1. laboratorijska vježba

Multivarijatna analiza podataka

ak. god. 2021/2022

Verzija: 1.0

1. Uvod i upute za predaju

Cilj ove laboratorijske vježbe je primijeniti osnovne koncepte multivarijatne analize podataka, istražiti podatke te ispitati hipoteze. Preduvjet za rješavanje vježbe je osnovno znanje programskog jezika *R* i rad s *R Markdown* dokumentima. Sama vježba je koncipirana kao projekt u kojem istražujete i eksperimentirate koristeći dane podatke - ne postoji nužno samo jedan točan način rješavanja svakog podzadatka.

Rješavanje vježbe svodi se na čitanje uputa u tekstu ovog dokumenta, nadopunjavanje blokova kôda (možete dodavati i dodatne blokove kôda ukoliko je potrebno) i ispisivanje rezultata (u vidu ispisa iz funkcija, tablica i grafova). Vježbu radite samostalno, a svoje rješenje branite na terminima koji su vam dodijeljeni u kalendaru. Pritom morate razumjeti teorijske osnove u okviru onoga što je obrađeno na predavanjima i morate pokazati da razumijete sav kôd koji ste napisali.

Vaše rješenje potrebno je predati u sustav *Moodle* u obliku dvije datoteke:

1. Ovaj .Rmd dokument s Vašim rješenjem (naziva IME_PREZIME_JMBAG.rmd),

2. PDF ili HTML dokument kao izvještaj generiran iz vašeg .Rmd rješenja (također naziva IME_PREZIME_JMBAG).

Rok za predaju je **3. travnja 2022. u 23:59h**. Podsjećamo da bodovi iz laboratorijskih vježbi ulaze i u bodove na ispitnom roku, te da je za polaganje predmeta potrebno imati barem 50% ukupnih bodova iz laboratorijskih vježbi. **Nadoknade laboratorijskih vježbi neće biti organizirane.** Za sva dodatna pitanja svakako se javite na email adresu predmeta: $map@fer.hr$.

2. Podatkovni skup

Podatkovni skup koji će biti razmatran u vježbi sadrži bodove studenata na jednom fakultetskom kolegiju. Svakom studentu upisani su bodovi iz dviju laboratorijskih vježbi (**LAB**), pet zadataka međuispita (**MI**), pet zadataka završnog ispita (**ZI**), pet zadataka ispitnog roka (**IR**) i kojoj grupi predavanja pripadaju (**Grupa**).

Studenti mogu položiti kolegij kontinuiranim putem ili na ispitnom roku. Kontinuirani put sastoji se od bodova s laboratorijskih vježbi, međuispita i završnog ispita. Kronološki, 1. laboratorijska vježba održana je prije međuispita, dok je 2. laboratorijska vježba održana između međuispita i završnog ispita. Ispitni rok održan je nakon završnog ispita. Ako student polaže predmet na ispitnom roku, gledaju se samo bodovi s ispitnog roka. Ukupan broj bodova je 100, a bodovi su raspodijeljeni na sljedeći način:

- Kontinuirana nastava:
  - **LAB**: 20 bodova (0-10 svaka vježba)
  - **MI** : 40 bodova (0-8 svaki zadatak)
  - **ZI** : 40 bodova (0-8 svaki zadatak)
- Ispitni rok:
  - **IR** : 100 bodova (0-20 svaki zadatak)

Za prolazak kolegija potrebno je skupiti **više** od 50 bodova i izaći na obje laboratorijske vježbe (izlazak na vježbe nužan je uvjet i za polaganje ispitnog roka, iako se bodovi ne prenose). Ako student nije pristupio pripadajućem ispitu/laboratorijskoj vježbi, nije upisan podatak (što nije isto kao i 0 bodova).

3. Priprema podataka i eksploratorna analiza

U ovom dijelu vježbe potrebno je učitati podatke i napraviti osnovnu eksploratornu analizu podataka.

3.1 Učitavanje podataka

Učitajte podatkovni skup iz datoteke *studenti.csv* i pripremite podatke za analizu. Pritom obratite pozornost na sljedeće:

- Provjerite jesu li sve varijable očekivanog tipa,
- Provjerite jesu li vrijednosti unutar zadanog raspona (s obzirom na gore opisano bodovanje),
- Provjerite zadovoljavaju li bodovi gore opisane uvjete predmeta,
- Za nedostajuće podatke ispitajte jesu li opravdani te odaberite i primijenite tehniku upravljanja nedostajućim podatcima.

Nakon što su podatci pripremljeni, analizirajte i ispišite deksriptivne statistike varijabli.

```
#  Vaš kôd ovdje
df <- read_csv("studenti.csv")
```

```
## Rows: 500 Columns: 18

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (4): MI_5, LAB_1, ZI_5, LAB_2
## dbl (14): MI_1, MI_2, MI_3, MI_4, ZI_1, ZI_2, ZI_3, ZI_4, IR_1, IR_2, IR_3, ...

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
summary(df)
```

```
##       MI_1            MI_2             MI_3            MI_4      
##  Min.   :4.000   Min.   : 0.000   Min.   :0.000   Min.   :0.500  
##  1st Qu.:6.500   1st Qu.: 4.500   1st Qu.:3.500   1st Qu.:3.500  
##  Median :7.000   Median : 6.000   Median :5.000   Median :4.000  
##  Mean   :6.912   Mean   : 5.847   Mean   :4.926   Mean   :4.007  
##  3rd Qu.:7.500   3rd Qu.: 7.500   3rd Qu.:6.500   3rd Qu.:4.500  
##  Max.   :8.000   Max.   :18.000   Max.   :8.000   Max.   :7.000  
##                                                                  
##      MI_5              LAB_1                ZI_1             ZI_2      
##  Length:500         Length:500         Min.   :-3.000   Min.   :3.000  
##  Class :character   Class :character   1st Qu.: 4.500   1st Qu.:5.500  
##  Mode  :character   Mode  :character   Median : 6.000   Median :6.000  
##                                        Mean   : 5.811   Mean   :5.985  
##                                        3rd Qu.: 7.500   3rd Qu.:6.500  
##                                        Max.   : 8.000   Max.   :8.000  
##                                                                        
##       ZI_3            ZI_4           ZI_5               LAB_2          
##  Min.   :0.000   Min.   :0.000   Length:500         Length:500        
##  1st Qu.:2.500   1st Qu.:2.500   Class :character   Class :character  
##  Median :4.000   Median :3.000   Mode  :character   Mode  :character  
##  Mean   :4.009   Mean   :2.997                                        
```

```
##  3rd Qu.:5.500   3rd Qu.:3.500
##  Max.   :8.000   Max.   :5.500
##
##       IR_1            IR_2            IR_3            IR_4
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   : 0.00
##  1st Qu.:13.50   1st Qu.:12.50   1st Qu.:13.50   1st Qu.: 8.00
##  Median :15.00   Median :14.50   Median :14.50   Median :11.25
##  Mean   :15.25   Mean   :14.07   Mean   :14.34   Mean   :11.14
##  3rd Qu.:17.62   3rd Qu.:16.00   3rd Qu.:15.50   3rd Qu.:14.12
##  Max.   :20.00   Max.   :20.00   Max.   :18.50   Max.   :20.00
##  NA's   :400     NA's   :400     NA's   :400     NA's   :400
##       IR_5            Grupa
##  Min.   : 0.000   Min.   :1.000
##  1st Qu.: 5.000   1st Qu.:1.000
##  Median : 6.500   Median :2.000
##  Mean   : 6.305   Mean   :2.006
##  3rd Qu.: 7.500   3rd Qu.:3.000
##  Max.   :11.500   Max.   :3.000
##  NA's   :400
```

```r
view(df)
```

```r
spec(df)
```

```
## cols(
##   MI_1 = col_double(),
##   MI_2 = col_double(),
##   MI_3 = col_double(),
##   MI_4 = col_double(),
##   MI_5 = col_character(),
##   LAB_1 = col_character(),
##   ZI_1 = col_double(),
##   ZI_2 = col_double(),
##   ZI_3 = col_double(),
##   ZI_4 = col_double(),
##   ZI_5 = col_character(),
##   LAB_2 = col_character(),
##   IR_1 = col_double(),
##   IR_2 = col_double(),
##   IR_3 = col_double(),
##   IR_4 = col_double(),
##   IR_5 = col_double(),
##   Grupa = col_double()
## )
```

MI_5, ZI_5 i LAB_1, LAB_2 treba pretvoriti u brojeve.

```r
df$MI_5 <- as.double(df$MI_5)
```

```
## Warning: NAs introduced by coercion
```

```r
df$ZI_5 <- as.double(df$ZI_5)
```

```
## Warning: NAs introduced by coercion
```

```r
df$LAB_1 <- as.double(df$LAB_1)
```

```
## Warning: NAs introduced by coercion
```

```r
df$LAB_2 <- as.double(df$LAB_2)
```

```
## Warning: NAs introduced by coercion
```

MI_2 i ZI_1 popraviti vrijednosti izvan specificiranih granica.

```r
df[df$MI_2 > 8, 'MI_2'] = 8
df[df$ZI_1 < 0, 'ZI_1'] = 0

df <- df %>% filter(!is.na(MI_5))
df <- df %>% filter(!is.na(ZI_5))

summary(df)
```

```
##       MI_1            MI_2           MI_3            MI_4           MI_5      
##  Min.   :4.000   Min.   :0.00   Min.   :0.000   Min.   :0.50   Min.   :0.000  
##  1st Qu.:6.500   1st Qu.:4.50   1st Qu.:3.500   1st Qu.:3.50   1st Qu.:1.500  
##  Median :7.000   Median :6.00   Median :5.000   Median :4.00   Median :3.000  
##  Mean   :6.918   Mean   :5.83   Mean   :4.924   Mean   :4.01   Mean   :3.055  
##  3rd Qu.:7.500   3rd Qu.:7.50   3rd Qu.:6.500   3rd Qu.:4.50   3rd Qu.:4.500  
##  Max.   :8.000   Max.   :8.00   Max.   :8.000   Max.   :7.00   Max.   :8.000  
##                                                                               
##      LAB_1          ZI_1            ZI_2            ZI_3            ZI_4      
##  Min.   :4.0   Min.   :0.000   Min.   :3.000   Min.   :0.000   Min.   :0.000  
##  1st Qu.:6.5   1st Qu.:4.500   1st Qu.:5.500   1st Qu.:2.500   1st Qu.:2.500  
##  Median :7.0   Median :6.000   Median :6.000   Median :4.000   Median :3.000  
##  Mean   :7.0   Mean   :5.818   Mean   :5.987   Mean   :4.014   Mean   :2.997  
##  3rd Qu.:7.5   3rd Qu.:7.500   3rd Qu.:6.500   3rd Qu.:5.500   3rd Qu.:3.500  
##  Max.   :9.5   Max.   :8.000   Max.   :8.000   Max.   :8.000   Max.   :5.500  
##  NA's   :2                                                                    
##       ZI_5           LAB_2            IR_1            IR_2            IR_3      
##  Min.   :0.00   Min.   :0.500   Min.   : 0.00   Min.   : 0.00   Min.   : 0.00  
##  1st Qu.:1.50   1st Qu.:2.500   1st Qu.:13.50   1st Qu.:12.50   1st Qu.:13.50  
##  Median :2.00   Median :3.000   Median :15.00   Median :14.50   Median :14.50  
##  Mean   :2.02   Mean   :3.002   Mean   :15.31   Mean   :14.16   Mean   :14.36  
##  3rd Qu.:2.50   3rd Qu.:3.500   3rd Qu.:17.88   3rd Qu.:16.00   3rd Qu.:15.50  
##  Max.   :5.50   Max.   :6.000   Max.   :20.00   Max.   :20.00   Max.   :18.50  
##                 NA's   :2       NA's   :400     NA's   :400     NA's   :400    
##       IR_4            IR_5            Grupa      
##  Min.   : 0.00   Min.   : 0.000   Min.   :1.000  
##  1st Qu.: 8.00   1st Qu.: 5.000   1st Qu.:1.000  
##  Median :11.25   Median : 6.500   Median :2.000  
##  Mean   :11.09   Mean   : 6.306   Mean   :2.006  
##  3rd Qu.:14.00   3rd Qu.: 7.500   3rd Qu.:3.000  
##  Max.   :20.00   Max.   :11.500   Max.   :3.000  
##  NA's   :400     NA's   :400
```

3.2 Korelacijska analiza

Razmotrimo studente koji su predmet položili kontinuirano. Izračunajte i vizualizirajte matricu korelacije za njihove bodove na nastavnim aktivnostima. Ponovite isto za studente koji su izašli na ispitni rok. Razmislite o zavisnosti različitih nastavnih aktivnosti koje vidite iz ovih korelacijskih matrica.

