SAP - Projektni zadatak

Case study *Analiza preferencija mladih ljudi*: Deskriptivna statistika, vizualizacija podataka, statističko zaključivanje i linarna regresija

Domagoj Marinello, Sven Skender, Ana Skukan, Matea Vasilj

1/13/2022

Case study: *Analiza preferencija mladih ljudi*

Interesi mladih ljudi (glazba, filmovi itd.), zdrave navike, odabir načina života i obrasci potrošnje vrlo su važni za različite industrije, kao i donositelje demografskih, poreznih ili mirovinskih politika svake zemlje. Upravo su te teme bile predmet istraživanja provedenog u Slovačkoj nad mladim osobama izmedu 15. i 30. godine života.

Shodno tome, obradili smo slijedeća istraživačka pitanja koja će biti detaljizirana u nastavku ove bilježnice:

1. Razlikuju li se izraženi strahovi ispitanih žena i muškaraca?
2. Možemo li predvidjeti obrazac potrošnje ovisno o žanru glazbe kojeg ispitanik preferira?
3. Možemo li temeljem danih varijabli predvidjeti dob ispitanika?
4. Kako su kategorije o ljudskom ponašanju povezane sa brojem prijatelja?

Podaci za analizu sadržani su u datoteci opis pitanja.csv koja se sastoji od informacija prikupljenih upitnikom koji je prezentiran gore navedenoj skupini ljudi. Podatci se sastoje od osobnih informacija ispitanika koje, medu ostalim, ukljucuju dob, visinu, tezinu i spol. Preferencije ispitanika prikupljene su kroz kategorije glazbenog i filmskog ukusa, hobija, strahova, (ne)zdravih navika, osobina licnosti i drugih. Skup podataka sastoji od vise pitanja koja pruzaju numericke podatke, primjerice ocjenu preferencije ispitanika na skali od 1 do 5 te od kategorijskih podataka, primjerice, spol.

Prije pregleda odgovora, pogledajmo postavljena pitanja s pojašnjenjima koja se nalaze u datoteci odgovori.csv:

```
odgovori = read.csv("./data/odgovori.csv")
odgovori
```

```
##                                                                            original
## 1                                                       I enjoy listening to music.
## 2                                                                         I prefer.
## 3                                                                Dance, Disco, Funk
## 4                                                                        Folk music
## 5                                                                           Country
## 6                                                                         Classical
## 7                                                                          Musicals
## 8                                                                               Pop
## 9                                                                              Rock
## 10                                                                 Metal, Hard rock
## 11                                                                             Punk
## 12                                                                     Hip hop, Rap
## 13                                                                      Reggae, Ska
## 14                                                                      Swing, Jazz
## 15                                                                      Rock n Roll
## 16                                                                Alternative music
## 17                                                                            Latin
## 18                                                                   Techno, Trance
## 19                                                                            Opera
## 20                                                  I really enjoy watching movies.
## 21                                                                    Horror movies
## 22                                                                  Thriller movies
## 23                                                                         Comedies
## 24                                                                  Romantic movies
## 25                                                                    Sci-fi movies
## 26                                                                       War movies
## 27                                                                            Tales
## 28                                                                         Cartoons
## 29                                                                    Documentaries
## 30                                                                   Western movies
## 31                                                                    Action movies
## 32                                                                          History
## 33                                                                       Psychology
## 34                                                                         Politics
## 35                                                                      Mathematics
## 36                                                                          Physics
## 37                                                                         Internet
## 38                                                            PC Software, Hardware
## 39                                                              Economy, Management
## 40                                                                          Biology
## 41                                                                        Chemistry
## 42                                                                   Poetry reading
## 43                                                                        Geography
## 44                                                                Foreign languages
## 45                                                                         Medicine
## 46                                                                              Law
## 47                                                                             Cars
## 48                                                                              Art
## 49                                                                         Religion
## 50                                                               Outdoor activities
## 51                                                                          Dancing
## 52                                                      Playing musical instruments
## 53                                                                   Poetry writing
## 54                                                     Sport and leisure activities
## 55                                                       Sport at competitive level
## 56                                                                        Gardening
## 57                                                              Celebrity lifestyle
## 58                                                                         Shopping
## 59                                                           Science and technology
## 60                                                                          Theatre
## 61                                                                      Socializing
## 62                                                                Adrenaline sports
## 63                                                                             Pets
## 64                                                                           Flying
## 65                                                               Thunder, lightning
## 66                                                                         Darkness
## 67                                                                          Heights
## 68                                                                          Spiders
## 69                                                                           Snakes
## 70                                                                      Rats, mice
## 71                                                                          Ageing
## 72                                                                Dangerous dogs
## 73                                                               Public speaking
## 74                                                                Smoking habits
## 75                                                                      Drinking
## 76                                               I live a very healthy lifestyle.
## 77                                      I take notice of what goes on around me.
## 78     I try to do tasks as soon as possible and not leave them until last minute.
## 79                                    I always make a list so I don't forget anything.
## 80                                        I often study or work even in my spare time.
## 81                       I look at things from all different angles before I go ahead.
## 82  I believe that bad people will suffer one day and good people will be rewarded.
## 83                    I am reliable at work and always complete all tasks given to me.
## 84                                                      I always keep my promises.
## 85          I can fall for someone very quickly and then completely lose interest.
## 86                            I would rather have lots of friends than lots of money.
## 87                                              I always try to be the funniest one.
## 88                                                     I can be two faced sometimes.
## 89                                        I damaged things in the past when angry.
## 90                                               I take my time to make decisions.
## 91                                              I always try to vote in elections.
## 92                                 I often think about and regret the decisions I make.
## 93                       I can tell if people listen to me or not when I talk to them.
## 94                                                          I am a hypochondriac.
## 95                                                        I am emphatetic person.
## 96           I eat because I have to. I don't enjoy food and eat as fast as I can.
## 97                       I try to give as much as I can to other people at Christmas.
## 98                                          I don't like seeing animals suffering.
## 99                                  I look after things I have borrowed from others.
## 100                                                         I feel lonely in life.
## 101                                                    I used to cheat at school.
## 102                                                     I worry about my health.
## 103               I wish I could change the past because of the things I have done.
## 104                                                              I believe in God.
## 105                                                    I always have good dreams.
## 106                                                   I always give to charity.
## 107                                                     I have lots of friends.
## 108                                                               Timekeeping.
## 109                                                        Do you lie to others?
## 110                                                           I am very patient.
## 111                                       I can quickly adapt to a new environment.
## 112                                                     My moods change quickly.
## 113                                I am well mannered and I look after my appearance.
## 114                                                    I enjoy meeting new people.
## 115                             I always let other people know about my achievements.
## 116                          I think carefully before answering any important letters.
## 117                                                   I enjoy childrens' company.
## 118         I am not afraid to give my opinion if I feel strongly about something.
## 119                                                   I can get angry very easily.
## 120                               I always make sure I connect with the right people.
## 121                                I have to be well prepared before public speaking.
## 122                            I will find a fault in myself if people don't like me.
## 123                             I cry when I feel down or things don't go the right way.
## 124                                                   I am 100% happy with my life.
## 125                                            I am always full of life and energy.
## 126                            I prefer big dangerous dogs to smaller, calmer dogs.
## 127                               I believe all my personality traits are positive.
## 128                 If I find something the doesn't belong to me I will hand it in.
## 129                              I find it very difficult to get up in the morning.
## 130                                    I have many different hobbies and interests.
## 131                                          I always listen to my parents' advice.
## 132                                                 I enjoy taking part in surveys.
## 133                                              How much time do you spend online?
## 134                                                     I save all the money I can.
## 135                                        I enjoy going to large shopping centres.
## 136                                       I prefer branded clothing to non branded.
## 137                            I spend a lot of money on  partying and socializing.
## 138                                        I spend a lot of money on my appearance.
## 139                                              I spend a lot of money on gadgets.
## 140                I will hapilly pay more money for good, quality or healthy food.
## 141                                                                             Age
## 142                                                                          Height
## 143                                                                          Weight
## 144                                                  How many siblings do you have?
## 145                                                                          Gender
## 146                                                                            I am
## 147                                                      Highest education achieved
## 148                                                             I am the only child
## 149                                               I spent most of my childhood in a
## 150                                               I lived most of my childhood in a
##                              short
## 1                            Music
## 2         Slow songs or fast songs
## 3                            Dance
## 4                             Folk
## 5                          Country
## 6                  Classical music
## 7                          Musical
## 8                              Pop
## 9                             Rock
## 10               Metal or Hardrock
## 11                            Punk
## 12                     Hiphop, Rap
## 13                     Reggae, Ska
## 14                     Swing, Jazz
## 15                     Rock n roll
## 16                     Alternative
## 17                          Latino
## 18                  Techno, Trance
## 19                           Opera
## 20                          Movies
## 21                          Horror
## 22                        Thriller
## 23                          Comedy
## 24                        Romantic
## 25                          Sci-fi
## 26                             War
## 27             Fantasy/Fairy tales
## 28                        Animated
## 29                     Documentary
## 30                         Western
## 31                          Action
## 32                         History
## 33                      Psychology
## 34                        Politics
## 35                     Mathematics
## 36                         Physics
## 37                        Internet
## 38                              PC
## 39              Economy Management
## 40                         Biology
## 41                       Chemistry
## 42                         Reading
## 43                       Geography
## 44               Foreign languages
## 45                        Medicine
## 46                             Law
## 47                            Cars
## 48                 Art exhibitions
## 49                        Religion
## 50           Countryside, outdoors
## 51                         Dancing
## 52             Musical instruments
## 53                         Writing
## 54                   Passive sport
## 55                    Active sport
## 56                       Gardening
## 57                     Celebrities
## 58                        Shopping
## 59          Science and technology
## 60                         Theatre
## 61                Fun with friends
## 62               Adrenaline sports
## 63                            Pets
## 64                          Flying
## 65                           Storm
## 66                        Darkness
## 67                         Heights
## 68                         Spiders
## 69                          Snakes
## 70                            Rats
## 71                          Ageing
## 72                  Dangerous dogs
## 73         Fear of public speaking
## 74                         Smoking
## 75                         Alcohol
## 76                  Healthy eating
## 77                    Daily events
## 78           Prioritising workload
## 79                   Writing notes
## 80                     Workaholism
## 81                 Thinking ahead
## 82                Final judgement
## 83                    Reliability
## 84               Keeping promises
## 85               Loss of interest
## 86           Friends versus money
## 87                      Funniness
## 88                           Fake
## 89                Criminal damage
## 90                Decision making
## 91                      Elections
## 92                 Self-criticism
## 93                 Judgment calls
## 94                   Hypochondria
## 95                        Empathy
## 96              Eating to survive
## 97                         Giving
## 98          Compassion to animals
## 99                 Borrowed stuff
## 100                    Loneliness
## 101            Cheating in school
## 102                        Health
## 103             Changing the past
## 104                           God
## 105                        Dreams
## 106                       Charity
## 107             Number of friends
## 108                   Punctuality
## 109                         Lying
## 110                       Waiting
## 111               New environment
## 112                   Mood swings
## 113       Appearence and gestures
## 114                   Socializing
## 115                  Achievements
## 116 Responding to a serious letter
## 117                      Children
## 118                 Assertiveness
## 119                 Getting angry
## 120      Knowing the right people
## 121               Public speaking
## 122                  Unpopularity
## 123                Life struggles
## 124             Happiness in life
## 125                 Energy levels
## 126              Small - big dogs
## 127                   Personality
## 128        Finding lost valuables
## 129                    Getting up
## 130          Interests or hobbies
## 131               Parents' advice
## 132       Questionnaires or polls
## 133                Internet usage
## 134                      Finances
## 135               Shopping centres
## 136               Branded clothing
## 137         Entertainment spending
## 138              Spending on looks
## 139            Spending on gadgets
## 140     Spending on healthy eating
## 141                            Age
## 142                         Height
## 143                         Weight
## 144             Number of siblings
## 145                         Gender
## 146            Left - right handed
## 147                      Education
## 148                     Only child
## 149                 Village - town
## 150         House - block of flats
```

Sada, kada smo saznali o kakvim se pitanjima radi, možemo pogledati kako su ona kodirana, kakve smo odgovore uspjeli prikupiti i koliko ih je uopće:

```
pitanja = read.csv("./data/opis pitanja.csv")
head(pitanja)
```

```
##   Music Slow.songs.or.fast.songs Dance Folk Country Classical.music Musical Pop
## 1     5                        3     2    1       2               2       1   5
## 2     4                        4     2    1       1               1       2   3
## 3     5                        5     2    2       3               4       5   3
## 4     5                        3     2    1       1               1       1   2
## 5     5                        3     4    3       2               4       3   5
## 6     5                        3     2    3       2               3       3   2
##   Rock Metal.or.Hardrock Punk Hiphop..Rap Reggae..Ska Swing..Jazz Rock.n.roll
## 1    5                 1    1           1           1           1           3
## 2    5                 4    4           1           3           1           4
## 3    5                 3    4           1           4           3           5
## 4    2                 1    4           2           2           1           2
## 5    3                 1    2           5           3           2           1
## 6    5                 5    3           4           3           4           4
##   Alternative Latino Techno..Trance Opera Movies Horror Thriller Comedy
## 1           1      1              1     1      5      4        2      5
## 2           4      2              1     1      5      2        2      4
## 3           5      5              1     3      5      3        4      4
## 4           5      1              2     1      5      4        4      3
## 5           2      4              2     2      5      4        4      5
## 6           5      3              1     3      5      5        5      5
##   Romantic Sci.fi War Fantasy.Fairy.tales Animated Documentary Western Action
## 1        4      4   1                   5        5           3       1      2
## 2        3      4   1                   3        5           4       1      4
## 3        2      4   2                   5        5           2       2      1
## 4        3      4   3                   1        2           5       1      2
## 5        2      3   3                   4        4           3       1      4
## 6        2      3   3                   4        3           3       2      4
##   History Psychology Politics Mathematics Physics Internet PC
## 1       1          5        1           3       3        5  3
## 2       1          3        4           5       2        4  4
## 3       1          2        1           5       2        4  2
## 4       4          4        5           4        1        3  1
## 5       3          2        3           2        2        2  2
## 6       5          3        4           2        3        4  4
##   Economy.Management Biology Chemistry Reading Geography Foreign.languages
## 1                  5       3         3       3         3                 5
## 2                  5       1         1       4         4                 5
## 3                  4       1         1       5         2                 5
## 4                  2       3         3       5         4                 4
## 5                  2       3         3       5         2                 3
## 6                  1       4         4       3         3                 4
##   Medicine Law Cars Art.exhibitions Religion Countryside..outdoors Dancing
## 1        3   1    1               1        1                     5       3
## 2        1   2    2               2        1                     1       1
## 3        2   3    1               5        5                     5       5
## 4        2   5    1               5        4                     1       1
## 5        3   2    3               1        4                     4       1
## 6        4   3    5               2        2                     5       1
##   Musical.instruments Writing Passive.sport Active.sport Gardening Celebrities
## 1                   3       2             1            5         5           1
## 2                   1       1             1            1         1           2
## 3                   5       5             5            2         1           1
## 4                   1       3             1            1         1           2
## 5                   3       1             3            1         4           3
## 6                   5       1             5            4         2           1
##   Shopping Science.and.technology Theatre Fun.with.friends Adrenaline.sports
## 1        4                      4       2                5                 4
## 2        3                      3       2                4                 2
## 3        4                      2       5                5                 5
## 4        4                      3       1                2                 1
## 5        3                      3       2                4                 2
## 6        2                      3       1                3                 3
##   Pets Flying Storm Darkness Heights Spiders Snakes Rats Ageing Dangerous.dogs
## 1    4      1     1        1       1       1      5    3      1              3
## 2    5      1     1        1       2       1      1    1      3              1
## 3    5      1     1        1       1       1      1    1      1              1
## 4    1      2     1        1       3       5      5    5      4              5
## 5    1      1     2        1       1       1      1    2      2              4
## 6    2      3     2        2       2       1      2    2      1              1
##   Fear.of.public.speaking       Smoking        Alcohol Healthy.eating
## 1                       2  never smoked    drink a lot              4
## 2                       4  never smoked    drink a lot              3
## 3                       2 tried smoking    drink a lot              3
## 4                       5 former smoker    drink a lot              3
## 5                       3 tried smoking social drinker              4
## 6                       3  never smoked          never              2
##   Daily.events Prioritising.workload Writing.notes Workaholism Thinking.ahead
## 1            2                     2             5           4              2
## 2            3                     2             4           5              4
## 3            1                     2             5           3              5
## 4            4                     4             4           5              3
## 5            3                     1             2           3              5
## 6            2                     2             3           3              3
##   Final.judgement Reliability Keeping.promises Loss.of.interest
## 1               5           4                4                1
## 2               1           4                4                3
## 3               3           4                5                1
## 4               1           3                4                5
## 5               5           5                4                2
## 6               1           3                4                3
##   Friends.versus.money Funniness Fake Criminal.damage Decision.making Elections
## 1                    3         5    1               1               3         4
## 2                    4         3    2               1               2         5
## 3                    5         2    4               1               3         5
## 4                    2         1    1               5               5         5
## 5                    3         3    2               1               3         5
## 6                    2         3    1               4               2         5
##   Self.criticism Judgment.calls Hypochondria Empathy Eating.to.survive Giving
## 1              1              3            1       3                 1      4
## 2              4              4            1       2                 1      2
## 3              4              4            1       5                 5      5
## 4              5              4            3       3                 1      1
## 5              5              5            1       3                 1      3
## 6              4              4            1       4                 2      3
##   Compassion.to.animals Borrowed.stuff Loneliness Cheating.in.school Health
## 1                     5              4          3                  2      1
## 2                     4              3          2                  4      4
## 3                     4              2          5                  3      2
## 4                     2              5          5                  5      1
## 5                     3              4          3                  5      3
## 6                     5              5          2                  4      3
##   Changing.the.past God Dreams Charity Number.of.friends
## 1                 1   1      4       2                 3
## 2                 4   1      3       1                 3
## 3                 5   5      1       3                 3
## 4                 5   4      3       3                 1
## 5                 4   5      3       3                 3
## 6                 3   3      3       2                 3
##              Punctuality                         Lying Waiting New.environment
## 1    i am always on time                         never       3               4
## 2       i am often early                     sometimes       3               4
## 3 i am often running late                     sometimes       2               3
## 4       i am often early only to avoid hurting someone       1               1
## 5    i am always on time         everytime it suits me       3               4
## 6       i am often early only to avoid hurting someone       3               4
##   Mood.swings Appearence.and.gestures Socializing Achievements
## 1           3                       4           3            4
## 2           4                       4           4            2
## 3           4                       3           5            3
## 4           5                       3           1            3
## 5           2                       3           3            3
## 6           3                       3           4            2
##   Responding.to.a.serious.letter Children Assertiveness Getting.angry
## 1                              3        5             1             1
## 2                              4        2             2             5
## 3                              4        4             3             4
## 4                              3        2             5             5
## 5                              3        5             4             2
## 6                              2        3             4             3
##   Knowing.the.right.people Public.speaking Unpopularity Life.struggles
## 1                        3               5            5              1
## 2                        4               4            4              1
## 3                        3               2            4              4
## 4                        4               5            3              3
## 5                        3               5            5              2
## 6                        4               4            4              3
##   Happiness.in.life Energy.levels Small...big.dogs Personality
## 1                 4             5                1           4
## 2                 4             3                5           3
## 3                 4             4                3           3
## 4                 2             2                1           2
## 5                 3             5                3           3
## 6                 3             4                4           3
##   Finding.lost.valuables Getting.up Interests.or.hobbies Parents..advice
## 1                      3          2                    3               4
## 2                      4          5                    3               2
## 3                      3          4                    5               3
## 4                      1          1                   NA               2
## 5                      2          4                    3               3
## 6                      3          3                    5               3
##   Questionnaires.or.polls  Internet.usage Finances Shopping.centres
## 1                       3 few hours a day        3                4
## 2                       3 few hours a day        3                4
## 3                       1 few hours a day        2                4
## 4                       4 most of the day        2                4
## 5                       3 few hours a day        4                3
## 6                       4 few hours a day        2                3
##   Branded.clothing Entertainment.spending Spending.on.looks Spending.on.gadgets
## 1                5                      3                 3                   1
## 2                1                      4                 2                   5
## 3                1                      4                 3                   4
## 4                3                      3                 4                   4
## 5                4                      3                 3                   2
## 6                3                      3                 1                   4
##   Spending.on.healthy.eating Age Height Weight Number.of.siblings Gender
## 1                          3  20    163     48                  1 female
## 2                          2  19    163     58                  2 female
## 3                          2  20    176     67                  2 female
## 4                          1  22    172     59                  1 female
## 5                          4  20    170     59                  1 female
## 6                          4  20    186     77                  1   male
##   Left...right.handed               Education Only.child Village...town
## 1        right handed college/bachelor degree         no        village
## 2        right handed college/bachelor degree         no           city
## 3        right handed        secondary school         no           city
## 4        right handed college/bachelor degree        yes           city
## 5        right handed        secondary school         no        village
## 6        right handed        secondary school         no           city
##   House...block.of.flats
## 1         block of flats
## 2         block of flats
## 3         block of flats
## 4         house/bungalow
## 5         house/bungalow
## 6         block of flats
```

```
# Dimenzije dataseta:
dim(pitanja)  # broj redaka, broj stupaca (broj primjera, broj varijabli)
```

```
## [1] 1010  150
```

```
## Pomoću summary-ja računamo statistike i doznajemo tipove podataka:
summary(pitanja)
```

```
##      Music       Slow.songs.or.fast.songs     Dance            Folk
##  Min.   :1.000   Min.   :1.000            Min.   :1.000   Min.   :1.000
##  1st Qu.:5.000   1st Qu.:3.000            1st Qu.:2.000   1st Qu.:1.000
##  Median :5.000   Median :3.000            Median :3.000   Median :2.000
##  Mean   :4.732   Mean   :3.328            Mean   :3.113   Mean   :2.289
##  3rd Qu.:5.000   3rd Qu.:4.000            3rd Qu.:4.000   3rd Qu.:3.000
##  Max.   :5.000   Max.   :5.000            Max.   :5.000   Max.   :5.000
##  NA's   :3       NA's   :2                NA's   :4       NA's   :5
##     Country      Classical.music    Musical            Pop
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:1.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:3.000
##  Median :2.000   Median :3.000   Median :3.000   Median :4.000
##  Mean   :2.123   Mean   :2.956   Mean   :2.762   Mean   :3.472
##  3rd Qu.:3.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
##  NA's   :5       NA's   :7       NA's   :2       NA's   :3
##       Rock       Metal.or.Hardrock      Punk        Hiphop..Rap
##  Min.   :1.000   Min.   :1.000     Min.   :1.000   Min.   :1.000
##  1st Qu.:3.000   1st Qu.:1.000     1st Qu.:1.000   1st Qu.:2.000
##  Median :4.000   Median :2.000     Median :2.000   Median :3.000
##  Mean   :3.762   Mean   :2.361     Mean   :2.456   Mean   :2.911
##  3rd Qu.:5.000   3rd Qu.:3.000     3rd Qu.:3.000   3rd Qu.:4.000
##  Max.   :5.000   Max.   :5.000     Max.   :5.000   Max.   :5.000
##  NA's   :6       NA's   :3         NA's   :8       NA's   :4
##   Reggae..Ska    Swing..Jazz    Rock.n.roll     Alternative        Latino
##  Min.   :1.00   Min.   :1.00   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.00   1st Qu.:2.00   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000
##  Median :3.00   Median :3.00   Median :3.000   Median :3.000   Median :3.000
##  Mean   :2.77   Mean   :2.76   Mean   :3.142   Mean   :2.829   Mean   :2.842
##  3rd Qu.:4.00   3rd Qu.:4.00   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :5.00   Max.   :5.00   Max.   :5.000   Max.   :5.000   Max.   :5.000
##  NA's   :7      NA's   :6      NA's   :7       NA's   :7       NA's   :8
##  Techno..Trance      Opera          Movies          Horror          Thriller
##  Min.   :1.000   Min.   :1.00   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:1.000   1st Qu.:1.00   1st Qu.:4.000   1st Qu.:1.000   1st Qu.:3.000
##  Median :2.000   Median :2.00   Median :5.000   Median :3.000   Median :4.000
##  Mean   :2.339   Mean   :2.14   Mean   :4.614   Mean   :2.794   Mean   :3.384
##  3rd Qu.:3.000   3rd Qu.:3.00   3rd Qu.:5.000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :5.000   Max.   :5.00   Max.   :5.000   Max.   :5.000   Max.   :5.000
##  NA's   :7       NA's   :1      NA's   :6       NA's   :2       NA's   :1
##      Comedy         Romantic        Sci.fi           War
##  Min.   :1.000   Min.   :1.00   Min.   :1.000   Min.   :1.000
##  1st Qu.:4.000   1st Qu.:3.00   1st Qu.:2.000   1st Qu.:2.000
##  Median :5.000   Median :4.00   Median :3.000   Median :3.000
##  Mean   :4.495   Mean   :3.49   Mean   :3.113   Mean   :3.156  
##  3rd Qu.:5.000   3rd Qu.:5.00   3rd Qu.:4.000   3rd Qu.:4.000  
##  Max.   :5.000   Max.   :5.00   Max.   :5.000   Max.   :5.000  
##  NA's   :3       NA's   :3      NA's   :2       NA's   :2      
##  Fantasy.Fairy.tales    Animated      Documentary       Western     
##  Min.   :1.00        Min.   :1.000   Min.   :1.000   Min.   :1.000  
##  1st Qu.:3.00        1st Qu.:3.000   1st Qu.:3.000   1st Qu.:1.000  
##  Median :4.00        Median :4.000   Median :4.000   Median :2.000  
##  Mean   :3.75        Mean   :3.788   Mean   :3.644   Mean   :2.126  
##  3rd Qu.:5.00        3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:3.000  
##  Max.   :5.00        Max.   :5.000   Max.   :5.000   Max.   :5.000  
##  NA's   :3           NA's   :3       NA's   :8       NA's   :4      
##      Action         History        Psychology       Politics    
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000  
##  1st Qu.:3.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:1.000  
##  Median :4.000   Median :3.000   Median :3.000   Median :2.000  
##  Mean   :3.537   Mean   :3.207   Mean   :3.138   Mean   :2.596  
##  3rd Qu.:5.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000  
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000  
##  NA's   :2       NA's   :2       NA's   :5       NA's   :1      
##   Mathematics       Physics         Internet           PC       
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000  
##  1st Qu.:1.000   1st Qu.:1.000   1st Qu.:4.000   1st Qu.:2.000  
##  Median :2.000   Median :2.000   Median :4.000   Median :3.000  
##  Mean   :2.335   Mean   :2.065   Mean   :4.176   Mean   :3.136  
##  3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:5.000   3rd Qu.:4.000  
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000  
##  NA's   :3       NA's   :3       NA's   :4       NA's   :6      
##  Economy.Management    Biology        Chemistry        Reading     
##  Min.   :1.000      Min.   :1.000   Min.   :1.000   Min.   :1.000  
##  1st Qu.:1.000      1st Qu.:2.000   1st Qu.:1.000   1st Qu.:2.000  
##  Median :2.000      Median :2.000   Median :2.000   Median :3.000  
##  Mean   :2.644      Mean   :2.665   Mean   :2.165   Mean   :3.159  
##  3rd Qu.:4.000      3rd Qu.:4.000   3rd Qu.:3.000   3rd Qu.:5.000  
##  Max.   :5.000      Max.   :5.000   Max.   :5.000   Max.   :5.000  
##  NA's   :5          NA's   :6       NA's   :10      NA's   :6      
##    Geography     Foreign.languages    Medicine          Law       
##  Min.   :1.000   Min.   :1.000     Min.   :1.000   Min.   :1.000  
##  1st Qu.:2.000   1st Qu.:3.000     1st Qu.:1.000   1st Qu.:1.000  
##  Median :3.000   Median :4.000     Median :2.000   Median :2.000  
##  Mean   :3.083   Mean   :3.778     Mean   :2.516   Mean   :2.257  
##  3rd Qu.:4.000   3rd Qu.:5.000     3rd Qu.:3.000   3rd Qu.:3.000  
##  Max.   :5.000   Max.   :5.000     Max.   :5.000   Max.   :5.000  
##  NA's   :9       NA's   :5         NA's   :5       NA's   :1      
##       Cars       Art.exhibitions    Religion     Countryside..outdoors
##  Min.   :1.000   Min.   :1.00    Min.   :1.000   Min.   :1.000        
##  1st Qu.:1.000   1st Qu.:1.00    1st Qu.:1.000   1st Qu.:3.000        
##  Median :3.000   Median :2.00    Median :2.000   Median :4.000        
##  Mean   :2.687   Mean   :2.59    Mean   :2.273   Mean   :3.687        
##  3rd Qu.:4.000   3rd Qu.:4.00    3rd Qu.:3.000   3rd Qu.:5.000        
##  Max.   :5.000   Max.   :5.00    Max.   :5.000   Max.   :5.000        
##  NA's   :4       NA's   :6       NA's   :3       NA's   :7            
##     Dancing      Musical.instruments    Writing      Passive.sport  
##  Min.   :1.000   Min.   :1.000       Min.   :1.000   Min.   :1.000  
##  1st Qu.:1.000   1st Qu.:1.000       1st Qu.:1.000   1st Qu.:2.000  
##  Median :2.000   Median :2.000       Median :1.000   Median :3.000  
##  Mean   :2.462   Mean   :2.324       Mean   :1.901   Mean   :3.388  
##  3rd Qu.:4.000   3rd Qu.:4.000       3rd Qu.:3.000   3rd Qu.:5.000  
##  Max.   :5.000   Max.   :5.000       Max.   :5.000   Max.   :5.000  
##  NA's   :3       NA's   :1           NA's   :6       NA's   :15     
##   Active.sport     Gardening      Celebrities       Shopping    
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000  
##  1st Qu.:2.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:2.000  
##  Median :3.000   Median :1.000   Median :2.000   Median :3.000  
##  Mean   :3.291   Mean   :1.907   Mean   :2.362   Mean   :3.277  
##  3rd Qu.:5.000   3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:4.000  
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000  
##  NA's   :4       NA's   :7       NA's   :2       NA's   :2      
##  Science.and.technology    Theatre      Fun.with.friends Adrenaline.sports
##  Min.   :1.000          Min.   :1.000   Min.   :2.000    Min.   :1.000    
##  1st Qu.:2.000          1st Qu.:2.000   1st Qu.:4.000    1st Qu.:2.000    
##  Median :3.000          Median :3.000   Median :5.000    Median :3.000    
##  Mean   :3.234          Mean   :3.025   Mean   :4.558    Mean   :2.948    
##  3rd Qu.:4.000          3rd Qu.:4.000   3rd Qu.:5.000    3rd Qu.:4.000    
##  Max.   :5.000          Max.   :5.000   Max.   :5.000    Max.   :5.000    
##  NA's   :6              NA's   :8       NA's   :4        NA's   :3        
##       Pets           Flying          Storm          Darkness    
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000  
##  1st Qu.:2.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000  
##  Median :4.000   Median :2.000   Median :2.000   Median :2.000  
##  Mean   :3.335   Mean   :2.062   Mean   :1.973   Mean   :2.251  
##  3rd Qu.:5.000   3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:3.000  
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000  
##  NA's   :4       NA's   :3       NA's   :1       NA's   :2      
##     Heights         Spiders          Snakes           Rats      
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000  
##  1st Qu.:2.000   1st Qu.:1.000   1st Qu.:2.000   1st Qu.:1.000  
##  Median :2.000   Median :3.000   Median :3.000   Median :2.000  
##  Mean   :2.616   Mean   :2.826   Mean   :3.028   Mean   :2.409  
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:3.000  
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000  
##  NA's   :3       NA's   :5                       NA's   :3      
##      Ageing      Dangerous.dogs  Fear.of.public.speaking   Smoking         
##  Min.   :1.000   Min.   :1.000   Min.   :1.000           Length:1010       
##  1st Qu.:1.000   1st Qu.:2.000   1st Qu.:2.000           Class :character  
##  Median :2.000   Median :3.000   Median :3.000           Mode  :character  
##  Mean   :2.581   Mean   :3.043   Mean   :2.804                             
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000                             
##  Max.   :5.000   Max.   :5.000   Max.   :5.000                             
##  NA's   :1       NA's   :1       NA's   :1                                 
##    Alcohol          Healthy.eating  Daily.events    Prioritising.workload
##  Length:1010        Min.   :1.000   Min.   :1.000   Min.   :1.000        
##  Class :character   1st Qu.:3.000   1st Qu.:2.000   1st Qu.:2.000        
##  Mode  :character   Median :3.000   Median :3.000   Median :3.000        
##                     Mean   :3.032   Mean   :3.075   Mean   :2.646        
##                     3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:3.000        
##                     Max.   :5.000   Max.   :5.000   Max.   :5.000        
##                     NA's   :3       NA's   :7       NA's   :5            
##  Writing.notes    Workaholism    Thinking.ahead  Final.judgement
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:3.000   1st Qu.:1.000
##  Median :3.000   Median :3.000   Median :3.000   Median :3.000
##  Mean   :3.083   Mean   :2.996   Mean   :3.414   Mean   :2.649
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
##  NA's   :3       NA's   :5       NA's   :3       NA's   :7
##   Reliability    Keeping.promises Loss.of.interest Friends.versus.money
##  Min.   :1.000   Min.   :1.000    Min.   :1.000    Min.   :1.000
##  1st Qu.:3.000   1st Qu.:3.000    1st Qu.:2.000    1st Qu.:3.000
##  Median :4.000   Median :4.000    Median :3.000    Median :4.000
##  Mean   :3.859   Mean   :3.987    Mean   :2.709    Mean   :3.779
##  3rd Qu.:5.000   3rd Qu.:5.000    3rd Qu.:4.000    3rd Qu.:5.000
##  Max.   :5.000   Max.   :5.000    Max.   :5.000    Max.   :5.000
##  NA's   :4       NA's   :1        NA's   :4        NA's   :6
##    Funniness          Fake       Criminal.damage Decision.making
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:3.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:2.000
##  Median :3.000   Median :2.000   Median :2.000   Median :3.000
##  Mean   :3.293   Mean   :2.131   Mean   :2.604   Mean   :3.198
##  3rd Qu.:4.000   3rd Qu.:3.000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
##  NA's   :4       NA's   :1       NA's   :7       NA's   :4
##    Elections     Self.criticism  Judgment.calls   Hypochondria
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000   1st Qu.:3.000   1st Qu.:3.000   1st Qu.:1.000
##  Median :4.000   Median :4.000   Median :4.000   Median :1.000
##  Mean   :3.415   Mean   :3.579   Mean   :3.987   Mean   :1.913
##  3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:3.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
##  NA's   :3       NA's   :5       NA's   :4       NA's   :4
##     Empathy      Eating.to.survive     Giving      Compassion.to.animals
##  Min.   :1.000   Min.   :1.000     Min.   :1.000   Min.   :1.000
##  1st Qu.:3.000   1st Qu.:1.000     1st Qu.:2.000   1st Qu.:3.000
##  Median :4.000   Median :2.000     Median :3.000   Median :4.000
##  Mean   :3.859   Mean   :2.229     Mean   :2.976   Mean   :3.971
##  3rd Qu.:5.000   3rd Qu.:3.000     3rd Qu.:4.000   3rd Qu.:5.000
##  Max.   :5.000   Max.   :5.000     Max.   :5.000   Max.   :5.000
##  NA's   :5                         NA's   :6       NA's   :7
##  Borrowed.stuff    Loneliness    Cheating.in.school     Health
##  Min.   :1.000   Min.   :1.000   Min.   :1.000      Min.   :1.000
##  1st Qu.:3.000   1st Qu.:2.000   1st Qu.:3.000      1st Qu.:3.000
##  Median :4.000   Median :3.000   Median :4.000      Median :3.000
##  Mean   :4.018   Mean   :2.887   Mean   :3.745      Mean   :3.251
##  3rd Qu.:5.000   3rd Qu.:4.000   3rd Qu.:5.000      3rd Qu.:4.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000      Max.   :5.000
##  NA's   :2       NA's   :1       NA's   :4         NA's   :1
##  Changing.the.past      God           Dreams         Charity
##  Min.   :1.000     Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000     1st Qu.:2.000   1st Qu.:3.000   1st Qu.:1.000
##  Median :3.000     Median :3.000   Median :3.000   Median :2.000
##  Mean   :2.952     Mean   :3.303   Mean   :3.297   Mean   :2.104
##  3rd Qu.:4.000     3rd Qu.:5.000   3rd Qu.:4.000   3rd Qu.:3.000
##  Max.   :5.000     Max.   :5.000   Max.   :5.000   Max.   :5.000  
##  NA's   :2         NA's   :2                       NA's   :3      
##  Number.of.friends Punctuality           Lying              Waiting     
##  Min.   :1.000     Length:1010        Length:1010        Min.   :1.000  
##  1st Qu.:3.000     Class :character   Class :character   1st Qu.:2.000  
##  Median :3.000     Mode  :character   Mode  :character   Median :3.000  
##  Mean   :3.344                                           Mean   :2.672  
##  3rd Qu.:4.000                                           3rd Qu.:3.000  
##  Max.   :5.000                                           Max.   :5.000  
##                                                          NA's   :3      
##  New.environment  Mood.swings    Appearence.and.gestures  Socializing   
##  Min.   :1.000   Min.   :1.000   Min.   :1.000           Min.   :1.000  
##  1st Qu.:3.000   1st Qu.:3.000   1st Qu.:3.000           1st Qu.:2.000  
##  Median :4.000   Median :3.000   Median :4.000           Median :3.000  
##  Mean   :3.475   Mean   :3.258   Mean   :3.598           Mean   :3.158  
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000           3rd Qu.:4.000  
##  Max.   :5.000   Max.   :5.000   Max.   :5.000           Max.   :5.000  
##  NA's   :2       NA's   :4       NA's   :3               NA's   :5      
##   Achievements   Responding.to.a.serious.letter    Children     Assertiveness 
##  Min.   :1.000   Min.   :1.000                  Min.   :1.000   Min.   :1.000  
##  1st Qu.:2.000   1st Qu.:2.000                  1st Qu.:3.000   1st Qu.:3.000  
##  Median :3.000   Median :3.000                  Median :4.000   Median :4.000  
##  Mean   :2.963   Mean   :3.071                  Mean   :3.621   Mean   :3.519  
##  3rd Qu.:4.000   3rd Qu.:4.000                  3rd Qu.:5.000   3rd Qu.:4.000  
##  Max.   :5.000   Max.   :5.000                  Max.   :5.000   Max.   :5.000  
##  NA's   :2       NA's   :6                      NA's   :4       NA's   :2      
##  Getting.angry   Knowing.the.right.people Public.speaking  Unpopularity  
##  Min.   :1.000   Min.   :1.000            Min.   :1.000   Min.   :1.000  
##  1st Qu.:2.000   1st Qu.:3.000            1st Qu.:3.000   1st Qu.:3.000  
##  Median :3.000   Median :4.000            Median :4.000   Median :3.000  
##  Mean   :3.015   Mean   :3.486            Mean   :3.522   Mean   :3.462  
##  3rd Qu.:4.000   3rd Qu.:4.000            3rd Qu.:5.000   3rd Qu.:4.000  
##  Max.   :5.000   Max.   :5.000            Max.   :5.000   Max.   :5.000  
##  NA's   :4       NA's   :2                NA's   :2       NA's   :3      
##  Life.struggles  Happiness.in.life Energy.levels   Small...big.dogs
##  Min.   :1.000   Min.   :1.000     Min.   :1.000   Min.   :1.000   
##  1st Qu.:2.000   1st Qu.:3.000     1st Qu.:3.000   1st Qu.:2.000   
##  Median :3.000   Median :4.000     Median :4.000   Median :3.000   
##  Mean   :3.032   Mean   :3.706     Mean   :3.634   Mean   :2.973   
##  3rd Qu.:4.000   3rd Qu.:4.000     3rd Qu.:4.000   3rd Qu.:4.000   
##  Max.   :5.000   Max.   :5.000     Max.   :5.000   Max.   :5.000   
##  NA's   :3       NA's   :4         NA's   :5       NA's   :4       
##   Personality    Finding.lost.valuables   Getting.up    Interests.or.hobbies
##  Min.   :1.000   Min.   :1.000          Min.   :1.000   Min.   :1.000       
##  1st Qu.:3.000   1st Qu.:2.000          1st Qu.:3.000   1st Qu.:3.000       
##  Median :3.000   Median :3.000          Median :4.000   Median :4.000       
##  Mean   :3.292   Mean   :2.872          Mean   :3.592   Mean   :3.551       
##  3rd Qu.:4.000   3rd Qu.:4.000          3rd Qu.:5.000   3rd Qu.:5.000       
##  Max.   :5.000   Max.   :5.000          Max.   :5.000   Max.   :5.000       
##  NA's   :4       NA's   :4              NA's   :5       NA's   :3          
##  Parents..advice Questionnaires.or.polls Internet.usage        Finances    
##  Min.   :1.000   Min.   :1.000           Length:1010        Min.   :1.000  
##  1st Qu.:3.000   1st Qu.:2.000           Class :character   1st Qu.:2.000  
##  Median :3.000   Median :3.000           Mode  :character   Median :3.000  
##  Mean   :3.266   Mean   :2.749                              Mean   :3.024  
##  3rd Qu.:4.000   3rd Qu.:3.000                              3rd Qu.:4.000  
##  Max.   :5.000   Max.   :5.000                              Max.   :5.000  
##  NA's   :2       NA's   :4                                  NA's   :3      
##  Shopping.centres Branded.clothing Entertainment.spending Spending.on.looks
##  Min.   :1.000    Min.   :1.000    Min.   :1.000          Min.   :1.000    
##  1st Qu.:2.000    1st Qu.:2.000    1st Qu.:2.000          1st Qu.:2.000    
##  Median :3.000    Median :3.000    Median :3.000          Median :3.000    
##  Mean   :3.234    Mean   :3.051    Mean   :3.202          Mean   :3.106    
##  3rd Qu.:4.000    3rd Qu.:4.000    3rd Qu.:4.000          3rd Qu.:4.000    
##  Max.   :5.000    Max.   :5.000    Max.   :5.000          Max.   :5.000    
##  NA's   :2        NA's   :2        NA's   :3              NA's   :3        
##  Spending.on.gadgets Spending.on.healthy.eating      Age           Height     
##  Min.   :1.00        Min.   :1.000              Min.   :15.00   Min.   : 62.0  
##  1st Qu.:2.00        1st Qu.:3.000              1st Qu.:19.00   1st Qu.:167.0  
##  Median :3.00        Median :4.000              Median :20.00   Median :173.0  
##  Mean   :2.87        Mean   :3.558              Mean   :20.43   Mean   :173.5  
##  3rd Qu.:4.00        3rd Qu.:4.000              3rd Qu.:22.00   3rd Qu.:180.0  
##  Max.   :5.00        Max.   :5.000              Max.   :30.00   Max.   :203.0  
##                      NA's   :2                  NA's   :7       NA's   :20     
##      Weight       Number.of.siblings    Gender          Left...right.handed
##  Min.   : 41.00   Min.   : 0.000     Length:1010        Length:1010        
##  1st Qu.: 55.00   1st Qu.: 1.000     Class :character   Class :character   
##  Median : 64.00   Median : 1.000     Mode  :character   Mode  :character   
##  Mean   : 66.41   Mean   : 1.298                                           
##  3rd Qu.: 75.00   3rd Qu.: 2.000                                           
##  Max.   :165.00   Max.   :10.000                                           
##  NA's   :20       NA's   :6                                                
##   Education          Only.child        Village...town    
##  Length:1010        Length:1010        Length:1010       
##  Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character  
##                                                          
##                                                          
##                                                          
##                                                          
##  House...block.of.flats
##  Length:1010           
##  Class :character      
##  Mode  :character      
##                        
##                        
##                        
##                        
```

