საქართველოს ტექნიკური უნივერსიტეტი

ელგუჯა ყუბანეიშვილი

მრავალგანზომილებიანი სტატისტიკური მეთოდები მედიცინაში

ლექციების კურსი

თბილისი 2016

მოცემული სახელმძღვანელო წარმოადგენს ბიოსტატისტიკის (http://gtu.ge/books/biostatistika . pdf) სალექციო კურსის გაგრძელებას, სადაც ბიოსტატისტიკიდან განსხვავებით, სადაც განიხილება მხოლოდ ორი ცვლადის სტატისტიკური მეთოდები, აქ წარმოდგენილია მრავალი ცვლადის ის სტატისტიკური მეთოდები, რომლებიც უფრო ხშირად გამოიყენებიან ბიოსამედიცინო ინფორმაციის დამუშავებისათვის. სახელმძღვანელო ძირითადად განკუთვნილია საქართველოს ტექნიკური უნივერსიტეტის ბიოსამედიცინო ინჟინერიის დეპარტამენტის მაგისტრანტების და დოქტორანტებისათვის.

1. შემთხვევით სიდიდეთა სისტემა და განაწილების კანონი

1.1 შემთხვევით სიდიდეთა სისტემა

შემთხვევითი მოვლენების შესწავლისას საქმე გვაქვს არა ერთ, არამედ რამდენიმე შემთხვევით სიდიდესთან, რომლებიც ქმნიან შემთხვევით სიდიდეთა სისტემას. შემთხვევით სიდიდეთა სისტემის განხილვისას ხელსაყრელია გეომეტრიული ინტერპრეტაციის გამოყენება. ასე მაგალითად, ორი შემთხვევითი სიდიდე შეიძლება განვიხილოთ, როგორც შემთხვევითი წერტილი სიბრტყეზე x და y კოორდინატებით, სამი შემთხვევითი სიდიდე, როგორც წერტილი სამგანზომილებიან სივრცეში და ა.შ. ზოგადად, n-რაოდენობის შემთხვევით სიდიდეთა სისტემა შეიძლება განვიხილოთ, როგორც წერტილი n - განზომილებიან სივრცეში ან როგორც n - განზომილებიანი ვექტორი.

შემთხვევით სიდიდეთა სისტემის განხილვისას შემოვიფარგლოთ ორი სიდიდით, რადგან ყველა ის შედეგი, რომელიც მიიღება ორი სიდიდის შემთხვევაში, ადვილად ვრცელდება ნებისმიერი რაოდენობის შემთხვევით სიდიდეებზე. განვიხილოთ ორი შემთხვევითი სიდიდის განაწილების ფუნქცია და განაწილების სიმკვრივის ფუნქცია.

განაწილების ფუნქცია. ორი შემთხვევითი X და Y სიდიდის განაწილების ფუნქცია ეწოდება ორარგუმენტიან $F(x, y)$ ფუნქციას, რომელიც ტოლია ორი $X < x$ და $Y < y$ უტოლობის ერთდროულად შესრულების ალბათობისა. ე.ი.

$$F(x, y) = P(X < x, Y < y)$$

და მას **ერთობლივი განაწილების ფუნქცია** ეწოდება. დაუმტკიცებლად მოვიყვანოთ ერთობლივი განაწილების ფუნქციის ძირითადი თვისებები.

1. $F(x, y)$ ფუნქცია ორივე არგუმენტისათვის ზრდადი ფუნქციაა, ე.ი. როცა $x_2 > x_1$, მაშინ $F(x_2, y) \geq F(x_1, y)$. ასევე, როცა $y_2 > y_1$, მაშინ $F(x, y_2) \geq F(x, y_1)$.

2. თუ $F(x, y)$ ფუნქციის ერთ-ერთი არგუმენტი მიისწრაფვის პლიუს უსასრულობისკენ, მაშინ ერთობლივი განაწილების ფუნქცია მიისწრაფვის

მეორე არგუმენტის შესაბამისი შემთხვევითი სიდიდის განაწილების ფუნქციისაკენ:

$$\lim_{y\to\infty} F(x,y) = F(x,\infty) = F_x(x), \quad \lim_{x\to\infty} F(x,y) = F(\infty,y) = F_y(y),$$

სადაც $F_x(x)$ და $F_y(y)$ უწოდებენ კერძო განაწილების ფუნქციებს.

3. თუ ორივე არგუმენტი მიისწრაფვის პლიუს უსასრულობისკენ, მაშინ ერთობლივი განაწილების ფუნქცია მიისწრაფვის ერთისკენ.

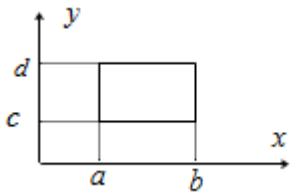
$$\lim_{x,y\to\infty} F(x,y) = 1 \quad \text{ან} \quad F(\infty,\infty) = 1.$$

4. თუ რომელიმე ერთი ან ორივე არგუმენტი მიისწრაფვის მინუს უსასრულობისკენ, მაშინ ერთობლივი განაწილების ფუნქცია მიისწრაფვის ნულისკენ.

$$\lim_{x\to-\infty} F(x,y) = \lim_{y\to-\infty} F(x,y) = \lim_{x,y\to-\infty} F(x,y) = 0,$$

ან $F(-\infty,y) = F(x,-\infty) = F(-\infty,-\infty) = 0.$

5. კოორდინატთა ღერძების პარალელურგვერდებიან ნებისმიერ ოთკუთხედში წერტილის მოხვედრის ალბათობა განისაზღვრება ფორმულით:



$$P(a \le X < b,\ c \le Y < d) = F(b,d) - F(a,d) - F(b,c) + F(a,c).$$

სიმკვრივის ფუნქცია. ზემოთ განხილული განაწილების ფუნქცია წარმოადგენს შემთხვევით სიდიდეთა სისტემის უნივერსალურ მახასიათებელს, რომელიც გამოიყენება როგორც დისკრეტული, ასევე უწყვეტი შემთხვევითი სიდიდეებისთვის. უნდა აღინიშნოს, რომ პრაქტიკული გამოყენება უფრო გააჩნია უწყვეტ შემთხვევით სისტემას, რომლის განაწილება ხასიათდება არა განაწილების ფუნქციით, არამედ განაწილების სიმკვრივით. განაწილების სიმკვრივის ფუნქცია წარმოადგენს სისტემის ამომწურავ მახასიათებელს, რომლის საშუალებითაც სისტემის განაწილების კანონის აღწერა გაცილებით თვალსაჩინოა, ვიდრე განაწილების ფუნქციით.

ორგანზომილებიანი სისტემის განაწილების სიმკვრივის ფუნქცია განისაზღვრება ისევე, როგორც ერთი შემთხვევითი სიდიდის დროს. კერძოდ, თუ ერთობლივი განაწილების ფუნქცია უწყვეტია და ორჯერ დიფერენცირებადი, მაშინ ერთობლივი განაწილების სიმკვრივის ფუნქცია განისაზღვრება შემდეგნაირად:

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = F''(x,y).$$

განვიხილოთ ერთობლივი განაწილების სიმკვრივის ფუნქციის ძირითადი თვისებები:

1. ერთობლივი განაწილების სიმკვრივე დადებითი ფუნქციაა $f(x,y) \ge 0.$

2. ორმაგი ინტეგრალი უსასრულო ზღვრებით ერთობლივი განაწილების სიმკვრივის ფუნქციიდან ერთის ტოლია.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1.$$

3. თუ ცნობილია ერთობლივი განაწილების სიმკვრივის ფუნქცია $f(x, y)$, მაშინ შემთხვევითი წერტილის (X, Y) ნებისმიერ D არეში მოხვედრის ალბათობა განისაზღვრება ფორმულით:

$$P[(X,Y)\in D]=\iint_D f(x,y)dxdy.$$

ერთობლივი განაწილების ფუნქცია შეიძლება გამოვსახოთ განაწილების სიმკვრივის ფუნქციით შემდეგნაირად:

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dxdy.$$

გეომეტრიულად $f(x, y)$ ფუნქცია შეიძლება წარმოვადგინოთ როგორც რაიმე ზედაპირი, რომელსაც განაწილების ზედაპირი ეწოდება.

მაგალითი. მოცემულია ორგანზომილებიანი სისტემის განაწილების სიმკვრივის ფუნქცია

$$f(x,y)=\frac{a}{1+x^2+x^2y^2+y^2}.$$

ვიპოვოთ a, ერთობლივი განაწილების ფუნქცია $f(x, y)$ და R კვადრატში შემთხვევითი წერტილის მოხვედრის ალბათობა.

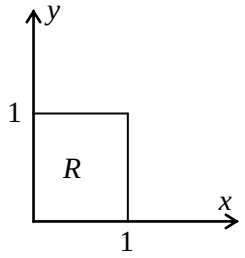
ერთობლივი განაწილების სიმკვრივის მეორე თვისებიდან გამომდინარე გვექნება:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{a}{1+x^2+x^2y^2+y^2}dxdy=a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{dxdy}{(1+x^2)(1+y^2)}=$$

$$=a\int_{-\infty}^{\infty}\frac{dx}{1+x^2}\int_{-\infty}^{\infty}\frac{dy}{1+y^2}=a\cdot\operatorname{arctg}x\Big|_{-\infty}^{\infty}\cdot\operatorname{arctg}y\Big|_{-\infty}^{\infty}=a\pi^2=1.$$

აქედან $a=1/\pi^2$ · ერთობლივი განაწილების ფუნქცია ტოლია:

$$F(x,y)=\frac{1}{\pi^2}\int_{-\infty}^{x}\int_{-\infty}^{y}\frac{dxdy}{(1+x^2)(1+y^2)}=\left(\frac{1}{\pi}\operatorname{arctg}x+\frac{1}{2}\right)\left(\frac{1}{\pi}\operatorname{arctg}y+\frac{1}{2}\right).$$



შემთხვევითი წერტილის R კვადრატში მოხვედრის ალბათობა ტოლია:

$$P[(X,Y)\in D]=\frac{1}{\pi^2}\int_0^1\int_0^1\frac{dxdy}{(1+x^2)(1+y^2)}=$$

$$=\frac{1}{\pi^2}\int_0^1\frac{dx}{1+x^2}\int_0^1\frac{dy}{1+y^2}=\frac{1}{\pi^2}\operatorname{arctg}x\Big|_0^1\cdot\operatorname{arctg}y\Big|_0^1=\frac{1}{\pi^2}\frac{\pi}{4}\frac{\pi}{4}=\frac{1}{16}.$$

1.2 პირობითი განაწილების სიმკვრივის ფუნქცია

შემთხვევით სიდიდეთა სისტემის დახასიათებისთვის არ არის საკმარისი ვიცოდეთ თითოეული შემთხვევითი სიდიდის განაწილების კანონი, საჭიროა ვიცოდეთ მათ შორის დამოკიდებულებაც. ეს დამოკიდებულება შეიძლება დახასიათდეს ე.წ. პირობითი განაწილების კანონით.

სისტემის ერთი X შემთხვევითი სიდიდის განაწილების კანონს, გამოთვლილს იმ პირობით, რომ მეორე, Y შემთხვევითმა სიდიდემ მიიღო გარკვეული $Y = y$ მნიშვნელობა, ეწოდება პირობითი განაწილების კანონი. იგი შეიძლება წარმოვადგინოთ, როგორც პირობითი განაწილების სიმკვრივის ფუნქცია $f(x \mid y)$, ან როგორც პირობითი განაწილების ფუნქცია $F(x \mid y)$. ერთობლივი განაწილების ფუნქციის თვისებიდან გამომდინარე, $F_1(x) = F(x, \infty)$ და $F_2(y) = F(\infty, y)$. აქედან გამომდინარე, კერძო განაწილების სიმკვრივის ფუნქციები ტოლია:

$$f_x(x) = F_1'(x) = \int_{-\infty}^{\infty} f(x,y)dy,$$

$$f_y(y) = F_2'(y) = \int_{-\infty}^{\infty} f(x,y)dx.$$

ადვილად მტკიცდება, რომ სისტემის ერთობლივი განაწილების სიმკვრივის ფუნქცია $f(x, y)$ მიიღება სისტემაში შემავალი ერთი შემთხვევითი სიდიდის კერძო განაწილების სიმკვრივისა $f_x(x)$ და მეორე შემთხვევითი სიდიდის პირობითი განაწილების სიმკვრივის $f(y \mid x)$ ერთმანეთზე გადამრავლებით. ე.ი.

$$f(x,y) = f_x(x)\, f(y \mid x) \quad \text{და} \quad f(x,y) = f_y(y)\, f(x \mid y).$$

ამ ტოლობებს ხშირად განაწილების კანონების გამრავლების თეორემას უწოდებენ. პირობითი განაწილების სიმკვრივის ფუნქციები ასე განისაზღვრება:

$$f(x \mid y) = \frac{f(x,y)}{f_y(y)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dx},$$

$$f(y \mid x) = \frac{f(x,y)}{f_x(x)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dy}.$$

პირობითი განაწილების სიმკვრივეს ახასიათებს იგივე თვისებები, რაც ჩვეულებრივი განაწილების სიმკვრივის ფუნქციას, კერძოდ:

$$\int_{-\infty}^{\infty} f(x \mid y)dx = \int_{-\infty}^{\infty} f(y \mid x)dy = 1.$$

პირობითი განაწილების სიმკვრივის ცნებიდან გამომდინარე, შეგვიძლია შემოვიტანოთ ალბათობის თეორიაში ერთ-ერთი უმნიშვნელოვანესი ცნება – შემთხვევით სიდიდეთა დამოუკიდებლობის ცნება.

X შემთხვევითი სიდიდის დამოუკიდებლობა Y შემთხვევით სიდიდესთან ნებისმიერი y-თვის შეიძლება ასე ჩაიწეროს:

$$f(x \mid y) = f_x(x).$$

თუ შემთხვევითი სიდიდე X დამოკიდებულია Y-ზე, მაშინ:

$$f(x \mid y) \neq f_x(x).$$

თუ X სიდიდე არ არის დამოკიდებული Y-ზე, მაშინ არც Y სიდიდეა დამოკიდებული X-ზე.

ამრიგად, უწყვეტ X და Y შემთხვევით სიდიდეებს ეწოდებათ დამოუკიდებელი, თუ თითოეული მათგანის განაწილების კანონი არ არის დამოკიდებული იმაზე, თუ რა მნიშვნელობა მიიღო მეორე შემთხვევითმა სიდიდემ. წინააღმდეგ შემთხვევაში, შემთხვევითი სიდიდეები დამოკიდებულნი არიან. დამოუკიდებული X და Y შემთხვევითი სიდიდეთა ერთობლივი განაწილების სიმკვრივის ფუნქცია ტოლია:

$$f(x, y) = f_x(x)\ f_y(y).$$

მაგალითი. ვთქვათ, მოცემულია ერთობლივი განაწილების სიმკვრივის ფუნქცია:

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

ეს ტოლობა წარმოვადგინოთ შემდეგნაირად:

$$f(x, y) = \frac{1}{\pi\left(x^2 + 1\right)} \cdot \frac{1}{\pi\left(y^2 + 1\right)}.$$

აქედან ჩანს, რომ შემთხვევითი სიდიდეები X და Y დამოუკიდებელი არიან, რადგან $f(x,y)$ გამოსახულების პირველი მამრავლი დამოკიდებულია მხოლოდ X-ზე, ხოლო მეორე მამრავლი – Y-ზე. ე.ი. $f(x, y) = f_x(x)\ f_y(y)$.

.

1.3 ორგანზომილებიანი შემთხვევითი სისტემის რიცხვითი მახასიათებლები

ორი X და Y შემთხვევით ვექტორთა (k, s) რიგის საწყისი მომენტი ეწოდება X^kY^s სიდიდეთა ნამრავლის მათემატიკურ ლოდინს და აღინიშნება $a_{k,s}$ სიმბოლოთი.

$$a_{ks} = M\left[X^kY^s\right]$$

დისკრეტულ შემთხვევით სიდიდეთა სისტემისათვის გვექნება:

$$a_{ks} = \sum_i \sum_j x_i^k y_j^s p_{ij},$$

სადაც, $p_{ij} = P(X = x_i, Y = y_i)$.

უწყვეტ შემთხვევით სიდიდეთა სისტემისთვის გვექნება:

$$a_{ks} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^s f(x, y)dxdy.$$

პრაქტიკაში ხშირად გამოიყენება პირველი რიგის საწყისი მომენტები, რომლებიც X და Y შემთხვევითი სიდიდეების მათემატიკურ ლოდინებს წარმოადგენენ.

$$a_{10} = M\left[X^1Y^0\right] = M[X] = m_x,$$
$$a_{01} = M\left[X^0Y^1\right] = M[Y] = m_y.$$

(k, s) რიგის ცენტრალური მომენტი ეწოდება $(X - m_x)^k(Y - m_y)^s$ ნამრავლის მათემატიკურ ლოდინს და აღინიშნება μ_{ks} სიმბოლოთი.

$$\mu_{ks} = M[(X - m_x)^k(Y - m_y)].$$

დისკრეტული შემთხვევითი სიდიდეებისათვის გვექნება:

$$\mu_{ks} = \sum_i \sum_j (x_i - m_x)^k (y_j - m_y)^s p_{ij},$$

ხოლო უწყვეტი შემთხვევითი სიდიდეებისათვის:

$$\mu_{ks} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - m_x)^k (y - m_y)^s f(x, y)dxdy.$$

პრაქტიკაში ყველაზე ხშირად გამოიყენება მეორე რიგის ცენტრალური მომენტი, რომელიც დისპერსიას წარმოადგენს.

$$D(x) = \mu_{20} = M[(X - m_x)^2(Y - m_y)^0] = M[(X - m_x)^2].$$
$$D(y) = \mu_{02} = M[(X - m_x)^0(Y - m_y)^2] = M[(Y - m_y)^2].$$

პრაქტიკულ კვლევებში განსაკუთრებულ როლს თამაშობს მეორე რიგის შერეული ცენტრალური მომენტი μ_{11}, რომელსაც X და Y შემთხვევითი სიდიდეთა **კორელაციურ მომენტს** ან კავშირის მომენტს უწოდებენ და მას k_{xy} სიმბოლოთი აღნიშნავენ.

$$k_{xy} = \mu_{11} = M[(X - m_x)(Y - m_y)].$$

დისკრეტული შემთხვევითი სიდიდეებისათვის გვექნება:

$$k_{xy} = \sum_i \sum_j (x_i - m_x)(y_j - m_y)p_{ij},$$

ხოლო უწყვეტი შემთხვევითი სიდიდეებისათვის:

$$k_{xy} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - m_x)(y - m_y)f(x, y)dxdy.$$

კორელაციური მომენტი ახასიათებს ორი შემთხვევითი სიდიდის კავშირს. კავშირის ხარისხის შეფასებისთვის გამოიყენება არა თვით კორელაციური მომენტი, არამედ უგანზომილებო სიდიდე

$$r_{xy} = \frac{k_{xy}}{s_x s_y},$$

რომელსაც **კორელაციის კოეფიციენტი** ეწოდება. s_x და s_y წარმოადგენენ საშუალო კვადრატულ გადახრებს.

ადვილად მტკიცდება, რომ დამოუკიდებელ შემთხვევით სიდიდეთა კორელაციური მომენტი და კორელაციის კოეფიციენტი ნულის ტოლია. წინააღმდეგ შემთხვევაში, შემთხვევითი სიდიდეები დამოკიდებულნი ანუ კორელირებულნი არიან.

1.4 ნორმალური განაწილება სიბრტყეზე

რადგან ორი შემთხვევითი სიდიდისგან შემდგარი სისტემა წარმოდგენილია შემთხვევითი წერტილით სიბრტყეზე, ამიტომ ორგანზომილებიანი სისტემის ნორმალური განაწილების კანონს ხშირად უწოდებენ ნორმალურ განაწილებას სიბრტყეზე.

დავუშვათ, X და Y დამოუკიდებელი შემთხვევითი სიდიდეებია, რომელთა ნორმალური განაწილების სიმკვრივის ფუნქციებია:

$$f_x(x)=\frac{1}{s_x\sqrt{2\pi}}\exp\left\{-\frac{(x-m_x)^2}{2s_x^2}\right\},$$

$$f_y(y)=\frac{1}{s_y\sqrt{2\pi}}\exp\left\{-\frac{(y-m_y)^2}{2s_y^2}\right\}.$$

მაშინ მათი ერთობლივი განაწილების სიმკვრივის ფუნქცია განისაზღვრება შემდეგი გამოსახულებით:

$$f(x,y)=f_x(x)f_y(y)=\frac{1}{2\pi s_x s_y}\exp\left\{-\frac{1}{2}\left[\frac{(x-m_x)^2}{s_x^2}+\frac{(y-m_y)^2}{s_y^2}\right]\right\}. \qquad (1.1)$$

თუ ორგანზომილებიანი სისტემის გაფანტვის ცენტრი ემთხვევა კოორდინატთა სათავეს, მაშინ $m_x = m_y = 0$. შესაბამისად გვექნება:

$$f(x,y)=\frac{1}{2\pi s_x s_y}\exp\left\{-\frac{1}{2}\left(\frac{x^2}{s_x^2}+\frac{y^2}{s_y^2}\right)\right\}$$

და მას ერთობლივი ნორმალური განაწილების კანონიკური ფორმა ეწოდება.

თუ X და Y შემთხვევითი სიდიდეები დამოკიდებულნი არიან, მაშინ ერთობლივი ნორმალური განაწილების სიმკვრივის ფუნქციას აქვს შემდეგი სახე:

$$f(x,y)=\frac{1}{2\pi s_x s_y\sqrt{1-r_{xy}^2}}\exp\left\{-\frac{1}{2(1-r_{xy}^2)}\left[\frac{(x-m_x)^2}{s_x^2}-\right.\right.$$
$$\left.\left.-2r_{xy}\frac{(x-m_x)(y-m_y)}{s_x s_y}+\frac{(y-m_y)^2}{s_y^2}\right]\right\} \qquad (1.2)$$

და იგი დამოკიდებულია ხუთ პარამეტრზე: $m_x, m_y, \sigma_x, \sigma_y$ და r_{xy}.

პირობითი ნორმალური განაწილების სიმკვრივის ფუნქცია განისაზღვრება შემდეგი ფორმულით:

$$f(y\mid x)=\frac{1}{\sqrt{2\pi}.s_y\sqrt{1-r_{xy}^2}}\exp\left\{-\frac{1}{2s_y^2(1-r_{xy}^2)}\left[y-m_y-r_{xy}\frac{s_y}{s_x}(x-m_x)\right]^2\right\},$$

$$f(x\mid y)=\frac{1}{\sqrt{2\pi}.s_x\sqrt{1-r_{xy}^2}}\exp\left\{-\frac{1}{2s_x^2(1-r_{xy}^2)}\left[x-m_x-r_{xy}\frac{s_x}{s_y}(y-m_y)\right]^2\right\}.$$

თუ X და Y სიდიდეები არაკორელირებულია ($r_{xy} = 0$), მაშინ (1.2) გამოსახულებიდან მივიღებთ (1.1) ფორმულას, რომელიც წარმოადგენს დამოუკიდებელი შემთხვევითი სიდიდეების ერთობლივი განაწილების სიმკვრივის ფუნქციას. აქედან გამომდინარე, თუ ნორმალურად განაწილებული შემთხვევითი სიდიდეები არაკორელირებულნი არიან, მაშინ ისინი დამოუკიდებელნი არიან.

1.5 მრავალგანზომილებიანი სისტემის ნორმალური განაწილების კანონი

ბიოსამედიცინო კვლევებში, როგორც წესი, საქმე გვაქვს არა ერთ ან ორ, არამედ საკმაოდ ბევრ პარამეტრთან. ვთქვათ მოცემულია მრავალგანზომილებიანი სისტემა, რომელიც აღიწერება $X_1,X_2,...,X_n$ ნორმალურად განაწილებული შემთხვევითი სიდიდეებით. მაშინ განაწილების სიმკვრივის ფუნქციას აქვს შემდეგი სახე:

$$f(X_1, X_2,..., X_n) = \frac{1}{(2\pi)^{\frac{n}{2}}|S|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(X - \overline{X}\right)' S^{-1}\left(X - \overline{X}\right)\right\}, \qquad (1.3)$$

სადაც, X საწყისი მონაცემების მატრიცაა

$$X = \left(X_1, X_2,..., X_n\right)' = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix},$$

$\overline{X}$ – საშუალო სიდიდეთა ვექტორია $\overline{X} = \left(\overline{x}_1, \overline{x}_2,..., \overline{x}_n\right)'$,

S – კოვარიაციული მატრიცაა, რომელიც განისაზღვრება შემდეგნაირად:

$$S = \frac{1}{m-1}\sum_{i=1}^{m}\left(X_i - \overline{X}\right)\left(X_i - \overline{X}\right)' = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1n} \\ s_{21} & s_{22} & \dots & s_{2n} \\ \dots & \dots & \dots & \dots \\ s_{n1} & s_{n2} & \dots & s_{nn} \end{bmatrix}$$

და რომლის მთავარ დიაგონალზე ($s_{11},s_{22},...,s_{nn}$) იმყოფება დისპერსიების მნიშვნელობები. მიღებული მატრიცა სიმეტრიულია, ე.ი. $s_{ij} = s_{ji}$.

თუ მოვახდენთ მოცემული სისტემის $X_1,X_2,...,X_n$ შემთხვევითი სიდიდეების სტანდარტიზირებას (ნორმირებას)

$$Z_i = \frac{X_i - \overline{X}_i}{\sqrt{s_{ii}}},$$

მაშინ მივიღებთ კორელაციურ მატრიცას

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{21} & 1 & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & \dots & 1 \end{bmatrix},$$

ხოლო განაწილების სიმკვრივის ფუნქციას ექნება შემდეგი სახე:

$$f(Z_1,Z_2,...,Z_n)=\frac{1}{(2\pi)^{\frac{n}{2}}|R|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}Z'R^{-1}Z\right\}.$$

ზოგადად, მრავალგანზომილებიანი სისტემის ნორმალურ განაწილებას $N_n(\bar{X},S)$ სიმბოლოთი აღნიშნავენ. უნდა შევნიშნოთ, რომ თუ $n=1$, მაშინ (1.3) გამოსახულება გარდაიქმნება ერთგანზომილებიან ნორმალურად განაწილებულ შემთხვევითი სიდიდის განაწილების სიმკვრივის გამოსახულებად. ამრიგად, (1.3) არის ნორმალური განაწილების განზოგადებული ფორმულა.

2. ჰიპოთეზების სტატისტიკური შემოწმება

2.1 ორზე მეტი ამონარჩევის ერთდროული შედარება

დისპერსიების ტოლობის ჰიპოთეზა. თუ მოცემულია k რაოდენობის ნორმალურად განაწილებული ამონარჩევი $x_{ij}, i=1,2,...,n_j,\ j=1,2,...,k$ და საჭიროა $H_0:\sigma_1^2=\sigma_2^2=...=\sigma_k^2$ ნულოვანი ჰიპოთეზის შემოწმება, მაშინ შეგვიძლია გამოვიყენოთ **ბარტლეტის კრიტერიუმი:**

$$\chi^2=\frac{2{,}303}{C}\left[(N-k)\lg\bar{\sigma}^2-\sum_{i=1}^{k}(n_i-1)\lg\sigma_i^2\right],$$

სადაც,

$$C=\frac{1}{3(k-1)}\left[\sum_{i=1}^{k}\frac{1}{n_i-1}-\frac{1}{N-k}\right]+1,$$

$$N=\sum_{i=1}^{k}n_i;\quad \sigma_j^2=\frac{1}{n_j-1}\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_j)^2$$

$$\bar{x}_j=\frac{1}{n_j}\sum_{i=1}^{n_j}x_{ij},\quad j=1,2,...,k.$$

$\bar{\sigma}^2-$ გაერთიანებული ანუ საშუალო დისპერსიაა, რომელიც ასე განისაზღვრება:

$$\bar{\sigma}^2=\frac{1}{N-k}\sum_{j=1}^{k}(n_j-1)\sigma_j^2=\frac{1}{N-k}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_j)^2.$$

თუ ამონარჩევები ერთნაირი განზომილებისაა, ე.ი. $n_1=n_2=\cdots=n_k=n_0$, მაშინ χ^2 სტატისტიკის გამოსახულება გამარტივდება და მიიღებს შემდეგ სახეს:

$$\chi^2=\frac{2{,}303}{C}\left[k(n_0-1)\left\{\lg\bar{\sigma}^2-\frac{1}{k}\sum_{i=1}^{k}\lg\sigma_i^2\right\}\right],$$

სადაც,

$$C=\frac{k+1}{3k(n_0-1)}+1;\quad \bar{\sigma}^2=\frac{1}{k}\sum_{i=1}^{k}\sigma_i^2.$$

χ^2 სტატისტიკას გააჩნია χ^2 განაწილება $\nu=k-1$ თავისუფლების ხარისხით. α მნიშვნელოვნების დონითა და ν სიდიდით χ^2 განაწილების ცხრილიდან მოიძებნება $\chi^2_{\alpha;\nu}$ კრიტიკული წერტილი. თუ $\chi^2\geq\chi^2_{\alpha;\nu}$, მაშინ ნულოვანი ჰიპოთეზა უარყოფილია ალტერნატიულის სასარგებლოდ. როცა $\chi^2<\chi^2_{\alpha;\nu}$, მაშინ ნულოვანი ჰიპოთეზა მიიღება. ე.ი. დისპერსიები არ განსხვავდებიან ერთმანეთისგან.

მაგალითი. მოცემულია $n_1 = 9$, $n_2 = 6$ და $n_3 = 5$ განზომილებიანი ამონარჩევები $\sigma_1^2=8{,}00$, $\sigma_2^2=4{,}67$, $\sigma_3^2=4{,}00$ დისპერსიებით. შევამოწმოთ დისპერსიების ტოლობის ნულოვანი ჰიპოთეზა $H_0:\sigma_1^2=\sigma_2^2=\sigma_3^2$. ავიღოთ α = 0,05.

$$\bar{\sigma}^2=\frac{1}{N-k}\sum_{j=1}^{k}(n_j-1)\sigma_j^2=\frac{1}{20-3}(8\cdot 8+5\cdot 4{,}67+4\cdot 4)=6{,}079;\quad \lg\bar{\sigma}^2=0{,}7838;$$

$$N=\sum_{i=1}^{k}n_i=9+6+5=20;$$

$$C=\frac{1}{3(k-1)}\left[\sum_{i=1}^{k}\frac{1}{n_i-1}-\frac{1}{N-k}\right]+1=\frac{1}{6}\left[\left(\frac{1}{8}+\frac{1}{5}+\frac{1}{4}\right)-\frac{1}{17}\right]+1=1{,}086;$$

$$\chi^2=\frac{2{,}303}{C}\left[(N-k)\lg\bar{\sigma}^2-\sum_{i=1}^{k}(n_i-1)\lg\sigma_i^2\right]=$$

$$=\frac{2{,}303}{1{,}086}\left[17\cdot 0{,}7838-(8\cdot 0{,}9031+5\cdot 0{,}6693+4\cdot 0{,}6021)\right]=0{,}731;\qquad \chi^2_{0{,}05;2}=5{,}99.$$

რადგან 0,731 < 5,99, ამიტომ ნულოვანი ჰიპოთეზა მიიღება, ე.ი. დისპერსიები არ განსხვავდებიან ერთმანეთისაგან.

