Estimation of Household Waste in the Republic of Serbia using R software

Melinda TOKAI (melinda.tokai@stat.gov.rs)
Statistical Office of the Republic of Serbia

ABSTRACT

This paper deals with the problem of estimation of annual amount of waste generated by households in Republic of Serbia. Waste generated by households is a part of municipal waste that also includes waste generated by trade and services activities as well as by tourists. In order to estimate pure household waste, regression analysis was preformed with reference to Cammarota et al. (2005) "A proposal for the estimation of household waste". In order to face this problem, regression models were constructed for municipal waste that are based on non domestic variables which are related to trade and services activities and tourism. The part of the municipal waste that could not be explained by a model based on non domestic variables was ascribed to pure household waste. In order to check validity of results, the model residuals were then related to domestic variables (usual population and the average number of inhabitants per occupied dwelling). The regression models were fitted using R software.

Keywords: *Household waste, regression models, R software*
JEL Classification: *C10, C88*

INTRODUCTION

The aim of this paper is to present the way household waste was estimated for reference year 2014 in SORS (Statistical Office of the Republic of Serbia). The need for this estimation came out of the Eurostat's demand to provide data on household waste amount in Serbia, so this was a first time in SORS that someone dealt with this issue with a notable time constrain. The key problem when estimating household waste in Serbia is that household waste is not directly detectable because municipal waste contains an unidentified amount of waste not generated by households. This paper presents the method how municipality waste was modeled with variables that are related to trade and services activities and tourism through linear regression. The part of municipality waste that could not be explained with this model is considered to be the pure household waste.

The collection system of municipal waste in Serbia is organized over the municipal territory with waste containers available to everyone with no mean to distinguish between different possible users (trade and services activities located near or in residential buildings). These containers are located along streets and near residential areas. Then municipalities report the entire amount of waste collected including both household waste and waste coming from trade and service activities. Moreover, tourists who spend their holydays in given cities can put their waste in the same containers. In order to face this problem, a simple regression model was set to estimate pure household waste starting from selected variables correlated with municipal waste. Models are fitted including non domestic variables; the outcoming residuals are then related to domestic variables in order to check results validity.

METHOD

Based on available data, we tried to identify a model which would allow estimation of the municipal waste amount that can be assigned uniquely to households. For this purpose we consider variables representing the most relevant sources of municipal waste, distinguishing them into non domestic and domestic variables. The non domestic variables are related to trade and services activities and tourism. We represented trade and services activities with the number of trade and services employees per inhabitant. As far as the tourism is concerned, it is represented by the ratio of tourist overnights stays and usual population.

Two variables are used to account for domestic waste's sources: usual population and the average number of inhabitants per occupied dwelling. The second indicator allows accounting for people not registered as usual residents but living in occupied dwellings. This last indicator is built on data coming from the Census of population in Serbia, in which information on occupied and unoccupied dwellings are collected.

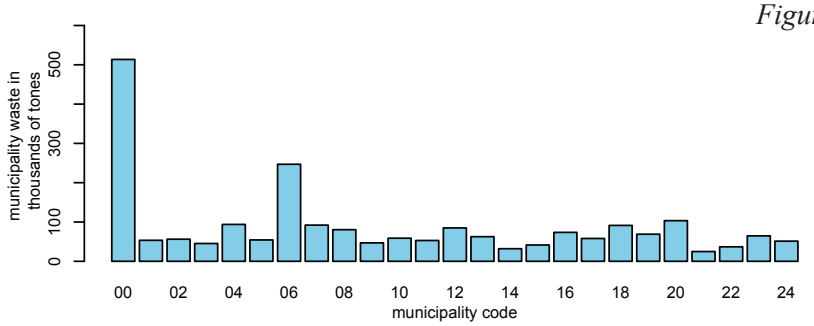
The part of the municipal waste that could not be explained by models based on non domestic variables can be ascribed to pure household waste. As a "validity check", the model residuals are than related to domestic variables.