Ako nisu na roku (is.na) onda su prosli kontinuirano.

```r
kont_nastava <- df[is.na(df$IR_1), c('MI_1', 'MI_2','MI_3', 'MI_4', 'MI_5', 'LAB_1', 'ZI_1', 'ZI_2','ZI_

summary(kont_nastava)
```

```
##       MI_1            MI_2            MI_3            MI_4      
##  Min.   :4.500   Min.   :0.500   Min.   :0.500   Min.   :1.500  
##  1st Qu.:6.500   1st Qu.:5.000   1st Qu.:4.000   1st Qu.:3.500  
##  Median :7.000   Median :6.000   Median :5.000   Median :4.000  
##  Mean   :7.076   Mean   :6.114   Mean   :5.226   Mean   :4.104  
##  3rd Qu.:7.500   3rd Qu.:7.500   3rd Qu.:6.500   3rd Qu.:5.000  
##  Max.   :8.000   Max.   :8.000   Max.   :8.000   Max.   :7.000  
##       MI_5           LAB_1            ZI_1            ZI_2            ZI_3     
##  Min.   :0.000   Min.   :5.000   Min.   :1.500   Min.   :3.000   Min.   :0.00  
##  1st Qu.:2.000   1st Qu.:6.500   1st Qu.:5.000   1st Qu.:5.500   1st Qu.:3.00  
##  Median :3.000   Median :7.000   Median :6.000   Median :6.000   Median :4.50  
##  Mean   :3.283   Mean   :7.143   Mean   :6.178   Mean   :6.136   Mean   :4.33  
##  3rd Qu.:4.500   3rd Qu.:8.000   3rd Qu.:8.000   3rd Qu.:7.000   3rd Qu.:5.50  
##  Max.   :8.000   Max.   :9.500   Max.   :8.000   Max.   :8.000   Max.   :8.00  
##       ZI_4            ZI_5           LAB_2      
##  Min.   :0.000   Min.   :0.000   Min.   :0.500  
##  1st Qu.:2.500   1st Qu.:1.500   1st Qu.:2.500  
##  Median :3.000   Median :2.000   Median :3.000  
##  Mean   :3.107   Mean   :2.105   Mean   :3.155  
##  3rd Qu.:4.000   3rd Qu.:3.000   3rd Qu.:4.000  
##  Max.   :5.500   Max.   :5.500   Max.   :6.000  
```

Na roku su ako nije NA na nekom od zadataka s roka.

```r
rok <- df[!is.na(df$IR_1), c('MI_1', 'MI_2','MI_3', 'MI_4', 'MI_5', 'LAB_1', 'ZI_1', 'ZI_2','ZI_3','ZI_

summary(rok)
```

```
##       MI_1           MI_2            MI_3            MI_4            MI_5      
##  Min.   :4.00   Min.   :0.000   Min.   :0.000   Min.   :0.500   Min.   :0.000  
##  1st Qu.:5.50   1st Qu.:3.500   1st Qu.:2.500   1st Qu.:3.000   1st Qu.:1.000  
##  Median :6.50   Median :4.500   Median :3.500   Median :3.500   Median :2.000  
##  Mean   :6.27   Mean   :4.673   Mean   :3.689   Mean   :3.628   Mean   :2.128  
##  3rd Qu.:7.00   3rd Qu.:5.875   3rd Qu.:5.000   3rd Qu.:4.000   3rd Qu.:3.375  
##  Max.   :8.00   Max.   :8.000   Max.   :8.000   Max.   :5.500   Max.   :5.000  
##                                                                                
##      LAB_1            ZI_1            ZI_2            ZI_3      
##  Min.   :4.000   Min.   :0.000   Min.   :3.500   Min.   :0.000  
##  1st Qu.:6.000   1st Qu.:3.000   1st Qu.:5.000   1st Qu.:1.500  
##  Median :6.500   Median :4.500   Median :5.500   Median :2.500  
##  Mean   :6.406   Mean   :4.352   Mean   :5.378   Mean   :2.724  
##  3rd Qu.:7.000   3rd Qu.:5.500   3rd Qu.:6.000   3rd Qu.:3.875  
##  Max.   :9.000   Max.   :8.000   Max.   :7.500   Max.   :7.000  
##  NA's   :2                                                      
##       ZI_4            ZI_5           LAB_2      
##  Min.   :0.500   Min.   :0.000   Min.   :0.500  
##  1st Qu.:2.000   1st Qu.:1.000   1st Qu.:1.875  
##  Median :2.500   Median :1.500   Median :2.500  
##  Mean   :2.546   Mean   :1.673   Mean   :2.365  
##  3rd Qu.:3.000   3rd Qu.:2.500   3rd Qu.:3.000  
##  Max.   :5.000   Max.   :4.500   Max.   :4.000  
```

```
##                                 NA's   :2
```

```
rok <- na.omit(rok)
summary(rok)
```

```
##       MI_1            MI_2            MI_3            MI_4      
##  Min.   :4.000   Min.   :0.000   Min.   :0.000   Min.   :0.500  
##  1st Qu.:5.500   1st Qu.:3.500   1st Qu.:2.500   1st Qu.:3.000  
##  Median :6.500   Median :4.500   Median :3.500   Median :3.500  
##  Mean   :6.279   Mean   :4.653   Mean   :3.747   Mean   :3.632  
##  3rd Qu.:7.000   3rd Qu.:5.750   3rd Qu.:5.250   3rd Qu.:4.250  
##  Max.   :8.000   Max.   :8.000   Max.   :8.000   Max.   :5.500  
##       MI_5           LAB_1         ZI_1            ZI_2            ZI_3      
##  Min.   :0.000   Min.   :4.0   Min.   :0.000   Min.   :3.500   Min.   :0.000  
##  1st Qu.:1.000   1st Qu.:6.0   1st Qu.:3.250   1st Qu.:5.000   1st Qu.:1.500  
##  Median :2.000   Median :6.5   Median :4.500   Median :5.500   Median :2.500  
##  Mean   :2.121   Mean   :6.4   Mean   :4.379   Mean   :5.363   Mean   :2.711  
##  3rd Qu.:3.250   3rd Qu.:7.0   3rd Qu.:5.500   3rd Qu.:6.000   3rd Qu.:3.750  
##  Max.   :5.000   Max.   :9.0   Max.   :8.000   Max.   :7.500   Max.   :7.000  
##       ZI_4            ZI_5           LAB_2      
##  Min.   :0.500   Min.   :0.000   Min.   :0.500  
##  1st Qu.:2.000   1st Qu.:1.000   1st Qu.:1.750  
##  Median :2.500   Median :1.500   Median :2.500  
##  Mean   :2.537   Mean   :1.632   Mean   :2.368  
##  3rd Qu.:3.000   3rd Qu.:2.250   3rd Qu.:3.000  
##  Max.   :5.000   Max.   :4.500   Max.   :4.000
```

```
cor(kont_nastava)
```

```
##              MI_1         MI_2         MI_3         MI_4        MI_5
## MI_1   1.00000000  0.367573619  0.051564425  0.207578404 -0.02698639
## MI_2   0.36757362  1.000000000  0.111133638  0.010834413 -0.01661229
## MI_3   0.05156443  0.111133638  1.000000000  0.003079974  0.04031700
## MI_4   0.20757840  0.010834413  0.003079974  1.000000000  0.10611246
## MI_5  -0.02698639 -0.016612285  0.040316997  0.106112460  1.00000000
## LAB_1  0.38073980  0.054516639  0.104296944  0.049035513 -0.06007239
## ZI_1  -0.08771760 -0.112132162  0.133702875 -0.071537453 -0.11071793
## ZI_2  -0.07514389 -0.028810952 -0.116586824 -0.054060822  0.02171840
## ZI_3  -0.11147291 -0.156369542 -0.096060575  0.016733495 -0.14554676
## ZI_4  -0.12510306 -0.107161796 -0.056232860 -0.057950758 -0.04406245
## ZI_5  -0.04862162 -0.024057915 -0.052628880 -0.011836361 -0.05826498
## LAB_2  0.15419084 -0.007685481 -0.134840300 -0.044123671 -0.10590933
##             LAB_1        ZI_1         ZI_2        ZI_3        ZI_4         ZI_5
## MI_1   0.38073980 -0.08771760 -0.075143891 -0.11147291 -0.12510306 -0.048621621
## MI_2   0.05451664 -0.11213216 -0.028810952 -0.15636954 -0.10716180 -0.024057915
## MI_3   0.10429694  0.13370287 -0.116586824 -0.09606058 -0.05623286 -0.052628880
## MI_4   0.04903551 -0.07153745 -0.054060822  0.01673349 -0.05795076 -0.011836361
## MI_5  -0.06007239 -0.11071793  0.021718405 -0.14554676 -0.04406245 -0.058264976
## LAB_1  1.00000000 -0.06211692 -0.091780609 -0.11051900 -0.08801703 -0.060388584
## ZI_1  -0.06211692  1.00000000  0.316275832  0.07908431  0.13150121 -0.026778495
## ZI_2  -0.09178061  0.31627583  1.000000000  0.21370340  0.16777173  0.006044679
## ZI_3  -0.11051900  0.07908431  0.213703396  1.00000000  0.11264642  0.032871720
## ZI_4  -0.08801703  0.13150121  0.167771733  0.11264642  1.00000000  0.054917273
## ZI_5  -0.06038858 -0.02677849  0.006044679  0.03287172  0.05491727  1.000000000
## LAB_2  0.29026106  0.41336620  0.080394391 -0.07361314 -0.05407846 -0.075268823
```

```
##              LAB_2
## MI_1   0.154190836
## MI_2  -0.007685481
## MI_3  -0.134840300
## MI_4  -0.044123671
## MI_5  -0.105909331
## LAB_1  0.290261056
## ZI_1   0.413366196
## ZI_2   0.080394391
## ZI_3  -0.073613141
## ZI_4  -0.054078458
## ZI_5  -0.075268823
## LAB_2  1.000000000
```

