ივ. ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტი

გიორგი გუმბერიძე

ნივთიერებების საშიშროების შეფასება რეგრესიული ანალიზის გამოყენებით

გამოყენებითი მათემატიკა

ნაშრომი შესრულებულია გამოყენებითი მათემატიკის მაგისტრის აკადემიური ხარისხის მოსაპოვებლად

ხელმძღვანელები: ბესარიონ დოჭვირი

გია ხატისაშვილი

თბილისი

2019

სარჩევი

ანოტაცია

დღესდღეისობით მსოფლიოს მრავალ ქვეყანაში ერთ-ერთი მნიშვნელოვანი პრობლემაა გარემოს დაცვა ისეთი ფაქტორისგან, როგორიც არის ტოქსიკანტები. აღნიშნული ფაქტორი აბინძურებს გარემოს და არახელსაყრელად მოქმედებს ადამიანზე. გარემოზე მოქმედი ამ ფაქტორის გაძლიერებას იწვევს: ინტენსიური ინდუსტიალიზაცია, ტექნოლოგიური პროცესების მექანიზაცია, მრეწველობაში ახალი ტექნოლოგიების დანერგვა, რეაქტიული ავიაციის განვითარება და მრავალი სხვა. ბუნების დამბინძურებლების წინაამღდეგ ბრძოლა საწარმოო და ბუნებრივი გარემოს დაცვის განუყოფელი ნაწილია. ეს ბრძოლა კომპლექსურია.

მოცემულ ნაშრომში გადმოცემულია კორელაციური და რეგრესიული ანალიზი და დროითი მწკრივების თეორია, რომელიც გამოყენებულია ბუნებაში არსებული ტოკსიკური ნივთიერებების საშიშროების შეფასების პრობლემაში. მიღებულია წრფივი რეგრესიული განტოლებები თევზის ორი სახეობისთვის ოქტანოლისა და წყლის შეფარდების კოეფიციენტსა და ბიოკონცენტრაციის ფაქტორის კავშირს შორის.

Summary

In recent times, in many countries of the world, one of the most important problems is to protect environment from toxicants. Toxicants pollute environment and doesn't play a good role on human beings. The things that enhance these toxicants are: Intensive industrialization, Mechanism of technological processes, Introduction of the new technologies in industry, Development of the reactive aviation and many more... It's an integral part of production and environmental protection to fight against these pollutants. The fight is complex.

In this work, there's a correlative, regressive and the theory of time rows set out, which are used to assess the danger of the Toxic substances in nature. There are linear regression equations for the two kind of species of fish. that's octanol and the link between water supply coefficient and the factor of bio-concentration.

შესავალი

ადამიანი ცხოვრობს, მონაწილეებს ბიოსფეროში მიმდინარე ნივთიერებათა წრებრუნვასა და გარემოსთან ენერგიის მიმოცვლაში. ბიოსფეროში ყველა პროცესი ურთიერთკავშირშია. კაცობრიობა ბიოსფეროს ნაწილია, ხოლო ადამიანი - ორგანული სიცოცხლის ერთ-ერთი სახე- Homo sapiens (გონიერი ადამიანი). გონებამ გამოყო ადამიანი ცოცხალი სამყაროდან და მისცა მას უზარმაზარი ძალა. ადამიანი საუკუნეების მანძილზე ცდილობდა არა ბუნებასთან შეგუებას, არამედ მის გარდაქმნას, რომ ბუნება გამხდარიყო მისი არსებობისათვის მოხერხებული. მაგრამ დღეისათვის კაცობრიობამ გააცნობიერა, რომ ადამიანის ნებისმიერი საქმიანობა ნეგატიურ გავლენას ახდენს გარემოზე, ხოლო ბიოსფეროს მდომარეობის გაუარესება საფრთხეს უქმნის ყოველ ცოცხალ არსებას, მათ შორის ადამიანს.

ამ ყოველივეს გათვალისწინებით აუცილებელი გახდა ისეთი ტოკსიკანტი ნივთიერებების აღწერა, და მათი დალაგება სკალაზე „საშიშროების“ მიხედვით.

რამდენიმე წელია აშშ-ში გაკეთეს ცხრილი, რომელშიც ტოქსიკანტი ნივთიერებების სათანდო სკალაზე დალაგება ხდება ტოქსიკურობის მაჩვენებლის მიხედვით.

ჩვენი ამოცანაა ნივთიერებების დალაგება არა მათი ტოქსიკურობის, არამედ „საშიშროების“ (თუ რომელ ნივთიერებას აქვს უფრო უაყოფითი გავლენა გარემოზე) მიხედვით.

ამ პრობლების სირთულიდან გამომდინარე, ნივთიერებების საშიშროებას განვიხილავთ პირველ შემთხვევაში წყალში მცხოვრები ცოცხალი ორგანიზმებისთვის (თევზებისათვის).

აღნიშნული პრობლემის გადაწყვეტაში ძირითად სტატისტიკურ მეთოდებთან ერთად ვიყენებთ რეგრესიულ ანალიზის მეთოდს. ჩვენი მიზანია მივიღოთ ტოქსიკანტი ნივთიერებების საშიშროების მიხედვით დალაგება, რომელიც არ იქნება დამოკიდებული კონკრეტული ქვეყნის გეოგრაფიულ მდებარეობაზე და იქნება მსოფლიოს ნებისმიერი ქვეყნისათვის საერთო.

თავი I

რეგრესია და კორელაცია

მათემატიკაში, როგორც მრავალ სხვა დარგში ხშირად აუცილებელი ხდება ორ სიდიდეს შორის დამოკიდებულების შესწავლა. ორ ცვლადს შორის კავშირის ყველაზე მკაფიო მაგალითია ფუნქციონალური კავშირი, როდესაც ერთი ცვლადის (არგუმენტის) რაიმე მნიშვნელობას შეესაბამება მეორე ცვლადის (ფუნქციის) ერთი და მხოლოდ ერთი მნიშვნელობა. ფუნქციის ცნობილი მნიშვნელობების მაგალითებია:

- $y = a + bx$ წრფივი ფუნქცია;
- $y = a + bx + cx^2$, კვადრატული ფუნქცია;
- $y = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$, პოლინომური ფუნქცია.

მათემატიკურ სტატისტიკაში განსაკუთრებული მნიშვნელობა აქვს ე.წ სტოქასტურ დამოკიდებულებას ორ ცვლადს შორის, ორ შემთხვევით სიდიდეს შორის. მათემატიკაში ტიპობრივია სწორედ ასეთი დამოკიდებულება. ასეთი დამოკიდებულებები აღიწერება ორი შემთხვევითი სიდიდის ერთობლივი განაწილების ფუნქციიის, განაწილების კანონის ან განაწილების სიმკვირივის საშუალებით. ეს ფუნქციები გვაძლევენ ორ შემთხვევით სიდიდეს შორის კავშირს, დამოკიდებულების ამომწურავ დახასიათებას. მაგრამ, ხშირად მოხერხებულია გამოვიყენოთ უფრო მარტივი ფუნქციები ან რიცხვითი კოეფიციენტები. ამ ამოცანას ემსახურება,სწორედ, რეგრესია და კორელაცია.

კორელაციის კოეფიციენტი. განვიხილოთ ორი შემთხვევითი სიდიდე, X და Y.

სიდიდეს რომელიც განისაზღვრება ფორმულით

$$cov(X,Y) = E(X - EX)(Y - EY) \qquad (1.1)$$

ეწოდება კოვარიაცია (X და Y სიდიდეებს შორის). კორელაციის კოეფიციენტი განისაზღვრება ფორმულით

$$\rho = \frac{cov(X,Y)}{\sigma_X * \sigma_Y} = \frac{E(X-EX)(Y-EY)}{\sigma_X * \sigma_Y} \qquad (1.2)$$

(1.1) და (1.2) ფორმულებით განსაზღვრული კოვარიაცია და კორელაციის კოეფიციენტები წარმოადგენენ თეორიულ მახასიათებლებს. კოვარიაციასა და კორელაციის კოეფიციენტებს და მათ სტატისტიკურ შეფასებებს განსაკუთრებული მნიშვნელობა ენიჭება რეგრესიულ ანალიზში.

ვთქვათ (X, Y) შემთხვევითი სიდიდეების წყვილის შესაძლო მნიშვნელობათა ერთობლიობიდან ამოღებული n მოცულობის შერჩევა

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n). \qquad (1.3)$$

შერჩევის ელემენტებია რიცხვითი წყვილები. თუ X და Y შემთხვევით სიდიდეებს შორის კავშირის შესწავლას ვაპირებთ, აუცილებებლად უნდა უნდა გაგვაჩნდეს X და Y წყვილზე მნიშვნელობები. შერჩევის მიხედვით ჩვენ მივიღებთ კავარიაციისა და კორელაციის კოეფიციენტების სტატისტიკურ შეფასებებს. (1.1) და (1.2) ფორმულების ანალოგიურად გვექნება გამოსახულებები:

$$\widetilde{cov}\ (X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y - \bar{y}), \qquad (1.4)$$

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \qquad (1.5)$$

სადაც $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

შემდგომში გამოვყენებთ შემდეგ აღნიშვნებს

$$SS_{XY} = \sum_{i=1}^{n}(x_i - \bar{x})(y - \bar{y}),\ SS_X = \sum_{i=1}^{n}(x_i - \bar{x})^2, SS_y = \sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad (1.6)$$

(SS პირველი ასოებია ინგლისური სიტყვების sum of squares - კვადრატების ჯამები).

ასე რომ გვექნება

$$\widetilde{cov}\ (X,Y) = \frac{SS_{XY}}{n-1}, \qquad (1.7)$$

$$r = \frac{SS_{XY}}{\sqrt{SS_X}\sqrt{SS_Y}} \qquad (1.8)$$

(1.8) ფორმულით განსაზღვრულ შერჩევით კორელაციის კოეფიციენტს თეორიული ρ კოეფიციენტის ანალოგიური თვესებები აქვს.

დასკვნები კორელაციის კოეფიციენტის შესახებ. როცა კორელაციის

კოეფიციენტი უცნობია, შერჩევითი კორელაციის კოეფიციენტის r-ის საშუალებით შეიძლება გაკეთდეს სტატისტიკური დასკვნები მისი მნიშვნელობის შესახებ, აიგოს ნდობის ინტერვალი ან შემოწმდეს ჰიპოთეზა კორელაციის კოეფიციენტის რაიმე რიცხვთან ტოლობის შესახებ. ზოგადად ეს ამოცანა საკმაოდ რთულია და აქ განვიხილავთ მხოლოდ შემდეგი ჰიპოთეზების შემოწმების საკითს:

$H_0\colon \rho = 0$,

$H_1\colon \rho \neq 0$,

თუ (X, Y) შემთხვევითი სიდიდეების ერთობლივი განაწილება ნორმალურია, მაშინ H_0 ჰიპოტეზის სამართლიანობისას სტატისტიკა

$$t = \frac{r}{\sqrt{1 - r^2}}\sqrt{n-2}$$

განაწილებულია სტიუდენტის კანონით n-2 თავისუფლების ხარისხით. ამგვარად სტიუდენტის განაწილების ცხრილის საშუალებით შეიძლება შემოწმდეს ზემოთ მოყვანილი ჰიპოთეზა.

სახელდობრ, H_0 ჰიპოთეზას უარვყოფთ α მნიშვნელობის დონით, თუ

$$|t| > t_{n-2,1-\alpha/2} \,.$$

წინაამღდეგ შემთხვევაში , H_0 ჰიპოთეზის უარყოფის საფუძველი არ გაგვაჩნია. ამავე განაწილების საშუალებით აიგება ნდობის ინტერვალი კორელაციის კოეფიციენტისათვის.

მარტივი წრფივი რეგრესია. უმცირეს კვადრატთა მეთოდი. როგორც უკვე აღვნიშნეთ Y შემთხვევითი სიდიდის დამოკიდებულება X შემთხვევითი სიდიდისგან შეიძლება აღიწეროს დეტერმინისტული ფუნქციით

$$Y = \varphi(X).$$

ამ კავშირის უმარტივესი ფორმაა

$$Y=B_0+B_1X. \tag{1.9}$$

შესაძლოა კავშირის სხვა ფორმებიც. მაგალითად

$$Y=B_0+B_1X+B_2X^2,\ Y=Ae^{BX},\ Y=AX^B,\ Y=A+BX^{-1}$$

და ა.შ

ტრადიციულ ეკონომიკურ და მათემატიკურ თეორიათა დიდ ნაწილში იგულისხმებოდა საზოგადოდ ზუსტი ფუნქციონალური კავშირების შესწავლა ცვლადებს შორის . მაგრამ ეკონომიკური მონაცემების ელემენტარული გაცნობაც კი გვიჩვენებს, რომ ცვლადებს შორის კავშირი უმეტეს შემთხვევაში არ არის დეტერმინისტული. მაგალითად, არ უნდა მოველოდეთ, ერთი და იგივე შემოსავლის (X) მქონე ოჯახების სამომხმარებლო ხარჯებიც (Y) ერთნაირი იქნება. ცალკეული უჯახებიც მოიხმარენ უფრო მეტს ვიდრე სხვები და პირიქით. ოღონდ მოსალოდნელია, რომ X სიდიდის შემოსავლის მქონე ოჯახების სამომხმარებლო ხარჯები დაჯგუფდება X-ზე დამოკიდებულ რაიმე მნიშვნელობის სიახლოვეში. ეს მოსაზრება გვთავაზობს დეტერმინისტული დამოკიდებულების განზოგადოებას შემდეგი ფორმით

$$Y = \varphi(X) + \varepsilon \,.$$

სადაც ε არის რაიმე შემთხვევითი სიდიდე, $E\varepsilon = 0,\ D\varepsilon = \sigma^2 > 0$. ამ კავშირის ყველაზე მარტივი ფორმაა ე.წ. წრფივი რეგრესიული კავშირი, როდესაც

$$Y = B_0 + B_1(x) + \varepsilon \tag{1.10}$$

და დამატებით იგულისხმება , რომ ε და X დამოუკიდებელი შემთხვევითი სიდიდეებია.

კერძოდ (1.10) თანაფარდობიდან გამომდინარეობს, რომ X-ის ფიქსირებული მნიშვნელობისათვის

$$Y(x) = B_0 + B_1(x) + \varepsilon \tag{1.11}$$

წარმოადგენს შემთხევით სიდიდეს, რომლის ლოდინი და დისპერსია გამოითვლება ფორმულებებით:

$$E(Y|X=x)=B_0+B_1x, \qquad D(Y|X=x)=\sigma^2$$

ამრიგად წრფივი რეგრესიული მოდელის ფარგლებში იგულისხმება, რომ Y-რეგრესია X-ზე წრფივია

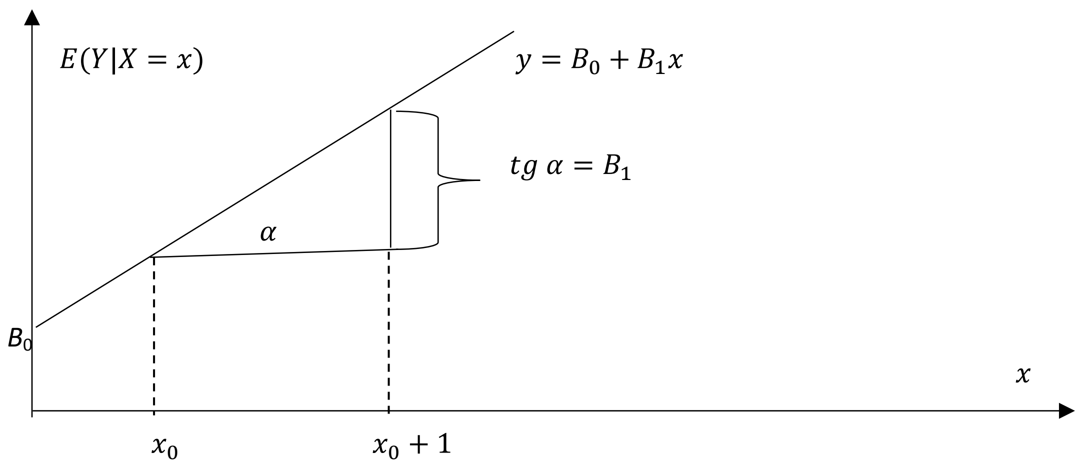
$$E(Y|X)=B_0+B_1X \tag{1.12}$$

და, შესაბამისად $y=B_0+B_1x$ წრფეს რეგერესიის წრფე, B_0 და B_1 კოეფიციენტებს - რეგრესიის კოეფიციენტები, ხოლო ε -ს - შემთხვევითი გადახრა ეწოდება.

შემდგომში ჩვენს მიერ პოსტულირებული იქნება მარტივი წრფივი რეგრესიის (1.10) მოდელი. X ცვლადს უწოდებენ დამოუკიდებელ (ამხსნელ) ცვლადს ანდა პრედიქტორს, ხოლო Y-ს კი - დამოკიდებულ (მოპასუხე) ცვლადს.

B_1 კოეფიციენტს შეიძლება მივცეთ შემდეგი ინტერპრეტაცია: ამხსნელ X ცვლადის ერთი ერთეულით ცვლილება დამოკიდებული Y ცვლადის ცვლილებას საშუალოდ B_1 სიდიდით. B_0 კოეფიციენტს თანაკვეთას უწოდებენ, რადგან ის წარმოადგენს რეგრესიის წრფის y ღერძთან გადაკვეთის წერტილს.

