

ლაბორატორია 3: მასობრივი გამოკითხვის მონაცემთა დამუშავება R-გარემოში

დავით სიჭინავა

8 აპრ. 2018 წ.

საკითხები

- მასობრივი გამოკითხვის მონაცემების წაკითხვა
- სიხშირის და შეუღლებული ცხრილები (კროსტაბულაცია)
- არითმეტიკული ოპერაციები

ინსტრუქციები:

თანმიმდევრობით შეასრულეთ მითითებული ამოცანები. თქვენს .rmd ფაილს სახელწოდება მიანიჭეთ შემდეგი ფორმით: თქვენი გვარი_lab3.rmd. მაგალითად:

```
sichinava_lab3.Rmd
```

დავალების და მონაცემების აღწერა

„კავკასიის ბარომეტრის“ ვებსაიტიდან გადმოწერეთ „ეროვნული დემოკრატიული ინსტიტუტის“ (NDI) გამოკითხვის 2017 წლის დეკემბრის კვლევის კითხვარი და მონაცემთა ბაზა თქვენთვის მოსაწონ ფორმატში (SPSS ან STATA). გახსენით ახალი ბლოკნოტი და ბაზის მონაცემები წაიკითხეთ R-ში. რომელი ბიბლიოთეკა დაგჭირდებათ ამ მიზნით?

წინამდებარე დავალების მიზანია, შეაფასოს, თუ რა კავშირშია ერთმანეთთან საქართველოში მოსახლეობის პოლიტიკური შეხედულებები და მათი დამოკიდებულება სხვადასხვა მედიასაშუალებებისადმი. პოლიტიკური კომუნიკაციების დისციპლინაში კვლევების მთელი სპექტრი არსებობს, რომლებიც ამტკიცებენ, რომ პარტიული შეხედულებების მქონე ადამიანები სულ უფრო ხშირად ირჩევენ ახალი ამბების ისეთ წყაროებს, რომლებიც მათ დამოკიდებულებებთან ყველაზე ახლოს დგანან [2] [1] [3] [5] [4]. შესაბამისად, მსგავსი ტიპის ე.წ. Echo Chamber ამომრჩეველთა კიდევ უფრო მეტ პოლარიზაციას უწყობს ხელს. „ეროვნული დემოკრატიული ინსტიტუტის“ გამოკითხვა საშუალებას გვაძლევს, ამ მოვლენას ქართველი ამომრჩევლების მაგალითზეც

ცხრილი 1: ცვლადების განმარტება

ცვლადი	კითხვის ნომერი კითხვარში	განმარტება
PARTYSUPP1	54	შეხედულებებთან ახლოს მყოფი პარტია: პირველი არჩევანი
TRUIMEDI	19	ენდობა „იმედს“ პოლიტიკის და მიმდინარე მოვლენების შესახებ ინფორმაციის მისაღებად
TRUR2	19	ენდობა „რუსთავი 2-ს“ პოლიტიკის და მიმდინარე მოვლენების შესახებ ინფორმაციის მისაღებად
RESPEDU	63	განათლების უმაღლესი საფეხური
AGEGROUP	დემოგრაფია	ასაკობრივი ჯგუფი
SETTYPE	დემოგრაფია	დასახლებული პუნქტის ტიპი

დავაკვირდეთ. დავალების ამოცანაა, პარტიული კუთვნილების მიხედვით გავაანალიზოთ, თუ რომელ მედია-საშუალებებს ენდობიან საქართველოს მცხოვრებლები, რომლებსაც ჩამოყალიბებული პარტიული შეხედულებები გააჩნიათ.

მართალია, ჩვენს ბაზაში ცვლადებს სტანდარტიზებული სახელები ჰქვიათ, თუმცა დავალების ტექსტში მითითებული მაქვს კითხვარში მათი შესაბამისი ნომრები, ასე, რომ კითხვის პოვნა არა მგონია, გაგიჭირდეთ. წინამდებარე დავალებისთვის დაგვირდება ცვლადები *PARTYSUPP1*, *AGEGROUP*, *RESPSEX*, *SETTYPE*, *SUBSTRATUM*, *TRUIMEDI*, *TRUR2*, *RESPEDU*. ცხრილში 1 წარმოდგენილია ცვლადების დახასიათება და მათი შესაბამისი ნომრები კითხვარში.