Analyzing the data which were aggregated on a county level, it has come to our attention that the recorded data for the City of Belgrade differs significantly (see bar plot on Figure 1) from the rest of the data. Belgrade's contribution in the total value of the observed variable (municipality waste) is nearly 25% (see pie chart on Figure 2), which is much more then for the other counties. Because of this, we decided to create two regression models for the estimation of the household waste: one for all the counties except the City of Belgrade and one for the Belgrade municipalities.

```
par(mar = c(4, 4, 1, 1), mgp = c(2, 1, 0), cex = 0.6)
barplot(okrug$kom.3/1000, col='skyblue',names.arg = okrug$okr,
        ylab='municipality waste in \n thousands of tones',
        ylim=c(0,600), xlab='municipality code',)
```

Municipality waste amount

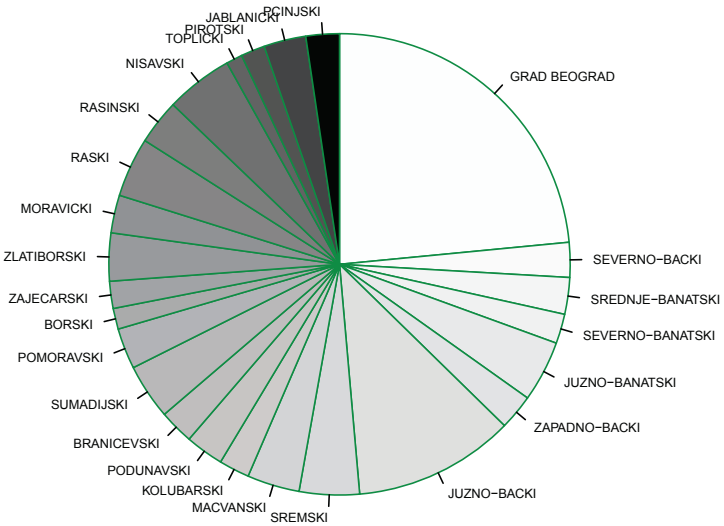
Figure 1




```
par(mar = c(1,1,5, 1), mgp = c(2, 1, 0), cex = 0.8)
pie(x=okrug$kom.3,init.angle=90,clockwise = T,labels=nazokr,cex=0.6,
    col=gray.colors(25,0.99,0.00001),lty=1, border='green4',lwd=6)
```

Pie chart of municipality waste amount

Figure 2



ESTIMATION PROCEDURE

As mentioned before, separate linear regression models were fitted for Serbian counties excluding the City of Belgrade and another for the Belgrade municipalities. More precisely, the weighted least squares method for fitting the best linear regression models was used. For this purpose the `lm()` function in the `stats R` package with a given `weights` argument was used.

The first model for all the counties except Belgrade is given by:

$$\log y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \qquad i = 1, \ldots, n \tag{1}$$

where n is the number of counties (except Belgrade) (n = 24), y_i is the amount of collected municipal waste in the counties (in tones), x_{1i} is the number of trade and service employees per inhabitant and x_{2i} is the ratio of tourist overnights stays and usual population. The model summary is in the Table 1.

```
okrug1=okrug[!okrug$nokr%in% "GRAD BEOGRAD",]
row.names(okrug1)=1:24
# linear model for Serbian counties excluding Belgrade
srb_bezBG=lm(log(kom.3)~I(zapos^2/br.stan)+I(nocenja/br.stan), data=okrug1,
        weights=okrug1$br.stan/sum(okrug1$br.stan))
```

Regression summary for model (1): municipality waste amount for Serbian counties without Belgrade

Table 1

Model (1)			
Coefficients	Estimate	Standard Error	t-value
$\hat{\alpha}$	1.084e+01	8.792e-02	123.311
$\hat{\beta}_1$	6.610e-05	7.451e-06	8.871
$\hat{\beta}_2$	6.758e-03	6.052e-02	0.112
Residual standard error = 0.05154 on 21 degrees of freedom Multiple R-Squared=0.791			

Once the linear model was fitted, the pure household waste was estimated as:

$$e^{\log y_i - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}}$$

For most of the counties the estimated pure household waste is approximately 70% of the municipal waste, which is in accordance with the estimates of the neighboring countries with similar economic situation where