ვთავათ, $(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)$ დამოუკიდებელი დაკვირვებებია (X,Y) შემთხვევით სიდიდეთა წყვილზე. იმ შემთხვევაში ,როდესაც Y და X შემთხევევით სიდიდეებს შორის არსებობს (1.9) ტიპის წრფივი დეტერმინისტული კავშირი, მაშინ დაკვირვებული წყვილები განლაგდება $y=B_0+B_1x$ წრფეზე,



$$y_i=B_0+B_1x_i, \quad i=1,2,\dots,n \tag{1.13}$$

და B_0 და B_1 კოეფიციენტების მოძებნას იოლად მოვახერხებდით.ავიღებდით ორ ნებისმიერი წერტილს და მოვძებნიდით ამ ორ წერტილზ გამავალი წრფის განტოლებას. იმ შემთხევაში, როდესაც X და Y შემთხევევით სიდიდეებს შორის კავშირი აიწერება (1.10) მოდელით, მაშინ ნათელია, რომ დაკვირვებული წყვილები აღარ განლაგდება B_0+B_1X წრფეზე, არამედ გაბნეული იქნება ამ, საზოგადოდ ჩვენთვის უცნობი, B_0+B_1X წრფის მიმართ.

ჩვენი ამოცანაა დაკვირვებული წყვილების მეშვეობით ავაგოთ უცნობი B_0, B_1 კოეფიციენტებისა და σ^2 დისპერსიის შეფასებები. პრაქტიკულ ამოცანებში ხშირად X ცვლადი არ არის შემთხვევითი, მაგრამ მაშინაც კი როცა ის შემთხვევითია, იგულისხმება, რომ X- ის მნიშვნელოებები ზუსტადაა ცნობილი მანამდე, სანამ დავაკვირდებით Y - ების შესაბამის მნიშვნელობებს, ამიტომ X- ის მნიშვნელობები განიხილება როგორც ზუსტი რიცხვები და გადახრები მიეწერება მხოლოდ Y ცვლადს. ამიტომ შემდეგში Y ცვლადის შესაბამის მნიშვნელობებს ავღნიშნავთ ასომთავრული ასოებით, ხოლო X -ისას პატარა ასოებით; ამავე დროს, როგორც ადრე შევთნხმდით, როდესაც საუბარი იქნება რაიმე სტატისტიკების განაწილბაზე, Y_i , $i = 1,2, \dots , n$ ცვლადების ჩათვლით შემთხვევით სიდიდეებად , ხოლო სატატისტიკების დაკვირვებული მნიშვნელობების გამოსათვლელად, Y_i- ს ქვეში იგულისხმება Y ცვლადის დაკვირვებული მნიშვნელობა.

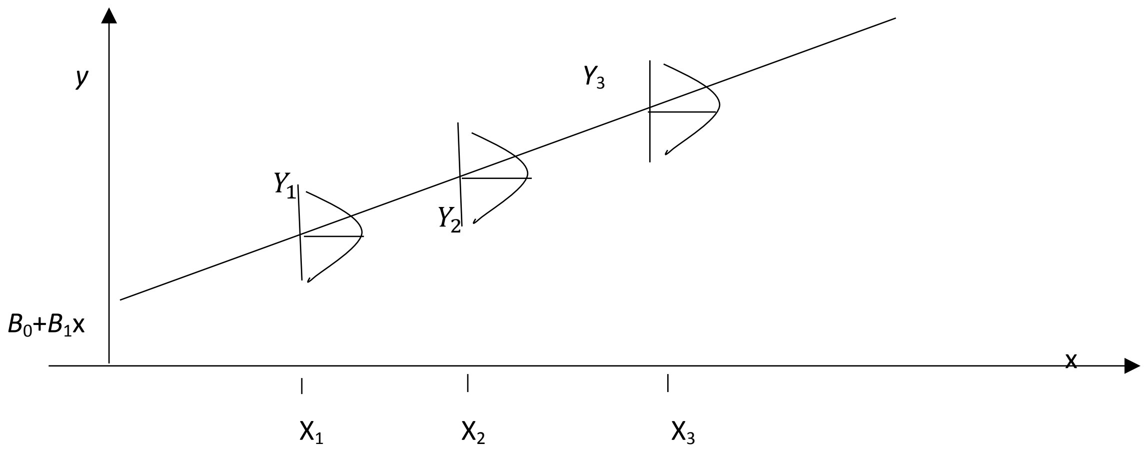
გაკეთებული შენიშვნების შემდეგ ამოცანა აღვწეროთ შემდეგნაირად: მოცემულია $(x_i, Y_i), i = 1,2, \dots , n$. წყვილები,რომელთა შორის შემდეგი დამოკიდებულებაა

$$Y_i = B_0 + B_1 x_1 + \varepsilon_i, \ \ i = 1,2, \dots , n, \qquad (1.14)$$

სადაც ε_i სწორედ ის შემთხვევითი სიდიდეა , რომელიც განაპირობებს წრფიდან გაბნევას.

ვუშვებთ, რომ ε_i განაწილებულია ნორმალურად ნულოვანი მათემატიკური ლოდინით და დისპერსიით, $\sigma^2 > 0$, რომელიც უცბობია; ვუშვებთ, აგრეთვე, რომ ε_i და ε_j, $i \neq j$ ურთიერთდამოუკიდებელია. B_0, B_1 საძიებელი წრფის ჭეშმარიტი (თეორიული) კოეფიციენტებია.

ჩვენი მიზანია შევაფასოთ B_0, B_1 კოეფიციენტები. შეფასებები აღვნიშნოთ ასოებით b_0 და b_1, შესაბამისად .



b_0 და b_1 შეფასებები მოძებნა ხდება ე.წ უმცირეს კვადრატთა მეთოდით. იგი დამუშავებული იყო ლაგრანჟისა და გაუსის მიერ XXI საუკუნის დასაწყისში . ამ მეთოდის თანახმად b_0 და b_1 შეფასებების აგება წარმოებს შემდეგი გამოსახულების მინიმიზაციით;

$$\sum_{i=1}^{n} (Y_i - b_0 - b_1 x_i)^2 \rightarrow min \qquad (1.15)$$

ე.ი ვეძებთ ისეთ $y = b_0 + b_1 x$ წრფეს, რომლიდანაც Yის ვერტიკალური გადახრების კვადრატების ჯამი $\sum e_i^2$, სადაც $e_i = Y_i - b_0 - b_1 x_i$ უმცირესია (ყველა სხვა წრფესთან შედარებით).

მინიმუმის მოსაძებნად გავაწარმოოთ (1.15) b_0და b_1 ცვლადებით და წარმოებულები გავუტოლოთ ნულს (მინიმუმის აუცილებელი პირობა). მივიღებთ;

$$\sum_{i=1}^{n} 2(Y_i - b_0 - b_1 x_i) = 0$$

$$\sum_{i=1}^{n} 2(Y_i - b_0 - b_1 x_i)^2 = 0 \tag{1.16}$$

შევკვეცოდ ორზე და ავჯამოთ წევრ-წევრად, შემდეგ ცნობილი და უცნობი წევრები დავალაგოთ ტოლობის სხვადასხვა მხარეს. მივიღებთ:

$$n b_0 + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} Y_i \,, \tag{1.17}$$

$$b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i Y_i \,, \tag{1.18}$$

მივიღეთ ორი განტოლებისაგან შემგარი სისტემა, ორი უცნობით (b_0 და b_1 უცნობებია).

ამ სისტემას ეწოდება ნორმალური განტოლებათა სისტემა. ამოვხსნათ ეს სისტემა

შეკრება-გამოკლების ხერხით.

$$b_1\left(n \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i \right) = n \sum_{i=1}^{n} x_i Y_i - \sum_{i=1}^{n} x_1 \sum_{i=1}^{n} Y_i$$

ამგვარად ,

$$b_1 = \frac{n \sum_{i=1}^{n} x_i Y_i - \sum_{i=1}^{n} x_1 \sum_{i=1}^{n} Y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{1.19}$$

და

$$b_0 = \frac{1}{n} \sum_{i=1}^{n} Y_i - b_1 \frac{1}{n} \sum_{n=1}^{n} x_i \tag{1.20}$$

იგულისხმება, რომ (1.20)-ში ჩასმულია b_1-ის გამოსახულება (1.19)-დან. ელემენტარული გარდაქმნების შემდეგ b_1-ის გამოსახულება (1.19) შეიძლება ჩაიწეროს შემდეგნაირად:

$$b_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{1.21}$$

ადრე შემოღებული აღნიშვნებში (იხ.(1.6)) (1.21) ფორმულა გადაიწერება შემდეგი ფორმით

$$b_1 = \frac{SS_{XY}}{SS_X} \,. \tag{1.22}$$

$\hat{Y} = b_0 + b_1 x$ წრფეს ეწოდება ეწოდება რეგრესიის წრფის შეფასება, ან უმცირეს კვადრატთა მეთოდით აგებული წრფე. თუ ამ წრფის განტოლებაში x-ის ნაცვლად

ჩავსვამთ ჩავსვამთ დაკვირვებულ $x_1, x_2, \dots, x_n$ მნიშვნელობებს, შესაბამის $\hat{Y} = b_0 + b_1 x_1, \dots$, $\hat{Y}_n = b_0 + b_1 x_n$ მნიშვნელობებს ეწოდება Y ცვლადის პროგნოზირებული მნიშვნელობები, ხოლო $e_1 = Y_1 - \hat{Y}_1, \dots, e_n = Y_n - \hat{Y}_n$ სიდიდეებს კი - ნაშთები.

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - (b_0 + b_1 x_i))^2 \qquad (1.23)$$

სიდიდეს ეწოდება კვადრატების ნარჩენები ჯამი (ნაშთა კვადრატების ჯამი). ის ხშირად აღინიშნება ასოებით SSE ინგლისური ფრაზის “sum of square for eror” სიტყვების პირველი ასოების მიხედვით ($SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$). მარტივი გარდაქმნების შემდეგ (1.23) შემდეგნაირად ჩაიწერება

$$SSE = SS_Y - \frac{SS_{XY}^2}{SS_X} \qquad (1.24)$$

ცხადია, რომ $\sum_{i=1}^{n} e_i^2$ წარმოადგენს $\sum_{i=1}^{n} \varepsilon_i^2$ -ის გარკვეულ შეფასებას. ამ სიდიდეს დიდი მნიშვნელობა აქვს რეგრესიის წრფის დასახასიათებლად.

მარტივი გარდაქმნების შემდეგ შეფასებული რეგრესიის წრფის განტოლება შეიძლება ჩავწეროთ ასე

$$y = \bar{Y} + b_1 (x - \bar{x}). \qquad (1.25)$$

ვგულისხმობთ, რომ ამმ განტოლებასი შეიძლება ჩაისვას x -სის ნებისმიერი მნიშვნელობა დაკვირვებული x_i, $i = 1,2, \dots, n$ სიდიდეების ცვალებადობის ინტერვალიდან, ე.ი ცვლადს ხშირად უწოდებენ ამხსნელ (განმმარტავ) ცვლადს.მისი ჩასმა რეგრესიის განტოლებაში ახსნის Y ცვლადის რაღაც ნაწილს, რაღაც ნაწილი კი Y ცვლადის აუხსნელი დარჩება.

შევნიშნოთ რეგრესიის წრფის შეფასების შემდეგი ორი თვისება:

1. ეს წრფე გადის $(\bar{x}, \bar{Y})$ წერტილზე.
2. $\sum_{i=1}^{n} (Y_i - \hat{Y}_i) = 0$.

თუ ამავე შერჩევის მონაცემების საშუალებთ განვსაზღვრავთ $X - ის$ y-ზე რეგრესიის წრფეს, მივიღებთ ასეთ განტოლებას

$$\hat{X} = \bar{X} + b_1 (y - \bar{y}) , \qquad (1.26)$$

სადაც $b_1 = \frac{SS_{XY}}{SS_Y}$. (აქ X ითვლება შემთხვევით სიდიდედ, y კი- ამხსნელ ცვლადად.)

უნდა გავითალისწინოთ, რომ მოდელი, რომელისც (1.14) ფორმულითაა მოცემული გულისხმობს , რომ გაბნევის დიაგრამაზე წერტილებს განევა განპირობებულია Y ცვლადის მნიშვნელობათა გაბნევით რეგრესიის მრუდიდან. x ცვლადის მნიშვნელობები განიხილება ზუსტ (დეტერმინისტულ) მნიშვნელობებად. დაშვება, რომ ჭეშმარიტი გადახრა ε_i ნორმალურად არის განაწილებული, თავისთავად,

აუცილებელი არ არის უმცირეს კვადრატთა მეთოდის გამოყენებისათვის, მაგრამ ეს დაშვება უაღრესად მნიშვნელოვანია სტატისტიკური დასკვნების მისაღებად.

შეიძლება ვაჩვენოთ, რომ b_0 და b_1 კოეფიციენტები განაწილებული არიან ნორმალურად

$$Eb_0 = B_0, \qquad Eb_1 = B_1, \tag{1.27}$$

მათემატიკური ლოდინებით

$$\sigma_{b_0}^2 = Db_0 = \sigma^2 \frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \tag{1.28}$$

და შესაბამისად

$$\sigma_{b_1}^2 = Db_1 = \sigma^2 \frac{n}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \tag{1.29}$$

დისპერსიით.

b_0 და b_1 შეფასებათა საშუალობი და დისპერსიები გამოითვლება ამ შეფასებათა (1.19) და (1.20) გამოსახულებებიდან, იმ ფაქტის გათვალისწინებით, რომ დაკვირვებული Y_i მნიშვნელობები დამოუკიდებელია, ნორმალურადაა განაწილებული $EY_i = b_0 + b_1x_i$ და $DY_i = \sigma^2$. ნათელია, რომ

$$\frac{b_0 - B_0}{\sigma_{b_0}} \sim N(0,1), \qquad \frac{b_1 - B_1}{\sigma_{b_1}} \sim N(0,1)$$

მაგრამ ეს ორი სტატისტიკა შეიცავს უცნობ, σ^2 სიდიდეს.

უცნობიი σ^2 დისპერსიის ჩაუნაცვლებელი შეფასება მოიცემა ფორმულით

$$S^2 = \frac{\sum_{i=1}^n e_i^2}{n-2} = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2} = \frac{SSE}{n-2} \tag{1.30}$$

უკანასკნელი ტოლობა შეიძლება ასეც ჩავწეროთ

$$S^2 = \frac{\sum_{i=1}^n Y_i^2 - b_0\sum_{i=1}^n Y_i - b_1\sum_{i=1}^n x_iY_i}{n-2} \tag{1.31}$$

ამ უკანასკნელის მრიცხველში $\sum_{i=1}^n Y_i$ და $\sum_{i=1}^n x_iY_i$ შესაკრებები უკვე გამოთვლილია b_0 და b_1-ის განსაზღვრის დროს . საჭიროა მხოლოდ $\sum_{i=1}^n Y_i^2$ -ის გამოთვლა .
S^2-ის გამოსახულებაში მნიშვნელში ფიგურირებს $n-2$ სიდიდე. ეს გამოწვეულია იმით, რომ შერჩევის მიხედვით შეფასებულია ორი მუდმივი (b_0 და b_1) და ამიტომ თავისუფლების ხარისხია $n-2$ (იხ.(1.16), რომელიც გვაძლევს ორ ბმას $\sum_{i=1}^n \left(Y_i - \hat{Y}_i\right) = 0$,

$\sum_{i=1}^n \left(Y_i - \hat{Y}_i\right)x_i = 0$, სადაც $\hat{Y} = b_0 + b_1x_i$).

მას შემდეგ, რაც გამოთვლილია S^2, შეიძლება დავწეროთ ${\sigma_{b_0}}^2$-ისა და ${\sigma_{b_1}}^2$-ის შეფასებები, შესაბამისად ${S_{b_0}}^2$ და ${S_{b_1}}^2$. გვექნება

$$S_{b_0}^2 = S^2 \frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} = S^2 \frac{\sum_{i=1}^n x_i^2}{n*SS_X} \tag{1.32}$$

$$S_{b_1}^2 = S^2 \frac{n}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} = S^2 \frac{1}{n*SS_X} \tag{1.33}$$

ამ შეფასებათა მეშვეობით უკვე შესაძლებელია

$$T_i = \frac{b_i - B_i}{S_{b_i}} \tag{1.34}$$

სტატისტიკების განხილვა. თითოეული მათგანი განაწილებულია სტიუდენტის, *n-2* თავისუფლების ხარისხის მქონე კანონით:

$$T_i \sim t(n-2) \qquad i = 0,1.$$

სწორედ T_0 და T_1 სტატისტიკების მეშვეობით ხდება დასკვნების გამოტანა. B_0 და B_1 პარამეტრების შესახებ (ჰიპოთეზების შემოწმება, ნდობის ინტერვალის აგება) .

