CSS
Center for Social Sciences
STUDIUM PRETIUM LIBERTATIS

სოციალურ მეცნიერებათა სერია

სოციოლოგიურ მონაცემთა ანალიზი

მომზადებულია სოციალურ მეცნიერებათა ცენტრის
აკადემიური სტიპენდიის საფუძველზე

სალექციო კურსი
სოციალური მეცნიერებების მაგისტრატურისათვის

ნინო დურგლიშვილი

თბილისი
2006

შედარებითი სკალირება 42
სკალირება წყვილური შედარების მეთოდით 42
სკალირება თანმიმდევრობით[1] დალაგების/რანჟირების მეთოდით 43
სკალირება მუდმივი ჯამური მაჩვენებლის მეთოდით 44
სკალირება Q-სორტირების მეთოდით 45
სკალირება გუტმანის მეთოდით 46
სკალირება მნიშვნელოვნების შეფასების მეთოდით 49
არაშედარებითი/დამოუკიდებლად გაზომვის სკალირება 49
სკალირება უწყვეტი რეიტინგული სკალით 49
დეტალიზებული რეიტინგული სკალების სპეციფიკა. 50
ლაიკერტის სკალა 50
სემანტიკური დიფერენციალი 52
სტეპელის სკალა 53
დეტალიზებული რეიტინგული სკალების მახასიათებლები და მათი მნიშვნელობა მონაცემთა ანალიზისას 54
ერთგანზომილებიანი სტატისტიკა 55
ერთგანზომილებიანი სიხშირული განაწილებები 56
ინტერვალებად დაჯგუფება 58
ემპირიული მონაცემების გრაფიკული წარმოდგენა 60
ცენტრალური/საშუალო ტენდენციების მაჩვენებლები (საშუალო არითმეტიკული, მოდა, მედიანა) 62
ვარიაციის მაჩვენებლები: ვარიაციის დიაპაზონი. დისპერსია. საშუალო კვადრატული გადახრა/სტანდარტული გადახრა, საშუალო აბსოლუტური გადახრა, ვარიაციის კოეფიციენტები 65
სტანდარტული მნიშვნელობები 67
განაწილების ფორმები სიმეტრიული და ასიმეტრიული განაწილებები, ექსცესი 68
სტატისტიკური ჰიპოთეზების შემოწმება 69
საშუალოთა შედარება 71
კორელაცია და რეგრესია (დამოუკიდებელი და დამოკიდებული, ფაქტორული და რეზულტატური ცვლადები) 73
ორგანზომილებიანი/კორელაციური განაწილება 74
კორელაციის სტატისტიკური მნიშვნელოვნების შეფასება 77
კორელაციის დახასიათება. კორელაციის კოეფიციენტები 81
სტატისტიკური დამოკიდებულების ინტერპრეტაცია 83
კორელაციური დამოკიდებულების სპეციფიკა 83
რეგრესიული ანალიზი 84
რეგრესიული ანალიზის ძირითადი ეტაპები 84
რეზულტატურ ცვლადზე მოქმედი არსებითი ფაქტორების გამოყოფა 86
რეგრესიის მოდელი 87
რეგრესიის წრფივი მოდელი 89
უმცირეს კვადრატთა მეთოდი 90

[1] ამ მეთოდს რიგის მიხედვით დალაგების მეთოდიც შეიძლება ვუწოდოთ.

რეგრესიის წრფივი მოდელის გრაფიკული გამოსახვა/სკატერგრამა 90
რეგრესიული დამოკიდებულების სტატისტიკური მნიშვნელოვნების შეფასება 91
რეგრესიის მოდელის ხარისხის შეფასება 93
რეგრესიის სტანდარტული ცდომილება 94
დეტერმინაციისა და მრავლობითი კორელაციის კოეფიციენტები 95
ცალკეული ფაქტორის მნიშვნელობა რეზულტატური ცვლადის ვარიაციაში 96
ბეტა კოეფიციენტები/ბეტა წონები 97
საკუთრივი/ნანილობითი კორელაციის კოეფიციენტი 98
მულტიკოლინეურობა 98
დისპერსიული ანალიზი 99
დისპერსიული ანალიზის ზოგადი დახასიათება 99
ერთფაქტორული დისპერსიული ანალიზი შემთხვევითი შერჩევისას 100
ორფაქტორული დისპერსიული ანალიზი შემთხვევითი შერჩევისას 103
დისპერსიული ანალიზის ინტერპრეტაცია ფაქტორის დისკრიმინაციული უნარის შეფასებისას 106
ფაქტორული ანალიზი 107
ფაქტორული ანალიზის იდეა 107
ზოგადი ფაქტორების გამოყოფის მეთოდები 108
ცენტროიდული მეთოდი 109
მთავარ კომპონენტთა მეთოდი 110
პარამეტრთა ექსტრემალური დაჯგუფების მეთოდი 111
ფაქტორების ბრუნვა 112
ზოგადი ფაქტორების შინაარსობრივი ინტერპრეტაცია 112
კლასტერული ანალიზი 118
კლასტერული ანალიზის ძირითადი მიზანი 118
კლასტერიზაციის მეთოდები 118
აგლომერატული იერარქიული კლასტერიზაციის მექანიზმი 119
კლასტერთა ურთიერთდაშორების ზომები 120
კლასტერების აგრეგაცია 121
ცვლადების სტანდარტიზება 122
არაიერარქიული კლასტერიზაციის მექანიზმი. 123
ოპტიმალური განაწილების მეთოდი
კლასტერიზაციის მეთოდის შერჩევის პრობლემა 123
დენდროგრამა 124
მრავალგანზომილებიანი სკალირება 125
მრავალგანზომილებიანი სკალირების ძირითადი მიზანი 125
მრავალგანზომილებიანი სკალირების გამოყენების სპეციფიკა 126
მრავალგანზომილებიანი სკალირების მოდელი 127
მრავალგანზომილებიანი სკალირების მოდელის ხარისხი 128
განზომილებათა რაოდენობის შერჩევა 128
წერტილთა კონფიგურაციის ინტერპრეტაცია სივრცულ რუკაზე 129
Analysis of Sociological Data 131

შესავალი

მონაცემთა ანალიზი, ზოგადი გაგებით, იმ მეთოდების სისტემას ეწოდება, რომელთა მეშვეობითაც ემპირიული მონაცემების ანალიზი ხორციელდება.

მონაცემთა ანალიზის მეთოდები უნივერსალურია. ისინი გამოიყენება ყველა სფეროში, სადაც მონაცემთა ერთობლიობასთან/მასივთან შეიძლება გვქონდეს საქმე (სოციოლოგია, ფსიქოლოგია, პოლიტოლოგია, ეკონომიკა, მედიცინა, ბიოლოგია, კრიმინალისტიკა და ა. შ.).

მონაცემთა ანალიზის მეთოდების გამოყენება ყოველ სფეროში გარკვეული სპეციფიკით ხასიათდება.

ჩვენი მიზანია მონაცემთა ანალიზის ძირითადი მეთოდების შესწავლა და ემპირიულ სოციოლოგიურ კვლევებში მათი გამოყენების დაუფლება.

ემპირიული სოციოლოგიური კვლევა რთული, ლოგიკურად სტრუქტურირებული პროცესია.

მონაცემთა ანალიზი კვლევის ერთ-ერთი დასკვნითი ეტაპია, მაგრამ იმ ძირითადი მეთოდების შერჩევა, რომელთა მეშვეობითაც ინფორმაცია უნდა დამუშავდეს, მონაცემთა მოპოვების დაწყებამდე ხდება **საკვლევი ობიექტის, კვლევის მიზნებისა და ამოცანების, სამუშაო ჰიპოთეზის შესაბამისად**.

მას შემდეგ, რაც ანალიზის რომელიმე მეთოდს შევარჩევთ, ინფორმაცია იმ წესით უნდა მოვიპოვოთ, რომ მისი დამუშავებისას შერჩეული მეთოდის გამოყენება შესაძლებელი გახდეს.

აქედან გამომდინარე, იმის მიუხედავად, რომ მონაცემები ჯერ უნდა მოვიპოვოთ და შემდეგ გავაანალიზოთ, სწორედ **ანალიზის მეთოდების შერჩევა განსაზღვრავს ინფორმაციის მოპოვების მეთოდების შერჩევას**.

მონაცემთა მოპოვებისა და მონაცემთა ანალიზის ეტაპების ურთიერთკავშირი ნათლად წარმოაჩენს კვლევის პროცესის რთულ, სისტემურ ხასიათს.

ემპირიული სოციოლოგიური კვლევის პროცესი სხვადასხვა ეტაპის ასეთი მრავალმხრივი ურთიერთკავშირებითაა გაერთიანებული.

სწავლების პროცესში ამ საკითხს არაერთხელ დავუბრუნდებით. შევეცდებით, რაც შეიძლება ნათლად დავინახოთ ემპირიული სოციოლოგიური კვლევის პროცესის ლოგიკური სტრუქტურა, გავარკვიოთ მონაცემთა ანალიზის, როგორც კვლევის ერთიანი ეტაპისა და მისი კონკრეტული მეთოდების ადგილი ამ სტრუქტურაში.

ემპირიული სოციოლოგიური კვლევის სტრუქტურა

❖ საკვლევი პრობლემა, კვლევის ობიექტი, საგანი, მიზანი და ამოცანები
რა? სად? რატომ?

ემპირიული სოციოლოგიური კვლევა საკვლევი პრობლემის, კვლევის ობიექტისა და საგნის განსაზღვრით იწყება.

საკვლევი პრობლემა ის საკითხია, რის შესწავლასაც სოციოლოგი ემპირიული კვლევის მეშვეობით გეგმავს.

საკვლევი პრობლემის ზუსტ, მკაფიო ფორმულირებასა და მისი **აქტუალობის განსაზღვრას** არსებითი მნიშვნელობა აქვს კვლევის პროცესისათვის.

სწორედ საკვლევი პრობლემის აქტუალური ასპექტების მიხედვით უნდა შეირჩეს **კვლევის ობიექტი** — ის სოციალური რეალობა, რომლის კვლევაც ჩვენთვის საინტერესო საკითხის შესწავლის ოპტიმალურ საშუალებას მოგვცემს.

ნებისმიერი სოციალური რეალობა კომპლექსური ხასიათისაა. მისი ყველა პარამეტრის აღწერა/დახასიათება შეუძლებელია და არც არის საჭირო.

არაარსებითი პარამეტრების შესწავლა მხოლოდ ზედმეტად გაართულებს, ხელს შეუშლის კვლევის პროცესს. ამიტომ, უნდა დაკონკრეტდეს **კვლევის საგანი** ანუ ის, თუ რა ასპექტით შევისწავლოთ შერჩეული ობიექტი.

დავუშვათ, საკვლევი პრობლემაა თბილისის სახელმწიფო უნივერსიტეტში მიმდინარე სასწავლო რეფორმების ეფექტურობა, პრობლემის აქტუალობა კი იმით განისაზღვრება, რომ რეფორმებმა უნდა უზრუნველყოს სასწავლო პროცესის გაუმჯობესება და საერთაშორისო სტანდარტებთან შესაბამისობაში მოყვანა.

კვლევის ობიექტად შეიძლება შეირჩეს სასწავლო პროცესი.

კვლევის საგნად — ამ პროცესის უშუალო აქტორების (მონაწილეების: სტუდენტებისა და მასწავლებლების) სასწავლო/შრომითი მაჩვენებლები და მათი დამოკიდებულება მიმდინარე რეფორმების მიმართ.

კვლევის ეფექტურად წარმართვისათვის არანაკლებ მნიშვნელოვანია **კვლევის მიზნის** ფორმულირება.

ემპირიული სოციოლოგიური კვლევის მიზანი შეიძლება იყოს გარკვეული სოციალური პროცესების რეგულირება, მართვა, პროგნოზირება და სხვა.

კვლევის მიზნის ზუსტი ფორმულირება მომდევნო ეტაპების მიზანმიმართულად წარმართვის, ცალკეული ასპექტების აქცენტირების საშუალებას იძლევა.

კვლევის მიზნის შესაბამისად ხდება იმ კონკრეტული ნაბიჯების ანუ **კვლევის ამოცანების** ჩამოყალიბება, რომელთა განხორციელებამაც უნდა უზრუნველყოს მიზნის მიღწევა.

ჩვენს მიერ მოტანილი მაგალითის შემთხვევაში, კვლევის მიზანს შეიძლება წარმოადგენდეს თბილისის სახელმწიფო უნივერსიტეტში სასწავლო რეფორმების ეფექტურობის ამაღლება.

ამ მიზნის განხორციელება შეიძლება შემდეგი ამოცანების გადაწყვეტით:

- აღვწეროთ რეფორმების მიზანი, ამოცანები, განხორციელების ფორმები, მეთოდები და განრიგი;
- დავაფიქსიროთ ის გარდაქმნები, რაც უკვე განიცადა სასწავლო პროცესმა რეფორმების გავლენით;
- შევისწავლოთ რეფორმების განმავლობაში სასწავლო პროცესის ობიექტური მაჩვენებლების ცვლილება (სტუდენტთა აკადემიური მოსწრება, მასწავლებელთა კვალიფიკაცია და ა. შ.);
- შევისწავლოთ სტუდენტებისა და მასწავლებლების დამოკიდებულება მიმდინარე რეფორმების მიმართ;

- გამოვყოთ არსებითი ფაქტორები, რომლებიც რეფორმებისადმი დადებით/უარყოფით განწყობებს განაპირობებს;
- გამოვავლინოთ სასწავლო პროცესის ცვლილებათა ის ტენდენციები, რასაც რეფორმების განხორციელება ინვევს;
- მოვახდინოთ რეფორმების პირობებში სასწავლო პროცესის განვითარების პროგნოზირება;
- გამოვავლინოთ სასწავლო პროცესის გაუმჯობესებისადმი მიმართული სტუდენტთა/მასწავლებელთა ინიციატივები;
- მიღებული ინფორმაციის ანალიზის საფუძველზე შევიმუშავოთ რეკომენდაციები რეფორმების ეფექტურობის ასამაღლებლად.

❖ ძირითადი ცნებების ოპერაციონალიზაცია

ცნებითი აპარატი, რომელსაც სოციოლოგი კვლევისას იყენებს, მრავალფეროვანია. მასში გამოიყოფა ცნებები, რომელთაც არსებითი მნიშვნელობა აქვთ კვლევისათვის. ასეთ ცნებებს **ძირითად ცნებებს** უწოდებენ.

ძირითადი ცნებების მნიშვნელობების მიხედვით დაკონკრეტდება ის მოვლენები, რომლებიც ემპირიულად შეიძლება დაფიქსირდეს და, რომლებზეც კვლევა ფოკუსირდება.

გარდა ამისა, ძირითადი ცნებები ერთი საზრისის მქონე და გასაგები უნდა იყოს სოციოლოგისათვის, ინტერვიუერისა თუ რესპონდენტისათვის. ისინი ერთი და იგივე მნიშვნელობით უნდა გამოვიყენოთ კვლევის ყველა ეტაპზე.

ამ ამოცანების გადაწყვეტა **ოპერაციული განსაზღვრებით** (ოპერაციონალიზაციით) ხდება.

ცნების ოპერაციონალიზაცია ახდენს იმის განსაზღვრას, თუ რა არსებითი ნიშან-თვისებები იგულისხმება მოცემულ ცნებაში.

ჩვენს მიერ განხილულ მაგალითებში ერთ-ერთი ძირითადი ცნებაა "სასწავლო რეფორმები", რადგან კვლევის მიზანი სწორედ სასწავლო რეფორმების ეფექტურობის ამაღლებაა.

ამ ცნების ოპერაციული განსაზღვრება შემდეგნაირად შეიძლება მოხდეს:

"სასწავლო რეფორმებში" ვიგულისხმებთ სასწავლო პროცესის მიზანმიმართულ გარდაქმნებს, რომლებსაც ახორციელებს თსუ ადმინისტრაცია ამ პროცესის გაუმჯობესებისა და საერთაშორისო სტანდარტებთან შესაბამისობაში მოყვანის მიზნით.

შედეგად, მოცემული ცნება ცალსახა და გასაგებ მნიშვნელობას იძენს, ხოლო კვლევის პროცესი შემდეგნაირად შეიძლება დაკონკრეტდეს:

- ჩვენ უნდა შევისწავლოთ სასწავლო პროცესში განხორციელებული ან მიმდინარე ყველა ის გარდაქმნა, რომელსაც თსუ ადმინისტრაცია მიზანმიმართულად ახორციელებს სასწავლო პროცესის გაუმჯობესებისა და საერთაშორისო სტანდარტებთან შესაბამისობაში მოყვანის მიზნით (მაგალითად, სწავლების ორი დონის — ბაკალავრიატისა და მაგისტრატურის შემოღება, გამოცდების ჩატარება ტესტების მეშვეობით, მასწავლებელთა კვალიფიკაციის ამაღლება და სხვა);
- ყველა სხვა სახის გარდაქმნა, რომელიც უნივერსიტეტში სასწავლო რეფორმების პარალელურად შეიძლება მიმდინარეობდეს (მაგალითად, ბუღალტერიის მუშაობის ოპტიმიზაცია) უნდა გაიმიჯნოს რეფორმული გარდაქმნებისაგან, რადგან მათი განურჩევლად განხილვა ხელს შეუშლის კვლევის პროცესს.

❖ კვლევის ობიექტის ლოგიკურ-სისტემური ანალიზი

სოციოლოგიის კვლევის ობიექტები გარკვეული სოციალური რეალობებია. მათი "ბუნებრივი" სახით შესწავლა შეუძლებელია:

ჯერ ერთი, იმიტომ, რომ ნებისმიერი რეალობა ამოუწურავია და კვლევის პროცესი უსასრულოდ გაგრძელდება.

მეორე, სოციოლოგიის უმთავრესი ამოცანაა სოციალურ კანონზომიერებათა კვლევა. კანონზომიერებები კი არ არის უშუალოდ დაკვირვებადი — ისინი კონკრეტულ ფაქტებში/ხდომილებებში შეიძლება გამოვლინდეს სხვადასხვა სახით.

შესაბამისად, აუცილებელია კვლევის ობიექტის ისე აღწერა, რომ:

- იგი **კვლევას** დაექვემდებაროს;
- მასში გამოიყოს ის არსებითი ელემენტები, რაც **შედეგების განზოგადებისა და კანონზომიერებების დადგენის** საშუალებას მოგვცემს.

ამიტომ, სოციოლოგი საკვლევი ობიექტის **მოდელირებას** (მოდელის აგებას) ახდენს:

ოპერაციულად განსაზღვრული ძირითადი ცნებების "ენით" ხორციელდება **საკვლევი ობიექტის ლოგიკურ-სისტემური ანალიზი** — საკვლევი ობიექტი განიხილება როგორც ერთიანი სისტემა. ხდება მისი მთლიანი სტრუქტურისა და ელემენტების აღწერა, მათი მნიშვნელობის, ურთიერთქმედების, განვითარების მექანიზმების გამოვლენა.

შედეგად, მიიღება საკვლევი ობიექტის **ფორმალიზებული მოდელი**[1] და განისაზღვრება კვლევის საგნის ძირითადი პარამეტრების/მახასიათებლების სისტემა.

მაგალითად, სასწავლო პროცესის დახასიათება შეიძლება ობიექტური და სუბიექტური მაჩვენებლებით.

სასწავლო პროცესის ობიექტური მახასიათებლების მაგალითებია:

- სტუდენტთა აკადემიური მოსწრება;
- ლექციებზე დასწრება;
- სამეცნიერო კონფერენციებში მონაწილეობის ინტენსივობა და წარმატების ხარისხი;
- მასწავლებელთა კვალიფიკაციის დონე;
- მასწავლებელთა შრომითი დისციპლინა (ლექციების დროული დაწყება, სალექციო დროის მიზნობრივი დატვირთვა და ა. შ.);

სასწავლო პროცესის სუბიექტური მახასიათებლების მაგალითებია:

- სტუდენტების დამოკიდებულება სასწავლო რეფორმების მიმართ;
- მასწავლებლების დამოკიდებულება სასწავლო რეფორმების მიმართ.

და ა. შ.

თუ მსჯელობის ამგვარ ლოგიკას გავყვებით, საბოლოოდ მივიღებთ სასწავლო პროცესის "ელემენტებად დაშლილ" **სქემატურ, ფორმალიზებულ** მოდელს, რომლის თითოეული ელემენტი გარკვეულ პარამეტრს (მახასიათებელს, ემპირიულ ინდიკატორს) გამოსახავს.

❖ ჰიპოთეზების როლი ემპირიულ სოციოლოგიურ გამოკვლევაში

გაზომვა ანუ საკვლევი პარამეტრის კონკრეტული მნიშვნელობის დადგენა კითხვა-პასუხის პროცესის ანალოგიურია.

შეკითხვა ინფორმაციის მოპოვების საშუალებაა და, ჩვეულებრივ, არცოდნას უკავშირდება. კითხვის დასმა ძირითადად მართლაც არცოდნითა და გაგების სურვილითაა

[1] ამ მოდელს კონცეპტუალურს, ჰიპოთეტურს, აპრიორულსაც უწოდებენ იმის აღსანიშნავად, რომ იგი ანალიტიკური მსჯელობის საფუძველზეა აგებული.

მოტივირებული — ვეკითხები იმიტომ, რომ არ ვიცი.

მაგრამ, მეორე მხრივ, იგი უკვე არსებული ცოდნიდან გამომდინარეობს — თუ არ ვიცი, ვერ ვეკითხავ: შეკითხვას ისე ვერ დავსვამ, თუ არ ვიცი, რა მაინტერესებს (რას ვეკითხები) და სად უნდა ვეძებო (ვის ვკითხო) ჩემთვის საინტერესო ინფორმაცია.

ეს ცხადია, ინფორმაციის მოპოვების იმ საშუალებებზეც ვრცელდება, რითაც სოციოლოგი გაზომვებს ანარმოებს:

- მან იცის, რა უნდა "იკითხოს" (ჰკითხოს, დააკვირდეს, ნაიკითხოს და ა. შ.) — როგორც უკვე ვნახეთ, საკვლევი პრობლემის განსაზღვრიდან დაწყებული, კვლევის თანდათანობითი დაკონკრეტება სწორედ ამის დადგენას ისახავს მიზნად.
- იცის, ვის უნდა "ჰკითხოს" — საიდან მიიღოს ინფორმაცია (ვინ გამოკითხოს, რას დააკვირდეს, რა ნაიკითხოს და ა. შ.).

ამ ცოდნისა და გამოცდილების, ესე იგი, უკვე არსებული ინფორმაციის საფუძველზე სოციოლოგი აყალიბებს **ჰიპოთეზებს**.

იმისათვის, რომ ნათლად დავინახოთ ჰიპოთეზის რაობა და როლი სოციოლოგიური კვლევის პროცესში, მივმართოთ ანალოგიას.

ხშირად გვაოცებს, საეჭვოდ ვთვლით ან საერთოდ დაუჯერებლად მიგვაჩნია ჩვენს კითხვაზე მიღებული პასუხი.

ასეთ რეაქციას ინვევს გარკვეული ვარაუდი, მოლოდინი იმის შესახებ თუ რა პასუხი **შეიძლებოდა** მიგვეღო.

მაგალითად, ვის არ მოუსმენია შემდეგი ტიპის დიალოგი:

გიორგი: რამდენი წლის ბრძანდებით?

ირაკლი: ორმოცის.

გიორგი: რას ბრძანებთ, ოცდაათზე მეტს არ მოგცემდით!

ამ მაგალითში (ვთვლით, რომ დიალოგის ორივე მონაწილე გულწრფელია) კარგად ჩანს **მოლოდინი, რომელიც პასუხის მიღებამდე არსებობდა**. კითხვის დამსმელს რომ მოლოდინი არ ჰქონოდა, არც გაოცდებოდა — გაოცებას იწვევს სწორედ ის, რომ მისი ვარაუდი მკვეთრად განსხვავდება მიღებული ინფორმაციისაგან.

შეიძლება ითქვას, რომ, ამ თვალსაზრისით, **შეკითხვა მოლოდინის შემოწმებაა.**

ამასთან, რაც უფრო კონკრეტული და მკაფიოდ განსაზღვრულია მოლოდინი, მით უფრო მიზანმიმართულია შეკითხვა.

მაგალითად, თუ ქალაქში ქუჩის მოძებნა გვინდა, როგორც წესი, ადრესატად უცხოელს არ შევარჩევთ, ადგილობრივი მაცხოვრებლის შემთხვევაში, ჯერ იმას დავაზუსტებთ, აქვს თუ არა მას დახმარების სურვილი და შესაძლებლობა, ცნობათა ბიუროში კი შეიძლება პირდაპირ ვიკითხოთ, როგორ მოვხვდეთ ამ ქუჩაზე.

ჩვენი ქცევა და კითხვების შინაარსი გარკვეულ მოლოდინებს შეესაბამება:

- უცხოელის მიმართ ნაკლებია მოლოდინი, რომ მისთვის უცხო ქალაქში გზა მიგვასწავლოს;
- უფრო მეტი მოლოდინი გვაქვს ადგილობრივი მაცხოვრებლის მიმართ, მაგრამ არც ისაა "ნარმოუდგენელი", რომ ადამიანმა არ იცოდეს ნებისმიერი ქუჩა საკუთარ ქალაქში ან არ სურდეს ჩვენთვის დროის დაკარგვა;
- ცნობათა ბიუროს ძირითადი ფუნქცია კი სწორედ მსგავსი ინფორმაციის გაცემაა. აქედან გამომდინარე, ამ შემთხვევაში ყველაზე მეტია მოლოდინი, რომ იქ ქუჩა მიგვასწავლონ.

ეს შესაბამისობა ემპირიული სოციოლოგიური კვლევის პროცესში არსებით როლს ასრულებს.

სოციოლოგს, ისევე როგორც ნებისმიერ სხვა ადამიანს, კითხვის ფორმულირებისას უჩნდება ვარაუდი იმის შესახებ, თუ რა პასუხია მოსალოდნელი არსებულ სიტუაციაში (პირობებში).

ზოგ შემთხვევაში შესაძლებელია ვარაუდის არა მხოლოდ გაცნობიერება და მკაფიო ფორმულირება, არამედ მისი არგუმენტირებაც.

ვარაუდის არგუმენტირების საფუძველზე ყალიბდება **ჰიპოთეზა** — დებულება, რომელშიც ფორმულირებულია ყველაზე უფრო მოსალოდნელი შედეგი (ყველაზე უფრო სავარუდო პასუხი დასმულ კითხვაზე).

ჰიპოთეზა შეიძლება ჩამოყალიბდეს არა მხოლოდ ცალკეული კითხვის, არამედ კითხვათა ჯგუფის (ბლოკის) და მთლიანად საკვლევი საკითხის შესახებაც, რადგან სოციალური პრობლემის დაყენება, ფაქტობრივად, კომპლექსური შეკითხვაა ამ პრობლემის რაობის, მნიშვნელობისა და გადაწყვეტის შესაძლებლობის თაობაზე.

ახლა უკვე შეგვიძლია ვიმსჯელოთ ჰიპოთეზის მნიშვნელობის შესახებ:

ჰიპოთეზა საშუალებას გვაძლევს მიზანმიმართულად ვარჩევთ გამოკითხვის პროცესი — ადეკვატურად შევარჩიოთ კითხვის ფორმა და შინაარსი, მისი მიცემის წესი.

ცხადია, ის ვარაუდი, რომლის გაცნობიერება, მკაფიო ფორმულირება ან არგუმენტირება არ ხერხდება, ჰიპოთეზის როლს ვერ შეასრულებს:

ყოველდღიურ ცხოვრებაში ჩვენი ქცევის მოტივაციის ზუსტი გაგება და დასაბუთება ხშირად არც არის საჭირო, მაგრამ ბუნდოვანი ან უსაფუძვლო ვარაუდების "გათვალისწინების" მცდელობამ კვლევა შეიძლება არასწორი მიმართულებით ნარმართოს.

და ბოლოს, თუ ყველა ჰიპოთეზას სისტემურად, თანმიმდევრულად გავაერთიანებთ, მივიღებთ **სამუშაო ჰიპოთეზას**.

ჰიპოთეზის მნიშვნელობიდან გამომდინარე, ძნელი არაა იმის მიხვედრა, რომ **სამუშაო ჰიპოთეზა შეასრულებს ათვლის წერტილის როლს, რომელთან შესაბამისობაშიც კოორდინირდება ინფორმაციის მოპოვების ძირითადი წესები და საშუალებები.**

მაგრამ, არსებობს უკუკავშირიც. ზემოთ უკვე აღვნიშნეთ, რომ შეკითხვა შეიძლება განვიხილოთ, აგრეთვე, როგორც მოლოდინის შემოწმების პროცესი, რადგან პასუხის მიღების შემდეგ "მოწმდება" (შესაძლებელი ხდება შემოწმდეს) **ჩვენი ვარაუდის ადეკვატურობა რეალობასთან**.

სოციოლოგიური კვლევებისას ეს ფაქტორი არანაკლებ მნიშვნელოვანია. ინფორმაციის მოპოვებისა და ანალიზის შემდეგ შესაძლებელი ხდება შემოწმდეს თვითონ სამუშაო ჰიპოთეზა.

რადგან სამუშაო ჰიპოთეზა ერთზე მეტი ცალკეული ჰიპოთეზისაგან შეიძლება შედგებოდეს, მისი შემოწმება ნიშნავს იმის განსაზღვრას, ჭეშმარიტია თუ მცდარი მასში შემავალი თითოეული ჰიპოთეზა.

ჰიპოთეზის ჭეშმარიტების დასაბუთებას ჰიპოთეზის **ვერიფიკაცია** ეწოდება, ხოლო იმის ჩვენებას, რომ ჰიპოთეზა მცდარია — **ფალსიფიკაცია**.

ჰიპოთეზების შემოწმებას არსებითი მნიშვნელობა აქვს კვლევის ძირითადი შედეგების დადგენისა და დასკვნითი ანალიზისათვის:

სწორად ნარმართული კვლევის პირობებში:

ა) ვერიფიცირებული ჰიპოთეზა დასაბუთებული დებულება, კვლევის ერთ-ერთი ძირითადი შედეგია.

ბ) ფალსიფიცირებული ჰიპოთეზა კი, ხშირ შემთხვევაში იმის მაჩვენებელია, რომ რეალური სიტუაცია არ შეესაბამება ლოგიკურად ყველაზე უფრო მოსალოდნელს და ამ შეუსაბამობის ახსნის ამოცანას აყენებს. ყალიბდება ვარაუდები იმის შესახებ, თუ რა ფაქტორებით შეიძლება იყოს გამოწვეული ასეთი სიტუაცია.

ახალი ვარაუდების საფუძველზე ფორმულირდება ახალი ჰიპოთეზები. მათ **ალტერნატიულ ჰიპოთეზებს** უწოდებენ.

ასე რომ, ჰიპოთეზის ფალსიფიცირების შემთხვევაში, კვლევა ალტერნატიული ჰიპოთეზის შემოწმებაზე ორიენტირდება.

მაგალითად, იქიდან გამომდინარე, რომ რეფორმების ერთ-ერთი ძირითადი მიზანია სასწაელო პროცესის გაუმჯობესება და რეფორმები უკვე მიმდინარეობს, შესაძლებელია

ჩამოვაყალიბოთ ჰიპოთეზა, რომ რეფორმირების პროცესში სასწავლო პროცესის ობიექტური მახასიათებლები (სტუდენტთა აკადემიური მოსწრება, ლექციებზე დასწრების მაჩვენებელი და სხვა) ამაღლდა.

თუ ჩვენი ჰიპოთეზა ჭეშმარიტი აღმოჩნდა, იგი კვლევის ძირითად შედეგებში შევა და შეიძლება რეფორმების წარმატებულობის ერთ-ერთ არგუმენტად გამოდგეს.

ხოლო მისი ფალსიფიცირების შემთხვევაში, უნდა დავიწყოთ იმ ფაქტორების კვლევა, რასაც შეიძლება ხელი შეეშალა ამ სიტუაციაში ლოგიკურად მოსალოდნელი შედეგისათვის.

ეს ფაქტორი შეიძლება იყოს თავად გარდაქმნების არაადეკვატურობა, ადეკვატური გარდაქმნების არასწორად გატარება, სუბიექტური დაინტერესება მასწავლებელთა გარკვეული ჯგუფის მხრიდან და ა. შ.

შედეგად მივიღებთ **ალტერნატიულ ვარაუდებს, მოხდება ალტერნატიული ჰიპოთეზის ფორმულირება** და დაიწყება მისი შემოწმება.

და ბოლოს, უნდა აღინიშნოს, რომ არის შემთხვევები, როდესაც ჰიპოთეზების ვერიფიკაცია მკვლევრის ღირსების, ფალსიფიკაცია კი ნაკლის მაჩვენებლად მიაჩნიათ.

ასეთი მიდგომა არასწორია:

ჯერ ერთი, ჭეშმარიტი დებულება არასწორი მსჯელობის შედეგადაც შეიძლება ჩამოყალიბდეს.

მეორე — ცხოვრება ლოგიკის კანონების მიხედვით არ ვითარდება და სავსებით დასაშვებია, რომ ყველაზე უფრო ლოგიკური ვარაუდი ძალიან შორს აღმოჩნდეს სოციალური რეალობისაგან.

❖ მოქმედების სტრატეგია

როგორ?

იმის შესაბამისად, თუ **რა, სად და რატომ** უნდა ვიკვლიოთ, იგეგმება **როგორ ვიმოქმედოთ.** აქედან გამომდინარე, კვლევის მომდევნო ეტაპებზე ხორციელდება:

- მონაცემთა ანალიზისა და პირველადი სოციოლოგიური ინფორმაციის მოპოვების მეთოდების განსაზღვრა.
- შერჩევის ტიპის განსაზღვრა.
- კვლევის ინსტრუმენტის შემუშავება.
- შერჩევითი ერთობლიობის განსაზღვრა.

ხოლო უშუალოდ ინფორმაციის მოპოვება, დამუშავება და ანალიზი შემდეგ ეტაპებს მოიცავს:

- პილოტაჟი (კვლევის ინსტრუმენტის შემოწმება);
- ინტერვიუერთა ტრენინგი (რომლის დროსაც ინტერვიუერი ინფორმაციის მოპოვების კონკრეტულ წესებს ეუფლება);
- საველე სამუშაოები (პირველადი სოციოლოგიური ინფორმაციის მოპოვება);
- მონაცემთა მომზადება ანალიზისათვის;
- მონაცემთა სტატისტიკური ანალიზი;
- მონაცემთა სოციოლოგიური ანალიზი, შედეგების ინტერპრეტაცია (შინაარსობრივი "გაგება", ახსნა).

❖ **შერჩევა. გენერალური ერთობლიობის არსებითი პარამეტრების შესახებ.**

ემპირიული სოციოლოგიური გამოკვლევის უმთავრესი მიზანია ინფორმაციის მოპოვება და გაანალიზება.

გენერალური ერთობლიობა, პოპულაცია (**population**) ეწოდება ყველა იმ ელემენტის[1] ერთობლიობას, რომლებსაც საკვლევი სოციალური პრობლემა მოიცავს და რომელთაც საერთო/ზოგადი პარამეტრები აქვთ[1].

ინფორმაცია გენერალური ერთობლიობის პარამეტრების შესახებ შესაძლებელია მივიღოთ სრული დაკვირვების/აღწერის მეშვეობით ან გენერალური ერთობლიობის გარკვეულ ნაწილზე — შერჩევით ერთობლიობაზე დაკვირვების შედეგად.

აღწერა (census) გულისხმობს ინფორმაციის მოპოვებას გენერალური ერთობლიობის **ყველა წევრის** შესახებ.

აღწერის შემთხვევაში, **გენერალური ერთობლიობის პარამეტრთა განსაზღვრა** (მათი მნიშვნელობების დადგენა) **ხდება უშუალოდ აღწერის შემდეგ**.

სოციალური პრობლემები ფართომასშტაბიანია, ამიტომ მათი კვლევა შეიძლება რეგიონის, მთელი ქვეყნის ან რამდენიმე ქვეყნის მოსახლეობას მოიცავდეს/ფარავდეს.

აქედან გამომდინარე, რიგ შემთხვევაში, საჭირო ხდება, **გენერალური ერთობლიობის** გარკვეული **ნაწილი/ქვესიმრავლე** შევარჩიოთ და კვლევის პროცესში მხოლოდ ეს ნაწილი ჩავრთოთ.

გენერალური ერთობლიობის იმ ნაწილს, რომელიც კვლევაში უშუალოდ მონაწილეობს, **შერჩევითი ერთობლიობა/შერჩევა (sample)** ეწოდება.

შერჩევის შემთხვევაში, გენერალური ერთობლიობის პარამეტრების განსაზღვრა **არ ხდება უშუალოდ ინფორმაციის/მონაცემების მოპოვების შემდეგ**:

თავდაპირველად განსაზღვრება **შერჩევითი ერთობლიობის პარამეტრები, რომლებსაც სტატისტიკებს (statistics)** უწოდებენ და მხოლოდ ამის შემდეგ, **სტატისტიკათა მნიშვნელობების საფუძველზე**, ხდება დასკვნების გამოტანა **გენერალური ერთობლიობის პარამეტრების მნიშვნელობების შესახებ**.

შერჩევითი ერთობლიობის ძირითად მახასიათებელია **რეპრეზენტატულობა** — შერჩევითი ერთობლიობის თვისება/უნარი, ასახოს გენერალური ერთობლიობის სტრუქტურა და თავისებურებები, გარკვეული (კვლევისათვის მნიშვნელოვანი) პარამეტრების პროპორციული განაწილება.

რეპრეზენტატულობა დამოკიდებულია შერჩევითი ერთობლიობის მოცულობაზე, შერჩევის პრინციპსა და განხორციელებაზე.

იქიდან გამომდინარე, რომ შერჩევითი ერთობლიობის ფარგლებში გენერალური ერთობლიობის პარამეტრთა მნიშვნელობების ზუსტი გამოთვლა შეუძლებელია, სტატისტიკების მნიშვნელობათა (ე. ი. შერჩევითი ერთობლიობის პარამეტრების მნიშვნელობათა) მიხედვით **ხდება იმის შეფასება**, თუ რა მნიშვნელობებს უნდა იღებდეს გენერალური ერთობლიობის შესაბამისი პარამეტრები.

ე. ი. შერჩევითი ერთობლიობის შემთხვევაში, საბოლოოდ ვიღებთ არა გენერალური ერთობლიობის პარამეტრების მნიშვნელობებს, არამედ, **პარამეტრების შეფასებით მნიშვნელობებს/შეფასებებს**.

შესაბამისად, დგება სიზუსტისა და სანდოობის მიხედვით სტატისტიკებისა და გენერალური ერთობლიობის პარამეტრთა მიმართების დახასიათების ამოცანა.

სიზუსტის ხარისხი (presicion level) — განისაზღვრება მაქსიმალურად დასაშვები განსხვა-

[1] წევრის, კომპონენტის

ვებით/გადახრით შერჩევითი ერთობლიობის **სტატისტიკასა და გენერალური ერთობლიობის პარამეტრის მნიშვნელობას** შორის.

მაქსიმალურად დასაშვებ განსხვავებას შერჩევითი ერთობლიობის **სტატისტიკასა და გენერალური ერთობლიობის პარამეტრის მნიშვნელობას** შორის, **შერჩევის შეცდომა/შერჩევის ცდომილება ეწოდება.**

სიზუსტის ხარისხი დამოკიდებულია შერჩევითი ერთობლიობის მოცულობაზე: რაც უფრო მეტია შერჩევითი ერთობლიობის მოცულობა, მით უფრო მაღალია სიზუსტის ხარისხი ანუ მით უფრო ზუსტად ასახავს ემპირიული შედეგები გენერალური ერთობლიობის პარამეტრებს.

სანდოობის ინტერვალი (confidence interval) — ეწოდება დიაპაზონს/ინტერვალს, რომელშიც უნდა მოთავსდეს გენერალური ერთობლიობის პარამეტრის შეფასებითი მნიშვნელობა/შეფასება სანდოობის დონის ფიქსირებული (მოცემული) მნიშვნელობის პირობებში.

რაც უფრო მცირეა სანდოობის ინტერვალის ზომა, მით მაღალია სიზუსტის ხარისხი. შესაბამისად, სანდოობის ინტერვალიც დამოკიდებულია შერჩევითი ერთობლიობის მოცულობაზე: რაც მეტია შერჩევითი ერთობლიობის მოცულობა, მით უფრო მცირეა სანდოობის ინტერვალის ზომა.

სანდოობის დონე (confidence level) — გამოხატავს იმის ალბათობას, რომ გენერალური ერთობლიობის პარამეტრის შეფასებითი მნიშვნელობა/შეფასება მოხვდება სანდოობის ფიქსირებულ/მოცემულ ინტერვალში.

შერჩევითი ერთობლიობის განსაზღვრა საკმაოდ რთული პროცესია. იგი წინასწარ იგეგმება და შემდეგი, თანმიმდევრული ეტაპების სახით ხორციელდება:

- გენერალური ერთობლიობის განსაზღვრა;
- შერჩევის საფუძვლის განსაზღვრა;
- შერჩევის მეთოდების განსაზღვრა;
- შერჩევის მოცულობის განსაზღვრა;
- შერჩევის პროცესის განხორციელება.

გენერალური ერთობლიობის განსაზღვრა ხდება საკვლევი პრობლემის, კვლევის ამოცანებისა და მიზნებიდან გამომდინარე.

შერჩევის საფუძველი (sampling frame) გენერალურ ერთობლიობას უნდა ასახავდეს. ეს შეიძლება იყოს გენერალური ერთობლიობის სია ან ინსტრუქცია, რომლის მიხედვითაც შესაძლებელია ერთმნიშვნელოვნად განვსაზღვროთ გენერალური ერთობლიობის ყველა ელემენტი.

შერჩევის საფუძვლის მაგალითებია: მოსახლეობის საყოველთაო აღწერის მონაცემები, სატელეფონო ცნობარი, ამომრჩეველთა სიები და ა. შ.

შერჩევის მეთოდები

შერჩევის მეთოდები ორ ძირითად ჯგუფად იყოფა:

- შერჩევის დეტერმინირებული (არაალბათური) მეთოდები (**nonprobability sampling**).
- შერჩევის ალბათური მეთოდები (**probability sampling**).

შერჩევის დეტერმინირებული (არაალბათური) მეთოდების ფარგლებში შერჩევითი ერთობ-

---

[1] ე. ი. რომელთა დახასიათებაც შესაძლებელია ერთიანად/"მთლიანობაში", ზოგადი პარამეტრების მეშვეობით.

ლიობის შემადგენლობა, ძირითადად, მკვლევრის შინაარსობრივი არგუმენტაციის საფუძველზე განისაზღვრება.

იქიდან გამომდინარე, რომ გენერალური ერთობლიობიდან ელემენტები ალბათური წესით არ ამოირჩევა, ე. ი. უცნობია გენერალური ერთობლიობის ცალკეული ელემენტის შერჩევით ერთობლიობაში მოხვედრის ალბათობა, კვლევის შედეგების სიზუსტე ობიექტურად ვერ ფასდება, არ ხდება მათი გენერალურ ერთობლიობაზე გავრცელება/განზოგადება.

შერჩევის დეტერმინირებული მეთოდების ძირითადი სახეებია:

- არარეპრეზენტატული შერჩევა;
- ზედაპირული შერჩევა;
- კვოტური შერჩევა;
- შერჩევა "თოვლის გუნდის" პრინციპით.

შერჩევის ალბათური მეთოდების ფარგლებში შერჩევითი ერთობლიობის ყოველი ელემენტი გენერალური ერთობლიობიდან შემთხვევითად ამოირჩევა, რაც შესაძლებლობას გვაძლევს, შევაფასოთ კვლევის შედეგების სიზუსტე და განვსაზღვროთ, რამდენად დასაშვებია მათი განზოგადება გენერალურ ერთობლიობაზე.

შერჩევის ალბათური მეთოდების ძირითადი სახეებია:

- მარტივი შემთხვევითი შერჩევა;
- სისტემატური შერჩევა;
- სტრატიფიცირებული შერჩევა;
- კლასტერული შერჩევა;
- თანმიმდევრული შერჩევა;
- ორმაგი კონტროლის შერჩევა.

შერჩევის დეტერმინირებული მეთოდები

არარეპრეზენტატული შერჩევა (convenience sampling).

არარეპრეზენტატული შერჩევის პრინციპია შერჩევითი ერთობლიობის კონსტრუირება რაც შეიძლება იოლად ხელმისაწვდომი/ამოსარჩევი ელემენტებისაგან. ელემენტთა ამორჩევას, უმეტეს შემთხვევაში, ინტერვიუერი ახორციელებს. საკმაოდ ხშირად, შერჩევითი ერთობლიობის ელემენტები ხდებიან ის ადამიანები, რომლებიც "საჭირო დროსა და საჭირო ადგილას" აღმოჩნდნენ.

არარეპრეზენტატული შერჩევის მაგალითებია: სტუდენტების გამოკითხვა მათი თავშეყრის ადგილებში, გამვლელების გამოკითხვა ქუჩაში, აქციის მონაწილეთა გამოკითხვა აქციის მსვლელობისას, გამოკითხვა ჟურნალ-გაზეთებში მოთავსებული ანკეტებით და სხვა.

ფინანსური და დროის რესურსების თვალსაზრისით, შერჩევის ეს მეთოდი, უდავოდ, ყველაზე უფრო ეკონომიურია, მაგრამ არარეპრეზენტატული შერჩევა, როგორც წესი, **არ წარმოადგენს გარკვეულ ერთიანობას/მთლიანობას** (ფაქტობრივად, არ არის განსაზღვრული გენერალური ერთობლიობა), რომლის ფარგლებშიც შესაძლებელია ვიმსჯელოთ მიღებული მონაცემების შესახებ.

მისი გამოყენება, ძირითადად, იმ შემთხვევებშია მიზანშეწონილი, როდესაც:

- კვლევა ევრისტიკულ ხასიათს ატარებს და ორიენტირებულია ახალი შეხედულებების, იდეების, ჰიპოთეზების გამოვლენაზე;
- ხდება ფოკუს-ჯგუფის მომზადება;
- ხორციელდება ექსპერიმენტული კვლევები.

ნებისმიერ შემთხვევაში, ამ მეთოდით მოპოვებული მონაცემების ანალიზის პროცესში უნდა გამოვიჩინოთ სიფრთხილე მიღებული შედეგების მნიშვნელოვნების, მათი მასშტაბურობის შეფასების თვალსაზრისით.

ზედაპირული შერჩევა (judgemental sampling).

ზედაპირული შერჩევა არარეპრეზენტატული შერჩევის სახესხვაობაა. შერჩევითი ერთობლიობის ელემენტების ამორჩევა ხორციელდება მკვლევრის მოსაზრებების საფუძველზე, რომელიც საკუთარი ცოდნიდან და გამოცდილებიდან გამომდინარე, თვლის, რომ **გარკვეული კატეგორიის ადამიანების გამოკითხვა უფრო ეფექტური იქნება კვლევის ამოცანების გადასაწყვეტად**.

შესაბამისად, ამ მეთოდის ნარმატებულობა უმთავრესად მკვლევრის კომპეტენტურობაზეა დამოკიდებული.

ზედაპირული შერჩევის მაგალითებად შეიძლება დავასახელოთ შერჩევითი ერთობლიობის ელემენტების ამორჩევა მმართველი პარტიის "ერთგული" ამომრჩევლებიდან (რომლებიც, მმართველ პარტიას აძლევენ ხმას, დავუშვათ, არანაკლებ 15 წლის განმავლობაში), მოსამართლეების ან მენარმეების გამოკითხვა და ა. შ.

კვოტური შერჩევა (quota sampling).

კვოტური შერჩევა შეიძლება განვიხილოთ, როგორც ორეტაპიანი შეზღუდული ზედაპირული შერჩევა.

პირველ ეტაპზე ხდება ეგრეთ წოდებული საკონტროლო ჯგუფების ანუ ქვოტების გამოყოფა კვლევის საგნის საკონტროლო მახასიათებლების/პარამეტრების საფუძველზე და განისაზღვრება მათი განაწილება გენერალურ ერთობლიობაში.

საკონტროლო მახასიათებელი შეიძლება იყოს ნებისმიერი ნიშან-თვისება: სქესი, ასაკი, ეროვნება, მიკუთვნებულობა ამა თუ იმ სოციალური ჯგუფისადმი და ა. შ., რომელთა გამოყოფაც, მკვლევრის აზრით, აუცილებელია კვლევის ადეკვატურად წარმართვისა და კვლევის საგნის უკეთ/უფრო დეტალურად შესწავლის თვალსაზრისით.

როგორც ნესი, ქვოტების დადგენა ხდება იმ პრინციპით, რომ შერჩევის კვოტური განაწილება პროპორციულად ასახავს გენერალური ერთობლიობის კვოტურ განაწილებას ანუ, შერჩევითი ერთობლიობის იმ ელემენტთა პროცენტული მაჩვენებელი (მთლიან შერჩევით ერთობლიობასთან ფარდობაში), რომლებიც კვოტური/საკონტროლო ნიშან-თვისების ამა თუ იმ მნიშვნელობით ხასიათდებიან, ემთხვევა გენერალური ერთობლიობის იმ ელემენტთა პროცენტულ მაჩვენებელს (მთლიან გენერალურ ერთობლიობასთან ფარდობაში), რომელთაც ეს მნიშვნელობა ასევე ახასიათებთ.

მეორე ეტაპზე ხდება შერჩევის ელემენტთა ამორჩევა. ამორჩევის პრინციპს ძირითადად განსაზღვრავს მკვლევრი იმ პირობის დაცვით, რომ ელემენტები საკონტროლო მახასიათებლებს შეესატყვისებოდეს.

შერჩევის ეს მეთოდი ხშირ შემთხვევაში ეფექტურია და თვალსაჩინოდ ასახავს გენერალური ერთობლიობის სტრუქტურას საკონტროლო მახასიათებლების გა-თვალისწინებით.

მეთოდის სუსტი მხარეა ის, რომ იგი არ იძლევა რეპრეზენტატულობის გარანტიებისა და შერჩევის ცდომილების/სიზუსტის ხარისხის განსაზღვრის საშუალებას.

შერჩევა "თოვლის გუნდის" პრინციპით (snowball sampling).

შერჩევა "თოვლის გუნდის" პრინციპით შემდეგნაირად ხორციელდება: თავდაპირველად შემთხვევითად ირჩევენ რესპონდენტთა საწყის ჯგუფს, შემდეგ კი საწყისი ჯგუფის წევრებს სთხოვენ მოაძებნინონ გამოკითხვის შესაფერისი კანდიდატები, შემდეგ ამ უკანასკნელთ მიმართავენ იგივე თხოვნით და ა. შ. რის შედეგადაც მიიღება თოვლის გუნდის (რომელიც

თოვლში ტრიალისას იზრდება) ეფექტი.

მართალია, საწყის ეტაპზე შერჩევა შემთხვევითად ხდება, მაგრამ მომდევნო ეტაპიდან დაწყებული, რესპონდენტთა ამორჩევა უკვე დეტერმინირებულად მიმდინარეობს. აქედან გამომდინარე, შერჩევა, **მთლიანობაში, დეტერმინირებულ** ხასიათს იძენს.

ამ მეთოდის გამოყენების ძირითადი მოტივია გენერალური ერთობლიობისათვის **არატიპური მახასიათებლების** გამოვლენა და მოხერხებულია **მცირე, დახურული ან მარგინალური ჯგუფების** კვლევის დროს.

შერჩევის ალბათური მეთოდები

მარტივი შემთხვევითი შერჩევა (Simple Random sampling – SRS).

მარტივი შემთხვევითი შერჩევის დროს, გენერალური ერთობლიობის ყოველ ელემენტს აქვს ცნობილი და თანაბარი შანსი იმისათვის, რომ იგი შერჩევით ერთობლიობაში მოხვდეს.

გარდა ამისა, **n** მოცულობის მქონე (**n** ელემენტიან) ყოველ შესაძლო შერჩევას ასევე გააჩნია ცნობილი და თანაბარი შანსი იმისათვის, რომ იგი რეალური შერჩევითი ერთობლიობა გახდეს.

მარტივი შემთხვევითი შერჩევა გულისხმობს:

- ყოველი ელემენტის ამორჩევას ერთმანეთისაგან დამოუკიდებლად;
- შერჩევითი ერთობლიობის კონსტრუირებას შერჩევის საფუძვლიდან ელემენტების შემთხვევითად/"ჩაურეველი" ამორჩევით.

ელემენტების შემთხვევითად ამორჩევის უზრუნველყოფა შესაძლებელია სტატისტიკური კომპიუტერული პროგრამების მეშვეობით.

მარტივი შემთხვევითი შერჩევა რიგი უპირატესობებით გამოირჩევა: იოლია მისი დაგეგმვა და იძლევა მიღებული შედეგების გენერალურ ერთობლიობაზე განზოგადების საშუალებას.

მარტივი შემთხვევითი შერჩევის უარყოფით მხარეებს შორის, მონაცემთა ანალიზის თვალსაზრისით, აღსანიშნავია ის, რომ ასეთი შერჩევის საფუძველზე მიღებული შედეგები ალბათური შერჩევის სხვა მეთოდებით მიღებულ შედეგებთან შედარებით, **ნაკლები სიზუსტითა და მეტი სტანდარტული ცდომილებით** ხასიათდება.

სისტემატური შერჩევა (systematic sampling).

სისტემატური შერჩევა გულისხმობს გარკვეული პრინციპით დალაგებული/"გადათვლილი" გენერალური ერთობლიობიდან თავდაპირველად სანყისი ნერტილის/პირველი ელემენტის შემთხვევითად გამოყოფას, შემდეგ კი, საწყისი ნერტილის მიმართ ყოველი მე-i-ურე წევრის ამორჩევას.

შერჩევის ინტერვალის ზომა/ბიჯი (**i**) განისაზღვრება გენერალური ერთობლიობის მოცულობისა (**N**) და შერჩევითი ერთობლიობის მოცულობის (**n**) თანაფარდობით (დამრგვალება ხდება უახლოეს მთელ რიცხვამდე).

სისტემატური შერჩევის განხორციელების მაგალითი:

დავუშვათ, გენერალური ერთობლიობის მოცულობაა 1 000 000, ხოლო შერჩევითი ერთობლიობის მოცულობა, 1 000.

მაშინ ინტერვალის/ბიჯის ზომა ტოლი იქნება 1 000 000/ 1000 = 1000.

ხოლო თუ საწყისი წერტილისათვის შემთხვევითად განისაზღვრა რიცხვი ოცდახუთი, მაშინ შერჩევის ელემენტების ამორჩევა უნდა მოხდეს შემდეგი წესით:

- შერჩევის პირველ ელემენტად ამოირჩევა გენერალური ერთობლიობის ოცდამეხუთე წევრი;
- ყოველი მომდევნო ელემენტი ნინას ათასით უნდა აღემატებოდეს. ე. ი. ამოირჩევა გენერალური ერთობლიობის 1025-ე, 2025-ე, 3025-ე, 4025-ე, 5025-ე და ა. შ. წევრები.

ისევე, როგორც მარტივი შემთხვევითი შერჩევის დროს, სისტემატური შერჩევის

პირობებში, გენერალური ერთობლიობის ყოველ ელემენტს შერჩევით ერთობლიობაში მოხვედრის ცნობილი და თანაბარი შანსი აქვს, მაგრამ n მოცულობიან შესაძლო შერჩევებს შორის რეალურ შერჩევად გადაქცევის ცნობილი და თანაბარი შანსი აქვთ მხოლოდ იმ შერჩევებს, რომლებიც საწყისი ნერტილის ამორჩევისა და შერჩევის ინტერვალის/ბიჯის მოცემულ პირობებს აკმაყოფილებს.

დანარჩენი n მოცულობიანი შერჩევების რეალურ შერჩევად გადაქცევის ალბათობა ნულის ტოლია.

სისტემატური შერჩევის პირობებში, როგორც წესი, საკმაოდ მაღალი რეპრეზენტატულობის მქონე შედეგები მიიღება.

სტრატიფიცირებული შერჩევა (stratified sampling).

სტრატიფიცირებული/ფენობრივი შერჩევა ორ ძირითად ეტაპს მოიცავს.

პირველ ეტაპზე ხდება გენერალური ერთობლიობის სტრატებად, ფენებად დაყოფა შემდეგი პრინციპების დაცვით:

- სტრატები უნდა იყოს ურთიერთგამომრიცხავი, ე. ი. არ უნდა არსებობდეს გენერალური ერთობლიობის ისეთი წევრი, რომელიც ერთდროულად ორ ან მეტ სტრატში შევა;
- სტრატების გაერთიანება უნდა გვაძლევდეს გენერალურ ერთობლიობას, ე. ი. არ უნდა არსებობდეს გენერალური ერთობლიობის ისეთი წევრი, რომელიც ერთ-ერთ სტრატში მაინც არ შევა.

მეორე ეტაპზე ხდება ელემენტების ამორჩევა თითოეული სტრატიდან შემთხვევითი შერჩევის წესით (თეორიულად, სტრატებიდან ელემენტების ამორჩევა მარტივი შემთხვევითი შერჩევით უნდა განხორციელდეს, მაგრამ, პრაქტიკაში, ზოგჯერ სისტემატურ შერჩევასაც მიმართავენ).

ცვლადებს, რომელთა მეშვეობითაც გენერალური ერთობლიობის სტრატებად დაყოფა ხორციელდება, **მასტრატიფიცირებელი ცვლადები** ეწოდება.

მასტრატიფიცირებელმა ცვლადებმა უნდა უზრუნველყოს, რომ:

- ერთ სტრატში შემავალი ელემენტები ხასიათდებოდნენ მაქსიმალური ერთგვაროვნებით (რაც შეიძლება მეტად უნდა "ჰგავდნენ" ერთმანეთს);
- განსხვავებულ სტრატებში შემავალი ელემენტები ხასიათდებოდნენ მაქსიმალური განსხვავებულობით (რაც შეიძლება მეტად უნდა განსხვავდებოდნენ ერთმანეთისაგან).

მასტრატიფიცირებელ ცვლადებად შეიძლება შეირჩეს დემოგრაფიული ცვლადები (ასაკი, სქესი და ა. შ.) ან კვლევისათვის მნიშვნელოვანი ნებისმიერი სხვა ცვლადი, რომელიც სტრატიფიკაციის მოთხოვნებს დააკმაყოფილებს.

სტრატიფიცირებული შერჩევის ერთ-ერთი არსებითი პრობლემა **სტრატების მოცულობის განსაზღვრაა**.

იმის მიხედვით, თუ რა პრინციპით განისაზღვრება სტრატების მოცულობა, სტრატიფიცირებული შერჩევა შეიძლება იყოს **პროპორციული ან არაპროპორციული**.

პროპორციული სტრატიფიცირებული შერჩევის დროს, შერჩევის სტრატებად განანილება პროპორციულად ასახავს გენერალური ერთობლიობის სტრატებად განანილებას.

ეს იმას ნიშნავს, რომ შერჩევითი ერთობლიობის ამა თუ იმ სტრატის ელემენტთა პროცენტული მაჩვენებელი (მთლიან შერჩევასთან ფარდობაში), ემთხვევა იმ ელემენტების პროცენტულ მაჩვენებელს (მთლიან გენერალურ ერთობლიობასთან ფარდობაში), რომლებიც გენერალური ერთობლიობის ანალოგიურ სტრატს მიეკუთვნებიან.

არაპროპორციული სტრატიფიცირებული შერჩევა ერთდროულად ორ გარემოებას

ითვალისწინებს:

- რაც უფრო დიდია სტრატი, მით მეტ გავლენას ახდენს იგი გენერალური ერთობლიობის პარამეტრთა **საშუალო მნიშვნელობების** ფორმირებაზე, შესაბამისად, იგი მეტი მოცულობით უნდა იქნეს წარმოადგენილი შერჩევით ერთობლიობაში;
- რაც უფრო მეტია საკვლევი პარამეტრის განაწილების საშუალო კვადრატული გადახრა სტრატის შიგნით, მით მეტი მოცულობით უნდა იქნეს წარმოდგენილი ეს სტრატი შერჩევაში, რათა შედეგების **სიზუსტის** ალბათობა ამაღლდეს.

აქედან გამომდინარე, არაპროპორციული სტრატიფიცირებული შერჩევის პირობებში შერჩევით ერთობლიობაში სტრატის მოცულობა ერთდროულად ორი მოცემულობის: გენერალურ ერთობლიობაში ამ სტრატის ხვედრითი წილისა და სტრატის ფარგლებში საკვლევი პარამეტრის საშუალო კვადრატული გადახრის პროპორციულად განისაზღვრება.

შესაბამისად, იმ შემთხვევაში, თუ საკვლევი პარამეტრის განაწილების საშუალო კვადრატული გადახრები ყველა სტრატისათვის ერთნაირი აღმოჩნდება, მაშინ პროპორციული და არაპროპორციული სტრატიფიცირებული შერჩევის რეზულტატები ერთმანეთს დაემთხვევა.

არაპროპორციული განაწილების მნიშვნელოვანი პრობლემა ისაა, რომ ხშირად, კვლევის ჩატარებამდე, არ არსებობს მონაცემები საკვლევი პარამეტრის განაწილების შესახებ.

ასეთ შემთხვევებში, ზოგიერთი მკვლევარი შინაარსობრივად არგუმენტირებული სავარაუდო განაწილების მოდელით ხელმძღვანელობს.

სტრატიფიცირებული შერჩევის არსებით უპირატესობებს შორის აღსანიშნავია, რომ იგი:

- ხელს უწყობს კვლევის შედეგების სიზუსტის გაზრდას;
- უზრუნველყოფს შერჩევით ერთობლიობაში ყველა მნიშვნელოვანი სტრატის ადეკვატურად ასახვას.

კლასტერული შერჩევა (cluster sampling).

კლასტერული შერჩევა, ისევე, როგორც სტრატიფიცირებული, ორ ძირითად ეტაპს მოიცავს.

პირველ ეტაპზე ხდება გენერალური ერთობლიობის კლასტერებად დაყოფა შემდეგი პრინციპების დაცვით:

- კლასტერები უნდა იყოს ურთიერთგამომრიცხავი, ე. ი. არ უნდა არსებობდეს გენერალური ერთობლიობის ისეთი წევრი, რომელიც ერთდროულად ორ ან მეტ კლასტერში შევა;
- კლასტერების გაერთიანება უნდა გვაძლევდეს გენერალურ ერთობლიობას, ე. ი. არ უნდა არსებობდეს გენერალური ერთობლიობის ისეთი წევრი, რომელიც ერთ-ერთ კლასტერში მაინც არ შევა.

მეორე ეტაპზე ხდება კლასტერების ამორჩევა და შერჩევითი ერთობლიობის ფორმირება მიმდინარეობს **მხოლოდ ამორჩეული კლასტერების ფარგლებში.**

შერჩევითი ერთობლიობის ფორმირების წესის მიხედვით, კლასტერული შერჩევა შეიძლება იყოს ერთსაფეხურიანი, ორსაფეხურიანი ან მრავალსაფეხურიანი.

ერთსაფეხურიანი კლასტერული შერჩევის დროს, შერჩევითი ერთობლიობის ელემენტი ხდება **თითოეული ამორჩეული კლასტერის ყოველი წევრი.**

ორსაფეხურიანი კლასტერული შერჩევის დროს, შერჩევითი ერთობლიობის ფორმირება ხდება ამორჩეული კლასტერებიდან ალბათური მეთოდით.

ორსაფეხურიანი კლასტერული შერჩევის ფარგლებში, თავის მხრივ, განასხვავებენ **მარტივ შერჩევასა და ალბათურ შერჩევას მოცულობის პროპორციულად (PPS).**

მარტივი ორსაფეხურიანი კლასტერული შერჩევის დროს, შერჩევითი ერთობლიობის

ელემენტების ამორჩევა ხდება მარტივი შემთხვევითი შერჩევის მეთოდით. კლასტერებიდან ამორჩეული ელემენტების რაოდენობა თანაბარია. ამ მეთოდს, ძირითადად, ტოლი მოცულობის კლასტერების შემთხვევებში მიმართავენ.

ალბათური შერჩევა მოცულობის პროპორციულად (Probabilty Proportionate to size Sampling — PPS) გულისხმობს შერჩევითი ერთობლიობის ელემენტების შემთხვევითად ამორჩევას კლასტერებიდან მათი მოცულობების პროპორციულად.

მრავალსაფეხურიანი კლასტერული შერჩევა, ფაქტობრივად, რამდენიმე შერჩევის ერთობლიობაა. იგი ხორციელდება თანმიმდევრულად, უფროდაუფრო მცირე მოცულობის კლასტერების ფორმირებით. კლასტერების ფორმირების ანუ კლასტერიზაციის პროცესი წყდება იმ საფეხურზე, როდესაც, გარკვეული კრიტერიუმის მიხედვით, ჩაითვლება, რომ შერჩევის საბოლოო მოდელი უკვე მიღებულია.

სტრატებისადმი ნაყენებული მოთხოვნის საპირისპიროდ, ერთ კლასტერში შემავალი ელემენტები რაც შეიძლება მრავალფეროვანი, ხოლო კლასტერები (ერთმანეთთან მიმართებაში) რაც შეიძლება ერთგვაროვანი უნდა იყოს.

საუკეთესო ვარიანტია, რომ თითოეული კლასტერი გენერალური ერთობლიობის მცირე მოდელს წარმოადგენდეს, რადგან ამით მნიშვნელოვნად გაიზრდება შერჩევის ეფექტურება.

აქ შეიძლება აღვნიშნოთ კიდევ ერთი არსებითი განსხვავება სტრატიფიცირებულ და კლასტერულ შერჩევებს შორის:

სტრატიფიცირებული შერჩევის ძირითადი მიზანია შედეგების სიზუსტის მაქსიმიზება (რაც შეიძლება გაზრდა), ხოლო კლასტერიზაციის ძირითად მიზანი — შერჩევის ეფექტურობის მაქსიმიზება მინიმალური ძალისხმევის პირობებში (იგულისხმება ადამიანური, ფინანსური, დროის, და ა. შ. რესურსების მინიმალური დანახარჯები).

