



```
library(dplyr)

rladies_global %>%
  filter(city == 'Tbilisi',
         Date == '20.11.2018',
         Presenter == 'Nutsa Abazadze'.)
```

R პაკეტის შექმნა კომპანიის  
შიდა გამოყენებისთვის



# რაზე ვისაუბრებთ?

1. შიდა გამოყენების R პაკეტის საჭიროება.
2. კითხვარების მოდულარიზაცია - **Split Questionnaire Design**.
3. R პაკეტი „**SQD**“.



1.

შიდა გამოყენების R  
პაკეტის საჭიროება

# რა უპირატესობები აქვს შიდა R პაკეტის არსებობას კომპანიაში?

- თავიდან იცილებს ველოსიპედის თავიდან გამოგონებას სხვადასხვა თანამშრომლის მიერ.
- ახდენს სხვადასხვა პროცესის სტანდარტიზებას და ავტომატიზაციას.
- სხვადასხვა ფუნქციას, მის დოკუმენტაციას და მაგალითებს ხელმისაწვდომს ხდის თანამშრომლების დიდი ნაწილისთვის რაც ზრდის სხვადასხვა საკითხსა თუ მეთოდზე ცნობიერების დონეს.
- კომპანიის შიდა Git პლატფორმის საშუალებით შეიძლება კოდის ვერსიონირება, მისი სხვადასხვა თანამშრომლის მიერ შესწორება და ოპტიმირება.
- შესაძლებელი ხდება პაკეტის, ანუ რაიმე სტატისტიკური მეთოდის პროდუქტის სახით შეფუთვა (მაგალითად Shiny აპლიკაციის შექმნა) და კლიენტებისთვის შეთავაზება.



# მაგალითები თუ რომელ კომპანიებს აქვთ შიდა გამოყენების პაკეტები?

## Airbnb

იყენებს R -ს და მათ მიერ შექმნილ შიდა პაკეტს „Rbnb“ კომპანიის სხვადასხვა ინიციატივის განსახორციელებლად.

## John Deere

სასოფლო სამეურნეო (ტრაქტორი)  
აღჭურვილობის მწარმოებელი, John Deere  
ავტომატიზირებული R კოდის მეშვეობით  
ახდენს მოკლე და გრძელვადიან  
პროგნოზირებას, მონაცემთა კოორდინაციას და  
საწარმოო ხაზის შეკვეთების ოპტიმიზაციას.

Facebook

Twitter

Microsoft

KANTAR



2.

კითხვარების მოდულარიზაცია  
**Split Questionnaire Design.**



<https://www.kantartns.de/kernkompetenzen/applied-marketing-science.asp>

# მოტივაცია ამ მეთოდის უკან

- რესპონდენტების შეზღუდული დრო.
- რესპონდენტების უარი გრძელ გამოკითხვებში მონაწილეობაზე.
- მონაცემების ცუდი ხარისხი კითხვარის სიდიდის გამო.
- ონლაინ გამოკითხვების სიჭარბე და მათ მიერ ქაღალდის კითხვარების ჩანაცვლება.
- მობილური ტელეფონი (Smartphone) - გამოკითხვების ჩატარების ახალი მედიუმი.



# საუეთესო გამოცდილებები

**Split Questionnaire Survey Design** მეთოდის გამოყენებას ადგილი ჰქონდა შემდეგ დიდ გამოითხვებში:

- ამერიკის Consumer Expenditure Survey.
- National Assessment of Educational Progress (NAEP).
- PISA.



### 3. R პაკეტი „SQD“.

საბაზისო მონაცემების მომზადება სრულ კითხვარზე დაყრდნობით  
(ძირითადი და არა ძირითადი კითხვების გამოვლენა)

კითხვების ბლოკების შედგენა სრული კითხვარიდან (B ბლოკის)

Multiple Matrix Sampling -  
შემთხვევითობის პრინციპით

კორელაციის კოეფიციენტები  
და კლასტერული ანალიზი

შინაარსიდან გამომდინარე

წინა ეტაპზე შედგენილ ბლოკებზე დაყრდნობით, შემცირებული (Q ზომის) კითხვარების ყველა შესაძლო  
კომბინაციის შექმენა და შემდეგ მათგან K რაოდენობის კითხვარის ამორჩევა

Multiple Matrix Sampling  
მეთოდით შექნილი B  
ბლოკიდან K რაოდენობის  
კითხვარების  
შემთხვევითობის პრინციპით  
შედგენა

კორელაციის კოეფიციენტების  
და კლასტერული ანალიზის  
მეთოდით შექმნილი B  
ბლოკიდან K რაოდენობის  
კითხვარების შემთხვევითობის  
პრინციპით შედგენა

