

```
RGui - [R Console]
File Edit View Misc Packages Windows Help

> #Parametric and Nonparametric Correlation Measures in R
> library(DAAG)
> attach(elasticband)
The following object(s) are masked from 'elasticband (position 3)':

    distance, stretch
> cor(stretch, distance) #izracunava koeficijent korelacije izmedju dve promenljive
[1] 0.7969995
> cor.test(stretch, distance) #Pearson correlation

Pearson's product-moment correlation

data: stretch and distance
t = 2.9507, df = 5, p-value = 0.03186
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1099050 0.9686728
sample estimates:
      cor
0.7969995

> cor.test(stretch, distance, method="spearman") #Spearman's rho

Spearman's rank correlation rho

data: stretch and distance
S = 8.5747, p-value = 0.0162
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.8468812

Warning message:
In cor.test.default(stretch, distance, method = "spearman") :
  Cannot compute exact p-values with ties
> |
> cor.test(stretch, distance, method="kendall") #Kendall's tau

Kendall's rank correlation tau

data: stretch and distance
z = 2.1268, p-value = 0.03344
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau
0.6831301

Warning message:
In cor.test.default(stretch, distance, method = "kendall") :
  Cannot compute exact p-value with ties
> detach(elasticband)
> |
```

```
RGui - [R Console]
File Edit View Misc Packages Windows Help

> #The Model Formula in Straight Line Regression
> library(DAAG)
> plot(distance ~ stretch, data=elasticband) #elasticband je baza podataka iz paketa DAAG, koja sadri 7 vrsta i dve promenljive
> elastic.lm<- lm(distance ~ stretch, data=elasticband) #kod za izracunavanje regresije; kao izlaz dobija se objekat tipa lm
> elastic.lm

Call:
lm(formula = distance ~ stretch, data = elasticband)

Coefficients:
(Intercept)      stretch
      -63.571         4.554

> abline(elastic.lm)
> options(digits=4)
> summary(elastic.lm)