კლასტერული შერჩევის ერთ-ერთი ყველაზე გავრცელებული ფორმაა **ტერიტორიული შერჩევა (area sampling)**.

ტერიტორიული შერჩევის ფარგლებში კლასტერები **გეოგრაფიული ერთეულებისაგან** შედგება. ეს შეიძლება იყოს მხარე, რეგიონი, რაიონი, საკრებულო, კვარტალი და ა. შ.

ერთსაფეხურიანი ტერიტორიული შერჩევის დროს, შერჩევითი ერთობლიობის ელემენტები შეიძლება გახდეს, მაგალითად, ამორჩეულ კვარტლებში მცხოვრები ყველა ოჯახი.

ორსაფეხურიანი შერჩევის დროს ამორჩეული კვარტლებიდან ხდება ოჯახების ალბათური ამორჩევა ტოლი რაოდენობით ან ამორჩეული კვარტლების მოცულობის (ანუ, ამ კვარტლებში მცხოვრები ოჯახების საერთო რაოდენობათა) პროპორციულად.

კლასტერული მეთოდი საშუალებას გვაძლევს გენერალური ერთობლიობა შევისწავლოთ კლასტერების და არა ელემენტების დონეზე, რაც, როგორც აღვნიშნეთ, ადამიანური, დროის და ფინანსური რესურსების მნიშვნელოვან ეკონომიასთანაა დაკავშირებული, მაგრამ მეთოდის სისუსტე ისაა, რომ პრობლემურია შერჩევის სიზუსტის მაღალი ხარისხის მიღწევა და მიღებული შედეგების შეფასება.

თანმიმდევრული შერჩევა (sequential sampling).

თანმიმდევრული შერჩევა მრავალეტაპობრივია. იგი თანმიმდევრულად ხორციელდება. ამასთან, ყოველ ეტაპზე მიმდინარეობს ინფორმაციის მოპოვება და ანალიზი, რის საფუძველზეც ხდება გადაწყვეტილების მიღება გენერალური ერთობლიობიდან ელემენტების დამატებითი ამორჩევის შესახებ.

აქედან გამომდინარე, თანმიმდევრული შერჩევის დროს, შერჩევითი ერთობლიობის მოცულობა კვლევის დასაწყისში განსაზღვრული არ არის: მისი დაზუსტება შერჩევის ეტაპების პარალელურად, თანმიმდევრულად მიმდინარეობს.

შერჩევის საწყის ეტაპზე განისაზღვრება შერჩევითი ერთობლიობის ფორმირების კრიტერიუმები. სწორედ ამ კრიტერიუმების მიხედვით ხდება შერჩევის (შესაბამისად, კვლევის პროცესის) გაგრძელება ან შეწყვეტა.

ასეთი კრიტერიუმის როლი შეიძლება შეასრულოს კვლევის კონკრეტულმა ამოცანამ. მაგალითად, დასაშვებია, რომ შერჩევითი ერთობლიობის მოცულობის თანმიმდევრული ზრდა გაგრძელდეს მანამდე, სანამ არ გადაწყდება კვლევის ამოცანა, დავუშვათ, სანამ არ ვუპასუხებთ კითხვას, რა ძირითადი ფაქტორები განსაზღვრავს დამოკიდებულებას თბილისის სახელმწიფო უნივერსიტეტში მიმდინარე რეფორმების მიმართ.

ორმაგი კონტროლის შერჩევა (double sampling).

ორმაგი კონტროლის შერჩევა (მას ორეტაპობრივ შერჩევის მეთოდსაც უწოდებენ) გულისხმობს გენერალური ერთობლიობიდან ზოგიერთი ელემენტის ორჯერ ან რამდენჯერმე ამორჩევას.

ორმაგი კონტროლის შერჩევის გამოყენება მიზანშეწონილია მაშინ, როდესაც ზუსტად არ არის ცნობილი შერჩევის საფუძველი, მაგრამ ვიცით, რომ იგი სხვა, უფრო დიდი მოცულობის შერჩევის საფუძველში შედის.

ასეთ შემთხვევაში, ეს მეთოდი გვეხმარება კვლევისათვის საჭირო ელემენტების გამოვლენასა და მათ შესახებ ინფორმაციის მოძიებაში:

- პირველ ეტაპზე განხორციელებული შერჩევის საფუძველზე ხდება შერჩევითი ერთობლიობის ელემენტების შესახებ ინფორმაციის მოპოვება;
- მომდევნო ეტაპზე ხორციელდება შერჩევითი ერთობლიობიდან ელემენტთა გარკვეული, **კვლევისათვის მნიშვნელოვანი ნაწილის გამოყოფა** და ხდება შერჩევითი ერთობლიობის მოცულობის შემცირება — ინფორმაციის მოპოვების შემდგომი პროცესი მხოლოდ გამოყოფილ (ანუ, კვლევისათვის მნიშვნელოვნად მიჩნეულ) ელემენტებს მოიცავს.

მაგალითად, თუ ჩვენ გვაინტერესებს დედაქალაქის მოსახლეობის იმ ნაწილის გამოკითხვა, რომელიც გარკვეულ პოლიტიკურ პარტიას (პირობითად მას **A** პარტია ვუწოდოთ) უჭერს მხარს, მაგრამ არ ვიცით, როგორ შევარჩიოთ ასეთი რესპონდენტები, შესაძლებელია, რომ:

- პირველ ეტაპზე ალბათური წესით ამოვარჩიოთ ელემენტები საერთოდ დედაქალაქში მცხოვრები მოსახლეობიდან და ჩავატაროთ მათი გამოკითხვა;
- კვლევის მეორე ეტაპისათვის პირველი გამოკითხვის შედეგად მიღებული მონაცემების საფუძველზე გამოვყოთ **A** პარტიის მხარდამჭერები და შემდგომი კვლევა თავდაპირველი შერჩევითი ერთობლიობის მხოლოდ ამ ნაწილზე გავავრცელოთ.

ეს პროცესი რამდენიმე ეტაპად შეიძლება მიმდინარეობდეს როგორც პარალელურად, ისე თანმიმდევრულად.

❖ მონაცემთა მომზადება ანალიზისათვის

მონაცემთა ანალიზის კონკრეტულ მეთოდებს უშუალოდ საველე სამუშაოების შედეგად მიღებული ინფორმაციის მიმართ ვერ გამოვიყენებთ.

ჯერ ის უნდა შევამოწმოთ, რამდენად შეესაბამება მოპოვებული მონაცემები კვლევის მოთხოვნებს (პირობებს): დაცულია თუ არა შერჩევის პრინციპები, შეკითხვათა თანმიმდევრობა (ეგრეთ წოდებული "გადასვლები" კითხვიდან კითხვაზე), პასუხების შინაარსობრივი ლოგიკა და ა. შ.

გარდა ამისა, საჭიროა შეიქმნას მონაცემთა ბაზა მოკვლეული ინფორმაციისა და კვლევის ამოცანების ადეკვატურ კომპიუტერულ პროგრამაში[1], რომლის მეშვეობითაც ამ ინფორმაციის შემდგომ დამუშავებასა და გაანალიზებას შევძლებთ.

[1] რაოდენობრივი სოციოლოგიური კვლევებისას, როგორც წესი, ამ მიზნით პროგრამა **SPSS**-ს იყენებენ.

ერთი სიტყვით, მანამ, სანამ მონაცემთა ანალიზზე გადავალთ, მიღებული ინფორმაცია ანალიზისათვის უნდა მოვამზადოთ.

მონაცემთა მომზადება ანალიზისათვის ემპირიული სოციოლოგიური გამოკვლევის დამოუკიდებელი ეტაპია, რომლის ამოცანებსაც პირველადი ინფორმაციის შინაარსობრივი დამუშავება, მონაცემთა კომპიუტერული ბაზის შექმნა და სტატისტიკური ანალიზისათვის მომზადებაა.

იგი შემდეგი მიმდევრობით ხორციელდება:

- კითხვარების/ანკეტების შემოწმება;
- მონაცემთა რედაქტირება;
- მონაცემთა კოდირება;
- მონაცემთა ბაზის შექმნა;
- მონაცემთა განმენდა;
- მონაცემთა სტატისტიკური კორექტირება;
- მონაცემთა ანალიზის სტრატეგიის შემუშავება.

❖ კითხვარების შემონმება

ამ ეტაპის ძირითადი ამოცანაა ინტერვიუს ჩატარების ხარისხის შეფასება ანუ იმის შემოწმება, თუ რამდენად აკმაყოფილებს კითხვარი კვლევის მოთხოვნებს.

თუ ინტერვიუს ჩატარების ხარისხი მოთხოვნებს ვერ აკმაყოფილებს, კითხვარი არ მიიღება/უქმდება — კვლევიდან უნდა ამოვიღოთ.

კითხვარის კვლევიდან ამოღება იმ შემთხვევებში ხდება, როდესაც:

- **არ არის დაცული შერჩევის პირობები;**
- **კითხვარი არასრულია ან არასრულადაა შევსებული:** კითხვარს აკლია ერთი ან რამდენიმე გვერდი, გამოტოვებულია ცალკეული ბლოკები (კითხვათა ჯგუფები) ან იმ კითხვების მნიშვნელოვანი ნაწილი, რომელთა გამოტოვებასაც მოცემულ შემთხვევაში ინტერვიუს ჩატარების წესები არ ითვალისწინებს;
- **დარღვეულია ინტერვიუს ჩატარების წესები:** არ არის დაცული პასუხთა გამოტოვების პირობები — კითხვიდან კითხვაზე, ბლოკიდან ბლოკზე გადასვლისა და სხვა ინსტრუქციები;
- პასუხების შინაარსიდან ირკვევა, რომ **რესპონდენტმა ვერ გაიგო კითხვების მნიშვნელოვანი ნაწილის საზრისი;**
- **კითხვარი დაგვიანებულია:** მოტანილია დადგენილი ვადის შემდეგ.

კითხვარების გარკვეული რაოდენობის კვლევიდან ამოღების შემთხვევაში, განსაკუთრებით, როდესაც ასეთი კითხვარების რაოდენობა დიდია, შესაძლებელია საჭირო გახდეს ინფორმაციის ხელახალი მოძიება (ახალი ინტერვიუების ჩატარება). ამიტომ, მიზანშეწონილი იქნება, თუ კითხვარების შემონმებას საველე სამუშაოების პარალელურად განვახორციელებთ.

❖ მონაცემთა რედაქტირება

მონაცემთა რედაქტირების ამოცანაა კითხვარების დამუშავება იმ მიზნით, რომ მათში მოცემული ინფორმაცია **იოლი, ზუსტი და ერთმნიშვნელოვნად აღსაქმელი გახდეს.**

ამისათვის:

I. ყოველ ანკეტაში მონმდება:

- რამდენად **გასაგებადაა მონიშნული** ანკეტაში მითითებული პასუხების ვარიანტები;
- **იკითხება** თუ არა რესპონდენტის მიერ დასახელებული პასუხების ჩანანერები;
- **დასრულებულია** თუ არა რესპონდენტის მიერ დასახელებული პასუხების ჩანანერები;
- **ერთმნიშვნელოვანია** თუ არა რესპონდენტის მიერ დასახელებული პასუხების ჩანანერები;
- რამდენად **ლოგიკურია** მიღებული პასუხები;
- შეესაბამება თუ არა ერთმანეთს ლოგიკურად სხვადასხვა კითხვაზე მიღებული პასუხები.

II. ხდება შესაბამისი შესწორებების შეტანა: პასუხების ვარიანტების გასაგებად მითითება/ჩანერა; ჩანერილი პასუხების დასრულება, დაზუსტება; შემოკლებული პასუხების/აბრევიატურების გაშლა; ლოგიკურ შეუსაბამობათა გაშიფვრა, შესაძლებლობის შემთხვევაში, გასწორება და ა. შ.

იმ შემთხვევაში, თუ ამ ეტაპზე გამოვლენილი კითხვარის ნაკლოვანებები მნიშვნელოვანია და მისი რედაქტირება არ ხერხდება, კითხვარი კვლევიდან უნდა ამოვიღოთ.

❖ კოდირება

კოდირება ყოველი პასუხისათვის/შინაარსობრივი მონაცემისათვის გარკვეული კოდის (შიფრის, სიმბოლოს) მინიჭების პროცედურას ეწოდება.

იმისათვის, რომ დაშიფვრამ ინფორმაციის შინაარსობრივი ინტერპრეტაცია არ გაართულოს, დაშიფვრის ანუ **კოდირების პრინციპები/წესები ერთმნიშვნელოვანი და რაც შეიძლება მარტივი უნდა იყოს.**

რაოდენობრივი სოციოლოგიური კვლევებისას, უმეტეს შემთხვევაში, ციფრული კოდები გამოიყენება, რადგან ციფრები ყველასათვის ნაცნობი სიმბოლოებია, რაც აიოლებს კოდების სახით მოცემული ინფორმაციის შინაარსობრივ "გაშიფვრას" (გაგებას, ინტერპრეტაცირას, შესაბამისობის დამყარებას კოდის რიცხვით და შინაარსობრივ მნიშვნელობებს შორის) და, ამასთან ერთად, მოხერხებულია მონაცემთა სტატისტიკური დამუშავებისათვის.

დავუშვათ, რომ კვლევის ერთ-ერთი ამოცანაა იმის განსაზღვრა, თუ რამდენად კმაყოფილები არიან სტუდენტები თბილისის სახელმწიფო უნივერსიტეტში მიმდინარე რეფორმებით.

ჩვენს მიერ გამოყოფილი ნიშან-თვისების მიხედვით, ორ ნებისმიერ სტუდენტს შორის შეიძლება არსებობდეს იგივეობის/განსხვავების მიმართება:

ორი სტუდენტი იგივეობის მიმართებაშია ერთმანეთთან, თუ უნივერსიტეტში მიმდინარე რეფორმებით მათი კმაყოფილება (შინაარსობრივი თვალსაზრისით), ერთნაირია, ხოლო განსხვავების მიმართებაშია, თუ განსხვავებულია — მაგალითად, ერთი უფრო კმაყოფილია, ვიდრე მეორე.

გამოკითხვამ ამ მიმართებების დადგენა უნდა უზრუნველყოს, ხოლო კოდირებამ, მიღებული შედეგების რაც შეიძლება მარტივი და გასაანალიზებლად მოხერხებული გამოსახვა.

ეს სხვადასხვაგვარად შეიძლება მოხდეს. მაგალითად, თითოეულ გამოკითხულ სტუდენტს შეიძლება შევუსაბამოთ ნატურალური რიცხვი ერთიდან ხუთის ფარგლებში შემდეგი წესით:

- ყოველ სტუდენტს, რომელიც საესებით კმაყოფილია, მივანიჭოთ კოდი 1.
- ყოველ სტუდენტს, რომელიც უფრო კმაყოფილია, ვიდრე უკმაყოფილო, მივანიჭოთ წინა კოდთან შედარებით ერთით მეტი რიცხვი, ე. ი. 2;
- ყოველ სტუდენტს, რომელიც კმაყოფილების საშუალო ხარისხით ხასიათდება,

მივანიჭოთ წინა კოდთან შედარებით ერთით მეტი რიცხვი, ე. ი. 3;

- ყოველ სტუდენტს, რომელიც უფრო უკმაყოფილოა, ვიდრე კმაყოფილი, მივანიჭოთ წინა კოდთან შედარებით ერთით მეტი რიცხვი, ე. ი. 4;
- ყოველ სრულიად უკმაყოფილო სტუდენტს მივანიჭოთ წინა კოდთან შედარებით ერთით მეტი რიცხვი, ე. ი. 5.

ჩვენს მიერ აღწერილი კოდირების პრინციპი შემდეგნაირად შეიძლება ჩამოყალიბდეს: კმაყოფილების ხარისხი პირობითად ხუთ გრადაციად იყოფა. კმაყოფილების მაქსიმალურ ხარისხს შეესაბამება კოდი 1, მინიმალურს — კოდი 5, ხოლო კმაყოფილების ხარისხის ერთი გრადაციით კლებას — კოდის რიცხვითი მნიშვნელობის ერთით მატება.

კოდების მინიჭების წესებმა უნდა უზრუნველყოს, რომ კოდირების დასრულების შემდეგ ერთ მონაცემს ერთი და მხოლოდ ერთი კოდი შეესაბამებოდეს ანუ:

- ყოველ მონაცემს უნდა შეესაბამებოდეს კოდი — არც ერთი მონაცემი არ უნდა დარჩეს კოდის გარეშე.
- ერთ მონაცემს მხოლოდ ერთი კოდი უნდა შეესაბამებოდეს.

დახურული კითხვების კოდირება ხდება კითხვარის შედგენისას. კითხვის ყოველი შესაძლო პასუხის ნომერი, ამავე დროს, მის კოდია.

ღია/ნახევრადღია კითხვების კოდირება კი მხოლოდ დასრულებული ინტერვიუს პირობებშია შესაძლებელი და მიზანშეწონილია, რომ მოხდეს საველე სამუშაოების დამთავრების შემდეგ, როდესაც რესპონდენტთა მიერ დასახელებული ყველა პასუხი გვექნება დაფიქსირებული.

ნახევრადღია/ნახევრადდახურული კითხვის შემთხვევაში, რესპონდენტთა მიერ დასახელებული პასუხების კოდირება უნდა გაგრძელდეს იმის გათვალისწინებით, თუ როგორაა კოდირებული კითხვარში შეტანილი შესაძლო პასუხები.

მაგალითად, თუ კითხვარში შეტანილი შესაძლო პასუხები კოდირებულია ნატურალური რიცხვების მზარდი მიმდევრობის პრინციპით და მათ შორის მაქსიმალური კოდია **n**, რესპონდენტთა მიერ დასახელებული პასუხების კოდირება უნდა გრძელდებოდეს **n+1**-დან.

მონაცემთა დამუშავებისა და ანალიზის გასაიოლებლად, მიზანშეწონილია, რომ იმ ტიპურ პასუხებს, რომლებიც არ შეიცავს **კონკრეტულ ინფორმაციას**, მაგალითად: "მიჭირს პასუხის გაცემა", "არ ვიცი", "არ მსურს პასუხის გაცემა" და ა. შ. (მათ ზოგჯერ ნეიტრალურ პასუხებს უწოდებენ) მიენიჭოს განსხვავებული და იოლად ამოსაცნობი სპეციფიკური კოდი. დავუშვათ, **99, 88, 77** და ა. შ.

იმ შემთხვევაში, როდესაც ღია/ნახევრადღია კითხვებისათვის კითხვარში შეტანილია და კოდირებულია ასეთი პასუხები, მხედველობიდან არ უნდა გამოგვრჩეს, რომ რესპონდენტთა მიერ დასახელებული პასუხების კოდირებისას ამ რიცხვების გამოყენება აღარ შეიძლება.

❖ მონაცემთა ბაზის შექმნა

კვლევის ამ ეტაპზე ხდება კოდებში ასახული ინფორმაციის ანუ კოდებად გარდაქმნილ მონაცემთა ბაზის შექმნა.

როგორც წესი, ეს ამოცანა იმ სტატისტიკური კომპიუტერული პროგრამის ფარგლებში ნყდება, რომელიც კვლევის ამოცანების ადეკვატურ სტატისტიკურ მეთოდთა გამოყენებას უზრუნველყოფს.

იმისათვის, რომ კომპიუტერული მონაცემთა ბაზით სარგებლობა შევძლოთ, შერჩეული პროგრამის მეშვეობით უნდა შევქმნათ კითხვარის სტრუქტურული ასლი: "ბადე"/უჯრების ერთობლიობა, რომელიც საშუალებას მოგცემს:

- თითოეული კითხვარის ყოველი მონაცემი შევიტანოთ "ბადის" რომელიმე უჯრაში;

- ერთი მონაცემი შევიტანოთ მხოლოდ ერთ უჯრაში;
- ყოველი "კომპიუტერული" მონაცემისათვის ერთმნიშვნელოვნად შეგვეძლოს მისი "ადეკვატის" მითითება კითხვარში ანუ შესაბამისი კითხვარის, კითხვისა და პასუხის მონახვა.

❖ მონაცემთა გაწმენდა

მონაცემთა კომპიუტერული ბაზის შექმნა ახალ პრობლემას — კომპიუტერში მონაცემთა შეტანისას დაშვებული შეცდომების ("შეტანის" შეცდომების) გასწორებას უკავშირდება. გარდა ამისა, არ არის გამორიცხული, რომ გარკვეული შეცდომები გამოგვრჩეს კითხვარების რედაქტირების დროს ან დავუშვათ შეცდომა კოდირების პროცესში. მონაცემთა საბოლოო "გასუფთავება" შეცდომებისაგან მონაცემთა გაწმენდის ეტაპზე ხდება.

მონაცემთა გაწმენდა გულისხმობს მონაცემთა ბაზის შემოწმება-კორექტირებასა და გამოტოვებული პასუხების დამუშავებას.

სტატისტიკურ პროგრამებს (მაგალითად, SPSS-ს) აქვს იმის შესაძლებლობა, რომ გამოვავლინოთ და გავასწოროთ ასეთი შეცდომები.

შეცდომების გამოვლენის მექანიზმი ყოველი სტატისტიკური პროგრამისათვის ინდივიდუალურია, მაგრამ გასწორების პრინციპი ერთი და იგივეა ნებისმიერი პროგრამით სარგებლობის შემთხვევაში:

- შეცდომის გამოვლენის შემდეგ უნდა მოინახის კითხვარი, რომელის კომპიუტერულ ვარიანტშიც შეცდომა დაფიქსირდა;
- თუ კითხვარის ვარიანტი სწორია და შეცდომა "შეტანის" შეცდომაა, მაშინ გასწორება კითხვარის მიხედვით უნდა მოხდეს;
- თუ საქმე გვაქვს კოდირების შეცდომასთან, უნდა მოხდეს ლოკალური მასშტაბის ხელახალი კოდირება/კოდის შეცვლა;
- ხოლო როდესაც შეცდომა კითხვარის ანუ "შევსების" შეცდომაა, გადაწყვეტილება იმის მიხედვით უნდა მივიღოთ, თუ რამდენად მნიშვნელოვანია ეს შეცდომა, რამდენად შესაძლებელია მისი კორექტირება და რა შედეგები შეიძლება მოჰყვეს ამ ეტაპზე მთლიანად კითხვარის ამოღებას კვლევიდან.

❖ მონაცემთა სტატისტიკური კორექტირება

მონაცემთა სტატისტიკურ კორექტირებაში შედის:

- ცვლადების აწონვა;
- ცვლადების გარდაქმნა;
- სკალების გარდაქმნა.

ცვლადების აწონვა (weighting).

ცვლადის აწონვა მონაცემთა სტატისტიკური კორექტირების მეთოდია. მისი მეშვეობით ყოველ პასუხს/ყოველ რესპონდენტს ენიჭება წონითი კოეფიციენტი, რომელიც გამოხატავს ამ პასუხის/რესპონდენტის მნიშვნელობის ხარისხს/"წონას" დანარჩენ პასუხებთან/რესპონდენტებთან მიმართებაში.

სტატისტიკის თვალსარისით, აწონვის ძირითადი დანიშნულებაა ის, რომ გავზარდოთ ან

შევამციროთ ცვლადის გარკვეული მნიშვნელობების რაოდენობა.

ამ მეთოდს ყველაზე ხშირად იმ მიზნით იყენებენ, რომ შერჩევითი მონაცემები მაქსიმალურად ასახავდეს გენერალური ერთობლიობის მახასიათებლებს ან მოხდეს შერჩევის კორექტირება: გაიზარდოს/შემცირდეს რესპონდენტთა გარკვეული ჯგუფის პასუხების მნიშვნელობა.

ანონვის პროცედურა, როგორც ვხედავთ, "ხელოვნურად" ცვლის (ზრდის ან ამცირებს) რესპონდენტთა ცალკეული ჯგუფის მნიშვნელობას. ამიტომ, მისი გამოყენებისას დიდი სიფრთხილეა საჭირო, რათა ამან რეალური სურათის დამახინჯება არ გამოიწვიოს. ყოველთვის, როდესაც ანონვის მეთოდს გამოვიყენებთ, დეტალურად უნდა წარმოვადგინოთ არგუმენტაცია იმის შესახებ, თუ რის საფუძველზე მივიღეთ ასეთი გადაწყვეტილება და მივუთითოთ ის კონკრეტული პროპორციები, რომლის მიხედვითაც განხორციელდა ცვლადის ანონვა.

ცვლადების გარდაქმნა (variable respecification).

ცვლადის გარდაქმნა ეწოდება მონაცემთა გარდაქმნას/შეცვლას **ახალი ცვლადის** შემოტანის მიზნით **ან მოცემული ცვლადის "გადაკეთებას"** კვლევის ამოცანების შესაბამისად.

ცვლადების გარდაქმნის ძირითადი მიზანია მონაცემთა ბაზის ისეთი სახით წარმოდგენა, რომელიც მაქსიმალურად შეესაბამება კვლევის ამოცანებს და ოპტიმალური იქნება სტატისტიკური ანალიზისათვის.

მაგალითად, საქართველოში ჩატარებული ერთ-ერთი გამოკვლევის დროს, იმ რესპონდენტთა რაოდენობა, რომელთაც საერთოდ არ უვლიათ სკოლაში, განისაზღვრა 0.1%-ით, ხოლო იმ რესპონდენტთა რაოდენობა, რომელთაც მხოლოდ ოთხწლიანი საშუალო განათლება აქვთ მიღებული, 0.5%-ით. ასეთი მცირე სიხშირეები სოციოლოგიური კვლევისათვის არააქტუალურია, ამიტომ მოხდა მიღებული ცვლადის გარდაქმნა: ეს მნიშვნელობები გაუგიდა მომდევნო მნიშვნელობას (არასრული საშუალო განათლება), რომელიც მნიშვნელოვანი სიხშირით იყო წარმოდგენილი.

თუ ჩვენს ამოცანას შეადგენს ცვლადის საშუალო მნიშვნელობის გამოთვლა, მაშინ ის პასუხები, რომელშიც კონკრეტული ინფორმაცია არ არის მოცემული, მაგალითად, "მიჭირს პასუხის გაცემა" სტატისტიკური ანალიზიდან უნდა ამოირითოს და ა. შ.

ცვლადის გარდაქმნის სპეციფიკურ სახეა ეგრეთ წოდებული **ფიქტიური ცვლადების (dummy variable)** შემოტანა.

ფიქტიური ცვლადი მიიღება მოცემული ცვლადის დიხოტომიურ ცვლადად გარდაქმნის შედეგად, მისი ერთ–ერთი მნიშვნელობის მიხედვით.

მაგალითად, ეროვნების ცვლადი მისი ერთ-ერთი მნიშვნელობის მიხედვით, დავუშვათ, ქართველი, შეიძლება გარდაიქმნას დიხოტომიურ ცვლადად, რომელსაც მხოლოდ ორი მნიშვნელობა ექნება: ქართველი-არაქართველი.

სკალების გარდაქმნა (scale respecification).

სკალის გარდაქმნა ეწოდება სკალის მნიშვნელობების შეცვლას ("მანიპულირებას" სკალის მნიშვნელობებით), იმ მიზნით, რომ იგი სხვა სკალებს შევადაროთ ან მოცემული სკალით მიღებული მონაცემები ანალიზისათვის ოპტიმალური სახით წარმოვადგინოთ.

მაგალითად:

შვიდდანაყოფიანი სკალის გარკვეულ მნიშვნელობათა გაერთიანება საშუალებას მოგვცემს, რომ იგი ხუთდანაყოფიან სკალას შევადაროთ.

ანალიზისათვის შეიძლება საჭირო გახდეს, რომ სკალის მნიშვნელობები სიმეტრიულად "შევატრიალოთ".

რაოდენობრივი სკალის მნიშვნელობათა გაერთიანება (დაეუშვათ, ასაკის მნიშვნელობების გარკვეული წესით დაჯგუფება) საშუალებას მოგვცემს, რომ იგი რიგის სკალად (ასაკობრივი ჯგუფების სკალად) "გადავაქციოთ" და ა. შ.

❖ მონაცემთა ანალიზის სტრატეგიის შემუშავება

მონაცემთა ანალიზის კონკრეტული მეთოდოლოგიის/სტრატეგიის შემუშავება ხდება შემდეგი ფაქტორების გათვალისწინებით:

- რა ტიპის სკალებით ჩატარდა გაზომვა;
- რა მეთოდებით მოხდა სკალირება;
- რა ტიპის ცვლადების საფუძველზე შეიქმნა მონაცემთა ბაზა;
- სტატისტიკის რომელი მეთოდებითაა შესაძლებელი მიღებული მონაცემების გაანალიზება;
- შესაძლო მეთოდებიდან, რომელია ოპტიმალური კვლევის ამოცანების განსახორციელებლად.

ჩვენ ზოგადად განვიხილეთ ემპირიული სოციოლოგიური კვლევის სტრუქტურა. ეს საკითხი რთული და მრავალმხრივია. მისი დაწვრილებითი ანალიზი ამ სალექციო კურსის მიზანი არაა.

განხილული მასალის შინაარსისა და გადმოცემის წესის შერჩევა შემდეგი მოსაზრებების გათვალისწინებით მოხდა:

- ყურადღება ძირითადად გამახვილებულია იმ ასპექტებზე, რომლებიც მომდევნო საკითხების შესწავლისას განსაკუთრებით მნიშვნელოვანია;
- სოციოლოგიის ენა მრავალფეროვანია. უნიფიცირებული ცნებითი აპარატი არ არსებობს:

სოციოლოგები ზოგჯერ სპეციფიკურად განსაზღვრავენ ამა თუ იმ ტერმინს, ხოლო რამდენიმე ტერმინს ერთი და იგივე/სინონიმური მნიშვნელობით იყენებენ.

ამიტომ, სიმარტივისათვის შერჩეულია ემპირიული სოციოლოგიის რაც შეიძლება "ნეიტრალური" ტერმინოლოგია — გავრცელებული, ტიპური ცნებები.

❖ მათემატიკის მეთოდების როლი ემპირიული სოციოლოგიური კვლევის პროცესში

რატომ მათემატიკა?

მონაცემთა ანალიზის მეთოდების განხილვისას ხშირად "მათემატიკის ენაზე" მოგვიწევს საუბარი. რატომ?

როგორც უკვე ვნახეთ, მონაცემთა მოპოვების დაწყებამდე, სოციოლოგი გამოყოფს შესასწავლი მოვლენის იმ ასპექტებს, რომლებიც კვლევისათვის არსებითია, დანარჩენები კი უკუიგდება, მხედველობაში არ მიიღება.

გარკვეული ასპექტების უკუგდების/აბსტრაქციის შედეგად მიიღება საკვლევი მოვლენის ***ფორმალიზებული მოდელი.***

მათემატიკური აპარატი ("მათემატიკის ენა") ფორმალიზებული მოდელების შესწავლის ოპტიმალური საშუალებაა.

კოდირებისა და გაზომვის საფუძველზე ფორმალიზებული მოდელი აისახება მათემატიკურ მოდელში, რომლის გაანალიზებაც მათემატიკის მეთოდებით ხდება.

რა საჭიროა სოციალურ მოვლენათა არითმეტიზაცია — "თარგმნა მათემატიკის" ენაზე და შემდეგ მათი შინაარსობრივი "უკუთარგმანი"? ხომ არ ართულებს ეს კვლევის პროცესს?

ასეთ შეკითხვას განსაკუთრებით ხშირად იძლევიან ჰუმანიტარულ მეცნიერებათა ის წარმომადგენლები (მათ შორის სოციოლოგები), რომელთაც სტერეოტიპული ბარიერი აქვთ მათემატიკის მიმართ.

ამ ბარიერის დაძლევა ბევრად უფრო იოლი და პროდუქტიულია, ვიდრე ინფორმაციის დამუშავება მათემატიკის მეთოდების გამოყენების გარეშე, რადგან:

- მათემატიკის მეთოდები უნივერსალურია. მათი გამოყენება უზრუნველყოფს იმას, რომ შედეგები დამოუკიდებელია მკვლევრის პიროვნული თავისებურებებისაგან, მისი პოზიციისაგან და ასევე უნივერსალურია/მისაღებია ყველასათვის;
- მათემატიკის მეთოდების გამოყენებისას დაზღვეული ვართ უზუსტობისაგან, რაც მკვლევრმა შინაარსობრივი ინფორმაციის დამუშავებისას შეიძლება დაუშვას;
- მათემატიკის მეთოდების გამოყენების შედეგად მიიღება რეზულტატები, რომლებიც მკვლევრმა შინაარსობრივად შეიძლება ვერ გამოიყვანოს;
- სოციოლოგიური კვლევების დროს ხშირად საქმე გვაქვს მონაცემთა დიდ მასივებთან. მათი შინაარსობრივი დამუშავება ან საერთოდ შეუძლებელია ან იმდენად დიდ დროს მოითხოვს, რომ საკვლევმა პრობლემამ შეიძლება აქტუალობა დაკარგოს;
- მათემატიკის მეთოდების გარეშე ვერც შერჩევით ერთობლიობას გამოვყოფთ და ვერც კვლევის შედეგებს განვაზოგადებთ გენერალური ერთობლიობის მასშტაბით;
- და ბოლოს, რამდენადაც სოციოლოგიურ მონაცემთა ანალიზის მთავარი ამოცანა სოციალური კანონზომიერებების გამოვლენაა, სოციოლოგისათვის განსაკუთრებით მნიშვნელოვანია მათემატიკური სტატისტიკა და მის საფუძვლადმდებარე ალბათობის თეორია, რადგან კანონზომიერებების დადგენა სწორედ მათემატიკური სტატისტიკის მეშვეობით ხდება.

ამიტომ, მათემატიკური მეთოდების გამოყენება მხოლოდ სოციოლოგის არჩევანი როდია — მისი განწყობის მიუხედავად, ემპირიული სოციოლოგია გვერდს ვერ აუვლის მათემატიკის სფეროს.

გაზომვა

❖ გაზომვისა და ცვლადის ცნებები სოციოლოგიაში

ცნება "გაზომვა" სოციოლოგიაში ემპირიული მონაცემების (სოციოლოგიური ინფორმაციის, პარამეტრების მნიშვნელობების) მოპოვებას უკავშირდება.

იგი სხვადასხვა კონტექსტში შეიძლება შეგვხვდეს. მაგალითად:

- ერთი რომელიმე **პარამეტრის გაზომვა** ნიშნავს მისი კონკრეტული/ემპირიული მნიშვნელობების დადგენას/პოვნას;
- ზოგადად, **საკვლევი პარამეტრების მნიშვნელობების დადგენა** ასევე გაზომვაა. ამ შემთხვევაში იგი აღნიშნავს სოციოლოგიური კვლევის ეტაპს — ემპირიული მონაცემების მოპოვების პროცესს;
- **გაზომვის მეთოდებში** იგულისხმება მონაცემების მოპოვების (პარამეტრების მნიშვნელობების დადგენის) წესები და საშუალებები.

ეს ჩამონათვალი სოციოლოგიაში "გაზომვის" ყველა გაგებას არ ამოწურავს. მაგრამ ცხადყოფს, რომ თუ ჩვენთვის ნათელია, რა იგულისხმება პარამეტრის მნიშვნელობების პოვნაში, "გაზომვის" გაგებას ნებისმიერ კონტექსტში შევძლებთ:

- თუ გავიგებთ, რას ნიშნავს ერთი პარამეტრის მნიშვნელობების პოვნა, იოლად გავიაზრებთ რა იგულისხმება ზოგადად, საკვლევი პარამეტრების მნიშვნელობების დადგენაში;
- თუ გავიგებთ, რას ნიშნავს პარამეტრის მნიშვნელობების პოვნა, ძნელი არ იქნება გაზომვის, როგორც მონაცემთა მოპოვების წესებისა და საშუალებების გაგება და ა. შ.

ახლა შევეცადოთ ვუპასუხოთ კითხვას — რას ნიშნავს პარამეტრის გაზომვა ანუ მისი კონკრეტული მნიშვნელობების პოვნა?

თავდაპირველად განვიხილოთ, რა ტიპის პარამეტრებთან შეიძლება ჰქონდეს საქმე სოციოლოგს.

ამას იოლად მოვახერხებთ, თუ რაოდენობრივი სოციოლოგიური კვლევის ინსტრუმენტს, კითხვარს გადავავლებთ თვალს.

კითხვარში გარკვეული კითხვებია დასმული. დავუშვათ:

"გთხოვთ, გვიპასუხოთ, რა ეროვნების ბრძანდებით?"

"გთხოვთ, გვიპასუხოთ, რა პროფესიის ბრძანდებით?"

"გთხოვთ, გვიპასუხოთ, რამდენად კმაყოფილი ხართ უნივერსიტეტში მიმდინარე რეფორმებით?"

მსგავსი კითხვების მეშვეობით სოციოლოგი ადგენს თითოეული რესპონდენტის ეროვნებას, პროფესიას, მიმდინარე რეფორმებით კმაყოფილებას და ა. შ.

ცნება "გაზომვა" ხელსაწყოსთან (მაგალითად, სახაზავი, სასწორი, სპიდომეტრი) გაზომვის შედეგი კი, რიცხვებში გამოხატულ გარკვეულ ზომასთან (მაგალითად, 85 სმ., 5 კგ., 250კმ/სთ) არის ასოცირებული.

ამიტომ ადამიანს შეიძლება გაუჩნდეს კითხვა:

კი მაგრამ როგორ უნდა ანარმოოს სოციოლოგმა გაზომვა, **რას ნიშნავს, მაგალითად, სქესის ან პროფესიის ზომა?**

პასუხი თავდაპირველად პარადოქსული შეიძლება მოგვეჩვენოს:

სოციოლოგიაში გაზომვები, ჩვეულებრივ, **ხელსაწყოთი** ხდება. შედეგადაც სწორედ **ზომებს** ვიღებთ.

ხელსაწყოს, რომელსაც სოციოლოგი გაზომვისას იყენებს, **სკალა** ჰქვია, გაზომვის შედეგად მიღებულ სიდიდეს (ზომას) კი — **პარამეტრის მნიშვნელობა**.

რადგან სკალები სოციოლოგიაში სპეციფიკური მოვლენების გაზომვის საშუალებაა, ამიტომ საკითხში გარკვევა სწორედ ამ მოვლენების აღწერითა და ანალიზით დავიწყოთ.

მივმართოთ ანალოგიას.

როდესაც ადამიანის დახასიათება გვსურს, მის გარკვეულ ნიშან-თვისებებს მივუთითებთ. მაგალითად:

გიორგი არის 30 წლის მამაკაცი, ფილოლოგი, მუშაობს სკოლის მასწავლებლად, გატაცებულია პოეზიით.

მაკა არის 22 წლის გოგონა, მუსიკოსი, მუშაობს ფილარმონიაში მომღერლად, გატაცებულია იაპონური კულტურით.

რა აქვთ საერთო გიორგისა და მაკას ამ დახასიათების მიხედვით?

ერთი შეხედვით, არაფერი. ისინი ყველაფრით ერთმანეთისაგან განსხვავდებიან. მაგრამ, მოდით, დავაზუსტოთ რას ვგულისხმობთ "ყველაფერში":

იმას, რომ გიორგი 30 წლისაა, მაკა კი — 22-ის, გიორგი მამრობითი სქესისაა, მაკა მდედრობითის და ა. შ.

მაგრამ ამ განსხვავებების მითითება შეუძლებელი იქნებოდა, გიორგისა და მაკას საერთო რომ არაფერი ჰქონდეთ.

ეს საერთოა ის, რომ ორივეს აქვს (შეიძლება ჰქონდეს) **ასაკი, სქესი, პროფესია, საქმიანობა, გატაცება**.

ესე იგი, ამა თუ იმ ადამიანის დახასიათებისას, ვახდენთ მის აღწერას ზოგადი ნიშან-თვისებების მიხედვით, რომლებიც განსაზღვრულ მნიშვნელობებს ამ ადამიანის მიხედვით იღებენ.

თვალსაჩინოებისათვის ჩვენი მონაცემები ცხრილის სახით წარმოვადგინოთ:

ცხრილი №1.

ნიშანთვისება	გიორგი	მაკა
ასაკი →	30 წ.	22 წ.
სქესი →	მამრობითი	მდედრობითი
პროფესია →	ფილოლოგი	მუსიკოსი
საქმიანობა →	სკოლის მასწავლებელი	ფილარმონიის მომღერალი
გატაცება →	პოეზია	იაპონური კულტურა

ამრიგად:

- გიორგისა და მაკას საერთო აქვთ ის, რომ თითოეული მათგანის დახასიათება შესაძლებელია ცხრილში მოცემული ნიშან-თვისებებით;
- ისინი განსხვავდებიან ამ ნიშან-თვისებების **კონკრეტული მნიშვნელობების მიხედვით.**

ის, რასაც სოციოლოგი "ზომავს", სწორედ ასეთი, ზოგადი ხასიათის ნიშან-თვისებებია.

ნიშან-თვისება (დავუშვათ, სქესი) **განსაზღვრულ მნიშვნელობას კონკრეტული შემთხვევის მიხედვით იღებს:**

მაგალითად, გია 18 წლისაა, თეონა — 22-ის ანი — 30-ის და ა. შ.

ეს პროცესი შემდეგნაირადაც შეიძლება აღვწეროთ: ამა თუ იმ ნიშან-თვისების მნიშვნელობები კონკრეტული შემთხვევების მიხედვით **იცვლება.**

ასეთი აღწერის შემდეგ აღარ გაგვიჭირდება იმის გაგება, რომ ნიშან-თვისება, ფაქტობრივად, **ცვლადია, გაზომვა — ცვლადის მნიშვნელობების პოვნა, გაზომვის შედეგად კი ემპირიულ მონაცემთა მასივს (ცვლადის მნიშვნელობების ერთობლიობას) ვიღებთ.**

გაზომვის საფუძველიცა და ძირითადი მიზანიც შედარებაა: ნებისმიერი სიდიდის გაზომვას გარკვეულ ეტალონთან შედარებით ვაწარმოებთ იმ მიზნით, რომ ამ სიდიდის კონკრეტული მნიშვნელობების ერთმანეთთან შედარება შევძლოთ.

მაგალითად, წონის დასადგენად, მოცემული ობიექტი წონის ეტალონს უნდა შევადაროთ. შედეგად, შეგვიძლია იმის განსაზღვრა, რომ, დავუშვათ, 5 კგ. წონის ობიექტი უფრო მსუბუქია, ვიდრე — 7 კგ.

❖ ცვლადების კლასიფიკაცია: ნომინალური, რიგის და რაოდენობრივი ცვლადები

მნიშვნელობათა შედარების სპეციფიკის მიხედვით, ცვლადების სამი ძირითადი ჯგუფი გამოიყოფა:

I. ცვლადები, რომელთა მნიშვნელობებს შორის **მხოლოდ** იგივეობის/განსხვავების მიმართება შეიძლება არსებობდეს.

სქესი, პროფესია, რელიგია, ეროვნება, საქმიანობა და ა. შ. მაგალითად, ჩვენ შეგვიძლია შევადაროთ ერთმანეთს ადამიანები პროფესიის მიხედვით:

ორი სოციოლოგი ერთმანეთისაგან პროფესიით არ განსხვავდება, მათემატიკოსი კი განსხვავდება სოციოლოგისაგან.

თუმცა, ვერ ვიტყვით, რომ მათემატიკოსობა მეტია ან ნაკლებია სოციოლოგობაზე, ვერ ვიმსჯელებთ მათ შორის სხვაობაზე ან მათ შეფარდებაზე.

ასეთ ცვლადებს **სახელდებითი (ნომინალური) ცვლადები** შეიძლება ვუწოდოთ, რადგან მათი მნიშვნელობების შედარება სახელწოდებების, ცნებების მიხედვით ხდება;

II. ცვლადები, რომელთა მნიშვნელობებს შორის შეიძლება დამყარდეს არა მხოლოდ იგივეობის/განსხვავების, არამედ, მეტობის/ნაკლებობის მიმართებაც, მაგრამ ვერ განვსაზღვრავთ, თუ რამდენით ან რამდენჯერ მეტია/ნაკლებია ერთი მეორეზე.

მაგალითად, ჩვენ შეგვიძლია ვთქვათ, რომ ლადო უფრო კმაყოფილია, ვიდრე გიგა, მაგრამ ზუსტად ვერ ვიტყვით, თუ რამდენით ან რამდენჯერ.

ასეთ ცვლადებს რიგის ცვლადები შეიძლება ვუწოდოთ, რადგან მათი მნიშვნელობების შედარება მეტობის/ნაკლებობის რიგის მიხედვით ხდება;

III. ცვლადები, რომელთა მნიშვნელობები საშუალებას გვაძლევს განვსაზღვროთ, თუ რამდენით ან რამდენჯერ მეტია/ნაკლებია ერთი მნიშვნელობა მეორეზე.

ასეთ ცვლადებს **რაოდენობრივი** ცვლადები შეიძლება ვუწოდოთ, რადგან მათი მნიშვნელობები **რაოდენობრიად ურთიერთშედარებადია**.

რაოდენობრივი ცვლადები, თავის მხრივ, **ინტერვალურ** და **შეფარდების** ცვლადებად იყოფა.

ინტერვალურია ცვლადი, თუ შესაძლებელია განვსაზღვროთ სხვაობის **ინტერვალი** მის ორ ნებისმიერ მნიშვნელობას შორის (რამდენით მეტია/ნაკლებია ერთი მეორესთან შედარებით). თუმცა, აბსოლუტური ნულის უქონლობის გამო, მნიშვნელობების შეფარდების განხილვა აზრს მოკლებულია.

ინტერვალური ცვლადების მაგალითია ასაკი. ჩვენ შეგვიძლია ვიმსჯელოთ სხვაობის ინტერვალზე ასაკის ორ მნიშვნელობას შორის (რამდენი წლით უმცროსია/უფროსია ერთი ადამიანი მეორეზე), მაგრამ, **მათი შეფარდების განხილვას (რამდენჯერ უმცროსია/უფროსია ერთი მეორეზე) აზრი არა აქვს,** რადგან **ასაკის განსაზღვრა აბსოლუტური ნულის მიმართ კი არ ხდება, არამედ მიმდინარე კალენდარული წლის მიხედვით, რომელიც იცვლება.**

თუ, მაგალითად, ერთი ადამიანი 18 წლისაა, მეორე კი 36-ის, წელს მეორე პირველზე ორჯერ უფროსია, მაგრამ მომავალ წელს, როდესაც პირველი 19 წლის იქნება, ხოლო მეორე 37-ის, "წლევანდელი" თანაფარდობა დაირღვევა: 37/19≠2.

რაოდენობრივ ცვლადს ეწოდება **შეფარდების ცვლადი**, თუ მისი ნებისმიერი ორი მნიშვნელობისათვის აზრი აქვს არა მხოლოდ სხვაობის, არამედ შეფარდების (რამდენჯერ მეტია/ნაკლებია ერთი მეორეზე) განხილვასაც.

შეფარდების ცვლადებს აბსოლუტურ ცვლადებსაც უწოდებენ.

ალტერნატიული სახელწოდება — "აბსოლუტური" იმ ფაქტს ასახავს, რომ შეფარდების ცვლადის კონკრეტული მნიშვნელობები აბსოლუტური ნულის მიმართ განისაზღვრება.

შეფარდების (აბსოლუტური) ცვლადის მაგალითია ხელფასი.

თუ, დავუშვათ, ერთი ადამიანის ხელფასი თვეში 100 ლ-ია, ხოლო მეორესი — 200 ლ. შეგვიძლია ვიმსჯელოთ მათ ხელფასებს შორის როგორც სხვაობაზე (100 ლ.), ისე შეფარდებაზე — პირველს მეორესთან შედარებით ორჯერ ნაკლები ხელფასი აქვს.

ამ შემთხვევაში გვაქვს აბსოლუტური ნული — ხელფასები განსაზღვრულია ნულთან მიმართებაში.

ახლა შეგვიძლია კიდევ უფრო ნათლად დავინახოთ რაოდენობრივი ცვლადების ინტერვალურ და შეფარდების ცვლადებად დაყოფის საფუძველი:

შეფარდების ცვლადისაგან განსხვავებით, ინტერვალური ცვლადის მნიშვნელობებს შორის **პროპორციული მიმართებების,** მათი **პროპორციული გარდაქმნების** შესახებ მსჯელობა, როგორც წესი, ინტერპრეტაციას არ ექვემდებარება.

მაგალითად, თუ ჩვენ ვიცით, რომ ერთ ადამიანს აქვს X ლარი მეორეს კი 2Xლარი და რომელიღაც ვალუტაში გადაყვანის შემდეგ პირველს ექნება ფულის 2X ერთეული, შეგვიძლია დავასკვნათ, რომ მაშინ მეორეს 2×2X ერთეული ექნება.

მაგრამ, თუ წელს ერთი ადამიანის ასაკია Xწელი, ხოლო მეორის 2X წელი, როდესაც პირველის ასაკი 2X წელი გახდება, მეორის ასაკი 2×2X წელი არ იქნება — ასაკის ორი განსხვავებული მნიშვნელობის ერთდროულად ორჯერ გაზრდა შეუძლებელია.

თუმცა, შეგვიძლია ვიმსჯელოთ იმის შესახებ, რომ, დავუშვათ, 10 წლის შემდეგ თითოეული მათგანის ასაკი ერთი და იგივე ინტერვალით: 10 წლით მოიმატებს.

ასაკის მნიშვნელობებს შორის პროპორციების ცვალებადობას ცვალებადი ათვლის წერტილი განაპირობებს, ხოლო სხვაობის მუდმივობას — ფიქსირებული რაოდენობრივი ერთეული/ინტერვალი (1 წელი).

❖ დიხოტომიური ცვლადები

ცვლადების სპეციფიკური სახეა დიხოტომიური ცვლადები. **დიხოტომიური ცვლადის** მნიშვნელობებს **ორი, ურთიერთგამომრიცხავი ალტერნატივა წარმოადგენს.** დიხოტომიური ცვლადის მაგალითებია: ფრიადოსნობა (მნიშვნელობები: არის ფრიადოსანი, არ არის ფრიადოსანი), სრულწლოვანება (მნიშვნელობები: არის სრულწლოვანი, არ არის სრულწლოვანი), ქართველობა/ქართველად ყოფნა (მნიშვნელობები: არის ქართველი, არ არის ქართველი), ბიოლოგიური სქესი (მნიშვნელობები: მდედრობითი, მამრობითი) და ა. შ.

დიხოტომიური ცვლადი ნებისმიერი ცვლადისაგან შეიძლება მივიღოთ მისი რომელიმე შესაძლო მნიშვნელობის ან მნიშვნელობათა კომბინაციის საფუძველზე.

მაგალითად, "ფრიადოსნობა" მიიღება რიგის ცვლადის (აკადემიური მოსწრების) ერთ-ერთი მნიშვნელობის (ფრიადი) საფუძველზე.

"ქართველად ყოფნა" მიიღება ნომინალური ცვლადის (ეროვნების) ერთ-ერთი მნიშვნელობის (ქართველი) საფუძველზე და ა. შ.

შესაბამისად, თუ ეს ანალიზის პროცესში დაგვჭირდა, ნებისმიერი ცვლადი დიხოტომიურზე შეიძლება "დავიყვანოთ".

❖ სკალების კლასიფიკაცია: ნომინალური, რიგის და რაოდენობრივი სკალები

რადგან სოციოლოგის ამოცანა ცვლადების გაზომვაა, "საზომი ხელსაწყოები" სწორედ ცვლადების სპეციფიკას უნდა ასახავდეს.

ეს ასეცაა. ჩვენ უკვე აღვნიშნეთ, რომ ცვლადების გასაზომი "ხელსაწყო" სოციოლგისათვის სკალაა.

ცვლადების კლასიფიკაციის შესაბამისად, სკალებიც სამ ძირითად ჯგუფად შეიძლება დავყოთ: ნომინალურ, რიგის და რაოდენობრივ (ამ უკანასკნელს მეტრულსაც უწოდებენ) სკალებად.

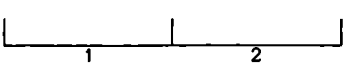
ნომინალური სკალა.

ნომინალური სკალა არის "სახაზავი", რომლის დანაყოფები (ერთეულები) გასაზომი ცვლადის შესაძლო მნიშვნელობებს შეესაბამება (იმ კონკრეტულ მნიშვნელობებს, რომლებიც ცვლადმა შეიძლება მიიღოს).

ასეთი სკალით გაზომვის შედეგად მიღებული მონაცემთა ერთობლიობა იმდენ ჯგუფად განაწილდება/დაიყოფა, რამდენი დანაყოფიც აქვს სკალას:

მაგალითად:

სქესი ორდანაყოფიანი სკალით იზომება. ერთი დანაყოფი პირობითად შეიძლება მდედრობით სქესს შევუსაბამოთ, მეორე — მამრობითს:



შესაბამისად, მონაცემთა ერთობლიობა ორ ჯგუფად — "მდედრობითი" და "მამრობითი" —

განანილდება.

ნომინალურ სკალაზე არც ათელის წერტილი განისაზღვრება, არც ცვლადის მნიშვნელობათა მიმდევრობა და არც საზომი ერთეულის/ინტერვალის სიდიდე.

იმის მიუხედავად, რომ, სკალის აგების, შესაბამისად, გაზომვის მოთხოვნები[1] მინიმალურია[2], ნომინალური სკალის აგება სულაც არ არის ტრივიალური:

გაზომვის შედეგად, მონაცემთა ერთობლიობა ჯგუფებად ისე უნდა განაწილდეს, რომ ყოველი მონაცემი ერთ და მხოლოდ ერთ ჯგუფში მოხვდეს ანუ:

- **ყოველი მონაცემი აუცილებლად უნდა შევიდეს რომელიმე ჯგუფში;**
- **არ შეიძლება, რომ ერთი და იგივე მონაცემი შევიდეს ერთზე მეტ ჯგუფში.**

ეს მოთხოვნა პრინციპულია. ნინააღმდეგ შემთხვევაში, ინფორმაციის ადეკვატურ დამუშავებას, მითუმეტეს, გაანალიზებას ვერ შევძლებთ:

- არ გვექნება სრული ინფორმაცია მონაცემთა ერთობლიობის განაწილების შესახებ;
- იარსებებს მონაცემთა განანილების ალტერნატიული ვარიანტები.

ბიოლოგიური სქესის შემთხვევაში ნომინალური სკალის აგების პირობების დაცვა იოლია.

მაგრამ პრობლემის სერიოზულობას მაშინვე მივხვდებით, თუ უფრო რთულ ცვლადს, ვთქვათ, საქმიანობას შევარჩევთ გასაზომად.

ხშირად, სკალაზე, რომლითაც საქმიანობას ზომავენ, დიფერენცირებულია "უმუშევარი" და "დროებით უმუშევარი".

თუ ამ ორი ცნების მკაფიო, ერთმნიშვნელოვანი განსაზღვრებები წინასწარ არ გვექნება შემუშავებული, "უმუშევრებისა" და "დროებით უმუშევრების" ჯგუფების დიფერენცირება შეუძლებელი გახდება.

მსგავს სიტუაციაში აღმოვჩნდებით, თუ, დავუშვათ, სკალაზე ერთდროულად გვექნება "ხელმძღვანელი მუშაკი" და "დირექტორი".

დირექტორი ამავე დროს ხელმძღვანელი მუშაკიცაა. მაგრამ ხელმძღვანელ მუშაკად სხვა თანამდებობის პირი, მაგალითად, დირექტორის მოადგილეც ითვლება.

შედეგად:

- დირექტორი შეიძლება მოხვდეს როგორც უშუალოდ დირექტორის, ისე ხელმძღვანელი მუშაკის შესაბამის ჯგუფში;
- შეუძლებელი გახდება იმის დიფერენცირება, ხელმძღვანელ მუშაკებს შორის რამდენია დირექტორი და რამდენი არ არის დირექტორი;
- ვერ დავადგენთ სულ რამდენი დირექტორია.

ნათელია, რომ ამგვარად მიღებული მონაცემთა ჯგუფები არათუ ანალიზს, არამედ აღწერასაც კი არ დაექვემდებარება.

ამრიგად, ნომინალური სკალის აგებისას მთავარი დატვირთვა იმ სახელწოდებების/ცნებების **მკაფიო და ზუსტ ვერბალურ ინტერპრეტაციაზე** მოდის, რომლებიც ნომინალური ცვლადის შესაძლო მნიშვნელობებია და რომლებსაც სკალის დანაყოფები შეესაბამება. სკალებთან მიმართებაში მათ უკვე **სკალის მნიშვნელობები** შეიძლება ვუწოდოთ.

კი მაგრამ, ძირითადი ცნებების ოპერაციონალიზაცია ხომ გაზომვამდე ხდება, რის გამო შეიძლება აღმოვჩნდეთ იგივე პრობლემის წინაშე?

საქმე ისაა, რომ ხშირ შემთხვევაში შეუძლებელია კვლევის დასანყისშივე განვსაზღვროთ ცვლადის ყველა შესაძლო მნიშვნელობა. ღია ან ნახევრადღია კითხვებზე მიღებული პასუხები სოციოლოგისათვის ის ინფორმაციაა, რომლის ზუსტი გათვალისწინება გაზომვამდე შეუძლებელია. ამიტომ ცნებების ინტერპრეტაციის ამოცანა ხელახლა შეიძლება დადგეს.

ეს ამოცანა კვლევის იმ ეტაპზე წყდება, რომელსაც "მონაცემთა ლოგიკური კონტროლი და კოდირება" ვუწოდეთ.

---

[1] გაზომვის პირობები.

[2] რის გამოც, ნომინალური სკალით გაზომვას გაზომვის ყველაზე დაბალ დონედ მიიჩნევენ.

კოდირება ცნებებში მოცემული ინფორმაციის "დაშიფვრის", მისთვის გარკვეული სიმბოლოს, კოდის მინიჭების პროცედურას ეწოდება.

რადგან ემპირიული სოციოლოგიური ინფორმაციის დამუშავებისას მათემატიკის მეთოდებს ვიყენებთ, უფრო მოხერხებულია კოდებად რიცხვების გამოყენება.

კვლევის ეს საფეხური განსაკუთრებული მნიშვნელობის მქონეა. სწორედ **კოდირებაა "ადაპტორი"/გადამყვანი, რომლის მეშვეობითაც კონცეპტუალური (ცნებებით აგებული) მოდელი მათემატიკურ მოდელში აისახება.**

კოდირება კვლევის ინსტრუმენტის შედგენისას იწყება.

კითხვების ნომრები ორმაგ დანიშნულებას ასრულებს: ნუმერაციის გარდა, თითოეული ნომერი შესატყვისი კითხვის კოდია.

გაზომვის შემდეგ ხდება ამ პროცესის დასრულება ანუ ცვლადების იმ მნიშვნელობებისათვის კოდების მინიჭება, რომლებიც კვლევის ინსტრუმენტში კი არ იყო შეტანილი, არამედ გაზომვის შემდეგ მივიღეთ.

ამ თვალსაზრისით, კოდირება გაზომვის შემადგენელ ნაწილადაც შეიძლება მივიჩნიოთ.

კოდირებისა და ინფორმაციის მოპოვების პროცედურების მიმართება კიდევ ერთხელ წარმოაჩენს სოციოლოგიური კვლევის ეტაპების ურთიერთგაპირობებულობას:

ერთი მხრივ, კოდირების ძირითადი პრინციპების განსაზღვრა და ცვლადების მნიშვნელობების დიდი ნაწილის "დაშიფვრა" გაზომვამდე ხდება, მეორე მხრივ, კოდირების პროცესი დამთავრებულად მხოლოდ გაზომვის შემდეგ შეიძლება ჩაითვალოს.

აქედან გამომდინარე, ერთი მხრივ, გაზომვამდე შერჩეული კოდირების პრინციპები განსაზღვრავს გაზომვის სპეციფიკას, მეორე მხრივ, გაზომვის შედეგად მიღებული ინფორმაცია ხშირად ახალი პრინციპების შემუშავების, ახალი დეფინიციების (განსაზღვრებების) შემოტანის აუცილებლობის წინაშე გვაყენებს.

რიგის სკალა

ინტერვიუერები ხშირად იძლევიან კითხვებს, რომლებიც შემდეგი ტიპის ფრაზებით იწყება:

— "რამდენად კმაყოფილი ხართ"...

— "რამდენად მოგწონთ"...

— "რამდენად ეთანხმებით"... და ა. შ.

ასეთ კითხვებზე მიღებული პასუხები უფრო რთულ შედარებას ექვემდებარება, ვიდრე ეს იგივეობის/განსხვავების მიმართებაა.

მაგალითად:

"რამდენად კმაყოფილი ხართ თქვენი ქორწინებით?"

ცხრილი №2

შესაძლო პასუხი/მნიშვნელობა	I	II	III
საკმაოდ კმაყოფილი	1	5	14
უფრო კმაყოფილი, ვიდრე უკმაყოფილო	2	4	11
კმაყოფილიც და უკმაყოფილოც (თანაბრად)	3	3	9
უფრო უკმაყოფილო, ვიდრე კმაყოფილი	4	2	8
სრულიად უკმაყოფილო	5	1	5

მიღებული პასუხების მიხედვით, შესაძლებელია **არა მხოლოდ განვასხვავოთ** მონაცემთა ხუთი **ჯგუფი**, არამედ **დავალაგოთ** კიდეც ისინი კმაყოფილების ზრდის ან კლების მიხედვით.

იმის მიუხედავად, რომ ამის შესაძლებლობას, როგორც აღვნიშნეთ, თავად კითხვა იძლევა, **მხოლოდ** კითხვის ადეკვატური ფორმულირება **ჯგუფების დალაგების საკმარისი პირობა არ არის.**

ერთ–ერთ არსებით როლს ამ ამოცანის გადაწყვეტისას კოდირება ასრულებს.

ჩვენს მაგალითში კითხვის შესაძლო პასუხების გასწვრივ მოცემულია პასუხების კოდირების ("დაშიფვრის") სამი ვარიანტი:

პირველი ვარიანტი შეიძლება გავიაზროთ როგორც **ადგილების მინიჭება**: "პირველი ადგილი ენიჭება" სავსებით კმაყოფილს, "მეორე" — უფრო კმაყოფილს, ვიდრე უკმაყოფილოს და ა. შ.

ამ წესით მიღებული მონაცემთა ჯგუფები შეიძლება დალაგდეს:

ა) კმაყოფილების ზრდის მიხედვით:

V ჯგ (Vადგილი) < IV ჯგ (IV ადგილი) < III ჯგ (III ადგილი) < II ჯგ (II ადგილი) < I ჯგ (I ადგილი).

ბ) კმაყოფილების კლების მიხედვით:

I ჯგ (I ადგილი) > II ჯგ (II ადგილი) > III ჯგ (III ადგილი) > IV ჯგ (IV ადგილი) > V ჯგ (V ადგილი).

მეორე ვარიანტი შეიძლება გავიაზროთ, როგორც **ქულების მინიჭება**: "მაქსიმალურ ქულას (5) ვანიჭებთ" სავსებით კმაყოფილს, ერთი საფეხურით დაბალს (4) — უფრო უკმაყოფილოს, ვიდრე კმაყოფილს და ა. შ.

ამ წესით მიღებული მონაცემთა ჯგუფებიც დალაგდებადია კმაყოფილების ზრდის ან კლების მიხედვით:

V ჯგ (1 ქულა) < IV ჯგ (2 ქულა) < III ჯგ (3 ქულა) < II ჯგ (4 ქულა) < I ჯგ (5 ქულა)

I ჯგ (5 ქულა) > II ჯგ (4 ქულა) > III ჯგ (3 ქულა) > IV ჯგ (2 ქულა) > V ჯგ (1 ქულა)

მესამე ვარიანტი შემდეგნაირად შეიძლება აღვწეროთ: **კმაყოფილების უფრო მაღალ ხარისხს უფრო მეტი რიცხვი შეესაბამება.** ამ წესით მიღებული მონაცემთა ჯგუფები, კმაყოფილების ხარისხის ზრდის ან კლების მიხედვით შემდეგნაირად დალაგება:

V ჯგ (5, კოდებს შორის მინიმალური რიცხვი) < IV ჯგ (8, 8>5) < III ჯგ (9, 9>8) < II ჯგ (11, 11>9) < I ჯგ (14, 14>9)

I ჯგ (14, კოდებს შორის მაქსიმალური რიცხვი) > II ჯგ (11, 11<14) > III ჯგ (9, 9<11) > IV ჯგ (8, 8<9) > V ჯგ (5, 5<8).

ამრიგად, იმის მიუხედავად, რომ კოდებს შორის სხვაობა არაერთგვაროვანია (მაგალითად, 14–11 ≠ 11-9), კოდირების ეს ვარიანტიც "მუშაობს" — იძლევა მონაცემთა ჯგუფების დალაგების საშუალებას, რადგან **კმაყოფილების რაოდენობრივი საზომი ერთეული** არ არსებობს და ინტერვალების ერთგვაროვნება არსებითი მოთხოვნა არ არის.

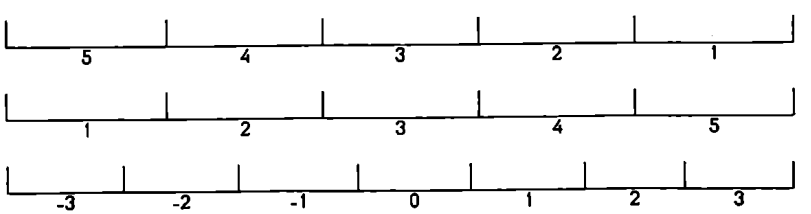
თუმცა, პირველი ორი ვარიანტი ბევრად უფრო მოხერხებულია — ჩვეულ მოვლენებთან ანალოგიის შესაძლებლობა მნიშვნელოვნად აიოლებს მონაცემების აღწერის, ინტერპრეტირებისა და ანალიზის პროცესს.

