Vilniaus Universitetas

Regresinė analizė

Laboratorinis darbas

Darbą atliko:

Vainius Gataveckas, Matas Gaulia, Dovydas Martinkus

Duomenų Mokslas

3 kursas 2 gr.

Vilnius, 2021

# Naudoti metodai

Darbas atliktas naudojant R, SAS ir Python.

Naudoti R paketai:

*tidyverse*.

*janitor*

*car*

*lmtest*

*RcmdrMisc*

*lm.beta*

*psych*

*ppcor*

# Duomenys ir jų šaltiniai

Šalių gyventojų vidutinė gyvenimo trukmė pagal sveikatos rodiklius.

Duomenų šaltinis - Kaggle. Prieiga per internetą: <https://www.kaggle.com/kumarajarshi/life-expectancy-who>

Originalus šaltinis – WHO.

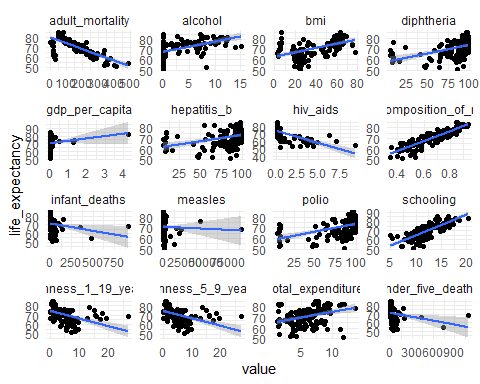
# Atliktos analizės aprašymas

**1. Naudojant R**

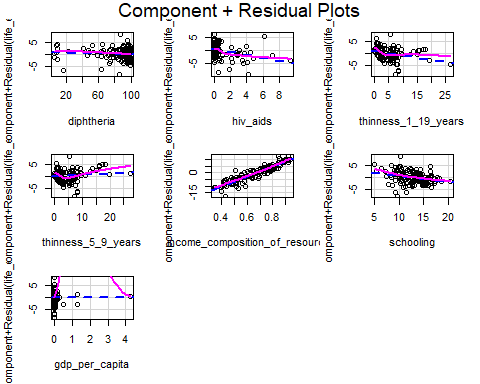
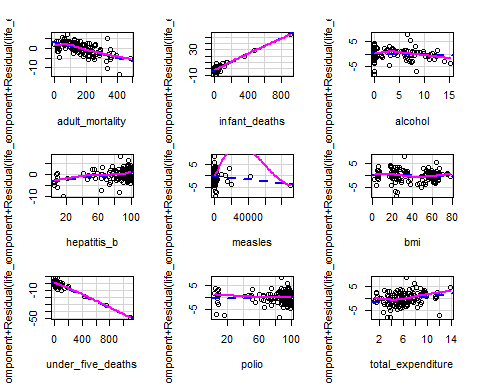
library(tidyverse)  
library(car)  
library(janitor)  
x <- read\_csv("life.csv") %>% clean\_names()

Tikslas: prognozuoti vidutinę gyvenimo trukmę šalyje pagal tam tikrus sveikatos rodiklius.

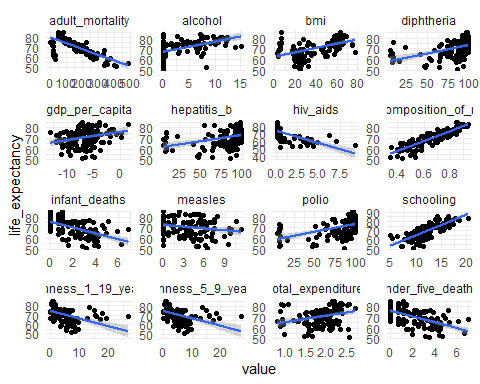
set.seed(100)  
transform\_1<- function(x) {  
 x %>%  
 mutate(gdp\_per\_capita = gdp / population) %>%  
 group\_by(country) %>%  
 fill(everything(), .direction = "up") %>%  
 dplyr::select(-c(1, 3), -population, -gdp, -percentage\_expenditure) %>%  
 drop\_na() %>%  
 ungroup() %>%  
 dplyr::select(-1)  
}  
  
x <- transform\_1(x)  
  
x\_1 <- x %>% filter(year == max(year)) %>% select(-1)  
  
# atskiri duomenys, patikrinti kaip gautas galutinis modelis progrnozuoja reikšmes  
x\_predict <- x %>% filter(year != max(year)) %>% slice\_sample(n=10) %>% select(-1)  
  
  
# kaikurių kovariančių priklausomybę nėra tiesinė  
x\_1 %>% pivot\_longer(-1) %>% ggplot(aes(x=value,y=life\_expectancy)) + facet\_wrap(vars(name),scales="free") + geom\_point() + geom\_smooth(method="lm") + theme\_minimal()



model <- lm(life\_expectancy ~ ., data = x\_1)  
crPlots(model)



transform\_2 <- function(x) {  
 x %>%   
 mutate(gdp\_per\_capita = log(gdp\_per\_capita),  
 infant\_deaths = log(infant\_deaths + 1),  
 measles = log(measles + 1),  
 total\_expenditure = log(total\_expenditure + 1),  
 under\_five\_deaths = log(under\_five\_deaths + 1)  
 )  
}  
  
# transformuojamos kaikurios kovariantės  
x\_2 <- transform\_2(x\_1)  
x\_predict <- transform\_2(x\_predict)  
  
  
# Kintamųjų tiesinis ryšys patikrinamas dar kartą  
x\_2 %>% pivot\_longer(-1) %>% ggplot(aes(x=value,y=life\_expectancy)) + facet\_wrap(vars(name),scales="free") + geom\_point() + geom\_smooth(method="lm") + theme\_minimal()

