1. ეკონომეტრიკის განმარტება

ეკონომეტრიკა არის დისციპლინა, რომელიც მე-20 საუკუნეში ეკონომიკისა და მათემატიკის შერწყმის შედეგად ჩამოყალიბდა. ეკონომეტრიკა შეიძლება განვმარტოთ როგორც დისციპლინა, რომელიც ცდილობს სტატისტიკური მეთოდებით დაადგინოს რაოდენობრივი კავშირურთიერთობები ეკონომიკურ ცვლადებს შორის.

თანამედროვე ეკონომეტრიკა ორი ნაწილისგან შედგება: ეკონომეტრიკული მეთოდები და ეკონომეტრიკული მეთოდების გამოყენება კონკრეტულ ეკონომიკურ ამოცანებში. ეკონომეტრიკული მეთოდები ეფუძნება რეგრესიულ ანალიზს, რომელიც სტატისტიკის თეორიისა და სტატისტიკის მეთოდების ჩვეულებრივი შემადგენელი ნაწილია. ამავე დროს, პრობლემის სპეციფიკა, რომლის გადაწყვეტისთვისაც გამოიყენება ეკონომეტრიკის მეთოდები, განაპირობებს სტატისტიკის ტრადიციულ მიდგომებში როგორც გარკვეულ ცვლილებების შეტანის აუცილებლობას, ისე ახალი მიდგომების შემუშავებას.

გამოყენებითი ნაწილი კი მოიცავს ეკონომიკურ თეორიაზე დაფუძნებული და შესაბამისი ეკონომეტრიკული მოდელების აგებით და გამოკვლევით ეკონომიკური კანონზომიერებების გამოვლენას ან ჰიპოთეტური თეორიული მოსაზრების ემპირიულ დადასტურებას ემსახურება. აქედან გამომდინარე ეკონომეტრიკა შეიძლება განვიხილოთ როგორც სტატისტიკისა და ეკონომიკური თეორიის მიმართულებათა სინთეზი. ამ სინთეზის მნიშვნელოვან როლს მათემატიკა ასრულებს. მის საფუძველზე ხდება თეორიული კავშირურთიერთობების მოდელირება და სტატისტიკური მეთოდების შემოწმება.

2. ეკონომეტრიკული მოდელების სახეები.

ეკონომეტრიკული მოდელები პირობითად შეიძლება სამ სახედ დავყოთ:

1. დროითი მწკრივების მოდელები - აქ შედის რეგრესიული მოდელთა განსაკუთრებული ჯგუფი, რომლებშიც რაიმე შესაბამისი y ცვლადის ქცევა

ასახვა ხდება ან ერთეული დროის ფაქტორის t მნიშვნელობით, ან კიდევ y ცვლადის დროში წინამსწრები მონაცემების მნიშვნელობით.

დროითი მწკრივების მოდელებიდან განსაკუთრებით მარტივი და გავრცელებულია ტრენდული, სეზონური და შერეული (ტრენდულ-სეზონური) მოდელები.

ტრენდული მოდელის ზოგადი სახეა

$$y(t) = T(t) + u(t),$$

სადაც $T(t)$ არის პარამეტრული მოცემული დროითი ტრენდი.

წმინდა სეზონური მოდელს წარმოადგენს

$$y(t) = S(t) + u(t),$$

სადაც $S(t)$ სეზონური კომპონენტია, ხოლო $u(t)$ შემთხვევითი კომპონენტი.

შერეულ ანუ ტრენდულ-სეზონურ მოდელს შეიძლება ჰქონდეს ადიტიური

$$y(t) = T(t) + S(t) + u(t),$$

ან, კიდევ, მულტიპლიკაციური სახე ჰქონდეს

$$y(t) = T(t) \cdot S(t) + u(t).$$

2. ერთი განტოლებით მოცემული რეგრესიული მოდელები – ეკონომეტრიკული მოდელის ენდოგენური ერთი რეგრესიული განტოლების საშუალებით აიხსნება. ზოგად შემთხვევაში ამ ტიპის მოდელს აქვს სახე

$$y = f(x_1, x_2, \dots, x_m) + u,$$

სადაც y დამოკიდებული ანუ შედეგობრივი ცვლადია; $x_1, x_2, \dots, x_m$ დამოუკიდებელი ცვლადებია, რომლებსაც, აგრეთვე, ფაქტორულ ცვლადებს უწოდებენ; u – შემთხვევითი სიდიდეა. როცა $m \geq 2$, მაშინ ამ მოცემულ განტოლებას მრავლობითი რეგრესიის მოდელი ეწოდება, მაგრამ თუ ეკონომეტრიკული მოდელი მხოლოდ ერთი დამოუკიდებელი ცვლადის შემცველია, ე.ი. თუ $y = f(x) + u$, მაშინ გვაქვს წყვილური ანუ მარტივი რეგრესიის მოდელი. როგორც წყვილური, ისე მრავლობითი რეგრესიის მოდელი შეიძლება, იყოს წრფივი ან არაწრფივი. წრფივი მოდელები შედარებით მარტივია და ამიტომ უფრო ხშირად შეისწავლიან მათ.

3. ერთდროულ განტოლებათა სისტემები – ამ ტიპის ეკონომეტრიკული მოდელი რამდენიმე რეგრესიული განტოლებისგან და შესაძლოა, აგრეთვე, რამდენიმე იგივეობისგანაც შედგება. ამასთან, ეს განტოლებები და იგივეობები ერთ-

მათთან ამ თვალსაზრისით დაკავშირებული, რომ ყოველი მათგანი საკუთარი ახსნადი ცვლადის გარდა აუხსნელ ცვლადებსაც შეიცავს, რომლებიც იმავდროულად სხვა განტოლებებისთვის ახსნად ცვლადებს წარმოადგენს.

სხვა ეკონომეტრიკული მოდელების მსგავსად, ერთდროული განტოლებათა სისტემის სახით აგებული ეკონომეტრიკული მოდელებიც ენდოგენურ და ეგზოგენურ ცვლადებს შეიცავს და საშუალებას იძლევა, აუხსნათ ენდოგენური ცვლადის ქცევა ეგზოგენური ცვლადების ქცევის საფუძველზე.

ერთდროული განტოლებათა სისტემები გამოიყენება როგორც მიკრო, ისე მაკრო დონეზე. მათ შორის კლასიკური მიწოდება-მოთხოვნისა და IS-LM მოდელი.

3. ეკონომეტრიკული მოდელების ეტაპები და პრობლემები.

ეკონომეტრიკული მოდელის შემუშავება, გამოკვლევა და გამოყენება ზოგადად ეტაპებად ხორციელდება:

აპრიორული დონე - აპრიორული დონე გულისხმობს შესასწავლი ობიექტის მახასიათებლების შედეგობრივი და ფაქტორული მაჩვენებლების გამოვლენას და მათ დადგენას, თუ რომელი მაჩვენებლის რომელ მაჩვენებელთან ან მაჩვენებელთა სისტემასთან დამოკიდებულება უნდა იქნეს შესწავლილი.

ამავე ეტაპზე ხდება გამოკვლევის მიზნის ფორმირება. მიზანი შეიძლება იყოს:

- ობიექტში არსებული რაოდენობრივი კანონზომიერებების გამოვლენა და გამოკვლევა;
- ობიექტის ეკონომიკური მაჩვენებლების პროგნოზირება;
- ობიექტის ან მოდელის განვითარების იმიტაცია სხვადასხვა ეგზოგენური ზემოქმედების პირობებში;
- მმართველობითი გადაწყვეტილებების შემუშავება.

ფაქტორთა შერჩევა - ფაქტორთა შერჩევისას უნდა განისაზღვროს ზოგადი სახით რაიმე ლოგიკური მოსაზრებით.

- მოდელში უნდა აისახოს მხოლოდ პრობლემის შესწავლისთვის არსებითი

ფაქტორები რეკომენდირებულია, რომ მათი რიცხვი არ იყოს ძალიან დიდი და რაოდენობრივი ნაწილი იყოს დაავადების რისკზე.

- გამოყენებული ფაქტორები საიმედოობით უნდა იზომებოდეს. ალტერნატ-იული შემთხვევაში დასაშვებია ალტერნატიული ფაქტორების გამოყენება.

- არ უნდა მოხდეს ფაქტორული ნიშნების დუბლირება. ყოველი ფაქტორი წარმოდგენილი უნდა იყოს ნატურალური ან ღირებულებითი, აბსოლუტური ან შეფარდებითი ფორმით და არა რომელიმე ორი მათგანით. არ შეიძლება ჯამური ნიშნების და მათი შემადგენელი ნაწილების ერთდროული ჩართვა მოდელში.

- მიზანშეწონილია, რომ მიზეზ-შედეგობრივი კავშირი ფაქტორები ერთ დონეზე იყოს განლაგებული, ანუ ახსნიდნენ და არა ურთიერთდაქვემდებარებულ წარმოადგენდნენ.

- ისეთი ფაქტორების ერთდროული გათვალისწინება, რომელთა შორისაც არსე-ბობს ფუნქციონალური ან მჭიდრო კორელაციური დამოკიდებულება არა-მიზანშეწონილია.

- მოდელში არ უნდა ჩართოთ ისეთი ფაქტორები, რომლებიც შესახებაც ჯგუფური მონაცემების მიღება შეუძლებელია ან ძალიან დიდ დანახარჯებთან დაკავშირე-ბული.

<u>საწყისი მონაცემების შეგროვება და დამუშავება</u>. მონაცემთა შეგროვების პროცესს სტატისტიკური დაკვირვება ეწოდება. დაკვირვების შედეგად მიიღე-ბა სტატისტიკური ინფორმაცია, რომელიც წარმოადგენს ეკონომიკური მოვ-ლენის საფუძველს. სტატისტიკური ინფორმაცია ერთი და იმავე სახის ელემენტების ანუ ერთეულების სიმრავლეა, რომელიც სხვადასხვა ნიშნების მიხედვით იცვლება.

სტატისტიკური ინფორმაცია შეიძლება იყოს სივრცობრივი ან დროითი. სივრცობრივი ინფორმაცია ასახავს ან აღწერს ერთი მომენტისთვის მოვლენ-ებისათვის ან შედეგებისათვის ახასიათებს, მაგალითად, რომელიც ასახა-ვს ერთი და იმავე პროდუქციის ფასს სხვადასხვა ბაზარზე.

დროითი სტატისტიკური ინფორმაცია ასახავს ან აღწერს დროის სხვადასხვა მომენტისათვის ან შედეგისათვის ახასიათებს. ასეთ ინფორმაციაში არსებითია არა მარტო დაკვირვების მნიშვნელობა, არამედ ამ მნიშვნელობათა თანმიმდევრობაც.

დაკვირვების მონაცემები ზოგჯერ პირდაპირ გამოიყენება, ზოგჯერ კი

საჭიროა მისი შინაარსობრივი დამუშავება, ანუ სისტემატიზაცია.

ეკონომეტრიკული მოდელის სპეციფიკაცია - ეკონომეტრიკული მოდელირების ეს ეტაპი გულისხმობს შესაბამისი დამოკიდებულების შესახებ რაიმე ჰიპოთეზის ზოგადი მათემატიკური სახით ჩამოყალიბებას. სხვანაირად, მოცემულ ეტაპზე ეკონომეტრიკული მოდელის სახის განსაზღვრა და საწყისი წინა-პირობების და შეზღუდვების ფორმირება ხდება, რისთვისაც ვიზუალუ-რი (გრაფიკული) ლოგიკური (ანალიტიკური) და ექსპერიმენტული (საცდელ-სატესტო) მეთოდები გამოიყენება. ვიზუალური მეთოდი გრაფიკების სახით შედეგის გრაფიკულ გამოსახვაზე დაყრდნობას გულისხმობს, ლოგიკური მეთოდი პროცესების შესახებ ეკონომიკური თეორიის ცოდნას ეყრდნობა, რთული ფუნქციის სახის დამოკიდებულების მოდე-ლირების კი საცდელ-სატესტო მეთოდს მიმართავენ, რომელიც ავტომატურ რეჟიმში კომპიუტერის გამოყენებას გულისხმობს.

ეკონომეტრიკული მოდელის იდენტიფიკაცია - ამ ეტაპზე განისაზღვრება ეკონომეტრიკული მოდელის პარამეტრების რიცხვითი მნიშვნელობები და რიგი მახასიათებლები, რომლებიც რეგრესიული ანალიზის სიზუსტეს ახასიათებს.

მოდელის ვერიფიკაცია - რადგან აგებულ მოდელს გარკვეული დასკ-ვნების გასაკეთებლად გამოიყენებენ, აუცილებელია, შევამოწმოთ მისი ადეკვატურობა. ვერიფიკაციის პროცესში მოდელში ჩართული ფაქტორების სისწორის, ცდომილებების შერჩევის ხარისხის და მოდელის ცალკეული პარამეტრების მნიშვნელოვნების შეფასება ხდება.

4. დისკრეტული შემთხვევითი სიდიდე და მისი რიცხვითი მახასიათებლები

ეკონომიკაში მიმდინარე პროცესების უმრავლესობა თავისი არსით შემთხვევითი ხასიათისაა, რადგანაც ამ პროცესებზე ზემოქმედებს სხვადასხვა ფაქტორთა დიდი რაოდენობა, რომელთა ნაწილი შეუმჩნეველია, ნაწილი - არარე-გულირებადი, ნაწილი კი კონტროლს არ ექვემდებარება. ისეთ სიდიდეს, რომლის მნიშვნელობის ზუსტი განჭვრეტა შეუძლებელია, შემთხვევითი სიდიდე ეწოდება. უფრო ზუსტად, შემთხვევითი სიდიდე X განიმარტება ფუნქციის

სახით, რომელიც მოცემულია ელემენტარულ ხდომილებათა Ω სიმრავლეზე

$$X = f(\omega)$$

სადაც ω არის Ω სიმრავლეში შემავალი ელემენტარული ხდომილება, ე.ი. $\omega \in \Omega$. თუკი X-ის შესაძლო მნიშვნელობები სასრულ ან თვლად სიმრავლეს წარმოქმნის, მაშინ გვაქვს დისკრეტული შემთხვევითი სიდიდე. დისკრეტული შემთხვევითი სიდიდის მთავარ მახასიათებელს მისი განაწილების კანონი წარმოადგენს. ეს უკანასკნელი ზოგად წარმოდგენას იძლევა იმ შესაძლო $X_1, X_2, \ldots, X_k$ მნიშვნელობების შესახებ, რომელთაც მიიღებს გარკვეული ალბათობით ჩვენი შემთხვევითი სიდიდე. დისკრეტული შემთხვევითი სიდიდისათვის განაწილების კანონი ცხრილის ფორმით ან გრაფიკის სახით მოიცემა

თუ მოცემულია დისკრეტული შემთხვევითი სიდიდის განაწილების კანონი, მაშინ ადვილად გამოითვლება მისი რიცხვითი მახასიათებლები.

მათემატიკური ლოდინი $E(X)$ გამოხატავს შემთხვევითი სიდიდის საშუალო მოსალოდნელ მნიშვნელობას და გამოითვლება, როგორც X-ის საშუალო შეწონილი სიდიდე:

$$E(X) = \sum_{i=1}^{k} X_i P_i$$

$E(X)$-ის თვისებები:

1. $E(C) = C$, სადაც C მუდმივი სიდიდეა;
2. $E(CX) = CE(X)$
3. $E(X \pm C) = E(X) \pm C$
4. $E(X \pm Y) = E(X) \pm E(Y)$
5. $E(X - \mu) = 0$, როცა $\mu = E(X)$
6. $E(XY) = E(X) \cdot E(Y)$ თუ X და Y დამოუკიდებელი შემთხვევითი სიდიდეებია.

დისპერსია $D(X)$ ახასიათებს X შემთხვევითი სიდიდის გაფანტულობას მისი საშუალო მნიშვნელობის მიმართ. გამოითვლება ფორმულით:

$$D(X) = E(X - E(X))^2 = E(X - \mu)^2 = \sum_{i=1}^{k} (X_i - \mu)^2 p_i$$

დისპერსიის თვისებები:

1. $D(C) = 0$, სადაც C არის მუდმივა
2. $D(CX) = C^2 D(X)$

3. $D(X) = E(X^2) - \mu^2$

4. $D(X \pm Y) = D(X) + D(Y)$, თუ X და Y დამოუკიდებელი შემთხვევითი სიდიდეებია.