ჩვენს მიერ განხილული თითოეული ვარიანტის გამოყენების შედეგად მივიღებთ ხუთსაფეხურიან რიგის სკალას.

რიგის სკალა შეიძლება განვიხილოთ როგორც n დანაყოფიანი "სახაზავი", რომლის ყოველი მომდევნო საფეხური მეტია/ნაკლებია წინაზე.

დანაყოფების რიცხვი იმდენია, რამდენი მნიშვნელობაც შეიძლება მიიღოს გასაზომმა ცვლადმა.

რიგის სკალაზე (ისევე, როგორც ნომინალურზე) **არც ათვლის წერტილი განისაზღვრება და არც საზომი ერთეულის სიდიდე, მაგრამ** (ნომინალურისაგან განსხვავებით) **ფიქსირებულია ცვლადის მნიშვნელობათა რიგი/მიმდევრობა:**



რიგის სკალით გაზომვა საშუალებას გვაძლევს ცვლადის მნიშვნელობებს შორის დავადგინოთ არა მხოლოდ იგივეობის/განსხვავების, არამედ მეტობის/ნაკლებობის მიმართებებიც.

ამიტომ, რიგის სკალით გაზომვა საშუალო — ნომინალურთან შედარებით მაღალ, მაგრამ რაოდენობრივთან შედარებით დაბალ დონედ ითვლება.

საყურადღებოა, რომ **ხუთზე მეტი გრადაციის შემთხვევაში, ინტერვალების მკაფიო ვერბალური ინტერპრეტაცია პრობლემურია.**

მაგალითად, შეუძლებელია, ვერბალურად დავაზუსტოთ კმაყოფილების ხარისხი სავსებით კმაყოფილსა და უფრო კმაყოფილს, ვიდრე უკმაყოფილოს შორის.

სკალის დანაყოფების დიფერენცირება, თავისთავად, სულაც არ არის **დიფერენცირებული ინფორმაციის** მიღების გარანტია.

ინტერპრეტაციის პრობლემურობის გამო, რესპონდენტებს შეიძლება ადეკვატური შეფასება გაუჭირდეთ, რაც "შეზობელი" დანაყოფების შესატყვისი ჯგუფების აღწერასა და ანალიზს გაართულებს.

ზოგ შემთხვევაში, სოციოლოგი იძულებულია გააერთიანოს ასეთი ჯგუფები და სკალა უკვე გაზომვის შემდეგ შეცვალოს.

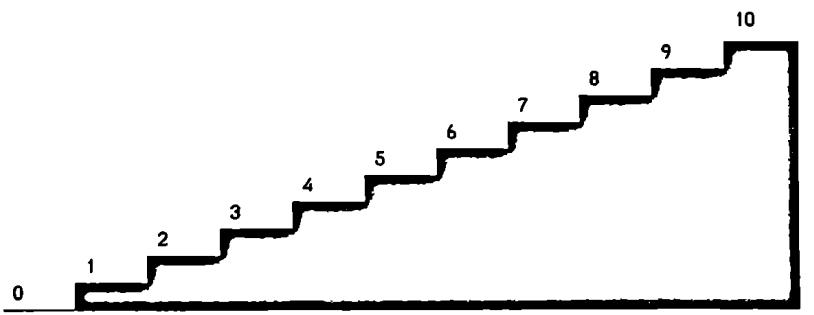
გაზომვის შემდეგ სკალების შეცვლამ კი ანალიზის მეთოდოლოგიისა და მთლიანად კვლევის სტრატეგიის შეცვლა შეიძლება გამოიწვიოს.

ამიტომ ხუთზე მეტი დანაყოფის შერჩევისას, აუცილებელია იმ პირობების მკაფიო განსაზღვრა, რამაც შეფასების სიზუსტე უნდა უზრუნველყოს.

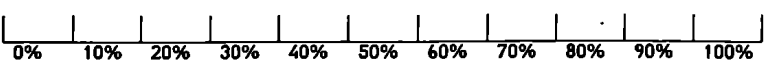
ამ თვალსაზრისით, ეფექტური შეიძლება აღმოჩნდეს:

ვიზუალური, **გრაფიკული ინტერპრეტაციების** გამოყენება. მაგალითად, სკალის წარმოდგენა კიბის სახით:

"გთხოვთ, გვიპასუხოთ, ბედნიერების რომელ საფეხურზე იმყოფებით თქვენი ქორწინების შედეგად?"



ანალოგია პროცენტებთან — **პროცენტული ინტერპრეტაცია.** მაგალითად, "გთხოვთ, გვიპასუხოთ, რამდენი პროცენტით გაკმაყოფილებთ ქალაქის მმართველობის საქმიანობა?"



და სხვა.

რაოდენობრივი სკალები

რაოდენობრივი სკალები, როგორც უკვე აღვნიშნეთ, რაოდენობრივი ცვლადების სპეციფიკას ასახავს. რაოდენობრივი ცვლადების თავისებურება კი ისაა, რომ მათი მნიშვნელობები

გარკვეული რიცხვითი სიდიდეებია[1].

ამიტომ რაოდენობრივი სკალების აგებისას კოდირება ნაკლებ პრობლემურია — ცვლადების რიცხვითი მნიშვნელობები პირდაპირ შეიძლება კოდებად გამოვიყენოთ.

რაოდენობრივი სკალების განსაზღვრისას არსებით როლს ასრულებს ათვლის წერტილისა და საზომი ერთეულის/ინტერვალის ცნებები.

რაოდენობრივ სკალას, რომელზეც განსაზღვრულია საზომი ერთეულის/ინტერვალის სიდიდე, მაგრამ არ არის განსაზღვრული აბსოლუტური ნული **ინტერვალური** ეწოდება.

რაოდენობრივ სკალას, რომელზეც განსაზღვრულია საზომი ერთეულის სიდიდეცა და აბსოლუტური ნულიც, **შეფარდების ან აბსოლუტური სკალა ეწოდება**.

რაოდენობრივი სკალით გაზომვა საშუალებას გვაძლევს ცვლადის მნიშვნელობებს შორის დავადგინოთ არა მხოლოდ იგივეობის/განსხვავებისა და ზრდის /კლების მიმართებები, არამედ ისიც, თუ:

ა) რამდენით მეტია/ნაკლებია ერთი მნიშვნელობა მეორეზე (ინტერვალური სკალა);

ბ) რამდენით და რამდენჯერ მეტია/ნაკლებია ერთი მნიშვნელობა მეორეზე (შეფარდების სკალა).

ამიტომ, რაოდენობრივი სკალით გაზომვა გაზომვის ყველაზე მაღალ დონედ ითვლება.

ინტერვალური სკალის მაგალითად შეიძლება მოვიტანოთ ნლებით გამოხატული ასაკის სკალა. ამ სკალაზე საზომი ერთეულია 1 წელი. აბსოლუტური ნული კი არ გვაქვს, რადგან ასაკი ცვალებად კალენდრულ წელთან მიმართებით განისაზღვრება.

შეფარდებითი/აბსოლუტური სკალის მაგალითია რომელიმე ვალუტაში, დავუშვათ, ლარებში გამოხატული შემოსავლების სკალა. საზომ ერთეულს ამ შემთხვევაში 1 ლარია, აბსოლუტური ნული კი — 0.

❖ სკალების გამოყენებისა და ანალიზის სპეციფიკა

ამრიგად, სოციოლოგიაში სკალები ნიშან-თვისებების/ცვლადების გასაზომად და მათი სპეციფიკის ადეკვატურად შეიქმნა.

სწორედ ამიტომ, სკალების გამოყენების შესაძლებლობები ცვლადების თავისებურებებს შეესაბამება.

ჩვენს მიერ განხილულ კლასიფიკაციაში გაზომვის ყოველი უფრო მაღალი დონის შესაბამისი ცვლადის მახასიათებლების ჩამონათვალი, როგორც ნაწილს, მოიცავს წინა დონეების შესაბამისი ცვლადების მახასიათებლებსაც:

ცხრილი №3

ცვლადი	ცვლადის მნიშვნელობების შედარების პრინციპი	გაზომვის დონე
ნომინალური	იგივეობა/განსხვავება	დაბალი
რიგის	იგივეობა/განსხვავება & მეტობა/ნაკლებობა	საშუალო
რაოდენობრივი (ინტერვალური)	იგივეობა/განსხვავება & მეტობა/ნაკლებობა & მნიშვნელობათა სხვაობა	მაღალი
რაოდენობრივი (შეფარდებითი)	იგივეობა/განსხვავება & მეტობა/ნაკლებობა & მნიშვნელობათა სხვაობა & მნიშვნელობათა შეფარდება	მაღალი

ეს კი იმას ნიშნავს, რომ

- **ცვლადი, რომელსაც გაზომვის უფრო მაღალი დონე შეესაბამება, შეიძლება გაზომვის უფრო დაბალი დონის სკალებითაც გაიზომოს.**

[1] ნინააღმდეგ შემთხვევაში, მათი რაოდენობრივი შედარება შეუძლებელი იქნებოდა.

განვიხილოთ მაგალითი:

I. "როგორ შეაფასებთ თბილისის სახელმწიფო უნივერსიტეტში გამოცდის ტესტურ ფორმაზე გადასვლას?"

1. დადებითად
2. უფრო დადებითად, ვიდრე უარყოფითად
3. საშუალოდ
4. უფრო უარყოფითად, ვიდრე დადებითად
5. უარყოფითად

II. "როგორ შეაფასებთ თბილისის სახელმწიფო უნივერსიტეტში გამოცდის ტესტურ ფორმაზე გადასვლას?"

1. დადებითად, გამოცდის ტესტური ფორმა პროგრესულია
2. დადებითად, მაგრამ თსუ-ში მისი შემოღება ნაადრევად მიმაჩნია
3. დადებითად, მაგრამ ვთვლი, რომ საკმარისი არ არის, საჭიროა გამოცდის სხვა ფორმებიც არსებობდეს
4. გამოცდის ტესტურ ფორმას აქვს როგორც დადებითი, ისე უარყოფითი მხარეები
5. უარყოფითად, გამოცდის ტესტური ფორმა სტუდენტის ცოდნას სათანადოდ ვერ ავლენს
6. უარყოფითად, გამოცდის ტესტური ფორმა არ იძლევა სტუდენტთა პირადი უნარების დიფერენცირების საშუალებას
7. უარყოფითად, გამოცდის ტესტური ფორმის შემოღება უკან გადადგმული ნაბიჯია

ჩვენს მიერ განხილული ორი შემთხვევიდან, პირველში რიგის ცვლადი შესაბამისი (რიგის) სკალით იზომება. იმის მიხედვით, თუ როგორ აფასებენ თსუ-ში გამოცდის ტესტურ ფორმაზე გადასვლას, მონაცემთა ხუთი ჯგუფი გამოიყოფა და ეს ჯგუფები შეიძლება შეფასების გრადაციების ზრდის ან კლების მიხედვით დავალაგოთ.

მეორეში იგივე ცვლადის გაზომვა ნომინალური სკალით ხდება: პასუხის შესაძლო ვარიანტები (სკალის ერთეულები, დანაყოფები) მხოლოდ გარკვეული პოზიციების ვერბალურ აღწერებს/ინტერპრეტაციებს ასახავს და არა შეფასების გრადაციებს.

შედეგად გამოიყოფა/ერთმანეთისაგან განსხვავდება მონაცემთა 7 ჯგუფი, მაგრამ ამ ჯგუფებს შეფასების გრადაციების მიხედვით ვერ დავალაგებთ.

მაგალითად, შეუძლებელია, განვსაზღვროთ, რომელი ჯგუფი უფრო დადებითად აფასებს თსუ-ში გამოცდის ტესტურ ფორმაზე გადასვლას: მეორე თუ მესამე ან რომელი უფრო უარყოფითად — მეხუთე თუ მეექვსე.

რიგის ცვლადის გაზომვა ნომინალური სკალით იმიტომაა შესაძლებელი, რომ, როგორც უკვე აღვნიშნეთ, რიგის ცვლადს აქვს ნომინალურის "შესაძლებლობებიც" ანუ **შეიცავს იმ ინფორმაციის მიღების საშუალებასაც, რომელიც ნომინალური (უფრო დაბალი დონის) სკალით იზომება.**

ანალოგიურად, შესაძლებელია რაოდენობრივი ცვლადები რიგის ან ნომინალური სკალებით გავზომოთ.

ხშირად ასაკს (ინტერვალური ცვლადი) რიგის სკალით — ასაკობრივი ჯგუფებით ზომავენ (მაგალითად: 18-21 წწ. 22-25 წწ. 26-29 წწ. და ა. შ.).

ინტენსიურად ხდება რიგის სკალით — გარკვეულ ვალუტაში გამოხატული ფულადი ინტერვალებით შემოსავლების (შეფარდების ცვლადი) გაზომვაც (მაგალითად: 100 ლარამდე, 101-200 ლარი, 201-300 ლარი და ა. შ.).

როგორც ვხედავთ, ცვლადის ტიპი სკალის ტიპის შერჩევას ერთმნიშვნელოვნად არ განსაზღვრავს — **ერთი და იგივე ცვლადი სხვადასხვა ტიპის სკალით შეიძლება გაიზომოს.**

რის მიხედვით ირჩევს სოციოლოგი სკალას?

ჩვენ უკვე ვისაუბრეთ იმის შესახებ, რომ გაზომვის მეთოდები ჯერ კიდევ გაზომვამდე შერჩეულ ანალიზის მეთოდებს უნდა შეესაბამებოდეს. სკალების შერჩევის სპეციფიკა ნათლად გვიჩვენებს, რატომაა კვლევის ასე დაგეგმვა ოპტიმალური:

თუ კითხვა იმისთვისაა გამიზნული, რომ **შეფასებითი დამოკიდებულების გრადაციები განვსაზღვროთ და მონაცემთა შესაბამისი ჯგუფების შედარებითი ანალიზი განვახორციელოთ**, რიგის სკალა უნდა ავირჩიოთ. ნომინალური სკალა შეფასების გრადაციებს საერთოდ ვერ გაზომავს.

ხოლო თუ კითხვას იმ მიზნით ვაძლევთ, რომ ის ძირითადი მოტივები გავაანალიზოთ, რაც დადებით/უარყოფით შეფასებებს განაპირობებს, მაშინ ნომინალური სკალა უკეთესია:

ჯერ ერთი, გრადაციების გამოყოფა ამ შემთხვევაში არ არის საჭირო, მეორე, გრადაციები მოტივაციებს არ ასახავს.

გასაგებია, რომ თუ სკალას ანალიზის ამოცანებისა და მეთოდების გაუთვალისწინებლად შევარჩევთ, მიღებული ინფორმაცია შეიძლება არაადეკვატური აღმოჩნდეს.

- **ცვლადი, რომელსაც გაზომვის უფრო დაბალი დონე შეესაბამება, გაზომვის უფრო მაღალი დონის სკალებით ვერ გაიზომება — ასეთი გაზომვა აზრს მოკლებულია.**

სხვაგვარად რომ ვთქვათ, უფრო მაღალი დონის "ხელსაწყო" მასთან შედარებით დაბალი დონის "ხელსაწყოს" დანიშნულებას ვერ შეასრულებს. რატომ? იმიტომ, რომ სოციოლოგიაში რიცხვები "უცნაურად იქცევიან":

ნომინალური, რიგის და რაოდენობრივი სკალები ვიზუალურად შეიძლება არც კი განსხვავდებოდეს ერთმანეთისაგან. მაგალითად, სკალა:

ა) ნომინალურია, თუ მასზე აღნიშნული რიცხვები გარკვეულ სახელწოდებებს, დავუშვათ, პროფესიებს შეესაბამება:

1 — ფილოსოფოსი
2 — მათემატიკოსი
3 — ფიზიკოსი
4 — სოციოლოგი
5 — ფილოლოგი

ბ) რიგის სკალაა, თუ მასზე აღნიშნული რიცხვები რიგის ცვლადის მნიშვნელობათა გრადაციებს გამოხატავს:

1 — სავსებით კმაყოფილი
2 — უფრო კმაყოფილი, ვიდრე უკმაყოფილო
3 — კმაყოფილიც და უკმაყოფილოც (თანაბრად)
4 — უფრო უკმაყოფილო, ვიდრე კმაყოფილი
5 — სრულიად უკმაყოფილო

გ) რაოდენობრივია, თუ მასზე აღნიშნული რიცხვები, დავუშვათ, საათობრივი ანაზღაურების ტარიფებს შეესაბამება: 1 ლარი საათში, 2 ლარი საათში და ა. შ. რადგან შესაძლებელია არა მხოლოდ მონაცემთა დალაგება საათობრივი ანაზღაურების ზრდის/კლების მიხედვით, არამედ მათ შორის რაოდენობრივი მიმართებების დადგენაც.

რა განსხვავებაა ამ სამ შემთხვევას შორის?

პირველში კოდებად აღებული რიცხვები საერთოდ "დაცლილია" არითმეტიკული შინაარსისაგან. მეტობა/ნაკლებობის მიმართებით მათი ერთმანეთთან შედარება ან მათზე

არითმეტიკული მოქმედებების ჩატარება არ შეიძლება. **ამას ცვლადის და არა რიცხვების სპეციფიკა განაპირობებს.**

ცხადია, თვითონ რიცხვები შეიძლება დავალაგოთ, მაგალითად, ზრდის მიხედვით: 1 < 2 < 3 < 4 < 5, მაგრამ როგორც კი კოდებს ნომინალური ცვლადის მნიშვნელობებს შევუსაბამებთ, უაზრო რეზულტატებამდე მივალთ:

ფილოსოფოსი < მათემატიკოსი < ფიზიკოსი < სოციოლოგი < ფილოლოგი.

ანალოგიურად, კოდების, როგორც რიცხვების სხვაობასა და შეფარდებაზეც შეიძლება საუბარი, მაგრამ იმის ინტერპრეტაცია, თუ რას ნიშნავს 5 – 4 (ფილოლოგს გამოკლებული სოციოლოგი) ან 1 : 2 (ფილოლოგი შეფარდებული მათემატიკოსთან) უკვე შეუძლებელია.

მეორე შემთხვევაში რიცხვები რიგის ცვლადის მნიშვნელობათა გრადაციებს გამოხატავს. ამ ცვლადის მიმართ აზრი აქვს მნიშვნელობების მეტობა/ნაკლებობაზე ვილაპარაკოთ, მაგრამ მათი სხვაობისა (1 – 2 ანუ კმაყოფილებას გამოკლებული უფრო კმაყოფილება, ვიდრე უკმაყოფილება) და შეფარდების (4 : 5 ანუ უფრო უკმაყოფილება, ვიდრე კმაყოფილება შეფარდებული უკმაყოფილებასთან) განხილვა ლოგიკას აღარ ექვემდებარება.

მესამე შემთხვევაში კოდები უშუალოდ ცვლადის რიცხვით მნიშვნელობებს გამოხატავს. მათ მიმართ არითმეტიკული მოქმედებების განხორციელება და მათ შორის რაოდენობრივი მიმართებების დამყარება იმიტომაა დასაშვები, რომ **თვითონ ცვლადის მნიშვნელობები ექვემდებარება** რაოდენობრივ გარდაქმნებსა და რაოდენობრივ ურთიერთშედარებას.

ახლა შეგვიძლია გავშიფროთ, რატომ "იქცევიან რიცხვები უცნაურად" სოციოლოგიაში:

რიცხვითი კოდები ჩვეულებრივ არითმეტიკის ერთეულებს კი არ წარმოადგენენ, არამედ, ცვლადის მნიშვნელობების სიმბოლურ გამოსახულებებს და **მათში მხოლოდ ის შინაარსი შეიძლება ჩაიდოს, რასაც მოცემული ცვლადი "გვკარნახობს"**[1].

სწორედ ამიტომაა უსაზრისო მაღალი დონის სკალის გამოყენება უფრო დაბალი დონის ცვლადის მიმართ.

- **ჩვენს მიერ განხილული სკალები თეორიული, "იდეალიზებული" ტიპებია. პრაქტიკაში მათი თვისებები და გამოყენება შეიძლება განსხვავებული იყოს.**

იმ ფაქტორებს შორის, რომლებიც სკალის თეორიული ტიპიდან "გადახრას" განაპირობებს, განსაკუთრებით აღსანიშნავია ეგრეთ ნოდებული "სოციოლოგიური ნული" და ინტერპრეტაციის შედეგად შეძენილი/დაკარგული თვისებები.

სოციოლოგიურ ნულს სკალის ისეთ მნიშვნელობას უწოდებენ, რომელიც ცვლადის რომელიმე მნიშვნელობას კი არ შეესაბამება, პირიქით, **იმას გამოხატავს, რომ ცვლადისათვის მნიშვნელობის მინიჭება ვერ ხერხდება.** მაგალითად: "მიჭირს პასუხის გაცემა", "არ ვიცი", "არ მსურს პასუხის გაცემა" და ა. შ.

როდესაც რიგის სკალაზე ასეთი მნიშვნელობა შეგვაქვს, ცხადია, მონაცემთა ყველა ჯგუფის დალაგებას ვერ შევძლებთ. "მიჭირს პასუხის გაცემა" მეტობა/ნაკლებობის მიმართებით შედარებას არ ექვემდებარება. ასეთ შემთხვევაში ამბობენ, რომ ადგილი აქვს **ნაწილობრივ დალაგებას.**

იგივე პრობლემა რაოდენობრივი სკალების მიმართაც არსებობს: დავუშვათ, ოჯახის საერთო შემოსავლებს ვზომავთ და გამოიყო ჯგუფი, რომელიც "არ ვიცის" შეესაბამება. ცხადია, "არ ვიცის" ვერაფერს გამოვაკლებთ და ვერც რაიმესთან შევაფარდებთ.

რიგ შემთხვევაში შეიძლება ასეთი მნიშვნელობა სკალაზე საერთოდ არ შევიტანოთ ან შესაბამისი ჯგუფი ანალიზის დროს გამოვყოთ და ჯგუფების ორი კატეგორია (სოციოლოგიური ნულის შესატყვისი და ცვლადის მნიშვნელობების შესატყვისი) ინდივიდუალურად გავაანალიზოთ.

- **ინტერპრეტაციის შედეგად სკალამ შეიძლება "შეიცვალოს" თვისებები.**

[1] სოციოლოგიაში რომ განმსაზღვრელ როლს კოდებად შერჩეული რიცხვების არითმეტიკული თვისებები ასრულებდეს და არა თვითონ ცვლადები, კოდებად სხვა სიმბოლოების, მაგალითად ანბანის ასოების გამოყენება შეუძლებელი იქნებოდა.

ა) სკალამ პრაქტიკულად შეიძლება შეიძინოს თვისება, რომელიც მის თეორიულ ეტალონს/ტიპს არ გააჩნია.

მაგალითად, საუნივერსიტეტო 5-ქულიანი შეფასების სისტემის ინტერპრეტაცია იმდენად "ჩვეული" და გასაგებია, რომ მას **პრაქტიკულად როგორც რაოდენობრივ/მეტრულ სკალას ისე იყენებენ:**

ნებისმიერი სტუდენტი, რომელმაც, დავუშვათ, მიიღო შეფასება "3" გეტყვით, რამდენით ნაკლებია მისი ნიშანი "5"-ზე. რეიტინგების დათვლისას ხდება აკადემიური მოსწრების საშუალო მაჩვენებლებს შორის სხვაობის გამოთვლა და ა. შ.

თეორიულად კი იგი რიგის სკალაა: ცოდნისათვის არც რაოდენობრივი საზომი ერთეულია განსაზღვრული და არც აბსოლუტური ნული.

ბ) სკალამ პრაქტიკულად შეიძლება დაკარგოს თვისება, რომელიც მის თეორიულ ეტალონს/ტიპს გააჩნია.

მაგალითად, შემოსავლების სკალა, როგორც აღვნიშნეთ რაოდენობრივია (შეფარდებითი).

მაგრამ როდესაც სოციოლოგი ამ სკალით მიღებულ მონაცემებს აანალიზებს, სკალის ერთეულმა შეიძლება უნივერსალურობა დაკარგოს. დავუშვათ, საკითხის შემდეგი ასპექტით განხილვისას:

იმ ადამიანისათვის, რომელსაც შემოსავალი 100 ლ. აქვს, შემოსავლის 50 ლ-ით შემცირება ბევრად უფრო მნიშვნელოვანია, ვიდრე 3 000 ლ. შემოსავლის მქონე ადამიანისათვის. თუმცა, ორივე შემთხვევაში შემოსავალი ერთი და იგივე რაოდენობით (50 ლ.) მოიკლებს.

ამ თვალსაზრისით, შეიძლება ითქვას, რომ მოცემულ შემთხვევაში შემოსავლების სკალა "ფსევდორაოდენობრივ" ხასიათს ატარებს. თუმცა, ინტერპრეტაციის გარეშე, ფსევდორაოდენობრიობაზე, ცხადია, ვერ ვიმსჯელებთ.

ფსევდორაოდენობრივი ხასიათი ასაკობრივმა სკალამაც შეიძლება "გამოავლინოს". 17 და 18 წელს შორის იგივე რაოდენობრივი განსხვავებაა, რაც 21-22 წელს შორის, მაგრამ როდესაც 18 წელი უსრულდება, ადამიანი სრულწლოვანი ხდება და არსებითად იცვლება მისი იურიდიული სტატუსი. ამგვარი ინტერპრეტაციისას, ეს ერთი წელი თავისი **მნიშვნელობით ბევრად უფრო "მეტია"**, ვიდრე 21-22 წწ. **რაოდენობრივად მისი ტოლი** ინტერვალი.

იგივე შეიძლება ითქვას 34-35 წლებს შორის ინტერვალზე. 35 წლის შემდეგ ადამიანი აღარ ითვლება ახალგაზრდად, რის გამოც იგი გარკვეულ უფლებებს კარგავს (მაგალითად, აღარ შეუძლია ახალგაზრდული სამეცნიერო გრანტის მიღება) და გარკვეულს იძენს (მაგალითად, შეუძლია მონაწილეობა მიიღოს საპრეზიდენტო არჩევნებში, პრეზიდენტობის კანდიდატად). ამიტომ ეს ინტერვალი შინაარსობრივად ბევრად უფრო მნიშვნელოვანი შეიძლება აღმოჩნდეს, ვიდრე, დავუშვათ 36-37 წწ. ინტერვალი.

მსგავსი ინტერპრეტაციები, ცხადია, საინტერესოა, მაგრამ მათი **ხელოვნური ძებნა სულაც არ არის საჭირო.** სოციოლოგმა მათ მხოლოდ იმ შემთხვევაში მიმართავს, როდესაც ეს კვლევისათვის აუცილებელია/მნიშვნელოვანია.

ჩვენს მიერ განხილული მაგალითები კიდევ ერთხელ ცხადყოფს სოციოლოგიაში ინტერპრეტაციების არსებით მნიშვნელობას. **ინტერპრეტაციის შედეგად ერთი და იგივე რაოდენობრივი მაჩვენებლები** (დავუშვათ, ინტერვალები) **შეიძლება თვისებრივად ერთმანეთისაგან განსხვავებული აღმოჩნდეს.**

ამიტომ ანალიზის დროს სკალებისა და მონაცემების თეორიული ან ტექნიკური შესაძლებლობებით "ბრმად" სარგებლობა მიზანშეწონილი არ არის. მათი ტიპი/ხასიათი იმის მიხედვით უნდა დაზუსტდეს, თუ რა ასპექტით ვაპირებთ ანალიზის წარმართვას.

❖ გაზომვის/სკალირების მეთოდები

სკალის ტიპთან ერთად მონაცემთა ანალიზისათვის არსებითი მნიშვნელობა აქვს იმასაც, თუ რა **მეთოდით** მოხდა სკალირება.

სკალირების მეთოდები ორ ძირითად ჯგუფად შეიძლება დაიყოს: **შედარებით** და **არაშედარებით/დამოუკიდებლად გაზომვის** მეთოდებად.

სკალებს, რომლებსაც შედარებითი მეთოდებით სარგებლობის შემთხვევაში იყენებენ, შედარების სკალებს (**comparative scales**) უწოდებენ.

შედარებითი სკალირებით მიღებული მონაცემები **ფარდობითი** ხასიათისაა. მათი გამოხატვა შესაძლებელია **რიგის/რანგული** სიდიდეებით.

რიგის/რანგული სიდიდეები ობიექტს სხვა ობიექტთან/ობიექტებთან შედარებით ახასიათებს, მაგრამ ასეთი სიდიდეების მეშვეობით შეუძლებელია მხოლოდ ერთი ობიექტის დამოუკიდებელი დახასიათება. ამიტომ შედარებით სკალირებას (შესაბამისად, შედარების სკალებს) **არამეტრულსაც** უწოდებენ.

შედარებითი სკალირება შეიძლება განხორციელდეს ობიექტების წყვილური (წყვილ-წყვილად) შედარების მეთოდით, ობიექტთა თანმიმდევრული დალაგების/რანჟირების მეთოდით, მუდმივი ჯამური მაჩვენებლის მეთოდით და ა. შ.

შედარებითი სკალირების სპეციფიკა საშუალებას გვაძლევს გამოვავლინოთ და შევისწავლოთ განსხვავებები მოცემულ ობიექტებს შორის, აიოლებს მათ **შედარებით ანალიზს.**

სკალებს, რომლებსაც არაშედარებითი მეთოდით სარგებლობის შემთხვევაში იყენებენ, არაშედარების/დამოუკიდებლად გაზომვის სკალები (**noncomparative scales**) ეწოდება.

არაშედარებითი სკალირების შედეგად მიღებული მონაცემები არ არის ფარდობითი ხასიათის. მათი მეშვეობით შესაძლებელია თითოეული ობიექტის დახასიათება **დამოუკიდებლად,** სხვა ობიექტებთან **შედარების გარეშე.**

ამიტომ არაშედარებით სკალირებას (შესაბამისად, არაშედარების სკალებს) **მონადურს** ან **მეტრულსაც** უწოდებენ.

არაშედარების სკალები იყოფა **უწყვეტ რეიტინგულ** და **დეტალიზებულ რეიტინგულ სკალებად.**

დეტალიზებულ რეიტინგულ სკალებში, თავის მხრივ, შეიძლება გამოვყოთ **ლაიკერტის (Likert), სტეპელისა (Stapel)** და **სემანტიკური დიფერენციალის** სკალები.

❖ შედარებითი სკალირება

სკალირება წყვილური შედარების მეთოდით (paired comparison scaling).

წყვილური შედარების მეთოდით სკალირება ობიექტთა წყვილ-წყვილად შედარებას გულისხმობს. ამ მეთოდის საფუძველია რესპონდენტის მიერ ორი მოვლენის (მოვლენათა წყვილის წევრების) ერთმანეთთან შედარება, გარკვეული კრიტერიუმის მიხედვით.

წყვილური შედარების მეთოდით მიღებული მონაცემები სხვადასხვა წესით შეიძლება გაანალიზდეს.

მაგალითად, ობიექტთა **ყოველი წყვილისათვის** შეიძლება გამოვთვალოთ იმ რესპონდენტთა ფარდობითი სიხშირე, რომლებიც წყვილის ერთ-ერთ წევრს (ერთ-ერთ ობიექტს) **ანიჭებენ უპირატესობას მეორესთან შედარებით** მოცემული კრიტერიუმის მიხედვით.

დასაშვებია, აგრეთვე, შედარების **ყველა** ობიექტის **ერთობლივი/ერთდროული შეფასება:**

- წყვილური შედარების შედეგების დალაგება ხდება **უპირატესობის მინიჭების ტრანზიტულობის (transitivity of preferences)** საფუძველზე: ითვლება, რომ თუ რესპონდენტი

ერთი და იგივე კრიტერიუმით **A** ობიექტს ანიჭებს უპირატესობას **B**-ს მიმართ, ხოლო **B** ობიექტს **C**-ს მიმართ, მაშინ იგი ამავე კრიტერიუმით **A** ობიექტს ანიჭებს უპირატესობას **C**-ს მიმართ.

- მონაცემები ლაგდება სვეტებად მინიჭებული უპირატესობების მიხედვით და ხდება იმის განსაზღვრა, თუ რამდენჯერ მიენიჭა უპირატესობა თითოეულ ობიექტს.

.მაგალითად:

	A ობიექტი	B ობიექტი	C ობიექტი	D ობიექტი	E ობიექტი	F ობიექტი
A ობიექტი	[illegible]	0	0	0	0	0
B ობიექტი	1	[illegible]	0	0	0	0
C ობიექტი	1	1	[illegible]	0	0	0
D ობიექტი	1	1	1	[illegible]	0	0
E ობიექტი	1	1	1	1	[illegible]	0
F ობიექტი	1	1	1	1	1	[illegible]
უპირატესობის მინიჭების შემთხვევათა რაოდენობა	5	4	3	2	1	0

მოცემულ ცხრილში ყოველ უჯრას შეესაბამება შედარების ორი განსხვავებული ობიექტი — ერთის დასახელება მითითებულია სვეტის დასაწყისში, მეორესი კი — სტრიქონის დასაწყისში. უჯრაში ინერება 1, თუ რესპონდენტმა სვეტის დასაწყისში მითითებულ ობიექტს მიანიჭა უპირატესობა სტრიქონის დასაწყისში მითითებულ ობიექტთან შედარებით და ინერება 0, თუ სტრიქონის დასაწყისში მითითებულ ობიექტს მიანიჭა უპირატესობა სვეტის დასაწყისში მითითებულ ობიექტთან შედარებით. ბოლო სტრიქონში მითითებულია, რამდენჯერ მიენიჭა უპირატესობა **თითოეულ** ობიექტს/მოვლენას.

ხშირია წყვილური შედარების მოდიფიცირებული/გარდაქმნილი მეთოდების გამოყენების შემთხვევებიც. მაგალითად, შესაძლებელია, რესპონდენტს ვთხოვოთ, რომ არა მხოლოდ მიანიჭოს უპირატესობა წყვილის ერთ-ერთ წევრს, არამედ განსაზღვროს/დააკონკრეტოს უპირატესობის ხარისხიც; შესაძლო პასუხებს შორის ჩავრთოთ ნეიტრალური პასუხები: არც ერთს არ ვანიჭებ უპირატესობას, მიჭირს პასუხის გაცემა და ა. შ.

წყვილური შედარების მეთოდების გამოყენების ეფექტურობა მნიშვნელოვნადაა დამოკიდებული შესადარებელი ობიექტების რაოდენობაზე — ობიექტთა დიდი რაოდენობის შემთხვევაში რთულდება როგორც შედარების[1], ისე ანალიზის პროცესი.

სკალირება თანმიმდევრობით[1] დალაგების/რანჟირების მეთოდით (rank-order scaling)

თანმიმდევრობით დალაგების (რანჟირების) მეთოდით სკალირება გულისხმობს რესპონდენტთა მიერ ობიექტების ერთდროულ, თანმიდევრულ დალაგებას (ანუ მოცემული ობიექტების რანჟირებას) გარკვეული კრიტერიუმის მიხედვით. ამ მეთოდით სკალირების შედეგად თითოეულ ობიექტს ენიჭება განსაზღვრული შეფასება ანუ რანგი წინასწარ

[1] იზრდება ინტერვიუს ჩატარების დრო, რესპონდენტი იღლება, ხდება ყურადღების მოდუნება, მატულობს უპირატესობათა მინიჭიების ტრანზიტულობის პირობის დარღვევის ალბათობა და ა. შ.

შემუშავებული ინსტრუქციის შესაბამისად. მაგალითად, n ობიექტის შემთხვევაში, რესპონდენტებს შეიძლება ვთხოვოთ, რომ:

- ჩაწერონ რიცხვი 1 იმ ობიექტის გასწვრივ, რომელსაც ანიჭებენ უპირატესობას ყველა დანარჩენის მიმართ მოცემული კრიტერიუმის მიხედვით.
- ჩაწერონ რიცხვი 2 იმ ობიექტის გასწვრივ, რომელსაც ანიჭებენ უპირატესობას ყველა დარჩენილ n-1 ობიექტს შორის მოცემული კრიტერიუმის მიხედვით.
- და ა. შ. ბოლო რიცხვი: n მიენიჭება იმ ობიექტს, რომელსაც მოცემული კრიტერიუმის მიხედვით რესპონდენტი ყველაზე ნაკლებ უპირატესობას ანიჭებს.

შედეგად, თითოეულ ობიექტს მიენიჭება რანგი, რომელიც გამოხატავს, თუ რამდენად უფრო უპირატესია იგი რესპონდენტისათვის სხვა ობიექტებთან შედარებით მოცემული კრიტერიუმის მიხედვით. მაგალითად:

M	ობიექტი	1
N	ობიექტი	4
K	ობიექტი	5
L	ობიექტი	6
I	ობიექტი	2
J	ობიექტი	3

სკალირების ეს მეთოდი, წყვილურთან შედარებით, უფრო იოლია, სწრაფია და გამორიცხავს უპირატესობის მინიჭების ტრანზიტულობის თვისების დარღვევას.

უპირატესობათა მინიჭების ტრანზიტულობის თვისების გამო იოლია, რომ რიგის მიხედვით დალაგების მონაცემები წარმოვადგინოთ წყვილური დალაგების მონაცემების სახით და პირიქით.

სკალირება მუდმივი ჯამური მაჩვენებლის მეთოდით (constant sum scaling)

მუდმივი ჯამური მაჩვენებლის მეთოდით სკალირება გულისხმობს რესპონდენტთა მიერ ობიექტებს შორის ქულების, ბარათების, ჟეტონების, რომელიმე ვალუტაში (მაგალითად, ლარებში) გამოხატული თანხის და ა. შ. მოცემული (ჯამური) რაოდენობის განაწილებას გარკვეული კრიტერიუმის მიხედვით.

მაგალითად:

I. სოციალურ მოვლენებს/რეალობებს შორის იმის შესაბამისად, თუ რამდენად ახასიათებს თითოეულ მოცემულ სოციალურ მოვლენას გარკვეული ნიშან-თვისება დანარჩენებთან შედარებით.

II. ნიშან-თვისებათა შორის იმის შესაბამისად, თუ რამდენად ახასიათებს თითოეული მოცემული ნიშან-თვისება გარკვეულ სოციალურ მოვლენას/რეალობას დანარჩენებთან შედარებით.

მაგალითად:

I. რესპონდენტებს შეიძლება ვთხოვოთ, რომ პოლიტიკურ შეხედულებათა სისტემებს შორის გაანაწილონ 100 ქულა შემდეგი პრინციპით:

- ყოველ სისტემას, რომელსაც, მათი აზრით, "ლიბერალურობის" ნიშან-თვისება უმნიშვნელოდ/მიზერულად ახასიათებს, მიანიჭონ 0 ქულა;
- თუ თვლიან, რომ ერთ სისტემას მოცემული ნიშან-თვისება n-ჯერ მეტად ახასიათებს, ვიდრე მეორეს, მაშინ პირველს მიანიჭონ მეორესთან შედარებით n-ჯერ მეტი ქულა (ან, რაც იგივეა, მეორეს მიანიჭონ პირველთან შედარებით n-ჯერ ნაკლები ქულა);

[1] ამ მეთოდს რიგის მიხედვით დალაგების მეთოდიც შეიძლება ვუწოდოთ.

- მინიჭებული ქულების საერთო რაოდენობა უნდა უდრიდეს 100-ს[1].

პოლიტიკურ შეხედულებათა ყოველი მოცემული სისტემისათვის გამოითვლება ზოგადი მაჩვენებელი სათითაო რესპონდენტის მიერ ამ სისტემისათვის მინიჭებული ქულების ჯამის შეფარდებით რესპონდენტების საერთო რაოდენობასთან.

შედეგად, შესაძლებელი გახდება, გამოვყოთ რესპონდენტთა ჯგუფები/სეგმენტები იმის შესაბამისად, თუ **რომელ სისტემას/სისტემებს ანიჭებენ ისინი უპირატესობას სხვებთან შედარებით (რომელი პოლიტიკური სისტემა/სისტემებია მათი აზრით უფრო ლიბერალური).**

II. რესპონდენტებს შეიძლება ვთხოვოთ, რომ პოლიტიკურ შეხედულებათა სისტემისათვის დამახასიათებელ რამდენიმე ნიშან-თვისებას შორის გაანაწილონ 100 ქულა იმის მიხედვით, თუ რამდენად ახასიათებს ეს ნიშან-თვისება, დავუშვათ, დემოკრატიულ სისტემას შემდეგი პრინციპით:

- ყოველ ნიშან-თვისებას, რომელიც, მათი აზრით, უმნიშვნელოდ/მიზერულად ახასიათებს პოლიტიკურ შეხედულებათა დემოკრატიულ სისტემას მიანიჭონ 0 ქულა;
- თუ ისინი თვლიან, პოლიტიკურ შეხედულებათა დემოკრატიულ სისტემას ერთი ნიშან-თვისება (დავუშვათ, ეგალიტარულობა) **n**-ჯერ მეტად ახასიათებს, ვიდრე მეორე (დავუშვათ, რადიკალურობა), მაშინ პირველს მიანიჭონ მეორესთან შედარებით **n**-ჯერ მეტი ქულა (ან, რაც იგივეა, მეორეს მიანიჭონ პირველთან შედარებით **n**-ჯერ ნაკლები ქულა);
- მინიჭებული ქულების საერთო რაოდენობა უნდა უდრიდეს 100-ს.

ყოველი ნიშან-თვისების ზოგადი მაჩვენებელი (რომელიც გამოხატავს თუ რამდენადაა იგი დამახასიათებელი პოლიტიკურ შეხედულებათა დემოკრატიული სისტემისათვის სხვა ნიშან-თვისებებთან შედარებით), გამოითვლება სათითაო რესპონდენტის მიერ ამ ნიშან-თვისებისათვის მინიჭებულ ქულათა ჯამის შეფარდებით რესპონდენტების საერთო რაოდენობასთან.

შედეგად, შესაძლებელი გახდება გამოვყოთ რესპონდენტთა ჯგუფები/სეგმენტები იმის შესაბამისად, თუ **რომელ ნიშან-თვისებას/ნიშან-თვისებებს ანიჭებენ ისინი უპირატესობას დანარჩენებთან შედარებით, როგორც პოლიტიკურ შეხედულებათა დემოკრატიული სისტემისათვის დამახასიათებელს.**

სკალირება მუდმივი ჯამური მაჩვენებლის მეთოდით იმ სირთულესთანაა დაკავშირებული, რომ რესპონდენტებს შეიძლება გაუძნელდეთ მოცემული ჯამური მაჩვენებლის პირობებში ადეკვატური პროპორციული განაწილება.

სკალირება Q-სორტირების მეთოდით (Q-sort scaling)

Q-სორტირების მეთოდით სკალირება გულისხმობს რესპონდენტთა მიერ ობიექტების **დალაგებულ** ჯგუფებად დაყოფას/განაწილებას გარკვეული კრიტერიუმის მიხედვით, ობიექტებს შორის არსებული მსგავსება-განსხვავების მიმართების საფუძველზე.

მაგალითად, რესპონდენტებს შეიძლება შევთავაზოთ 80 დებულება და ვთხოვოთ, რომ ისინი დაალაგონ 10 ჯგუფად იმის მიხედვით, თუ რამდენად უფრო ეთანხმებიან თითოეულს დანარჩენებთან შედარებით.

შეფასებათა (**ratings**) თანაბარგანაწილების უზრუნველყოფის მიზნით, გამოიყენება "იძულებით-ნორმალური" (**forced normal**) განაწილება, როდესაც რესპონდენტს სთხოვენ, რომ მან თითოეულ ჯგუფში მკაცრად განსაზღვრული რაოდენობის დებულებები შეიტანოს.

[1] **სწორედ ამ პირობის შესაბამისად მიიღო განსახილველმა მეთოდმა სახელად მუდმივი ჯამური მაჩვენებლის მეთოდი.**

ჯგუფების რაოდენობა და თითოეულ ჯგუფში შემავალ დებულებათა რაოდენობა დებულებების საერთო რაოდენობიდან გამომდინარე, ნორმალური განაწილების ცხრილის მიხედვით შეიძლება განისაზღვროს.

შედარების ნევრთა/ობიექტთა რიცხვი მოთავსებული უნდა იყოს 60-დან 140-ის ფარგლებში. ოპტიმალურად ითვლება შედარების ნევრთა რაოდენობის 60-90 ინტერვალში მოქცევა.

ეს მეთოდი განსაკუთრებით მოხერხებულია მაშინ, როდესაც ერთმანეთთან შესადარებელი ობიექტების რაოდენობა დიდია.

სკალირება გუტმანის (Guttman) მეთოდით

(ანუ სკალური დიაგრამების/სკალოგრამული ანალიზის მეთოდით).

სკალოგრამული ანალიზის მეთოდის ამოსავლად შეიძლება ჩაითვალოს მოსაზრება, რომ ხშირ შემთხვევაში, მაშიც კი, როდესაც რაიმე მოვლენას დამოუკიდებლად, სხვა მოვლენებთან უშუალო შედარების გარეშე ახასიათებს, ადამიანი განიცდის გარკვეული სტერეოტიპული ხასიათის იერარქიის გავლენას, რომლის არსებობა შეიძლება არც ჰქონდეს გაცნობიერებული.

ცხადია, რომ თუ ჩვენ ამ "შინაგანი", ლატენტური იერარქიის გამოვლენას შევძლებთ, საკვლევი საკითხის ანალიზი უფრო ღრმა და მრავალმხრივ ხასიათს შეიძენს.

ამ მოსაზრებებიდან გამომდინარე, სკალოგრამული ანალიზის მიზანია მოვლენათა ერთგანზომილებიანი კონტინუუმის, "შინაგანი"/ლატენტური ლოგიკით დალაგებული იერარქიის/სკალის გამოვლენა.

მისი გამოყენება განსაკუთრებით ეფექტურია, როდესაც არსებობს ჰიპოთზა იმის შესახებ, რომ შესასწავლი მოვლენების ერთობლიობა იერარქიული ხასიათისაა. ასეთ შემთხვევაში სკალოგრამული ანალიზის მეთოდი საშუალებას გვაძლევს შევამოწმოთ ჰიპოთეზა, ხოლო, მისი ვერიფიცირების შემთხვევაში, აღვწეროთ და გავაანალიზოთ საკვლევ მოვლენათა იერარქია.

მეთოდი სამ ეტაპად შეიძლება აღვწეროთ.

პირველ ეტაპზე ხდება იმ დებულებების შემუშავება, რომელთა მეშვეობითაც შესაძლებელია დავაფიქსიროთ რესპონდენტთა დამოკიდებულება საძიებელი/სავარაუდო იერარქიის/სკალის თითოეული კომპონენტის/მნიშვნელობის მიმართ.

დებულებები ერთმანეთისაგან დამოუკიდებელია. რესპონდენტმა მათი ურთიერთშედარება კი არა, თითოეული შეთავაზებული დებულების **დამოუკიდებელი** შეფასება უნდა განახორციელოს — უნდა მიუთითოს, ეთანხმება თუ არ ეთანხმება მოცემულ დებულებას.

დავუშვათ, კვლევის ამოცანაა ადამიანთა გარკვეული (A) ჯგუფისადმი არსებული დამოკიდებულების შესწავლა. საწყისი დებულებები შემდეგი სახით შეიძლება ჩამოყალიბდეს:

I. A ჯგუფის წარმომადგენლები შრომისმოყვარენი არიან
ვეთანხმები (1) არ ვეთანხმები (0)
II. A ჯგუფის წარმომადგენლები უპასუხისმგებლოები არიან
ვეთანხმები (0) არ ვეთანხმები (1)
III. A ჯგუფის წარმომადგენლები ნიჭიერები არიან
ვეთანხმები (1) არ ვეთანხმები (0)
IV. A ჯგუფის წარმომადგენლები მატყუარები არიან
ვეთანხმები (0) არ ვეთანხმები (1)
V. A ჯგუფის წარმომადგენლები სამართლიანები არიან
ვეთანხმები (1) არ ვეთანხმები (2)

და ა. შ. მოცემულ ჩამონათვალში ორი ტიპის: დადებითი და უარყოფითი დამოკიდებულებების გამომხატველი დებულებები შედის. ამასთან, არ არის აუცილებელი, რომ მათი რაოდენობა ტოლი იყოს.

მეორე ეტაპზე ხდება იმის გარკვევა, თუ რამდენად შესაძლებელია მოცემული

დებულებების დალაგება გარკვეული თანმიმდევრობით.

ყოველი რესპონდენტის პასუხების დაფიქსირების შემდეგ ხდება მინიჭებული ქულების — კოდების შეჯამება.

ერთი რესპონდენტისათვის ქულების ჯამის მაქსიმალური მნიშვნელობა უდრის დებულებათა რიცხვს, ხოლო მინიმალური — ნულს.

ხოლო ყველა რესპონდენტის გამოკითხვის შემდეგ ხდება ქულების შეჯამება თითოეული დებულებისათვის.

ერთი დებულებისათვის ქულების ჯამის მაქსიმალური მნიშვნელობა უდრის რესპონდენტთა საერთო რაოდენობას, ხოლო მინიმალური — ნულს.

გამოკითხვის შედეგები შემდეგი ცხრილის სახით შეიძლება წარმოვადგინოთ:

რესპონდენტის ნომერი	დებულება					ჯამი
	I	II	III	IV	V	
1	1	1	1	0	1	4
2	0	1	0	1	1	3
3	0	1	1	1	1	4
4	0	0	0	0	1	1
5	0	0	0	0	0	0
6	0	1	1	1	1	4
7	0	1	0	1	1	3
8	0	1	0	0	1	2
9	0	0	0	0	1	1
ჯამი	1	6	3	4	8	[illegible]

რესპონდენტთა მიერ მინიჭებული ქულების ჯამური მნიშვნელობების მიხედვით:

V დებულება > II დებულება > IV დებულება > III დებულება > I დებულება

მაგრამ, ამ შედეგის ინტერპრეტაცია და მისი ჩართვა შინაარსობრივ ანალიზში ჯერ კიდევ არ არის მიზანშეწონილი, რადგან არ არის გარკვეული, დებულებათა მიღებული იერარქია **რეალურ ტენდენციას/კანონზომიერებას ასახავს, თუ შერჩევითი ერთობლიობის** (გამოკითხული რესპონდენტების) ინდივიდუალურ პოზიციათა მექანიკური, **არაკანონზომიერი, შემთხვევითი რეზულტატია**.

ამიტომ, მესამე ეტაპზე ხდება მეორეზე გამოვლენილი იერარქიის/სკალის ხარისხის შემოწმება/შეფასება.

ამისათვის ცხრილი გარდაიქმნება შემდეგი წესით:

- რესპონდენტები ლაგდება მოცემული დებულებისადმი მინიჭებული ქულების ჯამური მნიშვნელობების კლების მიხედვით.

- მოცემული დებულებები ლაგდება რესპონდენტთა მიერ მინიჭებული ქულების ჯამური მნიშვნელობების კლების მიხედვით.

შედეგად, თავდაპირველი ცხრილი შემდეგ სახეს მიიღებს:

რესპონდენტის ნომერი	დებულება					ჯამი
	V	II	IV	III	I	
1	1	1	[illegible]	1	1	4
3	1	1	1	1	0	4
6	1	1	1	1	0	4
7	1	1	1	0	0	3
2	1	1	1	0	0	3
8	1	1	0	0	0	2
9	1	0	0	0	0	1
4	1	0	0	0	0	1
5	0	0	0	0	0	0
ჯამი	8	6	4	3	1	[illegible]

მიღებულ ცხრილს სკალოგრამა/სკალური დიაგრამა ეწოდება. საუკეთესო შემთხვევაში, სკალოგრამაზე ერთები და ნულები ცალსახად, იერარქიულად უნდა გაიმიჯნოს ერთმანეთისაგან: ერთები "გამყოფი, იერარქიული კიბის" ზემოთ, ხოლო ნულები — ქვემოთ უნდა დალაგდეს.

ჩვენს შემთხვევაში, იდეალურ სურათს პირველი რესპონდენტის მიერ მეოთხე დებულებისადმი მინიჭებული ნული არღვევს, მაგრამ ეს ჯერ კიდევ არ ნიშნავს, რომ მიღებული იერარქია არაკანონზომიერია/სკალის ხარისხი არ არის დამაკმაყოფილებელი. ემპირიული კვლევებისას იდეალური სკალების მიღების ალბათობა ძალიან მცირეა. ამიტომ, ხდება იმის შეფასება, თუ **რამდენად მნიშვნელოვანია გადახრა/ცდომილება** იდეალური ვარიანტიდან.

მაჩვენებელს, რომლის მიხედვითაც სკალოგრამის ხარისხის შეფასება ხდება, **კვლავწარმოების კოეფიციენტი** ეწოდება.

იგი ახასიათებს სკალოგრამის "გადახრილი"[1] უჯრების თანაფარდობას (რაოდენობრივ მიმართებას), უჯრების საერთო რაოდენობასთან.

თუ რესპონდენტთა რაოდენობას აღვნიშნავთ **n**-ით, დებულებათა რაოდენობას **m**-ით, ხოლო იმ უჯრების რაოდენობას, რომლებიც იერარქიას არღვევს, **k**-თი, მაშინ კვლავწარმოების კოეფიციენტის გამოსათვლელი ფორმულა შემდეგნაირად შეიძლება გამოვსახოთ:

$$R = \left(1 - \frac{k}{mn}\right) \times 100\%$$

ჩვენი მაგალითის შემთხვევაში:

$$R = \left(1 - \frac{1}{9 \times 5}\right) \times 100\%$$

სკალის ხარისხი არადამაკმაყოფილებლად ითვლება (ანუ სკალა უხარისხოდ ითვლება) მაშინ, თუ კვლავწარმოების კოეფიციენტი 90%-ზე ნაკლებია.

ჩვენი მაგალითის შემთხვევაში **R** ≈ 98%, ე. ი. სკალის ხარისხი მაღალია. შესაბამისად, შეიძლება დავასკვნათ, რომ სკალოგრამული ანალიზის შედეგად გამოვლენილი იერარქია **A** ჯგუფის წარმომადგენელთა მიმართ შეფასებითი დამოკიდებულების ზოგად ტენდენციას გამოხატავს.

[1] ე. ი. იმ უჯრების, რომლებიც არღვევს იერარქიას.

როგორც ვხედავთ, გუტმანის სკალით გაზომვა რთული პროცესია და გამორიცხული არაა, რომ გაზომვის შედეგად ჰიპოთეზა საკვლევი მოვლენების ერთგანზომილებიანი იერარქიის არსებობის შესახებ, ფალსიფიცირდეს.

ამიტომ, როგორც წესი, თავდაპირველად ხდება საწყის დებულებათა გამოცდა/პილოტაჟი რესპონდენტთა მცირე რაოდენობებზე და მათი კორექტირება, სანამ კვლავწარმოების კოეფიციენტის მიხედვით დებულებათა იერარქიის დამაკმაყოფილებელ ვარიანტს არ შევიმუშავებთ.

გუტმანის სკალით მიღებული მონაცემების ანალიზისას მხედველობიდან არ უნდა გამოგვრჩეს, რომ საწყის დებულებათა დამოუკიდებელი ხასიათის მიუხედავად, გამოკითხვის შედეგად მოპოვებული მონაცემები შედარებით ანალიზს ექვემდებარება და, დამაკმაყოფილებელი-/ხარისხიანი სკალის შემთხვევაში, საწყისი დებულებების ერთგანზომილებიან, თანმიმდევრულ დალაგებას/იერარქიას ვიღებთ.

შესაბამისად, სკალოგრამული ანალიზის შედეგები შედარებითი ხასიათისაა, ე. ი. მათი დამოუკიდებლად განხილვა არ არის მიზანშეწონილი.

სკალირება მნიშვნელოვნების შეფასების მეთოდით

მნიშვნელოვნების შეფასების მეთოდით სკალირება გულისხმობს რესპონდენტთა მიერ ობიექტებისათვის შეფასებების/"ფასების" მინიჭებას იმის შესაბამისად, თუ რამდენად უფრო **მნიშვნელოვანია** ერთი ობიექტი **დანარჩენებთან შედარებით,** გარკვეული კრიტერიუმის მიხედვით.

ობიექტისათვის მინიჭებული შეფასებები/"ფასები" განსაზღვრული პრინციპით შერჩეული ნატურალური რიცხვებით (დავუშვათ, ერთიდან ასის ფარგლებში) შეიძლება გამოიხატოს.

ჩვენს მიერ განხილული მეთოდები შედარებითი სკალირების ყველა მეთოდს არ ამონურავს, მაგრამ წარმოდგენილი მასალა **საკმარისია** იმისათვის, რათა შედარებითი სკალირების ნებისმიერი **სხვა მეთოდის ან მეთოდთა კომბინაციის** გაგება/ გამოყენება შევძლოთ.

და ბოლოს, კიდევ ერთხელ შეგახსენებთ, რომ შედარებითი სკალირების შედეგად მიღებული **შედეგებით** ობიექტები მხოლოდ **ერთმანეთთან მიმართებაში ხასიათდება** — ისინი **ობიექტების დამოუკიდებელ შეფასებებად არ უნდა გავიაზროთ.**

მაგალითად, რომელიმე რესპონდენტის მიერ, **მოცემულ n** ობიექტს შორის **საუკეთესოდ მიჩნეული** ობიექტი, იმავე რესპონდენტს, **ამ** შედარების ფარგლებს გარეთ[1], შეიძლება დამაკმაყოფილებლადაც კი არ მიაჩნდეს.

ანუ, სხვაგვარად თუ ვიტყვით, მოცემულ ობიექტებს შორის "საუკეთესო" ჯერ კიდევ არ ნიშნავს "კარგს", "უარესი" — "ცუდს" და ა. შ.

❖ **არაშედარებითი/დამოუკიდებლად გაზომვის სკალირება**

სკალირება უწყვეტი რეიტინგული სკალით (continuous rating scale)

უწყვეტი რეიტინგული სკალით გაზომვა გულისხმობს რესპონდენტის მიერ ობიექტის შეფასებას მოცემული კრიტერიუმის მიხედვით და ამ შეფასების აღნიშვნას კრიტერიუმის ზღვრული მნიშვნელობების შემაერთებელ უწყვეტ წირზე[2].

უწყვეტი რეიტინგული სკალა სხვადასხვა ფორმით შეიძლება გამოისახოს. მაგალითად:

გთხოვთ, შეაფასოთ, რამდენად ნარმატებულია თბილისის სახელმწიფო უნივერსიტეტში მიმდინარე სასწაელო რეფორმები:

---

[1] **თავისთავად, დამოუკიდებლად აღებული ან შედარების დანარჩენი წევრებისაგან განსხვავებულ ობიექტთან შედარებით**

[2] **ე. ი. უწყვეტი სკალის.**

სავსებით ნარმატებულია .....I..... სრულიად წარუმატებელია
სავსებით ნარმატებულია .....I..... სრულიად წარუმატებელია
სავსებით ნარმატებულია .....I..... სრულიად წარუმატებელია

და ა. შ.

უწყვეტი სკალების აგება, აღქმა და გამოყენება საკმაოდ იოლია, მაგრამ რთულია სიზუსტის დაცვა როგორც გაზომვის, ისე ანალიზის პროცესში.

დეტალიზებული რეიტინგული სკალების (itemized rating scale) სპეციფიკა

- დეტალიზებულ რეიტინგულ სკალაზე აღნიშნულია რიცხვები ან მოკლე ვერბალური აღწერები.
- თითოეული აღნიშვნა მოცემული ობიექტის მიმართ გარკვეული კრიტერიუმით განსაზღვრული დამოკიდებულების ერთ და მხოლოდ ერთ გრადაციას/მნიშვნელობას გამოხატავს.
- რესპონდენტმა უნდა მიუთითოს ის გრადაცია, რომელიც ყველაზე მეტად შეესაბამება მოცემული ობიექტისადმი მის დამოკიდებულებას.

ლაიკერტის სკალა[1] (Likert scale)

ლაიკერტის სკალის გამოყენებისას რესპონდენტს სთხოვენ შეაფასოს, რამდენად ეთანხმება იგი დებულებებს, რომლებშიც გამოხატულია შეფასებითი დამოკიდებულებები საკვლევი მოვლენის მიმართ, გარკვეული კრიტერიუმების მიხედვით.

ჩვეულებრივ, ლაიკერტის სკალის თითოეული პუნქტი წარმოდგენილია ხუთი გრადაციით (თითოეული დებულება ხუთი გრადაციით ფასდება/იზომება).

ლაიკერტის სკალის მაგალითი:

- ინსტრუქცია:

ახლა გაგაცნობთ სხვადასხვა მოსაზრებას თბილისის სახელმწიფო უნივერსიტეტში მიმდინარე სასწავლო რეფორმების შესახებ. გთხოვთ, მიუთითოთ, რამდენად ეთანხმებით თითოეულ მათგანს.

თქვენი პოზიციის გამოხატვისას ისარგებლეთ შემდეგი აღნიშვნებით:

5 — სავსებით ვეთანხმები
4 — უფრო ვეთანხმები, ვიდრე არ ვეთანხმები
3 — ნაწილობრივ ვეთანხმები და ნაწილობრივ არ ვეთანხმები (თანაბრად)
2 — უფრო არ ვეთანხმები, ვიდრე ვეთანხმები
1 — სრულიად არ ვეთანხმები

1. თბილისის სახელმწიფო უნივერსიტეტში მიმდინარე სასწავლო რეფორმების შედეგად სტუდენტთა აკადემიური მოსწრება გაუმჯობესდა	5	4	3	✓	2	1
2. თბილისის სახელმწიფო უნივერსიტეტში მიმდინარე სასწავლო რეფორმების შედეგად სტუდენტთა ლექციებზე დასწრებამ მოიმატა	5	✓	4	3	2	1
3. თბილისის სახელმწიფო უნივერსიტეტში მიმდინარე სასწავლო რეფორმების შედეგად სტუდენტები უფრო აქტიურები გახდნენ	✓	5	4	3	2	1

[1] სკალირების ეს მეთოდი შეიმუშავა რენსის ლაიკერტმა.

4. თბილისის სახელმწიფო უნივერსიტეტში მიმდინარე სასწავლო რეფორმების შედეგად მასწავლებელთა სამუშაო პირობები გაუარესდა	5	4	3	2	✓	1
5. თბილისის სახელმწიფო უნივერსიტეტში სასწავლო რეფორმები წარმატებით მიმდინარეობს	5	✓	4	3	2	1

და ა. შ.

"საკმაოდ ვეთანხმები" და "სრულიად არ ვეთანხმები" ზღვრული მნიშვნელობებია, "ნაწილობრივ ვეთანხმები და ნაწილობრივ არ ვეთანხმები (თანაბრად)" ნეიტრალურ ("ნულოვან", საშუალო, შუალედურ) მნიშვნელობას გამოხატავს (იგი არ არის გადახრილი რომელიმე ზღვრული მნიშვნელობისაკენ), ხოლო "უფრო ვეთანხმები, ვიდრე არ ვეთანხმები" და "უფრო არ ვეთანხმები, ვიდრე ვეთანხმები" გამოსახავს გრადაციებს საშუალო და ზღვრულ მნიშვნელობებს შორის.

ლაიკერტის სკალის თითოეული მნიშვნელობა/გრადაცია შეიძლება შეფასდეს ქულებით მინუს –2-დან 2-ის ჩათვლით ან 1-დან 5-ის ჩათვლით.

ლაიკერტის სკალით მიღებული მონაცემების გასაანალიზებლად იყენებენ **პროფილური ანალიზის** ან **ჯამური შეფასების** მეთოდებს.

პროფილური ანალიზი ხორციელდება თითოეული პუნქტის საშუალო მნიშვნელობის მიხედვით.

ამ ანალიზის ფარგლებში, შესაძლებელია, რომ ერთმანეთს შევადაროთ:

- როგორ აფასებს რესპონდენტთა ერთი და იგივე ჯგუფი განსხვავებულ დებულებებს;
- როგორ აფასებს რესპონდენტთა სხვადასხვა ჯგუფი ერთსა და იმავე დებულებებს.

ჯამური შეფასების მეთოდი გულისხმობს მონაცემთა გაანალიზებას ყოველი რესპონდენტის მიერ დებულებებისათვის მინიჭებული ქულების საერთო, ჯამური მაჩვენებლის საფუძველზე. ეს მეთოდი იმდენად გავრცელებულია, რომ ლაიკერტის სკალას ზოგჯერ ჯამურ სკალასაც (**summated scale**) უწოდებენ.

ჯამური შეფასების მეთოდის გამოყენების აუცილებელი პირობაა საწყისი სკალების ერთგვაროვნება.

ჩვენს მიერ განხილულ მაგალითში შეფასების სისტემა არაერთგვაროვანია იმ თვალსაზრისით, რომ რაც უფრო მაღალია მეოთხე დებულებისადმი თანხმობის ხარისხი, მით უარყოფითია დამოკიდებულება თბილისის სახელმწიფო უნივერსიტეტში მიმდინარე სასწავლო რეფორმების მიმართ, ყველა სხვა დებულებისადმი თანხმობის ხარისხის მატებას კი დადებითი დამოკიდებულების გრადაციის მატება შეესაბამება.

ბუნებრივია, ასეთ შემთხვევაში ქულების პირდაპირი/მექანიკური შეფასება არაადეკვატურ შედეგებს მოგვცემს. ამიტომ ანალიზისათვის მონაცემთა მომზადების დროს სკალები შესაბამისად უნდა გარდაიქმნას.

მოცემულ შემთხვევაში ეს ამოცანა შეიძლება გადაწყდეს მეოთხე სკალის მნიშვნელობათა ტრანსფორმაციით: საშუალო მნიშვნელობის მიმართ სკალის დანარჩენი მნიშვნელობების სიმეტრიული ურთიერთშენაცვლებით:

4. თბილისის სახელმწიფო უნივერსიტეტში მიმდინარე სასწავლო რეფორმების შედეგად მასწავლებელთა სამუშაო პირობები გაუარესდა	1	2	3	4	✓	5

სადაც:

1 — სავსებით ვეთანხმები
2 — უფრო ვეთანხმები, ვიდრე არ ვეთანხმები

3 — ნანილობრივ ვეთანხმები და ნაწილობრივ არ ვეთანხმები (თანაბრად)
4 — უფრო არ ვეთანხმები, ვიდრე ვეთანხმები
5 — სრულიად არ ვეთანხმები

სკალების ასეთი ტრანსფორმაციის საფუძველზე ითვლება, რომ რესპონდენტი, რომელიც დებულებათა მოცემულ სისტემას უფრო მეტ ქულას/ჯამურ მაჩვენებელს ანიჭებს, უფრო დადებითად აფასებს გაზომვის ობიექტს.

