

SOSIOLOJİ MƏLUMATLARIN R PROQRAMLAŞDIRMA DİLİNDƏ STATİSTİK ANALİZİ

Asya Əsgərova

*Azərbaycan Respublikası Prezidentinin yanında Dövlət İdarəçilik
Akademiyası İnzibati İdarəetmə fakültəsi, kompyuter təhsilə hazırlıq
magistr I kurs Rəhbər: f.r.e.d. Cəfərova H.Ə.*

Açar sözlər: R proqramlaşdırma dili, sosioloji məlumatların statistik analizi

Müasir dövrdə sosioloji məlumatların statistik analizində bir sıra müxtəlif proqram paketləri istifadə edilir. Son illərdə xüsusilə diqqət çəkən R proqramlaşdırma dili yaran-
dığı 1997-ci ildənfi bəri mütəxəssislər və tədqiqatçılar tərəfindən maraqla qarşılanan,
araşdırılan, inkişaf etdirilən və pulsuz olaraq yayımlan R proqramlaşdırma dili, bu gün
bir çox dünya ölkələrində iqtisadiyyatda, bank işində, maliyyə riyaziyyatında, sığorta
nəzəriyyəsində, xüsusən də sosial statistikada geniş istifadə olunur.

Məqalədə R proqramlaşdırma dili dili/haqqında ümumi məlumat verilmiş, sosioloji
məlumatların statistik analizi zamanı R proqramlaşdırma dilinin üstünlükləri göstərilmiş,
R proqramlaşdırma dilinin kodları vasitəsilə 2000-2012-ci illərdə Əmək bazarının əsas
sosial-iqtisadi göstəriciləri olan "Əhalinin orta illik sayı" (min nəfər) ilə "İşsizlər"in (min
nəfər) statistik verilənlərinə əsasən xətti reqresiya modelləri və uyğun qrafiklər qurul-
muşdur.

1. R proqramlaşdırma dili haqqında ümumi məlumat

Hal-hazırda yaşadığımız həyatı sosioloji məlumatların statistik analizi olmadan təsəv-
vür etmək mümkün deyil. Bildiyimiz kimi, sosial hadisələrin mövcud vəziyyətinin və
inkişaf qanunauyğunluqlarını öyrənilməsi, sosioloji məlumatların toplanılması və onların
statistik analizi dövlətin və cəmiyyətin inkişaf üçün, zəmin yaradır, əlverişli şərtlər for-
malaşdırır. Müasir dövrdə sosioloji məlumatların statistik analizinin səmərəliliyi istifa-
də olunan vasitələrdən birbaşa surətdə asılıdır. Doğrudur sosioloji məlumatların statistik

analizini ənənəvi qaydada yalnız kağız və qələmdən istifadə etməklə də həyata keçirmək mümkündür, lakin bunu xüsusi alətlərin köməyi ilə daha çevik və daha effektiv yerinə yetirmək olar, bir çox hallarda isə kompüter texnologiyasından, o cümlədən müxtəlif statistik analiz proqram paketlərindən istifadə etmədən doğru nəticələr

əldə etmək ümumiyyətlə mümkün deyil. Belə statistik proqramlaşdırma dillərindən biri R proqramlaşdırma dilidir.

R proqramlaşdırma dili məlumatların statistik analizi və həmçinin qrafik təsviri məqsədilə GNU layihəsi çərçivəsində yaradılan və hal-hazırda bir çox dünya ölkələrində müxtəlif sahələrdə geniş istifadə edilən proqramlaşdırma dilidir. [2] Bu dil hesablama əməliyyatları üçün açıq koda malik olması, bir çox əməliyyat sistemlərində (Unix, Linux, FreeBSD, Windows, MacOS) istifadə oluna bilinməsi, rahat istifadəçi qrafiki interfeysinə malik olması eyni zamanda da əmr sətiri istifadə oluna bilməsi imkanına görə istifadəçilərin sevimlisidir. [1] Təsədüfi deyildir ki, R 2010-cu ildə ən yaxşı açıq proqram təminatı nominasiyası üzrə InfoWorld jurnalının qalibləri siyahısına düşmüşdür.