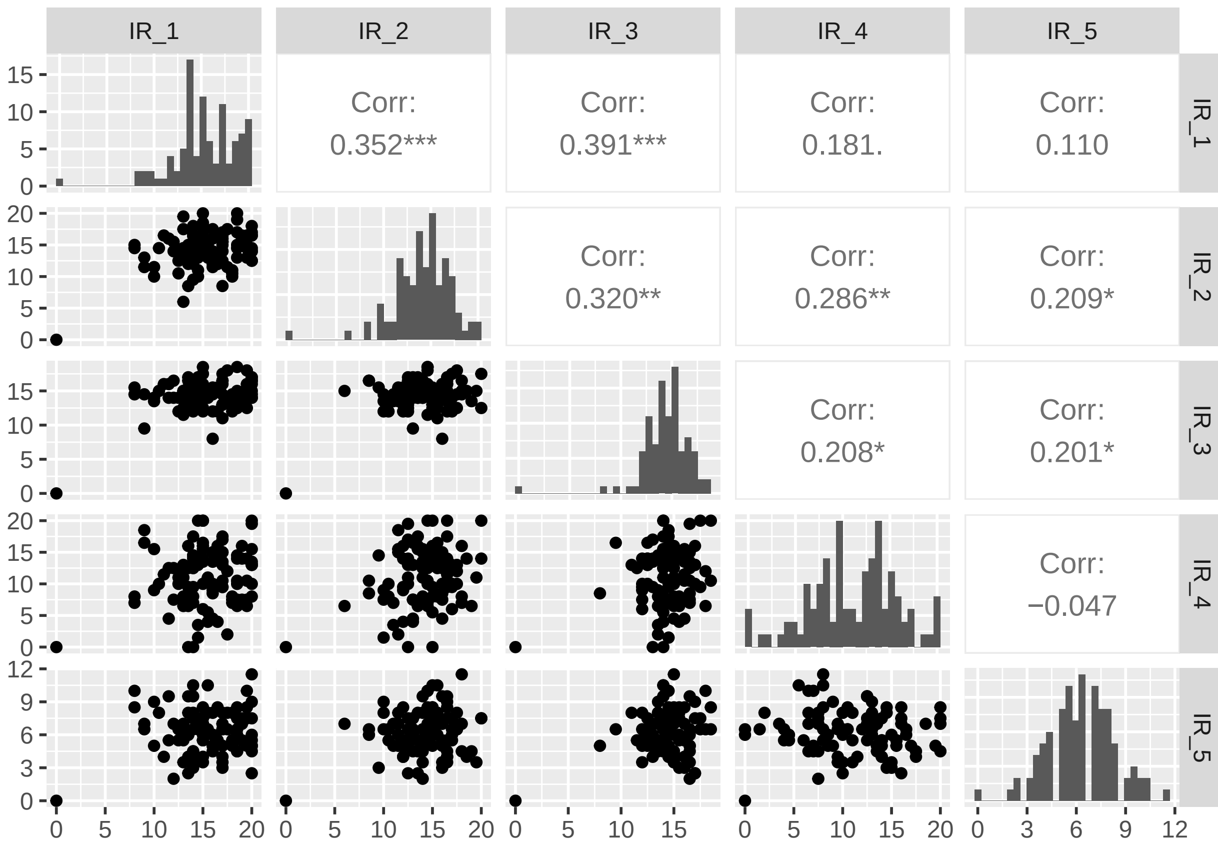
```
cor(rok, use="complete.obs")
```

```
##               MI_1         MI_2         MI_3         MI_4        MI_5
## MI_1   1.000000000  0.377360457 -0.010771416  0.014023348 -0.07448658
## MI_2   0.377360457  1.000000000 -0.168603163  0.064789672 -0.12665631
## MI_3  -0.010771416 -0.168603163  1.000000000 -0.009822388 -0.16335525
## MI_4   0.014023348  0.064789672 -0.009822388  1.000000000  0.15610481
## MI_5  -0.074486583 -0.126656313 -0.163355252  0.156104814  1.00000000
## LAB_1  0.397990988 -0.001591695 -0.122545094 -0.114967362  0.02412400
## ZI_1  -0.307352747 -0.357188976  0.127276265 -0.270127905 -0.17659707
## ZI_2  -0.080558559 -0.245407778 -0.017777209 -0.073876488 -0.07433605
## ZI_3  -0.302130563 -0.017153348 -0.285736503  0.009364368  0.05531232
## ZI_4   0.024323722 -0.045660473 -0.035855100 -0.039219888 -0.07114447
## ZI_5  -0.115440218 -0.132592599 -0.045865670 -0.117289034  0.04426251
## LAB_2 -0.008084342 -0.272164608 -0.053774180 -0.148830352  0.03263719
##              LAB_1        ZI_1        ZI_2         ZI_3        ZI_4        ZI_5
## MI_1   0.397990988 -0.30735275 -0.08055856 -0.302130563  0.02432372 -0.11544022
## MI_2  -0.001591695 -0.35718898 -0.24540778 -0.017153348 -0.04566047 -0.13259260
## MI_3  -0.122545094  0.12727627 -0.01777721 -0.285736503 -0.03585510 -0.04586567
## MI_4  -0.114967362 -0.27012791 -0.07387649  0.009364368 -0.03921989 -0.11728903
## MI_5   0.024123998 -0.17659707 -0.07433605  0.055312315 -0.07114447  0.04426251
## LAB_1  1.000000000 -0.35671775 -0.15827719 -0.169168201  0.05668903  0.01624348
## ZI_1  -0.356717750  1.00000000  0.25357564  0.270396400 -0.02172039  0.08851029
## ZI_2  -0.158277192  0.25357564  1.00000000  0.185817448  0.28103811 -0.01666102
## ZI_3  -0.169168201  0.27039640  0.18581745  1.000000000 -0.02872633 -0.06154188
## ZI_4   0.056689032 -0.02172039  0.28103811 -0.028726332  1.00000000 -0.10156006
## ZI_5   0.016243484  0.08851029 -0.01666102 -0.061541878 -0.10156006  1.00000000
## LAB_2  0.161287635  0.29433244  0.02962132  0.067230769 -0.04573784  0.03402904
##              LAB_2
## MI_1  -0.008084342
## MI_2  -0.272164608
## MI_3  -0.053774180
## MI_4  -0.148830352
## MI_5   0.032637194
## LAB_1  0.161287635
## ZI_1   0.294332443
## ZI_2   0.029621317
## ZI_3   0.067230769
## ZI_4  -0.045737838
## ZI_5   0.034029039
## LAB_2  1.000000000
```

Prikažite upareni graf za zadatke s ispitnog roka. Na dijagonalama prikažite empirijsku distribuciju podataka, a na elementima izvan dijagonala prikažite grafove raspršenja za parove varijabli. Razmislite o karakteristikama grafova i razmislite postoje li primjeri koji odskaču od ostalih.

```
rok_zadaci = df[, c('IR_1', 'IR_2','IR_3', 'IR_4', 'IR_5')]
rok_zadaci <- na.omit(rok_zadaci)
ggpairs(rok_zadaci, diag = list(continuous = "barDiag"))
```

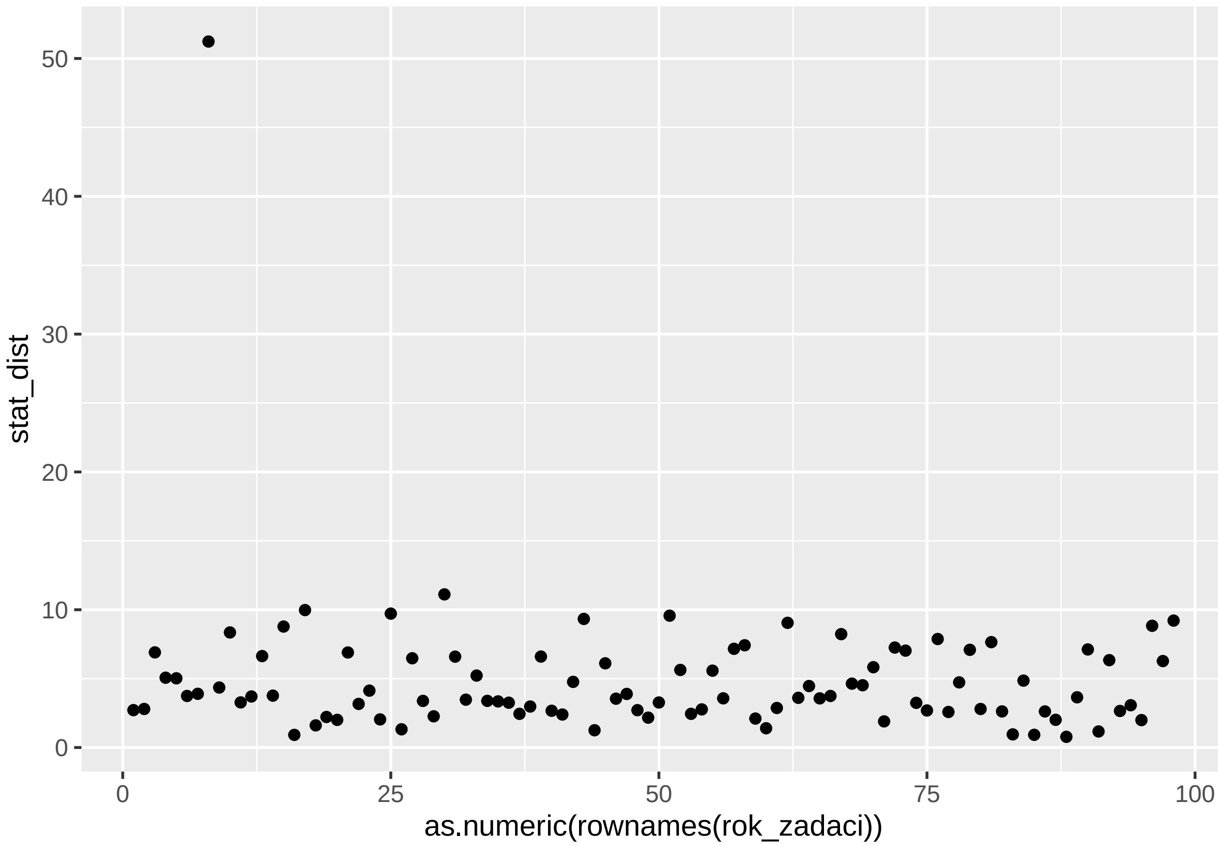
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



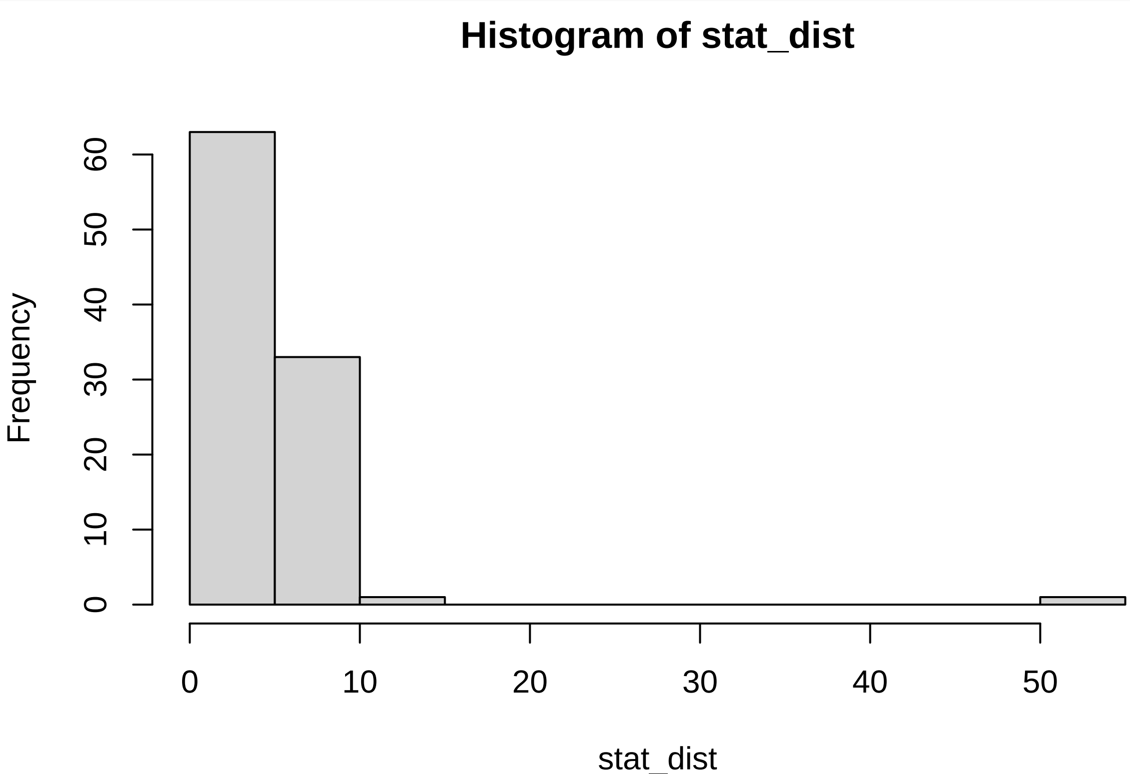
3.3 Statistička udaljenost

Izračunajte procjene vektora očekivanja i matrice kovarijance za zadatke s ispitnog roka, kao i statističke udaljenosti svih primjera u odnosu na procijenjeno očekivanje i kovarijancu. Ispitajte postoje li stršeće vrijednosti koje su statistički značajne.