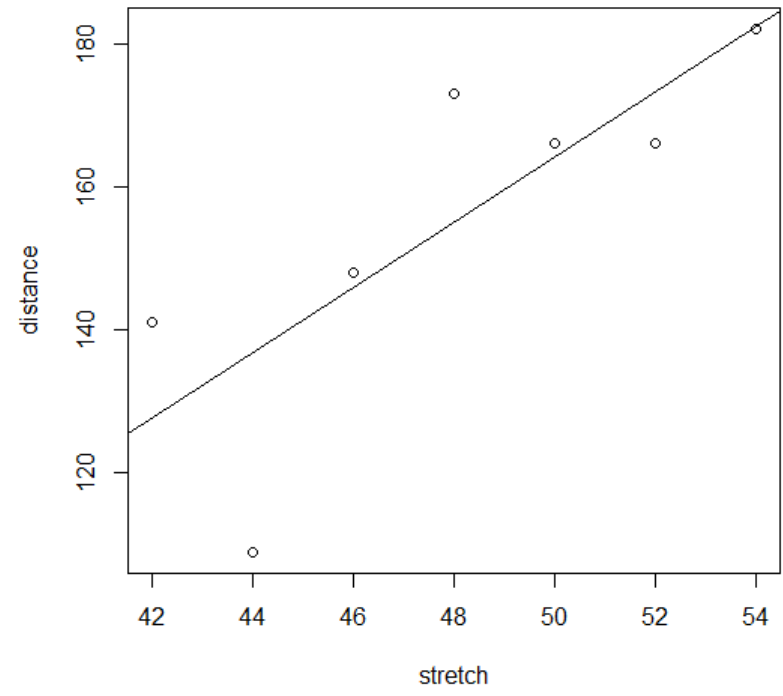
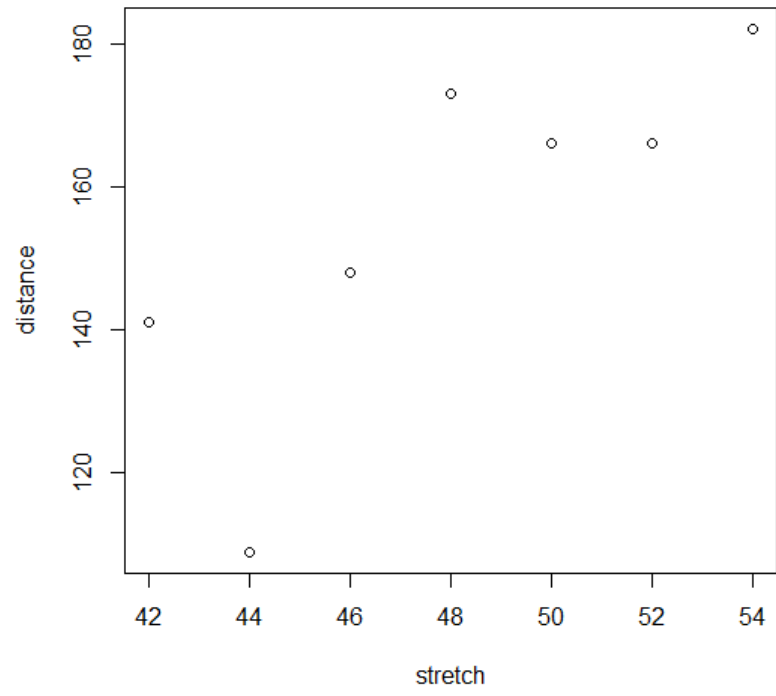
Call:
lm(formula = distance ~ stretch, data = elasticband)

Residuals:
    1     2     3     4     5     6     7
 2.107 -0.321 18.000  1.893 -27.786 13.321 -7.214

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -63.57     74.33   -0.86   0.431
stretch         4.55      1.54    2.95   0.032 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.3 on 5 degrees of freedom
Multiple R-squared:  0.635,    Adjusted R-squared:  0.562
F-statistic: 8.71 on 1 and 5 DF,  p-value: 0.0319

> |
```

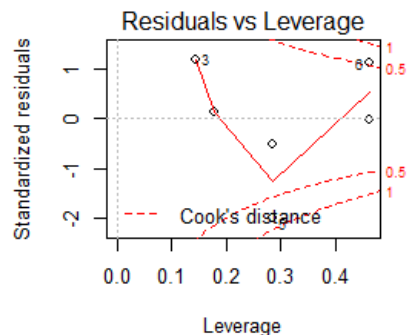
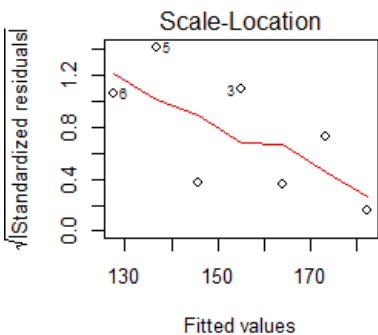
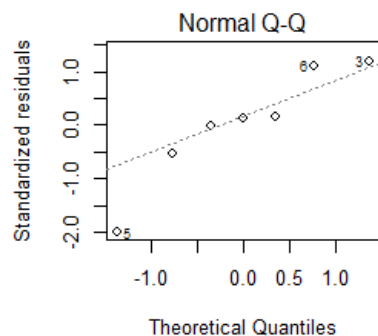
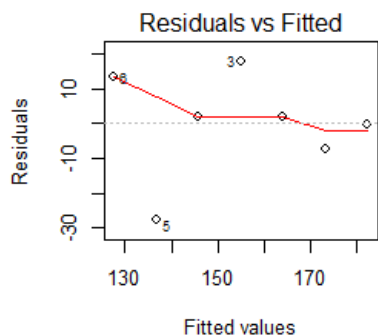


```

RGui - [R Console]
File Edit View Misc Packages Windows Help

> #objekat tipa lm je lista imenovanih komponenti
> names(elastic.lm) #dobijaju se imena komponenti
[1] "coefficients" "residuals" "effects" "rank" "fitted.values" "assign" "qr" "df.residual" "xlevels" "call"
[11] "terms" "model"
> #dostupne su razne f-je kojima se mogu preuzeti potrebne informacije iz ove liste; primeri su:
> coef(elastic.lm)
(Intercept) stretch
-63.571 4.554
> resid(elastic.lm)
 1 2 3 4 5 6 7
2.1071 -0.3214 18.0000 1.8929 -27.7857 13.3214 -7.2143
> #najcesce korisca f-ja za razmatranje regresije je summary()
> #postoji plot metod namenjen bas lm objektima, koji daje "dijagnosticke" informacije
> #on se dobija sledecim kodom:
> par(mfrow=c(2,2))
> plot(elastic.lm)
> par(mfrow=c(1,1))
> |

```



<code>print</code>	returns a brief summary
<code>summary</code>	returns a detailed summary
<code>df.residual</code>	returns the number of residual degrees of freedom
<code>coef</code>	returns the estimated coefficients (sometimes with their standard-errors)
<code>residuals</code>	returns the residuals
<code>deviance</code>	returns the deviance
<code>fitted</code>	returns the fitted values