Prije nego počnemo s uporabom deskriptivne statistike i manipulacijom podataka, neka nam misao vodilja budu predložena istraživačka pitanja:

Istraživačko pitanje 1: Razlikuju li se izrazeni strahovi ispitanih žena i muškaraca?

Počinjemo odvajanjem skupa podataka na skup podataka gdje su ispitanici žene i drugog gdje su ispitanici muškarci.

```
zene = pitanja[pitanja$Gender == "female", ]
muskarci = pitanja[pitanja$Gender == "male", ]
```

Pogledajmo sada o kojim se to strahovima radi, odnosno o kojim strahovima imamo prikupljene podatke.

```
for(column_name in names(pitanja[64:73]))
  print(column_name)
```

```
## [1] "Flying"
## [1] "Storm"
## [1] "Darkness"
## [1] "Heights"
## [1] "Spiders"
## [1] "Snakes"
## [1] "Rats"
## [1] "Ageing"
## [1] "Dangerous.dogs"
## [1] "Fear.of.public.speaking"
```

Kako bismo vidjeli kako su ispitanici rangirali razinu straha (ocjena 1-5) moramo proći kroz svaki pojedini strah za naše dvije skupine, vizualizirati podatke i i mjere centralne tendencije za stjecanje boljeg uvida u podatke; žene i muškarce.

```
zene = zene[complete.cases(zene['Rats']),]
muskarci = muskarci[complete.cases(muskarci['Rats']),]

cat('Srednja vrijednost iskazanog straha od štakora kod žena iznosi ', mean(zene$Rats),'\n')
```

```
## Srednja vrijednost iskazanog straha od štakora kod žena iznosi  2.721284
```

```
cat('Srednja vrijednost iskazanog straha od štakora muškaraca iznosi ', mean(muskarci$Rats), '\n')
```

```
## Srednja vrijednost iskazanog straha od štakora muškaraca iznosi  1.963325
```

```
cat('Podrezana srednja vrijednost iskazanog straha od štakora kod žena iznosi ', mean(zene$Rats, trim =
```

```
## Podrezana srednja vrijednost iskazanog straha od štakora kod žena iznosi  2.651899
```

```
cat('Podrezana srednja vrijednost iskazanog straha od štakora muškaraca iznosi ', mean(muskarci$Rats, t
```

```
## Podrezana srednja vrijednost iskazanog straha od štakora muškaraca iznosi  1.787234
```

```
cat('Medijan iskazanog straha od štakora kod žena iznosi ', median(zene$Rats),'\n')
```

```
## Medijan iskazanog straha od štakora kod žena iznosi  3
```

```
cat('Medijan iskazanog straha od štakora muškaraca iznosi ', median(muskarci$Rats), '\n')
```

```
## Medijan iskazanog straha od štakora muškaraca iznosi  2
```

```
cat('Standardna devijacija iskazanog straha od štakora kod žena iznosi ', sd(zene$Rats),'\n')
```

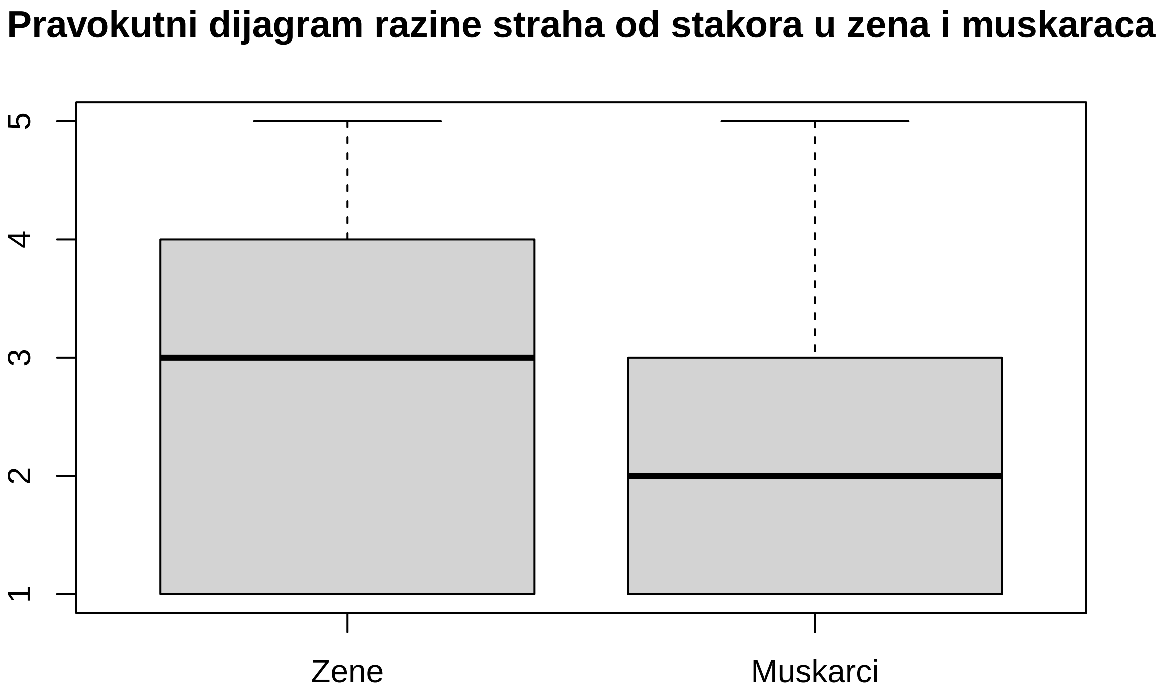
```
## Standardna devijacija iskazanog straha od štakora kod žena iznosi  1.468802
```

```
cat('Standardna devijacija iskazanog straha od štakora muškaraca iznosi ', sd(muskarci$Rats), '\n')
```

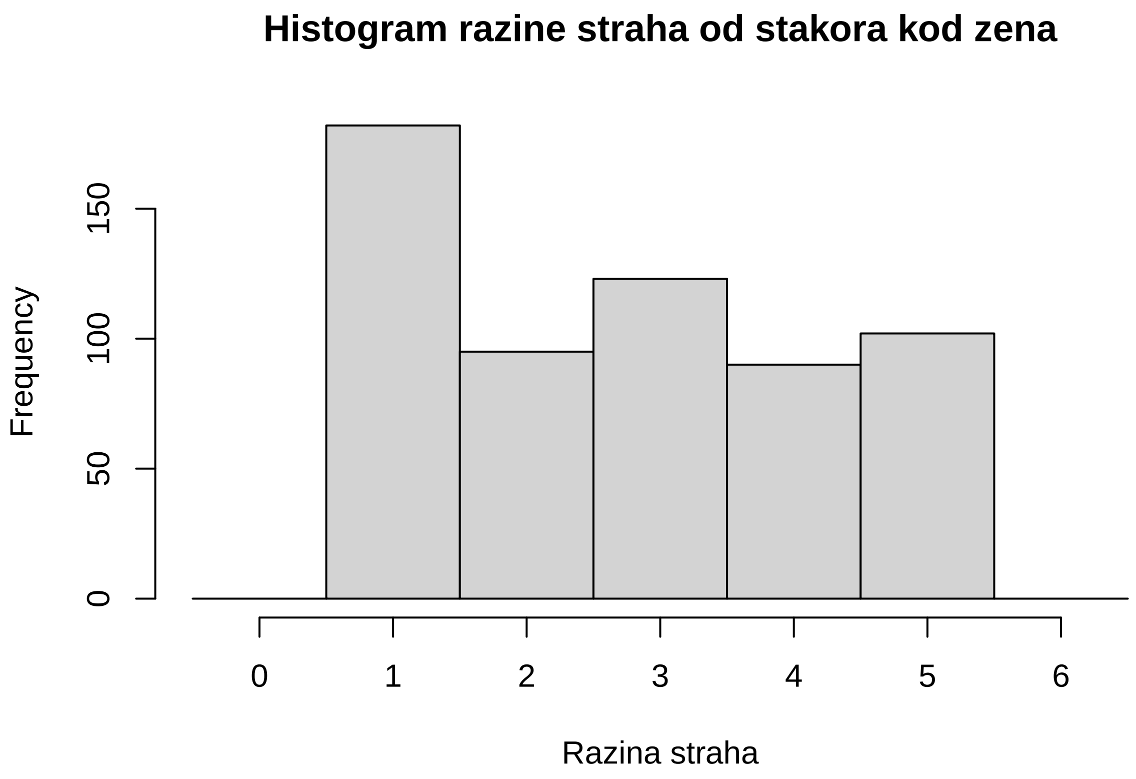
```
## Standardna devijacija iskazanog straha od štakora muškaraca iznosi  1.161526
```

Nakon sto smo vidjeli koliko iznose mjere centralne tendencije, zanima nas kako su te vrijednosti raspoređene, imamo li vrijednosti koje odskaču i sl., a to najbolje možemo uvidjeti vizualizacijom podataka. Nadalje, histograma ćemo koristit kako bismo doznali oblik naše distribucije i gustoće podataka.

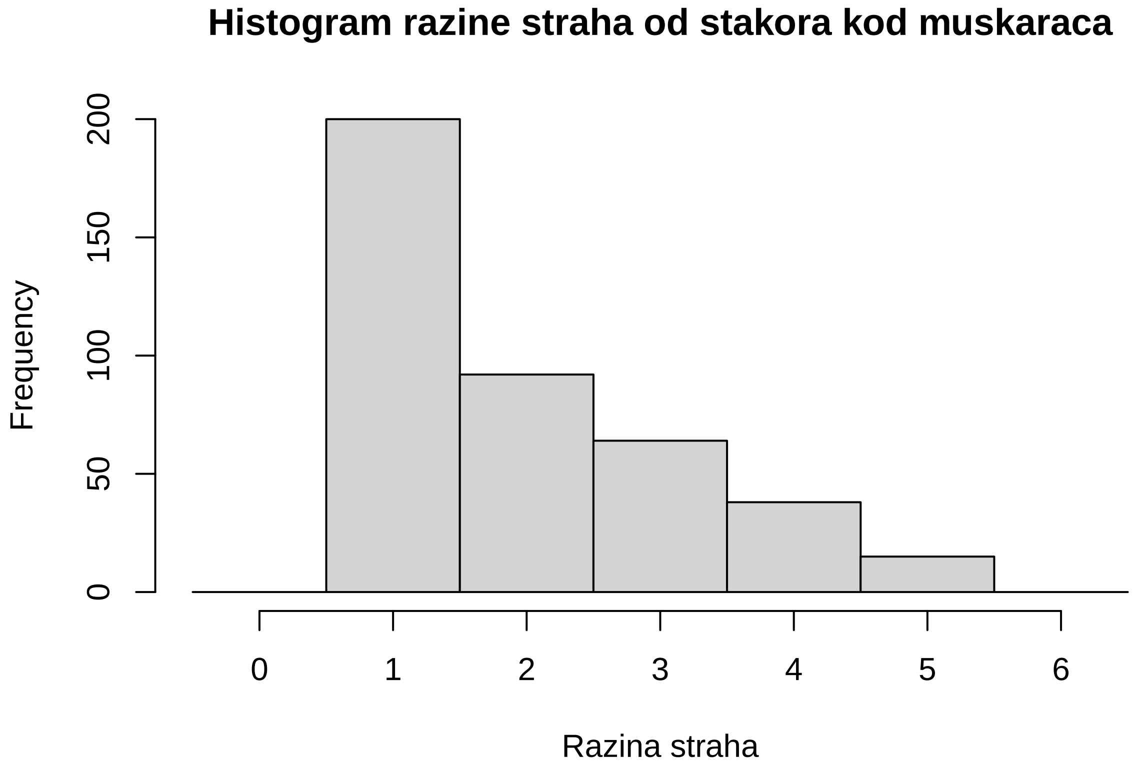
```
# Pravokutni dijagrami dviju skupina za strah od stakora:
boxplot(zene$Rats, muskarci$Rats,
        names = c('Zene','Muskarci'),
        main='Pravokutni dijagram razine straha od stakora u zena i muskaraca')
```



```
hist(zene$Rats,
     breaks=seq(min(zene$Rats)-1.5,max(zene$Rats)+1.5,1),
     main='Histogram razine straha od stakora kod zena',
     xlab='Razina straha')
```



```
hist(muskarci$Rats,
     breaks=seq(min(muskarci$Rats)-1.5,max(muskarci$Rats)+1.5,1),
     main='Histogram razine straha od stakora kod muskaraca',
     xlab='Razina straha')
```



Vizualno možemo pretpostaviti da zaista postoji razlika među iskazanih strahom od štakora u žena i muškaraca, ali kako bismo to zaista i dokazali potrebno je provesti statistički test koji će testirati jednakost srednjih vrijednosti dviju populacija.