(<http://www.infoworld.com/d/open-source/bossie-awards-2010-the-best-open-source-software-the-year-115>)

R dili və mühiti R Foundation təşkilatı tərəfindən dəstəklənir və inkişaf etdirilir. 1993-cü ildə Yeni Zelandiyanın Auckland Universitetinin statistika fakültəsinin əməkdaşları Ayhek Ross (Ihaka Ross) və Robert Gentleman tərəfindən hazırlanmış bu dilin adı müəlliflərin adlarının baş hərfinə uyğun adlandırılmışdır. R dili S dilinə alternativ olaraq yaradılmışdır. Lakin bu dillər arasında əhəmiyyətli fərqlərə baxmayaraq S dilinin kodları R mühitində də işləyir. [1]

2. Sosioloji məlumatların statistik analizi zamanı R proqramlaşdırma dilinin üstünlükləri

Bu proqramlaşdırma dili vasitəsilə aktual olan bir sıra sosial məsələlərin statistik analizi (İnzibati - ərazi bölgüsü, əhalinin sayı və tərkibi, əhalinin təkrar istehsalının ümumi göstəriciləri, doğum, ölüm, nikah və boşanmalar, miqrasiya, iqtisadi və inzibati rayon və şəhərlər üzrə müxtəlif göstəricilər, dünya ölkələri üzrə demografik göstəricilər və s.) çox rahatlıqla yerinə yetirilə bilər. Funksional dil olan R dili kod yazmağı asanlaşdıran bir sıra xüsusiyyətlərə malikdir. Ələxsus, hər bir sosioloji statistika mütəxəssisinin asanlıqla istifadə edə biləcəyi məlumatların analizi və qrafiki təqdimat funksiyaları diqqəti cəlb edir. [3] Sosioloji məlumatların analizi zamanı istifadə SPSS, SAT, STATA kimi statistik proqram paketlərindən fərqli olaraq R yalnız statistik proqram paketi deyil, statistik proqram yaratma mühitidir [4] Bununla əlaqədar R-in aşağıdakı diqqət çəkən xüsusiyyətlərini qeyd edə bilərik:

- Məlumatların aktiv emalı və saxlanması;
- Ardıcılıq, massiv və matrislərlə hesablamalar zamanı istifadə oluna bilən xüsusi

operatorlardan ibarətdir;

• Məlumatların analizi zamanı qrafiki təsvirə imkan verən əhəmiyyətli vasitələr mövcuddur;

• C və S proqramlaşdırma dilinin xüsusiyyətlərini özündə birləşdirən əlverişli proqramlaşdırma mühiti; [5]

Bundan başqa işdə lazım olan faylların R mühitinə uyğunlaşa bilməsi üçün fərqli seçimlər var. Lazımı məlumat paketlərini yükləmək üçün isə müxtəlif

verilənlər bazalarından (MySQL, Access, Microsoft SQL Server, PostgreSQL, Oracle, IBM DB2) verilənləri almaq mümkündür. [2]

Bu dilin güclü tərəfləri:

• \

• Pulsuz olması;

• Obyektyönlü bir proqramlaşdırma dili olması;

• 2D, 3D və daha çox ölçülü qrafiki vasitələrə sahib olması;

2012-ci ilin fevral ayından Oracle Advanced Analytics şirkəti böyük məlumatlar bazası üzərində R dilinin imkanlarından istifadə etmək məqsədilə platformasına Oracle R Enterprise - yeni komponentin yaranmasına təminat vermişdir. [1]

3. R proqramlaşdırma dilində sosioloji məsələnin həlli

Hal-hazırda R özündə bir sıra müxtəlif tədqiqat metodlarını birləşdirir. Bunların sırasına xətti və qeyri xətti proqramlaşdırma, zaman sıralarının analizi, sinifləndirmə, klasterizasiya və s. daxildir.

Azərbaycan Respublikası Dövlət Statistika Komitəsinin rəsmi saytından alınmış məlumata görə (<http://www.stat.rov.az/source/labour/>) 2000-2012-ci ilə qədər Əmək bazasının əsas sosial-iqtisadi göstəriciləri olan əhəlinin orta illik sayı (min nəfər) ilə işsizlər (min nəfər) göstəriciləri arasında xətti reqressiya asılılığına baxaq. Göstəricilərin cədvəli aşağıdakı şəkildədir:

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Əhəlinin orta illik sayı (min nəfər)	8073.6	8309.2	8500.3	8609.6	8723.0	8838.5	8947.3	9054.3	9173.1	9295.8			
İşsizlər (min nəfər)	514.7	400.9	317.8	291.2	281.1	262.2	260.2	258.3	250.9	243.1			