<code>logLik</code>	computes the logarithm of the likelihood and the number of parameters
<code>AIC</code>	computes the Akaike information criterion or AIC (depends on <code>logLik()</code>)

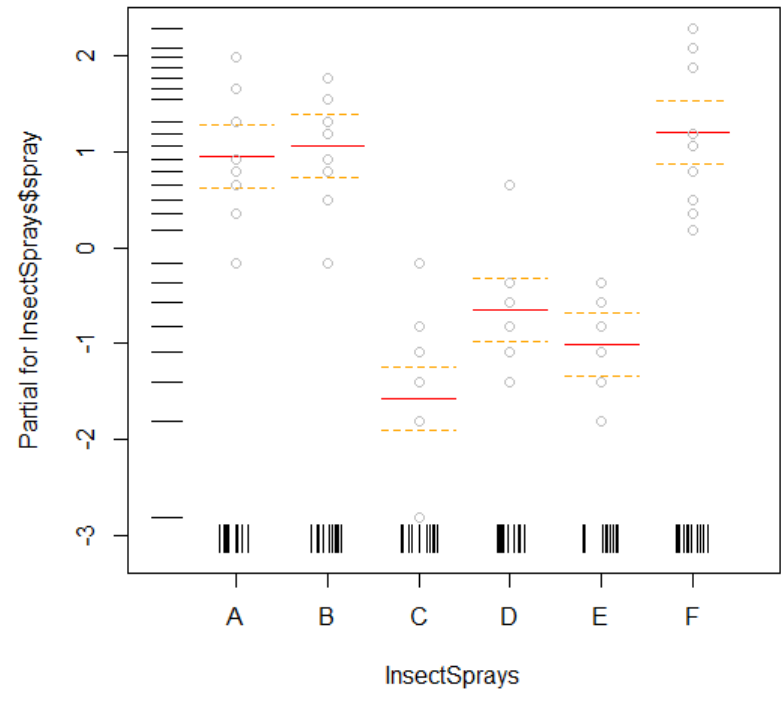
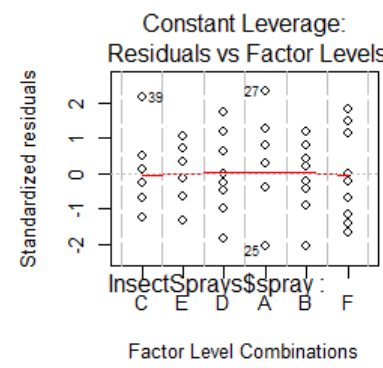
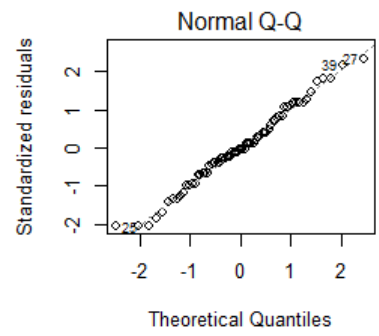
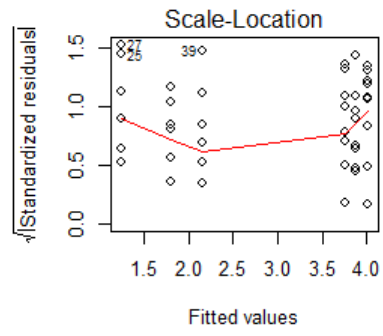
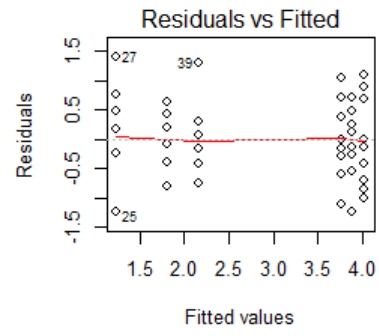
Ovo su glavne generičke f-je, koje se mogu koristiti za preuzimanje potrebnih informacija iz objekata, dobijenih analizom

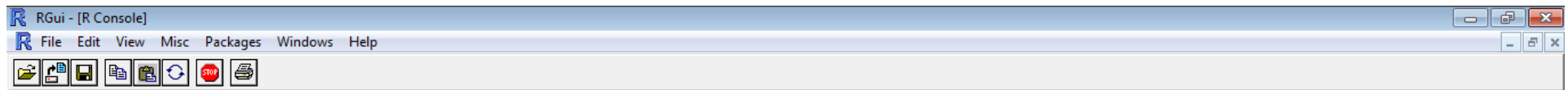
```
RGui - [R Console]
File Edit View Misc Packages Windows Help

> #The Analysis of Variance
> data(InsectSprays) #baza sadrzi 72 observacije
> #sest insekticida je tretirano u poljskim uslovima, svaki insekticid po 12 puta; jedna promenljiva je tip insekticida (spray), druga broj tretiranih insekata (count)
> aov.spray <- aov(sqrt(count) ~ spray, data=InsectSprays) #drugi arg. precizira da se promenljive, na koje se primenjuje formula i analiza, nalaze u bazi InsectSprays
>
> aov.spray <- aov(sqrt(InsectSprays$count) ~ InsectSprays$spray) #ovakva sintaksa ekvivalentna je onoj u prethodnom redu
> aov.spray #ekvivalentno ovom pozivu bilo bi da se otkuca: print(aov.spray)
Call:
aov(formula = sqrt(InsectSprays$count) ~ InsectSprays$spray)

Terms:
                InsectSprays$spray Residuals
Sum of Squares      88.43787   26.05798
Deg. of Freedom         5         66

Residual standard error: 0.6283453
Estimated effects may be unbalanced
>
> opar<-par() #promenljiva opar cuva trenutna podesavanja u vezi sa crtanjem
> par(mfcol=c(2,2))
> plot(aov.spray)
> par(opar)
Warning messages:
1: In par(opar) : graphical parameter "cin" cannot be set
2: In par(opar) : graphical parameter "cra" cannot be set
3: In par(opar) : graphical parameter "csi" cannot be set
4: In par(opar) : graphical parameter "cxy" cannot be set
5: In par(opar) : graphical parameter "din" cannot be set
>