საშუალოების ტოლობის ჰიპოთეზა. თუ ამონარჩევების რაოდენობა $k > 2$, მაშინ საშუალოების ტოლობის ჰიპოთეზა ეფუძნება დისპერსიულ ანალიზს, კერძოდ, ერთფაქტორიანი დისპერსიული ანალიზის შედეგებს. იდეა მდგომარეობს საერთო დისპერსიის ორ დამოუკიდებელ მდგენელად: ფაქტორულ (ჯგუფთაშორისო) და ნარჩენ (შიგაჯგუფური) დისპერსიებად წარმოდგენაში. ამრიგად, $\sigma^2=\sigma_f^2+\sigma_{nar}^2$. ფიშერის კრიტერიუმის $F=\frac{\sigma_f^2}{\sigma_{nar}^2}$ გამოყენებით მიდიან დასკვნამდე, არის თუ არა საშუალოებს შორის განსხვავება.

ვთქვათ, მოცემულია k რაოდენობის n_i განზომილებიანი ამონარჩევები x_{ij}, $i=1,2,\ldots,n_j$, $j=1,2,\ldots,k$, რომლებიც ნორმალურად არიან განაწილებული $N(a_i, s_i)$, სადაც, a_i და s_i პარამეტრები უცნობია, მაგრამ გულისხმობენ, რომ $s_1^2=s_2^2=\cdots=s_k^2$. ამ ტოლობის ჰიპოთეზა შეიძლება შევამოწმოთ ბარტლეტის კრიტერიუმით.

საშუალოების ტოლობის $H_0:\bar{x}_1=\bar{x}_2=\cdots=\bar{x}_k$ ნულოვანი ჰიპოთეზის შესამოწმებლად განვიხილოთ სტატისტიკა:

$$F=\frac{\frac{1}{k-1}\sum_{i=1}^{k}n_i\left(\bar{x}_i-\bar{\bar{x}}\right)^2}{\frac{1}{N-k}\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{ij}-\bar{x}_j\right)^2},$$

ან

$$F=\frac{\frac{1}{k-1}\sum_{i=1}^{k}n_i\left(\bar{x}_i-\bar{\bar{x}}\right)^2}{\frac{1}{N-k}\sum_{j=1}^{k}\left(n_j-1\right)\sigma_j^2},$$

სადაც,

$$\bar{x}_j=\frac{1}{n_j}\sum_{i=1}^{n_j}x_{ij},\ \ j=1,2,...,k$$

$$\bar{\bar{x}}=\frac{1}{N}\sum_{i=1}^{k}n_i\bar{x}_i;\qquad N=\sum_{i=1}^{k}n_i.$$

თუ ამონარჩევების განზომილებები ერთმანეთის ტოლია, ე.ი. $n_1=n_2=\cdots=n_k=n_0$, მაშინ გვექნება:

$$F=\frac{\frac{n_0}{k-1}\sum_{i=1}^{k}\left(\bar{x}_i-\bar{\bar{x}}\right)^2}{\bar{\sigma}^2},$$

სადაც,

$$\bar{\bar{x}}=\frac{1}{k}\sum_{i=1}^{k}\bar{x}_i,\quad \bar{\sigma}^2=\frac{1}{k}\sum_{i=1}^{k}\sigma_i^2.$$

F სიდიდეს გააჩნია ფიშერის განაწილება $\nu_1=k-1$ და $\nu_2=N-k$ თავისუფლების ხარისხებით. α,ν_1 და ν_2 სიდიდეებით ფიშერის განაწილების ცხრილიდან შეირჩევა $F_{\alpha;\nu_1\nu_2}$ კრიტიკული წერტილი. თუ $F\geq F_{\alpha;\nu_1\nu_2}$, მაშინ ნულოვანი ჰიპოთეზა უარყოფილია, ხოლო როცა $F<F_{\alpha;\nu_1\nu_2}$, მაშინ ნულოვანი ჰიპოთეზა მიიღება.

მაგალითი. მოცემულია n_0 = 26 განზომილებიანი სამი ამონარჩევი $\sigma_1^2=1{,}69;\ \sigma_2^2=4{,}41;\ \sigma_3^2=5{,}76.$ დისპერსიებისა და $\bar{x}_1=11{,}5;\ \bar{x}_2=10{,}1$ და $\bar{x}_3=9{,}1$ საშუალო არითმეტიკულების შეფასებებით.

უნდა შევამოწმოთ $H_0:\bar{x}_1=\bar{x}_2=\bar{x}_3$ ნულოვანი ჰიპოთეზა. ავიღოთ $\alpha=0{,}05.$

$$\bar{\bar{x}}=\frac{1}{k}\sum_{i=1}^{k}\bar{x}_i=\frac{1}{3}\left(11{,}5+10{,}1+9{,}1\right)=10{,}2;\qquad \bar{\sigma}^2=\frac{1}{k}\sum_{i=1}^{k}\sigma_i^2=\frac{1}{3}\left(1{,}69+4{,}41+5{,}76\right)=3{,}95;$$

$$F=\frac{\frac{n_0}{k-1}\sum_{i=1}^{k}\left(\bar{x}_i-\bar{\bar{x}}\right)^2}{\bar{\sigma}^2}=\frac{\frac{16}{3-1}\left[\left(11{,}5-10{,}2\right)^2+\left(10{,}1-10{,}2\right)^2+\left(9{,}1-10{,}2\right)^2\right]}{3{,}95}=9{,}58;$$

$\nu_1=3-1=2;\quad \nu_2=3\cdot 26-3=75;\quad F_{0{,}05;2;75}=3{,}15.$

რადგან 9,58 > 3,15, ამიტომ ნულოვანი ჰიპოთეზა უარყოფილია, ე.ი. საშუალო სიდიდეები განსხვავდება ერთმანეთისგან.

შევამოწმოთ $H_0 : \bar{x}_1 = \bar{x}_2 = \bar{x}_3$ ნულოვანი ჰიპოთეზა P-მნიშვნელობით. სტანდარტიზირებული ნორმალური განაწილების ფუნქციის ცხრილიდან (იხ. დანართი) $F(3{,}15) = 0{,}9992$. $P = 1 - F(z) = 1 - F(3{,}15) = 1 - 0{,}9992 = 0{,}0008$ რადგან $0{,}0008 < 0{,}05$, ამიტომ ნულოვანი ჰიპოთეზა უარყოფილია.

უნდა აღვნიშნოთ, რომ თუ შესადარებელი ამონარჩევების რაოდენობა $k = 2$, მაშინ ადვილად მტკიცდება, რომ სტიუდენტის კრიტერიუმი წარმოადგენს დისპერსიული ანალიზის კერძო შემთხვევას და სამართლიანია $F = t^2$ ტოლობა. მართლაც, თუ განვიხილავთ ერთნაირი განზომილების ორ ამონარჩევს $\bar{x}_1, \bar{x}_2$ საშუალოებითა და σ_1^2, σ_2^2 დისპერსიებით, მაშინ

$$F = \frac{n\left((\bar{x}_1 - \bar{\bar{x}})^2 + (\bar{x}_2 - \bar{\bar{x}})^2\right)}{\frac{1}{2}\left(\sigma_1^2 + \sigma_2^2\right)},$$

სადაც, $\bar{\bar{x}} = \frac{1}{2}(\bar{x}_1 + \bar{x}_2)$. F-ის გამოსახულებიდან გამოვრიცხოთ $\bar{\bar{x}}$.

$$(\bar{x}_1 - \bar{x})^2 + (\bar{x}_2 - \bar{\bar{x}})^2 = \left[\bar{x}_1 - \frac{1}{2}(\bar{x}_1 + \bar{x}_2)\right]^2 + \left[\bar{x}_2 - \frac{1}{2}(\bar{x}_1 + \bar{x}_2)\right]^2 =$$

$$= \left(\frac{1}{2}\bar{x}_1 - \frac{1}{2}\bar{x}_2\right)^2 + \left(\frac{1}{2}\bar{x}_2 - \frac{1}{2}\bar{x}_1\right)^2 = \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2,$$

რადგან $\left(\frac{1}{2}\bar{x}_2 - \frac{1}{2}\bar{x}_1\right)^2 = \left(\frac{1}{2}\bar{x}_1 - \frac{1}{2}\bar{x}_2\right)^2$.

$$F = \frac{\frac{n}{2}(\bar{x}_1 - \bar{x}_2)^2}{\frac{1}{2}\left(\sigma_1^2 + \sigma_2^2\right)} = \frac{(\bar{x}_1 - \bar{x}_2)^2}{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}} = \left[\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}}\right]^2 = t^2.$$

ამრიგად, ორი ამონარჩევის შედარების დროს, სტიუდენტის კრიტერიუმი და დისპერსიული ანალიზი წარმოადგენს ერთი და იგივე კრიტერიუმს. თუ ამონარჩევების რაოდენობა $k > 2$, მაშინ ეს ასე არ არის.

2.2 საშუალოების მრავლობითი შედარება

პარამეტრული მეთოდები. თუ დისპერსიული ანალიზის გამოყენების შემდეგ აღმოჩნდება, რომ საშუალო სიდიდეები განსხვავდებიან ერთმანეთისგან, მაშინ შეუძლებელია დავადგინოთ, კერძოდ რომელი ამონარჩევები განსხვავდებიან. ასეთ შემთხვევაში, საჭიროა საშუალოები წყვილ-წყვილად შევადაროთ ერთმანეთს. უნდა გვახსოვდეს, რომ სტიუდენტის კრიტერიუმი გამოიყენება მხოლოდ ორი ამონარჩევის შედარებისთვის. თუ მას გამოვიყენებთ მრავლობითი შედარებისთვის, მაშინ ადგილი ექნება მრავლობითი შედარების ეფექტს, რომელიც იწვევს პირველი გვარის შეცდომის მოხდენის ალბათობის ზრდას, რადგან იზრდება მნიშვნელოვნების დონის სიდიდე, რომელიც განისაზღვრება შემდეგი ფორმულით:

$$\alpha' = 1 - (1 - \alpha)^k,$$

სადაც, k – შედარებათა რაოდენობაა. თუ k სიდიდე არც ისე დიდია, მაშინ შეიძლება გამოვიყენოთ მიახლოებითი ფორმულა $\alpha' \approx \alpha k$. მაგალითად, თუ $k = 3$ და ავიღებთ 5% მნიშვნელოვნების დონეს, ე.ი. $\alpha = 0{,}05$, მაშინ $\alpha' = 0{,}15$ და პირველი გვარის შეცდომის მოხდენის ალბათობა შეიძლება 15%-მდე გაიზარდოს. როცა $k = 6$, მაშინ იგი 30%-ის ტოლია და ა.შ.

ამ ეფექტის შესუსტება შეგვიძლია **ბონფერონის შესწორებით**, რომლის თანახმად, თითოეული შედარების მნიშვნელოვნების დონედ უნდა ავიღოთ $\frac{\alpha'}{k}$ სიდიდე. მაგალითად, სამჯერადი შედარებისას, მნიშვნელოვნების დონე უნდა იყოს $\frac{0{,}05}{3} \approx 1{,}7\%$. ბონფერონის შესწორება შედარებით კარგად მუშაობს მცირე რაოდენობის შედარებისას ($k < 8$). უფრო დიდი რაოდენობის შედარებისათვის, უმჯობესია, გამოვიყენოთ **ნიუმენ-კეილსის კრიტერიუმი**, რომელსაც აქვს შემდეგი სახე:

$$q_{ij} = \frac{\left|\bar{x}_j - \bar{x}_i\right|}{\sqrt{\frac{\bar{\sigma}^2}{2}\left(\frac{1}{n_j} + \frac{1}{n_i}\right)}},$$

სადაც, $\bar{\sigma}^2$ – შესადარებელი ამონარჩევების საშუალო დისპერსიაა, n_j, n_i – ამონარჩევების განზომილება. მიღებული q მნიშვნელობა უნდა შევადაროთ ცხრილიდან აღებულ $q_{\alpha;\nu,l}$ კრიტიკულ მნიშვნელობას α მნიშვნელოვნების დონით, $\nu = N - m$ თავისუფლების ხარისხითა და l შედარების ინტერვალით (იხ. დანართი). აქ $N = \sum_{i=1}^{m} n_i$, m – ამონარჩევების რაოდენობაა. თუ აღმოჩნდება, რომ $q_{ij} < q_{\alpha;\nu,l}$, მაშინ ნულოვანი ჰიპოთეზა მიიღება, ე.ი. საშუალოები არ განსხვავდებიან ერთმანეთისგან, წინააღმდეგ შემთხვევაში, როცა $q_{ij} \geq q_{\alpha;\nu,l}$ – საშუალოები განსხვავდებიან.

შედარების ინტერვალი l განისაზღვრება შემდეგნაირად: საშუალო სიდიდეები $\bar{x}_i$, $i = 1,2,...,m$ უნდა დავალაგოთ ზრდადობის მიხედვით. მაგალითად, თუ ვადარებთ $\bar{x}_{(j)}$ და $\bar{x}_{(i)}$ საშუალოებს, რომლებსაც რანჟირებულ მწკრივში უკავიათ j-ური და i-ური ადგილი, მაშინ $l = j - i + 1$. მაგალითად, თუ ვადარებთ $\bar{x}_{(4)}$ და $\bar{x}_{(1)}$, მაშინ $l = 4 - 1 + 1 = 4$; $\bar{x}_{(2)}$ და $\bar{x}_{(1)}$ შედარებისას: $l = 2 - 1 + 1 = 2$ და ა.შ.

ნიუმენ-კეილსის კრიტერიუმის გამოყენების შედეგი დამოკიდებულია შედარების გარკვეულ რიგზე. კერძოდ, თუ საშუალო სიდიდეებს დავალაგებთ ზრდადობით $1,2,\ldots,m$, მაშინ ჯერ უნდა შევადაროთ მწკრივის კიდურა (მაქსიმალური და მინიმალური) სიდიდეები, ე.ი. m-ური და 1-ლი საშუალო სიდიდეები, შემდეგ m-ური და მე-2 და ა.შ. მაგალითად ოთხი საშუალო სიდიდისათვის გვექნება შედარების ასეთი თანმიმდევრობა: 4 – 1; 4 – 2; 4 – 3; 3 – 1; 3 – 2 და 2 – 1. აქვე უნდა შევნიშნოთ, რომ შედარება ყველა წყვილისთვის არაა საჭირო. იმ შემთხვევაში, როცა რომელიმე საშუალოების წყვილი არ განსხვავდება ერთმანეთისგან, მაშინ შედარების ამოცანა წყდება, რადგან დანარჩენები მით უფრო არ იქნება განსხვავებული. მაგალითად, თუ აღმოჩ-

ნდება, რომ 3–1 წყვილი არ განსხვავდება ერთმანეთისგან, მაშინ აღარაა საჭირო $3-2$ და $2-1$ წყვილების შედარება.

მაგალითი. §2.1-ში მოყვანილი მაგალითისთვის, სადაც მივიღეთ, რომ საშუალოები განსხვავდებიან, ჩავატაროთ წყვილ-წყვილად შედარება. დავალაგოთ საშუალოები ზრდადობით. $\bar{x}_{(1)}=9{,}1;\ \bar{x}_{(2)}=10{,}1;\ \bar{x}_{(3)}=11{,}5$. შევამოწმოთ $H_0:\bar{x}_{(3)}=\bar{x}_{(1)}$ ნულოვანი ჰიპოთეზა

$$q_{31}=\frac{\left|\bar{x}_{(3)}-\bar{x}_{(1)}\right|}{\sqrt{\frac{\bar{\sigma}^2}{2}\left(\frac{1}{n_3}+\frac{1}{n_1}\right)}}=\frac{11{,}5-9{,}1}{\sqrt{\frac{3{,}95}{2}\left(\frac{1}{26}+\frac{1}{26}\right)}}=6{,}16.$$

ამ შემთხვევაში, $l = j - i + 1 = 3 - 1 + 1 = 3$. q განაწილების ცხრილიდან $\alpha=0{,}05;\ \nu=3\cdot 26-3=75$; ვღებულობთ $q_{0{,}05;75;3}=3{,}39$. რადგან 6,16 > 3,39, ამიტომ ნულოვანი ჰიპოთეზა უარყოფილია, ე.ი. განსხვავება სარწმუნოა. შევამოწმოთ $H_0:\bar{x}_{(3)}=\bar{x}_{(2)}$ ნულოვანი ჰიპოთეზა

$$q_{32}=\frac{11{,}5-10{,}1}{\sqrt{\frac{3{,}95}{2}\left(\frac{1}{26}+\frac{1}{26}\right)}}=3{,}59.$$

აქ, $l = 3 - 2 + 1 = 2$, $q_{0{,}05;75;2}=2{,}82$. რადგან 3,59 > 2,82, ამიტომ ნულოვანი ჰიპოთეზა უარყოფილია, ე.ი. განსხვავება სარწმუნოა.

$H_0:\bar{x}_2=\bar{x}_1$ ჰიპოთეზის შემოწმებისთვის გვექნება:

$$q_{21}=\frac{10{,}1-9{,}1}{\sqrt{\frac{3{,}95}{2}\left(\frac{1}{26}+\frac{1}{26}\right)}}=2{,}57.$$

აქ, $l = 2 - 1 + 1 = 2$, $q_{0{,}05;75;2}=2{,}82$. რადგან 2,57 < 2,82, ამიტომ ნულოვანი ჰიპოთეზა მიიღება, ე.ი. საშუალოები არ განსხვავდება ერთმანეთისგან.

არაპარამეტრული მეთოდები. მრავლობითი შედარების პარამეტრული მეთოდები ადვილად ადაპტირდება არაპარამეტრულ მეთოდებში. კერძოდ, როდესაც ამონარჩევების განზომილება ერთნაირია, მაშინ შეგვიძლია გამოვიყენოთ ნიუმენ-კეილსის არაპარამეტრული (რანგული) კრიტერიუმი, ხოლო როდესაც გვაქვს სხვადასხვა განზომილებიანი ამონარჩევები – დანას კრიტერიუმი.

ვთქვათ, მოცემულია m რაოდენობის n-განზომილებიანი ამონარჩევი. მოვახდინოთ ამონარჩევების ერთდროული რანჟირება და ყოველ მათგანს მივანიჭოთ რანგი. მაშინ ნიუმენ-კეილსის რანგულ კრიტერიუმს აქვს შემდეგი სახე:

$$q=\frac{\left|R_i-R_j\right|}{\sqrt{\frac{n^2 l(nl+1)}{12}}},$$

სადაც, R_i და R_j შესადარებელი i-ური და j-ური ამონარჩევების რანგების ჯამია. l – შედარების ინტერვალი, რომელიც განისაზღვრება ისევე, როგორც პარამეტრული მეთოდის დროს. α მნიშვნელოვნების დონით, $\nu = \infty$ და l სიდიდეებით ცხრილიდან შეირჩევა $q_{\alpha;\nu;l}$ კრიტიკული მნიშვნელობა. თუ

$q \geq q_{\alpha;v;l}$, მაშინ ნულოვანი ჰიპოთეზა უარყოფილია, ე.ი. განსხვავება ამონარჩევებს შორის სარწმუნოა. წინააღმდეგ შემთხვევაში, როცა $q < q_{\alpha;v;l}$, ნულოვანი ჰიპოთეზა მიიღება და ამონარჩევები ერთმანეთისგან არ განსხვავდება.

თუ ამონარჩევები სხვადასხვა განზომილებისაა, მაშინ მათი წყვილ-წყვილად შედარებისათვის უნდა გამოვიყენოთ **დანას კრიტერიუმი** (Q კრიტერიუმი):

$$Q_{ij} = \frac{\left|\overline{R}_i - \overline{R}_j\right|}{\sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}},$$

სადაც, $\overline{R}_i$ და $\overline{R}_j$ i-ური და j-ური ამონარჩევების საშუალო რანგებია, n_i, n_j – ამონარჩევების განზომილებები, N – ყველა ამონარჩევის ჯამური განზომილება, ე.ი. $N = \sum_{i=1}^{m} n_i$. Q კრიტიკული მნიშვნელობები მოცემულია სპეციალურ ცხრილში (იხ. დანართი). α მნიშვნელოვნების დონითა და m სიდიდით ცხრილიდან შეირჩევა $Q_{\alpha m}$ კრიტიკული მნიშვნელობა. თუ $Q_{ij} \geq Q_{\alpha;m}$, მაშინ ამონარჩევები განსხვავდება ერთმანეთისგან, წინააღმდეგ შემთხვევაში, როცა $Q_{ij} < Q_{\alpha;m}$, ისინი არ განსხვავდებიან ერთმანეთისგან. უნდა აღინიშნოს, რომ დანას კრიტერიუმი შეიძლება გამოვიყენოთ მაშინაც, როცა ამონარჩევებს გააჩნიათ ერთნაირი განზომილება.

მაგალითი. ცხრილში მოცემულია პულსის სიხშირის მნიშვნელობები 5 წლამდე ბავშვების სამი ჯგუფისთვის (x,y,z). დავადგინოთ, არის თუ არა განსხვავება ჯგუფებს შორის.

№	x	y	z	R_x	R_y	R_z
1	112	90	110	2,5	2,5	9,5
2	116	78	117	16	1	17,5
3	108	109	112	6	7,5	12,5
4	120	92	115	19	4	14,5
5	109	100	118	7,5	5	19
6	90	115	117	2,5	14,5	17,5
7	110		125	9,5		20
8	111			11		
Σ				84,0	34,5	110,5

N = 8 + 6 + 7 = 21; საშუალო რანგებია $\overline{R}_x$ = 10,5; $\overline{R}_y$ = 5,75; $\overline{R}_z$ = =15,79. კრიტიკული მნიშვნელობა $Q_{0,05;3}$ =2,39. დანას კრიტერიუმით მოვახდინოთ ჯგუფების წყვილ-წყვილად შედარება. მივიღებთ:

$$Q_{xy} = \frac{\left|\overline{R}_x - \overline{R}_y\right|}{\sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}} = \frac{\left|10{,}5 - 5{,}75\right|}{\sqrt{\frac{21 \cdot 22}{12}\left(\frac{1}{8} + \frac{1}{6}\right)}} = 1{,}42;$$

. $Q_{xz} = \dfrac{|10{,}5 - 15{,}79|}{\sqrt{\dfrac{21 \cdot 22}{12}\left(\dfrac{1}{8} + \dfrac{1}{7}\right)}} = 1{,}65;$ $\qquad Q_{yz} = \dfrac{|5{,}75 - 15{,}79|}{\sqrt{\dfrac{21 \cdot 22}{12}\left(\dfrac{1}{6} + \dfrac{1}{7}\right)}} = 2{,}91$

როგორც ვხედავთ, მხოლოდ *y* და *z* ჯგუფები განსხვავდებიან ერთმანეთისგან, რადგან 2,91 > 2,39. დანარჩენ ორ შემთხვევაში ჯგუფებს შორის განსხვავება არ შეიმჩნევა.

2.3 მრავალგანზომილებიანი სისტემის ზოგიერთი ჰიპოთეზების შემოწმება

ორი ვექტორის ტოლობის ჰიპოთეზა. ვთქვათ, მოცემულია საშუალოების ორი ვექტორი $\overline{X}_i$ და $\overline{X}_j$, რომლებიც მიღებულია ორი ნორმალურად განაწილებული *n*-განზომილებიანი სისტემიდან. საჭიროა შევამოწმოთ $\overline{X}_i$ და $\overline{X}_j$ ვექტორების ტოლობის ნულოვანი ჰიპოთეზა, ე.ი. $H_0 : \overline{X}_i = \overline{X}_j$. ასეთი ნულოვანი ჰიპოთეზის შესამოწმებლად განვიხილოთ ჰოტელინგის კრიტერიუმი:

$$T^2 = \frac{m_i m_j}{m_i + m_j}\left(\overline{X}_i - \overline{X}_j\right)' S^{-1}\left(\overline{X}_i - \overline{X}_j\right),$$

სადაც, m_i,m_j შესაბამისად X_i და X_j ვექტორების განზომილებაა; S – გაერთიანებული კოვარიაციის მატრიცაა, რომელიც გამოითვლება შემდეგნაირად:

$$S = \frac{1}{m_i + m_j - 2}\left[(m_i - 1)S_i + (m_j - 1)S_j\right].$$

თუ ნულოვანი ჰიპოთეზა სამართლიანია, მაშინ განვიხილოთ სტატისტიკა

$$F = \frac{m_i + m_j - n - 1}{(m_i + m_j - 2)n} T^2,$$

ღომელსაც გააჩნია ფიშერის განაწილება $\nu_1 = n$ და $\nu_2 = m_i + m_j - n - 1$ თავისუფლების ხარისხებით. თუ $F < F_{\alpha;\nu_1\nu_2}$, მაშინ ნულოვანი ჰიპოთეზა მიიღება, წინააღმდეგ შემთხვევაში, როცა $F \geq F_{\alpha;\nu_1\nu_2}$, ნულოვანი ჰიპოთეზა უარყოფილია, ე.ი. განსხვავება ორ ვექტორს შორია სარწმუნოა.

კოვარიაციული მატრიცების ტოლობის ჰიპოთეზა. განვიხილოთ ორი კოვარიაციული მატრიცის ტოლობის ჰიპოთეზა $H_0 : S_1 = S_2$. ამისათვის უნდა გამოვთვალოთ შემდეგი სტატისტიკა:

$$W = b(-2 \ln V_1),$$

სადაც,

$$b = 1 - \left(\sum_{j=1}^{2} \frac{1}{\nu_j} - \frac{1}{\sum_{j=1}^{2} \nu_j} \right) \left(\frac{2n^2 + 3n - 1}{6(n+1)} \right),$$

$$-2 \ln V_1 = \left(\sum_{j=1}^{2} \nu_j \right) \ln|S| - \sum_{j=1}^{2} (\nu_j \ln|S_j|).$$

$$\nu_1 = m_1 - 1; \qquad \nu_2 = m_2 - 1,$$

რომელსაც გააჩნია χ^2 განაწილება $\nu = \frac{n(n+1)}{2}$ თავისუფლების ხარისხით.

თუ $W < \chi^2_{\alpha;\nu}$, მაშინ ნულოვანი ჰიპოთეზა მიიღება და კოვარიაციული მატრიცები ერთმანეთის ტოლია, წინააღმდეგ შემთხვევაში, როცა $W \geq \chi^2_{\alpha;\nu}$, მაშინ კოვარიაციული მატრიცები განსხვავდება ერთმანეთისგან.

არტეფაქტური ვექტორის გამოვლენა. ვთქვათ, მოცემულია ნორმალურად განაწილებული $X_1, X_2, ..., X_n$ შემთხვევით ვექტორთა სისტემა, რომლის სტრიქონები დაკვირვებათა ვექტორებია: $X_i = (x_{i1}, x_{i2}, ..., x_{in}),\ i = 1,2,...., m.$

არტეფაქტური ვექტორის გამოვლენის პროცედურა შემდეგია: თოთოეული X_i, $i=1,2,...,m$ დაკვირვების ვექტორისთვის გამოითვლება საშუალო არითმეტიკულის ვექტორი

$$\overline{X}_i = \frac{1}{n} \sum_{k=1}^{n} x_{ik}, \quad i = 1,2,....,m$$

და კოვარიაციული მატრიცა S ყველა $(m-1)$ დაკვირვების ვექტორით, გარდა X_i ვექტორისა. შემდეგ გამოითვლება მახალანობისის მანძილი X_i და $\overline{X}_i$ ვექტორებს შორის S კოვარიაციული მატრიცის საშუალებით:

$$D_i^2 = \left(X - \overline{X}_i\right)' S^{-1} \left(X - \overline{X}_i\right).$$

ამის შემდეგ განისაზღვრება F_i მნიშვნელობა $k = m - 1$ სიდიდისთვის.

$$F_i = \frac{(m-n)m}{(m^2-1)n} D_i^2,$$

რომელსაც გააჩნია ფიშერის განაწილება $\nu_1 = n$ და $\nu_2 = m - n$ თავისუფლების ხარისხებით.

თუ $F_i > F_{\alpha;\nu_1\nu_2}$, მაშინ X_i ვექტორი ითვლება არტეფაქტად და იგი უნდა გამოირიცხოს ამონარჩევიდან. პროცედურა გრძელდება დარჩენილ $(n-1)$ დაკვირვებისთვის.

2.4 კონკორდაციის კოეფიციენტი

პრაქტიკაში ხშირად იქმნება ისეთი სიტუაცია, როდესაც საჭირო ხდება ექსპერტთა ჯგუფის გამოყენება ამა თუ იმ საკითხის გადასაჭრელად (კონსილიუმის მოწვევა სამედიცინო დიაგნოზის დასასმელად, სხვადასხვა ნორმატივების დადგენა, პრეპარატების შეფასება და სხვ.). აქედან გამომდინარე, სასურველია დავადგინოთ, რამდენად ემთხვევა ექსპერტთა აზრი ერთი და იმავე საკითხის განხილვას. თუ გვყავს მხოლოდ ორი ექსპერტი, მაშინ თანხმობის ზომად შეგვიძლია მივიღოთ სპირმენის რანგობრივი კორელაციის კოეფიციენტის სიდიდე. მაგრამ როდესაც ექსპერტთა რაოდენობა დიდია, მაშინ სპირმენის რანგობრივი კორელაციის კოეფიციენტის გამოყენება მიზანშეწონილი არ არის.