pure household waste makes 60-85% of the total municipal waste. According to this model, for some counties the resulting estimate of the part of the municipality waste that can be ascribed to pure household waste is notably below (< 20%) or above (> 90%) the expected value. This is probably due to deficiencies of the data on which the model was based. For these counties imputation was applied, instead of the resulting estimates, the average of all the estimated values was used.

```
#average of the part of the municipality waste that can be ascribed
#to pure household (in all the counties except Belgrade)
sum(exp(residuals(srb_bezBG)+coef(srb_bezBG)[1]))/sum(okrug1$kom.3)

## [1] 0.7025566

#part of the municipality waste that can be ascribed to pure household
exp(residuals(srb_bezBG)+coef(srb_bezBG)[1])/okrug1$kom.3

##         1         2         3         4         5         6         7
## 0.7461787 0.8403478 0.8572623 0.7684268 0.8461123 0.1908959 0.7529099
##         8         9        10        11        12        13        14
## 0.7797441 0.8220857 0.8738667 0.8425341 0.7018723 0.8183730 0.8776869
##        15        16        17        18        19        20        21
## 0.8827785 0.7465805 0.7788016 0.7549611 0.8415400 0.5775235 0.9291225
##        22        23        24
## 0.9282892 0.8514118 0.8671652
```

As it can be seen from the previus output, results are somewhat suspicious for couinties Juzno Backi(6), Nisavski(20), Toplicki(21), Pirotski(22). There is were imputation come in handy.

```
#average of the part of the municipality waste that can be ascribed
#to pure household waste(in all the counties except Belgrade)
a=sum(exp(residuals(srb_bezBG)+coef(srb_bezBG)[1]))/sum(okrug1$kom.3)
#part of the municipality waste that can be ascribed to pure household waste
srb_bezBG_hw=exp(residuals(srb_bezBG)+coef(srb_bezBG)[1])/okrug1$kom.3
srb_bezBG_hw[c(6,20,21,22)]=a
#part of the municipality waste that can be ascribed to pure
#household waste after imputation
srb_bezBG_hw

##         1         2         3         4         5         6         7
## 0.7461787 0.8403478 0.8572623 0.7684268 0.8461123 0.7025566 0.7529099
##         8         9        10        11        12        13        14
## 0.7797441 0.8220857 0.8738667 0.8425341 0.7018723 0.8183730 0.8776869
##        15        16        17        18        19        20        21
## 0.8827785 0.7465805 0.7788016 0.7549611 0.8415400 0.7025566 0.7025566
##        22        23        24
## 0.7025566 0.8514118 0.8671652
```

For the City of Belgrade among all the fitted models the municipal waste was best estimated with the linear regression model given by:

$$\log y_i = a + \beta_3 x_{3i} + \beta_4 x_{4i} + e_i \qquad i = 1, \ldots, n \tag{2}$$

where n is the number of Belgrade municipalities ($n = 17$), y_i is the amount of collected municipal waste (in tones), x_{3i} is the number of trade and service employees and x_{4i} is the number of tourist overnights stays. The model summary is in Table 2.

```
#linear model for the municipalities of Belgrade
Bg=lm(log(kom.3)~zapos+nocenja, data=Beograd,
      weights=Beograd$br.stan/sum(Beograd$br.stan))
# the estimated part of the municipality waste in Belgrade that can be
#ascribed to the pure household waste
(bg_hw=sum(exp(residuals(Bg)+coef(Bg)[1]))/sum(Beograd$kom.3))

## [1] 0.6914103
```

Regression summary for model (2): municipality waste amount of Belgrade munici- pality

Table 2

Model (2)			
Coefficients	Estimate	Standard Error	t-value
â	1.006e+01	1.813e-01	55.457
$\hat{\beta}_3$	3.268e-05	8.716e-06	3.750
$\hat{\beta}_4$	-6.274e-06	2.333e-06	-2.689
Residual standard error = 0.1072 on 14 degrees of freedom Multiple R-Squared=0.5222			

The pure household waste is estimated as

$$e^{\log y_i - \hat{\beta}_3 x_{3i} - \hat{\beta}_4 x_{4i}}$$

The estimated pure household waste in the City of Belgrade is approximately 69% of the municipal waste.

We came to a conclusion that from 302 kilos of municipal waste per inhabitant, 76% is pure household waste that is 230 kilos. For the City of Belgrade this is a bit smaller 69% which can be explained with a large number of tourist overnights stays and the large number of trade and service employees.

VALIDITY CHECK

In order to validate the proposed models, we created another linear regression model where the response variable is the estimated amount of pure household waste and the predictors are the average number of inhabitants per occupied dwelling and the usual population. We should stress it here that the validation of the results was done on the county level with the full data (Belgrade is included as a county).