> termplot(aov.spray, se=T, partial.resid=T, rug=T)|
```





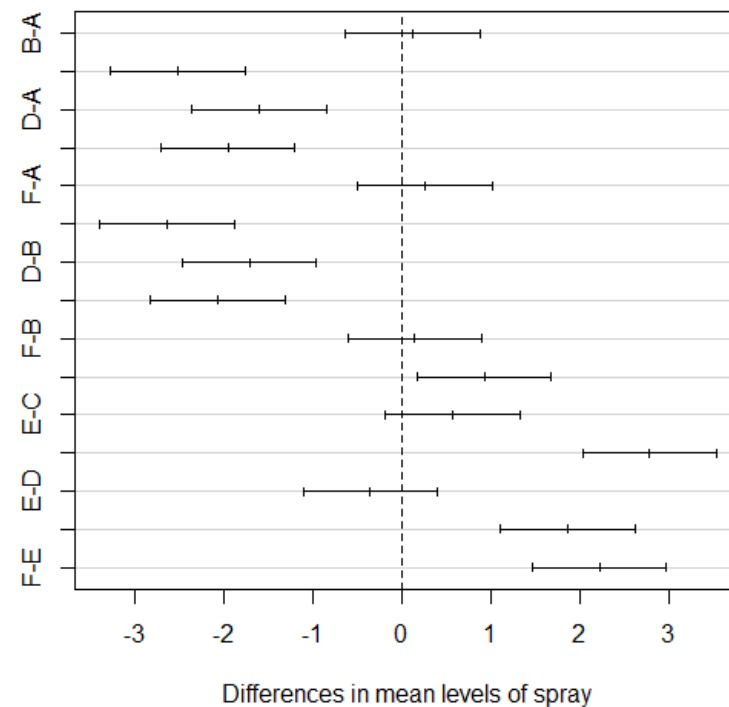
```
> aov.spray <- aov(sqrt(count)~ spray,data=InsectSprays)
> spray_t <- TukeyHSD(aov.spray) #spravodi se post-hoc test
> #post-hoc analiza sastoji se u tome da se posmatraju podaci, nakon sto je experiment završen, a u cilju otkrivanja sablona, koji nisu uoceni a priori
> spray_t
  Tukey multiple comparisons of means
    95% family-wise confidence level
```

Fit: aov(formula = sqrt(count) ~ spray, data = InsectSprays)

\$spray	diff	lwr	upr	p adj
B-A	0.1159530	-0.6369601	0.8688661	0.9975245
C-A	-2.5158217	-3.2687349	-1.7629086	0.0000000
D-A	-1.5963245	-2.3492377	-0.8434114	0.0000006
E-A	-1.9512174	-2.7041305	-1.1983042	0.0000000
F-A	0.2579388	-0.4949744	1.0108519	0.9144964
C-B	-2.6317747	-3.3846879	-1.8788616	0.0000000
D-B	-1.7122775	-2.4651907	-0.9593644	0.0000001
E-B	-2.0671704	-2.8200835	-1.3142572	0.0000000
F-B	0.1419858	-0.6109274	0.8948989	0.9935788
D-C	0.9194972	0.1665841	1.6724103	0.0080813
E-C	0.5646043	-0.1883088	1.3175175	0.2512638
F-C	2.7737605	2.0208474	3.5266736	0.0000000
E-D	-0.3548928	-1.1078060	0.3980203	0.7366389
F-D	1.8542633	1.1013502	2.6071764	0.0000000
F-E	2.2091561	1.4562430	2.9620693	0.0000000

```
> plot(spray_t)
> |
```

95% family-wise confidence level



```
RGui - [R Console]
File Edit View Misc Packages Windows Help

> #One-sample t-test
> #pp. da podaci imaju normalnu raspodelu; testira se hipoteza (H0) da je mu=mu0
> salm <- c(0.593, 0.142, 0.329, 0.691, 0.231, 0.793, 0.519, 0.392, 0.418) #meren je nivo Sallmonela u 9 slucajno odabranih sladoleda (jedinica je MPN/g)
> mean(salm)
[1] 0.4564444
> sd(salm)
[1] 0.2128439
> quantile(salm)
 0%  25%  50%  75% 100%
0.142 0.329 0.418 0.593 0.793
> qqnorm(salm)
> qqline(salm)
> t.test(salm, mu=0.3) #two-sided test

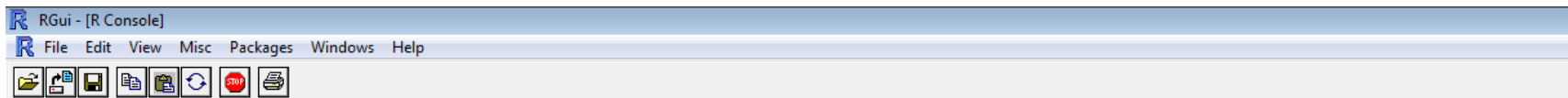
      One Sample t-test

data:  salm
t = 2.2051, df = 8, p-value = 0.05853