Bu məsələni R-da həll etmək üçün ilk öncə verilənləri daxil edirik,

0is<-c(8073.6, 8309.2, 8500.3, 8609.6, 8723.0, 8838.5, 8947.3, 9054.3, 9173.1, 9295.8)

ISH<-c(514.7, 400.9, 317.8, 291.2, 281.1, 262.2, 260.2, 258.3, 250.9, 243.1)

R-da regressiya analizi üçün əsas sintaksis hissəsi

lm(Y ~ model)

təşkil edir. (<http://stat.ethz.ch/R-manual/R-patched/library/stats/html/lm.html>) Burada Y obyektidir, özündə asılı dəyişən saxlayır, model isə seçilmiş riyazi model üçün istifadə olunan formuldur. **lm ()** komandası modelin koefisientlərini təmin

edir. **OIS**- əhalinin orta illik sayı, **ISH**- işsizləri göstərir. Onda **lm(OIS ~ISH)** işarə edə bilərik.

```
> lm(OIS ~ISH)
```

Call:

```
lm(formula = OIS ~ ISH)
```

Coefficients:

```
(Intercept)    ISH
```

```
9990.100  -4.018
```

Növbəti mərhələlərdə rahatlıq üçün aşağıdakı işarələri aparırıq **XETTI_MODEL**
DEL<-lm(OIS~ISH)

XETTI_MODEL

İndi biz bu obyektə digər bütün əmrlər üçün argument kimi istifadə edə bilərik. Bizə lazım olan qrafikləri quraq

```
lm(formula = OIS ~ ISH)
```

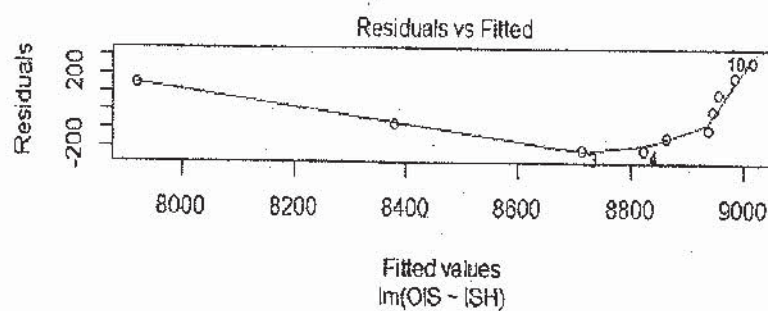
Coefficients:

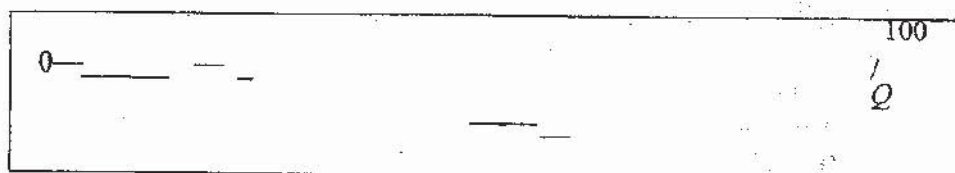
```
(Intercept)    ISH
```

```
9990.100  -4.018
```

```
plot(XETTI_MODEL)
```

Bu əməliyyatı yerinə yetirsək aşağıdakı qrafik təsviri alarıq.