დავუშვათ, გვაქვს ობიექტების n რაოდენობა და გვყავს m ექსპერტი, რომლებიც აფასებენ ამ ობიექტებს და ახდენენ მათ რანჟირებას (უკეთესიდან უარესისკენ). რანჟირების შედეგად ვიღებთ ასეთ ცხრილს:

ობიექტი / ექსპერტი	1	2	3	...	n
1	x_{11}	x_{12}	x_{13}	...	x_{1n}
2	x_{21}	x_{22}	x_{23}	...	x_{2n}
⋮	⋮	⋮	⋮	⋮	⋮
j	x_{j1}	x_{j2}	x_{j3}	...	x_{jn}
⋮	⋮	⋮	⋮	⋮	⋮
m	x_{m1}	x_{m2}	x_{m3}	...	x_{mn}

ექსპერტთა თანხმობის ზომის დასადგენად, შეგვიძლია გამოვიყენოთ კენდელის მიერ შემოთავაზებული თანხმობის ანუ კონკორდაციის კოეფიციენტი W, რომელიც ასე განისაზღვრება:

$$W = \frac{12\sum_{i=1}^{n} S_i^2}{m^2(n^3 - n)},$$

სადაც S არის სხვაობა ობიექტების რანგების ჯამსა და რანგების საერთო საშუალო არითმეტიკულს შორის, ე.ი.

$$S_i = R_i - \overline{R}, \quad R_i = \sum_{k=1}^{m} x_{ki}, i = 1,2,...,n .$$

თუ ექსპერტების რანჟირებულ მწკრივში გვხვდება ერთი და იგივე რანგის მნიშვნელობა (ეს ის შემთხვევაა, როცა ექსპერტი ვერ ანიჭებს უპირატესობას), მაშინ კონკორდაციის კოეფიციენტი გამოითვლება შემდეგი ფორმულით:

$$W = \frac{\sum_{i=1}^{n} S_i^2}{\frac{1}{12} m^2 \left(n^3 - n\right) - m \sum_{j=1}^{m} T_j},$$

სადაც, $T_j = \frac{1}{12} \sum_j \left(t_j^3 - t_j\right)$, t_j – j-ურ მწკრივში ერთნაირი რანგების რაოდენობაა.

ზოგადად, $0 \le W \le 1$. თუ ექსპერტთა შეფასებები ერთმანეთს მთლიანად ემთხვევა, მაშინ $W = 1$, ხოლო თუ მათი აზრები მკვეთრად განსხვავდებიან ერთმანეთისგან, მაშინ $W = 0$.

კონკორდაციის კოეფიციენტის სარწმუნოების დასადგენად უნდა შევამოწმოთ $H_0: W = 0$ ნულოვანი ჰიპოთეზა. ასეთი ნულოვანი ჰიპოთეზის შესამოწმებლად უნდა გამოვთვალოთ შემდეგი სტატისტიკა: $\chi^2 = Wm(n-1)$, რომელსაც გააჩნია χ^2 განაწილება $\nu = n - 1$ თავისუფლების ხარისხით.

თუ $\chi^2 < \chi^2_{\alpha;\nu}$, მაშინ ნულოვანი ჰიპოთეზა მიიღება, ე.ი. ექსპერტთა აზრები განსხვავდებიან ერთმანეთისაგან, ხოლო როცა $\chi^2 \ge \chi^2_{\alpha;\nu}$, მაშინ ექსპერტთა აზრები ერთმანეთს ემთხვევა.

მაგალითი. 6 ექსპერტი აფასებს 4 ფარმაცევტული ფირმის მიერ გამოშვებულ პრეპარატს. შედეგები მოყვანილია შემდეგ ცხრილში:

ფირმები / ექსპერტები	1	2	3	4	სულ
1	1	3	2	4	
2	2	1	4	3	
3	1	3	2	4	
4	3	2	1	4	
5	1	2	4	3	
6	2	3	1	4	
R	10	14	14	22	60
S	–5	–1	1	7	
S^2	25	1	1	49	76

$$\overline{R} = \frac{60}{4} = 15; \quad W = \frac{12 \sum_{i=1}^{n} S_i^2}{m^2 \left(n^3 - n\right)} = \frac{12 \cdot 76}{36(4^3 - 4)} = 0{,}42; \quad .$$

$$\chi^2 = Wm(n-1) = 0{,}42 \cdot 6 \cdot 3 = 7{,}56; \qquad \chi^2_{0{,}05;5} = 11{,}07$$

რადგან 7,56 < 11,07, ამიტომ ნულოვანი ჰიპოთეზა მიიღება, ე.ი. ექსპერტთა აზრები განსხვავდება ერთმანეთისგან.

3. მრავლობითი რეგრესიული ანალიზი

3.1. მრავლობითი რეგრესიის მოდელი

წრფივი მოდელი. ორცვლადიანი რეგრესიის დროს y მნიშვნელობა დამოკიდებულია მხოლოდ ერთ x ცვლადზე. საზოგადოდ, y შეიძლება იყოს მრავალი ცვლადის ფუნქცია. განვიხილოთ ეს შემთხვევა. ვთქვათ, დამოკიდებულ Y-სა და დამოუკიდებელ $X_1,X_2,...,X_n$ ცვლადებს შორის არსებობს ასეთი წრფივი დამოკიდებულება:

$$Y = a_1X_1 + a_2X_2 + ... + a_n X_n,$$

სადაც, $Y, X_1, X_2, ... , X_n$ ვექტორებია. წარმოვადგინოთ ეს დამოკიდებულება მატრიცული სახით $Y = AX$, სადაც, $A = [a_i]$, $i = 1,2,...,n$ საძიებელი კოეფიციენტების ვექტორია, $Y = [y_i]$, $i = 1,2,...,m$ დამოკიდებული ცვლადის ვექტორია, ხოლო $X = [x_{ij}]$, $i = 1,2,...,m, j = =1,2,...,n$ დამოუკიდებელი ცვლადების მატრიცაა.

a_i კოეფიციენტები მოვძებნოთ უმცირეს კვადრატთა მეთოდით, რომლის თანახმად $Q = \sum_{i=1}^{m} [y_i - \hat{y}_i]^2 = \min$.

ჩავწეროთ ეს გამოსახულება მატრიცული სახით

$$Q = (Y - XA)'(Y - XA) = Y'Y - A'X'Y - Y'XA + A'X'XA.$$

რადგან $A'X'Y = Y'XA$, ამიტომ გვექნება:

$$Q = Y'Y - 2A'X'Y + A'X'XA.$$

გავაწარმოოთ ეს გამოსახულება

$$\frac{\partial Q}{\partial a} = -2X'Y + 2(X'X)A = 0,$$

მაშინ ნორმალურ განტოლებათა სისტემას ექნება შემდეგი სახე:

$$X'Y = X'XA,$$

საიდანაც, $A = (X'X)^{-1}X'Y$.

რადგან X მატრიცა შეიცავს n წრფივად დამოუკიდებელ ვექტორ-სვეტებს, ამიტომ, როგორც ცნობილია, X მატრიცის რანგი $(m–1)$-ის ტოლია. ეს თავის მხრივ, მიგვანიშნებს იმაზე, რომ $|X'X| \neq 0$, ე.ი. მატრიცას გააჩნია შებრუნებული მატრიცა $(X'X)^{-1}$.

ხშირად რეგრესიის განტოლება შეიცავს თავისუფალ წევრს

$$\hat{Y} = a_0 + a_1X_1 + a_2X_2 + ... + a_n X_n.$$

მაშინ a_0 პარამეტრის შეფასებისთვის საჭიროა X მატრიცას დაემატოს ერთეულოვანი ვექტორ-სვეტი

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & ... & x_{1n} \\ 1 & x_{21} & x_{22} & ... & x_{2n} \\ ... & ... & ... & ... & ... \\ 1 & x_{m1} & x_{m2} & ... & x_{mn} \end{bmatrix},$$

მაშინ გვექნება:

$$X'X = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{m1} \\ \dots & \dots & \dots & \dots \\ x_{1n} & x_{2n} & \dots & x_{mn} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1n} \\ 1 & x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} =$$

$$= \begin{bmatrix} n & \sum_i x_i & \dots & \sum_i x_{im} \\ \sum_i x_i & \sum_i x_i^2 & \dots & \sum_i x_i x_{im} \\ \dots & \dots & \dots & \dots \\ \sum_i x_{im} & \sum_i x_{i1} x_{im} & \dots & \sum_i x_{im}^2 \end{bmatrix}.$$

ნარჩენი დისპერსია გამოითვლება შემდეგი ფორმულით:

$$\sigma_e^2 = \frac{1}{m-n-1} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 ,$$

სადაც, $n-$ ცვლადების რაოდენობაა, $m-$ დაკვირვებათა რაოდენობა.

რეგრესიის განტოლების ადეკვატურობის შესამოწმებლად საჭიროა გამოითვალოს შემდეგი სტატისტიკა:

$$F = \frac{Q}{\sigma_e^2}, \quad Q = \frac{1}{n} \sum_{i=1}^{m} (\hat{y}_i - \bar{y})^2 .$$

α მნიშვნელოვნების დონითა და $\nu_1 = n$, $\nu_2 = m–n–1$ თავისუფლების ხარისხებით ფიშერის განაწილების ცხრილიდან შეირჩევა $F_{\alpha;\nu_1\nu_2}$ კრიტიკული მნიშვნელობა. როცა $F \geq F_{\alpha;\nu_1\nu_2}$, მაშინ რეგრესიის განტოლება ადეკვატურია, წინააღმდეგ შემთხვევაში, როცა $F < F_{\alpha;\nu_1\nu_2}$, განტოლება არაადეკვატურია.

რეგრესიის განტოლების კოეფიციენტების შემოწმება სარწმუნოებაზე ხდება შემდეგი $H_0 : a_i = 0$ ნულოვანი ჰიპოთეზის საშუალებით. ამისათვის საჭიროა გამოითვალოს სტატისტიკა:

$$t_i = \frac{a_i}{\sigma_i}, \quad \sigma_i = \sqrt{\sigma_e^2 \cdot s_{ii}}, \quad i = 1,2,...,n,$$

სადაც s_{ii} წარმოადგენს $(X'X)^{-1}$ მატრიცის მთავარ დიაგონალზე მყოფი ელემენტის მნიშვნელობას. α მნიშვნელოვნების დონითა და $\nu = m - n - 1$ თავისუფლების ხარისხით სტიუდენტის განაწილების ცხრილიდან მოიძებნება $t_{\alpha;\nu}$ კრიტიკული მნიშვნელობა. თუ $|t_i| < t_{\alpha;\nu}$ მაშინ ნულოვანი ჰიპოთეზა მიიღება, ე.ი. i-ური კოეფიციენტის მნიშვნელობა ახლოსაა ნულთან და შესაძლებელია მისი გამორიცხვა რეგრესიის განტოლებიდან. თუ $|t_i| \geq t_{\alpha;\nu}$, მაშინ a_i კოეფიციენტი სარწმუნოა.

მაგალითი. ცხრილში მოცემულია y დამოკიდებული და x_1, x_2 დამოუკიდებელი ცვლადების მნიშვნელობები.

	y	x_1	x_2	$\hat{y}$	$(y-\hat{y})$	$(y-\hat{y})^2$	$(\hat{y}-\bar{y})$	$(\hat{y}-\bar{y})^2$
1	10	2	1	10,256	–0,256	0,066	–3,444	11,861
2	12	2	2	10,868	1,132	1,281	–2,832	8,020
3	17	8	10	16,532	0,468	0,219	2,832	8,020
4	13	2	4	12,091	0,909	0,826	–1,609	2,589
5	15	6	8	15,052	–0,052	0,003	1,352	1,828
6	10	3	4	12,22	–2,22	4,928	–1,480	2,190
7	14	5	7	14,312	–0,312	0,098	0,612	0,375
8	12	3	3	11,608	0,392	0,154	–2,092	4,377
9	16	9	10	16,661	–0,661	0,437	2,961	8,768
10	18	10	11	17,401	0,599	0,359	3,701	13,697
$\sum$	137	50	60			8,371		61,725

$\bar{y}=13{,}7; \quad \bar{x}_1=5{,}0; \quad \bar{x}_2=6{,}0.$

შევარჩიოთ რეგრესიის განტოლება $\hat{y}=a_0+a_1x_1+a_2x_2$.

ამრიგად გვაქვს:

$$Y=\begin{bmatrix}10\\12\\17\\...\\18\end{bmatrix} \quad X=\begin{bmatrix}1 & 2 & 1\\1 & 2 & 2\\1 & 8 & 10\\... & ... & ...\\1 & 10 & 11\end{bmatrix};$$

$$X'X=\begin{bmatrix}1 & 1 & 1 & ... & 1\\2 & 2 & 8 & ... & 10\\1 & 2 & 10 & ... & 11\end{bmatrix}\begin{bmatrix}1 & 2 & 1\\1 & 2 & 2\\1 & 8 & 10\\... & ... & ...\\1 & 10 & 11\end{bmatrix}=\begin{bmatrix}10 & 50 & 60\\50 & 336 & 398\\60 & 398 & 480\end{bmatrix};$$

$$(X'X)^{-1}=\begin{bmatrix}0{,}40168 & -0{,}01676 & -0{,}03631\\-0{,}01676 & 0{,}16760 & -0{,}13687\\-0{,}03631 & -0{,}13687 & 0{,}12011\end{bmatrix};$$

$$X'Y=\begin{bmatrix}1 & 1 & 1 & ... & 1\\2 & 2 & 8 & ... & 10\\1 & 2 & 10 & ... & 11\end{bmatrix}\begin{bmatrix}10\\12\\17\\...\\18\end{bmatrix}=\begin{bmatrix}137\\756\\908\end{bmatrix};$$

$$A=(X'X)^{-1}X'Y=\begin{bmatrix}0{,}40168 & -0{,}01676 & -0{,}03631\\-0{,}01676 & 0{,}16760 & -0{,}13687\\-0{,}03631 & -0{,}13687 & 0{,}12011\end{bmatrix}\begin{bmatrix}137\\756\\908\end{bmatrix}=$$

$$= \begin{bmatrix} 9{,}3872 \\ 0{,}1285 \\ 0{,}6174 \end{bmatrix}.$$

ამრიგად, რეგრესიის განტოლებას ღებულობს შემდეგ სახეს:

$$\hat{y} = 9{,}3872 + 0{,}1285x_1 + 0{,}6174x_2$$

თუ x_1 და x_2 სიდიდეები იზომება ერთი და იგივე ფიზიკური ერთეულით, მაშინ ადვილი შესამჩნევია, რომ x_2-ის ზეგავლენა y-ზე x_1-თან შედარებით, დაახლოებით ხუთჯერ უფრო ძლიერია. თუ x_1 და x_2 სხვადასხვა ფიზიკურ ერთეულებშია წარმოდგენილი, ასეთი შედარება უაზრობაა.

გამოვთვალოთ ნარჩენი დისპერსია

$$\sigma_e^2 = \frac{1}{m-n-1}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 = \frac{8{,}371}{10-2-1} = 1{,}1959; \quad Q = \frac{1}{n}\sum_{i=1}^{m}(\hat{y}_i - \bar{y})^2 = \frac{61{,}725}{2} = 30{,}863.$$

შევამოწმოთ რეგრესიის განტოლების ადეკვატურობა

$$F = \frac{Q}{\sigma_e^2} = \frac{30{,}863}{1{,}1959} = 25{,}807; \qquad F_{0{,}05;2;7} = 4{,}74.$$

რადგან 25,807 > 4,74, რეგრესიის განტოლება ადეკვატურია. კოეფიციენტების შესამოწმებლად განვიხილოთ სტატისტიკა:

$$t_i = \frac{a_i}{\sigma_i};\ i = 0{,}1{,}2; \quad \sigma_0 = \sqrt{\sigma_e^2 \cdot s_{11}} = \sqrt{1{,}1959 \cdot 0{,}40168} = 0{,}6931$$

$$\sigma_1 = \sqrt{1{,}1959 \cdot 0{,}1676} = 0{,}4477; \ \sigma_2 = \sqrt{1{,}1959 \cdot 0{,}12011} = 0{,}379\,0;$$

$$t_0 = \frac{9{,}3872}{0{,}6931} = 13{,}54; \ t_1 = \frac{0{,}1285}{0{,}4477} = 0{,}287; \ t_2 = \frac{0{,}6174}{0{,}379} = 1{,}629;$$

$t_{0{,}05;7} = 2{,}37$. როგორც ვხედავთ, გარდა a_0-სა, დანარჩენი a_1 და a_2 კოეფიციენტები სარწმუნონი არ არიან.

არაწრფივი მოდელი. პრაქტიკულ კვლევებში, ზოგჯერ, შეუძლებელი ხდება მრავლობითი წრფივი რეგრესიის განტოლების გამოყენება მისი არაადეკვატურობის გამო. ამ შემთხვევაში, უნდა გამოვიყენოთ ისეთი არაწრფივი მოდელი, რომელიც არაწრფივია ცვლადების მიმართ, მაგრამ წრფივია რეგრესიის კოეფიციენტების მიმართ. ამ შემთხვევაში, კოეფიციენტების შეფასება ხდება საკმაოდ ადვილად. კერძოდ, დამოკიდებული ცვლადები შეიცვლება ახალი პირველი ხარისხის ცვლადებით და მის მიმართ გამოიყენება წრფივი უმცირეს კვადრატთა მეთოდი. მაგალითისთვის განვიხილოთ შემდეგი არაწრფივი რეგრესიის განტოლება:

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_2^2.$$

შემოვიტანოთ ახალი ცვლადები $z_1 = x_1, z_2 = x_2, z_3 = x_1^2, z_4 = x_2^2$, მაშინ გვექნება:

$$y = a_0 + a_1z_1 + a_2z_2 + a_3z_3 + a_4z_4.$$

უმცირეს კვადრატთა მეთოდის გამოყენებით მივიღებთ შემდეგ ნორმალურ განტოლებათა სისტემას:

$$Z'Y = Z'ZA, \text{ საიდანაც } \quad A = (Z'Z)^{-1}Z'Y.$$

მიღებული რეგრესიის განტოლებისა და კოეფიციენტების შეფასებები ხდება ისევე, როგორც მრავლობითი წრფივი რეგრესიის დროს.

3.2. მრავლობითი წრფივი რეგრესიის განტოლების ნდობის ინტერვალი

განვსაზღვროთ მრავალგანზომილებიანი წრფივი რეგრესიის განტოლების **ნდობის ინტერვალი**. ამისათვის გამოვთვალოთ $\hat{y} = a_0 + a_1x_1 + a_2x_2 + ... + a_nx_n$ გამოსახულების დისპერსია

$$D(\hat{y}) = D(a_0 + a_1x_1 + a_2x_2 + ... + a_nx_n) .$$

თუ გავიხსენებთ დამოკიდებული ცვლადების ჯამის დისპერსიის თვისებას, მივიღებთ:

$$\begin{aligned}\sigma^2_{\hat{y}} = D(\hat{y}) = D(a_0) + x_1^2 D(a_1) + x_2^2 D(a_2) + ... + x_n^2 D(a_n) + \\ + 2x_1 \operatorname{cov}(a_0 a_1) + 2x_1 x_2 \operatorname{cov}(a_1 a_2) + ... + 2x_{n-1} x_n \operatorname{cov}(a_{n-1} a_n)\end{aligned} \qquad (3.1)$$

გავიხსენოთ, რომ $\operatorname{cov}(a_i,a_j) = M(a_i,a_j)$. ჩავწეროთ (3.1) გამოსახულება მატრიცული სახით

$$D(\hat{Y}) = X'_p \operatorname{cov}(a) X_p , \qquad (3.2)$$

სადაც, $X_p = (1, X_{p_1}, X_{p_2}, ..., X_{p_n})$ მოცემული დამოუკიდებელი ცვლადის ვექტორია. $\operatorname{cov}(a)$ – a შეფასების კოვარიაციული მატრიცაა. ვიპოვოთ a პარამეტრის დისპერსია

$$\operatorname{cov}(a) = M[(a - \alpha)(a - \alpha)'],$$

$$a = (X'X)^{-1} X'Y = (X'X)^{-1} X'(X\alpha + \varepsilon) = \alpha + (X'X)^{-1} X\varepsilon .$$

აქედან,

$$\begin{aligned}a - \alpha = (X'X)^{-1} X'\varepsilon , \\ \operatorname{cov}(a) = M[(X'X)^{-1} X'\varepsilon\varepsilon' X (X'X)^{-1}] = M(\varepsilon\varepsilon')(X'X)^{-1}\end{aligned} \qquad (3.3)$$

$M(\varepsilon\varepsilon')$ წარმოადგენს მატრიცას, რომლის ყველა ელემენტი, გარდა მთავარ დიაგონალზე მყოფისა, ნულის ტოლია, რადგან ჩვენი დაშვებით, შემთხვევითი ცდომილებები ერთმანეთთან არ არიან კორელირებული, ე.ი. $M(\varepsilon\varepsilon') = 0$. რაც შეეხება მთავარ დიაგონალზე მყოფ ელემენტებს, ისინი წარმოადგენენ დისპერსიებს და რადგან ჩვენივე დაშვებით, შემთხვევით ცდომილებებს გააჩნიათ ერთი და იგივე დისპერსია, ამიტომ

$$M(\varepsilon\varepsilon') = \sigma^2_{\text{nar}} I ,$$

სადაც, I – ერთეულოვანი მატრიცაა. მიღებული შედეგი ჩავსვათ (3.3)-ში:

$$\operatorname{cov}(a) = \sigma^2_{\text{nar}} (X'X)^{-1} ,$$

ხოლო ეს უკანასკნელი კი – (3.2)-ში:

$$D(\hat{y}) = \sigma^2_{\text{nar}} X_p (X'X)^{-1} X_p .$$

ამრიგად, $\hat{y}$ მნიშვნელობისთვის გვექნება შემდეგი ნდობის ინტერვალი:

$$\hat{y} \pm t_{\alpha;\nu} \sigma_{\text{nar}} \sqrt{X'_p (X'X)^{-1} X_p} .$$

3.3. ცვლადების შერჩევა

რეგრესიული ანალიზი საშუალებას გვაძლევს მოცემულ $X_1, X_2, ..., X_n$ დამოუკიდებელ ცვლადებიდან, რომლებსაც ზოგჯერ კოვარიანტებს უწოდებენ, შევარჩიოთ ის ცვლადები, რომლებიც კავშირშია დამოკიდებულ Y ცვლადთან და გამოვრიცხოთ რეგრესიის განტოლებიდან არაინფორმატიული ანუ ის ცვლადები, რომლებიც ნაკლებად ან სულაც არ არიან კავშირში Y ცვლადთან. განვიხილოთ რამდენიმე მეთოდი.

ყველა შესაძლო რეგრესიათა მეთოდი. თუ დამოუკიდებელ ცვლადთა რაოდენობა n არც ისე დიდია, მაშინ რეკომენდებულია, მოისინჯოს ყველანაირი კომბინაცია რეგრესიის წრფივი მოდელის ასაგებად და რაიმე კრიტერიუმით არჩეულ იქნეს მათ შორის საუკეთესო. მაგალითად, თუ $n = 4$, მაშინ უნდა მოისინჯოს $2^4 = 16$ მოდელი. კერძოდ, 4 – ერთცვლადიანი, 6 – ორცვლადიანი, 4 – სამცვლადიანი, 1 – ოთხცვლადიანი, და ბოლოს ერთი, რომელიც არ შეიცავს არც ერთ ცვლადს, გარდა თავისუფალი წევრისა. საუკეთესოდ ითვლება ის ადეკვატური მოდელი, რომელსაც აქვს უმცირესი ნარჩენი დისპერსია ან, რაც იგივეა, უდიდესი დეტერმინაციის კოეფიციენტი.

გამორიცხვის მეთოდი. თუ დამოუკიდებელ ცვლადთა რაოდენობა საკმაოდ დიდია, მაშინ ყველა შესაძლო რეგრესიათა მეთოდის გამოყენება შეუძლებელია. არსებობს რამდენიმე ალტერნატიული მეთოდი, მათ შორის ცვლადების გამორიცხვის ანუ ცვლადების შერჩევის მიმდევრობითი მეთოდი, რომლის არსი შემდეგში მდგომარეობს: დასაწყისში განიხილება მოდელი, რომელიც შეიცავს ყველა განსახილველ ცვლადებს. რეგრესიის განტოლების კოეფიციენტების შემოწმება სარწმუნოებაზე ხდება $H_0 : a_i = 0$ ნულოვანი ჰიპოთეზის საშუალებით. ამისათვის, თითოეული კოეფიციენტისათვის უნდა განისაზღვროს შემდეგი სტატისტიკა:

$$t_i = \frac{a_i}{\sigma_i},\ i = 1,2,....,n,$$

სადაც $\sigma_i = \sqrt{\sigma_e^2 \cdot s_{ii}}$, s_{ii} წარმოადგენს $(X'X)^{-1}$ მატრიცის მთავარ დიაგონალზე მყოფი ელემენტის მნიშვნელობას. თუ გამოთვლილ სტატისტიკებს შორის მინიმალური t_i ფარდობის აბსოლუტური სიდიდე $|t_i| < t_{\alpha;\nu}$, სადაც $t_{\alpha;\nu}$ სიდიდე აღებულია სტიუდენტის განაწილების ცხრილიდან α მნიშვნელოვნების დონითა და $\nu = m - n - 1$ თავისუფლების ხარისხით, მაშინ ნულოვანი ჰიპოთეზა მიიღება, ე.ი. i-ური კოეფიციენტის მნიშვნელობა ახლოსაა ნულთან და შესაძლებელია მისი გამორიცხვა რეგრესიის განტოლებიდან. შემდეგ ბიჯზე განიხილება მოდელი დარჩენილი $(n - 1)$ ცვლადით და მოწმდება რეგრესიის განტოლების ადეკვატურობა. თუ აღმოჩნდება, რომ რეგრესიის განტოლება არაადეკვატურია, მაშინ გამორიცხული i-ური კოეფიციენტი, ანუ X_i პარამეტრი უნდა დავაბრუნოთ რეგრესიის განტოლებაში.

პროცედურა გაგრძელდება მანამ, სანამ ყველა გამოთვლილი t_i ფარდობის აბსოლუტური სიდიდე არ აღმოჩნდება $\geq t_{\alpha;\nu}$. საბოლოოდ, რეგრესიის განტოლებაში რჩება ის ცვლადები, რომლებიც არ გამოირიცხა პროცედურის ჩატარებისას.

ჩართვის მეთოდი. გამორიცხვის მეთოდის ალტერნატივაა ჩართვის მეთოდი, რომელსაც ბიჯურ რეგრესიას უწოდებენ. პირველ ბიჯზე განიხილება ერთცვლადიანი მოდელები. თითოეულ მოდელში ხდება a_0 და a_i კოეფიციენტების შეფასება და $t_i = \frac{a_i}{\sigma_i}$ სტატისტიკის განსაზღვრა ისე, როგორც ჩართვის მეთოდის დროს. ის ცვლადი, რომელსაც შეესაბამება $|t_i|$ სიდიდის მაქსიმალური მნიშვნელობა, ჩაირთვება მოდელში მხოლოდ იმ პირობით, რომ ეს მაქსიმალური მნიშვნელობა აღემატება $t_{\alpha;\nu}$ კრიტიკულ მნიშვნელობას. დაუშვათ, ასეთი ცვლადია X_1. ამის შემდეგ განიხილება ორცვლადიანი მოდელი X_1 ცვლადის დანარჩენ ცვლადებთან: (X_1,X_2), $(X_1,X_3),\ldots,(X_1,X_n)$. ხდება თითოეული მოდელის შეფასება და დარჩენილ $X_2,X_3,\ldots,X_n$ ცვლადებიდან მოდელში ჩაირთვება ის ცვლადი, რომლისთვისაც $|t_i|$ სიდიდე მაქსიმალურია და აღემატება $t_{\alpha;\nu}$. დაუშვათ ასეთი ცვლადია X_2. შემდეგ განიხილება (X_1,X_2,X_3), (X_1,X_2,X_4), … ,(X_1,X_2,X_n) სამცვლადიანი მოდელები, სადაც ტარდება იგივე პროცედურა და ა.შ. პროცედურა გრძელდება მანამ, სანამ რომელიმე ბიჯზე არ აღმოჩნდება, რომ ყველა t_i ფარდობის აბსოლუტური სიდიდე $\le t_{\alpha;\nu}$ მნიშვნელობაზე.

არსებობს ჩართვის სხვა მეთოდებიც. მაგალითად, კორელაციის კერძო კოეფიციენტების გამოყენების მეთოდი. ამ შემთხვევაში განისაზღვრება Y ცვლადის ყველა დამოუკიდებულ $X_1,X_2,\ldots,X_n$ ცვლადებს შორის კორელაციის კერძო კოეფიციენტები. ის დამოუკიდებული ცვლადი, რომელსაც გააჩნია Y-თან უდიდესი კერძო კორელაციის კოეფიციენტი, ჩაირთვება რეგრესიის მოდელში და რეგრესიის განტოლება მოწმდება ადეკვატურობაზე. თუ აღმოჩნდება, რომ განტოლება არაადეკვატურია, მაშინ მოდელში ჩაირთვება შემდეგი ცვლადი, რომელსაც დამოკიდებულ Y ცვლადთან გააჩნია უდიდესი კორელაციის კერძო კოეფიციენტი, კვლავ მოწმდება განტოლების ადეკვატურობა და ა.შ. მანამ, სანამ არ მივიღებთ რეგრესიის ადეკვატურ განტოლებას.

კომბინირებული მეთოდი. იგი წარმოადგენს ჩართვისა და გამორიცხვის მეთოდების კომბინაციას. პროცედურა იწყება ისევე, როგორც ჩართვის მეთოდი, ე.ი. ცვლადების მიმდევრობით ჩართვით მოდელში, ოღონდ ყოველი ახალი ცვლადის დამატების შემდეგ მოწმდება ადრე ჩართული ცვლადები, კერძოდ, ხომ არ მოიძებნა მათ შორის გამოსარიცხი. მაგალითად, ვთქვათ, მოდელში ჩართულია X_3, X_5, X_6 და X_7 ცვლადები, რომელთაგან X_7 წარმოადგენს ამ ბიჯზე დამატებულ ცვლადს, მაშინ მოწმდება X_3, X_5 და X_6 ცვლადები ისევე, როგორც გამორიცხვის მეთოდშია აღწერილი. თუ ამ ცვლადებიდან რომელიმე აღმოჩნდა ისეთი, რომ სრულდება $|t_i| < t_{\alpha;\nu}$ პირობა, მაშინ ეს ცვლადი მოდელიდან გამოირიცხება და ა.შ.