საშუალოდან გადახრის ხარისხის შესაფასებლად გამოიყენება ასევე საშუალო კვადრატული გადახრა $\sigma = \sqrt{D(X)}$ და ვარიაციის კოეფიციენტი, რომელიც პროცენტული მაჩვენებელია და გამოითვლება ფორმულით

$$V_x = \frac{\sigma}{|E(X)|} \cdot 100\%$$

5. უწყვეტი შემთხვევითი სიდიდე და მისი რიცხვითი მახასიათებლები

თუ X შემთხვევითი სიდიდის შესაძლო მნიშვნელობების სიმრავლე უსასრულოა და არათვლადია, მაშინ X უწყვეტია. უწყვეტი შემთხვევითი სიდიდის დასახასიათებლად განაწილების ფუნქცია და განაწილების სიმკვრივე გამოიყენება.

ვთქვათ, x ნამდვილი რიცხვია და X შემთხვევითი სიდიდეა. მაშინ ნამდვილი რიცხვის $F(x)$ ფუნქციას, რომელიც შემდეგნაირად განისაზღვრება

$$F(x) = P(X \le x)$$

განაწილების ფუნქცია ეწოდება. მაშასადამე, $F(x)$ გამოსახავს იმის ალბათობას, რომ შემთხვევითი სიდიდე X მიიღებს მნიშვნელობას, რომელიც არ აღემატება x-ს.

განაწილების ფუნქციის თვისებები:

1. $0 \le F(x) \le 1$ ნებისმიერი $x \in R$, სადაც R ნამდვილ რიცხვთა სიმრავლეა.
2. $F(x)$ არაკლებადია რიცხვით ღერძზე. მაშასადამე, თუ $x_2 > x_1$, მაშინ $F(x_2) \ge F(x_1)$.
3. $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$, $F(+\infty) = \lim_{x \to +\infty} F(x) = 1$.
4. თუ a და b სასრული რიცხვებია, მაშინ $P(a \le X \le b) = F(b) - F(a)$
5. $P(X \ge x) = 1 - F(x)$
6. თუ X შემთხვევითი სიდიდის შესაძლო მნიშვნელობები მიეკუთვნება $[a, b]$ შუალედს, მაშინ

$$F(x)=\begin{cases}0, & \text{როცა } x\le a\\ 1, & \text{როცა } x>b.\end{cases}$$

დავუშვათ, არსებობს $F(x)$ ფუნქციის წარმოებული

$$f(x)=F'(x)$$

მაშინ $f(x)$ ფუნქციას X შემთხვევითი სიდიდის განაწილების სიმკვრივე ან ალბათობის სიმკვრივე ეწოდება. შესაბამისად ალბათობის სიმკვრივე უსასრულოდ მცირე Δx სიგრძის ინტერვალის ერთეულზე შემთხვევითი სიდიდის მოხვედრის ალბათობას გამოსახავს. მას შემდეგი თვისებები გააჩნია

1. $f(x)\ge 0$ ე.ი ალბათობის სიმკვრივე არაუარყოფითი ფუნქციაა.
2. $P(a\le X\le b)=\int_a^b f(x)\cdot dx.$
3. $P(a\le X\le b)=P(a<X<b)=P(a\le X<b)=P(a<X\le b).$
4. $F(x)=P(X\le x)=\int_{-\infty}^{x} f(t)\,dt$
5. $\int_{-\infty}^{+\infty} f(x)\cdot dx=1.$

$$E(X)=\int_{-\infty}^{+\infty} x\cdot f(x)\cdot dx$$

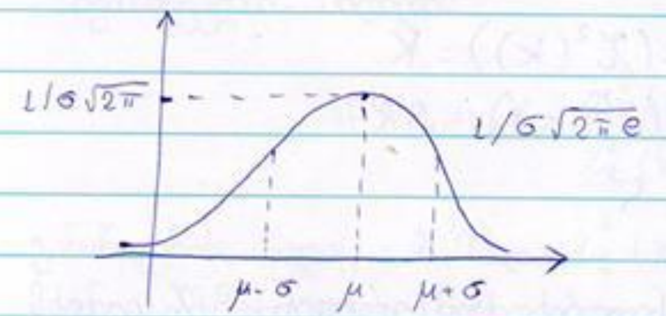
$$D(X)=\int_{-\infty}^{+\infty}(x-\mu)^2 f(x)\cdot dx.$$

6. ნორმალური და ხი-კვადრატ განაწილება.

თუ შემთხვევითი სიდიდის განაწილების სიმკვრივეს შემდეგი სახე აქვს

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

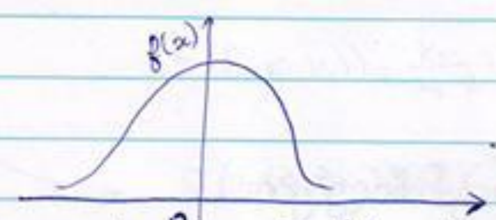
სადაც μ და σ^2 ნამდვილი რიცხვებია, მაშინ ვამბობთ, რომ X განაწილებულია ნორმალურად. განაწილების ამ ფორმის აღსანიშნავად გამოიყენება ჩანაწერი $X\sim N(\mu,\sigma^2)$



$\mu = E(X), \quad \sigma^2 = D(X)$

სხვა თანაბარ პირობებში ფუნქციაში μ პარამეტრის ცვლილება ნორმალური მრუდის Ox ღერძის გასწვრივ გადაადგილებას, ხოლო σ^2 პარამეტრის ცვლილება კი მრუდის ფორმის შეცვლას იწვევს.

ნორმალური განაწილებას $\mu = 0$ და $\sigma^2 = 1$ პარამეტრებით სტანდარტული, ანუ ნორმირებული ჰქვია. ეს შემთხვევა შემდეგნაირად ჩაიწერება $X \sim N(0,1)$



თუ X შემთხვევითი სიდიდე ნორმალურადაა განაწილებული, მაშინ მოცემულ $[x_1, x_2]$ შუალედში X-ის მოხვედრის ალბათობა გამოითვლება ფორმულით

$$P(x_1 \le X \le x_2) = \Phi(z_2) - \Phi(z_1)$$

სადაც z_1 და z_2 ნორმალიზებული სიდიდეებია

$$z_1 = \frac{x_1 - \mu}{\sigma} \quad \text{და} \quad z_2 = \frac{x_2 - \mu}{\sigma}$$

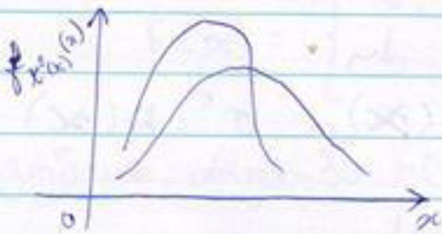
ხოლო $\Phi(z)$ ლაპლასის ფუნქციაა $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{x^2}{2}} dx$.

ნორმალურთან დაკავშირებულ განაწილებას χ^2 განაწილება წარმოადგენს.

ვთქვათ, $X_1, X_2, \dots, X_k$ სტანდარტული ნორმალური განაწილების მქონე დამოუკიდებელი შემთხვევითი სიდიდეებია. მაშინ შემდეგი სახით განსაზღვრული

$$\chi^2(k) = X_1^2 + X_2^2 + \dots + X_k^2$$

შემთხვევითი სიდიდის განაწილებას ეწოდება χ^2 განაწილება თავისუფლების k ხარისხით.



$$E(\chi^2(k)) = K$$
$$D(\chi^2(k)) = 2K$$

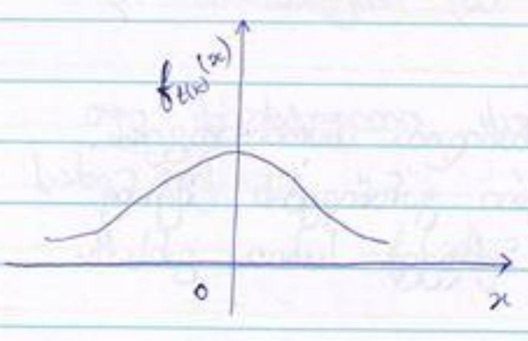
χ^2 განაწილების მქონე შემთხვევითი სიდიდის სიმკვრივის გრაფიკი დეკარტის საკოორდინატო სიბრტყის I მეოთხედში მოთავსებული. თუ $K \le 2$, $f_{\chi^2(k)}(x)$ კლებადი ფუნქციაა, თუ $k > 2$ $f_{\chi^2(k)}(x)$-ს გააჩნია მაქსიმუმი $x = k-2$ წერტილში. აქედან გამომდინარე $f_{\chi^2(k)}(x)$-ს გრაფიკი ასიმეტრიული და მარჯვნივ წაგრძელებულია, მაგრამ k-ს მნიშვნელობის ზრდისას ასიმპ-ტოტურად ნორმალურ სახეს უახლოვდება.

7. სტიუდენტის და ფიშერის განაწილებები

სტიუდენტის განაწილება თავისუფლების K ხარისხით გააჩნია t შემთხვევით სიდიდეს, რომელიც შემდეგი სახით განისაზღვრება

$$t = \frac{X}{\sqrt{\chi^2(k)/K}}.$$

სადაც X სტანდარტული ნორმალური კანონით განაწილებული შემთხვევითი სიდიდეა, $\chi^2(k)$ კი X-საგან დამოუკიდებელი შემთხვევითი სიდიდეა, რომელსაც გააჩნია χ^2 განაწილება თავისუფლების K ხარისხით.



$$E(t(k)) = 0, \quad D(t(k)) = \begin{cases} \frac{k}{k-2}, & \text{თუ } k > 2 \\ \infty, & \text{თუ } 0 < k \le 2. \end{cases}$$

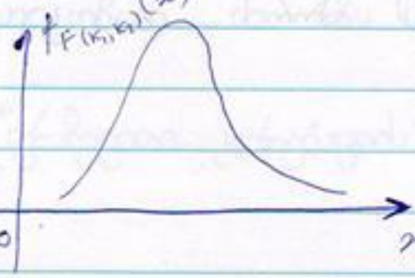
თუ $k \to +\infty$, განაწილება ნორმალურს უახლოვდება. პრაქტიკაში, თუ $k \ge 30$, სტიუდენტის განაწილება შეიძლება ნორმალური განაწილებით შევცვალოთ.

შემთხვევითი სიდიდის

$$F(k_1, k_2) = \frac{\chi_1^2(k_1)/k_1}{\chi_2^2(k_2)/k_2}$$

განაწილებას, სადაც $\chi_1^2(k_1)$ და $\chi_2^2(k_2)$ დამოუკიდებელი χ^2 განაწილების მქონე შემთხვევითი სიდიდეებია თავისუფლების k_1 და k_2 ხარისხებით, ეწოდება ფიშერის F განაწილება თავისუფლების k_1 და k_2 ხარისხებით. სამართლიანია ტოლობა

$$F(k_1, k_2) = \frac{\chi_1^2(k_1)/k_1}{\chi_2^2(k_2)/k_2} = \left(\frac{\chi_2^2(k_2)/k_2}{\chi_2^{1}(k_1)/k_1}\right)^{-1} = \frac{1}{F(k_2, k_1)}.$$



$$E(F(k_1,k_2)) = \frac{k_2}{k_2-2}, \text{ თუკი } k_2 > 2$$

$$D(F(k_1,k_2)) = \frac{2k_2^2(k_1+k_2-2)}{k_1(k_2-2)^2(k_2-4)}, \text{ თუკი } k_2 > 4.$$

ეკონომეტრიკაში F განაწილება განსაკუთრებით ფართოდ სტატისტიკური ჰიპოთეზების შემოწმებისას გამოიყენება.

8. კოვარიაცია და კორელაცია.

ხშირ შემთხვევაში ეკონომიკაში საქმე გვაქვს სიტუაციასთან, როცა ობიექტის, მოვლენის ან პროცესის მახასიათებელი შემთხვევითი ვექტორის შემადგენელი ელემენტები ურთიერთს განსაზღვრავს, ანუ ურთიერთზე დამოკიდებულებას წარმოადგენს.

ორ შემთხვევით სიდიდეს შორის სტატისტიკური კავშირის არსებობის ერთ-ერთი მნიშვნელოვანი მახასიათებელია კოვარიაცია, X და Y შემთხვევითი სიდიდეების კოვარიაცია ეწოდება ამ სიდიდეთა მათემატიკური ლოდინებიდან გადახრების ნამრავლის მათემატიკურ ლოდინს.

$$\sigma_{x,y} = \mathrm{cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

კოვარიაციის შემდეგი თვისებები გააჩნია:

1. $Cov(X,Y)=0$, თუ X და Y დამოუკიდებელია;
2. $cov(X,Y)=Cov(Y,X)$ - კოვარიაცია სიმეტრიულია;
3. $Cov(X,C)=0$, თუ C რაიმე მუდმივია;
4. $cov(X,Y)=Cov(X,U)+cov(X,V)$, თუ $Y=U+V$;
5. $cov(aX+b,Y)=a\cdot cov(X,Y)$, თუ a და b მუდმივებია.

კოვარიაციას, განსხვავებით დისპერსიისგან, შეუძლია მიიღოს ნებისმიერი მნიშვნელობა $(-\infty,+\infty)$ ინტერვალიდან. თუ კოვარიაცია დადებითია, შემთხვევითი სიდიდეები ერთი და იმავე მიმართულებით იცვლება, თუ უარყოფითია - საწინააღმდეგო მიმართულებით.

კოვარიაცია განსხვავებულ შემთხვევით სიდიდეებს შორის დამოკიდებულების ძალის განსაზღვრის საშუალებას არ იძლევა, ამ ნაკლისგან თავისუფალია კორელაცია, რომელიც, ასევე, არის შემთხვევით სიდიდეებს შორის დამოკიდებულების მახასიათებელი.

$$cor(X,Y)=\rho_{x,y}=\frac{cov(X,Y)}{\sqrt{\sigma_x^2\,\sigma_y^2}}.$$

სადაც $\sigma_x^2=D(X)$ და $\sigma_y^2=D(Y)$

თვისებები:

1. $-1\le\rho_{x,y}\le 1$;
2. $\rho_{x,y}=0$, თუ X და Y დამოუკიდებელია;
3. როცა $|\rho_{x,y}|=1$, მაშინ X და Y შორის არსებობს წრფივი ფუნქციონალური დამოკიდებულება. ამასთან, როცა $\rho_{x,y}=-1$ ეს დამოკიდებულება უარყოფითია, ხოლო როცა $\rho_{x,y}=1$ - დადებითი;
4. $\rho_{x,y}=\rho_{y,x}$, კორელაციის კოეფიციენტი სიმეტრიულია.

9. შემთხვევითი შერჩევა და მისი რიცხვითი მახასიათებლები.

განასხვავებენ გენერალური და შერჩევითი ერთობლიობის ცნებას. გენერალური ერთობლიობა შემთხვევითი სიდიდის რეალიზაციის ყველა შესაძლო მნიშვნელობას მოიცავს, მაგრამ, როგორც წესი, მთლიანი გენერალური ერთობლიობის გამოკვლევა შეუძლებელია, ან ხელმისაწვდომია მხოლოდ მისი ნაწილი, რომელსაც შერჩევით ერთობლიობას უწოდებენ.

ეწოდება, თუ $X_1, X_2, \dots, X_n$ მოცემულია, როგორც $F(x)$ განაწილების მქონე რაიმე X შემთხვევითი სიდიდის დამოუკიდებელი რეალიზაციები.

შემთხვევითი შერჩევიდან მიღებულ მახასიათებლებს შერჩევითი სტატისტიკები ან შეფასებები ეწოდება. ეს მახასიათებლებია:

შერჩევითი საშუალო $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} X_i$

შერჩევითი დისპერსია $Var(X) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})^2$

შერჩევითი კოვარიაცია $Cov(X,Y) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$, სადაც $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ და $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$. კოვარიაციის ფორმულა შეგვიძლია, ჩავწეროთ შემდეგნაირად:

$$cov(X,Y) = \frac{1}{n} \sum_{i=1}^{n} X_i Y_i - \bar{X} \cdot \bar{Y}.$$

შერჩევითი კორელაციის კოეფიციენტი $r_{x,y} = \dfrac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}}$

10. შეფასებისა და შეფასებების ცნება

ზოგად შემთხვევაში შეფასება (სტატისტიკა) ეწოდება $X = (X_1, X_2, \dots, X_n)$ შერჩევაზე რეალიზებულ ნებისმიერ $T(X) = T(X_1, X_2, \dots, X_n)$ ფუნქციას, რომელიც მნიშვნელობებს მოცემულ Θ პარამეტრულ სივრცეზე იღებს. შეფასება შეიძლება იყოს წერტილოვანი და ინტერვალური.

წერტილოვანი შეფასების ამოცანაა მოიძებნოს θ-ს ისეთი სტატისტიკა $\hat{\theta}_n = T(X_1, X_2, \dots, X_n)$, რომლის მნიშვნელობა შეიძლება ჩაითვალოს θ პარამეტრის ჭეშმარიტი მნიშვნელობის მიახლოებით რიცხვად და მას ნაკლები აქვს გაბნეულობა. ინტერვალური შეფასების დროს $X = (X_1, X_2, \dots, X_n)$ შერჩევის საფუძველზე განისაზღვრება ისეთი რიცხვითი ინტერვალი, რომელიც შეიძლება შეიცავდეს θ პარამეტრს.