სანამ მონაცემთა ანალიზზე გადავალთ, საჭიროა მონაცემების გარდაქმნა. გამოიტანეთ ცვლად *PARTYSUPP1*-ის ცხრილი. ქართული პოლიტიკური ლანდშაფტის მრავალფეროვნებიდან გამომდინარე, ამ ცვლადშიც საკმაო სიჭრელე შეინიშნება, ამიტომ ანალიზისთვის აჯობებს, მსხვილი პარტიები და პასუხის ვარიანტები („ქართული ოცნება“, „ნაციონალური მოძრაობა“, „არ ვიცი“, „უარი პასუხზე“, „არცერთი“) ცალ-ცალკე კატეგორიებად გამოვყოთ, ხოლო დანარჩენი პარტიები დავაჯგუფოთ კატეგორიაში *other*. გაითვალისწინეთ, რომ ნებისმიერი კატეგორიული ცვლადი ფაქტორად უნდა აქციოთ, ანუ - შესაბამისი პასუხის ვარიანტებიც უნდა მიანიჭოთ. ყურადღება მი-აქციეთ იმასაც, რომ ჩვენს ბაზაში კოდები -9 და -3 სტანდარტულია და აღნიშნავს გამოტოვებულ პასუხებს. შესაბამისად, ეს კატეგორიები უმჯობესია, *NA*-ად გადავაკეთოთ.

ანალოგიურად მოიქცეთ განათლების შემთხვევაში. პასუხის ვარიანტები დააჯგუფეთ სამ კატეგორიად: საშუალო ან უფრო დაბალი, საშუალო ტექნიკური, სრული ან არასრული უმაღლესი. განათლების ცვლადიც ფაქტორი უნდა იყოს.

ცვლადები *TRUIMEDI* და *TRUR2* ე.წ. ბინარული ცვლადებია, სადაც 1 აღნიშნავს, რომ რეს-პონდენტი ენდობა ტელეარხს. ჩვენი ანალიზისთვის უმჯობესია, რომ სხვა პასუხის ვარიანტები 0-ად გადავაკეთოთ, პასუხის ვარიანტს 1 დავარქვათ "Mentioned 0 - "Not Mentioned" და შესაბამისად, გავაფაქტოროთ.

რაც შეეხებათ ცვლადებს *RESPSEX*, *AGEGROUP*, *SETTYPE*, მათი კოდების შესაბამისი პასუხის ვარიანტებია:

- AGEGROUP: 1- 18-35, 2 - 36-55, 3 - 56+
- SETTYPE: 1- თბილისი, 2 - ქალაქები, 3 - სოფლები, 4 - ეთნიკური უმცირესობები

ეს თითოეული ცვლადიც აქციეთ ფაქტორად და მიაწიკეთ შესაბამისი მნიშვნელობები

სიხშირის ცხრილების აგება

ააგეთ ამ თითოეული ცვლადის ცხრილი *table* ფუნქციის გამოყენებით. როგორც ხედავთ, შედეგებში მოცემულია *აბსოლუტური* სიხშირე, ანუ თითოეულ კატეგორიაში რესპონდენტთა რაოდენობა. იმისთვის, რათა გამოვიტანოთ *ფარდობითი სიხშირე*, საჭიროა, *table*-თან ერთად, გამოვიყენოთ *prop.table* ფუნქცია:

```
prop.table(table(ndidec$SETTYPE))
```