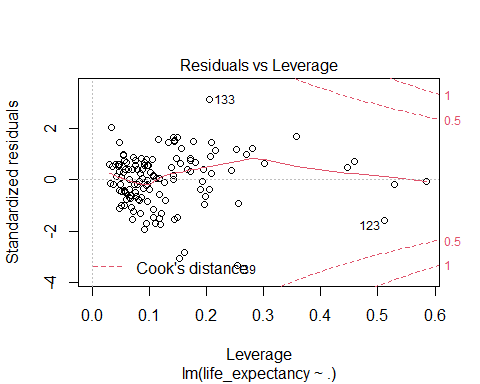
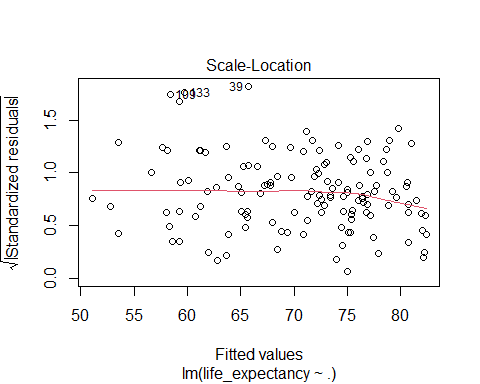
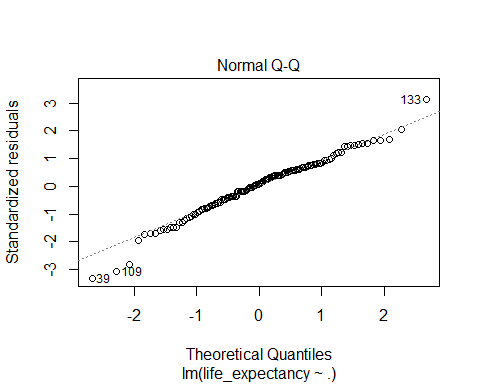
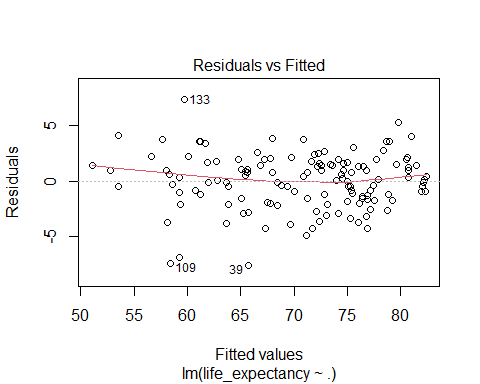


Modifikuoti duomenys išsaugomi faile „life\_modified.csv“.

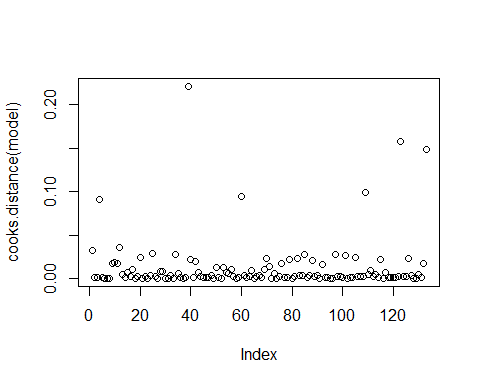
write.csv(x\_2, "life\_modified.csv")  
  
# Sukuriamas modelis  
model <- lm(life\_expectancy ~ ., data = x\_2)

Modelio prielaidos

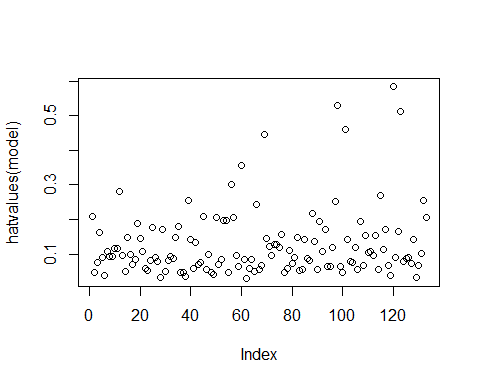
# Tikrinamas liekanų normalumas, homoskadiškumas, liekanų nepriklausomumas, išskirtys  
plot(model)



plot(cooks.distance(model))



plot(hatvalues(model))



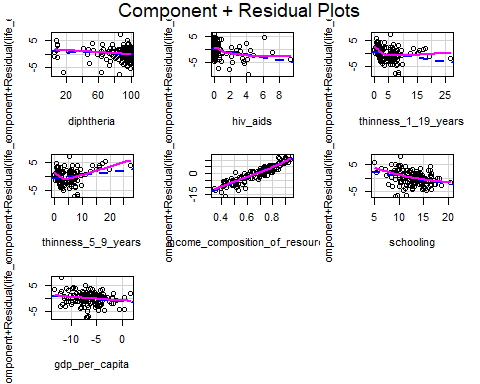
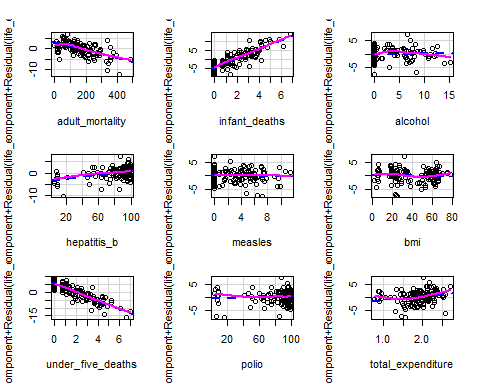
# Liekanų normalumo testas  
shapiro.test(residuals(model))