alternative hypothesis: true mean is not equal to 0.3
95 percent confidence interval:
 0.2928381 0.6200508
sample estimates:
mean of x
0.4564444

> t.test(salm, alternative="greater", mu=0.3) #one-sided alternativa moze biti 'greater' ili 'less'

      One Sample t-test

data:  salm
t = 2.2051, df = 8, p-value = 0.02927
alternative hypothesis: true mean is greater than 0.3
95 percent confidence interval:
 0.3245133      Inf
sample estimates:
mean of x
0.4564444
|
```

```
> #Two-sample t-test
> #testira se hipoteza da dva uzorka dolaze iz raspodela sa jednakim ocekivanjem  $\mu_1 - \mu_2 = 0$ 
> Control <- c(91, 87, 99, 77, 88, 91)
> Treat <- c(101, 110, 103, 93, 99, 104)
> t.test(Control, Treat, alternative="less", var.equal=TRUE) #alternativna hipoteza je  $\mu_1 - \mu_2 < 0$ ; pp. se da su disperzije raspodela jednake
```

Two Sample t-test

```
data: Control and Treat
t = -3.4456, df = 10, p-value = 0.003136
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -6.082744
sample estimates:
mean of x mean of y
 88.83333 101.66667
```

```
> t.test(Control, Treat, alternative="less") #ne pretpostavlja se jednakost disperzija
```

Welch Two Sample t-test

```
data: Control and Treat
t = -3.4456, df = 9.48, p-value = 0.003391
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -6.044949
sample estimates:
mean of x mean of y
 88.83333 101.66667
```

```
> #testiranje jednakosti disperzija
> var.test(Control, Treat)
```

F test to compare two variances

```
data: Control and Treat
F = 1.6119, num df = 5, denom df = 5, p-value = 0.6131
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2255582 11.5194293
sample estimates:
ratio of variances
 1.611925
```

```
> |
```

```
RGui - [R Console]
File Edit View Misc Packages Windows Help

> #Wilcoxon signed-rank test
> #spada u 'distribution-free' testove
> wilcox.test(salm, mu=0.3)

      Wilcoxon signed rank test

data:  salm
V = 38, p-value = 0.07422