Multiple Matrix Sampling მეთოდით  
შექნილი B ბლოკიდან K  
რაოდენობის ოპტიმალური  
კითხვარის შედგენა BIBD და D-  
optimality კრიტერიუმებზე  
დაყრდნობით

კორელაციის კოეფიციენტების და  
კლასტერული ანალიზის მეთოდით  
შექმნილი B ბლოკიდან K რაოდენობის  
ოპტიმალური კითხვარის შედგენა  
BIBD და D-optimality კრიტერიუმებზე  
დაყრდნობით

ოთხივე სცენარის შემთხვევაში ფუნქცია ქმნის List-ს რომელიც შედგება შემდეგი ობიექტებისგან:

1. კითხვების ბლოკები
2. დაყოფილი (შემცირებული) კითხვარების დიზაინი
3. მონაცემთა ბაზა რომელიც შეიცავს არარსებულ მონაცემებს შექმნილ დიზაინებზე დაყრდნობით

იმ არარსებული მონაცემების იმპუტაცია რომელიც კითხვარების გაყოფის შედეგად წარმოიქმნა

იმპუტაციის შედეგად აღდგენილი მონაცემების შეფასება

# კითხვების ბლოკების შედგენა



ბლოკების შედგენამდე უნდა გადაწყდეს რამდენი ბლოკის შედგენა იგეგმება (B)  
(მაგ: 10 ბლოკი 100 კითხვიდან)

რაოდენობის შესახებ გადაწყვეტილების მიღების შემდეგ, ბლოკების შესადგენად შესაძლოა მოცემული ორი მეთოდის გამოყენება:

## 1. Multiple matrix sampling (MMS)

- ეს მეთოდი გამოიყენება ყოველთვის როდესაც ხდება კითხვარის გაყოფა, რადგან ამ შემთხვევაში ბლოკების შედგენა ხდება შემთხვევითობის პრინციპით.

**მაგალითი:**

- კითხვარი შედგება 100 კითხვისგან.
- გვინდა შევქმნათ 10 ბლოკი, თითოეული 10 შეკითხვით რომელსაც შევარჩევთ მარტივი შემთხვევითი შერჩევის გზით (Simple random sampling).

## 2. კორელაციის კოეფიციენტები და კლასტერული ანალიზი

- ამ მეთოდის გამოყენების წინაპირობა არის დიდი რაოდენობის საბაზისო მონაცემები, რომელზე დაყრდნობითაც მოხდება ცვლადთა შორის კორელაციის ანალიზი.
- იდეა : ცვლადები რომლებიც ერთმანეთთან მაღალ კორელაციაში იმყოფებიან უნდა განთავსდნენ სხვადასხვა ბლოკში.
- შედეგად ცვლადები რომლებიც ერთმანეთის კარგი ამხსნელები არიან ნაწილდებიან სხვადასხვა ბლოკში, რაც ქმნის ხელოვნურად შექმნილი არარსებული მონაცემების კარგ წინაპირობას.

# ბლოკების შედგენა კორელაციის კოეფიციენტების და კლასტერული ანალიზის გამოყენებით

საბაზისო მონაცემებზე დაყრდნობით კორელაციის მატრიცის შედგენა და ანალიზი.



ბლოკების შესაქმნელად კლასტერული ანალიზის გამოყენება



დაკვირვებების ნაცვლად ცვლადების კლასტერიზაცია



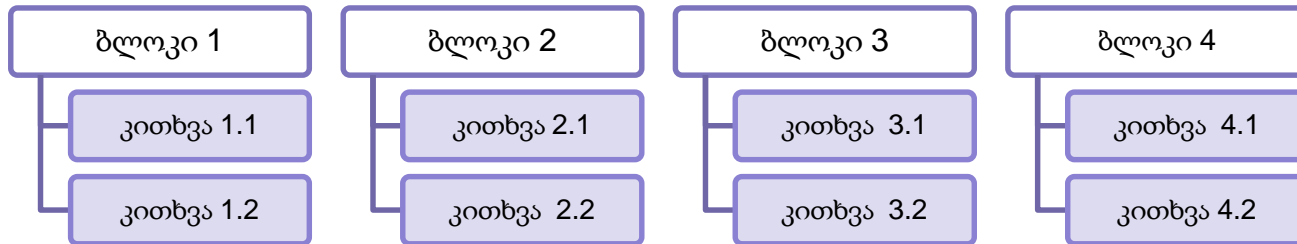
Distance matrix (დაშორებების მატრიცის) ნაცვლად კორელაციის მატრიცის გამოყენება