8000

8200

8400

8600

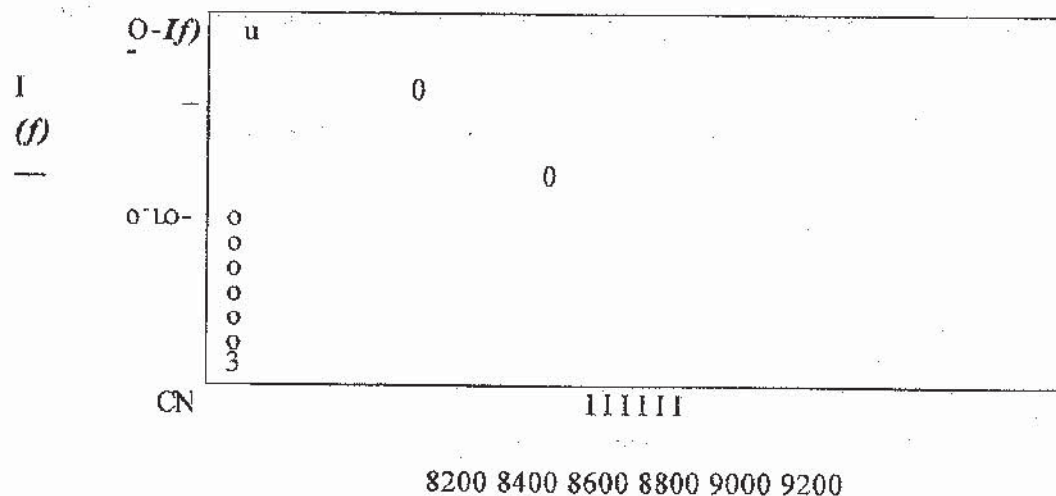
8800

9000

Fitted values $\text{lm}(\text{OIS} - \text{ISH})$

Korelyasiya əmsalı $\text{cor}(\text{OISJSH})$

[1] -0.8933367 $\text{plot}(\text{OISJSH})$



OIS

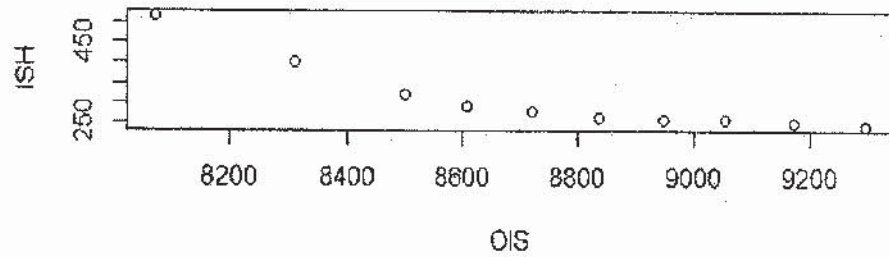
$\text{lines}(\text{lowess}(\text{OIS}, \text{ISH}))$

$\text{lines}(\text{OIS}, \text{fitted}(\text{XETTI_MODEL}))$

OIS

Daha sonra burada “Residuals” bölməsi vasitəsilə tətbiqi və nəzəri kəmiyyətlər arasındakı fərq tapılır. Modelin konfisientlərinin qiymətləndirilməsi və səhvlərin aşkarlanması üçün isə “Std error” bölməsi mövcuddur.

$\text{summary}(\text{XETTI_MODEL})$



Call:

lm(formula = OIS ~ ISH)

Residuals:

Min IQ Median 3Q Max -212.96 -127.82 -33.78 139.08 282.42

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 9990.1003 227.7305 43.868 8.05e-11 *** ISH -4.0178 0.7146
-5.623 0.000497 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 184.7 on 8 degrees of freedom Multiple R-squared: 0.7981,
Adjusted R-squared: 0.7728 F-statistic: 31.61 on 1 and 8 DF, p-value: 0.000497

8200 8400 8600 8800 9000 9200

CHS

Regressiya təhlili vasitəsilə tapılan araşdırmalar göstərir ki, göstəricilər arasında tərs mütənəşib asılılıq mövcuddur. Bu isə Azərbaycanda düzgün siyasətin nəticəsidir, işsizliklə mülbürizə tədbirlərinin doğru istiqamətdə aparılmasını göstərir.

Nəticə

Beləliklə, R akademik araşdırmalarla bərabər digər bir çox sahədə, xüsusilə sosioloji statistikada geniş istifadə oluna biləcək güclü bir proqramlaşdırma mühitidir. Günbəgün istifadəçi sayı artan bu proqrama Avropada böyük maraq var. R-in sosioloji statistika imkanlarının koorportiv istifadəçiləri sırasına Google, Facebook, Pfizer, Bank of America kimi dünya nəhərləri daxildir.

1. <http://www.r-project.org/>

2. Cəfərova H., Əliyev R., "R proqramlaşdırma dilinin riyazi statistikanın elementlərinə tətbiqi", Bakı 2013

3. Braun W.J., Murdoch D.J., "A First course in statistical programming", Cambridge University press, England, 2007

4. Statistical Computing Group, "Very Basics of R (Windows)", Research Data Services, University of Pennsylvania, 2008

5. R Development Core Team, "R language definition", 2008