##   
## Shapiro-Wilk normality test  
##   
## W = 0.98195, p-value = 0.07493

# Homoskadiškumo testas  
library(lmtest)  
bptest(model)

## studentized Breusch-Pagan test  
## BP = 13.511, df = 16, p-value = 0.6351

crPlots(model)



anova(model) # Tikrinama hipotezė H0: beta\_1 = beta\_2 = ... = 0

## Analysis of Variance Table  
##   
## Response: life\_expectancy  
## Df Sum Sq Mean Sq F value Pr(>F)   
## adult\_mortality 1 4541.4 4541.4 658.0923 < 2.2e-16 \*\*\*  
## infant\_deaths 1 714.3 714.3 103.5021 < 2.2e-16 \*\*\*  
## alcohol 1 631.9 631.9 91.5693 2.427e-16 \*\*\*  
## hepatitis\_b 1 278.4 278.4 40.3488 4.305e-09 \*\*\*  
## measles 1 0.2 0.2 0.0300 0.8628941   
## bmi 1 152.7 152.7 22.1288 7.095e-06 \*\*\*  
## under\_five\_deaths 1 238.6 238.6 34.5813 4.022e-08 \*\*\*  
## polio 1 78.7 78.7 11.4067 0.0009967 \*\*\*  
## total\_expenditure 1 33.3 33.3 4.8273 0.0300005 \*   
## diphtheria 1 9.6 9.6 1.3904 0.2407448   
## hiv\_aids 1 50.6 50.6 7.3376 0.0077755 \*\*   
## thinness\_1\_19\_years 1 53.1 53.1 7.6883 0.0064776 \*\*   
## thinness\_5\_9\_years 1 6.9 6.9 0.9952 0.3205464   
## income\_composition\_of\_resources 1 766.0 766.0 110.9948 < 2.2e-16 \*\*\*  
## schooling 1 9.0 9.0 1.3108 0.2546025   
## gdp\_per\_capita 1 19.2 19.2 2.7882 0.0976592 .   
## Residuals 116 800.5 6.9   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Modelio parinkimas

# Požinksninė regresija  
library(RcmdrMisc)  
model\_2 <- stepwise(model)

##   
## Direction: backward/forward  
## Criterion: BIC   
##   
## Step: AIC=278.2  
## life\_expectancy ~ adult\_mortality + hepatitis\_b + total\_expenditure +   
## hiv\_aids + income\_composition\_of\_resources  
##   
## Df Sum of Sq RSS AIC  
## <none> 863.91 278.20  
## - total\_expenditure 1 37.46 901.37 278.96  
## + measles 1 11.09 852.82 281.37  
## + schooling 1 8.38 855.52 281.79  
## + thinness\_1\_19\_years 1 8.26 855.65 281.81  
## + under\_five\_deaths 1 6.98 856.93 282.01  
## + gdp\_per\_capita 1 6.83 857.08 282.04  
## + thinness\_5\_9\_years 1 5.20 858.71 282.29  
## + infant\_deaths 1 5.00 858.90 282.32  
## - hiv\_aids 1 61.54 925.45 282.46  
## + polio 1 2.30 861.60 282.74  
## + alcohol 1 2.23 861.68 282.75  
## + bmi 1 0.30 863.61 283.04  
## + diphtheria 1 0.17 863.73 283.06  
## - hepatitis\_b 1 89.00 952.91 286.35  
## - adult\_mortality 1 248.42 1112.32 306.92  
## - income\_composition\_of\_resources 1 2064.50 2928.40 435.67

Parametrų vertinimas ir interpretacija

# Koeficientai  
summary(model\_2)

##   
## Call:  
## lm(formula = life\_expectancy ~ adult\_mortality + hepatitis\_b +   
## total\_expenditure + hiv\_aids + income\_composition\_of\_resources,   
## data = x\_2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.1512 -1.5507 0.2728 1.6248 8.3196   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 46.015816 1.879961 24.477 < 2e-16 \*\*\*  
## adult\_mortality -0.019823 0.003280 -6.043 1.56e-08 \*\*\*  
## hepatitis\_b 0.035768 0.009888 3.617 0.000428 \*\*\*  
## total\_expenditure 1.383667 0.589638 2.347 0.020491 \*   
## hiv\_aids -0.608046 0.202160 -3.008 0.003174 \*\*   
## income\_composition\_of\_resources 33.937181 1.948050 17.421 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.608 on 127 degrees of freedom  
## Multiple R-squared: 0.897, Adjusted R-squared: 0.8929   
## F-statistic: 221.1 on 5 and 127 DF, p-value: < 2.2e-16