Ovakvo ispitivanje možemo provesti t-testom.

Testiranje jednakosti srednjih vrijednosti dvije populacije

Neka su $X_1^1, X_1^2, \ldots, X_1^{n_1}$ i $X_2^1, X_2^2, \ldots, X_2^{n_2}$ dva nezavisna slučajna uzorka koji dolaze iz normalnih distribucija s očekivanjima μ_1 i μ_2 te s nepoznatim, ali jednakim varijancama σ. Zajednička disperzija uzorka se računa kao težinska sredina disperzija S_{X_1} i S_{X_2}:

$$S_X^2 = \frac{1}{n_1 + n_2 - 2}[(n_1 - 1)S_{X_1}^2 + (n_2 - 1)S_{X_2}^2].$$

Slučajna varijabla

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

ima jediničnu normalnu distribuciju. Slučajna varijabla

$$W^2 = \frac{(n_1 - 1)S_{X_1}^2 + (n_2 - 1)S_{X_2}^2}{\sigma^2}$$

ima χ^2 razdiobu s $n_1 + n_2 - 2$ stupnja slobode. Zato slučajna varijabla

$$T = \frac{Z\sqrt{n_1 + n_2 - 2}}{W} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_X\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

ima egzaktnu t distribuciju s $n_1 + n_2 - 2$ stupnja slobode.

Ukoliko imamo 2 nezavisno normalo distribuirana uzorka, ali ovoga puta sa različitim varijancama, tada koristimo testnu statistiku

$$T' = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_{X_1}^2}{n_1} + \frac{s_{X_2}^2}{n_2}}}$$

koja ima aproksimativnu t-distribuciju sa stupnjevima slobode

$$v = \frac{(s_{X_1}^2/n_1 + s_{X_2}^2/n_2)^2}{(s_{X_1}^2/n_1)^2/(n_1 - 1) + (s_{X_2}^2/n_2)^2/(n_2 - 1)}$$

gdje je

$$s_{X_i}^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(X_i^j - \bar{X}_i)^2$$

za $i = 1, 2$.

Hipoteze tada glase:

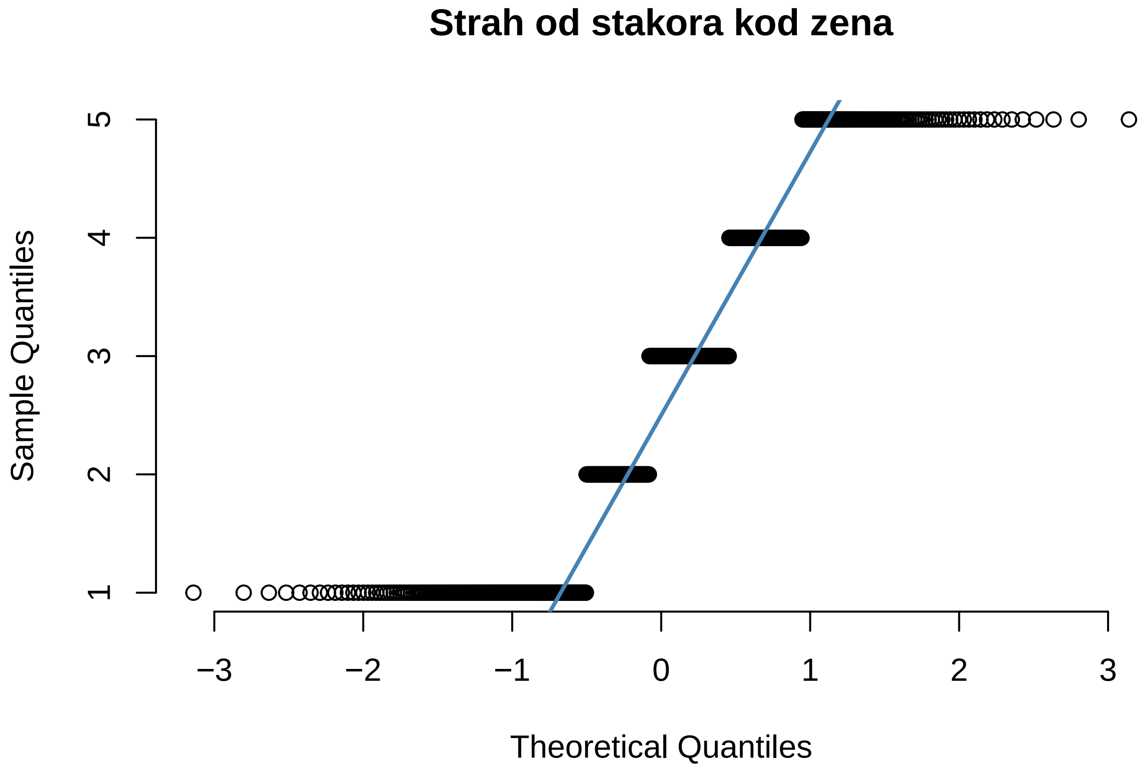
$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Kako bismo mogli provesti test, moramo najprije provjeriti pretpostavke normalnosti i nezavisnosti uzorka. Već iz histograma možemo vidjeti da bismo mogli imati problem s normalnošću naših podataka, ali ono što nam ide u prilog da će naša statistika ipak biti robusna jest veličina skupa podataka.

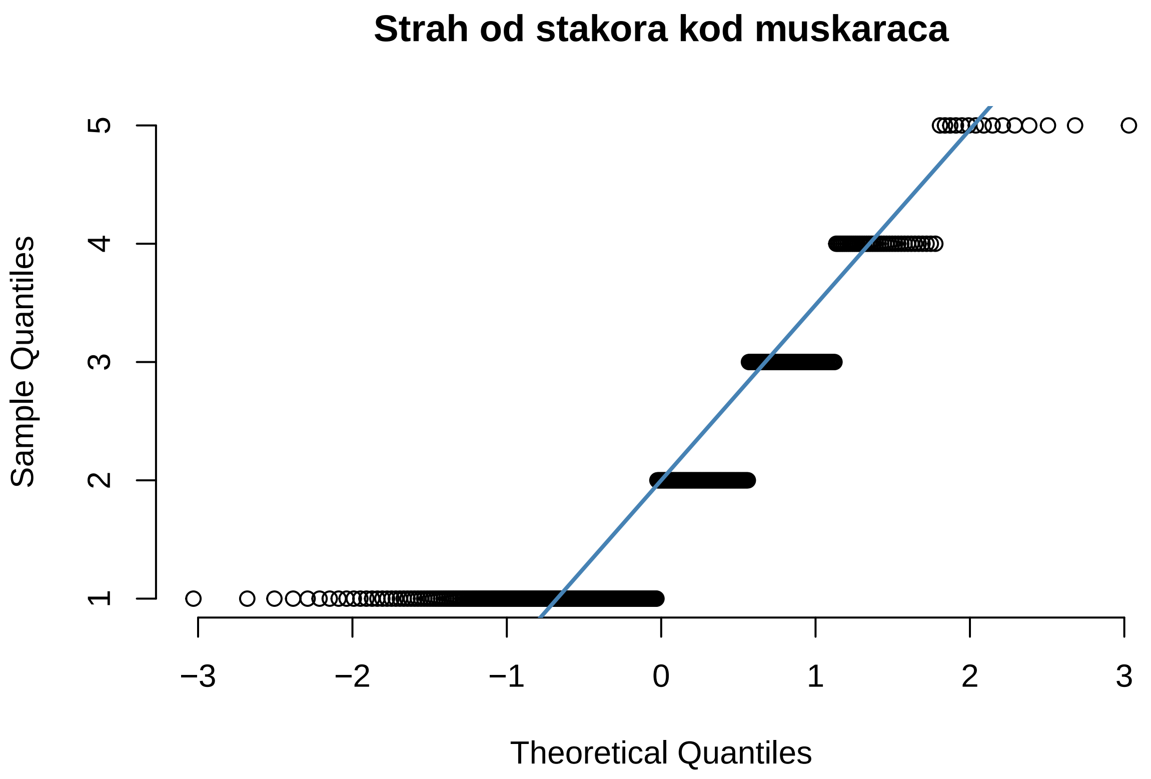
Obzirom da razmatramo dva uzoraka dvaju različitih spolova, možemo pretpostaviti njihovu nezavisnost.

Sada, dakle, trebamo provjeriti normalnost podataka koju najčešće provjeravamo: histgoramom (kojeg smo prethodno već iscrtali), qq-plotom te KS-testom (kojim provjeravamo pripadnost podataka distribuciji).

```
qqnorm(zene$Rats, pch = 1, frame = FALSE,main='Strah od stakora kod zena')
qqline(zene$Rats, col = "steelblue", lwd = 2)
```



```
qqnorm(muskarci$Rats, pch = 1, frame = FALSE,main='Strah od stakora kod muskaraca')
qqline(muskarci$Rats, col = "steelblue", lwd = 2)
```



Na temelju qq-plota možemo vidjeti da su naše sumnje u normalnost podataka zaista opravdane. Ali, kao što smo već prije spomenuli, statistika bi idalje mogla biti robusna obzirom na veličinu našeg skupa podataka. Provjerimo zadovoljavamo li ostale preduvjete za provođenje t-testa, uzevši u obzirom već izračunate varijance naših podataka. Testirajmo prvo jesu li naše varijance značajno različite.

Test o jednakosti varijanci

Ako imamo dva nezavisna slučajna uzorka $X_1^1, X_1^2, \ldots X_1^{n_1}$ i $X_2^1, X_2^2, \ldots, X_2^{n_2}$ koji dolaze iz normalnih distribucija s varijancama σ_1^2 i σ_2^2, tada slučajna varijabla

$$F = \frac{S_{X_1}^2/\sigma_1^2}{S_{X_2}^2/\sigma_2^2}$$

ima Fisherovu distribuciju s $(n_1 - 1, n_2 - 1)$ stupnjeva slobode, pri čemu vrijedi:

$$S_{X_1}^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_1^i - \bar{X}_1)^2, \quad S_{X_2}^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(X_2^i - \bar{X}_2)^2.$$

Hipoteze testa jednakosti varijanci glase:

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

U programskom paketu R test o jednakosti varijanci je implementiran u funkciji `var.test()`, koja prima uzorke iz dvije populacije čije varijance uspoređujemo.

Dakle, ispitajmo jednakost varijanci naših danih uzoraka.

```
var.test(zene$Rats, muskarci$Rats)
```

```
##
##  F test to compare two variances
```

```
##
## data:  zene$Rats and muskarci$Rats
## F = 1.5991, num df = 591, denom df = 408, p-value = 4.314e-07
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.335516 1.908986
## sample estimates:
## ratio of variances
##           1.599076
```

p-vrijednost od 4.314e-07 nam govori da nećemo odbaciti hipotezu H_0 da su varijance naša dva uzorka jednaka.

Provedimo sada dvostrani t-test uz pretpostavku jednakosti varijanci.

```
# Uvijek se držimo istog poretka
t.test(zene$Rats, muskarci$Rats, alt = "two.sided", var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  zene$Rats and muskarci$Rats
## t = 8.7206, df = 999, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.5874004 0.9285168
## sample estimates:
## mean of x mean of y
##  2.721284  1.963325
```

Zbog jako male p-vrijednost možemo odbaciti H_0 hipotezu o jednakosti izrazenih strahov u korist H_1, odnosno možemo reći da se izraženi strahovi kod muškaraca i žena razlikuju.

Pokušajmo sada nešto robusnije od t-testa kako bismo dokazali svoju pretpostavku, vidjeli smo da naši podaci ne odgovaraju normalnosti, što ćemo i dodatno potvrditi Shapiro testom. Provedimo dakle neparametarski test kako bismo ustanovili razlikuju se izraženi strahovi žena i muškaraca, dakle koristeći iste hipoteze. Koristit ćemo Mann-Whitney i Wilcox signed-rank test.

```
shapiro.test(muskarci$Rats) #podaci nisu normalno distribirani - studentov t-test nije prikladan
```

```
##
##  Shapiro-Wilk normality test
##
## data:  muskarci$Rats
## W = 0.78849, p-value < 2.2e-16
```

```
wilcox.test(muskarci$Rats, zene$Rats, paired = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  muskarci$Rats and zene$Rats
## W = 86003, p-value = 5.698e-16
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(zene$Rats, muskarci$Rats, paired = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
```

```
##
## data:  zene$Rats and muskarci$Rats
## W = 156125, p-value = 5.698e-16
## alternative hypothesis: true location shift is not equal to 0
```

Zaključno, nakon što smo Shapiro testom potvrdili da se podaci ne ravnaju po normalnoj distribuciji, uz 95% interval povjerenja, možemo odbaciti H_0 hipotezu o jednakosti izrazenih strahova u korist H_1, odnosno možemo reći da se izraženi strahovi kod muškaraca i žena razlikuju.

Možemo primjetiti i da se različitim redoslijedom argumenata W vrijednost mijenja u Mann-Whitney u-testu. Uzevši u obzir uobičajne prakse, zadržavamo vrijednost manje statistile i bilježimo da je W = 86003 iako se p-vrijednosti pritom, naravno, ne mijenjaju.

Provedimo još sada Wilcox signed-rank test s našim uparenim uzorcima. Potrebno će dakle biti uskladiti veličine uzoraka te ćemo se zadržati pri tome da veći uzorak smanjimo na veličinu manjeg uzorka, iako postoje sofisticiranije metode, ovdje ćemo se ipak zadovoljiti ovakvim pristupom.

```r
uzorak = pitanja[complete.cases(pitanja['Rats']),]
zene = uzorak[uzorak$Gender == "female", ]
muskarci = uzorak[uzorak$Gender == "male", ]

dim(zene)
```

```
## [1] 592 150
```

```r
dim(muskarci)
```

```
## [1] 409 150
```

```r
zene2 = zene[1:409,] #broj muškaraca u skupu podataka

wilcox.test(muskarci$Rats, zene2$Rats, paired = TRUE)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  muskarci$Rats and zene2$Rats
## V = 13676, p-value = 2.198e-14
## alternative hypothesis: true location shift is not equal to 0
```

```r
wilcox.test(zene2$Rats, muskarci$Rats, paired = TRUE)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  zene2$Rats and muskarci$Rats
## V = 39299, p-value = 2.198e-14
## alternative hypothesis: true location shift is not equal to 0
```

Provođenjem Wilcox signed-rank test, uz 95% interval povjerenja, također, možemo odbaciti H_0 hipotezu o jednakosti izrazenih strahova od štakora u korist H_1, odnosno možemo reći da se izraženi strahovi od štkaora kod muškaraca i žena razlikuju. Ovdje također bilježimo vrijednost manje statistike, tj, V = 13676.

Istovjetne postupke, vizualizacije i testiranja proveli smo i na svim drugim strahovima te su rezultati više-manje istovjetno možemo ustvrditi da se strahovi zaista razlikuju između žena i muškaraca, ali se zbog opsežnosti ove bilježnice nećemo upuštati u provođenje testova. Valja naglasiti da se i kod ostalih skupina podaci ne ravnaju po normalnoj distribuciji te je učinkovitije bilo provoditi neparametarske testove. Samo ćemo ukratko navedeno dokazati provođenjem neparamtarskih testova za svaki strah.

```
shapiro.test(muskarci$Flying) #podaci nisu normalno distribirani - studentov t-test nije prikladan
```

```
##
##  Shapiro-Wilk normality test
##
## data:  muskarci$Flying
## W = 0.75116, p-value < 2.2e-16
```

```
wilcox.test(muskarci$Flying, zene$Flying, paired = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  muskarci$Flying and zene$Flying
## W = 102696, p-value = 2.909e-05
## alternative hypothesis: true location shift is not equal to 0
```

```
uzorak = pitanja[complete.cases(pitanja['Flying']),]

zene = uzorak[uzorak$Gender == "female", ]
muskarci = uzorak[uzorak$Gender == "male", ]
zene2 = zene[1:409,] #broj muškaraca u skupu podataka

wilcox.test(muskarci$Flying, zene2$Flying, paired = TRUE)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  muskarci$Flying and zene2$Flying
## V = 16030, p-value = 0.001086
## alternative hypothesis: true location shift is not equal to 0
```

```
shapiro.test(muskarci$Storm) #podaci nisu normalno distribirani - studentov t-test nije prikladan
```

```
##
##  Shapiro-Wilk normality test
##
## data:  muskarci$Storm
## W = 0.65773, p-value < 2.2e-16
```

```
wilcox.test(muskarci$Storm, zene$Storm, paired = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  muskarci$Storm and zene$Storm
## W = 78790, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```
uzorak = pitanja[complete.cases(pitanja['Storm']),]

zene = uzorak[uzorak$Gender == "female", ]
muskarci = uzorak[uzorak$Gender == "male", ]
zene2 = zene[1:411,] #broj muškaraca u skupu podataka

wilcox.test(muskarci$Storm, zene2$Storm, paired = TRUE)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  muskarci$Storm and zene2$Storm
## V = 8718.5, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```
shapiro.test(muskarci$Darkness) #podaci nisu normalno distribirani - studentov t-test nije prikladan
```

```
##
##  Shapiro-Wilk normality test
##
## data:  muskarci$Darkness
## W = 0.749, p-value < 2.2e-16
```

```
wilcox.test(muskarci$Darkness, zene$Darkness, paired = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  muskarci$Darkness and zene$Darkness
## W = 76749, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```
uzorak = pitanja[complete.cases(pitanja['Darkness']),]

zene = uzorak[uzorak$Gender == "female", ]
muskarci = uzorak[uzorak$Gender == "male", ]

zene2 = zene[1:410,] #broj muškaraca u skupu podataka

wilcox.test(muskarci$Darkness, zene2$Darkness, paired = TRUE)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  muskarci$Darkness and zene2$Darkness
## V = 8965, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```
shapiro.test(muskarci$Heights) #podaci nisu normalno distribirani - studentov t-test nije prikladan
```

```
##
##  Shapiro-Wilk normality test
##
## data:  muskarci$Heights
## W = 0.89395, p-value = 3.292e-16
```

```
wilcox.test(muskarci$Heights, zene$Heights, paired = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  muskarci$Heights and zene$Heights
## W = 120292, p-value = 0.9505
## alternative hypothesis: true location shift is not equal to 0
```

```
uzorak = pitanja[complete.cases(pitanja['Heights']),]

zene = uzorak[uzorak$Gender == "female", ]
muskarci = uzorak[uzorak$Gender == "male", ]

zene2 = zene[1:409,] #broj muškaraca u skupu podataka

wilcox.test(muskarci$Heights, zene2$Heights, paired = TRUE)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  muskarci$Heights and zene2$Heights
## V = 25932, p-value = 0.4767
## alternative hypothesis: true location shift is not equal to 0
```

```
shapiro.test(muskarci$Spiders) #podaci nisu normalno distribirani - studentov t-test nije prikladan
```

```
##
##  Shapiro-Wilk normality test
##
## data:  muskarci$Spiders
## W = 0.81886, p-value < 2.2e-16
```

```
wilcox.test(muskarci$Spiders, zene$Spiders, paired = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  muskarci$Spiders and zene$Spiders
## W = 74371, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```
uzorak = pitanja[complete.cases(pitanja['Spiders']),]

zene = uzorak[uzorak$Gender == "female", ]
muskarci = uzorak[uzorak$Gender == "male", ]

zene2 = zene[1:409,] #broj muškaraca u skupu podataka

wilcox.test(muskarci$Spiders, zene2$Spiders, paired = TRUE)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  muskarci$Spiders and zene2$Spiders
## V = 12052, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```
shapiro.test(muskarci$Snakes) #podaci nisu normalno distribirani - studentov t-test nije prikladan
```

```
##
##  Shapiro-Wilk normality test
##
## data:  muskarci$Snakes
## W = 0.86623, p-value < 2.2e-16
```

```r
wilcox.test(muskarci$Snakes, zene$Snakes, paired = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  muskarci$Snakes and zene$Snakes
## W = 88962, p-value = 4.946e-13
## alternative hypothesis: true location shift is not equal to 0
```

```r
uzorak = pitanja[complete.cases(pitanja['Snakes']),]

zene = uzorak[uzorak$Gender == "female", ]
muskarci = uzorak[uzorak$Gender == "male", ]

zene2 = zene[1:411,] #broj muškaraca u skupu podataka

wilcox.test(muskarci$Snakes, zene2$Snakes, paired = TRUE)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  muskarci$Snakes and zene2$Snakes
## V = 16746, p-value = 9.238e-13
## alternative hypothesis: true location shift is not equal to 0
```

```r
shapiro.test(muskarci$Ageing) #podaci nisu normalno distribirani - studentov t-test nije prikladan
```

```
##
##  Shapiro-Wilk normality test
##
## data:  muskarci$Ageing
## W = 0.83926, p-value < 2.2e-16
```

```r
wilcox.test(muskarci$Ageing, zene$Ageing, paired = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  muskarci$Ageing and zene$Ageing
## W = 98651, p-value = 1.555e-07
## alternative hypothesis: true location shift is not equal to 0
```

```r
uzorak = pitanja[complete.cases(pitanja['Ageing']),]

zene = uzorak[uzorak$Gender == "female", ]
muskarci = uzorak[uzorak$Gender == "male", ]

zene2 = zene[1:411,] #broj muškaraca u skupu podataka

wilcox.test(muskarci$Ageing, zene2$Ageing, paired = TRUE)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  muskarci$Ageing and zene2$Ageing
## V = 18157, p-value = 1.39e-06
## alternative hypothesis: true location shift is not equal to 0
```

```r
shapiro.test(muskarci$Dangerous.dogs) #podaci nisu normalno distribirani - studentov t-test nije prikla
```

```
##
##  Shapiro-Wilk normality test
##
## data:  muskarci$Dangerous.dogs
## W = 0.90161, p-value = 1.209e-15
```

```r
wilcox.test(muskarci$Dangerous.dogs, zene$Dangerous.dogs, paired = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  muskarci$Dangerous.dogs and zene$Dangerous.dogs
## W = 93088, p-value = 1.265e-10
## alternative hypothesis: true location shift is not equal to 0
```

```r
uzorak = pitanja[complete.cases(pitanja['Dangerous.dogs']),]

zene = uzorak[uzorak$Gender == "female", ]
muskarci = uzorak[uzorak$Gender == "male", ]

zene2 = zene[1:411,] #broj muškaraca u skupu podataka

wilcox.test(muskarci$Dangerous.dogs, zene2$Dangerous.dogs, paired = TRUE)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  muskarci$Dangerous.dogs and zene2$Dangerous.dogs
## V = 14790, p-value = 1.7e-12
## alternative hypothesis: true location shift is not equal to 0
```

```r
shapiro.test(muskarci$Fear.of.public.speaking) #podaci nisu normalno distribirani - studentov t-test ni
```

```
##
##  Shapiro-Wilk normality test
##
## data:  muskarci$Fear.of.public.speaking
## W = 0.90292, p-value = 1.642e-15
```

```r
wilcox.test(muskarci$Fear.of.public.speaking, zene$Fear.of.public.speaking, paired = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  muskarci$Fear.of.public.speaking and zene$Fear.of.public.speaking
## W = 103495, p-value = 4.52e-05
## alternative hypothesis: true location shift is not equal to 0
```

```r
uzorak = pitanja[complete.cases(pitanja['Fear.of.public.speaking']),]

zene = uzorak[uzorak$Gender == "female", ]
muskarci = uzorak[uzorak$Gender == "male", ]

zene2 = zene[1:410,] #broj muškaraca u skupu podataka
```

```
wilcox.test(muskarci$Fear.of.public.speaking, zene2$Fear.of.public.speaking, paired = TRUE)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  muskarci$Fear.of.public.speaking and zene2$Fear.of.public.speaking
## V = 21660, p-value = 0.0001985
## alternative hypothesis: true location shift is not equal to 0
```

Zaključak

Provođenjem Wilcox signed-rank test, uz 95% interval povjerenja, možemo odbaciti H_0 hipotezu o jednakosti izraženih strahova u žena i muškaraca u korist H_1, odnosno možemo reći da se izraženi strahovi kod muškaraca i žena različiti.

Istraživačko pitanje 2: Možemo li predvidjeti obrazac potrošnje ovisno o žanru glazbe kojeg ispitanik preferira?

Pogledajmo prvo koje sve žanrove glazbe imamo na raspolaganju i koje obrasce potršnje smo opažali.

```
for(column_name in names(pitanja[4:18]))
  print(column_name)
```

```
## [1] "Folk"
## [1] "Country"
## [1] "Classical.music"
## [1] "Musical"
## [1] "Pop"
## [1] "Rock"
## [1] "Metal.or.Hardrock"
## [1] "Punk"
## [1] "Hiphop..Rap"
## [1] "Reggae..Ska"
## [1] "Swing..Jazz"
## [1] "Rock.n.roll"
## [1] "Alternative"
## [1] "Latino"
## [1] "Techno..Trance"
```

```
for(column_name in names(pitanja[137:140]))
  print(column_name)
```

```
## [1] "Entertainment.spending"
## [1] "Spending.on.looks"
## [1] "Spending.on.gadgets"
## [1] "Spending.on.healthy.eating"
```

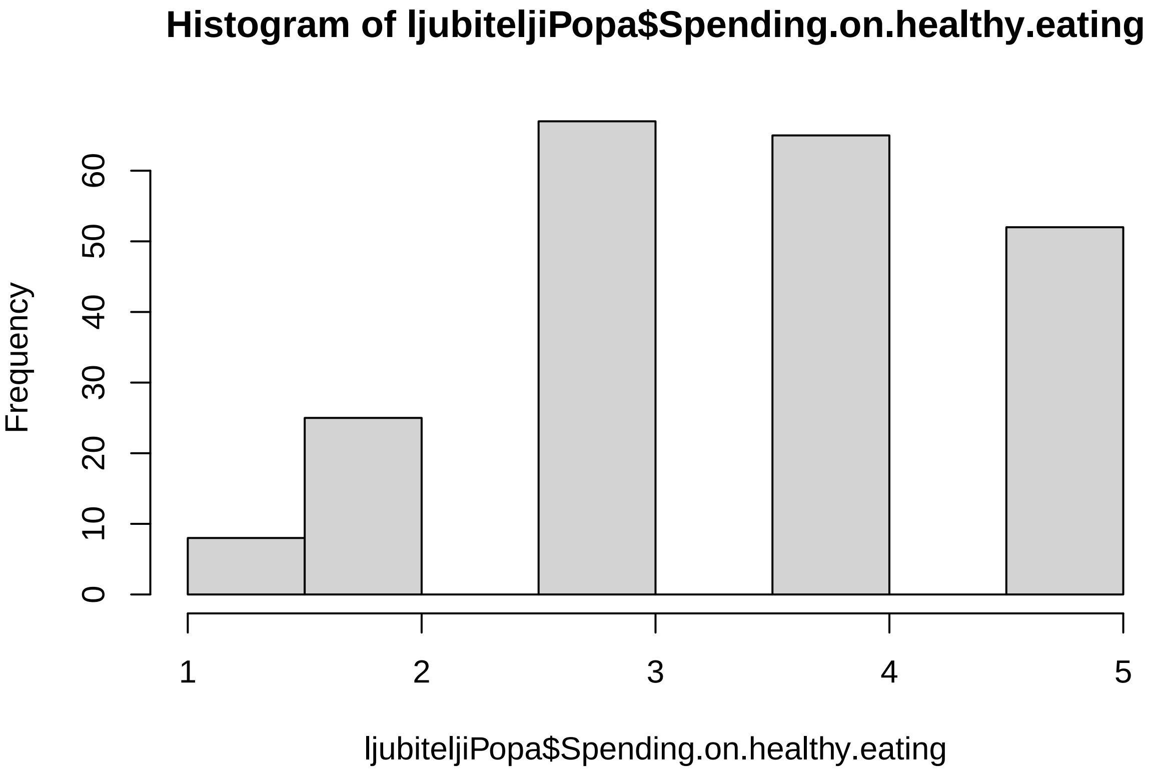
Kako bismo dokučili postoji li veza između ulaznih varijabli, preferencija određene glazbe ili možda čak više njih i izlazne varijable, tj. obrasca potrošnje upotrijebit ćemo upravo linearnu regresiju. Upravo nam ona odgovara na pitanje koje ulazne varijable najviše utječu na izlaznu te, posljedično, možemo li predvidjeti izlaz za pojedine vrijednosti ulaznih varijabli.