ერთმანეთთან მაღალ კორელაციაში მყოფი ცვლადების სხვადასხვა ბლოკებში განთავსება

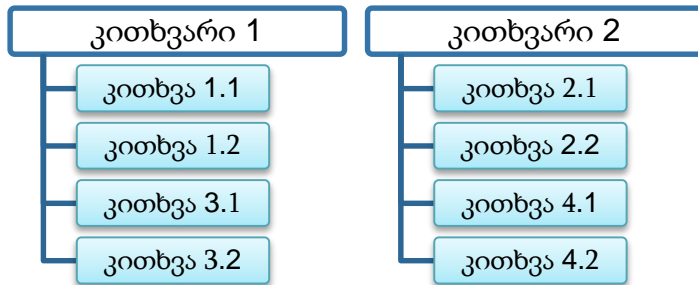
# Between-block & within-block designs

არსებობს ორი ძირითადი მიდგომა თუ როგორ შეიძლება კითხვების ბლოკებისგან კითხვარების შექმნა



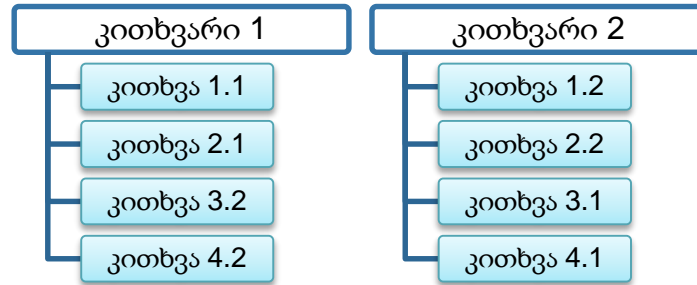
## Between-block design:

- რესპონდენტი პასუხობს შერჩეულბლოკში შემავალ ყველა შეკითხვას



## Within-block design:

- რესპონდენტი პასუხობს შერჩეულ ბლოკში შემავალ რამდენიმე შეკითხვას



# ბლოკებზე დაყრდნობით მცირე ზომის კითხვარების შედგენა

ამ ეტაპზე მისაღებია ორი მნიშვნელოვანი გადაწყვეტილება:

1. რამდენი კითხვარი უნდა შევადგინოთ ერთი დიდი კითხვარიდან? (K)
2. რა ზომის უნდა იყოს ახლად შედგენილი კითხვარები? (N)  
(რამდენი ბლოკისგან უნდა შედგებოდეს თითოეული)

მას შემდეგ რაც ეს ორი გადაწყვეტილება მიიღება, გვაქვს ასეთი ზომის კითხვარის შედგენის იმაზე მეტი შესაძლებლობა ვიდრე გვჭირდება. ასეთ დროს, ყველა შესაძლო ვარიანტიდან შეგვიძლია შევარჩიოთ K ყველაზე ოპტიმალური **Balanced Incomplete Block Design** და **D-Optimality** კრიტერიუმის გამოყენებით.



1. პირველ ეტაპზე შევადგინეთ 10 ბლოკი და თითოეული მათგანი შედგება 10 კითხვისგან ( $B=10$ ).
2. გადაწყვეტით რომ სულ გვინდა ერთი დიდი კითხვარიდან 5 პატარა კითხვარის შედგენა ( $K=3$ ).
3. თითოეულ ბლოკში უნდა შედიოდეს 10 დან მხოლოდ 4 ბლოკი ( $N=4$ ).

$$C_4(10) = \binom{10}{4} = \frac{10!}{4!(10-4)!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2 \cdot 1} = 210$$

როგორც ვხედავთ სულ გვაქვს 210 ვარიანტი  
საიდანაც უნდა შევარჩიოთ 5 ყველაზე ოპტიმალური

[illegible]





შეიძლება თუ არა ამ პაკეტის გამოყენება  
ყველა ზომის კითხვარის გაყოფისთვის?

# Build thematic blocks based on variable labels

Apply between-block design for generating splits

Apply within-block design for generating splits

Apply the mixture of within-block and between-block designs

Decide how many splits/questionnaires shall be created

Decide which thematic blocks shall be used as a whole and for which thematic blocks shall be created sub-blocks

Create splits by choosing blocks at random

Create splits using balanced incomplete block designs and D-optimality

Create sub-blocks for each thematic block – decide how many sub-blocks shall be built

Create sub-block for chosen thematic blocks – decide how many sub-blocks per block shall be built

Create sub-blocks randomly, using multiple matrix sampling

Impute missing data in the dataset

Create splits by choosing blocks from all sub-blocks at random

Create splits by using balanced incomplete block designs and D-optimality for all sub-blocks

Create sub-blocks by using correlation coefficients and cluster analysis

Create splits by using greedy approach

Create splits by choosing blocks from all sub-blocks at random

Create splits by using balanced incomplete block designs and D-optimality for all sub-blocks

Create splits by using greedy approach



დიდი მადლობა ყურადღებისთვის!

ნუცა აბაზაძე



[nucaabazadze@gmail.com](mailto:nucaabazadze@gmail.com)



<https://www.linkedin.com/in/nutsaabazadze/>