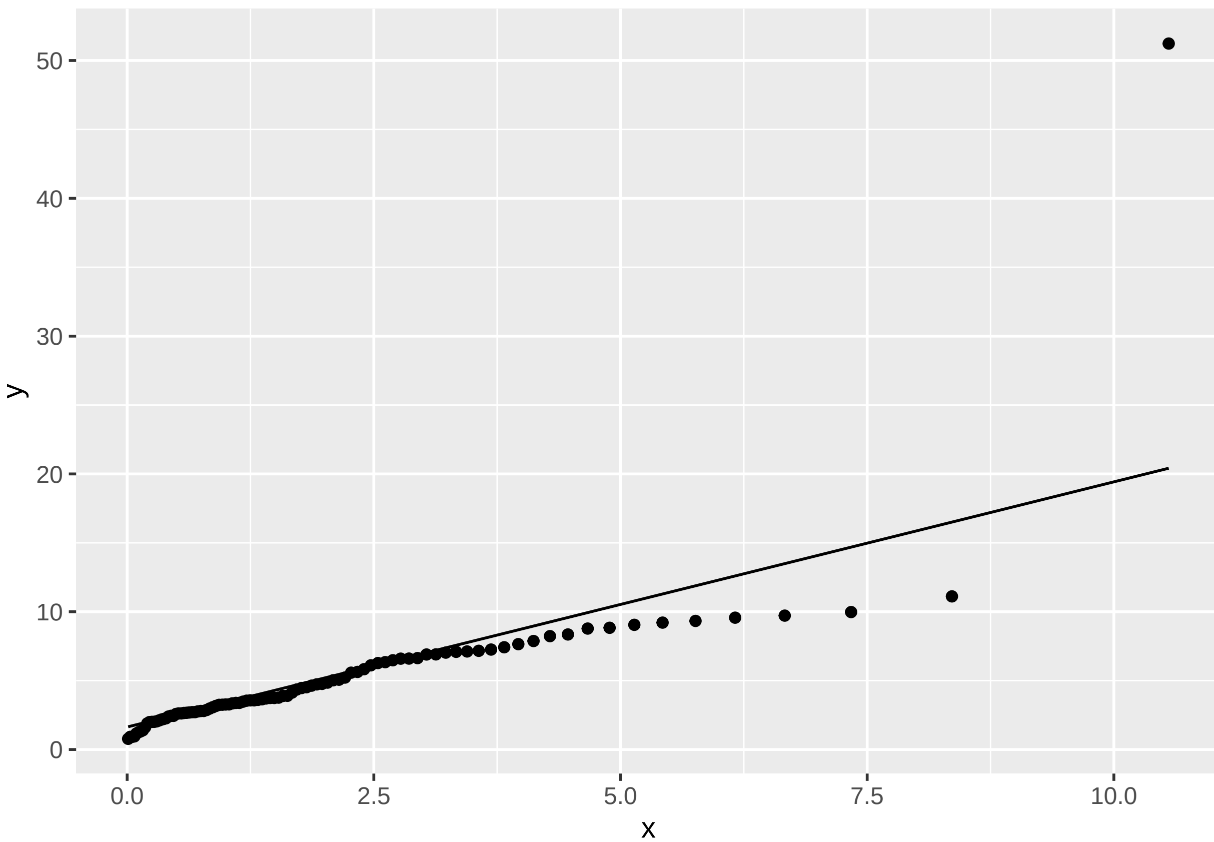
```
# Vaš kôd ovdje
stat_dist <- mahalanobis(rok_zadaci, colMeans(rok_zadaci), cov(rok_zadaci))

pval <- pchisq(stat_dist, df=dim(rok_zadaci)[2], lower.tail=FALSE)
rok_zadaci$stat_dist <- stat_dist

ggplot(rok_zadaci, aes(as.numeric(rownames(rok_zadaci)), stat_dist))+geom_point()
```




```
hist(stat_dist)
```



```
ggplot(rok_zadaci, aes(sample = stat_dist)) + stat_qq(distribution = qchisq, dparams = 2) +stat_qq_line
```



4. Analiza podataka

4.1 Vizualizacija i deskriptivna statistika

Analizirajte u podatcima sljedeća istraživačka pitanja, koristeći odgovarajuće vizualizacije i deskriptivne statistike ili druge tehnike (dodatno možete provesti i statistički test - nije obavezno).

- Imaju li grupe utjecaj na ukupne bodove iz kontinuirane nastave (postoje li grupe koje su uspješnije od ostalih)? Vrijedi li isto za bodove na roku?

```
# Vaš kôd ovdje

grupe_bodovi <- df[is.na(df$IR_5), c('Grupa')]
grupe_bodovi$ukupno <- kont_nastava$MI_1 + kont_nastava$MI_2 + kont_nastava$MI_3 + kont_nastava$MI_4 + I

grupe_bodovi[, c('Grupa')] <- lapply(grupe_bodovi[,  c('Grupa')], as.factor)

grupe_bodovi <- na.omit(grupe_bodovi)

basicStats(grupe_bodovi[grupe_bodovi$Grupa == 1, 'ukupno'])
```

```
##                  ukupno
## nobs         142.000000
## NAs            0.000000
## Minimum       50.500000
## Maximum       71.500000
```

```
## 1. Quartile   56.000000
## 3. Quartile   63.500000
## Mean          60.119718
## Median        61.000000
## Sum         8537.000000
## SE Mean        0.417500
## LCL Mean      59.294349
## UCL Mean      60.945088
## Variance      24.751523
## Stdev          4.975090
## Skewness      -0.132702
## Kurtosis      -0.770022
```

```r
basicStats(grupe_bodovi[grupe_bodovi$Grupa == 2, 'ukupno'])
```

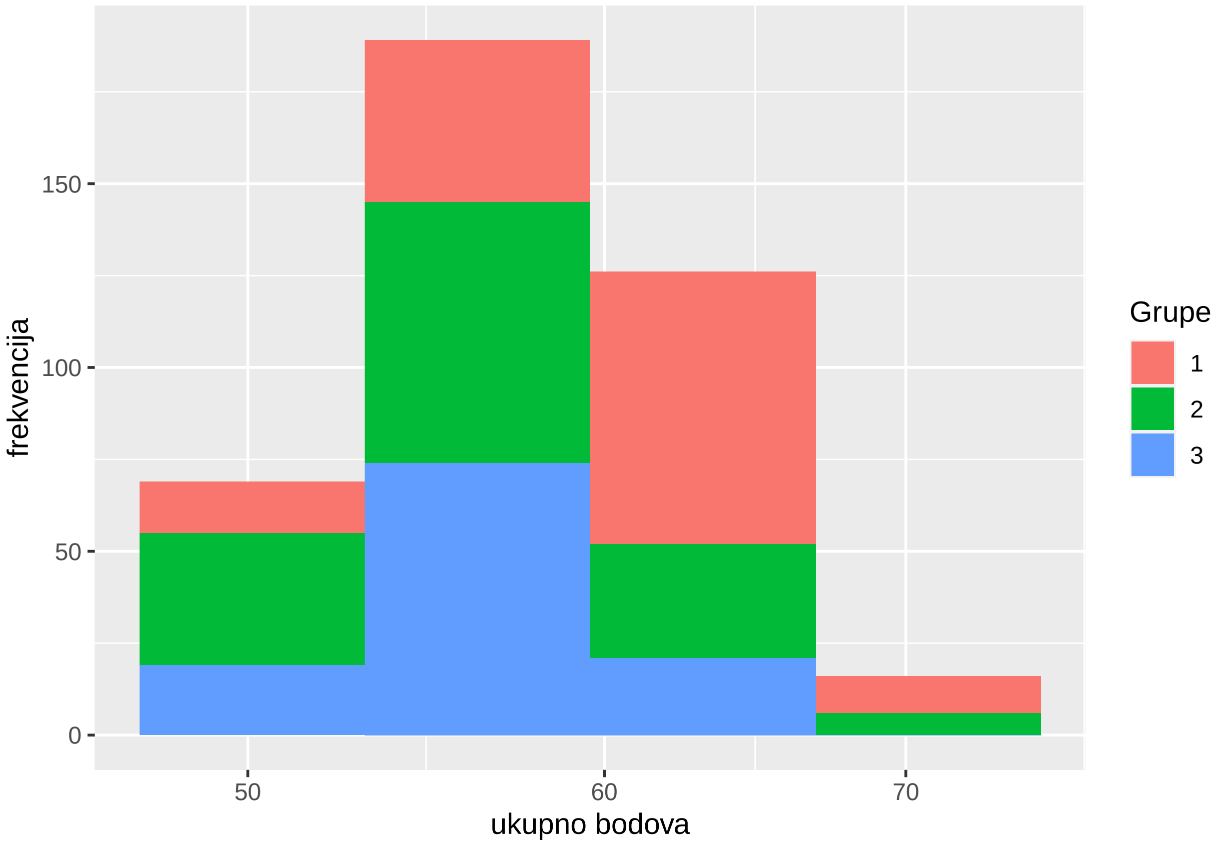
```
##                  ukupno
## nobs         144.000000
## NAs            0.000000
## Minimum       50.500000
## Maximum       71.500000
## 1. Quartile   53.375000
## 3. Quartile   60.000000
## Mean          56.875000
## Median        56.000000
## Sum         8190.000000
## SE Mean        0.391969
## LCL Mean      56.100198
## UCL Mean      57.649802
## Variance      22.124126
## Stdev          4.703629
## Skewness       0.808712
## Kurtosis       0.025157
```

```r
basicStats(grupe_bodovi[grupe_bodovi$Grupa == 3, 'ukupno'])
```

```
##                  ukupno
## nobs         114.000000
## NAs            0.000000
## Minimum       51.000000
## Maximum       64.500000
## 1. Quartile   54.500000
## 3. Quartile   59.000000
## Mean          56.627193
## Median        57.000000
## Sum         6455.500000
## SE Mean        0.287585
## LCL Mean      56.057436
## UCL Mean      57.196950
## Variance       9.428369
## Stdev          3.070565
## Skewness       0.013288
## Kurtosis      -0.634674
```

```r
ggplot(aes(x = ukupno), data = grupe_bodovi) +
  geom_histogram(aes(fill = Grupa), binwidth = 0.05) +
  scale_x_log10() +
```

```
  ylab("frekvencija") +
  xlab("ukupno bodova") +
  labs(fill = "Grupe")
```




```
grupe_bodovi_rok <- df[!is.na(df$IR_5),c('Grupa') ]
grupe_bodovi_rok$ukupno <- rok_zadaci$IR_1 + rok_zadaci$IR_2 + rok_zadaci$IR_3 + rok_zadaci$IR_4 + rok_z

grupe_bodovi_rok[, c('Grupa')] <- lapply(grupe_bodovi_rok[, c('Grupa')], as.factor)

grupe_bodovi_rok <- na.omit(grupe_bodovi_rok)

basicStats(grupe_bodovi_rok[grupe_bodovi_rok$Grupa == 1, 'ukupno'])
```

```
##                  ukupno
## nobs          15.000000
## NAs            0.000000
## Minimum        0.000000
## Maximum       66.500000
## 1. Quartile   54.250000
## 3. Quartile   64.000000
## Mean          55.566667
## Median        59.500000
## Sum          833.500000
## SE Mean        4.220378
## LCL Mean      46.514857
## UCL Mean      64.618476
## Variance     267.173810
```

```
## Stdev        16.345452
## Skewness     -2.547678
## Kurtosis      5.953262
```

```
basicStats(grupe_bodovi_rok[grupe_bodovi_rok$Grupa == 2, 'ukupno'])
```

```
##                  ukupno
## nobs          37.000000
## NAs            0.000000
## Minimum       47.000000
## Maximum       75.500000
## 1. Quartile   55.500000
## 3. Quartile   66.000000
## Mean          61.432432
## Median        62.000000
## Sum         2273.000000
## SE Mean        1.155214
## LCL Mean      59.089549
## UCL Mean      63.775316
## Variance      49.377252
## Stdev          7.026895
## Skewness      -0.102844
## Kurtosis      -0.926061
```