მემდგომში ანალიზისათვის დაგვჭირდება შემდეგი მარტივი შესამოწმებელი თანაფარდობები

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2. \tag{1.35}$$

ავშიშნოთ, რომ (1.35) ფორმულაში $\sum_{i=1}^n (Y_i - \bar{Y})^2$ არის Y- ის დაკვირვებული Y_i

$i = 1,2, \dots, n$, მნიშვნელობის მათი საშუალო მნიშვნელობებიდან გადახრების კვადრატების სრული ჯამური სიდიდე (total sum of squares – *SST*, $SST = \sum_{i=1}^n (Y_i - \bar{Y})^2$),

რომელსაც სრულ ვარიაციას ვუწოდებთ, $\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ - რეგრესიის მრუდის მიმართ Y–ის მნიშვნელობების გადახრების კვადრეატების გაბნევის ჯამური სიდიდეა (error sum of squares – *SSE*, $SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$), *SSE* არის სრული გაბნევის ის ნაწილი, რომელიც არ აიხსნება რეგრესიის საშუალებით. $\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$ სრული გაბნევის ის ნაწილია, რომელიც ახსნილია რეგრესიის წრფით (regression sum of squares- *SSR*, $SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$). ამრიგად,

$$SST = SSE + SSR$$

რეგრესიის ხარისხის დასახასიათებლად შემოჰყავთ ე.წ. შერჩევითი დეტერმინაციის კოეფიციენტი. იგი განისაზღვრება ფორმულით

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{SSR}{SST}, \ \left(\frac{SSR}{SST} = 1 - \frac{SSE}{SST}\right) \tag{1.36}$$

ანუ R^2 წარმოადგენს რეგრესიით ახსნილ გაბნევის წილს სრულ გაბნევაში. შეიძლება ვაჩვენოთ, რომ R^2 ახალს არაფერს გვაძლევს, ვინაიდან

$$R^2 = r^2 \quad (1.37)$$

სადაც r შერჩევითი კორელაცის კოეფიციენტია.

თუ r^2 მცირეა, ე.ი. რეგრესიით ახსნილია სრული გაბნევის მცირე წილი, მაშინ მასაძებნია სხვა ალტერნატიული მოდელი (ვთქვათ, არაწრფივი ან მრავლობითი წრფივი რეგრესიის მოდელი და სხვა), რომელიც უფრო ეფექტურად ახსნიდა Y ცვლადის დაკვირვებული მნიშვნელობების სრულ გაბნევას თავისი საშუალო მნიშვნელობის მიმართ.

კორელაციის კოეფიციენტის საშუალებით შეიძლება ასეთი ფრომულა დავწეროთ:

$$\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = R^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = r^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2 \,, \quad (1.38)$$

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \left(1 - r^2\right) \sum_{i=1}^{n}(Y_i - \bar{Y})^2 \,. \quad (1.39)$$

სტატისტიკური დასკვნები რეგრესიის კოეფიციენტების შესახებ. მოდელის ვარგისიანობის შემოწმება. ზემოთ მოყვანილი ყველა ძირითადი ფორმულა, რომელთა მიხედვითაც სრულდება მარტივი წრფივი რეგრესიასთან დაკავშირებული გამოთვლები.

ბოლო ეტაპზე მოწმდება ჰიპოთეზები B_0 და B_1 კოეფიციენტების შესახებ. პირველ რიგში მოწმდება B_1 კოეფიციენტის ნულთან ტოლობის ჰიპოთეზა $H_0 : B_1 = 0$. B_1 კოეფიციენტს ვიხილავთ იმიტომ, რომ ეკონომიკური შინაარსი გააჩნია მხოლოდ ამ კოეფიციენტს, მის ნულთან ტოლობას იმიტომ, რომ თუ $B_1 = 0$ მაშინ რეგრესიის წრფე X ღერძის პარალელურია, ე.ი. X-ის ცვლილებით Y არ იცვლება.

თუ გვსურს α მნიშვნელობის დონით დავრწმუნდეთ, რომ არ არსებობს რაიმე (დადებითი ან უარყოფითი) წრფივი კავშირი, ნულოვან ჰიპოთეზა H_0 უნდა შემოწმდეს ორმხრივი

$H_1 : B_1 \neq 0$

ალტერნატივის წინაამღდეგ, თუ გვსურს დავწმუნდეთ დადებითი (უარყოფითი) წრფივი კავშირის არსებობაში, მაშინ ვიხილავთ ცალმხრივ

$H_1 : B_1 > 0 \; (B_1 < 0)$

ალტერნატივას და ყველა შემთხვევაში ჰიპოთეზის შემოწმება ხდება

$$T_1 = \frac{b_1}{S_{b_1}}$$

სტატისტიკაზე დაყრდნობით, სტანდარტული წესით (იმ ფაქტის გათვალისწინებით, რომ ეს უკანასკნელი განაწილებული სტიუდენტის კანონით, (n-2) თავისუფლების ხარისხით).

მაგალითად, შევამოწოთ ჰიპოთეზა

$H_0: \ B_1 = 0,$

$H_1 : B_1 \neq 0 \, .$

გადაწყვეტილების მიღების წესი შემდეგია:

H_0-ს ვიწუნებთ, თუ$|t_1| > t_{n-2,\alpha/2}$,

სადაც t_1 არის T_1 სტატისტიკის დაკვირვებული მნიშვნელობა, ხოლო $t_{n-2,\alpha/2}$ სტიუდენტის $t(n-2)$ განაწილების α-კრიტიკული წერტილი.
მაშინ α მნიშვნელობის დონისათვის ვიღებთ სტიუდენტის განაწილებას α-კრიტიკულ წერტილს $t_{n-2,\alpha}$ და

H_0-ს ვიწუნებთ, თუ $t > t_{n-2,\alpha}, (t < t_{n-2,\alpha})$

პროგნოზირება და სტატისტიკური დასკვნები წრფივი რეგრესიის საშუალებით. ვთქვათ x^* აღნიშნავს დამოუკიდებლი X ცვლადის რაიმე წინასწარ ფიქსირებულ მნიშვნელობას. მას შემდეგ რაც გამოთვლილია B_0-ისა და B_1-ის შეფასებები, b_0 და b_1, შეფასებული რეგრესიის წრფის $\hat{Y}(x^*) = b_0 + b_1 x^*$ მნიშვნელობა შეიძლება ჩაითვალოს ჭეშმარიტი რეგრესიის წრფის $\mu^* = E(Y|X = x^*) = B_0 + B_1 x^*$ მნიშვნელობის წერტილოვან შეფასებად და ამავე დროს $Y(x^*) = B_0 + B_1 x^* + \varepsilon$ შემთხვევითი სიდიდის პროგნოზირებულ მნიშვნელობად. თავისთავად $\hat{Y}(x^*) = b_0 + b_1 x^*$ არ იძლევა არანაირ იმფორმაციას იმის შესახებ, თუ რა სიზუსტით იყო შეფასებული ჭეშმარიტი რეგრესიის წრფე და რა სიზუსტით იყო პროგნოზირებული x^*- ის შესაბამისი $Y(x^*)$ სიდიდე. ამ იმფორმაციის მოპოვება შესაძლებელია μ^*-თვის ნდობის ინტერვალისა და $\hat{Y}(x^*)$-ისათვის საპროგნოზო ინტერვალების აგების მეშვეობით.

ცხადია, რომ $\hat{Y}(x^*) = b_0 + b_1 x^*$ შემთხვევითი სიდიდეა, ამასთან

1. $E(b_0 + b_1 x^*) = B_0 + B_1 x^*$;

2. $$D(b_0 + b_1 x^*) = \sigma^2_{b_0 + b_1 x^*} = \sigma^2 \left[\frac{1}{n} + \frac{n(x^* - \bar{x})^2}{n \sum x_i^2 - (\sum x_i)^2} \right];$$

3. $b_0 + b_1 x^*$ შემთხევევითი სიდიდე ნორმალურადაა განაწილებული, საიდანაც ვასკვნით, რომ

$$T = \frac{b_0 + b_1 x^* - (B_0 + B_1 x^*)}{S_{b_0 + b_1 x^*}} = \frac{\hat{Y}(x^*) - \mu^*}{S_{b_0 + b_1 x^*}}$$

სტატისტიკას აქვს სტიუდენტის განაწილება n-2 თავისუფლების ხარისხით, სადაც $S^2_{b_0 + b_1 x^*}$ წარმოადგენს $\sigma^2_{b_0 + b_1 x^*}$ -ის შეფასებას

$$S^2_{b_{0+b_1 x^*}} = S^2\left[\frac{1}{n} + \frac{n(x^* - \bar{x})^2}{n\sum x_i^2 - (\sum x_i)^2}\right]$$

ყოველივე ზემოთქმულიდან გამომდინარე $1-\alpha$ ნდობის მქონე ნდობის ინტერვალს პირობითი მათემატიკური ლოდინისთვის აქვს სახე:

$$\hat{Y}(x^*) - t_{n-2,\frac{\alpha}{2}} S\sqrt{\frac{1}{n} + \frac{(x^*-\bar{x})^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}} < E(\mu^*) < \hat{Y}(x^*) + t_{n-2,\frac{\alpha}{2}} S\sqrt{\frac{1}{n} + \frac{(x^*-\bar{x})^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}}\,. \qquad (1.41)$$

ანალოგიურად მსჯელობიდან შეიძლება დავასკვნათ, რომ $1-\alpha$ ნდობის მქონე საპროგნოზო ინტერვალს $\hat{Y}(x^*)$ -ისათვის აქვს შემდეგი სახე

$$\hat{Y}(x^*) - t_{n-2,\frac{\alpha}{2}} S\sqrt{\frac{1}{n} + \frac{(x^*-\bar{x})^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} + 1} < Y(x^*) < \hat{Y}(x^*) + t_{n-2,\frac{\alpha}{2}} S\sqrt{\frac{1}{n} + \frac{(x^*-\bar{x})^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} + 1}\,. \quad (1.42)$$

ეს ფორმულა განსხვავდება (1.41)-საგან 1-ის ტოლი დამატებითი შესაკრებით რადიკალების ქვეშ. ეს აიხსნება გაბნევის დამატებითი წყაროთი $Y(x^*)$-ისათვის. ესაა $Y(x^*)$-ის გაბნევა $E(Y|x^*)$-ის მიმართ.

შევნიშნოთ, რომ $S^2_{b_{0+b_1 x}}$ უმცირესია მაშინ, როცა $x = \bar{x}$ და იზრდება, როდესაც x შორდება $\bar{x}$-ს ნებისმიერი მიმართულებით.ეს იწვევს იმას, რომ ნდობისა და საპროგნოზო ინტერვალები ყველაზე ვიწროა, როდესაც $x = \bar{x}$.

ნაშთთა ანალიზი. რეგრესიული ანალიზის ბოლო სტადიაზე უნდა ჩატარდეს ნაშთთა ანალიზი. ნაშთთა (ანუ ნარჩენი სხვაობა) ჩვენ განვმარტეთ შემდეგი ფორმულის დახმარებით:

$$e_i = Y_i - \hat{Y}_i\,,$$

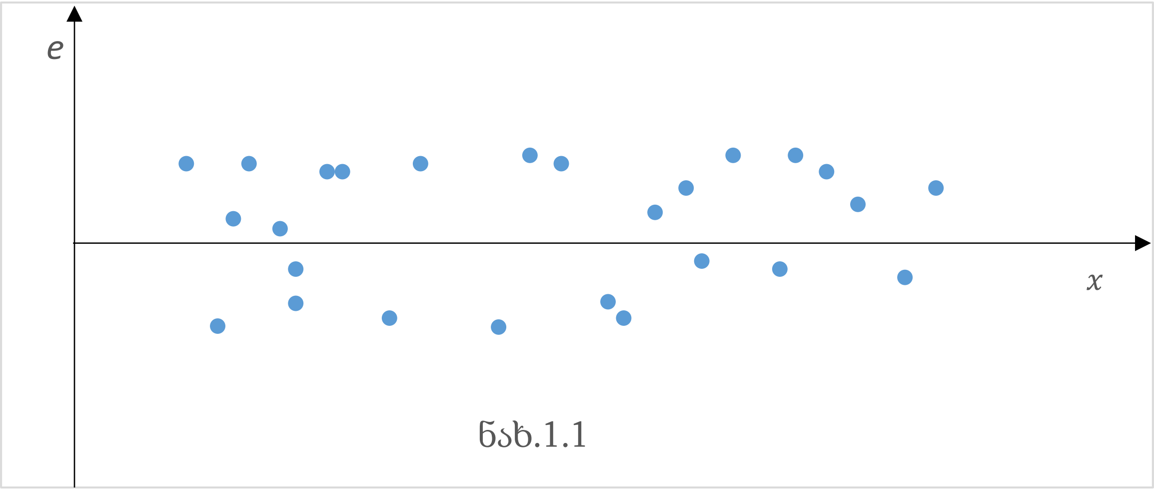
ეს არის სხვაობა Y_iმნიშვნელობას და რეგრესიის ჩვენს მიერ აგებულ წრფის მიხედვით გამოთვლილ $\hat{Y}_i$ მნიშვნელობას შორის. ასეთი ნაშთი უნდა გამოვთვალოთ i-ს ყველა მნიშვნელობისათვის, მივიღებთ $(e_1, e_2, \dots, e_n)$ ნაშთთა მიმდევრობას ყველა შერჩევითი წერტილისათვის.

ნაშშთა ანალიზის ამოცანაა შემოწმდეს ის ძირითადი დაშვებები, რომელთა საფუძველზეც იყო განხორციელებული გამოთვლები.

ჩამოვთვალოთ ძირითადი დაშვებები

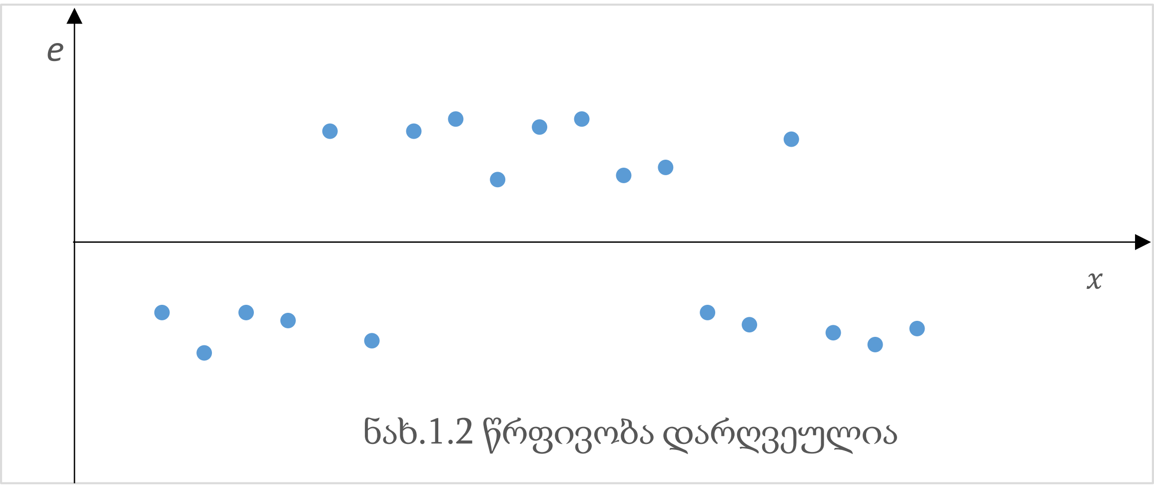
- რეგრესიის მრუდი წრფეა;
- გადახრათა (ε_i) დისპერსიები მუდმივია ყველა შერჩევითი წერტილისათვის;
- გადახრები დამოუკიდებელია შერჩევის ელემენტის ნომერზე;
- გადახრები განაწილებულია ნორმალურად.

ნაშთა ანალიზი არსებითად ხორციელდება გრაფიკულად. x_i-ს მნიშვნელობათა მიხედვით გრაფიკზე დავსვამთ e_i წერტილებს. თუ მიღებულ წერტილთა ერთობლიობა დაახლოებით ერთნაირადაა განაწილებული Ox ღერძის გასწვრივ (თანაბრად წრფის ზემოთ და ქვემოთ), უნდა ჩავთვალოთ, რომ ყველა დაშვება ასე თუ ისე, შესრულებულია (იხ.ნახ.1.1).