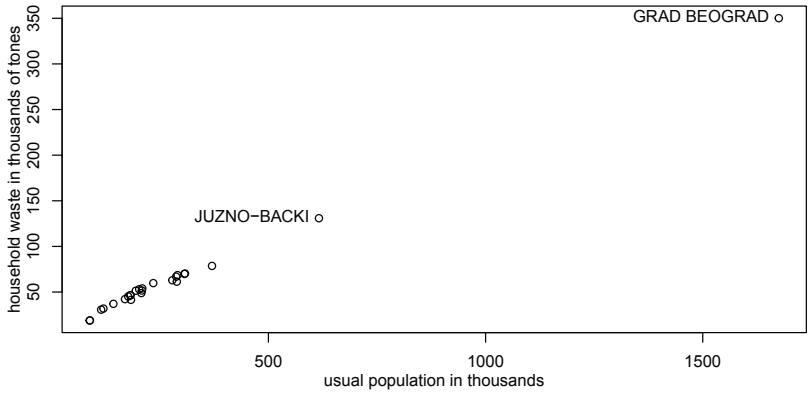
```
rezultat=data.frame(okrug[,c(1,2,4,5,6,8,9)],koef=c(bg_hw,srb_bezBG_hw))
rezultat$domaci=rezultat$koef*rezultat$agen.otpad
colnames(rezultat)
```

```
## [1] "okr"        "nokr"       "br.stan"    "zapos"      "nocenja"
## [6] "kom.otpad"  "agen.otpad" "koef"       "domaci"
```

```
par(mar = c(4, 4, 4, 1), mgp = c(2, 1, 0), cex = 0.8)
plot((rezultat$br.stan/1000),(rezultat$domaci/1000),
     xlab='usual population in thousands',
     ylab='household waste in thousands of tones')
text(rezultat$br.stan/1000,rezultat$domaci/1000, labels=naz,cex=1,pos=2)
```

Scatter plot of the relationship between household waste and number of usual popu- lation by counties

Figure 3




```
par(mar = c(4, 4, 4, 1), mgp = c(2, 1, 0), cex = 0.8)
plot(rezultat$br.stan[-c(1,7)]/1000,rezultat$domaci[-c(1,7)]/1000,
     xlab='usual population in thousands',
     ylab='household waste in thousands of tones')
```

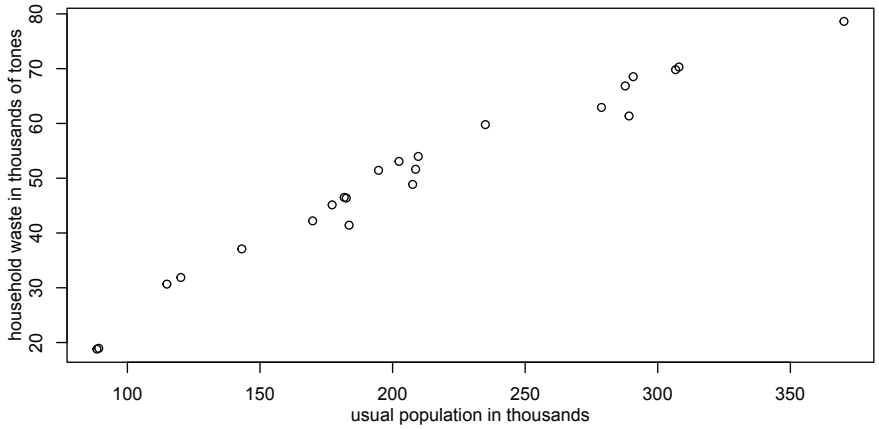
From the plots on Figure 3 and Figure 4 it is obvious that a very strong linear relationship exist between the household waste amount and the usual population. The regression model is given by:

$$z_i = b + \beta_5 x_{5i} + \beta_5 x_{5i} + \xi_i \qquad i = 1, \ldots, n \tag{3}$$

where z_i is the estimated amount of pure household waste (in tones), x_{5i} is the average number of inhabitants per dwelling and x_{6i} is the usual population. This model's summary is in Table 3.