11. შეფასების გადაუადგილებლობა

გადაუადგილებელია შეფასება, თუ მისი მათემატიკური ლოდინი ემთხვევა შესაფასებელ პარამეტრს $E(\hat{\theta}_n)=\theta$. წინააღმდეგ შემთხვევაში შეფასება გადაადგილებულია.

გადაადგილების ზომას განსაზღვრავენ სიდიდით $b(\theta)=E(\hat{\theta}_n)-\theta$ როცა $b(\theta)>0$, მაშინ $\hat{\theta}_n$ შეფასება გადაადგილებულია ზემოთ, ხოლო როცა $b(\theta)<0$ - ქვემოთ.

$\bar{X}$ არის μ_x-ის გადაუადგილებელი შეფასება. მართლაც

$$E(\bar{X})=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)$$

ვინაიდან, განმარტების თანახმად X_i შემთხვევითი სიდიდეა, რომელსაც იგივე განაწილება აქვს, რაც X-ს, ამიტომ $E(X_i)=E(X)=\mu_x$. აქედან გამომდინარეობს ტოლობა $E(\bar{X})=\mu_x$. მაშასადამე, $\bar{X}$ ყოფილა საშუალოს გადაუადგილებელი შეფასება. მაგრამ უნდა ითქვას, რომ $\bar{X}$ μ_x-სათვის ეს ერთადერთ შეფასებას არ წარმოადგენს. ნებისმიერი შეწონილი საშუალო აგრეთვე აკმაყოფილებს გადაუადგილებლობის პირობას, რაც კი იმაზე მიუთითებს, რომ μ_x-ის გადაუადგილებელ შეფასებათა სიმრავლე უსასრულოა

$$Z=\sum_{i=1}^{n}\lambda_i X_i,\quad \lambda\geq 0\quad i=1,\dots,n\quad \sum_{i=1}^{n}\lambda_i=1$$

$$E(Z)=\sum_{i=1}^{n}\lambda_i E(X_i)=\sum_{i=1}^{n}\lambda_i\mu_x=\mu_x.$$

12. შეფასების ეფექტურობა (ოპტიმალურობა)

მოცემული შერჩევითი ერთობლიობით გამოთვლილ გადაუადგილებელ $\hat{\theta}_n$ შეფასებას θ პარამეტრის ეფექტური, ანუ ოპტიმალური შეფასება ეწოდება, თუ იგივე შერჩევითი ერთობლიობით გამოთვლილ ნებისმიერ სხვა გადაუადგილებელ შეფასებებს შორის მას მინიმალური დისპერსია $D(\hat{\theta}_n)$ შეესაბამება.

μ_x-ისათვის წრფივად განსაზღვრულ გადაუადგილებელ შეფასებათა შორის ეფექტურ შეფასებას წარმოადგენს საშუალო არითმეტიკული $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}X_i$.

13. შეფასების ძალმოსილობა

θ პარამეტრის $\hat{\theta}_n$ შეფასებას ძალმოსილი ეწოდება, თუ იგი $n \to \infty$ პირობებში ალბათობით თეორიული მნიშვნელობისაკენ მიისწრაფვის.

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \le \varepsilon) = 1$$

$$\operatorname{P}\lim_{n\to\infty} \hat{\theta}_n = \theta \quad \text{ან კიდევ} \quad \hat{\theta}_n \xrightarrow{P} \theta$$

თუ $E(\hat{\theta}_n) \to \theta$ და $V(\hat{\theta}_n) \to 0$, როცა $n \to \infty$, მაშინ $\hat{\theta}_n$ არის θ-ს ძალმოსილი შეფასება.

ნორმალურად განაწილებული შემთხვევითი შერჩევისას $Var(X)$ არის σ^2-ის ძალმოსილი შეფასება. ზოგად შემთხვევაში თუ შეფასება გადაუნაცვლებელია, მაშინ ის ძალმოსილია

14. ნდობის ინტერვალი. ნდობის ინტერვალის აგების ზოგადი სქემა.

ინტერვალური შეფასების დროს წერტილოვანი შეფასების, ანუ ერთი კონკრეტული $\hat{\theta}_n$ რიცხვის დადგენის ამოცანა იცვლება ისეთი $(\hat{\theta}_n^{(1)}, \hat{\theta}_n^{(2)})$ ინტერვალის დადგენის ამოცანით, რომელშიც რაიმე წინასწარ განსაზღვრული γ ალბათობით შეიძლება უცნობი ჭეშმარიტი θ პარამეტრის მნიშვნელობა იმყოფებოდეს. ასეთ ინტერვალს ნდობის ინტერვალი ეწოდება, ხოლო γ ალბათობას კი - ნდობის ალბათობა, ანუ შეფასების საიმედოობა. $\hat{\theta}_n^{(1)} = T_1(X_1, \dots, X_n)$ და $\hat{\theta}_n^{(2)} = T_2(X_1, \dots, X_n)$ ნდობის ზედა და ქვედა საზღვრებია, რომლებიც შემთხვევითი სიდიდეებია, შემთხვევითია ნდობის ინტერვალიც.

ნდობის ინტერვალის აგების ზოგადი სქემა:

1. ცნობილი განაწილების მქონე გენერალური ერთობლიობიდან აიღება n მოცულობის შერჩევა და მის საფუძველზე განისაზღვრება უცნობი θ პარამეტრის წერტილოვანი $\hat{\theta}_n$ შეფასება.
2. აიგება θ პარამეტრთან დაკავშირებული ისეთი შემთხვევითი სიდიდე

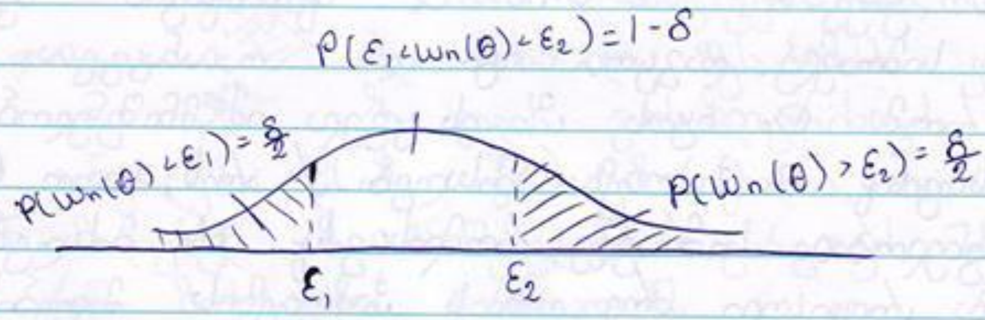
$W_n(\theta)$, რომლისთვისაც ცნობილია ალბათობის სიმკვრივის ფუნქცია $f(w)$.

3. შეირჩევა ნდობის ალბათობა γ (ან მნიშვნელოვნების დონე $\delta = 1-\gamma$). მისი არჩევა კონკრეტულ გარემოებებზეა დამოკიდებული.

4. შემთხვევითი $W_n(\theta)$ სიდიდის სიმკვრივის ფუნქციის გამოყენებით და მნიშვნელოვნების მოცემული $\delta = (1-\gamma)$ დონის გათვალისწინებით ვეძებთ ისეთ ε_1 და ε_2 რიცხვებს, რომელთათვისაც შესრულდება ტოლობა

$$P(\varepsilon_1 < W_n(\theta) < \varepsilon_2) = \int_{\varepsilon_1}^{\varepsilon_2} f(w)\cdot dw = \gamma = 1-\delta.$$

უმეტეს შემთხვევაში, ε_1 და ε_2 რიცხვების მოსაძებნად იხსნება განტოლებები შემდეგი წყვილები: $P(W_n(\theta) < \varepsilon_1) = \frac{\delta}{2}$ და $P(W_n(\theta) > \varepsilon_2) = \frac{\delta}{2}$.



5. მიღებული $\varepsilon_1 < W_n(\theta) < \varepsilon_2$ უტოლობა საბოლოოდ გარდაიქმნება ისეთ სახის $\hat{\theta}_n^{(1)} < \theta < \hat{\theta}_n^{(2)}$ უტოლობად, რომელიც γ ალბათობით ფარავს θ პარამეტრს და წარმოადგენს მის ინტერვალურ შეფასებას.

15. ჰიპოთეზის სტატისტიკური შემოწმება. ძირითადი ცნებები.

სტატისტიკური ჰიპოთეზის შემოწმების არსი იმაში მდგომარეობს, რომ დავადგინოთ თანხმდება თუ არა დაკვირვების შედეგები წამოყენებულ ჰიპოთეზას, შეიძლება თუ არა განსხვავება ჰიპოთეზასა და დაკვირვების

შედეგებს შორის შემთხვევითი ცდომილებების მიუხედავად, რომელიც გამოწვეულია შემთხვევითი შერჩევის მექანიზმით.

ჰიპოთეზას, რომლის შემოწმებაც გვინდა, ძირითადი ეწოდება და აღინიშნება H_0 ასოთი.

განასხვავებენ მარტივ და რთულ ჰიპოთეზებს. ჰიპოთეზა მარტივია, როდესაც აფიქსირებს უცნობი θ პარამეტრის რაიმე სიდიდესთან ტოლობას: $H_0: \theta=0$. რთულ ჰიპოთეზებში, როგორიცაა, მაგალითად $H_0: \theta>\theta_0$, მიიღება უამრავი მნიშვნელობა დაშვებული არის.

ყოველ ნულოვან ანუ ძირითად ჰიპოთეზას ალტერნატიული ჰიპოთეზა შეესაბამება და იგი H_1-ით აღინიშნება. ალტერნატიული ჰიპოთეზა შეიძლება იყოს მარტივი $H_1: \theta=\theta_1$, $\theta_0 \neq \theta_1$ და რთული $H_1: \theta \neq \theta_0$, $H_1: \theta>\theta_0$, $H_1: \theta<\theta_0$.

ჰიპოთეზის შესამოწმებლად ვირჩევთ სტატისტიკას, რომლის განაწილება ცნობილია. ვთქვათ, Θ არის ϑ_n სტატისტიკის შესაძლო მნიშვნელობათა სიმრავლე. დავყოთ იგი ორ არათანაკვეთ ქვესიმრავლედ ისე, რომ Θ_0 იყოს ϑ_n-ის ყველა იმ მნიშვნელობათა სიმრავლე, რომლისთვისაც H_0 ჰიპოთეზა მისაღები არ არის, ხოლო Θ_1 კი ϑ_n-ის მნიშვნელობათა სიმრავლე, რომლისთვისაც H_0 მისაღებია. Θ_0-ს ეწოდება კრიტიკული არე, Θ_1-ს ჰიპოთეზის მიღების არე. წერტილებს, რომლებიც კრიტიკულ არეს და ჰიპოთეზის მიღების არეს ერთმანეთისგან ყოფს კრიტიკული წერტილები ეწოდება.

δ-ს ეწოდება კრიტერიუმის მნიშვნელოვნების დონე და გვიჩვენებს ϑ_n-ის კრიტიკულ არეში მოხვედრის ალბათობას.

16. ნულოვანი ჰიპოთეზის $H_0: \theta=\theta_n$ შემოწმების სქემა რთულ ალტერნატივების პირობებში.

ა) როცა $H_0: \theta=\theta_0$ ჰიპოთეზა $H_1: \theta<\theta_0$ ალტერნატივის საწინააღმდეგოდ მოწმდება, უნდა შევარჩიოთ სტატისტიკა, რომლის განაწილება ცნობილია. ვთქვათ, ნულოვანი ჰიპოთეზის სამარ-

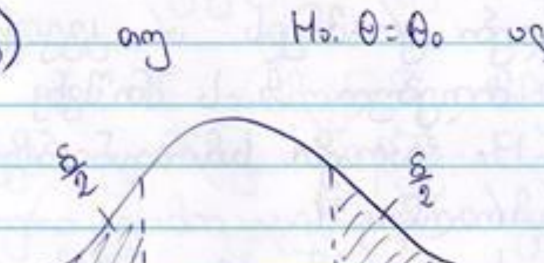
სამართლიანობის დროს θ_n-ის ფუნქციას აქვს სახე:

ასეთ შემთხვევაში, კრიტიკული არე და შესაბამისი კრიტიკული წერტილი შემდეგი პირობებით განისაზღვრება:
$P(\theta_n < \theta^{*}_{\text{კრ}}) = \delta$

თუ $X_1, X_2, \dots, X_n$ შერჩევიდან მიღებული θ_n მნიშვნელობა $\theta_n < \theta^{*}_{\text{კრ}}$ არეში აღმოჩნდება, ნულოვან ჰიპოთეზას უარვყოფთ და სამართლიანად ალტერნატიულ ჰიპოთეზას მივიჩნევთ. წინააღმდეგ შემთხვევაში ნულოვან ჰიპოთეზას ვღებულობთ.

ბ) ანალოგიური მსჯელობის $H_0: \theta = \theta_0$ ჰიპოთეზის შემოწმება $H_1: \theta > \theta_0$ ალტერნატიული პირობებში. განსხვავება ისაა, რომ კრიტიკული არის და შესაბამისი კრიტიკული წერტილის განსაზღვრისათვის გამოიყენება შემდეგი პირობა: $P(\theta_n > \theta^{**}_{\text{კრ}}) = \delta$. ცხადია, $X_1, X_2, \dots, X_n$ შერჩევიდან მიღებული θ_n მნიშვნელობის უარყოფის პირობას $\theta_n > \theta_{\text{კრ}}$ არეში მოხვედრა წარმოადგენს.

გ) თუ $H_0: \theta = \theta_0$ ალტერნატივაა $H_1: \theta \neq \theta_0$ მაშინ ორმხრივი კრიტიკული არე იქნება. თუკი შერჩევა θ_n სიმეტრიულია, რომელიც განაწილების კანონი ნულოვანი ჰიპოთეზის სამართლიანობის დროს θ_n სიმკვრივის ფუნქციის ნახაზზე მოცემული სახე ექნება. კრიტიკული წერტილები კი შემდეგი პირობებიდან განისაზღვრება:

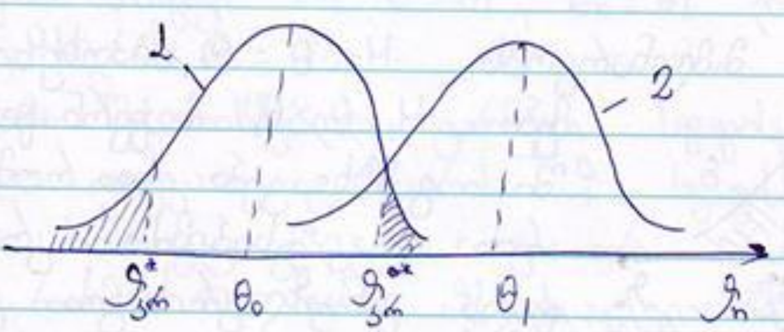


$$P(\theta_n < \theta^{*}_{\text{კრ}}) = \frac{\delta}{2} \qquad P(\theta_n > \theta^{**}_{\text{კრ}}) = \frac{\delta}{2}.$$

ნულოვანი ჰიპოთეზის უარყოფის პირობას $X_1, X_2, \dots, X_n$ შერჩევიდან მიღებული θ_n-ის მარცხენა და მარჯვენა დაშორებული კრიტიკული არე-ებში მოხვედრა წარმოადგენს.

17. მარტივ ჰიპოთეზათა $H_0: \theta = \theta_0$, $H_1: \theta = \theta_1$, $\theta_1 \neq \theta_0$ შემოწმების ლოგიკა.

დავუშვათ, რომ მოცემულ ჰიპოთეზებში θ აღნიშნავს ნორმალური განაწილების საშუალოს μ-ს, ხოლო θ_0 და θ_1 მისი წინასწარ მოცემული სავარაუდო მნიშვნელობებია. ამასთან, $\theta_1 > \theta_0$. წამოყენებული ჰიპოთეზათა წყვილის შემოწმებისათვის ავირჩიოთ სტატისტიკა $\vartheta_n = \bar{X}$. როგორც ვიცით, n-ის დიდი მნიშვნელობისათვის $\bar{X}$ ექვემდებარება ასიმპტოტურად ნორმალურ განაწილებას პარამეტრებით $E(\bar{X}) = \mu$ და $D(\bar{X}) = \sigma^2/n$.