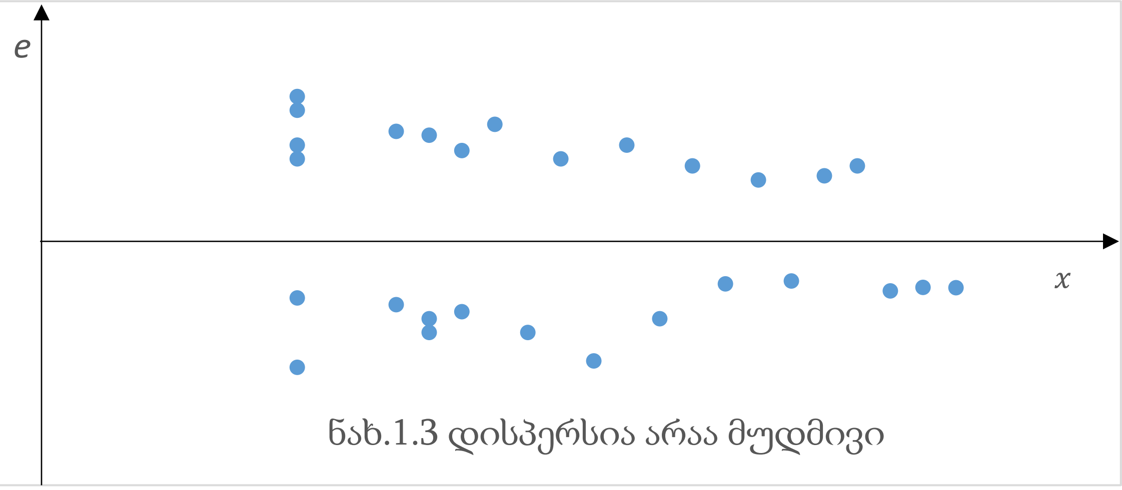
ნახ.1.1

თუ დარღვეულია რეგრესიის მრუდის წრფივობის დაშვება, მაშინ ნაშთა გრაფიკს დაახლოებით ასეთი სახე ექნება, როგორიც ეს მოცემულია ნახ 1.2-ზე



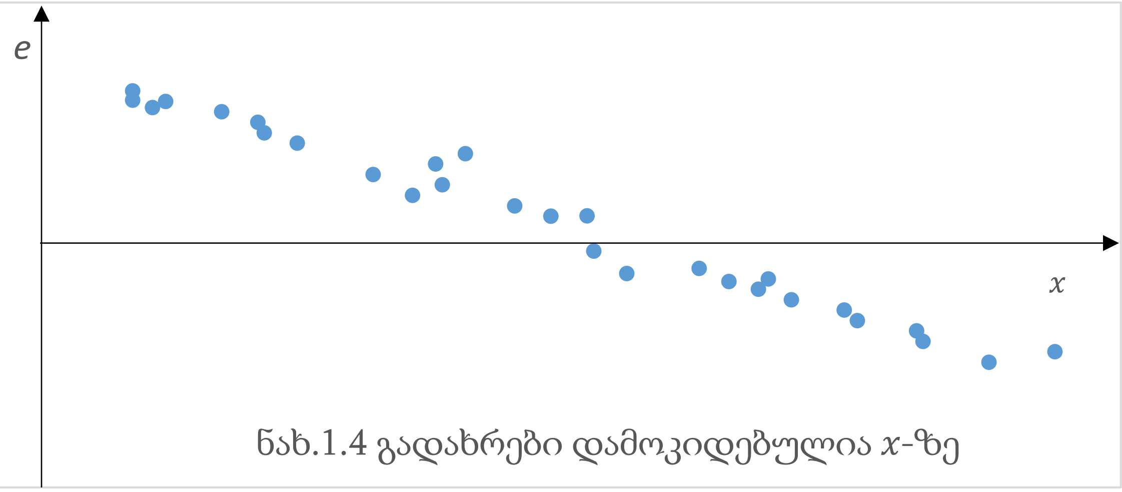
ნახ.1.2 წრფივობა დარღვეულია

თუ დარღვეულია დაშვება σ^2-ის მუდმივობის შესახებ, მაშინ გვექნება დაახლოებით ასეთი სურათი, როგორც მოცემულია ნახ. 1.3-ზე



ნახ.1.3 დისპერსია არაა მუდმივი

თუ გადახრები დამოკიდებულია x_i-ზე მაშინ გვქნება დაახლოებით შემდეგი სურათი (იხ.ნახ.1.4)



ნახ.1.4 გადახრები დამოკიდებულია x-ზე

გადახრათა ნორმალურობა მოწმდება ნორმალური განაწილების მიხედვით დახაზული სპეციალური ქაღალდის საშუალებით.

პოლინომური რეგრესია. ამოცანებში ხშირად დამოკიდებულება ორ X და Y ცვლადებს შორის არაა წრფივი, არამედ უფრო რთული ხასიათისაა. მაგალითად, თეორიული რეგრესიის მრუდს Y-ისა X-ზე შეიძლება ჰქონდეს სახე

$$E(Y|X = x) = B_0 + B_1x + B_2x^2 \,, \tag{1.13}$$

(მეორე ხარისხის პარაბოლა) ან

$$E(Y|X = x) = B_0 + B_1 x + \cdots + B_m x^m \,, \tag{1.14}$$

(m-ური რიგის პოლინომი), სადაც m არც თუ ისე დიდი ნატურალური რიცხვია.პრაქტიკულად $m \leq 4$.

სიმარტივისვის განვიხილოთ (1.13) შემთხვევა. აქ ჩვენ გვაქვს სამი უცნობი

B_0, B_1, B_2.

თუ მიღებული შერჩევა $(x_1, Y_1), (x_2, Y_2), \dots, (x_n, Y_n)$, მაშინ შეიძლება ჩაიწეროს ასეთი მათემატიკური მოდელი:

$$Y_i = B_0 + B_1 x + B_2 x^2 + \varepsilon_i; \tag{1.15}$$

ε_i-ის მიმართ დაშვება ანალოგიურია იმისა, რაც გვქონდა წრფივი რეგრესიის შემთხვევაში.

რა თქმა უნდა, B_0, B_1, B_2-ის განსასაზღვრად შეიძლება გამოვიყენოთ უმცირეს კვადრატთა მეთოდი. არსებითად აქ ერთი ამხსნელი ცვლადია, მაგრამ უცნობთა რაოდენობა 3, x^2 თამაშობს ახალი ამხსნელი ცვლადის როლს. ასე, რომ არსებითად საქმე გვაქვს 2 ამხსნელ ცვლადთან, ე.ი. პოლინომური რეგრესია არ არის მარტივი. B_0, B_1, B_2 კოეფიციენტების შესაფასებლად გამოდგება ის მიდგომა, რაც მრავლობითი რეგრესიის ამოცანაში გვაქვს.

დასკვნები

რეგრესია და კორელაცია შეისწავლის ორ ცვლადს შორის სტოქასტურ კავშირის მნიშვნელოვან ასპექტებს. კორელაციის კოეფიციენტი რიცხვია, რომელიც ახასიათებს წრფივი კავშირის სიმჭიდროვეს. მარტივი, წრფივი რეგრესია იძლევა ორ ცვლადს შორის სტოქასტური კავშირის წრფივ აპროქსიმაციას. შერჩევითი მონაცემებით რეგრესიის საუკეთესო წრფის განსასაზღვრავად გამოიყენება უმცირეს კვადრატთა მეთოდი. იგი მდგომარეობს კოეფიციენტების ისეთ მნიშვნელობათა განსაზღვრაში, რომლებიც უზრუნველყოფენ გადახრათა კვადრატების ჯამის მინიმალურობას. ამ პრინციპის გამოყენებით მიიღება ნორმალურ განტოლებათა სისტემა, რომლის ამოხსნები წარმოადგენს კოეფიციენტთა საძიებელ მნიშვნელობებს. გამოითვლება რეგრესიის წრფის კოეფიციენტთა დისპერსიების შეფასებები. მნიშვნელოვანია რეგრესიის კოეფიცენტი, რომელიც წარმოადგენს რეგრესიის წრფის საკუთხო კოეფიციენტს.

მოწმდება რეგრესიის კოეფიციენტის ნულთან ტოლობის ჰიოთეზა.

ტექსტში მოცემული ფორმულებით აიგება ნდობის ინტერვალები პირობითი მათემატიკური ლოდინის თეორიული მნიშვნელობისათვის და Y ცვლადის კონკრეტული თეორიული მნიშვნელობისათვის.

თავი II

დროითი მწკრივები

დაკვირვებები რაიმე მოვლენაზე, რომლის ხასიათიც დროში ცვალებადია, წარმოქმნის დალაგებულ მიმდებრობას, რომელსაც დროით მწკრივს უწოდებენ.

ესა თუ ის მოვლენა ხასიათდება გარკვეული მაჩვენებლით, რომელიც როგორც წესი ჩვენს მიერ ინტერპრეტირებული იქნება როგორც შემთხვევითი სიდიდე.

ამრიგად, დროის დისკრეტულ $t_1, t_2, \dots, t_n, \dots$ მომენტებში დაკვირვებებით წარმოქმნილი დროითი მწკრივი, ჩვენ გვესმის როგორც დამოკიდებულ შემთხვევით სიდიდეთა მიმდევრობა

$$Y(t_1), Y(t_2), \dots, Y(t_n), \dots$$

უნდა ავღნიშნოთ, რომ დროითი მწკრივის მთავარი დამახასიათებელი ნიშანი ის არის, რომ მნიშვნელოვანია დაკვირვებათა მიმდევრობა, ე.ი. დროში დალაგება, განსხვავებით შემთხვევითი ამოკრეფისაგან $X_1, X_2, \dots, X_n, \dots$ სადაც ინდექსისაცია მხოლოდ პირობითია და გადანომვრა პირობის არს არ ცვლის.

ჩვენ განვიხილავთ მხოლოდ იმ შემთხვევას, როდესაც დაკვირვებები ხდება ერთმანეთისაგან თანაბრად დაშორებულ დროის მომენტებში და ჩვენ დროის ერთეულად ავირჩევთ ინტერვალს ორ მეზობელ მომენტს შორის, მაშინ მწკრივის წევრები შეიძლება აღინიშნოს სიმბოლოებით

$$Y_1, Y_2, \dots, Y_n, \dots,$$

მოვიყვანოთ დროითი მწკრივის რამდენიმე მაგალითი ადამიანის მოღვაწეობის სხვადასხვა სფეროდან. ესენია: საქონლის მოხმარების მონაცემები წლების მანძილზე, ერთობლივი ეროვნული პროდუქტის ცვლილება წლების მიხედვით, მრავალწლიანი დაკვირვებები ტემპერატურაზე, სხვადახსვა სამეურნეო კულტურის მოსავლიანობა წლების მიხედვით, სხვადასხვა კომპანიების პროდუქციაზე მოთხოვნის ცვლილება, ინდივიდუალური მოხმარება წლების მიხედვით, გადაზიდული ტვირთის რაოდენობა და სხვა.

შევეხოთ ერთ საკითხს დროითი მწკრივის დისკრეტულობის ტიპის შესახებ.

1. დროითი მწკრივის მაგალითია რაიმე სასოფლო-სამეურნეო კულტურის მოსავლიანობა ფართობის ერთეულზე წლების მიხედვით. მართლაც მოცემულ ფართობის ერთეულზე მოსავლის აღება ხდება წელიწადში ერთხელ, დაახლოებით ერთსა და იმავე დროს. ასეთივე ბუნებრივი დისკრეტულობის მაგალითია წლიური მთლიანი შიდა პროდუქტი და ა.შ.

2. მოსახლეობის რაოდენობის ცვლილების შესწავლისას მოსახლეობის აღწერა ხდება გარკვეული პერიოდულობით და არა უწყვეტად.

3. დისკრეტული დროითი მწკრივი შეიძლება მიღებული იქნას დროში უწყვეტად მიმდინარე პროცესებზე დისკრეტულ მომენტებში დაკვირვებებით. ასეთი დისკრეტული დროითი მწკრივის მაგალითად შეიძლება მოვიყვანოთ ბანკის მცურავ საპროცენტო განაკვეთზე დისკრეტული დროის მომენტებში დაკვირვებები. მეორე მაგალითია აქციის ფასზე გარკვეული დროის მომენტებში (ყველდღიური, ყოველკვირეული) დაკვირვებები

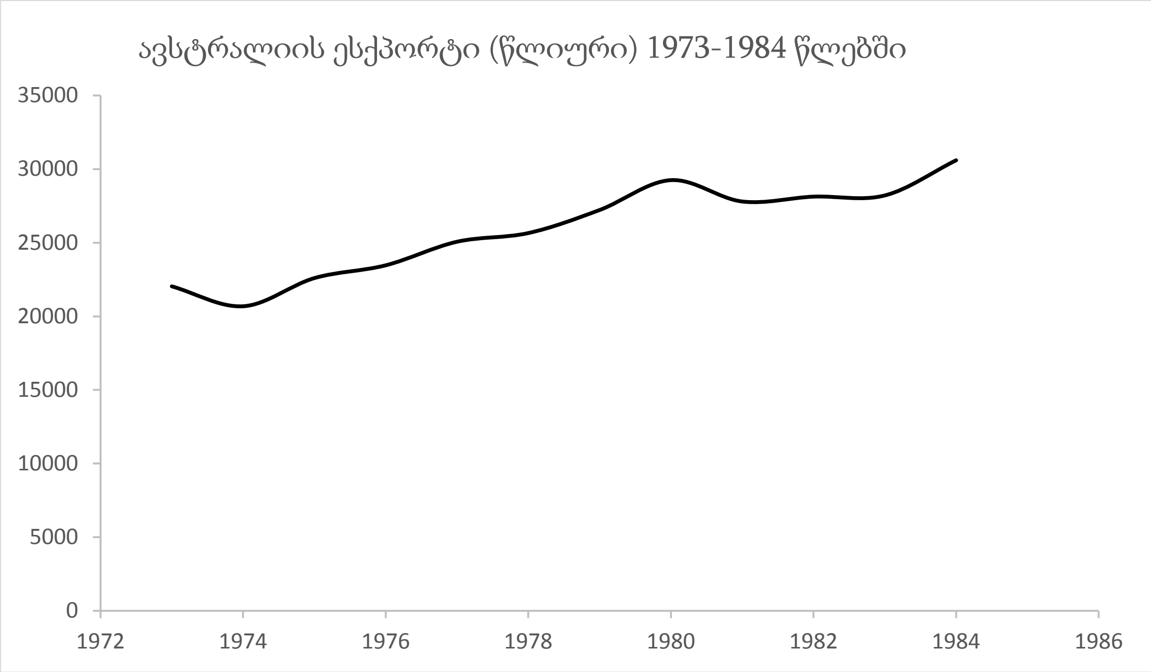
4. დისკრეტული დროითი მწკრივის გენერირების ერთ-ერთი წყაროა ცვლადი სიდიდის მნიშვნელობის დაგროვება დროის გარკვეული პერიოდის განმავლობაში. ასეთი დროის მწკრივების მაგალითებია: ა)ნალექების რაოდენობა, რომლებიც გროვდება ისეთი პერიოდების განმავლობაში, როგორებიცაა დღე ან თვე; ბ) გარკვეული პერიოდის განმავლობაში წარმოებული ესა თუ ის პროდუქცია; გ) გარკვეულ პროდუქციაზე თვიური, წლიური ან სხვა პერიოდის მოთხოვნა.

პრაქტიკაში, განსაკუთრებით ეკონომიკაში დროის ის მომენტები, როდესაც უნდა ჩატარდეს დაკვირვებები, ხშირად წინასწ არ არის მოცემული. ისეთ სიტუაციებში კი რომელსაც ექსპერიმენტული ხასიათი გააჩნია გადაწყვეტილება დაკვირვებების მომენტების შერჩევის შესახებ ჩვენ თვითონვე შეგვიძლია მივიღოთ. ეს განსაკუთრებით მნიშვნელოვანია ეკონომიკური მაჩვენებლების შესწავლისას.

დროითი მწკრივების წარმოდგენა მოხერხებულია როგორც ცხრილების, ისევე გრაფიკების მეშვეობით.

ცხრილი 2.1. ავსტრალიის ესქპორტი (წლიური) 1973-1984 წლებში

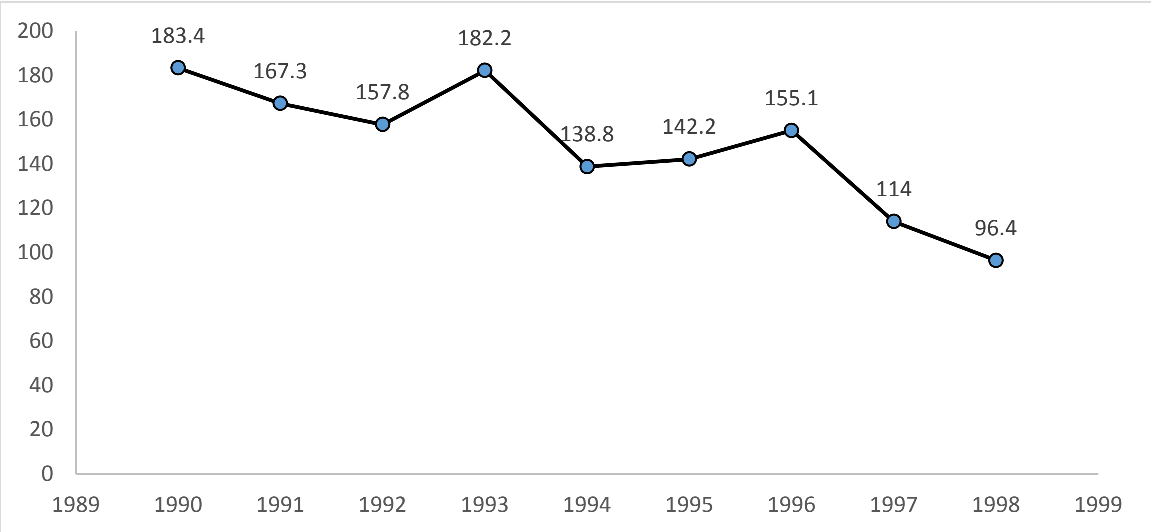
წელი	t	ექსპორტი ($ მილიონებში)	წელი	t	ექსპორტი ($ მილიონებში)
1973	1	22040	1979	7	27225
1974	2	20686	1980	8	29256
1975	3	22600	1981	9	27804
1976	4	23468	1982	10	28135
1977	5	25070	1983	11	28216
1998	6	25659	1984	12	30605



ნახ .1.1. ცხრილი 2.2-ის შესაბამისი გრაფიკი

ცხრილ 2.2. თბილისის მეტროპოლიტენით მგზავრთა გადაყვანა 1990-1998 წლებში (მლნ.კაცი)

1990	1991	1992	1993	1994	1995	1996	1997	1998
183.4	167.3	157.8	182.2	138.8	142.2	155.1	114.0	96.4



ნახ .1.2. ცხრილი 2.2-ის შესაბამისი გრაფიკი

დროითი მწკრივის ანალიზის მიზანია იმ მექანიზმების დადგენა, რომლებიც განაპირობებს ამ მწკრივების წარმოქმნას. ამის მიღწევა შეუძლებელია ცალკეული (ერთეული) დროითი მწკრივის ანალიზის საფუძველზე, რადგანაც ეს უკანასკნელი შეიძლება დახასიათდეს რამდენიმე დროითი მწკრივით, რომლებიც ურთიერკავშირში იმყოფება. სრული ანალიზის ჩატარებისას და იქედან გამომდინარე დასკვნების მისაღებად საჭიროა მრავალგანზომილებიანი სისტემების შესწავლა. აქ ჩვენ შემოვიფარგლებით მხოლოდ ცალკეულუი (ერთგანზომილებიანი) დროითი მწკრივების ქცევის სხვადასხვა ტიპების შესწავლით.