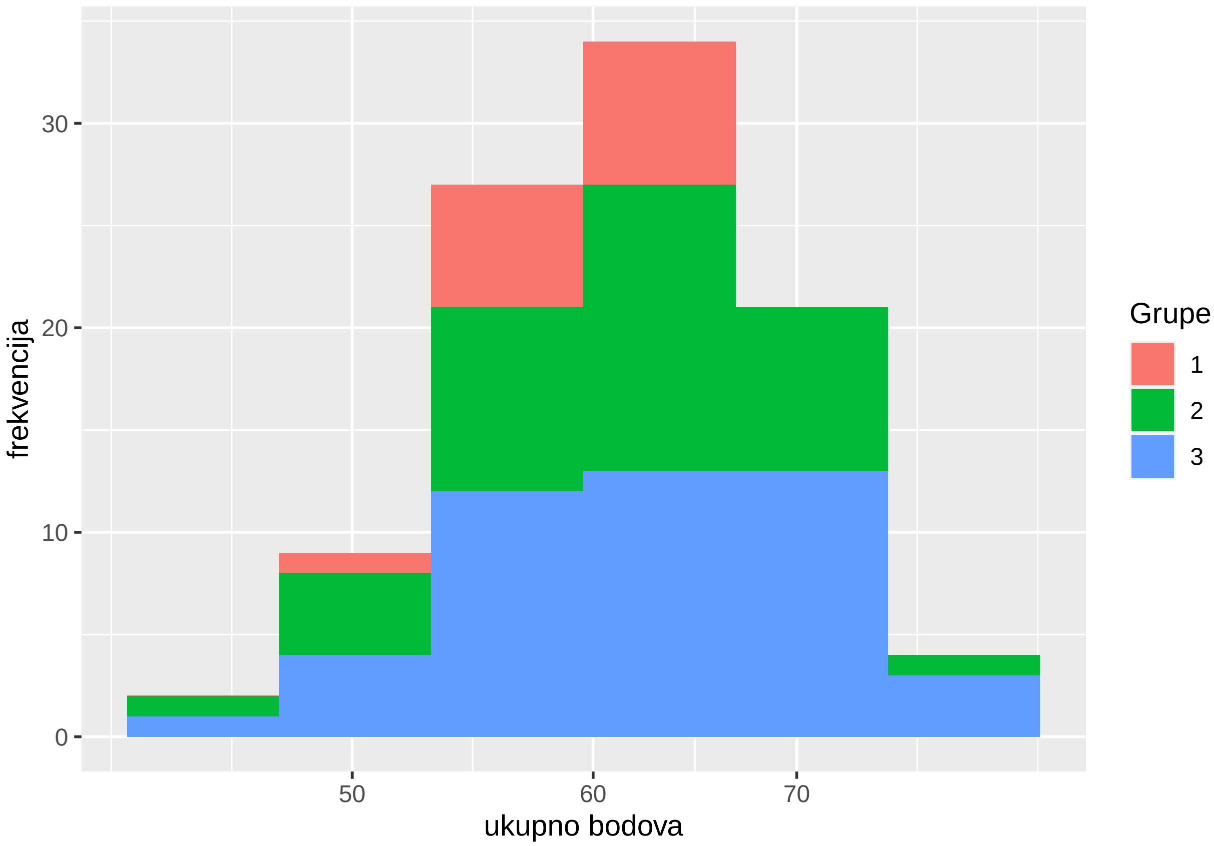
```
basicStats(grupe_bodovi_rok[grupe_bodovi_rok$Grupa == 3, 'ukupno'])
```

```
##                  ukupno
## nobs          46.000000
## NAs            0.000000
## Minimum       47.000000
## Maximum       80.000000
## 1. Quartile   55.625000
## 3. Quartile   68.875000
## Mean          62.913043
## Median        63.000000
## Sum         2894.000000
## SE Mean        1.170832
## LCL Mean      60.554868
## UCL Mean      65.271219
## Variance      63.058937
## Stdev          7.940966
## Skewness       0.004224
## Kurtosis      -0.888210
```

```
ggplot(aes(x = ukupno), data = grupe_bodovi_rok) +
  geom_histogram(aes(fill = Grupa), binwidth = 0.05) +
  scale_x_log10() +
  ylab("frekvencija") +
  xlab("ukupno bodova") +
  labs(fill = "Grupe")
```

```
## Warning: Transformation introduced infinite values in continuous x-axis
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```



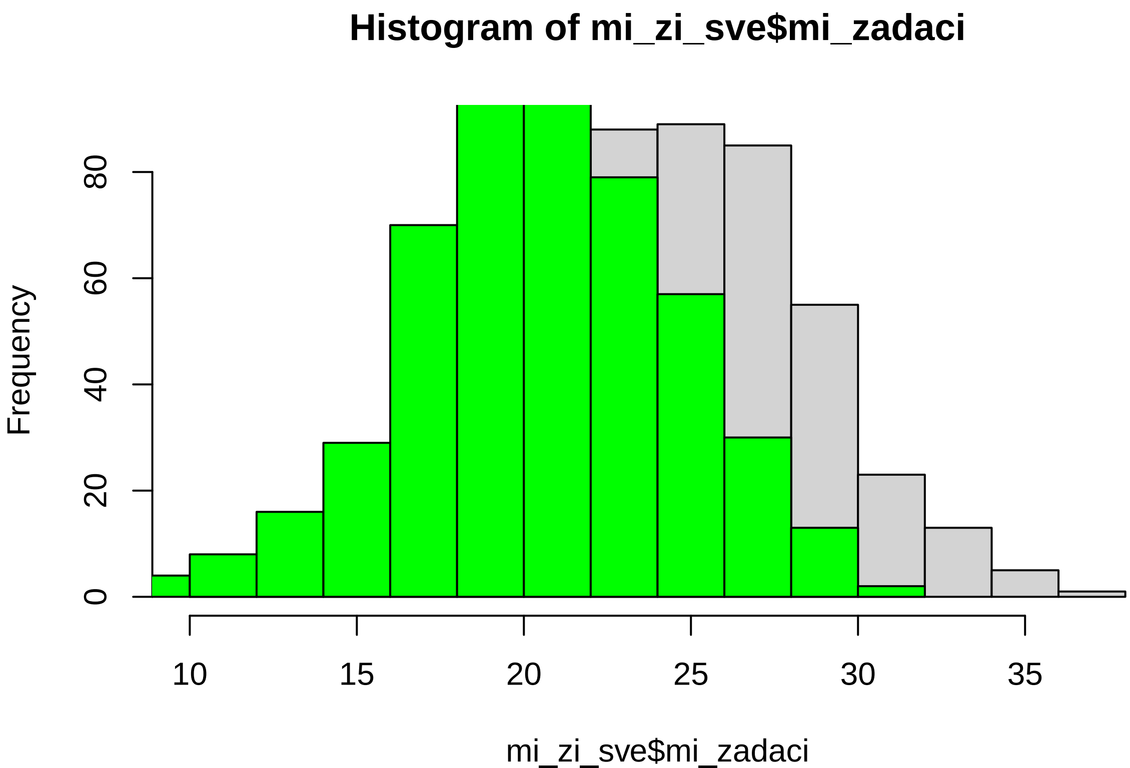
- Postoji li povezanost između uspjeha studenata na međuispitu i završnom ispitu (vrijedi li da su uspješniji studenti na MI ujedno uspješniji i na ZI)?

```
# Vaš kôd ovdje

mi_zadaci <- c(df$MI_1 + df$MI_2 + df$MI_3 + df$MI_4 + df$MI_5)

zi_zadaci <- c(df$ZI_1 + df$ZI_2 + df$ZI_3 + df$ZI_4 + df$ZI_5)

mi_zi_sve <- data.frame(mi_zadaci, zi_zadaci)
basicStats(mi_zi_sve)
```

```
##                 mi_zadaci    zi_zadaci
## nobs          498.000000   498.000000
## NAs             0.000000     0.000000
## Minimum        10.500000     8.000000
## Maximum        37.000000    31.000000
## 1. Quartile    22.000000    18.000000
## 3. Quartile    27.500000    24.000000
## Mean           24.736948    20.836345
## Median         24.500000    21.000000
## Sum         12319.000000 10376.500000
## SE Mean         0.186020     0.179705
## LCL Mean       24.371466    20.483271
## UCL Mean       25.102430    21.189420
## Variance       17.232475    16.082318
## Stdev           4.151202     4.010277
## Skewness       -0.002103    -0.197503
```

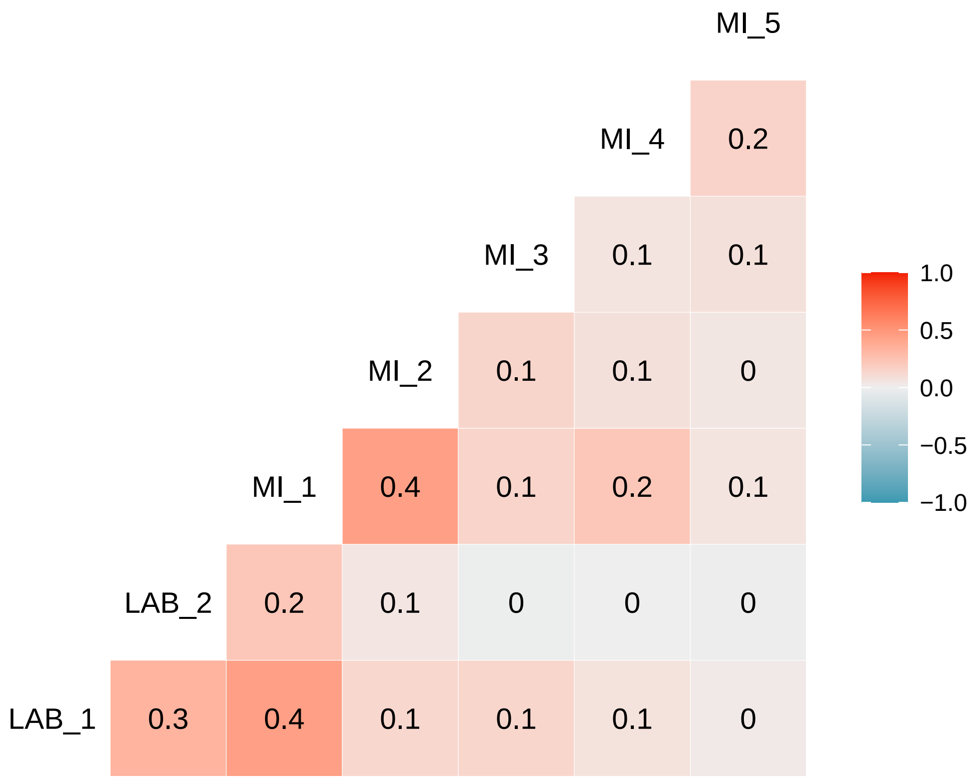
```
## Kurtosis        -0.187827    -0.097357
```