როგორც მიხვდით, მაგალითში *ndidec* ჩემი ცხრილის სახელწოდებაა. როგორც ვხედავთ, ცხრილების აგება საკმაოდ მარტივი ამოცანაა. როდესაც საქმე მასობრივ გამოკითხვასთან გვაქვს, როგორც წესი, მონაცემები *შეწონილია*, ანუ კორექტირებულია მოსახლეობის პარამეტრების მიხედვით. მაგალითად, როგორც წესი, გამოკითხვებში ოდნავ უფრო მეტად არიან ქალები წარმოდგენილნი. იმის მიხედვით, თუ ქალების როგორი განაწილებაა მოსახლეობაში, მკვლევარს შეუძლია, მათ პასუხებს უფრო ნაკლები კოეფიციენტი მიაწიკოს, ვიდრე - მამაკაცებისას, რათა კვლევის შედეგები მეტნაკლებად ასახავდეს დემოგრაფიულ სტრუქტურას. *xtabs* ფუნქციის გამოყენებით, მარტივად შეგვიძლია, გამოვთვალოთ შეწონილი პროპორციები.

```
xtabs(WTIND~party, data=ndidec, na.action = "na.omit")
```

xtabs ფუნქციაში ამ ნიშნამდე (ტილდამდე) ვწერთ წონის ცვლადს, ტილდის შემდეგ - ვუთითებთ ჩვენთვის საინტერესო ცვლადს, შემდეგ - მონაცემთა ცხრილის სახელწოდებას, დაბოლოს, *R*-ს ვატყობინებთ, რომ ცარიელი მნიშვნელობები თუ შეხვდა, გამოტოვოს.

როგორც ხედავთ, ეს *ჩვეულებრივი ოპერაციაა*. ასევე მარტივად შეგვიძლია, გამოვთვალოთ შეწონილი მონაცემების ფარდობითი პროპორციის ცხრილიც, *prop.table* ფუნქციის გამოყენებით:

```
prop.table(xtabs(WTIND~party, data=ndidec, na.action = "na.omit"))
```

კროსტაბულაციის ცხრილების აგება

კროსტაბულაციები, ანუ *ნიშანთა შეუღლების ცხრილები* აღწერითი სტატისტიკური ანალიზის საკმაოდ მოხერხებული მეთოდია. ასეთი ცხრილები საშუალებას გვაძლევს, მონაცემები შევაჯამოთ ჯგუფების მიხედვით, მაგალითად - როგორაა განაწილებული პარტიული კუთვნილება ასაკობრივ ჯგუფებში და ა.შ.

მარტივი კროსტაბულაციის აგება *table* ფუნქციითაცაა შესაძლებელი. მაგალითისთვის, თუ გვინტერესებს, დასახლებული პუნქტის ტიპის მიხედვით პარტიული კუთვნილების გაანალიზება, უნდა გამოვიყენოთ შემდეგი სინტაქსი:

```
table(ndidec$party, ndidec$SETTYPE)
```

ჯერ ვწერთ იმ ცვლადს, რომლის სიხშირეც გვაინტერესებს, ხოლო შემდეგ - იმ ცვლადს, რომლის კატეგორიების მიხედვითაც უნდა დავითვალოთ სიხშირეები. როგორც ხედავთ, შედეგში წარმოდგენილია აბსოლუტური სიხშირეები. ჩვენ კი სვეტების მიხედვით პროპორციები გვჭირდება. ამიტომ, ისევ უნდა მივმართოთ *prop.table* ფუნქციას:

```
prop.table(table(ndidec$party, ndidec$SETTYPE), 1)
```

სინტაქსში რიცხვი 1 მიუთითებს, რომ პროპორციები გამოთვლილია სვეტების მიხედვით. თუ 2 ეწერებოდა, მაშინ პროგრამა პროპორციებს რიგების მიხედვით დაითვლიდა, თუ საერთოდ არაფერს მიუთითებდით, მაშინ 100