Linearan model ima slijedeće pretpostavke:

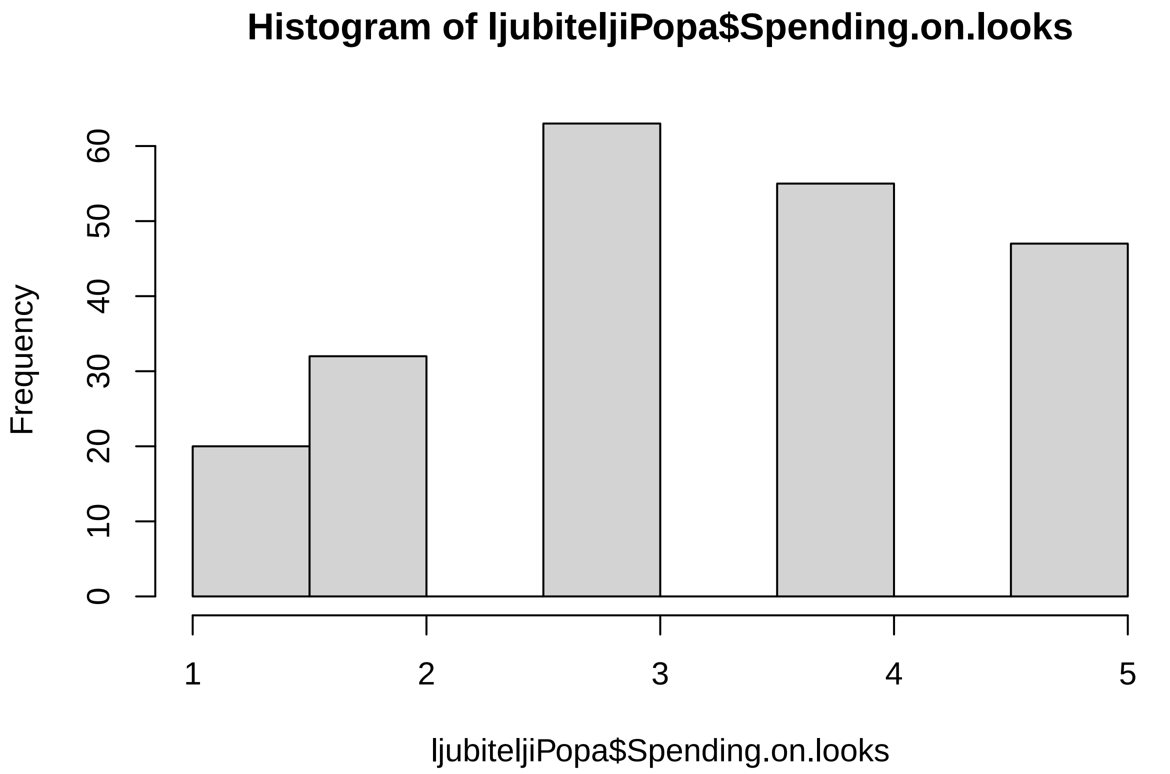
- linearnost veze X i Y
- pogreške nezavisne, homogene i normalno distribuirane s $\epsilon \sim \mathcal{N}(0, \sigma^2)$

Općenito, promatranje utjecaja pojedine nezavisne varijable na neku zavisnu, moguće he grafički dobiti dobar dojam o njihovom odnosu, ali kako su naši podaci ocijene od 1 do 5 za sve naše varijable, takav pristup ne bi imao previše smisla. Umjesto toga, iz svakog smo žanra glazbe izdvojili one sudionike koji su ga označili odličnom ocjenom te smo za takve pogledali kakve imaju obrasce potrošnje. Nećemo ovdje prolaziti kroz sve opcije jer ih ima 60, nego ćemo samo izdvojiti one koje smo smatrali da pokazuju upravo sklonost određenoj vrsti potrošnje i prokomentirati zaključke.

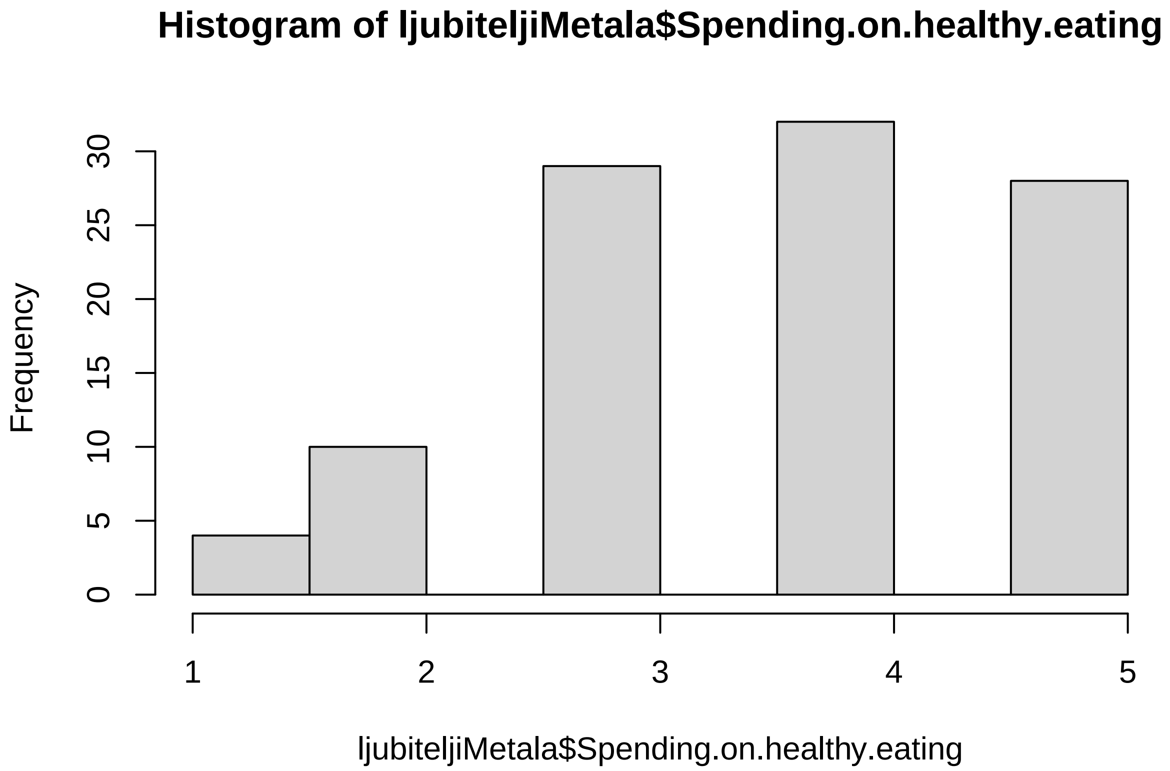
```
ljubiteljiPopa = pitanja[pitanja$Pop == 5, ]
hist(ljubiteljiPopa$Spending.on.healthy.eating)
```



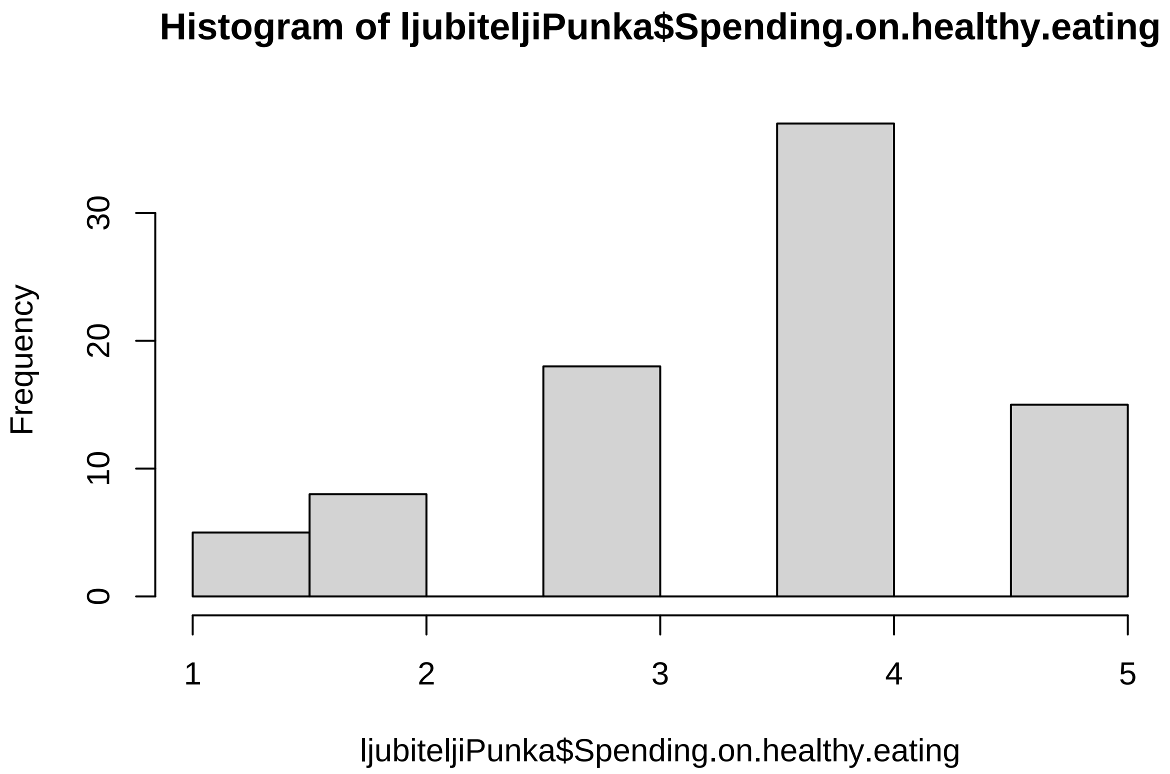
```
hist(ljubiteljiPopa$Spending.on.looks)
```



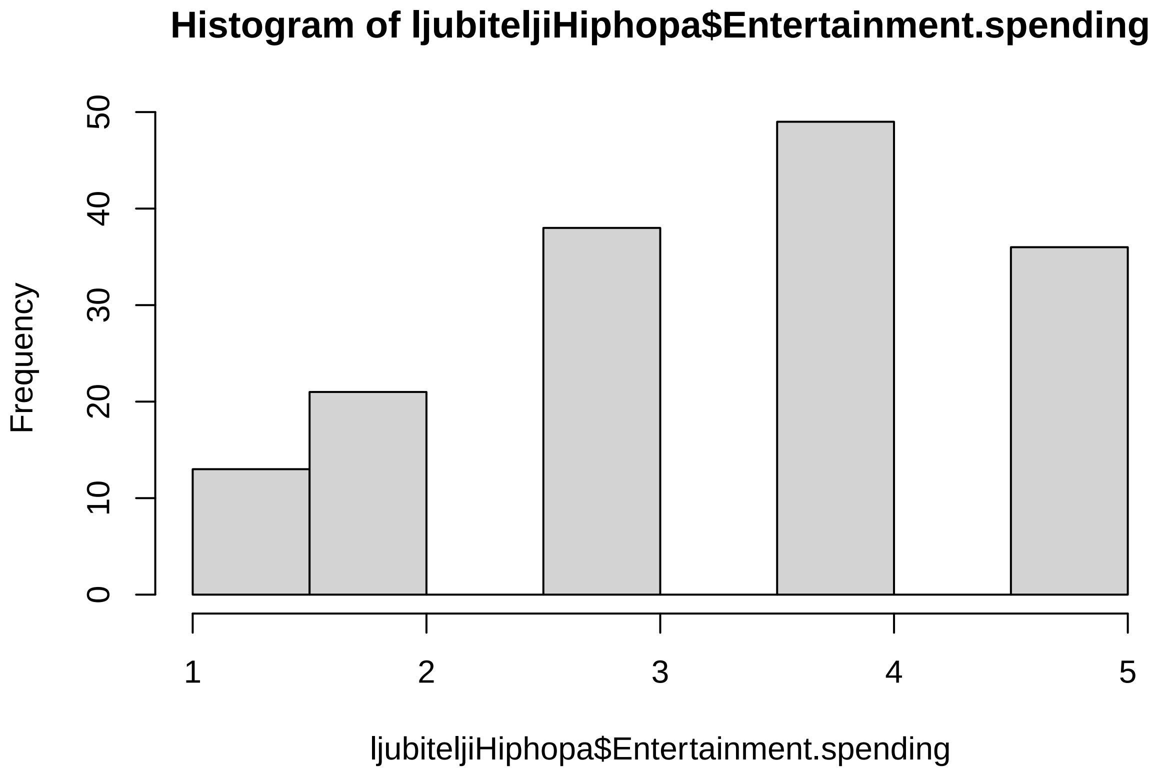
```
ljubiteljiMetala = pitanja[pitanja$Metal.or.Hardrock == 5, ]
hist(ljubiteljiMetala$Spending.on.healthy.eating)
```



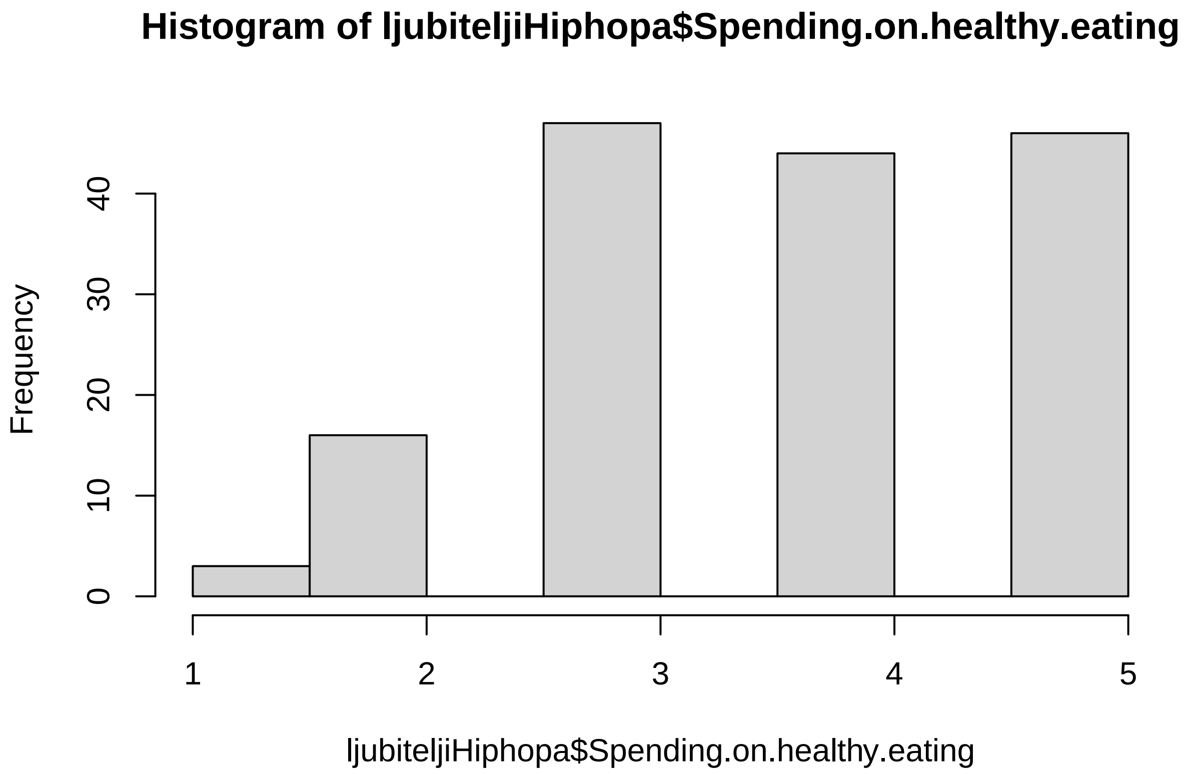
```
ljubiteljiPunka = pitanja[pitanja$Punk == 5, ]
hist(ljubiteljiPunka$Spending.on.healthy.eating)
```



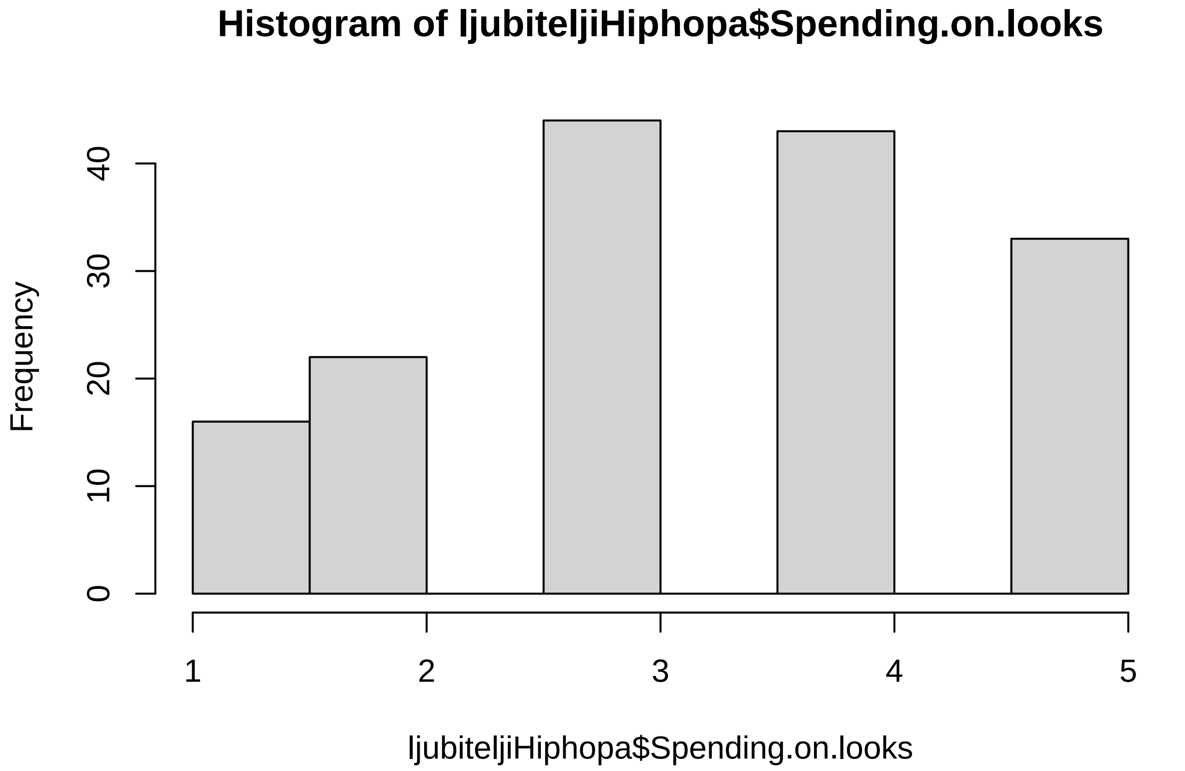
```
ljubiteljiHiphopa = pitanja[pitanja$Hiphop..Rap == 5, ]
hist(ljubiteljiHiphopa$Entertainment.spending)
```



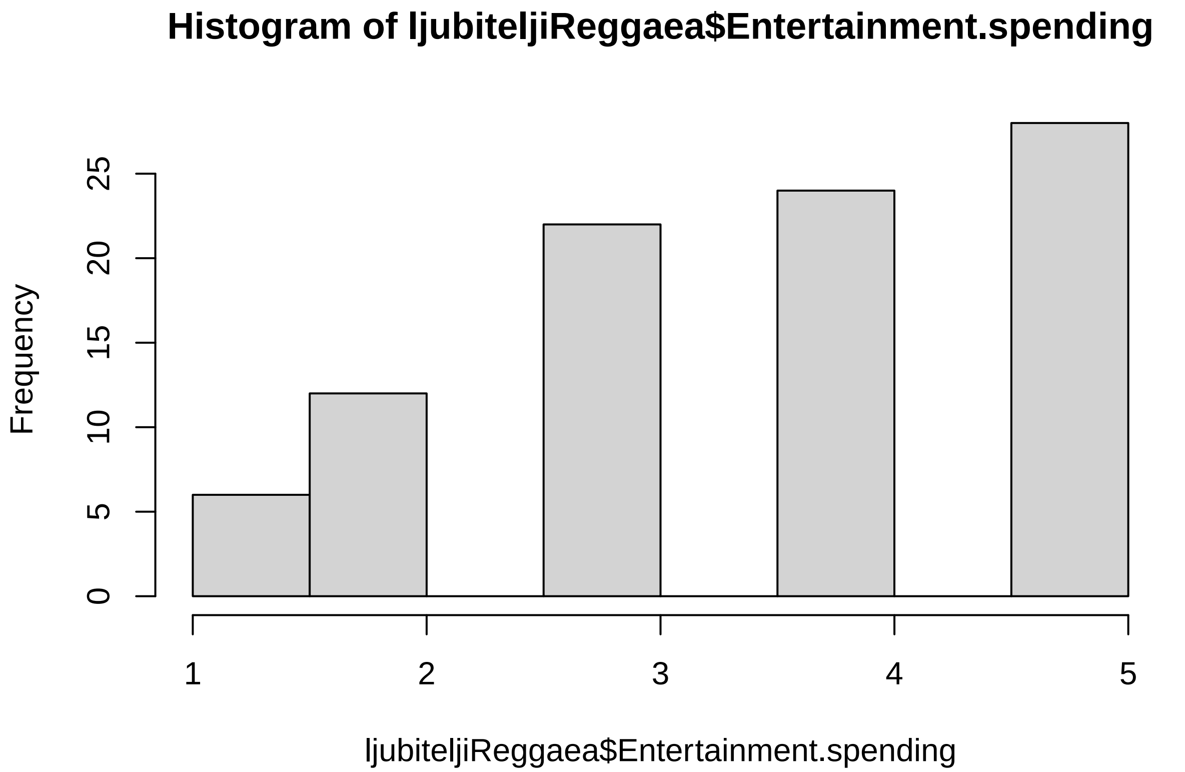
```
hist(ljubiteljiHiphopa$Spending.on.healthy.eating)
```



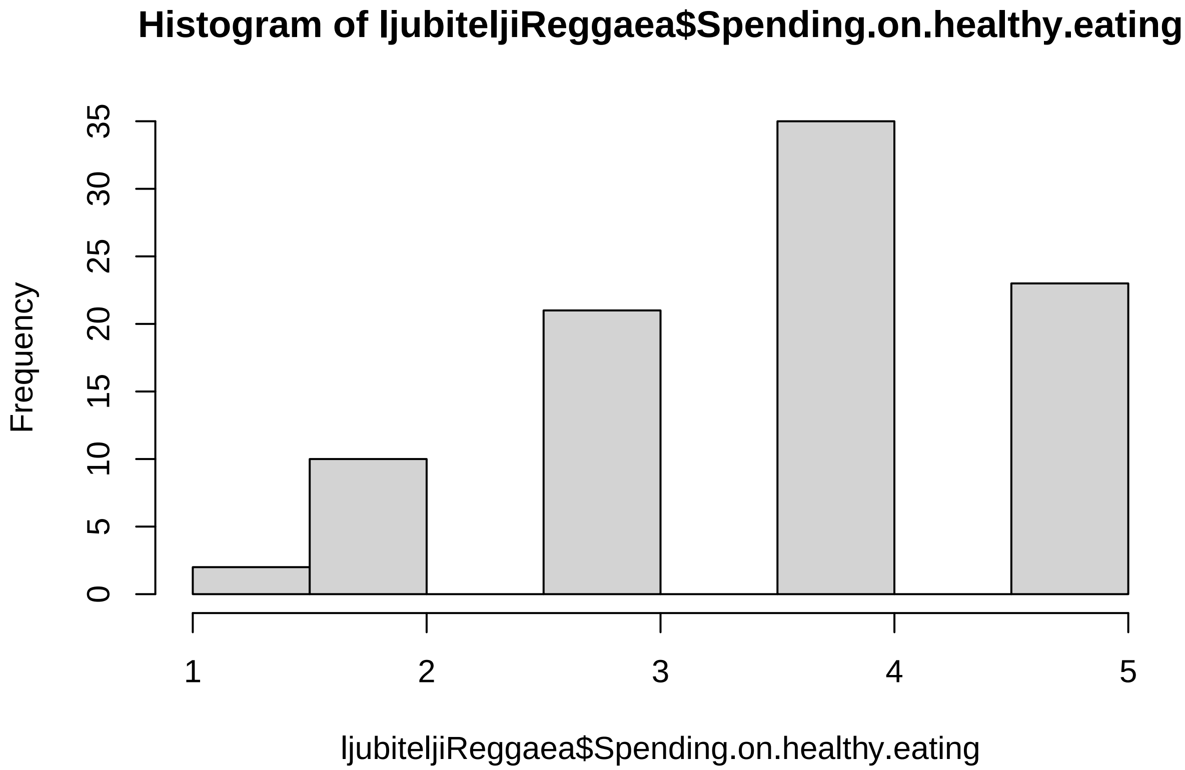
```
hist(ljubiteljiHiphopa$Spending.on.looks)
```



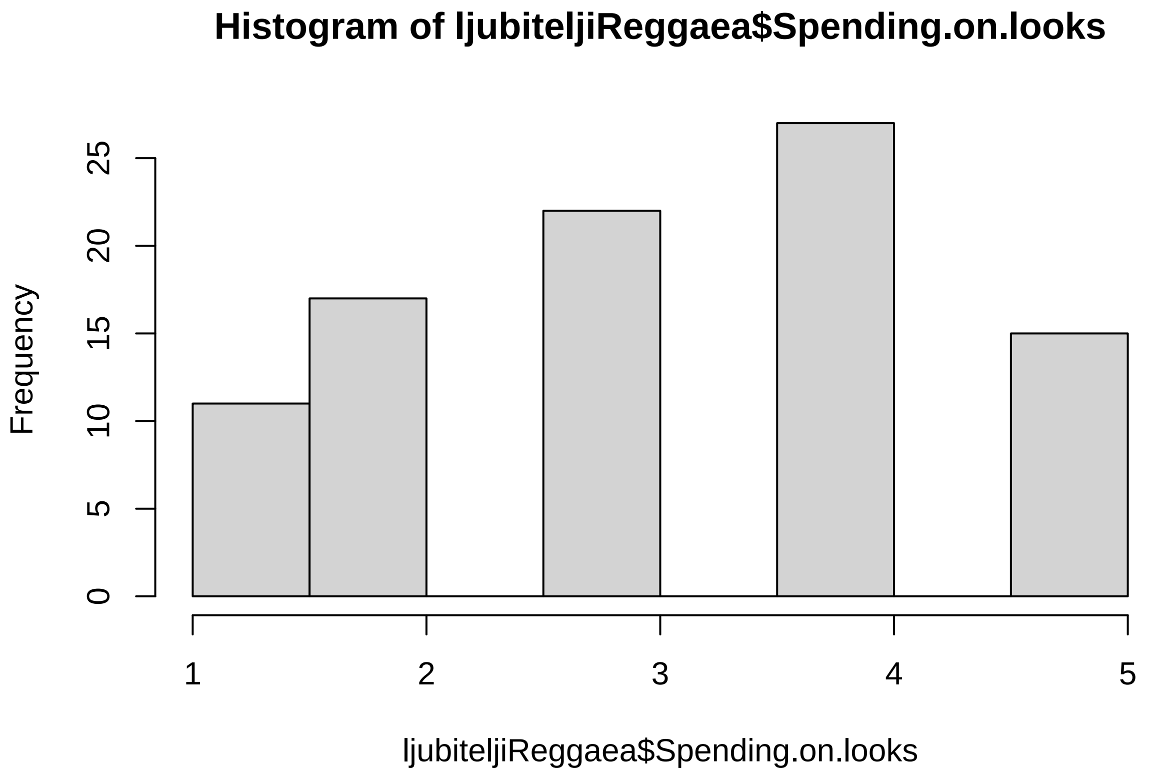
```
ljubiteljiReggaea = pitanja[pitanja$Reggae..Ska == 5, ]
hist(ljubiteljiReggaea$Entertainment.spending)
```