დროითი მწკრივების ანალიზის მრავალი მოდელიდან ჩვენ თავდაპირველად შევჩერდებით ეკონომიკასა და მათემატიკაში ფართოდ გამოყენებული კლასიკურ მოდელზე, რომელიც გულისხმობს დროითი მწკრივის ოთხი ძირითადი ურთიერთდაკავშირებული კომპონენტების სახით წარმოდგენას.

ეს კომპონენტებია:

1. გრძელვადიანი ტრენდი, ანუ სისტემური ძრაობა - T_1;

2 . ციკლური ეფექტი - მეტნაკლებად რეგულარული ხასიათის რხევითი მოძრაობა ტრენდის მიმართ -C_t;

3. სეზონურობის ეფექტი S_t ;

4. შემთხვევითი ანუ არარეგულარული, არასისტემური კომპონენტა -I_t ;

დავახასიათოთ თითოეული კომპონენტა ცალ-ცაკლე.

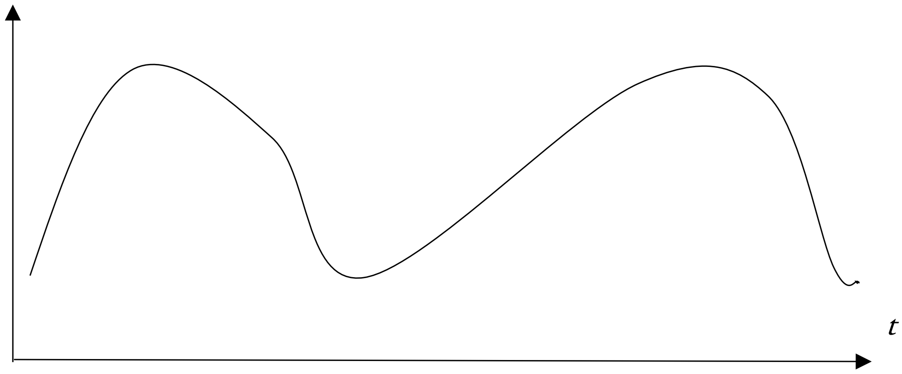
ტრენდი. ტრენდის განსაზღვრა საკმაოდ რთული პრობლემაა. საზოგადოდ, ტრენდის ქვეშ იგულისხმება რაიმე მდგრადი სისტემური ძრაობა საკმაოდ ხანგრძლივი პერიოდის გამნავლობაში. სხვა სიტყვებით რომ ვთქვათ, ტრენდი წარმოადგენს გრძელვადიან (ხანგრძლივ) შედარებით გლუვი ხასიათს ძრაობას, რომელსაც ავლენს დროითი მწკრივი. ავღნიშნოთ რომ ცნება „ხანგრძლივი“ პირობითია და დამოკიდებულია შესასწავლი მოვლენის არსზე. ის, რაც „ხანგრძლივია“ ერთი თვალსაზრისით, შეიძლება ხანმოკლე აღმოჩნდეს სხვა თვალსაზრისით. მაგალითად, თუ ჩვენ შევისწავლით ვალუტის კურსის ცვლილებას რამდენიმე (ორი-სამი წლის განმავლობაში, ის მდგრადი ძრაობა, რომელიც შეიძლება აღქმული იყოს როგორც ტრენდი, სინამდვილეში შეიძლება აღმოჩნდეს მხოლოდ ნაწილი ნელად ხრევადი პროცესისა, რომელიც აღწერს ვალუტის კურსის ცვლილებას ათეული წლების განმავლობაში.

ციკლური ეფექტი. ციკლური ძრაობა დროით მწკრივში შეიძლება დახასიათდეს როგორც ფართო ხრევა ანუ ტალღისებრი ძრაობა ტრენდის წირის (ხაზის) მიმართ ზევით და ქვემოთ. ამასთან თითოეული ციკლის ხანგრძლივობა, როგორც წესი, წელიწადზე მეტია, შეიძლება იყოს რამდენიმე წელიც.

ციკლურობის მაგალითებია კარგად ცნობილი საქმიანი ციკლები, რომლებიც შეესაბამება (ანდა იმეორებს) ეკონომიკური რეაქციებისა და ინფლაციების ციკლებს, საქონელზე გრძელვადიანი მოთხოვნის ციკლებს ფინანსურ და ფულად სექტორებში.

ამ ბოლო ციკლის მაგალითად შეიძლება დავასახელოთ საპროცენტო განაკვეთის ცვლილების კარგად ცნობილი ციკლები.

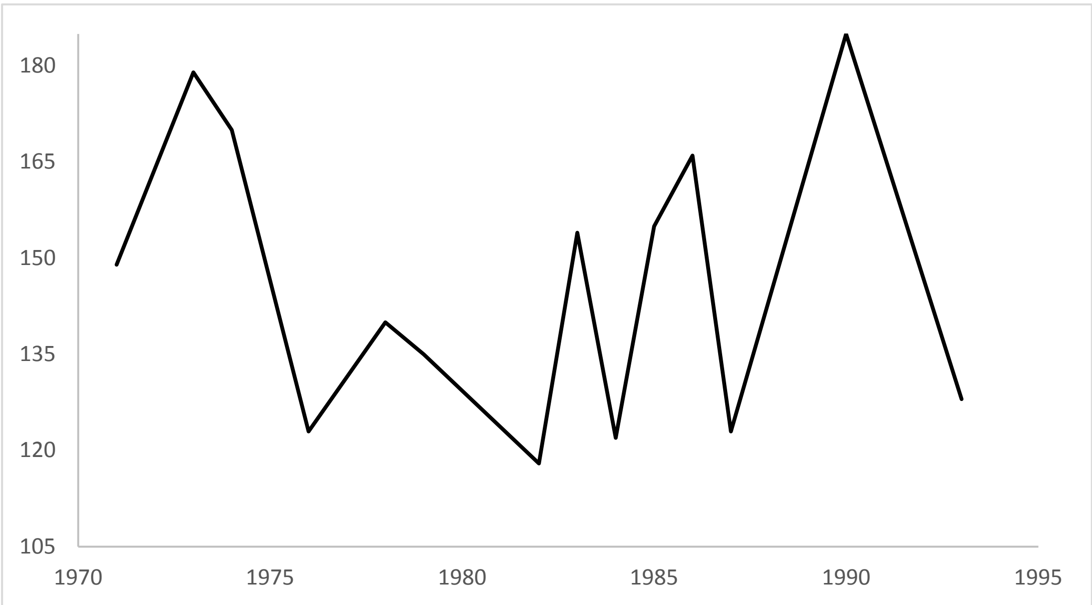
ქვემოთ მოყვანილ ნახაზზე ნაჩვენებია დროით მწკრივში ციკლური მრაობის ერთ-ერთი მაგალითი.



ნახ.2.3

სამწუხაროდ, პრაქტიკაში ციკლები არ ატარებს ისეთ რეგულარულ ხასიათს, როგორც ეს ნახაზზეა ნაჩვენები. ისინი ძალიან იშვიათად არიან რეგულარული და ჩნდებიან სხვა კომპონენტებთან ერთობლიობაში.

მოვიყვანოთ არარეგულარული ციკლური მოძრაობის კიდევ ერთი მაგალითი.



ნახ.2.4. არარეგულარული ციკლური მოძრაობა

ციკლური მწკრივების გამოყოფა დროითი მწკრივების ერთ-ერთი ურთულესი ამოცანაა.

სეზონური ეფექტი (სეზონური ვარიაციები). შეიძლება ითქვას, რომ ყველაზე ადვილი აღმოსაჩენად, გამოსაყოფად და შესასწავლად არის სეზონური ეფექი. ეს გამოწვეული იმით, რომ სეზონური ცვლილებები განპირობებულია შესასწავლ სისტემაზე გარე მექანიზმების და არა ძირითადი მექანიზმების ზემოქმედებით.

სეზონური ვარიაციები ციკლურის მსგავსია. განსხვავება მათ შორის ის არის, რომ სეზონურების პერიოდები უფრო ხანმოკლეა. მათი ხანგრძლივობა ერთი წელია ან ერთ წელზე ნაკლებია (კვარტალი, თვე, კვირა, დღე და სხვა). სეზონური ვარიაცია შეიძლება დაკავშირებული იყოს წელიწადის ოთხ სეზონთან ამ ტრენდის მიმართ სისტემური ხასიათის ძრაობასთან, რომელიც წარმოიქმნება ერთი კვირის პერიოდით ან უფრო მეტიც ერთი დღის პერიოდით. მაგალითად, აქციათა ბაზარზე აქციების ფასების დღის განმავლობაში დროის გარკვეულ მომენტებში აღწევს უმაღლეს და უდაბლეს მნიშვნელობებს. სეზონური ვარიაციის კლასიკური მაგალითია ბინათმფლობელთა მიერ საწვავის (ნავთობის ან ბუნებრივი გაზის) მოხმარება.

შემთხვევითი ვარიაციები, ანუ შემთხვევითი კომპონენტა მოიცავს ყველა იმ არარეგულარულ ცვლილებას დროით მწკრივში, რომლებიც არაა განპირობებული სხვა კომპონენტებით. ისინი აძნელებენ სხვა უფრო გაჭვრეტადი კომპონენტების გამოვლენას. თითქმის ყველა დროითი მწკრივი შეიცავს შემთხვევით იმფორმაციებს. საჭიროა ისეთი მეთოდების ფლობა, რომლებიც საშუალებას მოგვცემდა მოგვეშორებინა ისინი დანარჩენი კომპონენტების აღწერისა და აღრიცხვის მიზნით.

დროითი მწკრივის კლასიკური მოდელი

მას შემდეგ, რაც დავახასიათეთ დროითი მწკრივის კლასიკური მოდელის ძირითადი შემადგენელი კომპონენტები: ტრენდი - T_t ; ციკლური ეფექტი - C_t ; სეზონურობის ეფექტი - S_t ; არარეგულარული შემთხვევითი კომპონენტა - I_t , შევეხოთ საკითხს, თუ ამ ოთხი კომპონენტის როგორ კომბინაციად შეიძლება იქნას წარმოდეგნილი ეს მოდელი.

მათემატიკაში ფართოდ გავრცელებული მოდელებია ან ადიციური მოდელი, რომლის ფარგლებში დროითი მწკრივის მნიშვნელობა t-ურ მომენტში (აღნიშვნა Y_t) განსაზღვრულია როგორც ამ კომპონენტების ჯამი

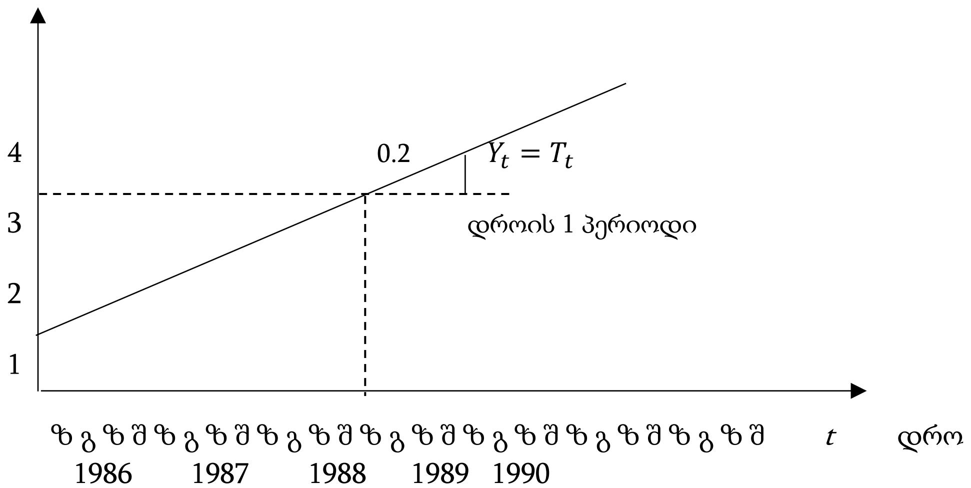
$$Y_t = T_t + C_t + S_t + R_t$$

ანდა მულტიპლიკაციური მოდელი, რომლის ფარგლებში Y_t-ს გენერირება ხდება კომპონენტების გადამრავლებით

$$Y_t = T_t * C_t * S_t * R_t$$

მულტიპლიკაციური მოდელი განსაკუთრებით მოსახერხებელია დროითი მწკრივის ანალიზისათვის, მისი კომპონენტებად დაშლისათვის და შემდგომში ჩვენი განხილვა შეეხაბა მხოლოდ ამ მოდელს.

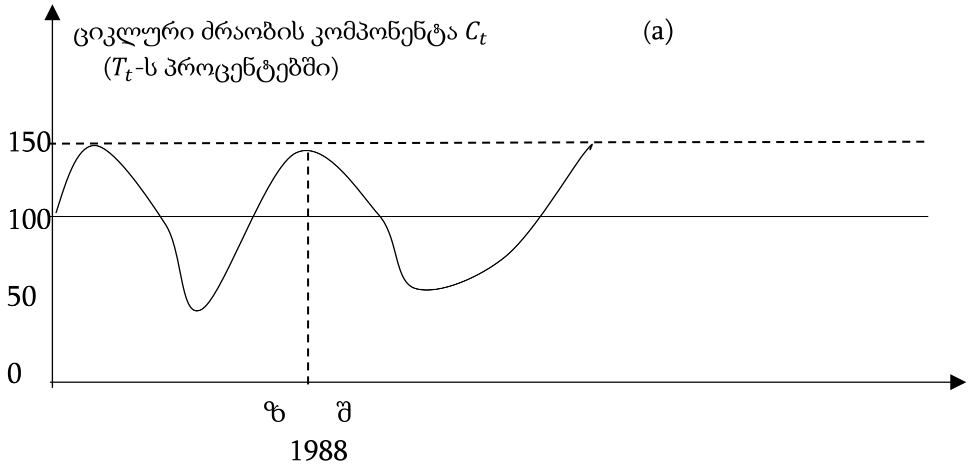
ავხსნათ ეს მოდელი ხელოვნური დროითი მწკრივის აგებით იმ პირობით, რომ მისი კომპონენტები მოცემულია. მილიონებში გამოხატული პირობითი საცალო საქონლის გაყიდვის მოცულობისათვის.

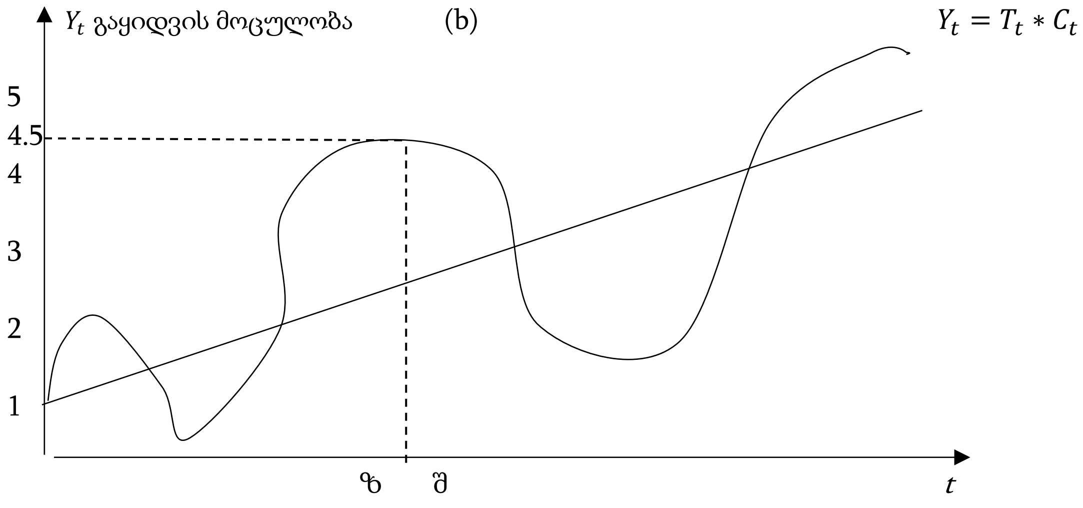


ნახ.2.5.ტრენდის ხაზი (გზშზ-გაზაფხული,ზაფხული, შემოდგომა,ზამთარი)

პირველ რიგში შევნიშნოთ, რომ ტრენდის ხაზი წრფეა (წრფივი ტრენდი). ეს წრფე მიგვითითებს, რომ 1986 წლის პირველ კვარტლის დასაწყისში გაყიდული საქონლის რაოდენობა იყო ერთი მილიონი და ყოველ მომდევნო კვარტალში გაყიდული საქონლის რაოდენობა იზდება 0.2 მილიონით და 1988 წლის ზაფხულისათვის გაიზარდა სამი მილიონამდე.