მსგავსად სიხშირისა, კროსტაბულაციაც შეგვიძლია, შევწონოთ. ამისთვის ვიყენებთ ფუნქციას *xtabs*:

```
prop.table(xtabs(WTIND~party+SETTYPE, data=ndidec, na.action = "na.omit"), 1)
```

xtabs ფუნქციაში ამ ნიშნამდე (ტილდამდე) ვწერთ წონის ცვლადს, ტილდის შემდეგ - ვუთითებთ რიგის და სვეტის აღმნიშვნელ ცვლადებს, შემდეგ - მონაცემთა ცხრილის სახელწოდებას. რადგან საბოლოო ჯამში, კორიზონტალური პროპორციები გვაინტერესებს, ვუთითებთ *prop.table*-ს და რიცხვ 1-ს. თუ ცარიელი მნიშვნელობების გამოტოვება გვსურს, დავამატებთ *na.action = "na.omit"*

მონაცემთა ანალიზი

პირველ რიგში, გამოიტანეთ თქვენს მიერ შექმნილი და დამუშავებული ცვლადების შეწონილი სიხშირეები. ძალიან კარგი, ახლა, ეცადეთ, უპასუხოთ, ქვემოთ ჩამოთვლილ კითხვებს:

- როგორია პარტიული კუთვნილება დასახლებული პუნქტების ტიპის მიხედვით?
- როგორია პარტიული კუთვნილება ასაკობრივი ჯგუფების მიხედვით?
- როგორია „რუსთავი 2-ის“ და „იმედის“ ნდობა განათლების დონის მიხედვით?
- მოსახლეობის რა ნაწილია, რომელიც პოლიტიკური ინფორმაციის მისაღებად ენდობა როგორც „რუსთავი 2-ს“, ასევე - „იმედს“?
- როგორია პარტიული კუთვნილების განაწილება იმ მოსახლეობაში, რომელიც ენდობა „რუსთავი 2-ს“?
- როგორია პარტიული კუთვნილების განაწილება იმ მოსახლეობაში, რომელიც ენდობა „იმედს“?

ეს რა იყო? როგორ ჩავაბარო?!

დაზიპეტ ფოლდერი და დაარქვით სახელი შემდეგი ფორმატით: *surname.lab3.zip*. ატვირთეთ თქვენი დავალება ამ [ლინკზე](#) მომდევნო შესვედრის დაწყებამდე. წარმატებები!

References

- [1] Levi Boxell, Matthew Gentzkow, and Jesse M. Shapiro. *Is the Internet Causing Political Polarization? Evidence from Demographics*. Working Paper 23258. National Bureau of Economic Research, Mar. 2017. DOI: [10.3386/w23258](https://doi.org/10.3386/w23258). URL: <http://www.nber.org/papers/w23258> (visited on 04/07/2018).
- [2] Matthew Levendusky. "Partisan Media and Polarization: Challenges for Future Work". In: *Oxford Research Encyclopedia of Politics* (Jan. 25, 2017). DOI: [10.1093/acrefore/9780190228637.013.50](https://doi.org/10.1093/acrefore/9780190228637.013.50). URL: <http://politics.oxfordre.com/view/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-50> (visited on 04/07/2018).
- [3] Matthew Levendusky and Neil Malhotra. "Does Media Coverage of Partisan Polarization Affect Political Attitudes?" In: *Political Communication* 33.2 (Apr. 2, 2016), pp. 283–301. ISSN: 1058-4609. DOI: [10.1080/10584609.2015.1038455](https://doi.org/10.1080/10584609.2015.1038455). URL: <https://doi.org/10.1080/10584609.2015.1038455> (visited on 04/07/2018).
- [4] Natalie Stroud. "Polarization and Partisan Selective Exposure". In: *Journal of Communication* 60.3 (Aug. 19, 2010), pp. 556–576. ISSN: 0021-9916. DOI: [10.1111/j.1460-2466.2010.01497.x](https://doi.org/10.1111/j.1460-2466.2010.01497.x). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1460-2466.2010.01497.x> (visited on 04/07/2018).
- [5] Natalie Jomini Stroud. "Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure". In: *Political Behavior* 30.3 (Sept. 1, 2008), pp. 341–366. ISSN: 0190-9320, 1573-6687. DOI: [10.1007/s11109-007-9050-9](https://doi.org/10.1007/s11109-007-9050-9). URL: <https://link.springer.com/article/10.1007/s11109-007-9050-9> (visited on 04/07/2018).