_n - \overline{X}_n^2) = \frac{1}{n}\sum_{i=1}^{n} E(X_kX_i) - E(\overline{X}_n^2) = \\ &= \frac{1}{n}((n-1)m^2 + (\sigma^2 + m^2)) - (\frac{\sigma^2}{n} + m^2) = 0, \end{aligned}$$

од каде следи дека $Y_n = \overline{X}_n$ не зависи од $Y_1, Y_2, ..., Y_{n-1}$. Потоа,

$$n\overline{S}_n^2 = \sum_{k=1}^{n}(X_k - \overline{X}_n)^2 = \sum_{k=1}^{n-1} Y_k^2 + (X_n - \overline{X}_n)^2 = \sum_{k=1}^{n-1} Y_k^2 + (\sum_{k=1}^{n-1} Y_k)^2,$$

што значи дека $\overline{S}_n^2$ зависи само од $Y_1, Y_2, ..., Y_{n-1}$ за кои претходно покажавме дека се независни од $Y_n = \overline{X}_n$. Значи, $\overline{X}_n$ и $\overline{S}_n^2$ се независни случајни променливи, што требаше да се докаже. ∎

Теорема 3.4. *Нека $(X_1, X_2, ..., X_n)$ е примерок кој одговара на обележјето X кое има $\mathcal{N}(m, \sigma^2)$ распределба. Тогаш, случајната променлива*

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

Доказ. Од Теорема 3.3 в), следи дека случајните променливи $\overline{X}_n$ и $\overline{S}_n^2$ се независни. Тогаш, независни се и случајните променливи $\frac{n\overline{S}_n^2}{\sigma^2}$ и $\frac{\overline{X}_n - m}{\sigma/\sqrt{n}}$. Од друга страна заради (3.10) и Својство 1.4 имаме дека

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1). \tag{3.12}$$

Сега, од (3.11) и (3.12), според Својство 1.10, имаме дека

$$\frac{\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{n\overline{S}_n^2}{\sigma^2}/(n-1)}}=\frac{\overline{X}_n-\mu}{\overline{S}_n}\sqrt{n-1}\sim t_{n-1}, \tag{3.13}$$

што требаше да се покаже. ■

Теорема 3.5. *Нека се дадени два независни примероци, првиот $(X_1, X_2, ..., X_n)$ кој одговара на обележјето $X \sim \mathcal{N}(m_X, \sigma^2)$ и вториот $(Y_1, Y_2, ..., Y_k)$ кој одговара на обележјето $Y \sim \mathcal{N}(m_Y, \sigma^2)$. Тогаш,*

а) Случајната променлива $\frac{\overline{X}_n-m_X}{\sigma}$ има $\mathcal{N}(0,\frac{1}{n})$ распределба,

б) Случајната променлива $\frac{\overline{Y}_k-m_Y}{\sigma}$ има $\mathcal{N}(0,\frac{1}{k})$ распределба,

в) Случајните променливи $\frac{n\overline{S}_{Xn}^2}{\sigma^2}\sim\chi^2_{n-1}$ и $\frac{k\overline{S}_{Yk}^2}{\sigma^2}\sim\chi^2_{k-1}$,

г) Случајната променлива $\frac{\overline{X}_n-m_X}{\sigma}-\frac{\overline{Y}_k-m_Y}{\sigma}$ има $\mathcal{N}(0,\frac{1}{n}+\frac{1}{k})$ распределба,

д) Случајната променлива $\frac{1}{\sigma^2}(n\overline{S}_{Xn}^2+k\overline{S}_{Yk}^2)$ има χ^2_{n+k-2} распределба од каде следува дека

$$\frac{(\overline{X}_n-m_X)-(\overline{Y}_k-m_Y)}{\sqrt{n\overline{S}_{Xn}^2+k\overline{S}_{Yk}^2}}\sqrt{\frac{nk}{n+k}(n+k-2)}\sim t_{n+k-2}.$$

Доказ. а)-в) следат директно од Теорема 3.3, г) следи од Својство 1.6.