alternative hypothesis: true location is not equal to 0.3

>
> #Two-sample Wilcoxon test ili, u literaturi, Mann-Whitney sample Wilcoxon test
> wilcox.test(Control, Treat)

      Wilcoxon rank sum test with continuity correction

data:  Control and Treat
W = 1.5, p-value = 0.01014
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(Control, Treat) :
  cannot compute exact p-value with ties
> wilcox.test(Control, Treat, alternative="l")

      Wilcoxon rank sum test with continuity correction

data:  Control and Treat
W = 1.5, p-value = 0.005069
alternative hypothesis: true location shift is less than 0

Warning message:
In wilcox.test.default(Control, Treat, alternative = "l") :
  cannot compute exact p-value with ties
> |
```

```
RGui - [R Console]
File Edit View Misc Packages Windows Help

> #The paired t-test
> #koristi se kada postoje dva merenja za istu experimentalnu jedinicu
> reg <- c(16, 20, 21, 22, 23, 22, 27, 25, 27, 28)
> prem <- c(19, 22, 24, 24, 25, 25, 26, 26, 28, 32)
> t.test(prem, reg, alternative="greater", paired=TRUE)

      Paired t-test

data:  prem and reg
t = 4.4721, df = 9, p-value = 0.0007749
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.180207      Inf
sample estimates:
mean of the differences
                2

> t.test(prem-reg, alternative="greater") #daje izlaz ekvivalentan onom koji se dobija prethodnom komandom

      One Sample t-test

data:  prem - reg
t = 4.4721, df = 9, p-value = 0.0007749
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 1.180207      Inf
sample estimates:
mean of x
                2

> #POGRESNO! t.test(prem,reg)
> |
```

```
RGui - [R Console]
File Edit View Misc Packages Windows Help

> #Tabelarni podaci
> #Single sample proportion test
> #bazira se na binomnoj raspodeli B(n,p)
> #Primer: Pp. baca se novcic baca 20 puta i da je od toga 12 puta pala G. Sa pragom znacajnosti od .05 testirati hipotezu da je ovaj novcic fer (regularan).