ცხადია, H_0 ჰიპოთეზის სამართლიანობის შემთხვევაში უნდა შესრულდეს პირობა $E(\vartheta_n) = \theta_0$ ამ სიტუაციას (1) მრუდი შეესაბამება. როდესაც სამართლიანია ალტერნატიული ჰიპოთეზა, მაშინ $E(\vartheta_n) = \theta_1$ – (2) მრუდი. კერძო შემთხვევაში, როცა ეს მრუდები ერთმანეთს არ კვეთენ, ჰიპოთეზის შემოწმება ტრივიალურია: შერჩევიდან მიღებული ϑ_n-ს მნიშვნელობის მოხვედრა 1 მრუდის ქვეშ მიგვანიშნებს H_0 ჰიპოთეზის სამართლიანობაზე, ხოლო თუ ϑ_n მეორე წირის ქვეშ აღმოჩნდება სამართლიანია H_1.

თუ θ_0 და θ_1 მნიშვნელობები იმდენად ახლოსაა, რომ (1) და (2) მრუდები ერთმანეთს კვეთს, მნიშვნელოვანი ჰიპოთეზის შემოწმება ხდება მხოლოდ კრიტერიუმით. როცა $\theta_1 > \theta_0$, უნდა განვიხილოთ მარჯვენა კრიტიკული არე, როცა $\theta_1 < \theta_0$ – მარცხენა კრიტიკული არე.

18. პირველი და მეორე გვარის შეცდომები

როგორც ვიცით, ნულოვანი ჰიპოთეზის მიღება ან უარყოფა მნიშვნელოვნების დონეზეა დამოკიდებული. რაც მცირეა α-ს სიდიდე, მით ნაკლებალბათურია $\bar{x}_n$ სტატისტიკის კრიტიკულ არეში მოხვედრა, თუმცა მაინც შესაძლებელია. გამორიცხული არ არის, რომ სამართლიანი H_0 ჰიპოთეზის დროს $\bar{x}_n$-ის მნიშვნელობა კრიტიკულ არეში აღმოჩნდეს და H_0 უარვყოთ, ასეთ შემთხვევაში ვამბობთ, რომ ვუშვებთ პირველი გვარის შეცდომას, რომელიც სამართლიანი H_0 ჰიპოთეზის უარყოფას გულისხმობს, ცხადია, რაც მცირეა α, მით ნაკლებალბათურია პირველი გვარის შეცდომის დაშვება, მაგრამ α-ს შემცირებით მცირდება კრიტიკული არე და ნაკლებშესაძლებელი ხდება მასში არსებული $\bar{x}_n$-ის მოხვედრა, მაშინაც კი, როდესაც H_0 არასამართლიანია, ამიტომ α-ს შემცირებას არასწორი ჰიპოთეზის მიღების ალბათობის, ანუ მეორე გვარის შეცდომის დაშვების შესაძლებლობის ზრდა მოსდევს.

19. ფუნქციონალური და სტატისტიკური დამოკიდებულება.

ორ ან რამდენიმე მახასიათებელს შორის დამოკიდებულებას შეიძლება ჰქონდეს ფუნქციონალური ან სტატისტიკური სახე. ფუნქციონალური დამოკიდებულება გამოვლინდება ყოველ კონკრეტულ შემთხვევაში და ყოველ ცალკეულ დაკვირვებაში მკაცრად და განსაზღვრულად. ამ სახის დამოკიდებულების ცოდნა საშუალებას იძლევა ზუსტად დავადგინოთ შედეგი, როდესაც ცნობილია მისი განმსაზღვრელი ფაქტორული ცვლადის მნიშვნელობა. სხვანაირად, ფუნქციონალური დამოკიდებულება სრული დამოკიდებულებაა, რადგანაც შედეგი მთლიანად განისაზღვრება მოცემული ფაქტორული ცვლადით. მაგალითად, როდესაც რომელიმე პროდუქტზე ფასი p კანონმდებლობით არის დაფიქსირებული, სხვა თანაბარ პირობებში, შემოსავალი სრულიად ცალსახად

განისაზღვრება რეალიზებული პროდუქციის რაოდენობით.

ეკონომიკაში ფუნქციონალური დამოკიდებულებები იშვიათია, რადგანაც ეკონომიკური სიდიდეები, როგორც წესი, მრავალი ფაქტორის ზემოქმედებით ყალიბდება. მათი ნაწილი აღრიცხვადია, ნაწილი - არააღრიცხვადი. ზოგიერთი მათგანი შეიძლება ცალსახად განისაზღვროს, ზოგიერთი კი მხოლოდ მიახლოებით აღწერას ექვემდებარება ან საერთოდ რაოდენობრივად არ გამოისახება. ასეთი სახის დამოკიდებულებებს სტოქასტური ეწოდება. სხვანაირად რომ ვთქვათ, სტოქასტური დამოკიდებულების დროს ერთი ცვლადის მნიშვნელობას შეესაბამება მეორე ცვლადის გარკვეული განაწილება. სტოქასტური დამოკიდებულების მაგალითად შეიძლება გავიხსენოთ დამოკიდებულება გამოშვებულ პროდუქციის რაოდენობასა და დახარჯული რესურსის მოცულობას შორის, უმუშევრობასა და ხელფასის დონეებს შორის.

20. წყვილური წრფივი რეგრესიის მოდელი.

წყვილური რეგრესიის მოდელი ჩაიწერება შემდეგნაირად:

$$y = \alpha + \beta x + u.$$

მოცემულ ~~დამოკიდებულებაში~~ განტოლებაში დამოკიდებული ცვლადი y წარმოდგენილია ორი ნაწილით: სისტემატური ანუ არაშემთხვევითი შემადგენელი - $(\alpha + \beta x)$-ით და შემთხვევითი შემადგენელი - u-თი. ამ უკანასკნელის სახელწოდებად გამოიყენებენ ტერმინებს: შემთხვევითი წევრი, შემთხვევითი შეშფოთება, შემთხვევითი გადახრა.

რეგრესიის მოდელში u-ს არსებობას რამდენიმე მიზეზი განაპირობებს:

1. X-ის გარდა არსებობს მრავალი ფაქტორი, რომლებიც ზემოქმედებს y-ზე, მაგრამ მოდელში გათვალისწინებული არ არის.
2. დაზღვეული არ არის გაზომვის შეცდომებისგან, რომელიც შემდეგ u-ში აისახება.
3. არსებობს რომ ამ მოდელის განტოლების არასწორი სპეციფიკაცია,

ანუ განტოლების სახის შერჩევა.

შემთხვევითი u წევრი ერთდროულად ყველა ზემოთ აღნიშნული ფაქტორების ჯამურ გამოვლინებას წარმოადგენს.

რეგრესიული ანალიზის ერთ-ერთი ამოცანაა რეგრესიის განტოლების აგება, იდენტიფიკაცია, ე.ი. α და β პარამეტრების განსაზღვრას ნიშნავს. ამ მიზნით, სტატისტიკურად ვაგროვებთ მონაცემებს გენერალური ერთობლიობის შესახებ. მაგრამ პრაქტიკაში, როგორც წესი, ხელმისაწვდომია მხოლოდ X-ის და Y-ის გარკვეული შერჩევა. შერჩევითი მონაცემების საფუძველზე რეგრესიის განტოლების აგება განტოლების პარამეტრების შეფასებას გულისხმობს.

დავუშვათ, საკვლევი ოჯახების მთელი ერთობლიობის მაგივრად, მხოლოდ მათი ნაწილის შესწავლა მოხერხდა. რის შედეგადაც მივიღეთ შემთხვევითი დაკვირვების n წყვილი

$$(X_1, Y_1), (X_2, Y_2) \ldots (X_n, Y_n)$$

მოცემული წერტილებს გადავიტანთ სიბრტყეზე და გარკვეული წესით გავავლებთ წრფეს, რომელსაც შეესაბამება ფუნქცია

$$\hat{y} = a + bX.$$

ამ ფუნქციას რეგრესიის ემპირიული ფუნქცია ეწოდება. მასში შემავალი a და b პარამეტრები წარმოადგენს უცნობი α და β პარამეტრების სტატისტიკურ შეფასებებს.

21. უმცირეს კვადრატთა მეთოდი წყვილური რეგრესიისათვის.

დავუშვათ, გვაქვს x და y ცვლადები, რომელთა შორის წრფივი სტოქასტიკური დამოკიდებულება არსებობს. x დამოუკიდებელი, y-დამოკიდებული. მათ შორის დამოკიდებულების წრფივი სახე აქვს, რომელიც შემდეგი რეგრესიული მოდელით გამოისახება:

$$y = \alpha + \beta x + u.$$

ვთქვათ, ცნობილია x და y ცვლადების მნიშვნელობა n დაკვირვებისას: (x_1, y_1); $(x_2, y_2) \ldots (x_n, y_n)$. ამ წერტილებით აგებული შერჩევითი განტოლება იქნება შემდეგი

$$\hat{y} = a + bx + e$$

თუ x და y-ს შორის ხაზოვანი ... წრფივი დამოკიდე-ბულების პარამეტრები იქნება.

ისმის კითხვა, თუ რას მივიღებთ პარამეტრების სხვადასხვა კოეფიციენტებით მოცემული შეიძლება სხვადასხვა ვარიანტი მოვიღებთ. ჩვენი ამოცანა, რომელიც კოეფიციენტის შერჩევის ისეთ საშუალება, წრფის იყოს დაკვირვების და ... ხაზის, რომ შეძლებისდაგვარად ახლოს იყოს დაკვირვების წერტილებს. ამ მიზნით, გამოიყენებენ უმცირეს კვადრატთა მეთოდს.

უმცირეს კვადრატთა მეთოდის მოთხოვნა შემდეგნაირად ჩაიწე-რება

$$G = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \min.$$

თუ ჩავსვამთ $\hat{y}_i$-ს მნიშვნელობას მივიღებთ:

$$G = \sum_{i=1}^{n} (y_i - a - bx_i)^2 = \min.$$

განვსაზღვროთ a და b-ს ისეთი მნიშვნელობები, რომლისთვისაც G მინიმალური იქნება. G უწყვეტი, ამოზნექილი ფუნქცია და შემოსაზღვრული ქვემოდან $(G \geq 0)$ ამიტომ მას მინიმუმი გააჩნია. მინიმუმის აუცილებელი პირობაა $\frac{\partial G}{\partial a}$ და $\frac{\partial G}{\partial b}$ კერძო წარმოებულების ნულის ტოლობა.

$$\begin{cases} \frac{\partial G}{\partial a} = -2 \sum_{i=1}^{n} (y_i - a - bx_i) = 0 \\ \frac{\partial G}{\partial b} = -2 \sum_{i=1}^{n} (y_i - a - bx_i) x_i = 0 \end{cases}$$

გარდაქმნის შემდეგ სისტემა მიიღებს სახეს:

$$\begin{cases} \sum_{i=1}^{n} y_i = na + b \cdot \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i = a \cdot \sum_{i=1}^{n} x_i + b \cdot \sum_{i=1}^{n} x_i^2 \end{cases}$$

თუ განტოლებებს გავყოფთ n-ზე და გავამარტივებთ, მივიღებთ a და b პარამეტრების გამოსათვლელ ფორმულებს:

$$b = \frac{Cov(X,Y)}{Var(X)} \qquad a = \bar{y} - b\bar{x}$$

b-ს გამოსათვლელი ფორმულიდან, ნორმალურ განტოლებათა სისტემას ამონახსნი აქვს მხოლოდ მაშინ, თუ $\sum_{i=1}^{n}(X_i - \bar{X})^2 \neq 0$. რაც წარმოადგენს წყვილი რეგრესიის მოდელის იდენტიფიცირებადობის პირობას.

22. წყვილი რეგრესიის მოდელის და რეგრესიის კოეფიციენტების ეკონომიკური ინტერპრეტაცია.

რეგრესიული ანალიზი ნებისმიერ შემთხვევაში გულისხმობს მიღებული მოდელის და მისი კოეფიციენტების ეკონომიკური შინაარსის ახსნას და ინტერპრეტაციას. როგორც მოდელის, ასევე მისი კოეფიციენტების შინაარსი ძირითადად იმ კონკრეტული ამოცანით განისაზღვრება, რომლის გადაწყვეტისთვისაც მიღებული რეგრესიული მოდელი გამოიყენება. ასევე დიდ როლს ასრულებს ზოგადი შინაარსობრივი მხარე. პირველრიგში, ეს ეხება a და b კოეფიციენტების შინაარსს.

როცა ფაქტორული ცვლადის მნიშვნელობა $x=0$, მაშინ $\hat{y}=a$, მაშასადამე, ფორმალური თვალსაზრისით a გამოხატავს შედეგობრივი ცვლადის საშუალო მნიშვნელობას იმ შემთხვევისას, როცა $x=0$. მიუხედავად ამისა, რომ a-ს ასეთი ინტერპრეტაცია ყოველთვის ვერ არის ლოგიკური, ზოგჯერ მას ეკონომიკური შინაარსი გააჩნია.

ზოგად შემთხვევაში b კოეფიციენტი შედეგობრივი ცვლადის მოსალოდნელი მნიშვნელობის ნაზრდს გამოსახავს, რომელიც გამოწვეულია ფაქტორული x ცვლადის მნიშვნელობის ერთი ერთეულით ცვლილებით.

23. რეგრესიული ანალიზის ძირითადი დაშვებები

დაშვება I – მოდელის წრფივი ფორმა სპეციფიკაციას შეესაბამება $y_i=\alpha+\beta x_i+u_i$, $i=1,2,\dots,n$. სადაც u_i და y_i შემთხვევითი სიდიდეებია, ხოლო ახსნადი ცვლადი არა-შემთხვევითია. ეს უკანასკნელი (x_i) ყოველ დაკვირვებაში მუდ-

მუდმივი იქნება და მხოლოდ განსხვავებები გარეშე ფაქტორების, რომ-ლებიც რეგრესიის განტოლებაში არ გაითვალისწინება. ფორმალურად ეს დაშვება შემდეგნაირად ჩაიწერება:

$$cov(x_i, u_i) = E[(x_i - \bar{x})(u_i - E(u_i))] = 0$$

დაშვება II - შემთხვევითი u_i წევრის მათემატიკური ლოდინი ნულის ტოლია ყველა i-სათვის.

$$E(u_i) = 0 \quad \forall_i \in \{1, 2, \dots, n\}$$

მაშასადამე, ეს დაშვება ნიშნავს, რომ შეგვიძლია ვთქვათ y_i ცვლადის მთავარი განმსაზღვრელი x_i ცვლადი და მრავალი შემთხვევითობითი ცვლადი, რომელთა ერთი ნაწილი y_i სიდიდეზე დადებითად, ხოლო მეორე ნაწილი კი უარყოფითად გავლენას ისეთნაირად ახდენს, რომ შემთხვევითი u_i წევრის მინუსისკენ ან პლიუსისკენ გადახრას სისტემატური ხასიათი არ ექნება. რადგანაც მოცემული პირობა სრულდება, მაშინ

$$E(y_i) = E(\alpha + \beta x_i + u_i) = \alpha + \beta x_i$$

ეს კი ნიშნავს, რომ ფიქსირებული x_i-სთვის y_i-ს საშუალო მოსალოდნელი მნიშვნელობა მუდმივია და $(\alpha + \beta x_i)$-ს ტოლია.

დაშვება III - u_i წევრის დისპერსია $D(u_i)$ ერთი და იგივეა ყველა დაკვირვებისათვის.

$$D(u_i) = E(u_i - E(u_i))^2 = E(u_i^2) = \sigma_u^2 \quad \forall_i = \{1, 2, \dots, n\}$$

ეს დაშვება მათემატიკურად ნიშნავს, რომ თითოეული მეტი ან ნაკლები არ უნდა იყოს დამოუკიდებელი x_i ცვლადი, u_i-ს განაწილება ყველა შემთხვევაში ერთი და იგივე იქნება. შემთხვევითი წევრის დისპერსიის დაკვირვების ნომრისგან დამოუკიდებლობას ჰომოსკედასტიურობა ეწოდება, მაგრამ, თუ u_i წევრის განაწილება ცვალებადია, მაშინ ადგილი აქვს ჰეტეროსკედასტიურობას.

დაშვება IV - სხვადასხვა დაკვირვების შემთხვევითი წევრები u_i და u_j ერთმანეთისგან დამოუკიდებელია ე.ი.

$$cov(u_i, u_j) = E[(u_i - E(u_i))(u_j - E(u_j))] = E(u_i \cdot u_j) = 0 \quad i \neq j. \quad (1)$$

აგრეთვე $cov(y_i, y_j) = 0$

მოცემული დაშვების თანახმად, გათვალისწინებული ფაქტორები, რომლებიც ერთ რომელიმე დაკვირვებაში y-ს მნიშვნელობის ცდომილებას გამოიწვევს, ავტომატურად არ გამოიწვევს y-ს მნიშვნელობის ცდომილებას სხვა დაკვირ-

ვექტორზე. როდესაც (1) პირობა სრულდება, მინიმუმის წერტილი ერთადერთია, წინააღმდეგ შემთხვევაში გვაქვს მულტიკოლინეარობა.