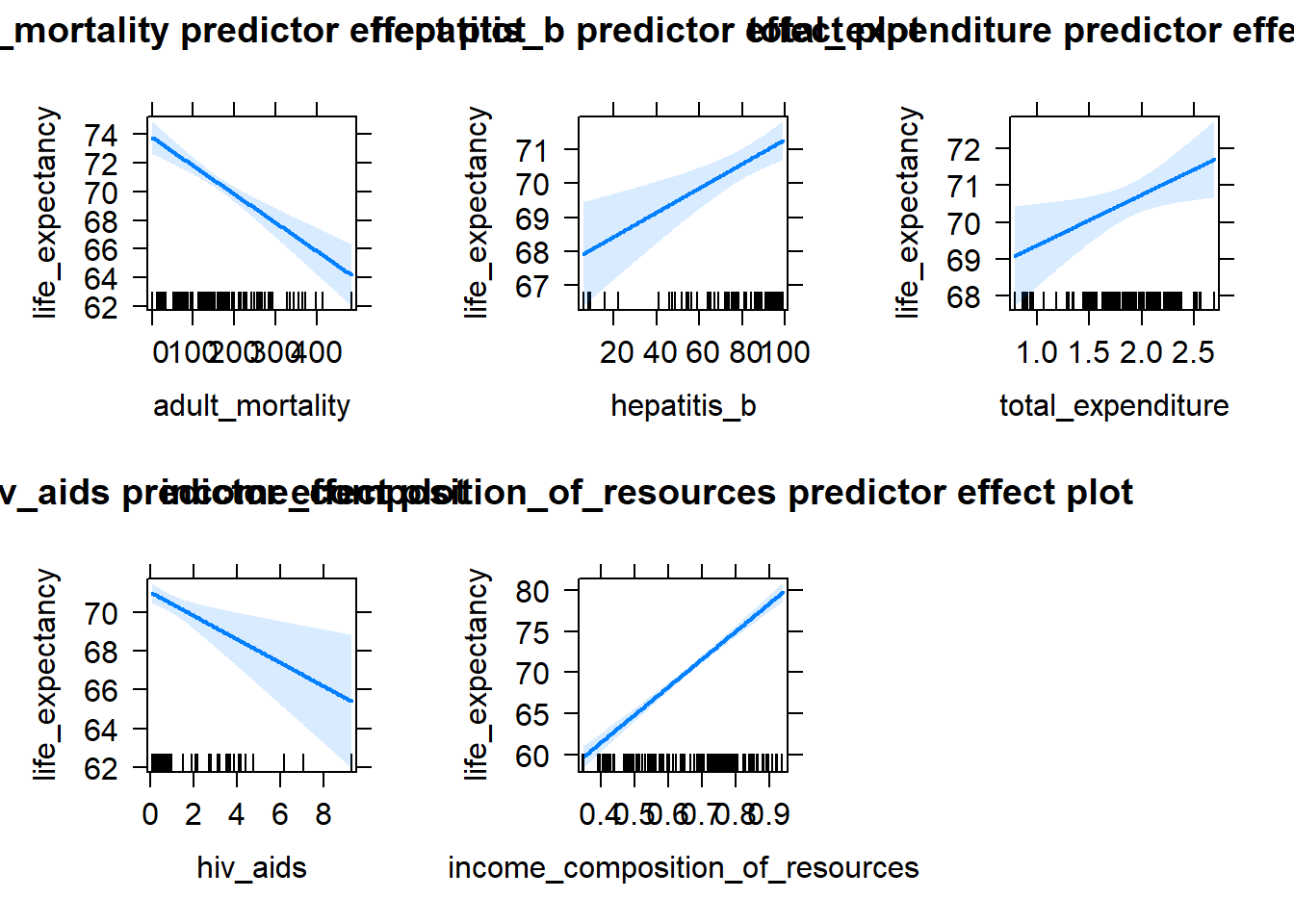
# Visų koeficientų interpretacija paprasta,  
 # nes pažinksnine regresija neišrinkti transformuoti kintamieji  
library(lm.beta)  
# Standartizuoti koeficientai  
lm.beta(model\_2)

##   
## Call:  
## lm(formula = life\_expectancy ~ adult\_mortality + hepatitis\_b +   
## total\_expenditure + hiv\_aids + income\_composition\_of\_resources,   
## data = x\_2)  
##   
## Standardized Coefficients::  
## (Intercept) adult\_mortality   
## 0.00000000 -0.24840840   
## hepatitis\_b total\_expenditure   
## 0.11222105 0.06927302   
## hiv\_aids income\_composition\_of\_resources   
## -0.11477877 0.64768318

# Pasikliovimo interalai  
confint(model\_2)

## 2.5 % 97.5 %  
## (Intercept) 42.29571386 49.73591902  
## adult\_mortality -0.02631364 -0.01333173  
## hepatitis\_b 0.01620110 0.05533575  
## total\_expenditure 0.21687917 2.55045417  
## hiv\_aids -1.00808384 -0.20800885  
## income\_composition\_of\_resources 30.08234193 37.79202074

# Kovariancių įtaka vizualizuota  
library(effects)  
plot(predictorEffects(model\_2))



Multikolinearumo tikrinimas

vars <- dplyr::select(x\_2, c(adult\_mortality, hepatitis\_b, total\_expenditure,  
 hiv\_aids, income\_composition\_of\_resources, life\_expectancy))  
  
#library(psych)  
#corr.test(vars)  
  
#dalinės koreliacijos  
library(ppcor)  
pcor(vars)$estimate

## adult\_mortality hepatitis\_b total\_expenditure  
## adult\_mortality 1.00000000 0.284752689 0.031114658  
## hepatitis\_b 0.28475269 1.000000000 -0.007076189  
## total\_expenditure 0.03111466 -0.007076189 1.000000000  
## hiv\_aids 0.30378653 -0.187990543 0.103610440  
## income\_composition\_of\_resources 0.18178399 -0.156298047 -0.086817301  
## life\_expectancy -0.47258053 0.305618694 0.203857631  
## hiv\_aids income\_composition\_of\_resources  
## adult\_mortality 0.3037865 0.1817840  
## hepatitis\_b -0.1879905 -0.1562980  
## total\_expenditure 0.1036104 -0.0868173  
## hiv\_aids 1.0000000 0.1721392  
## income\_composition\_of\_resources 0.1721392 1.0000000  
## life\_expectancy -0.2578685 0.8396372  
## life\_expectancy  
## adult\_mortality -0.4725805  
## hepatitis\_b 0.3056187  
## total\_expenditure 0.2038576  
## hiv\_aids -0.2578685  
## income\_composition\_of\_resources 0.8396372  
## life\_expectancy 1.0000000