მიუხედავად ცვლადების შერჩევის მრავალფეროვნებისა, უნდა გვახსოვდეს, რომ არ არსებობს იმის გარანტია, რომ ყოველ ცალკეულ შემთხვევაში მიიღება ჩვენთვის სასურველი ადეკვატური მოდელი. აქედან გამომდინარე, სასურველია თავიდან გამოირიცხოს ის დამოუკიდებული ცვლადები, რომელთა ჩართვა რეგრესიის მოდელში, ამა თუ იმ მოსაზრებით, აზრს მოკლებულია.

3.4 ნაშთთა ანალიზი

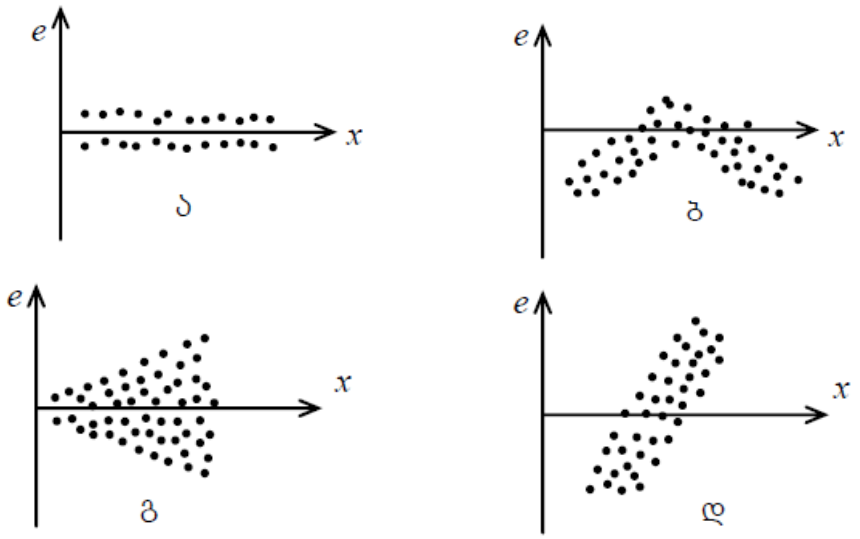
რეგრესიის განტოლების ადეკვატურობისა და კოეფიციენტების შემოწმების გარდა, ხშირად მიმართავენ ნაშთთა ანალიზს. როგორც ვიცით, მოცემულ y_i მნიშვნელობების რეგრესიის განტოლებით გამოთვლილ $\hat{y}_i$ მნიშვნელობებთან გადახრას

$$e_i = y_i - \hat{y}_i, \ i = 1,2,....,n$$

ეწოდება ნარჩენი მნიშვნელობები ანუ ნაშთები.

ნაშთთა ანალიზით შესაძლებელია შემოწმდეს ის ძირითადი დაშვებები გადახრების (შეცდომების) მიმართ, რომლებსაც ეფუძნება წრფივი რეგრესია. ჩვენ დავუშვით, რომ რეგრესიის მრუდი წრფივია, გადახრები e_i არიან დამოუკიდებელნი, გააჩნიათ ნულოვანი საშუალო, ერთნაირი (მუდმივი) დისპერსია და განაწილებულნი არიან ნორმალურად. თუ შერჩეული რეგრესიის მოდელი ადეკვატურია, მაშინ იგი მეტ-ნაკლებად უნდა აკმაყოფილებდეს დაშვების პირობებს. სწორედ ეს იდეა უდევს საფუძვლად ნაშთთა ანალიზს.

ნაშთთა ანალიზი ხორციელდება გრაფიკულად. აიგება $e = f(x)$ გრაფიკული გამოსახულება და თუ მიღებულ წერტილთა ერთობლიობა დაახლოებით ერთნაირადაა განლაგებული x ღერძის მიმართ (თანაბრად ღერძის ზემოთ და ქვემოთ), მაშინ რეგრესიის განტოლება ადეკვატურია და ყველა დაშვება დაახლოებით შესრულებულია (ნახ. ა).



თუ დარღვეულია რეგრესიის მრუდის წრფივობის დაშვება, მაშინ ნაშთთა გრაფიკს დაახლოებით ექნება ისეთი სახე, როგორც ეს ნაჩვენებია ბ ნახაზზე. თუ დარღვეულია დისპერსიის მუდმივობა, მაშინ დაახლოებით გვექნება ბ ნახაზზე წარმოდგენილი სურათი. თუ გადახრები დამოკიდებულია x_i-ზე, მაშინ გვექნება დაახლოებით დ ნახაზზე წარმოდგენილი სურათი.

3.5 ნარჩენების ავტოკორელაციისა და მულტიკოლინეარნობის პრობლემები

რეგრესიულ ანალიზში, უმცირეს კვადრატთა მეთოდის გამოყენებისას, ჩვენ დავუშვით, რომ ε_i ცდომილებები შემთხვევითი დამოუკიდებელი (არაკორელირებული) სიდიდეებია ნულოვანი საშუალოთი. პრაქტიკაში ამ მოთხოვნილების შესრულება ძნელია, განსაკუთრებით დროითი მწკრივებისათვის.

აღმოჩნდა, რომ თუ e_i ნარჩენები ერთმანეთში კორელირებენ, მაშინ ამბობენ, რომ ადგილი აქვს ცდომილების ავტოკორელაციას. მიუხედავად იმისა, რომ უმცირეს კვადრატთა მეთოდი ამ შემთხვევის დროსაც გვაძლევს გადაუადგილებად და საფუძვლიან შეფასებებს, რეგრესიის პარამეტრების განსაზღვრისას ნდობის ინტერვალის განსაზღვრა კარგავს აზრს მისი არასაიმედობის გამო. ამიტომ, თუ აღმოჩნდება ნარჩენების ავტოკორელაციის ეფექტი, საჭიროა გადაისინჯოს რეგრესიის განტოლების მოდელი.

არსებობს ავტოკორელაციის აღმოჩენის მთელი რიგი მეთოდები. ჩვენ განვიხილავთ პირველი რიგის ავტოკორელაციის არსებობის ჰიპოთეზის შემოწმების შედარებით მარტივ და საკმაოდ საიმედო მეთოდს, რომელიც შემოგვთავაზეს დარბინმა და უიტსონმა. ამ ჰიპოთეზის შესამოწმებლად გამოვთვალოთ შემდეგი სტატისტიკა:

$$d = \frac{\sum_{i=1}^{n}(e_i - e_{i+1})^2}{\sum_{i=1}^{n} e_i^2}.$$

n-ის დიდი რაოდენობის დროს $\sum_{i=1}^{n} e_i \approx \sum_{i=1}^{n} e_{i-1}$, მაშინ

$$d \approx \frac{2\sum_{i=2}^{n} e_i^2 - 2\sum_{i=2}^{n} e_i e_{i-1}}{\sum_{i=1}^{n} e_i^2} = 2\left(1 - \frac{\sum_{i=2}^{n} e_i e_{i-1}}{\sum_{i=1}^{n} e_i^2}\right) = 2(1-R),$$

სადაც, R წარმოადგენს პირველი რიგის ავტოკორელაციის კოეფიციენტს და თუ იგი ნულის ტოლია, მაშინ ავტოკორელაცია არ არსებობს. თუ ავტოკორელაცია მთლიანად არსებობს, მაშინ $R = \pm 1$. აქედან გამომდინარე, თუ ავტოკორელაცია არ არსებობს, მაშინ d სიდიდის მნიშვნელობა მიახლოებით 2-ის ტოლია და მთლიანი ავტოკორელაციის არსებობის დროს იგი 0-ის ან 4-ის ტოლია.

ავტოკორელაციის სარწმუნოების დასადგენად სპეციალური ცხრილიდან დამოკიდებული n პარამეტრების რაოდენობისა და m დაკვირვებათა რაოდენობის საშუალებით მოიძებნება d კრიტერიუმის ქვედა d_e და ზედა d_u ზღვრების მნიშვნელობები. თუ გამოთვლილი d სტატისტიკა მოთავსებულია d_u და $(4 - d_u)$ საზღვრებში, მაშინ ჰიპოთეზა ავტოკორელაციის არარსებობის შესახებ მიიღება. თუ d მოთავსებულია d_e და d_u შორის ან $(4 - d_u)$ და $(4 - d_e)$ შორის, მაშინ ჩვენ არა გვაქვს საფუძველი ჰიპოთეზის უარსაყოფად და არც მისაღებად, ე.ი. საქმე გვაქვს განუსაზღვრელობასთან. თუ $d < d_e$, საქმე გვაქვს დადებით ავტოკორელაციასთან, ხოლო როცა $d > (4 - d_e)$ – უარყოფით ავტოკორელაციასთან.

რეგრესიის განტოლების ფორმირებისას, ხშირად ვაწყდებით მულტიკოლინეარნობის პრობლემას. როგორც აღვნიშნეთ, დამოუკიდებელი ცვლადები $X_1, X_2, ..., X_n$ რეგრესიის განტოლებაში უნდა იყვნენ ურთიერთდამოუკიდებელი, მაგრამ ამ პირობის შესრულება პრაქტიკულად საკმაოდ რთულია, განსაკუთრებით, ბიოსამედიცინო კვლევებში. ამ მოვლენას მულტიკოლინეარნობა ეწოდება. თუ ცვლადებს შორის დამოკიდებულება ფუნქციონალურია, მაშინ საქმე გვაქვს მკაცრ მულტიკოლინეარნობასთან, ხოლო თუ დამოკიდებულება არც ისე მკაცრია და გამოვლინდება მიახლოებით, მაშინ მულტიკოლინეარნობა არ არის მკაცრი.

უმცირეს კვადრატთა მეთოდის ერთ-ერთი მოთხოვნა ისაა, რომ დამოუკიდებელ ცვლადებს შორის არ უნდა არსებობდეს წრფივი კავშირი. მულტიკოლინეარნობის არსებობა იწვევს ამ მოთხოვნის დარღვევას. ფორმალურად რეგრესიის განტოლება ამ შემთხვევაში მიიღება, მაგრამ იგი არ არის საიმედო იმ გაგებით, რომ საწყისი მონაცემების უმნიშვნელო ცვლილებამ შეიძლება გამოიწვიოს პარამეტრთა შეფასების მკვეთრი ცვლილებები.

მულტიკოლინეარნობის აღმოჩენის მეთოდებიდან შეგვიძლია განვიხილოთ კორელაციური მატრიცის მეთოდი. კორელაციის კოეფიციენტები, რომლებიც ახლოს არიან ±1 სიდიდესთან, მიგვანიშნებენ მულტიკოლინეარნობის არსებობაზე. უფრო საიმედო მეთოდია $X'X$ მატრიცის დეტერმინატის განსაზღვრა. თუ ეს სიდიდე ნულთან ახლოსაა, მაშინ საქმე გვაქვს მულტიკოლინეარნობასთან. მულტიკოლინეარნობის აღმოსაჩენად შეგვიძლიაგამოვიყენოთ შემდეგი სტატისტიკა:

$$\chi^2 = -\left[m - 1 - \frac{1}{6}(2n+5) \right] \lg\left(\det\left[\tilde{X}'\tilde{X}\right]\right),$$

რომელსაც გააჩნია χ^2 განაწილება $\nu = \frac{n(n-1)}{2}$ თავისუფლების ხარისხით. აქ, m–დაკვირვებათა რაოდენობაა, n– დამოუკიდებელი ცვლადების რაოდენობა. $\left[\tilde{X}'\tilde{X}\right]$ მატრიცის ელემენტები განისაზღვრება საწყისი $[X'X]$ მატრიციდან შემდეგნაირად:

$$\tilde{x}_{ik} = \frac{x_{ik} - \bar{x}_k}{\sigma_k \sqrt{m}},$$

სადაც, $\bar{x}, \sigma_i$ – შესაბამისად i-ური ცვლადის საშუალო არითმეტიკული და საშუალო კვადრატული გადახრაა.

თუ $\chi^2 \geq \chi^2_{\alpha;\nu}$, მაშინ მულტიკოლინეარნობა არ არსებობს, წინააღმდეგ შემთხვევაში, მისი არსებობა სარწმუნოა.

მულტიკოლინეარნობის გამორიცხვა შესაძლებელია რეგრესიის განტოლების სტრუქტურის გადასინჯვით, კერძოდ, ორი დამოკიდებული ცვლადიდან უნდა გამოირიცხოს ერთი. მეორე გზაა ცვლადების გარდაქმნა ისე, რომ ისინი გახდნენ ურთიერთდამოუკიდებელი, მაგალითად, მთავარი კომპონენტების მეთოდის გამოყენებით.

4. მრავალფაქტორიანი დისპერსიული ანალიზი

4.1. დისპერსიული ანალიზის არსი

საშუალოების წყვილ-წყვილი შედარება, თუ ამონარჩევების (ჯგუფების) რაოდენობა დიდია, მიზანშეუწონელია არა მარტო მეთოდოლოგიური შეცდომების, არამედ დიდი გამოთვლითი სამუშაოს გამო. მაგალითად, თუ გვაქვს 7 ამონარჩევი, მაშინ ჩასატარებელი იქნება $C_7^2 = 21$ საშუალოების შედარება. ცხადია, რომ ჯგუფების რაოდენობის ზრდისას, შესადარებელ წყვილთა რაოდენობა ძალზე სწრაფად იზრდება. ასე მაგალითად, 14 ჯგუფისათვის გვექნება 91 შედარება.

გაითვალისწინა რა ეს პრობლემა, რ. ფიშერმა შემოგვთავაზა საშუალოების კომპლექსური შედარების მეთოდი, რომელსაც დისპერსიული ანალიზი (ANOVA – *Analysis of Variance*) ეწოდება. დისპერსიული ანალიზის მეთოდი ეფუძნება საერთო დისპერსიის დაშლაში დამოუკიდებელ მდგენელებად, რომლებიც გამოწვეულია როგორც რეგულირებადი, ისე არარეგულირებადი ფაქტორებით. შემოვიტანოთ რამდენიმე განმარტება.

პარამეტრებს, რომლებიც რაიმე მიზეზით იცვლებიან, ეწოდებათ შედეგობრივი. მიზეზებს, რომლებიც იწვევენ შედეგობრივი პარამეტრების ცვლილებას, ეწოდებათ ფაქტორები. მაგალითად, სხეულის მასა, მოსავლიანობა, მოსწავლეთა მოსწრება და ა.შ. წარმოადგენს შედეგობრივ პარამეტრებს, რომლებზედაც სხვადასხვა ფაქტორები ახდენენ ზემოქმედებას: წამლის ან ტოქსიკური ნივთიერებათა დოზები, სასუქის რაოდენობა, კვების რეჟიმი, ფიზიკური და გონებრივი სავარჯიშოები და სხვა. არსებობს ერთი და იმავე პარამეტრზე მოქმედი მრავალი ფაქტორი, რომელთაგან ცდაში (ექსპერიმენტში) რეგულირებადია მხოლოდ ზოგიერთი მათგანი და მათ რეგულირებადი ფაქტორები ეწოდებათ. ისეთ ფაქტორებს, რომლებიც არ ექვემდებარებიან რეგულირებას, ეწოდებათ არარეგულირებადი, თუმცა მათაც გააჩნიათ გარკვეული ზემოქმედება შედეგობრივ პარამეტრებზე. არარეგულირებად ფაქტორებს მიეკუთვნებიან აგრეთვე სხვა დაუფიქსირებადი ანუ გაუთვალისწინებელი ფაქტორები. ჩვეულებრივ, ყოველი რეგულირებადი ფაქტორი წარმოდგენილია დამოუკიდებელ გრადაციებად (ჯგუფებად), რომელთა რაოდენობა დამოკიდებულია ცდის (ექსპერიმენტის) პირობებზე.

თუ რეგულირებადი ფაქტორი იწვევს მნიშვნელოვან ზეგავლენას შედეგობრივ პარამეტრზე, მაშინ ეს ზეგავლენა აისახება ჯგუფურ საშუალოებზე, რომლებიც სტატისტიკურად ერთმანეთისგან განსხვავებული იქნებიან. თითოეული ჯგუფის შიგნითაც აღინიშნება ცვალებადობა, რომელიც გამოწვეულია არარეგულირებადი ფაქტორით ან ფაქტორებით. ცვალებადობის ეს დამოკიდებულება შეიძლება ასე წარმოვადგინოთ:

$$\sigma^2 = \sigma_f^2 + \sigma_e^2,$$

სადაც, σ^2 – მთელი კომპლექსის საერთო დისპერსიის შეფასებაა; σ_f^2 – ჯგუფთაშორისო (დონეთაშორისო) დისპერსიის შეფასებაა, რომელიც გამოწვეულია რეგულირებადი ფაქტორის ზეგავლენით; σ_e^2 – შიგაჯგუფური (ნარჩენი) დისპერსიის შეფასებაა, რომელიც გამოწვეულია გაუთვალისწინებელი ანუ არარეგულირებადი ფაქტორებით.

რეგულირებადი ფაქტორის ზეგავლენის დასადგენად საჭიროა ფიშერის კრიტერიუმით შემოწმდეს $H:\sigma_f^2=\sigma_e^2$ ნულოვანი ჰიპოთეზა. თუ ნულოვანი ჰიპოთეზა უარყოფილია α მნიშვნელოვნების დონით, მაშინ რეგულირებადი ფაქტორის ზეგავლენა შედეგობრივ პარამეტრზე სარწმუნოა, წინააღმდეგ შემთხვევაში რეგულირებადი ფაქტორის ზეგავლენა შედეგობრივ პარამეტრზე არ შეიმჩნევა ან უმნიშვნელოა.

ამრიგად, დისპერსიული ანალიზი, გარდა საშუალოების ერთდროული შედარებისა, შეისწავლის დაკვირვებებზე (ექსპერიმენტზე) მოქმედი სხვადასხვა ფაქტორების ზეგავლენას, მნიშვნელოვანი ფაქტორების ამორჩევასა და მათი მოქმედებების შეფასებას.

დისპერსიული ანალიზის ჩასატარებლად საჭიროა, რომ დაკვირვებები იყოს შემთხვევითი სიდიდეები, რომელთაც გააჩნიათ ნორმალური განაწილება და ერთნაირი დისპერსია. მხოლოდ ამ შემთხვევაშია შესაძლებელი დისპერსიისა და მათემატიკური ლოდინის სარწმუნოების შეფასება და ნდობის ინტერვალების დადგენა და საერთოდ, დისპერსიული ანალიზის ჩატარება. განვიხილოთ ორფაქტორიანი დისპერსიული ანალიზი

4.2 ორფაქტორიანი დისპერსიული ანალიზი

ვთქვათ, ექსპერიმენტალურ შედეგებზე მოქმედებს ორი A და B ფაქტორი, რომელთაც გააჩნიათ შესაბამისად r და v დონეები. დაკვირვებათა მატრიცა შეიძლება ასე წარმოვადგინოთ:

A \ B	B_1	B_2	. . .	B_j	. . .	B_v	საშ.
A_1	x_{11}	x_{12}	. . .	x_{1j}	. . .	x_{1v}	$\bar{x}_{1\bullet}$
A_2	x_{21}	x_{22}	. . .	x_{2j}	. . .	x_{2v}	$\bar{x}_{2\bullet}$
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
A_i	x_{i1}	x_{i2}	. . .	x_{ij}	. . .	x_{iv}	$\bar{x}_{i\bullet}$
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
A_r	x_{r1}	x_{r2}	. . .	x_{rj}	. . .	x_{rv}	$\bar{x}_{r\bullet}$
$\bar{x}_{\bullet j}$	$\bar{x}_{\bullet 1}$	$\bar{x}_{\bullet 2}$	. . .	$\bar{x}_{\bullet j}$	. . .	$\bar{x}_{\bullet v}$	$\bar{x}$

i-ური დონის A ფაქტორის გადაკვეთა j-ური დონის B ფაქტორთან ქმნის ij-ურ უჯრედს, სადაც ჩაწერილია დაკვირვების შედეგი x_{ij}, მიღებული A და B ფაქტორების ერთდროული მოქმედების დროს. სიმარტივისთვის დავუშვათ, რომ უჯრედში გვაქვს მხოლოდ ერთი დაკვირვება და ფაქტორები ურთიერთდამოუკიდებლები არიან, ე.ი. ურთიერთქმედება გამორიცხულია. მაშინ ორფაქტორიანი დისპერსიული ანალიზის მოდელი შეიძლება ასე წარმოვადგინოთ:

$$x_{ij}=\mu+\gamma_i+g_j+e_{ij},\quad i=1,2,...,r\quad j=1,2,....,v,$$

სადაც μ – საერთო საშუალოა;

γ – ეფექტი, გამოწვეული i-ური დონის A ფაქტორის მიერ;

g_j – ეფექტი, გამოწვეული j-ური დონის B ფაქტორის მიერ;
e_{ij} – ij-ური უჯრედში შედეგების ვარიაცია.

თუ უჯრედში ერთი მნიშვნელობაა, მაშინ $e_{ij} = 0$. განვიხილოთ μ, γ და g შეფასებები. განვსაზღვროთ შემდეგი სიდიდეები:

საერთო საშუალო: $\bar{x} = \frac{1}{r\nu}\sum_{i=1}^{r}\sum_{j=1}^{\nu} x_{ij}$;

საშუალოები დონეების მიხედვით:

$$\bar{x}_{i\bullet} = \frac{1}{\nu}\sum_{j=1}^{\nu} x_{ij} \ , \quad \bar{x}_{\bullet j} = \frac{1}{r}\sum_{i=1}^{r} x_{ij} \ .$$

დისპერსიების შეფასებისათვის განვიხილოთ შემდეგი გამოსახულება:

$$Q = \sum_{i=1}^{r}\sum_{j=1}^{\nu}\left(x_{ij} - \bar{\bar{x}}\right)^2 = \sum_{i=1}^{r}\sum_{j=1}^{\nu}\left(x_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{\bar{x}} + \bar{x}_{i\bullet} - \bar{\bar{x}} + \bar{x}_{\bullet j} - \bar{\bar{x}}\right)^2 =$$

$$= \underbrace{\nu\sum_{i=1}^{r}\left(\bar{x}_{i\bullet} - \bar{\bar{x}}\right)^2}_{Q_A} + \underbrace{r\sum_{j=1}^{\nu}\left(\bar{x}_{\bullet j} - \bar{\bar{x}}\right)^2}_{Q_B} + \underbrace{\sum_{i=1}^{r}\sum_{j=1}^{\nu}\left(x_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{\bar{x}}\right)^2}_{Q_e} \qquad \text{ე.ი.}$$

$$Q = Q_A + Q_B + Q_e,$$

სადაც, Q_A ახასიათებს პარამეტრის ცვლილებას, გამოწვეულს A ფაქტორის მიერ, Q_B – B ფაქტორის მიერ და Q_e – სხვა გაუთვალისწინებელი ფაქტორების მიერ. თუ ცნობილია Q, Q_A, Q_B და Q_e, მაშინ დისპერსიების შეფასებები იქნება:

$$\sigma^2 = \frac{Q}{r\nu - 1}; \ \sigma_A^2 = \frac{Q_A}{r-1}; \ \sigma_B^2 = \frac{Q_B}{\nu - 1}; \ \sigma_e^2 = \frac{Q_e}{(r-1)(\nu-1)} \ .$$

A და B ფაქტორების ზეგავლენის დასადგენად საჭიროა დისპერსიების შედარება ფიშერის კრიტერიუმის გამოყენებით. ამისათვის უნდა განისაზღვროს

$$F_A = \frac{\sigma_A^2}{\sigma_e^2}$$

სიდიდე, რომელსაც გააჩნია, $\nu_1 = r - 1$ და $\nu_3 = (r - 1)(\nu - 1)$ თავისუფლების ხარისხები. ეს სიდიდე უნდა შედარდეს $F_{\alpha;\nu_1\nu_3}$ კრიტიკულ მნიშვნელობას. თუ $F_A \geq F_{\alpha;\nu_1\nu_3}$, მაშინ A ფაქტორის ზეგავლენა სარწმუნოა. წინააღმდეგ შემთხვევაში A ფაქტორის ზეგავლენა უმნიშვნელოა. ანალოგიურად ტარდება B ფაქტორის ზეგავლენის დადგენა.

დისპერსიული ანალიზი უმჯობესია წარმოვადგინოთ შემდეგი ცხრილების სახით:

ისპერსია	კვადრატების ჯამი	თავისუფლების ხარისხი
A ფაქტორი	$Q_A = \nu \sum_i (\bar{x}_{i\bullet} - \bar{x})^2$	$\nu_1 = r - 1$
B ფაქტორი	$Q_B = r \sum_i (\bar{x}_{\bullet j} - \bar{x})^2$	$\nu_2 = \nu - 1$
ნაშთი	$Q_e = \sum_{i,j} (x_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{x})$	$\nu_3 = (r-1)(\nu-1)$
საერთო	$Q = \sum_{i,j} (x_{ij} - \bar{x})^2$	$\nu = r\,\nu - 1$

ცხრილის გაგრძელება

დისპერსია	დის-პერსიების შეფასება	F ფარდ.	F კრიტ.
A ფაქტორი	$\sigma_A^2 = \frac{Q_A}{\nu_1}$	$F_A = \frac{\sigma_A^2}{\sigma_e^2}$	$F_{\alpha;\nu_1\nu_2}$
B ფაქტორი	$\sigma_B^2 = \frac{Q_B}{\nu_2}$	$F_B = \frac{\sigma_B^2}{\sigma_e^2}$	$F_{\alpha;\nu_2\nu_3}$
ნაშთი	$\sigma_e^2 = \frac{Q_e}{\nu_3}$		
საერთო	$\sigma^2 = \frac{Q}{\nu}$		

ჩვენ განვიხილეთ კერძო შემთხვევა, როცა უჯრედში იყო ერთი მონაცემი და ფაქტორებს შორის ურთიერთქმედება გამორიცხული იყო. ზოგადად, უჯრედში შეიძლება იყოს რამდენიმე, როგორც თანაბარი, ისე არათანაბარი რაოდენობის მონაცემები და ფაქტორებს შორის შეიძლება ადგილი ჰქონდეს ურთიერთზემოქმედებას. სასურველია, რომ უჯრედებში მონაცემები იყვნენ ერთი და იგივე რაოდენობის.

ამრიგად, ზოგადი შემთხვევისთვის – ორფაქტორიანი დისპერსიული ანალიზის დროს – ერთი დაკვირვება შეიძლება წარმოვადგინოთ შემდეგნაირად

$$x_{ijk} = \mu + \gamma_i + g_j + \nu_{ij} + e_{ijk},$$

სადაც, μ – საერთო საშუალოა;

γ_i – ეფექტი, გამოწვეული i-ური დონის A ფაქტორის ზეგავლენით;

g_j – ეფექტი, გამოწვეული j-ური დონის B ფაქტორის ზეგავლენით;

ν_{ij} – ეფექტი, გამოწვეული A და B ფაქტორების ურთიერთქმედების შედეგად მიღებული ზეგავლენით;

e_{ij} – უჯრედშიგა ვარიაცია.

თუ უჯრედებში ერთნაირი რაოდენობის მონაცემებია, მაშინ დაკვირვების მატრიცა შეიძლება ასე წარმოვადგინოთ:

A \ B	B_1	B_2	. . .	B_v	$\bar{x}_{i\bullet\bullet}$
A_1	$\bar{x}_{11\bullet}$ $x_{111}, x_{112},..,x_{11n}$	$\bar{x}_{12\bullet}$ $x_{121},x_{122},..,x_{12n}$	. . .	$\bar{x}_{1v\bullet}$ $x_{1v1},x_{1v2},..,x_{1vn}$	$\bar{x}_{1\bullet\bullet}$
A_2	$\bar{x}_{21\bullet}$ $x_{211}, x_{212},..,x_{21n}$	$\bar{x}_{22\bullet}$ $x_{221},x_{222},..,x_{22n}$	. . .	$\bar{x}_{2v\bullet}$ $x_{2v1},x_{2v2},..,x_{2vn}$	$\bar{x}_{2\bullet\bullet}$
. . .	. . .	. . .	. . .	. . .	. . .
A_r	$\bar{x}_{r1\bullet}$ $x_{r11},x_{r12},...,x_{r1n}$	$\bar{x}_{r2\bullet}$ $x_{r21},x_{r22},...,x_{r2n}$	. . .	$\bar{x}_{rv\bullet}$ $x_{rv1},x_{rv2},...,x_{rvn}$	$\bar{x}_{r\bullet\bullet}$
$\bar{x}_{\bullet j\bullet}$	$\bar{x}_{\bullet 1\bullet}$	$\bar{x}_{\bullet 2\bullet}$	. . .	$\bar{x}_{\bullet v\bullet}$	$\bar{x}$

აქ $x_{111},x_{112},...,x_{rvn}$ გამოსაკვლევი პარამეტრის დაკვირვებებია. დისპერსიული ანალიზის ჩატარებისათვის საჭიროა გამოვთვალოთ შემდეგი მნიშვნელობები:

– უჯრედის საშუალო მნიშვნელობა

$$\bar{x}_{ij\bullet} = \frac{1}{n}\sum_{k=1}^{n} x_{ijk} \ ;$$

– სტრიქონების (A ფაქტორი) საშუალო მნიშვნელობები

$$\bar{x}_{i\bullet\bullet} = \frac{1}{v}\sum_{j=1}^{v} \bar{x}_{ij\bullet} \ ;$$

– სვეტების (B ფაქტორი) საშუალო მნიშვნელობები

$$\bar{x}_{\bullet j\bullet} = \frac{1}{r}\sum_{i=1}^{r} \bar{x}_{ij\bullet} \ ;$$

– საერთო საშუალო

$$\bar{x} = \frac{1}{r\,v}\sum_{i=1}^{r}\sum_{j=1}^{n} \bar{x}_{ij\bullet} \ ,$$

სადაც, r, v – შესაბამისად A და B ფაქტორების დონეების რაოდენობებია. გარდა ამისა, განისაზღვრება შემდეგი კვადრატების ჯამი, ანუ დევიატები:

$$Q_A = v\,n\sum_{i=1}^{r}\left(\bar{x}_{i\bullet\bullet} - \bar{x}\right)^2;\ Q_B = rn\sum_{j=1}^{v}\left(\bar{x}_{\bullet j\bullet} - \bar{x}\right)^2;\ Q_e = \sum_{i=1}^{r}\sum_{j=1}^{v}\sum_{k=1}^{n}\left(x_{ijk} - \bar{x}_{ij\bullet}\right)^2;$$

$$Q_{AB} = n\sum_{i=1}^{n}\sum_{j=1}^{v}\left(\bar{x}_{ij\bullet} - \bar{x}_{i\bullet\bullet} - \bar{x}_{\bullet j\bullet} + \bar{x}\right)^2;\quad Q = \sum_{i=1}^{r}\sum_{j=1}^{v}\sum_{k=1}^{n}\left(x_{ijk} - \bar{x}\right)^2;$$

ე.ი.
$$Q = Q_A + Q_B + Q_{AB} + Q_e,$$

სადაც, Q_A-სა და Q_B-ს გააჩნია იგივე მნიშვნელობები, რაც წინა შემთხვევის დროს. Q_{AB} – არის კვადრატების ჯამი, რომელიც აფასებს A და B ფაქტორების ურთიერთქმედებას, Q_e – კვადრატების ჯამი, რომელიც აფასებს უჯრედშიგა ვარიაციას.