დავუშვათ V-შემთხვევითი u_i ცდომილები განაწილებულია ნორმალურად. $u_i \sim N(0, \sigma_i^2)$ ასეთი დაშვება ერთი მხრივ იმ ფაქტს ეფუძნება, რომ რეგრესიის მოდელში შემთხვევითი ცდომილება მოიცავს ბევრ მცირე ფაქტორებს, რომელთა ეფექტების ჯამი გამოიხატება u-ში. მეორე მხრივ ცნობილია ცენტრალური თეორემა, რომელიც ამბობს, რომ თუ შემთხვევითი სიდიდე დიდი რაოდენობის სხვა შემთხვევითი სიდიდეების ზემოქმედების შედეგია, მაშინ მოცემული შემთხვევითი სიდიდეს შეიძლება ჰქონდეს მიახლოებით ნორმალური განაწილება, მაშინაც კი, როცა ცალკეული ზემოქმედებები არ ემორჩილებიან ნორმალურ განაწილებას.

როდესაც u_i წევრი ნორმალურად განაწილებულია, მაშინ ნორმალური განაწილება ჰქონია აგრეთვე y_i-ს. ასე, ჩვენი მოდელი დაშვების პირობებში $y_i \sim N(\alpha+\beta x_i, \sigma^2)$.

24. გაუს-მარკოვის თეორემა. a და b შეფასებების წრფივობა.

დავუშვათ, მოცემული გვაქვს წრფივი რეგრესიის მოდელი

$$y_i = \alpha+\beta x_i + u_i, \quad i=1,2,\dots,n.$$

და უმცირეს კვადრატთა მეთოდით შეფასებული მისი აპროქსიმაცია

$$\hat{y}_i = a + b x_i + e_i, \quad i=1,2,\dots,n.$$

შეფასებული a და b კოეფიციენტები შემთხვევით სიდიდეთა ჯგუფს მიეკუთვნება და მათი მნიშვნელობები დამოკიდებულია შემთხვევითი u_i წევრის თვისებებზე. ჩავთვალოთ, რომ u_i წევრისათვის გაუს-მარკოვის ოთხივე პირობა სრულდება. მაშინ სამართლიანია შემდეგი დებულება, რომელსაც გაუს მარკოვის თეორემა ეწოდება: a და b კოეფიციენტები წრფივად დამოკიდებულია y-ს დაკვირვებულ მნიშვნელობებზე, გადაუხრელი შეფასებებია და ყველა წრფივ გადაუხრელ შეფასებებს შორის ეფექტურია.

a და b კოეფიციენტების წრფივობა უშუალოდ გამომდინარეობს ფორმულებიდან, რომლებიც ნორმალური განტოლებათა სისტემიდან მიიღება:

$$b = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$ შემოვიღოთ აღნიშვნა:

$$w_i = \frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{x_i-\bar{x}}{n\cdot var(x)} \quad i=1,2,\dots,n$$

$$b = w_i \cdot \sum_{i=1}^{n}(y_i-\bar{y}) = \sum_{i=1}^{n} w_i y_i - \bar{y}\sum_{i=1}^{n} w_i,$$ მაგრამ

$$\sum_{i=1}^{n} w_i = \frac{\sum_{i=1}^{n}(x_i-\bar{x})}{n\cdot var(x)} = \frac{n\cdot\bar{x}-n\cdot\bar{x}}{n\cdot var(x)} = 0,$$ ამიტომ $b = \sum_{i=1}^{n} w_i y_i$

$$a = \bar{y} - b\bar{x} = \frac{1}{n}\sum_{i=1}^{n} y_i - \bar{x}\cdot\sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n}\left(\frac{1}{n}-\bar{x}w_i\right)y_i$$

25. აჩვენეთ a და b კოეფიციენტების გადაუადგილებლობა (დამტკიცება)

a და b გადაუადგილებელი შეფასებებია. ანუ დასამტკიცებელია უნდა ვაჩვენოთ, რომ $E(b)=\beta$ და $E(a)=\alpha$.

როგორც ვნახეთ, $\sum_{i=1}^{n} w_i(x_i-\bar{x}) = 1 = \sum_{i=1}^{n} w_i x_i$.

$$b = \sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n} w_i(\alpha+\beta x_i+u_i) = \alpha\sum_{i=1}^{n} w_i + \beta\sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} w_i u_i$$

$\sum_{i=1}^{n} w_i = 0$ და $\sum_{i=1}^{n} w_i x_i = 1$ ე.ი $b = \beta + \sum_{i=1}^{n} w_i u_i$

როგორც ვხედავთ, b-ს ორი შემადგენელი აქვს, არაშემთხვევითი β და შემთხვევითი $\sum_{i=1}^{n} w_i u_i$ წევრები. უცნობი β პარამეტრია და მისი გადახრის ხარისხი დამოკიდებულია u_i, $i=1,2,\dots,n$ შემთხვევითი სიდიდეების მნიშვნელობებზე, ამიტომ გავიხსენოთ გაუს-მარკოვის II დაშვება, u_i წევრის მათემატიკური ლოდინი ნულის ტოლი ყველა i-სთვის. გვექნება:

$$E(b) = \beta + \sum_{i=1}^{n} w_i E(u_i) = \beta$$

$$a = \sum_{i=1}^{n}\left(\frac{1}{n} - \bar{x} w_i\right) y_i = \sum_{i=1}^{n}\left(\frac{1}{n} - \bar{x} w_i\right)(\alpha + \beta x_i + u_i) =$$

$$= \alpha + \beta \bar{x} - \alpha x \sum_{i=1}^{n} w_i - \beta \bar{x} \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n}\left(\frac{1}{n} - \bar{x} w_i\right) u_i$$

$$a = \alpha + \sum_{i=1}^{n}\left(\frac{1}{n} - \bar{x} w_i\right) u_i$$

ის, რაც b-სთვის აღვნიშნეთ, სამართლიანი იქნება a-სთვისაც, მართლაც:

$$E(a) = \alpha + \sum_{i=1}^{n}\left(\frac{1}{n} - \bar{x} w_i\right) E(u_i) = \alpha .$$

26. რეგრესიის a და b კოეფიციენტების შეფასების ეფექტიანობა

$$a = \bar{y} - b\bar{x} \quad \text{და} \quad b = \frac{\text{cov}(x,y)}{\text{Var}(x)}$$

ახლა ვაჩვენოთ a-ს და b-ს გააჩნია მინიმალური დისპერსიები α-ს და β-ს წრფივ გადაუადგილებელ შეფასებებს შორის.

დასაწყისში გამოვთვალოთ a-ს და b-ს დისპერსიები

$$D(b) = D\left(\sum_{i=1}^{n} w_i y_i\right) = \sum_{i=1}^{n} w_i^2 D(y_i) + \sum_{i<j} \text{cov}(w_i y_i, w_j y_j)$$

გაუს მარკოვის IV დაშვების თანახმად $\text{cov}(w_i y_i; w_j y_j) = w_i w_j \text{cov}(y_i, y_j) = 0, i \neq j$ ხოლო III დაშვების თანახმად $D(y_i) = \sigma_u^2$ ამიტომ:

$$D(b) = \sigma_u^2 \sum_{i=1}^{n} w_i^2 = \sigma_u^2 \cdot \frac{1}{n \cdot \text{Var}(x)}$$

ანალოგიურად:

$$D(a) = D\left(\sum_{i=1}^{n}\left(\frac{1}{n} - \bar{x} w_i\right) y_i\right) = \sum_{i=1}^{n}\left(\frac{1}{n} - \bar{x} w_i\right)^2 \cdot D(y_i) =$$

$$= \sigma_u^2 \left(\frac{1}{n} - 2\bar{x} \sum_{i=1}^{n} w_i + \bar{x}^2 \sum_{i=1}^{n} w_i^2 \right)$$ რადგანაც $\sum_{i=1}^{n} w_i = 0$ და

$\sum_{i=1}^{n} w_i^2 = \frac{1}{n \cdot Var(x)}$ გვექნება:

$$D(a) = \frac{\sigma_u^2}{n} \left(1 + \frac{\bar{x}^2}{Var(x)} \right)$$

ვთქვათ, $\tilde{b}$ ($\tilde{b} \neq b$) არის β პარამეტრის რაიმე ჩაუნაცვლებელი წრფივი შეფასება

$$\tilde{b} = \sum_{i=1}^{n} c_i y_i \qquad E(\tilde{b}) = \beta \qquad c_i = w_i + d_i$$

გამოვთვალოთ $D(\tilde{b})$ დისპერსიის თვისების და გაუს-მარკოვის მე-3, მე-4 დაშვებების გამოყენებით მივიღებთ

$$D(\tilde{b}) = D\left(\sum_{i=1}^{n} c_i y_i\right) = \sum_{i=1}^{n} c_i^2 \cdot D(y_i) = \sigma_u^2 \sum_{i=1}^{n} c_i^2 = \sigma_u^2 \sum_{i=1}^{n} (w_i + d_i)^2 =$$

$$= \sigma_u^2 \left(\sum_{i=1}^{n} w_i^2 + 2\sum_{i=1}^{n} w_i d_i + \sum_{i=1}^{n} d_i^2 \right)$$

რადგანაც $\sum_{i=1}^{n} w_i d_i = 0$ საბოლოოდ გვექნება

$$D(\tilde{b}) = D(b) + \sigma_u^2 \sum_{i=1}^{n} d_i^2$$

რადგანაც $\sigma_u^2 \sum_{i=1}^{n} d_i^2 \geq 0$ ამიტომ $D(\tilde{b}) \geq D(b)$

ანალოგიურად ვაჩვენებთ, რომ $D(\tilde{a}) \geq D(a)$, მაშასადამე a და b ეფექტური შეფასებებია.

27. შემთხვევითი წევრის და რეგრესიის კოეფიციენტების დისპერსიების შეფასებები (წყვილური რეგრესიის შემთხვევა)

a და b შეფასებების თეორიული დისპერსიები:

$$D(b) = \sigma_b^2 = \sigma_u^2 \frac{1}{n \cdot Var(x)} \quad (1) \qquad D(a) = \sigma_a^2 = \frac{\sigma_u^2}{n}\left(1 + \frac{\bar{x}^2}{Var(x)}\right) \quad (2)$$

მიღებული დისპერსიები სხვა წრფივ გადაუადგილებელ შეფასებების დისპერსიებს შორის მინიმალურია, მაგრამ შეფასების ხარისხი პრაქტიკაში იმაზეა დამოკიდებული, თუ რის ტოლია კონკრეტულად σ_b^2-ს და σ_a^2-ს მნიშვნელობა. რაც მცირეა σ_b^2 და σ_a^2, მით მაღალია რეგრესიის კოეფიციენტის შეფასების სიზუსტე.

- a-ს და b-ს დისპერსიები პირდაპირპროპორციულია შემთხვევითი გადახრის დისპერსიის σ_u^2-ის. მაშასადამე, რეგრესიის მოდელში რაც დიდია შემთხვევითობის ფაქტორის როლი, მით ნაკლებად ზუსტი შეფასებები მიიღება.
- a-ს და b-ს დისპერსიები დაკვირვებების რიცხვის n-ის უკუპროპორციულია, ე.ი საწყისი მონაცემების მოცულობის გაზრდა რეგრესიის კოეფიციენტის შეფასებას სიზუსტის ამაღლების დადებით როლს ასრულებს.
- a-ს და b-ს დისპერსიები მით მცირეა, რაც დიდია დამოუკიდებელი x ცვლადის შერჩევითი დისპერსია $Var(x)$.

პრაქტიკაში a-ს და b-ს თეორიული დისპერსიების გამოთვლა შეუძლებელია, რადგანაც σ_u^2 თეორიული სიდიდეა და მისი გამოთვლა შეუძლებელია. ამიტომ, u_i გადახრების შერჩევითი მნიშვნელობების ნაცვლად უნდა გამოვიყენოთ e_i შეფასებები:

$$e_i = y_i - a - b x_i \qquad i = 1, 2, \ldots, n.$$

e_i-ს შერჩევითი დისპერსია $Var(e)$-თი აღვნიშნოთ $Var(e) = \frac{1}{n}\sum_{i=1}^{n} e_i^2$

თუმცა σ_u^2-ის გადაუადგილებელი შეფასება რომ $Var(e)$ არ გამოდგება. ამ შეფასების კორექტირებული $Var(e)$ წარმოადგენს:

$$S_e^2 = \frac{n}{n-2} Var(e) = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2$$

შევცვალოთ (1) და (2) ფორმულებში σ_u^2 და მივიღებთ

$$S_b^2 = \frac{S_e^2}{n \cdot Var(x)} = \frac{Var(e)}{(n-2)Var(x)}$$

$$S_a^2 = \frac{S_e^2}{n}\left(1 + \frac{\bar{x}^2}{Var(x)}\right) = \frac{Var(e)}{n-2}\left(1 + \frac{\bar{x}^2}{Var(x)}\right)$$

28. რეგრესიის a და b კოეფიციენტების სტანდარტული ცდომილებები. t_b და t_a სტატისტიკები

წყვილური რეგრესიის ნორმალური კლასიკური მოდელში:

1) a და b შეფასებები ნორმალურად განაწილებული შემთხვევითი სიდიდეებია

$$b \sim N(\beta, \sigma_b^2) \qquad a \sim N(\alpha, \sigma_a^2)$$

2) შემთხვევითი სიდიდე:

$$(n-2) S_e^2 / \sigma_u^2 = \sum_{i=1}^{n} \left(\frac{e_i}{\sigma_u}\right)^2$$

თავისუფლების (n-2) ხარისხის მქონე ხი-კვადრატულ განაწილებას შეესაბამება:

$$(n-2) \cdot S_e^2 / \sigma_u^2 \sim \chi^2(n-2)$$

3) a და b შეფასებები S_e^2 შეფასებისგან სტატისტიკურად დამოუკიდებელია.

შესაბამისად ნორმალური შემთხვევითი სიდიდეები

$$t_b = \frac{(b-\beta)/\sigma_b}{\sigma_e/\sigma_u} \qquad t_a = \frac{(a-\alpha)/\sigma_a}{\sigma_e/\sigma_u}$$

t_b და t_a სიდიდეების განაწილება თავისუფლების $(n-2)$ ხარისხის მქონე სტიუდენტის t განაწილების კანონით ხასიათდება.

სწორედ a და b კოეფიციენტების მნიშვნელობების შესაფასებლად გამოიყენება t_b და t_a სტატისტიკები, რომლებიც შემდეგნაირად შეგვიძლია ჩავწეროთ:

$$t_b = \frac{b-\beta}{S_b} \qquad t_b \sim t(n-2) \qquad S_b = S_e\sqrt{\frac{1}{n \cdot Var(x)}}$$

$$t_a = \frac{a-\alpha}{S_a} \qquad t_a \sim t(n-2) \qquad S_a = S_e\sqrt{\frac{1}{n}\left(1+\frac{\bar{x}^2}{Var(x)}\right)}$$

29. ცალკეული რეგრესიის კოეფიციენტის შესახებ $\beta=\beta_0$ და $\alpha=\alpha_0$ ჰიპოთეზის შემოწმება.

b კოეფიციენტის მნიშვნელობების შესამოწმებლად ჩამოყალიბდა ნულოვანი ჰიპოთეზა იმის შესახებ, რომ $\beta=\beta_0$, სადაც β_0 რაიმე ცნობილი რიცხვითი სიდიდეა. უმეტესად $\beta_0=0$. ალტერნატიულად შეიძლება განვიხილოთ $H_1: \beta>0$, $H_1: \beta<0$ და $H_1: \beta\neq0$ ჰიპოთეზები.

დავუშვათ, ვამოწმებთ ჰიპოთეზას:

$H_0: \beta=0$

$H_1: \beta\neq0$

ვიპოვოთ t_b სტატისტიკა $\quad t_b = \frac{b-\beta}{S_b} = \frac{b}{S_b} \quad t_b \sim t(n-2)$

თავისუფლების ხარისხისა და მნიშვნელოვნების დონის საფუძველზე სტიუდენტის განაწილების ცხრილში ვპოულობთ $t_{კრ}$-ს და ვადარებთ t_b-ს. თუ $|t_b|>t_{კრ}$, უარვყოფთ H_0 ჰიპოთეზას და მივიღებთ H_1-ს.

ანალოგიურად მოწმდება $\alpha=\alpha_0$ ჰიპოთეზა

$H_0: \alpha=\alpha_0=0 \qquad H_1: \alpha=\alpha_0\neq0$

$t_a = \frac{a-\alpha}{S_a} = \frac{a}{S_a} \sim t(n-2)$ თუ $|t_a|>|t_{კრ}|$ უარვყოფთ H_0 ჰიპოთეზას.

30. რეგრესიის კოეფიციენტების ნდობის ინტერვალები.