Scatter plot of the relationship between household waste and number of usual population by counties (a closer look - without City of Belgrade and Juzna-Backa)

Figure 4



Regression summary for model (3): valdity check

Table 3

Coefficients	Estimate	Standard Error	t-value
$\hat{b}$	-4.823e+03	5.721e+03	-0.4083
$\hat{\beta}_5$	4.022e+03	1.905e+03	2.111
$\hat{\beta}_6$	2.046e-01	1.989e-03	102.886
Residual standard error = 0. 3013 on 22 degrees of freedom Multiple R-Squared=0. 9979			

The negative sign of the intercept in model (3) is most likely due to the wide range of the response variable.

CONCLUSION

Due to Serbian collection system of municipal waste information data related to pure house- hold waste in not available. Starting from the total amount of municipal waste in the Serbian counties we presented a simple procedure to estimate the pure household waste considering only the selected non domestic variables. Note, that this was the first time for SORS to conduct this kind of estimation and this was the best we could come up with in the given time, but there are plans in the future to continue model development.

References

1. **Cammarota M., Jona Lasinio G., Di Sarro T.,** 2005, *"A proposal for the estima- tion of household waste"*, Atti del Convergno intermedio SIS 2005, Statistica e ambiente, Messina, 21-23rd September 2005, pp. 215-218
2. **Cammarota M., Jona Lasinio G., Di Sarro T.**, 2006, *"Methods for the Analysis and Estimation of Household Waste"*
3. **R Core Team**, 2015, *R: A language and environment for statistical computing*, R Foun- dation for Statistical Computing, Vienna, Austria. ***https://www.R-project.org/***