```
hist(mi_zi_sve$mi_zadaci)
hist(mi_zi_sve$zi_zadaci, col='green', add=TRUE)
```

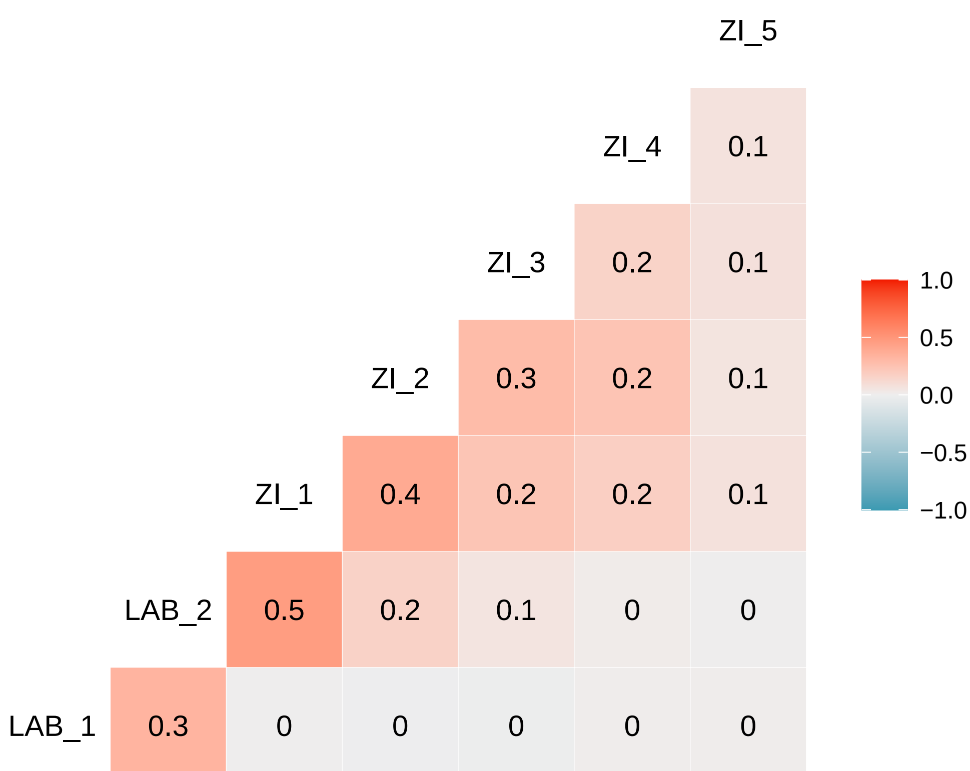


- Postoji li povezanost između uspjeha studenata na nekim zadatcima na ispitima i pojedinim laboratorijskim vježbama? Razmislite koji su mogući uzroci ovakvih zavisnosti, ako postoje.

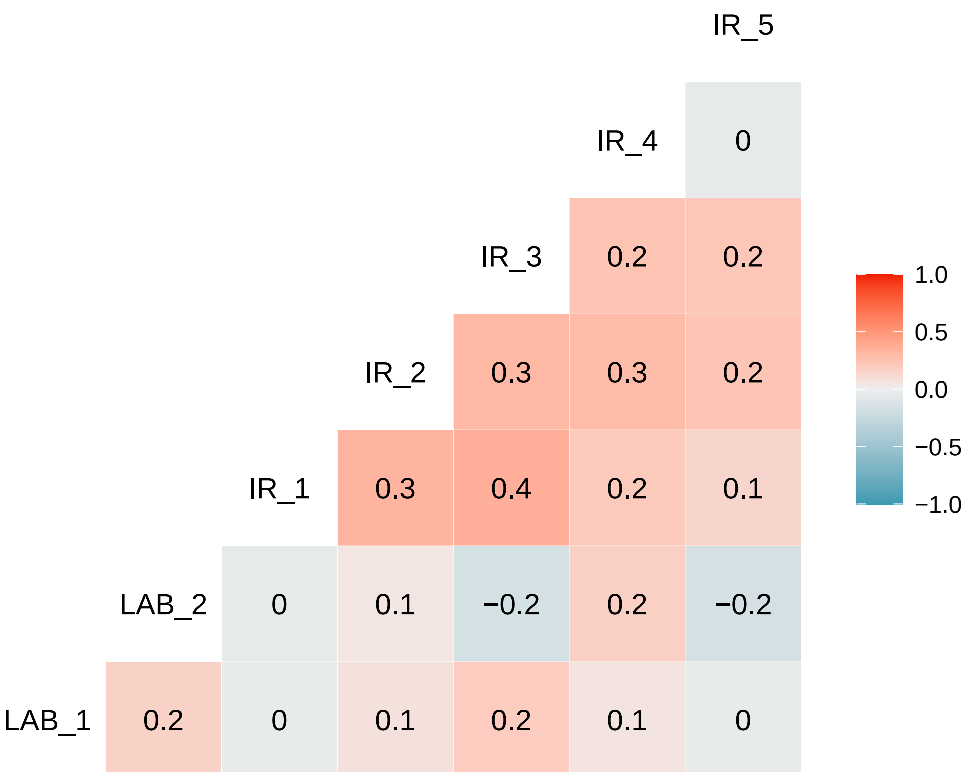
```
# Vaš kôd ovdje

mi_lab_kont = df[, c('LAB_1', 'LAB_2', 'MI_1', 'MI_2','MI_3', 'MI_4', 'MI_5')]
mi_lab_kont <- na.omit(mi_lab_kont)
ggcorr(mi_lab_kont, label = T)
```




```
zi_lab_kont = df[, c('LAB_1', 'LAB_2', 'ZI_1', 'ZI_2','ZI_3', 'ZI_4', 'ZI_5')]
zi_lab_kont <- na.omit(zi_lab_kont)
ggcorr(zi_lab_kont, label = T)
```



```r
rok_sve = df[, c('LAB_1', 'LAB_2', 'IR_1', 'IR_2','IR_3', 'IR_4', 'IR_5')]

rok_sve <- na.omit(rok_sve)
ggcorr(rok_sve, label = T)
```



Postavite i analizirajte na ovaj način još barem jedno vlastito istraživačko pitanje.

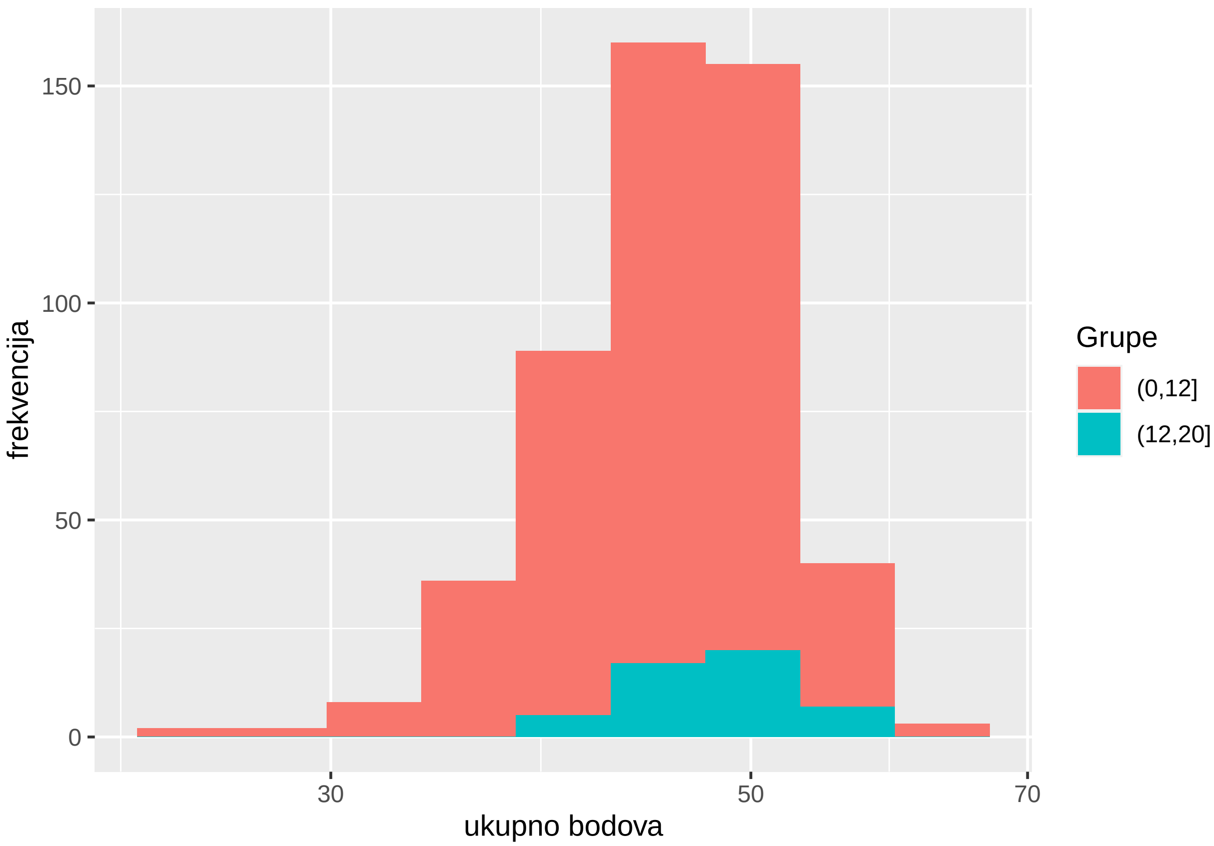
Kako uspjeh na kontinuiranoj nastavi ovisi o laboratorijskim vjezbama?

```r
# Vaš kôd ovdje

lab_bodovi <- c(df$LAB_1 + df$LAB_2)
kontinuirano <-  c(df$MI_1 + df$MI_2 + df$MI_3 + df$MI_4 + df$MI_5 + df$ZI_1 + df$ZI_2 + df$ZI_3 + df$Z

lab_analiza <- data.frame(lab_bodovi, kontinuirano)
lab_analiza <- na.omit(lab_analiza)
basicStats(lab_analiza)
```

```
##              lab_bodovi kontinuirano
## nobs         495.000000   495.000000
## NAs            0.000000     0.000000
## Minimum        4.500000    26.000000
## Maximum       14.000000    62.000000
## 1. Quartile    9.000000    42.000000
## 3. Quartile   11.000000    50.000000
## Mean          10.004040    45.623232
## Median        10.000000    46.000000
## Sum         4952.000000 22583.500000
## SE Mean        0.073694     0.267468
## LCL Mean       9.859248    45.097717
```

```
## UCL Mean      10.148833    46.148748
## Variance       2.688243    35.411909
## Stdev          1.639586     5.950791
## Skewness      -0.024680    -0.243555
## Kurtosis       0.015911     0.073125
```

```
lab_analiza$group = cut(lab_analiza$lab_bodovi, c(0, 12, 20))

ggplot(aes(x = kontinuirano), data = lab_analiza) +
  geom_histogram(aes(fill = group), binwidth = 0.05) +
  scale_x_log10() +
  ylab("frekvencija") +
  xlab("ukupno bodova") +
  labs(fill = "Grupe")
```



4.2. Regresijska analiza

Razmotrimo u kakvom su odnosu zadatci ispitnog roka s ostalim aktivnostima iz kontinuirane nastave. Istražite odnos koristeći model multivarijatne linearne regresije. Procijenite model gdje su zavisne varijable bodovi zadataka s ispitnog roka, odaberite konačni skup ulaznih varijabli i provjerite adekvatnost modela.