ამის შემდეგ უნდა განისაზღვროს დისპერსიების შეფასებები:

$$\sigma^2 = \frac{Q}{r\,\nu\,n-1};\ \sigma_A^2 = \frac{Q_A}{r-1};\ \sigma_B^2 = \frac{Q_B}{\nu-1};$$
$$\sigma_{AB}^2 = \frac{Q_{AB}}{(\nu-1)(r-1)};\ \sigma_e^2 = \frac{Q_e}{r\,\nu(n-1)}$$

და სათანადო ფიშერის კრიტერიუმები:

$$F_A = \frac{\sigma_A^2}{\sigma_e^2};\quad F_B = \frac{\sigma_B^2}{\sigma_e^2};\quad F_{AB} = \frac{\sigma_{AB}^2}{\sigma_e^2}.$$

შესაბამისი ნულოვანი ჰიპოთეზების შემოწმებით დგინდება *A, B* და *AB* ფაქტორების ზეგავლენის ან უმნიშვნელო ზეგავლენის ფაქტები.

ორფაქტორიანი დისპერსიული ანალიზი უმჯობესია წარმოვადგინოთ შემდეგი ცხრილების სახით:

დისპერსია	კვადრატების ჯამი	თავისუფლების ხარისხი
A ფაქტორი	$Q_A = \nu n\sum_i \left(\bar{x}_{i\bullet\bullet} - \bar{x}\right)^2$	$\nu_1 = r-1$
B ფაქტორი	$Q_B = rn\sum_j \left(\bar{x}_{\bullet j\bullet} - \bar{x}\right)^2$	$\nu_2 = \nu - 1$
AB ფაქტორები	$Q_{AB} = n\sum_{i,j} \left(\bar{x}_{ij\bullet} - \bar{x}_{i\bullet\bullet} - \bar{x}_{\bullet j\bullet} + \bar{x}\right)^2$	$\nu_3 = (\nu-1)(r-1)$
ნაშთი	$Q_e = \sum_{i,j,k} \left(x_{ijk} - \bar{x}_{ij\bullet}\right)^2$	$\nu_4 = r\,\nu(n-1)$
საერთო	$Q = \sum_{i,j,k} \left(x_{ijk} - \bar{x}\right)^2$	$\nu = r\,\nu\,n - 1$

ცხრილის გაგრძელება

დისპერსია	დისპერსიების შეფასება	*F* ფარდობა	*F* კრიტიკ.
A ფაქტორი	$\sigma_A^2 = \frac{Q_A}{\nu_1}$	$F_A = \frac{\sigma_A^2}{\sigma_e^2}$	$F_{\alpha;\nu_1\nu_4}$
B ფაქტორი	$\sigma_B^2 = \frac{Q_B}{\nu_2}$	$F_B = \frac{\sigma_B^2}{\sigma_e^2}$	$F_{\alpha;\nu_2\nu_4}$
AB ფაქტორ-ფაქტორები	$\sigma_{AB}^2 = \frac{Q_{AB}}{\nu_3}$	$F_{AB} = \frac{\sigma_{AB}^2}{\sigma_e^2}$	$F_{\alpha;\nu_3\nu_4}$
ნაშთი	$\sigma_e^2 = \frac{Q_e}{\nu_4}$		
საერთო	$\sigma^2 = \frac{Q}{\nu}$		

შედეგობრივ პარამეტრზე ამა თუ იმ ფაქტორის ან ფაქტორების ერთობლივი ზემოქმედების სიძლიერე შეიძლება განისაზღვროს შემდეგი ფორმულებით:

$$h_A^2 = \frac{\widehat{\sigma}_A^2}{\sigma_y^2}; \quad h_B^2 = \frac{\widehat{\sigma}_B^2}{\sigma_y^2}; \quad h_{AB}^2 = \frac{\widehat{\sigma}_{AB}^2}{\sigma_y^2},$$

სადაც,

$$\widehat{\sigma}_A^2 = \frac{\sigma_A^2 - \sigma_e^2}{\nu n}; \quad \widehat{\sigma}_B^2 = \frac{\sigma_B^2 - \sigma_e^2}{rn}; \quad \widehat{\sigma}_{AB}^2 = \frac{\sigma_{AB}^2 - \sigma_e^2}{n};$$

$$\sigma_y^2 = \widehat{\sigma}_A^2 + \widehat{\sigma}_B^2 + \widehat{\sigma}_{AB}^2 + \sigma_e^2,$$

σ_e^2 – ნარჩენი დისპერიაა.

თუ რომელიმე რეგულირებადი ფაქტორის ან ფაქტორთა ერთობლივი ზეგავლენა შედეგობრივ პარამეტრზე არ დასტურდება, მაშინ მისი შესაბამისი კომპონენტი σ_y^2 გამოსახულებაში უნდა გამოირიცხოს.

მაგალითი. გამოკვლეულ იქნა სამი ტიპის მიკროელემენტის ზეგავლენა ძროხის რძის ცხიმიანობაზე. ექსპერიმენტი ჩატარდა ერთნაირი ასაკის ოთხი სხვადასხვა ჯიშის ცხოველთა ჯგუფზე. მონაცემები მოყვანილია შემდეგ ცხრილში:

ძროხ. ჯიში	რძის ცხიმიანობა %-ში								
	A_1			A_2			A_3		
B_1	2,1	2,0	3,4	2,8	2,6	3,0	2,4	2,1	2,8
B_2	4,0	3,2	4,1	3,9	4,1	4,5	3,0	3,9	4,2
B_3	3,0	2,8	2,7	3,5	4,0	2,8	4,8	3,1	2,9
B_4	3,4	3,0	2,9	3,0	2,9	3,0	3,3	2,8	3,0

აქ *A*-თი აღნიშნულია მიკროელემენტები, ხოლო *B*-თი – სხვადასხვა ჯიშის ძროხები. *A* ფაქტორის გრადაციების რაოდენობაა $r = 3$, ხოლო *B* ფაქტორის – $\nu = 4$. დისპერსიული ანალიზის შედეგები მოყვანილია შემდეგ ცხრილში:

დისპერსია	კვადრატების ჯამი	თავისუფლების ხარისხი	დისპერსიების შეფასება	*F* ფარდობა	*F* კრიტიკ.
A ფაქტორი	$Q_A = 0,51$	$\nu_1 = 2$	$\sigma_A^2 = 0,26$	$F_A = 1,0$	$F_{0,05;2,24}=3,4$
B ფაქტორი	$Q_B = 7,94$	$\nu_2 = 3$	$\sigma_B^2 = 2,65$	$F_B = 10,2$	$F_{0,05;3,24}=3,0$
AB ფაქტორები	$Q_{AB} = 1,10$	$\nu_3 = 6$	$\sigma_{AB}^2 = 0,18$	$F_{AB} = 1,4$	$F_{0,05;6,24}=3,8$
ნაშთი	$Q_e = 6,23$	$\nu_4 = 24$	$\sigma_e^2 = 0,26$		
საერთო	$Q = 15,78$	$\nu = 35$			

როგორც ამ ცხრილიდან ჩანს, მხოლოდ *B* ფაქტორის დროს ხდება ჰიპოთეზის უარყოფა. ეს იმას ნიშნავს, რომ ამ ჯიშის ცხოველებს გააჩნიათ რძის ცხიმიანობაზე მიდრეკილება, რომელიც ალბათ შთამომავლობით უნდა აიხსნას და ამიტომ მიკროელემენტების ზეგავლენას იგი არ ექვემდებარება. ალბათ, ამიტომაა, რომ *A* და *B* ფაქტორების ურთიერთქმედებაც ვერ ახდენს რძის ცხიმიანობაზე ზეგავლენას.

რადგან განხილული ორი ფაქტორიდან მხოლოდ *B* ფაქტორი (ძროხის ჯიში) მოქმედებს რძის ცხიმიანობაზე, ამიტომ შეგვიძლია რძის ცხიმიანობაზე ძროხის ჯიშის ზემოქმედების სიძლიერის განსაზღვრა. ამისათვის გვაქვს: $\sigma_B^2 = 2{,}65$; $\sigma_e^2 = 0{,}26$; $n = 3$; $r = 3$. თუ გამოვიყენებთ ზემოთ მოყვანილ ფორმულებს, მაშინ მივიღებთ:

$$\hat{\sigma}_B^2 = \frac{\sigma_B^2 - \sigma_e^2}{rn} = \frac{2{,}65 - 0{,}26}{3 \cdot 3} = 0{,}266;$$

$$h_B^2 = \frac{\hat{\sigma}_B^2}{\sigma_y^2} = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_B^2 + \sigma_e^2} = \frac{0{,}266}{0{,}266 + 0{,}26} = 0{,}50 \qquad \text{ანუ} \quad 50\%.$$

ამრიგად, რძის ცხიმიანობაზე ძროხის ჯიშის ზემოქმედების სიძლიერე 50%-ის ტოლია.

ანალოგიურად ტარდება სამი, ოთხი და ზოგადად, მრავალფაქტორიანი დისპერსიული ანალიზი.

5. მთავარი კომპონენტების მეთოდი

პრაქტიკაში ძალიან ხშირად გვხვდება ისეთი სიტუაცია, როცა პარამეტრების რაოდენობა ძალზე დიდია და, მიუხედავად ამისა, საჭიროა საწყისი მონაცემების სტატისტიკური დამუშავება და გარკვეული გადაწყვეტილების მიღება. აქედან გამომდინარე, საჭიროა საწყისი ინფორმაციის შეკუმშული სახით წარმოდგენა, ანუ შესასწავლი ობიექტის აღწერა მცირე რაოდენობის განზოგადებული მაჩვენებლებით, მაგალითად, მთავარი კომპონენტებით ან ფაქტორებით. მთავარი კომპონენტები წარმოადგენენ მეტად მოსახერხებელ გამსხვილებულ მაჩვენებლებს, რომლებიც ასახავენ ობიექტის (პროცესის) იმ შინაგან კანონზომიერების აღწერას, რაც შეუძლებელია დაკვირვებების საშუალებით.

მთავარი კომპონენტების მეთოდით შესაძლებელია შემდეგი ამოცანების გადაწყვეტა.

1. შესასწავლ მოვლენაში ობიექტურად არსებული ფარული კანონზომიერების გამოვლენა;

2. შესასწავლი პროცესის აღწერა მცირე რაოდენობის მთავარი კომპონენტებით, რომელთა რიცხვი გაცილებით ნაკლებია საწყისი ცვლადების რაოდენობაზე. ამ შემთხვევაში, მთავარი კომპონენტები პროცესს ადეკვატურად ასახავენ უფრო კომპაქტური ფორმით და შეიცავენ საშუალოდ უფრო მეტ ინფორმაციას, ვიდრე უშუალოდ გაზომვადი ცვლადები;

3. ცვლადების მთავარ კომპონენტებთან სტატისტიკური კავშირის გამოვლენა და შესწავლა, რაც საშუალებას იძლევა უფრო აქტიურად ვიმოქმედოთ პროცესზე მისი ეფექტური ფუნქციონირებისთვის;

4. პროცესის განვითარების ტენდენციის პროგნოზირება რეგრესიის განტოლებით, რომელიც აგებულია მთავარი კომპონენტების საშუალებით. პროგნოზირების ასეთ მეთოდს გააჩნია გარკვეული უპირატესობა კლასიკურ რეგრესიულ ანალიზთან შედარებით, განსაკუთრებით იმ შემთხვევის დროს, როცა საქმე გვაქვს მულტიკოლინეარნობის პრობლემასთან.

მთავარი კომპონენტების მოდელი. ვთქვათ, მოცემულია n შემთხვევითი ცვლადები $X_1,X_2,...,X_n$, რომლებსაც გააჩნიათ $\mu=(\mu_1,\mu_2,...,\mu_n)'$ საშუალოების ვექტორი და კოვარიაციული მატრიცა $S_{n,n}$. საჭიროა, განისაზღვროს ამ ცვლადების ურთიერთკავშირი, ანუ სტრუქტურული დამოკიდებულება. სტრუქტურული დამოკიდებულების ერთ-ერთ მეთოდს წარმოადგენს მთავარი კომპონენტების მეთოდი, რომლის ძირითადი ამოცანა შეიძლება ასე ჩამოვაყალიბოთ: უნდა მოიძებნოს საწყისი $X_1,X_2,...,X_n$ ცვლადების ისეთი წრფივი კომბინაცია

$$Y_i=\sum_{j=1}^{n}a_{ij}x_{ij}=\sum_{j=1}^{n}a_{ij}X_j,\quad i=1,2,...,m \tag{5.1}$$

სადაც, a_{ij} – წარმოადგენს წონით კოეფიციენტებს, როცა სრულდება შემდეგი პირობები:

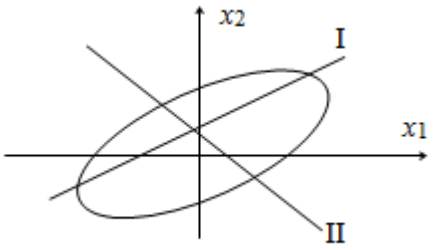
$$\operatorname{corr}(Y_iY_j)=0,\quad i,j=1,2,...,n,\ i\neq j$$

$$\operatorname{var}(Y_1)\geq\operatorname{var}(Y_2)\geq...\geq\operatorname{var}(Y_m);$$

$$\sum_{i=1}^{n}\operatorname{var}(Y_i)=\sum_{i=1}^{n}s_{ii}\ .$$

როგორც ამ ფორმულიდან ჩანს, ახალი ცვლადები $Y_1,Y_2,...,Y_m$, რომლებსაც მთავარ კომპონენტებს უწოდებენ, ერთმანეთის მიმართ არიან არაკორელირებული და რანჟირებული – დისპერსიის კლებადობის მიხედვით. უნდა აღინიშნოს, რომ ჯამური დისპერსია გარდაქმნის შემდეგ არ იცვლება. აქედან გამომდინარე, Y_i ცვლადების პირველ q ქვესიმრავლეზე მოდის საერთო დისპერსიის ძირითადი ნაწილი და ამიტომ შესაძლებელი ხდება საწყისი ცვლადების დამოკიდებულების სტრუქტურის შეკვეცილი აღწერა.

იმისათვის, რომ უკეთ გავერკვეთ მთავარი კომპონენტების მეთოდის არსში, განვიხილოთ მისი **გეომეტრიული ინტერპრეტაცია.** დავუშვათ, რომ ორი X_1 და X_2 შემთხვევითი ცვლადი ნორმალურად არის განაწილებული $\mu=(\mu_1,\mu_2)$ საშუალოების ვექტორითა და S კოვარიაციული მატრიცით. ამ განაწილების სიმკვრივის ელიფსოიდი, რომლის ცენტრი მოთავსებულია (μ_1,μ_2) წერტილში, წარმოდგენილია შემდეგ ნახაზზე:



პირველ მთავარ კომპონენტს $Y_1 = \alpha_{11}x_1 + \alpha_{12}x_2$ შეესაბამება ელიფსოიდის I ღერძი, რომლის სიგრძის ნახევარი ტოლია $\sqrt{\lambda_1}$ სიდიდისა, სადაც, λ_1 არის კოვარიაციული მატრიცის მაქსიმალური საკუთრივი მნიშვნელობა. რადგან X გააჩნია ორგანზომილებიანი ნორმალური განაწილება, ამიტომ მას გააჩნია აგრეთვე II პატარა ღერძი, რომელიც I ღერძის პერპენდიკულარულია და იგი შეესაბამება მეორე მთავარ კომპონენტს $Y_2 = \alpha_{21}x_1 + \alpha_{22}x_2$, რომლის სიგრძე პროპორციულია $\sqrt{\lambda_2}$ სიდიდისა. ამრიგად, ელიფსოიდის, როგორც I, ასევე II ღერძი, განისაზღვრება $X\alpha_1$ და $X\alpha_2$ სიდიდეებით, სადაც, α_1 და α_2 საკუთრივი ვექტორებია, რომლებიც შეესაბამება S მატრიცის λ_1 და λ_2 საკუთრივ მნიშვნელობებს.

თუ X_1 და X_2 ერთმანეთის მიმართ დადებით კორელაციურ დამოკიდებულებაშია, მაშინ რაც უფრო იზრდება კორელაცია ცვლადებს შორის, მით უფრო უახლოვდება ელიფსოიდი I წრფეს და თუ X_1 და X_2 შორის დამოკიდებულება ფუნქციონალურია, მაშინ ელიფსოიდი გარდაიქმნება წრფედ.

მთავარი კომპონენტების განსაზღვრა. მთავარი კომპონენტის მეთოდი მდგომარეობს α_{ij} კოეფიციენტების მოძებნაში. (5.1) გამოსახულება მატრიცული სახით ასე ჩაიწერება: $Y = X\alpha$. მოცემული α-ს დროს ამ გამოსახულების დისპერსია ტოლია:

$$\mathrm{var}(Y) = \mathrm{var}(X\alpha) = \alpha' \mathrm{var}(X)\alpha = \alpha' S\alpha ,$$

სადაც, S მოცემული საწყისი ცვლადების კოვარიაციული მატრიცაა. თუ საწყისი მონაცემები ნორმირებულია, მაშინ გვექნება კორელაციური მატრიცა

$$R = \frac{1}{n-1} XX' .$$

მთავარი კომპონენტების მეთოდის პირველი ამოცანა მდგომარეობს Y_1 კომპონენტის მოძებნაში, რომელსაც გააჩნია უდიდესი დისპერსია. ზოგადად, ამ ამოცანის გადაწყვეტა სათანადო შეზღუდვების შემოტანის გარეშე შეუძლებელია. მაგალითად, თუ ფიქსირებულ α-თვის რაღაც c მუდმივის დროს მივიღებთ $\alpha^* = c\alpha$ სიდიდეს, სადაც c -ს ზრდასთან ერთად უსასრულოდ იზრდება დისპერსიაც. ეს მოვლენა თავიდან რომ ავიცილოთ, საჭიროა α ვექტორის ნორმირება ისე, რომ

$$\alpha'\alpha = \alpha_1^2 + \alpha_2^2 + \ldots + \alpha_n^2 = 1 .$$

მაშინ ამოცანა ჩამოყალიბდება შემდეგნაირად: მოვახდინოთ $\alpha' S\alpha$ გამოსახულების მაქსიმიზაცია, როცა $\alpha'\alpha = 1$. დავუშვათ

$$\varphi = \alpha' S\alpha - \lambda(\alpha'\alpha - 1) ,$$

სადაც, λ – ლაგრანჟის მამრავლია. φ-ის კერძო წარმოებულის ნულთან გატოლების შემდეგ

$$\frac{\partial \varphi}{\partial \alpha} = 2S\alpha - 2\lambda\alpha = 0$$

მივიღებთ შემდეგ განტოლებას:

$$(S - \lambda I)\alpha = 0,$$

რომელიც წარმოადგენს კლასიკური ტიპის განტოლებას და მას გააჩნია ამონახსნი მხოლოდ იმ შემთხვევაში, როცა

$$\det|S - \lambda I| = 0.$$

ამრიგად, მიღებული განტოლების ამოხსნისთვის საჭიროა მოიძებნოს S მატრიცის $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ მახასიათებელი ფესვები.

იმისათვის, რომ განვსაზღვროთ, ამ ფესვებიდან რომელი გამოვიყენოთ მახასიათებელი (საკუთრივი) ვექტორის შესარჩევად, რომელიც მოახდენს $\alpha' S\alpha$ გამოსახულების მაქსიმიზაციას, საჭიროა, განტოლების მარცხენა მხარე გავამრავლოთ α'-ზე, მაშინ, თუ მხედველობაში მივიღებთ $\alpha' S\alpha = 1$, გვექნება:

$$\alpha'(S - \lambda I)\alpha = \alpha' S\alpha - \lambda = 0, \qquad \text{ე.ი.} \qquad \alpha' S\alpha = \lambda,$$

მაგრამ ჩვენ ვიცით, რომ $\alpha' S\alpha = \mathrm{var}(Y)$. ე.ი. λ წარმოადგენს დისპერსიას.

ამრიგად, იმისათვის, რომ მოვახდინოთ Y კომპონენტის დისპერსიის მაქსიმიზაცია, უნდა ავიღოთ მახასიათებელი ფესვებიდან უდიდესი მნიშვნელობის ფესვი, კერძოდ λ_1 და მისი შესაბამისი საკუთრივი ვექტორი α_1. მაშინ პირველი მთავარი კომპონენტი მიიღებს შემდეგ სახეს:

$$Y_1 = X\alpha_1,$$

რომლის დისპერსია იქნება λ_1.

ზოგადად, როცა საქმე გვაქვს n ცვლადთან, პირველი მთავარი კომპონენტი Y_1 წარმოადგენს n ცვლადების წრფივ კომბინაციას, რომელთა კოეფიციენტები ტოლია ნორმირებული საკუთრივი ვექტორის კომპონენტებისა, რომელიც, თავის მხრივ, შეესაბამება R ან S მატრიცის უდიდეს საკუთრივ მნიშვნელობას.

მეორე მთავარი კომპონენტი Y_2 წარმოადგენს საწყისი n ცვლადების წრფივ კომბინაციას კოეფიციენტებით, რომლებიც ნორმირებული საკუთრივი ვექტორის კომპონენტების ტოლია და იგი შეესაბამება λ_2 მახასიათებელ მნიშვნელობას, რომელიც წარმოადგენს λ_1-ის შემდეგ უდიდეს მნიშვნელობას. ანალოგიურად განისაზღვრება მესამე მთავარი კომპონენტი და ა.შ. n-ური კომპონენტის ჩათვლით. თითოეული კომპონენტის დისპერსია შესაბამისად ტოლია მახასიათებელი λ_i, $i = 1,2,\ldots,n$ ფესვებისა და თოთოეული კომპონენტი არ არის ერთმანეთის მიმართ დამოკიდებული. ვაჩვენოთ ეს ბოლო დამოკიდებულება.

ცნობილია თეორემა, რომლის თანახმად, ნებისმიერი სიმეტრიული S მატრიცისთვის არსებობს ისეთი ორთოგონალური მატრიცა α, რომლისთვისაც სრულდება შემდეგი ტოლობა:

$$\alpha' S\alpha = \begin{bmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \lambda_n \end{bmatrix}. \qquad (5.2)$$

ამასთან, თუ S დადებითად განსაზღვრული მატრიცაა, მაშინ ყველა $\lambda_i > 0$ და $|S| > 0$. რადგან α წარმოადგენს S მატრიცის საკუთრივ ვექტორებს და როგორც მიღებული (5.2) გამოსახულებიდან ჩანს, $\mathrm{cov}(\alpha_{ij}) = 0$, $i = 1,2,\ldots,n$, $j = 1,2,\ldots,n$, $i \neq j$, ამიტომ კომპონენტები ურთიერთდამოუკიდებელი არიან.

თითოეული კომპონენტის ჯამური დისპერსიის ფარდობითი წილი განისაზღვრება შემდეგი გამოსახულებით:

$$q_i = \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i}, \qquad i = 1,2,\ldots,n. \tag{5.3}$$

პრაქტიკულად, თუ ჯამური დისპერსიის 80-85% მოდის პირველი k რაოდენობის კომპონენტებზე, მაშინ დანარჩენი კომპონენტები $k + 1, k + 2, \ldots, n$ შეიძლება მხედველობაში არ მივიღოთ და ამით მოვახდინოთ სივრცის განზომილების შემცირება. აქედან გამომდინარე, საჭიროა ჩამოვაყალიბოთ ისეთი კრიტერიუმი, რომელიც ნაკლებდისპერსიანი კომპონენტების ანალიზიდან გამორიცხვის საშუალებას მოგვცემს.

მთავარი კომპონენტის მეთოდის გამოყენება მიზანშეწონილია იმ შემთხვევაში, როცა საწყის ცვლადებს გააჩნიათ საერთო ფიზიკური ბუნება და იზომებიან ერთი და იმავე ფიზიკურ ერთეულებში. თუ ცვლადები სხვადასხვა ფიზიკური ბუნებისაა, მაშინ აუცილებელია მათი ნორმირება, მაგალითად, შემდეგნაირად:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}, \quad i = 1,2,\ldots,m, \quad j = 1,2,\ldots,n,$$

სადაც, $\bar{x}_i$ საშუალო მნიშვნელობებია, ხოლო σ_i – საშუალო კვადრატული გადახრები. ასეთი ნორმირების შემდეგ კოვარიაციული მატრიცა გარდაიქმნება კორელაციურ მატრიცად.

კორელაცია X_i ცვლადსა და Y_j მთავარ კომპონენტს შორის განისაზღვრება შემდეგნაირად:

$$\mathrm{corr}(Y_i X_j) = \frac{\alpha_{ij}\sqrt{\lambda_i}}{\sigma_j}, \tag{5.4}$$

სადაც, σ_i – სტანდარტული გადახრაა. ამრიგად, თუ გვინდა შევადაროთ თითოეული X_j ცვლადის წილი Y_i კომპონენტის ფორმირებაში, საჭიროა შევადაროთ $\frac{\alpha_{ij}}{\sigma_j}$ მნიშვნელობები. თუ კორელაციური მატრიცა ცნობილია, მაშინ საკმარისია α_{ij} კოეფიციენტების შედარება. კერძოდ, იმ X_j ცვლადს, რომელსაც გააჩნია უდიდესი α_{ij} კოეფიციენტი, მიუძღვის უდიდესი წილი Y_i მთავარი კომპონენტის ფორმირებაში.

მთავარი კომპონენტების ინტერპრეტაციისთვის, კერძოდ, მასში არსებული ინფორმაციის გამოსავლენად, ხშირად (5.1) გამოსახულებაში a_{ij} წონითი კოეფიციენტების მაგივრად იყენებენ კორელაციის კოეფიციენტებს, რომლებიც განისაზღვრება (5.4) ფორმულით.

ჰიპოთეზების შემოწმება. დავუშვათ, რომ კოვარიაციული (კორელაციური) მატრიცის მახასიათებელი რიცხვები ერთმანეთის ტოლია, ანუ (5.3) გამოსახულებიდან მიღებული q_i მნიშვნელობები ერთმანეთის ტოლია. ორი X_1 და X_2 ცვლადების დროს ეს ნიშნავს ელიფსოიდის გარდაქმნას წრედ, ე.ი. ორივე I და II ღერძები ერთმანეთის ტოლია. მრავალგანზომილებიანი სისტემის დროს საქმე გვაქვს სფეროსთან. ამრიგად, თუ ნულოვანი ჰიპოთეზა

$$H_0 : \lambda_1 = \lambda_2 = \cdots = \lambda_n$$

სამართლიანია, მაშინ მუდმივი სიმკვრივის ელიფსოიდი გარდიქმნება მუდმივი სიმკვრივის სფეროდ და მაშინ ნულოვანი ჰიპოთეზის შემოწმება შეიძლება გავიხილოთ როგორც სფეროზე შემოწმების ჰიპოთეზად. ასეთი ჰიპოთეზის შესამოწმებლად განვიხილოთ შემდეგი სტატისტიკა:

$$\chi^2 = -(m-1)\left[\sum_{i=1}^{n}\ln\lambda_i - n\ln\left(\frac{1}{n}\sum_{i=1}^{n}\lambda_i\right)\right], \qquad (5.5)$$

რომელსაც გააჩნია χ^2 განაწილება $\nu=0{,}5n(n+1)-1$ თავისუფლების ხარისხით. თუ $\chi^2 < \chi^2_{\alpha;\nu}$, მაშინ ნულოვანი ჰიპოთეზა მიიღება.

დავუშვათ, რომ პირველ k მთავარ კომპონენტზე მოდის ჯამური დისპერსიის უდიდესი ნაწილი. ჩვენ გვაინტერესებს, დარჩენილი კომპონენტები განსხვავდებიან თუ არა ერთმანეთისგან. თუ ისინი არ განსხვავდებიან, მაშინ მათი გამორიცხვა შემდგომი ანალიზიდან მიზანშეწონილია. ამრიგად, ჩამოვაყალიბოთ შემდეგი ნულოვანი ჰიპოთეზა:

$$H_0 : \lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_n ,$$

მაშინ (5.5) სტატისტიკას აქვს შემდეგი სახე:

$$\chi^2 = -(m-1)\left[\sum_{j=k+1}^{n}\ln\lambda_j - q\ln\left(\frac{1}{q}\sum_{j=K+1}^{n}\lambda_j\right)\right],$$

სადაც, $q = n - k$. თუ $\chi^2 < \chi^2_{\alpha;\nu}$, მაშინ ნულოვანი ჰიპოთეზა მიიღება და ბოლო q რაოდენობის მთავარი კომპონენტები შეიძლება გამოვრიცხოთ ანალიზიდან.

უნდა გვახსოვდეს, რომ ამ ჰიპოთეზების შემოწმების კრიტერიუმები მეტად მგრძნობიარეა ნორმალური განაწილების მიმართ. განსაკუთრებით საეჭვოა მათი გამოყენება დროითი მწკრივების (მაგალითად, ბიოსიგნალების) მიმართ, რადგან მონაცემების დამოუკიდებლობა ამ შემთხვევაში იშვიათობას წარმოადგენს.

მთავარი კომპონენტების მეთოდი გამოიყენება იმ შემთხვევაში, როდესაც საწყისი ცვლადები ურთიერთდამოკიდებული არიან. როცა ცვლადები ურთიერთდამოუკიდებელია, მაშინ ამ მეთოდის გამოყენებას აზრი არა აქვს, რადგან ამ დროს ფაქტიურად ხდება საწყისი ცვლადების რანჟირება დისპერსიების კლების მიხედვით. აქედან გამომდინარე, სასურველია შევამოწმოთ პარამეტრების დამოუკიდებლობის ჰიპოთეზა. ამისათვის განვიხილოთ სტატისტიკა:

$$\gamma = -\left(m - \frac{2n+1}{6}\right)\ln|R|,$$

სადაც, $|R|$ – კორელაციური მატრიცის დეტერმინანტია, რომელიც შეიძლება ასე განისაზღვროს: $|R| = \prod_{i=1}^{n}\lambda_i$, სადაც, λ_i კორელაციური მატრიცის საკუთრივი მნიშვნელობებია.