> prop.test(12, 20, p=.5, correct=F)

      1-sample proportions test without continuity correction

data: 12 out of 20, null probability 0.5
X-squared = 0.8, df = 1, p-value = 0.3711
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.3865815 0.7811935
sample estimates:
 p
0.6

>
> binom.test(12, 20, 0.5)

      Exact binomial test

data: 12 and 20
number of successes = 12, number of trials = 20, p-value = 0.5034
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.3605426 0.8088099
sample estimates:
probability of success
      0.6

> |
```

```
RGui - [R Console]
File Edit View Misc Packages Windows Help

> #Comparison of two population proportions test
> library(MASS)
> data(quine)
> str(quine)
'data.frame':  146 obs. of  5 variables:
 $ Eth : Factor w/ 2 levels "A","N": 1 1 1 1 1 1 1 1 1 1 ...
 $ Sex : Factor w/ 2 levels "F","M": 2 2 2 2 2 2 2 2 2 2 ...
 $ Age : Factor w/ 4 levels "F0","F1","F2",...: 1 1 1 1 1 1 1 1 2 2 ...
 $ Lrn : Factor w/ 2 levels "AL","SL": 2 2 2 1 1 1 1 1 2 2 ...
 $ Days: int  2 11 14 5 5 13 20 22 6 6 ...
> table(quine$Eth, quine$Sex)

      F M
A 38 31
N 42 35
> prop.test(table(quine$Eth, quine$Sex), correct=FALSE)

      2-sample test for equality of proportions without continuity correction

data:  table(quine$Eth, quine$Sex)
X-squared = 0.0041, df = 1, p-value = 0.9491
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.1564218  0.1669620
sample estimates:
 prop 1    prop 2 
0.5507246 0.5454545

> fisher.test(table(quine$Eth, quine$Sex)) #kao argument prima matrice ili tabele kontingencije

      Fisher's Exact Test for Count Data

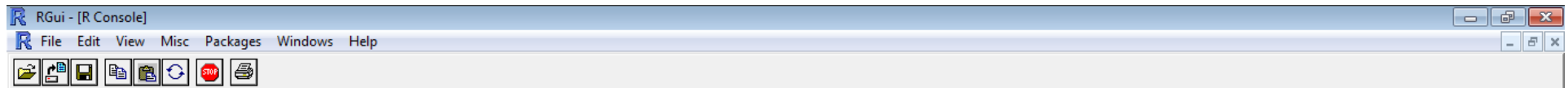
data:  table(quine$Eth, quine$Sex)
p-value = 1
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.505342 2.066744
sample estimates:
odds ratio
 1.021363

> chisq.test(table(quine$Eth, quine$Sex)) #kao argument prima matrice ili tabele kontingencije

      Pearson's Chi-squared test with Yates' continuity correction

data:  table(quine$Eth, quine$Sex)
X-squared = 0.0105, df = 1, p-value = 0.9182

> |
```



```
> #Multiple Proportions test
> smokers <- c(83, 90, 129, 70)
> patients <- c(86, 93, 136, 82)
> prop.test(smokers, patients)

      4-sample test for equality of proportions without continuity correction

data: smokers out of patients
X-squared = 12.6004, df = 3, p-value = 0.005585
alternative hypothesis: two.sided
sample estimates:
 prop 1   prop 2   prop 3   prop 4 
0.9651163 0.9677419 0.9485294 0.8536585

> prop.trend.test(smokers, patients)

      Chi-squared Test for Trend in Proportions

data: smokers out of patients ,
      using scores: 1 2 3 4
X-squared = 8.2249, df = 1, p-value = 0.004132

> prop.trend.test(smokers, patients, c(0,0,0,1)) #poslednji argument je 'score' (Group score)

      Chi-squared Test for Trend in Proportions

data: smokers out of patients ,
      using scores: 0 0 0 1
X-squared = 12.1731, df = 1, p-value = 0.0004848

> data(survey, package="MASS")
> str(survey)
'data.frame':  237 obs. of  12 variables:
 $ Sex   : Factor w/ 2 levels "Female","Male": 1 2 2 2 2 1 2 1 2 2 ...