# Variance inflation factor  
vif(model\_2)

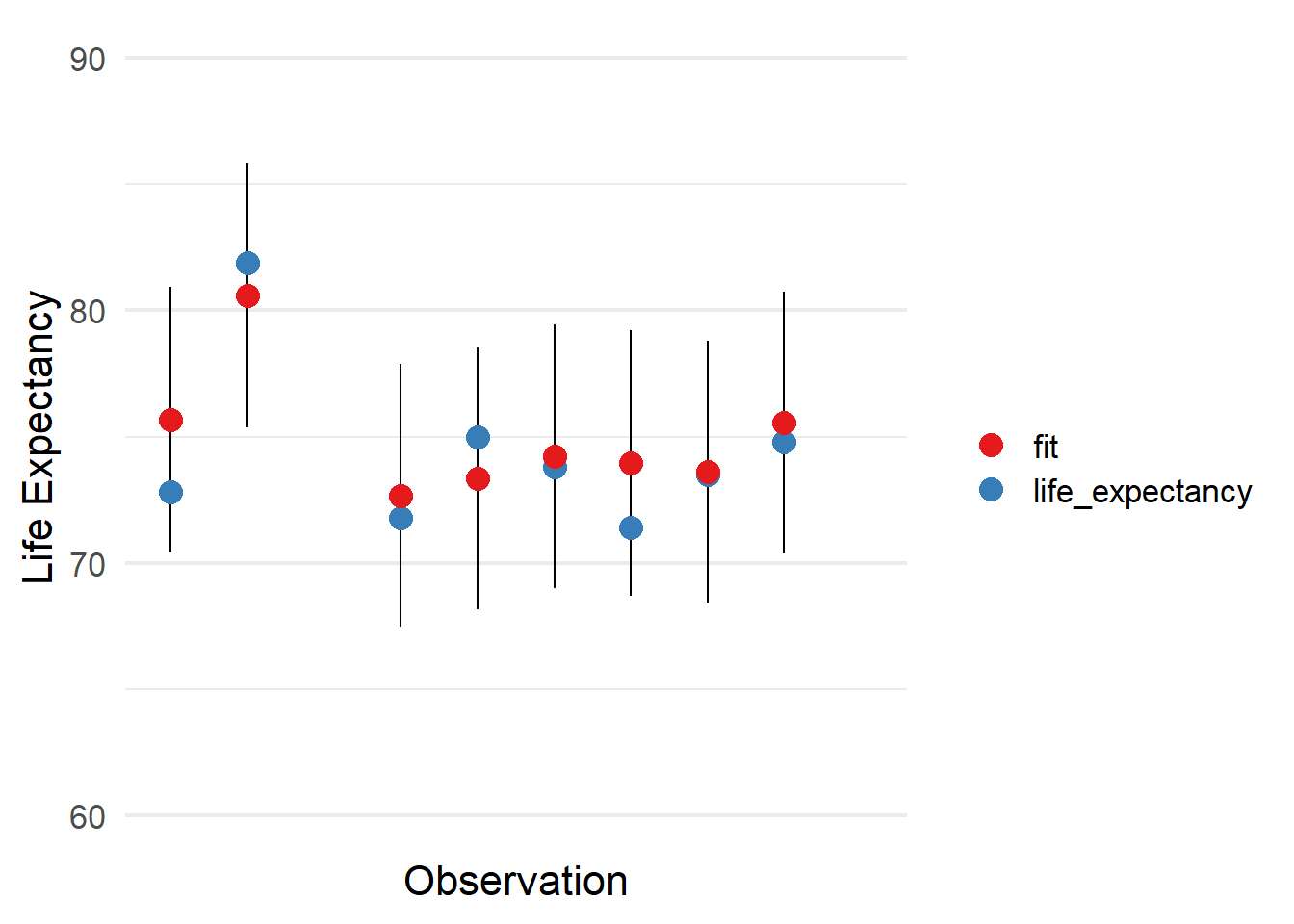
## adult\_mortality hepatitis\_b   
## 2.082698 1.186351   
## total\_expenditure hiv\_aids   
## 1.074114 1.794951   
## income\_composition\_of\_resources   
## 1.703679

Modelio tinkamumo analizė

summary(model\_2)

##   
## Call:  
## lm(formula = life\_expectancy ~ adult\_mortality + hepatitis\_b +   
## total\_expenditure + hiv\_aids + income\_composition\_of\_resources,   
## data = x\_2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.1512 -1.5507 0.2728 1.6248 8.3196   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 46.015816 1.879961 24.477 < 2e-16 \*\*\*  
## adult\_mortality -0.019823 0.003280 -6.043 1.56e-08 \*\*\*  
## hepatitis\_b 0.035768 0.009888 3.617 0.000428 \*\*\*  
## total\_expenditure 1.383667 0.589638 2.347 0.020491 \*   
## hiv\_aids -0.608046 0.202160 -3.008 0.003174 \*\*   
## income\_composition\_of\_resources 33.937181 1.948050 17.421 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.608 on 127 degrees of freedom  
## Multiple R-squared: 0.897, Adjusted R-squared: 0.8929   
## F-statistic: 221.1 on 5 and 127 DF, p-value: < 2.2e-16

# R-squared = 0.897  
 # Adj R-squared = 0.892  
  
plot\_predictions <- function(x,y) {  
 predictions <- predict(x,newdata = y, interval = "prediction")  
 predictions <- as\_tibble(predictions) %>% mutate(n = 1:nrow(predictions))  
  
   
 predictions\_points <- y %>%  
 mutate(pred = predictions) %>%   
 unnest(pred) %>%  
 dplyr::select(1,last\_col(3),last\_col(2),last\_col(1),last\_col(0)) %>%  
 pivot\_longer(c(1,2))  
   
  
 ggplot(predictions) +   
 geom\_linerange(aes(x=n,ymin=lwr,ymax=upr)) +   
 geom\_point(data=predictions\_points,aes(x=n,y=value,color=name),size = 4) +   
 scale\_x\_discrete("Observation") +  
 scale\_y\_continuous("Life Expectancy",limits = c(60,90)) +   
 theme\_minimal(base\_size = 16) +   
 scale\_color\_brewer("",palette = "Set1")   
}  
  
# Atliekamos kelios pavyzdinės prognozės  
plot\_predictions(model\_2,x\_predict)



Rezultatai

Siekiant ištirti gyvenimo trukmės ryšį su sveikata susijusiais kriterijais naudota daugelio kintamųjų tiesinė regresija.