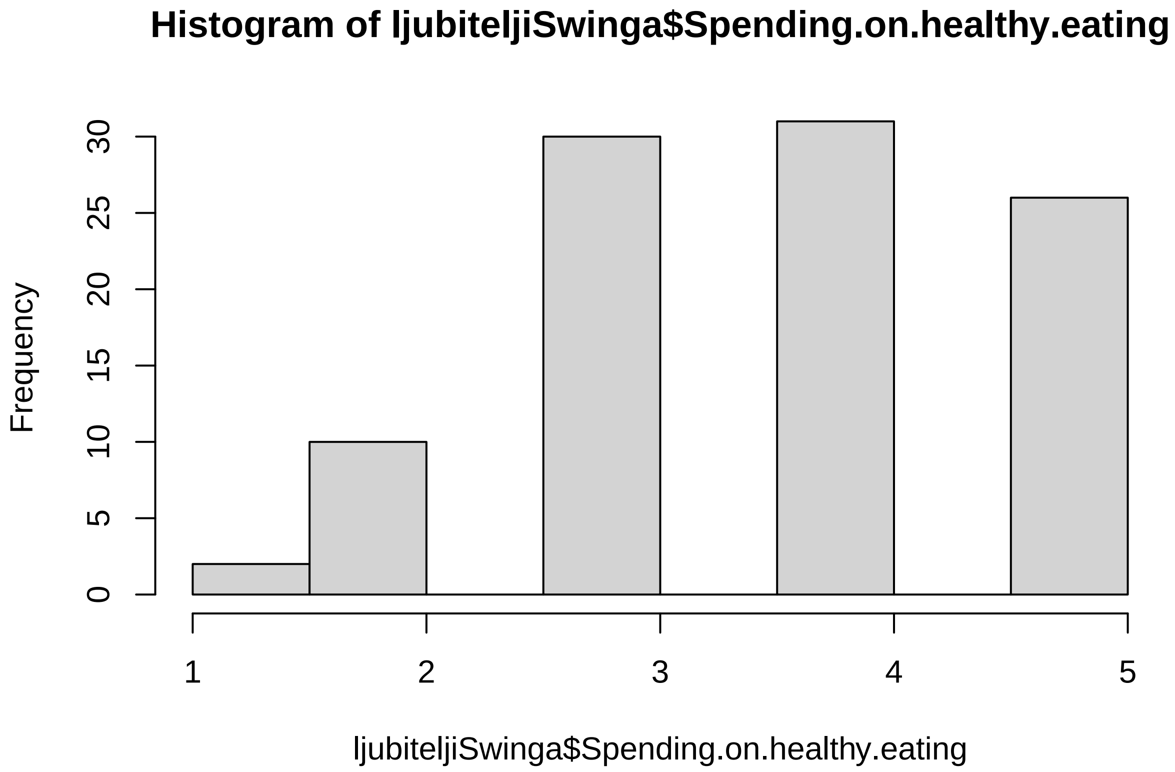
```
hist(ljubiteljiReggaea$Spending.on.healthy.eating)
```



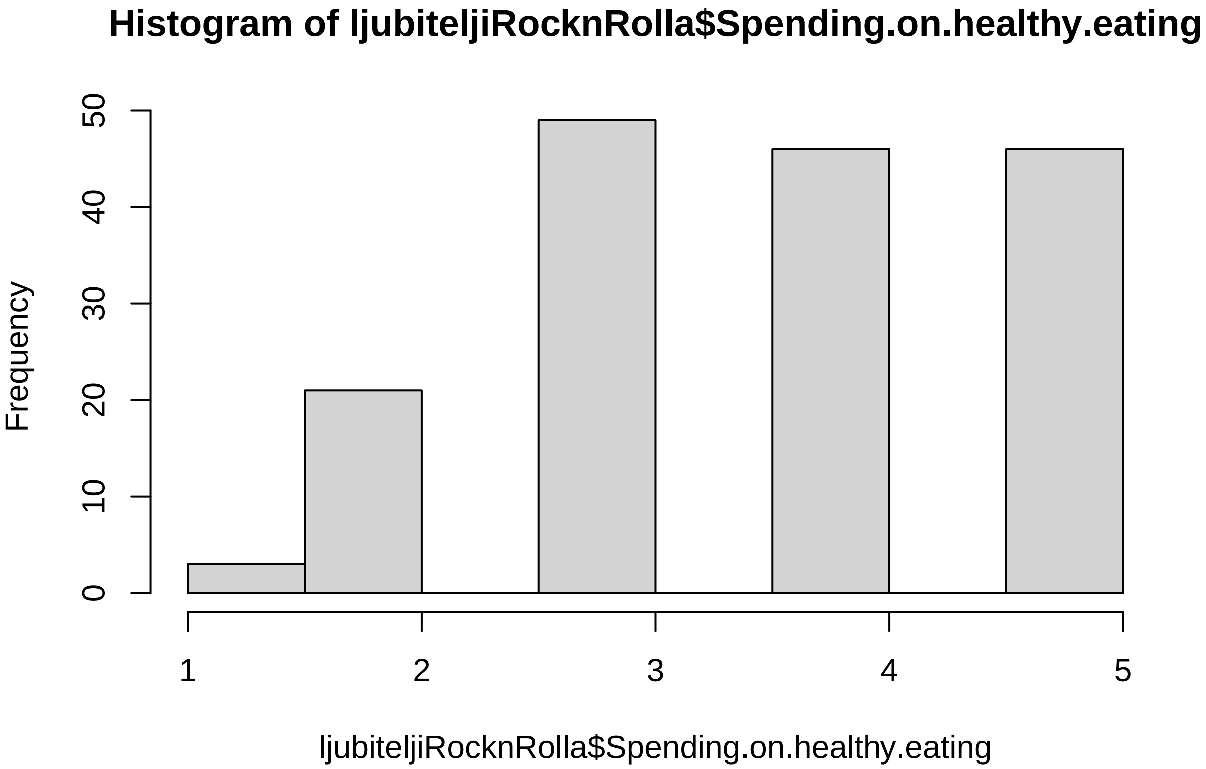
```
hist(ljubiteljiReggaea$Spending.on.looks)
```



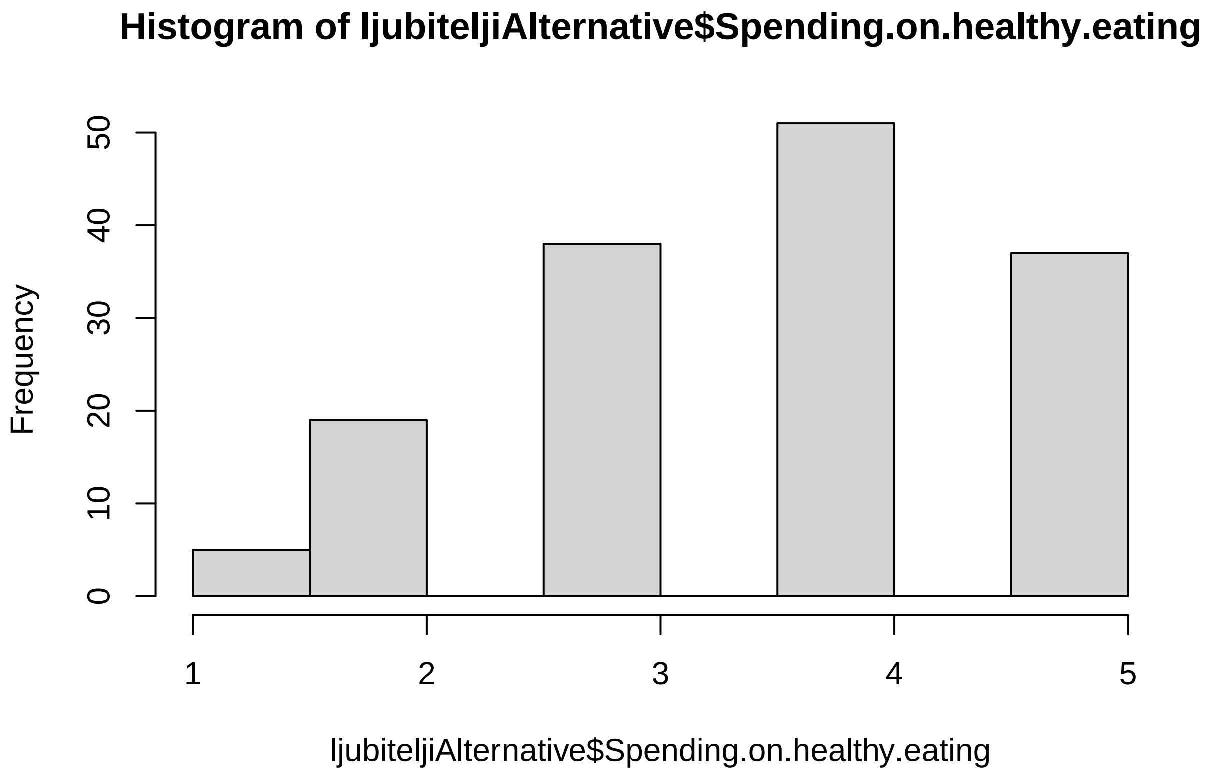
```
ljubiteljiSwinga = pitanja[pitanja$Swing..Jazz == 5, ]
hist(ljubiteljiSwinga$Spending.on.healthy.eating)
```



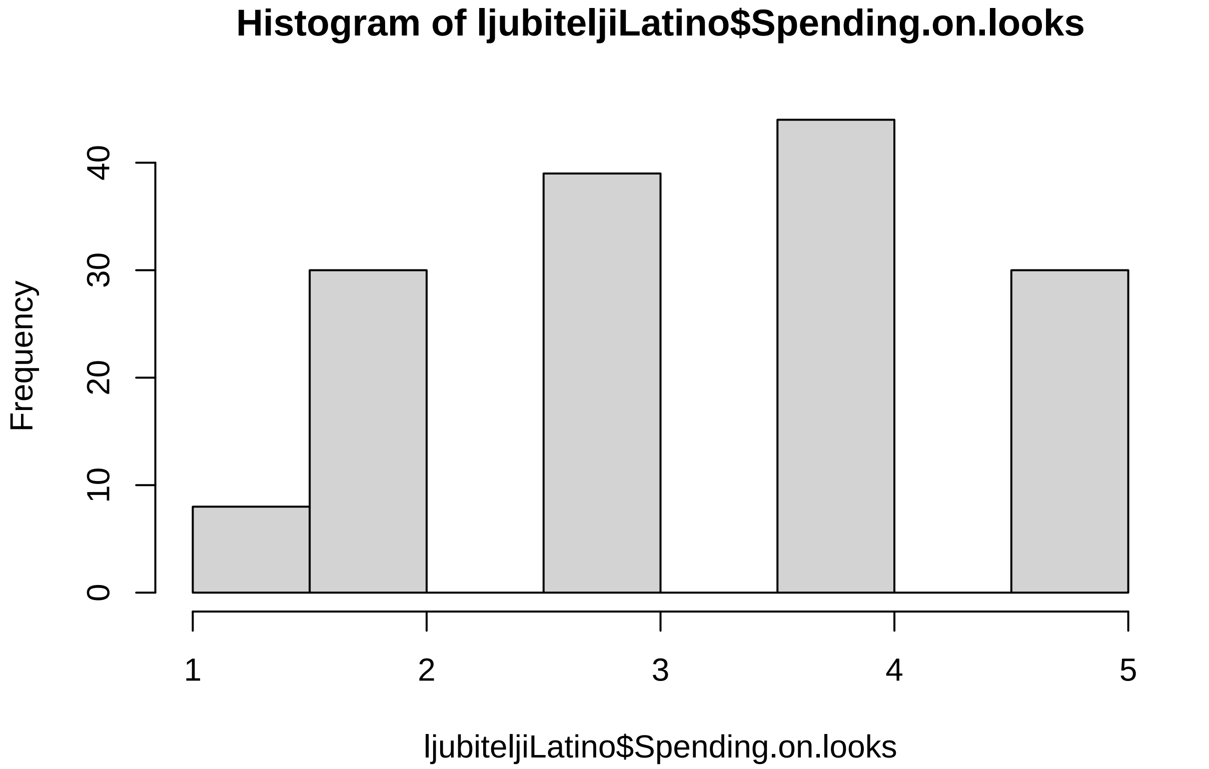
```
ljubiteljiRocknRolla = pitanja[pitanja$Rock.n.roll == 5, ]
hist(ljubiteljiRocknRolla$Spending.on.healthy.eating)
```



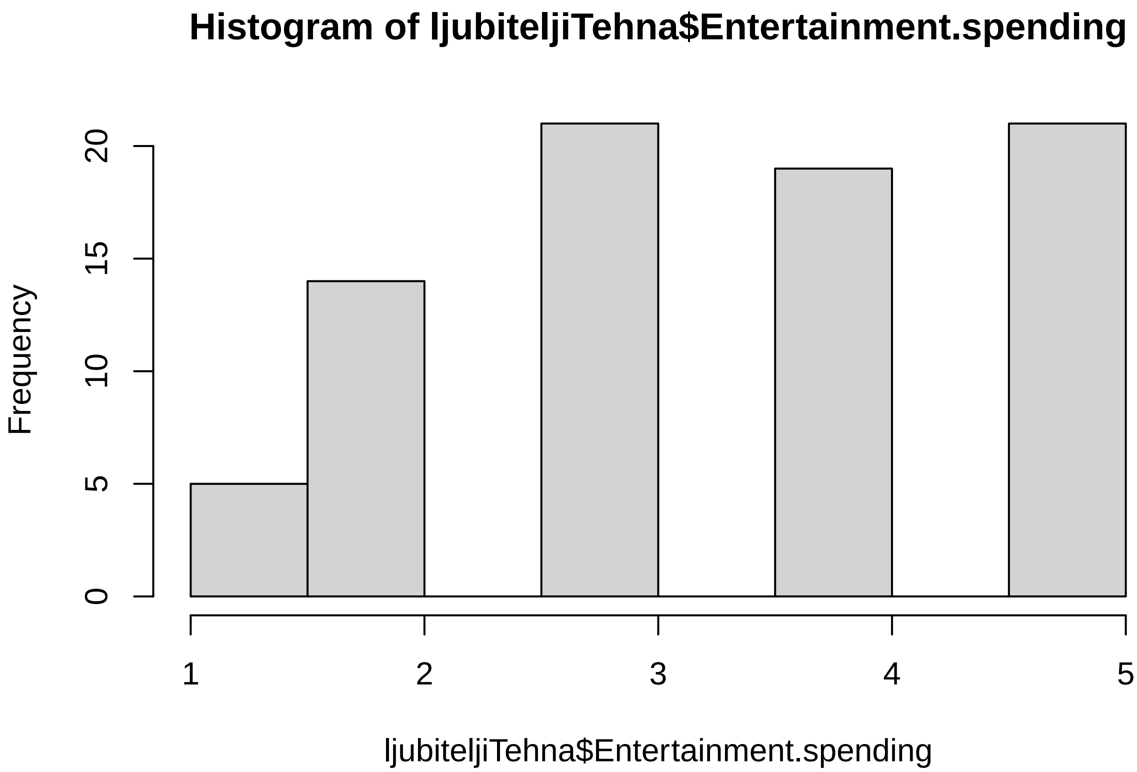
```
ljubiteljiAlternative = pitanja[pitanja$Alternative == 5, ]
hist(ljubiteljiAlternative$Spending.on.healthy.eating)
```



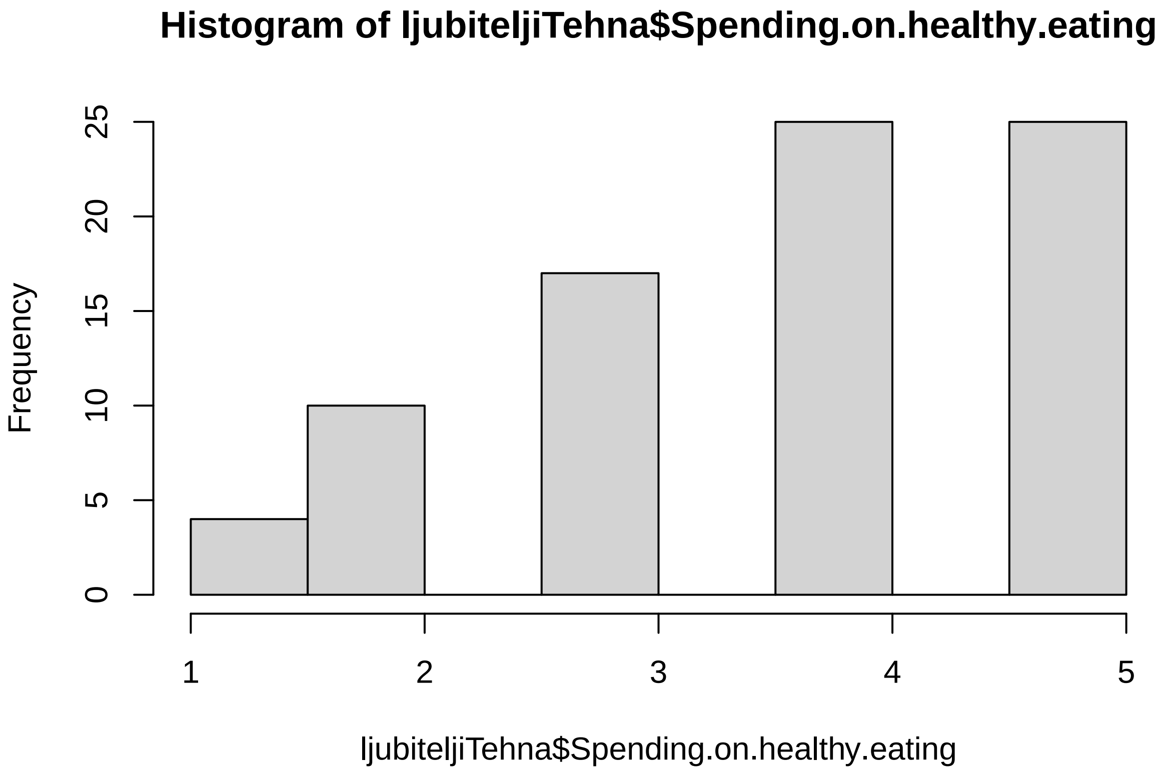
```
ljubiteljiLatino = pitanja[pitanja$Latino == 5, ]
hist(ljubiteljiLatino$Spending.on.looks)
```



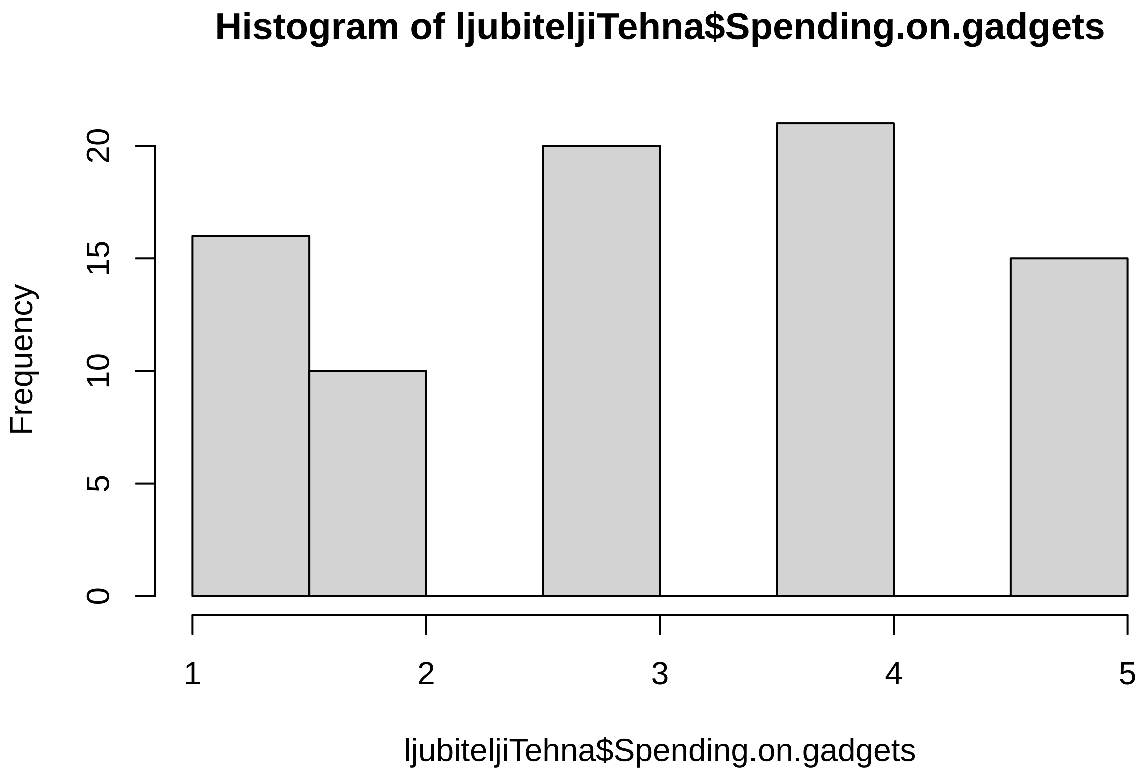
```
ljubiteljiTehna = pitanja[pitanja$Techno..Trance == 5, ]
hist(ljubiteljiTehna$Entertainment.spending)
```



```
hist(ljubiteljiTehna$Spending.on.healthy.eating)
```



```
hist(ljubiteljiTehna$Spending.on.gadgets)
```



```
nrow(ljubiteljiMetala)
```

```
## [1] 107
```

```
nrow(ljubiteljiPopa)
```

```
## [1] 221
```

```
nrow(ljubiteljiPunka)
```

```
## [1] 92
```

```
nrow(ljubiteljiHiphopa)
```

```
## [1] 162
```

```
nrow(ljubiteljiReggaea)
```

```
## [1] 99
```

```
nrow(ljubiteljiSwinga)
```

```
## [1] 106
```

```
nrow(ljubiteljiRocknRolla)
```

```
## [1] 174
```

```
nrow(ljubiteljiAlternative)
```

```
## [1] 158
```

```
nrow(ljubiteljiLatino)
```

```
## [1] 159
```

```
nrow(ljubiteljiTehna)
```

```
## [1] 89
```

Pogledavši histograme kojima je mod težio prema višim ocjenama za pojedini obrazac potrošnje, ili barem preko trojke, odlučili smo upravo te žanrove uvrstiti kao nezavisne varijable u predviđanju pojedinog obrasca potrošnje.

```
potrosnjazdravahrana = pitanja[complete.cases(pitanja[, c('Spending.on.healthy.eating', 'Pop', 'Metal.o

fit.spending.on.healthy.eating = lm(Spending.on.healthy.eating ~ Pop + Metal.or.Hardrock + Punk + Hiphop
summary(fit.spending.on.healthy.eating)
```

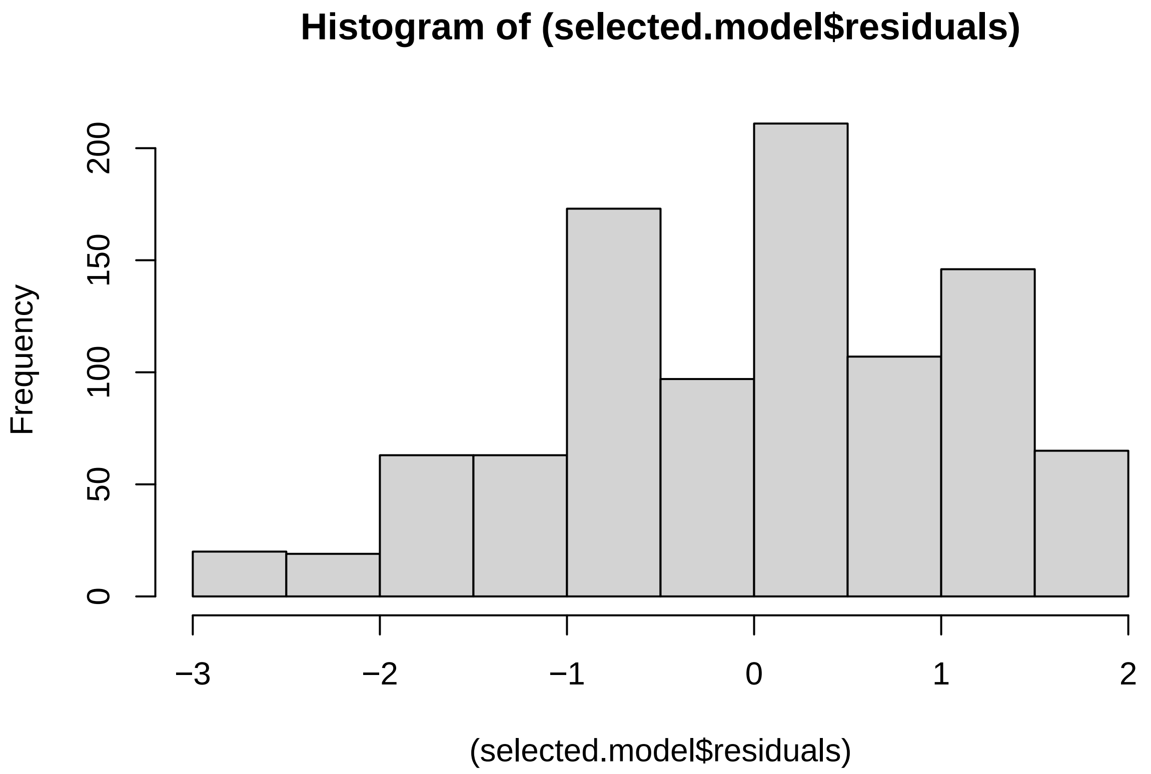
```
##
## Call:
## lm(formula = Spending.on.healthy.eating ~ Pop + Metal.or.Hardrock +
##     Punk + Hiphop..Rap + Reggae..Ska + Swing..Jazz + Rock.n.roll +
##     Alternative + Techno..Trance, data = potrosnjazdravahrana)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8166 -0.6339  0.2622  0.6329  1.9639
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        3.224571   0.187474  17.200  < 2e-16 ***
## Pop               -0.059094   0.033038  -1.789  0.07398 .
## Metal.or.Hardrock  0.011499   0.031977   0.360  0.71923
## Punk              -0.042627   0.034870  -1.222  0.22184
## Hiphop..Rap        0.093711   0.029545   3.172  0.00156 **
## Reggae..Ska       -0.008562   0.034056  -0.251  0.80154
## Swing..Jazz        0.070227   0.033622   2.089  0.03699 *
## Rock.n.roll        0.005340   0.035032   0.152  0.87888
## Alternative        0.029140   0.030602   0.952  0.34123
## Techno..Trance     0.033562   0.027977   1.200  0.23058
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.08 on 954 degrees of freedom
## Multiple R-squared:  0.02658,    Adjusted R-squared:  0.01739
## F-statistic: 2.894 on 9 and 954 DF,  p-value: 0.00221
```

Rezultati nisu ni blizu onima koje smo očekivali, R^2 je izrazito malen, što ukazuje na to da nije moguće predvidjeti tko će potrošiti više na zdravi prehranu na temelju glazbe koje sluša. Ali, vrijednost F-testa uz stupnjeve slobode 9, 954 ipak upućuje da je model značajan jer prelazi vrijednost od 1.88. Prije nego pokušamo još neku kombinaciju provjerimo normalnost dobivenih reziduala i homogenost varijance.