ნდობის ინტერვალის განსაზღვრის საფუძველი გახლავთ:

$$P\left(-t_{n-2,\,\delta/2} < \frac{b-\beta}{S_b} < t_{n-2,\,\delta/2}\right) = 1-\delta$$

$$P\left(-t_{n-2,\,\delta/2} < \frac{a-\alpha}{S_a} < t_{n-2,\,\delta/2}\right) = 1-\delta$$

სადაც $1-\delta$ ნდობის ალბათობაა, ხოლო $t_{n-2,\,\delta/2}$ t განაწილების მნიშვნელობა სტუდენტის ცხრილში თავისუფლების $n-2$ ხარისხისათვის და მნიშვნელოვნების $\delta/2$ დონისათვის.

ამ გამოსახულებიდან გამომდინარეობს, რომ α და β კოეფიციენტები $1-\delta$ ალბათობით შემდეგ ინტერვალში იქნებიან:

$$b - t_{კრ} \cdot S_b < \beta < b + t_{კრ} \cdot S_b \quad (1)$$
$$a - t_{კრ} \cdot S_a < \alpha < a + t_{კრ} \cdot S_a \quad (2)$$

β-სა და α-ს ნდობის ინტერვალის სიგრძეები δ-ით განისაზღვრება. რაც მცირეა δ, მით ფართოა ნდობის ინტერვალი. პრაქტიკულ გამოყენებაში δ მნიშვნელობად ძირითადად 5%-იანი ან 1%-იანი დონეები გამოიყენება. მაგალითად, 99%-იანი ნდობის ინტერვალი უფრო ფართოა, ვიდრე 95%-იანი. ეს ნიშნავს, რომ:

ა) თუ ნულოვანი ჰიპოთეზა მიიღება 5%-იანი დონით, ის მიიღება 1%-იანი დონით.

ბ) თუ H_0 უარყოფილი იქნება 1%-იანი დონით, ის უარყოფილი იქნება 5%-იანი დონით.

გ) H_0 შეიძლება უარვყოთ 5%-იანი დონით, მაგრამ მივიღოთ 1%-იანი დონით.

ინტერვალების განსაზღვრა შეიძლება შევამოწმოთ β-ს და α-ს ყოველი შესაძლებელი მნიშვნელობისთვის შეფასების გარშემო სტატისტიკური მნიშვნელობებით. თუ (1) შუალედში აღმოჩნდება ნული, მაშინ ჰიპოთეზა $H_0: \beta=0$ აღიარებული იქნება b-ს მიღებულ მნიშვნელობასთან და რეგრესიის განტოლებაში b კოეფიციენტი უმნიშვნელო იქნება. ანალოგიურად მოწმდება

a კოეფიციენტის მნიშვნელობას (2) ინტერვალის მეშვეობით.

31. შედეგობრივი ცვლადის ვარიაციის ანალიზი (დისპერსიული ანალიზი რეგრესიის მოდელში) (გამოყვანით)

რეგრესიის მოდელის ვერიფიკაციის პროცესი არ შემოიფარგლება მხოლოდ აგებული განტოლების კოეფიციენტების მნიშვნელობების შეფასებით. აუცილებელია მნიშვნელოვანია მთლიანი მოდელის შეფასება, ანუ იმის დადგენა, შეესაბამება თუ არა რეალობაში მიღებული მათემატიკური მოდელი დაკვირვების მონაცემებს და საკმარისია თუ არა მოდელში ჩართული ახსნადი ცვლადი შედეგობრივი ცვლადის აღწერისათვის. ასეთი შეფასება შეიძლება განხორციელდეს დისპერსიული ანალიზის საშუალებით, რომელშიც ცენტრალური ადგილი საშუალოდან გადახრის კვადრატების ჯამის გაშლის შემდეგ ფორმულას უკავია:

$$\underbrace{\sum_{i=1}^{n}(y_i-\bar{y})^2}_{Q}=\underbrace{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}_{Q_R}+\underbrace{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}_{Q_E}. \quad (1)$$

y_i - შედეგობრივი ცვლადის დაკვირვების მნიშვნელობა.
$\hat{y}_i$ - შედეგობრივი ცვლადის რეგრესიით შეფასებული მნიშვნელობა.
$\bar{y}$ - შედეგობრივი ცვლადის მნიშვნელობების საშუალო არითმეტიკული
Q - შედეგობრივი ცვლადის დაკვირვების მნიშვნელობების საშუალოს ირგვლივ მთლიანი გაბნეულობის მაჩვენებელი.
Q_R - რეგრესიის მოდელით ახსნილი გაბნეულობის ჯამი.
Q_E - რეგრესიის მოდელით აუხსნელი გაბნეულობის ჯამი.

დავამტკიცოთ (1)-ის სამართლიანობა. გვაქვს:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}[(y_i-\hat{y}_i)+(\hat{y}_i-\bar{y})]^2=$$

$$= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2.$$

შუა წევრი შესაძლოა შემდეგნაირად გარდავქმნათ:

$$2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 2\sum_{i=1}^{n} e_i(a + b x_i - \bar{y}) =$$

$$= 2a\sum_{i=1}^{n} e_i + 2b\sum_{i=1}^{n} e_i x_i - 2\bar{y}\sum_{i=1}^{n} e_i.$$

უმცირეს კვადრატთა მეთოდის ფუნქციის ექსტრემუმის აუცილებელი პირობების თანახმად $\sum_{i=1}^{n} e_i = 0$ და $\sum_{i=1}^{n} e_i x_i = 0$ მაშასადამე $2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$ და (1) ტოლობა მართებულია.

32. F სტატისტიკით რეგრესიული მოდელის მნიშვნელოვნების შემოწმება (წყვილური რეგრესიის მოდელში დისპერსიული ანალიზის შეფასება)

დეტერმინაციის R^2 კოეფიციენტი შედეგობრივი ცვლადის მთლიან ვარიაციაში რეგრესიით ახსნილი ვარიაციის წილს გამოხატავს.

$$R^2 = \frac{Q_R}{Q} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{Var(\hat{y})}{Var(y)}.$$

რეგრესიის მოდელის მნიშვნელოვნების შემოწმება დისპერსიული ანალიზით შემდეგი სქემის მიხედვით ხდება: ვაყალიბებთ ნულოვან და ალტერნატიულ ჰიპოთეზებს

$H_0: R^2 = 0$

$H_1: R^2 \neq 0$

ფიშერის განაწილების ცხრილში ვპოულობთ $F_{კრ}$ წერტილს და ვადარებთ F-ის მნიშვნელობას

$$F = \frac{(n-2)R^2}{1-R^2} \sim F(1, n-2)$$

თუ $F > F_{კრ}$, ნულოვანი ჰიპოთეზა უნდა უარვყოთ და მივიღოთ H_1. ამ შემთხვევაში R^2-ს იმდენად დიდი მნიშვნელობა აქვს, რომ იგი შეუძლებელია, გამოწვეული იყოს შემთხვევითობით.

34. კორელაციის კოეფიციენტი და მისი გამოყენება წყვილი რეგრესიის მოდელის სტატისტიკური მნიშვნელოვნობის შესამოწმებლად.

კორელაციის კოეფიციენტი არის ცვლადს შორის წრფივი სტატისტიკური დამოკიდებულების სიმჭიდროვის მაჩვენებელი, მისი საზღვრებია [-1, 1]. მინუსი ნიშნავს, რომ მათ შორის უკუ-პროპორციული დამოკიდებულებაა. [0,7; 0,9] საზღვრებში მოქცეული მაღალი ხარისხის სიმჭიდროვეს გამოხატავს. [0,2; 0,4] საზღვრებში კავშირი სუსტია.

კორელაციის კოეფიციენტსა და დეტერმინაციის კოეფიციენტს შორის არსებობს კავშირი

$$R^2 = (r_{xy})^2$$

ხოლო $t_b = t_r$ არის კავშირი კორელაციის კოეფიციენტსა და b კოეფიციენტის t სტატისტიკას შორის.

r_{yx} შეიძლება რეგრესიის განტოლების პარამეტრის შესამოწმებლად გამოვიყენოთ. ჩამოვაყალიბოთ ნულოვანი და მისი ალტერნატიული ჰიპოთეზა

$$H_0: r_{yx} = 0$$
$$H_1: r_{yx} \neq 0.$$

ვითვლით სტატისტიკას: $t_r = \frac{r_{yx}\sqrt{n-2}}{\sqrt{1-r_{yx}^2}} \sim t(n-2)$

და სტიუდენტის განაწილების ცხრილში α მნიშვნელოვნობის დონისა და $(r-2)$ თავისუფლების ხარისხით ვპოულობთ $t_{კრ}$-ს.

თუ $|t_r| > t_{კრ}$, ნულოვან ჰიპოთეზას უარვყოფთ და ვღებულობთ H_1-ს.

35. კავშირი კოეფიციენტებს შორის წყვილურ ანალიზში.

წყვილური რეგრესიის ანალიზის დროს დეტერმინაციის კოეფიციენტი X და Y-ს შორის კორელაციის კოეფიციენტის კვადრატს ემთხვევა.

$$R^2 = r_{yx}^2.$$

ასევე, თუ ერთმანეთს შევადარებთ R^2-სა და r_{yx}-ს მნიშვნელობების შესამოწმებლად გამოყენებულ F და t სტატისტიკებს, შევნიშნავთ, რომ F სტატისტიკა t_r სტატისტიკის კვადრატს ემთხვევა. ორივე ტესტი ყოველთვის ერთ და იმავე შედეგს მოგვცემს:

$$F = \frac{(n-2) \cdot R^2}{1 - R^2} \qquad t_r = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1 - r_{xy}^2}}$$

კორელაციის კოეფიციენტის სტატისტიკასა და b კოეფიციენტის სტატისტიკას შორის არსებობს კავშირი: $t_r = t_b$.

36. ეკონომიკური პროგნოზირება. პროგნოზის შეცდომა.

ჩვეულებრივ ტერმინი „პროგნოზირება" დროით მწკრივებში გამოიყენება და გულისხმობს რაიმე ცვლადის ან ცვლადების მომავალი

მოვლენების მახასიათებლის განჭვრეტას. რეგრესიული მოდელით პროგნოზირების მიზანი შემთხვევითი ცვლადის შედეგის იმ ახსნილი ცვლადის მნიშვნელობის საფუძველზე, რომელიც სხვის შერჩევაში არ შედის.

განასხვავებენ წერტილოვან და ინტერვალურ პროგნოზირებას. პირველ შემთხვევაში შემთხვევითი ცვლადის შედეგის კონკრეტული რიცხვი, მეორე შემთხვევაში ინტერვალი, რომელშიც მოცემული ნდობის ალბათობით შეიძლება იმყოფებოდეს შესაბამისი სიდიდე. არსებობს აგრეთვე უპირობო და პირობითი პროგნოზირება. როდესაც სხვის შერჩევის გარეთ მოცემული ახსნილი ცვლადის მნიშვნელობა ზუსტად ცნობილია, მაშინ მისი შესაბამისი დამოკიდებული ცვლადის შედეგის პროცესს უპირობო პროგნოზირება ეწოდება, პირობითი პროგნოზირების დროს კი ახსნილი ცვლადის მნიშვნელობა მიახლოებით არის ცნობილი.

თუ x^* შერჩევის გარეთაა მოცემული, მაშინ $y^* = a + bx^*$.

პროგნოზის შეცდომა ორი მიზეზით წარმოიქმნება. ერთი მიზეზი პროგნოზის გაკეთების რეგრესიის ფორმული კოეფიციენტების ნაცვლად a და b შეფასებების გამოყენებას უკავშირდება. ამიტომ, რაც უფრო ეს შეფასებები ფაქტიური მნიშვნელობიდან განსხვავდება, მით მეტია პროგნოზის შეცდომა. მეორე მიზეზი საპროგნოზო მნიშვნელობის განსაზღვრისას შემთხვევითი u^* წევრის გაუთვალისწინებლობაა. დაშვების თანახმად $E(u^*) = 0$.

37. პროგნოზის ნდობის ინტერვალი.

პროგნოზის შეცდომის დისპერსიის გამოთვლის შედეგი:

$$S(\hat{y}^* - y^*) = Se^2 \left[1 + \frac{1}{n} + \frac{(x^* - \bar{X})^2}{n \cdot Var(X)} \right)^2 \quad (1)$$

პროგნოზის ნდობის ინტერვალს აქვს სახე:

$$\hat{y}^* - t_{კრ} \cdot S(\hat{y}^* - y^*) < y^* < \hat{y}^* + t_{კრ} \cdot S(\hat{y}^* - y^*)$$

სადაც t კრიტიკული არის t ~~[illegible]~~ სტუდენტის კრიტიკული მნიშვნელობა, რომელსაც ყოველ კუდში შეესაბამება $\delta/2$ დონე და თავისუფლების $n-2$ ხარისხი.

პროგნოზის ნდობის ინტერვალის სიგრძე ცალსახად დამოკიდებულია სტანდარტულ შეცდომაზე $S(\hat{y}^* - y^*)$-ზე. (1) ფორმულის თანახმად, იგი მინიმალურ $S_e\sqrt{1+1/n}$ მნიშვნელობას ღებულობს $x^* = \bar{x}$ სიტუაციაში და იზრდება $(x^* - \bar{x})^2$-ის ზრდის კვალობაზე. მაშასადამე, რაც მეტად განსხვავდება x^*-ის მნიშვნელობა შერჩევის საშუალოსგან, მით დიდია პროგნოზის ნდობის ინტერვალი და მით მეტად განუსაზღვრელია პროგნოზი.

38. წრფივი რეგრესიის ორფაქტორიანი მოდელი.

ხშირად შედეგობრივ ცვლადზე ერთდროულად და შესაძლო სხვადასხვა მიმართულებით რამდენიმე მნიშვნელოვანი ფაქტორი ზემოქმედებს, რომელთა გათვალისწინებაც აუცილებელია.

დამოუკიდებელი ცვლადის შესწავლის პროცესში რამდენიმე ფაქტორის გათვალისწინებას მრავლობითი რეგრესიის მოდელი გვთავაზობს, რომელიც, თავის მხრივ, წყვილური რეგრესიის მოდელის განზოგადებას წარმოადგენს. ორივე ვარიანტი განსხვავდება პრინციპებით: შედეგობრივ ცვლადზე დამოუკიდებელ ცვლადთა გავლენის ერთობლივად გამოვლენის პრინციპებს და მოდელის სპეციფიკაციის პრინციპებს, ანუ იმის გარკვევას დამოუკიდებელი ცვლადებიდან რეგრესიის განტოლებაში რომელი ჩავრთოთ და რომელი არა.

წრფივი რეგრესიის ორფაქტორიან მოდელს, რომელშიც y შედეგობრივი ცვლადი ორი ფაქტორული ცვლადით განისაზღვრება, შემდეგი სახე აქვს: $y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_i \quad i = 1, 2, \ldots, n.$

სადაც i დაკვირვების ნომერია, β_0, β_1 და β_2 რეგრესიის უცნობი პარამეტრები, u_i შემთხვევითი წევრი. ასევე, როგორც წყვილური რეგრესიის მოდელის შემთხვევაში, შეიძლება ვთქვათ, რომ Y ორი, სისტემატური $(\beta_0+\beta_1X_1+\beta_2X_2)$ და u_i შემთხვევითი, ნაწილით არის წარმოდგენილი.

39. უმცირეს კვადრატთა მეთოდი ორფაქტორიანი მოდელისათვის (გამოყვანა)

$y=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+u_i$, $i=1,2,\ldots,n$ თეორიული მოდელის შესაბამისი შერჩევითი განტოლებაა $\hat{y}_i=b_0+b_1x_{i1}+b_2x_{i2}+e_i$ $i=1,2,\ldots,n$, სადაც e_i ნარჩენობითი წევრია u_i-ს შესაბამისი.