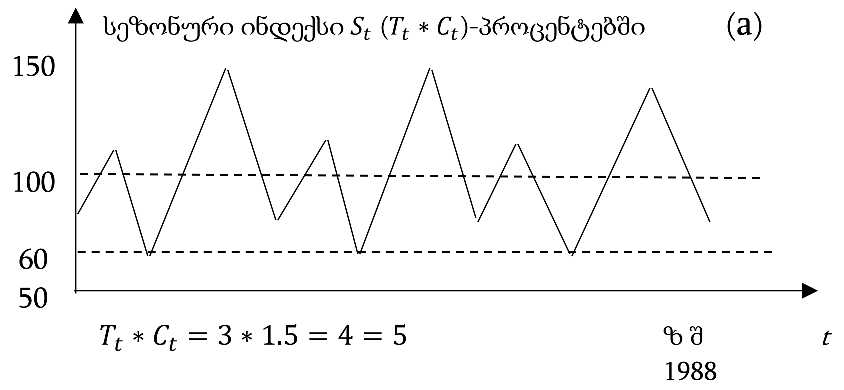
ახლა განვიხილოთ ციკლური ძრაობის C_t-ს ზემოქმედება)გავლენა) დროით მწკრივზე. ეს კომპონენტა აღწერს პროდუქციის გაყიდვის მოცულობის რხევას ტრენდის ხაზის მიმართ აწევას კარგ წლებში და დაწევას უარს წლებში. ციკლურ ეფექტს გამოვხატათ ტრენდის პროცენტებში. მაგალითად $C_t = 150\%$ ნიშნავს, რომ $Y_t = T_t * 1.5$, $C_t = 50\%$ იმას, რომ $Y_t = T_t * 0.5$, ხოლო $C_t = 100\%$ ნიშნავს, რომ დროითი მწკრივის მნიშვნელობა ემთხვევა ტრენდის მნიშვნელობას $Y_t = T_t$ (შევნიშნოთ, რომ ამ ნაბიჯზე იგულისხმება, რომ დროითი მწკრივი მხოლოდ ამ ორ კომპონენტისაგან შედგება).

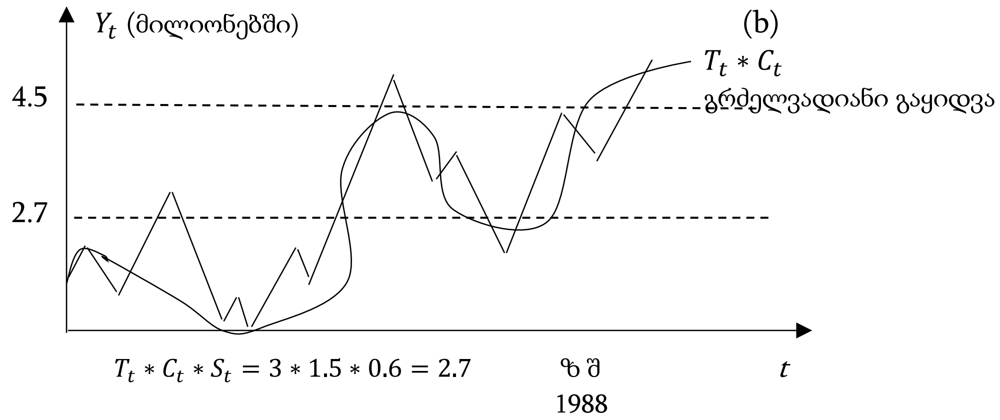




ნახ.2.6. (a) ციკლური კომპონენტა და (b) ტრენდისა და ციკლური კომპონენტის კომბინაცია 1988 წლის ზაფხულისათვისა $C_t = 150\%$, $T_t = 3$ (მილიონი),$Y_t = 3 * 1.5$=4.5

ახლა გავითვალისწინოთ სეზონური ეფექტი.სეზონური ეფექტი არის ტალღისებრი ძრაობა ტრენდისა და ციკლური კომონენტის კომბინირებით მიღებული დონის მიმართ. გაყიდვის მოცულობა ამ დონის ზევითაა აქტიური სეზონის განმავლობაში და ეშვება ამ დონეზე დაბლა პასიური სეზონის განმავლობაში. თუ ფორმალურად ტრენდისა და ციკლური კომპონენტის კომბინირებით მიღებულ დონეს ვუწოდებთ „ნორმალურს“, მაშინ სეზონური ეფექტი გამოისახება „ნორმალური“ დონის პროცენტებში და მას ეწოდება სეზონური ინდექსი S_t. მაგალითად $S_t = 60\%$ ნიშნავს $Y_t = T_t * C_t * 0.6$. და ა.შ.

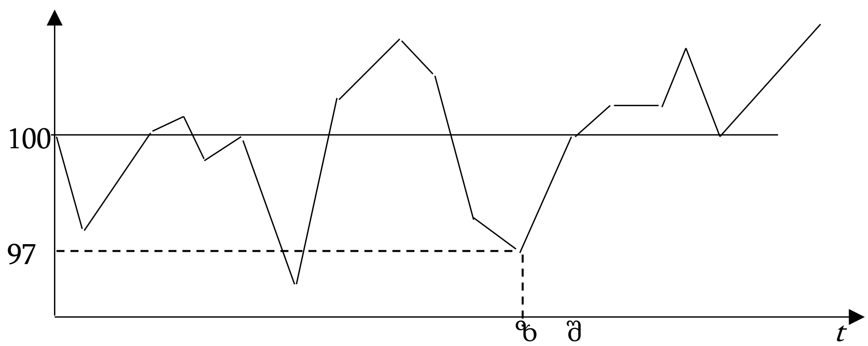




ნახ.2.7. (a) სეზონური ინდექსი ხელოვნური დროითი მწკრივისათვის (b) კომბინირებული ტრენდი, ციკლური და სეზონური კომპონენტები

დაგვჩა მეოთხე - შემთხვევითი ანუ არარეგულარული კომპონენტის ჩართვა ხელოვნურ დროითი მწკრივის კომბინირების პროცესში. თუ კვლავ, „ნორმალურ" დონედ ჩავთვლით წინა ეტაზე აგებულ მწკრივს, მაშინ არერეგულარული კომპონენტა R_t გამოისახება ამ დონის პროცესებში. ერთ-ერთი შესაძლო რეალიზაციაა ასეთი კომპონენტისა მოცემულია ქვემოთ მოყვანილ ნახაზზე. მაგალითად, თუ $R_t = 97\%$, გვექნება $Y_t = T_t * C_t * S_t * 0.97$, $R_t = 50\%$ ნიშნავს, რომ $Y_t = T_t * C_t * S_t * 0.5$. მაგალითად, 1988 წლის ზაფხულისათვის $T_t = 3, C_t = 150\%, S_t = 60\%, R_t = 97\%$. ამიტომ :

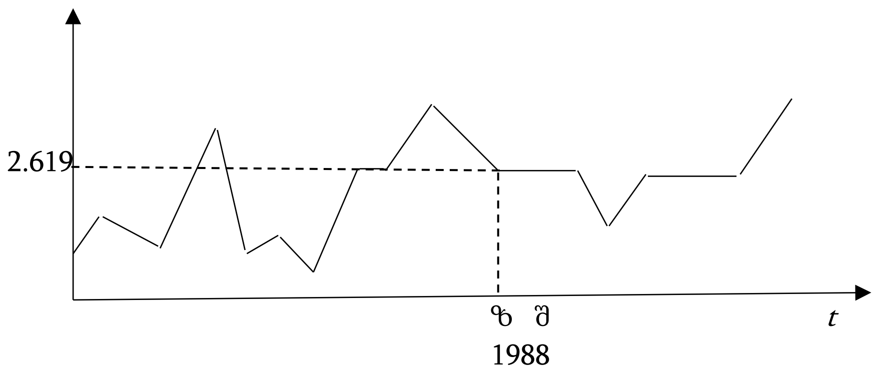
$$Y_t = 3 * 1.5 * 0.6 * 0.97 = 1.619.$$



ნახ.2.8.არარეგულარული (შემთხვევითი) ფლუქტუაციის კომპონენტა

საბოლოოდ,სრულ დროით მწკრივს ექნება შემდეგი სახე:

$$Y_t = T_t * C_t * S_t * 0.97 * R_t = 3 * 1.5 * 0.6 * 0.97 = 2.619$$



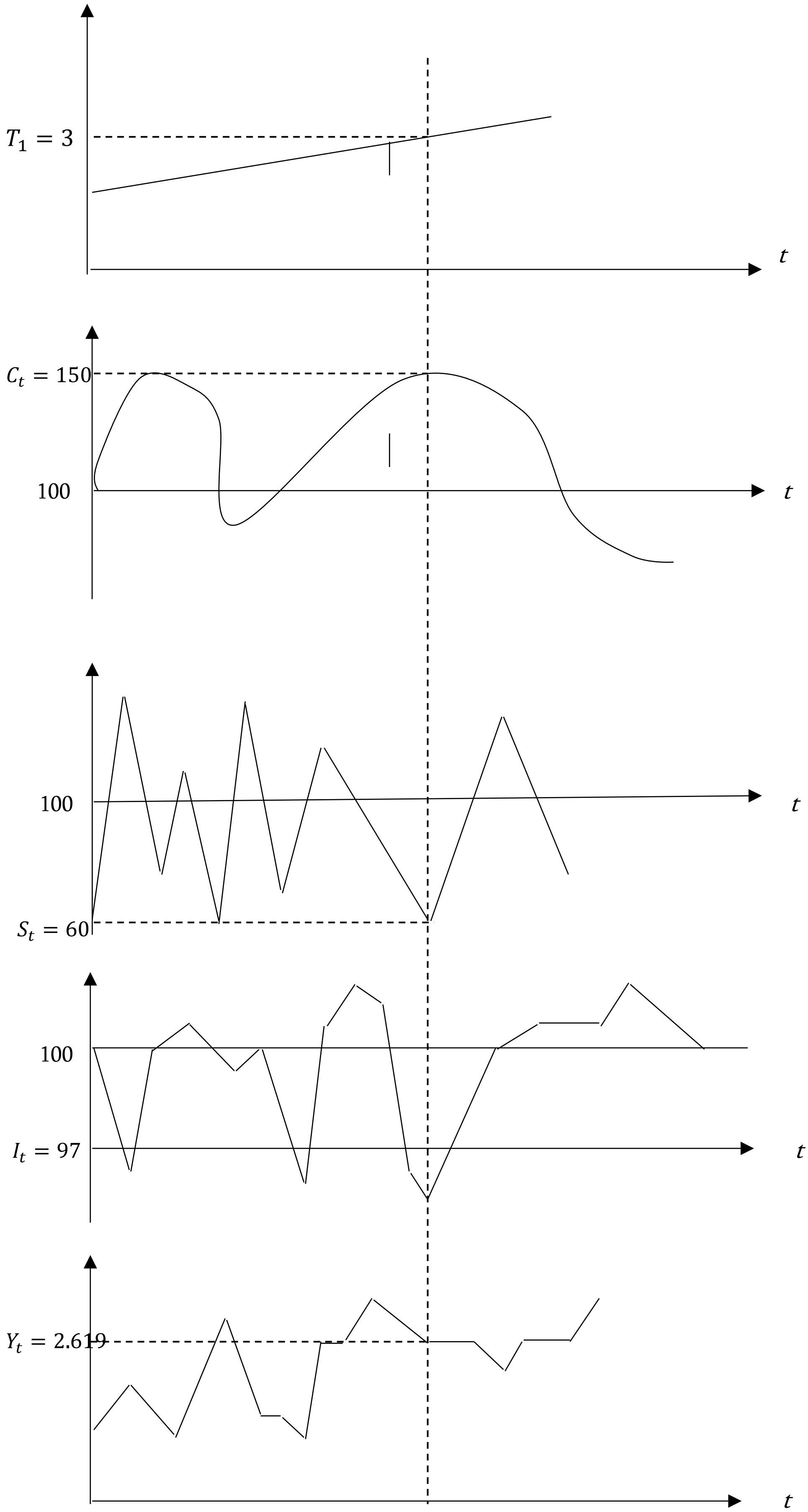
ნახ.2.9.პირობითი საქონლის გაყიდვის მოცულობის სრული ხელოვნურად კომბინირებული დროითი მწკრივი

პრაქტიკაში საქმე გვაქვს შებრუნებულ ამოცანასთან. ჩვენი მიზანია არა დროითი მწკრივის აგება მისი შემადგენელი ნაწილების მიხედვით, არამედ დროითი მწკრივიდან მისი კომპონენტების გამოყოფა და ამის საფუძველზე პროგნოზირების ტექნიკის შემუშავება.

უნდა ავღნიშნოთ, რომ მულტიპლიკაციური ურთიერთკავშირი (მულტიპლიკაციური მოდელი) შეიძლება ზედმეტად გამარტივებული აღმოჩნდეს. დროითი მწკრივის ზოგიერთი მოდელი, გარდა ზევით აღწერილი ოთხი კომპონენტისა , შეიცავს დამატებით ფაქტორებს. მაგალითად,ჩვენ შეგვიძლია დავუშვათ, რომ Y_t დამოკიდებულია მის წარსულ მნიშვნელობებზეც, სახელდობრ Y_{t-1}-ზე.

ზოგიერთი მოდელი, რომელიც გამოიყენება დროითი მწკრივების ანალიზისათვის, დროითი მწკრივებისს ყოფაქცევას ხსნის მხოლოდ მის წარსულ მნიშვნელობებზე დაყრდნობით, სხვა სიტყვებით რომ ვთქვათ, მისი წარსული ყოფაქცევით, ციკლური და სეზონური კომპონენტების გამოყოფის გარეშე. ყურადღება შევაჩეროთ მულტიპლიკაციურ მოდელზე და მის საფუძველზე ჩავატარებთ დროითი მწკრივების ანალიზს.

შევაჯამოთ, გრაფიკულად დროითი მწკივის მულტიპლიკაციური მოდელის ანალიზის შედეგები.



ნახ.2.10. სხვადასხვა კომპონენტების ზედდებით შედგენილი ხელოვნური დროითი მწკრივი

დროითი მწკრივის ანალიზი კლასიკურ მოდელზე დაყრდნობით

ტრენდის ანალიზის მიზანია ისეთი წირის (ტრენდის მრუდისს) შერჩევა, რომელიც ადეკვატურად ასახავს დროითი წირის მრაობის ძირითად ტენდენციას დროის ხანგრძლივ პერიოდსი (დროის საკმაოდ გრძელ ინტერვალში). ტრენდის კომპონენტა განსაკუთრებით მნიშვნელოვანია პროგნოზირებისათვის.

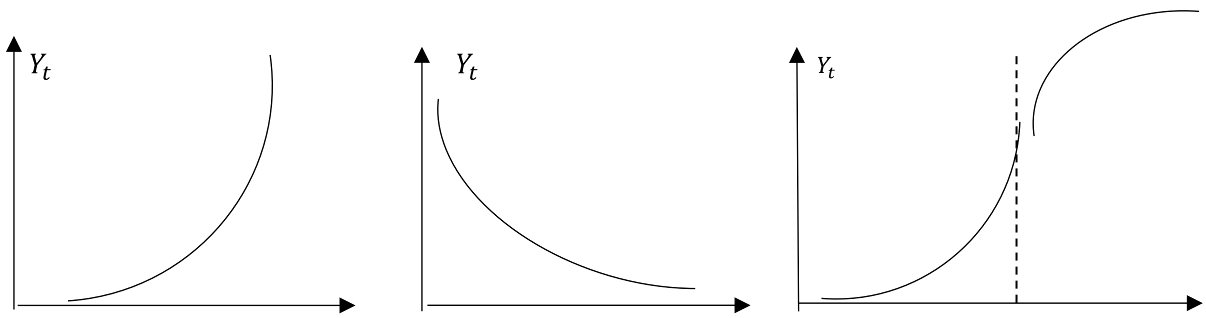
ტრენდი შეიძლება იყოს წრფივი ან არაწრფივი, მაგალითად პოლინომური, ლოგარითმული, ექსპონენციალური და სხვა. ტრენდის გამოყოფა ძირითადად რეგრესიული ანალიზის მეთოდებით ხდება. დროითი მწკრივის შემთხვევაში გაბნევის დიაგრამის როლს ასრულებს დროითი მწკრივის გრაფიკული გამოსახულება, რომელიც გვიქმნის საწყის წარმოდგენას, იმის თაობაზე თუ რა ტიპის მოდელი უნდა გამოვიყნოთ.