სემანტიკური დიფერენციალი[1] **(semantic differential)**

სემანტიკური დიფერენციალი ეწოდება შვიდდანაყოფიან/შვიდქულიან სკალას, რომლის ზღვრულ/კიდურა დანაყოფებსაც კონტრადიქტორული ანუ ურთიერთგამომრიცხავი მნიშვნელობები შეესაბამება.

როგორც წესი, სემანტიკური დიფერენციალის ზღვრული მნიშვნელობები ურთიერთ-საპირისპირო ზედსართავი სახელებით გამოიხატება.

მაგალითად: "ნიჭიერი" — "უნიჭო", "ერთგული" — "მოღალატე" და ა. შ.

სკალის დანაყოფები/მნიშვნელობები მოცემული ობიექტის მიმართ გარკვეული კრიტერიუმის შესაბამისი დამოკიდებულების გრადაციებს გამოხატავს. რესპონდენტმა უნდა შეაფასოს, თუ რომელი მათგანი გამოხატავს უფრო მეტად საკვლევი მოვლენისადმი მის დამოკიდებულებას.

შეფასებისა და ანალიზის გასამარტივებლად მიზანშეწონილია, რომ ნეგატიური/უარყოფითი ზღვრული მნიშვნელობები ერთ მხარეს (დავუშვათ, მარჯვნივ) დალაგდეს, ხოლო პოზიტიური/დადებითი — მეორე მხარეს (მარცხნივ).

სემანტიკური დიფერენციალური სკალის მაგალითი:

- ინსტრუქცია:

გთხოვთ, ქვემოთ ჩამოთვლილ ნიშან-თვისებებსა და მათ შორის არსებულ "გარდამავალ საფეხურებს" შორის მიუთითოთ ის, რომელიც ყველაზე მეტად შეეფერება **თქვენს** შეხედულებას:

1. საქართველოში მცხოვრები **A** ეროვნების წარმომადგენლების მიმართ:

თანამედროვე	_____X_____	**ჩამორჩენილი**
ნიჭიერი	____X_____	**უნიჭო**
სამართლიანი	__X_____	**უსამართლო**
მამაცი	X_____	**მშიშარა**
ერთგული	____X_____	**მოღალატე**

და ა. შ.

სემანტიკური დიფერენციალური სკალის თითოეული დანაყოფი/გრადაცია შეიძლება შეფასდეს ქულებით მინუს –3-დან 3-ის ჩათვლით ან 1-დან 7-ის ჩათვლით.

მიღებული შედეგები შეიძლება გაანალიზდეს:

- **პროფილური ანალიზის** მეთოდით, სკალის თითოეული პუნქტისათვის **საშუალო** ან **მედიანური** მნიშვნელობების საფუძველზე, რაც საშუალებას გვაძლევს, ერთმანეთს შევადაროთ: როგორ აფასებს რესპონდენტთა ერთი და იგივე ჯგუფი განსხვავებულ

[1] იგივე სემანტიკური დიფერენციალური სკალა. ეს მეთოდიკა შეიმუშავეს ჩ. ოსგუდმა და მისმა თანამშრომლებმა (**Osgood, Suci, & Tannenbaum, 1957**), რის გამოც სემანტიკურ დიფერენციულ სკალას ზოგჯერ ოსგუდის სკალას/ოსგუდის სემანტიკურ დიფერენციალს უწოდებენ.

პუნქტებს და როგორ აფასებს რესპონდენტთა სხვადასხვა ჯგუფი ერთსა და იმავე პუნქტს;

- **ჯამური შეფასების მეთოდით.**

სტეპელის სკალა[1] (Stapel scale)

სტეპელის სკალა ათდანაყოფიანია. დანაყოფებს შეესაბამება მნიშვნელობები/ქულები 1-დან 10-ის ჩათვლით ან –5-დან 5-ის ჩათვლით. ამ სკალის ერთ-ერთი არსებითი თავისებურება ისაა, რომ მნიშვნელობებს შორის არ არის შეტანილი ნეიტრალური/"ნულოვანი"[2] წერტილი.

სკალირება ხორციელდება გარკვეული კრიტერიუმების შესატყვისი ტერმინების ან ლაკონიური ფრაზების საფუძველზე. რესპონდენტს სთხოვენ შეაფასოს, თუ რამდენად ზუსტად ახასიათებს თითოეული ტერმინი/ფრაზა მოცემულ ობიექტს.

სტეპელის სკალას, როგორც წესი, ვერტიკალურად გამოსახავენ.

მაგალითი:

- ინსტრუქცია:

გთხოვთ, —5-დან +5-ის ფარგლებში შეაფასოთ, რამდენად ზუსტად ახასიათებს ქვემოთ მოცემული ტერმინები/გამოთქმები თბილისის სახელმწიფო უნივერსიტეტში მიმდინარე რეფორმებს (+5 ნიშნავს, რომ ტერმინი/გამოთქმა თსუ-ში მიმდინარე რეფორმებს მაქსიმალური სიზუსტით ახასიათებს, ხოლო —5 ნიშნავს, რომ მოცემული ტერმინი/გამოთქმა თსუ-ში მიმდინარე რეფორმებს საერთოდ არ ახასიათებს):

+5	+5
+4	+4
+3	+3
+2	+2
+1	+1
ნარმატებული	**უსარგებლო**
—1	—1
—2	—2
—3	—3
—4	—4
—5	—5

და ა. შ.

მიღებული შედეგები შეიძლება გაანალიზდეს:

- **პროფილური ანალიზის** მეთოდით, სკალის თითოეული პუნქტისათვის **საშუალო** ან **მედიანური** მნიშვნელობების საფუძველზე, რაც საშუალებას გვაძლევს, ერთმანეთს შევადაროთ: როგორ აფასებს რესპონდენტთა ერთი და იგივე ჯგუფი განსხვავებულ პუნქტებს და როგორ აფასებს რესპონდენტთა სხვადასხვა ჯგუფი ერთსა და იმავე პუნქტს;
- **ჯამური შეფასების მეთოდით.**

[1] სკალირების ეს მეთოდი შეიმუშავა ჟან სტეპელმა.

[2] საშუალო, შუალედური პოზიციის გამომხატველი.

დეტალიზებული რეიტინგული სკალების მახასიათებლები და მათი მნიშვნელობა მონაცემთა ანალიზისას

I. სკალის დანაყოფთა რაოდენობა.

სკალის დანაყოფთა/მნიშვნელობათა რაოდენობის პრობლემას უკვე შევეხეთ. ამ საკითხის ძირითადი სირთულე ისაა, რომ, ერთი მხრივ, რაც უფრო მეტი დანაყოფია სკალაზე, მით უფრო **დეტალურია** ობიექტის დახასიათება, მეორე მხრივ კი, რაც მეტია დანაყოფთა რაოდენობა, მით მაღალია იმის ალბათობა, რომ რესპონდენტებს შეიძლება გაუჭირდეთ განსხვავების აღქმა "ახლო მეზობელ" დანაყოფებს შორის. "ტრადიციულად", სკალის დანაყოფების ოპტიმალურ რაოდენობად მიჩნეულია შვიდი, პლუს-მინუს ხუთი დანაყოფი. მაგრამ ეს პრობლემა ყოველი კვლევის ფარგლებში ინდივიდუალურად უნდა გადაწყდეს შემდეგი, ზოგადი ფაქტორების გათვალისწინებით:

- **რამდენად იძლევა ობიექტი სკალის დეტალიზების საშუალებას** (ანუ, **რამდენად დეტალურადაა** შესაძლებელი თვითონ გაზომვის ობიექტის დახასიათება);
- **რა მეთოდებია დაგეგმილი მონაცემთა გასაანალიზებლად** (რთული სტატისტიკური ანალიზების ჩასატარებლად ოპტიმალურად ითვლება შვიდ ან მეტდანაყოფიანი სკალით მოპოვებული მონაცემები);
- **რა მეთოდებით ვაპირებთ მონაცემთა მოპოვებას** (მაგალითად, სატელეფონო გამოკითხვების დროს რესპონდენტისათვის მაქსიმალურად დეტალიზებული სკალის აღქმა/გამოყენება რთულია);
- **რამდენად კარგად იცნობენ რესპონდენტები ობიექტს** (იმ შემთხვევაში, როდესაც რესპონდენტი საკვლევ ობიექტს კარგად იცნობს, იზრდება ალბათობა, რომ იგი ამ ობიექტის დეტალიზებულ შეფასებას იოლად და ადეკვატურად შეძლებს);
- **რამდენად დაინტერესებულნი არიან რესპონდენტები კვლევაში თანამონაწილეობით** (იმ შემთხვევაში, როდესაც რესპონდენტი დაინტერესებულია მიიღოს მონაწილეობა კვლევაში, იზრდება ალბათობა, რომ იგი მეტ გულმოდგინებას გამოიჩენს, **შეეცდება**, რაც შეიძლება ადეკვატურად შეასრულოს ინსტრუქცია. რესპონდენტთა დაინტერესების გაზრდის მიზნით, შეიძლება გამოვიყენოთ საჩუქრის ან წახალისების სხვა მეთოდები).

II. დაბალანსებული და დაუბალანსებელი სკალები

დაბალანსებულ სკალაზე (**balanced scale**) დადებითი და უარყოფითი მნიშვნელობების რაოდენობა ტოლია, დაუბალანსებელზე — განსხვავებული.

მონაცემთა ობიექტუროებისა და ადეკვატური ანალიზისათვის ოპტიმალურია დაბალანსებული სკალების გამოყენება, მაგრამ დაუბალანსებელი სკალა შეიძლება გამოვიყენოთ მაშინ, როდესაც დიდია ალბათობა, რომ მნიშვნელობები დადებითი ან უარყოფითი მხარისაკენ წანაცვლდება.

მსგავს შემთხვევაში, შესაძლებელია, სკალის დანაყოფები ერთი მხარის მიმართულებით უფრო დიფერენცირებულად წარმოვადგინოთ, ვიდრე მეორე მხარის მიმართულებით.

მაგალითად, როდესაც წინასწარი მონაცემებით, ამა თუ იმ მოვლენის მიმართ დამოკიდებულება ძირითადად დადებითია, შესაძლებელია, რომ მკვლევრი სწორედ დადებითი დამოკიდებულების გრადაციების **დეტალიზებით** დაინტერესდეს, უარყოფითი დამოკიდებულების გრადაციათა დიფერენცირება კი, **ნაკლებად დეტალურად** განახორციელოს.

დაბალანსებული სკალის მაგალითი:

გთხოვთ, შეაფასოთ **A** ჯგუფის საქმიანობა:

საუკეთესო

ძალიან კარგი

კარგი

ცუდი

ძალიან ცუდი
ერთ-ერთი ყველაზე უარესი

დაუბალანსებელი სკალის მაგალითი:
გთხოვთ, შეაფასოთ **A** ჯგუფის საქმიანობა:
საუკეთესო
ძალიან კარგი
კარგი
ცუდი
ძალიან ცუდი

დაუბალანსებელი სკალის გამოყენების შედეგად მიღებული მონაცემების ანალიზის დროს, გაზომვის ეს თავისებურება მხედველობიდან არ უნდა გამოგვრჩეს.

III. იძულებითი და ნებაყოფლობითი შეფასების სკალები

იძულებითი შეფასების სკალა (**forced rating scale**) არ იძლევა ისეთი პასუხების დაფიქსირების საშუალებას, რომლებიც უშუალოდ შეფასებას არ გამოხატავს ("მიჭირს პასუხის გაცემა", "არ ვიცი" და სხვა). ამით რესპონდენტს ვაიძულებთ დააფიქსიროს შეფასების რომელიმე მნიშვნელობა.

იძულებითი შეფასების სკალები კონკრეტული შეფასებების მიღებაში გვეხმარება, მაგრამ უნდა გავითვალისწინოთ, რომ უმეტეს შემთხვევაში, ის რესპონდენტები, რომლებსაც შეფასება ვაიძულეთ, შუალედურ პასუხებს ირჩევენ, რამაც, მონაცემთა ანალიზის დროს, ცენტრალური ტენდენციებისა და ვარიაციის მაჩვენებლების ხელოვნური გადახრა/დამახინჯება შეიძლება გამოიწვიოს.

ნებაყოფლობითი შეფასების სკალაზე შესაძლებელია "მიჭირს პასუხის გაცემა", "არ ვიცი" და სხვა მსგავსი პასუხების დაფიქსირება.

ერთგანზომილებიანი სტატისტიკა

გავლილი მასალიდან უკვე ვიცით რა არის ცვლადი და როგორ ხდება მისი მნიშვნელობების პოვნა — გაზომვა.

მაგრამ კვლევის დროს სოციოლოგისათვის ცვლადებისა თუ სკალების თეორიული განსაზღვრა მხოლოდ საშუალებაა. მისი ძირითადი მიზანი ემპირიული მონაცემების მოპოვება და ანალიზია.

ემპირიულ მონაცემთა ანალიზის პროცესში **მონაცემთა რაოდენობრივი მახასიათებლები** არსებით როლს ასრულებს.

ცვლადის თეორიული/შესაძლო მნიშვნელობების დიდი ნაწილი ემპირიული კვლევის გარეშეც შეიძლება განვსაზღვროთ. ემპირიულ მონაცემთა მოპოვება ძირითადად იმ მიზნით ხდება, რომ დავადგინოთ, **რამდენჯერ მიიღებს ცვლადი თითოეულ თეორიულ მნიშვნელობას**.

მაგალითად, როდესაც რაოდენობრივი სოციოლოგიური კვლევის დროს ბიოლოგიურ სქესს ვზომავთ, ცვლადის შესაძლო მნიშვნელობების განსაზღვრა (მდედრობითი, მამრობითი) ტრივიალურია. ჩვენი მიზანი ამ შემთხვევაში იმის დადგენაა, თუ რამდენჯერ მიიღებს ცვლადი მნიშვნელობა "მდედრობითს" და რამდენჯერ — "მამრობითს" (რამდენი **რესპონდენტი** — კვლევისათვის შერჩეული პირი აღმოჩნდება მდედრობითი და რამდენი — მამრობითი სქესის).

გასაგებია, რომ რაოდენობრივი მახასიათებლების აღწერა/ანალიზს სტატისტიკის სფეროსთან მივყავართ. ამიტომ სტატისტიკის ტერმინებს უკვე ინტენსიურად გამოვიყენებთ, თუმცა, ვაჩვენებთ, რომ ყოველი მათგანი გარკვეული სოციოლოგიური შინაარსის ადეკვატურია, რაც

სტატისტიკის ცნებითი აპარატის ათვისებასა და გამოყენებას მნიშვნელოვნად გააიოლებს.

ამ ადეკვატურობის ჩვენება მხოლოდ სასწავლო პროცესის გამარტივებას როდი ისახავს მიზნად. როგორც წინა მასალაში აღვნიშნეთ, სოციოლოგისათვის განმსაზღვრელი სოციოლოგიური ინტერპრეტაციაა და არა საკუთრივ რიცხვების სტატისტიკური თავისებურებები.

ამიტომ, **სტატისტიკური ანალიზის** ანუ, ემპირიული მონაცემების სტატისტიკის მეთოდებით დამუშავების დროს, სოციოლოგიურ ინტერპრეტაციასთან ადეკვატურობის კრიტერიუმით უნდა ვიხელმძღვანელოთ.

❖ ერთგანზომილებიანი სიხშირული განაწილებები

სოციოლოგიური გამოკითხვის შედეგად ათასობით მონაცემს ვიღებთ. მათ პირველადი მონაცემები ეწოდება. პირველადი მონაცემები **მოუწესრიგებელი და მრავალრიცხოვანია.** აქედან გამომდინარე, **სტატისტიკური ანალიზის საწყისი ეტაპი** მონაცემთა მასივების **სისტემატიზება და კომპაქტური** აღწერაა. ეს ამოცანა **სტატისტიკურ ერთობლიობათა დაჯგუფების** მეშვეობით წყდება.

სტატისტიკური ერთობლიობა შეიძლება განვმარტოთ, როგორც ნებისმიერი ერთმანეთისაგან გარჩევადი/თვლადი ობიექტების ერთობლიობა.

რაოდენობრივი სოციოლოგიური კვლევისას ეს შეიძლება იყოს რესპონდენტების ერთობლიობა, მონაცემთა ერთობლიობა და ა. შ.

ობიექტებს, რომლებიც სტატისტიკურ ერთობლიობას ქმნის, **სტატისტიკური ერთობლიობის წევრები, ერთეულები** ეწოდება.

სტატისტიკური ერთობლიობის წევრების რაოდენობას **სტატისტიკური ერთობლიობის მოცულობა ეწოდება.**

სტატისტიკური დაჯგუფება სტატისტიკური ერთობლიობის ჯგუფებად დაყოფის შემდეგი წესია:

- გარკვეული ნიშნის მიხედვით ხდება სტატისტიკური ერთობლიობის წევრებს შორის იგივეობა/განსხვავების მიმართების დამყარება;
- იგივეობრივ წევრთა ყოველი ერთობლიობა ცალკე **ჯგუფად** მოიაზრება. ყოველი ასეთი **ჯგუფი ერთგვაროვანია** იმ აზრით, რომ ჯგუფის ნებისმიერი ორი წევრი მოცემული არსებითი ნიშან-თვისებების ერთი და იგივე მნიშვნელობებით ხასიათდება;
- ნებისმიერი ორი ჯგუფი განსხვავდება ერთმანეთისაგან იმ აზრით, რომ ერთი ჯგუფის ნებისმიერი წევრი განსხვავდება მეორე ჯგუფის ნებისმიერი წევრისაგან მოცემული არსებითი ნიშან-თვისებებიდან ერთ-ერთის მნიშვნელობით მაინც;
- ყველა სხვა ნიშან-თვისება არაარსებითად ითვლება და ჯგუფებად დაყოფაზე/კლასიფიკაციაზე არ აისახება.

მაგალითად, თუ ადამიანებს ეროვნების მიხედვით დავყოფთ, მივიღებთ ეროვნების თითოეული მნიშვნელობის მიხედვით წარმოქმნილ ჯგუფებს:

- ყოველი ასეთი ჯგუფი ერთგვაროვანია იმ აზრით, რომ ამ ჯგუფში შემავალი ნებისმიერი ორი წევრის ეროვნება ერთი და იგივეა;
- ნებისმიერი ორი ჯგუფი განსხვავდება ერთმანეთისაგან იმ აზრით, რომ ერთი ჯგუფის ნებისმიერი წევრის ეროვნება განსხვავებული იქნება სხვა ჯგუფის ნებისმიერი ნევრის ეროვნებისაგან;
- ადამიანთა ნებისმიერი სხვა ნიშან-თვისება (მაგალითად, თვალების ფერი, სიმაღლე, განათლება, საქმიანობა, ჩაცმის სტილი და ა. შ.) ამ კლასიფიკაციის ფარგლებში **არაარსებითად ითვლება** და უგულებელყოფილია.

სოციოლოგიაში სტატისტიკური დაჯგუფების ნიშნად შეიძლება ავიღოთ ცვლადის მნიშვნელობები. ასეთ დაჯგუფებას **მარტივ დაჯგუფებას** უწოდებენ.

მარტივი დაჯგუფების შედეგად მიღებულ ინფორმაციას **ერთგანზომილებიანი განაწილება** ეწოდება.

მაგალითად, თუ რესპონდენტებს სქესის მიხედვით განვასხვავებთ/დავყოფთ, მივიღებთ რესპონდენტთა **ერთგანზომილებიან განაწილებას სქესის ნიშნით.**

ერთგანზომილებიანი განაწილება შეიძლება ნარმოვადგინოთ სტატისტიკური მწკრივის, რანჟირებული მწკრივის ან ვარიაციული მწკრივის სახით.

ერთგანზომილებიანი განაწილება ნარმოდგენილია **სტატისტიკური მწკრივის** სახით, თუ ცვლადის განსხვავებული მნიშვნელობები არ არის დალაგებული ზრდის/კლების მიხედვით. სტატისტიკურ მწკრივში შესაძლებელია მხოლოდ ცვლადის მნიშვნელობების ერთმანეთისაგან განსხვავება/გარჩევა ან მათ შორის იგივეობის დამყარება.

ერთგანზომილებიანი განაწილება ნარმოდგენილია **რანჟირებული მწკრივის** სახით, თუ ცვლადის მნიშვნელობები დალაგებულია ზრდის/კლების მიხედვით.

ერთგანზომილებიანი განაწილება ნარმოდგენილია **ვარიაციული მწკრივის**[1] სახით, თუ მწკრივის წევრებია ცვლადის იგივეობრივ მნიშვნელობათა ჯგუფები. ყოველი ჯგუფი აღინერება გარკვეული რიცხვით, რომელიც ჯგუფის სიდიდეს ახასიათებს/აღწერს.

ამრიგად, სტატისტიკური მწკრივი ეწოდება, უბრალოდ, მონაცემთა ჩამონათვალს, რომელიც **საშუალებას გვაძლევს ერთმანეთისაგან გავარჩიოთ მოცემული ნიშნით განსხვავებული წევრები;** რანჟირებული მწკრივი მიიღება სტატისტიკური მწკრივის წევრების ზრდის/კლების მიხედვით დალაგების შედეგად, ხოლო ვარიაციული მწკრივი — რანჟირებული მწკრივის იგივეობრივი წევრების ჯგუფებად გაერთიანების შედეგად.

დავუშვათ, კითხვას: "რამდენად კმაყოფილი ხართ უნივერსიტეტში მიმდინარე რეფორმებით?" 1000-მა რესპონდენტმა უპასუხა, ხოლო პასუხის ვარიანტებია:

1. სავსებით კმაყოფილი
2. უფრო კმაყოფილი, ვიდრე უკმაყოფილო
3. კმაყოფილიც და უკმაყოფილოც (თანაბრად)
4. უფრო უკმაყოფილო, ვიდრე კმაყოფილი
5. უკმაყოფილო

თუ ანკეტებიდან შემთხვევითად ამოვწერთ რესპონდენტთა პასუხებს, **სტატისტიკურ მწკრივს** მივიღებთ, თუ მიღებული მწკრივის მნიშვნელობებს დავალაგებთ ზრდის/კლების მიხედვით (მაგალითად, ჯერ ამოვწერთ ყველა "1"-ს, შემდეგ ყველა "2"-ს და ა. შ.) — **რანჟირებულს,** ხოლო თუ ამოვწერთ პასუხის თითოეულ ვარიანტს და გვერდით მივუწერთ რიცხვს, რომელიც ჯგუფის სიდიდეს გამოხატავს — **ვარიაციულს.**

განვიხილოთ ვარიაციული მწკრივის მაგალითი:

ცხრილი №4

ცვლადის მნიშვნელობა	აბსოლუტური სიხშირე	ფარდობითი სიხშირე	დაგროვილი სიხშირე
1. სავსებით კმაყოფილი	100	10%	10%
2. უფრო კმაყოფილი, ვიდრე უკმაყოფილო	250	25%	35%
3. კმაყოფილიც და უკმაყოფილოც (თანაბრად)	350	35%	70%
4. უფრო უკმაყოფილო, ვიდრე კმაყოფილი	150	15%	85%
5. სრულიად უკმაყოფილო	150	15%	100%
სულ	*1000*	*100%*	

რიცხვს, რომელიც გვიჩვენებს, თუ რამდენჯერ გვხვდება ცვლადის მოცემული

მნიშვნელობა ერთგანზომილებიან განაწილებაში, **აბსოლუტური სიხშირე** ეწოდება. 100, 250 და ა. შ. აბსოლუტური სიხშირეებია.

მაგრამ, მხოლოდ აბსოლუტური სიხშირეების განსაზღვრა ნაკლებად ინფორმატიულია. დავუშვათ, ვიცით, რომ ცვლადმა რომელიმე მნიშვნელობა 200-ჯერ მიიღო. ცოტაა ეს თუ ბევრი? ცვლადის მნიშვნელობების საერთო რაოდენობასთან შედარების გარეშე ამ კითხვას ვერ ვუპასუხებთ.

თუ ცვლადის მნიშვნელობათა საერთო რაოდენობა არის: ა) 2 000 — ცოტაა, ბ) 400 — საშუალო, გ) 300 — ბევრი.

ამიტომ, ერთგვაროვანი ჯგუფის სიდიდის მახასიათებლად ხშირად იყენებენ რიცხვს, რომელიც გვიჩვენებს შესაბამისი აბსოლუტური სიხშირის ხვედრით წილს ცვლადის ყველა მნიშვნელობის რაოდენობასთან ფარდობაში. ამ რიცხვს **ფარდობითი სიხშირე** ეწოდება.

10%, 25% და ა. შ. ფარდობითი სიხშირეებია. ფარდობითი სიხშირე პირდაპირ შეფარდების სახითაც (100/1000, 250/1000 და ა. შ.) შეიძლება გამოისახოს. მაგრამ წილადი რიცხვების სახით მოცემული ფარდობითი სიხშირეების ერთმანეთთან შედარება დამატებით არითმეტიკულ მოქმედებებთანაა დაკავშირებული. ამიტომ, სიმარტივისათვის მათ ხშირად პროცენტული სახით წარმოადგენენ.

დავუშვათ, სტატისტიკური ერთობლიობა n წევრისაგან შედგება და იგი განაწილებულია k ჯგუფად, მაშინ ყოველ i-ურ ჯგუფს შეესაბამება n_i აბსოლუტური სიხშირე ($i = 1, 2, ..., k$).

აბსოლუტური სიხშირის (n_i) ხვედრითი წილი (w_i) სტატისტიკურ ერთობლიობაში ანუ ფარდობითი სიხშირე გამოითვლება ფორმულით:

$$w_i = \frac{n_i}{n} \times 100\%$$

მაგალითად, *ცხრილი №4*-ის პირველი შემთხვევის მიხედვით:

$$w_1 = \frac{100}{1000} \times 100\% = 10\%$$

ცხრილის ბოლო სვეტში მითითებულია დაგროვილი/კუმულიატური (აკუმულირებული) სიხშირეები. **დაგროვილი** ეწოდება სიხშირეს, რომელიც ვარიაციული მწკრივის ყოველი i-რი მნიშვნელობისათვის $w_1 + w_2 + ... + w_i$ ფარდობითი სიხშირეების ჯამს გამოსახავს.

❖ ინტერვალებად დაჯგუფება

ინტერვალებად დაჯგუფების დროს სიხშირე ცვლადის გარკვეულ ინტერვალში მოქცეულ მნიშვნელობათა ერთობლიობას შეესაბამება და არა თითოეულ მნიშვნელობას ცალ-ცალკე.

მაგალითად, როდესაც რესპონდენტთა ასაკს რიგის სკალით — ასაკობრივი ჯგუფებით ვზომავთ: 18-22 წწ., 23-27 წწ. და ა.შ., 18, 19, 20, 21 და 22 წლის ყველა რესპონდენტი ერთ ინტერვალში მოექცევა, 23, 24, 25, 26, 27 წლის მეორეში და ა. შ.

სიხშირეები ინტერვალებს (ინტერვალში მოქცეულ ყველა მნიშვნელობას ერთად აღებულს) შეესაბამება და არა ასაკის თითოეულ მნიშვნელობას.

---

[1] ტერმინი "ვარიაციული" შერჩეულია იმ მოსაზრებით, რომ თითოეული ჯგუფი ცვლადის მნიშვნელობის ამა თუ იმ თეორიულ ვარიანტს (პასუხის შესაძლო მნიშვნელობას, ვარიანტს) შეესაბამება.

ინტერვალებად დაჯგუფებულ ვარიაციულ მწკრივს **ინტერვალური ვარიაციული მწკრივი** ეწოდება.

ინტერვალებად დაჯგუფების დროს იგულისხმება, რომ ცვლადის მნიშვნელობები ინტერვალის შიგნით **თანაბრად** ნაწილდება. ამიტომ ინტერვალურ ვარიაციულ მწკრივს სოციოლოგიაში **უწყვეტად** მიიჩნევენ.

ვარიაციულ მწკრივს, რომელიც ინტერვალებად არ არის დაჯგუფებული, **დისკრეტული ვარიაციული მწკრივი** ეწოდება.

ინტერვალებად დაჯგუფება შეიძლება იყოს ფორმალური ან არაფორმალური ხასიათის.

არაფორმალური დაჯგუფების დროს ცვლადის მნიშვნელობების განაწილება ინტერვალებში (ინტერვალების განსაზღვრა) ხდება სოციოლოგის მიერ გარკვეული შინაარსობრივი მოსაზრებებით (მაგალითად, თუ მას აქვს არგუმენტები იმის სასარგებლოდ, რომ საკვლევი საკითხისათვის არაარსებითია ასაკობრივი განსხვავება 18-დან 22 წლის ჩათვლით, 23-დან 27 წლის ჩათვლით და ა. შ.).

ფორმალური დაჯგუფების დროს ინტერვალები ფორმალური წესების (კრიტერიუმების) მიხედვით განისაზღვრება და სოციოლოგის მოსაზრებებზე აღარ არის დამოკიდებული.

როდესაც დისკრეტულ ვარიაციულ მწკრივში ცვლადის გარკვეულ მნიშვნელობას/მნიშვნელობებს ნულოვანი სიხშირე შეესაბამება, ფორმალური დაჯგუფება საშუალებას გვაძლევს, რომ იგი უწყვეტი მწკრივის სახით წარმოვადგინოთ.

ინტერვალებად ფორმალური დაჯგუფების ძირითადი კრიტერიუმები:

1. **ინტერვალების რიცხვი** განისაზღვრება სტატისტიკური ერთობლიობის სიდიდის შესაბამისად (სტატისტიკური თვალსაზრისით ინტერვალების ოპტიმალური რაოდენობაა 10-15);
2. **მნიშვნელობათა დიაპაზონი** განისაზღვრება ცვლადის მაქსიმალურ და მინიმალურ მნიშვნელობებს შორის სხვაობით;
3. **ინტერვალის დიაპაზონი** განისაზღვრება მნიშვნელობათა დიაპაზონის შეფარდებით ინტერვალების რაოდენობასთან;
4. **ინტერვალების საზღვრები** ზუსტდება ისე, რომ ყოველი მნიშვნელობა შევიდეს ერთ და მხოლოდ ერთ ინტერვალში;
5. ხდება თითოეული ჯგუფის ცენტრის განსაზღვრა: **ჯგუფის ცენტრი** გამოითვლება მისი ზუსტი ქვედა და ზედა საზღვრების საშუალო არითმეტიკულით.

❖ **ემპირიული მონაცემების გრაფიკული წარმოდგენა**

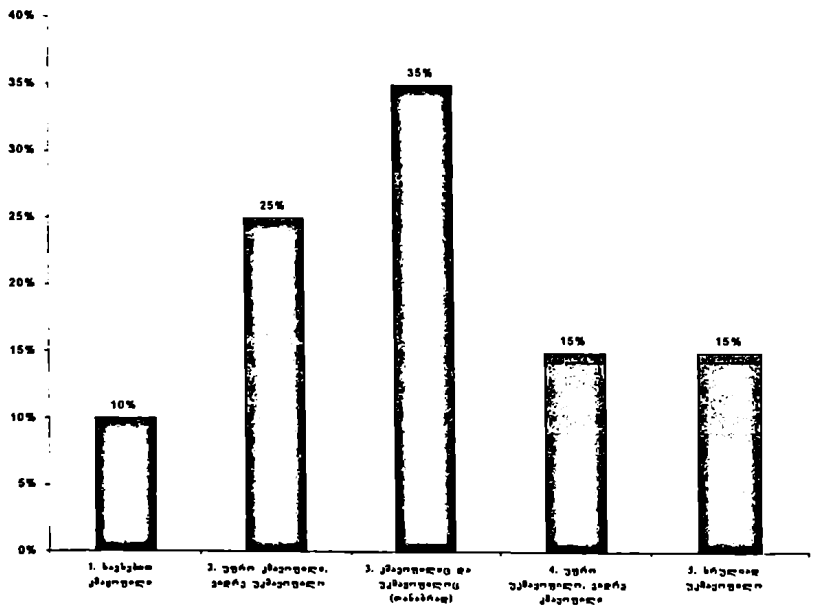
ცვლადის მნიშვნელობათა განაწილება/ვარიაციული მწკრივი ხშირად ვიზუალური, გრაფიკული სახითაა ნარმოდგენილი.

განაწილების გრაფიკულ სახეს **ჰისტოგრამას** უწოდებენ.

ჰისტოგრამა თვალსაჩინოს ხდის სტატისტიკური ერთობლიობის სტრუქტურას, **აიოლებს მონაცემთა აღქმას, პრეზენტაციასა და მათ ანალიზს.** ჰისტოგრამების აგება გარკვეული ნესების მიხედვით ხდება.

განაწილების დიაგრამა. განვიხილოთ განაწილების დიაგრამა ***ცხრილი №4***-ში მოცემული განაწილების მიხედვით:

გრაფიკი №1.



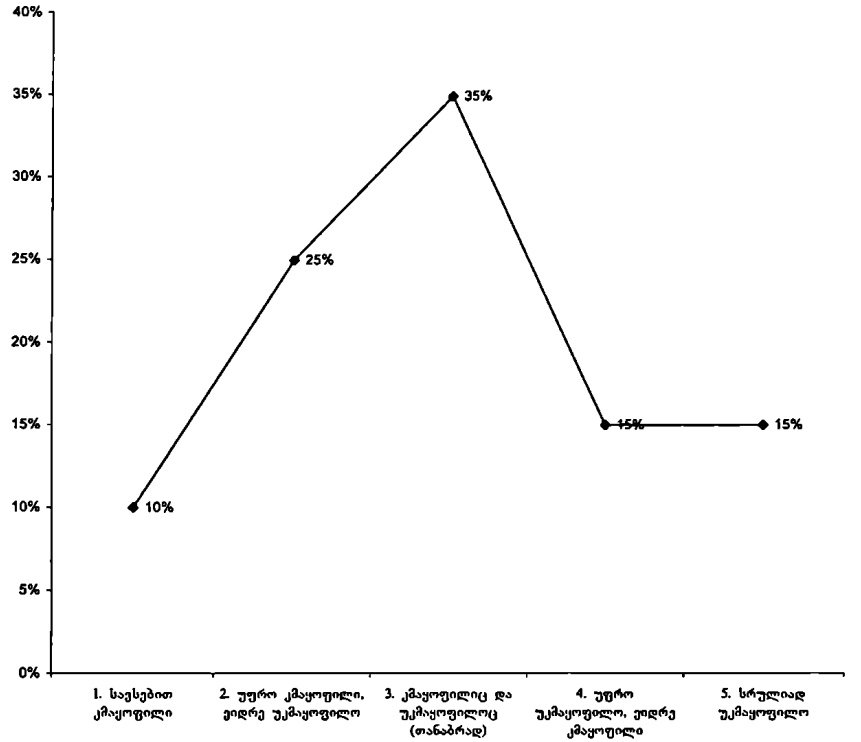
გრაფიკი №1-ის აბსცისაზე ცვლადის მნიშვნელობები, ხოლო ორდინატაზე მათი ფარდობითი სიხშირეებია გამოსახული. ცვლადის ყოველი მნიშვნელობისათვის აგებულია მართკუთხედი, რომლის სიმაღლეც ამ მნიშვნელობის ფარდობითი სიხშირის ტოლია.

განაწილების დიაგრამაზე ერთდროულად სხვადასხვა ვარიაციული მწკრივი შეიძლება ნარმოვადგინოთ. მაგალითად, ერთი ცვლადის მნიშვნელობების განაწილება სხვადასხვა ერთობლიობაში. მხოლოდ ერთი ვარიაციული მწკრივის ნარმოდგენისას მართკუთხედების ფართობების ჯამი უდრის 100%-ს.

განაწილების დიაგრამის გამოყენება აიოლებს ვარიაციული მწკრივის მნიშვნელობათა ფარდობითი **სიხშირეების აღწერასა და ურთიერთშედარებას.**

განაწილების პოლიგონი. იგივე მონაცემებისათვის შეიძლება ავაგოთ განაწილების მრუდი — განაწილების პოლიგონი:

გრაფიკი №2

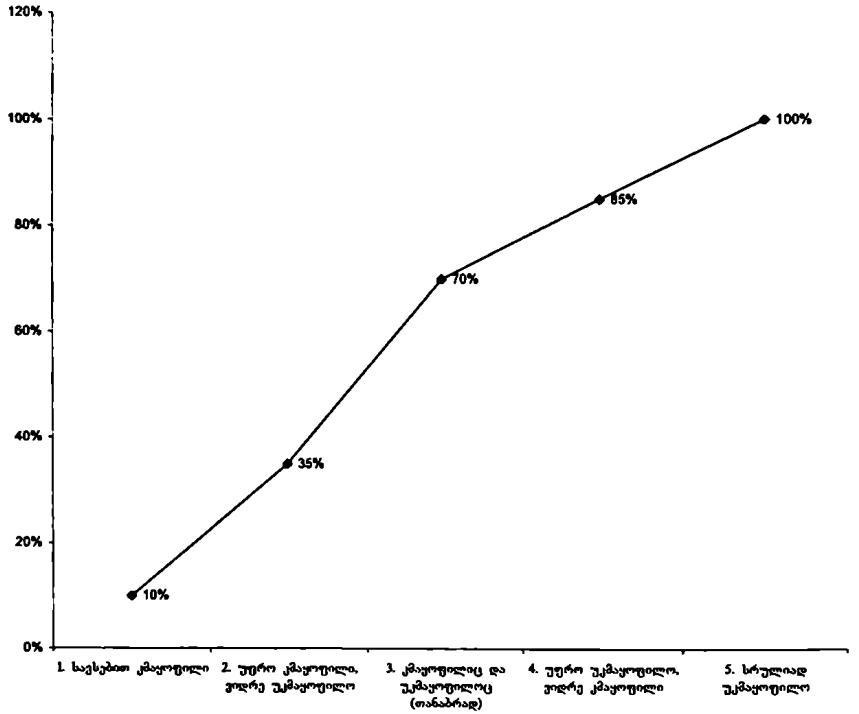


გრაფიკი №2-ის აბსცისაზე ცვლადის მნიშვნელობები, ხოლო ორდინატაზე მათი **ფარდობითი სიხშირეებია** გამოსახული (განაწილების მრუდის მიღება შეიძლება განაწილების დიაგრამიდან მართკუთხედების ზედა გვერდების შუანერტილების შეერთებით).

განაწილების პოლიგონი განაწილებას **ერთიანი, უწყვეტი პროცესის** სახით გამოხატავს და მის **მთლიანობაში წარმოდგენის** საშუალებას გვაძლევს.

კუმულიატა. რიგის და მეტრული სკალებისათვის ვარიაციული მწკრივი კუმულიატას/კუმულიატური მრუდის სახითაც შეიძლება წარმოვადგინოთ:

გრაფიკი №3



გრაფიკი №3-ის აბსცისაზე ცვლადის მნიშვნელობები, ორდინატაზე კი დაგროვილი სიხშირეებია გამოსახული.

კუმულიატის მეშვეობით თვალსაჩინო ხდება სიხშირეთა დაგროვების/აკუმულირების პროცესის სპეციფიკა.

❖ ცენტრალური/საშუალო ტენდენციებისა და ვარიაციის მაჩვენებლები

მონაცემთა სისტემატიზებული და შეკუმშული (დაჯგუფებული) სახით წარმოდგენა უკვე გვაძლევს იმის შესაძლებლობას, რომ მთლიანობაში აღვნეროთ თითოეული მასივის სტრუქტურა, ვიმსჯელოთ მის შესახებ ზოგადი რაოდენობრივი მაჩვენებლების/მახასიათებლების მიხედვით.

ცვლადის მნიშვნელობათა განაწილების ზოგადი სტატისტიკური მაჩვენებლები ორ ჯგუფად შეიძლება დავყოთ: **ცენტრალური ტენდენციებისა და ვარიაციის მაჩვენებლებად.**

ცენტრალური ტენდენციების მაჩვენებლები/ზომები (measures of location):
საშუალო არითმეტიკული, მოდა, მედიანა.

ცენტრალური ტენდენციების მაჩვენებლები: საშუალო არითმეტიკული, მოდა, მედიანა ახასიათებს ცვლადის მნიშვნელობათა განაწილების ცენტრალურ ტენდენციას. მათი მეშვეობით შესაძლებელია განვსაზღვროთ **განაწილების ცენტრი**, რომლის **ირგვლივაც კონცენტრირდება ცვლადის ემპირიული მნიშვნელობები**.

თუ ცვლადის თითოეულ ემპირიულ მნიშვნელობას ერთი და იგივე სიდიდით შევცვლით (დავუშვათ, მივუმატებთ n-ს), მაშინ ცენტრალური ცენტენდენციების მაჩვენებლებიც იგივე სიდიდით შეიცვლება (მოცემული მაგალითის შემთხვევაში, გაიზრდება n-ით).

საშუალო არითმეტიკული (mean). ცვლადის მნიშვნელობათა ჯამის შეფარდებას მათ რაოდენობასთან, **საშუალო არითმეტიკული** ეწოდება.

საშუალო არითმეტიკულს ზოგჯერ **შერჩევით საშუალოს, საშუალო მნიშვნელობას ან პირდაპირ, საშუალოს** უწოდებენ.

X ცვლადის საშუალო არითმეტიკული ($X_{საშ}$) გამოითვლება ფორმულით:

$$X_{საშ} = \frac{\sum_{i=1}^{n} x_i}{n}$$

სადაც x_i ცვლადის i-ური მნიშვნელობაა, ხოლო n ცვლადის მნიშვნელობათა საერთო რაოდენობა.

სოციოლოგიურ ანალიზებში ხშირად იყენებენ სიდიდეებს, რომლებსაც "საშუალო ასაკს", "საშუალო შემოსავალს", "დანახარჯების საშუალო მაჩვენებელს" და ა. შ. უწოდებენ. ეს სიდიდეები შესატყვისი ცვლადების მნიშვნელობათა საშუალო არითმეტიკულის გამოთვლით მიიღება. შესაბამისად, ცვლადის საშუალო მნიშვნელობა წარმოებული სიდიდე, **თეორიული** მნიშვნელობაა და შეიძლება, რომ ცვლადის არც ერთ ემპირიულ მნიშვნელობას არ ემთხვეოდეს.

საშუალო არითმეტიკულის გამოთვლა, თავისთავად, რთული არაა, მაგრამ რამდენად თავისუფალია სოციოლოგი ამ მარტივი ფორმულის გამოყენებისას?

ანალიზის დროს საშუალო არითმეტიკულის გამოყენებას აზრი მაშინ აქვს, როდესაც ჯგუფი ერთგვაროვანია. ჯგუფის რაოდენობრივი ერთგვაროვნების აღწერა ცენტრალური ტენდენციებისა და ვარიაციების მახასიათებლების შედარებითი ანალიზის საფუძველზე ხდება.

მაგრამ, სანამ რაოდენობრივ ერთგვაროვნებას შეეხება, სოციოლოგმა საკვლევი ჯგუფის თვისებრივი ერთგვაროვნება უნდა განსაზღვროს. თვისებრივად არაერთგვაროვანი ჯგუფების მიმართ საშუალო მნიშვნელობები ნაკლებად ინფორმატიულია.

მაგალითად, შემოსავლების საშუალო მნიშვნელობა ბევრს არაფერს გვეუბნება, როდესაც რესპონდენტების სამსახურებრივი სტატუსი, შესაბამისად ანაზღაურება მკვეთრად განსხვავდება ერთმანეთისაგან ან კვლევა მთელი ქვეყანის მასშტაბით ტარდება, მაგრამ შეიძლება ეფექტური აღმოჩნდეს, თუ მას ერთგვაროვანი სამსახურებრივი სტატუსის მქონე პირების ანდა ერთ-ერთი სოციალური სტრატის შესაბამის ჯგუფში გამოვთვლით.

საშუალო არითმეტიკულის გამოყენება შეზღუდულია გაზომვის დონითაც:

საშუალო არითმეტიკული უპრობლემოდ ინტერპრეტირდება გაზომვის მეტრული დონისათვის.

გაზომვის ნომინალურ დონეზე საშუალო მნიშვნელობა, როგორც წესი, ინტერპრეტაციას არ ექვემდებარება.

გაზომვის საშუალო (რიგის) დონეზე საშუალო მნიშვნელობის გამოთვლა იმ დაშვებით

ხდება, რომ რიგის სკალის ინტერვალები ერთგვაროვანია. ეს დაშვება ანალიზისა და ინტერპრეტაციის დროს მხედველობიდან არ უნდა გამოგვრჩეს[1].

მოდა (mode). ამ ცნებას სოციოლოგიაში თითქმის იგივე შინაარსი აქვს, რაც ყოველდღიურ ცხოვრებაში: **მოდა/მოდური** ეწოდება ცვლადის იმ მნიშვნელობას, რომელიც მოცემულ განანილებაში ყველაზე უფრო ხშირად გვხვდება (რომლის სიხშირეც ყველაზე მეტია).

ცხრილი №4-ის მიხედვით, მოდური მნიშვნელობაა "კმაყოფილიც და უკმაყოფილოც (თანაბრად)", რადგან ამ მნიშვნელობის სიხშირე (35%) ყველა დანარჩენს აღემატება.

ცვლადს შეიძლება ერთზე მეტი მოდური მნიშვნელობა ჰქონდეს. მაგალითად, განაწილებაში:

ცხრილი №5

ცვლადის მნიშვნელობა	ფარდობითი სიხშირე
1	30%
2	30%
3	25%
4	15%

მოდური მნიშვნელობებია 1 და 2.

გაზომვის რაოდენობრივ დონეზე სკალა (შესაბამისად, ცვლადის მნიშვნელობები) ხშირ შემთხვევაში იმდენად დიფერენცირებულია, რომ **დაბალი** (მცირე სიდიდის) და **ერთგვაროვანი** სიხშირეების გამო, მოდური მნიშვნელობის განხილვა აზრს კარგავს.

მაგალითად, შემოსავლების რაოდენობრივი სკალით გაზომვისას შეიძლება შემდეგნაირი განანილება მივიღოთ: 140 ლარი — 7%, 150 ლ. — 8%, 160 ლ. — 5%, 170 ლ. — 6% და ა. შ. ასეთ შემთხვევაში ყველაზე ხშირი მნიშვნელობა ვერ გამოხატავს ცენტრალურ ტენდენციას, არ იძლევა სხვა მნიშვნელობებთან შედარებითი ანალიზის ჩატარების საშუალებას, ამიტომ მიზანშეწონილია მონაცემთა ინტერვალებად დაჯგუფება და მოდური ინტერვალის განსაზღვრა.

მედიანა (median). ცვლადის იმ მნიშვნელობას, რომელიც რანჟირებული მწკრივის შუაშია, **მედიანა** ეწოდება.

შესაბამისად, ცვლადის მნიშვნელობათა ნახევარი მედიანის მარცხნივაა განლაგებული, ნახევარი — მარჯვნივ.

მედიანა გამოითვლება ფორმულით:

$$Me = l + \frac{\frac{N}{2} - F_b}{f_p} \times i$$

სადაც **l** მედიანური ინტერვალის ქვედა საზღვარია, F_b — სიხშირეების ჯამი **მედიანურ ინტერვალამდე** (მედიანურ ინტერვალამდე დაგროვილი სიხშირე), **N** — ცვლადის მნიშვნელობათა რაოდენობა, f_p — მედიანური ინტერვალის სიხშირე, ხოლო **i** — მედიანური ინტერვალის დიაპაზონი.

[1] რიგის სკალით მიღებული მონაცემების საშუალო მნიშვნელობების დათვლის დროს, სწორედ იმ შემთხვევასთან გვაქვს საქმე, როდესაც სკალის პრაქტიკული გამოყენება მის თეორიულ ჩარჩოებს არ ემთხვევა.

მედიანური ინტერვალი ეწოდება ინტერვალს, რომელშიც შედის მედიანური მნიშვნელობა. მედიანური ინტერვალის პოვნა შეგვიძლია დაგროვილი სიხშირეების მიხედვით:

რადგან მედიანა რანჟირებული მწკრივის შუაშია, ამიტომ იგი იმ ინტერვალში შედის, რომლის შესაბამისი დაგროვილი სიხშირე მეტია ან ტოლი 50%-ზე (მოიცავს მონაცემთა ნახევარს) და მასზე წინ არ არსებობს ინტერვალი, რომელიც იგივე პირობას აკმაყოფილებს.

დავუშვათ, პირველი ინტერვალის შესაბამისი დაგროვილი სიხშირეა 5%, მეორე ინტერვალის — 19%, მესამე ინტერვალის — 32%, მეოთხე ინტერვალის — 59%, მეხუთე ინტერვალის — 68% და ა. შ. მედიანური ინტერვალი იქნება მეოთხე, რადგან სწორედ ამ ინტერვალში გროვდება **პირველად** ცვლადის მნიშვნელობების 50%.

ცვლადის მნიშვნელობების 50% ამის შემდეგ უკვე ყოველი მომდევნო ინტერვალის შესაბამის დაგროვილ სიხშირეში შევა, მაგრამ, ცხადია, რომ ისინი მედიანურ ინტერვალს არ წარმოადგენენ.

მედიანის გამოყენება ოპტიმალურია გაზომვის ინტერვალურ ან შეფარდების დონეზე.

მედიანა, ისევე, როგორც საშუალო არითმეტიკული, წარმოებული სიდიდეა და შეიძლება, რომ ცვლადის არც ერთ ემპირიულ მნიშვნელობას არ ემთხვეოდეს.

ვარიაციის მაჩვენებლები:

ვარიაციის დიაპაზონი. დისპერსია. საშუალო კვადრატული გადახრა/სტანდარტული გადახრა, საშუალო აბსოლუტური გადახრა, ვარიაციის კოეფიციენტები.

ვარიაციის მაჩვენებლები გამოხატავს ცვლადის მნიშვნელობათა გაფანტვის (ცვალებადობის, ვარიაბელურობის) ხარისხს/ზომას.

ვარიაციის დიაპაზონი/გაფანტვის არე (range) ეწოდება ცვლადის მნიშვნელობათა გაფანტვის/ცვლილების დიაპაზონის ზომას. იგი ცვლადის უდიდეს/მაქსიმალურ და უმცირეს/მინიმალურ ემპირიულ მნიშვნელობებს შორის სხვაობით გამოითვლება:

$$\text{ვარიაციის დიაპაზონი} = X_{\text{მაქსიმალური}} - X_{\text{მინიმალური}}$$

თუ ცვლადის თითოეულ ემპირიულ მნიშვნელობას მუდმივაზე/კონსტანტაზე გავამრავლებთ/გავყოფთ, მაშინ ვარიაციის დიაპაზონიც იგივენაირად/იმდენჯერვე შეიცვლება.

დისპერსია (variance). დისპერსია ცვლადის თითოეულ ემპირიულ მნიშვნელობასა და საშუალო მნიშვნელობას შორის სხვაობის კვადრატების საშუალო არითმეტიკულს ეწოდება:

$$D = \frac{\sum_{i=1}^{n}(x_i - x_{საშ})^2}{n}$$

სადაც x_i **x** ცვლადის **i**-ური მნიშვნელობაა, $x_{საშ}$ — საშუალო მნიშვნელობა, ხოლო **n** — მნიშვნელობათა საერთო რაოდენობა.

დისპერსიის გამოთვლის დროს ცვლადის საშუალო და ემპირიულ მნიშვნელობებს შორის სხვაობა კვადრატში იმიტომ ადის, რომ ემპირიული მნიშვნელობა საშუალოდან შეიძლება გადახრილის იყოს მარცხნივ ან მარჯვნივ.

შესაბამისად, სხვაობა შეიძლება მივიღოთ უარყოფითი ან დადებითი ნიშნით. განსხვავებულნიშნიანი სიდიდეების შეკრება (საშუალო არითმეტიკულის გამოთვლის დროს) მათ გაბათილებას გამოიწვევს და არარეალურ რეზულტატებამდე მიგვიყვანს.

მაგალითად, თუ $x_{საშ} = 90$, $x_k = 70$, ხოლო $x_m = 120$, $x_k - x_{საშ}$ უდრის: 70 — 90 = —20, ხოლო $x_m - x_{საშ}$ უდრის: 120 — 90 = 30. (x_k და x_m გამოსახავს X ცვლადის ემპირიულ მნიშვნელობებს)

x_k მნიშვნელობა საშუალოდან 20 ერთეულით მარცხნივაა გადახრილი, x_m კი 30 ერთეულით

მარჯვნივ. სხვადასხვანიშნიანი რიცხვების შეკრებით გადახრის ერთეულების ჯამი (50) კი არ დაფიქსირდება, არამედ განსხვავებულნიშნიანი ერთეულები გაბათილდება და მივიღებთ —20 + 30 = 10. რაც არაადეკვატური მაჩვენებელია. სხვაობების კვადრატში აყვანა ასეთ არაადეკვატურობას თავიდან აგვაცილებს.

შერჩევით ანუ უშუალოდ კვლევაში ჩართულ ერთობლიობაზე განსაზღვრული დისპერსია ნანაცვლებულია/განსხვავდება გენერალურ ერთობლიობაზე (მთელ საკვლევ ერთობლიობაზე) განსაზღვრული დისპერსიისაგან (σ^2გენ):

$$\sigma^2\text{გენ} = \frac{n \times M(D)}{n - 1}$$

ამ განსხვავების თავიდან აცილების მიზნით, ხშირად დისპერსიას შემდეგი ფორმულით ითვლიან:

$$D_0 = \frac{\sum_{i=1}^{n}(x_i - x_{i-1})^2}{n - 1}$$

D_0-ს ნაუნაცვლებელი დისპერსია ეწოდება.

ნაუნაცვლებელი დისპერსიის გამოყენება მნიშვნელოვანია მონაცემთა მცირე მასივების შემთხვევაში. მაგრამ, როდესაც შერჩევითი ერთობლიობა: n დიდი რიცხვია:

$$\frac{n - 1}{n} \to 1$$

$n - 1$ შეფარდებული n-თან 1-ისაკენ მიისწრაფის, ამიტომ რაოდენობრივი სოციოლოგიური კვლევების დროს, სადაც მონაცემთა დიდ მასივებთან გვაქვს საქმე, ხშირ შემთხვევაში, განსხვავება ნანაცვლებულ და ნაუნაცვლებელ დისპერსიებს შორის არსებითი არ არის.

საშუალო კვადრატული გადახრა/სტანდარტული გადახრა (standard deviation). კვადრატულ ფესვს დისპერსიიდან საშუალო კვადრატული გადახრა (S_x) ეწოდება. იგი გამოითვლება ფორმულით:

$$S_x = \sqrt{D}$$

საშუალო კვადრატული გადახრა ცვლადის საშუალო მნიშვნელობის მიმართ ემპირიული მნიშვნელობების **ვარიაციის** (გაფანტვის, განსხვავების, გადახრის, ცვლილების ხარისხის) **საშუალო მაჩვენებელია.**

გარდა ამისა, იგი საშუალებას გვაძლევს ერთმანეთს შევადაროთ **სხვადასხვა ცვლადის მნიშვნელობათა ვარიაციის ხარისხები** ან **ერთი ცვლადის მნიშვნელობების ვარიაციის ხარისხები სხვადასხვა ჯგუფისათვის.**

ვარიაციის კოეფიციენტი. სხვადასხვა ცვლადის/ჯგუფის ვარიაციის ხარისხების შედარება საშუალო კვადრატული გადახრის მიხედვით, პრობლემურია. მაგალითად, როდესაც:

- საშუალო მნიშვნელობებიც და საშუალო კვადრატული გადახრებიც ერთმანეთისაგან განსხვავებულია;
- საშუალო კვადრატული გადახრები ერთნაირია, მაგრამ განსხვავებულია საშუალო მნიშვნელობები.

ასეთ შემთხვევაში შედარებისათვის ვარიაციის კოეფიციენტს იყენებენ.

ვარიაციის კოეფიციენტი (coefficient of variation) ეწოდება ვარიაციის ხვედრით წილს, რომელიც საშუალო მნიშვნელობის ერთეულზე მოდის და გამოითვლება ფორმულით:

$$V_{S_x} = \frac{S_x}{X_{საშ}} \times 100\%$$

ცხადია, როდესაც საშუალო მნიშვნელობის ერთეულზე მოსული ვარიაციის ხვედრითი წილია ცნობილი, საშუალო მნიშვნელობებს ან საშუალო კვადრატულ მნიშვნელობებს შორის არსებული განსხვავებები შედარებას ხელს აღარ შეუშლის.

საშუალო აბსოლუტური გადახრა. საშუალო მნიშვნელობიდან ცვლადის მნიშვნელობების გადახრათა საშუალო მნიშვნელობის გამოთვლისას ნიშნის (+, —) სირთულე სხვაობების კვადრატებში აყვანის გარდა, სხვა გზით — გადახრის აბსოლუტური მნიშვნელობების (მოდულების) შეკრებითაც შეიძლება დაიძლიოს.

ვარიაციის ხარისხის ამ წესით გამოთვლილ მაჩვენებლს ($d_{საშ}$) საშუალო აბსოლუტურ გადახრას უწოდებენ. იგი გამოითვლება ფორმულით:

$$d_{საშ} = \frac{\sum_{i=1}^{n} |x_i - x_{საშ}|}{n}$$

საშუალო აბსოლუტური გადახრის კოეფიციენტი (V_d) გამოითვლება ფორმულით:

$$V_d = \frac{d_{საშ}}{X_{საშ}} \times 100\%$$

საშუალო აბსოლუტური გადახრისა და შესაბამისი ვარიაციის კოეფიციენტის ფუნქციები იგივეა, რაც საშუალო კვადრატული გადახრის და მისი შესაბამისი კოეფიციენტის. განსხვავებულია მხოლოდ მათი გამოთვლის წესები. უფრო ინტენსიურად საშუალო კვადრატული გადახრით სარგებლობენ, რადგან იგი გამოიყენება სტატისტიკური ანალიზის რთულ ეტაპებზე.

რაც მცირეა ვარიაციის კოეფიციენტი, მით უფრო ერთგვაროვანია ჯგუფი.

ვარიაციის მაჩვენებლები გაზომვის ინტერვალურ და შეფარდების დონეზე გამოიყენება.

❖ სტანდარტული მნიშვნელობები

თუ რაოდენობრივი ცვლადის მნიშვნელობებს სტანდარტული გადახრის ერთეულებში გამოვხატავთ, შედეგად ამ ცვლადის **სტანდარტულ მნიშვნელობებს** მივიღებთ.

ცვლადის სტანდარტულ მნიშვნელობებს **Z-მნიშვნელობებს** უწოდებენ.

ცვლადის მნიშვნელობების მიხედვით, Z-მნიშვნელობის გამოთვლა შემდეგი წესით ხდება:

$$Z_i = \frac{X_i - X_{საშ}}{S_x}$$

შესაბამისად, Z-მნიშვნელობის მიხედვით ცვლადის მნიშვნელობის პოვნა შეიძლება ფორმულით:

$$X_i = X_{საშ} + Z_i \times S_x$$

თუ ცვლადის მნიშვნელობა საშუალოდან მარჯვნივაა გადახრილი (საშუალოზე მეტია), მაშინ მისი Z-მნიშვნელობა დადებითია, ხოლო თუ ცვლადის მნიშვნელობა საშუალოდან მარცხნივაა გადახრილი (საშუალოზე ნაკლებია), Z-მნიშვნელობა უარყოფითია.

მაგალითი: თუ განაწილებაში საშუალო ყოველთვიური პირადი შემოსავალი არის 200 ლარი,

ხოლო სტანდარტული გადახრაა 25 ლარი, მაშინ, ცვლადის მნიშვნელობა X_k=250 ლარი სტანდარტული გადახრის ერთეულებში იქნება:

$$Z_k = (250-200)/25=2$$

ე. ი. ცვლადის მოცემული მნიშვნელობა (250 ლ.) საშუალოდან სტანდარტული გადახრის 2 ერთეულით მარჯვნივაა გადახრილი:

$$X_k = 200 + 25\times2 = 200 + 50 = 250.$$

ხოლო, იგივე განაწილებაში ცვლადის მნიშვნელობა X_m=150 ლ. შესაბამისი სტანდარტული გადახრა იქნება:

$$Z_m = (150-200)/25= -2.$$

ცვლადის მოცემული მნიშვნელობა (150 ლ.) საშუალოდან სტანდარტული გადახრის 2 ერთეულით მარცხნივაა გადახრილი:

$$X_m = 200 + 25\times(-2) = 200 - 50 = 150.$$

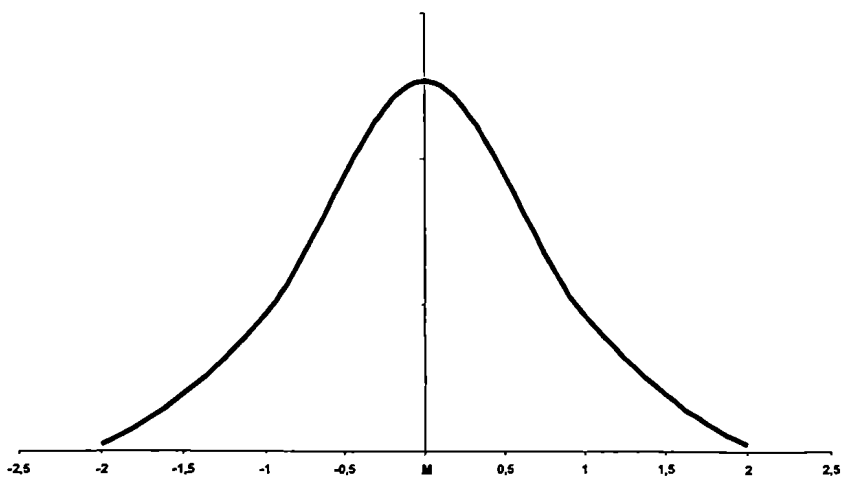
Z-მნიშვნელობები ინტენსიურად გამოიყენება სტატისტიკური ანალიზის რთულ მეთოდებში.

❖ განაწილების ფორმები
სიმეტრიული და ასიმეტრიული განაწილებები, ექსცესი

თუ ცვლადის მნიშვნელობათა განაწილებას საკოორდინატო სისტემაში გამოვსახავთ, მივიღებთ მრუდს, რომლის ფორმა შეიძლება იყოს **სიმეტრიული ან ასიმეტრიული.**

განაწილებას ეწოდება **სიმეტრიული,** თუ ცვლადის ნებისმიერი ორი მნიშვნელობა, რომლებიც განაწილების **ცენტრიდან** ერთი და იგივე მანძილითაა დაშორებული, ერთმანეთის ტოლია. ე. ი. **განაწილების მრუდი სიმეტრიულია განაწილების ცენტრის** მიმართ. გარდა ამისა, ერთმანეთს ემთხვევა ცენტრალური ტენდენციების მაჩვენებლები (საშუალო არითმეტიკული = მოდა = მედიანა).

სიმეტრიული განაწილების ფორმა "ზარისებრი" მრუდით გამოისახება:



სიმეტრიული განაწილების მაგალითია **ნორმალური** ანუ გაუსის განაწილება (**normal**

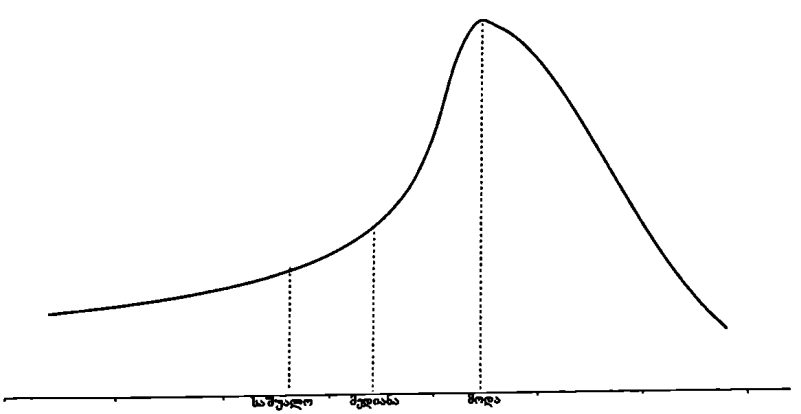
distribution). იგი ალბათობის თეორიის საფუძველია. მასზე დაყრდნობით მრავალი მნიშვნელოვანი სტატისტიკური მახასიათებელია გამოყვანილი. მაგალითად, სტანდარტული გადახრა, კორელაციის კოეფიციენტების განაწილების კანონები და ა. შ.

ემპირიული განაწილების მრუდის ფორმა (განსაკუთრებით, დიდი მოცულობის შერჩევითი ერთობლიობის შემთხვევებში) ხშირად უახლოვდება ჩვენს მიერ თეორიულად აღწერილ ნორმალური განაწილების მრუდის ფორმას.

შემთხვევითი სიდიდე **X**, რომელიც ნორმალური განაწილების კანონზომიერებას ექვემდებარება, მოთავსებულია უსასრულო ინტერვალში: $-\infty < x < +\infty$.

ცვლადის მნიშვნელობათა განაწილებას ეწოდება **ასიმეტრიული**, თუ საშუალო არითმეტიკულიდან თანაბრად დაშორებული მნიშვნელობები სხვადასხვა სიხშირითაა წარმოდგენილი.

ასიმეტრიული განაწილების ფორმა შემდეგნაირია:



ასიმეტრიული განაწილების ცენტრალური ტენდენციების მაჩვენებლები ერთმანეთს არ ემთხვევა

ექსცესი (kurtosis). გვიჩვენებს თუ რამდენად უფრო ციცაბოა/ბლაგვია (მახვილწვეროიანია/ბლაგვწვეროიანია) მოცემული ცვლადის განაწილების მრუდი ნორმალური განაწილების მრუდთან შედარებით.