Appendix

```
library(xlsx)
otpad=read.xlsx("../data/komunalni(2).xls", sheetIndex=1,
                colIndex=c(1,2,3,4,6), startRow=2, endRow=206, stringsAsFactors=F)
colnames(otpad)=c('mbops','opstina', 'kom.otpad','br.stan','agen.otpad')

otpad$kom.otpad=as.numeric(otpad$kom.otpad)
otpad$kom.otpad=round(otpad$kom.otpad)

otpad=otpad[substr(otpad$mbops, 1,2)!='RS',]
otpad=otpad[substr(otpad$opstina,1,4)!='Grad',]

#identifikacije okruga
library(sas7bdat)
sifrarnik=read.sas7bdat('../data/ops_2014.sas7bdat')
sifrarnik$okr=as.character(sifrarnik$okr)
sifrarnik$okr=as.character(sifrarnik$okr)
sifrarnik$nokr=as.character(sifrarnik$nokr)
sifrarnik=sifrarnik[c(1:122, 124:169),c(1,5,6)]
#dodajemo id. okr
otpad=merge(otpad,sifrarnik,by='mbops')
okr=c(paste0(0, as.character(0:9)), as.character(10:24))
nokr=c()
for(i in c(paste0(0, as.character(0:9)), as.character(10:24))){
  nokr=c(nokr, sifrarnik$nokr[min(which(sifrarnik$okr==i))])
}
# agregimo podatke na nivou okruga
okrug=data.frame(cbind(okr, nokr))
for(i in 1:25){
  okrug$br.stan[i]=sum(otpad$br.stan[otpad$okr==okr[i]])
  okrug$kom.otpad[i]=sum(otpad$kom.otpad[otpad$okr==okr[i]])
}
```

```
#########################################################################
#prosecna kolicina otpada po okruzima
okrug$koef=okrug$kom.otpad/okrug$br.stan

otpad2=merge(otpad, okrug[,c(1,5)], by='okr')
otpad2$kom.2=otpad2$koef*otpad2$br.stan
#mbops,brzap, nocenja
pomocni=read.xlsx("../data/otpad_modelnov_2014.xls", sheetIndex=2,
                  colIndex=c(1,3,6),endRow=169)

otpad2=merge(otpad2, pomocni, by='mbops')

library(data.table)
setnames(otpad2, 'totBr_zaposlenih','zapos')
setnames(otpad2,'br_nocenja_turista','nocenja')
otpad2=otpad2[,c(1,3,2,7,6,4,8,5,10,11,9)]

#########################################################################
#imputacije zbog NA u nocenju#
okrug$min=sapply(okrug$okr,
                 function(x) return(min(otpad2$nocenja[otpad2$okr==x], na.rm=T)))

otpad2=merge(otpad2,okrug[,c(1,6)], by='okr')
otpad2$nocenja[is.na(otpad2$nocenja)]=otpad2$min[is.na(otpad2$nocenja)]
#u Toplickom okrugu samo je za jednu opstinu dostupan podatak
#(Kursumlija - Djavolja varos)
otpad2$nocenja[otpad2$okr==21 & otpad2$mbops!='70688']=min(okrug$min)
otpad2=otpad2[,c(2,3,1,4,5,6,7,8,9,10,11,12)]
save(otpad2,file="otpad.Rda")

#########################################################################
okrug$agen.otpad=sapply(okrug$okr,
                  function(x) return(sum(otpad2$agen.otpad[otpad2$okr==x])))
okrug$zapos=sapply(okrug$okr,
              function(x) return(sum(otpad2$zapos[otpad2$okr==x])))
okrug$nocenja=sapply(okrug$okr,
               function(x) return(sum(otpad2$nocenja[otpad2$okr==x])))
okrug$kom.2=sapply(okrug$okr,
              function(x) return(sum(otpad2$kom.2[otpad2$okr==x])))

okrug=okrug[,c(1,2,5,3,8,9,6,4,7,10)]

okrug$a.otpad.stan=okrug$agen.otpad/okrug$br.stan
okrug$kom.3=((okrug$koef+okrug$a.otpad.stan)/2*okrug$br.stan)
save(okrug,file="okrug.Rda")
#########################################################################

okrug1=okrug[-1,]

####SRB BEZ BG
srb_bezBG=lm(log(kom.3)~I(zapos^2/br.stan)+I(nocenja/br.stan), data=okrug1,
        weights=okrug1$br.stan/sum(okrug1$br.stan))
summary(srb_bezBG)
sum(exp(residuals(srb_bezBG)+coef(srb_bezBG)[1]))/sum(okrug1$kom.3)
exp(residuals(srb_bezBG)+coef(srb_bezBG)[1])/okrug1$kom.3
```

```
#za 6,20,21,22  uzimam prosek
(a=sum(exp(residuals(srb_bezBG)+coef(srb_bezBG)[1]))/sum(okrug1$kom.3))
(srb_bezBG_hw=exp(residuals(srb_bezBG)+coef(srb_bezBG)[1])/okrug1$kom.3)
srb_bezBG_hw[c(6,20,21,22)]=a
srb_bezBG_hw
### Beogradske opstine
Beograd=otpad2[otpad2$okr=='00',]

save(Beograd,file="Beograd.Rda")

Beograd$kom.3=((Beograd$koef+Beograd$agen.otpad/Beograd$br.stan)/2)*Beograd$br.stan
Bg=lm(log(kom.3)~zapos+nocenja, data=Beograd,
      weights=Beograd$br.stan/sum(Beograd$br.stan))
summary(Bg)
(bg_hw=sum(exp(residuals(Bg)+coef(Bg)[1]))/sum(Beograd$kom.3))
#########################################################################
# rezultat
rezultat=data.frame(okrug[,c(1,2,4,5,6,8,9)],koef=c(bg_hw,srb_bezBG_hw))
write.xlsx(rezultat, 'C:/Melinda/otpad/v_2/rezultati.xls',
           showNA=F, sheetName='rezultat',col.names=T, row.names=F)

rezultat$domaci=rezultat$koef*rezultat$agen.otpad
# plot za odnos za proveru rezultata
naz=c("GRAD BEOGRAD", rep("",5),"JUZNO-BACKI", rep("",18))
plot(rezultat$domaci~rezultat$br.stan,
     main='Odnos domacinskog otpada i broja stanovnika \n po okruzima',
     xlab='broj stanovnika', ylab='kolicina otpada iz domacinstava')
text(rezultat$br.stan,rezultat$domaci, labels=naz,cex=0.5,pos=2)
plot(rezultat$domaci[-c(1,7)]~rezultat$br.stan[-c(1,7)],
     main='Odnos domacinskog otpada i broja stanovnika \n po okruzima',
     sub='kada se Uklone Beograd i Juzna Backa',
     xlab='broj stanovnika', ylab='kolicina otpada iz domacinstava')

summary(lm(domaci~br.stan, data=rezultat))
plot(residuals(lm(domaci~br.stan, data=rezultat)))


rezultat$dom=rezultat[,3]/rezultat[,10]
save(rezultat, file="rezultat.Rda")
summary(lm(domaci~dom+br.stan, data=rezultat))
plot(residuals(lm(domaci~dom+br.stan, data=rezultat)))
plot(rezultat$domaci~rezultat$dom)
```