γ სტატისტიკას გააჩნია χ^2 განაწილება $\nu = \frac{m(m-1)}{2}$ თავისუფლების ხარისხით. თუ აღმოჩნდება, რომ $\gamma < \chi^2_{\alpha;\nu}$, მაშინ პარამეტრები დამოუკიდებელია, წინააღმდეგ შემთხვევაში, როცა $\gamma \geq \chi^2_{\alpha;\nu}$, ისინი დამოკიდებული არიან.

მაგალითი. შესწავლილი იქნა მექანიკური ტრავმის 2300 შემთხვევის ავადმყოფობის ისტორია [8]. სტატისტიკური დამუშავებისთვის გამოყენებულ იქნა შემდეგი 11 მაჩვენებელი:

1. მდგომარეობის სიმძიმის სუბიექტური შეფასება (დამაკმაყოფილებელი, საშუალო სიმძიმის, მძიმე და უკიდურესად მძიმე);
2. ცნობიერების მდგომარეობა (ნათელი, არეული, არ ჰქონდა);
3. არტერიული წნევის სიდიდე;
4. ჰიპერტონიის ხანგრძლივობა (არტერიული წნევის სიდიდე 100მმ ვერცხლის სვეტის სიმაღლეზე ნაკლებია);
5. პულსის სიხშირე;
6. სისხლკარგვის სიდიდე;
7. სიცოცხლისთვის მნიშვნელოვანი დაზიანებული ორგანოების რაოდენობა;
8. გადასხმული სისხლის რაოდენობა;
9. გადასხმული სისხლის შემცვლის რაოდენობა;
10. ოპერატიული ჩარევის ხანგრძლივობა;
11. ოპერატიული ჩარევის რაოდენობა.

ჩვენ შემთხვევაში კორელაციურ მატრიცას აქვს შემდეგი სახე:

$$R = \begin{array}{c} \begin{array}{ccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \end{array} \\ \left[\begin{array}{ccccccccccc} 1 & 0{,}714 & -0{,}624 & 0{,}426 & 0{,}525 & 0{,}748 & 0{,}741 & 0{,}503 & 0{,}588 & 0{,}434 & 0{,}507 \\ & 1 & -0{,}558 & 0{,}354 & 0{,}326 & 0{,}533 & 0{,}672 & 0{,}315 & 0{,}454 & 0{,}265 & 0{,}428 \\ & & 1 & -0{,}600 & -0{,}382 & 0{,}706 & 0{,}674 & 0{,}538 & 0{,}601 & 0{,}296 & 0{,}386 \\ & & & 1 & 0{,}304 & 0{,}467 & 0{,}378 & 0{,}510 & 0{,}436 & 0{,}364 & 0{,}387 \\ & & & & 1 & 0{,}482 & 0{,}406 & 0{,}351 & 0{,}341 & 0{,}273 & 0{,}293 \\ & & & & & 1 & 0{,}750 & 0{,}691 & 0{,}747 & 0{,}419 & 0{,}539 \\ & & & & & & 1 & 0{,}529 & 0{,}617 & 0{,}368 & 0{,}553 \\ & & & & & & & 1 & 0{,}737 & 0{,}430 & 0{,}515 \\ & & & & & & & & 1 & 0{,}421 & 0{,}556 \\ & & & & & & & & & 1 & 0{,}605 \\ & & & & & & & & & & 1 \end{array}\right] \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \end{array} \end{array}$$

მიღებული კორელაციური მატრიცისათვის განისაზღვრა საკუთრივი მნიშვნელობები და საკუთრივი ვექტორები, რომელთა საშუალებით მივიღეთ მთავარი კომპონენტების კოეფიციენტების მნიშვნელობები, რომლებიც მიღებულია (5.4) ფორმულით. პირველი ოთხი მთავარი კომპონენტის საკუთრივი მნიშვნელობები და კოეფიციენტები მოცემულია შემდეგ ცხრილში:

ცვლადები	მთავარი კომპონენტები			
	1	2	3	4
1	0,84	–0,27	0,21	0,08
2	0,69	–0,44	0,24	–0,30
3	–0,79	0,23	0,32	0,09
4	0,62	0,15	–0,44	0,04
5	0,57	–0,20	0,20	0,76
6	0,86	0,06	–0,09	0,04
7	0,84	–0,25	0,12	–0,17
8	0,76	0,30	–0,13	0,11
9	0,81	0,15	–0,19	–0,04
10	0,59	0,58	0,36	–0,02
11	0,72	0,40	0,32	–0,16
დისპერსიის აბსოლუტური მნიშვნელობა (λ)	6,087	1,058	0,849	0,754
კომპონენტის წილი მთელ დისპერსიაში (%)	55,3	9,6	7,7	6,9
ჯამური წილი (%)	55,3	64,3	72,0	78,9

როგორც ამ ცხრილიდან ჩანს, პირველ ოთხ მთავარ კომპონენტზე მოდის მთელი დისპერსიის დაახლოებით 80% და აქედან მხოლოდ პირველ კომპონენტზე მოდის 55%.

განვიხილოთ უფრო დეტალურად პირველი მთავარი კომპონენტის $Z_1 = 0{,}84X_1 + 0{,}69X_2 - 0{,}79X_3 + \ldots + 0{,}72X_{11}$ კოეფიციენტები, რომლებიც წარმოადგენენ კორელაციურ კოეფიციენტებს საწყის მაჩვენებლებთან.

1. პირველი მთავარი კომპონენტის დადებითი კორელაცია მდგომარეობის სიმძიმის სუბიექტურ შეფასებასთან გასაგებია, რადგან რაც უფრო მძიმეა ტრავმა, მით უფრო მაღალია სუბიექტური შეფასება.

2. რაც უფრო მძიმეა ტრავმა, მით უფრო გამოკვეთილია ცნობიერების დარღვევა, რომელიც ბალური სისტემით არის შეფასებული. ამასთან, უდიდესი ქულა ენიჭებოდა იმ შემთხვევაში, როცა პაციენტს ცნობიერება არ გააჩნდა. სწორედ აქედან გამომდინარეობს მეორე მაჩვენებლის დადებითი კორელაცია პირველ კომპონენტთან.

3. რაც უფრო მძიმეა ტრავმა, მით უფრო მცირეა არტერიული წნევა, რაც ობიექტურად დასტურდება უარყოფითი კორელაციით მესამე მაჩვენებელისა პირველ კომპონენტთან.

4. ხანგრძლივი ჰიპოტონია მძიმე ტრავმების დამახასიათებელი ნიშანია, რასაც მიუთითებს მეოთხე მაჩვენებლის დადებითი კორელაცია პირველ მთავარ კომპონენტთან.

5–6. მძიმე ტრავმის დროს პულსის სიხშირე მატულობს, ხოლო სისხლკარგვა ხელს უწყობს ორგანიზმის მდგომარეობის გართულებას. ამით აიხსნება პირველი კომპონენტის დადებითი კორელაცია მე–5 და მე–6 მაჩვენებლებთან.

7. რაც უფრო მეტი სიცოცხლისთვის მნიშვნელოვანი ორგანოებია დაზიანებული, მით უფრო დიდია ტრავმის სიმძიმის ხარისხი. აქედან გამომდინარეობს დადებითი კორელაცია მე–7 და პირველ მთავარ კომპონენტებს შორის.

8–11. ტრავმის სიმძიმის ზრდასთან ერთად იზრდება ჩარევის რაოდენობებიც. ამიტომაა ამ მაჩვენებლების დადებითი კორელაცია პირველ მთავარ კომპონენტთან.

6. ფაქტორული ანალიზი

ფაქტორული ანალიზის მიზანია მარტივი სტრუქტურის დადგენა, რომელიც გამოავლენს შესასწავლ მოვლენაში ობიექტურად არსებულ ფარულ კანონზომიერებებს. გარდა ამისა, ფაქტორული ანალიზი საშუალებას იძლევა განისაზღვროს ახალი ცვლადები ანუ ფაქტორები და მოახდინოს ისეთი სიდიდეების შეფასება, რომელთა უშუალო გაზომვა შეუძლებელია. ფაქტორული ანალიზით შეგვიძლია საწყისი მონაცემების გარდაქმნა, განზომილების შემცირება და სხვა სპეციფიკური ამოცანების გადაწყვეტა.

ფაქტორული ანალიზის ძირითადი მოდელი. ვთქვათ გვაქვს m ობიექტი, რომლებიც აღწერილი არიან $X_1,X_2,...,X_n$ პარამეტრებით. ფაქტორული ანალიზის ჩასატარებლად ინფორმაცია წარმოდგენილი უნდა იყოს $m\times n$ განზომილებიანი მატრიცის სახით. იმისათვის, რომ გამოირიცხოს სხვადასხვა ფიზიკურ ერთეულებში გაზომილი პარამეტრების ეფექტი, საჭიროა საწყისი მონაცემები წარმოვადგინოთ სტანდარტიზირებული სახით, ე.ი. გადავიდეთ უგანზომილებო ცვლადებზე

$$Z_{ij}=\frac{x_{ij}-\bar{x}_i}{\sigma_i},\ i=1,2,...,m,\ j=1,2,...,n.$$

მაშინ ფაქტორული ანალიზის ძირითადი მოდელი შეიძლება ასე წარმოვადგინოთ:

$$Z_i=\sum_{j=1}^{n}\lambda_{ij}F_j+e_i,\ \ i=1,2,...,m, \qquad (6.1)$$

სადაც, Z_i – i-ური შემთხვევითი ცვლადია, $F_1,F_2,...,F_n$ – ზოგადი ფაქტორებია, რომლებიც შემთხვევით სიდიდეებს წარმოადგენენ და გააჩნიათ ნორმალური განაწილება; e_i – სპეციფიური ანუ მახასიათებელი ფაქტორებია, რომლითაც ხასიათდებიან თითოეული საწყისი Z_i ცვლადები (იგულისხმება, რომ მახასიათებელი ფაქტორები არაკორელირებული არიან); λ_{ij} – ფაქტორული დატვირთვებია, რომლითაც ხასიათდება თითოეული ფაქტორი და რომლებიც უნდა იყვნენ განსაზღვრულნი.

ამრიგად, ფაქტორული ანალიზის ძირითადი ამოცანაა განისაზღვროს ფაქტორული დატვირთვები. თუ ზოგადი და მახასიათებელი ფაქტორები ურთიერთარაკორელირებული არიან, მაშინ i-ური ფაქტორის დისპერსია შეიძლება ასე წარმოვადგინოთ:

$$\sigma_i^2=1=\lambda_{i1}^2+\lambda_{i2}^2+...+\lambda_{in}^2+\tau_i,\ \ i=1,2,...,m,$$

სადაც, λ_i^2 – Z_i პარამეტრის დისპერსიის წილია, რომელიც მოდის i-ურ ფაქტორზე, ხოლო მახასიათებელი ვექტორისათვის $\mathrm{var}(e_i) = \tau_i$, $i = 1,2,...,n$, სადაც, τ_i-ს უწოდებენ სპეციფიურ დისპერსიას. შემოვიღოთ აღნიშვნა

$$h_i^2 = \lambda_{i1}^2 + \lambda_{i2}^2 + ... + \lambda_{in}^2 \ ,$$

რომელიც წარმოადგენს საერთო დისპერსიის წილს, გამოწვეულს ზოგადი ფაქტორებით და მას უწოდებენ i-ური საწყისი პარამეტრის ერთიანობას. k-ური ფაქტორის სრული წილი საერთო დისპერსიაში იქნება:

$$V_k = \sum_{j=1}^{m} \lambda_{jk}^2 , \ \ k = 1,2,...,n .$$

ფაქტორების დამოუკიდებლობის შემთხვევაში, ადვილია პარამეტრებს შორის კორელაციის კოეფიციენტის განსაზღვრა

$$r'_{ik} = \lambda_{i1}\lambda_{k1} + \lambda_{i2}\lambda_{k2} + ... + \lambda_{in}\lambda_{kn}, \ \ i \neq k \, , k = 1,2,..., n .$$

შემოვიტანოთ ნარჩენი კორელაციისა და ნარჩენი კორელაციური მატრიცის ცნებები. (6.1) მოდელის ჩაწერისათვის საწყის ინფორმაციას წარმოადგენს სპირმენის კორელაციური მატრიცა. თუ გამოვიყენებთ ფაქტორულ მოდელს და ხელახლა გამოვთვლით კორელაციურ მატრიცას, მაშინ მათ შორის სხვაობა იქნება ნარჩენი კორელაციის კოეფიციენტი, ე.ი.

$$\bar{r}_{jk} = r_{jk} - r'_{jk} \, ,$$

ხოლო ნარჩენი კორელაციის კოეფიციენტებისგან შედგენილი მატრიცა

$$\bar{R} = R - R' .$$

ფაქტორული ანალიზის ამოცანაა, შეაფასოს λ_{ij} ფაქტორული დატვირთვების τ_i სპეციფიური დისპერსიები და ფაქტორული მნიშვნელობები. როდესაც ფაქტორული დატვირთვები ცნობილი იქნება, შემდეგ რჩება კიდევ ერთი ამოცანა, კერძოდ, ფაქტორების ინტერპრეტაციის პრობლემა. ამისათვის გამოიყენება ფაქტორული ბრუნვა.

მთავარი ფაქტორების განსაზღვრა. ფაქტორული ანალიზის პირველი ამოცანაა R კორელაციური (ან S კოვარიაციული) მატრიცის საშუალებით განისაზღვროს λ_{ij} ფაქტორული დატვირთვების l_{ij} შეფასებები და τ_i სპეციფიურ დისპერსიის t_i შეფასებები. ამისათვის არსებობს მრავალი მეთოდი, მაგრამ ჩვენ განვიხილავთ მთავარი ვექტორის განსაზღვრის მეთოდს, რომელიც მდგომარეობს n მთავარი კომპონენტის განსაზღვრაში:

$$Y_i = \sum_{j=1}^{n} a_{ij} X_j \ , \ i = 1,2,...,n .$$

გავიხსენოთ, რომ n მთავარი კომპონენტები ერთმანეთის მიმართ არაკორელირებული არიან და i-ური კომპონენტის დისპერსია $\mathrm{var}(Y_i)$ ტოლია კორელაციური მატრიცის i-ური საკუთრივი მნიშვნელობისა, ე.ი. $\mathrm{var}(Y_i) = \lambda_i$.

მთავარი ფაქტორების მეთოდიდან გამომდინარე, ზოგად ფაქტორებად მიიღება m პირველი მთავარი კომპონენტი, რომლებიც განისაზღვრება შემდეგნაირად:

$$F_j = \frac{Y_j}{\sqrt{\mathrm{var}(Y_j)}}, \ \ j = 1,2,...,m \, ,$$

ხოლო ფაქტორული დატვირთვების შეფასებები ტოლია:

$$l_{ij} = a_{ji}\sqrt{\mathrm{var}(Y_i)}, \ \ i = 1,2,...,n, \ \ j = 1,2,...,m .$$

რაც შეეხება მახასიათებელი ფაქტორების შეფასებებს, ისინი ასე განისაზღვრება:

$$e_i = \sum_{j=m+1}^{n} a_{ji} Y_j, \;\; i = 1,2,...,n .$$

ამრიგად, მივიღებთ ფაქტორული მოდელის შემდეგ შეფასებას:

$$Z_i = \sum_{j=1}^{m} e_{ij} F_j + e_i, \;\; i = 1,2,...,n .$$

აქ ყველა ფაქტორი ერთმანეთის მიმართ არაკორელირებულია და დისპერსიები ერთის ტოლია. სპეციფიური დისპერსიებისა და ერთიანობების შეფასებები იქნება:

$$h_i^2 = \sum_{j=1}^{n} a_{ji}^2 \operatorname{var}(Y_j) = \sum_{j=1}^{m} e_{ji}^2;$$

$$t_i = \sum_{i=m+1}^{n} a_{ji} \operatorname{var}(Y_j).$$

ფაქტორების ბრუნვა. ფაქტორული დატვირთვების განსაზღვრის შემდეგ საჭიროა თითოეული ფაქტორის ინტერპრეტაცია. ამისათვის საჭიროა ფაქტორების ბრუნვა ახალი ორთოგონალური ფაქტორების მისაღებად, რომლებიც ერთმანეთის მიმართ არაკორელირებულია და გააჩნიათ ერთეულოვანი დისპერსია. ბრუნვის შემდეგ ფაქტორული მოდელი ასე ჩაიწერება:

$$Z_i = \sum_{j=1}^{m} C_{ij} F^{(R)} + e_i, \;\; i = 1,2,...,n,$$

სადაც, C_{ij} წარმოადგენს ახალი ფაქტორების დატვირთვებს. აქვე უნდა აღვნიშნოთ, რომ ორთოგონალური ბრუნვის შედეგად, თითოეული საწყისი Z_i ცვლადის ერთიანობა უცვლელი რჩება, ე.ი.

$$h_i^2 = \sum_{j=1}^{m} C_{ij}^2 = \sum_{j=1}^{m} l_{ij}^2, \;\; i = 1,2,...,n .$$

C_{ij} მუდმივები ისე უნდა შეირჩეს, რომ დატვირთვები იყვნენ მარტივი სტრუქტურის. ზოგადად, ფაქტორული დატვირთვების სტრუქტურა ითვლება მარტივად, როცა C_{ij} კოეფიციენტების უმრავლესობა ახლოსაა ნულთან და მხოლოდ ერთ ან რამდენიმე მათგანს გააჩნია ნულისგან შედარებით დიდი მნიშვნელობები. ამრიგად, ბრუნვის მიზანია, თითოეული საწყისი ცვლადი წარმოადგინოს ერთი ან მცირე რაოდენობის ფაქტორებით, ხოლო სხვა დანარჩენი ფაქტორების დატვირთვა ნულთან უნდა იყოს ახლოს.

არსებობს ფაქტორების ბრუნვის მრავალი მეთოდი, როგორც გრაფიკული, ასევე ანალიზური. ანალიზური მეთოდებიდან გამოირჩევა ე.წ. მიზნობრივი ფუნქციის მინიმიზაციის მეთოდი, რომელიც დამოკიდებულია C_{ij} სიდიდეებზე. ორთოგონალური ბრუნვისთვის, ძირითადად, იყენებენ შემდეგ ფუნქციას:

$$G = \sum_{k=1}^{m} \sum_{\substack{j=1 \\ j \neq k}}^{m} \left[\sum_{i=1}^{n} C_{ij}^2 C_{ik}^2 - \frac{\gamma}{n} \left(\sum_{i=1}^{n} C_{ij}^2 \right) \left(\sum_{i=1}^{n} C_{ik}^2 \right) \right], \qquad (6.2)$$

სადაც, $0 \le \gamma \le 1$.

როცა $\gamma = 0$, ბრუნვას, რომელიც მიიღება G ფუნქციის მინიმიზაციით, ეწოდება „კვარტიმაქსის“ მეთოდი. ამ შემთხვევაში, G ფუნქციის მინიმიზაცია ექვივალენტურია

$$\frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n}\left(C_{ij}^2-\overline{C}_{\bullet\bullet}^2\right)^2 \qquad (6.3)$$

გამოსახულების მაქსიმიზაციისა. აქ, $\overline{C}_{\bullet\bullet}^2=\frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n}C_{ij}^2$.

როგორც (6.3) გამოსახულებიდან ჩანს, „კვარტიმაქსის“ მეთოდით ხდება ფაქტორული დატვირთვების კვადრატების დისპერსიის მაქსიმიზაცია. ამ შემთხვევაში, ის ფაქტორები, რომლებთაც აქვთ დიდი დატვირთვის მნიშვნელობები, კიდევ უფრო იზრდებიან, ხოლო მცირე მნიშვნელობები კიდევ უფრო მცირენი ხდებიან.

როცა $\gamma = 1$, ბრუნვის მეთოდს „ვარიმაქს“ უწოდებენ. ეს მეთოდი ყველაზე უფრო ხშირად გამოიყენება პრაქტიკაში. ამ შემთხვევაში G ფუნქციის მინიმიზაცია ექვივალენტურია

$$\frac{1}{n}\sum_{j=1}^{m}\sum_{i=1}^{n}\left(C_{ij}^2-\overline{C}_{\bullet j}^2\right)^2 \qquad (6.4)$$

გამოსახულების მაქსიმიზაციისა. აქ,

$$\overline{C}_{\bullet j}^2=\frac{1}{n}\sum_{i=1}^{n}C_{ij}^2,\ j=1,2,\ldots,m.$$

(6.4) გამოსახულება წარმოადგენს ფაქტორული დატვირთვების კვადრატების დისპერსიების ჯამს სვეტების მიხედვით და იგი იწვევს თითოეული ფაქტორის დატვირთვების კვადრატების დისპერსიის მაქსიმიზაციას. ეს უკანასკნელი, თავის მხრივ, დატვირთვების დიდ მნიშვნელობებს კიდევ უფრო ზრდის, ხოლო დატვირთვების მცირე მნიშვნელობებს უფრო ამცირებს. ამრიგად, ამ შემთხვევაში უბრალო სტრუქტურა მიიღება თითოეული ფაქტორისათვის ცალ-ცალკე, ხოლო წინა „კვარტიმაქსის“ მეთოდში უბრალო სტრუქტურა განისაზღვრებოდა ყველა ფაქტორისათვის ერთდროულად.

აქამდე ჩვენ განვიხილეთ ფაქტორების ბრუნვის მხოლოდ ორთოგონალური მეთოდები. არსებობს მოსაზრება, რომ მნიშვნელოვანია მივიღოთ ფაქტორული დატვირთვების უბრალო სტრუქტურა, ვიდრე შევინარჩუნოთ მათი ორთოგონალურობა. ასეთი ფაქტორების მიღების მეთოდს ირიბკუთხა ბრუნვა ეწოდება. აღვნიშნოთ $R = (r_{ik})$ ფაქტორების მეორადი „უბრალო“ სტრუქტურა $i = 1,2,\ldots,n$, $k = 1,2,\ldots,m$, სადაც, r_{ik} არის კორელაციის კოეფიციენტი i-ური საწყისი ცვლადის k-ურ მეორად ფაქტორთან. ირიბკუთხა ბრუნვის დროს ხდება

$$G=\sum_{k=1}^{m}\sum_{\substack{j=1\\ j\neq k}}^{m}\left[\sum_{i=1}^{n}r_{ij}^2r_{ik}^2-\frac{\gamma}{n}\left(\sum_{i=1}^{n}r_{ij}^2\right)\left(\sum_{i=1}^{n}r_{ik}^2\right)\right]$$

მიზნობრივი ფუნქციის მინიმიზაცია, სადაც, $r_{ik} = \mathrm{corr}(X_i, G_j)$ და იგი იცვლება 0-დან 1-მდე. ანალიტიკურ მეთოდებს, სადაც იძებნება უბრალო მეორადი სტრუქტურა, ეწოდება არაპირდაპირი მეთოდი „ობლიმინი“. აქ, $0 \le \gamma \le 1$. როცა $\gamma = 0$, მაშინ საქმე გვაქვს მძლავრ ირიბკუთხა ბრუნვასთან, როცა $\gamma = 0{,}5$ – ნაკლებად ირიბკუთხა ბრუნვასთან, ხოლო როცა $\gamma = 1$ – ყველაზე მცირე ირიბკუთხა ბრუნვასთან.

მაგალითი. 113 პაციენტზე ჩატარებული გამოკვლევების შედეგად გაზომილ იქნა შემდეგი პარამეტრები: 1) სისტოლური წნევა; 2) დიასტოლური წნევა; 3) საშუალო არტერიული წნევა; 4) გულის შეკუმშვის სიხშირე; 5)

საშუალო ვენოზური წნევა; 6) გულის ინდექსი; 7) გამოვლენის დრო; 8) ცირკულირების საშუალო დრო; 9) დიურეზი; 10) პლაზმის მოცულობის ინდექსი; 11) ერითროციტული ინდექსი; 12) ჰემოგლობინი და 13) ჰემოტოკრიტი.

მოცემული 13 მაჩვენებლის კორელაციური R მატრიცას აქვს შემდეგი სახე:

$$R = \begin{bmatrix}
1{,}00 & & & & & & & & & & & & \\
0{,}90 & 1{,}00 & & & & & & & & & & & \\
0{,}10 & -0{,}07 & 1{,}00 & & & & & & & & & & \\
-0{,}81 & 0{,}95 & 0{,}00 & 1{,}00 & & & & & & & & & \\
-0{,}03 & -0{,}07 & 0{,}05 & -0{,}13 & 1{,}00 & & & & & & & & \\
0{,}12 & 0{,}03 & -0{,}05 & -0{,}07 & -0{,}05 & 1{,}00 & & & & & & & \\
-0{,}16 & -0{,}11 & -0{,}15 & -0{,}04 & -0{,}01 & -0{,}49 & 1{,}00 & & & & & & \\
-0{,}17 & -0{,}11 & 0{,}02 & -0{,}00 & 0{,}14 & -0{,}68 & 0{,}84 & 1{,}00 & & & & & \\
0{,}13 & 0{,}15 & -0{,}12 & 0{,}12 & -0{,}23 & 0{,}09 & -0{,}21 & -0{,}18 & 1{,}00 & & & & \\
0{,}08 & -0{,}17 & -0{,}13 & -0{,}27 & 0{,}13 & 0{,}54 & -0{,}16 & -0{,}28 & 0{,}04 & 1{,}00 & & & \\
0{,}09 & 0{,}11 & -0{,}02 & 0{,}14 & -0{,}06 & -0{,}11 & 0{,}20 & 0{,}21 & -0{,}05 & 0{,}04 & 1{,}00 & & \\
0{,}09 & 0{,}21 & 0{,}09 & 0{,}33 & -0{,}09 & -0{,}48 & 0{,}39 & 0{,}47 & -0{,}07 & -0{,}49 & 0{,}38 & 1{,}00 & \\
0{,}09 & 0{,}21 & 0{,}06 & 0{,}32 & -0{,}08 & -0{,}48 & 0{,}40 & 0{,}49 & -0{,}09 & -0{,}50 & 0{,}39 & 0{,}97 & 1
\end{bmatrix}$$

R მატრიცის საკუთრივი რიცხვებისა და ჯამური დისპერსიების წილს აქვთ შემდეგი მნიშვნელობები:

ფაქტორი	საკ. მნიშვნელობა	ჯამური დისპერსიის წილი (%)
1	3,875	30
2	2,980	53
3	1,269	62
4	1,233	72
5	1,095	80
6	0,766	86
7	0,711	92
8	0,507	96
9	0,290	98
10	0,150	99
11	0,084	99
12	0,023	100
13	0,019	100

როგორც ამ ცხრილიდან ჩანს, პირველ ხუთ ფაქტორზე მოდის მთელი დისპერსიის 80%, ამიტომ შევჩერდეთ ამ ხუთ ფაქტორზე, რომელთა ფაქტორული დატვირთვების შეფასებები მოცემულია შემდეგ ცხრილში:

ცვლა-დები	ფაქტორები				
	1	2	3	4	5
1	0,21	0,88	– 0,22	0,15	– 0,09
2	0,33	0,90	– 0,13	0,14	– 0,09
3	0,05	– 0,08	0,59	0,48	0,33
4	0,47	0,83	– 0,04	0,13	– 0,07
5	– 0,07	– 0,18	– 0,35	0,71	– 0,03
6	– 0,70	0,33	– 0,10	– 0,06	0,34
7	0,61	– 0,44	– 0,42	– 0,12	– 0,20
8	0,71	– 0,48	– 0,26	0,05	– 0,21
9	– 0,13	0,31	0,18	– 0,59	– 0,22
10	– 0,61	– 0,03	– 0,52	– 0,07	0,31
11	0,40	0,03	– 0,32	– 0,23	0,69
12	0,87	– 0,00	0,15	– 0,09	0,26
13	0,88	– 0,01	0,13	– 0,08	0,26

როგორც აღვნიშნეთ, ფაქტორული დატვირთვები – ესაა კორელაცია საწყის ცვლადსა და ფაქტორს შორის. მაგ. l_{11} = 0,21 ეს არის კორელაციის კოეფიციენტი სისტოლურ წნევასა და პირველ ფაქტორს შორის; l_{12} = 0,88 კი არის იგივე ცვლადისა და მეორე ფაქტორს შორის და ა.შ.

ფაქტორების ინტერპრეტაციისთვის განვიხილოთ ის დატვირთვები, რომელთა სიდიდე მეტია რაიმე ზღვრულ მნიშვნელობაზე, მაგ. r = 0,4. ცხრილში მოცემული პირველი ფაქტორისათვის ასეთი მაჩვენებლებია რვა, ე.ი. პირველი ფაქტორი, ძირითადად, დამოკიდებულია ამ რვა მაჩვენებელზე, მეორე ფაქტორი კი მხოლოდ ხუთ ცვლადზე და ა.შ. როგორც ვხედავთ, ფაქტორების ინტერპრეტაცია ძნელდება. ამიტომ მიზანშეწონილია ჩავატაროთ ფაქტორების ბრუნვა „ვარიმაქსის“ მეთოდის გამოყენებით. მიღებული შედეგები მოყვანილია ცხრილში:

ცვლა-დები	ფ ა ქ ტ ო რ ე ბ ი				
	1	2	3	4	5
1	– 0,11	0,94	– 0,99	0,02	0,03
2	– 0,01	0,98	– 0,00	– 0,04	0,04
3	– 0,08	– 0,10	0,81	0,17	0,04
4	0,10	0,95	0,09	– 0,08	0,09
5	0,03	– 0,00	– 0,00	0,81	– 0,14
6	– 0,85	0,02	– 0,14	0,01	0,07
7	0,78	– 0,13	– 0,36	0,14	0,19
8	0,88	– 0,12	– 0,14	0,21	0,14
9	– 0,15	0,14	– 0,19	– 0,67	– 0,14
10	– 0,61	– 0,21	– 0,49	0,27	0,18
11	0,08	0,07	– 0,09	0,02	0,88
12	0,64	0,21	0,32	– 0,16	0,54
13	0,65	0,21	0,30	– 0,15	0,54

მიღებული ფაქტორული დატვირთვებით უკვე შესაძლებელია ფაქტორების ინტერპრეტაცია. კერძოდ, F_1 ფაქტორი წარმოადგენს სისხლდენის ფაქტორს, F_2 – არტერიული წნევის, F_3 – გულის შეკუმშვის სიხშირისა და პლაზმის, F_4 – დიურეზის და F_5 – სისხლის შედგენილობის.