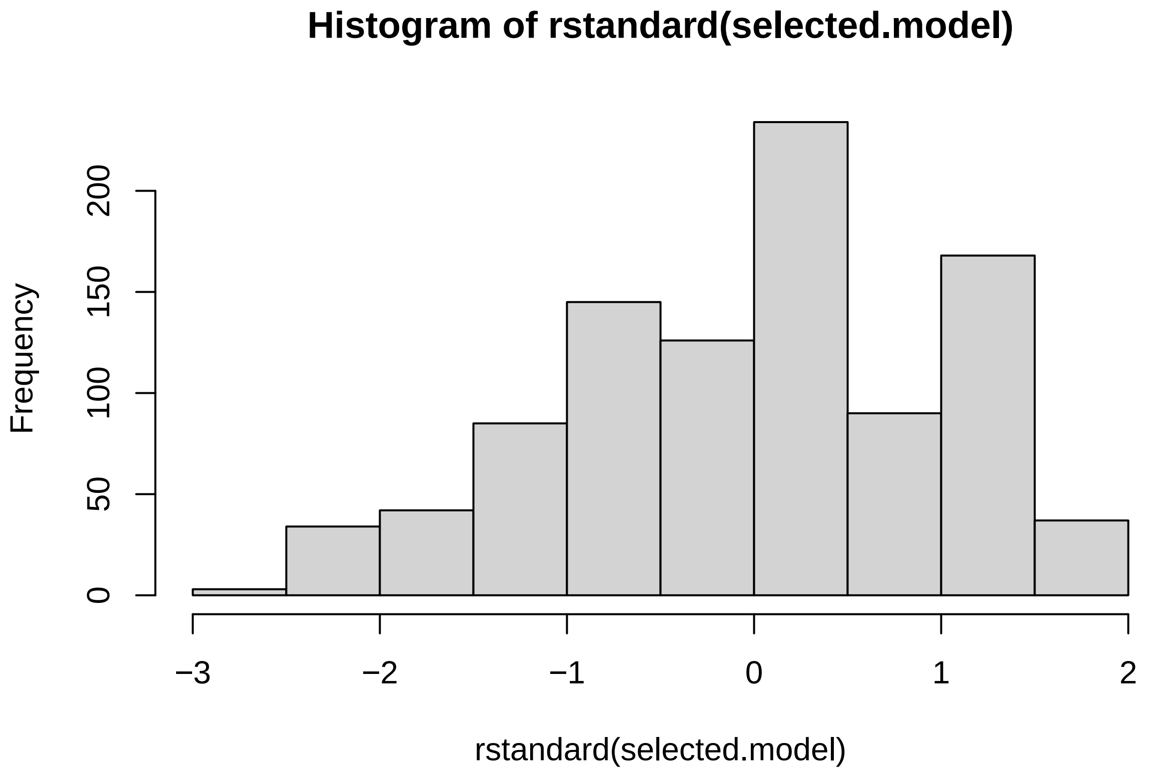
Normalnost reziduala i homogenost varijance

Normalnost reziduala moguće je provjeriti grafički, pomoću kvantil-kvantil plota (usporedbom s linijom normalne razdiobe), te statistički pomoću Kolmogorov-Smirnovljevog testa.

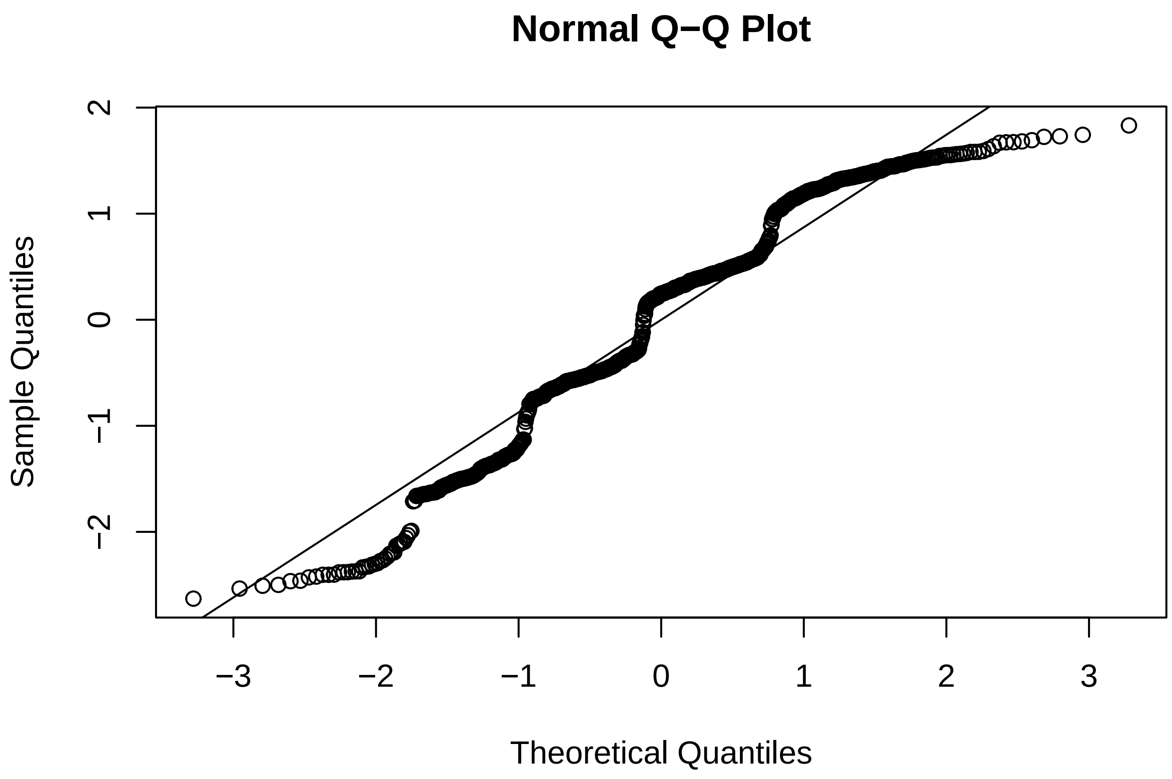
```
selected.model = fit.spending.on.healthy.eating

hist((selected.model$residuals))
```



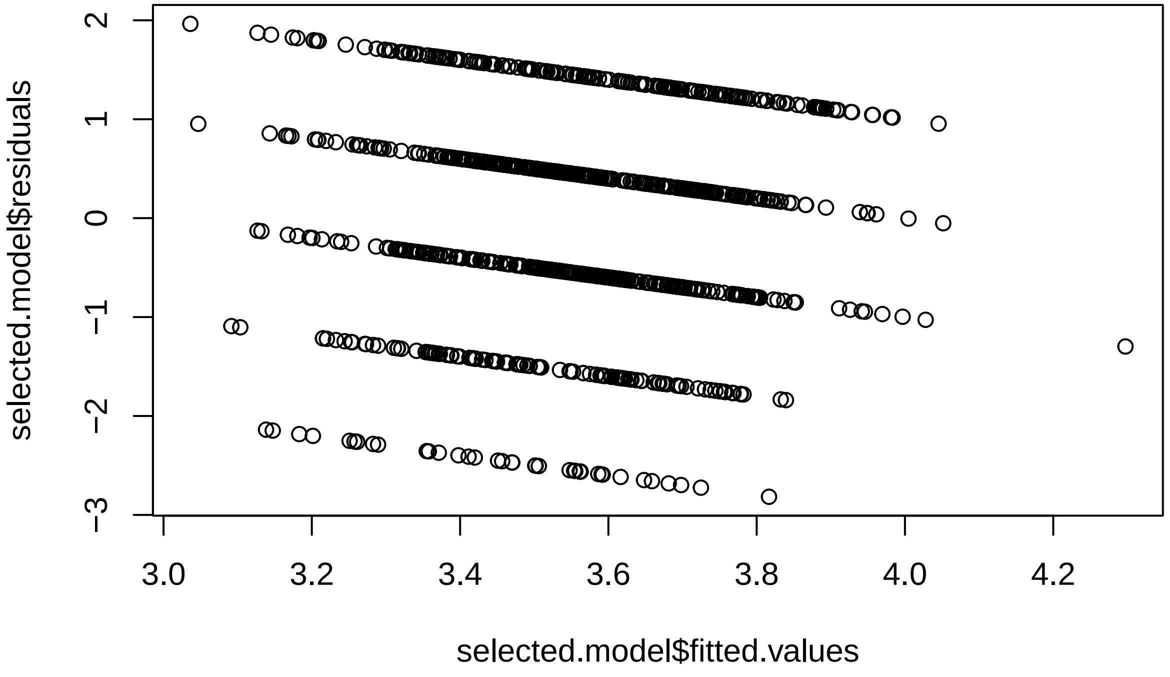
```
hist(rstandard(selected.model))
```



```
#q-q plot reziduala s linijom normalne distribucije
qqnorm(rstandard(selected.model))
qqline(rstandard(selected.model))
```



```
#reziduale je dobro prikazati u ovisnosti o procjenama modela
plot(selected.model$fitted.values,selected.model$residuals)
```



```
ks.test(rstandard(fit.spending.on.healthy.eating),'pnorm')
```

```
## Warning in ks.test(rstandard(fit.spending.on.healthy.eating), "pnorm"): ties
## should not be present for the Kolmogorov-Smirnov test

##
```

```
##  One-sample Kolmogorov-Smirnov test
##
## data:  rstandard(fit.spending.on.healthy.eating)
## D = 0.10307, p-value = 2.548e-09
## alternative hypothesis: two-sided
```

```
require(nortest)
```

```
## Loading required package: nortest
```

```
lillie.test(rstandard(fit.spending.on.healthy.eating))
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstandard(fit.spending.on.healthy.eating)
## D = 0.10296, p-value < 2.2e-16
```

```
ad.test(rstandard(fit.spending.on.healthy.eating))
```

```
##
##  Anderson-Darling normality test
##
## data:  rstandard(fit.spending.on.healthy.eating)
## A = 10.121, p-value < 2.2e-16
```

Grafički, iz histograma možemo zaključiti da se naša distribucija razlikuje od normalne, što dodatno potvrđuje q-q plot koji je dosta zakrivljen i ima "prekide".

Prokomentirajmo sada KS test; ovdje imamo dva problema, Kolmogorov-Smirnovljev test je test za kontinuiranu distribuciju i stoga skup podataka ne bi trebao sadržavati nikakve veze, odnsono ponovljene vrijednosti, što je upravo slučaj u našem skupu podataka (ocijene 1 do 5). Umjesto toga potrebno je koristiti Lilliefors test, kao što smo i učinili, ali smo posegnuli i za jačom alternativom kada je u pitanju točnost na ovakvom skupu podataka, a to je Anderson-Darling test na normalnost koji nam je doduše dao istu p vrijednost, ali vrijednost statistike A svakako je veća. Izvedimo sada zaključak o normlanosti na temelju Anderson-Darlingovog testa;

Test odbacuje hipotezu normalnosti jer je p-vrijednost manja od 2.2e-16, dok je dovoljan uvjet za odbacivanje normalnosti p-vrijednost manja ili jednaka 0,05. Zaključujemo dakle da naši reziduali nisu normalno distribuirani.

Korelacijski koeficijent

Zanimalo nas je, je li možda uzrok lošim rezultatima modela velika korelacija među varijablama pa smo stoga proveli slijedeći testove.

```
cor(cbind(potrosnjazdravahrana$Spending.on.healthy.eating, potrosnjazdravahrana$Pop, potrosnjazdravahran
```

```
##               [,1]         [,2]         [,3]        [,4]        [,5]       [,6]
##  [1,]  1.000000000 -0.029642841  0.002522815 -0.02689210  0.10316969 0.04728458
##  [2,] -0.029642841  1.000000000 -0.291497609 -0.15265410  0.28661060 0.02002766
##  [3,]  0.002522815 -0.291497609  1.000000000  0.54054236 -0.20198184 0.11448128
##  [4,] -0.026892102 -0.152654098  0.540542357  1.00000000 -0.08893575 0.29458762
##  [5,]  0.103169690  0.286610597 -0.201981843 -0.08893575  1.00000000 0.28415213
##  [6,]  0.047284581  0.020027655  0.114481280  0.29458762  0.28415213 1.00000000
##  [7,]  0.088147967 -0.028885665  0.145434404  0.10722263 -0.01617556 0.33777469
##  [8,]  0.026929998 -0.003076951  0.298359145  0.32180290 -0.11778341 0.23563368
##  [9,]  0.045391114 -0.211581284  0.292688550  0.34463663 -0.15285520 0.19195259
## [10,]  0.065639533  0.155471085 -0.053960271 -0.08788903  0.29072552 0.05140808
##              [,7]         [,8]         [,9]        [,10]
##  [1,]  0.08814797  0.026929998  0.045391114  0.065639533
##  [2,] -0.02888567 -0.003076951 -0.211581284  0.155471085
##  [3,]  0.14543440  0.298359145  0.292688550 -0.053960271
##  [4,]  0.10722263  0.321802902  0.344636628 -0.087889026
##  [5,] -0.01617556 -0.117783412 -0.152855196  0.290725525
##  [6,]  0.33777469  0.235633681  0.191952587  0.051408082
##  [7,]  1.00000000  0.471473652  0.335475527 -0.023470326
##  [8,]  0.47147365  1.000000000  0.388549620 -0.080696203
##  [9,]  0.33547553  0.388549620  1.000000000 -0.005626447
## [10,] -0.02347033 -0.080696203 -0.005626447  1.000000000
```

Rezultati pokuzauju kako korelacija među varijablama nije uzrok problema, odnsono nemogućnosti predviđanja pa naše rješenje neće biti nestabilno zbog korelacija jer nigdje nemamo visoki koeficijent korelacije. Zaključno, ne možemo kvalitetno donijeti zaključak ljubitelji kojih glazbenih žanrova će će kako trošiti na zdravu hranu.

Kako se ova bilježnica ne bi odveć oduljila, za druge ćemo kategorije samo provesti linearnu regresiju prema gore navedenom principu i komentirati vrijednosti koeficijenta korelacije iR^2; ukoliko bude potrebno, na ispitivanju možemo predočiti dulju verziju bilježnice.

```
potrosnjaizgled = pitanja[complete.cases(pitanja[, c('Spending.on.looks', 'Pop', 'Hiphop..Rap', 'Reggae

cor(cbind(potrosnjaizgled$Spending.on.looks, potrosnjaizgled$Pop, potrosnjaizgled$Hiphop..Rap, potrosnja
```

```
##             [,1]       [,2]      [,3]        [,4]       [,5]
## [1,] 1.000000000 0.16995614 0.2267351 0.006597866 0.05948717
## [2,] 0.169956143 1.00000000 0.2844186 0.022186180 0.29911584
## [3,] 0.226735082 0.28441856 1.0000000 0.282849840 0.14282992
## [4,] 0.006597866 0.02218618 0.2828498 1.000000000 0.19079113
## [5,] 0.059487174 0.29911584 0.1428299 0.190791128 1.00000000
```

```
fit.spending.on.looks = lm(Spending.on.looks ~ Pop + Hiphop..Rap + Reggae..Ska + Latino, data = potrosnj
summary(fit.spending.on.healthy.eating)
```

```
## 
## Call:
## lm(formula = Spending.on.healthy.eating ~ Pop + Metal.or.Hardrock + 
##     Punk + Hiphop..Rap + Reggae..Ska + Swing..Jazz + Rock.n.roll + 
##     Alternative + Techno..Trance, data = potrosnjazdravahrana)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -2.8166 -0.6339  0.2622  0.6329  1.9639 
## 
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)    
## (Intercept)        3.224571   0.187474  17.200  < 2e-16 ***
## Pop               -0.059094   0.033038  -1.789  0.07398 .  
## Metal.or.Hardrock  0.011499   0.031977   0.360  0.71923    
## Punk              -0.042627   0.034870  -1.222  0.22184    
## Hiphop..Rap        0.093711   0.029545   3.172  0.00156 ** 
## Reggae..Ska       -0.008562   0.034056  -0.251  0.80154    
## Swing..Jazz        0.070227   0.033622   2.089  0.03699 *  
## Rock.n.roll        0.005340   0.035032   0.152  0.87888    
## Alternative        0.029140   0.030602   0.952  0.34123    
## Techno..Trance     0.033562   0.027977   1.200  0.23058
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.08 on 954 degrees of freedom
## Multiple R-squared:  0.02658,    Adjusted R-squared:  0.01739
## F-statistic: 2.894 on 9 and 954 DF,  p-value: 0.00221
```

Pri predviđanju potrošenja na izgled dobili smo istu vrijednost R^2 kao i za potrošnju na zdravu hranu. a F-statistika nam opet ukazuje da je model značajan. Ovdje, također, ne opažamo velike korelacije pa možemo ustvrditi da one ne smanjuju kvalitetu naše predikcije. Zaključno, ne možemo kvalitetno donijeti zaključak ljubitelji kojih glazbenih žanrova će kako trošiti na izgled.

```
potrosnjazabava = pitanja[complete.cases(pitanja[, c('Entertainment.spending', 'Reggae..Ska', 'Hiphop..

cor(cbind(potrosnjazabava$Entertainment.spending, potrosnjazabava$Reggae..Ska, potrosnjazabava$Hiphop..
```

```
##           [,1]      [,2]      [,3]
## [1,] 1.0000000 0.1220722 0.1337346
## [2,] 0.1220722 1.0000000 0.2781312
## [3,] 0.1337346 0.2781312 1.0000000
```

```
fit.entertainment.spending = lm(Entertainment.spending ~ Pop + Hiphop..Rap + Reggae..Ska,data=potrosnja
summary(fit.entertainment.spending)
```

```
##
## Call:
## lm(formula = Entertainment.spending ~ Pop + Hiphop..Rap + Reggae..Ska,
##     data = potrosnjazabava)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.80701 -0.97459 -0.07538  0.84302  2.26331
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.89577    0.14590  19.848  < 2e-16 ***
## Pop         -0.07111    0.03357  -2.118  0.03441 *
## Hiphop..Rap  0.11132    0.02949   3.774  0.00017 ***
## Reggae..Ska  0.08515    0.03198   2.663  0.00787 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.174 on 990 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.02975,    Adjusted R-squared:  0.02681
## F-statistic: 10.12 on 3 and 990 DF,  p-value: 1.441e-06
```

Ponovno smo došli do sličnih opservacija, pri predviđanju potrošenja na zabavu dobili smo istu vrijednost R^2 nešto veću od dosadašnjih, a F-statistika je ovaj puta također veća i veća od granične, koja iznosi 2.6 pa ukazuje da je model značajan. Ovdje, također, ne opažamo velike korelacije pa možemo ustvrditi da one ne smanjuju kvalitetu naše predikcije. Zaključno, ne možemo kvalitetno donijeti zaključak ljubitelji kojih glazbenih žanrova će kako trošiti na izgled.

```
potrosnjatehnologija = pitanja[complete.cases(pitanja[, c('Spending.on.gadgets', 'Techno..Trance')]),]

cor(potrosnjatehnologija$Spending.on.gadgets, potrosnjatehnologija$Hiphop..Rap)
```

```
## [1] NA
```

```
cor.test(potrosnjatehnologija$Spending.on.gadgets, potrosnjatehnologija$Techno..Trance)
```

```
##
##  Pearson's product-moment correlation
##
## data:  potrosnjatehnologija$Spending.on.gadgets and potrosnjatehnologija$Techno..Trance
## t = 4.9594, df = 1001, p-value = 8.303e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.0938605 0.2147032
## sample estimates:
##      cor
## 0.154861
```

```
fit.spending.on.gadgets = lm(Spending.on.gadgets ~ Techno..Trance,data=potrosnjatehnologija)
summary(fit.spending.on.gadgets)
```

```
##
## Call:
## lm(formula = Spending.on.gadgets ~ Techno..Trance, data = potrosnjatehnologija)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.26742 -0.96677  0.03323  1.03323  2.33389
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     2.51579    0.08146  30.884  < 2e-16 ***
## Techno..Trance  0.15033    0.03031   4.959  8.3e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.27 on 1001 degrees of freedom
## Multiple R-squared:  0.02398,    Adjusted R-squared:  0.02301
## F-statistic:  24.6 on 1 and 1001 DF,  p-value: 8.303e-07
```

Došli smo tako i do posljednje kategorije, odnosno predviđanja potrošnje na tehnologiju u odnosu na glazbeni žanr.Vrijednost R^2 nešto je niža od posljednje, a F-statistika je ovaj puta još veća, čak i gledajući komparativu u odnosu na granicu koja sada iznosi 3.84 pa ukazuje da je model značajan. Ovdje, imamo korelaciju samo dviju varijabli i ona također nije značajna, što potvrđuje i Pearsonov test. Zaključno, ne možemo kvalitetno donijeti zaključak kako će trošiti na izgled.

```
fit.Spending.on.looks = lm(potrosnjaizgled$Spending.on.looks ~ factor(potrosnjaizgled$Hiphop..Rap) + fac
summary(fit.Spending.on.looks)
```

```
##
## Call:
## lm(formula = potrosnjaizgled$Spending.on.looks ~ factor(potrosnjaizgled$Hiphop..Rap) +
##     factor(potrosnjaizgled$Techno..Trance), data = potrosnjaizgled)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.48142 -0.92969  0.05407  0.82615  2.37644
##
## Coefficients:
```

```
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                               2.62356    0.08591  30.538  < 2e-16
## factor(potrosnjaizgled$Hiphop..Rap)2      0.30613    0.11779   2.599  0.00949
## factor(potrosnjaizgled$Hiphop..Rap)3      0.48354    0.12151   3.979 7.43e-05
## factor(potrosnjaizgled$Hiphop..Rap)4      0.77916    0.11970   6.509 1.21e-10
## factor(potrosnjaizgled$Hiphop..Rap)5      0.68035    0.13194   5.156 3.05e-07
## factor(potrosnjaizgled$Techno..Trance)2   0.06676    0.10588   0.630  0.52852
## factor(potrosnjaizgled$Techno..Trance)3   0.01624    0.10952   0.148  0.88216
## factor(potrosnjaizgled$Techno..Trance)4   0.07870    0.12247   0.643  0.52063
## factor(potrosnjaizgled$Techno..Trance)5   0.17213    0.15004   1.147  0.25157
##
## (Intercept)                              ***
## factor(potrosnjaizgled$Hiphop..Rap)2     **
## factor(potrosnjaizgled$Hiphop..Rap)3     ***
## factor(potrosnjaizgled$Hiphop..Rap)4     ***
## factor(potrosnjaizgled$Hiphop..Rap)5     ***
## factor(potrosnjaizgled$Techno..Trance)2
## factor(potrosnjaizgled$Techno..Trance)3
## factor(potrosnjaizgled$Techno..Trance)4
## factor(potrosnjaizgled$Techno..Trance)5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.176 on 971 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.06051,    Adjusted R-squared:  0.05276
## F-statistic: 7.817 on 8 and 971 DF,  p-value: 3.262e-10
```

```
fit.Spending.on.looks.nonfactor = lm(Spending.on.looks ~ Hiphop..Rap + Techno..Trance, data=potrosnjaizg
summary(fit.Spending.on.looks.nonfactor)
```

```
##
## Call:
## lm(formula = Spending.on.looks ~ Hiphop..Rap + Techno..Trance,
##     data = potrosnjaizgled)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.59980 -0.88809  0.07745  0.85170  2.30321
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     2.47104    0.09840  25.112  < 2e-16 ***
## Hiphop..Rap     0.19130    0.02878   6.646 4.99e-11 ***
## Techno..Trance  0.03446    0.02998   1.149    0.251
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.177 on 977 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.05361,    Adjusted R-squared:  0.05168
## F-statistic: 27.67 on 2 and 977 DF,  p-value: 2.038e-12
```

Pokušali smo još metodu, odlučili smo faktoririzati vrijednosti prije ugradnje u model što je doista rezultairali višom vrijendošću R^2, nego kada to nismo učinili. Vrijenodsti smo odlučili faktorizirati jer ipak ovdje rukujemo s kategorijama koje, istina, imaju uređaj, ali nas je zainmalo do kakvih rezultata možemo ovim putem doći. Možemo primjetiti da se to velika rzlika očituje u razini F-statistike koje je znatno, čak i gledavši omjere, veća u slučaju sa nefaktoriziranim vrijednostima te bismo stoga ipak odbacili opciju faktoriziranja. Valja još spomenuti kako su isprobane i svi parovi vrijednosti žanrova i obrazaca potrošnje, ali to nam nije dalo bolje razultate od ovdje izvedenih.

Zaključak

Nije moguće kvalitetno predvidjeti obrazac potrošnje ovisno o žanru glazbe kojeg ispitanik preferira.

Istraživačko pitanje 3: Možemo li temeljem danih varijabli predvidjeti dob ispitanika?

Za ovo istraživačko pitanje koristiti ćemo model linearne regresije.

Model linearne regresije pretpostavlja linearnu vezu između ulaznih i izlaznih varijabli:

$$Y = \beta_0 + \sum_{j=1}^{p} \beta_j x_j + \epsilon$$

Pretpostavke modela:

- linearnost veze X i Y
- pogreške nezavisne, homogene i normalno distribuirane s $\epsilon \sim \mathcal{N}(0, \sigma^2)$

Iz podataka je moguće dobiti procjenu modela:

$$\hat{Y} = b_0 + \sum_{j=1}^{p} b_j x_j + e,$$

Procjena je zasnovana na metodi najmanjih kvadrata, tj. minimizaciji tzv. "sum of squared errors":

$$SSE = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = (\mathbf{y} - \mathbf{X}\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b})$$

Derivacijom se dobije:

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Da bi se ova jednadžba mogla riješiti potrebno je invertirati matricu $\mathbf{X}^T\mathbf{X} \in \mathrm{R}^{p \times p}$ (složenost $O(n^3)$), uz pretpostavku da je matrica **punog ranga**.

Estimacija parametara linearne regresije dostupni su u funkciji `lm` u paketu `stats` koji ćemo dalje i koristiti.

Počinjemo sa biranjem kategorija koje ćemo ispitivati u korelaciji sa dobi ispitanika funkcijom lm().

```
fit.economy_mgnt = lm(pitanja$Age~pitanja$Economy.Management,data=pitanja)
fit.Gardening = lm(pitanja$Age~pitanja$Gardening,data=pitanja)
fit.Celebrities = lm(pitanja$Age~pitanja$Celebrities,data=pitanja)
fit.Fun_with_friends = lm(pitanja$Age~pitanja$Fun.with.friends,data=pitanja)
fit.Adrenaline.sports = lm(pitanja$Age~pitanja$Adrenaline.sports,data=pitanja)
fit.Ageing = lm(pitanja$Age~pitanja$Ageing,data=pitanja)
fit.Fear.of.public.speaking = lm(pitanja$Age~pitanja$Fear.of.public.speaking,data=pitanja)
fit.Prioritising.workload = lm(pitanja$Age~pitanja$Prioritising.workload,data=pitanja)
fit.Thinking.ahead = lm(pitanja$Age~pitanja$Thinking.ahead,data=pitanja)
fit.Loss.of.interest = lm(pitanja$Age~pitanja$Loss.of.interest,data=pitanja)
fit.Decision.making = lm(pitanja$Age~pitanja$Decision.making,data=pitanja)
fit.Giving = lm(pitanja$Age~pitanja$Giving,data=pitanja)
```

```r
fit.Changing.the.past = lm(pitanja$Age~pitanja$Changing.the.past,data=pitanja)
fit.Waiting = lm(pitanja$Age~pitanja$Waiting,data=pitanja)
fit.Socializing  = lm(pitanja$Age~pitanja$Socializing ,data=pitanja)
fit.Unpopularity = lm(pitanja$Age~pitanja$Unpopularity,data=pitanja)
fit.Life.struggles = lm(pitanja$Age~pitanja$Life.struggles,data=pitanja)
fit.Energy.levels = lm(pitanja$Age~pitanja$Energy.levels,data=pitanja)
fit.Entertainment.spending = lm(pitanja$Age~pitanja$Entertainment.spending,data=pitanja)
fit.Education = lm(pitanja$Age~pitanja$Education)
fit.Height = lm(pitanja$Age~pitanja$Height)
fit.Weight = lm(pitanja$Age~pitanja$Weight)
fit.number.of.siblings = lm(pitanja$Age~pitanja$Number.of.siblings)
```

Bitno: Budući da vrijedi $B_i \sim N(\mu_{B_i}, \sigma_{B_i})$, $\mu_{B_i} = \beta_i$, statistika

$$T = \frac{B_i - \beta_i}{SE(B_i)}$$

ima t-distribuciju s $n - k - 1$ stupnjeva slobode, gdje je k broj parametara. Većina programskih paketa, pa tako i R, pri estimiranju koeficijenata linearne regresije automatski testira $\beta_i = 0$. One koeficijente za koje možemo odbaciti $H_0 : \beta_i = 0$ u korist $H_1 : \beta_i \neq 0$ zovemo **značajni koeficijenti**.