```
rok_regresija <- na.omit(df)
rok_regresija <- select(rok_regresija, -Grupa)
mlm.fit <- lm(cbind(IR_1, IR_2, IR_3, IR_4, IR_5) ~ ., data = rok_regresija)
summary(mlm.fit)
```

```
## Response IR_1 :
##
```

```
## Call:
## lm(formula = IR_1 ~ MI_1 + MI_2 + MI_3 + MI_4 + MI_5 + LAB_1 +
##     ZI_1 + ZI_2 + ZI_3 + ZI_4 + ZI_5 + LAB_2, data = rok_regresija)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.7884  -1.3134  -0.0987   1.4226   5.3250
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.544430   4.953401   1.927   0.0575 .
## MI_1         2.138095   0.477136   4.481 2.38e-05 ***
## MI_2        -0.489640   0.214732  -2.280   0.0252 *
## MI_3         0.213703   0.186342   1.147   0.2548
## MI_4         0.291496   0.363211   0.803   0.4246
## MI_5        -0.271616   0.222206  -1.222   0.2251
## LAB_1       -0.758607   0.368609  -2.058   0.0428 *
## ZI_1        -0.122925   0.245884  -0.500   0.6185
## ZI_2        -0.091031   0.392946  -0.232   0.8174
## ZI_3        -0.009219   0.218738  -0.042   0.9665
## ZI_4        -0.142278   0.351937  -0.404   0.6871
## ZI_5         0.065407   0.350730   0.186   0.8525
## LAB_2       -0.154722   0.427607  -0.362   0.7184
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.008 on 82 degrees of freedom
## Multiple R-squared:  0.2708, Adjusted R-squared:  0.1641
## F-statistic: 2.538 on 12 and 82 DF,  p-value: 0.006735
##
##
## Response IR_2 :
##
## Call:
## lm(formula = IR_2 ~ MI_1 + MI_2 + MI_3 + MI_4 + MI_5 + LAB_1 +
##     ZI_1 + ZI_2 + ZI_3 + ZI_4 + ZI_5 + LAB_2, data = rok_regresija)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.273  -1.565   0.233   1.636   4.917
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.3858     4.5392   1.847 0.068296 .
## MI_1         -1.0873     0.4372  -2.487 0.014923 *
## MI_2          0.2129     0.1968   1.082 0.282413
## MI_3         -0.0412     0.1708  -0.241 0.809937
## MI_4          1.2586     0.3328   3.781 0.000295 ***
## MI_5          0.4120     0.2036   2.024 0.046278 *
## LAB_1         0.7336     0.3378   2.172 0.032759 *
## ZI_1          0.1967     0.2253   0.873 0.385229
## ZI_2          0.2079     0.3601   0.577 0.565284
## ZI_3         -0.1504     0.2004  -0.750 0.455206
## ZI_4         -0.2805     0.3225  -0.870 0.386977
```

```
## ZI_5          0.2177     0.3214   0.677 0.500032
## LAB_2         0.2068     0.3918   0.528 0.599038
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.756 on 82 degrees of freedom
## Multiple R-squared:  0.2641, Adjusted R-squared:  0.1564
## F-statistic: 2.452 on 12 and 82 DF,  p-value: 0.008773
##
##
## Response IR_3 :
##
## Call:
## lm(formula = IR_3 ~ MI_1 + MI_2 + MI_3 + MI_4 + MI_5 + LAB_1 +
##     ZI_1 + ZI_2 + ZI_3 + ZI_4 + ZI_5 + LAB_2, data = rok_regresija)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.3777  -0.7981   0.1147   0.9154   3.4616
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.80333    3.12106   3.141  0.00234 **
## MI_1        -0.06572    0.30064  -0.219  0.82750
## MI_2         0.11596    0.13530   0.857  0.39391
## MI_3        -0.36394    0.11741  -3.100  0.00265 **
## MI_4         0.44812    0.22885   1.958  0.05362 .
## MI_5         0.12642    0.14001   0.903  0.36919
## LAB_1        0.99309    0.23225   4.276 5.11e-05 ***
## ZI_1         0.88372    0.15493   5.704 1.80e-07 ***
## ZI_2        -0.35330    0.24759  -1.427  0.15739
## ZI_3        -0.16485    0.13782  -1.196  0.23509
## ZI_4        -0.44956    0.22175  -2.027  0.04588 *
## ZI_5        -0.04814    0.22099  -0.218  0.82808
## LAB_2       -1.13574    0.26943  -4.215 6.38e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.895 on 82 degrees of freedom
## Multiple R-squared:  0.4182, Adjusted R-squared:  0.333
## F-statistic: 4.911 on 12 and 82 DF,  p-value: 4.743e-06
##
##
## Response IR_4 :
##
## Call:
## lm(formula = IR_4 ~ MI_1 + MI_2 + MI_3 + MI_4 + MI_5 + LAB_1 +
##     ZI_1 + ZI_2 + ZI_3 + ZI_4 + ZI_5 + LAB_2, data = rok_regresija)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.8426 -2.7981  0.0664  2.5251 10.7818
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.00216    6.94332  -0.720   0.4733
## MI_1        -0.06197    0.66881  -0.093   0.9264
## MI_2        -0.31860    0.30100  -1.058   0.2929
## MI_3         0.20559    0.26120   0.787   0.4335
## MI_4         0.87433    0.50912   1.717   0.0897 .
## MI_5        -0.07301    0.31147  -0.234   0.8153
## LAB_1        0.47476    0.51669   0.919   0.3609
## ZI_1        -0.37358    0.34466  -1.084   0.2816
## ZI_2         2.43891    0.55080   4.428 2.91e-05 ***
## ZI_3         0.12188    0.30661   0.397   0.6920
## ZI_4        -1.01360    0.49332  -2.055   0.0431 *
## ZI_5        -0.25692    0.49163  -0.523   0.6027
## LAB_2        0.96985    0.59939   1.618   0.1095
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.216 on 82 degrees of freedom
## Multiple R-squared:  0.2932, Adjusted R-squared:  0.1898
## F-statistic: 2.835 on 12 and 82 DF,  p-value: 0.002679
##
##
## Response IR_5 :
##
## Call:
## lm(formula = IR_5 ~ MI_1 + MI_2 + MI_3 + MI_4 + MI_5 + LAB_1 +
##     ZI_1 + ZI_2 + ZI_3 + ZI_4 + ZI_5 + LAB_2, data = rok_regresija)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7666 -0.7346 -0.0134  1.0709  3.9694
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.8774100  2.9669495   0.633   0.5286
## MI_1        -0.1107108  0.2857909  -0.387   0.6995
## MI_2         0.1916578  0.1286184   1.490   0.1400
## MI_3         0.0099311  0.1116134   0.089   0.9293
## MI_4         0.1718900  0.2175530   0.790   0.4317
## MI_5         0.1041226  0.1330952   0.782   0.4363
## LAB_1        0.0167144  0.2207864   0.076   0.9398
## ZI_1        -0.0004482  0.1472779  -0.003   0.9976
## ZI_2         0.1871975  0.2353638   0.795   0.4287
## ZI_3         0.0284725  0.1310178   0.217   0.8285
## ZI_4         0.3595076  0.2108005   1.705   0.0919 .
## ZI_5         1.1935463  0.2100778   5.681 1.98e-07 ***
## LAB_2       -0.3118738  0.2561249  -1.218   0.2268
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.801 on 82 degrees of freedom
## Multiple R-squared:  0.3295, Adjusted R-squared:  0.2314
## F-statistic: 3.359 on 12 and 82 DF,  p-value: 0.000525
```

```r
coef(mlm.fit)
```

```
##                    IR_1        IR_2        IR_3        IR_4          IR_5
## (Intercept)  9.544429783  8.38577891  9.80332557 -5.00216171  1.8774099565
## MI_1         2.138095230 -1.08728305 -0.06572073 -0.06197477 -0.1107108324
## MI_2        -0.489640029  0.21291839  0.11596042 -0.31859885  0.1916577537
## MI_3         0.213703308 -0.04120192 -0.36394029  0.20558638  0.0099311112
## MI_4         0.291496076  1.25855327  0.44811762  0.87433062  0.1718899807
## MI_5        -0.271616245  0.41204058  0.12642265 -0.07300889  0.1041226294
## LAB_1       -0.758606965  0.73361334  0.99308518  0.47475607  0.0167144450
## ZI_1        -0.122925052  0.19670187  0.88372100 -0.37357570 -0.0004482244
## ZI_2        -0.091030977  0.20789906 -0.35329915  2.43891489  0.1871974944
## ZI_3        -0.009219295 -0.15040178 -0.16485470  0.12187576  0.0284724990
## ZI_4        -0.142278329 -0.28050312 -0.44956074 -1.01359521  0.3595075556
## ZI_5         0.065406843  0.21773342 -0.04814382 -0.25692124  1.1935462602
## LAB_2       -0.154722080  0.20683474 -1.13573913  0.96985499 -0.3118737536
```

```r
resid(mlm.fit)
```

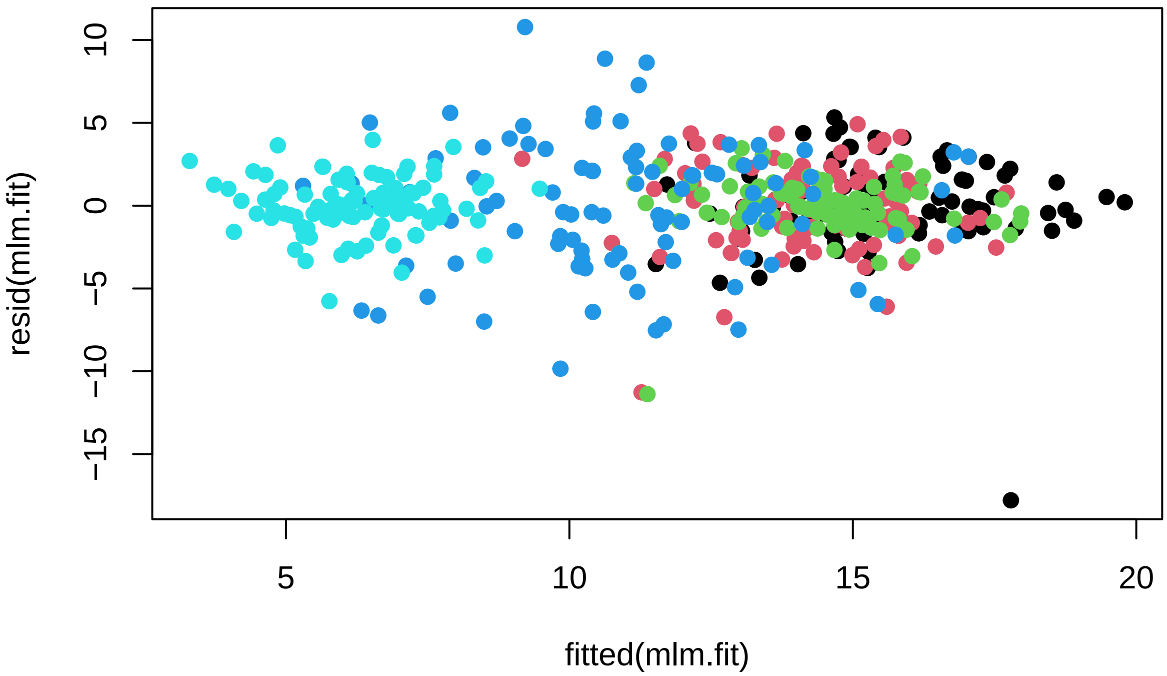
```
##           IR_1          IR_2         IR_3        IR_4        IR_5
## 1   -1.17756382   3.209433612  -0.71388362  2.63027534 -0.01342843
## 2    4.34098639   3.740420244   0.84842479 -3.26005102  0.20640080
## 3    3.55690484   4.917186154  -1.34179091  1.81852272  0.78205621
## 4   -0.82029337   1.050542981   0.75068054  0.81514241  1.02335606
## 5   -1.54659261  -1.031240175  -0.38238734 -0.99137153  1.07284890
## 6    1.43999547  -2.464071729   2.69568269  1.35769856 -2.75780516
## 7   -0.97907049   1.425729837   0.79690689  2.08741249  1.88606613
## 8  -17.78835078 -11.272532804 -11.37774529 -9.84256426 -5.76664514
## 9    1.90650161   1.185814223   1.77013966  0.70260257 -4.04341899
## 10   3.33800036  -1.250850187   1.26022376 -0.39386135 -2.66332761
## 11  -0.83006929  -1.948505148   0.15583083 -3.62112738 -0.34484892
## 12   1.21435404   0.302886508   0.66154938  3.65668296  1.01542503
## 13  -0.44464530  -2.848754560  -0.98658950 -2.71488554  0.74880484
## 14   3.53779630   0.728516805   2.58614063  0.79454311  0.33739658
## 15   1.27766319   3.962339971   0.32234108 -0.71406525 -2.59948161
## 16  -0.06988915   0.336494252  -1.16863431 -1.83979483  1.86085907
## 17  -4.65583762  -1.371771903  -1.40157040 -0.90685650  1.46853090
## 18  -0.20793063   2.281675761  -0.11194530 -0.01215008  0.21212444
## 19  -1.80135865  -1.975276030  -1.37643324  0.75897899 -1.37781425
## 20  -1.68850893  -0.973579083   0.76881729 -0.39012354 -0.74177948
## 21   0.25086471  -3.097915595   0.73476514  2.85880146 -0.49982581
## 22   4.37433347  -0.084024341   1.40327495  1.34077668 -0.57737188
## 23   4.10062128   2.653340598  -0.56786527  1.83700382 -0.83238507
## 24  -0.47852711   0.039326382  -0.98741994 -1.53915068 -3.34711824
## 25   0.48155828   1.138200161  -0.58158221 -0.53117558  0.74757551
## 26   4.10877140   2.887416310   2.40661485  3.22046027 -2.39811393
## 27   1.82321940   0.232955960   1.02594918  1.67537275 -1.81254177
## 28  -2.78004283  -3.250516349  -0.68733958 -3.14239662  0.71046262
## 29   1.15457126   1.984973979  -2.87687110 -3.32954269 -0.48786936
## 30  -2.01183030   2.382549993  -0.78668836  5.01852621 -2.41229144
## 31  -0.91557519   1.545216049   1.35546607  1.01300299 -1.03132299
## 32  -0.05931295   2.441133809  -0.93603598  4.05330654  0.27592636
## 33  -0.73393907  -1.809456772  -0.20278312 -0.68252329  2.35487694
## 34   1.81983351   0.840374532   2.56261147 -2.30412847  0.47452227
## 35   1.40527261   1.957019146   1.48697042  5.08211466 -0.68652900
```