Pažinksnine regresija išrinktas modelis paaiškina 89.7% duomenų sklaidos (*F(5,127) = 221.1, R2 = 0.897, p<0.01).*

Rastos 5 statistiškai reikšmingos kovariantės gyvenimo trukmės prognozavimui (pateikti standartizuoti krypties koeficientai):

Suaugusių mirtingumas (tikimybė mirti tarp 15 ir 60 metų 1000 gyventojų) (stulp. adult\_*mortality β=-0.25, p<0.001*)

Imunizacija nuo Hepatito B tarp 1 metų vaikų % (stulp. hepatitis\_b *β=0.11, p<0.001*)

Dalis visų vyriausybės išlaidų sveikatos apsaugai (stulp. total\_expenditure *β=0.07, p=0.02*)

Mirtys nuo ŽIV/AIDS nuo 0 iki 4 metų 1000 gimimų (stulp. hiv\_aids *β=-0.11, p=0.003*)

HDI pagal pajamų parametrą (stulp. income\_composition\_of\_resources *β=0.65, p<0.001*)

**2. Naudojant SAS**

Naudojamas anksčiau sukurtas duomenų failas.

PROC IMPORT DATAFILE='/home/u45871880/life\_modified.csv'

DBMS=CSV

OUT=data;

GETNAMES=YES;

RUN;

/\* Modelio prielaidos \*/

PROC REG data=data simple corr plots=(diagnostics(stats=none) RStudentByLeverage(label)

CooksD(label) Residuals(smooth) ObservedByPredicted(label));

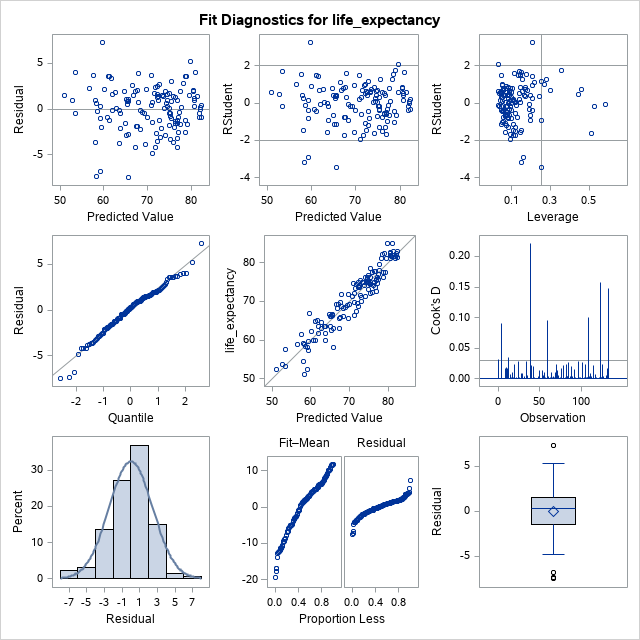
MODEL life\_expectancy = adult\_mortality infant\_deaths alcohol hepatitis\_b measles

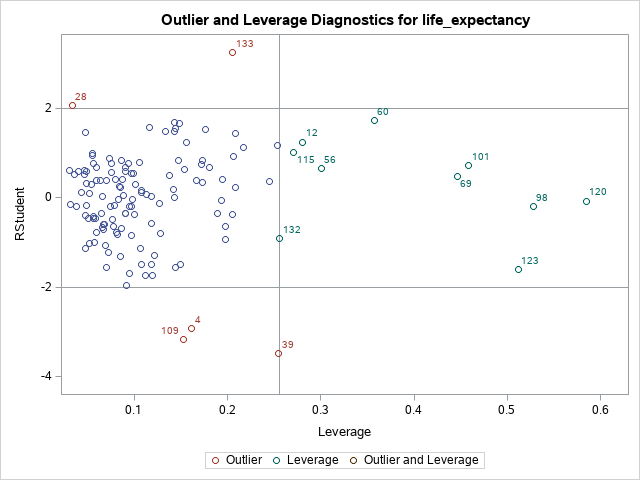
bmi under\_five\_deaths polio total\_expenditure diphtheria hiv\_aids

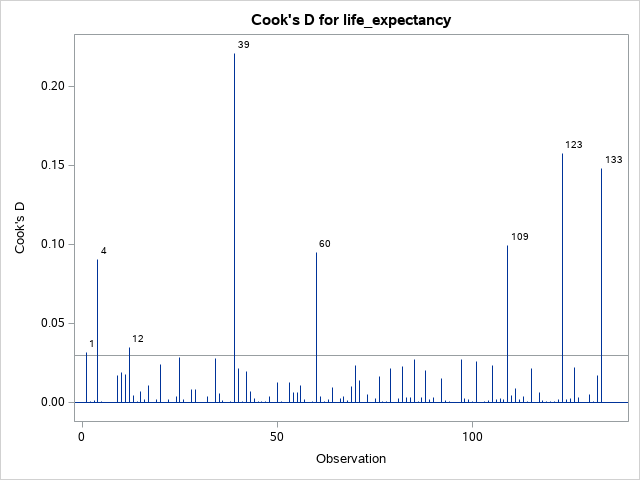
thinness\_1\_19\_years thinness\_5\_9\_years income\_composition\_of\_resources

schooling gdp\_per\_capita;

run;







/\* Normalumo testas \*/

proc univariate data=rez normal;

var liekanos;

run;

| **Tests for Normality** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Shapiro-Wilk** | **W** | 0.981952 | **Pr < W** | 0.0749 |
| **Kolmogorov-Smirnov** | **D** | 0.060241 | **Pr > D** | >0.1500 |
| **Cramer-von Mises** | **W-Sq** | 0.100101 | **Pr > W-Sq** | 0.1135 |
| **Anderson-Darling** | **A-Sq** | 0.63253 | **Pr > A-Sq** | 0.0979 |

/\* Modelio parinkimas naudojant pažinksninę regresiją\*/

/\* Parametrų vertinimas \*/

PROC REG data=data plots=none outest=summary;

MODEL life\_expectancy = adult\_mortality infant\_deaths alcohol hepatitis\_b measles

bmi under\_five\_deaths polio total\_expenditure diphtheria hiv\_aids

thinness\_1\_19\_years thinness\_5\_9\_years income\_composition\_of\_resources

schooling gdp\_per\_capita / stb vif cli clb pcorr2 slentry=0.05 slstay=0.05 selection=stepwise aic bic;

run;

proc print data=summary;

run;