 $ Wr.Hnd: num  18.5 19.5 18 18.8 20 18 17.7 17 20 18.5 ...
 $ NW.Hnd: num  18 20.5 13.3 18.9 20 17.7 17.7 17.3 19.5 18.5 ...
 $ W.Hnd : Factor w/ 2 levels "Left","Right": 2 1 2 2 2 2 2 2 2 2 ...
 $ Fold  : Factor w/ 3 levels "L on R","Neither",...: 3 3 1 3 2 1 1 3 3 3 ...
 $ Pulse : int  92 104 87 NA 35 64 83 74 72 90 ...
 $ Clap  : Factor w/ 3 levels "Left","Neither",...: 1 1 2 2 3 3 3 3 3 3 ...
 $ Exer  : Factor w/ 3 levels "Freq","None",...: 3 2 2 2 3 3 1 1 3 3 ...
 $ Smoke : Factor w/ 4 levels "Heavy","Never",...: 2 4 3 2 2 2 2 2 2 2 ...
 $ Height: num  173 178 NA 160 165 ...
 $ M.I   : Factor w/ 2 levels "Imperial","Metric": 2 1 NA 2 2 1 1 2 2 2 ...
 $ Age   : num  18.2 17.6 16.9 20.3 23.7 ...
```

```
RGui - [R Console]
File Edit View Misc Packages Windows Help

>
> #izvlacenje podataka iz izlaza testa
> reg <- c(16, 20, 21, 22, 23, 22, 27, 25, 27, 28)
> prem <- c(19, 22, 24, 24, 25, 25, 26, 26, 28, 32)
> t.test(prem-reg, alternative="greater")

      One Sample t-test

data:  prem - reg
t = 4.4721, df = 9, p-value = 0.0007749
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 1.180207      Inf
sample estimates:
mean of x
      2

> names(t.test(prem-reg, alternative="greater")) #ovom komandom stice se uvid u to koje sve podatke sadrzi izlaz testa
[1] "statistic" "parameter" "p.value" "conf.int" "estimate" "null.value" "alternative" "method" "data.name"
> t.test(prem-reg, alternative="greater")$p.value #dobija se p-vrednost testa
[1] 0.000774943
> t.test(prem-reg, alternative="greater")$conf.int #dobija se interval poverenja sacuvan u atomickom objektu,
[1] 1.180207      Inf
attr(,"conf.level")
[1] 0.95
> #koji se f-jom 'as.numeric()' moze konvertovati u num vektor
> |
```

```

> (table(survey$Smoke)->smokers)
Heavy Never Occas Regul
  11  189   19   17
> chisq.test(smokers, p=c(.1, .7, .1, .1))

      Chi-squared test for given probabilities

data: smokers
X-squared = 12.8983, df = 3, p-value = 0.004862

> chisq.test(smokers, p=c(.1, .7, .1, .1))$expected
Heavy Never Occas Regul
 23.6 165.2  23.6  23.6
> chisq.test(smokers, p=c(.1, .7, .1, .1))$observed

Heavy Never Occas Regul
  11  189   19   17
> chisq.test(smokers, p=c(.1, .7, .1, .1))$resid

      Heavy      Never      Occas      Regul
-2.593669  1.851706 -0.946895 -1.358588
> |

> ?HairEyeColor
> data(HairEyeColor)
> (HairEye <- margin.table(HairEyeColor, c(1, 2)))
      Eye
Hair  Brown Blue Hazel Green
Black   68  20   15    5
Brown  119  84   54   29
Red     26  17   14   14
Blond   7   94  10   16
> chisq.test(HairEye)

      Pearson's Chi-squared test

data: HairEye
X-squared = 138.2898, df = 9, p-value < 2.2e-16

> fisher.test(HairEye)
Error in fisher.test(HairEye) : FEXACT error 6.
LDKEY is too small for this problem.
Try increasing the size of the workspace.
> #POGLEDAI 'prop.table()'
> |

```