ამრიგად, საწყისი 13 ცვლადის მაგივრად ჩვენ შევჩერდით ხუთ ძირითად ფაქტორზე, რომლებიც ერთმანეთის მიმართ არაკორელირებული არიან და შესაძლებელია მათი ინტერპრეტაცია.

7. გადარჩენის ანალიზი

7.1 მეთოდის არსი

გადარჩენის ანალიზის (***Survival Analysis***) მეთოდები პირველად გამოყენებული იყო სამედიცინო-ბიოლოგიურ კვლევებში და სადაზღვევო საქმეში. შემდგომ მისი გამოყენების სფერო გაიზარდა და დღეისათვის გამოიყენება როგორც სოციოლოგიურ და ეკონომიკურ კვლევებში, აგრეთვე წარმოებაში სხვადასხვა საინჟინერო ამოცანების გადასაწყვეტად.

დაუშვათ, შეისწავლება მკურნალობის ახალი მეთოდი ან რომელიმე სამკურნალო პრეპარატის ეფექტიანობა. ცხადია, რომ ყველაზე მნიშვნელოვან და ობიექტური მახასიათებელს წარმოადგენს პაციენტის სიცოცხლის საშუალო ხანგრძლივობა დაწყებული კლინიკაში შემოსვლიდან ან დაავადების რემისიის საშუალო ხანგრძლივობა. სიცოცხლის საშუალო ხანგრძლივობის აღწერისათვის შეგვიძლია გამოვიყენოთ მათემატიკური სტატისტიკის სტანდარტული პარამეტრული ან არაპარამეტრული მეთოდები. მაგრამ, შესასწავლ მონაცემებში არსებობს მნიშვნელობანი თავისებურებები. კერძოდ, შეიძლება მოიძებნონ ისეთი პაციენტები, რომლებიც დაკვირვების პერიოდში გამოჯამრთელდენ. ხოლო ზოგიერთი მათგანისათვის დაავადება კვლავ რჩება რემისიის სტადიაში. ასევე შეიძლება შეიქმნას პაციენტების ისეთი ჯგუფი, რომლებთანაც ექსპერიმენტის დამთავრებამდე გარკვეული მიზეზების გამო კონტაქტი დაიკარგა, მაგალითად, მათი სხვადასხვა კლინიკებში გადაყვანის გამო. ამასთან ერთად დაკვირვების პერიოდში ამ პაციენტთა უმრავლესობა გამოჯამრთელდა, რითაც დასტურდება მკურნაალობის ახალი მეთოდის ან სამკურნალო პრეპარატის ეფექტიანობა.

ისეთ ინფორმაციას, როცა ჩვენთვის საინტერესო მოვლენის (ხდომილობის) დადგომის შესახებ არ არსებობს, ეწოდება არასრული. მოვიყვანოთ არასრული ინფორმაციის მაგალითი: „*A* პაციენტი ცოცხალი იყო 4 თვის განმავლობაში, ანუ იმ მომენტამდე, როცა ის გადაიყვანეს სხვა კლინიკაში და ამის გამო მასთან კონტაქტი დაიკარგა“, ან „დღემდე *A* პაციენტი ცოცხალია“.

თუ არსებობს მონაცემები ჩვენთვის სასურველი მოვლენის დადგომის შესახებ, მაშინ ასეთ ინფორმაციას ეწოდება სრული. მაგალითად: „ჩატარებული მკურნალობის შემდეგ *A* პაციენტმა იცოცხლა 5 წელი“, ან „მკურნალობის შემდეგ *A* პაციენტი 3 თვეში კვლავ დაავადდა“.

დაკვირვებები, რომლებიც შეიცავენ არასრულ ინფორმაციას ეწოდებათ **ცენზურირებული**. ცენზურირებული დაკვირვებები ტიპიურია როცა დაკვირვებულ სიდიდეს წარმოადგენს დრო რაიმე კრიტიკული მოვლენის დადგომამდე, ხოლო დაკვირვების ხანგრძლივობა დროით შეზღუდულია.

ცენზურირებული დაკვირვებები, გარდა მედიცინისა, გვხდება სხვა სფეროებშიც. მაგალითად, სოციალურ მეცნიერებაში ჩვენ შეგვიძლია შევისწავლოთ ქორწინების „ხარგრძლივობა“, სტუდენტების უმაღლეს სასწავლებლიდან გარიცხვის ინტენსიობა, ზოგიერთი დაწესებულის თანამშრომლების რაოდენობის დინამიკა და სხვა. ეკონომიკაში შეგვიძლია შევისწავლოთ ახალი დაწესებულების გადარჩენის ანუ მისი ფუნქციონერების ხარგრძლივობა, ან წარმოების პროდუქტის „სიცოცხლის“ ხანგრძლივობა. ხარისხის კონტროლის ამოცანებში ტიპიური პროდუქციის ელემენტების დატვირთვის დროს გადარჩენის (ხარგრძლივობის) შესწავლა (მტყუნების დროის ანალიზი) და სხვა.

ცენზურირებული დაკვირვებების გამოყენება წარმოადგენს გადარჩენის ანალიზის სპეციფიკას, სადაც ხდება კრიტიკული მოვლენების წაემოქმნის დროითი ინტერვალების ალბათური მახასიათებლების შესწავლა.

ამრიგად გადარჩენის ანალიზის მეთოდები ძირითადად გამოიყენება ცენზურირებული დაკვირვებების სტატისტიკური ამოცანების გადასაწყვეტად.

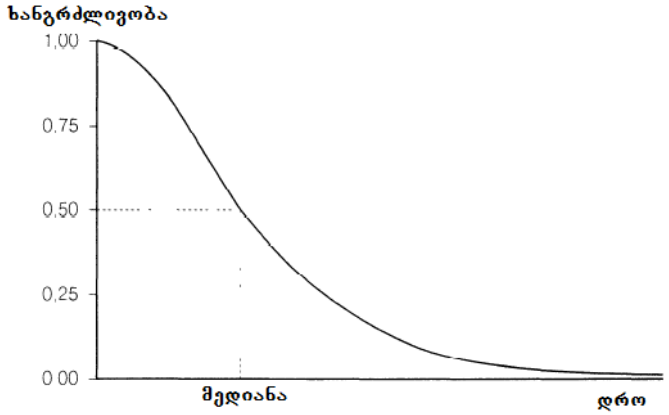
7.2 გადარჩენის ფუნქცია

განაწილების $F(t)$ ფუნქციის და განაწილების სიმკვრივის $f(t)$ ფუნქციების მაგივრად გადარჩენის ანალიზში გამოიყენება ე.წ. გადარჩენის $S(t)$ ფუნქცია, რომელიც წარმოადგენს იმის ალბათობას, რომ ობიექტი იცოცხლებს მოცემულ t დროზე მეტად, ანუ მტყუნება დადგება t დროის მომენტის შემდეგ:

$$S(t) = P(X > t).$$

გადარჩენის ფუნქცია გამოიყენება სიცოცხლის ხანგრძლივობის შესაფასებლად. გარდა ამისა, იგი გამოიყენება სხვა არასასურველი მოვლენებისთვისაც. ასე მაგალითად, შეიძლება შევისწავლოთ რომელიმე დაავადების მკურნალობის ხანგრძლივობა (შედეგი – რემისია), პროტეზის ხანდრძლივობა (შედეგი – დამტვრევა) და ბევრი სხვა.

ტიპიური გადარჩენის ფუნქცია წარმოდგენილია შემდეგ ნახაზზე:



დასაწყისში გადარჩენის ფუნქცია $S(t) = 1$. შემდგომ ის კლებულობს და თანდათანობით უახლოვდება ნულს. დროის იმ მომენტს, როცა $S(t) = 0{,}5$ ეწოდება გადარჩენის მედიანა.

თუ d_t აღვნიშნავთ t დროის მომენტის დროს მტყუნების რაოდენობას, r_t - იმ ობიექტა რაოდენობას, რომელთა მტყუნება დადგება t დროის მომენტის შემდეგ, ხოლო n - ით მონაცემების რაოდენობას და თუ ჩავთვალოთ, რომ $r_1 = n$, მაშინ დისკრეტული შემთხვევითი სიდიდეებისათვის გადარჩენის ფუნქცია განისაზღვრება შემდეგი გამოსახულებით:

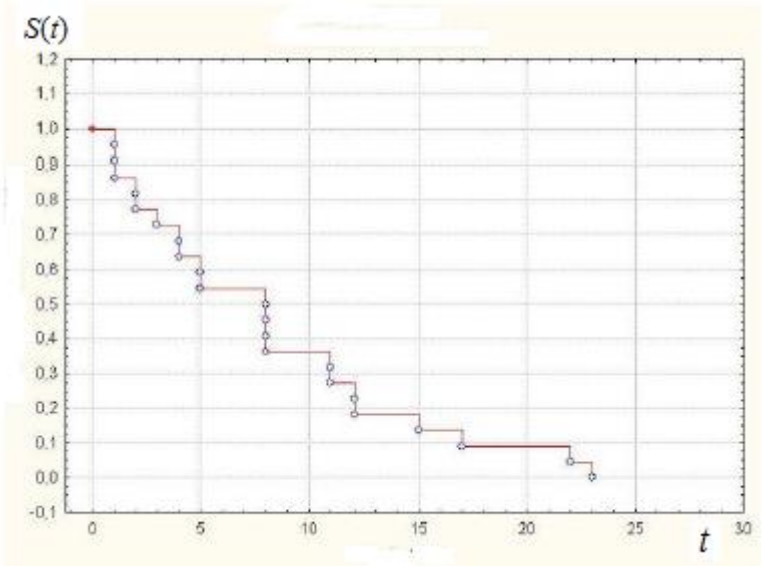
$$S(t) = \frac{r_t - d_t}{n} = \frac{r_{t+1}}{n} \qquad (7.1)$$

მაგალითი. ვთქვათ მოცემულია საწყისი მონაცემები, რომლებიც წარმოადგენენ თითოეული ინდივიდუმისათვის მტყუნების დადგენის მომენტებს:

τi : 1 1 1 2 2 3 4 4 5 5 8 8 8 8 11 11 12 12 15 17 22 23

განვსაზღვროთ გადარჩენის ფუნქციის მნიშვნელობები და ავაგოთ შესაბამისი გრაფიკული გამოსახულება.

t	r_t	d_t	$r_t - d_t$	$\widehat{F}_t$
0	22	0	22	1
1	22	3	19	$\frac{19}{22}$
2	19	2	17	$\frac{17}{22}$
3	17	1	16	$\frac{16}{22}$



რისკის ფუნქცია. გადარჩენის $S(t)$ ფუნქციის წარმოებულს, აღებული შებრუნებული ნიშნით, ეწოდება გადარჩენის სიმკვრივის ***f(t)*** ფუნქცია: $f(t) = -S'(t)$. აზრობრივად გადარჩენის სიმკვრივე ტოლია t დროის მომენტში დაღუპვის ალბათობისა.

t დროის მომენტში დაღუპვის პირობითი ალბათობა გამოისახება h(t) ინტენსივობის ფუნქციით, რომელიც განისაზღვრება შემდეგნაირად:

$$H(t)=\frac{-S'(t)}{S(t)}=\left(-\ln S(t)\right)'$$

რომელსაც რისკის $H(t)$ ფუნქცია ეწოდება.

$$H(t)=-\ln S(t) \qquad S(t)=e^{-H(t)}.$$

რადგან $S(0)=1$, ამიტომ $H(0)=0$. როგორც მიღებული გამოსახულებიდან ჩანს, თუ რისკის $H(t)$ ფუნქცია წარმოადგენს მუდმივ სიდიდეს, მაშინ გადარჩენის $S(t)$ ფუნქცია წარმოადგენს ექსპონენციალურად კლებად ფუნქციას.

გადარჩენის ცხრილები. გადარჩენის $S(t)$ ფუნქციის წარმოდგენის (აღწერის) ერთ-ერთ ბუნებრივ მეთოდს წარმოადგენს გადარჩენის ცხრილები. ცხრილი შეიძლება განვიხილოთ, როგორც სტატისტიკაში არსებული სიხშირული ცხრილის „გაფართოება“. კრიტიკული ხდომილების (მტყუნება, დაღუპვა) შესაძლო დადგომის დროს არე იყოფა გარკვეული რაოდენობის ინტერვალებად. თითოეული ინტერვალისათვის განისაზღვრება: ინტერვალის დასაწყისი, ინტერვალის საშუალო სიდიდე, ინტერვალის სიგრძე, იმ ობიექტების რაოდენობა და ფარდობითი სიხშირე, რომლებიც „ცოცხლები“ იყვნენ განსახილველი დროის ინტერვალის დასაწყისში, ობიექტების რაოდენობა და ფარდობითი სიხშირე, რომლებიც „დაიღუპნენ“ ამ დროის ინტერვალში და სხვა მაჩვენებლები.

7.3 კაპლან - მეიერის მამრავლების შეფასების მეთოდი

გადარჩენის განაწილების ცხრილები წარმოადგენენ ძველ, მაგრამ ყველაზე უფრო გამოყენებად გადარჩენის ფუნქციის შეფასების მეთოდს. მაგრამ ცხრილისათვის ზუსტი შეფასება დამოკიდებულია გადარჩენის ინტერვალების რაოდენობაზე და სიგრძეზე. კაპლან-მეიერის მეთოდით გადარჩენის ფუნქციის შეფასება საშუალებას იძლევა შეფასებული იყოს ცენზურირებული მონაცემების გადარჩენის ფუნქცია, გადარჩენის დროის გამოყენებით და მონაცემების დაჯგუფების გარეშე.

თუ ცენზურირებულ მონაცემებში c_i-ით აღვნიშნავთ ცენზურირების მომენტებს, მაშინ საწყისი მონაცემები შეიძლება ასე წარმოვადგინოთ:

$$(X_i, V_i), \text{ სადაც } X_i=\min(\tau_i, c_i), \quad i=1,2,...,n$$

$$V_i=0 \text{ , როცა } \tau_i \le c_i \text{ (დაღუპვა),}$$

$$V_i=1 \text{ , როცა } \tau_i > c_i \text{ (ცენზურირება).}$$

ცენზურირებული მონაცემებისათვის გადარჩენის ფუნქციის შეფასება განისაზღვრება ფორმულით:

$$\tilde{S}(t)=\prod_{i=1}^{t}\left(\frac{r_i-d_i}{r_i}\right)=\prod_{i=1}^{t}\left(1-\frac{d_i}{r_i}\right)=\prod_{i=1}^{t}(1-h(t)), \qquad (7.2)$$

სადაც r_i - t_i მომენტამდე „ცოცხალი“ობიექტების რაოდენობა, გამოკლებული ობიექტების გათვალისწინებით,

d_i - t_i მომენტში დაღუპულ ობიექტა რაოდენობა.

მიღებულ (7.2) გამოსახულებას კაპლან - მეიერის მამრავლების შეფასების ფორმულა ეწოდება. ამ ფორმულაში მნიშვნელობების გადამრავლება შესაძლებელია მხოლოდ დროის იმ მომენტამდე, როდესაც მოხდა თუნდაც ერთი

ობიექტის დაღუპვა იმიტომ, რომ თუ d_i=0, მაშინ $\frac{r_i - d_i}{r_i} = 1$ და ერთზე გამრავლება შედეგს არ ცვლის.

არაცენზურირებული მონაცემებისათვის t_{i+1} მომენტში $r_{i+1} = r_i - d_i$, ამიტომ (7.2) ფორმულაში მოსაზღვრე ელემენტები იკვეცებიან და ვღებულობთ:

$$\widetilde{S}(t) = \frac{r_t - d_t}{r_1} = \frac{r_{t+1}}{r_1}$$

რაც შეესაბამება (7.1) შეფასებას.

კაპლან - მეიერის ფუნქციის განსაზღვრისათვის განვიხილოთ მაგალითი. ვთქვათ მოცემულია ცენზურირებული მონაცემები:

j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
X_j	6	6	6	6	7	9	10	10	11	13	16	17	19	20	22	23	25	32	32	34	35
V_j	1	0	0	0	0	1	1	0	1	0	0	1	1	1	0	0	1	1	1	1	1

i	d_i დაღუპვა	z_i ცენზურირება	$r_i = r_{i-1} - d_{i-1} - z_{i-1}$ დარჩენილები	$\widehat{p}_i = 1 - h_i$ პირ.ალბათ.	$\widetilde{S}(t)$ $\prod_{j=1}^{i}(1-h_j)$
1	0	0	21	1	1
2	0	0	21	1	1
3	0	0	21	1	1
4	0	0	21	1	1
5	0	0	21	1	1
6	3	1	21	0.857	0.857
7	1	0	17	0.941	0.807
8	0	0	16	1	0.807
9	0	1	16	1	0.807
10	1	1	15	0.933	0.753
11	0	1	13	1	0.753
12	0	0	12	1	0.753
13	1	0	12	0.917	0.690
14	0	0	11	1	0.690
15	0	0	11	1	0.690

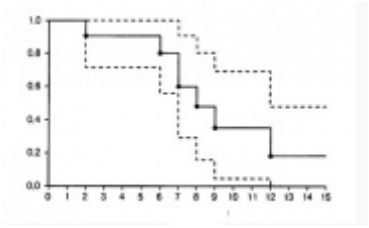
გადარჩენის ფუნქციის ნდობის ინტერვალი. გადარჩენის ფუნქციის სტანდარტული ცდომილება შეიძლება განისაზღვროს გრინვულდის ფორმულით:

$$\sigma_{\tilde{S}} = \tilde{S}(t)\sqrt{\sum_{i=1}^{t}\frac{d_i}{r_i(r_i-d_i)}}.$$

t დროის მომენტში ნდობის ინტერვალი $1-\alpha$ ნდობის ალბათობით განისაზღვრება შემდეგნაირად:

$$\tilde{S}(t)-\sigma_{\tilde{S}}\Phi_\alpha < S(t) < \tilde{S}(t)+\sigma_{\tilde{S}}\Phi_\alpha,$$

სადაც Φ_α- ნორმალური განაწილების კვანტილია. საზოგადოდ იღებენ $\alpha = 0{,}05$.



ნახაზიდან ჩანს, რომ t დროის ზრდასთან ერთად იზრდება ნდობის ინტერვალიც. ეს დამოკიდებულია იმაზე, რომ რაც უფრო ნაკლებია ექსპერიმენტის ბოლოს ობიექტები, მით უფრო დიდია ცდომილება. ამიტომ არსებობს შეზღუდვა ხანგრძლივობის ფუნქციის ნდობის ინტერვალის განსაზღვრაზე.

7.4 ორ ჯგუფს შორის გადარჩენის ფუნქციების შედარება

გადარჩენის სტატისტიკური მახასიათებლების შეფასების შემდეგ კანონზომიერად ისმება ორ და უფრო მეტ ჯგუფებს შორის გადარჩენის ფუნქციების შედარება. ორ და უფრო მეტ ჯგუფს შორის გადარჩენის ფუნქციების შედარება ხდება სტატისტიკური კრიტერიუმის გამოყენებით ანუ უნდა შევადაროთ ჯგუფების გადარჩენის ფუნქციის ტოლობის ნულოვანი ჰიპოთეზა $H_0 : S_1(t) = S_2(t)$.

რადგან, როგორც წესი, გადარჩენის ფუნქცია ნორმალურად არ არის განაწილებული, ამიტომ უნდა გამოვიყენოთ არაპარამეტრული კრიტერიუმები. განვიხილოთ ზოგიერთი მათგანი.

გეხანის კრიტერიუმი. გეხანის კრიტერიუმი წარმოადგენს უილკოკსონის განზოგადოებულ კრიტერიუმს, რომელიც გეხმანმა წარმოადგინა 1967წ. ვთქვათ მოცემულია ორი ამონარჩევი (ჯგუფი): $X^1 = \left(t_i^1, d_i^1, r_i^1\right)$, $i = 1,2,...,n_1$ და $X^2 = \left(t_i^2, d_i^2, r_i^2\right)$, $i = 1,2,...,n_2$, სადაც $t_i^1, t_i^2 -$ დროის მომენტებია, $d_i^1, d_i^2 -$ t_i დროის მომენტში დაღუპულ ობიექტთა რაოდენობაა და $r_i^1, r_i^2 -$ t_i დროის მომენტამდე „ცოცხალი“ ობიექტების რაოდენობაა.

დაუშვათ, რომ ორივე ამონარჩევი დამოუკიდებელნი არიან. ნულოვანი ჰიპოთეზის შესამოწმებლად 1-ლი ჯგუფის თითოეული ობიექტი უნდა შევადაროთ 2-რე ჯგუფის თითოეულ ობიექტს. შედარების λ_{ij} შედეგი ტოლია:

$$\lambda_{ij} = \begin{cases} 1, & t_i > t_j \\ -1, & t_i < t_j \\ 0 \end{cases}$$

$\lambda_{ij} = 0$ შესაძლებელია იმ შემთხვევაში, როცა ორივე ობიექტი გამოირიცხა ამონარჩევიდან ან i-ური ობიექტი გამოირიცხა t_j დროის მომენტამდე, ან j-ური ობიექტი გამოირიცხა t_i დროის მომენტამდე, ან $t_i = t_j$. თითოეული ობიექტის შედეგები აიჯამება

$$U = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \lambda_{ij} .$$

სტანდარტული გადახრა განისაზღვრება შემდეგი ფორმულით:

$$\sigma = \sqrt{\frac{n_1 n_2}{(n_1 + n_2)(n_1 + n_2 - 1)} \sum_{i=1}^{n_1} \left(\sum_{j=1}^{n_2} \lambda_{ij} \right)^2} .$$

განვიხილოთ სტატისტიკა $Z_\sigma = \dfrac{U}{\sigma}$, რომელსაც მიახლოებით გააჩნია ნორმალური განაწილება. თუ $Z_\sigma > \Phi_\alpha$, სადაც Φ_α არის ნორმალური განაწილების კვანტილი, მაშინ ნულობანი ჰიპოთეზა უარყოფილია. ე.ი. გადარჩენის ფუნქციები განსხვავდებიან ერთმანეთისაგან.

ლოგარითმული რანგული (ლოგრანგული) კრიტერიუმი. ლოგრანგული კრიტერიუმის გამოყენება ეფუძნება შემდეგ სამ დაშვებას:

– ორი შესადარებელი ამონარჩევი (ჯგუფები) შემთხვევითი და დამოუკიდებელნი არიან;

– ორივე ჯგუფისათვის ობიექტების გამოთიშვის რაოდენობა ერთნაირია;

– გადარჩენის ფუნქციები დაკავშირებულნი არიან შემდეგი გამოსახულებით: $S_2(t) = (S_1(t))^\psi$.

ψ სიდიდეს ობიექტის დაღუპვის ფარდობა ეწოდება. თუ $\psi = 1$, მაშინ გადარჩენის ფუნქციები ერთმანეთს ემთხვევიან. თუ $\psi < 1$, მაშინ ობიექტები მე-2 ჯგუფში იღუპებიან მოგვიანებით, ვიდრე პირველ ჯგუფში. როცა $\psi > 1$, მაშინ მოგვიანებით იღუპებიან პირველ ჯგუფში.

პირველი ჯგუფისათვი t_i დროის მომენტისათვის ობიექტის დაღუპვის რაოდენობა განისაზღვრება ფორმულით:

$$E_t^1 = \frac{r_i^1}{r_i^1 + r_i^2} \left(d_i^1 + d_i^2 \right) ,$$

სადაც $d_i^1, d_i^2 -$ t_i დროის მომენტში დაღუპულ ობიექტთა რაოდენობაა,

$r_i^1, r_i^2 -$ t_i დროის მომენტამდე „ცოცხალი“ ობიექტების რაოდენობაა, გამოკლებული ობიექტების გათვალისწინებით.

ანალოგიურად განისაზღვრება მეორე ჯგუფისათვის:

$$E_t^2 = \frac{r_i^2}{r_i^1 + r_i^2} \left(d_i^1 + d_i^2 \right) ,$$

ხოლო დისპერსია t_i დროის მომენტისათვის განისაზღვრება შემდეგი გამოსახულებით:

$$D_i = \frac{r_i^1 r_i^2 \left(d_i^1 + d_i^2\right)\left(r_i^1 + r_i^2 - d_i^1 - d_i^2\right)}{\left(r_i^1 + r_i^2\right)^2 \left(r_i^1 + r_i^2 - 1\right)} \quad .$$

განვიხილოთ სტატისტიკა

$$LR = \frac{\max\limits_k \left(\sum\limits_i d_i^k - \sum\limits_i E_i^k \right)^2}{\sum\limits_i V_i} \quad ,$$

რომელსაც მიახლოებით გააჩნია ხი-კვადრატ განაწილება $\nu = 1$ თავისუფლების ხარისხით. თუ $LR > \chi^2_{1,\alpha}$, მაშინ ნულოვანი ჰიპოთეზა უარყოფილია.

თუ განვიხილავთ სტატისტიკას

$$Z = \frac{\sum\limits_i \left(d_i^1 - E_i^1\right)}{\sqrt{\sum\limits_i D_i}} \quad ,$$

მაშინ მას გააჩნია მიახლოებით ნორმალური განაწილება. თუ ორივე ჯგუფს გააჩნიათ ერთნაირი გადარჩენის ფუნქციები, მაშინ Z სიდიდეს გააჩნია სტანდარტული ნორმალური განაწილება. როცა $Z > \Phi_\alpha$, სადაც Φ_α ნორმალური განაწილების კვანტილია, მაშინ ნულოვანი ჰიპოთეზა უარყოფილია.

უნდა აღინიშნოს, რომ სულერთია რომელ ჯგუფისათვის განისაზღვრება Z სტატისტიკა, რადგან მეორე ჯგუფის Z აბსოლუტური სიდიდით პირველი ჯგუფის Z აბსოლუტური სიდიდის ტოლია, მაგრამ გააჩნია საწინააღმდეგო ნიშანი.

იეიტსის შესწორება. რადგან Z მნიშვნელობას მიახლოებით გააჩნია ნორმალური განაწილება, ამიტომ ეს იწვევს კრიტერიუმის ე.წ. „დარბილებას“, ანუ ჩვენ უფრო ხშირად, ვიდრე ეს საჭიროა, ვუარყოფთ ნულოვან ჰიპოთეზას. ამ მოვლენის კომპესირებისათვის იყენებენ იეიტსის შესწორებას. მაშინ გვექნება:

$$Z = \frac{\sum\limits_i \left(d_i^1 - E_i^1\right) - 0{,}5}{\sqrt{\sum\limits_i D_i}} \quad .$$

7.5 რეგრესიული მოდელები

გადარჩენის ანალიზის საკმაოდ რთულ ამოცანას წარმოადგენს მყისიერი რისკის ფუნქციის შეფასება, რომელიც წარმოადგენს მცირე დროის მონაკვეთში ობიექტის დაღუპვის ალბათობას, იმის გათვალისწინებით, რომ განსახილველი დროის ინტერვალის დასაწყისში პაციენტი ცოცხალი იყო. ეს წარმოადგენს დავადების განვითარების მნიშვნელობან მახასიათებელს.

მყისიერი რისკის ფუნქციის შეფასების უშუალო განსაზღვრა მოითხოვს დაკვირვებათა დიდ რაოდენობას, ამიტომ გამოიყენება სპეციალური რეგრესიული მოდელები, რომელთა შორის კოკსის პროპორციული ინტესიობის მოდელს განსაკუთრებული ადგილი უკავია.

ბიოსამედიცინო კვლევებში ყველაზე დიდი პრობლემა მდგომარეობს იმის დადგენა არსებობს თუ არა კავშირი ზოგიერთ ცვლადებსა და სიცოცხლის ხანგრძლივობას შორის და თუ არსებობს, მაშინ საჭიროა მათი რიცხობრივი შეფასება. ამ ამოცანის გადაწყვეტა შესაძლებელია რეგრესიის მოდელის აგებით.

არსებობს ორი მთავარი მიზეზი რის გამოც არ შეიძლება კლასიკური მრავლობითი რეგრესიული ანალიზის გამოყენება: 1) ჩვეულებრივ გადარჩენის ფუნქცია არ წარმოადგენს ნორმალურად განაწილებულ ფუნქციას, რის გამოც კლასიკური მრავლობითი რეგრესიული ანალიზის გამოყენება არ შეიძლება, რადგან ამან შეიძლება მიგვიყვანოს მცდარ შედეგებამდე. მაგალითად, არ იქნეს აღმოჩენილი მნიშვნელოვანი რეგრესორები, რომლებიც ხანგრძლივობის დროსთან წრფივად არ არიან დამოკიდებულნი. 2) გარკვეული პრობლემები იქმნება ცენზურირებული მონაცემების დროს.

გადარჩენის ანალიზში ცენზურირებული მონაცემებისათვის გამოიყენება ხუთი რეგრესიული მოდელი:

- კოკსის პროპორციული ინტესიობის მოდელი;
- კოვარიანტების დროზე დამოკიდებული კოკსის მოდელი;
- ექსპონენციალური რეგრესიული მოდელი;
- ნორმალური წრფივი რეგრესიული მოდელი;
- ლოგნორმალური წრფივი რეგრესიული მოდელი.

კოკსის მოდელი. 1972წ კოკსმა შემოგვთავაზა რეგრესიის მოდელი, რომელიც ფართოდ გამოიყენება მედიცინაში და სადაზღვევო საქმეში. მოდელი გამოიყენება ტექნიკურ კვლევებშიც, მაგალითად, ხელსაწყოების მტყუნებოის ინტესიურობის შეფასებისათვის.