გრძელვადიანი ტრენდის წრფივი მოდელი. თუ ვარაუდობთ , რომ გრძელვადიანი ტრენდი წრფივია, სხვა სიტყვებით რომ ვთქვათ, იგი დროში მუდმივი სიჩქარით იცვლება, უნდა გამოვიყენოთ მოდელი

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t,$$

სადაც ε_t შემთხვევითი ხმაურია. ცხადია, რომ დროის ერთეულში ტრენდის ცვლილების სიდიდე β_1-ის ტოლია.

ხშირ შემთხვევაში დაშვება, რომ ტრენდის ცვლილების სიჩქარე მუდმივია, არარეალურია. თუ დავაკვირდებით, დროითი მწკრივის რეალიზაციებს , შეიძლება აღმოჩნდეს, რომ არაწრფივი მოდელები უფრო მოსახერხებელია. ეს ნიშნავს, რომ დროითი მწკრივის ცვლილების სიჩქარე დროში ცვალებადია. კერძოდ, თავად სიჩქარე შეიძლება აღმოჩნდეს ზრდადი ან კლებადი, ანდა დროის რაღაც მომენტში ზრდადი სიჩქარე შეიცვალოს კლებადი სიჩქარით. ქვემოთ მოყვანილი ნახაზები აღწერს ამ სიტუაციაებს.

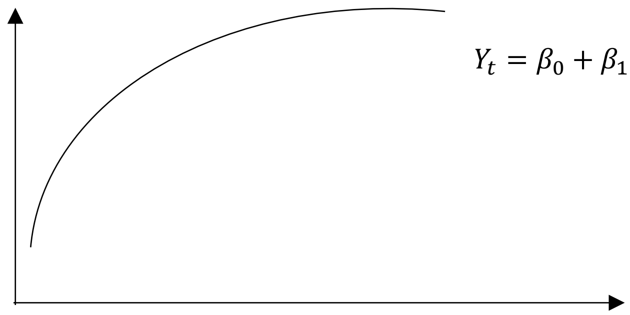


Y_tიზრდება ზრდადი სიჩქარით | Y_tკლებადია კლებადი სიჩქარით | ზრდა ზრდადი სიჩქარით | ზრდა კლებადი სიჩქარით

ნახ.2.11.

არაწრფივი ტრენდები შეიძლება აღიწეროს პოლინომური, ლოგარითმული და სხვა არაწრფივი მოდელებით.

პოლინომური მოდელი. პოლინომური მოდელები აღიწერება რაიმე რიგირს პოლინომების მეშვეობით: Y_t გამოისახება t ცვლადის რაიმე $p > 1$ რიგის პოლინომით. მოვიყვანოთ $p = 2$ შემთხვევის მაგალითი:



$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t,$$

ნახ.2.12. პოლინომური მოდელი

გრძელვადიანი ტრენდის ლოგარითმული მოდელი. ლოგარითმული ანუ ექსპონენციალური მოდელი გამოიყენება ისეთი დროითი მწკრივის აღსაწერად, რომელიც ავლენს ექსპონენციალურ ზრდას ან (კლებას).

$$Y_t = \beta_0(\beta_1)^t \varepsilon_t, \quad \text{ან} \quad Y_t = \beta_0(\beta_1)^{-t} \varepsilon_t,$$

სადაც β_1 ყოველთვის დადებითია, ამასთან პირველი მოდელი შეესაბამება ექსპონენციალურ ზრდას მეორე კი - კლებას.

ტრენდის გამოყოფა უმცირეს კვადრატთა მეთოდით. ტრენდის ანალიზი ანუ ტრენდის გამოყოფა შესაძლებელია რეგრესიული ანალიზის მეთოდების გამოყენებით. ამ შემთხვევაში, დამოუკიდებელ ცვლადად ჩაითვლება დროითი ცვლადი, ხოლო დამოკიდებულ ცვლადად კი -t-ურ მომენტში დროითი მწკრივის მნიშვნელობა.

რეგრესიულიული ანალიზისათვის განხილული შემთხვევვისაგან განსხვავებით , დროითი მწკრივების ანალიზი რეგრესიული ანალიზის მეთოდების გამოყენებით გარკვეულად შემოფარგლულია. ეს განპირობებულია იმით, რომ დაშვება, რომელის გამოიყენება რეგრესიულ ანალიზში დასკვნების მისაღებად, დროითი მწკრივების შემთხვევაში ხშირად არ სრულდება. ეს დაშვებებია: შეცდომების, ε_t-ების, დამოუკიდებლობა ნორმალურობა და ჰომოსკედატურობა. ამრიგად, დასკვნების სტატისტიკური საიმედოობის შეფასება ხშირად არ ხერხდება.

როგორც წრფივი, ასევე პოლიომური ტრენდის განსაზღვრა უმცირეს კვადრატთა მეთოდით (ანუ რაც იგივეა, რეგრესიის განტოლების კოეფიციენტების განსაზღვრა უმცირეს კვადრატთა მეთოდით) სავსებით იდენტურია წინა თავში მოყვანილი მეთოდებისა, ისევე როგორც შესაბამისი სტატისტიკური დასკვნების გაკეთება, იმ შემთხვევაში, როდესაც რეგრესიული ანალიზის დაშვებები სრულდება.

მაგალითისათვის, განვიხილოთ ლოგარითმული ანუ ექსპონენციალური მოდელის შემთხვევა:

$$Y_t = \beta_0(\beta_1)^t \varepsilon_t,$$

რომელიც ადვილად დაიყვანება ისეთ ფორმულამდე, რომელიც მოგვცემს რეგრესიული ანალიზის გამოყენების შესაძლებლობას. ეს დაყვანა ხდება გალოგარითმებით.

$$logY_t = log\beta_0 + (log\ \beta_1)t + log\varepsilon_t,$$

ბოლო განტოლება შეიძლება ჩაიწეროს რეგრესიული ანალიზისათვის უფრო მოსახერხებელი ფორმით

$$Y_t' = \beta_0' + \beta_1' t + \varepsilon_t^i \quad t = 0,1,2,\dots,n.$$

ამ განტოლებიდან უკვე ადვილად განისაზღვრება რეგრესიის კოეფიციენტები

$$\hat{\beta}_1' = \frac{\sum_{t=1}^{n} tY' - \frac{1}{n}\sum_{t=1}^{n} t \sum_{t=1}^{n} Y_t'}{\sum_{t=1}^{n} t^2 - n(\frac{1}{n}\sum_{t=1}^{n} t)^2}, \qquad \hat{\beta}_0' = \bar{Y}' - (\frac{1}{n}\sum_{t=1}^{n} t)\hat{\beta}_1'.$$

თავი III

ნივთიერებების საშიშროების შეფასება

გარემოზე ადამიანის ზემოქმედებამ საგანგაშო მასშტაბებს მიაღწია. ამ მდომარეობის მიზანმიმართულად და გააზრებულად გამოსწორება შესაძლებელია გარემოს თანამედროვე მდგომარეობის შესახებ საიმედო მონაცემების შეგროვებით და სისტემის შემოღებით, რომლების გააკონტროლებს გარემოში მოხვედრილ საშიშ ნივთიერებების რაოდენობას.

საშიშროებს მიხედვით ნივთიერებათა რეიტინგი, ჩვენი აზრით, უნდა ეფუძნებოდეს სამ კრიტერიუმს:

● ტოქსიკურობა - ორგანიზმზე მავნე ზემოქმედების უნარი,

LD_{50}-ლეტალური(სასიკვდილო) დოზა, ან LC_{50} ლეტალური კონცენტრაცია,

LD_{50}-დოზა, რომელიც 100 გამოსაცდელი ობიექტიდან 50-ის სიკვდილს იწვევს

ერთეული-გ/კგ ან მგ/კგ (გ ან მგ -ტოქსიკანტის მასაა, კგ ობიექტის მასა)

● ორგანიზმში შეღწევადობა - ნივთიერებების შესაძლებლობას შეაღწიოს ორგანიზმში და მოახდინოს მასზე ზემოქმედება, უგანზომილებო სიდიდეა.

● მდგრადობა - რამდენად მდგრადია ნივთიერება ბუნებაში, ფასდება ნახევარდაშლის პერიოდით, რომელიც გვიჩვენებს, თუ რა დროა საჭირო ნივთიერების მასის ნახევრსის დასაშლელად გარემოში ბიოტური და აბიოტური ფაქტორების გავლენით.

თითოეულ ბლოკს მივანიჭოთ ქულა [0;600] ჩაკეტილი ინტერვალიდან.

ნივთიერების საშიშროება Y_n

ტოქსიკურობა X_{n1} ორგანიზმში შეღწევადობა X_{n2} მდგრადობა
X_{n1} X_{n2} X_{n3}

საბოლოო ქულაY_n = ტოქსიკურობა(X_{n1})+ორგანიზმში შეღწევადობა (X_{n2})+მდგრადობა(X_{n3})

(1800 მაქს ქულა) (600 ქულა) (600 ქულა) (600 ქულა)

n-კონცკრეტულ ნივთიერებას მიუთითებს, რომელიც გარემოს დამბინძურებლად ითვლება.

თითოეული მდგენელის შესაბამისი ქულის მინიჭებისათვის განსხვავებულ ალგორითმს ვიყენებ.

პირველი ბლოკი (მდგენელი) არის ტოქსიკურობა

ტოქსიკურობა

განვიხილავთ მეთოდს რომელსაც გამოვიყენებთ ნივთიერების ტოქსიკურობის ქულის დასადგენად.

თითოეული ნივთიერებას განხვიხილავთ შედეგი ხუთი თვისების მიხედვით ესენია:
●ქიმიური აქტიურობა (ანუ რეაქციის უნარი),

●წყალხსნარის ტოქსიკურობა,

●ძუძუმწოვრების მიმართ ტოქსიკურობა,

●ქრონიკული ტოქსიკურობა,

●კარცინოგენურობა(სიმსივნის გამოწვევის უნარი).

ცალკე განვიილოთ თითოეული კომპონენტი. დავიწყოთ განხილვა ქიმიური აქტიურობით (ანუ ნივთიერების რეაქციის უნარი) . ავღწეროთ აალების ტეპერატურით და დუღილის ტემპერატურით

ქიმიური აქტიურობა

კატეგორია	აალებადობის ხარისხი	ქიმიური აქტიურობა		ქულა
		წყლალთან ურთიერთქმედება	რეაქციის უნარიანობა	
D	F.P 100^0-140^0F	საშუალოდ რეაქციის უნარი მაგ: ამიაკი	ადვილად პოლიმერიზდება	5000
C	F.P<100^0,B.P∃100F^0	მაღალი რეაქციის უნარი მაგ: ოლეუმი	საშუალოდ პოლიმერიზდება	1000
B	FP<100^0,B.P<100^0	ძალიან მაღალი რეაქციის უნარი მაგ:გოგირდის ტრიოქსენი	მაღალი პოლიმერიზების უნარი	100
A	თვითაალებადი ან სპეცაალებადი	აალებადი მაგ:ნატრიუმი	ძალიან მაღალი, აფეთქება	10
X	არ ამჟღავნებს			1

FP (Flash Point) არის (აალების ტემპერატურა)

BP (Boiling Point) არის (დუღილის ტემპერატურის ქულა).

შემდეგი კომპონენტი არის წყალხსნარის ტოქსიკურობა

AQUATIC TOXICITY SCALE (წყალხსნარის ტოქსიკურობის საზომი)

კატეგორია	წყლალში ტოქსიკურობა	ქულა
D	100 მგ/ლ ≤ LC50 < 500 მგ/ლ	5000
C	10 მგ/ლ ≤ LC50 < 100 მგ/ლ	1000
B	1 მგ/ლ ≤ LC50 < 10 მგ/ლ	100
A	0.1 მგ/ლ ≤ LC50 < 1 მგ/ლ	10
X	LC50 < 0.1 მგ/ლ	1

LC_{50}=სასიკვდილო კონცენტრაცია (50% კვდება) Lethal Concentration (50% kill)

შემდეგი მნიშვნელოვანი მდგენელი არის ძუძუმწოვართა ტოქსიკურობა, მის დასადგენად გამოვიყენებთ ცხრილს რომელის არის წარმოდგენილი

MAMMALIAN TOXICITY SCALE (ძუძუმწოვართა მიმართ ტოქსიკურობის საზომი)

კატეგორია	ორალურად შეღწევა	კანიდან შეღწევადობა	ინჰალაცია (ჩასუნთქვა)	ქულა
D	100 მგ/კგ ≤ LD50 < 500მგ/კგ	40 მგ/კგ ≤ LD_{50} < 200მგ/კგ	400 ppm ≤ LC_{50} < 2000 ppm	5000
C	10 მგ/კგ ≤ LD_{50} < 100 მგ/კგ	4 მგ/კგ ≤ LD_{50} < 40 მგ/კგ	40 ppm ≤ LC_{50} < 400 ppm	1000
B	1 მგ/კგ ≤ LD_{50} < 10 მგ/კგ	0.4 მგ/კგ≤ LD_{50} < 4 მგ/კგ	4 ppm ≤ LC_{50} < 40 ppm	100
A	0.1 მგ/კგ ≤ LD_{50} < 1 მგ/კგ	0.04 მგ/კგ ≤ LD_{50} < 0.4 მგ/კგ	0.4 ppm ≤ LC_{50} < 4 ppm	10
X	LD_{50} < 0.1 მგ/კგ	LD_{50}< 0.04 მგ/კგ	LC_{50}< 0.4 ppm	1

შემდეგი კომპონენტი რომელიც ნივთიერების ტოქსიკურობსი ქულის დადგენაში მონაწილეობას ღებულობს არის ქრონიკული ტოქსიკურიბა.

ქრონიკული ტოქსიკურობა (chronic toxicity)

თითოეული ნივთიერებისთვის ვიხილავთ ორ ძირითად ნიშანს:

● მინიმალური ეფექტური დოზა ქრონიკული ზემოქმედებისთვის (მგ/დღე 70 კგ ადამიანისთვის)

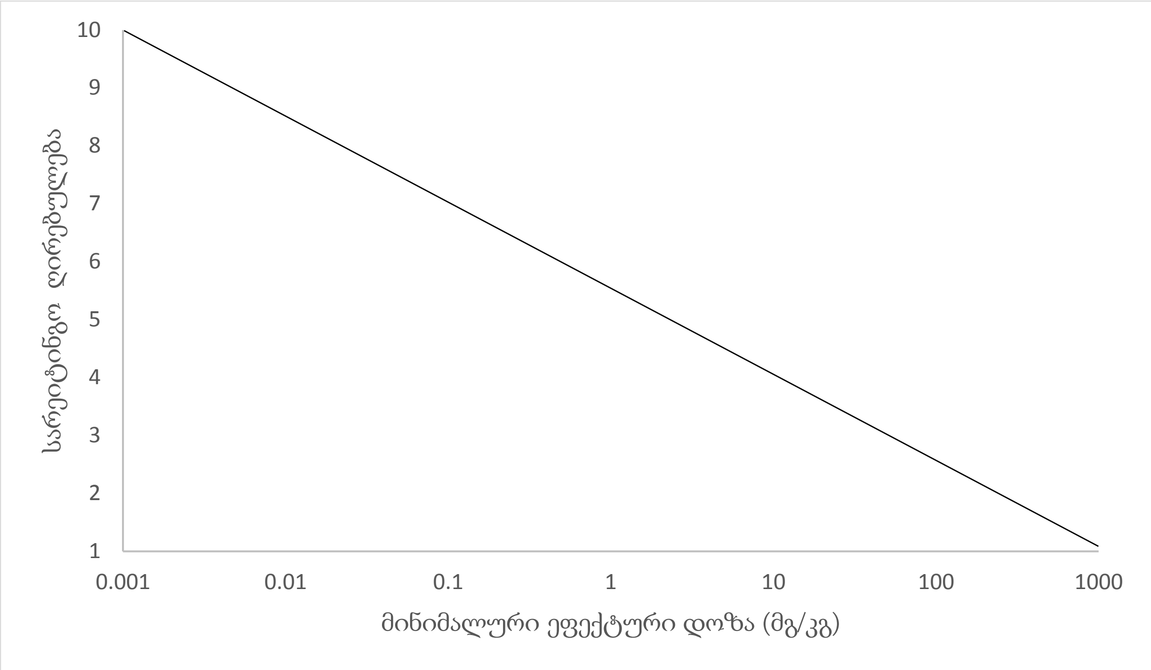
●რა ტიპის ზემოქმედებას იწვევს (ღვიძლის ნეკროზი, ტერატოგენურობა და ა.შ)

მინიმალური ეფქტურობის დოზა ეს არის მინიმალური დოზა, რომელსაც აქვს ეფექტი ორგანიზმზე. ადამიანის ორგანიზმზე მინიმალური ეფექტური დოზის დადგენისთვის ცხოველების მინიმალური ეფექტური დოზა წონის მიხედვით გადაკეთებულია ადამიანზე, ადამიანის წონა 70 კგ.

ცხოველისთის დოზა (მგ/დღე):

$$ცხოველის\ დოზა(მგ/დღე)x\left(\frac{70\ კგ}{ცხოველის\ წონა}\right)^{\frac{2}{3}}$$

სარეიტინგო ღირებულება მინიმალურ ეფექტურ დოზასა და და მის ღირებულებას შორის არის მოცემული შემდეგი სქემით :



შენიშვნა: 1000 ან მეტი მინიმალური ეფექტურობის დოზას ენიჭება ღირებულება 1, ხოლო 0.001 ან ნაკლებს მინიმალური დოზა 10.