Mjere kvalitete prilagodbe modela podatcima

SSE Mjera koju minimiziramo estimiranjem parametara modela ("fitanjem na podatke") je SSE:

$$SSE = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

$\mathbf{R^2}$ Vrlo česta mjera kvalitete prilagodbe modela je koeficijent deteminacije, definiran kao:

$$R^2 = 1 - \frac{SSE}{SST},$$

gdje je: $SST = \sum_{i=1}^{N} (y_i - \bar{y}_i)^2$ tzv. "total corrected sum of squares". Koeficijent determinacije R^2 je za linearne modele po definiciji $R^2 \in [0, 1]$ i opisuje koji postotak varijance u izlaznoj varijabli Y je estimirani linearni model objasnio/opisao.

Adjusted $\mathbf{R^2}$ Prilagođeni koeficijent determinacije penalizira dodatne parametre u modelu:

$$R^2_{adj} = 1 - \frac{SSE/(n - k - 1)}{SST/(n - 1)}.$$

```r
summary(fit.economy_mgnt)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Economy.Management, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.6552 -1.6552 -0.4817  1.1714  9.8652
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 19.96137    0.19608 101.800  < 2e-16 ***
## pitanja$Economy.Management  0.17345    0.06609   2.625  0.00881 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.814 on 996 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.006868,   Adjusted R-squared:  0.005871
## F-statistic: 6.888 on 1 and 996 DF,  p-value: 0.008809
```

```
summary(fit.Gardening)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Gardening, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.1919 -1.6990 -0.4525  0.8697  9.7940
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       19.95952    0.16973 117.596  < 2e-16 ***
## pitanja$Gardening  0.24649    0.07574   3.254  0.00117 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.811 on 994 degrees of freedom
##   (14 observations deleted due to missingness)
## Multiple R-squared:  0.01054,    Adjusted R-squared:  0.009547
## F-statistic: 10.59 on 1 and 994 DF,  p-value: 0.001175
```

```
summary(fit.Fun_with_friends)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Fun.with.friends, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.8214 -1.5677 -0.3140  0.9249  9.6860
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               21.5825     0.5620  38.401   <2e-16 ***
## pitanja$Fun.with.friends  -0.2537     0.1216  -2.087   0.0372 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.808 on 997 degrees of freedom
##   (11 observations deleted due to missingness)
## Multiple R-squared:  0.004349,   Adjusted R-squared:  0.00335
## F-statistic: 4.355 on 1 and 997 DF,  p-value: 0.03716
```

```
summary(fit.Adrenaline.sports)
```

```
##
```

```
## Call:
## lm(formula = pitanja$Age ~ pitanja$Adrenaline.sports, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.5351 -1.5351 -0.4819  1.4649  9.6776
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               20.58830    0.20721  99.359   <2e-16 ***
## pitanja$Adrenaline.sports -0.05318    0.06317  -0.842      0.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.832 on 998 degrees of freedom
##   (10 observations deleted due to missingness)
## Multiple R-squared:  0.0007096,  Adjusted R-squared:  -0.0002917
## F-statistic: 0.7086 on 1 and 998 DF,  p-value: 0.4001
```

```
summary(fit.Ageing)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Ageing, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.5281 -1.4897 -0.4514  1.4719  9.6253
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    20.33632    0.18910 107.540   <2e-16 ***
## pitanja$Ageing  0.03835    0.06466   0.593    0.553
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.831 on 1000 degrees of freedom
##   (8 observations deleted due to missingness)
## Multiple R-squared:  0.0003516,  Adjusted R-squared:  -0.0006481
## F-statistic: 0.3517 on 1 and 1000 DF,  p-value: 0.5533
```

```
summary(fit.Fear.of.public.speaking)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Fear.of.public.speaking, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.5996 -1.5047 -0.4099  1.4004  9.7798
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     20.69444    0.22456  92.156   <2e-16 ***
## pitanja$Fear.of.public.speaking -0.09485    0.07341  -1.292    0.197
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.824 on 1000 degrees of freedom
##   (8 observations deleted due to missingness)
## Multiple R-squared:  0.001667,   Adjusted R-squared:  0.0006686
## F-statistic:  1.67 on 1 and 1000 DF,  p-value: 0.1966
```

```
summary(fit.Prioritising.workload)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Prioritising.workload, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.9067 -1.8353 -0.5496  1.1647 10.1647
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   19.47818    0.21004   92.73  < 2e-16 ***
## pitanja$Prioritising.workload  0.35714    0.07215    4.95  8.7e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.786 on 996 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.02401,    Adjusted R-squared:  0.02303
## F-statistic:  24.5 on 1 and 996 DF,  p-value: 8.702e-07
```

```
summary(fit.Thinking.ahead)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Thinking.ahead, data = pitanja)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.722 -1.722 -0.538  1.278  9.830
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            19.80209    0.28190  70.246   <2e-16 ***
## pitanja$Thinking.ahead  0.18398    0.07837   2.348   0.0191 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.815 on 998 degrees of freedom
##   (10 observations deleted due to missingness)
## Multiple R-squared:  0.005492,   Adjusted R-squared:  0.004495
## F-statistic: 5.511 on 1 and 998 DF,  p-value: 0.01909
```

```
summary(fit.Loss.of.interest)
```

```
##
## Call:
```

```
## lm(formula = pitanja$Age ~ pitanja$Loss.of.interest, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4742 -1.4742 -0.4464  1.5258  9.6372
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              20.50207    0.20010 102.460   <2e-16 ***
## pitanja$Loss.of.interest -0.02785    0.06596  -0.422    0.673
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.819 on 997 degrees of freedom
##   (11 observations deleted due to missingness)
## Multiple R-squared:  0.0001788,  Adjusted R-squared:  -0.000824
## F-statistic: 0.1783 on 1 and 997 DF,  p-value: 0.6729
```

```
summary(fit.Decision.making)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Decision.making, data = pitanja)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.532 -1.532 -0.425  1.468  9.682
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             20.26477    0.25527  79.386   <2e-16 ***
## pitanja$Decision.making  0.05340    0.07478   0.714    0.475
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.833 on 997 degrees of freedom
##   (11 observations deleted due to missingness)
## Multiple R-squared:  0.0005112,  Adjusted R-squared:  -0.0004913
## F-statistic: 0.5099 on 1 and 997 DF,  p-value: 0.4754
```

```
summary(fit.Giving)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Giving, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.5270 -1.4753 -0.4235  1.4730  9.6800
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    20.26829    0.22132   91.58   <2e-16 ***
## pitanja$Giving  0.05174    0.06809    0.76    0.448
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.81 on 995 degrees of freedom
##   (13 observations deleted due to missingness)
## Multiple R-squared:  0.0005799,  Adjusted R-squared:  -0.0004245
## F-statistic: 0.5774 on 1 and 995 DF,  p-value: 0.4475
```

```
summary(fit.Changing.the.past)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Changing.the.past, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0737 -1.7472 -0.7472  0.9263  9.9060
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               21.40034    0.22320  95.879  < 2e-16 ***
## pitanja$Changing.the.past -0.32659    0.06936  -4.709 2.84e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.802 on 999 degrees of freedom
##   (9 observations deleted due to missingness)
## Multiple R-squared:  0.02171,    Adjusted R-squared:  0.02073
## F-statistic: 22.17 on 1 and 999 DF,  p-value: 2.842e-06
```

```
summary(fit.Waiting)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Waiting, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.8534 -1.6738 -0.4942  1.3262  9.6854
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     19.95534    0.25469  78.352   <2e-16 ***
## pitanja$Waiting  0.17961    0.08948   2.007    0.045 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.828 on 998 degrees of freedom
##   (10 observations deleted due to missingness)
## Multiple R-squared:  0.004021,   Adjusted R-squared:  0.003023
## F-statistic: 4.029 on 1 and 998 DF,  p-value: 0.045
```

```
summary(fit.Socializing)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Socializing, data = pitanja)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4870 -1.4870 -0.4309  1.4569  9.6813
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         20.59916    0.27367  75.269   <2e-16 ***
## pitanja$Socializing -0.05609    0.08181  -0.686    0.493
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.826 on 996 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.0004717,  Adjusted R-squared:  -0.0005319
## F-statistic:  0.47 on 1 and 996 DF,  p-value: 0.4931
```

```
summary(fit.Unpopularity)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Unpopularity, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.6882 -1.5081 -0.4823  1.4147  9.7235
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          20.79110    0.29077  71.505   <2e-16 ***
## pitanja$Unpopularity -0.10292    0.07995  -1.287    0.198
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.832 on 998 degrees of freedom
##   (10 observations deleted due to missingness)
## Multiple R-squared:  0.001658,   Adjusted R-squared:  0.0006572
## F-statistic: 1.657 on 1 and 998 DF,  p-value: 0.1983
```

```
summary(fit.Life.struggles)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Life.struggles, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.6500 -1.5681 -0.4317  1.3500  9.7865
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            20.75908    0.21615  96.042   <2e-16 ***
## pitanja$Life.struggles -0.10912    0.06488  -1.682   0.0929 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2.823 on 998 degrees of freedom
##   (10 observations deleted due to missingness)
## Multiple R-squared:  0.002826,   Adjusted R-squared:  0.001827
## F-statistic: 2.829 on 1 and 998 DF,  p-value: 0.09291
```

```
summary(fit.Energy.levels)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Energy.levels, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4323 -1.4323 -0.4298  1.5677  9.5727
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)           20.419684   0.336160  60.744   <2e-16 ***
## pitanja$Energy.levels  0.002524   0.089165   0.028    0.977
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.825 on 996 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  8.046e-07,  Adjusted R-squared:  -0.001003
## F-statistic: 0.0008014 on 1 and 996 DF,  p-value: 0.9774
```

```
summary(fit.Entertainment.spending)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Entertainment.spending, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.5949 -1.5215 -0.4481  1.4051  9.6988
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     20.66837    0.25760  80.235   <2e-16 ***
## pitanja$Entertainment.spending -0.07344    0.07536  -0.975     0.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.833 on 998 degrees of freedom
##   (10 observations deleted due to missingness)
## Multiple R-squared:  0.0009507,  Adjusted R-squared:  -5.037e-05
## F-statistic: 0.9497 on 1 and 998 DF,  p-value: 0.33
```

```
summary(fit.Celebrities)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Celebrities, data = pitanja)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4837 -1.4837 -0.4499  1.5163  9.6514
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         20.51742    0.18889  108.62   <2e-16 ***
## pitanja$Celebrities -0.03377    0.07035   -0.48    0.631
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.831 on 999 degrees of freedom
##   (9 observations deleted due to missingness)
## Multiple R-squared:  0.0002307,  Adjusted R-squared:  -0.0007701
## F-statistic: 0.2305 on 1 and 999 DF,  p-value: 0.6313
```

```
summary(fit.Education)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Education)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8205 -1.0952  0.0692  1.0692 10.0692
##
## Coefficients:
##                                                   Estimate Std. Error t value
## (Intercept)                                         17.000      2.136   7.958
## pitanja$Educationcollege/bachelor degree             4.095      2.141   1.913
## pitanja$Educationcurrently a primary school pupil   -0.500      2.240  -0.223
## pitanja$Educationdoctorate degree                    8.400      2.340   3.590
## pitanja$Educationmasters degree                      8.820      2.150   4.103
## pitanja$Educationprimary school                      0.500      2.150   0.233
## pitanja$Educationsecondary school                    2.931      2.138   1.371
##                                                   Pr(>|t|)
## (Intercept)                                       4.73e-15 ***
## pitanja$Educationcollege/bachelor degree          0.056088 .
## pitanja$Educationcurrently a primary school pupil 0.823445
## pitanja$Educationdoctorate degree                 0.000347 ***
## pitanja$Educationmasters degree                   4.41e-05 ***
## pitanja$Educationprimary school                   0.816134
## pitanja$Educationsecondary school                 0.170718
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.136 on 996 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.4332, Adjusted R-squared:  0.4298
## F-statistic: 126.9 on 6 and 996 DF,  p-value: < 2.2e-16
```

```
summary(fit.Height)
```

```
##
## Call:
```

```
## lm(formula = pitanja$Age ~ pitanja$Height)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.9745 -1.8064 -0.5195  1.0580 10.0005
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    14.799536   1.556198    9.51  < 2e-16 ***
## pitanja$Height  0.032500   0.008953    3.63 0.000298 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.821 on 986 degrees of freedom
##   (22 observations deleted due to missingness)
## Multiple R-squared:  0.01319,    Adjusted R-squared:  0.01219
## F-statistic: 13.18 on 1 and 986 DF,  p-value: 0.0002979
```

```
summary(fit.Weight)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Weight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3428 -1.8321 -0.5144  1.1192 10.5090
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.200521   0.430504  39.954  < 2e-16 ***
## pitanja$Weight  0.048733   0.006345   7.681  3.8e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.759 on 985 degrees of freedom
##   (23 observations deleted due to missingness)
## Multiple R-squared:  0.05651,    Adjusted R-squared:  0.05555
## F-statistic:    59 on 1 and 985 DF,  p-value: 3.799e-14
```

```
summary(fit.number.of.siblings)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Number.of.siblings)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.8674 -1.6155 -0.3636  1.1326  9.8884
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                20.11163    0.14466 139.030  < 2e-16 ***
## pitanja$Number.of.siblings  0.25193    0.08784   2.868  0.00422 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.818 on 999 degrees of freedom
##   (9 observations deleted due to missingness)
## Multiple R-squared:  0.008168,   Adjusted R-squared:  0.007175
## F-statistic: 8.227 on 1 and 999 DF,  p-value: 0.004215
```

Nakon ispisa možemo zaključiti da postoje varijable koje su povezane sa varijablom godine, te želimo istražiti koje kombinacije su najbolje korištenjem višestruke regresije, zato biramo one sa najvišim parametrima R squared i najnižim p vrijednostima.

```
fit.GardeningAndWorkload = lm(pitanja$Age~pitanja$Gardening + pitanja$Prioritising.workload,data=pitanj

summary(fit.GardeningAndWorkload)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Gardening + pitanja$Prioritising.workload,
##     data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0873 -1.6889 -0.5541  1.2446 10.3111
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   19.15572    0.24006  79.796  < 2e-16 ***
## pitanja$Gardening              0.20126    0.07564   2.661  0.00792 **
## pitanja$Prioritising.workload  0.33195    0.07265   4.569 5.52e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.773 on 988 degrees of freedom
##   (19 observations deleted due to missingness)
## Multiple R-squared:  0.03143,    Adjusted R-squared:  0.02947
## F-statistic: 16.03 on 2 and 988 DF,  p-value: 1.411e-07
```

```
fit.ChangingThePastAndWorkload = lm(pitanja$Age~pitanja$Changing.the.past + pitanja$Prioritising.worklo

summary(fit.ChangingThePastAndWorkload)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Changing.the.past + pitanja$Prioritising.workload,
##     data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.4488 -1.8245 -0.5267  1.1264 10.4242
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   20.44028    0.30583  66.834  < 2e-16 ***
## pitanja$Changing.the.past     -0.29776    0.06910  -4.309 1.80e-05 ***
## pitanja$Prioritising.workload  0.32657    0.07201   4.535 6.46e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.764 on 993 degrees of freedom
##   (14 observations deleted due to missingness)
## Multiple R-squared:  0.04215,    Adjusted R-squared:  0.04022
## F-statistic: 21.85 on 2 and 993 DF,  p-value: 5.173e-10
```

```
fit.GardeningAndChangingThePast = lm(pitanja$Age~pitanja$Gardening + pitanja$Changing.the.past,data=pita

summary(fit.GardeningAndChangingThePast)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Gardening + pitanja$Changing.the.past,
##     data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.1785 -1.7745 -0.5285  1.1400 10.1400
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               20.95081    0.26575  78.837  < 2e-16 ***
## pitanja$Gardening          0.24606    0.07522   3.271  0.00111 **
## pitanja$Changing.the.past -0.33420    0.06916  -4.832 1.57e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.782 on 991 degrees of freedom
##   (16 observations deleted due to missingness)
## Multiple R-squared:  0.03348,    Adjusted R-squared:  0.03153
## F-statistic: 17.17 on 2 and 991 DF,  p-value: 4.691e-08
```

```
fit.ChangingThePastAndWorkloadAndGardening = lm(pitanja$Age~pitanja$Changing.the.past + pitanja$Priorit

summary(fit.ChangingThePastAndWorkloadAndGardening)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Changing.the.past + pitanja$Prioritising.workload +
##     pitanja$Gardening, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.2403 -1.7361 -0.5316  1.0742 10.5784
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   20.14540    0.32692  61.622  < 2e-16 ***
## pitanja$Changing.the.past     -0.30690    0.06902  -4.446 9.73e-06 ***
## pitanja$Prioritising.workload  0.29933    0.07247   4.130 3.93e-05 ***
## pitanja$Gardening              0.20447    0.07525   2.717   0.0067 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 2.748 on 985 degrees of freedom
##   (21 observations deleted due to missingness)
## Multiple R-squared:  0.05083,    Adjusted R-squared:  0.04794
## F-statistic: 17.58 on 3 and 985 DF,  p-value: 3.991e-11
```

```
fit.ChangingThePastAndWorkloadAndGardeningAndEducation = lm(pitanja$Age~pitanja$Changing.the.past + pita

summary(fit.ChangingThePastAndWorkloadAndGardeningAndEducation)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Changing.the.past + pitanja$Prioritising.workload +
##     pitanja$Gardening + pitanja$Education, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8918 -1.2307 -0.1616  1.0418 10.4285
##
## Coefficients:
##                                                   Estimate Std. Error t value
## (Intercept)                                       16.81769    2.10828   7.977
## pitanja$Changing.the.past                         -0.15291    0.05308  -2.881
## pitanja$Prioritising.workload                      0.13241    0.05560   2.382
## pitanja$Gardening                                  0.12190    0.05763   2.115
## pitanja$Educationcollege/bachelor degree           4.17211    2.09768   1.989
## pitanja$Educationcurrently a primary school pupil -0.38361    2.19541  -0.175
## pitanja$Educationdoctorate degree                  8.38602    2.29288   3.657
## pitanja$Educationmasters degree                    8.74944    2.10640   4.154
## pitanja$Educationprimary school                    0.69104    2.10644   0.328
## pitanja$Educationsecondary school                  2.95820    2.09434   1.412
##                                                   Pr(>|t|)
## (Intercept)                                       4.17e-15 ***
## pitanja$Changing.the.past                         0.004055 **
## pitanja$Prioritising.workload                     0.017431 *
## pitanja$Gardening                                 0.034662 *
## pitanja$Educationcollege/bachelor degree          0.046988 *
## pitanja$Educationcurrently a primary school pupil 0.861326
## pitanja$Educationdoctorate degree                 0.000268 ***
## pitanja$Educationmasters degree                   3.56e-05 ***
## pitanja$Educationprimary school                   0.742937
## pitanja$Educationsecondary school                 0.158128
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.093 on 979 degrees of freedom
##   (21 observations deleted due to missingness)
## Multiple R-squared:  0.4532, Adjusted R-squared:  0.4481
## F-statistic: 90.14 on 9 and 979 DF,  p-value: < 2.2e-16
```

```
fit.ChangingThePastAndWorkloadAndGardeningAndWeight = lm(pitanja$Age~pitanja$Changing.the.past + pitanja

summary(fit.ChangingThePastAndWorkloadAndGardeningAndWeight)
```

```
##
## Call:
```

```
## lm(formula = pitanja$Age ~ pitanja$Changing.the.past + pitanja$Prioritising.workload +
##     pitanja$Gardening + pitanja$Weight, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.8202 -1.7267 -0.4116  1.0840 10.0595
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   16.617026   0.534635  31.081  < 2e-16 ***
## pitanja$Changing.the.past     -0.296240   0.067418  -4.394 1.24e-05 ***
## pitanja$Prioritising.workload  0.326592   0.070963   4.602 4.74e-06 ***
## pitanja$Gardening              0.244098   0.073877   3.304 0.000988 ***
## pitanja$Weight                 0.050577   0.006177   8.188 8.34e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.668 on 968 degrees of freedom
##   (37 observations deleted due to missingness)
## Multiple R-squared:  0.1133, Adjusted R-squared:  0.1096
## F-statistic: 30.91 on 4 and 968 DF,  p-value: < 2.2e-16
```

```
fit.EducationAndWeight = lm(pitanja$Age~pitanja$Education + pitanja$Weight,data=pitanja)

summary(fit.EducationAndWeight)
```

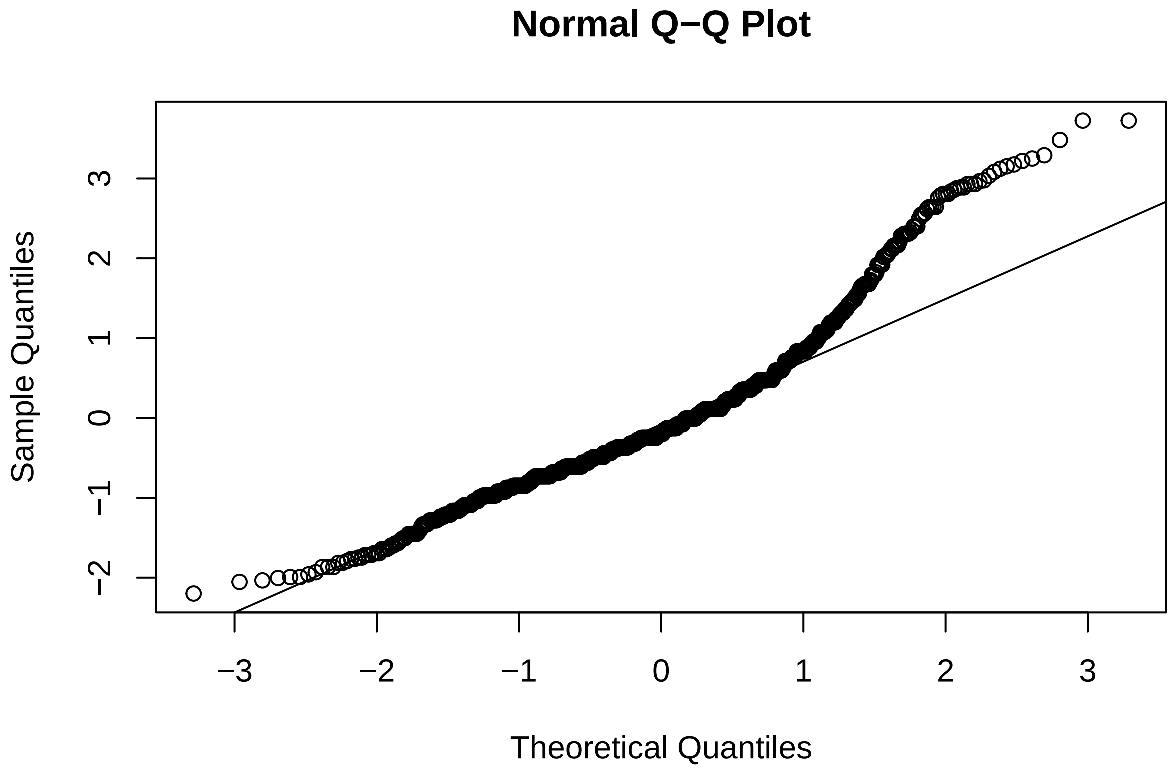
```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Education + pitanja$Weight,
##     data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.2860 -1.2994 -0.2922  1.0675  9.9844
##
## Coefficients:
##                                                   Estimate Std. Error t value
## (Intercept)                                       15.42326    2.11204   7.303
## pitanja$Educationcollege/bachelor degree           3.56695    2.10412   1.695
## pitanja$Educationcurrently a primary school pupil -0.85634    2.20063  -0.389
## pitanja$Educationdoctorate degree                  6.79172    2.31151   2.938
## pitanja$Educationmasters degree                    8.19143    2.11324   3.876
## pitanja$Educationprimary school                    0.09678    2.11197   0.046
## pitanja$Educationsecondary school                  2.41640    2.10083   1.150
## pitanja$Weight                                     0.03154    0.00494   6.383
##                                                   Pr(>|t|)
## (Intercept)                                       5.84e-13 ***
## pitanja$Educationcollege/bachelor degree          0.090352 .
## pitanja$Educationcurrently a primary school pupil 0.697260
## pitanja$Educationdoctorate degree                 0.003379 **
## pitanja$Educationmasters degree                   0.000113 ***
## pitanja$Educationprimary school                   0.963460
## pitanja$Educationsecondary school                 0.250337
## pitanja$Weight                                    2.67e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.098 on 979 degrees of freedom
##   (23 observations deleted due to missingness)
## Multiple R-squared:  0.4581, Adjusted R-squared:  0.4543
## F-statistic: 118.2 on 7 and 979 DF,  p-value: < 2.2e-16
```

```
fit.EducationAndWorkload = lm(pitanja$Age~pitanja$Education + pitanja$Prioritising.workload,data=pitanj

summary(fit.EducationAndWorkload)
```

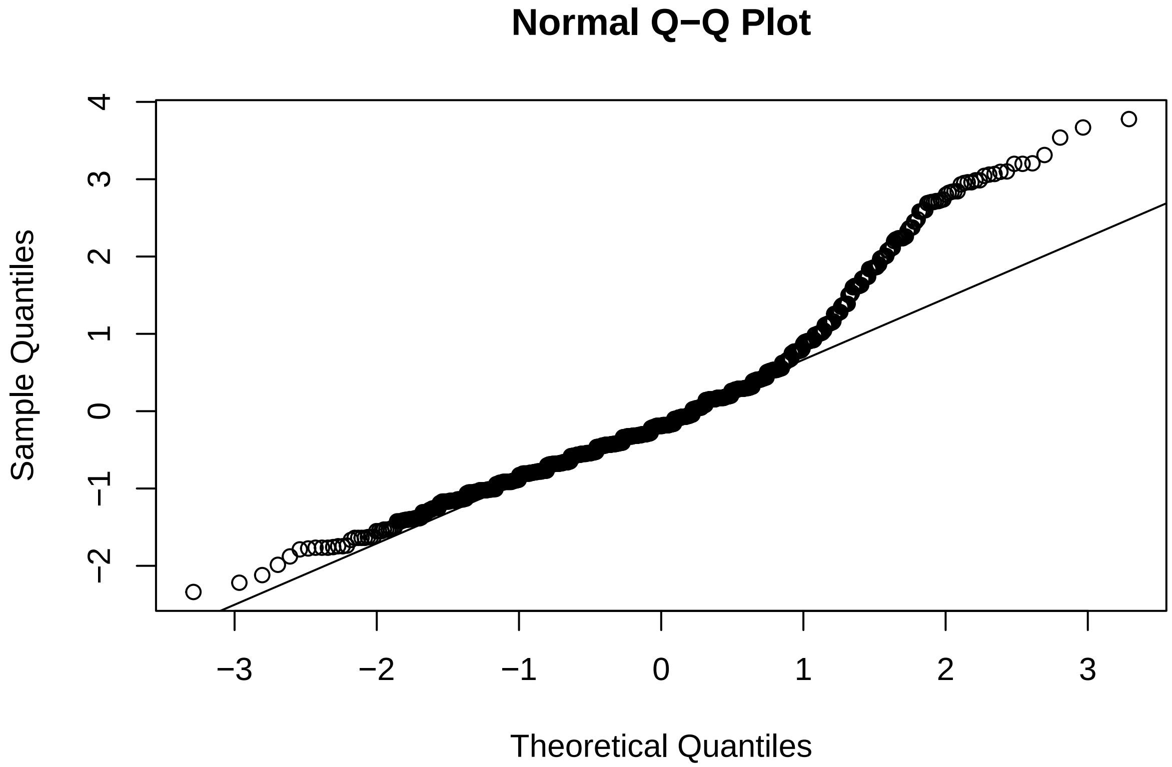
```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Education + pitanja$Prioritising.workload,
##     data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.6800 -1.1521 -0.1402  1.0171 10.3319
##
## Coefficients:
##                                                   Estimate Std. Error t value
## (Intercept)                                       16.52783    2.12065   7.794
## pitanja$Educationcollege/bachelor degree           4.14995    2.11926   1.958
## pitanja$Educationcurrently a primary school pupil -0.45278    2.21740  -0.204
## pitanja$Educationdoctorate degree                  8.46296    2.31604   3.654
## pitanja$Educationmasters degree                    8.83736    2.12804   4.153
## pitanja$Educationprimary school                    0.60089    2.12795   0.282
## pitanja$Educationsecondary school                  2.98285    2.11597   1.410
## pitanja$Prioritising.workload                      0.15739    0.05532   2.845
##                                                   Pr(>|t|)
## (Intercept)                                       1.64e-14 ***
## pitanja$Educationcollege/bachelor degree          0.050486 .
## pitanja$Educationcurrently a primary school pupil 0.838242
## pitanja$Educationdoctorate degree                 0.000272 ***
## pitanja$Educationmasters degree                   3.57e-05 ***
## pitanja$Educationprimary school                   0.777711
## pitanja$Educationsecondary school                 0.158946
## pitanja$Prioritising.workload                     0.004531 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.114 on 990 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.4414, Adjusted R-squared:  0.4374
## F-statistic: 111.8 on 7 and 990 DF,  p-value: < 2.2e-16
```

Normalnost reziduala moguće je provjeriti grafički, pomoću kvantil-kvantil plota (usporedbom s linijom normalne razdiobe), te statistički pomoću Kolmogorov-Smirnovljevog testa.