ნორმალური განაწილების მრუდის ექსცესი ნულის ტოლადაა მიჩნეული.

თუ ექსცესი დადებითია, მაშინ მოცემული განაწილების მრუდი უფრო მახვილია, ვიდრე ნორმალური განაწილების მრუდი.

თუ ექსცესი უარყოფითია, მაშინ მოცემული განაწილების მრუდი უფრო ბლაგვია, ვიდრე ნორმალური განაწილების მრუდი.

❖ სტატისტიკური ჰიპოთეზების შემოწმება

შერჩევითი ერთობლიობის კვლევის შედეგად მიღებული მონაცემებიდან ხშირად არაერთმნიშვნელოვანი შედეგები გამომდინარეობს. გარდა ამისა, დაზღვეული არ ვართ შემთხვევითი შეცდომებისაგან/გადახრებისაგან. ამიტომ, დასკვნები, რომლებიც გენერალური ერთობლიობის შესახებ შერჩევითი ერთობლიობის კვლევის მონაცემებიდან გამოგვყავს, არა

კატეგორიული, არამედ ალბათური ხასიათისაა. ისინი **სავარაუდო** დებულებებია.

სავარაუდო დებულებებს, რომელებშიც გამოთქმულია მოსაზრება გენერალური ერთობლიობის ნიშან-თვისებების, მისი სტატისტიკური პარამეტრების განაწილებათა შესახებ, **სტატისტიკური ჰიპოთეზები** ეწოდება.

ბუნებრივია, რომ სტატისტიკური ჰიპოთეზა, ისევე როგორც ნებისმიერი სხვა სავარაუდო დებულება, შემოწმებას საჭიროებს.

სტატისტიკური ჰიპოთეზის შემოწმება ნიშნავს იმის შეფასებას, თუ **სტატისტიკურად რამდენად მნიშვნელოვანია განსხვავება ჰიპოთეზასა და შერჩევითი ერთობლიობის მონაცემებს შორის.**

როდესაც სტატისტიკურ ჰიპოთეზასა და შერჩევითი ერთობლიობის მონაცემებს შორის არსებული განსხვავება ფასდება როგორც სტატისტიკურად უმნიშვნელო, ითვლება, რომ ეს განსხვავება **შემთხვევითი გადახრების შედეგია** ე. ი. სტატისტიკური ჰიპოთეზა ვერიფიცირდება და ხდება მისი მიღება.

როდესაც სტატისტიკურ ჰიპოთეზასა და შერჩევითი ერთობლიობის მონაცემებს შორის არსებული განსხვავება ფასდება როგორც სტატისტიკურად მნიშვნელოვანი, ითვლება, რომ ეს განსხვავება **კანონზომიერია** ე. ი. სტატისტიკური ჰიპოთეზა ფალსიფიცირდება და ხდება მისი უკუგდება.

სტატისტიკური ჰიპოთეზის შემოწმების ზოგადი სქემა შემდეგნაირია:

- ნულოვანი (H_0) და ალტერნატიული (H_A) ჰიპოთეზების ფორმულირება;
- ჰიპოთეზის სტატისტიკური შემოწმების მეთოდის: სტატისტიკური კრიტერიუმის/სტატისტიკის შერჩევა;
- მნიშვნელოვნების დონის (α) შერჩევა;
- სტატისტიკური კრიტერიუმის/სტატისტიკის შერჩევითი (შეფასებითი) მნიშვნელობის გამოთვლა;
- სტატისტიკური კრიტერიუმის/სტატისტიკის კრიტიკული მნიშვნელობის განსაზღვრა;
- სტატისტიკური კრიტერიუმის/სტატისტიკის შერჩევითი და კრიტიკული მნიშვნელობების შედარება;
- გადაწყვეტილების მიღება ნულოვანი ჰიპოთეზის ვერიფიცირების/მიღების ან ფალსიფიცირების/უკუგდების შესახებ.

როგორ წესი, ნულოვან ჰიპოთეზაში (**null hypothesis**) ფორმულირებულია მოსაზრება იმის შესახებ, რომ გენერალური ერთობლიობის გარკვეულ პარამეტრებს შორის არ არსებობს კავშირი/დამოკიდებულება.

ალტერნატიული ჰიპოთეზა (**alternative hypothesis**) ნულოვანი ჰიპოთეზის გამომრიცხავ დებულებას ეწოდება. ე. ი. მასში ფორმულირებულია მოსაზრება იმის შესახებ, რომ გენერალური ერთობლიობის მოცემულ პარამეტრებს შორის არსებობს კავშირი/დამოკიდებულება.

შემოწმებას ნულოვანი ჰიპოთეზა ექვემდებარება. მისი ვერიფიცირების/მიღების შემთხვევაში ხდება მასში ფორმულირებული მოსაზრების გაზიარება, ხოლო ფალსიფიცირების/უკუგდების შემთხვევაში — ალტერნატიულის.

სტატისტიკური ჰიპოთეზის შესამოწმებლად უნდა ავირჩიოთ ადეკვატური სტატისტიკური მეთოდი — სტატისტიკური კრიტერიუმი/სტატისტიკა.

სტატისტიკური კრიტერიუმი (**test statistic**) შემთხვევითი სიდიდეა. მისი განაწილება ხშირად ექვემდებარება "ცნობილ"/გავრცელებულ განაწილებებს (მაგალითად, ნორმალური განაწილება, სტიუდენტის განაწილება/t-განაწილება, ხი-კვადრატ განაწილება).

სტატისტიკური კრიტერიუმის განაწილების კრიტიკული მნიშვნელობა გვიჩვენებს ზღვარს, რომელსაც, მოცემულ პირობებში, **არ უნდა გადააჭარბოს შერჩევითმა მნიშვნელობამ, თუ ნულოვანი ჰიპოთეზა ჭეშმარიტია.**

ნულოვანი ჰიპოთეზა როგორც ნებისმიერი დებულება, **ჭეშმარიტია ან მცდარი.** ამიტომ,

მისი შემოწმების დროს არსებობს ორი ტიპის შეცდომის დაშვების საფრთხე:

პირველი ტიპის შეცდომა (**Type I error**) შეიძლება დავუშვათ იმ შემთხვევაში, როდესაც **ნულოვანი ჰიპოთეზა ჭეშმარიტია, მაგრამ ჩვენ მას უარვყოფთ** (ჩავთვლით მცდარად). ასეთი შეცდომის დაშვების ალბათობას **მნიშვნელოვნების დონე (level of significance)** ეწოდება. მის აღსანიშნავად გამოიყენება სიმბოლო **α.**

როგორც პრაქტიკა გვიჩვენებს, ყველაზე ხშირია/ტიპურია **α = 0.05** მნიშვნელოვნების დონის შერჩევა.

მეორე ტიპის შეცდომას (**Type II error**) შეიძლება დავუშვათ იმ შემთხვევაში, როდესაც ნულოვანი ჰიპოთეზა მცდარია, მაგრამ ჩვენ მას არ უარვყოფთ (ჩავთვლით ჭეშმარიტად). ასეთი შეცდომის დაშვების ალბათობა აღინიშნება **β**-თი.

შესაბამისად, იმის ალბათობა, რომ მოხდება ნულოვანი ჰიპოთეზის უკუგდება მაშინ, როდესაც იგი რეალურად მცდარია (ანუ, მოხდება მისი კორექტული, "სამართლიანი" უკუგდება) განისაზღვრება სხვაობით: **(1 — β)**. ამ სიდიდეს **სტატისტიკური კრიტერიუმის სიმძლავრე (power of test)** ეწოდება.

α-ს შერჩევა ხდება მკვლევრის მიერ, **იმის მიხედვით, თუ რამდენად მნიშვნელოვნად შეიძლება ჩაითვალოს პირველი ტიპის შეცდომა.** ამასთან, მისი შერჩევისას, აუცილებელია გავითვალისწინოთ ისიც, **რომ პირველი ტიპის შეცდომის ალბათობის შემცირება ზრდის ნულოვანი ჰიპოთეზის მიღების შანსს.**

იმ შემთხვევაში, როდესაც ნულოვანი ჰიპოთეზა ჭეშმარიტია, ეს იწვევს ჭეშმარიტი დებულების მცდარად ჩათვლის ალბათობის შემცირებას, მაგრამ, იმ შემთხვევაში, როდესაც ნულოვანი ჰიპოთეზა მცდარია, იზრდება მცდარი დებულების ჭეშმარიტად ჩათვლის (β-შეცდომის დაშვების) ალბათობა.

შესაბამისად, რადგან ჩვენ არ ვიცით ნულოვანი ჰიპოთეზა ჭეშმარიტია თუ მცდარი, აუცილებელია ბალანსის დაცვა ორი ტიპის შესაძლო შეცდომის ალბათობებს შორის

სტატისტიკური კრიტერიუმის/სტატისტიკის შერჩევითი მნიშვნელობა განისაზღვრება ემპირიული მონაცემების მიხედვით.

სტატისტიკური კრიტერიუმის/სტატისტიკის კრიტიკული მნიშვნელობა შეიძლება მოვნახოთ სპეციალურ ცხრილებში.

სტატისტიკური კრიტერიუმის/სტატისტიკის შერჩევითი და კრიტიკული მნიშვნელობების შედარების საფუძველზე, მნიშვნელოვნების გარკვეულ, წინასწარშერჩეულ დონეზე ხდება გადაწყვეტილების მიღება იმის შესახებ, უნდა მივიღოთ, თუ უნდა უკუვაგდოთ სტატისტიკური ჰიპოთეზა:

თუ სტატისტიკური კრიტერიუმის შერჩევითი მნიშვნელობა ნაკლებია კრიტიკულზე, მაშინ ნულოვანი ჰიპოთეზა ვერიფიცირდება/უნდა მივიღოთ, ხოლო თუ მეტი — ფალსიფიცირდება/უნდა უკუვაგდოთ, შესაბამისად, უნდა მივიღოთ ალტერნატიული ჰიპოთეზა.

❖ საშუალოთა შედარება

სოციოლოგიურ ანალიზებში ხშირად გვხვდება დასკვნები, რომლებიც ცვლადის საშუალო მნიშვნელობათა შედარების საფუძველზეა გამოტანილი.

მაგალითად, შეიძლება ერთმანეთს შევადაროთ საკუთარი საქმიანობით კმაყოფილების საშუალო მაჩვენებლები ქალებსა და მამაკაცებს შორის, აკადემიური მოსწრების საშუალო მაჩვენებლები სხვადასხვა სპეციალობის სტუდენტთა შორის, ცხოვრების პირობებით კმაყოფილების ხარისხის საშუალო მაჩვენებლები სხვადასხვა ასაკობრივ ჯგუფს შორის და ა. შ.

მაგრამ, რა გარანტიაა, რომ **განსხვავება საშუალო მნიშვნელობებს შორის შემთხვევითი კი არაა, არამედ, გენერალური ერთობლიობისათვის დამახასიათებელ რეალურ სოციალურ**

ტენდენციას, გარკვეულ კანონზომიერებას გამოხატავს?

ამ კითხვაზე პასუხის გასაცემად, უნდა შევამოწმოთ, **სტატისტიკურად რამდენად მნიშვნელოვანია განსხვავება საშუალოებს შორის:**

თუ ცვლადის საშუალოებს შორის განსხვავება სტატისტიკურად მნიშვნელოვანია, მაშინ ითვლება, რომ იგი გენერალური ერთობლიობის თავისებურებას გამოხატავს, ხოლო თუ საშუალოებს შორის განსხვავება სტატისტიკურად უმნიშვნელოა, ითვლება, რომ იგი შემთხვევითი/არაარსებითია.

საშუალოებს შორის განსხვავების შეფასება ხდება სტატისტიკური ჰიპოთეზების შემოწმების მეთოდით:

ნულოვანი ჰიპოთეზა (H_0): გენერალური ერთობლიობის საშუალოები ერთმანეთის ტოლია/შერჩევით საშუალოებს შორის განსხვავება სტატისტიკურად მნიშვნელოვანი არ არის.

ალტერნატიული ჰიპოთეზა (H_1): გენერალური ერთობლიობის საშუალოები ერთმანეთისაგან განსხვავებულია/შერჩევით საშუალოებს შორის განსხვავება სტატისტიკურად მნიშვნელოვანია.

სტატისტიკური კრიტერიუმები:

I. დამოუკიდებელი შერჩევებისათვის:

გამოიყენება ორი დამოუკიდებელი შერჩევისათვის საშუალო მნიშვნელობებს შორის განსხვავების სტატისტიკური მნიშვნელოვნების შესაფასებლად.

ორ შერჩევას ეწოდება დამოუკიდებელი, თუ ისინი ერთმანეთზე არ არიან დამოკიდებული ექსპერიმენტულად ანუ, როდესაც **ერთი შერჩევის ფარგლებში განხორციელებული გაზომვები გავლენას არ ახდენს მეორე შერჩევის ფარგლებში განხორციელებულ გაზომვებზე.**

1. t-კრიტერიუმი დამოუკიდებელი შერჩევებისათვის **(Independent-Sample T Test)** გამოიყენება იმ შემთხვევაში, როდესაც ცნობილია, რომ გენერალურ ერთობლიობათა დისპერსიები ტოლია.

2. გარდაქმნილი/აპროქსიმირებული t-კრიტერიუმი დამოუკიდებელი შერჩევებისათვის გამოიყენება იმ შემთხვევაში, როდესაც ცნობილია, რომ გენერალურ ერთობლიობათა დისპერსიები განსხვავებულია.

იმ შემთხვევაში, როდესაც არ არის ცნობილი, განსხვავდება თუ არა გენერალურ ერთობლიობათა დისპერსიები ერთმანეთისაგან, ჯერ **F**-კრიტერიუმის (ფიშერის კრიტერიუმი/**F Test**) მეშვეობით განისაზღვრება ტოლია, თუ არა დისპერსიები და მიღებული შედეგების მიხედვით ხდება შესაბამისი t-კრიტერიუმის შერჩევა.

3. t-კრიტერიუმი დამოკიდებული შერჩევებისათვის **(Paired-Sample T Test)** გამოიყენება ერთი და იგივე შერჩევისათვის სხვადასხვა გაზომვის შედეგად მიღებულ საშუალო მნიშვნელობებს შორის განსხვავების სტატისტიკური მნიშვნელოვნების შესაფასებლად.

4. t-კრიტერიუმი ერთი შერჩევისათვის **(One-Sample T Test)** გამოიყენება ცვლადის საშუალო მნიშვნელობის შესადარებლად რაიმე ეტალონთან.

შეფასება ხდება მნიშვნელოვნების წინასწარშერჩეულ დონეზე. მაგალითად, თუ თუ t-კრიტერიუმის კრიტიკულ მნიშვნელობას შევარჩევთ α=0.05 დონის შესაბამისად და მივიღებთ, რომ:

I. t-კრიტერიუმის შერჩევითი/შეფასებითი მნიშვნელობა არ აღემატება კრიტიკულს, მაშინ ითვლება, რომ 5%-იანი ალბათობით, შერჩევით საშუალოებს შორის განსხვავება სტატისტიკურად უმნიშვნელოა და ნულოვანი ჰიპოთეზა ვერიფიცირდება/უნდა მივიღოთ;

II. t-კრიტერიუმის შერჩევითი/შეფასებითი მნიშვნელობა აღემატება კრიტიკულს, მაშინ ითვლება, რომ 5%-იანი ალბათობით, შერჩევით საშუალოებს შორის განსხვავება სტატისტიკურად მნიშვნელოვანია და ნულოვანი ჰიპოთეზა ფალსიფიცირდება/უნდა უკუვაგდოთ.

კორელაცია და რეგრესია

სოციალური მოვლენები განხოლოებული სახით არ არსებობს. ისინი რთული სოციალური სინამდვილის შემადგენელი ნაწილებია და, ბუნებრივია, ურთიერთქმედებენ, გარკვეულ დამოკიდებულებაში იმყოფებიან ერთმანეთთან.

ემპირიული სოციოლოგია სოციალურ მოვლენებს მათი **ნიშან-თვისებების/ცვლადების** შესწავლით იკვლევს.

ნიშან-თვისება/ცვლადი აბსტრაქციის შედეგად მიიღება. აბსტრაქცია კი სხვა ნიშან-თვისებებთან დამოკიდებულებებს არ ასახავს. ამიტომ, თუ თითოეული ცვლადის იზოლირებული შესწავლით შემოვიფარგლებით და არ გვეცოდინება რა გავლენას ახდენს ცვლადები ერთმანეთზე, არა მხოლოდ საკვლევი პრობლემის შესწავლას ვერ შევძლებთ, არამედ თვითონ ნიშან-თვისებებიც/ცვლადებიც, ფაქტობრივად, შეუსწავლელი დაგვრჩება.

სწორად ჩატარებული კვლევის პირობებში, **საკვლევ სოციალურ მოვლენათა შორის არსებული დამოკიდებულებები ცვლადებს შორის სტატისტიკურ დამოკიდებულებებში აისახება.**

ცვლადებს შორის სტატისტიკური დამოკიდებულების შესახებ ცვლადების **მნიშვნელობათა განაწილებების** მიხედვით ვმსჯელობთ.

მოცემული ცვლადის მნიშვნელობათა განაწილება შეიძლება ერთი ცვლადის მნიშვნელობათა განაწილებაზე დამოკიდებული იყოს, ხოლო მეორე ცვლადის მნიშვნელობათა განაწილებაზე — არა.

ამიტომ, ცვლადებს შორის დამოკიდებულებას ზოგადად ვერ მოვიაზრებთ: არ არსებობს "თავისთავად" დამოკიდებელი და დამოუკიდებელი ცვლადები.

ცვლადებს შორის დამოკიდებულება ყოველი მოცემული შემთხვევისათვის ინდივიდუალურად უნდა დახასიათდეს:

თუ ცვლადების **A** ჯგუფის: $A = \{A_1, A_2, \ldots a_n\}$ ერთი წევრი მაინც დამოკიდებულია **B** ჯგუფის: $B = \{B_1, B_2, \ldots, B_k\}$ ერთ წევრზე მაინც, მაშინ **A** ჯგუფს **B** ჯგუფზე **დამოკიდებული** ეწოდება (**n** და **k** სასრული ნატურალური რიცხვებია).

თუ ცვლადების **A** ჯგუფის ნებისმიერი წევრი დამოუკიდებელია **B** ჯგუფის ნებისმიერი წევრისაგან, მაშინ **A** ჯგუფს **B** ჯგუფისაგან **დამოუკიდებული** ეწოდება.

როდესაც $n=1$ და $k=1$, მივიღებთ დებულებებს ერთი ცვლადის მეორეზე დამოკიდებულების შესახებ:

თუ ერთი ცვლადის (A_1) მნიშვნელობათა განაწილება დამოკიდებულია მეორე ცვლადის (B_1) მნიშვნელობათა განაწილებაზე, მაშინ A_1-ს B_1-ზე **დამოკიდებული ცვლადი** ეწოდება.

თუ A_1 ცვლადის მნიშვნელობათა განაწილება არ არის დამოკიდებული B_1 ცვლადის მნიშვნელობათა განაწილებაზე, მაშინ A_1-ს B_1-საგან **დამოუკიდებული ცვლადი** ეწოდება.

შინაარსობრივი მოსაზრებით, დამოკიდებულ ცვლადს **რეზულტატურს** უწოდებენ, ხოლო დამოუკიდებელ ცვლადს ან ცვლადების ჯგუფს, რომელიც რეზულტატური ცვლადის განაწილებაზე ზემოქმედებს, **ფაქტორულს**[1].

ფაქტორული ცვლადი ინტერპრეტირდება როგორც რეზულტატური ცვლადის ვარიაციის **მიზეზი**[1].

რეზულტატური ცვლადი ინტერპრეტირდება როგორც ფაქტორული ცვლადის ან ცვლადების ზეგავლენის ობიექტი, რომლის ვარიაციაც ამ ზეგავლენის **შედეგი/რეზულტატია.**

შესაბამისად, ფაქტორულ და რეზულტატურ ცვლადებს შორის დამოკიდებულება ინტერპრეტირდება როგორც **დეტერმინაციული/მიზეზ-შედეგობრივი** ხასიათის დამოკიდებულება.

---

[1] **ინგ.: Factor — მამოძრავებელი, მამოქმედებელი ძალა. ფაქტორი ზოგადი ტერმინია და სხვადასხვა გაგებით შეიძლება შეგვხვდეს ყველა იმ ტექსტში, სადაც მოცემულია მსჯელობა მოვლენების/ცვლადების სხვა მოვლენაზე/ცვლადზე ზეგავლენის შესახებ.**

კორელაციური ანალიზი

❖ ორგანზომილებიანი/კორელაციური განაწილება

სტატისტიკურ დამოკიდებულებას ორი ცვლადის მნიშვნელობათა ემპირიულ განაწილებებს შორის **კორელაცია** ეწოდება.

კორელაციური ანალიზი იმ მეთოდების ერთობლიობაა, რომელთა მეშვეობითაც კორელაციის გამოვლენა და დახასიათება ხორციელდება.

განვიხილოთ ერთი და იგივე შერჩევითი ერთობლიობის ორი განაწილება:

ცხრილი №6

სტუდენტთა სპეციალობა	აბსოლუტური სიხშირე
1. სოციოლოგი	125
2. პოლიტოლოგი	240
3. ფსიქოლოგი	110
4. მათემატიკოსი	165
5. ფიზიკოსი	220
6. ქიმიკოსი	140
სულ	*1000*

ცხრილი №7

რამდენად კმაყოფილი ხართ უნივერსიტეტში მიმდინარე რეფორმებით?	აბსოლუტური სიხშირე
1. სავსებით კმაყოფილი	170
2. უფრო კმაყოფილი, ვიდრე უკმაყოფილო	190
3. კმაყოფილიც და უკმაყოფილოც (თანაბრად)	310
4. უფრო უკმაყოფილო, ვიდრე კმაყოფილი	180
5. სრულიად უკმაყოფილო	150
სულ	*1000*

ეს ცხრილები ვერ ასახავს იმ ფაქტს, რომ თითოეული ცვლადის განაწილების ყოველი ჯგუფი **განაწილებულია მეორე ცვლადის მნიშვნელობების მიხედვით.** მაგალითად:

პირველ განაწილებაში 125 რესპონდენტი სოციოლოგიის სპეციალობის სტუდენტია. ამ 125 სტუდენტი-სოციოლოგის ერთმა ნაწილმა მეორე ცვლადს მიანიჭა მნიშვნელობა "სავსებით კმაყოფილი", მეორე ნაწილმა: "უფრო კმაყოფილი, ვიდრე უკმაყოფილო" და ა. შ.

თავის მხრივ, "სავსებით კმაყოფილი" 170 სტუდენტი-რესპონდენტი განაწილებულია სპეციალობის მიხედვით: ერთი ნაწილი სოციოლოგია, მეორე ნაწილი პოლიტოლოგი და ა. შ.

თუ შერჩევით ერთობლიობას ამ **ორი ცვლადის მნიშვნელობათა ყოველი წყვილის** შესაბამისად გავანაწილებთ, მივიღებთ ორი ცვლადის დამოკიდებულების ანუ **კორელაციის ცხრილს (contingency tables), რომელსაც კროს-ტაბულაციასაც (cross-tabulation)** უწოდებენ:

[1] კონკრეტული სიტუაციის მიხედვით, იგი შეიძლება იყოს ერთ-ერთი ან ერთადერთი მიზეზი.

ცხრილი №8

I. სტუდენტთა სპეციალობა	II. უნივერსიტეტში მიმდინარე რეფორმებით სტუდენტთა კმაყოფილების ხარისხი					
	1	2	3	4	5	$n_{i\theta}$
1. სოციოლოგი	n_{11}=14	n_{12}=36	n_{13}=53	n_{14}=13	n_{15}=9	*125*
2. პოლიტოლოგი	n_{21}=48	n_{22}=58	n_{23}=93	n_{24}=23	n_{25}=18	*240*
3. ფსიქოლოგი	n_{31}=10	n_{32}=37	n_{33}=23	n_{34}=27	n_{35}=13	*110*
4. მათემატიკოსი	n_{41}=21	n_{42}=33	n_{43}=33	n_{44}=36	n_{45}=42	*165*
5. ფიზიკოსი	n_{51}= 56	n_{52}=17	n_{53}=57	n_{54}=48	n_{55}=42	*220*
6. ქიმიკოსი	n_{61}=21	n_{62}=9	n_{63}=51	n_{64}=33	n_{65}=26	*140*
$n_{\theta j}$	*170*	*190*	*310*	*180*	*150*	*n=1000*

კორელაციის ცხრილში ერთდროულად არის წარმოდგენილი მნიშვნელობათა რამდენიმე განაწილება:

- სტუდენტთა სპეციალობის განაწილება ($n_{i\theta}$-ის ქვემოთ);
- უნივერსიტეტში მიმდინარე რეფორმებით სტუდენტთა კმაყოფილების ხარისხის განაწილება ($n_{\theta j}$-ის გასწვრივ);
- თითოეული ცვლადის ყოველი მნიშვნელობის გასწვრივ მითითებულია სტუდენტთა იმ ჯგუფის განაწილება, რომლის წევრებმაც ეს მნიშვნელობა დაასახელეს, **მეორე ცვლადის მნიშვნელობების მიხედვით.**

განაწილებათა ასეთი **ჯვარედინი** წარმოდგენის შედეგად, მნიშვნელობების **ყოველი წყვილის გადაკვეთაზე** ვიღებთ იმ შემთხვევების რაოდენობას, როდესაც ეს ორი მნიშვნელობა ერთდროულად დასახელდა ანუ **ორგანზომილებიან განაწილებას.**

n_{ij} არის იმ სტუდენტების რაოდენობა, რომლებმაც დაასახელეს პირველი ცვლადის **i**-ური მნიშვნელობა და მეორე ცვლადის **j**-ური მნიშვნელობა.

n_{ij}-ის ფარდობითი სიხშირე იმ სტუდენტებთან ფარდობაში, რომლებმაც პირველი ცვლადის **i**-ური მნიშვნელობა დაასახელეს, იქნება: $\varpi^1_{ij} = n_{ij}/n_{i\theta}$

ხოლო იმ სტუდენტებთან ფარდობაში, რომლებმაც მეორე ცვლადის **j**-ური მნიშვნელობა დაასახელეს: $\varpi^2_{ij} = n_{ij}/n_{\theta j}$

წარმოვადგინოთ კორელაციის ცხრილი ამ თანაფარდობების მიხედვით:

ცხრილი №9

I. სტუდენტთა სპეციალობა	II. უნივერსიტეტში მიმდინარე რეფორმებით სტუდენტთა კმაყოფილების ხარისხი					
						$n_{i\theta}$
1. სოციოლოგი	14/125 =*0.11*	36/125 =*0.29*	53/125 =*0.42*	13/125 =*0.10*	9/125 =*0.07*	*125*
2. პოლიტოლოგი	48/240 = *0.2*	58/240 =*0.24*	93/240 =*0.3*	23/240 =*0.10*	18/240 =*0.08*	*240*
3. ფსიქოლოგი	10/110 =*0.09*	37/110 =*0.34*	23/110 =*0.21*	27/110 =*0.25*	13/110 =*0.12*	*110*
4. მათემატიკოსი	21/165 =*0.13*	33/165 =*0.2*	33/165 =*0.2*	36/165 =*0.22*	42/165 =*0.26*	*165*
5. ფიზიკოსი	56/220 =*0.25*	17/220 =*0.08*	57/220 =*0.26*	48/220 =*0.22*	42/220 =*0.19*	*220*
6. ქიმიკოსი	21/140 =*0.15*	9/140 =*0.06*	51/140 =*0.36*	33/140 =*0.24*	26/140 =*0.19*	*140*
						n=1000

სტუდენტთა სპეციალობის ყოველი მნიშვნელობის გასწვრივ, თეთრ გრაფებში

მითითებულია ამ სპეციალობის სტუდენტთა ჯგუფის განაწილება მიმდინარე რეფორმებით კმაყოფილების ხარისხის მიხედვით.

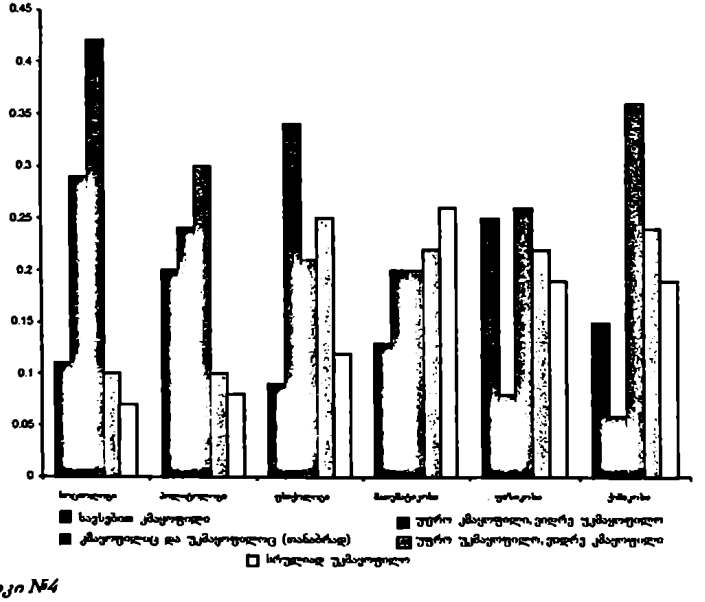
მაგალითად, ფსიქოლოგიის სპეციალობის სტუდენტების ჯგუფის აბსოლუტური სიხშირეა 110. ამ ჯგუფის 10/110 სავსებით კმაყოფილია უნივერსიტეტში მიმდინარე რეფორმებით, 37/110 — უფრო კმაყოფილი, ვიდრე უკმაყოფილო, 23/110 — კმაყოფილიც და უკმაყოფილოც (თანაბრად), 27/110 — უფრო უკმაყოფილო, ვიდრე კმაყოფილი, ხოლო 13/110 — სრულიად უკმაყოფილო.

მიმდინარე რეფორმებით სტუდენტთა კმაყოფილების ხარისხის ყოველი მნიშვნელობის ქვემოთ, ნაცრისფერ გრაფებში მოცემულია სტუდენტთა შესაბამისი ჯგუფის განაწილება სპეციალობის მიხედვით.

მაგალითად, უნივერსიტეტში მიმდინარე რეფორმებით უფრო კმაყოფილი, ვიდრე უკმაყოფილო სტუდენტების ჯგუფის აბსოლუტური სიხშირეა 190. ამ ჯგუფის 36/190 სოციოლოგია, 58/190 — პოლიტოლოგი, 37/190 — ფსიქოლოგი, 33/190 — მათემატიკოსი, 17/190 — ფიზიკოსი, ხოლო 9/190 — ქიმიკოსი.

მონაცემთა ასეთი სახით წარმოდგენა ჯგუფების **სტრუქტურირებული აღწერისა და მათი ერთმანეთთან შედარების საშუალებას გვაძლევს.**

ჩვენი მაგალითის მიხედვით, შესაძლებელი ხდება ვიმსჯელოთ მიმდინარე რეფორმებით კმაყოფილების ხარისხზე სხვადასხვა სპეციალობის სტუდენტთა ჯგუფებში. განვიხილოთ განაწილების დიაგრამა:



გრაფიკი №4

გრაფიკი №4 გამოსახავს სხვადასხვა სპეციალობის სტუდენტთა ჯგუფების განაწილებას მიმდინარე რეფორმებით კმაყოფილების ხარისხის მიხედვით.

ვიზუალური შთაბეჭდილება უკვე საკმარისია იმაში დასარწმუნებლად, რომ თითოეული

სპეციალობის სტუდენტთა ჯგუფის სტრუქტურა სპეციფიკურია. თუმცა, გარკვეული ჯგუფების სტრუქტურებს შორის შეინიშნება საერთო თვისებებიც. მაგალითად:

- რადიკალურად უარყოფითი შეფასების ("სრულიად უკმაყოფილო") სიხშირეები მათემატიკის, ფიზიკისა და ქიმიის სპეციალობების სტუდენტთა ჯგუფებში უფრო მაღალია, ვიდრე სოციოლოგიის, პოლიტოლოგიისა და ფსიქოლოგიის სპეციალობების სტუდენტთა ჯგუფებში;
- მნიშვნელობა "უფრო კმაყოფილი, ვიდრე უკმაყოფილო" სოციოლოგიის, პოლიტოლოგიისა და ფსიქოლოგიის სტუდენტთა ჯგუფებში უფრო მაღალი სიხშირეებითაა წარმოდგენილი, ვიდრე მათემატიკის, ფიზიკისა და ქიმიის სპეციალობების სტუდენტთა ჯგუფებში და ა. შ.

შეიძლება თუ არა დავასკვნათ, რომ მიმდინარე რეფორმებით სტუდენტთა კმაყოფილების ხარისხი **დამოკიდებულია** სტუდენტთა სპეციალობაზე?

ამ კითხვას გრაფიკიდან მიღებული ვიზუალური შთაბეჭდილებისა და ცხრილებში წარმოდგენილი განაწილებების საფუძველზე ვერ ვუპასუხებთ.

ჯერ ის უნდა დავადგინოთ, ჯგუფებს შორის არსებული სტრუქტურული იგივეობა/განსხვავებები **გენერალური ერთობლიობისათვის დამახასიათებელი კანონზომიერებაა, თუ მხოლოდ შერჩევითი ერთობლიობის შემთხვევითი თავისებურება?**

მივაქციოთ ყურადღება, რომ ეს პრობლემა **თვისებრივად განსხვავდება** იმ საკითხებისაგან, რომლებიც უკვე შევისწავლეთ.

ერთგანზომილებიანი სიხშირული განაწილებებისა თუ ცენტრალური ტენდენციებისა და ვარიაციის მახასიათებლების მეშვეობით ჩვენ მხოლოდ **შერჩევითი ერთობლიობის** შესახებ ვმსჯელობდით.

სტატისტიკის იმ მეთოდების ერთობლიობას, რომელთა მეშვეობითაც შერჩევითი ერთობლიობას აღწერას ვახდენთ, **აღწერითი სტატისტიკა** ეწოდება.

ახლა კი საკითხი **გენერალური ერთობლიობის** შესახებ დაისვა, რაც იმას ნიშნავს, რომ სტატისტიკის სპეციფიკური მეთოდები დაგვჭირდება.

სტატისტიკის იმ მეთოდების ერთობლიობას, რომლებიც აღწერითი სტატისტიკის შედეგების შეფასებისა და გენერალურ ერთობლიობაზე მათი განზოგადების საშუალებას გვაძლევს, **შეფასებითი/დასკვნითი** სტატისტიკა ეწოდება.

❖ კორელაციის სტატისტიკური მნიშვნელოვნების შეფასება

შერჩევითი ერთობლიობის ფარგლებში გამოვლენილი კორელაცია გენერალურ ერთობლიობაზე მხოლოდ მაშინ შეიძლება განზოგადდეს, თუ იგი იმდენად მნიშვნელოვანია, რომ შემთხვევითად არ ჩაითვლება.

მაგრამ, რის მიხედვით დავადგინოთ, **რამდენად მნიშვნელოვანია** კორელაცია, როგორ განვსაზღვროთ მისი **სტატისტიკური მნიშვნელოვნება?**

ამისათვის ჯერ კორელაციის გარკვეული სტატისტიკური მახასიათებლის, **მაჩვენებლის** პოვნაა საჭირო.

ორ ცვლადს შორის დამოკიდებულების სტატისტიკური მაჩვენებელი ყველაზე მარტივად ამ ცვლადებს შორის ნულოვან დამოკიდებულებასთან მიმართებით შეიძლება განისაზღვროს.

ცვლადებს შორის დამოკიდებულება ნულოვანია, თუ ისინი ერთმანეთისაგან **დამოუკიდებელია.**

დამოუკიდებლობის შემთხვევაში, ერთი ცვლადის ვარიაციული მწკრივის ნებისმიერ ჯგუფში მეორე ცვლადის ყოველი მნიშვნელობა იმ პროპორციით უნდა შედიოდეს, რა პროპორციითაც ეს უკანასკნელი შერჩევით ერთობლიობაშია წარმოდგენილი:

$$\frac{n_{ij}}{n_{0j}} = \frac{n_{i0}}{n}$$

საიდანაც:

$$n_{ij} = \frac{n_{i0}\, n_{0j}}{n}$$

ემპირიული სიხშირისაგან განსასხვავებლად, მას თეორიული სიხშირე ვუწოდოთ და აღვნიშნოთ n'_{ij}-ით.

ჩვენი მაგალითის მიხედვით:

სოციოლოგები შერჩევითი ერთობლიობის 125/1000=1/8-ს წარმოადგენენ, პოლიტოლოგები — 240/1000=6/25 და ა. შ.

მიმდინარე რეფორმებით "სავსებით კმაყოფილი" რესპონდენტები შერჩევითი ერთობლიობის 170/1000=17/100-ს წარმოადგენენ, "უფრო კმაყოფილი, ვიდრე უკმაყოფილო" — 190/1000=19/100 და ა. შ.

თუ დავუშვებთ, რომ მიმდინარე რეფორმებით სტუდენტთა კმაყოფილების ხარისხი და სტუდენტთა სპეციალობა ერთმანეთისაგან დამოუკიდებელია, მაშინ:

ა) 170 "სავსებით კმაყოფილი რესპონდენტიდან" 1/8 უნდა იყოს სოციოლოგი, 6/25 — პოლიტოლოგი და ა. შ.

190 "უფრო კმაყოფილი, ვიდრე უკმაყოფილო" რესპონდენტიდან ასევე 1/8 უნდა იყოს სოციოლოგი, 6/25 — პოლიტოლოგი და ა. შ.

ბ)[1] და პირიქით, 125 სოციოლოგიდან 170/1000=17/100 "სავსებით კმაყოფილი" უნდა იყოს, 190/1000=19/100 — "უფრო კმაყოფილი, ვიდრე უკმაყოფილო" და ა. შ.

240 პოლიტოლოგიდან 17/100 "სავსებით კმაყოფილი" უნდა იყოს, 19/100 — "უფრო კმაყოფილი, ვიდრე უკმაყოფილო" და ა. შ.

განვიხილოთ ცხრილი:

ცხრილი №10

I. სტუდენტთა სპეციალობა	II. მიმდინარე რეფორმებით სტუდენტთა კმაყოფილების ხარისხი					n_{i0}
1. სოციოლოგი	n_{11}=14	n_{12}=36	n_{13}=53	n_{14}=13	n_{15}=9	*125*
2. პოლიტოლოგი	n_{21}=48	n_{22}=58	n_{23}=93	n_{24}=23	n_{25}=18	*240*
3. ფსიქოლოგი	n_{31}=10	n_{32}=37	n_{33}=23	n_{34}=27	n_{35}=13	*110*
4. მათემატიკოსი	n_{41}=21	n_{42}=33	n_{43}=33	n_{44}=36	n_{45}=42	*165*
5. ფიზიკოსი	n_{51}= 56	n_{52}=17	n_{53}=57	n_{54}=48	n_{55}=42	*220*
6. ქიმიკოსი	n_{61}=21	n_{62}=9	n_{63}=51	n_{64}=33	n_{65}=26	*140*
n_{0j}	*170*	*190*	*310*	*180*	*150*	*n=1000*

ცხრილი №10-ში ემპირიული სიხშირეების ქვემოთ, ნაცრისფერ გრაფებში მითითებულია **თეორიული — ნულოვანი კორელაციის შესაბამისი პროპორციებით გამოთვლილი სიხშირეები.** ანუ, სიხშირეები, რომლებიც უნდა დაფიქსირებულიყო მოცემულ ცვლადებს შორის, თუ ისინი სტატისტიკურად ერთმანეთისაგან დამოუკიდებლები იქნებოდნენ.

თეორიულ განაწილებაში თითოეული ცვლადის **ვარიაციული მწკრივის ყოველი ჯგუფი** მეორე ცვლადის მიხედვით განაწილებული **შერჩევითი ერთობლიობის სტრუქტურული ასლია.**

ბუნებრივია, რომ კორელაციის **სტატისტიკურ მახასიათებლად** შეიძლება მივიჩნიოთ **ნულოვანი კორელაციიდან გადახრა** — განსხვავება, რომელიც ემპირიულ და თეორიულ (ნულოვანი კორელაციის შესაბამის) სიხშირეებს შორის არსებობს.

ეს განსხვავება გამოითვლება ფორმულით:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{s} \frac{(n_{ij} - n'_{ij})^2}{n'_{ij}}$$

მიღებულ სიდიდეს **ხი-კვადრატს** უწოდებენ. ხი-კვადრატის გამოსათვლელ ფორმულაში:

- **k** ერთი ცვლადის მნიშვნელობათა რაოდენობაა, **s** — მეორე ცვლადის;
- ემპირიულ და თეორიულ მნიშვნელობებს შორის სხვაობის კვადრატში[2] აყვანა ჯამის წევრების გაერთნიშნიანების მიზნით ხდება;
- სხვაობის კვადრატის შეფარდებით თეორიულ სიხშირესთან კი ფარდობითი სიდიდე მიიღება[3].

ჩვენი მაგალითის შემთხვევაში $\chi^2 \approx 142.9$.

მაგრამ χ^2-ის კონკრეტული მნიშვნელობა (მოცემულ შემთხვევაში, ≈142.9), თავისთავად აღებული, არაფერს გვეუბნება იმის შესახებ, თუ **რამდენად მნიშვნელოვანია კორელაცია.**

იმისათვის, რომ ამ კითხვას ვუპასუხოთ, უფრო ღრმად გავიაზროთ, რა სიდიდე მივიღეთ χ^2-ის სახით.

χ^2-ის ემპირიული მნიშვნელობების გამოთვლა შერჩევითი ერთობლიობის ფარგლებში ხდება, ამიტომ, არც თვითონ ეს მნიშვნელობებია "დაზღვეული" შემთხვევითი გადახრებისაგან:

დავუშვათ, ერთი და იგივე გენერალური ერთობლიობიდან, შემთხვევითი წესით, მრავალჯერ შევარჩიეთ ერთი და იგივე რაოდენობის რესპონდენტები და გამოვთვალეთ χ^2-ის მნიშვნელობა შერჩევის თითოეული შემთხვევისათვის.

ცხადია, რომ რაც არ უნდა ზუსტად დავიცვათ შერჩევისა და გამოკითხვის წესები, χ^2-ის მნიშვნელობებს ერთმანეთისაგან განსხვავებულს მივიღებთ.

მაგრამ შემთხვევითი შერჩევა შემთხვევით გადახრებთან ერთად სტატისტიკურ კანონზომიერებებსაც ასახავს.

ამიტომ, χ^2-ის მნიშვნელობების განაწილებაც განსაზღვრულ სტატისტიკურ კანონზომიერებას ექვემდებარება.

ამ კანონზომიერებას χ^2-ის **განაწილების კანონს უწოდებენ.**

იგი დამოკიდებულია პარამეტრზე, რომელსაც თავისუფლების ხარისხს უწოდებენ. თავისუფლების ხარისხი გამოითვლება ფორმულით **(k-1)(s-1)**, სადაც **k** ერთი ცვლადის შესაძლო მნიშვნელობათა რაოდენობაა, ხოლო **s** — მეორის.

χ^2-ის განაწილების კანონის საფუძველზე შესაძლებელია განვსაზღვროთ მისი თეორიული მნიშვნელობები მოცემული თავისუფლების ხარისხისა და დასაშვები ცდომილების (α) დონის მიხედვით, როდესაც ორ ცვლადს შორის დამოკიდებულება ნულოვანია.

---

[1] თეორიული განაწილების პროპორციებიდან გამომდინარე, ა) და ბ) დებულებები, ფაქტობრივად, ეკვივალენტურია: ა) პირობის შესრულებისას ბ) თავისთავად სრულდება და პირიქით — ბ) პირობის შესრულება ა) პირობის შესრულებასაც ნიშნავს.

[2] ისევე, როგორც ვარიაციის მახასიათებლების გამოთვლისას.

[3] გავიხსენოთ ფარდობითი სიხშირეების შემოტანის იდეა.

α შეიძლება უდრიდეს 0.01; 0.05; 0.10 და ა. შ[1].

ამ მნიშვნელობებს α დონის კრიტიკული მნიშვნელობები ეწოდება და წარმოდგენილია ცხრილის — ე. წ. t-ტესტის (t-კრიტერიუმის) სახით.

მაგალითისათვის მოვიტანთ t-ტესტის ფრაგმენტს:

თავისუფლების ხარისხი	α=0.01	α=0.05	α=0.10
1	0.000157	0.00393	0.0158
2	0.0201	0.103	0.211
3	0.115	0.352	0.584
4	0.297	0.711	1.06
5	0.554	1.15	1.61
6	0.872	1.64	2.20
7	1.24	2.17	2.83
8	1.65	2.73	3.49
9	2.09	3.33	4.17
10	2.56	3.94	4.87
11	3.05	4.57	5.58
12	3.57	5.23	6.30
13	4.11	5.89	7.04
14	4.66	6.57	7.79
15	5.23	7.26	8.55
16	5.81	7.96	9.31
17	6.41	8.67	10.1
18	7.01	9.39	10.9
19	7.63	10.1	11.7
20	8.26	10.9	12.4

ახლა უკვე შეგვიძლია აღვწეროთ, თუ როგორ ფასდება კორელაციის მნიშვნელოვნება χ^2-ის მიხედვით.

როდესაც შერჩევითი ერთობლიობის ფარგლებში გამოვლენილი კორელაციის მნიშვნელოვნების შეფასებას ვცდილობთ, ფაქტობრივად, ვამოწმებთ **სტატისტიკურ ჰიპოთეზას** გენერალურ ერთობლიობაში ორ ცვლადს შორის დამოკიდებულების შესახებ.

შესაბამისად, ორ ცვლადს შორის კორელაციის მნიშვნელოვნების შეფასება ხდება სტატისტიკური ჰიპოთეზების შემოწმების მეთოდით:

- ვარაუდს, რომ ორ ცვლადს შორის ნულოვანი დამოკიდებულებაა, **ნულოვანი ჰიპოთეზა** (H_0) ეწოდება.
- χ^2-ის ემპირიული მნიშვნელობა უნდა შევადაროთ კრიტიკულს, რომელიც ნულოვან ჰიპოთეზას (ნულოვან დამოკიდებულებას) შეესაბამება და t-ტესტში შემთხვევითი ერთობლიობის თავისუფლების ხარისხისა და α დონის მიხედვით ვიპოვით.

ჩვენი მაგალითის შემთხვევაში, ერთი ცვლადის შესაძლო მნიშვნელობათა რაოდენობაა 6, მეორის კი 5. შესაბამისად, თავისუფლების ხარისხი იქნება: (6-1)(5-1)=20.

[1] მათემატიკურ სტატისტიკაში შერჩევითი ერთობლიობის კვლევის შედეგად მიღებული დასკვნები არ არის აბსოლუტურად ზუსტი. დასაშვები ცდომილების დონე გვიჩვენებს ზუსტი მაჩვენებლიდან გადახრის ზომას. მაგალითად, α= 0.05-ს ნიშნავს, რომ ჩვენი დასკვნები სწორია 100-დან 95 შემთხვევაში. დასაშვები ცდომილების განსხვავებულ დონეს, ცხადია, განსხვავებული განაწილებები შეესაბამება. ამიტომ, როდესაც შედარებას ვახდენთ, აუცილებლად უნდა მივუთითოთ დასაშვები ცდომილების რომელი დონის შესაბამის განაწილებას ვირჩევთ. ჩვენი შედარების შედეგად მიღებული დასკვნაც, ცხადია, ამ დონის შესაბამისი იქნება.

t-ტესტში მოვძებნოთ შემთხვევა, როდესაც თავისუფლების ხარისხია 20, ცდომილების, დავუშვათ, 0.05 დონისათვის[1]. ეს მნიშვნელობაა **10.9**. ე. ი. $\chi t^2 = 10.9$.

- თუ ემპირიული/შემთხვევითი მნიშვნელობა აღემატება კრიტიკულს, ითვლება, რომ ნული-დან გადახრა მნიშვნელოვანია და კორელაცია გენერალური ერთობლიობისათვის დამახასიათებელ სტატისტიკურ კანონზომიერებას წარმოადგენს.
- თუ ემპირიული მნიშვნელობა არ აღემატება კრიტიკულს, ითვლება, რომ ცვლადების მოცემული წყვილი გენერალური ერთობლიობის მასშტაბით ერთმანეთისაგან დამოუკიდებელია.

ჩვენს შემთხვევაში, $\chi^2 \approx 142.9$, ხოლო $\chi t^2 \approx 10.9$. აქედან გამომდინარე:

$$\chi^2 > \chi t^2$$

რაც იმას ნიშნავს, კორელაცია სტატისტიკურად მნიშვნელოვანია და მოცემულ ცვლადებს შორის გამოვლენილი დამოკიდებულება გენერალური ერთობლიობისათვის დამახასიათებელ სტატისტიკურ კანონზომიერებად უნდა მივიჩნიოთ.

❖ კორელაციის დახასიათება. კორელაციის კოეფიციენტები

ამრიგად, χ^2-ის მნიშვნელობის მიხედვით შეიძლება განვსაზღვროთ არსებობს თუ არა ორ ცვლადს შორის კორელაცია, მაგრამ იგი არანორმირებული სიდიდეა და ნებისმიერი დადებითი მნიშვნელობა შეიძლება მიიღოს.

ამიტომ, χ^2-ის მიხედვით კორელაციას ვერ **დავახასიათებთ**. ეს ამოცანა კორელაციის კოეფიციენტების შემოტანით წყდება.

კორელაციის კოეფიციენტები, ისევე, როგორც χ^2, კორელაციის სტატისტიკური მაჩვენებლებია.

თუმცა, χ^2-ისაგან განსხვავებით, მათი მეშვეობით **კორელაციის მნიშვნელოვნების შეფასებასთან ერთად**, შესაძლებელი ხდება კორელაციის **დახასიათებაც**.

კორელაციის სხვადასხვა კოეფიციენტის არსებობას შემდეგი ფაქტორები განაპირობებს:

- კორელაციის კოეფიციენტის იდეის მათემატიკური რეალიზება ანუ ისეთი ფორმულის გამოყვანა, რომელიც ორ ცვლადს შორის დამოკიდებულების ხარისხის შეფასებისა და ამ დამოკიდებულების აღწერის საშუალებას მოგვცემს, სხვადასხვა მეთოდითაა შესაძლებელი.
- გაზომვის დონეების სპეციფიკა ცვლადებს შორის დამოკიდებულებაზე აისახება. ამიტომ, იმის მიხედვით, თუ გაზომვის რომელ დონეზე უნდა შეფასდეს და დახასიათდეს ორ ცვლადს შორის დამოკიდებულება, განსხვავებული კოეფიციენტების შემოტანაა საჭირო.
- კორელაციის კოეფიციენტები ადაპტირებულია ანალიზის ამოცანებთან და ერთმანეთისაგან განსხვავდებიან, აგრეთვე, იმის მიხედვით ანალიზი ერთი შერჩევითი ერთობლიობის ფარგლებში ტარდება, თუ დაგვჭირდება შედარება სხვა შერჩევითი ერთობლიობის მონაცემებთან.

ანალიზის დროს კორელაციის კოეფიციენტის შერჩევა, ცხადია, ამ ფაქტორების გათვალისწინებით უნდა მოხდეს.

განსხვავებების მიუხედავად, კორელაციის კოეფიციენტების შემოტანის **იდეისა და ძირითადი ფუნქციების იდენტურობა** მათი პარალელური დახასიათების საშუალებას გვაძლევს:

- კორელაციის კოეფიციენტები ნორმირებული სიდიდეებია. მათი მნიშვნელობები -1-დან +1 იცვლება;

[1] α=0.05 დონის შერჩევა ტიპურია.

- თუ კორელაციის კოეფიციენტის მნიშვნელობა ნულია, ითვლება, რომ ორი ცვლადი ერთმანეთისაგან დამოუკიდებელია;
- რაც უფრო უახლოვდება კოეფიციენტის მნიშვნელობა ზღვრულ მნიშვნელობებს (-1 ან 1), მით უფრო ძლიერია დამოკიდებულება (მჭიდროა კავშირი) ცვლადებს შორის;
- კოეფიციენტის დადებითი მნიშვნელობა ორ ცვლადს შორის პირდაპირპროპორციული დამოკიდებულების არსებობას მიუთითებს, ხოლო უარყოფითი მნიშვნელობა — უკუპროპორციული დამოკიდებულების არსებობას.

გაზომვის რაოდენობრივი დონის კოეფიციენტებს შორის ინტენსიურად იყენებენ პირსონის **r** კოეფიციენტს, რომლის მეშვეობითაც მოწმდება სტატისტიკური ჰიპოთეზა გენერალურ ერთობლიობაში ორ ცვლადს შორის წრფივი დამოკიდებულების არსებობის შესახებ.

როდესაც პირსონის **r** კოეფიციენტი იღებს ზღვრულ მნიშვნელობას (-1 ან 1), ითვლება, რომ ორ ცვლადს შორის წრფივი ფუნქციური დამოკიდებულებაა.

გაზომვის რიგის დონის კოეფიციენტებს მიეკუთვნება რანგული კოეფიციენტები: კენდალის τ_k, სპირმენის r_s, გუდმენისა და კრასკალის γ (გამა) კოეფიციენტები და სხვა.

რანგული კოეფიციენტები შესაძლებლობას გვაძლევს შევაფასოთ ორ ცვლადს შორის დამოკიდებულების მნიშვნელობა გაზომვის რიგის დონეზე და **დავახასიათოთ რანგების დალაგება:** თუ ერთი რიგის ცვლადის მნიშვნელობების ზრდა მეორე რიგის ცვლადის მნიშვნელობების ზრდას იწვევს, დალაგებას **პირდაპირი** ეწოდება, ხოლო თუ ერთი რიგის ცვლადის მნიშვნელობათა ზრდა მეორე რიგის ცვლადის მნიშვნელობათა კლებას იწვევს — **უკუდალაგება**.

სპირმენის r_s კოეფიციენტის მიხედვით, შესაძლებელია რანგებს შორის დაშორებაც დახასიათდეს.

გაზომვის ნომინალური დონის კოეფიციენტების მაგალითია **პირსონის ხი-კვადრატ კოეფიციენტი:**

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

სადაც **n** შერჩევითი ერთობლიობის რაოდენობაა.

ჩვენი მაგალითის შემთხვევაში, პირსონის ხი-კვადრატ კოეფიციენტი $C \approx 0.35$. იგი საკვლევ ცვლადებს შორის სუსტი დამოკიდებულების არსებობაზე მიუთითებს.

კორელაციის კოეფიციენტების მიხედვით კორელაციის მნიშვნელოვნება ისევე ფასდება, როგორც χ^2-ის შემთხვევაში: მოცემული კორელაციის კოეფიციენტის განაწილების კანონის საფუძველზე განისაზღვრება კოეფიციენტის თეორიული მნიშვნელობები შესაბამისი პარამეტრებისა და ცდომილების (α) დონის მიხედვით.

კრიტიკული მნიშვნელობები წარმოდგენილია ცხრილების სახით.

კორელაციის კოეფიციენტის მიხედვით, კორელაციის მნიშვნელობის შეფასება ხდება სტატისტიკური ჰიპოთეზების შემოწმების მეთოდით:

შერჩევითი ერთობლიობის მიხედვით მიღებული კოეფიციენტის მნიშვნელობა უნდა შევადაროთ კრიტიკულს, რომელსაც ცხრილში შემთხვევითი ერთობლიობის პარამეტრებისა და α დონის მიხედვით ვიპოვით.

თუ ემპირიული/შემთხვევითი მნიშვნელობა აღემატება კრიტიკულს, ითვლება, რომ ნულიდან გადახრა მნიშვნელოვანია და კორელაცია გენერალური ერთობლიობისათვის დამახასიათებელ სტატისტიკურ კანონზომიერებას წარმოადგენს.

თუ ემპირიული მნიშვნელობა არ აღემატება კრიტიკულს, ითვლება, რომ ცვლადების მოცემული წყვილი გენერალური ერთობლიობის მასშტაბით ერთმანეთისაგან დამოუკიდებელია.

❖ სტატისტიკური დამოკიდებულების ინტერპრეტაცია

ამრიგად, ცვლადებს შორის დამოკიდებულების შესახებ შინაარსობრივ მსჯელობასა და სტრუქტურულ-შედარებით ანალიზს მხოლოდ იმ შემთხვევაში აქვს აზრი, როდესაც მათ შორის სტატისტიკური კანონზომიერება არსებობს. მაგრამ **არა ყოველი სტატისტიკური დამოკიდებულება ექვემდებარება ინტერპრეტაციას.**

ცვლადებს შორის სტატისტიკური დამოკიდებულება სიმეტრიულია:

თუ ერთი ცვლადი სტატისტიკურად დამოკიდებულია მეორეზე, მაშინ მეორეც დამოკიდებული იქნება პირველზე, რადგან თუ პირველის განაწილება გარკვეული კანონზომიერებით არის დაკავშირებული მეორესთან, მაშინ, ბუნებრივია, მეორეც კანონზომიერად იქნება დაკავშირებული პირველთან.

სტატისტიკური დამოკიდებულებისაგან განსხვავებით, **შინაარსობრივი დამოკიდებულება შეიძლება არ იყოს სიმეტრიული**[1].

ჩვენი მაგალითის მიხედვით მიღებული რეზულტატი შინაარსობრივად შემდეგნაირად შეიძლება ინტერპრეტირდეს: მიმდინარე რეფორმებით სტუდენტთა კმაყოფილების ხარისხი დამოკიდებულია სტუდენტთა სპეციალობაზე.

მაგრამ აზრი არა აქვს მსჯელობას იმის შესახებ, რომ სპეციალობა, რომელიც რესპონდენტებმა უნივერსიტეტში შესვლამდე აირჩიეს, დამოკიდებულია უნივერსიტეტში ამჟამად მიმდინარე, ჯერ კიდევ დაუსრულებელი რეფორმებით სტუდენტთა კმაყოფილების ხარისხზე. ე. ი. ორი მოცემული ცვლადიდან ერთი დამოკიდებულია მეორეზე როგორც სტატისტიკურად, ისე შინაარსობრივად, მაგრამ მეორის პირველზე დამოკიდებულების შინაარსობრივი ახსნა შეუძლებელია.

ჩვენი მსჯელობა ზოგადი დასკვნის სახით შეიძლება შევაჯამოთ:

ცვლადებს შორის სტატისტიკური დამოკიდებულება მათ შორის შინაარსობრივი დამოკიდებულების არსებობის აუცილებელი, მაგრამ არა საკმარისი პირობაა.

❖ კორელაციური დამოკიდებულების სპეციფიკა

B ცვლადი **ფუნქციურად** არის დამოკიდებული **A** ცვლადზე, თუ **A** ცვლადის ყოველ მნიშვნელობას **B**-ს *ერთი და მხოლოდ ერთი მნიშვნელობა* შეესაბამება.

კორელაციის განხილვისას ჩვენ ცვლადებს შორის ისეთ დამოკიდებულებას დავაკვირდით, როდესაც დამოუკიდებელი ცვლადის **(A)** ერთსა და იმავე მნიშვნელობას დამოკიდებულის **(B)** სხვადასხვა მნიშვნელობა შეიძლება შეესაბამებოდეს:

მაგალითად, სტუდენტთა სპეციალობის თითოეულ მნიშვნელობას (დავუშვათ, 1/სოციოლოგს) შეესაბამებოდა უნივერსიტეტში მიმდინარე რეფორმებით კმაყოფილების ხარისხის **ერთ-ერთი მნიშვნელობა: 1 ან 2 ან 3 ან 4 ან 5** (სავსებით კმაყოფილი ან უფრო კმაყოფილი, ვიდრე უკმაყოფილო და ა. შ.).

ასეთ შემთხვევაში, ცხადია, ცვლადებს შორის ფუნქციური დამოკიდებულების შესახებ ვერ ვიმსჯელებთ. თუმცა, **არ არის გამორიცხული, რომ თავი იჩინოს გარკვეულმა კანონზომიერებამ/ტენდენციამ ფაქტორული და რეზულტატური ცვლადების მნიშვნელობათა ემპირიულ განაწილებებს შორის.**

ამ კანონზომიერების/ტენდენციის გამოვლენა შემდეგი მეთოდით შეიძლება:

1. ფაქტორული ცვლადის *თითოეული შესაძლო* მნიშვნელობისათვის გამოვყოთ რეზულტატური ცვლადის იმ **ემპირიული მნიშვნელობების** მწკრივი, რომლებიც ფაქტორული ცვლადის მოცემულ მნიშვნელობას გაზომვის სხვადასხვა შემთხვევაში შეესაბამება.

მაგალითად:

ფაქტორული ცვლადის შესაძლო მნიშვნელობა	1/სოციოლოგი
რეზულტატური ცვლადის ემპირიული მნიშვნელობები	4 1 1 1 2 1 3 4 5 3 2 1 2 5 4 3 2 1 2 3 2 2 1 2 3 5 4 1 3 1 გაზომვის 30 შემთხვევა

1. გამოვთვალოთ ამ მწკრივების საშუალო არითმეტიკულები. მათ **პირობითი საშუალოები** ენოდება[2].

ჩვენი მაგალითის მიხედვით, პირობითი საშუალო უდრის:

$$\Sigma (4\,1\,1\,1\,2\,1\,3\,4\,5\,3\,2\,1\,2\,5\,4\,3\,2\,1\,2\,3\,2\,2\,1\,2\,3\,5\,4\,1\,3)/30 = 2.5$$

2. ფაქტორული ცვლადის ყოველ მნიშვნელობას შევუსაბამოთ პირობითი საშუალო[1].

შედეგად, ფაქტორული ცვლადის ყოველ მნიშვნელობას შეესაბამება **ერთი და მხოლოდ ერთი მნიშვნელობა — პირობითი საშუალო. ე. ი. მივიღებთ ფუნქციური ხასიათის შესაბამისობას/ფუნქციას.**

ეს ფუნქცია, გარკვეული ალბათობით, ფაქტორული და რეზულტატური ცვლადების ემპირიულ განაწილებათა შორის არსებული დამოკიდებულების სპეციფიკასაც ასახავს, რადგან პირობითი საშუალოები რეზულტატური ცვლადის ემპირიული მნიშვნელობების მიხედვით არის გამოთვლილი.

ჩვენი მსჯელობა შემდეგნაირად შეიძლება შევაჯამოთ:

- კორელაცია ცვლადებს შორის ერთმნიშვნელოვან/ზუსტად განსაზღვრულ დამოკიდებულებას კი არ გამოხატავს, არამედ **დამოკიდებულების ზოგად კანონზომიერებას/ტენდენციას.**
- ამ ტენდენციის გამოვლენა ხდება არა **უშუალოდ** ცვლადების მნიშვნელობათა განაწილებების მიხედვით, არამედ, **სტატისტიკური მახასიათებლის — პირობითი საშუალოს შემოტანით.**
- შესაბამისად, კორელაციური დამოკიდებულება **ალბათური** ხასიათისაა.

რეგრესიული ანალიზი

❖ რეგრესიული ანალიზის ძირითადი ეტაპები

სოციალური **მოვლენები ურთიერთქმედებენ.** ეს იმას ნიშნავს, რომ შესასწავლი **მოვლენის** იმ სახით არსებობა, როგორც იგი კვლევის პროცესში ფიქსირდება, მნიშვნელოვანწილად **სხვა მოვლენების გავლენით** შეიძლება იყოს გაპირობებული.

როგორც უკვე აღვნიშნეთ, მოვლენათა შორის არსებული ურთიერთდამოკიდებულებები ცვლადებს შორის ურთიერთდამოკიდებულებებში აისახება.

სტატისტიკის მეთოდები საშუალებას გვაძლევს:

- განვსაზღვროთ მოცემული ცვლადის **ვარიაციის** სტატისტიკური მაჩვენებლები;
- დავადგინოთ, მოცემული ცვლადის ვარიაცია სრულიად შემთხვევითია, თუ ნაწილო-

[1] **დამოუკიდებელი და დამოკიდებული ცვლადების ცნებების პარალელურად ფაქტორული და რეზულტატური ცვლადების ცნებების შემოტანა, მნიშვნელოვანწილად, სწორედ ამ პრობლემის გათვალისწინებით მოხდა.**

[2] **რადგან თითოეული მწკრივი გამოიყოფა იმ პირობით, რომ მისი წევრები ფაქტორული ცვლადის ერთსა და იმავე/მოცემულ მნიშვნელობას შეესაბამება.**

ბრივ მაინც, სხვა ცვლადის ან ცვლადების ჯგუფის გავლენის **კანონზომიერ შედეგია;**

- **განვსაზღვროთ, კონკრეტულად რა სახის** კანონზომიერი დამოკიდებულება არსებობს ცვლადებს შორის.

ცვლადებს შორის სტატისტიკური დამოკიდებულების ანალიზითა და ინტერპრეტაციით კი მოვლენათა ურთიერთდამოკიდებულების გაგებას შევძლებთ.

კორელაციური ანალიზის მეთოდებით შესაძლებელია გამოვავლინოთ და ხარისხობრივად დავახასიათოთ **სტატისტიკური დამოკიდებულება** ორ ცვლადს შორის.

მაგრამ, **როგორ, რა სახის კანონზომიერებით** შეიძლება იყოს რეზულტატური ცვლადი **სტატისტიკურად დამოკიდებული** ფაქტორული ცვლადების ჯგუფზე?

ეს პრობლემა **რეგრესიული**[2] **ანალიზის** მეთოდებით წყდება.

ტერმინოლოგიის ლაკონიურობის მიზნით, ფაქტორულ ცვლადს **ფაქტორი,** ხოლო ფაქტორული ცვლადების ჯგუფს **ფაქტორების ჯგუფი** ვუწოდოთ.