**Stepwise Selection: Step 5**

**Variable total\_expenditure Entered: R-Square = 0.8970 and C(p) = 4.1881**

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 5 | 7520.59056 | 1504.11811 | 221.12 | <.0001 |
| **Error** | 127 | 863.90673 | 6.80242 |  |  |
| **Corrected Total** | 132 | 8384.49729 |  |  |  |

| **Variable** | **Parameter Estimate** | **Standard Error** | **Type II SS** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Intercept** | 46.01582 | 1.87996 | 4075.49110 | 599.12 | <.0001 |
| **adult\_mortality** | -0.01982 | 0.00328 | 248.41814 | 36.52 | <.0001 |
| **hepatitis\_b** | 0.03577 | 0.00989 | 89.00457 | 13.08 | 0.0004 |
| **total\_expenditure** | 1.38367 | 0.58964 | 37.45889 | 5.51 | 0.0205 |
| **hiv\_aids** | -0.60805 | 0.20216 | 61.53858 | 9.05 | 0.0032 |
| **income\_composition\_of\_resources** | 33.93718 | 1.94805 | 2064.49803 | 303.49 | <.0001 |

**Bounds on condition number: 2.0827, 39.209**

**All variables left in the model are significant at the 0.0500 level.**

**No other variable met the 0.0500 significance level for entry into the model.**

| **Summary of Stepwise Selection** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Step** | **Variable Entered** | **Variable Removed** | **Number Vars In** | **Partial R-Square** | **Model R-Square** | **C(p)** | **F Value** | **Pr > F** |
| **1** | income\_composition\_of\_resources |  | 1 | 0.8052 | 0.8052 | 107.730 | 541.34 | <.0001 |
| **2** | adult\_mortality |  | 2 | 0.0619 | 0.8671 | 34.4953 | 60.56 | <.0001 |
| **3** | hepatitis\_b |  | 3 | 0.0187 | 0.8857 | 13.8226 | 21.07 | <.0001 |
| **4** | hiv\_aids |  | 4 | 0.0068 | 0.8925 | 7.6163 | 8.04 | 0.0053 |
| **5** | total\_expenditure |  | 5 | 0.0045 | 0.8970 | 4.1881 | 5.51 | 0.0205 |

3**. Naudojant Python**

**import** warnings

warnings**.**filterwarnings("ignore")

**import** pandas **as** pd

**import** seaborn **as** sns

**import** numpy **as** np

**import** matplotlib.pyplot **as** plt

**from** scipy.stats **import** levene

**import** statsmodels.api **as** sm

**from** statsmodels.formula.api **import** ols

**import** pylab

**import** scipy.stats **as** stats

**from** bioinfokit.analys **import** stat

**def** split(df, col):

**return** [

df[df["handedness"] **==** "left"][df["sex"] **==** "female"][col],

df[df["handedness"] **==** "left"][df["sex"] **==** "male"][col],

df[df["handedness"] **==** "right"][df["sex"] **==** "female"][col],

df[df["handedness"] **==** "right"][df["sex"] **==** "male"][col]

]

**def** vartest(df,col):

s **=** split(df,col)

stat, p **=** levene(s[0],s[1],s[2],s[3])

print("F value:", round(stat,4), "Pr(>F)", round(p,4))

**def** anova(df, col):

stats**.**probplot(df[col], dist**=**"norm", plot**=**pylab)

pylab**.**show()

model **=** ols(col **+** ' ~ sex \* handedness', data**=**df)**.**fit()

anova\_table **=** sm**.**stats**.**anova\_lm(model, typ**=**3)

**return** anova\_table

data **=** pd**.**read\_csv("data.csv")

data **=** data**.**sort\_values(["sex", "handedness"])

mypal **=** {sex: '#f9665e' **if** sex **==** "female" **else** '#799fcb' **for** sex **in** data["sex"]**.**unique()}

ft **=** data

ft["group"] **=** ft["handedness"] **+** ft["sex"]

fig, axes **=** plt**.**subplots(2, 2,figsize**=**(15,12))

fig**.**suptitle("Tiriamieji grafikai")

sns**.**boxplot(ax **=** axes[0,0],x**=**"handedness", y**=**"t1", hue**=**"sex", data**=**data, palette**=**mypal)

sns**.**boxplot(ax **=** axes[0,1],x**=**"handedness", y**=**"t2", hue**=**"sex", data**=**data, palette**=**mypal)

sns**.**boxplot(ax **=** axes[1,0],x**=**"handedness", y**=**"t3", hue**=**"sex", data**=**data, palette**=**mypal)

sns**.**boxplot(ax **=** axes[1,1],x**=**"handedness", y**=**"t4", hue**=**"sex", data**=**data, palette**=**mypal)

Paveikslėlis, kuriame yra žinutė, kryžiažodis, pirmosios pagalbos rinkinys, rezultatų lentelė

Automatiškai sugeneruotas aprašymas

means **=** data**.**groupby(['sex','handedness'])**.**mean()

fig, axes **=** plt**.**subplots(2, 2,figsize**=**(15,12))