კოკსის პროპორციული ინტესიურობის ანუ რისკის მოდელი წარმოადგენს ყველაზე ზოგად რეგრესიულ მოდელს, სადაც რისკის ფუნქცია წარმოდგენილია ორი ფუნქციის ნამრავლის სახით:

$$h(t) = h_0(t)\Psi(z_1, z_2, ..., z_m),$$

სადაც $h_0(t)$– ინტენსივობის საბაზისო ფუნქციაა, რომელიც დამოკიდებულია პაციენტის ასაკზე ან გამოკვლევის შემდეგ გასულ დროზე. $\Psi(z_1, z_2, ..., z_m)$– შესასწავლი პარამეტრების ფუნქცია, მაგალითად, პაციენტის სქესი, ოჯახური მდგომარეობა ან წელიწადის დრო, სასტუმროს კლასი, გამგზავრების ქვეყანა და სხვა. ტექნიკური სისტემებისათვის ეს შეიძლება იყოს მაგალითად, გარემოს ტემპერატურა, სეზონი, ტენიანობა და სხვა. ხშირად კოკსის მოდიელი შეიძლება ასე წარმოვადგინოთ:

$$h[(t),(z_1, z_2, ..., z_m)] = h_0(t)\exp\{b_1 z_1 + \cdots + b_m z_m\}.$$

ინტენსივობის $h_0(t)$ საბაზისო ფუნქცია შეიძლება განვიხილოთ, როგორც ინტესივობის ფუნქცია, როდესაც ყველა პრედიქტორები ნულის ტოლია. ე.ი. ამოცანა მდგიმარეობს შევაფასოდ h_0 და უცნობი $b_1, b_2 \ldots b_m$ კოეფიციენტები.

მოვიყვანოთ ასეთი მაგალითი: დაუშვათ შეისწავლება პაციენტზე გარკვეული ახალი პრეპარატის ზემოქმედება. z წარმოადგენს ბინარულ ცვლადს მნიშვნელობით 1, როცა პაციენტი ღებულობს ახალ პრეპარატს და 0, როცა პაციენტი არ ღებულობს ამ პრეპარატს. მაშინ რისკის ფუნქცია შეგვიძლია ასე ჩავწეროთ:

$$h(t, z) = h_0(t)\exp\{b_1 z_1 + b_2[z\log(t) - 100]\}.$$

დროის ლოგარითმზე კოვარიანტების ნამრავლი საშუალებას იძლევა გავითვალიაწინოთ მაგალითად, ახალი პრეპარატის მიღების დროის ფაქტორი. კონსტანტა 100 გამოიყენება მხოლოდ ნორმირებისათვის, რადგან ამ სიმრავლისათვის გადარჩენის ანალიზის მონაცემების საშუალო ლოგარითმი 100-ს ტოლია. თუ ცნობილია b_1, b_2 კოეფიციენტები და h_0 ინტესიობის ფუნქციის შეფასება, მაშინ ოპერაციის შემდგომი t დროის გასვლის შემდეგ შესაძლებელია მყისიერი რისკის ფუნქციის შეფასება.

ამრიგად, მყისიერი რისკის ფუნქცია კოკსის მოდელში წარმოდგენილია როგორც ორი ფუნქციის ნამრავლი, სადაც ერთი ფუნქცია ახასიათებს ობიექტს, ხოლო მეორე – მყისიერი რისკის ბაზურ ფუნქციას. მოდელის დამოკიდებული ცვლადები ანუ პრედიქტორები განისაზღვრებიან ამოცანიდან გამომდინარე, მაგალითად, პაციენტის სქესი, ასაკი, არსებული დაავადება ან ახალი წამლის მიღება და სხვა. პრედიქტორების შერჩევა ხდება მკვლევარის ინტუიციის დონეზე.

კოკსის მოდელი შეგვიძლია გავაწრფივოთ თუკი მის ორივე ნაწილს გავყოფთ $h_0(t)$ სიდიდეზე და გავალოგარითმებთ:

$$\log\left\{\frac{h[(t),(z_1,z_2,...)]}{h_0(t)}\right\} = b_1z_1 + b_2z_2 + \cdots + b_mz_m \quad .$$

ასეთი მოდელის უპირატესობა იმაში მდგომარეობს, რომ მოდელის პარამეტრების შეფასებისათვის საჭიროა უფრო მცირე მონაცემები, ვიდრე არაწრფივი მოდელის დროს.

ამრიგად, კოკსის მოდელი ეფუძნება ორ მოსაზრებას: 1) ინტესივობის კოვარიანტების ლოგწრფივი ფუნქციების მიმართ წარმოადგენს მულტიპლიკატიურს. ამ მოსაზრებას ეწოდება პროპორციულობიოს ფუნქციის დამოკიდებულება ჰიპოთეზა. რეალურად ეს იმას ნიშნავს, რომ დამოუკიდებელი ცვლადების ორი სხვადასხვა მნიშვნელობის დაკვირვების ინტესივობის ფუნქციების შეფარდება დროზე დამოკიდებული არ არის. 2) მეორე მოსაზრება მდგომარეობს ინტესივობის ფუნქციის რეგრესორებთან ლოგწრფივ დამოკიდებულებაში.

მოსაზრება რისკების პროპორციულობაზე ხშირად საეჭვო ხდება. შეიძლება მოვიყვანოთ მრავალი მაგალითი სადაც რისკების პროპორციულობის ჰიპოთეზა მიუღებელია. ასე მაგალითად, ფიზიკური წამრთელობის შესწავლისას ასაკი თამაშობს მნიშვნელობან როლს ქირურგიული ოპერაციის შემდგომ პაციენტის გადარჩენაში. ცხადია, რომ ასაკი უფრო მნიშვნელობანი პრედიქტორია ოპერაციის შემდეგ რისკის განვითარებაში, ვიდრე ოპერაციის შემდეგ გასული დროის გარკვეული ინტერვალის შემდეგ. აქედან გამომდინარე, რეალურია სხვა მოდელი, როცა ოპერაციის შემდგომ მაშინვე პაციენტის რისკი მაღალია, მაგრამ ოპერაციის წარმატების შემთხვევაში დროთა განმავლობაში რისკი მცირდება. ამ შემთხვევაში გამოიყენება კოვარიანტების დროზე დამოკიდებული კოკსის მოდელი.

ექსპონენციალური რეგრესია. ექსპონენციალური რეგრესიის მოდელი შეიძლება ასე წარმოვადგინოთ:

$$S(z) = \exp\{a + b_1z_1 + b_2z_2 + \cdots + b_mz_m\}$$

სადაც $S(z)$ წარმოადგენს სიცოცხლის ხანგრძლივობას, a – დამოუკიდებელი კონსტანტაა, b_i – რეგრესიის პარამეტრები.

მოდელის ადეკვატურობის დასადგენად გამოიყენება χ^2 კრიტერიუმი. სტატისტიკა χ^2 შეიძლება განისაზღვროს როგორც მოდელის ყველა შეფასებული პარამეტრების დასაჯერობის ფუნქციის ლოგარითმი (L_1) და მოდელის დასაჯერობის ლოგარითმი (L_0), სადაც კოვარიანტა ნულის ტოლია. თუ χ^2 მნიშვნელობა სარწმუნოა, მაშინ ნულოვანი ჰიპოთეზა უარყოფილია და გამოგვაქვს დასკვნა: დამოუკიდებელი ცვლადები მნიშვნელოვნად მოქმედებენ სიცოცხლის ხანგრძლივობაზე.

ნორმალური და ლოგნორმალური რეგრესია. აქ იგულისხმება, რომ ხანგრძლივიობის ფუნქცია (ან მისი ლოგარითმი) ნორმალურად არის განაწილებული. ამ შემთხვევაში მოდელი ემთხვევა მრავლობითი რეგრესიის მოდელს და შესაძლებელია მისი ჩაწერა შემდეგნაირად:

$$S(z) = a + b_1 z_1 + b_2 z_2 + \cdots + b_m z_m ,$$

თუ საქმე გვაქვს ლოგნორმალურ რეგრესიასთან, მაშინ $S(z)$ იცვლება $\ln S(z)$-ით. ნორმალური რეგრესიის მოდელი განსაკუთრებით სასარგებლოა, რადგან მონაცემთა ნაწილი შეიძლება გარდავქმნათ მიახლოებით ნორმალურ განაწილებაზე. აქედან გამომდინარე, გარკვეული მოსაზრებით ეს მოდელი წარმოადგენს პარამეტრულს კოკსის მოდელთან განსხვავებით, რომელიც არაპარამეტრულ მოდელს წარმოადგენს.

ლოგნორმალური განაწილების რეგრესიის დროს რისკის ფუნქცია დასაწყისში იზრდება, ხოლო შემდეგ ეცემა ნულამდე. შეიძლება ვიგულისხმოდ, რომ ეს მოდელი ადეკვატური იქნება კერძოდ, ტრამვების შედეგად სიკვდილი დადგომის დროისთვის. მართლაც, თუ კრიტიკული პერიოდის გასვლის შემდეგ პაციენტი გადარჩა, მაშინ ტრამვის შედეგად გამოწვეული მისი სიკვდილის შანსი დროის განმავლობაში სულ უფრო და უფრო მცირდება.

დ ა ნ ა რ თ ი

სტანდარტიზირებული ნორმალური განაწილების ფუნქციის $F(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{t^2}{2}}dt$ მნიშვნელობები

z	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0,1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5754
0,2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0,4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5	6915	6950	6985	7020	7054	7088	7123	7157	7190	7224
0,6	7258	7291	7324	7357	7389	7422	7454	7486	7518	7549
0,7	7580	7612	7642	7673	7704	7734	7764	7794	7823	7852
0,8	7881	7910	7939	7967	7996	8023	8051	8079	8106	8133
0,9	8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1,0	0,8413	8438	8461	8485	8508	8531	8554	8577	8599	8621
1,1	8643	8665	8686	8708	8729	8749	8770	8790	8810	8830
1,2	8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
1,3	9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
1,4	9192	9207	9222	9236	9251	9265	9279	9292	9306	9319
1,5	9332	9345	9357	9370	9382	9394	9407	9418	9430	9441
1,6	9452	9463	9474	9485	9495	9505	9515	9525	9535	9545
1,7	9554	9564	9573	9582	9591	9599	9608	9616	9625	9633
1,8	9641	9649	9656	9664	9671	9678	9686	9693	9700	9706
1,9	9713	9720	9726	9732	9738	9744	9750	9756	9762	9767
2,0	0,9773	9778	9783	9788	9793	9798	9803	9808	9812	9817
2,1	9821	9826	9830	9834	9838	9842	9846	9850	9854	9857
2,2	9861	9865	9868	9871	9875	9878	9881	9984	9887	9890
2,3	9893	9896	9898	9901	9904	9906	9909	9911	9913	9916
2,4	9918	9920	9922	9925	9927	9929	9931	9932	9934	9936
2,5	9938	9940	9941	9943	9945	9946	9948	9949	9951	9952
2,6	9953	9955	9956	9957	9959	9960	9961	9962	9963	9964
2,7	9965	9966	9967	9968	9969	9970	9971	9972	9973	9974
2,8	9974	9975	9976	9977	9977	9978	9979	9980	9980	9981
2,9	9981	9982	9983	9983	9984	9984	9985	9985	9986	9986
3,0	0,9987	9987	9987	9988	9988	9989	9989	9989	9990	9990
3,1	9990	9991	9991	9991	9992	9992	9992	9992	9993	9993
3,2	9993	9993	9994	9994	9994	9994	9994	9995	9995	9995
3,3	9995	9995	9996	9996	9996	9996	9996	9996	9996	9997
3,4	9997	9997	9997	9997	9997	9997	9997	9997	9998	9998
3,5	9998	9998	9998	9998	9998	9998	9998	9998	9998	9998
3,6	9998	9999	9999	9999	9999	9999	9999	9999	9999	9999

სტანდარტიზირებული ნორმალური განაწილების სიმკვრივის ფუნქციის $f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$ მნიშვნელობები

z	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1694	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0904	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

ლაპლასის $\Phi(U_i) = \frac{2}{\sqrt{2\pi}} \int_0^{u_i} e^{-\frac{x^2}{2}} dx$ ფუნქციის მნიშვნელობები

U_i	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0080	0160	0239	0319	3999	0478	0558	0638	0717
0,1	0797	0878	0955	1034	1113	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2960	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	1778	4843	4907	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6679	6729	6778
1,0	0,6827	6875	6923	6970	7017	7063	7109	7154	7199	7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7994	8029
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8358	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9090
1,7	9109	9127	9146	9164	9181	9199	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9421	9439	9451	9464	9476	9488	9500	9512	9523	9534
2,0	0,9545	9556	9566	9576	9586	9596	9606	9616	9625	9634
2,1	9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9849	9756	9762	9768	9774	9780
2,3	9786	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9841	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9910	9912	9915	9917	9920	9922	9924	9926	9928
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9951	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	0,9973	9974	9975	9976	9976	9977	9978	9979	9979	9980
3,1	9981	9981	9982	9983	9983	9984	9984	9985	9985	9986
3,2	9986	9987	9987	9988	9988	9989	9989	9989	9990	9990
3,3	9990	9991	9991	9991	9992	9992	9992	9992	9993	9993
3,4	9993	9994	9994	9994	9994	9994	9995	9995	9995	9995
3,5	9995	9996	9996	9996	9996	9996	9996	9996	9997	9997
3,6	9997	9997	9997	9997	9997	9997	9997	9998	9998	9998
3,7	9998	9998	9998	9998	9998	9998	9998	9998	9998	9998
3,8	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
3,9	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999

სტიუდენტის განაწილება

α / ν	ცალმხრივი კრიტერიუმი							
	0,30	0,20	0,10	0,05	0,025	0,01	0,005	0,001
1	0,727	1,376	3,078	6,314	12,71	31,82	63,66	318,3
2	0,617	1,061	1,886	2,920	4,303	6,965	9,925	22,33
3	0,584	0,978	1,638	2,353	3,182	4,541	5,841	10,22
4	0,569	0,941	1,533	2,132	2,776	3,747	4,604	7,173
5	0,559	0,906	1,440	1,943	2,447	3,143	3,707	5,208
6	0,553	0,920	1,476	2,015	2,571	3,365	5,032	5,893
7	0,549	0,896	1,415	1,895	2,365	2,998	3,499	4,785
8	0,546	0,889	1,397	1,860	2,306	2,896	3,355	4,501
9	0,543	0,883	1,383	1,833	2,262	2,821	3,250	4,297
10	0,542	0,879	1,372	1,812	2,228	2,764	3,169	4,144
11	0,540	0,876	1,363	1,796	2,201	2,718	3,106	4,025
12	0,539	0,873	1,356	1,782	2,179	2,681	3,055	3,930
13	0,538	0,870	1,350	1,771	2,160	2,650	3,012	3,852
14	0,537	0,868	1,345	1,761	2,145	2,624	3,977	3,787
15	0,536	0,866	1,341	1,753	2,131	2,602	2,947	3,733
16	0,535	0,865	1,337	1,746	2,120	2,583	2,921	3,686
17	0,534	0,863	1,333	1,740	2,110	2,567	2,898	3,646
18	0,534	0,862	1,330	1,734	2,101	2,552	2,878	3,611
19	0,533	0,861	1,328	1,729	2,093	2,539	2,861	3,579
20	0,533	0,860	1,325	1,725	2,086	2,528	2,845	3,552
21	0,532	0,859	1,323	1,721	2,080	2,518	2,831	3,527
22	0,532	0,858	1,321	1,717	2,074	2,508	2,819	3,505
23	0,532	0,858	1,319	1,714	2,069	2,500	2,807	3,485
24	0,531	0,857	1,318	1,711	2,064	2,492	2,797	3,467
ν	0,60	0,40	0,20	0,10	0,05	0,02	0,01	0,002
α	ორმხრივი კრიტერიუმი							

სტიუდენტის განაწილება (გაგრძელება)

α / ν	ცალმხრივი კრიტერიუმი							
	0,30	0,20	0,10	0,05	0,025	0,01	0,005	0,001
25	0,531	0,856	1,316	1,708	2,060	2,485	2,787	3,450
26	0,531	0,856	1,315	1,706	2,056	2,479	2,779	3,435
27	0,531	0,855	1,314	1,703	2,052	2,473	2,771	3,421
28	0,530	0,855	1,313	1,701	2,048	2,467	2,763	3,408
29	0,530	0,854	1,311	1,699	2,045	2,462	2,756	3,398
30	0,530	0,854	1,310	1,697	2,042	2,457	2,750	3,385
40	0,529	0,851	1,303	1,684	2,021	2,423	2,704	3,307
50	0,528	0,849	1,298	1,676	2,009	2,403	2,678	3,262
60	0,527	0,848	1,296	1,671	2,000	2,390	2,660	3,232
80	0,527	0,846	1,292	1,664	1,990	2,374	2,639	3,195
100	0,526	0,845	1,290	1,660	1,984	2,365	2,626	3,174
200	0,525	0,843	1,286	1,653	1,972	2,345	2,601	3,131
500	0,525	0,842	1,283	1,648	1,965	2,334	2,586	3,106
∞	0,524	0,842	1,282	1,645	1,960	2,326	2,576	3,090
ν / α	0,60	0,40	0,20	0,10	0,05	0,02	0,01	0,002
	ორმხრივი კრიტერიუმი							

χ^2 განაწილება

ν \ σ	0,50	0,30	0,20	0,10	0,05	0,025	0,01	0,001
1	0,455	1,07	1,64	2,71	3,84	5,02	6,63	10,83
2	1,39	2,41	3,22	4,61	5,99	7,38	9,21	13,82
3	2,37	3,66	4,64	6,25	7,81	9,35	11,34	16,27
4	3,36	4,88	5,99	7,78	9,49	11,14	13,28	18,47
5	4,35	6,06	7,29	9,24	11,07	12,83	15,09	20,52
6	5,35	7,23	8,56	10,64	12,59	14,45	16,81	22,46
7	6,35	8,38	9,80	12,02	14,07	16,01	18,48	24,32
8	7,34	9,52	11,0	13,36	15,51	17,53	20,09	26,12
9	8,34	10,7	12,2	14,68	16,92	19,02	21,67	27,88
10	9,34	11,8	13,4	15,99	18,31	20,48	23,21	29,59
11	10,3	12,9	14,6	17,28	19,68	21,92	24,73	31,26
12	11,3	14,0	15,8	18,55	21,03	23,34	26,22	32,91
13	12,3	15,1	17,0	19,81	22,36	24,74	27,69	34,53
14	13,3	16,2	18,2	21,06	23,68	26,12	29,14	36,12
15	14,3	17,3	19,3	22,31	25,00	27,49	30,58	37,70
16	15,3	18,4	20,5	23,54	26,30	28,85	32,00	39,25
17	16,3	19,5	21,6	24,77	27,59	30,19	33,41	40,79
18	17,3	20,6	22,8	25,99	28,87	31,53	34,81	42,31
19	18,3	21,7	23,9	27,20	30,14	32,85	36,19	43,82
20	19,3	22,8	25,0	28,41	31,41	34,17	37,57	45,32
22	21,3	24,9	27,3	30,81	33,92	36,78	40,29	48,27
24	23,3	27,1	29,6	33,20	36,42	39,36	42,98	51,18
26	25,3	29,2	31,8	35,56	38,88	41,92	45,64	54,05
28	27,3	31,4	34,0	37,92	41,34	44,46	48,28	56,89
30	29,3	33,5	36,2	40,26	43,77	46,98	50,89	59,70
35	34,3	38,9	41,8	46,06	49,80	53,20	57,34	66,62
40	39,3	44,2	47,3	51,81	55,76	59,34	63,69	73,40
50	49,3	54,7	58,2	63,17	67,50	71,42	76,15	86,66
60	59,3	65,2	69,0	74,40	79,08	83,30	88,38	99,61
80	79,3	86,1	90,4	96,58	101,88	106,63	112,33	124,84
100	99,3	106,9	111,7	118,50	124,34	129,56	135,81	149,45
120	119,3	127,6	132,8	140,23	146,57	152,21	158,95	173,62
150	149,3	158,6	164,6	172,6	179,6	185,8	193,2	209,3
200	199,3	210,0	216,6	226,0	234,0	241,1	249,4	267,5

ფიშერის განაწილება (F განაწილება) α = 0,05

ν_2 \ ν_1	1	2	3	4	5	6	8	12	20	24	∞
1	161	200	216	225	230	234	239	244	248	249	254
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,44	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,66	8,64	8,53
4	7,71	6,95	6,59	6,39	6,26	6,16	6,04	5,91	5,80	5,77	5,63
5	6,61	5,79	5,41	5.19	5,05	4,95	4,82	4,68	4,56	4,53	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,87	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,44	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,15	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,93	2,90	2,71
10	4,97	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,77	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,65	2,61	2,41
12	4,75	3,89	3,49	3,26	3,11	3,00	2,85	2,69	2,54	2,51	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,77	2,60	2,46	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,39	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,33	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,28	2,24	2,01
17	4,45	3,59	3,20	2,97	2,81	2,70	2,55	2,38	2,23	2,19	1,96
18	4,41	3,56	3,16	2,93	2,77	2,66	2,51	2,34	2,19	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,15	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,12	2,08	1,84
21	4,33	3,47	3,07	2,84	2,69	2,57	2,42	2,25	2,09	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,07	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,04	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	2,02	1,98	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,34	2,17	2,00	1,97	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,99	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,31	2,13	1,97	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,29	2,12	1,96	1,92	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,28	2,10	1,94	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,93	1,89	1,62
40	4,09	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,84	1,79	1,51
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,75	1,70	1,39
120	3,92	3,07	2,68	2,45	2,29	2,18	2,02	1,83	1,66	1,61	1,25
∞	3,84	3,00	2,61	2,37	2,21	2,10	1,94	1,75	1,57	1,52	1,00

ფიშერის განაწილება (F განაწილება) α = 0,01

ν_2 \ ν_1	1	2	3	4	5	6	8	12	20	24	∞
1	4052	4999	5403	5625	5764	5859	5981	6106	6208	6234	6366
2	98,49	99,00	99,17	99,25	99,30	99,33	99,37	99,42	99,45	99,46	99,50
3	34,12	30,82	29,46	28,71	28,24	27,91	27,49	27,05	26,69	26,60	26,12
4	21,20	18,00	16,69	15,98	15,52	15,21	14,80	14,37	14,02	13,93	13,42
5	16,26	13,27	12,06	11,39	10,97	10,67	10,29	9,89	9,55	9,47	9,02
6	13,74	10,92	9,78	9,15	8,75	8,47	8,10	7,72	7,39	7,31	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,84	6,47	6,15	6,07	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,03	5,67	5,36	5,28	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,47	5,11	4,80	4,73	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,06	4,71	4,41	4,33	3,91
11	9,65	7,20	6,22	5,67	5,32	5,07	4,74	4,40	4,10	4,02	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,86	3,78	3,36
13	9,07	6,70	5,74	5,20	4,86	4,62	4,30	3,96	3,67	3,59	3,16
14	8,86	6,51	5,56	5,03	4,69	4,46	4,14	3,80	3,51	3,43	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,00	3,67	3,36	3,29	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,55	3,25	3,18	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,79	3,45	3,16	3,08	2,65
18	8,28	6,01	5,09	4,58	4,25	4,01	3,71	3,37	3,07	3,00	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,63	3,30	3,00	2,92	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,23	2,94	2,86	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,51	3,17	2,88	2,80	2,36
22	7,94	5,72	4,82	4,31	3,99	3,76	3,45	3,12	2,83	2,75	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,41	3,07	2,78	2,70	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,36	3,03	2,74	2,66	2,21
25	7,77	5,57	4,68	4,18	3,86	3,63	3,32	2,99	2,70	2,62	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,29	2,96	2,66	2,58	2,13
27	7,68	5,49	4,60	4,11	3,79	3,56	3,26	2,93	2,63	2,55	2,10
28	7,64	5,45	4,57	4,07	3,76	3,53	3,23	2,90	2,60	2,52	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,20	2,87	2,57	2,49	2,03
30	7,56	5,39	4,51	4,02	2,70	3,47	3,17	2,84	2,55	2,47	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	2,99	2,66	2,37	2,29	1,61
60	7,08	4,98	4,13	3,65	3,34	3,12	2,82	2,36	2,20	2,12	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,66	2,34	2,03	1,95	1,38
∞	6,63	4,60	3,78	3,32	3,02	2,80	2,51	2,18	1,87	1,79	1,00

q კრიტიკული მნიშვნელობები ($\alpha' = 0,05$)

ν \ *l*	2	3	4	5	6	7	8	9	10
1	17,97	26,98	32,82	37,08	40,41	43,40	45,40	47,36	49,07
2	6,09	8,33	9,80	10,88	11,74	12,44	13,03	13,54	13,99
3	4,50	5,91	6,83	7,50	8,04	8,48	8,85	9,18	9,46
4	3,93	5,04	5,76	6,29	6,71	7,05	7,35	7,60	7,83
5	3,64	4,60	5,22	5,67	6,03	6,33	6,58	6,80	7,00
6	3,46	4,34	4,90	5,31	5,63	5,90	6,12	6,32	6,49
7	3,34	4,17	4,68	5,06	5,36	5,61	5,82	6,00	6,16
8	3,26	4,04	4,53	4,89	5,17	5,40	5,60	5,77	5,92
9	3,20	3,95	4,42	4,76	5,02	5,24	5,43	5,60	5,74
10	3,15	3,88	3,33	4,65	4,91	5,12	5,31	5,46	5,60
11	3,11	3,82	4,26	4,57	4,82	5,03	5,20	5,35	5,49
12	3,08	3,77	4,20	4,51	4,75	4,95	5,12	5,27	5,40
13	3,06	3,74	4,15	4,46	4,69	4,89	5,05	5,19	5,32
14	3,03	3,70	4,11	4,41	4,64	4,83	5,00	5,13	5,25
15	3,01	3,67	4,08	4,37	4,60	4,78	4,94	5,08	5,20
16	3,00	3,65	4,05	4,33	4,56	4,74	4,90	5,03	5,15
17	2,98	3,63	4,02	4,30	4,52	4,71	4,86	4,99	5,11
18	2,97	3,61	4,00	4,28	4,50	4,67	4,82	4,96	5,07
19	2,96	3,59	3,98	4,25	4,47	4,65	4,79	4,92	5,04
20	2,95	3,58	3,96	4,23	4,45	4,62	4,77	4,90	5,01
24	2,92	3,53	3,90	4,17	4,37	4,54	4,68	4,81	4,92
30	2,89	3,49	3,85	4,10	4,30	4,46	4,60	4,72	4,82
40	2,86	3,44	3,79	4,04	4,23	4,39	4,52	4,64	4,74
60	2,83	3,40	3,74	3,98	4,16	4,31	4,44	4,55	4,65
120	2,80	3,36	3,69	3,92	4,10	4,24	4,36	4,47	4,57
∞	2,77	3,31	3,63	3,86	4,03	4,17	4,29	4,39	4,47

q კრიტიკული მნიშვნელობები ($\alpha' = 0{,}01$)

l / ν	2	3	4	5	6	7	8	9	10
1	90,03	135,0	164,3	185,6	202,2	215,8	227,2	237,0	245,5
2	14,04	90,02	22,29	24,72	26,63	28,20	29,53	30,68	31,69
3	8,26	10,62	12,17	13,33	14,24	15,00	15,64	16,20	16,69
4	6,51	8,12	9,17	9,96	10,58	11,10	11,55	11,93	12,27
5	5,70	6,98	7,80	8,42	8,91	9,32	9,70	9,97	10,24
6	5,24	6,33	7,03	7,56	7,97	8,32	8,61	8,87	9,10
7	4,95	5,92	6,54	7,01	7,37	7,68	7,94	8,17	8,37
8	4,75	5,64	6,20	6,63	6,96	7,24	7,47	7,68	7,86
9	4,60	5,43	5,96	6,35	6,66	6,92	7,13	7,33	7,50
10	4,48	5,27	5,77	6,14	6,43	6,67	6,88	7,06	7,21
11	4,39	5,15	5,62	5,97	6,25	6,48	6,67	6,84	6,99
12	4,32	5,05	5,50	5,84	6,10	6,32	6,51	6,67	6,81
13	4,26	4,96	5,40	5,73	5,98	6,19	6,37	6,53	6,67
14	4,21	4,90	5,32	5,63	5,88	6,09	6,26	6,41	6,54
15	4,17	4,84	5,25	5,56	5,80	5,99	6,16	6,31	6,44
16	4,13	4,79	5,19	5,49	5,72	5,92	6,08	6,22	6,35
17	4,10	4,74	5,14	5,43	5,66	5,85	6,01	6,15	6,27
18	4,07	4,70	5,09	5,38	5,60	5,79	5,94	6,08	6,20
19	4,05	4,67	5,05	5,33	5,55	5,74	5,89	6,02	6,14
20	4,02	4,64	5,02	5,29	5,51	5,69	5,84	5,97	6,09
24	3,96	4,55	4,91	5,17	5,37	5,54	5,69	5,81	5,92
30	3,89	4,46	4,80	5,05	5,24	5,40	5,54	5,65	5,76
40	3,83	4,37	4,70	4,93	5,11	5,27	5,39	5,50	5,56
60	3,76	4,28	4,60	4,82	4,99	5,13	5,25	5,36	5,45
120	3,70	4,20	4,50	4,71	4,87	5,01	5,12	5,21	5,30
∞	3,64	4,12	4,40	4,60	4,76	4,88	4,99	5,08	5,16

Q კრიტიკული მნიშვნელობები

m	2	3	4	5	6	7	8	9
$\alpha = 0{,}05$	1,96	2,39	2,64	2,81	2,94	3,04	3,12	3,20
$\alpha = 0{,}01$	2,58	2,94	3,14	3,29	3,40	3,49	3,57	3,64

m	10	11	12	13	14	15	16	17
$\alpha = 0{,}05$	3,26	3,32	3,37	3,41	3,46	3,49	3,53	3,56
$\alpha = 0{,}01$	3,69	3,74	3,79	3,83	3,87	3,90	3,94	3,97

m	18	19	20	21	22	23	24	25
$\alpha = 0{,}05$	3,59	3,62	3,65	3,68	3,70	3,72	3,74	3,77
$\alpha = 0{,}01$	3,99	4,02	4,04	4,07	4,09	4,11	4,13	4,15

ლიტერატურა

1. ე. ყუბანეიშვილი ბიომრტრია. დამხმარე სახელმძღვანელო. თბილისი, 2005
2. ე. ყუბანეიშვილი. ბიოსტატისტიკა. ლექციების კურსი, სტუ, 3013. http://gtu.ge/books/ biostatistika . pdf
3. Болч Б. Хуань К.Дж. многомерные статистические методы для экономики. М.,1979.
4. Гланц С. Медиико-биологическая статистика, М.Практика,1999
5. Реброва О.Ю. Статистический анализ медицинских данных М.,Медио Сфера, 2002 .
6. Юнкеров В.И. Григорьев С.Г. Математико-статистическая обработка данных Медцинск Исследований. СПб. ВМедА, 2002
7. Дубров А.М. и др. Многомерные статистические методы.,М.Финансы и статистика,2003.
8. Максимов Г.К.,Синицин А.Н. Статистическое моделирование многомерных Систем в медицине., М."Медицина"б 1983.