იგულისხმება, რომ **ფაქტორების ჯგუფი** არაცარიელია. ე. ი. მასში ერთი მაინც (ერთი ან ერთზე მეტი) ფაქტორი შედის.

თუ ფაქტორების ჯგუფი სტატისტიკურად მნიშვნელოვან კანონზომიერ ზეგავლენას ახდენს რეზულტატურ ცვლადზე, ამბობენ, რომ ფაქტორების ჯგუფსა და რეზულტატურ ცვლადს შორის არსებობს რეგრესიული[3] **დამოკიდებულება.**

დავუშვათ, რომ **მასწავლებელთა კვალიფიკაციის დონე,** ლექციების **ხარისხი** და **სასწავლო დატვირთვის ხანგრძლივობა** სტატისტიკურად მნიშვნელოვანი კანონზომიერებით აისახება **სასწავლო პროცესით სტუდენტთა კმაყოფილების ხარისხზე.**

სასწავლო პროცესით სტუდენტთა კმაყოფილების ხარისხი **რეზულტატური** ცვლადია, მასწავლებელთა კვალიფიკაციის დონე — **ერთი ფაქტორი,** ლექციების ხარისხი — მეორე, სასწავლო დატვირთვის ხანგრძლივობა — მესამე, **ერთიანად** განხილული სამივე ფაქტორი — **ფაქტორების ჯგუფი,** ხოლო რეზულტატურ ცვლადსა და ფაქტორების ჯგუფს შორის არსებობს **რეგრესიული დამოკიდებულება.**

რეზულტატურ ცვლადზე ფაქტორების ჯგუფური ზეგავლენა სტატისტიკურად მნიშვნელოვნად იმ შემთხვევაში ითვლება, თუ ფაქტორების ჯგუფსა[4] **და რეზულტატურ ცვლადს შორის არსებობს სტატისტიკურად მნიშვნელოვანი კორელაცია.**

შესაბამისად, კორელაციური დამოკიდებულების თავისებურებიდან გამომდინარე, რეგრესიული დამოკიდებულებაც ზოგად კანონზომიერებას/ტენდენციას გამოხატავს და ალბათური ხასიათისაა.

რეგრესიული ანალიზი იმ მეთოდების ერთობლიობას ეწოდება, რომელთა მეშვეობითაც ცვლადებს შორის რეგრესიული დამოკიდებულების გამოვლენა და გაანალიზება ხდება.

რეგრესიული ანალიზის ძირითადი მიზანია ანალიზურად გამოისახოს რეგრესიული დამოკიდებულება ფაქტორების ჯგუფსა და რეზულტატურ ცვლადს შორის[5].

რეგრესიული ანალიზის მეშვეობით შესაძლებელია:

- განვსაზღვროთ **დამოკიდებულების ფორმა და შინაარსი** ფაქტორების ჯგუფსა და

---

[1] ჩვენს შემთხვევაში: 2.5.

[2] ლათ.: **regressio** — უკუსვლა, დაბრუნება.

[3] ე. ი. როდესაც შესაძლებელია, რომ ფაქტორების მნიშვნელობებსა და რეზულტატური ცვლადის მნიშვნელობებს შორის დამოკიდებულება გარკვეული ფუნქციის სახით წარმოვადგინოთ.

[4] ფაქტორების ჯგუფი ამ შემთხვევაში განიხილება როგორც კომპლექსური — სხვადასხვა ცვლადისაგან **შედგენილი ცვლადი.**

[5] ე. ი. მათემატიკური ანალიზის განტოლების სახით ჩამოყალიბდეს ფაქტორების მნიშვნელობების მიხედვით რეზულტატური ცვლადისათვის მნიშვნელობების მინიჭების წესი/შეფასებითი ფუნქცია.

რეზულტატურ ცვლადს შორის[1];

- სტატისტიკურად დავახასიათოთ დამოკიდებულება რეზულტატურ ცვლადსა და ფაქტორების ჯგუფს შორის;
- სტატისტიკურად დავახასიათოთ დამოკიდებულება რეზულტატურ ცვლადსა და ცალკეულ ფაქტორს შორის;
- განვსაზღვროთ რეზულტატური ცვლადის **ფარდობითი ცვლილება**, რომელიც ფაქტორების ჯგუფის **ფარდობითი ცვლილების ერთეულზე** მოდის;
- განვსაზღვროთ ცალკეული ფაქტორის **წვლილი** რეზულტატური ცვლადის ვარიაციაში.

შესაბამისად:

- რეგრესიული ანალიზი გვეხმარება ცვლადებს შორის **დეტერმინაციული** დამოკიდებულებების/კავშირების გამოვლენაში[2];
- შესაძლებელი ხდება ფაქტორების მნიშვნელობების მიხედვით განვსაზღვროთ/**"ვიწინასწარმეტყველოთ"** რეზულტატური ცვლადის მნიშვნელობები.

რეგრესიული ანალიზის ძირითადი ეტაპებია:

- იმ ფაქტორების გამოყოფა, რომლებიც **არსებითად** ზემოქმედებს რეზულტატურ ცვლადზე;
- **ცვლადებს შორის რეგრესიული დამოკიდებულების მოდელის — რეგრესიის** მოდელის აგება;
- შერჩევითი მაჩვენებლების სტატისტიკური **მნიშვნელოვნების** შეფასება;
- **რეგრესიის მოდელის ხარისხის** განსაზღვრა;
- რეზულტატურ ცვლადზე **ფაქტორების ჯგუფისა და ცალკეული ფაქტორის გავლენის სტატისტიკური შეფასება.**

❖ რეზულტატურ ცვლადზე მოქმედი არსებითი ფაქტორების გამოყოფა

სოციალურ მოვლენათა ურთიერთდამოკიდებულება **მრავალმხრივია** — ნებისმიერი სოციალური მოვლენა **მრავალი** სხვა მოვლენის ერთობლივ ზემოქმედებას შეიძლება განიცდიდეს.

შესაბამისად, **რეზულტატური** ცვლადის ემპირიული მნიშვნელობები (რომლებსაც ის უშუალოდ გაზომვის პროცესში იღებს) **სხვადასხვა მიზეზის** კომპლექსური/**ერთობლივი მოქმედების** შედეგს — **რეზულტატს** შეიძლება წარმოადგენდეს.

ეს მიზეზები იყოფა/კლასიფიცირდება იმის მიხედვით, თუ რამდენად მნიშვნელოვანია, არსებითია მათი გავლენა რეზულტატურ ცვლადზე.

მაგალითად:

- ყოველთვიური პირადი შემოსავლის საერთო რაოდენობაზე ხელფასი **რეგულარულად, კანონზომიერად** აისახება. მაგრამ იმას, თუ რა შემოსავალი ექნება რეალურად (სინამდვილეში) კონკრეტულ პიროვნებას ამა თუ იმ თვის განმავლობაში, **შემთხვევითი** მიზეზებიც განსაზღვრავს: მისი შემოსავალი შეიძლება შემცირდეს გარკვეული თანხის დაკარგვის, გასესხების, გაჩუქების გამო ან გაიზარდოს ნაპოვნი, ნასესხები, ნაჩუქარი თანხის გამო და ა. შ.

---

[1] **ე. ი. განვსაზღვროთ რეგრესიული დამოკიდებულების ფორმა და შინაარსი.**

[2] **შესაბამისად, გაზომვისას დაფიქსირებულ პირობებში მოვლენებს შორის მიზეზ-შედეგობრივი დამოკიდებულებების/კავშირების გამოვლენაში.**

- მასწავლებელთა კვალიფიკაციის დონე, ლექციების ხარისხი და სასწავლო დატვირთვის ხანგრძლივობა, როგორც ნესი, კანონზომიერ ზემოქმედებას/გავლენას ახდენს სასწავლო პროცესით სტუდენტთა კმაყოფილების ხარისხზე, მაგრამ სასწავლო პროცესით კონკრეტული სტუდენტის კმაყოფილებაზე შეიძლება რომელიმე მასწავლებელთან ან ჯგუფელთან მისმა პირადმა ურთიერთობამაც (განსაკუთრებით კარგი/ცუდი) იმოქმედოს ა. შ.

შემთხვევითი/არაკანონზომიერი მიზეზები რეზულტატურ ცვლადზე **კანონზომიერ** გავლენას ვერ მოახდენს.

ამიტომ, ცვლადებს შორის დამოკიდებულებათა შესწავლა იმის გარკვევით უნდა დავიწყოთ, თუ რომელი ცვლადები **შეიძლება** ახდენდეს რეზულტატურ ცვლადზე **არსებით/კანონზომიერ** გავლენას ანუ, რომელი ცვლადები უნდა შევიტანოთ **ფაქტორების** ჯგუფში, რათა **შევაფასოთ რა კანონზომიერება უდევს საფუძვლად რეზულტატური ცვლადისათვის მნიშვნელობების მინიჭებას/მის ვარიაციას.**

ფაქტორების გამოყოფა **შინაარსობრივი ანალიზის** საფუძველზე ხდება.

რადგან ცვლადების შინაარსობრივი ანალიზი და რეზულტატურ ცვლადზე მათი **შესაძლო გავლენის შეფასება** კვლევის დასაწყისში **პრობლემურია,** ამიტომ, არსებითი ფაქტორების გამოყოფისას ხშირად იყენებენ თანმიმდევრული რეგრესიის მეთოდს.

თანმიმდევრული რეგრესიის მეთოდის იდეა ისაა, რომ თანმიმდევრულად განვიხილოთ ფაქტორების ყველა შესაძლო კომბინაცია და ყოველი კომბინაციისა და მასში ჩართული თითოეული ფაქტორის მნიშვნელობა სტატისტიკურ მახასიათებლებზე დაყრდნობით, ინდივიდუალურად შევაფასოთ.

ეს ამოცანა სტატისტიკური კომპიუტერული პაკეტების (მაგალითად, **SPSS**-ის) მეშვეობით შეიძლება გადაწყდეს.

❖ რეგრესიის მოდელი

რეზულტატური ცვლადი ემპირიულ მნიშვნელობებს ($Y_{ემპ}$) მასზე **კანონზომიერი** და **შემთხვევითი** მიზეზების **ერთობლივი** ზემოქმედების შედეგად იღებს. აქედან გამომდინარე, იგი **ორ შემადგენელ ცვლადად** შეიძლება დავშალოთ:

$$Y_{ემპ} = \hat{Y} + \varepsilon \quad (*)$$

განტოლება (*) რეგრესიის მოდელის ზოგად სახეა.

ჯამის პირველი ცვლადი ($\hat{Y}$) რეზულტატური ცვლადის კანონზომიერი ნაწილია.

იგი მნიშვნელობებს ფაქტორების ჯგუფის კანონზომიერი ზემოქმედების ანუ **რეგრესიული დამოკიდებულების** შედეგად იღებს.

ამიტომ, შესაძლებელია, რომ $\hat{Y}$-ის მნიშვნელობები **გარკვეული წესით**/განტოლებით გამოვთვალოთ **ფაქტორების მნიშვნელობებიდან.**

$\hat{Y}$-ის მნიშვნელობებს რეზულტატური ცვლადის **რეგრესიული მნიშვნელობები** ეწოდება.

ვთქვათ, **რეზულტატური ცვლადია** "საკუთარი სამსახურით კმაყოფილება", ხოლო **ფაქტორების ჯგუფი** შედგება ორი ცვლადისაგან: "ხელფასი" და "დაკავებული თანამდებობის პრესტიჟულობა".

დავუშვათ:

- რეზულტატური ცვლადი იზომება პროცენტებით (0% ნიშნავს, რომ რესპონდენტი საკუთარი სამსახურით სრულიად უკმაყოფილოა, ხოლო 100% ნიშნავს, რომ რესპონდენტი საკუთარი სამსახურით სავსებით კმაყოფილია);
- ხელფასი იზომება ლარებით;
- დაკავებული თანამდებობის პრესტიჟულობა კი იზომება ათქულიანი სკალით (1 ქულა ნიშნავს, რომ რესპონდენტი საკუთარ თანამდებობას ყველაზე ნაკლებპრესტიჟულად მიიჩნევს, ხოლო 10 ქულა ნიშნავს, რომ რესპონდენტი საკუთარ თანამდებობას უმაღლესი პრესტიჟის მქონედ მიიჩნევს).

რეგრესიული დამოკიდებულება ანუ ხელფასისა და დაკავებული თანამდებობის პრესტიჟულობის **კანონზომიერი ზემოქმედება** საკუთარი სამსახურით კმაყოფილებაზე, **შეიძლება გამოიხატოს განტოლებით**, რომლის მეშვეობითაც გამოითვლება:

რამდენი პროცენტით **უნდა იყოს** კმაყოფილი საკუთარი სამსახურით ადამიანი, თუ ის, მაგალითად, თვეში 200 ლარს იღებს და მის მიერ დაკავებული თანამდებობის პრესტიჟულობას 6 ქულით აფასებს.

რეზულტატური ცვლადის რეგრესიული მნიშვნელობების გამოთვლა, ფაქტობრივად, ნიშნავს იმის **პროგნოზირებას/"წინასწარმეტყველებას"**, რა მნიშვნელობებს **მიიღებდა ან მიიღებს** რეზულტატური ცვლადი **მხოლოდ მოცემული რეგრესიული დამოკიდებულების შედეგად. ე. ი. თუ თეორიულად გამოვრიცხავთ, რომ იგი რაიმე სხვა სახის ზემოქმედებას განიცდის.**

მეორე ცვლადი (ε) რეზულტატური ცვლადის არაკანონზომიერი — შემთხვევითი ნაწილია.

შემთხვევითი მიზეზების გათვალისწინება შეუძლებელია. ამიტომ მათი ზემოქმედება **"წინასწარმეტყველების" გაუთვალისწინებელ/შემთხვევით "შეცდომას" იწვევს**: რეგრესიული მნიშვნელობები, რომლებსაც რეზულტატური ცვლადი **კანონზომიერად უნდა იღებდეს**, შეიძლება **არ ემთხვეოდეს ემპირიულს** ანუ იმ მნიშვნელობებს, რომლებსაც ის **სინამდვილეში** იღებს.

მაგალითად, შესაძლებელია, რომ თვეში 200 ლ. ხელფასის მქონე ადამიანი, რომელიც მის მიერ დაკავებული თანამდებობის პრესტიჟულობას 6 ქულით აფასებს, რეგრესიული დამოკიდებულების შედეგად საკუთარი სამსახურით 40%-ით კმაყოფილი **უნდა იყოს, მაგრამ მას პირადი კონფლიქტი** აქვს ადმინისტრაციის ერთ-ერთ წარმომადგენელთან და ამიტომ მხოლოდ 30%-ითაა კმაყოფილი.

შემთხვევითი მიზეზის გამო (პირადი კონფლიქტი) ჩვენი პროგნოზი/"წინასწარმეტყველება" ემპირიულ მონაცემთან მიმართებაში[1] **არაზუსტი**, "მცდარი" აღმოჩნდება.

აქედან გამომდინარე, **ε-ს რეგრესიის ცდომილებას** (ან შემთხვევით ცდომილებას) უწოდებენ.

რეგრესიის ცდომილების მნიშვნელობები რეზულტატური ცვლადის ემპირიულ და რეგრესიულ მნიშვნელობებს შორის სხვაობით გამოითვლება:

$$\varepsilon = Y_{\text{ემპ}} - \hat{Y}$$

ჩვენს მიერ განხილული მაგალითის შემთხვევაში, **ε = 10%.**

ε-ის მნიშვნელობათა განაწილება სტანდარტულ ნორმალურ განაწილებას უნდა წარმოადგენდეს.

თუ რეგრესიის ცდომილების მნიშვნელობათა **განაწილება** ამ პირობას არ აკმაყოფილებს, სავარაუდოა, რომ რეზულტატურ ცვლადზე ზემოქმედებს **არსებითი ფაქტორი/ფაქტორების ერთობლიობა**, რომელიც ფაქტორების ჯგუფში არ არის შეტანილი და ამიტომ, მისი გავლენა ε-ის განაწილებაზე აისახება.

[1] ანუ "ნამდვილ", გაზომვის შედეგად მიღებულ მონაცემთან მიმართებაში.

ასეთი მოდელის ფარგლებში კი რეზულტატური ცვლადის ვარიაციის ახსნის ცდა მიზანშეწონილი არ არის, რადგან რეზულტატურ ცვლადზე **კანონზომიერი ფაქტორის/ფაქტორების მოქმედების ეფექტი შემთხვევითი ცდომილების ეფექტად ჩაითვლება, რაც არაადეკვატურ შედეგებს მოგვცემს.**

❖ რეგრესიის წრფივი მოდელი

რეგრესიული მნიშვნელობები (Ŷ) შეიძლება გამოვთვალოთ **წრფივი ან არაწრფივი**[1] **სახის განტოლებით.**

შესაბამისად, განასხვავებენ რეგრესიის წრფივ და არაწრფივ მოდელებს.

რეგრესიის ზოგადი მოდელის განმარტებიდან გამომდინარე ($Y_{ემპ} = Ŷ + ε$), რეგრესიის წრფივი მოდელი ეწოდება რეზულტატური ცვლადის ემპირიული მნიშვნელობების გამოთვლის წესს/ფორმულას, რომელშიც რეგრესიული მნიშვნელობები (Ŷ) გამოითვლება ფაქტორების მნიშვნელობებიდან წრფივი განტოლებით:

$$Y_{ემპ} = B_0 + B_1X_1 + B_2X_2 + ... + B_kX_k + ε \qquad (1)$$

სადაც $Y_{ემპ}$ რეზულტატური ცვლადია, X_1, X_2 ,..., X_k ფაქტორული ცვლადები, B_0, B_1 ,..., B_k — მუდმივები, რომლებსაც **რეგრესიის კოეფიციენტები** ეწოდება, ε — **რეგრესიის ცდომილება,** ხოლო Ŷ გამოითვლება წრფივი განტოლებით, რომელსაც **რეგრესიის განტოლება** ეწოდება და შემდეგი ზოგადი ფორმით აღიწერება:

$$Ŷ = B_0 + B_1X_1 + B_2X_2 + ... + B_kX_k \qquad (2)$$

ჩვენს მიერ განხილული მაგალითის შემთხვევაში:

$$Ŷ = B_0 + B_1X_1 + B_2X_2$$

სადაც Ŷ საკუთარი სამსახურით კმაყოფილების ხარისხია, X_1 ხელფასის რაოდენობა, ხოლო X_2 დაკავებული თანამდებობის პრესტიჟულობა (k=2).

განტოლება მოცემულად მხოლოდ მაშინ ითვლება, როდესაც განსაზღვრულია მასში შემავალი მუდმივების მნიშვნელობები.

იმისათვის, რომ, მაგალითად, X_1 (ხელფასი) და X_2 (დაკავებული თანამდებობის პრესტიჟულობა) ცვლადების მნიშვნელობების მიხედვით, Ŷ-ის (საკუთარი სამსახურით კმაყოფილების ხარისხი) გამოთვლა შეძლოთ, B_0, B_1 და B_2 მნიშვნელობები უნდა ვიცოდეთ.

აქედან გამომდინარე, რეგრესიის მოდელის აგების მომდევნო საფეხურს (განტოლების სახის შერჩევის შემდეგ) — რეგრესიის კოეფიციენტების მნიშვნელობების პოვნაა.

რეგრესიის კოეფიციენტების **ზუსტი** მნიშვნელობების პოვნა მხოლოდ გენერალური ერთობლიობის სრული შესწავლით შეიძლება მოხდეს. **შერჩევითი ერთობლიობის შემთხვევაში შესაძლებელია მათი შეფასება**[2] **ამ ერთობლიობის ფარგლებში ანუ** შეფასებითი მნიშვნელობების/შეფასებების (აღვნიშნოთ ისინი b_0, b_1 ,..., b_k-თი) გამოთვლა.

---

[1] ლოგარითმული, ხარისხობრივი და სხვა.

[2] გარკვეული ალბათობით.

უმცირეს კვადრატთა მეთოდი

რეგრესიის კოეფიციენტების შეფასებების გამოსათვლელად გამოიყენება **უმცირეს კვადრატთა მეთოდი.**

ამ მეთოდის იდეა შემდეგშია:

რაც უფრო უახლოვდება რეზულტატური ცვლადის **რეგრესიული და ემპირიული მნიშვნელობები ერთმანეთს,** მით **უფრო ადეკვატურია/ზუსტია** რეგრესიის მოდელი.

შესაბამისად, მოდელის აგებისას, საუკეთესო ვარიანტია, $b_0, b_1, \ldots, b_k$ ისე შევარჩიოთ, რომ რეგრესიის განტოლებით გამოთვლილი რეზულტატური ცვლადის მნიშვნელობები **მაქსიმალურად უახლოვდებოდეს** მის ემპირიულ მნიშვნელობებს ანუ რეგრესიის ცდომილება **მინიმალური** იყოს.

ამიტომ $b_0, b_1, \ldots, b_k$-ს გამოსათვლელად შეიძლება გამოვიყენოთ რეგრესიის ცდომილების ($\varepsilon = y_i - \hat{y}_i$) კვადრატების ჯამის მინიმიზაცია:

$$\sum_{i=1}^{n} (y_i - b_0 - b_1x_i^1 - b_2x_i^2 - \ldots - b_kx_i^k)^2 \rightarrow \min$$

რის შედეგადაც მივიღებთ წრფივ განტოლებათა სისტემას, რომელშიც ცვლადებს $b_0, b_1, b_2, \ldots, b_k$ ნარმოადგენს. ე. ი. მისი ამოხსნა $b_0, b_1, b_2, \ldots, b_k$ მნიშვნელობებს ანუ რეგრესიის კოეფიციენტების შეფასებებს მოგვცემს.

ცხადია, რაც უფრო დიდია n (შერჩევითი ერთობლიობის მოცულობა), მით უფრო ზუსტია შეფასებები.

$b_0, b_1, \ldots, b_k$-ს შეტანით რეგრესიის განტოლების ზოგად ფორმაში/განტოლება (2), მივიღებთ **რეგრესიის კონკრეტულ განტოლებას, რომელიც მოცემულ რეგრესიულ დამოკიდებულებას აღწერს ანუ გამოხატავს ფაქტორების მნიშვნელობების მიხედვით რეზულტატური ცვლადის რეგრესიული მნიშვნელობების გამოთვლის წესს:**

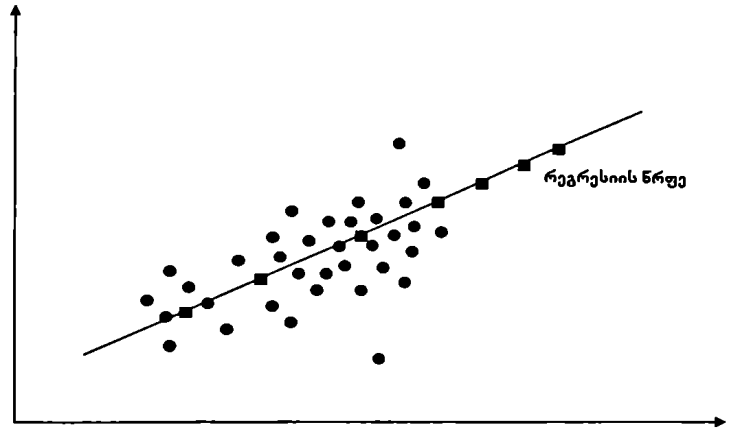
$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k \quad (3)$$

რეგრესიის წრფივი მოდელის გრაფიკული გამოსახვა/სკატერგრამა

რეგრესიის წრფივ მოდელში რეგრესიული მნიშვნელობები წრფივი განტოლებით გამოითვლება, ამიტომ, გრაფიკული გამოსახვისას ისინი ერთ წრფეზე დალაგდება.

წრფეს, რომელზეც რეგრესიული მნიშვნელობებია დალაგებული, **რეგრესიის წრფე ეწოდება.**

ემპირიული მნიშვნელობები კი რეგრესიის წრფის მიმართ სხვადასხვა პოზიციაზე იქნება განლაგებული/"გაფანტული":



—■ რეგრესიული მნიშვნელობები.
—● ემპირიული მნიშვნელობები.

დეკარტის საკოორდინატო სისტემაში მიმოფანტული/მიმობნეული წერტილების სახით წარმოდგენილ გრაფიკს **სკატერგრამა**[1] ეწოდება.

❖ რეგრესიული დამოკიდებულების სტატისტიკური მნიშვნელოვნების შეფასება

რეგრესიის კოეფიციენტების გამოთვლაში შემთხვევითი სიდიდეები მონაწილეობს. ამიტომ, თვითონ რეგრესიის კოეფიციენტებიც შემთხვევითი სიდიდეებია. აქედან გამომდინარე, იბადება კითხვა:

რის მიხედვით შეიძლება ვიმსჯელოთ იმის შესახებ, რომ **რეგრესიის განტოლება ცვლადებს შორის რეალურად არსებულ კანონზომიერ დამოკიდებულებას ასახავს** — რა გარანტიაა, რომ **გენერალურ ერთობლიობაში ყველა კოეფიციენტი ნულოვანი** არაა და მათი არანულოვანი შეფასებები მონაცემთა **შემთხვევითი გადახრის/ვარიაციის** შედეგი არ არის[2]?

ამ კითხვაზე პასუხის გასაცემად, ბუნებრივია, ჯერ რეზულტატური ცვლადის ვარიაცია უნდა გავაანალიზოთ.

ემპირიული მნიშვნელობების შესაბამისად, **რეზულტატური ცვლადის ვარიაციაც არაერთგვაროვანია**:

- მისი ერთი ნაწილი **რეგრესიული მნიშვნელობების** მიხედვით განისაზღვრება, ესე იგი, **ცვლადებს შორის რეგრესიული დამოკიდებულებით აიხსნება**;
- მეორე ნაწილის განსაზღვრა კი **რეგრესიის ცდომილების** მიხედვით ხდება. ამიტომ, ცვლადებს შორის რეგრესიული დამოკიდებულებით ვერ აიხსნება/**აუხსნელი რჩება**.

შესაბამისად, **საშუალო მნიშვნელობიდან ემპირიული მნიშვნელობების კვადრატული**

---

[1] ინგ.: **scatter** — გაფანტვას/მიმობნევა.

[2] **ნულოვანი კოეფიციენტების შემთხვევაში, $\gamma=0$ და $Y_{\text{ას}}=\varepsilon$,** ე. ი. რეზულტატური ცვლადის მნიშვნელობები მთლიანად შემთხვევითი აღმოჩნდება.

გადახრის (ანუ გადახრათა კვადრატების) ჯამი ორი შესაკრების სახით შეიძლება წარმოვადგინოთ:

$$SS_{მთ} = SS_{რეგ} + SS_{ნაშ}$$

სადაც:

$SS_{მთ} = \sum(y_{ემპ} - y_{საშ})^2$ არის რეზულტატური ცვლადის საშუალო მნიშვნელობიდან ემპირიული მნიშვნელობების გადახრათა კვადრატების ჯამი.

$SS_{მთ}$ შეიძლება განვიხილოთ, როგორც რეზულტატური ცვლადის მთლიანი ანუ კანონზომიერი და შემთხვევითი მიზეზების ერთობლივი მოქმედების შედეგად გამოწვეული ვარიაციის მახასიათებელი სიდიდე. ამიტომ აღნიშვნისას ინდექსად მივუთითოთ "მთ" ("მთლიანი").

$SS_{რეგ} = \sum(\hat{y}_i - y_{საშ})^2$ არის რეზულტატური ცვლადის საშუალო მნიშვნელობიდან რეგრესიული მნიშვნელობების გადახრათა კვადრატების ჯამი.

ეს გადახრები ცვლადებს შორის რეგრესიული დამოკიდებულებით აიხსნება. შესაბამისად, $შშ_{რეგ}$ შეიძლება განვიხილოთ როგორც რეზულტატური ცვლადის კანონზომიერი/რეგრესიული ვარიაციის მახასიათებელი სიდიდე. ამიტომ, აღნიშვნისას ინდექსად მივუთითოთ "რეგ" ("რეგრესიული").

$SS_{ნაშ} = \sum(y_{ემპ} - \hat{y}_i)^2$ — არის რეზულტატური ცვლადის ემპირიული მნიშვნელობებიდან რეგრესიული მნიშვნელობების გადახრათა ანუ რეგრესიის ცდომილების კვადრატების ჯამი.

ეს გადახრები ცვლადებს შორის რეგრესიული დამოკიდებულებით ვერ იხსნება. ამ გაგებით, ისინი რეგრესიის აუხსნელ "ნაშთად" რჩება. ამიტომ, მათი კვადრატების ჯამს რეგრესიის ნაშთი ვუწოდოთ და აღნიშვნისას ინდექსად მივუთითოთ "ნაშ" ("ნაშთი").

$შშ_{ნაშ}$ შეიძლება განვიხილოთ როგორც რეზულტატური ცვლადის შემთხვევითი ვარიაციის მახასიათებელი სიდიდე.

რეგრესიის სტატისტიკური მნიშვნელოვნების შეფასება რეზულტატური ცვლადის რეგრესიული და შემთხვევითი ვარიაციების მახასიათებლების საფუძველზე ხდება, სტატისტიკური ჰიპოთეზების შემოწმების მეთოდით.

მოწმდება ნულოვანი ჰიპოთეზა (H_0) ანუ, ვარაუდი იმის შესახებ, რომ რეგრესიის თითოეული კოეფიციენტი უდრის ნულს და ცვლადებს შორის რეგრესიული დამოკიდებულება არ არსებობს.

შემოწმება ხდება F (ფიშერის) სტატისტიკის მეშვეობით:

$$F = \frac{SS_{რეგ}}{SS_{ნაშ}} \left\{ \frac{k}{N - k - 1} \right\}$$

შემოწმების კრიტერიუმი:

ფიშერის სტატისტიკის განაწილების კანონის მიხედვით გამოთვლილია α დონის კრიტიკული მნიშვნელობები თავისუფლების ხარისხების შესაბამისად.

თუ ფიშერის სტატისტიკის ემპირიული მნიშვნელობა ($F_{ემპ}$) მეტია კრიტიკულზე ($F_{კრ}$), მაშინ ნულოვანი ჰიპოთეზა ფალსიფიცირდება: ითვლება, რომ რეზულტატურ ცვლადსა და ფაქტორების ჯგუფს შორის არსებობს მოცემული მოდელით გამოხატული წრფივი რეგრესიული დამოკიდებულება.

თუ $F_{ემპ}$ არ არის მეტი $F_{კრ}$-ზე, მაშინ ნულოვანი ჰიპოთეზა ვერიფიცირდება: ითვლება, რომ რეზულტატურ ცვლადსა და ფაქტორების ჯგუფს შორის არ არსებობს მოცემული მოდელით გამოხატული წრფივი რეგრესიული დამოკიდებულება.

❖ რეგრესიის მოდელის ხარისხის შეფასება

რეგრესიის მოდელი მრავალი სოციალური მოვლენის ცვალებადი ურთიერთქმედების შედეგს ასახავს.

- **ერთი და იგივე მოვლენა, სხვადასხვა პირობებში განსხვავებული მნიშვნელობის გავლენას შეიძლება ახდენდეს მეორე მოვლენაზე;**
- **აქედან გამომდინარე, რეზულტატური ცვლადი სხვადასხვა პირობებში განსხვავებულ ფაქტორებზე ან ერთსა და იმავე ფაქტორებზე, მაგრამ განსხვავებული კანონზომიერებით შეიძლება იყოს დამოკიდებული.**

მაგალითად, **განათლება და პროფესიონალიზმი შრომის ანაზღაურებაზე** სხვადასხვა სახელმწიფოში განსხვავებულ გავლენას ახდენს.

მოცემული ფაქტორების გავლენა შრომის ანაზღაურებაზე, ჩვეულებრივ, **მაღალი მნიშვნელობით გამოირჩევა** იმ სახელმწიფოებში, რომლებიც **დემოკრატიულობისა** და, შესაბამისად, **შრომითი უფლებების დაცვის მაღალი დონით** ხასიათდება.

არადემოკრატიულ ან შრომითი უფლებების დაცვის დაბალი დონის მქონე სახელმწიფოებში კი **შრომის ანაზღაურებას** უფრო მეტად სხვა ფაქტორები: **პარტიულობა, კორუფცია, პრო-ტექცია** და ა. შ. შეიძლება განსაზღვრავდეს, რაც შრომის ანაზღაურების თვალსაზრისით, **განათლებისა და პროფესიონალიზმის მნიშვნელობის დევალვაციას/გაუფასურებას უწყობს ხელს.**

გარდა ამისა, **ზოგად კანონზომიერებათა პარალელურად,** შრომის ანაზღაურებაზე განათლებისა და პროფესიონალიზმის გავლენის ხვედრითი წილი (მნიშვნელობის ხარისხი), როგორც წესი, ყოველ სახელმწიფოში **ინდივიდუალურია.**

ამ მსჯელობიდან ნათელია, რომ **უნივერსალური და სრულყოფილი რეგრესიის მოდელის აგების მექანიზმი არ არსებობს.** ამიტომ, რეგრესიის მოდელის **შერჩევა** ხდება მკვლევრის ან მკვლევართა ჯგუფის მიერ, **სიტუაციური ანალიზისა**[1] და **შინაარსობრივი** არგუმენტაციის გამოყენებით.

შესაბამისად, რეგრესიის ნებისმიერი მოდელი **ჰიპოთეტური** ხასიათისაა და **ალტერნატიული მოდელების არსებობას არ გამორიცხავს.**

ეს დებულება პირველ რიგში შემდეგი პრობლემის წინაშე გვაყენებს:

- თუ ერთი და იგივე რეზულტატური ცვლადისათვის რეგრესიის სხვადასხვა მოდელი შეიძლება არსებობდეს, როგორ შევადაროთ ისინი ერთმანეთს, რის მიხედვით გავაკეთოთ არჩევანი მათ შორის?

ამასთან, ამ პრობლემის წარმატებით გადაწყვეტა, ე. ი. ერთი და იგივე რეზულტატური ცვლადისათვის რეგრესიის სხვადასხვა მოდელის ერთმანეთთან შედარების კრიტერიუმის პოვნა საკითხის სირთულეს არ ამოწურავს:

- "უკეთესის" ძებნა უსასრულოდ შეიძლება გაგრძელდეს. რის მიხედვით ვიმსჯელოთ იმის შესახებ, რომ **ალტერნატიულ მოდელებს შორის საუკეთესო** კვლევის ამოცანებისათვის დამაკმაყოფილებელია და მასზე უკეთესის ძებნა აღარ არის საჭირო?
- რის მიხედვით ვიმსჯელოთ იმის შესახებ, რომ შერჩეული მოდელი კვლევის ამოცანებისათვის დამაკმაყოფილებელია, თუ მისი ალტერნატიული მოდელის მონახვა არ ხერხდება?
- შეიძლება თუ არა, ხარისხის მიხედვით ერთმანეთს შევადაროთ სხვადასხვა რეზულტატური ცვლადისათვის აგებული რეგრესიის მოდელები?

[1] ე. ი. იმ კონკრეტული სიტუაციის/პირობების ანალიზით, რომლებიც საკვლევ პრობლემაზე აისახება.

ჩამოთვლილი კითხვები **რეგრესიის მოდელის ხარისხობრივი დახასიათების/შეფასების** აუცილებლობას იწვევს.

რეგრესიის მოდელის ხარისხის ერთ-ერთ მახასიათებელი მისი **სტატისტიკური მნიშვნელოვნებაა**: მოდელი, რომელიც სტატისტიკურად მნიშვნელოვანი არ არის, როგორც ვნახეთ, **უკუიგდება, ე. ი. უხარისხოდ ითვლება.**

მაგრამ **როგორ განვსაზღვროთ სტატისტიკურად მნიშვნელოვანი მოდელის** ხარისხი და როგორ შევადაროთ იგი სხვა, **ასევე სტატისტიკურად მნიშვნელოვან** მოდელებს?

რეგრესიის მოდელის ხარისხი შემდეგი არსებითი მახასიათებლების მიხედვით შეიძლება განისაზღვროს:

1. ფაქტორების ჯგუფის მოცულობა/ფაქტორების რაოდენობა;
2. რეზულტატური ცვლადის ახსნილი ვარიაციის ხვედრითი წილი;
3. კორელაცია რეზულტატური ცვლადის რეგრესიულ და ემპირიულ მნიშვნელობებს შორის.

თითოეულ მათგანს სპეციფიკური შინაარსობრივი დატვირთვა აქვს:

1. რაც უფრო ნაკლებ მიზეზზე **"დავიყვანთ"** შესასწავლი მოვლენის ახსნას, მით უფრო **მნიშვნელოვანი/ძლიერი** იქნება მოკვლეული მიზეზები და **ნათელი** — მათი მოქმედების კანონზომიერება;

შესაბამისად, ითვლება, რომ რაც ნაკლებია ფაქტორების რიცხვი, მით მაღალია რეგრესიის მოდელის ხარისხი.

2. რაც უფრო მეტია რეზულტატური ცვლადის ვარიაციის ახსნილი ნაწილი, მით უკეთ შევძლებთ შესასწავლი მოვლენის დახასიათებას;

შესაბამისად, ითვლება, რომ რაც მეტია რეზულტატური ცვლადის ახსნილი ვარიაციის ხვედრითი წილი, მით მაღალია რეგრესიის მოდელის ხარისხი;

3. რეგრესიული ანალიზი პროგნოზირების საშუალებას იძლევა და რიგ შემთხვევაში სწორედ ამ მიზნით იყენებენ.

რაც უფრო **ზუსტია რეგრესიული მნიშვნელობები**, ე. ი. რაც უფრო **ნაკლებ გადახრილია** ისინი რეზულტატური ცვლადის ემპირიული მნიშვნელობებიდან, მით **უფრო სანდოა რეგრესიის პროგნოზი.**

რეგრესიული მნიშვნელობების ემპირიული მნიშვნელობებიდან გადახრა კი მათ შორის კორელაციით შეიძლება დახასიათდეს: რაც უფრო ძლიერია კორელაცია რეზულტატური ცვლადის რეგრესიულ და ემპირიულ მნიშვნელობებს შორის, მით უფრო ნაკლებია გადახრა, ე. ი. მით უფრო ზუსტია/სანდოა რეგრესიის მოდელი.

შესაბამისად, ითვლება, რომ რაც უფრო ძლიერია კორელაცია რეზულტატური ცვლადის რეგრესიულ და ემპირიულ მნიშვნელობებს შორის, მით უფრო მაღალია რეგრესიის მოდელის ხარისხი.

რეგრესიის მოდელის ხარისხის განსაზღვრა და ინტერპრეტაცია ხდება სტატისტიკური მაჩვენებლებით — კოეფიციენტებით, რომლებიც რეგრესიის არსებითი მახასიათებლების მნიშვნელობებს ასახავს.

I. რეგრესიის სტანდარტული ცდომილება.

რეგრესიის სტანდარტული ცდომილება — S გამოითვლება ფორმულით:

$$S = \sqrt{(\sum(Y_{\text{ემპ}} - \hat{Y})^2)/(N - k - 1)}$$

ფესქვეშ მოქცეული გამოსახულება ე. ი.

$$S^2 = (\Sigma(Y_{\text{ემპ}} - \hat{Y})^2)/(N - k - 1)$$

რეგრესიის ცდომილების დისპერსიას გამოსახავს.

მივაქციოთ ყურადღება, რომ რაც ნაკლებია განსხვავება რეზულტატური ცვლადის რეგრესიულ და ემპირიულ მნიშვნელობებს შორის, მით ნაკლებია ჯამის თითოეული წევრის მრიცხველი და რაც ნაკლებია ფაქტორების რიცხვი (k) მით მეტია ჯამის თითოეული წევრის მნიშვნელი.

ე. ი. რაც **უფრო ზუსტია რეგრესიული** მნიშვნელობები და რაც ნაკლებია **ფაქტორების რაოდენობა,** მით მცირეა **რეგრესიის სტანდარტული ცდომილება.**

ამიტომ, ეს მაჩვენებელი/კოეფიციენტი რეგრესიის მოდელის მახასიათებლად შეიძლება გამოდგეს. მისი მეშვეობით შესაძლებელია ერთმანეთს შევადაროთ ერთი და იგივე რეზულტატური ცვლადისათვის აგებული **ალტერნატიული მოდელები:**

თუ ერთი მოდელის რეგრესიის სტანდარტული ცდომილება ნაკლებია მეორე მოდელის სტანდარტულ ცდომილებაზე, მაშინ პირველი მოდელის ხარისხი უკეთესია/უფრო მაღალია, ვიდრე მეორე მოდელის.

II. დეტერმინაციისა და მრავლობითი კორელაციის კოეფიციენტები

დეტერმინაციის კოეფიციენტი — R^2 ეწოდება რეზულტატური ცვლადის რეგრესიული მნიშვნელობების ვარიაციის ხვედრითი წილის მაჩვენებელს მთლიან ვარიაციასთან ფარდობაში:

$$R^2 = 1 — SS_{\text{ცდ}}/SS_{\text{ემ}}$$

იგი შეფასებითი სიდიდეა და გვიჩვენებს, თუ მთლიანი ვარიაციის მერამდენედი აიხსნა რეგრესიის მოდელის მეშვეობით.

დეტერმინაციის კოეფიციენტი ნულსა და ერთს შორის მოთავსებული ნორმირებული სიდიდეა. შესაბამისად, საშუალებას გვაძლევს, რომ რეგრესიის მოდელი დავახასიათოთ **დამოუკიდებლად,** მისივე (ე. ი. R^2-ის) **რეალური და ზღვრული (0&1) მნიშვნელობების თანაფარდობის მიხედვით, სხვა მოდელებთან** შედარების გარეშე:

რაც მეტია ნულიდან გადახრა, მით უფრო მაღალია ახსნილი ვარიაციის ხვედრითი წილი, შესაბამისად, მოდელის ხარისხი.

შინაარსობრივი ინტერპრეტაციის გაიოლების მიზნით, მას ზოგჯერ პროცენტული სახით გამოხატავენ.

დავუშვათ, ჩვენს მიერ განხილული მაგალითის შემთხვევაში, R^2 = 82%. ეს იმას ნიშნავს, რომ საკუთარი სამსახურით კმაყოფილების **82%** ხელფასის რაოდენობითა და დაკავებული თანამდებობის პრესტიჟულობით **იხსნება.**

თუ რეგრესიის მოდელში რეზულტატური და მხოლოდ ერთი ფაქტორული ცვლადი შედის, დეტერმინაციის კოეფიციენტი ორ ცვლადს შორის დამოკიდებულების ანუ კორელაციის **პირსონის R^2 კოეფიციენტს** ემთხვევა.

კვადრატულ ფესვს დეტერმინაციის კოეფიციენტიდან ($\sqrt{R^2}=R$) **მრავლობითი კორელაციის კოეფიციენტი** ეწოდება.

იგი რეზულტატური ცვლადის **რეგრესიულ და ემპირიულ მნიშვნელობებს შორის** დამოკიდებულების/კორელაციის სტატისტიკური მაჩვენებელია. შესაბამისად, **რეგრესიის**

მოდელის სიზუსტის/სანდოობის შეფასების საშუალებას გვაძლევს.

R ნორმირებული კოეფიციენტია. მისი ზღვრული მნიშვნელობებია 0 და ±1.

ხოლო მნიშვნელობები, რომლებსაც იგი 0-სა და ±1-ს შორის იღებს, ზღვრულ მნიშვნელობებთან შედარებით ინტერპრეტირდება: **რაც მეტია გადახრა ნულიდან ±1-საკენ, მით უფრო მაღალია დამოკიდებულების ხარისხი ე. ი. მით უფრო სანდოა/ზუსტია მოდელი, შესაბამისად, უფრო მაღალია მისი ხარისხი.**

კოეფიციენტის ნიშანი კორელაციის მიმართულებას შეესაბამება: უარყოფითი მნიშვნელობა უარყოფითი დამოკიდებულების მაჩვენებელია, დადებითი ნიშანი — დადებითი დამოკიდებულების.

III. ცალკეული ფაქტორის მნიშვნელობა რეზულტატური ცვლადის ვარიაციაში

კორელაციის ინტერპრეტაცია მონაცემთა ანალიზის ერთ-ერთი არსებითი პრობლემაა.

ძირითადი სირთულე ისაა, რომ იგი სტატისტიკურად მნიშვნელოვანი შეიძლება აღმოჩნდეს არა მხოლოდ მოცემული ორი ცვლადის უშუალო დამოკიდებულების შედეგად, არამედ, მათ **"ურთიერთობაში"** მესამე ცვლადის **"ჩარევის"** გამოც.

მაგალითად, გამორიცხული არაა, რომ კორელაცია ტრანზიტულად, გავლით გადაეცეს: თუ **A** ცვლადი სტატისტიკურად ძლიერ გავლენას ახდენს **B**-ზე, ხოლო **B** — **C**-ზე, ამან შეიძლება **A** და **C** ცვლადების მნიშვნელობებს შორის გადახრა ისე შეამციროს, რომ კორელაციის კოეფიციენტმა სტატისტიკურად მნიშვნელოვანი დამოკიდებულების არსებობა გვიჩვენოს.

ანდა, **A** და **C** ცვლადებს შორის კორელაციის კოეფიციენტის მიხედვით, სტატისტიკურად მნიშვნელოვანი დამოკიდებულება შეიძლება დაფიქსირდეს იმის გამო, რომ თითოეული მათგანი **B** ცვლადის ძლიერ ზეგავლენას განიცდის.

ამასთან, კორელაციის კოეფიციენტის მიხედვით ვერ ვიტყვით **მესამე ცვლადის მოქმედების ეფექტის გამორიცხვის შემთხვევაში**, მოცემულ ცვლადებს შორის დამოკიდებულება **ისევ სტატისტიკურად მნიშვნელოვანი დარჩება თუ არა**[1].

ეს პრობლემა რეგრესიული ანალიზის კონტექსტში განსაკუთრებულ მნიშვნელობას იძენს, რადგან ფაქტორები ინტერპრეტირდება, როგორც რეზულტატური ცვლადის ვარიაციის ძირითადი მიზეზები.

მიზეზ-შედეგობრივი კავშირების კვლევისას კი ერთ-ერთი პირველხარისხოვანი როლი ენიჭება **მიზეზების რანჟირებას იმის მიხედვით, თუ რამდენად ძლიერია თითოეული მათგანის გავლენა შედეგზე**.

ფაქტორები არა მხოლოდ რეზულტატურ ცვლადზე, არამედ ერთმანეთზეც შეიძლება ზემოქმედებდნენ.

როგორ განვსაზღვროთ მოცემული ფაქტორის გავლენა რეზულტატურ ცვლადზე ისე, რომ ჩვენი შეფასება რეალურ დამოკიდებულებას ასახავდეს და არა სხვა ფაქტორის/ფაქტორების "ჩარევის" შედეგად დამახინჯებულ, მოჩვენებით სურათს?

ამ ამოცანის გადაწყვეტა ბეტა კოეფიციენტებისა და ნაწილობითი კორელაციის კოეფიციენტებით ხდება.

[1] ე. ი. ვერ განვსაზღვრავთ, მოცემული კორელაცია ნამდვილია თუ "ცრუ"/მოჩვენებითი.

რეგრესიის სტანდარტიზებულ კოეფიციენტებს (β_0, β_1 ,..., β_k) **ბეტა კოეფიციენტები ან ბეტა წონები** ეწოდება.

ბეტა კოეფიციენტების მიხედვით ფასდება ცალკეული ფაქტორის წვლილი რეზულტატური ცვლადის მთლიან ვარიაციაში.

უშუალოდ რეგრესიის კოეფიციენტების შეფასებებს (b_0, b_1 ,..., b_k) ამ მიზნით[1] ვერ გამოვიყენებთ, რადგან ფაქტორები განსხვავებული ცვლადებია. მათი მნიშვნელობები, უმეტეს შემთხვევაში, განსხვავებული ერთეულებით იზომება. ამიტომ, ურთიერთიერთშედარებას არ ექვემდებარება.

ჩვენს მიერ განხილულ მაგალითში საკუთარი სამსახურით კმაყოფილება პროცენტებით არის გამოხატული, ხელფასი ლარებით, ხოლო დაკავებული თანამდებობის პრესტიჟულობა — 10 ქულიანი სკალით.

შესაბამისად, იმისათვის, რომ რეგრესიის კოეფიციენტების შეფასებების მიხედვით გავარკვიოთ **რომელი ფაქტორის** (ხელფასისა თუ დაკავებული თანამდებობის პრესტიჟულობის) **ცვლილება** იწვევს **უფრო მეტად რეზულტატური ცვლადის** (საკუთარი სამსახურით კმაყოფილების) **ცვლილებას**, ჯერ ის უნდა განვსაზღვროთ, რამდენი ლარი შეესაბამება 1 ქულა პრესტიჟს, რომელსაც თითოეული რესპონდენტი ინდივიდუალურად აფასებს. საკითხის ასე დაყენება კი აზრს მოკლებულია.

ეს განსხვავება შეიძლება ელიმინირდეს[2], თუ ცვლადების სტანდარტიზებას მოვახდენთ, რადგან სტანდარტული გადახრის ერთეული უნივერსალურია (საერთოა ყველა სტანდარტიზებული ცვლადისათვის).

ბეტა კოეფიციენტები გამოითვლება ფორმულით:

$$\beta_j = b_j(\sigma x_j/\sigma y)$$

სადაც β_j არის **j**-ური ბეტა კოეფიციენტი, b_j — რეგრესიის **j**-ური კოეფიციენტი, σx_j — **j**-ური ფაქტორის საშუალო კვადრატული გადახრა, ხოლო σy — რეზულტატური ცვლადის საშუალო კვადრატული გადახრა.

ბეტა კოეფიციენტების მიხედვით, შესაძლებელია ვიმსჯელოთ რეზულტატური ცვლადის ვარიაციაში **ცალკეული ფაქტორის წვლილზე/წონაზე**:

X_j ფაქტორის ($1 \le j \le k$) **სტანდარტული გადახრის ერთი ერთეულით ცვლილება**[3], **დანარჩენი ფაქტორების ფიქსირებული მნიშვნელობების შემთხვევაში**, რეგრესიულ მნიშვნელობას **სტანდარტული გადახრის β_j ერთეულით**[4] **ცვლის**.

შესაბამისად, თუ $\beta_i > \beta_j$, მაშინ რეზულტატური ცვლადის ვარიაციაში X_i ფაქტორის **წვლილი/წონა** უფრო მეტია, ვიდრე – X_j-ის.

ბეტა კოეფიციენტების მიხედვით, შესაძლებელია ვიმსჯელოთ იმის შესახებ, თუ **რამდენით და რამდენჯერ მეტია** რეზულტატური ცვლადის ვარიაციაში **ერთი ფაქტორის (X_i) მიერ შეტანილი წვლილი მეორესთან (X_j) შედარებით**.

---

[1] **ე. ი. რეზულტატური ცვლადის მთლიან ვარიაციაში ცალკეული ფაქტორის ხვედრითი წილის შეფასების მიზნით.**

[2] გამოირიცხოს.

[3] კლება ან მატება.

[4] ანუ, β_j კოეფიციენტის ტოლი სტანდარტული გადახრით ცვლილებას იწვევს.

საკუთრივი/ნაწილობითი კორელაციის კოეფიციენტი (r) საშუალებას გვაძლევს სტატისტიკურად შევაფასოთ და დავახასიათოთ რეზულტატური ცვლადის დამოკიდებულება ფაქტორების ჯგუფის თითოეულ წევრზე — თითოეულ ფაქტორულ ცვლადზე ისე, რომ ელიმინირდეს ნებისმიერი სხვა ფაქტორული ცვლადის ზეგავლენის ეფექტი.

დავუშვათ, Y რეზულტატური ცვლადია, X_1 და X_2 კი — ფაქტორები.

ნაწილობითი/საკუთრივი კორელაციის კოეფიციენტი $r_{yx_1 . x_2}$ გამოითვლება ფორმულით:

$$r_{yx_1 . x_2} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_2})(1 - r^2_{x_1x_2})}}$$

იგი **საკუთრივ/მხოლოდ** რეზულტატურ ცვლადსა (Y) და ერთ ფაქტორს (X_1) შორის დამოკიდებულების სტატისტიკური მახასიათებელია. მეორე ფაქტორის (X_2) მოქმედების ეფექტი **ელიმინირებულია — გამორიცხულია/გამოკლებულია**[1].

ნაწილობითი კორელაციის კოეფიციენტის გამოთვლა ფაქტორების მეტი რაოდენობისათვის ხდება მოცემული ფორმულის ტრივიალური განზოგადებით.

r ნორმირებული კოეფიციენტია. მისი ზღვრული მნიშვნელობებია 0 და ±1.

მნიშვნელობები, რომლებსაც იგი 0-სა და ±1-ს შორის იღებს, ზღვრულ მნიშვნელობებთან შედარებით ინტერპრეტირდება: **რაც მეტია გადახრა ნულიდან ±1-საკენ, მით უფრო მაღალია დამოკიდებულების ხარისხი.**

კოეფიციენტის ნიშანი კორელაციის მიმართულებას შეესაბამება: უარყოფითი მნიშვნელობა უარყოფითი დამოკიდებულების მაჩვენებელია, დადებითი ნიშანი — დადებითი დამოკიდებულების.

მულტიკოლინეურობა

ფაქტორებს შორის სტატისტიკური დამოკიდებულების საკითხს უკავშირდება მულტიკოლინეურობის პრობლემაც.

მულტიკოლინეურობას ადგილი აქვს მაშინ, როდესაც ფაქტორების ერთ წყვილს შორის მაინც არსებობს ნრფივი დამოკიდებულება.

მისი გამოვლენა სტატისტიკური კომპიუტერული პაკეტების (მაგალითად, **SPSS**-ის) მეშვეობით შეიძლება.

მულტიკოლინეურობა უარყოფითად აისახება რეგრესიის მოდელის სანდოობაზე, ართულებს შედეგების ინტერპრეტაციას. ამიტომ, მიზანშეწონილია, რომ ის რეგრესიის მოდელის აგებამდე გამოვავლინოთ და თავიდან ავიცილოთ. მაგალითად, ერთ-ერთი ფაქტორის უკუგდებით ან მნიშვნელობების ტრანსფორმაციით/განსხვავებული ფორმით წარმოდგენით და სხვა.

გაზომვის დონე

რეგრესიის მოდელი ფაქტორების მნიშვნელობების ზრდის/კლების მიხედვით რეზულტატური ცვლადის ზრდის/კლების ტენდენციას გამოხატავს.

ზრდადობა/კლებადობის ინტერპრეტაცია კი, გაზომვის ნომინალურ დონეზე აზრს მოკლებულია.

ამ თვალსაზრისით, გამონაკლისია **დიხოტომიური ნომინალური** ცვლადი, რადგან მას მხოლოდ ორი შესაძლო მნიშვნელობა აქვს და **ერთ-ერთი მნიშვნელობის შესაბამისი კოდის**

[1] **ამას საკუთრივი/ნაწილობითი კორელაციის კოეფიციენტის გამოთვლის წესი უზრუნველყოფს.**

ზრდა/კლება შეიძლება ინტერპრეტირდეს როგორც მეორე მნიშვნელობაზე გადასვლა.

ამიტომ, რეგრესიული ანალიზის გამოყენება გაზომვის ნომინალურ დონეზე არ ხდება, მაგრამ რეგრესიის განტოლებაში შეიძლება დიხოტომიური ნომინალური ცვლადი შედიოდეს.

დისპერსიული ანალიზი

❖ დისპერსიული ანალიზის ზოგადი დახასიათება

ცვლადის საშუალო მნიშვნელობა ხშირად ინტერპრეტირდება როგორც გენერალური ერთობლიობისათვის დამახასიათებელი მაჩვენებელი, "ნორმა". ვარიაცია კი, როგორც "ნორმიდან გადახრა".

რა ფაქტორები იწვევს რეზულტატური ცვლადის "ნორმიდან გადახრას"?

ამ კითხვაზე პასუხის გაცემა ანუ **რეზულტატური ცვლადის ვარიაციის ახსნა** სოციოლოგიურ მონაცემთა ანალიზის ერთ-ერთი არსებითი და რთული პრობლემაა. ეს პრობლემა სხვადასხვა სახის სტატისტიკური ანალიზის გამოყენებით შეიძლება გადაწყდეს[1]. მათ შორის ცენტრალური ადგილი უკავია დისპერსიულ ანალიზს.

დისპერსიული ანალიზი იმ მეთოდების ერთობლიობაა, რომელთა მეშვეობითაც შესაძლებელია **შევაფასოთ ცალკეული ფაქტორისა და ფაქტორების ურთიერთქმედების** ზეგავლენის მნიშველობა რეზულტატურ ცვლადზე.

შესაბამისად, იგი გამოიყენება **დამხმარე მეთოდის სახით** იმ სტატისტიკურ ანალიზებში, რომელთა ამოცანაც ცვლადის **ვარიაციის ახსნაა**[2].

მაგალითად, **რეგრესიის მოდელის ფარგლებში** რეზულტატური ცვლადის ვარიაციის აღწერა და მთლიან ვარიაციაში ცალკეული ფაქტორის წვლილის შეფასება სწორედ დისპერსიული ანალიზის გამოყენებით ხორციელდება.

დისპერსიული ანალიზის მეშვეობით შესაძლებელია რეზულტატური ცვლადის **მთლიანი ვარიაციის** დაშლა/დანაწევრება ერთმანეთისაგან **დამოუკიდებელ ნაწილებად**. ესენია:

- მთლიანი ვარიაციის ის ნაწილი, რომელსაც **თითოეული** მოცემული **ფაქტორის ინდივიდუალური** ზეგავლენა იწვევს;
- მთლიანი ვარიაციის ის ნაწილი, რომელსაც მოცემული **ფაქტორების ურთიერთქმედების** ზეგავლენა იწვევს;
- მთლიანი ვარიაციის ის ნაწილი, რომელსაც სხვა, შემთხვევითი/**შემთხვევითად მიჩნეული** ფაქტორების ზეგავლენა იწვევს.

ფაქტორების რაოდენობის მიხედვით, დისპერსიული ანალიზის სახეებია:

- **ერთფაქტორული** — როდესაც რეზულტატური ცვლადის ვარიაციაზე მხოლოდ ერთი ფაქტორის ზეგავლენას ვაანალიზებთ **(ANOVA[3])**;
- **მრავალფაქტორული** — როდესაც რეზულტატური ცვლადის ვარიაციაზე ორი ან მეტი ფაქტორის ზეგავლენას ვაანალიზებთ **(MANOVA[4])**.

გაზომვის დონე:

- **დისპერსიული ანალიზი ფაქტორების მნიშვნელობათა გრადაციების/ფიქსირე-**

[1] რეზულტატური ცვლადის ვარიაციის ასახსნელად სტატისტიკური ანალიზის კონკრეტული სახის შერჩევა გაზომვის დონისა და კვლევის ამოცანების გათვალისწინებით ხდება.

[2] მაგალითად, რეგრესიული ანალიზი, დისკრიმინანტული ანალიზი, ფაქტორული ანალიზი, კლასტერული ანალიზი, მრავალგანზომილებიანი სკალირება და სხვა.

[3] ANalysis Of VAriance.

[4] Maltiple ANalysis Of VAriance.

ბული მნიშვნელობების მიხედვით ხორციელდება და ამიტომ შესაძლებელია, რომ იგი გაზომვის ნომინალურ დონეზეც გამოვიყენოთ.

დისპერსიული ანალიზის ძირითადი შეზღუდვა:

- დისპერსიული ანალიზის გამოყენება ხდება იმ შემთხვევაში, როდესაც რეზულტატური ცვლადის განაწილება ნორმალურია.

❖ **ერთფაქტორული დისპერსიული ანალიზი შემთხვევითი შერჩევისას**

ერთფაქტორული დისპერსიის მოდელი აღინერება ფორმულით:

$$X = X_{საშ} + \alpha + \varepsilon$$

X რეზულტატური ცვლადია, $X_{საშ}$ — რეზულტატური ცვლადის საშუალო მნიშვნელობა, α — რეზულტატურ ცვლადზე მოცემული ფაქტორის ზეგავლენის შედეგად გამოწვეული ვარიაციის მაჩვენებელი, ხოლო ε — ყველა სხვა მიზეზით გამოწვეული რეზულტატური ცვლადის ვარიაციის მაჩვენებელი.

დავუშვათ, მოცემულია, რომ:

- რეზულტატური ცვლადი (X) ფაქტორის (Y) p გრადაციის (ფიქსირებული მნიშვნელობის) მიხედვით, განაწილებულია/დაჯგუფებულია p — ჯგუფად;
- კვლევის ამოცანაა ფაქტორის (Y) მნიშვნელობის გაანალიზება/შეფასება რეზულტატური ცვლადის (X) მთლიან ვარიაციაში.

მსჯელობის გამარტივების მიზნით, გაზომვისა და ერთგანზომილებიანი ანალიზის რეზულტატები წარმოვადგინოთ ცხრილის სახით:

ცხრილი №2.1

ჯგუფის №	რეზულტატური ცვლადის ემპირიულ მნიშვნელობათა ჯგუფები ფაქტორის თითოელი გრადაციის მიხედვით ↕ დისპერსიული ჯგუფები	ჯგუფის წევრების ჯამი	ჯგუფური საშუალო
1	$X_{11}, X_{21}, ..., X_{i1}, ..., Xn_1^1$	$\Sigma X_{i1}=T_1$	$X_{საშ1}=\Sigma X_{i1}/n_1$
...	...	...	...
j	$X_{1j}, X_{2j}, ..., X_{ij}, ..., Xn_j^j$	$\Sigma X_{ij}=T_j$	$X_{საშj}=\Sigma X_{ij}/n_j$
...	...	...	...
p	$X_{1p}, X_{2p}, ..., X_{ip}, ..., Xn_p^p$	$\Sigma X_{ip}=T_p$	$X_{საშp}=\Sigma X_{ip}/n_p$
ჯამი		$G=\Sigma\Sigma X_{ij}$	$X_{საშ}=\Sigma\Sigma X_{ij}/N$

რეზულტატური ცვლადის ემპირიულ მნიშვნელობათა ჯგუფებს, რომლებიც ფაქტორის გრადაციების მიხედვით გამოიყოფა, **დისპერსიული ჯგუფები** ენოდება.

- დისპერსიული **ჯგუფის ნომერი** განისაზღვრება ფაქტორის გრადაციების მიხედვით: j-ური ჯგუფი შეესაბამება ფაქტორის j-ურ გრადაციას/მნიშვნელობას.
- j-ური დისპერსიული **ჯგუფის წევრებია** რეზულტატური ცვლადის იმ მნიშვნელობებს, რომლებიც ემპირიულ განაწილებაში ფაქტორის j-ურ გრადაციას/მნიშვნელობას შეესაბამება.
- $\Sigma X_{ij}=T_j$ გამოსახავს j-ური დისპერსიული ჯგუფის წევრების ჯამს.
- $X_{\text{საშ}j}=\Sigma X_{ij}/n_j$ გამოსახავს j-ური დისპერსიული ჯგუფის წევრების საშუალო მნიშვნელობას. მას **ჯგუფური საშუალო**[1] ენოდება.
- **N** მთლიანი შერჩევითი ერთობლიობის მოცულობაა.
- $G=\Sigma\Sigma X_{ij}$ გამოსახავს რეზულტატური ცვლადის ყველა ემპირიული მნიშვნელობის ჯამს.
- $X_{\text{საშ}} = \Sigma\Sigma X_{ij}/N$ გამოსახავს რეზულტატური ცვლადის მთლიან საშუალო მნიშვნელობას.

მაგალითად, თუ:

- ფაქტორული ცვლადის (სტუდენტთა სპეციალობა) გრადაციებია: 1 — სოციოლოგი, 2 — პოლიტოლოგი, 3 — ფსიქოლოგი;
- რეზულტატური ცვლადის (თსუ-ში მიმდინარე სასწავლო რეფორმებით კმაყოფილება) შესაძლო მნიშვნელობებია: 1 — სავსებით კმაყოფილი, 2 — უფრო კმაყოფილი, ვიდრე უკმაყოფილო, 3 — კმაყოფილიც და უკმაყოფილოც (თანაბრად), 4 — უფრო უკმაყოფილო, ვიდრე კმაყოფილი, 5 — სრულიად უკმაყოფილო;
- სტუდენტთა საერთო რაოდენობაა 1000.

მაშინ ცხრილი №2.1 შემდეგ სახეს მიიღებს:

ჯგუფის №	რეზულტატური ცვლადის ემპირიული მნიშვნელობების ჯგუფები ფაქტორის თითოელი გრადაციის მიხედვით	ჯგუფის წევრების ჯამი	ჯგუფური საშუალო
1	2, 3 ,..., 5 ,..., 1, 4	750	2.5
2	1, 3 ,..., 4 ,..., 5, 1	1240	3.1
3	3, 2 ,..., 5 ,..., 1, 5	540	1.8
ჯამი		2530	$X_{\text{საშ}} = 2.5$

ფაქტორის თითოეული გრადაცია რეზულტატური ცვლადის ვარიაციის **განსხვავებული პირობაა.**

ჩვენი მაგალითის შემთხვევაში, სასწავლო რეფორმებით კმაყოფილების ხარისხის ვარიაცია სხვადასხვა პირობებში — სხვადასხვა სპეციალობის სტუდენტების მიხედვით ხდება. თითოეულ ამ პირობას ფაქტორის ერთი გრადაცია/ფიქსირებული მნიშვნელობა შეესაბამება: სოციოლოგი,

[1] ჯგუფისშიდა საშუალოს "პირობით საშუალოსაც" უწოდებენ, რადგან თითოეული ჯგუფი გამოიყოფა გარკვეული პირობის ანუ ფაქტორის გრადაციის შესაბამისად.

პოლიტოლოგი, ფსიქოლოგი.

ცხრილი №2.1-ის მიხედვით შესაძლებელია **ჯგუფთაშორისი და ჯგუფისშიდა ვარიაციების სტატისტიკური მახასიათებლების:** კვადრატული გადახრების ჯამების პოვნა.