ასევე, როგორც წყვილურის შემთხვევაში, b_0, b_1, b_2 პარამეტრების შესაბამისი ძირითადად უმცირეს კვადრატთა მეთოდით გამოყე-ნება, რომელიც შემდეგ კრიტერიუმზეა დაფუძნებული:

$$G=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}e_i^2=\min.$$

ჩავსვათ $\hat{y}_i$-ს მნიშვნელობა და ჩავწეროთ პირველი რიგის აუცილებელი პირობა, რომლის დროსაც G მინიმალურია.

$$\frac{\partial G}{\partial b_0}=-2\sum_{i=1}^{n}(y_i-b_0-b_1x_{i1}-b_2x_{i2})=\sum_{i=1}^{n}e_i=0$$

$$\frac{\partial G}{\partial b_1}=-2\sum_{i=1}^{n}(y_i-b_0-b_1x_{i1}-b_2x_{i2})x_{i1}=\sum_{i=1}^{n}e_ix_{i1}=0$$

$$\frac{\partial G}{\partial b_2}=-2\sum_{i=1}^{n}(y_i-b_0-b_1x_{i1}-b_2x_{i2})x_{i2}=\sum_{i=1}^{n}e_ix_{i2}=0$$

მარტივი გარდაქმნებით მივიღებთ:

$$\begin{cases} nb_0+b_1\sum_{i=1}^{n}x_{i1}+b_2\sum_{i=1}^{n}x_{i2}=\sum_{i=1}^{n}y_i \\ b_0\sum_{i=1}^{n}x_{i1}+b_1\sum_{i=1}^{n}x_{i1}^2+b_2\sum_{i=1}^{n}x_{i1}\cdot x_{i2}=\sum_{i=1}^{n}x_{i1}\cdot y_i \\ b_0\sum_{i=1}^{n}x_{i2}+b_1\sum_{i=1}^{n}x_{i1}x_{i2}+b_2\sum_{i=1}^{n}x_{i2}^2=\sum_{i=1}^{n}x_{i2}y_i \end{cases}$$

მიღებულ განტოლებას ვყოფთ n-ზე, გამოვსახავთ $b_0 = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2$ და ჩავსვამთ მის მნიშვნელობას მეორე და მესამე განტოლებაში, მივიღებთ:

$$\begin{cases} b_1 \, var(x_1) + b_2 \, cov(x_1, x_2) = cov(x_1, y) \\ b_1 \, cov(x_1, x_2) + b_2 \, var(x_2) = cov(x_2, y) \end{cases}$$

სადაც: $Var(x_1) = \frac{1}{n} \sum x_{i1}^2 - \bar{x}_1^2 \qquad Var(x_2) = \frac{1}{n} \sum_{i=1}^{n} x_{i2}^2 - \bar{x}_2^2$.

$$cov(x_1, x_2) = \frac{1}{n} \sum_{i=1}^{n} x_{i1} \, x_{i2} - \bar{x}_1 \cdot \bar{x}_2$$
$$cov(x_1, y) = \frac{1}{n} \sum_{i=1}^{n} x_{i1} \, y - \bar{x}_1 \cdot \bar{y}.$$
$$cov(x_2, y) = \frac{1}{n} \sum_{i=1}^{n} x_{i2} \, y - \bar{x}_2 \cdot \bar{y}.$$

საბოლოოდ გვაქვს:

$$b_1 = \frac{cov(x_1, y) \cdot Var(x_2) - cov(x_2, y) \, cov(x_1, x_2)}{Var(x_1) \, Var(x_2) - (cov(x_1, x_2))^2},$$

$$b_2 = \frac{cov(x_2, y) \cdot Var(x_1) - cov(x_1, y) \cdot cov(x_1, x_2)}{var(x_1) \, Var(x_2) - (cov(x_1, x_2))^2},$$

40. მრავლობითი რეგრესიის კოეფიციენტების ინტერპრეტაცია

მრავლობითი რეგრესიის განტოლებაში b_1 გამოსახავს x_1 ფაქტორული ცვლადის ერთეული ცვლილების შედეგად საშუალოდ შედეგობრივი y ცვლადის საშუალო მნიშვნელობის ნაზრდს იმ პირობებში, როცა x_2-ის მნიშვნელობა ფიქსირებულია. ანალოგიურად, b_2 გამოსახავს x_2 ფაქტორის ერთეული ნაზრდის წმინდა გავლენას y ცვლადის საშუალო მნიშვნელობაზე, როცა x_1 ცვლადის მნიშვნელობა ფიქსირებულია.

41. მრავლობითი რეგრესიის წრფივი მოდელი (ზოგადი შემთხვევა)

დავუშვათ, რომ შედეგობრივ y ცვლადზე მოქმედებს m ახსნადი $x_1, x_2, \dots, x_m$ ცვლადი ფაქტორები. მაშინ მრავლობითი რეგრესიის ზოგადი წრფივი მოდელი შემდეგ სახით ჩაიწერება.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + u.$$

სადაც $\beta_0, \beta_1, \dots, \beta_m$ რეგრესიის უცნობი კოეფიციენტებია. რეგრესიული ანალიზის მთავარი ამოცანაა $b_0, b_1, \dots, b_m$ შეფასებების განსაზღვრა და შერჩევითი განტოლების აგება

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_m x_m.$$

b_0, b_1, b_m-ის განსაზღვრისათვის შევადგინოთ უმცირეს კვადრატთა მეთოდის კვადრატული ფორმით

$$G = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_1 - b_2 x_2 - \dots - b_m x_m) = \min$$

ვიპოვოთ G-ის კერძო წარმოებულები და გავუტოლოთ ნულს. შესაბამისი გარდაქმნებით ნორმალური განტოლებათა სისტემა მიიღებს სახეს:

$$\begin{aligned}
&b_0 \cdot n + b_1 \sum_{i=1}^{n} x_{i1} + b_2 \sum_{i=1}^{n} x_{i2} + \dots + b_m \sum_{i=1}^{n} x_{im} = \sum_{i=1}^{n} y_i \\
&b_0 \cdot \sum_{i=1}^{n} x_{i1} + b_1 \sum_{i=1}^{n} x_{i1}^2 + b_2 \sum_{i=1}^{n} x_{i1} x_{i2} + \dots + b_m \sum_{i=1}^{n} x_{i1} x_{im} = \sum_{i=1}^{n} y_i x_{i1} \\
&b_0 \cdot \sum_{i=1}^{n} x_{i2} + b_1 \sum_{i=1}^{n} x_{i2} \cdot x_{i2} + b_2 \sum_{i=1}^{n} x_{i2}^2 + \dots + b_m \sum_{i=1}^{n} x_{i2} x_{im} = \sum_{i=1}^{n} y_i x_{i2} \\
&\dots \\
&b_0 \cdot \sum_{i=1}^{n} x_{im} + b_1 \sum_{i=1}^{n} x_{im} x_{i1} + b_2 \sum_{i=1}^{n} x_{im} x_{i2} + \dots + b_m \sum_{i=1}^{n} x_{im}^2 = \sum_{i=1}^{n} y_i x_{im}
\end{aligned}$$

გამოვიყენოთ მატრიცულ ჩანაწერის ფორმით

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ y_n \end{pmatrix} \qquad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} \qquad B = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$$

$$u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \qquad \hat{B} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{pmatrix} \qquad e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

$$Y = XB + u$$
$$Y = X\hat{B} + e$$

თუ ავაგებთ A მატრიცს ნორმალური განტოლებათა სისტემის b_j, $j=1,2,\dots,m$ კოეფიციენტებთან მდგომი სიდიდეებისგან შევადგენთ;

$$A = X^T \cdot X$$

$$X^T \cdot X = \begin{pmatrix} n & \sum_{i=1}^n x_{i1} & \dots & \sum_{i=1}^n x_{im} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \dots & \sum_{i=1}^n x_{i1} x_{im} \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^n x_{im} & \sum_{i=1}^n x_{im} x_{i1} & \dots & \sum_{i=1}^n x_{im}^2 \end{pmatrix}$$

$$X^T \cdot Y = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_{i1} \\ \vdots \\ \sum_{i=1}^n y_i x_{im} \end{pmatrix}$$

მაშასადამე, ნორმალური განტოლებათა სისტემის მატრიცული სახე: $X^T \cdot X \cdot \hat{B} = X^T \cdot Y$

$$\hat{B} = (X^T X)^{-1} \cdot X^T \cdot Y$$

42. მრავლობითი რეგრესიის განტოლება და კოეფიციენტები სტანდარტიზებულ მასშტაბში (განტოლება ნატურალური ფორმით.

დავუშვათ, რომ ემპირიულ ანალიზით მიღებულია მრავლობითი შერჩევითი განტოლება:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_m x_m$$

ნორმალური განტოლებათა სისტემის I განტოლების n-ზე გაყოფის, რომელსაც ასეთი სახე აქვს $\bar{y} = b_0 + b_1\bar{x}_1 + b_2\bar{x}_2 + \dots + b_m\bar{x}_m$ გამოვაკლებთ (1) განტოლებას და მივიღებთ:

$$\hat{y} - \bar{y} = b_1(x_1 - \bar{x}) + b_2(x_2 - \bar{x}_2) + \dots + b_m(x_m - \bar{x}_m).$$

ამ განსხვავების ორივე წევრი გავყოთ S_y-ზე, ამასთან პირველი შესაკრები გავყოთ და გავამრავლოთ S_{x_1}-ზე, მეორე S_{x_2}-ზე და ა.შ. მივიღებთ:

$$\frac{\hat{y} - \bar{y}}{S_y} = b_1 \frac{S_{x_1}}{S_y} \cdot \frac{(x_1 - \bar{x}_1)}{S_{x_1}} + b_2 \cdot \frac{S_{x_2}}{S_y} \cdot \frac{(x_2 - \bar{x}_2)}{S_{x_2}} + \dots + \frac{b_m \cdot S_{x_m}(x_m - \bar{x}_m)}{S_y \cdot S_{x_m}}$$

სადაც $\frac{\hat{y} - \bar{y}}{S_y}$; $\frac{x_1 - \bar{x}_1}{S_{x_1}}$; $\frac{x_2 - \bar{x}_2}{S_{x_2}}$; $\frac{x_m - \bar{x}_m}{S_{x_m}}$

ეწოდება სტანდარტიზებული ცვლადები. ხოლო $b_j \cdot \frac{S_{x_j}}{S_y}$ - სტანდარტიზებული კოეფიციენტები.

შემოვიღოთ აღნიშვნები: $\frac{\hat{y} - \bar{y}}{S_y} \to w$ $\frac{x_j - \bar{x}_j}{S_{x_j}} \to z_j$ $j = 1, 2, \dots, m$.

მაშინ რეგრესიის განტოლება სტანდარტიზებულ მასშტაბში შეგვიძლია ჩავწეროთ:

$$w = \tilde{b}_1 z_1 + \tilde{b}_2 z_2 + \dots + \tilde{b}_m z_m$$

$\tilde{b}_j = b_j \frac{S_{x_j}}{S_y}$, $j = 1, 2, \dots m$.

43. მრავლობითი რეგრესიის განტოლება და კოეფიციენტები სტანდარტიზებულ მასშტაბში (რეგრესიის განტოლება ნატურალური ფორ-მით და არის მიღებული)

რეგრესიის სტანდარტიზებული განტოლება და კოეფიციენტები შეიძლება უშუალოდ სხვა მონაცემებიდან რეგრესიის ნატურ-ალური განტოლების იგივე გარდაქმნით მივიღოთ.

$$b_1 \, var(X_1) + b_2 \, cov(X_1, X_2) + \ldots + b_m \, cov(X_1, X_m) = cov(X_1, y)$$
$$b_1 \, cov(x_1, X_2) + b_2 \, Var(X_2) + \ldots + b_m \, cov(X_2, X_m) = cov(X_2, y)$$
$$\ldots$$
$$b_1 \, cov(X_m, X_1) + b_2 \, cov(X_m, X_2) + \ldots + b_m \, Var(X_m) = cov(X_m, y)$$

სისტემის პირველი განტოლების ყველა წევრი გავყოთ $S_y S_{x_1}$-ზე, მეორე განტოლების $S_y S_{x_2}$-ზე და ა.შ. გავითვალისწინოთ, რომ $\frac{cov(X_j, y)}{S_{x_j} S_y} = r_{yx_j}$ და $\frac{cov(X_j, X_k)}{S_{x_j} \cdot S_{x_k}} = r_{x_j x_k}$

მივიღებთ:

$$\tilde{b}_1 + \tilde{b}_2 r_{x_1, x_2} + \ldots + \tilde{b}_m r_{x_1 x_m} = r_{y x_1}$$
$$\tilde{b}_1 \cdot r_{x_2, x_1} + \tilde{b}_2 + \ldots + \tilde{b}_m r_{x_2 x_m} = r_{y x_2}$$
$$\ldots$$
$$\tilde{b}_1 r_{x_m, x_1} + \tilde{b}_2 r_{x_m x_2} + \ldots + \tilde{b}_m = r_{y x_m}$$

როდესაც წყვილური კორელაციის r_{yx_j}, r_{x_j, x_k} კოეფიციენტები მიღებულია, სისტემის ამოხსნა სტანდარტიზებულ კოეფიციენტებს მარ-ტივად ვპოვით.

44. მრავლობითი რეგრესიული ანალიზის ძირითადი დაშვებები

დაშვება I - დამოკიდებულება შედეგობრივ და ფაქტორულ ცვლადებს შორის $Y = XB + U$ - ში მოყვანილი წრფივი სახე გააჩნია, სადაც X დეტერმინირებული მატრიცაა, U - შემთხვევითი ვექტორი. ამასთან, U ვექტორის ელემენტები და x_j $j=1,2,\ldots m$ ცვლადები ერთმანეთისგან დამოუკიდებელია, ანუ $cov(u_i, x_j)=0$ $i=1,2,\ldots n$. $j=1,2,\ldots,m$.

დაშვება II - $E(u_i)=0$, $i=1,2,\ldots,n$.
დაშვება III - $E(u_i^2)=\sigma_u^2$, $i=1,2,\ldots,n$.
დაშვება IV - $E(u_i\cdot u_j)=0$ $i\neq j$.
დაშვება V - X მატრიცის რანგი ემთხვევა მისი სვეტების რიცხვს $(m+1)$-ს და ნაკლებია დაკვირვებების რიცხვზე n-ზე.
დაშვება VI - u_i, $i=1,2,\ldots,n$ ნორმალურად განაწილებული შემთხვევითი სიდიდეა $u_i \sim N(0, \sigma_u^2)$

45. გაუს მარკოვის თეორემა მრავლობითი რეგრესიის შემთხვევაში. რეგრესიის კოეფიციენტების შეფასებათა გადაუადგილებლობა (ა)

თუ მრავლობითი რეგრესიის მოდელისათვის $Y=XB+U$ ადგილი აქვს მრავლობითი რეგრესიული ანალიზის ძირითად 1-5 დაშვებას, მაშინ უმცირეს კვადრატთა მეთოდით მიღებული შეფასებები $\hat{B}=(X^TX)^{-1}\cdot X^T\cdot Y$ (1) წრფივ გადაუადგილებელ შეფასებებს შორის ოპტიმალურია.

(1)-დან უშუალოდ გამომდინარეობს, რომ შეფასებების $\hat{B}$ ვექტორი Y-ზე წრფივად დამოკიდებული. ვაჩვენოთ, რომ $\hat{B}$ არის გადაუადგილებელი შეფასების ვექტორი B-სთვის.

$$E(\hat{\beta}_j)=\beta_j \quad j=0,1,\ldots,m$$

ამ უკანასკნელის ვექტორული ჩანაწერი - $E(\hat{B})=B$.
გარდავქმნათ იგი, მივიღებთ:

$$\hat{B}=(X^TX)^{-1}X^TY=(X^TX)^{-1}X^T(XB+U)=(X^TX)^{-1}\cdot(X^TX)B+$$
$$+(X^TX)^{-1}X^T\cdot U.$$

რადგანაც X^TX და $(X^TX)^{-1}$ ურთიერთშებრუნებული მატრიცებია, მათი ნამრავლი ერთეულოვანი მატრიცაა, ე.ი.:

$$\hat{B}=B+(X^TX)^{-1}X^T\cdot U.$$

წყვილური რეგრესიის მოდელის ანალოგიურად მრავლობითი რეგრესიის მოდელის ყოველი b_j, $j=0,1,\dots m$ არის ნაწილობრივი შეფასება, რადგან მისი მოლოდინი და β_j-ს ტოლია, რადგან X-შემთხვევითი და U ვექტორია განსაზღვრული. I და II დაშვების თანახმად გვექნება:

$$E(\hat{B})=B+(X^TX)^{-1}X^T\cdot E(U)=B$$

მაშასადამე, B გადაუნაცვლებელი შეფასების ვექტორია.

46. მრავლობითი რეგრესიის კოეფიციენტების კოვარიაციული მატრიცა.

კოვარიაციული მატრიცა $cov(\hat{B})$ შეფასების $\hat{B}$ ვექტორის ელემენტების დისპერსიების განსაზღვრისათვის გამოიყენება. იგი შემდეგნაირად განისაზღვრება:

$$cov(\hat{B})=E[\hat{B}-E(\hat{B}))(\hat{B}-E(\hat{B}))^T]=E[(\hat{B}-B)(\hat{B}-B)^T]=$$

$$=E\left[\begin{pmatrix} b_0-\beta_0 \\ b_1-\beta_1 \\ \vdots \\ b_m-\beta_m \end{pmatrix}\begin{pmatrix} b_0-\beta_0 & b_1-\beta_1 & \cdots & b_m-\beta_m \end{pmatrix}\right]=$$

$$=\begin{pmatrix} \sigma_{00} & \sigma_{01} & \cdots & \sigma_{0m} \\ \sigma_{10} & \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & \vdots & & \vdots \\ \sigma_{m0} & \sigma_{m1} & \cdots & \sigma_{mm} \end{pmatrix}$$

მატრიცის მთავარ დიაგონალზე მდგომი ელემენტები $\sigma_{00}, \sigma_{11}, ... \sigma_{mm}$ $b_0, b_1, ... b_m$ შეფასებების დისპერსიებია. არადიაგონალური ელემენტები კი რეგრესიის კოეფიციენტების შეფასებების ურთიერთ-დამოკიდებულების მახასიათებლებს წარმოადგენს.