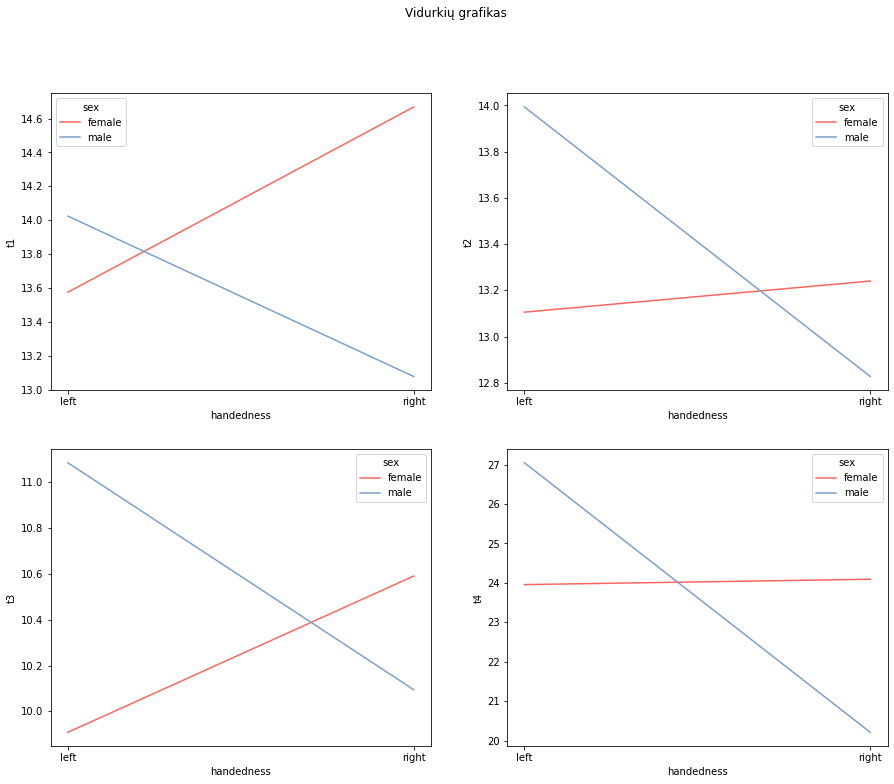
fig**.**suptitle("Vidurkių grafikas")

sns**.**lineplot(ax **=** axes[0,0],x**=**"handedness", y**=**"t1", hue**=**"sex", data**=**means, palette**=**mypal)

sns**.**lineplot(ax **=** axes[0,1],x**=**"handedness", y**=**"t2", hue**=**"sex", data**=**means, palette**=**mypal)

sns**.**lineplot(ax **=** axes[1,0],x**=**"handedness", y**=**"t3", hue**=**"sex", data**=**means, palette**=**mypal)

sns**.**lineplot(ax **=** axes[1,1],x**=**"handedness", y**=**"t4", hue**=**"sex", data**=**means, palette**=**mypal)



vartest(data,"t1")

F value: 0.8848 Pr(>F) 0.4572

vartest(data,"t2")

F value: 0.4793 Pr(>F) 0.6985

vartest(data,"t3")

F value: 1.7389 Pr(>F) 0.1745

vartest(data,"t4")

F value: 1.8124 Pr(>F) 0.1604

sum\_sq df F PR(>F)

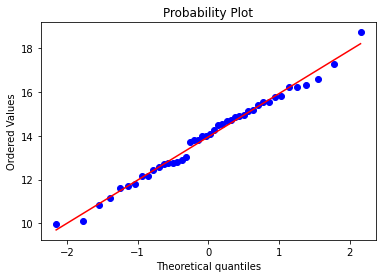
Intercept 1842.757078 1.0 507.629040 2.428428e-24

sex 1.102086 1.0 0.303594 5.847024e-01

handedness 6.979565 1.0 1.922679 1.732417e-01

sex:handedness 10.967304 1.0 3.021191 8.987678e-02

Residual 145.205016 40.0 NaN NaN



sum\_sq df F PR(>F)

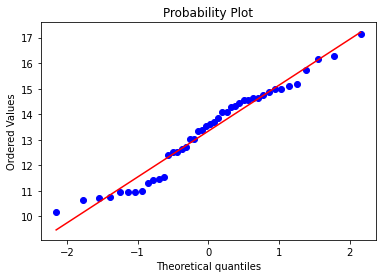
Intercept 1717.564555 1.0 538.700233 8.037321e-25

sex 4.301995 1.0 1.349286 2.522877e-01

handedness 0.105805 1.0 0.033185 8.563713e-01

sex:handedness 4.459198 1.0 1.398591 2.439417e-01

Residual 127.533975 40.0 NaN NaN



sum\_sq df F PR(>F)

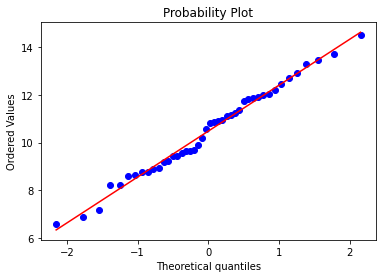
Intercept 981.812039 1.0 276.846236 1.413343e-19

sex 7.541214 1.0 2.126432 1.525851e-01

handedness 2.710706 1.0 0.764351 3.871895e-01

sex:handedness 7.358940 1.0 2.075035 1.575086e-01

Residual 141.856657 40.0 NaN NaN



sum\_sq df F PR(>F)

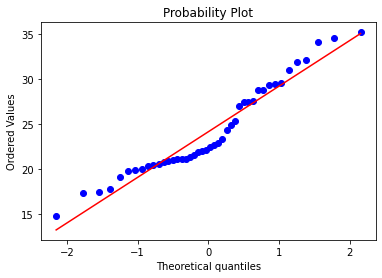
Intercept 5738.425909 1.0 261.107038 3.929689e-19

sex 52.051729 1.0 2.368432 1.316855e-01

handedness 0.108554 1.0 0.004939 9.443204e-01

sex:handedness 127.945606 1.0 5.821718 2.050468e-02

Residual 879.091725 40.0 NaN NaN



res = stat()

res.tukey\_hsd(df=ft, res\_var='t4', xfac\_var='group', anova\_model='t4 ~ group')

res.tukey\_summary

group1 group2 Diff Lower Upper q-value p-value

0 leftfemale rightfemale 0.136416 -5.066654 5.339486 0.099392 0.900000

1 leftfemale leftmale 3.089145 -2.291556 8.469846 2.176434 0.425942

2 leftfemale rightmale 3.745010 -2.215856 9.705877 2.381714 0.345613

3 rightfemale leftmale 2.952729 -1.990949 7.896407 2.264224 0.390736

4 rightfemale rightmale 3.881426 -1.688128 9.450981 2.641903 0.257923

5 leftmale rightmale 6.834156 1.098309 12.570002 4.516826 0.013984