- **ჯგუფთაშორისი ვარიაციის** სტატისტიკური მახასიათებელი გამოითვლება **ჯგუფებს "შორის",** რეზულტატური ცვლადის საშუალო მნიშვნელობიდან ჯგუფური საშუალოების გადახრების მიხედვით და განიხილება როგორც მოცემული ფაქტორის მოქმედების შედეგი.
- **ჯგუფისშიდა ვარიაციის** სტატისტიკური მახასიათებელი გამოითვლება **თითოეული ჯგუფის "შიგნით",** ჯგუფური საშუალოდან ჯგუფის წევრების გადახრების მიხედვით და განიხილება როგორც ყველა სხვა/**შემთხვევითად მიჩნეული** მიზეზების მოქმედების შედეგი.

შესაბამისად, რეზულტატური ცვლადის კვადრატული გადახრების ჯამი დაიშლება/დანაწევრდება ორ შესაკრებად: $SS_{ზ} = SS_{ფ} + SS_{შ}$, სადაც:

- $SS_{ზ}$ (მთლიანი) არის რეზულტატური ცვლადის ემპირიული მნიშვნელობების საშუალო მნიშვნელობიდან გადახრების კვადრატების ჯამი:

$$SS_{ზ} = \Sigma(X_{ij} - X_{საშ})^2$$

- $SS_{ფ}$ (ფაქტორული) არის რეზულტატური ცვლადის საშუალო მნიშვნელობიდან ჯგუფური საშუალოების გადახრების კვადრატების ჯამი ანუ ჯგუფთაშორისი ვარიაციის მახასიათებელი.

$$SS_{ფ} = \Sigma(X_{საშ\,j} - X_{საშ})^2$$

- $SS_{შ}$ (შემთხვევითი) არის ჯგუფური საშუალოებიდან ჯგუფის წევრების გადახრების კვადრატების ჯამი ანუ ჯგუფისშიდა ვარიაციის მახასიათებელი.

$$SS_{შ} = \Sigma(X_{ij} - X_{საშj})^2$$

ახდენს თუ არა, ფაქტორი სტატისტიკურად მნიშვნელოვან ზეგავლენას რეზულტატური ცვლადის მთლიან ვარიაციაზე?

ამ კითხვაზე პასუხის გასაცემად უნდა შევაფასოთ **სტატისტიკურად რამდენად მნიშვნელოვანია ფაქტორული დისპერსია.**

ფაქტორული დისპერსიის სტატისტიკური მნიშვნელოვნება კი **შემთხვევით დისპერსიასთან თანაფარდობით შეიძლება განისაზღვროს.**

თუ ფაქტორული დისპერსია აღემატება შემთხვევითს, ითელება, რომ რეზულტატური ცვლადის ვარიაცია დამოკიდებულია ფაქტორზე.

თუ ფაქტორული დისპერსია არ აღემატება შემთხვევითს, ითელება, რომ რეზულტატური ცვლადის ვარიაცია არ არის დამოკიდებული ფაქტორზე.

შეფასება ხდება სტატისტიკური ჰიპოთეზის შემოწმების მეთოდით.

ნულოვანი ჰიპოთეზა (H_0): ფაქტორული ვარიაცია არ აღემატება შემთხვევითს.

შემოწმება ხდება F (ფიშერის) სტატისტიკის მეშვეობით:

$$F = \frac{SS_{ფ}}{SS_{შ}} \left\{ \frac{p - 1}{N - p} \right\}$$

შემოწმების კრიტერიუმი:

ფიშერის სტატისტიკის განაწილების კანონის მიხედვით, გამოთვლილია α დონის კრიტიკული მნიშვნელობები თავისუფლების ხარისხების შესაბამისად.

თუ ფიშერის სტატისტიკის ემპირიული მნიშვნელობა ($F_{ემპ}$) მეტია კრიტიკულზე ($F_{კრ}$), მაშინ ნულოვანი ჰიპოთეზა ფალსიფიცირდება და ითვლება, რომ რეზულტატური ცვლადის ვარიაცია დამოკიდებულია ფაქტორზე.

თუ ($F_{ემპ}$) არ არის მეტი ($F_{კრ}$), მაშინ ნულოვანი ჰიპოთეზა ვერიფიცირდება და ითვლება, რომ რეზულტატური ცვლადის ვარიაცია არ არის დამოკიდებული ფაქტორზე.

❖ **ორფაქტორული დისპერსიული ანალიზი შემთხვევითი შერჩევისას**

ორფაქტორული დისპერსიის მოდელი აღიწერება ფორმულით:

$$X = X_{საშ} + \alpha + \beta + (\alpha \times \beta) + \varepsilon$$

სადაც **X** რეზულტატური ცვლადია, $X_{საშ}$ — რეზულტატური ცვლადის საშუალო მნიშვნელობა, α — რეზულტატურ ცვლადზე **A** ფაქტორის ზეგავლენის შედეგად გამოწვეული ვარიაციის მაჩვენებელი, β — რეზულტატურ ცვლადზე **B** ფაქტორის ზეგავლენის შედეგად გამოწვეული ვარიაციის მაჩვენებელი, $(\alpha\times\beta)$ — რეზულტატურ ცვლადზე **A** და **B** ფაქტორების ურთიერთქმედების ზეგავლენის შედეგად გამოწვეული ვარიაციის მაჩვენებელი, ხოლო ε — ყველა სხვა მიზეზით გამოწვეული ვარიაციის მაჩვენებელი.

შესაბამისად, რეზულტატური ცვლადის გადახრების კვადრატების ჯამი/მთლიანი ვარიაცია შემდეგ ნაწილებად შეიძლება დავშალოთ:

$$SS_{ჯ} = SS_{ფA} + SS_{ფB} + SS_{ფ(A\times B)} + SS_{შ}$$

მაგალითად, თუ რეზულტატური ცვლადია სასწავლო პროცესით კმაყოფილება (იზომება 5 ქულიანი სკალით. 1 მაქსიმალური, ხოლო 5 — მინიმალური შეფასებაა), ხოლო ფაქტორები სტუდენტთა სპეციალობა (1 — სოციოლოგი, 2 — პოლიტოლოგი, 3 — ფსიქოლოგი) და სტუდენტთა აკადემიური მოსწრება (1 — საშუალო, 2 — საშუალოზე მაღალი, 3 — მაღალი), დისპერსიული ანალიზი საშუალებას გვაძლევს შევაფასოთ სასწავლო პროცესით სტუდენტთა კმაყოფილების ვარიაციაში სტუდენტთა სპეციალობის ხვედრითი წილი, სტუდენტთა აკადემიური მოსწრების ხვედრითი წილი და სტუდენთა სპეციალობისა და აკადემიური მოსწრების ურთიერთდამოკიდებულების ხვედრითი წილი.

მსჯელობის გასამარტივებლად, გაზომვისა და ერთგანზომილებიანი ანალიზის რეზულტატები წარმოვადგინოთ ცხრილის სახით:

ცხრილი №2.2

B ფაქტორის გრადაცია	**A ფაქტორის გრადაცია**				**ჯამი**
	A_1	A_2	...	A_p	
B_1	$X_{11}^1, X_{11}^2, ..., X_{11}^n$	$X_{21}^1, X_{21}^2, ..., X_{21}^n$	...	$X_{p1}^1, X_{p1}^2, ..., X_{p1}^n$	ΣX_{i1}^k
B_2	$X_{12}^1, X_{12}^2, ..., X_{12}^n$	$X_{22}^1, X_{22}^2, ..., X_{22}^n$	...	$X_{p2}^1, X_{p2}^2, ..., X_{p2}^n$	ΣX_{i2}^k
...	...	...	...	...	...
B_q	$X_{1q}^1, X_{1q}^2, ..., X_{1q}^n$	$X_{2q}^1, X_{2q}^2, ..., X_{2q}^n$	...	$X_{pq}^1, X_{pq}^2, ..., X_{pq}^n$	ΣX_{iq}^k
ჯამი	ΣX_{1j}^k	ΣX_{2j}^k	...	ΣX_{pj}^k	ΣX_{ij}

- $A_1, A_2 \_ A_p$ ერთი (A) ფაქტორის გრადაციებია, ხოლო $B_1, B_2 \_ B_p$ მეორე (B) ფაქტორის გრადაციები;

- A_i და B_j გადაკვეთაზე ჩაწერილი დისპერსიული ჯგუფის წევრებია რეზულტატური ცვლადის ის მნიშვნელობები, რომლებიც ემპირიულ განაწილებაში **A** ფაქტორის **i**-ურ და **B** ფაქტორის **j**-ურ გრადაციებს შეესაბამება;
- ΣX_i გამოსახავს რეზულტატური ცვლადის ყველა იმ მნიშვნელობის ჯამს, რომლებიც **A** ფაქტორის **i**-ურ გრადაციას შეესაბამება, იმისგან დამოუკიდებლად, თუ **B** ფაქტორის რომელ გრადაციას შეესაბამება ის.
- **A** ფაქტორის ყოველი **i**-ური გრადაციის სვეტის ბოლოს შეჯამებულია რეზულტატური ცვლადის ის მნიშვნელობები, რომლებიც ამ გრადაციას შეესაბამება, იმისგან დამოუკიდებლად, თუ **B** ფაქტორის რომელ გრადაციას შეესაბამება თითოეული მათგანი;
- **B** ფაქტორის ყოველი **j**-ური გრადაციის სტრიქონის ბოლოს შეჯამებულია რეზულტატური ცვლადის ის მნიშვნელობები, რომლებიც ამ გრადაციას შეესაბამება, იმისგან დამოუკიდებლად, თუ **A** ფაქტორის რომელ გრადაციას შეესაბამება თითოეული მათგანი;
- ΣX_{ij} გამოსახავს თითოეული რეზულტატური ცვლადის ყველა მნიშვნელობის ჯამს.

ცხრილში წარმოდგენილი მონაცემების მიხედვით, შეიძლება გამოვთვალოთ:

1. რეზულტატური ცვლადის საშუალო მნიშვნელობა:
$$X_{საშ} = \Sigma X_{ijk}/pqn$$

2. საშუალო მნიშვნელობები **A** ფაქტორის ყოველი გრადაციისათვის:
$$T_{საშi} = \Sigma X_{ijk}/qn$$

3. საშუალო მნიშვნელობები **B** ფაქტორის ყოველი გრადაციისათვის:
$$T_{საშj} = \Sigma X_{ijk}/pn$$

4. საშუალო მნიშვნელობები ფაქტორების გრადაციების ყოველი წყვილისათვის:
$$X_{საშij} = \Sigma X_{ijk}/pq$$

ამ მაჩვენებლების მიხედვით, რეზულტატური ცვლადის მთლიანი ვარიაციის სტატისტიკური მახასიათებელი შემდეგი ჯამის სახით შეიძლება წარმოვადგინოთ:

$$(\Sigma X_{ijk} — X_{საშ})^2 = (T_{საშi} — X_{საშ})^2 + (T_{საშj} — X_{საშ})^2 + (X_{საშij} — T_{საშi} — T_{საშj} + X_{საშ})^2 + (\Sigma X_{ijk} — X_{საშij})^2$$

მოცემული ჯამი ოთხი შესაკრებისაგან შედგება: პირველი ასახავს რეზულტატური ცვლადის ვარიაციაზე **A** ფაქტორის ზეგავლენის შედეგს, მეორე — **B** ფაქტორის ზეგავლენის შედეგს, მესამე **A** და **B** ფაქტორების ურთიერთქმედების ზეგავლენის შედეგს, ხოლო მეოთხე — ყველა სხვა მიზეზის ზეგავლენის შედეგს.

თავისუფლების ხარისხები:

- საერთო რიცხვი: **N—1**, სადაც **N** გაზომვათა საერთო რაოდენობაა;
- სვეტებს შორის: **p—1**;
- სტრიქონებს შორის: **q—1**;
- **A×B**-სათვის: **(p—1)(q—1)**;
- ცხრილის თითოეული უჯრის შიგნით: **(n—1)pq = N—pq**.

რადგან დისპერსიული ანალიზის ძირითადი მიზანია რეზულტატურ ცვლადზე თითოეული ფაქტორისა და ფაქტორების ურთიერთქმედების ზეგავლენის მნიშველობის შეფასება, ამიტომ, სამი ნულოვანი ჰიპოთეზა უნდა შევამოწმოთ:

H_0^1: **A** ფაქტორით გამოწვეული რეზულტატური ცვლადის დისპერსია არ აღემატება შემთხვევითს.

H_0^2: **B** ფაქტორით გამოწვეული რეზულტატური ცვლადის დისპერსია არ აღემატება შემთხვევითს.

H_0^3: **A** და **B** ფაქტორების ურთიერთქმედებით გამოწვეული რეზულტატური ცვლადის დისპერსია არ აღემატება შემთხვევითს.

შემოწმება ხდება ფიშერის სტატისტიკის მეშვეობით. შესაბამისად, მივიღებთ ფიშერის სტატისტიკის სამ მნიშვნელობას.

I. **A** ფაქტორის ეფექტის შესაფასებლად:

$$F_A = \frac{SS_{ფA}}{SS_შ} \left\{ \frac{p-1}{N-pq} \right\}$$

II. **B** ფაქტორის ეფექტის შესაფასებლად:

$$F_A = \frac{SS_{ფB}}{SS_შ} \left\{ \frac{p-1}{N-pq} \right\}$$

III. **A** და **B** ფაქტორების ურთიერთქმედების ეფექტის შესაფასებლად:

$$F_{(A\times B)} = \frac{SS_{ფ(A\times B)}}{SS_შ} \left\{ \frac{(p-1)(q-1)}{N-pq} \right\}$$

შემოწმება ხდება სტანდარტულად, $F_{ემპ}$ და $F_{კრ}$ შედარებით.

ნულოვანი ჰიპოთეზების შემოწმების შედეგების მიხედვით გამოგვაქვს დასკვნა იმის შესახებ, თუ მთლიანი ვარიაციის რომელი კომპონენტებია სტატისტიკურად მნიშვნელოვანი და რომელი — უმნიშვნელო.

მაგალითად, არ არის გამორიცხული, რომ სასწავლო პროცესით სტუდენტთა კმაყოფილების ვარიაციაში მხოლოდ სტუდენტთა სპეციალობის ხვედრითი წილი აღმოჩნდეს სტატისტიკურად მნიშვნელოვანი.

ან, სტატისტიკურად მნიშვნელოვანი აღმოჩნდეს ორივე ფაქტორის: სტუდენტთა სპეციალობისა და აკადემიური მოსწრების ხვედრითი წილი, მაგრამ უმნიშვნელო — მათი ურთიერთქმედების ხვედრითი წილი.

ორფაქტორული დისპერსიული ანალიზი ტრივიალურად შეიძლება განზოგადდეს ფაქტორების ნებისმიერი სასრული რაოდენობისათვის.

❖ დისპერსიული ანალიზის ინტერპრეტაცია ფაქტორის დისკრიმინაციული უნარის[1] შეფასებისას

სოციოლოგიის ერთ-ერთი უმნიშვნელოვანესი ამოცანა იმ **მიზეზების** გამოვლენაა, რომლებიც **განსხვავებული სოციალური ჯგუფების წარმოქმნას ინვევს**.

ფაქტორის უნარს, ზეგავლენა მოახდინოს რეზულტატური ცვლადის მნიშვნელობების განსხვავებულ ჯგუფებად დაყოფაზე/დისკრიმინაციაზე, **დისკრიმინაციული უნარი** ეწოდება.

თუ ფაქტორის მოქმედება, ე. ი. მისი **გრადაციების ცვლილება** რეზულტატური ცვლადის მნიშვნელობების **განსხვავებულ ჯგუფებად დაყოფას იწვევს**, მაშინ, ცხადია, რომ რეზულტატური ცვლადის მნიშვნელობათა **ჯგუფები ფაქტორის გრადაციების მიხედვით ერთმანეთისაგან არსებითად უნდა განსხვავდებოდეს.**

ჯგუფურ საშუალოებს შორის განსხვავების/**ჯგუფთაშორისი დისპერსიის** შეფასება საშუალებას გვაძლევს განვსაზღვროთ **რამდენად მნიშვნელოვანია ფაქტორის ზემოქმედების შედეგად გამოწვეული განსხვავება რეზულტატური ცვლადის ჯგუფებს შორის.**

ე. ი. შეიძლება **შევაფასოთ ფაქტორის დისკრიმინაციული უნარი რეზულტატურ ცვლადთან მიმართებაში:**

თუ ჯგუფურ საშუალოებს შორის **განსხვავება/ჯგუფთაშორისი დისპერსია** სტატისტიკურად **მნიშვნელოვანია**, ითვლება, რომ მოცემული ფაქტორი **ზეგავლენას ახდენს განსხვავებული ჯგუფების წარმოქმნაზე**.

ხოლო თუ ჯგუფურ საშუალოებს შორის **განსხვავება/ჯგუფთაშორისი დისპერსია სტატისტიკურად მნიშვნელოვანი არ არის**, ითვლება, რომ განსხვავებულ ჯგუფებად დაყოფას მოცემული ფაქტორის ზეგავლენა არ განაპირობებს.

დისპერსიული ანალიზის ასეთი ინტერპრეტაცია მნიშვნელოვანი შინაარსობრივი დასკვნების გამოტანის საშუალებას გვაძლევს. განვიხილოთ მაგალითი:

დავუშვათ, გამოკითხვის შედეგად, ამომრჩევლები ოთხ ჯგუფად დაიყო: პირველი ჯგუფი მმართველ პარტიას აძლევს ხმას, მეორე ოპოზიციურს, მესამეს არჩევანის გაკეთება უჭირს, ხოლო მეოთხეს არ სურს საკუთარი არჩევანის დაფიქსირება.

თუ კვლევა რომელიმე პარტიისათვის რეკომენდაციების მიცემის მიზნით ტარდება, ეს ინფორმაცია საკმარისი არ არის: ვერ განვსაზღვრავთ, ვის დაუჭერენ მხარს მესამე ან მეოთხე ჯგუფის წარმომადგენლები, ვერ ვიმსჯელებთ იმის შესახებ, თუ რა მეთოდებით შეიძლება ამ პარტიისათვის ხმის მიმცემთა ჯგუფის გაზრდა და სხვა.

ამისათვის აუცილებელია ვიცოდეთ **რა არსებითი ნიშან-თვისებებით განსხვავდებიან რეზულტატური ცვლადის მნიშვნელობათა ჯგუფები ანუ რომელი ფაქტორები განაპირობებს ამომრჩევლების განაწილებას მოცემულ ჯგუფებად.**

სავარაუდო ფაქტორების მიხედვით დისპერსიული ანალიზის ჩატარების შედეგად შეიძლება გამოვლინდეს, რომ, მაგალითად, ამომრჩეველთა ჯგუფები **ერთმანეთისაგან განსხვავდებიან** სახელმწიფოს საშინაო ეკონომიკური პოლიტიკით კმაყოფილების ხარისხით.

ეს იმას ნიშნავს, რომ სახელმწიფოს საშინაო ეკონომიკური პოლიტიკით კმაყოფილების ხარისხი ამომრჩეველთა დისკრიმინაციის/**დისკრიმინაციულ ჯგუფებად დაყოფის ერთ-ერთი არსებითი მიზეზია.**

მიღებული შედეგი შემდგომი შინაარსობრივი ანალიზის საფუძველზე პროგნოზებსა და რეკომენდაციებში აისახება. მაგალითად:

- სახელმწიფოს საშინაო ეკონომიკური პოლიტიკით კმაყოფილი ამომრჩევლები ძირითადად მმართველ პარტიას მისცემენ ხმას, ხოლო უკმაყოფილო — ოპოზიციურს;
- წინასაარჩევნო კამპანიაში აქცენტი საშინაო ეკონომიკურ პოლიტიკაზე უნდა გაკეთდეს და ა. შ.

---

[1] ლათ.: discrimen — განსხვავების უნარი.

ფაქტორული ანალიზი

❖ ფაქტორული ანალიზის იდეა

სოციალურ მოვლენათა უმრავლესობა რთული, **კომპლექსური** ხასიათისაა და **უშუალო გაზომვას არ ექვემდებარება.**

შეუძლებელია უშუალოდ გავზომოთ დემოკრატიულობის, ტრადიციულობის, რელიგიურობის, პოლიტიკური თუ სოციალური აქტივობის ხარისხები და ა. შ.

მსგავსი მოვლენების შესახებ ლოგიკურ-სისტემური ანალიზის საფუძველზე გამოყოფილი **იმ ნიშან-თვისებების** მიხედვით ემსჯელობთ, რომელთა უშუალო გაზომვაც შესაძლებელია.

მაგალითად, დემოკრატიულობის ხარისხის განსაზღვრისას შეიძლება გავზომოთ შეფასებითი დამოკიდებულებები სიტყვისა და რელიგიური მწრამსის თავისუფლების, ეროვნული თუ სხვა უმცირესობების მდგომარეობის, ძალაუფლების ორგანოთა ფორმირების წესის მიმართ და სხვა.

ცხადია, რაც უფრო მეტ ნიშან-თვისებას შევისწავლით ანუ რაც უფრო მეტ ცვლადს ჩავრთავთ კვლევაში, მით უფრო **მომცველი და მრავალფეროვანი** იქნება საკვლევი მოვლენის შესახებ მიღებული ინფორმაცია.

მეორე მხრივ, **მრავალრიცხოვანი და მრავალფეროვანი ინფორმაცია** წარმოშობს **ახალ პრობლემას:** როგორ გაეანალიზოთ მიღებული მონაცემები ისე, რომ **საკვლევი მოვლენის შესახებ ზოგადი მსჯელობა, მისი მთლიანობაში განხილვა შეგვძლოთ?**

ნებისმიერი მრავალგანზომილებიანი ანალიზი, ფაქტობრივად, ამ პრობლემის გადაჭრის **თავისებური მიდგომაა.**

ერთ-ერთი მათგანია **ფაქტორული ანალიზი.**

იგი არსებითად განსხვავდება ჩვენს მიერ უკვე განხილული მრავალგანზომილებიანი ანალიზებისაგან — არ გულისხმობს, რომ საწყისი[1] **ცვლადები უნდა კლასიფიცირდეს** ფაქტორულ[2] და რეზულტატურ[3] ცვლადებად.

ფაქტორული ანალიზის **ძირითადი მიდგომა** შემდეგნაირია: ჩვენს მიერ **კვლევაში ჩართული ცვლადები ერთი და იგივე მოვლენის ნიშან-თვისებებია.** ამიტომ, სავარაუდოა, რომ ამ ცვლადების მნიშვნელობათა განაწილებებზე **სპეციფიკურთან ერთად, ზოგადი/საერთო მიზეზებიც[4] ზემოქმედებს.**

მაგალითად, თუ სახელმწიფოში საერთოდ ირღვევა დემოკრატიის პრინციპები, სიტყვის თავისუფლების უარყოფითი შეფასება იმ **სპეციფიკურ მიზეზებთან ერთად,** რომლებიც **ძირითადად მხოლოდ სიტყვის თავისუფლებას უკავშირდება:** ზენოლა მასმედიაზე, დევნა განსხვავებული აზრის საჯაროდ გამოთქმის გამო და ა. შ. **ზოგადი მიზეზის** — საერთო დარღვევების შედეგიც შეიძლება იყოს.

აქედან გამომდინარე, ფაქტორული ანალიზის ამოსავალია **არა საწყისი ცვლადების კლასიფიკაცია ფაქტორულ და რეზულტატურ ცვლადებად, არამედ მათზე მოქმედი მიზეზების — ფაქტორების კლასიფიკაცია ზოგად და სპეციფიკურ ფაქტორებად.**

შესაბამისად, საწყისი ცვლადის ემპირიული მნიშვნელობა განიხილება როგორც არაერთგვაროვანი სიდიდე: მისი **ნაწილი ზოგადი მიზეზებით/ზოგადი ფაქტორებით აიხსნება,** ნაწილს კი — **სპეციფიკური მიზეზების ერთობლიობა/სპეციფიკური ფაქტორი[5] განაპირობებს.**

ზოგადი/საერთო მიზეზები, ცხადია, **უშუალოდ არ ვლინდება** — ისინი საწყისი

---

[1] უშუალოდ გაზომილი.
[2] მიზეზობრივ.
[3] შედეგობრივ.
[4] **ერთი ან რამდენიმე.**
[5] ეს მიზეზები, ძირითადად, მხოლოდ მოცემულ ცვლადზე მოქმედებს.

ცვლადებიდან **გენერალიზაციით**[1] მიიღება. ამ გაგებით, **ზოგადი ფაქტორები ლატენტური**, ფარული ხასიათისაა. ამიტომ, მათ ხშირად **ლატენტურ ფაქტორებსაც** უწოდებენ.

თუ ზოგად ფაქტორებს გამოვაელენთ და დამოუკიდებელი ცვლადების სახით შემოვიტანთ, შევძლებთ, რომ მოვახდინოთ **საწყისი ცვლადების რედუქცია**[2] **ზოგად ცვლადებზე/ლატენტურ ფაქტორებზე**.

ახლა უკვე შესაძლებელია ჩამოვაყალიბოთ ფაქტორული ანალიზის დეფინიცია[3]:

ფაქტორული ანალიზი ეწოდება იმ მეთოდების ერთობლიობას, რომელთა მეშვეობითაც შესაძლებელია:

- **საწყის მონაცემთა რედუქცია ზოგად მონაცემებზე.**
- **ზოგადი ფაქტორების გამოვლენა და შინაარსობრივი ახსნა.**

ზოგადი ფაქტორები ორთოგონალურია/დამოუკიდებელია[4] როგორც ერთმანეთისაგან, ისე სპეციფიკური ფაქტორებისაგან.

შესაბამისად, საწყისი ცვლადების რედუქცია ზოგად ფაქტორებზე თავიდან აგვაცილებს **ცვლადების ურთიერთქმედების აღწერისა და ახსნის სირთულეს**.

გარდა ამისა, ზოგადი ფაქტორების **მაქსიმალური რაოდენობა** საწყისი ცვლადების რაოდენობას ვერ გადააჭარბებს, ხშირ შემთხვევაში კი, მასზე მცირეა.

საწყის ცვლადებთან შედარებით მცირერიცხოვანი ფაქტორების შემოტანას ის უპირატესობა გააჩნია, რომ შემცირდება გასაანალიზებელი ცვლადების რაოდენობა, მსჯელობა უფრო ზოგად ხასიათს მიიღებს.

ამიტომ ფაქტორული ანალიზის ერთ-ერთ მთავარ ამოცანად მონაცემთა რიცხვის შემცირება ითვლება და ხშირია შემთხვევა, როდესაც მას ძირითადად **საწყისი ცვლადების რაოდენობის მინიმიზაციის**[5] მიზნით იყენებენ.

აქედან გამომდინარე, ზოგადი ფაქტორების რაოდენობის განსაზღვრას არსებითი მნიშვნელობა ენიჭება.

რამდენი ზოგადი ფაქტორის გამოყოფაა მიზანშეწონილი?

ამ კითხვაზე ერთმნიშვნელოვანი პასუხი არ არსებობს. ზოგადი ფაქტორების რაოდენობის დასადგენად, სტატისტიკურ მეთოდებთან ერთად, **შინაარსობრივი მოსაზრებებიც უნდა გამოვიყენოთ (კვლევის მიზანი და ამოცანები, გასაანალიზებელი მასალის სპეციფიკა და სხვა).**

გაზომვის დონე: ფაქტორული ანალიზის გამოყენება გაზომვის ნომინალურ დონეზე არ ხდება.

❖ ზოგადი ფაქტორების გამოყოფის მეთოდები

ზოგადი ფაქტორები სხვადასხვა სტატისტიკური მეთოდით შეიძლება გამოვყოთ. თითოეული მათგანი განსხვავებულ მიდგომას ემყარება და სპეციფიკური ფაქტორული მოდელის აგების საფუძველია.

---

[1] გენერალიზაცია — განზოგადება.
[2] დაყვანა.
[3] განმარტება.
[4] იგულისხმება დამოუკიდებლობა პირსონის r^2კოეფიციენტის მიხედვით.
[5] რაც შეიძლება მცირე რაოდენობაზე დაყვანის.

ცენტროიდული მეთოდი

ცენტროიდული მეთოდის ამოსავალია დაშვება, რომ ზოგადი ფაქტორები საერთოა საწყისი ცვლადებისათვის ანუ **ყოველი ზოგადი ფაქტორი მეტ-ნაკლებად თითოეულ საწყის ცვლადზე ზემოქმედებს,** ხოლო **i-ური საწყისი ცვლადის შესატყვისი სპეციფიკური ფაქტორის** ცალკეულება მხოლოდ ამ ცვლადზე აისახება, ე. ი. i-ური სპეციფიკური ფაქტორი i-ური საწყისი ცვლადის უნიკალურობას, მის იმ თავისებურებას გამოხატავს, რომლის გაწყვეტაც მოცემულ პირობებში შეუძლებელია.

ამ მეთოდის საფუძველზე აგებული ფაქტორული მოდელის[1] ძირითადი ამოცანაა **საწყისი** ცვლადების ასახვა **მინიმალური** რაოდენობის **ზოგადი ფაქტორების** წრფივ კომბინაციაში:

$$X_1, X_2, ..., X_n \rightarrow F_1, F_2, ..., F_m \quad (m \leq n)$$

სადაც:

$$X_i = \Sigma a_{ik} F_k + U_i$$

- F_k არის **k-ური "ლატენტური"** ცვლადი/ **k-ური ზოგადი ფაქტორი;**
- a_{ik} არის **k-ური ზოგადი ფაქტორის დატვირთვა** X_i ცვლადთან მიმართებაში. იგი შეიძლება განვიხილოთ როგორც იმის მაჩვენებელი, თუ რამდენად მნიშვნელოვანია F_k ზოგადი ფაქტორის გავლენა საწყის X_i ცვლადზე;
- $a_{ik}F_k$ არის საწყისი X_i ცვლადის ემპირიული მნიშვნელობების ის ნაწილი, რომელსაც იგი **k-ური ზოგადი ფაქტორის გავლენით იღებს;**
- U_i არის X_i ცვლადის ემპირიული მნიშვნელობების ის ნაწილი, რომელსაც იგი **სპეციფიკური ფაქტორის გავლენით იღებს და, შესაბამისად,** X_i ცვლადის დამახასიათებელ თავისებურებად/სპეციფიკად ითვლება[2].

ცვლადები სტანდარტიზებულია, ხოლო **ზოგადი ფაქტორები — ორთოგონალური. ამ პირობებში ფაქტორული დატვირთვები ემთხვევა კორელაციის კოეფიციენტებს საწყისი ცვლადებისა და ზოგადი ფაქტორების წყვილებს შორის.**

a_{ik} ფაქტორული დატვირთვის დადებითი ნიშანი X_i ცვლადსა და F_k ფაქტორს შორის პირდაპირპროპორციული დამოკიდებულების არსებობაზე მიუთითებს, ხოლო უარყოფითი ნიშანი — უკუპროპორციული დამოკიდებულების მაჩვენებელია.

შესაბამისად, X_i ცვლადის დისპერსია შეიძლება წარმოვადგინოთ ჯამის სახით, რომლის ერთი **შესაკრები ფაქტორული ნონების კვადრატების ჯამია,** მეორე კი **სპეციფიკური ფაქტორის დისპერსია:**

$$S^2x_i = H^2_i + S^2u_i$$

სადაც:

$$H^2_i = \Sigma a_{ik}^2$$

- S^2x_i არის საწყისი X_i ცვლადის **მთლიანი/ემპირიული** დისპერსია;
- H^2_i არის საწყისი X_i ცვლადის დისპერსიის ის ნაწილი, რომელსაც ზოგადი ფაქტორები ინვევს;
- S^2u_i არის საწყისი X_i ცვლადის დისპერსიის ის ნაწილი, რომელსაც **სპეციფიკური ფაქტორი ინვევს.**

შესაბამისად, H^2_i გამოხატავს საწყისი X_i ცვლადის დისპერსიის იმ ნაწილს, რომელიც

[1] ცენტროიდული მეთოდის საფუძველზე აგებულ მოდელს ცენტროიდული ფაქტორული მოდელი ეწოდება.
[2] ინგლ.: Unique — უნიკალური, ერთადერთი, სპეციფიკური.

ზოგადი/ლატენტური ფაქტორებით იხსნება, ხოლო $S^2_{u_i}$ — საწყისი X_i ცვლადის დისპერსიის იმ ნაწილს, რომელიც ზოგადი **ფაქტორებით ვერ იხსნება.**

რაც უფრო ნაკლებია ზოგადი ფაქტორების რაოდენობა და მეტია ზოგადი ფაქტორებით ახსნილი საწყისი X_i ცვლადის დისპერსიის ხვედრითი წილი მით უკეთესია ცენტროიდული ფაქტორული მოდელი.

მთავარ კომპონენტთა მეთოდი

მთავარ კომპონენტთა მეთოდის ძირითადი იდეა შემდეგში მდგომარეობს:

ზოგადი ფაქტორები იმ წესით უნდა გამოიყოს, რომ აუხსნელი დისპერსია სტატისტიკურად უმნიშვნელო აღმოჩნდეს. აქედან გამომდინარე, შესაბამის ფაქტორულ მოდელში სპეციფიკური ფაქტორი ელიმინირდება, ხოლო საწყისი ცვლადების ჯამური დისპერსია განიხილება როგორც, ძირითადად, მხოლოდ ზოგადი ფაქტორების გავლენის შედეგი.

ზოგადი ფაქტორების გამოყოფის წესი:

თავდაპირველად **გამოიყოფა პირველი ზოგადი ფაქტორი**, რომლის მეშვეობითაც აიხსნება საწყისი ცვლადების **ჯამური დისპერსიის ის მაქსიმალური ნაწილი, რომელიც მხოლოდ ერთმა ზოგადმა ფაქტორმა შეიძლება გამოიწვიოს.**

შემდეგ ხდება **მეორე ზოგადი ფაქტორის** გამოყოფა, რომლის მეშვეობითაც აიხსნება **დარჩენილი ჯამური დისპერსიის ის მაქსიმალური ნაწილი, რომელიც მხოლოდ ერთმა ფაქტორმა შეიძლება გამოიწვიოს** და ა. შ.

ამ წესით მიღებულ ზოგად ფაქტორებს **მთავარი კომპონენტები** ეწოდება.

თვალსაჩინოებისათვის მთავარ კომპონენტთა გამოყოფის მეთოდი შეიძლება გეომეტრიულად აღიწეროს (იხ. დიაგრამა F1):

საკოორდინატო სისტემის ცენტრსა და "მონაცემთა ღრუბელზე" გამავალი ნებისმიერი ღერძი შეიძლება განვიხილოთ როგორც ახალი ცვლადი.

ამ ღერძზე თითოეულ საწყის მონაცემს შეესაბამება "ღერძული მნიშვნელობა" — მანძილი მის პროექციასა და საკოორდინატო სისტემის ცენტრს შორის[1].

თავდაპირველად ვეძებთ ცენტრზე გამავალ ისეთ ღერძს, რომელზეც მონაცემთა "ღერძული მნიშვნელობების" დისპერსია მაქსიმალურია.

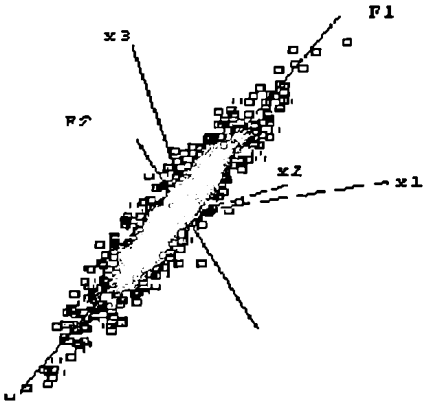
მისი ნორმირების შედეგად მივიღებთ **პირველ ფაქტორს.**

შემდეგ ვეძებთ საკოორდინატო სისტემის ცენტრზე პირველი ღერძის პერპენდიკულარულად გამავალ ისეთ ღერძს, რომელზეც მონაცემთა "ღერძული მნიშვნელობების" დისპერსია მაქსიმალურია და მისი ნორმირების შედეგად ვიღებთ **მეორე ფაქტორს.**

მესამე ღერძი პირველი ორის პერპენდიკულარი უნდა იყოს და ა. შ.

[1] ნიშანი განისაზღვრება მონაცემის მდებარეობით საკოორდინატო სისტემის ცენტრის/ნულის მიმართ.

დიაგრამა F1.



X1, X2, X3 — საწყისი ცვლადები. F1, F2 — ფაქტორები

მთავარი კომპონენტები ორთოგონალურია.

მთავარი კომპონენტების რაოდენობის განსაზღვრა კაიზერის კრიტერიუმით:

კაიზერის კრიტერიუმის მიხედვით, უმნიშვნელოდ/მინიმალურად ითვლება ჯამური დისპერსიის ის ნაწილი, რომელიც ერთ საწყის ცვლადზე მოდის[1] ანუ 1 და მთავარ კომპონენტად გამოიყოფა ყველა ის ზოგადი ფაქტორი, რომლის მიერ ახსნილი დისპერსიაც 1-ზე მეტია.

პარამეტრთა ექსტრემალური დაჯგუფების მეთოდი

პარამეტრთა ექსტრემალური დაჯგუფების მეთოდის ძირითადი იდეა:

- საწყისი ცვლადები შეიძლება დაჯგუფდეს ისე, რომ ერთ ჯგუფში შემავალი ცვლადები უფრო მეტად/ძლიერ იყოს ერთმანეთთან დაკავშირებული, ვიდრე სხვა ჯგუფის ცვლადებთან;
- ასეთი დაჯგუფების შედეგად, თითოეულმა ჯგუფმა შეიძლება ერთი გარკვეული ზოგადი მიზეზის/ზოგადი ფაქტორის მოქმედება გამოხატოს.

ე. ი. "მაქსიმალურად ურთიერთკორელირებადი" ცვლადების ჯგუფი, ამავე დროს, მაქსიმალურად უნდა კორელირებდეს ერთ რომელიმე ზოგად ფაქტორთან.

ამ წესით გამოყოფილი ზოგადი ფაქტორები, ცხადია, ზოგადი/საერთო იქნება მხოლოდ ერთ ჯგუფში შემავალი საწყისი ცვლადებისათვის და არა ყველა საწყისი ცვლადისათვის ერთდროულად.

შესაბამისად, პარამეტრთა ექსტრემალური დაჯგუფების მეთოდის საფუძველზე აგებული ფაქტორული მოდელის ფარგლებში თითოეული საწყისი ცვლადი განიხილება

1 იგი ჯამური დისპერსიის ერთეულადაა მიჩნეული.

როგორც ერთ და მხოლოდ ერთ ზოგად მიზეზზე დამოკიდებული.

ანუ ყოველი ცვლადი აუცილებლად შევა ერთ რომელიმე ზოგად ფაქტორში და გამორიცხულია, რომ იგი სხვა ფაქტორშიც შევიდეს.

პარამეტრთა ექსტრემალური დაჯგუფების მეთოდი გამონაკლისია იმ თვალსაზრისით, რომ ამ მეთოდით მიღებული ზოგადი ფაქტორები შეიძლება არ იყოს ორთოგონალური.

❖ ფაქტორების ბრუნვა

ზოგადი ფაქტორების შინაარსობრივი ინტერპრეტაციის ამოსავალია ფაქტორული დატვირთვები.

კონტრასტული/ერთმანეთისაგან განსხვავებული ფაქტორული დატვირთვები ამარტივებს ზოგად ფაქტორებსა და საწყის ცვლადებს შორის არსებულ დამოკიდებულებათა ანალიზს, ხელს უწყობს ფაქტორების შინაარსობრივი ინტერპრეტაციის მონახვას.

ფაქტორულ დატვირთვათა კონტრასტულობის გაზრდა შესაძლებელია ფაქტორების ბრუნვის მეთოდით:

თუ საკოორდინატო სისტემაში, რომლის ღერძებიც ზოგად ფაქტორებს გამოსახავს, საწყისი ცვლადების სკატერგრამას[1] ავაგებთ, მაშინ, i-ური ცვლადის თითოეული კოორდინატი აღმოჩნდება შესაბამისი ღერძის, ე. ი. ზოგადი ფაქტორის ფაქტორული დატვირთვა i-ურ ცვლადთან მიმართებაში.

ამ სისტემის ღერძების/ზოგადი **ფაქტორების** ნებისმიერი მიმართულებითა და კუთხით ბრუნვის შემთხვევაში, ნერტილთა/**საწყისი ცვლადების** ურთიერთგანლაგება იგივე დარჩება, მაგრამ შეიცვლება მათი კოორდინატები/**ფაქტორული დატვირთვები**.

ეს თავისებურება ზოგადი ფაქტორების კონტრასტულობის გასაზრდელად შეიძლება გამოვიყენოთ: უნდა მოინახოს ფაქტორების ბრუნვის ისეთი წესი, რომელიც თითოეული ცვლადისათვის ფაქტორული დატვირთვების ერთი ნაწილის ზრდასა და მეორე ნაწილის კლებას გამოიწვევს.

ასეთი წესით ზოგადი ფაქტორების ბრუნვის შედეგად, ყოველი საწყისი ცვლადისათვის გამოიყოფა მაქსიმალური დატვირთვის მქონე ზოგადი ფაქტორი და შესაძლებელი გახდება, რომ თითოეული საწყისი ცვლადი **ერთ** ზოგად ფაქტორში შევიდეს.

გარდა ამისა, შესაძლებელი გახდება ზოგადი ფაქტორების რანჟირება ზოგადობის ხარისხის მიხედვით[2].

შესაბამისად, მნიშვნელოვნად გამარტივდება როგორც საწყისი ცვლადების, ისე ზოგადი ფაქტორების სტრუქტურა. ეს კი ზოგადი ფაქტორების შინაარსობრივი ინტერპრეტაციის კიდევ ერთი ხელშემწყობი პირობაა.

❖ ზოგადი ფაქტორების შინაარსობრივი ინტერპრეტაცია

ზოგადი ფაქტორების შინაარსობრივი ინტერპრეტაცია ფაქტორული ანალიზის საბოლოო და ერთ-ერთი ყველაზე რთულ ეტაპია.

ზოგადი ფაქტორები, როგორც უკვე აღვნიშნეთ, ლატენტური ხასიათისაა. ფაქტორული მოდელის აგების შემდეგ, ჩვენ გვაქვს სტატისტიკური ინფორმაცია იმის შესახებ, თუ რამდენი ზოგადი ფაქტორი/მიზეზი განსაზღვრავს საწყისი ცვლადების ვარიაციას და რომელი ზოგადი ფაქტორი რომელი საწყისი ცვლადისათვისაა არსებითი, მაგრამ არ ვიცით

[1] თითოეულ ცვლადს ერთი ნერტილი შეესაბამება.

2 რაც უფრო მეტი ცვლადის მიმართ გააჩნია მაქსიმალური დატვირთვა, მით უფრო ზოგადად ითვლება ფაქტორი.

მთავარი — რა არის ეს მიზეზები.

ზოგადი ფაქტორების შინაარსობრივი ინტერპრეტაცია გულისხმობს:

- **ზოგადი ფაქტორისათვის "სახელის დარქმევას" — ისეთი ტერმინის შემოტანას, რომელიც მის შინაარსობრივ საზრისს გამოხატავს.**
- **ზოგად ფაქტორებსა და საწყის ცვლადებს შორის დამოკიდებულების ახსნას.**

ზოგადი ფაქტორების შინაარსობრივი ინტერპრეტაციის უნივერსალური/საყოველთაო წესები არ არსებობს. რეკომენდაციის სახით შეიძლება მივუთითოთ, რომ უნდა დავეყრდნოთ არა მხოლოდ სტატისტიკურ შედეგებს, არამედ სიტუაციურ ანალიზსაც და გამოვიყენოთ **ნებისმიერი ინფორმაცია**, რაც მიღებული სურათის ახსნაში დაგვეხმარება.

ამიტომ, იმისათვის, რომ ნათლად დავინახოთ პრობლემის სპეციფიკა და მისი გადაჭრის ძირითადი ეტაპების მნიშვნელობა, განვიხილოთ რამდენიმე მაგალითი[1].

I. ხუთქულიანი სკალით, რომელზეც 1 ქულა შეესაბამება "სავსებით კმაყოფილს", ხოლო 5 "სრულიად უკმაყოფილოს", შეფასდა რესპონდენტთა კმაყოფილება:

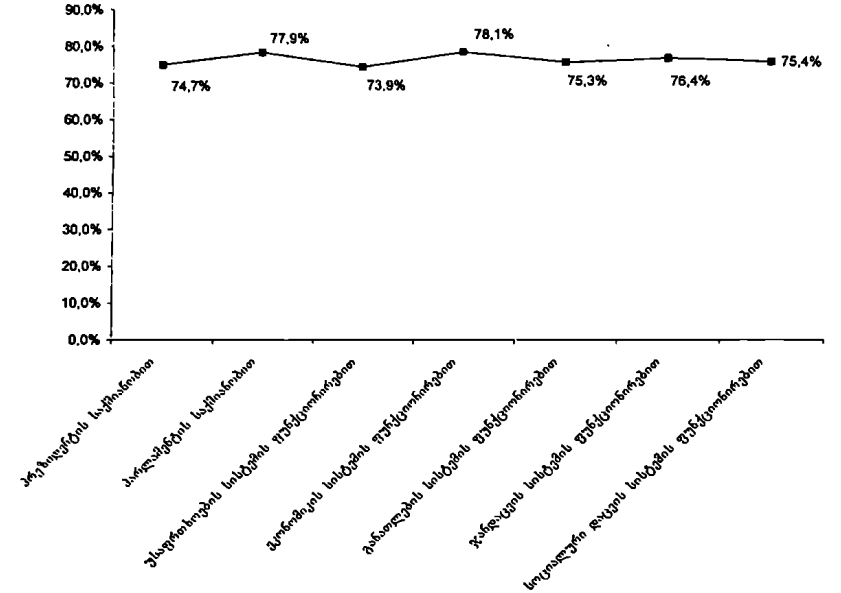
1. პრეზიდენტის საქმიანობით
2. პარლამენტის საქმიანობით
3. უსაფრთხოების სისტემის ფუნქციონირებით
4. ეკონომიკის სისტემის ფუნქციონირებით
5. განათლების სისტემის ფუნქციონირებით
6. ჯანდაცვის სისტემის ფუნქციონირებით
7. სოციალური დაცვის სისტემის ფუნქციონირებით

ფაქტორული ანალიზის შედეგად გამოიყო მხოლოდ ერთი ფაქტორი. ეს იმის მაჩვენებელია, რომ შვიდივე საწყისი ცვლადის მნიშვნელობათა განაწილებას ერთი და იგივე ზოგადი ფაქტორი/ერთი ძირითადი მიზეზი განსაზღვრავს იხ. დიაგრამა **F2**.

ფაქტორის შესატყვის ნირზე მითითებულია **პროცენტებით გამოხატული ფაქტორული წონები** თითოეული საწყისი ცვლადისათვის.

[1] მაგალითები აღებულია ერთ-ერთ პოსტსაბჭოურ სახელმწიფოში რამდენიმე წლის წინ ჩატარებული სოციოლოგიური კვლევიდან.

დიაგრამა F2. "გთხოვთ, გვიპასუხოთ, რამდენად კმაყოფილი ბრძანდებით..."



რა ზოგადი ფაქტორი შეიძლება იყოს ეს?

აშკარაა, რომ მიღებული სურათი მთელი სახელმწიფოებრივი სისტემისადმი ერთგვაროვანი დამოკიდებულების შედეგია.

ამის შემდეგ იბადება კითხვა: ერთგვაროვნად კარგი, თუ ერთგვაროვნად ცუდი?

იმისათვის, რომ ამ კითხვას ვუპასუხოთ, საჭიროა განვიხილოთ საწყისი ცვლადების საშუალო შეფასებები.

საშუალო შეფასებები 4-სა და 5-ს შორის აღმოჩნდა.

ე. ი. დაფიქსირდა ერთგვაროვნად უარყოფითი განწყობა მთელი სახელმწიფოებრივი სისტემის მიმართ. რა შეიძლება ვუწოდოთ მიღებულ ფაქტორს, როგორ გამოვხატოთ მისი შინაარსობრივი საზრისი?

ფაქტორი ზოგადი ცვლადი/ნიშან-თვისებაა და მის სახელში უნდა აისახოს ის, თუ რა ნიშან-თვისებაა იგი.

მოცემულ შემთხვევაში გამოყოფილ ზოგად ფაქტორს შეიძლება ვუწოდოთ "დამოკიდებულება **სახელმწიფოებრივი სისტემის** მიმართ" ან, პირდაპირ, "სახელმწიფოებრივი სისტემის ფაქტორი".

შემდეგ ეტაპზე, ბუნებრივია, დგება საკითხი იმის შესახებ, თუ რა იწვევს ასეთ ზოგად და რადიკალურად უარყოფით დამოკიდებულებას.

ამის გასარკვევად სხვა მონაცემები უნდა განვიხილოთ.

სხვა მონაცემების მიხედვით გაირკვა, რომ შემოსავლებისა და დასაქმების მაჩვენებლები ძალიან დაბალია, არადამაკმაყოფილებელია სოციალური დაცვისა და უსაფრთხოების სისტემების მუშაობა, ხშირია კანონდარღვევები. ე. ი. სახელმწიფოებრივი სისტემა, ზოგადად, მოშლილია.

აქედან გამომდინარე, ფაქტორსა და საწყის ცვლადებს შორის არსებული კავშირი იმით შეიძლება აიხსნას, რომ სახელმწიფოებრივი სისტემის ნებისმიერი რგოლი მოშლილი, მწყობრიდან გამოსული მთელის ნაწილად აღიქმება. შესაბამისად, თითოეული მათგანის შეფასება, ძირითადად, მასზე მთლიანი სისტემის შეფასების გავრცელებით ხდება და არა დიფერენცირებული მიდგომის შედეგად.

ჩვენი მსჯელობა შემდეგნაირად შეიძლება შევაჯამოთ: სახელმწიფოებრივი სისტემის მოშლილობა მის მიმართ საზოგადოების ზოგად უარყოფით განწყობას განაპირობებს, რაც, თავის მხრივ, ყველა სახელმწიფოებრივი ქვესისტემის ერთგვაროვან დეევალვაციას იწვევს.

II. ხუთქულიანი სკალით, რომელზეც 1 ქულა შეესაბამება "სავსებით კმაყოფილს", ხოლო 5 "სრულიად უკმაყოფილოს", შეფასდა რესპონდენტთა კმაყოფილება:

1. სასმელი წყლის ხარისხით.
2. ჰაერის სისუფთავით.
3. კლიმატით.
4. ელექტროენერგიის მიწოდებით.
5. საკუთარი საქმიანობით.
6. იმ პირობებით, რომელშიც საქმიანობა უწევს.
7. ოჯახის შემოსავლით.
8. ცხოვრების პირობებით საერთოდ.

ფაქტორული ანალიზი ჩატარდა პირველი შვიდი ცვლადის მიმართ. შედეგად, სამი ფაქტორი გამოიყო (იხ. დიაგრამა F3).

ფაქტორული წონების მიხედვით, პირველი ფაქტორი არსებით გავლენას ორ საწყის ცვლადზე: რესპონდენტთა **საქმიანობისა და საქმიანობის პირობების** მიმართ დამოკიდებულებაზე ახდენს. თითოეული მათგანის საშუალო მნიშვნელობა "სრულიად უკმაყოფილოს" უახლოვდება.

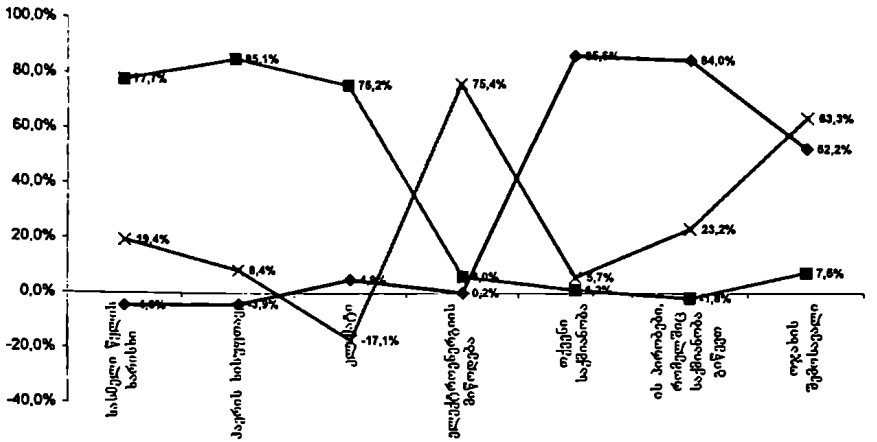
სხვა მონაცემების მიხედვით, უმუშევრობა ერთ-ერთ აქტუალურ პრობლემად სახელდება, რესპონდენტთა მნიშვნელოვანი ნაწილი მისი პროფესიისა და კვალიფიკაციისათვის შეუფერებლად დაბალი დონის სამუშაოს ასრულებს, ხოლო ანაზღაურება, უმეტეს შემთხვევაში, ძალიან დაბალია.

ფაქტორსა და საწყის ცვლადებს შორის არსებული კავშირი იმით შეიძლება აიხსნას, რომ ასეთ პირობებში, ზოგადად **საქმიანობის სფერო** უარყოფითად ფასდება და ეს შეფასება განსაზღვრავს დამოკიდებულებას მისი სხვადასხვა ასპექტის მიმართ.

შესაბამისად, პირველი ფაქტორის შინაარსობრივი საზრისი/მისი "სახელი" შეიძლება გამოიხატოს, როგორც "დამოკიდებულება **საქმიანობის სფეროს** მიმართ"/"საქმიანობის სფეროს ფაქტორი".

დიაგრამა F3.

გთხოვთ, გვიპასუხოთ, რამდენად გაკმაყოფილებთ...



მეორე ფაქტორი არსებით გავლენას სამ საწყის ცვლადზე: სასმელი წყლის ხარისხით, ჰაერის სისუფთავითა და კლიმატით კმაყოფილებაზე ახდენს. ამასთან, თითოეული მათგანის საშუალო მნიშვნელობა "სავსებით კმაყოფილს" უახლოვდება.

ამ ფაქტორის შინაარსობრივი საზრისი შეიძლება გამოვხატოთ, როგორც "დამოკიდებულება **ბუნებრივი გარემოს** მიმართ"/"ბუნებრივი გარემოს ფაქტორი", ხოლო ზოგად ფაქტორსა და საწყის ცვლადებს შორის არსებული კავშირი იმით შეიძლება აიხსნას, რომ რესპონდენტები, ზოგადად, კმაყოფილნი არიან იმ ბუნებრივი გარემოთი, რომლებშიც ცხოვრობენ და შეფასებისას მისი ასპექტების დიფერენცირებას არ ახდენენ.

მესამე ფაქტორი არსებით გავლენას ორ საწყის ცვლადზე: ელექტროენერგიის მიწოდებითა და ოჯახის შემოსავლით კმაყოფილებაზე ახდენს. თითოეული მათგანის შეფასება "სრულიად უკმაყოფილოს" უახლოვდება.

სხვა მონაცემების მიხედვით, შემოსავლები ძალიან დაბალია და, უმეტეს შემთხვევაში, საარსებო მინიმუმს ვერ აკმაყოფილებს. ელექტროენერგიის მიწოდება კი ხანგრძლივი და ქაოტური წყვეტებით ხდება, რაც ხელს უშლის როგორც ელემენტარული ყოფითი მოთხოვნილებების დაკმაყოფილებას, ისე სამეურნეო საქმიანობას.

ეკონომიკური გასაჭირი და ელექტროენერგია პირველი რიგის პრობლემებს შორის ყველაზე მაღალი სიხშირეებით ხასიათდება.

აქედან გამომდინარე, ამ ერთმანეთისაგან განსხვავებულ მოვლენებში ის **საერთო** შეიძლება დავინახოთ, რომ თითოეული მათგანი საზოგადოებისათვის საყოველთაო და უმწვავესი პრობლემაა.

შესაბამისად, მეორე ფაქტორის შინაარსობრივი საზრისი შეიძლება გამოვხატოთ, როგორც "დამოკიდებულება **პირველი რიგის პრობლემების** მიმართ"/"პირველი რიგის პრობლემების ფაქტორი".

ფაქტორსა და საწყის ცვლადებს შორის არსებული კავშირი კი იმით აიხსნება, რომ პირველი რიგის პრობლემების სიმწვავე და საყოველთაობა, მათ მიმართ ერთგვაროვან დამოკიდებულებას იწვევს.

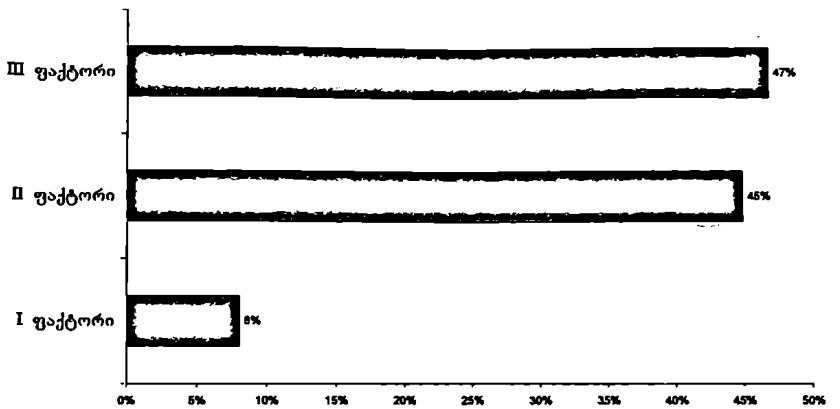
თუ მერვე საწყის ცვლადს — საერთოდ, ცხოვრების პირობებით კმაყოფილებას განვიხილავთ როგორც ზოგად ფაქტორებზე დამოკიდებულს და რეგრესიულ ანალიზს ჩავატარებთ, შესაძლებელი გახდება განისაზღვროს თითოეული ფაქტორის ხვედრითი წილი საერთოდ, ცხოვრების პირობების მიმართ დამოკიდებულებაში (იხ. დიაგრამა F4).

ჩვენი მაგალითის შემთხვევაში, ბუნებრივი გარემოს მიმართ დამოკიდებულება 8%-ით, საქმიანობის სფეროს მიმართ დამოკიდებულება 45%-ით, ხოლო პირველი რიგის პრობლემების მიმართ დამოკიდებულება 47%-ით ხსნის იმას, თუ რა დამოკიდებულება აქვთ რესპონდენტებს საერთოდ, ცხოვრების მიმართ.

საერთოდ, ცხოვრების პირობების მიმართ დამოკიდებულების საშუალო მაჩვენებელი "სრულიად უკმაყოფილოს" უახლოვდება.

აქედან გამომდინარე, შეიძლება დავასკვნათ, რომ **რესპონდენტთა უკმაყოფილებას ცხოვრების პირობებით ძირითადად საქმიანობის სფერო და დაბალი შემოსავლები განსაზღვრავს.** ხოლო ბუნებრივი გარემო, რომლის მიმართაც დადებითი დამოკიდებულება ფიქსირდება, ზოგადად ცხოვრების პირობების მიმართ დამოკიდებულებაზე არსებით გავლენას ვერ ახდენს.

დიაგრამა F4.



❖ კლასტერული ანალიზის ძირითადი მიზანი

კლასტერული ანალიზი ევრისტიკული ხასიათის მეთოდია. მის ძირითად მიზანს **მონაცემთა სტრუქტურის გამოვლენა და შესწავლა წარმოადგენს.**

კლასტერული ანალიზის ფარგლებში მონაცემთა სტრუქტურის კვლევა კლასტერიზაციის — დაჯგუფების საფუძველზე ხდება.

კლასტერიზაციის შედეგად გამოყოფილ ობიექტთა ჯგუფებს **კლასტერები**[1] ეწოდება. ცნება "კლასტერის" სინონიმებად იყენებენ "ტაქსონს", "კლასს", "ჯგუფს", "სიმრავლეს", "ერთობლიობას".

კლასტერული ანალიზის მეშვეობით ხორციელდება:

- **ობიექტთა კლასიფიკაცია მოცემული ნიშან-თვისებების მიხედვით.**
  მაგალითად, შესაძლებელია, გამოვყოთ ამომრჩეველთა სხვადასხვა კლასტერი მათი ასაკის, საცხოვრებელი ადგილის, საქმიანობის, ეკონომიკური მდგომარეობის და ა. შ. ნიშნით.
- **სტატისტიკური ერთობლიობის სტრუქტურის შესახებ შემუშავებული ჰიპოთეზის შემოწმება.**
  მაგალითად, შესაძლებელია, შევამოწმოთ ჰიპოთეზა იმის შესახებ, რომ ამომრჩევლები **k** კლასტერად იყოფიან.
- **სტატისტიკური ერთობლიობის სტრუქტურის აღწერა/ანალიზი.**

შესაძლოა დავახასიათოთ, რამდენი კლასტერისაგან შედგება სტატისტიკური ერთობლიობა, რა მოცულობისაა (რამდენი ობიექტისაგან შედგება) თითოეული კლასტერი, განვსაზღვროთ ცალკეული ნიშან-თვისების დისკრიმინაციული მნიშვნელობა კლასტერების მიმართ და ა. შ.

იმისათვის, რომ კლასტერული ანალიზის მეშვეობით მონაცემთა სტრუქტურის **ოპტიმალური აღწერა და ინტერპრეტაცია** შევძლოთ, კლასტერიზაციის პროცესი საკვლევ ობიექტებს შორის არსებულ **მსგავსება-განსხვავებათა ადეკვატურად (შესაბამისად) უნდა ნარიმართოს.**

თუ ობიექტებს/მოცემული სტატისტიკური ერთობლიობის წევრებს **საკოორდინატო სისტემაში განლაგებული წერტილების** სახით მოვიაზრებთ, მაშინ:

- ობიექტების **მსგავსება-განსხვავების** შესახებ შესაძლებელია ვიმსჯელოთ **წერტილთა შორის მანძილების** ანუ იმის მიხედვით, რამდენად ახლოსაა/შორსაა წერტილები ერთმანეთისაგან;
- ობიექტებსა და წერტილებს შორის იგულისხმება შემდეგი სახის შესაბამისობა: რაც უფრო **მსგავსია** ობიექტები, მით **მცირეა მანძილი** საკოორდინატო სისტემაში მათ შესატყვის წერტილებს შორის;
- საკოორდინატო სისტემაში **კლასტერები წერტილთა არაცარიელი სიმრავლეების სახით იქნება წარმოდგენილი.**

❖ კლასტერიზაციის მეთოდები

კლასტერიზაცია სხვადასხვა მეთოდის მეშვეობით შეიძლება განხორციელდეს. ეს მეთოდები ორ ძირითად ჯგუფად იყოფა: **იერარქიულად და არაიერარქიულად.**

იერარქიული კლასტერიზაციის მეთოდის გამოყენების შედეგად მიიღება მონაცემთა

[1] ინგლ.: cluster — კონა, ჯგუფი, შენადედი.

იერარქიული, "ხისებრი" სტრუქტურა.

იერარქიული მეთოდები, თავის მხრივ, იყოფა **გამაერთიანებელ/აგლომერატულ და დამყოფ/დივიზიურ** მეთოდებად.

გამაერთიანებელი/აგლომერატული იერარქიული მეთოდების გამოყენებისას, კლასტერიზაციის საწყის ეტაპზე ითვლება, რომ თითოეული ობიექტი ცალკე კლასტერია და ხდება კლასტერების გაერთიანება ანუ ობიექტების უფროდაუფრო დიდი მოცულობის ჯგუფებში თავმოყრა მანამ, სანამ საბოლოოდ ყველა ობიექტი ერთი კლასტერში არ აღმოჩნდება.

დამყოფი/დივიზიური მეთოდების გამოყენებისას, კლასტერიზაციის საწყის ეტაპზე ითვლება, რომ ყველა ობიექტი ერთ კლასტერში შედის და შემდეგ ხდება მისგან მცირე კლასტერების გამოყოფა ანუ ობიექტების განაწილება უფროდაუფრო პატარა მოცულობის ჯგუფებში მანამ, სანამ საბოლოოდ სათითაო ობიექტი ცალკე კლასტერად არ გამოიყოფა.

ემპირიული სოციოლოგიური კვლევებისას, იერარქიული კლასტერიზაციის დროს უფრო ხშირად აგლომერატული მეთოდები გამოიყენება.

არაიერარქიული კლასტერიზაცია/k-საშუალოთა მეთოდი გულისხმობს თავდაპირველად **კლასტერის ცენტრის** ანუ საწყისი წერტილის განსაზღვრას/შერჩევას, ხოლო შემდეგ ობიექტთა თავმოყრას ამ ცენტრის ირგვლივ — მათ დაჯგუფებას ცენტრიდან ზღვრული დაშორების/მანძილის მიხედვით.

არაიერარქიული კლასტერიზაციის მეთოდებია:

თანმიმდევრული ზღვრული/"ზღურბლის" მეთოდი. ამ მეთოდის გამოყენებისას ხდება კლასტერის ცენტრის შერჩევა და ობიექტების გაერთიანება ცენტრის ირგვლივ ზღვრული დაშორების/მანძილის მიხედვით.

პარალელური ზღვრული/"ზღურბლის" მეთოდი. ამ მეთოდის გამოყენებისას ხდება ერთდროულად რამდენიმე კლასტერის ცენტრის შერჩევა და ობიექტების გაერთიანება ცენტრების ირგვლივ ზღვრული დაშორების/მანძილის მიხედვით.

ოპტიმალური განაწილების მეთოდი. ამ მეთოდის გამოყენებისას ობიექტები გადანაწილდება ისეთი წესით, რომ მოხდეს კლასტერებისშიდა და კლასტერებშორისი დაშორებების/მანძილების ჯამური მაჩვენებლების ოპტიმიზება.