47. σ_u^2 დისპერსიის შეფასება მრავლობითი რეგრესიის მოდელში.

რეგრესიის კოეფიციენტების შეფასებათა კოვარიაციული მატრიცის გამო-სახულებაში $cov(\hat{B}) = \sigma_u^2 (X^T X)^{-1}$ $(X^T X)^{-1}$ მატრიცასთან ერთად მამრავლის სახით შედის შემთხვევითი წევრის დისპერსია σ_u^2, რომლის მნიშვნელობა უცნობია. რომ გამოვთვალოთ $\hat{B}$ ვექტორის ელემენ-ტების დისპერსიები σ_u^2 უნდა შევაფასოთ, რისთვისაც ვსარგებ-ლობთ ნარჩენობითი წევრების ვექტორით $e = Y - \hat{Y} = Y - X\hat{B} = (e_1\ e_2\ \dots\ e_n)^T$ გამოვთვლით შერჩევით დისპერსიას

$$Var(e) = \frac{1}{n} \sum_{i=1}^{n} e_i^2 = \frac{1}{n} e^T e$$

მაგრამ უშუალოდ $Var(e)$ არ გამოდგება σ_u^2-ის შეფასებად, ვინაიდან იგი გადაადგილებულ შეფასებას წარმოადგენს. ამიტომაც, σ_u^2-ის გადაუადგილებელი შეფასების როლში $Var(e)$-ს ნაცვლად უნდა განვიხილოთ კორექტირებული ნარჩენობითი დისპერსია S_e^2.

$$S_e^2 = \frac{n}{n-m-1} \cdot Var(e) = \frac{1}{n-m-1} \cdot \sum_{i=1}^{n} e_i^2 = \frac{1}{n-m-1} \cdot e^T e$$

48. მრავლობითი რეგრესიის წრფივი მოდელის კოეფიციენტების მნიშვნელოვნების შეფასება. ნდობის ინტერვალი

რეგრესიის კოეფიციენტების შეფასებათა დისპერსიების $S^2_{b_j}$ $j=0,1,\dots m$ ცოდნა საშუალებას გვაძლევს, შევაფასოთ კვლევით მეთოდით მიღე-ბული b_j, $j=0,1,\dots,m$ პარამეტრების სტატისტიკური მნიშვნე-ლობის შემოწმება მოვახდინოთ.
მრავლობითი რეგრესიის კოეფიციენტების მნიშვნელოვნობის შესამოწმებლად ვიყენებთ t_{b_j} სტატისტიკას.

$$t_{b_j} = \frac{b_j - \beta_j}{S_{b_j}} \sim t(n-m-1) \qquad j=0,1,\dots,m.$$

ვუშვებთ ნულოვან და ალტერნატიულ ჰიპოთეზებს:
$H_0: \beta_1 = \beta_{j0}$ $\quad H_1: \beta_j \neq \beta_{j0}$
პრაქტიკაში β_{j0}-ის როლში უმეტესად გამოიყენება $\beta_{j0}=0$.
სტუდენტის განაწილების ცხრილიდან მნიშვნელოვნების δ დონით და თავისუფლების $n-m-1$ ხარისხით ვპოულობთ კრიტიკულ მნიშვნელობას $t_{კრ}$-ს. თუ $|t_{b_j}| > t_{კრ}$, მაშინ ნულოვან ჰიპოთეზას უარვყოფთ და მივიღებთ H_1-ს.
β_j კოეფიციენტი $t_{n-m-1,\delta/2}$-ს მნიშვნელობისათვის $1-\delta$ ნდობის ალბათობით შემდეგ საზღვრებში შეიძლება იმყოფებოდეს:

$$b_j - t_{n-m-1,\delta/2} S_{b_j} \le \beta_j \le b_j + t_{n-m-1,\delta/2} S_{b_j}.$$

49 ერთი წრფივი შეზღუდვის ტესტირება

ზოგჯერ ეკონომიკაში მოქმედებს მოსაზრების მიხედვით მოდელის პარამეტრების ეფექტი, რაც ფორმალიზებული ასეთნაირი ჩანაწერი:

$$\beta_1 + \beta_2 = C$$

ჩამოვაყალიბოთ ნულოვანი ჰიპოთეზა:
$H_0: \beta_1 + \beta_2 = C$
$H_0: \pi_0\beta_0 + \pi_1 \cdot \beta_1 + \dots + \pi_m\beta_m = C$, სადაც C და $\pi_0, \pi_1 \dots \pi_m$ წინასწარ მოცემული სკალარული სიდიდეებია, ხოლო $\beta_1, \beta_2 \dots, \beta_m$

რეგრესიის განტოლების კოეფიციენტები.
შემოვიღოთ აღნიშვნები: $\pi = (\pi_0, \pi_1 \dots \pi_m)^T$ $B = (\beta_0, \beta_1, \dots \beta_m)^T$
$\pi^T = (0, 1, 1)$.
ნულოვანი ჰიპოთეზის მიხედვით ეს სტატისტიკა t ტესტის საშუალებით, რომელიც სტუდენტის განაწილების კანონს ექვემდებარება

$$t = \frac{\pi^T \hat{B} - C}{S_e \sqrt{\pi^T (X^T X)^{-1} \pi}} \sim t(n-m-1)$$

თუ აღმოჩნდება, რომ მიღებული t ნაკლებია $t_{კრ}$-ზე, ე.ი H_0 მიიღება. t ტესტის გამოყენებით წრფივი შეზღუდვის შემოწმება სხვა გზითაც შეიძლება ამ შემთხვევაში განტოლების გარდაქმნით შემდეგნაირად:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \rightarrow y = \beta_0 + \beta_1 (x_1 - x_2) + (\beta_1 + \beta_2) x_2$$
$$\hat{y} = b_0 + b_1 (x_1 - x_2) + (b_1 + b_2) x_2 + e$$

მიღებული ფორმით x_2-ის კოეფიციენტი უკვე არის ნულოვანი ჰიპოთეზის პირობა. აღვნიშნოთ $\beta_1 + \beta_2 = \tilde{\beta}_2$ და $b_1 + b_2 = \tilde{b}_2$. $\tilde{b}_2$-ის მნიშვნელობა შევამოწმოთ t ტესტით

$$t_{\tilde{b}_2} = \frac{\tilde{b}_2 - \tilde{\beta}_2}{S_{\tilde{b}_2}}$$

თუ აღმოჩნდა, რომ $t_{\tilde{b}_2} < t_{კრ}$, ეს მიგვანიშნებს იმაზე, რომ ნულოვანი ჰიპოთეზა მიიღება.

50. დისპერსიული ანალიზი წრფივი რეგრესიის მოდელში.

ისევე როგორც წყვილური რეგრესიის განტოლებაში, ამ შემთხვევაშიც დისპერსიული ანალიზის ძირითადი ვარიანტი არის y-ის საშუალოს ირგვლივ მიღებული გადახრების ჯამის დაშლა:

$$\underbrace{\sum_{i=1}^{n} (y_i - \bar{y})^2}_{Q} = \underbrace{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}_{Q_R} + \underbrace{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}_{Q_E}$$

$$Q = Q_R + Q_E$$

Q - შედეგობრივი ცვლადის ჯამური ვარიაცია.
Q_R - რეგრესიის განტოლებით ახსნილი ვარიაცია
Q_E - რეგრესიის განტოლებით აუხსნელი ვარიაცია.

დავამტკიცოთ განტოლების სამართლიანობა:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}\left[(y_i-\hat{y}_i)+(\hat{y}_i-\bar{y})\right]^2 =$$

$$= \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 + 2\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y}) + \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$$

შუა წევრი შესაკრები შემდეგნაირად გარდავქმნათ

$$A = \sum_{i=1}^{n} e_i(y_i-\hat{y}_i) = \sum_{i=1}^{n} e_i\hat{y}_i - \bar{y}\sum_{i=1}^{n} e_i =$$

$$= b_0\sum_{i=1}^{n} e_i + b_1\sum_{i=1}^{n} e_i x_{i1} + \dots + b_m\sum_{i=1}^{n} e_i x_{im} - \bar{y}\sum_{i=1}^{n} e_i,$$ სადაც

$e_i = y_i - \hat{y}_i$ ნორმალური განტოლებათა სისტემის თანახმად $\sum_{i=1}^{n} e_i = 0$ და $\sum_{i=1}^{n} e_i x_{i1} = 0$, ამიტომ $A=0$ და მოყვანილი ტოლობა სამართლიანია.

51. დეტერმინაციის კოეფიციენტი მრავლობითი რეგრესიის მოდელში.

დეტერმინაციის კოეფიციენტი R^2 y-ის დაკვირვებული მნიშვნელობების ჯამურ ვარიაციაში რეგრესიის შეფასებული განტოლებით ახსნილი ვარიაციის წილს გამოსახავს.

$$R^2 = 1 - \frac{Q_E}{Q} = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} = 1 - \frac{Var(e)}{Var(y)}$$

$$R^2 = \frac{Q_R}{Q} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{Var(\hat{y})}{Var(y)}.$$

რაც დიდია R^2-ის მნიშვნელობა, მით მაღალია რეგრესიის განტოლებით შეფასებული ცვლადის ახსნის ხარისხი. $0 \le R^2 \le 1$

52. დეტერმინაციის კორექტირებული კოეფიციენტი

დეტერმინაციის კოეფიციენტი ფაქტორთა რიცხვის მიმართ მგრძნობიარეა. იზრდება ფაქტორთა რაოდენობა და შესაბამისად იზრდება დეტერმინაციის კოეფიციენტი, რაც სულაც არ ნიშნავს რეგრესიის განტოლების გაუმჯობესებას. ამ ნაკლის აღმოსაფხვრელად შემოღებულია კორექტირებული კოეფიციენტი $\bar{R}^2$.

$$\bar{R}^2 = 1 - \frac{Q_E/(n-m-1)}{Q/(n-1)} = R^2 - \frac{m}{n-m-1}(1-R^2)$$

მოცემული გამოსახულება გვიჩვენებს:

1. კორექტირებული კოეფიციენტი არ აღემატება დეტერმინაციის კოეფიციენტს.

2. m-ის გაზრდისას ერთად იზრდება შეფარდება $\frac{m}{n-m-1}$, მაშასადამე, კოეფიციენტი მცირდება.

ე.ი. ახსნელი ცვლადის რეგრესიის მოდელში ჩართვის შემთხვევაში R^2-სგან განსხვავებით კორექტირებული კოეფიციენტი შეიძლება შემცირდეს. რეგრესიაში ახალი ცვლადის X_{m+1} დამატება მხოლოდ მაშინ ზრდის $\bar{R}^2$-ს, თუ ამ ცვლადის კოეფიციენტის b_{m+1} შესაბამისი $t_{b_{m+1}}$ სტატისტიკა აღემატება ერთიანს $|t_{b_{m+1}}| > 1$. მაგრამ, როგორც ეს ხშირად ჯერ კიდევ არ ნიშნავს X_{m+1} ცვლადის b_{m+1} კოეფიციენტის მნიშვნელოვნებას, ამიტომ $\bar{R}^2$-ის ზრდა რეგრესიის განტოლების ხარისხს ყოველთვის არ აუმჯობესებს.

53. მრავლობითი კორელაციის კოეფიციენტი

მრავლობითი კორელაციის კოეფიციენტი ახასიათებს დამოკიდებულებას შედეგობრივ ცვლადსა და განტოლებაში შემავალ ყველა ამხსნელ ცვლადს შორის.

წყვილური კორელაციისგან განსხვავებით ის მოქცეულია შემდეგ საზღვრებში $0 \le R_{yx} \le 1$. ვინაიდან y-ზე მოქმედი ფაქტორების ერთობლივი და ცალკეული ზეგავლენა შეიძლება განვიხილოთ და მრავლობითი კორელაციის კოეფიციენტი 0-დან 1-მდე საზღვრებში მოექცეს.

მრავლობითი კორელაციის კოეფიციენტის გამოთვლის საფუძვლად უდევს წყვილური კორელაციის კოეფიციენტების სიმეტრიული მატრიცა.

$$R = \begin{pmatrix} 1 & r_{yx_1} & \dots & r_{yx_m} \\ r_{yx_1} & 1 & \dots & r_{x_1x_m} \\ \dots & \dots & \dots & \dots \\ r_{x_my} & r_{x_mx_1} & \dots & 1 \end{pmatrix}$$

მატრიცის პირველი სტრიქონი და სვეტი წარმოადგენს y-ის ყველა ფაქტორთან კორელაციის კოეფიციენტებს. დიაგონალზე მოცემულია ერთიანები, დანარჩენი არის ფაქტორებს შორის კორელაციის კოეფიციენტები.

მრავლობითი კორელაციის კოეფიციენტი გამოითვლება ფორმულით

$$R_{yx} = \sqrt{1 - \frac{D}{D_{11}}}$$, სადაც D არის R მატრიცის შესაბამისი დეტერმინანტი, ხოლო D_{11} კი R მატრიციდან პირველი სვეტის და სტრიქონის ამოღების შემდეგ მიღებული მატრიცის დეტერმინანტი. ეს ფორმულა ორფაქტორიანის შემთხვევაში ასე ჩაიწერება:

$$R_{yx} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2 \cdot r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}}$$

არსებობს დამოკიდებულება $(R_{yx})^2 = R^2$

54. 55. დეტერმინაციის კოეფიციენტის და მრავლობითი რეგრესიის სტატისტიკური მნიშვნელობის ანალიზი.

დეტერმინაციის კოეფიციენტის სტატისტიკური მნიშვნელოვნების შეფასება შეიძლება ორი ერთმანეთის ეკვივალენტური ნულოვანი ჰიპოთეზის გამოყენებით. ერთი მათგანი შეიძლება ასე ჩამოყალიბდეს:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_m = 0$$

მეორე ჰიპოთეზის სახე:

$$H_0: R^{*2} = 0.$$

სადაც R^{*2} დეტერმინაციის ერთობლივი კოეფიციენტია. პირველი ჰიპოთეზა რეგრესიის მოდელში შემავალი m ფაქტორული ცვლადის ერთდროული ამხსნელი უნარის მნიშვნელოვნებას, მეორე ჰიპოთეზა კი დეტერმინაციის მიღებული შერჩევითი კოეფიციენტის ნულისგან განსხვავების სტატისტიკური მნიშვნელობის შემოწმება.

ორი ჰიპოთეზისათვის F სტატისტიკა გამოიყენება:

$$F = \frac{(n-m-1)R^2}{m(1-R^2)} \sim F(m, n-m-1)$$

თუ ანგარიშით მიღებული $F > F_{კრ}$, რეგრესიის მოდელში ჩართულ ფაქტორთა ერთობლივი ამხსნელი უნარი, ასევე R^2-ის ნულისგან განსხვავებაც მნიშვნელოვანია.

56. ~~ნაწილობრივი F სტატისტიკის გამოყენებით.~~

განვიხილოთ m ფაქტორიანი მოდელი, რომელსაც ვუწოდოთ სრული მოდელი და $m-1$ ფაქტორიანი მოდელი, რომელსაც ვუწოდებთ შეკვეცილ მოდელს და რომელშიც x_j ფაქტორი არ ფიგურირებს.

m ფაქტორიანი მოდელის ახსნილი ვარიაცია აღვნიშნოთ Q_R^m-ით, შეკვეცილის კი - Q_R^{m-1}. მათ შორის სხვაობა გვაძლევს პასუხს კითხვაზე,

გააუმჯობესა თუ არა განტოლების ხარისხი x_j ცვლადის შეყვანამ.
ჩამოვაყალიბოთ ნულოვანი ჰიპოთეზა:

$$H_0 = R_m^{*2} - R_{m-1}^{*2} = 0$$
$$H_1 = R_m^{*2} - R_{m-1}^{*2} \neq 0.$$

შევადგინოთ F სტატისტიკა:

$$F_{x_j} = \frac{(R_m^2 - R_{m-1}^2)/1}{(1-R_m^2)/n-m-1} \sim F(1, n-m-1)$$

თუ $F_{x_j} > F_{კრ}$ ნულოვანი ჰიპოთეზა უნდა უარვყოთ ანუ x_j-ს დამატებამ გააუმჯობესა განტოლების ხარისხი.
არსებობს დამოკიდებულება $F_{x_j} = (t_{b_j})^2$.