```
## 36  -1.53888435  -1.278348484  -0.97647790 -6.41486538  2.07271964
## 37  -0.50647599  -1.189186550  -0.35922842  3.34873135 -0.47183810
## 38  -0.64886393  -6.097993384   0.42534959  3.31007869 -2.97998088
## 39   0.67744718  -2.371221629   0.83493868  3.42052108 -0.89082596
## 40  -1.41233111   1.311012718   0.31049208  0.06643296  0.79292919
## 41  -0.32001459  -0.746760218  -3.46673794 -5.93970333 -1.92097250
## 42  -0.43776131  -1.386414553   1.16871171  2.41986815 -0.61731658
## 43   2.75124961  -1.505237020   1.36691261 -5.49986578  1.92631899
## 44   2.95759897   4.358592881   0.80485040  3.72063858  1.91285396
## 45   1.31711506  -2.992861113   0.43424958 -7.52740235 -0.06786594
## 46   2.40694232   2.406247130   0.08182690  4.81334253 -0.27260770
## 47  -1.10911342   2.818442093   0.15292882  1.32135328  0.70422292
## 48   2.22406583   0.367076892   3.08625054  5.60237818 -0.83002428
## 49  -1.62383108  -6.733521383   0.18575428 -3.66369731 -0.61107436
## 50   1.02835311  -0.648685953  -0.85446719 -7.48290469  3.96942485
## 51   0.52443715   0.003489933   0.59595392 -0.26759242  0.66653889
## 52  -0.25184030   0.234189144   1.15313204  1.89542222  1.71617271
## 53  -1.66601146  -2.859053114  -1.44720334 -6.99725748  0.28762373
## 54  -1.51340595   1.749905456   0.06298289  0.93074595 -0.48628712
## 55  -1.69496658   0.908450712  -3.04781413 -6.33383114  0.08277105
## 56  -1.32455348   0.122810001  -0.95201083  5.09604823 -1.58271819
## 57   0.35395301   1.387423631  -0.01088626  1.98519370  2.35599049
## 58   0.36682284   0.768054156   0.05007992  2.27666136  1.57037862
## 59  -0.29614305   1.095820054  -0.46970704 -2.19988359 -0.22203106
## 60  -4.35123898  -3.714126576   0.17264862  7.27694895 -0.15825780
## 61  -0.90358744  -2.087772181  -1.43784617 -2.05921586 -1.19076247
## 62  -1.37315288  -2.128748783   0.28230730  1.74121846  1.08025906
## 63   0.51071511  -2.056206441  -0.30842436 -4.03771983 -0.72740297
## 64  -3.53262132   0.105149313  -0.44124908 -3.56894915  2.34421017
## 65  -0.01187520  -1.095454180   2.66300758  8.63773035  1.98387629
## 66   0.03655134   0.787871195  -0.80948648  2.91522933 -1.26279591
## 67  -1.66288428  -1.040595520  -0.03873085  2.95986836  0.26816456
## 68  -1.22901562  -3.447183231  -0.55128806 -6.62923771 -0.47418851
## 69  -2.73933778  -0.347939351  -0.57251608  2.10483204 -0.70602615
## 70   0.20325794   1.002893895   3.46155538  8.87163634  1.26709282
## 71  -3.76182525   0.437087187  -1.77411154 -3.49540149 -1.80184837
## 72   2.63367869   0.855944668  -1.47697327 -1.10843866 -0.27055414
## 73   0.70828666  -1.623847191  -0.42463488 -1.75530697 -0.44140248
## 74   1.57549677   3.595228550  -1.17544995 -3.78082654 -0.66909010
## 75  -0.57499564  -0.794502072  -0.75579914 -7.16404447  2.69759784
## 76  -0.37688649   1.694229446  -2.67630469 -5.19921613 -1.78515440
## 77  -0.76660864   2.280828808   0.61307817 -3.22542629 -0.28237591
## 78   3.53684936   0.181924781   0.72601614  5.56570671  1.41004276
## 79   0.29389415   4.155186001   1.79271975 10.78178653  1.42974581
## 80   2.95143671   3.831575807   1.68177886 -1.11545584  3.64302676
## 81  -2.18833686  -2.611656028  -1.32202761 -0.57067945 -3.00449989
## 82  -3.27091156  -2.812716778  -0.07394828  3.74350251  0.36649398
## 83  -0.09870482  -0.379260094   1.55026495  2.03655767  0.45281075
## 84   0.65472745  -2.457494662   0.99870038 -1.80033950  1.10018679
## 85  -1.30223624   2.831447171  -1.38738225  0.28369056 -1.63018355
## 86  -1.43593834  -0.261140845   0.98243898 -0.98219104  1.06894083
## 87   1.50634030   2.380769671  -0.97104890 -5.09794396 -0.39640854
## 88   4.72703726   0.705330659   1.76445294  1.19789685  2.38449101
## 89  -0.35013787  -0.394471255   0.50601996  2.32411417  0.28701234
```

```
## 90  -0.09064412  -2.249840982   1.12778683 -0.04084796  0.03468811
## 91   0.86735732   1.578644261   0.63722922  3.68441464  0.05863280
## 92   2.82750954   2.349380641   0.37601788  3.52258518 -0.17587543
## 93   3.78285390   1.327564845  -0.56188668 -0.59888710 -0.18726630
## 94  -3.52572714  -2.528215757   0.11470558 -4.92223359  1.85386565
## 95   5.32497549   4.342831034   1.07367467 -2.88122583  3.54523863
```

```r
## define our own "rstandard" method for "mlm" class
rstandard.mlm <- function (model) {
  Q <- with(model, qr.qy(qr, diag(1, nrow = nrow(qr$qr), ncol = qr$rank)))  ## Q matrix
  hii <- rowSums(Q ^ 2)  ## diagonal of hat matrix QQ'
  RSS <- colSums(model$residuals ^ 2)  ## residual sums of squares (for each model)
  sigma <- sqrt(RSS / model$df.residual)  ##  ## Pearson estimate of residuals (for each model)
  pointwise_sd <- outer(sqrt(1 - hii), sigma)  ## point-wise residual standard error (for each model)
  model$residuals / pointwise_sd  ## standardised residuals
}
```

```r
plot(fitted(mlm.fit), resid(mlm.fit), col = as.numeric(col(fitted(mlm.fit))), pch = 19)
```




```r
sigma(mlm.fit)
```

```
##      IR_1     IR_2     IR_3     IR_4     IR_5
## 3.007643 2.756166 1.895068 4.215900 1.801495
```

```r
mlm2.fit <- update(mlm.fit, . ~ . - ZI_3 - ZI_4)
anova(mlm.fit, mlm2.fit)
```

```
## Analysis of Variance Table
##
## Model 1: cbind(IR_1, IR_2, IR_3, IR_4, IR_5) ~ MI_1 + MI_2 + MI_3 + MI_4 +
##     MI_5 + LAB_1 + ZI_1 + ZI_2 + ZI_3 + ZI_4 + ZI_5 + LAB_2
## Model 2: cbind(IR_1, IR_2, IR_3, IR_4, IR_5) ~ MI_1 + MI_2 + MI_3 + MI_4 +
##     MI_5 + LAB_1 + ZI_1 + ZI_2 + ZI_5 + LAB_2
##   Res.Df Df Gen.var.  Pillai approx F num Df den Df  Pr(>F)
## 1     82      4.8631
```

```
## 2     84  2   4.9751 0.21692   1.9222     10    158 0.04577 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
Anova(mlm2.fit)
```

```
##
## Type II MANOVA Tests: Pillai test statistic
##       Df test stat approx F num Df den Df    Pr(>F)
## MI_1   1   0.51819  17.2082      5     80 1.588e-11 ***
## MI_2   1   0.29180   6.5926      5     80 3.477e-05 ***
## MI_3   1   0.32922   7.8528      5     80 4.687e-06 ***
## MI_4   1   0.20821   4.2075      5     80 0.0019074 **
## MI_5   1   0.22531   4.6534      5     80 0.0008855 ***
## LAB_1  1   0.53389  18.3267      5     80 4.405e-12 ***
## ZI_1   1   0.57013  21.2208      5     80 1.900e-13 ***
## ZI_2   1   0.43651  12.3944      5     80 6.536e-09 ***
## ZI_5   1   0.28356   6.3325      5     80 5.312e-05 ***
## LAB_2  1   0.38358   9.9565      5     80 1.975e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
lh.out <- linearHypothesis(mlm.fit, hypothesis.matrix = c("ZI_3 = 0", "ZI_4 = 0"))
lh.out
```

```
##
## Sum of squares and products for the hypothesis:
##           IR_1      IR_2       IR_3      IR_4       IR_5
## IR_1  1.479504  2.976657   4.736212  10.45521  -3.740739
## IR_2  2.976657  9.311814  12.939832  16.32167  -7.656210
## IR_3  4.736212 12.939832  18.662792  28.63121 -12.108465
## IR_4 10.455210 16.321675  28.631214  80.56959 -26.250120
## IR_5 -3.740739 -7.656210 -12.108465 -26.25012   9.463078
##
## Sum of squares and products for error:
##          IR_1     IR_2      IR_3        IR_4       IR_5
## IR_1 741.7653 408.1852 306.20688  246.452803 118.182197
## IR_2 408.1852 622.9091 139.62684  280.232637 102.673232
## IR_3 306.2069 139.6268 294.48505  340.271877  92.178845
## IR_4 246.4528 280.2326 340.27188 1457.452690   5.389921
## IR_5 118.1822 102.6732  92.17885    5.389921 266.121401
##
## Multivariate Tests:
##                  Df test stat approx F num Df den Df   Pr(>F)
## Pillai            2 0.2169237 1.922181     10    158 0.045768 *
## Wilks             2 0.7910374 1.939858     10    156 0.043620 *
## Hotelling-Lawley  2 0.2540984 1.956558     10    154 0.041689 *
## Roy               2 0.2050062 3.239098      5     79 0.010338 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```