❖ აგლომერატული იერარქიული კლასტერიზაციის მექანიზმი

აგლომერატული იერარქიული კლასტერიზაციის მექანიზმი შემდეგნაირია:

- კლასტერიზაციის დაწყებამდე, მოცემული სტატისტიკური ერთობლიობის ყოველი წევრი/ობიექტი ითვლება ცალკე კლასტერად. ე. ი. თუ სტატისტიკურ ერთობლიობაში შედის **n** ობიექტი, ითვლება, რომ გვაქვს **n** კლასტერი;
- თავდაპირველად ხდება **ორი უახლოესი კლასტერის ამორჩევა და მათი ერთ კლასტერად გაერთიანება.** შედეგად, კლასტერების რაოდენობა ერთით მოიკლებს, ე. ი. დაგვრჩება **n-1** კლასტერი;
- შემდეგ **n-1** კლასტერს შორის ხდება ორი უახლოესი კლასტერის ამორჩევა და მათი გაერთიანება, მივიღებთ **n-2** კლასტერს. და ა. შ.
- იერარქიული კლასტერიზაციის პროცესი მთავრდება **ერთი** კლასტერის მიღებით, მაგრამ იგი შეიძლება შეწყდეს **ნებისმიერ** ეტაპზე, როდესაც რაიმე **კრიტერიუმის** მიხედვით ჩაითვლება, რომ მნიშვნელოვანი კლასტერები უკვე გამოყოფილია.

მაგალითად, ასეთი კრიტერიუმი შეიძლება იყოს **კლასტერთა ურთიერთდაშორების/კლასტერებს შორის მანძილის ზღვრულად მიჩნეული სიდიდე.** ე. ი. კლასტერების გაერთიანების პროცესი შეიძლება შევწყვიტოთ მაშინ, როდესაც ორი უახლოესი კლასტერის

ურთიერთდაშორება გარკვეულ **ზღვარს** გადააჭარბებს.

კლასტერთა გაერთიანებას **აგრეგაცია** ენოდება.

აგლომერატული იერარქიული კლასტერიზაციის შედეგად მიღებული, ეგრეთ წოდებული **"იერარქიული ხე"** მოცემული სტატისტიკური ერთობლიობის **სტრუქტურას** გამოხატავს.

❖ კლასტერთა ურთიერთდაშორების ზომები

როგორც უკვე აღვნიშნეთ, საკოორდინატო სისტემაში კლასტერები წერტილთა არა-ცარიელი სიმრავლეების სახით მოიაზრება.

საკოორდინატო სისტემაში წერტილთა ურთიერთდაშორების ზომის/წერტილთა შორის მანძილის განსაზღვრა სხვადასხვა წესით/მეთოდით შეიძლება მოხდეს.

ასეთი მეთოდების მაგალითებია:

- **ევკლიდეს მანძილი.** ევკლიდეს მანძილი ეწოდება წერტილთა **გეომეტრიულ** დაშორებას მრავალგანზომილებიან საკოორდინატო სისტემაში და გამოითვლება ფორმულით:

$$\text{მანძილი } (x,y) = \sqrt{\Sigma_i (x_i - y_i)^2}$$

- **ევკლიდეს მანძილის კვადრატი.** ევკლიდეს მანძილის კვადრატი მიიღება ევკლიდეს მანძილის კვადრატში აყვანით, რაც დაშორების/მანძილის დატვირთვას/მნიშვნელობას ზრდის:

$$\text{მანძილი } (x,y) = \Sigma_i (x_i - y_i)^2$$

- **ხარისხობრივი მანძილი.** ხარისხობრივი მანძილი გამოითვლება ფორმულით:

$$\text{მანძილი } (x,y) = \Sigma_i \left(\left| x_i - y_i \right| ^p \right)^{1/r}$$

P და **r** პარამეტრების შერჩევა ხდება მკვლევრის მიერ.

როდესაც **P = r =2**, ხარისხობრივი მანძილი ევკლიდეს მანძილს ემთხვევა.

ხარისხის მაჩვენებელთა ცვლილება საშუალებას გვაძლევს მოვახდინოთ დაშორების დატვირთვის/მნიშვნელობის ვარირება.

- **მანჰეტენის მანძილი.** მანჰეტენის მანძილი გამოითვლება ფორმულით:

$$\text{მანძილი } (x,y) = \Sigma_i \left| x_i - y_i \right|$$

- **ჩებიშევის მანძილი.** ჩებიშევის მანძილი გამოიყენება მაშინ, როდესაც საჭიროა განისაზღვროს ორ ობიექტს შორის განსხვავება ერთი კოორდინატის (ერთი განზომილების/ნიშან-თვისების მიხედვით):

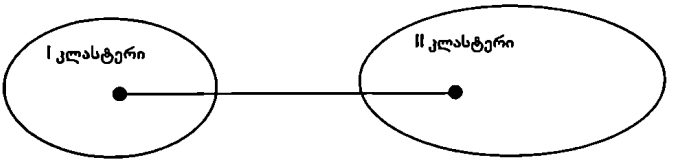
$$\text{მანძილი } (x,y) = \mathbf{maximum} \left| x_i - y_i \right|$$

- **უთავსობის პროცენტი.** უთავსობის პროცენტი გამოიყენება მაშინ, როდესაც მონაცემები კატეგორიული ფორმითაა წარმოდგენილი:

$$\text{მანძილი } (x,y) = (\text{რაოდენობა } x_i \neq y_i)/i$$

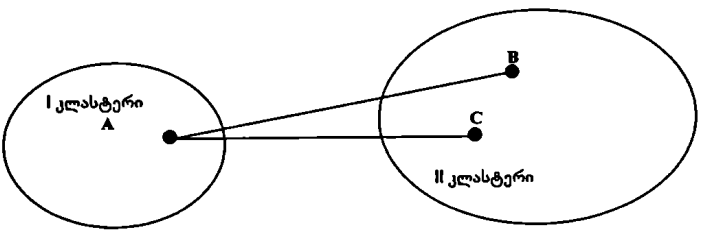
❖ **კლასტერების აგრეგაცია**

კლასტერიზაციის საწყის ეტაპზე, როდესაც ერთი კლასტერი ერთი ობიექტით, ე. ი. საკოორდინატო სისტემაში ერთი წერტილითაა წარმოდგენილი, კლასტერების აგრეგაცია რთული არაა, რადგან ორ კლასტერს შორის მანძილის ემთხვევა მანძილს/დაშორებას შესაბამის ორ წერტილს შორის:



ასეთ შემთხვევაში, როგორც ზემოთ აღვნიშნეთ, აგრეგაციის კრიტერიუმი წერტილთა წყვილებს შორის უმცირესი მანძილია.

მაგრამ სიტუაცია რთულდება, როდესაც ერთ-ერთ კლასტერში მაინც ერთზე მეტი წერტილი შედის:



ამ შემთხვევაში კლასტერებში შემავალ წერტილთა წყვილებს შორის სხვადასხვა მანძილია. რომელი მათგანი შეიძლება ჩაითვალოს მოცემულ კლასტერებს შორის მანძილად? **რა კრიტერიუმით გავაერთიანოთ კლასტერები?**

ეს პრობლემა რამდენიმე მეთოდით შეიძლება გადაიჭრას.

ასეთი მეთოდების მაგალითებია:

- **უახლოესი მეზობლების მეთოდი.** ორ კლასტერს შორის მანძილად მიჩნეულია მანძილი/დაშორება მათ უახლოეს წერტილებს შორის (ჩვენი მაგალითის შემთხვევაში, უახლოესი წერტილებია **A** და **C**). შესაბამისად, ამ მეთოდის ფარგლებში აგრეგაციის კრიტერიუმია უმცირესი მანძილი უახლოეს წერტილებს/"მეზობლებს" შორის;
- **უშორესი მეზობლების მეთოდი.** ორ კლასტერს შორის მანძილად მიჩნეულია მანძილი/დაშორება მათ უშორეს წერტილებს შორის (ჩვენი მაგალითის შემთხვევაში, უშორესი წერტილებია **A** და **B**). შესაბამისად, ამ მეთოდის ფარგლებში აგრეგაციის კრიტერიუმია უმცირესი მანძილი/დაშორება უშორეს წერტილებს/"მეზობლებს" შორის;
- **აუწონავი წყვილური საშუალოს მეთოდი.** ორ კლასტერს შორის მანძილი გამოითვლება როგორც წერტილების ყველა წყვილს შორის არსებულ მანძილთა/დაშორებათა სა-

შუალო. შესაბამისად, ამ მეთოდის ფარგლებში აგრეგაციის კრიტერიუმია **უმცირესი საშუალო მანძილი/დაშორება;**

- **აწონილი წყვილური საშუალო.** კლასტერებს შორის მანძილი გამოითვლება აუწონავი წყვილური საშუალოს ანალოგიურად, მაგრამ გამოთვლებისას თითოეული კლასტერის მოცულობა (ე. ი. კლასტერში შემავალი ობიექტების/წერტილების რაოდენობა) განიხილება როგორც წონითი კოეფიციენტი;
- **აუწონავი ცენტროიდული მეთოდი.** კლასტერებს შორის მანძილი განისაზღვრება როგორც მანძილი/დაშორება მათ ცენტრებს შორის. შესაბამისად, ამ მეთოდის ფარგლებში აგრეგაციის კრიტერიუმია უმცირესი მანძილი/დაშორება კლასტერების ცენტრებს შორის;
- **აწონილი ცენტროიდული მეთოდი (მედიანა).** აუწონავი ცენტროიდული მეთოდის ანალოგიურია, მაგრამ გაერთიანებული კლასტერის ცენტრის გამოთვლისას თითოეული კლასტერის მოცულობა განიხილება როგორც წონითი კოეფიციენტი;
- **ვარდის მეთოდი.** კლასტერებს შორის მანძილის შეფასება ხდება დისპერსიული ანალიზის მეთოდების გამოყენებით.

კლასტერთა ურთიერთდაშორების/კლასტერებს შორის მანძილის ზომები და აგრეგაციის მეთოდები ერთმანეთს უნდა შეესაბამებოდეს.

მაგალითად, თუ აგრეგაციას ვარდის ან ცენტროიდული მეთოდის მეშვეობით ვაწარმოებთ, კლასტერთა ურთიერთდაშორება ევკლიდეს მანძილის კვადრატით უნდა გავზომოთ.

❖ ცვლადების სტანდარტიზება

კლასტერიზაცია რამდენიმე ნიშან-თვისების/ცვლადის მიხედვით იმ სირთულესთანაც არის დაკავშირებული, რომ საწყისი ცვლადების მნიშვნელობები სხვადასხვა ერთეულით არის წარმოდგენილი და ნებისმიერად შეიძლება იყოს "გაფანტული", რაც შედეგების ანალიზისა და ინტერპრეტაციის სირთულეს იწვევს. ეს პრობლემა ცვლადების სტანდარტიზებით წყდება.

კლასტერული ანალიზის დროს ცვლადების სტანდარტიზება სხვადასხვა მეთოდითაა შესაძლებელი:

- **Z-სკალები:** Z-სკალების მეთოდის გამოყენება გულისხმობს თითოეული ცვლადის საშუალო მნიშვნელობის გამოთვლასა და ამ მნიშვნელობის გაყოფას სტანდარტულ გადახრაზე;
- **ნორმირება (-1; 1):** ნორმირება −1-დან 1-ის ფარგლებში გულისხმობს წრფივი გარდაქმნების მეშვეობით ცვლადების მნიშვნელობების მოქცევას **−1-სა და 1-ს შორის;**
- **ნორმირება (0; 1):** ნორმირება 0-დან 1-ის ფარგლებში გულისხმობს წრფივი გარდაქმნების მეშვეობით ცვლადების მნიშვნელობების მოქცევას **0-სა და 1-ს შორის;**
- **მაქსიმუმზე გაყოფა:** მაქსიმუმზე გაყოფის მეთოდი გულისხმობს ცვლადების მნიშვნელობების გაყოფას მათ მაქსიმალურ მნიშვნელობაზე;
- **საშუალოზე გაყოფა:** ცვლადების მნიშვნელობები იყოფა მათ საშუალო მნიშვნელობაზე;
- **სტანდარტული გადახრა:** სტანდარტული გადახრის მეთოდი გულისხმობს ცვლადების მნიშვნელობების გაყოფას მათ სტანდარტულ გადახრაზე.

❖ არაიერარქიული კლასტერიზაციის მექანიზმი. ოპტიმალური განაწილების მეთოდი.

K საშუალოთა მეთოდი ძირითადად მაშინ გამოიყენება, როდესაც ხდება **ჰიპოთეზის შემოწმება სტატისტიკური ერთობლიობის სტრუქტურის** შესახებ.

დავუშვათ, შემუშავებულია ჰიპოთეზა, რომლის მიხედვითაც, სტატისტიკური ერთობლიობის სტრუქტურა **k** კლასტერისაგან შედგება.

K საშუალოთა მეთოდის მეშვეობით შესაძლებელია სტატისტიკური ერთობლიობა დავყოთ ზუსტად **k** კლასტერად, რომლებიც **ერთმანეთისაგან რაც შეიძლება მეტადაა დაშორებული/რომელთა შორისაც რაც შეიძლება მეტი მანძილია.**

K საშუალოთა მეთოდით კლასტერიზაციის მექანიზმი შემდეგნაირია:

- თავდაპირველად ხდება **k** კლასტერის **შემთხვევითი** შერჩევა;
- შემდეგ კლასტერებს შორის ობიექტები **გადანაცვლდება/გადანაწილდება** იმ წესით, რომ მოხდეს:
  - **ა) კლასტერებისშიდა მანძილების ვარიაციათა მინიმიზება.**
  - **ბ) კლასტერებსშორისი მანძილების ვარიაციათა მაქსიმიზება.**

K საშუალოთა მეთოდით კლასტერების გამყოფის შემდეგ უნდა შეფასდეს **რამდენად მნიშვნელოვანია განსხვავება/დისკრიმინაცია მიღებულ კლასტერებს** შორის.

კლასტერებს შორის განსხვავების **სტატისტიკური მნიშვნელოვნება** შეიძლება შეფასდეს **ფიშერის კრიტერიუმის** გამოყენებით კლასტერებისშიდა და კლასტერებშორისი მანძილების ვარიაციების შედარებითი ანალიზის საფუძველზე ან T-კრიტერიუმის (სტიუდენტის განაწილების) მეშვეობით: **კლასტერული ცენტროიდების[1] შედარებითი ანალიზის** საფუძველზე.

თუ კლასტერებს შორის განსხვავება შეფასდება როგორც მნიშვნელოვანი, ჰიპოთეზა სტატისტიკური ერთობლიობის k კლასტერად დაყოფის შესახებ ვერიფიცირებულად ჩაითვლება, ხოლო წინააღმდეგ შემთხვევაში — ფალსიფიცირებულად.

❖ კლასტერიზაციის მეთოდის შერჩევის პრობლემა

კლასტერიზაციის სხვადასხვა მეთოდის არსებობა, ბუნებრივია, ბადებს კითხვას: რის მიხედვით შევარჩიოთ კლასტერიზაციის კონკრეტული მეთოდი მონაცემთა ანალიზისას?

ამ კითხვაზე ერთმნიშვნელოვანი პასუხის გაცემა შეუძლებელია. გადაწყვეტილება მკვლევარმა უნდა მიიღოს კვლევის პირობებისა და კლასტერიზაციის სხვადასხვა მეთოდის სპეციფიკის მიხედვით, მაგრამ მიზანშეწონილია შემდეგი, ზოგადი სახის რეკომენდაციების გათვალისწინება:

- იერარქიული კლასტერიზაციის გამოყენება განსაკუთრებით პროდუქტიულია მაშინ, როდესაც სტატისტიკური ერთობლიობის სტრუქტურა უცნობია. ამ შემთხვევაში, კლასტერიზაცია ერთგვარი "დაზვერვითი" საშუალებაა იმ თვალსაზრისით, რომ ხდება მონაცემთა სტრუქტურის გამოვლენა, რის საფუძველზეც შესაძლებელია შემდგომი ანალიზის სწორი მიმართულებით წარმართვა/კორექტული დაგეგმვა;
- მონაცემთა სტრუქტურის შესახებ არსებული ჰიპოთეზის შემოწმების ოპტიმალური საშუალება არაიერაქიული კლასტერიზაციაა;

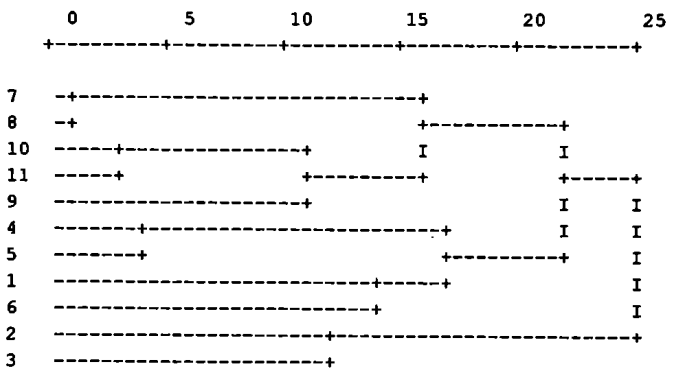
[1] **კლასტერული ცენტროიდი ეწოდება განზომილების/ნიშან-თვისების საშუალო მნიშვნელობას მოცემული კლასტერის ფარგლებში.**

- არაიერარქიული კლასტერიზაციის მეთოდები იერარქიულთან შედარებით უფრო სწრაფია და მისი გამოყენება მოხერხებულია ობიექტთა დიდი რაოდენობის მიმართ;
- რიგ შემთხვევაში, მიზანშეწონილია იერარქიული და არაიერარქიული მეთოდების ერთობლივი გამოყენება: თავდაპირველად, იერარქიული მეთოდის მეშვეობით შეიძლება განისაზღვროს კლასტერების რაოდენობა, შემდეგ კი, არაიერაქიული მეთოდით მოხდეს დადგენილი რაოდენობის კლასტერებად ობიექტთა ოპტიმალური განაწილება.

❖ დენდროგრამა

დენდროგრამა ანუ ეგრეთ წოდებული იერარქიული ხე კლასტერიზაციის შედეგების გრაფიკული გამოსახვის/ვიზუალიზაციის ერთ-ერთი გავრცელებული მეთოდია.

დენფროგრამის ზოგადი სახე შემდეგნაირია:



- დენდროგრამის გარჩევა (გაშიფვრა, "წაკითხვა") ხდება მარცხნიდან მარჯვნივ. + -ებს შორის არსებული ვერტიკალური დაშორებების მიხედვით შესაძლებელია ვიმსჯელოთ იმის შესახებ, თუ როგორ მიმდინარეობს აგრეგაციის პროცესი (როგორ ერთიანდება კლასტერები): დავადგინოთ, რომელი კლასტერები ერთიანდება თითოეულ მიღებულ კლასტერში, განვსაზღვროთ კლასტერების რაოდენობა და მათი აგრეგაციის მიმდევრობა.
- კლასტერების ზემოთ მოთავსებული ჰორიზონტალური მონაკვეთი დაშორებების/მანძილების ზომებს გამოსახავს. შესაბამისად, მასზე მითითებული ინტერვალების მეშვეობით შესაძლებელია განვსაზღვროთ რა დაშორების მიხედვით/რა მანძილზე მოხდა კლასტერთა აგრეგაცია.
- ეს ინფორმაცია შესაძლებელია გამოვიყენოთ როგორც წანამძღვარი გადაწყვეტილების მისაღებად იმის შესახებ, თუ როდის დავასრულოთ კლასტერიზაციის პროცესი ანუ **რამდენი კლასტერი გამოვყოთ** იმისათვის, რომ მოცემული სტატისტიკური ერთობლიობის სტრუქტურა რაც შეიძლება ოპტიმალურად აღიწეროს: კლასტერიზაცია შეიძლება დასრულებულად ჩავთვალოთ იმ ეტაპზე, რომლის შემდეგაც **მკვეთრად იზრდება მანძილი/დაშორება კლასტერებს შორის**.
- დენდროგრამის მეშვეობით, აგრეთვე, ვიღებთ თვალსაჩინო ინფორმაციას იმის შესახებ, ეკუთვნის თუ არა მოცემულ კლასტერს სტატისტიკური ერთობლიობის ესა თუ ის წევრი/ობიექტი.

მრავალგანზომილებიანი სკალირება

❖ მრავალგანზომილებიანი სკალირების ძირითადი მიზანი

მრავალგანზომილებიანი სკალირება, კლასტერული ანალიზის მსგავსად, ევრისტიკული ხასიათის მეთოდია და მისი ძირითადი მიზანია **მონაცემთა სტრუქტურის გამოვლენა/შესწავლა.**

მრავალგანზომილებიანი სკალირების ამოსავალია დაშვება იმის შესახებ, რომ ობიექტთა წყვილებს შორის მსგავსება-განსხვავების მიმართებებს განაპირობებს ძირითადი/საბაზისო მახასიათებლების (ნიშან-თვისებების) ლატენტური მეტრული სივრცის არსებობა. აქედან გამომდინარე, შესაძლებელია, რომ ობიექტები მოვიაზროთ როგორც ამ სივრცის წერტილები.

ამასთან, ობიექტებსა და წერტილებს შორის იგულისხმება შემდეგი სახის შესაბამისობა: რაც უფრო მსგავსია ობიექტები, მით მცირეა დაშორება/მანძილი საბაზისო მახასიათებლების სივრცეში მათ შესატყვის წერტილებს შორის.

შესაბამისად, მრავალგანზომილებიანი სკალირება შეიძლება განიმარტოს, როგორც იმ მეთოდების ერთობლიობა, რომელთაც მეშვეობითაც ხორციელდება:

- არსებით მახასიათებელთა (ნიშან-თვისებათა) სივრცის აგება;
- არსებით მახასიათებელთა სივრცეში ობიექტების/წერტილების კონფიგურაციის კონსტრუირება.

არსებით მახასიათებელთა სივრცე, ფაქტობრივად, წარმოადგენს მრავალგანზომილებიან სკალას, რომელიც ჩვეულებრივი სკალის ანალოგიურია იმ თვალსაზრისით, რომ ობიექტებს შეესაბამება გარკვეული პროექციები/კოორდინატები ამ სივრცის ღერძებზე.

მრავალგანზომილებიანი სკალირების კონკრეტული ამოცანებია:

- მრავალგანზომილებიანი მეტრული სივრცის აგება, რომელიც მონაცემთა სტრუქტურის ასახვას უზრუნველყოფს;
- მრავალგანზომილებიან მეტრულ სივრცეში ობიექტთა სკალირება, ე. ი. მათი ღერძებზე პროეცირება;
- საწყის მონაცემთა სტრუქტურის ვიზუალიზაცია — მისი წარმოდგენა მეტრულ სივრცეში წერტილთა კონფიგურაციის სახით;
- მეტრული სივრცისა და წერტილთა კონფიგურაციის ინტერპრეტაცია.

მრავალგანზომილებიანი სკალირების საბოლოო პროდუქტია მონაცემთა **სტრუქტურის აღწერა და ობიექტთა მსგავსება-განსხვავების დახასიათება,** რაც, თავის მხრივ, მათი შინაარსობრივი ინტერპრეტაციის საფუძველია და მონაცემთა ანალიზის შემდგომი ეტაპების მიმართულებას განსაზღვრავს.

მრავალგაზომილებიანი სკალირების მეთოდები ორ ძირითად ჯგუფად იყოფა: არამეტრულ მრავალგანზომილებიან სკალირებად და მეტრულ მრავალგანზომილებიან სკალირებად.

არამეტრული მრავალგანზომილებიანი სკალირება გამოიყენება გაზომვის რიგის დონეზე, ხოლო მეტრული — რაოდენობრივზე.

ემპირიულ სოციოლოგიურ კვლევებში მრავალგანზომილებიანი სკალირების მეთოდის გამოყენება შეიძლება გავიაზროთ, როგორც **"სოციალური სივრცის" კარტოგრაფირება,** ხოლო მიღებული შედეგი — როგორც **რუკა,** რომელიც ამ სივრცის სტრუქტურას ასახავს. ამიტომ, მრავალგანზომილებიანი სკალირების შედეგს ხშირად **სივრცულ რუკას** უწოდებენ.

❖ მრავალგანზომილებიანი სკალირების გამოყენების სპეციფიკა

მრავალგანზომილებიანი სკალირების გამოყენებით სოციოლოგიაში სხვადასხვა ამოცანა შეიძლება გადაიჭრას.

მისი გამოყენება განსაკუთრებით ეფექტურია იმ შემთხვევებში, როდესაც გვაინტერესებს:

- საკვლევ სოციალურ მოვლენათა აღქმის, მათდამი არსებული განწყობებისა და სხვა სახის დამოკიდებულებების შესწავლა;
- საკვლევი სოციალური მოვლენების რანჟირება რესპონდენტთა მიერ მინიჭებული უპირატესობების (პრიორიტეტულობის) მიხედვით და უპირატესობათა მინიჭების მექანიზმის შესწავლა.

სივრცულ რუკაზე საკოორდინატო სისტემის ღერძები ძირითად მახასიათებლებს შეესაბამება. ისინი, ფაქტობრივად, ის ლატენტური განზომილებები/ნიშან-თვისებებია, რომელთა მიხედვითაც არსებითად განისაზღვრება საკვლევი მოვლენების აღქმა, ძირითადი დამოკიდებულებები მათ მიმართ, არჩევანი უპირატესობების მინიჭების დროს[1] და ა. შ.

მრავალგანზომილებიანი სკალირების შედეგების ანალიზსა და ინტერპრეტაციას მნიშვნელოვნად აიოლებს მონაცემთა მოპოვება ადეკვატური მეთოდებით.

აღქმის, განწყობებისა და სხვა სახის დამოკიდებულებების კვლევისას მონაცემთა მოპოვების მეთოდები შეიძლება ეყრდნობოდეს პირდაპირ ან არაპირდაპირ მიდგომას.

პირდაპირი მიდგომა გულისხმობს მონაცემთა მოპოვებას ისეთი წესით, რომლის მეშვეობითაც დაფიქსირდება რესპონდენტთა მიერ **პირდაპირ** (უშუალოდ) **მსგავსება-განსხვავების შეფასება** საკვლევი მოვლენების/ობიექტების წყვილებს შორის.

მაგალითად, ეროვნული სტერეოტიპების კვლევის დროს შეიძლება შეფასდეს მსგავსება-განსხვავება საკვლევ ეროვნებათა წარმომადგენელებს შორის:

	სრულიად არ ჰგავს						ძალიან ჰგავს
ქართველი-აფხაზი:	1	2	3	4	5	6	7
ოსი–ქართველი	1	2	3	4	5	6	7
ქართველი-რუსი	1	2	3	4	5	6	7
და ა. შ.							

არაპირდაპირი მიდგომა გულისხმობს მონაცემთა მოპოვებას ისეთი წესით, რომელის დროსაც რესპონდენტები აფასებენ საკვლევ ობიექტებს გარკვეული მახასიათებლების/ნიშან-თვისებების მიხედვით, დავუშვათ, სემანტიკურ-დიფერენციალური სკალის ან ლაიკერტის სკალის გამოყენებით.

მაგალითად, თითოეული ეროვნება შეიძლება შეფასდეს სხვადასხვა მახასიათებლის მიხედვით:

ქართველი

სამართლიანი	1	2	3	4	5	6	7	უსამართლო
მამაცი	1	2	3	4	5	6	7	მშიშარა
შრომისმოყვარე	1	2	3	4	5	6	7	ზარმაცი
და ა. შ.								

[1] მივაქციოთ ყურადღება, რომ ამ ასპექტით, მრავალგანზომილებიანი სკალირების ამოცანა ფაქტორული ანალიზის ამოცანასთან იკვეთება.

უპირატესობათა მინიჭების მექანიზმების კვლევისას მონაცემთა მოპოვება, ჩვეულებრივ, გარკვეული ნიშან-თვისებების მიხედვით მინიჭებული უპირატესობების რანჟირებით ხდება.

უპირატესობათა რანჟირება სხვადასხვა წესით შეიძლება განხორციელდეს.

მაგალითად:

- რესპონენტებს ვთხოვოთ, რომ დაალაგონ პოლიტიკური პარტიები იმის მიხედვით, თუ რომელი მათგანი უკეთ შეძლებს ამომრჩეველთა ამა თუ იმ პრობლემის მოგვარებას;
- პარტიათა სათითაო წყვილს შორის გააკეთონ არჩევანი იმის მიხედვით, თუ რომელი მათგანი უკეთ შეძლებს ამომრჩეველთა ამა თუ იმ პრობლემის მოგვარებას;
- შეაფასონ თითოეული პარტია იმის მიხედვით, თუ რამდენად შეძლებს იგი ამომრჩეველთა ამა თუ იმ პრობლემის მოგვარებას და ა. შ.

❖ მრავალგანზომილებიანი სკალირების მოდელი

მრავალგანზომილებიან სივრცეში ყოველი ობიექტი წარმოდგენილია $\mathbf{x}_i$ წერტილის სახით. უმარტივეს შემთხვევაში, $\mathbf{x}_i$ და $\mathbf{x}_j$ წერტილებს შორის მანძილის მნიშვნელობა $\mathbf{d}_{ij}$ გამოითვლება ფორმულით:

$$\mathbf{d}_{ij} = \mathbf{d}(\mathbf{x}_i, \mathbf{x}_j) + \text{ცდომილება},$$

სადაც $\mathbf{d}(\mathbf{x}_i, \mathbf{x}_j)$ არის ევკლიდეს მანძილი $\mathbf{x}_i$ და $\mathbf{x}_j$ წერტილებს შორის და გამოითვლება ფორმულით:

$$\mathbf{d}_{ij} = \sqrt{\Sigma_i(x^k_i - y^k_j)^2}$$

ხშირ შემთხვევაში, გამოთვლებისას საკმარისია ვისარგებლოთ ორ ან სამგანზომილებიანი ევკლიდეს სივრცით.

მეტრული მრავალგანზომილებიანი სკალირების მოდელი, ზოგადად, შემდეგი სახის ფუნქციით შეიძლება აღინეროს:

$$\mathbf{f}(\mathbf{d}_{ij}) = \mathbf{d}(\mathbf{x}_i, \mathbf{x}_j) + \text{ცდომილება}.$$

ტორგერსონის მოდელი.

ტორგერსონის მოდელი **მეტრული** მრავალგანზომილებიანი სკალირების მოდელია. იგი გამოისახება ფორმულით:

$$\mathbf{L(S)} = \mathbf{D^2} + \mathbf{E},$$

სადაც **L(S)** არის მანძილთა/დაშორებათა თავდაპირველი/საწყისი მატრიცის **წრფივი გარდაქმნა**, $\mathbf{D^2}$ მატრიცა, რომელიც სკალირების შედეგად მიიღება, ხოლო **E** მოდელის გადახრა საწყისი მონაცემებიდან.

ტორგერსონის მოდელის ფარგლებში მრავალგანზომილებიანი სკალირების ძირითად ამოცანაა საწყის მანძილთა/დაშორებათა მატრიცის **წრფივი** გარდაქმნისა და სკალის ისეთი ოპტიმალური ვარიანტების პოვნა, რომელთა მეშვეობითაც მოხდება **E** ცდომილების მინიმიზება.

შეპარდისა და კრასკალის მოდელი.

შეპარდისა და კრასკალის მოდელი არამეტრული მრავალგანზომილებიანი სკალირების მეთოდია. იგი გამოისახება ფორმულით:

$$\mathbf{M(S)} = \mathbf{D^2} + \mathbf{E},$$

სადაც **M(S)** არის მანძილთა/დაშორებათა თავდაპირველი/საწყისი მატრიცის **მონოტონური გარდაქმნა**, $\mathbf{D^2}$ მატრიცა, რომელიც სკალირების შედეგად მიიღება, ხოლო **E** მოდელის გადახრა საწყისი მონაცემებიდან.

შეპარდისა და კრასკალის მოდელის ფარგლებში მრავალგანზომილებიანი სკალირების

ძირითადი ამოცანაა სანყის მანძილთა/დაშორებათა მატრიცის მონოტონური გარდაქმნისა და სკალის ისეთი ოპტიმალური ვარიანტების პოვნა, რომელთა მეშვეობითაც მოხდება **E** ცდომილების მინიმიზება.

❖ მრავალგანზომილებიანი სკალირების მოდელის ხარისხი

მრავალგანზომილებიანი სკალირების მოდელის ხარისხი სხვადასხვა მეთოდით შეიძლება შეფასდეს. ასეთი მეთოდების მაგალითებია:

S-Stress მაჩვენებელი. იგი გამოითვლება ფორმულით:

$$\text{S-Stress} = \sqrt{\frac{||E||}{||Y||}}$$

სადაც || || წარმოადგენს მატრიცის ნორმას და გამოითვლება მატრიცის ელემენტთა კვადრატების ჯამით, ხოლო სიტყვა **Stress** გამოიყენება "დატვრითვის" მნიშვნელობით.

S-Stress მაჩვენებლის ზღვრული მნიშვნელობებია 0 და 1.

0 მაქსიმალური სიზუსტის/ ადეკვატურობის მაჩვენებელია, ხოლო 1 — უზუსტობის/არაადეკვატურობის.

შესაბამისად, რაც უფრო ახლოსაა **S-Stress** ნულთან, მით ადეკვატურია მოდელი და რაც უფრო უახლოვდება 1-ს მით არაადეკვატურია იგი.

კრასკალის **S-Stress index** მაჩვენებელი. მისი გამოთვლისას ხდება უშუალოდ დაშორებებების და არა მათი კვადრატების გამოყენება.

RSQ მაჩვენებელი. იგი **T** და **E** მატრიცებს შორის კორელაციის კოეფიციენტის კვადრატია. შესაბამისად, შეიძლება ინტერპრეტირდეს როგორც გარდაქმნილ **T** მანძილთა/დაშორებათა ხვედრითი წილი, რომელიც **D** დაშორებათა მატრიცით აიხსნება. მოდელის ხარისხი დამაკმაყოფილებლად ითვლება, როდესაც **RSQ** მაჩვენებელი უდრის 0.06.

❖ განზომილებათა რაოდენობის შერჩევა

რაც უფრო მცირეა სივრცული რუკის განზომილებათა რაოდენობა და იოლია მისი მეშვეობით მონაცემთა გაანალიზება, მით უკეთესია მრავალგანზომილებიანი სკალირების მოდელი.

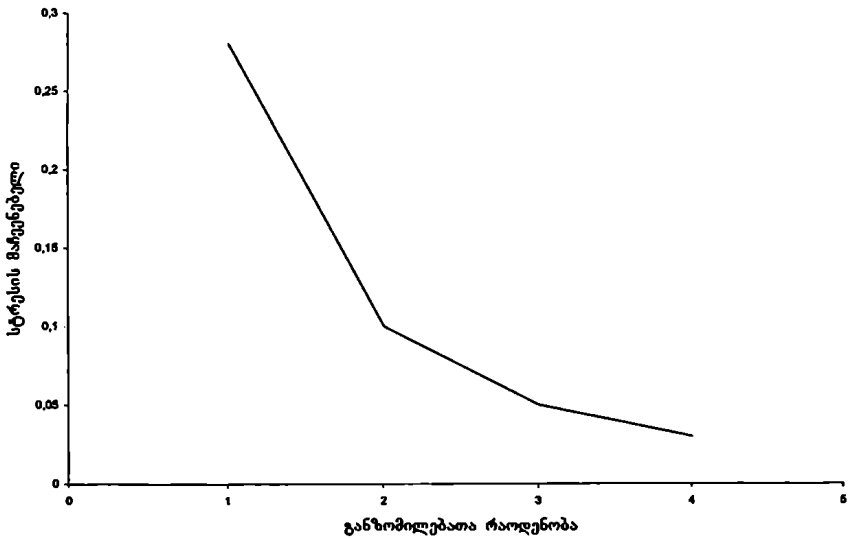
შესაბამისად, განზომილებათა რაოდენობის განსაზღვრა მრავალგანზომილებიანი სკალირების ერთ-ერთი მნიშვნელოვანი პრობლემაა.

ამ პრობლემის გადაწყვეტა ხდება მკვლევრის მიერ კვლევის კონკრეტული პირობებისა და სტატისტიკური მაჩვენებლების მიხედვით, მაგრამ მიზანშეწონილია შემდეგი, ზოგადი ხასიათის შინაარსობრივი რეკომენდაციების გათვალისწინებაც:

- განზომილებათა რაოდენობის განსაზღვრისას შეიძლება გამოვიყენოთ წინა კვლევების შედეგები;
- მიღებული შედეგები ინტერპრეტაციას უნდა ექვემდებარებოდეს. უმეტეს შემთხვევაში, სამზე მეტი განზომილების შემთხვევაში ინტერპრეტაცია შესამჩნევად რთულდება;
- ანალიზისათვის უფრო იოლია ორგანზომილებიანი სივრცული რუკის გამოყენება.

განზომილებათა რაოდენობის განსაზღვრისას ხშირად იყენებენ სივრცის განზომილებებზე

სტრესის დამოკიდებულების გრაფიკს. ეს გრაფიკი გვიჩვენებს **რა მინიმალურ რაოდენობაზე შეიძლება დავიყვანოთ განზომილებათა რიცხვი.** იგი, როგორც წესი, ჩაზნექილი მრუდია და გადანყვეტილება განზომილებათა რაოდენობის განსაზღვრის შესახებ, შესაძლებელია მივიღოთ იმ წერტილის მიხედვით, სადაც გრაფიკის მოხრილობა/სიმრუდე მკვეთრად მატულობს.



ჩვენი მაგალითის შემთხვევაში, მრუდის ის ნაწილი, ე. ი. სტრესის მაჩვენებლის ის მნიშვნელობები, რომლებიც შეესაბამება 2, 3, და 4 განზომილებას მკვეთრად არ იცვლება. ეს იმას ნიშნავს, რომ მოდელის ხარისხი არ არის არსებითად დამოკიდებული იმაზე, ორ, სამ თუ ოთხ განზომილებიან სივრცეს ავაგებთ, ამიტომ მათ შორის **უმცირესს** შეიძლება მივანიჭოთ უპირატესობა.

ერთი განზომილების შემთხვევაში კი სიმრუდე მკვეთრად მატულობს, ე. ი. სტრესის მაჩვენებელი მკვეთრად იზრდება. ეს იმას ნიშნავს, რომ ერთი განზომილების შერჩევა მოდელის ხარისხის მკვეთრ გაუარესებას გამოიწვევს.

შესაბამისად, მოცემული გრაფიკის მიხედვით, განზომილებათა მინიმალური რაოდენობა შეიძლება განისაზღვროს ორით.

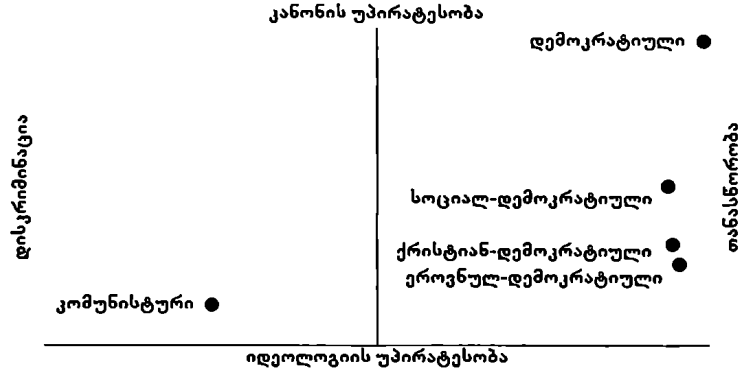
ე. ი. ჩვენს მიერ ჩატარებული მსჯელობის რეზულტატი ორგანზომილებიანი სივრცის/სისტემის აგების სასარგებლოდ მეტყველებს.

❖ წერტილთა კონფიგურაციის ინტერპრეტაცია სივრცულ რუკაზე

როგორც უკვე აღვნიშნეთ, მრავალგანზომილებიანი სკალირების შედეგად აგებული სივრცული რუკის ინტერპრეტაცია გულისხმობს საკოორდინატო სისტემის ღერძებისა და ამ სისტემაში წერტილთა კონფიგურაციის ინტერპრეტაციას.

განვიხილოთ სივრცული რუკის მაგალითი, რომელიც ერთ-ერთი კვლევის დროს სხვადასხვა მახასიათებლების (ნიშან-თვისებების) მიხედვით პოლიტიკურ შეხედულებათა სისტემების შე-

ფასების საფუძველზე იქნა მიღებული[1].



საკოორდინატო სისტემის ღერძების ინტერპრეტაცია ხდება ზოგადი ფაქტორების ინტერპრეტაციის ანალოგიურად, ამიტომ, ამ საკითხზე დაწვრილებით აღარ შევჩერდებით. დავუშვათ, რომ ჰორიზონტალური ღერძი ინტერპრეტირდება როგორც ოპოზიცია: დისკრიმინაცია–თანასწორობა, ხოლო ვერტიკალური ღერძი, როგორც ოპოზიცია: იდეოლოგიის უპირატესობა–კანონის უპირატესობა. მაშინ:

- პოლიტიკურ შეხედულებათა დემოკრატიული სისტემა რესპონდენტთა მიერ აღიქმება როგორც თანასწორობისა და კანონის უპირატესობის გამომხატველი;
- პოლიტიკურ შეხედულებათა კომუნისტური სისტემა აღიქმება როგორც დისკრიმინაციისა და იდეოლოგიის უპირატესობის გამომხატველი;
- დემოკრატიული და კომუნისტური სისტემები აღიქმება არა მხოლოდ როგორც განსხვავებული, არამედ, როგორც ურთიერთგამომრიცხავი;
- სოციალ-დემოკრატიული, ქრისტიან-დემოკრატიული და ეროვნულ-დემოკრატიული სისტემები აღიქმება როგორც თანასწორობისა და იდეოლოგიის უპირატესობის გამომხატველი;
- ქრისტიან-დემოკრატიული და ეროვნულ-დემოკრატიული სისტემები აღიქმება როგორც უფრო ერთმანეთის მსგავსი, ვიდრე სხვა სისტემების;
- სოციალ-დემოკრატიული სისტემა აღიქმება როგორც უფრო განსხვავებული სოციალ-დემოკრატიული და ქრისტიან-დემოკრატიული სისტემებისაგან, ვიდრე ეს უკანასკნელნი ერთმანეთისაგან და ა. შ.

[1] მაგალითის თვალსაჩინოებისათვის პოლიტიკურ შეხედულებათა სისტემების ჩამონათვალი შემცირებულია.

ANALYSIS OF SOCIOLOGICAL DATA
Expanded Syllabus

Summary
The main purpose of the Course is the acquaintance with the methodology of analysis of sociological information. It includes 8 chapters intended for learning the basic methods, techniques and formal languages of analysis for different kinds of sociological information.

The Course begins with the introduction into the principles and basic concepts of the methodology of analysis of empirical sociologic data. A general survey of basic methods of analysis of sociological information is given; the logic and program of sociological research are outlined. Further, the methods of measurement in sociology are considered, different scales and the specificity of their application in sociology are described.

Two chapters of the Course deal with the methods of construction and analysis of one-dimensional and two-dimensional frequency distributions.

An important part of the Course is devoted to the description of various models used in sociology, viz.: regression models, dispersion analysis models and factor analysis models. The methods of cluster analysis and multidimensional scaling are also considered.

Chapter 1.1. Structure of empiric sociological investigation
Logic and program of sociological investigation. General survey of the methods of analysis of sociological information. The role and place of hypotheses and their classification. Methodological part of the program. Sampling. Preparation of data for analysis.

The following issues are considered in detail:

Structure of empiric sociological study. Correlation of stages of empiric sociological study.

- **Analysis of basic conceptions.**
- **Role of hypotheses in empiric sociological study.**

Sampling.

Information about the characteristics of a population can be obtained either by sampling or by a census.

Sampling techniques may be classified as non-probability and probability techniques.

Non-probability sampling.

Non-probability sampling techniques depend on the aim of investigation and rely on the researcher's judgment.
1. Convenience sampling.
2. Judgmental sampling.
3. Quota sampling.
4. Snowball sampling.

Probability sampling.

In probability sampling techniques, sampling units are selected by chance. One also can determine the precision of the sample estimates and inferences and make projections to the target population.

1. Simple Random sampling – SRS.
2. Systematic sampling.
3. Stratified sampling.
4. Cluster sampling/ area sampling.
5. Sequential sampling.
6. Double sampling.

Preparation of data for analysis.

Data preparation begins with a preliminary check of all questionnaires for completeness and interviewing quality.

- **Editing.** It consists of screening questionnaires to identify illegible, incomplete, inconsistent, or ambiguous responses.
- **Treatment of unsatisfactory responses**
- **Assigning missing values.**
- **Discarding unsatisfactory respondents.**
- Coding questions. Codebook. Coding questionnaires. **A numeric code is assigned to represent a specific response to a specific question, along with the column position that code will occupy.**
- **Transcribing.**
- **Data cleaning.** Cleaning the data requires consistency checks and treatment of missing response.
  1. **Consistency checks.**
  2. **Treatment of missing responses.**
  3. **Casewise deletion.**
  4. **Pairwise deletion.**
- Weighting.
- Variable respecification.
- Dummy variable.
- Scale transformation.
- **Selecting a data analysis strategy.** The selection of a data analysis strategy should be based on the earlier steps of the research process, known characteristics of the data, properties of statistical techniques, and the background and judgment of the researcher.

Chapter 1.2. The methods of measurement in sociology.

The problem of measurement of the sociological data. The scales used in sociology: naming scale, order scale, interval scale, absolute scale. The fields of application of the scales, their properties and the specificity of their use in sociology.

The following issues are considered in detail:

The concepts of measurement **and** variable **in sociology.**

- **Classification of variables**: nominal variable, ordinal variable, quantitative variables: interval variable and absolute variable.
- **Dichotomous variables.**
- **Classification of scales**: nominal scale, ordinal scale, quantitative scales: interval scale and absolute scale.
- **Specificity** of the usage of scales.
- **Methods of scaling:**

I. Comparative scales:

1. Paired comparison scaling.
2. Rank-order scaling.
3. Constant sum scaling.
4. Q-sort scaling.
5. Scaling by Guttmann's method.

II. Non-comparative scales:

1. Continuous rating scale.
2. Itemized rating scale.
3. Likert scale/ summated scale.
4. Semantic differential.
5. Stapel scale.

Chapter 2. The one-dimensional frequency distributions.

Methods of construction of one-dimensional distributions. Statistics of one-dimensional distributions. One-dimensional frequency distribution and one-dimensional statistics: measures of central tendency and measures of dispersion.

"Descriptive" and "conjecturing" strategies of data analysis. The concept of mathematical model. Statistics of one-dimensional distribution as indicators of the quality of a model.

Hypothesis testing.

The following issues are considered in detail:

- **Frequency distributions for discrete and continuous variables.**
- **Relative frequency distributions, or proportions and percentage distributions.**
- **Grouped and cumulative distributions.**
- **Measurement interval or measurement class.**
- **Statistical table.**
- **Visual display** of quantitative information. Diagrams or graphs. Histogram. Polygon.
- **Measures of central tendency.**
- **Mean.** The arithmetic mean/average is the most common measure of central tendency. It can be calculated only for continuous distributions. The values of all observations are added together and divided by the number of observations.
- **Median:** the outcome that divides an ordered distribution exactly into halves. The median applies only to variables whose categories can be ordered from lowest to highest.
- **Mode:** the category with the largest number/highest percentage of observations.
- **Bimodal distribution.**
- **Measures of dispersion.**
- **Index of diversity and qualitative variation.** The index of diversity, D, measures whether two observations selected randomly from a population are likely to fall into the same or into different categories. To calculate D, we simply square the proportion of cases in each of the N discrete categories, sum these squares, and subtract from 1.
- **Range.** A distribution's range is defined as the difference between the largest and smallest scores.
- **Average absolute deviation:** the mean of the absolute values of the difference between a set of continuous measured and their mean.
- **Variance:** the mean square deviation of a continuous distribution.
- **Standard deviation:** the positive square root of the variance.
- **Coefficients of variation.**
- **Symmetric distributions and nonsymmetrical distribution.**
- **Normal distribution.**
- Skewness. **Skewed distribution – a distribution that is nonsymmetrical about its median value, having many categories with small frequencies at one end.**
- Positive skew – **the tail of a skewed distribution is to the right of the median/mean greater than median.**
- Negative skew – **the tail of a skewed distribution is to the left of the median/median greater than mean.**
- Standardized scores/Z Scores: **a transformation of the scores of a continuous frequency distribution by subtracting the mean from each outcome and dividing by the standard deviation.**

Hypothesis testing:

- **A general procedure for hypothesis testing.**
- **Formulating the hypothesis.** Null hypothesis – a statistical hypothesis that one usually expects to reject. Symbolized H_0. Alternative hypothesis – a secondary hypothesis about the

value of a population parameter that often reverses the research or operational hypothesis. Symbolized H_a.

- Selecting an appropriate test.
- Type I error/false rejection error: **a statistical decision error that occurs when a true null hypothesis is rejected, its probability is alpha.**
- **Probability (alpha) level:** the probability selected for rejection of a null hypothesis, which is the likelihood of making a Type I error.
- **Type II error/false acceptance error:** a statistical decision error that occurs when a false null hypothesis is not rejected, its probability is beta.
- **Power of a test:** a probability of correctly rejecting H_0 when H_0 is false.
- **Suspending judgment:** a position taken by a researcher when the results of a statistical test permit neither clear rejection nor clear acceptance of the null hypothesis or alternative hypothesis.
- **t test:** a test of significance for continuous variables where the population variances is unknown and the sample is assumed to have been drawn from a normally distributed population.
- **One-tailed hypothesis test:** a hypothesis test in which the alternative is stated in such a way that the probability of marking a Type I error is entirely in one tail of a probability distribution.
- **Two-tailed hypothesis test:** a hypothesis test in which the region of rejection falls equally within both tails of the sampling distribution.

Chapter 3. The two-dimensional frequency distributions.

The models of statistical relation in two-dimensional tables. The idea of relation coefficient. χ-square coefficient and its derivative coefficients.

The idea of range correlation coefficient. Gamma-coefficient, Spearman's and Candle's coefficients.

Cross-tabulation are tables that reflect the joint distribution of two or more variables. In cross-tabulation, the percentages can be computed either columnwise, based on column totals, or rowwise, based on row totals. The general rule is to compute the percentages in the direction of the independent variable, across the dependent variables.

Often the introduction of a third variable can provide additional insights. The chi-square statistic is a test of the statistical significance of the observed association in a cross-tabulation.

Parametric and nonparametric tests are available for testing hypothesis related to differences. In the parametric case, the *t test* is used to examine hypothesis related to the population mean. Different forms of the *t test* are suitable for testing hypotheses based on one sample, two independent samples, or paired samples.

The following issues are considered in detail:

Cross-tabulation/contingency tables.

- **Chi-square statistic.**
- **Chi-square distribution:** a family of distributions, each of which has different degrees of freedom, on which the chi-square test statistics is based.
- **Contingency coefficient.**
- **Parametric tests and nonparametric tests.**
- ***F-test*/F-distribution:** a theoretical probability distribution for one of a family of F ratios having v_1 and v_2 *df* in the numerator and denominator, respectively.

Chapter 4. Regression models.

Simple and multiple linear regression model. Coefficient of multiple correlation and determination coefficient. Regression coefficients. Restrictions of the regression model.

Bivariate regression derives a mathematical equation between a single metric criterion variable and a single metric predictor variable. The equation is derived in the form of a straight line by using the least-squares procedure.

When the regression is run on standardized data, the intercept assumes a value of 0, and the regression coefficients are called beta weights.

The strength of association is measured by the coefficient of determination, r^2, which is obtained by computing a ratio of SS_{reg} to SS_y.

The standard error of estimate is used to assess the accuracy of prediction and may be interpreted as a kind of average error made in predicting Y from the regression equation.

Multiple regression involves a single dependent variable and two or more independent variables.

Nominal or categorical variables may be used as predictors by coding them as dummy variables.

The following issues are considered in detail:

- Dependent or **criterion** variable and independent or **predictor** variable/variables.
- **Scatterplot:** a type of diagram that displays the co-variation of two continuous variables as a set of points on a coordinate system.
- **Bivariate regression:** a regression of Y on X.
- **Prediction equation:** a regression equation without the error term, useful for predicting the score on the dependent variable from the independent variable, or variables.
- **Linear regression model:** a model that takes into account deviations from the linear prediction by showing the linear relationship between a continuous dependent variable and one or more independent variables, plus error term.
- **Error term:** the difference between an observed score and a score predicted by the model.
- **Residual:** the amount that remains after subtracting the prediction equation from the linear regression model.
- **Least squares:** a method for obtaining estimates of regression equation coefficients that minimizes error sum of squares.
- **Bivariate regression coefficient:** a parameter estimate of a bivariate regression equation that measures the amount of increase or decrease in the dependent variable for a one-unit difference in the independent variable.
- **Covariance:** the sum of the product of deviations of the Xs and Ys about their respective means, divided by N-1 in the sample and N in the population.
- **Partitioning the sum of squares:** $SS_{TOTAL} = SS_{REGRESSION} + SS_{ERROR}$.
- **Coefficient of determination** expresses the amount of variation in the dependent variable explained or accounted for by the independent variable or variables in a regression equation.
- **Correlation coefficient:** a measure of association between two continuous variables that estimates the direction and strength of linear relationship.
- **Homoscedasticity:** a condition in which the variances of the prediction errors are equal at every outcome of the predictor variable.
- **Comparing two regression models.**
- **Correlation difference test:** a statistical test to determine whether two correlation coefficients differ in the population.
- **Regression difference test:** a statistical test to determine whether two correlation coefficients differ in the population.
- **Multiple regression analysis:** a statistical technique for estimating the relationship between a continuous dependent variable and two or more continuous or discrete independent/predictor variables.

- **Index:** a variable that is a summed composite of other variables that are assumed to reflect some underlying construct.
- **Construct:** unobserved concept used by social scientists to explain observation.
- **Indicator:** observable measure of underlying unobservable theoretical construct.
- **Multiple regression coefficient:** a measure of association showing the amount of increase or decrease in a continuous dependent variable for a one-unit different variable, controlling for the other independent variable/variables.
- **Part correlation:** a measure of the proportion of variance in a dependent variable that an independent variable can explain, when squared, after independent variable in a multiple regression equation.
- **Multiple correlation coefficient:** the coefficient for a multiple regression equation, which, when squared, equals the ratio of the sum of squares due to regression to the total sum of squares.
- **Standardization. Partial regression coefficient/β coefficients:** the effect of regressing a dependent variable on an independent variable, controlling for one or more other independent variable.
- **Multiple regression with K independent variable.**
- **Multicollinearity.**

Chapter 5. Dispersion analysis models.

Dispersion analysis. Possibilities and restrictions. Single-factor (ANOVA) and multi-factor (MANOVA) dispersion models. Methods of pair comparisons in the dispersion analysis. Possibilities of visualization of the results of a single-factorial dispersion analysis.

In ANOVA the dependent variable is metric and the independent variables are all categorical, or combinations of metric and categorical variables.

One-way ANOVA involves a single independent categorical variable.

The total variation in the dependent variable is decomposed into two components: variation related to the independent variable and variation related to error.

The null hypothesis of equal means is tested by an *F statistic*, which is the ratio of the mean square related to the independent variable to the mean square related to error.

N-way analysis of variance involves the simultaneous examination of two or more categorical independent variables. A major advantage is that the interactions between the independent variables can be examined. The significance of the overall effect, interaction terms, and main effects of individual factors is examined by appropriate ***F tests***. It is meaningful to test the significance of main effects only if the corresponding interaction terms are not significant.

In repeated measures analysis of variance, observations on each subject are obtained under each treatment condition. This design is useful for controlling for the differences in subjects that exist prior to the experiment. Non-metric analysis of variance involves examining the differences in the central tendencies of two or more groups when the dependent variable is measured on an ordinal scale.

Multivariate analysis of variance (MANOVA) involves two or more metric dependent variables.

The following issues are considered in detail:

- **The ANOVA model:** an ANOVA asks what proportion of the total variation in dependent variable Y can be attributed to individual *i*'s membership in the *j*-th group classification.
- **Error term in ANOVA:** that part of an observed score that cannot be attributed to either the common component or the group component.
- **Total sum of squares:** a number obtained by subtracting the scores of a distribution from their mean, squaring, and summing these values.

- **Between sum of squares:** a value obtained by subtracting the grand mean from each group mean, squaring these differences for all individuals and summing them.
- **Within sum of squared:** a value obtained by subtracting each subgroup mean from each observed score, squaring, and summing .
- **Homoscedasticity:** a condition in which the variances of two or more population distributions are equal.
- **One-way analysis of variance.**
- **N-way analysis of variance.**
- **Multivariate analysis of variance – MANOVA.**
- **Analysis of covariance – ANCOVA.**
- **Interaction.** Ordinal interaction. Disordinal interaction.
- **Multiple η^2.**
- **Significance of the overall effect.**
- **Significance of the interaction effect.**
- **Significance of the main effect of each factor.**
- **Multiple comparison tests.**

Chapter 6. Factor analysis models.

Introduction into latent variables models. Basic ideas of factor analysis. The possibilities and restrictions of a factor model. The problem of definition of number of the factors. The problem of omitted data and possibilities of its solution.

In analysis of variance and multiple regression, one variable is considered as the dependent or criterion variable and the others as independent or predictor variable/variables. However, no such distinction is made in factor analysis. Rather, factor analysis is an interdependent technique in that an entire set of interdependent relationship is examined.

Factor analysis is a class of procedures used for reducing and summarizing data. Each variable is expressed as a linear combination of the underlying factors. Likewise, the factors themselves can be expressed as linear combinations of the observed variables. The factors are extracted in such a way that the first factor accounts for the highest variance in the data, the second the next highest, and so on. Additionally, it is possible to extract the factors so that the factors are uncorrelated, as in principal components analysis.

In formulating the factor analysis problem, the variables to be included in the analysis should be specified based on past research, theory, and the judgment of the researcher. These variables should be measured on an interval or ratio scale. Factor analysis is based on a matrix of correlation between the variables.

The two basic approaches to factor analysis are principal components analysis and common factor analysis. In principal components analysis, the total variance in the data is considered. Principal components analysis is recommended when the researcher's primary concern is to determine the minimum number of factors that will account for maximum variance in the data for use in subsequent multivariate analysis.

In common factor analysis, the factors are estimated based only on the common variance. This method is appropriate when the primary concern is to identify the underlying dimensions, and when the common variance is of interest. This method is also known as principal axis factoring.

Factor scores can be computed for each respondent. Alternatively, surrogate variables may be selected by examining the factor matrix and selecting for each factor a variable with me highest or near-highest loading. The differences between the observed correlations and the reproduced correlations, as estimated from the factor matrix, can be examined to determine model fit.

The following issues are considered in detail:

- **Factor analysis model.**
- **Statistics** associated with factor analysis: correlation matrix, communality, eigenvalue, factor loading, factor loading plot, factor matrix, factor scores, Kaiser-Meyer-Oklin measure of sampling adequacy, percentage of variance, residuals, scree plot.
- **Principal components analysis:** an approach to factor analysis that considers the total variance in the data.
- **Common factor analysis:** an approach to factor analysis that estimates the factors based only on the common variance.
- **Determination of the number of factors:** a priori determination, determination based on eigenvalues, determination based on scree plot, determination based on percentage of variance, determination based on Split-Half reliability, determination based on significance tests.
- **Rotation of factors:**
  1. **Orthogonal rotation:** a rotation of factors in which the axes are maintained at right angles.
  2. **Varimax procedure:** an orthogonal method of factor rotation that minimizes the number of variables with high loading on a factor, thereby enhancing the interpretability of the factors.
  3. **Oblique rotation:** a rotation of factors when the axes are not maintained at right angles.
- **Factor scores:** a composite scores estimated for each respondent on the derived factors.
- **Selection of surrogate variables.** Surrogate variables are a subset of original variables selected for use in subsequent analysis.
- **Interpretation** of factors.

Chapter 7. Methods of cluster analysis.

Basic idea of the cluster analysis method. The concepts of "cluster form" and distance function. Hierarchic and non-hierarchic cluster methods. Cluster analysis of objects and variables.

Cluster analysis is used for classifying objects or cases, and sometimes variables, into relatively homogenous groups. The groups or clusters are suggested by the data and are not denned a priori.

The variables on which the clustering is done should be selected based on past research, theory, the hypotheses being tested, or the judgment of the researcher.

An appropriate measure of distance or similarity should be selected. The most commonly used measure is the Euclidean distance or its square.

Clustering procedures may be hierarchical or nonhierarchical.

Hierarchical clustering is characterized by the development of a hierarchy or tree-like structure. Hierarchical methods can be agglomerative or divisive.

Agglomerative methods consist of linkage methods, variance methods, and centroid methods. Linkage methods comprise single linkage, complete linkage, and average linkage. A commonly used variance method is Ward's procedure.

The nonhierarchical methods are frequently referred to as k-means clustering. These methods can be classified as sequential threshold, parallel threshold, and optimizing partitioning.

Hierarchical and nonhierarchical methods can be used in tandem. The choice of a clustering procedure and the choice of a distance measure arc interrelated.

The number of clusters may be based on theoretical, conceptual, or practical considerations. In hierarchical clustering, the distances at which the clusters are being combined is an important criterion.

The relative sizes of the clusters should be meaningful. The clusters should be interpreted in terms of cluster centroids. Often it is helpful to profile the clusters in terms of variables that were not used for clustering. The reliability and validity of the clustering solutions may be assessed in different ways.

The following issues are considered in detail:

- **Hierarchical clustering:** a clustering procedure characterized by the development of a hierarchy or tree-like structure.
- **Agglomerative clustering:** a hierarchical clustering procedure in which each object starts out in a separate cluster. Clusters are formed by grouping objects into bigger and bigger cluster.
- **Divisive clustering:** a hierarchical clustering procedure in which all objects start out in one giant cluster. Clusters are formed by dividing this cluster into smaller clusters.
- **Nonhierarchical clustering:** a procedure that first assigns or determines a cluster center and then groups all objects within from the center.
- Selection of a distance measure:
  1. **Euclidean distance:** the square root of the sum of the squared differences in values for each variable.
  2. **Square of the Euclidean distance.**
  3. **The city-block/Manhattan distance.** Manhattan distance between two objects is the sum of the absolute differences in value for each variable.
  4. **The Chebyshev distance.** The Chebyshev distance between two objects is the maximum absolute difference in values for any variable.
- Selection of a clustering procedure for hierarchical clustering:
  1. **Linkage methods:** agglomerative methods of hierarchical clustering that cluster objects based on a computation of the distance between them.
  2. **Single linkage method:** a method that is based on maximum distance or the further neighbor approach.
  3. **Average linkage method:** a linkage method based on the average distance between all pair is from each of the cluster.
  4. **Variance method:** an agglomerative method of hierarchical clustering in which clusters are generated to minimize the within-cluster variance.
  5. **Ward's procedures:** a variance method in which the squared euclidean distance to the cluster means is minimized.
  6. **Centroid method:** a variance method of hierarchical clustering in which the distance between two clusters is the distance between their centroids (means for all the variables).
- Selection of a clustering procedure for nonhierarchical clustering:

  1. **Segmental threshold method:** a nonhierarchical clustering method in which a cluster center is selected and all objects within a prespecified threshold value from the center are grouped together.
  2. **Parallel threshold method:** a nonhierarchical clustering method that specifies several cluster centers at once. All objects that within a prespecified threshold value from the center are grouped together.
  3. **Optimizing partitioning method:** a nonhierarchical clustering method that allows for later reassignment of objects to clusters to optimize an overall criterion.
- **Making decision on the number of clusters.**
- **Interpretation and profiling of clusters.**
- **Assessment of the validity of clustering.**

Chapter 8. Multidimensional scaling.

Basic ideas of multidimensional scaling model. Scaling of objects and scaling of variables. The indicators of model quality. The possibilities of two-dimensional and multidimensional projection of modeling results.

Multidimensional scaling is used for obtaining spatial representations of respondents perceptions and preferences. Perceived or psychological relationships among stimuli are represented as geometric relationships among points in a multidimensional space. Formulating the MDS problem requires a specification of the brands or stimuli to be included. The number and nature of brands selected influences the resulting solution. Input data obtained from the respondents can be related to perceptions or preferences. Perception data can be direct or derived. The direct approaches are more common in marketing research.

The selection of an MDS procedure depends on the nature (metric or non-metric) of the input data and whether perceptions or preferences are being scaled. Another determining factor is whether the analysis will be conducted at the individual or aggregate level. The decision about the number of dimensions in which to obtain a solution should be based on theory, interpretability, elbow criterion, and ease-of-use considerations. Labeling the dimensions is a difficult task that requires subjective judgment. Several guidelines are available for assessing the reliability and validity of MDS solutions. Preference data can be subjected to either internal or external analysis. If the input data are of a qualitative nature, they can be analyzed via correspondence analysis. If the attribute-based approaches are used to obtain input data, spatial maps can also be obtained by means of factor or discriminant analysis.

The following issues are considered in detail:

- **Multidimensional scaling – MDS:** a class of procedures for representing perceptions and preferences of respondents spatially by means of a visual display.
- **Basic concepts** in multidimensional scaling.
- **Statistics and terms** associated with multidimensional scaling: similarity judgments, preference rankings, stress, spatial map, coordinates, unfolding.
- **Input data** for multidimensional scaling:
  1. **Perceptions:** direct/similarity judgments and derived/attribute ratings.
  2. **Preferences.**
- **Selection of a multidimensional scaling procedure.**
- **Metric** multidimensional scaling: a multidimensional scaling method that assumes input data are metric.
- **Non-metric** multidimensional scaling: a type of multidimensional scaling method that assumes the input data are ordinal.
- **Making decision on the number of dimensions:**
  1. **A priori knowledge:** a theory or past research may suggest a particular number of dimension.
  2. **Interpretability** of the spatial map.
  3. **Elbow criterion:** a plot of stress versus dimensionality used in MDS. The point at which an elbow or a sharp bend occurs indicates an appropriate dimensionality.
  4. **Ease of use.** It is generally easier to work with two-dimensional maps or configurations than with those involving more dimensions.
  5. **Statistical approaches** are also available for determining the dimensionality.
- **Interpretation of the configuration.**
- **Assess of reliability and validity.**