д) Од в) и Својство 1.7 следи дека

$$\frac{n\overline{S}_{Xn}^2}{\sigma^2}+\frac{k\overline{S}_{Yk}^2}{\sigma^2}\sim\chi^2_{(n-1)+(k-1)}\equiv\chi^2_{n+k-2},$$

потоа од г) следи дека

$$\frac{\frac{\overline{X}_n-m_X}{\sigma}-\frac{\overline{Y}_k-m_Y}{\sigma}}{\sqrt{\frac{1}{n}+\frac{1}{k}}}\sim\mathcal{N}(0,1),$$

и конечно од Својство 1.10 (дефиниција на студентова распределба) имаме

$$\frac{\frac{\frac{\overline{X}_n-m_X}{\sigma}-\frac{\overline{Y}_k-m_Y}{\sigma}}{\sqrt{\frac{1}{n}+\frac{1}{k}}}}{\sqrt{\frac{\frac{n\overline{S}_{Xn}^2}{\sigma^2}+\frac{k\overline{S}_{Yk}^2}{\sigma^2}}{n+k-2}}}\sim t_{n+k-2},$$

од каде следи бараното тврдење. ■

Теорема 3.6. *Нека се дадени два независни примероци, првиот* $(X_1, X_2, ..., X_n)$ *кој одговара на обележјето* $X \sim \mathcal{N}(m_X, \sigma_X^2)$ *и вториот* $(Y_1, Y_2, ..., Y_k)$ *кој одговара на обележјето* $Y \sim \mathcal{N}(m_Y, \sigma_Y^2)$*. Тогаш, случајната променлива*

$$\frac{n(k-1)\sigma_Y^2 \overline{S}_{Xn}^2}{k(n-1)\sigma_X^2 \overline{S}_{Yk}^2} \sim F_{n-1,k-1}.$$

Доказ. Од Теорема 3.5 в) следи дека $\frac{n\overline{S}_{Xn}^2}{\sigma^2} \sim \chi_{n-1}^2$ и $\frac{k\overline{S}_{Yk}^2}{\sigma^2} \sim \chi_{k-1}^2$, и од нивната независност, според Својство 1.11 (дефиниција на Фишерова распределба) имаме

$$\frac{\frac{n\overline{S}_{Xn}^2}{\sigma_X^2}/(n-1)}{\frac{k\overline{S}_{Yk}^2}{\sigma_Y^2}/(k-1)} \sim F_{n-1,k-1},$$

од каде следи бараното тврдење. ■

При големи вредности на обемот n на примерокот $(X_1, X_2, ..., X_n)$ кој одговара на произволно обележје X со конечни математичко очекување m и дисперзија σ^2, распределбата на средината на примерокот $\overline{X}_n$ се стреми кон распределбата на средината на примерокот кој одговара на обележје со нормална $\mathcal{N}(m, \sigma^2)$ распределба. Во пракса доволно е да $n > 30$ па да се користи резултатот од следната теорема.

Теорема 3.7. *Нека* $(X_1, X_2, ..., X_n)$ *е примерок кој одговара на обележјето* X *со конечни математичко очекување* $EX = m < +\infty$ *и дисперзија* $DX = \sigma^2 < +\infty$*. Тогаш,*

$$\overline{X}_n \xrightarrow{\text{с.с.}} m, \ n \to \infty,$$
$$\frac{\sqrt{n}}{\sigma}(\overline{X}_n - m) \xrightarrow{dist.} \mathcal{N}(0,1), \ n \to \infty.$$

Доказ. Првото тврдење, $\overline{X}_n \xrightarrow{\text{с.с.}} m, \ n \to \infty$, следи од законот на големите броеви. Второто тврдење, $\frac{\sqrt{n}}{\sigma}(\overline{X}_n - m) \xrightarrow{\text{dist.}} \mathcal{N}(0,1), \ n \to \infty$, следи од Централната гранична теорема. ■