ანალოგიურად, განვიხილავთ სახიფათო ნივთიერებების თვისებას მიაყენოს ორგანიზმს ზიანი, ზიანის ტიპს ვანიჭებთ ქულას ერთიდან ათამდე სიმძიმის მიხედვით.

ტოქსიკურობის ეფექტის შეფასების ცხრილი

რეიტინგი	ეფექტი
1	იწვევს ფერმენტის ინდუქციას ან სხვა ბიოქიმიური ცვლილებას არ იწვევს პათოლოგიური ცვლილებას და არ ცვლის ორგანოს მასას.
2	იწვევს ფერმენტების ინდუქციურ და სუბკულარულ პროლიფერაციას ან სხვა ორგანოების ცვლილებებ, მაგრამ არ ამჟღავნებს ხილული ეფექტს.
3	იწვევს ჰიპერპლაციას, ჰიპერტროფიას ან ატროფიას, მაგრამ არ იწვევს ორგანოს მასის ცვლილებას.
4	იწვევს ჰიპერპლაციას, ჰიპერტროფიას ან ატროფიას, ასევე ორგანოს იწვევს ორგანოს მასის ცვლილებას.
5	იწვევს შეუქცევად უჯრედულ ცვლილებებს, შეშუპებას ან ცხიმოვან ცვლილებებს.
6	იწვევს ნეკროზს ან მეტაპლაზის,მაგრამ არ არის აშკარა ორგანოების ფუნქციის შემცირება, სხვა ნეიროპათია ხილულუ ჩვენების გარეშე,იწვევს სენსორულ ან ფიზიოლოგიურ ცვლილებებს.
7	იწვევს ნეკროზს, ატროფიას, ჰიპერტროფიას ან მეტაპლაზიას, აშკარაა ორგანოს ფუნქციის შემცირება , სხვა ნეიროპათია ხილული ჩვენებებით, სენსორული ან ფიზიოლოგიური ცვლილებები.
8	იწვევს ნეკროზს, ატროფიას, ან მეტაპლაზიას საბოლოოდ ორგანოს დისფუნქციით, სხვა ნეიროპათიური ცვლილებები ,იწვევს სენსორული ან მოტორული ფუნქციების დისფუქნციას.
9	იწვევს პათოლოგიურ ცვლილებებს, ორგანოების მძიმე დისფუნქციას,სენსორული უნარის დაკარგვას, რეპროდუქციული დისფუნქცის.
10	იწვევს სიკვდილი ან სიცოცხლისთვის შეუთავსებელ ორგანულ დაზიანებებს, ორგანოების დისფუნქციას.

თითოეული ნივთიერებისთვის „ტოქსიკურობის ეფექტის" შესაბამის ქულას ვამრავლებთ „სარეიტინგო ღირებულების" ქულაზე. მიღებული ქულა იქნება ერთსა და ასს შორის, მიღებულ ქულას ვუწოდოთ "შედგენილი ქულა", ხოლო მიღებული „შედგენილი ქულის" შესაბამის საბოლოო ქულა (შესაბამისობას ვახდენთ შემდეგი ცხრილის გამოყენებით) ვუწოდებთ მოცემული ნივთიერების ქრონიკული ტოქსიკურობის ქულას

ქრონიკული ტოქსიკურობის ქულის დადგენაში გამოვიყენოთ შემდეგი ცხრილი

ქრონიკული ტოქსიკურობის ქულა

კატეგორია	შედგენილი ქულა	ქულა
D	1-5	5000
C	6-20	1000
B	21-40	100
A	41-80	10
X	81-100	1

შემდეგი მდგენელი ტოქსიკურობის ქულის დადგენისთვის არის კარცეროგენურობა

კანცეროგენურობა, რაც კიბოს (ავთვისებიან სიმსივნეს) იწვევს.

გარემოსდაცვითი სააგენტოს მტკიცებულების საზომი კატეგორიები	გარემოს დაცვითი სააგენტოს კიბოს კლასიფიკციის კატეგორიები და ქვეკატეგორიები	განმარტება
ჯგუფი A	ცნობილი კარცინოგენი ადამიანებში	ნივთიერებები, რომლისთვისაც საკმარისი იმფორმაცია არსებობს ადამიანური ეპიდემიოლოგიური კვლევების მიხედვით.
ჯგუფი B1	სავარაუდო კარცენოგენები ადამიანებში (მტკიცებულება ზუსტი არ არის)	ადამიანის კარცინოგენიის მტკიცებულების სიმყარე, რომელიც დამყარებულია ეპიდემიოლოგიურ კვლევებზე შეზღუდულია.
ჯგუფი B2	სავარაუდო კარცინოგენი ადამიანებში(სამხილი მტკიცებულება არ არსებობს)	ამ კლასში შეყვანილია ნივთიერებები, რომლისთვისაც არანაირი მტკიცებულება ან ინფორმაცია არ არსებობს ადამიანურ ეპიდემიოლოგიურ კველევებზე დაყრდნობით, მაგრამ ამ ნივთიერებებისთვის კარციოგენიურობის შესახებ მტკიცებულება, რომელიც დამყარებულია ცხოველების კვლევებზე, სრულიად საკმარისია.

ჯგუფი C	შესაძლო კარცინოგენი ადამიანებში	ამ კლასში შედის ის ნივთიერებები, რომლებზეც კარციოგენურობის არასაკმარისი(შეზღუდული) მტკიცებულება მოიპოვება ცხოველებში. და ასევე არანაირი ინფორმაცია, და მიზანშეწონილი სამხილი ადამიანური ეპიდემიოლოგიურიკ ვლევებიდან.
ჯგუფი D	არ არის შეფასებული	არასაკმარისი იმფორმაცია კარციოგენირობის კუთხით.
ჯგუფიE	არაკარცინოგენური	მტკიცდება, რომ ადამიანებში კარცეროგენური არ არის.

ნივთიერებები განლაგდება ჯგუფებში, შემდეგი სქემა კი გვიჩვენებს თუ როგორ უნდა შევუსაბამოთ ჯგუფებს ქულები.

მტკიცებულების საზომი ჯგუფი	სიმძლავრის ჯგუფი		
	1(უმაღლესი)	2	3(უდაბლესი)
A	მაღალი(1ქ)	მაღალი(1ქ)	საშვალო(10ქ)
B	მაღალი(1ქ)	საშუალო(10ქ)	დაბალი(100ქ)
C	საშუალო(10ქ)	დაბალი(100ქ)	დაბალი(100ქ)
D	არ არის საშიშროების რანჟირებაში		
E	არ არის საშიშროების რანჟირებაში		

ნივთიერების „საბოლოო ქულა“ არის ქულა რომელიც არის მინიმალური ამ ნივთიერებისთვის თითოული კრიტერიუმში, მას ვუწოდებთ მოცემული ნივთიერების საბოლოო ქულას (რომელის იქნება 1 ;10;100;1000;5000 დან ერთ-ერთი)

საბოლოო ქულა=min*(ქიმიური ტოქსიკურობა, წყალში ტოქსიკურობა, ძუძუმწოვართა მიმართ ტოქსიკურობა, ქრონიკული ტოქსიკურობა, კარცეროგენურობა)*

საბოლოოდ ჩვენი მიზანია სახიფათო ნივთიერებას შევუსაბამოთ შესაბამისი ქულა რომელიც იქნება [0:600] ჩაკეტილი ინტერვალიდან, რომელსაც შემდეგი ცხრილის მიხედვით გავაკეთებთ.

ტოქსიკანტის ქულა

საბოლოო ქულა	რიგობითი ქულა	კორექტირებული რიგობითი ქულა (COR)	2/3 ხარისხად კორექტირებული რიგობითი ქულა $(2/3)^{COR}$	ტოქსიკანტის ქულა $(2/3)^{COR}$x600
1	0	0	1	600
10	1	1	0.6667	400
100	2	3	0.2963	178
1000	3	6	0.0878	53
5000	4	10	0.0173	10

ორგანიზმში შეღწევადობა

ორგანიზმში შეღწევადობა- ორგანიზმში ნივთიერების შეღწევადობა გვიჩვენებს ნივთიერების შესაძლებლობას შეაღწიოს ცოცხალი ორგანიზმში, ნივთიერებისთვის შესაბამისი ქულის მინიჭებისთვის ვიყენებთ ბიოკონცენტრაციის ფაქტორს(BCF), რომელსაც განვსაზღვრავთ ოქტანოლის და წყალის განაწილების კოეფიციენტით(K_{ow}), რომელთა შორისაც კავშირს ამყარებს შემდეგი ფორმულა

$$log\ (BCF) = mlog(K_{ow}) + b$$

ბიოკონცენტრაციის ფაქტორი (BCF) – სიდიდე, რომელიც ასახავს ცოცხალ ორგანიზმში (თევზი) ნივთიერების დაგროვების უნარს. მიიღება ცოცხალ ორგანიზმში ნივთიერების კონცენტრაციის (მგ/კგ) შეფარდებით ნივთიერების კონცენტრაციასთან (მგ/ლ) იმ გარემომცველ გარემოში (წყალი), რომელშიც ცოცხალი ორგანიზმი ბინადრობს

$$BCF=\frac{concentrration\ \ (biota)}{concentration\ (water)}$$

ოქტანოლ/წყალი განაწილების კოეფიციენტი – (K_{ow}): ეკოტოქსიკოლოგიური მაჩვენებელი, რომელსაც ხშირად იყენებენ $logK_{ow}$-ს სახით ორგანული ქიმიური ნივთიერებების გარემოში გადაადგილების შესაფასებლად. ასახავს დამოკიდებულებას ოქტანოლსა და წყალში განაწილებულ ნივთიერების რაოდენობას შორის. განისაზღვრება ოქტანოლში გახსნილი ნივთიერების კონცენტრაციის ($C_{octanol}$) შეფარდებით ამ ნივთიერების კონცენტრაციასთან წყალში (C_{water})

$$K_{ow}=\frac{concentrration\ \ (octanol)}{concentration\ (water)}$$

ჩვენი კლევა ეხება წყალში მცხოვრებ ორგანიზმებს (თევზებს).

Kow -ოქტანოლსა და წყალს შორის განაწილების კოეფიციენტი ნებისმიერი ტოქსიკანტისთვის მუდმივი სიდიდეა და არ არის დამოკიდებული სახეობაზე განსხვავებით BCF -ისგან (ბიოკონცენტრაციის ფაქტორისგან), რომელიც სახეობასდამიხედვით სხვადასხვაა ნივთიერებისთვის განსხვავეული სიდიდეა. ჩვენი მიზანი არის რეგრესიული ანალიზის გამოყენებით ვაჩვენოთ,რომ ამ ორ სიდიდეს შორის (BCF -სა და Kow) არსებობს წრფივი რეგრესიული კავშირი. ამის საჩვენებლად ავიღეთ თევზის ორი სახეობა (rainbow trout-ცისარტყელა კალმახი და sunfish-მზისთევზა) რომელთა შესაბამისი BCF- ები გარკვეული რაოდენობა ნივთიერებებისთვის ცნობილია.

განვიხილავთ ორი სახეობას შესაბამისად ჩვენს განტოლებებს ექნება სახე

$lnBCF^{(1)}{}_n = m_s lnK_{OW_n} + c_s$ n=1,…76 კალმახისთვის (rainbow trout)

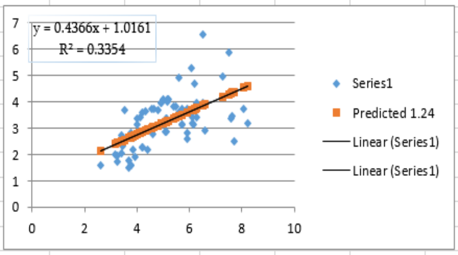
$lnBCF^{(2)}{}_s = m_s lnK_{OW_s} + c_s$ s=1,…41 მზისთევზისთვის (sunfish)

გვაქვს წყვილების ერთობლიობა

$(K_n{}^{(1)}; BCF_n{}^{(1)})$ n=1,… 72 და $(K_s{}^{(2)}; BCF_s{}^{(2)})$ s=1,…41

ცისარტყელა კალმახის მონაცემების ერთობლიობაზე რეგრესიულ ანალიზის გამოყენებისას მივიღებთ შემდეგ სურათს

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.579094							
R Square	0.33535							
Adjusted R	0.325576							
Standard E:	0.812558							
Observatio:	70							
ANOVA								
	df	*SS*	*MS*	*F*	*ignificance F*			
Regression	1	22.65291	22.65291	34.30954	1.51E-07			
Residual	68	44.89708	0.660251					
Total	69	67.54999						
	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	1.016112	0.392896	2.586213	0.011849	0.232101	1.800124	0.232101	1.800124
2.83	0.436615	0.07454	5.857435	1.51E-07	0.287872	0.585358	0.287872	0.585358



ბიოკონცენტრაციის ფაქტორებს(BCF) ოქტანოლის და წყლის შეფარდების კოეფიციენტს (Kow) შორის არის შემდეგი დამოკიდებულება

$lnBCF^{(1)} = 0.4366 lnK_{ow} + 1.0161$ R^2=0.33535 r=0.58

თევზის მეორე სახეობისთვის, მზისთევზასთვის (sunfish) რეგრესიული ანალიზის გამოყებით გვექნება შემდეგი სურათი

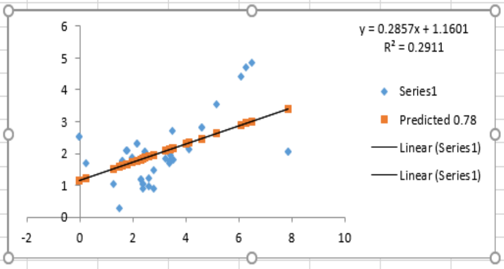
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.539543
R Square	0.291107
Adjusted R	0.271947
Standard E	0.855507
Observatio	39

ANOVA

	df	SS	MS	F	Significance F
Regression	1	11.1204	11.1204	15.19404	0.000393
Residual	37	27.08001	0.731892		
Total	38	38.2004			

	Coefficients	Standard Erro	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	1.160123	0.265768	4.36518	9.82E-05	0.621627	1.698619	0.621627	1.698619
1.9	0.285677	0.073289	3.897953	0.000393	0.137179	0.434175	0.137179	0.434175



ფაქტორებს (BCF)-სა და (K_{ow})-ს შორის არის შემდეგი დამოკიდებულება

$lnBCF^{(2)} = 0.2857lnK_{ow} + 1.1601$ R^2=0.2911 r=0.54

ამ ორ სიდიდეს შორის არსებობს წრფივი დადებითი კავშირი.

მივიღეთ ორი განტოლება

$lnBCF^{(1)} = 0.4366lnK_{ow} + 1.0161$ m_1=0.2857 b_1=1.1601

$lnBCF^{(2)} = 0.2857lnK_{ow} + 1.1601$ m_1=0.4366 b_2=1.0161

თევზებისთვის ავიღოთ განტოლება რომელიც იქნება მოცემული ორი განტოლების საშუალო არითმეტიკული

მივიღებთ ახალ განტოლებას რომლისთვისაც

$$m = \frac{m_1+m_2}{2} = \frac{0.2857+0.4366}{2} = 0.36115$$

$$b = \frac{b_1+b_2}{2} = \frac{1.1601+1.0161}{2} = 1.30855$$

და განტოლებას ექნება შემდეგი სახე

$$lnBCF = 0.36115logK_{ow} + 1.30855$$

საიდანაც მივიღებთ განტოლებას,

ორგანიზმში შეღწევადობის განტოლება

$$e^{lnBCF} = e^{0.36115lnK_{ow}+}e^{1.30855}$$

$$BCF = K_{ow}{}^{036115} * e^{1.30855}$$

დასკვნა

სამაგისტრო ნაშრომში გადმოცემულია რეგრესიისა და კორელაციის და დროითი მწკრივების თეორემები, მოყვანილია თეორემები რეგრესიულ ანალიზიდან და დროითი მწკრივების თეორემებიდან. მიღებული ნივთიერებათა ცხრილი, ნივთიერებათა საშიშროების მიხედვით. გამოყვანილია რეგრესიული განტოლებები თევზის ორი სახეობისთვის, რომლებიც გვიჩვენებენ კავშირს ოქტანოლის და წყლის გაყოფის კოეფიციენტსა და ბიოკონცენტრაციის ფაქტორს შორის, რომლებიც გამოიყენება ნივთიერებათა საშიშროების შეფასებაში.

ლიტერატურა

1. ალბათობის თეორია და მათემატიკური სტატისტიკა . ნ.ლაზრიევა, გ.მარია, ა.მოსიძე, ა.ტორონჯაძე, თ.ტორონჯაძე, თ.შერვაშიძე

2. ზ.დოჭვირი, ალბათობის თეორია და მათემატიკური სტატისტიკა,ლექციების კურსი ეკონომიკური ფაკულტეტის სტუდენტებისათვის.

3. გარემოს ფიზიკური დაბინძურება. ლუცინდა ჩხეიძე

4.Support document to the 2017 (Candidates for toxicological profiles)

5.The ATSDR 2017 Substance Priority List

6. Physical-Chemical Properties and Environmental Fate for Organic Chemicals, Volume I, Volume II, Volume III.