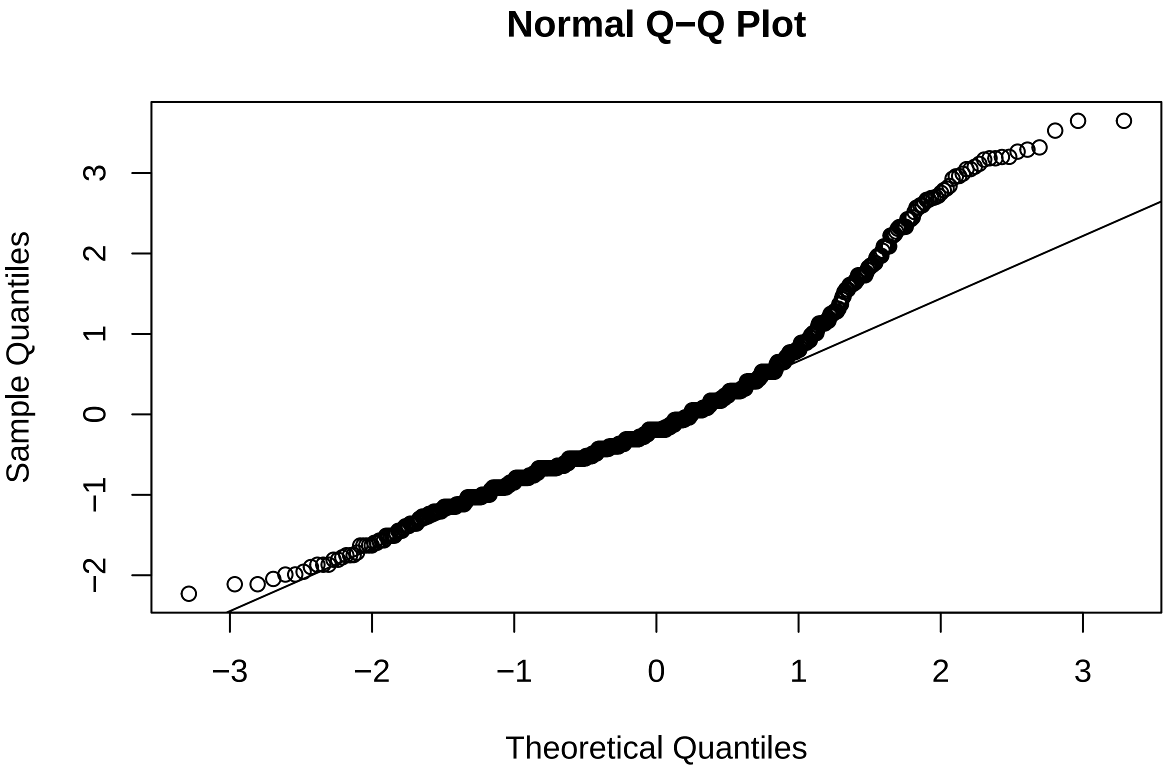
```
require(nortest)
qqnorm(rstandard(fit.GardeningAndWorkload))
qqline(rstandard(fit.GardeningAndWorkload))
```



```
qqnorm(rstandard(fit.ChangingThePastAndWorkload))
qqline(rstandard(fit.ChangingThePastAndWorkload))
```



```
qqnorm(rstandard(fit.GardeningAndChangingThePast))
qqline(rstandard(fit.GardeningAndChangingThePast))
```



```
lillie.test(rstandard(fit.GardeningAndWorkload))
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstandard(fit.GardeningAndWorkload)
## D = 0.11769, p-value < 2.2e-16
```

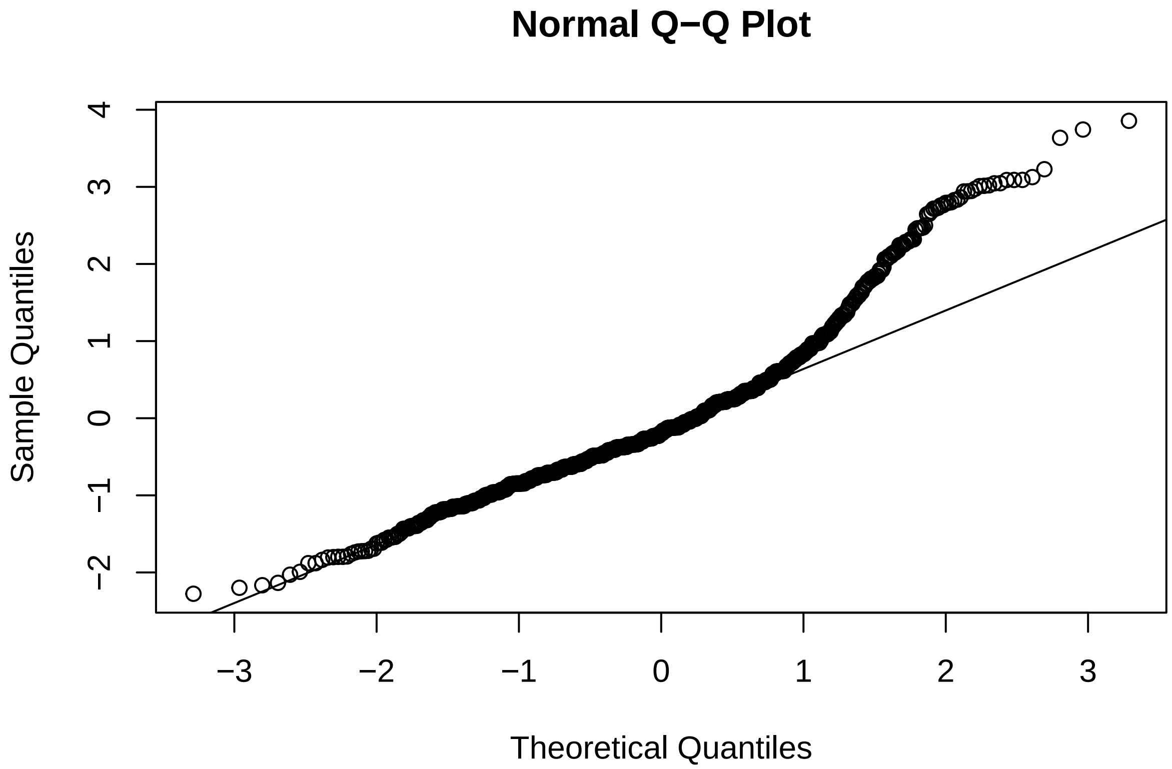
```
lillie.test(rstandard(fit.ChangingThePastAndWorkload))
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstandard(fit.ChangingThePastAndWorkload)
## D = 0.11523, p-value < 2.2e-16
```

```
lillie.test(rstandard(fit.GardeningAndChangingThePast))
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstandard(fit.GardeningAndChangingThePast)
## D = 0.10937, p-value < 2.2e-16
```

```
qqnorm(rstandard(fit.ChangingThePastAndWorkloadAndGardening))
qqline(rstandard(fit.ChangingThePastAndWorkloadAndGardening))
```



```
lillie.test(rstandard(fit.ChangingThePastAndWorkloadAndGardening))
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstandard(fit.ChangingThePastAndWorkloadAndGardening)
## D = 0.10082, p-value < 2.2e-16
```

```
cor(pitanja$Changing.the.past, pitanja$Prioritising.workload)
```

```
## [1] NA
```

```
cor(pitanja$Changing.the.past, pitanja$Gardening)
```

```
## [1] NA
```

```
cor(pitanja$Gardening, pitanja$Prioritising.workload)
```

```
## [1] NA
```

Top 3 kombinacija - Education + Weight + Workload ima najveći R squared i najmanju p vrijednost te nju obrađujemo za daljnje provjere

```
cor(pitanja$Weight,pitanja$Prioritising.workload)
```

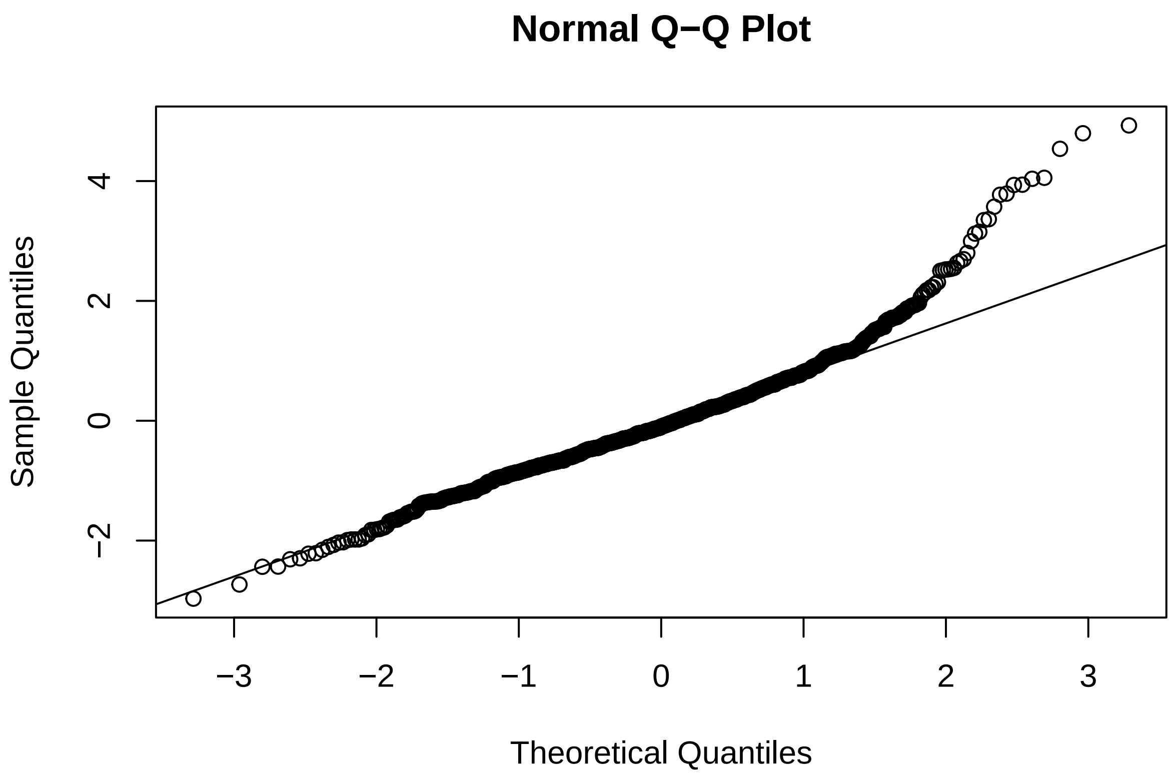
```
## [1] NA
```

```
fit.EducationAndWeightAndWorkload = lm(pitanja$Age~pitanja$Education + pitanja$Weight + pitanja$Priorit
summary(fit.EducationAndWeightAndWorkload)
```

```
##
## Call:
## lm(formula = pitanja$Age ~ pitanja$Education + pitanja$Weight +
##     pitanja$Prioritising.workload, data = pitanja)
```

```
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -6.0945 -1.2979 -0.1994  1.0424 10.1795 
## 
## Coefficients:
##                                                    Estimate Std. Error t value
## (Intercept)                                       14.789649   2.090895   7.073
## pitanja$Educationcollege/bachelor degree           3.603054   2.075667   1.736
## pitanja$Educationcurrently a primary school pupil -0.822628   2.170846  -0.379
## pitanja$Educationdoctorate degree                  6.771323   2.280242   2.970
## pitanja$Educationmasters degree                    8.175302   2.084958   3.921
## pitanja$Educationprimary school                    0.192107   2.083605   0.092
## pitanja$Educationsecondary school                  2.448527   2.072445   1.181
## pitanja$Weight                                     0.033354   0.004886   6.826
## pitanja$Prioritising.workload                      0.180890   0.054707   3.307
##                                                   Pr(>|t|)    
## (Intercept)                                       2.89e-12 ***
## pitanja$Educationcollege/bachelor degree          0.082906 .  
## pitanja$Educationcurrently a primary school pupil 0.704812    
## pitanja$Educationdoctorate degree                 0.003055 ** 
## pitanja$Educationmasters degree                   9.43e-05 ***
## pitanja$Educationprimary school                   0.926559    
## pitanja$Educationsecondary school                 0.237706    
## pitanja$Weight                                    1.53e-11 ***
## pitanja$Prioritising.workload                     0.000979 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2.069 on 973 degrees of freedom
##   (28 observations deleted due to missingness)
## Multiple R-squared:  0.4695, Adjusted R-squared:  0.4651 
## F-statistic: 107.6 on 8 and 973 DF,  p-value: < 2.2e-16
```

```
qqnorm(rstandard(fit.EducationAndWeightAndWorkload))
qqline(rstandard(fit.EducationAndWeightAndWorkload))
```



```
lillie.test(rstandard(fit.EducationAndWeightAndWorkload))
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstandard(fit.EducationAndWeightAndWorkload)
## D = 0.06577, p-value = 8.933e-11
```

H0 - dob ispitanika se ne može predvidjeti H1 - dob ispitanika se može predvidjeti Distribucija reziduala teži ka normalnoj na što i ciljamo, a q-q plot reziduala ne varira daleko od normalne distribucije tj. nalikuje normalnoj Uz ovoliko mali p i veliki R možemo zaključiti da se H0 odbacuje (cilj nam je imati što manji p radi testa). R squared nam treba biti što veći obzirom da on opisuje koji postotak varijance u izlaznoj varijabli Y je estimirani linearni model objasnio tj. opisao.

Dodatni zadatak: Kako su kategorije o ljudskom ponašanju povezane sa brojem prijatelja?

```
fit.FriendsvsFake = lm(pitanja$Number.of.friends~pitanja$Fake,data=pitanja)
summary(fit.FriendsvsFake)
```

```
##
## Call:
## lm(formula = pitanja$Number.of.friends ~ pitanja$Fake, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4956 -0.4956 -0.2238  0.6403  2.0480
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.63150    0.07467  48.635  < 2e-16 ***
## pitanja$Fake -0.13590    0.03145  -4.321 1.71e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.045 on 1007 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.0182, Adjusted R-squared:  0.01723
## F-statistic: 18.67 on 1 and 1007 DF,  p-value: 1.71e-05
```

```
fit.FriendsvsMoodSwings = lm(pitanja$Number.of.friends~pitanja$Mood.swings,data=pitanja)
summary(fit.FriendsvsMoodSwings)
```

```
##
## Call:
## lm(formula = pitanja$Number.of.friends ~ pitanja$Mood.swings,
##     data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6351 -0.5062 -0.2483  0.7517  1.8806
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          3.76409    0.10838  34.731  < 2e-16 ***
## pitanja$Mood.swings -0.12894    0.03167  -4.071 5.05e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.049 on 1004 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.01624,    Adjusted R-squared:  0.01526
## F-statistic: 16.57 on 1 and 1004 DF,  p-value: 5.051e-05
```

```
fit.FriendsvsLying = lm(pitanja$Number.of.friends~pitanja$Lying,data=pitanja)
summary(fit.FriendsvsLying)
```

```
##
## Call:
## lm(formula = pitanja$Number.of.friends ~ pitanja$Lying, data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3679 -0.3679 -0.3259  0.6741  1.7255
##
## Coefficients:
##                                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                                  3.5000     0.7474   4.683 3.22e-06
## pitanja$Lyingeverytime it suits me          -0.1957     0.7528  -0.260    0.795
## pitanja$Lyingnever                          -0.2255     0.7619  -0.296    0.767
## pitanja$Lyingonly to avoid hurting someone  -0.1741     0.7502  -0.232    0.817
## pitanja$Lyingsometimes                      -0.1321     0.7488  -0.176    0.860
##
## (Intercept)                                ***
## pitanja$Lyingeverytime it suits me
## pitanja$Lyingnever
## pitanja$Lyingonly to avoid hurting someone
## pitanja$Lyingsometimes
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.057 on 1005 degrees of freedom
## Multiple R-squared:  0.0008139,  Adjusted R-squared:  -0.003163
## F-statistic: 0.2047 on 4 and 1005 DF,  p-value: 0.9359
```

```
fit.FriendsvsPunctuality = lm(pitanja$Number.of.friends~pitanja$Punctuality,data=pitanja)
summary(fit.FriendsvsPunctuality)
```

```
##
## Call:
## lm(formula = pitanja$Number.of.friends ~ pitanja$Punctuality,
##     data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4858 -0.4858 -0.1040  0.5564  1.8960
##
## Coefficients:
##                                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                                  2.5000     0.7373   3.391 0.000725
## pitanja$Punctualityi am always on time       0.9436     0.7392   1.277 0.202060
## pitanja$Punctualityi am often early          0.6040     0.7396   0.817 0.414336
## pitanja$Punctualityi am often running late   0.9858     0.7400   1.332 0.183077
##
## (Intercept)                                ***
## pitanja$Punctualityi am always on time
## pitanja$Punctualityi am often early
## pitanja$Punctualityi am often running late
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.043 on 1006 degrees of freedom
## Multiple R-squared:  0.0266, Adjusted R-squared:  0.0237
## F-statistic: 9.164 on 3 and 1006 DF,  p-value: 5.512e-06
```

```
fit.FriendsvsGettingAngry = lm(pitanja$Number.of.friends~pitanja$Getting.angry,data=pitanja)
summary(fit.FriendsvsGettingAngry)
```

```
##
## Call:
## lm(formula = pitanja$Number.of.friends ~ pitanja$Getting.angry,
##     data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3765 -0.3584 -0.3221  0.6598  1.6960
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              3.39461    0.09171  37.014   <2e-16 ***
## pitanja$Getting.angry -0.01813    0.02835  -0.639    0.523
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.055 on 1004 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.000407,   Adjusted R-squared:  -0.0005886
## F-statistic: 0.4088 on 1 and 1004 DF,  p-value: 0.5227
```

```
fit.FriendsvsCheatingInSchool = lm(pitanja$Number.of.friends~pitanja$Cheating.in.school,data=pitanja)
summary(fit.FriendsvsCheatingInSchool)
```

```
##
## Call:
## lm(formula = pitanja$Number.of.friends ~ pitanja$Cheating.in.school,
##     data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4794 -0.4794 -0.2636  0.7364  1.9522
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 2.93995    0.10403  28.261  < 2e-16 ***
## pitanja$Cheating.in.school  0.10789    0.02635   4.095 4.56e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.047 on 1004 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.01643,    Adjusted R-squared:  0.01545
## F-statistic: 16.77 on 1 and 1004 DF,  p-value: 4.56e-05
```

```
fit.FriendsvsCriminalDamage = lm(pitanja$Number.of.friends~pitanja$Criminal.damage,data=pitanja)
summary(fit.FriendsvsCriminalDamage)
```

```
##
## Call:
## lm(formula = pitanja$Number.of.friends ~ pitanja$Criminal.damage,
##     data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3804 -0.3804 -0.3123  0.6650  1.7104
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              3.40304    0.06671  51.013   <2e-16 ***
## pitanja$Criminal.damage -0.02268    0.02218  -1.023    0.307
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.056 on 1001 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.001044,   Adjusted R-squared:  4.554e-05
## F-statistic: 1.046 on 1 and 1001 DF,  p-value: 0.3068
```

```
fit.FriendsvsLoneliness = lm(pitanja$Number.of.friends~pitanja$Loneliness,data=pitanja)
summary(fit.FriendsvsLoneliness)
```

```
##
## Call:
## lm(formula = pitanja$Number.of.friends ~ pitanja$Loneliness,
##     data = pitanja)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.90191 -0.60620 -0.01479  0.68951  2.28092
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         4.19762    0.08645   48.55   <2e-16 ***
## pitanja$Loneliness -0.29571    0.02788  -10.61   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.002 on 1007 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.1005, Adjusted R-squared:  0.09958
## F-statistic: 112.5 on 1 and 1007 DF,  p-value: < 2.2e-16
```

```
fit.FriendsvsInternetUsage = lm(pitanja$Number.of.friends~pitanja$Internet.usage,data=pitanja)
summary(fit.FriendsvsInternetUsage)
```

```
##
## Call:
## lm(formula = pitanja$Number.of.friends ~ pitanja$Internet.usage,
##     data = pitanja)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3750 -0.3750 -0.3597  0.6250  1.8468
##
## Coefficients:
##                                              Estimate Std. Error t value
## (Intercept)                                   3.37500    0.03863  87.358
## pitanja$Internet.usageless than an hour a day -0.01529    0.09737  -0.157
## pitanja$Internet.usagemost of the day         -0.22177    0.10222  -2.170
## pitanja$Internet.usageno time at all          -0.70833    0.60963  -1.162
##                                              Pr(>|t|)
## (Intercept)                                    <2e-16 ***
## pitanja$Internet.usageless than an hour a day   0.8753
## pitanja$Internet.usagemost of the day           0.0303 *
## pitanja$Internet.usageno time at all            0.2456
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.054 on 1006 degrees of freedom
## Multiple R-squared:  0.005907,   Adjusted R-squared:  0.002943
## F-statistic: 1.993 on 3 and 1006 DF,  p-value: 0.1134
```

```
fit.FriendsvsInternet = lm(pitanja$Number.of.friends~pitanja$Internet,data=pitanja)
summary(fit.FriendsvsInternet)
```

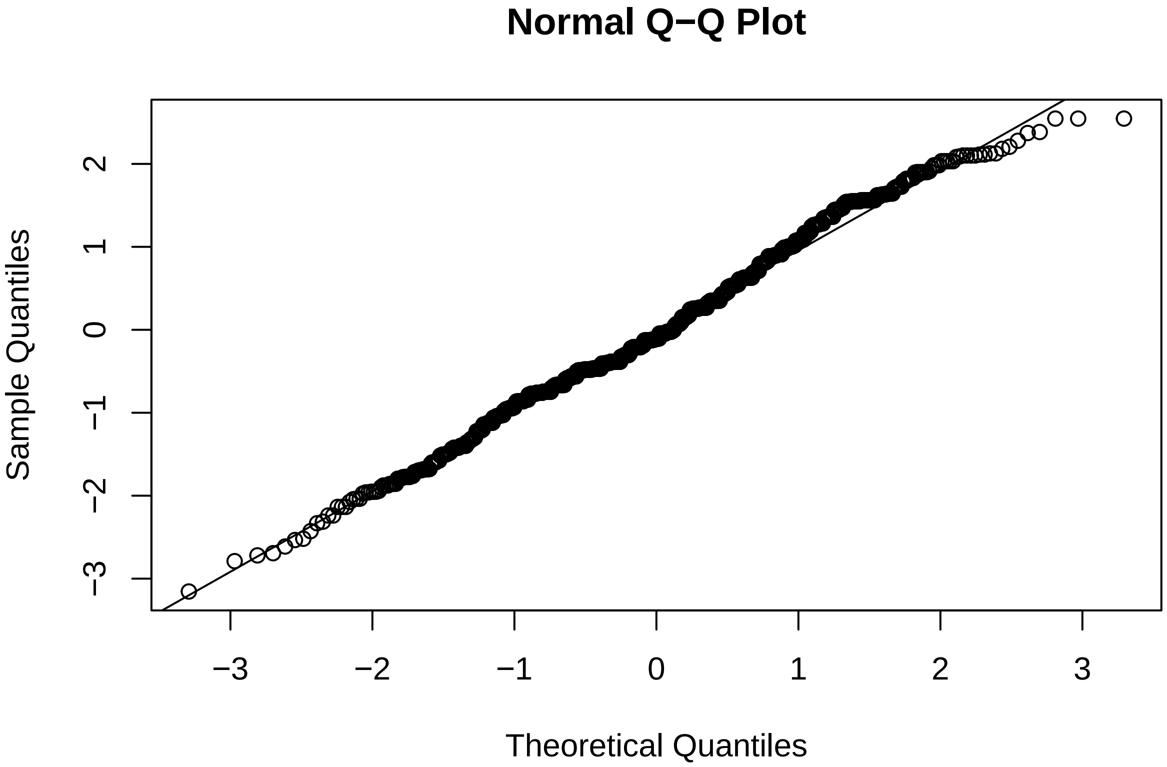
```
## 
## Call:
## lm(formula = pitanja$Number.of.friends ~ pitanja$Internet, data = pitanja)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -2.4021 -0.4021 -0.2609  0.6685  1.8802 
## 
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)    
## (Intercept)       3.04920    0.15438  19.751   <2e-16 ***
## pitanja$Internet  0.07058    0.03610   1.955   0.0509 .  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.054 on 1004 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.003792,   Adjusted R-squared:  0.0028 
## F-statistic: 3.822 on 1 and 1004 DF,  p-value: 0.05087
```

```
fit.FriendsvsFakeLonelyandPunctuality = lm(pitanja$Number.of.friends~pitanja$Fake + pitanja$Punctuality
summary(fit.FriendsvsFakeLonelyandPunctuality)
```

```
## 
## Call:
## lm(formula = pitanja$Number.of.friends ~ pitanja$Fake + pitanja$Punctuality + 
##     pitanja$Loneliness, data = pitanja)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -3.1047 -0.6556 -0.1045  0.6320  2.5064 
## 
## Coefficients:
##                                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                                 3.26991    0.70196   4.658 3.62e-06
## pitanja$Fake                               -0.08065    0.03037  -2.656  0.00804
## pitanja$Punctualityi am always on time      1.09845    0.70087   1.567  0.11737
## pitanja$Punctualityi am often early         0.84415    0.70132   1.204  0.22901
## pitanja$Punctualityi am often running late  1.19113    0.70172   1.697  0.08992
## pitanja$Loneliness                         -0.27570    0.02828  -9.749  < 2e-16
##                                               
## (Intercept)                                ***
## pitanja$Fake                               ** 
## pitanja$Punctualityi am always on time        
## pitanja$Punctualityi am often early           
## pitanja$Punctualityi am often running late .  
## pitanja$Loneliness                         ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.9876 on 1002 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.1281, Adjusted R-squared:  0.1238
## F-statistic: 29.45 on 5 and 1002 DF,  p-value: < 2.2e-16
```

Kombinacija top 3 kategorije

```
qqnorm(rstandard(fit.FriendsvsFakeLonelyandPunctuality))
qqline(rstandard(fit.FriendsvsFakeLonelyandPunctuality))
```



```
lillie.test(rstandard(fit.FriendsvsFakeLonelyandPunctuality))
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstandard(fit.FriendsvsFakeLonelyandPunctuality)
## D = 0.051476, p-value = 1.318e-06
```

```
cor(pitanja$Loneliness, pitanja$Fake)
```

```
## [1] NA
```

Distribucija rezudiala je normalna, q-q plot prikazuje da distribucija nalikuje normalnoj. R squared nije visok kao u prošlom primjeru no ovakav tip pitanja može jako varirati od osobe do osobe te ipak dokazuje povezanost ovih kategorija (Fake, Loneliness i Punctuality) sa brojem prijatelja.