SAP - projekt

Procjena kreditnog rizika

Vedran Knežević, Andre Medvedić, Jan Celin, Ante Čavar

20.10.2021.

Uvod

U našem projektu obrađujemo veliki skup podataka kreditnog stanja korisnika neke banke. Naš je zadatak procijentiti koji čimbenici utječu na sposobnost otplate kredita u zadanome roku.

Skup podataka - statistika

```
data <- read.csv("procjena_kreditnog_rizika.csv")


cat("number of missing values: ", sum(is.na(data[,])), "\n")
```

```
## number of missing values:  0
```

```
data$AccountStatus <- factor(
  data$AccountStatus, levels = c(
    "no checking account",
    "... < 0",
    "0 <= ... < 200",
    "... >= 200")
)
data$CreditHistory <- factor(
  data$CreditHistory, levels = c(
    "delay in paying off in the past",
    "critical account/ other credits existing (not at this bank)",
    "no credits taken/ all credits paid back duly",
    "existing credits paid back duly till now",
    "all credits at this bank paid back duly"
  )
)
data$Purpose <- factor(data$Purpose)
data$Account <- factor(
  data$Account, levels = c(
    "unknown/ no savings account",
    "... < 100",
    "100 <= ... < 500",
    "500 <= ... < 1000",
    "... >= 1000"
  )
)
data$EmploymentSince <- factor(
```

```
  data$EmploymentSince, levels = c(
    "unemployed",
    "... < 1 year",
    "1 <= ... < 4 years",
    "4 <= ... < 7 years",
    "... >= 7 years"
  )
)
data$PercentOfIncome <- factor(
  data$PercentOfIncome, levels = c(
    "... < 20%",
    "20% <= ... < 25%",
    "25% <= ... < 35%",
    "... >= 35%"
  )
)
data$PersonalStatus <- factor(
  data$PersonalStatus
)
data$OtherDebtors <- factor(
  data$OtherDebtors, levels = c(
    "none",
    "guarantor",
    "co-applicant"
  )
)
data$ResidenceSince <- factor(
  data$ResidenceSince, levels = c(
    "... < 1 year",
    "1 <= ... < 4 years",
    "4 <= ... < 7 years",
    ".. >= 7 years"
  )
)
data$Property <- factor(
  data$Property, levels = c(
    "unknown / no property",
    "building society savings agreement/ life insurance",
    "car or other, not in attribute Account",
    "real estate"
  )
)
data$OtherInstallPlans <- factor(
  data$OtherInstallPlans, levels = c(
    "none",
    "stores",
    "bank"
  )
)
data$Housing <- factor(
  data$Housing, levels = c(
    "for free",
    "rent",
```

```
    "own"
  )
)
data$NumExistingCredits <- factor(
  data$NumExistingCredits, levels = c(
    "1",
    "2 or 3",
    "4 or 5",
    "above 6"
  )
)
data$Job <- factor(
  data$Job, levels = c(
    "unemployed/ unskilled - non-resident",
    "unskilled - resident",
    "management/ self-employed/highly qualified employee/ officer",
    "skilled employee / official"
  )
)
data$NumberOfDependents <- factor(
  data$NumberOfDependents, levels = c(
    "less than 3",
    "3 or more"
  )
)
data$Telephone <- factor(
  data$Telephone, levels = c(
    "none",
    "yes, registered under the customers name"
  )
)
data$ForeignWorker <- factor(
  data$ForeignWorker, levels = c(
    "no",
    "yes"
  )
)
data$Default <- factor(
  data$Default,
  levels = c(0,1),
  labels = c(FALSE, TRUE)
)
summary(data)
```

```
##              AccountStatus    Duration
##  no checking account:394   Min.   : 4.0
##  ... < 0            :274   1st Qu.:12.0
##  0 <= ... < 200     :269   Median :18.0
##  ... >= 200         : 63   Mean   :20.9
##                            3rd Qu.:24.0
##                            Max.   :72.0
##
##                                                    CreditHistory
##  delay in paying off in the past                          : 88
```

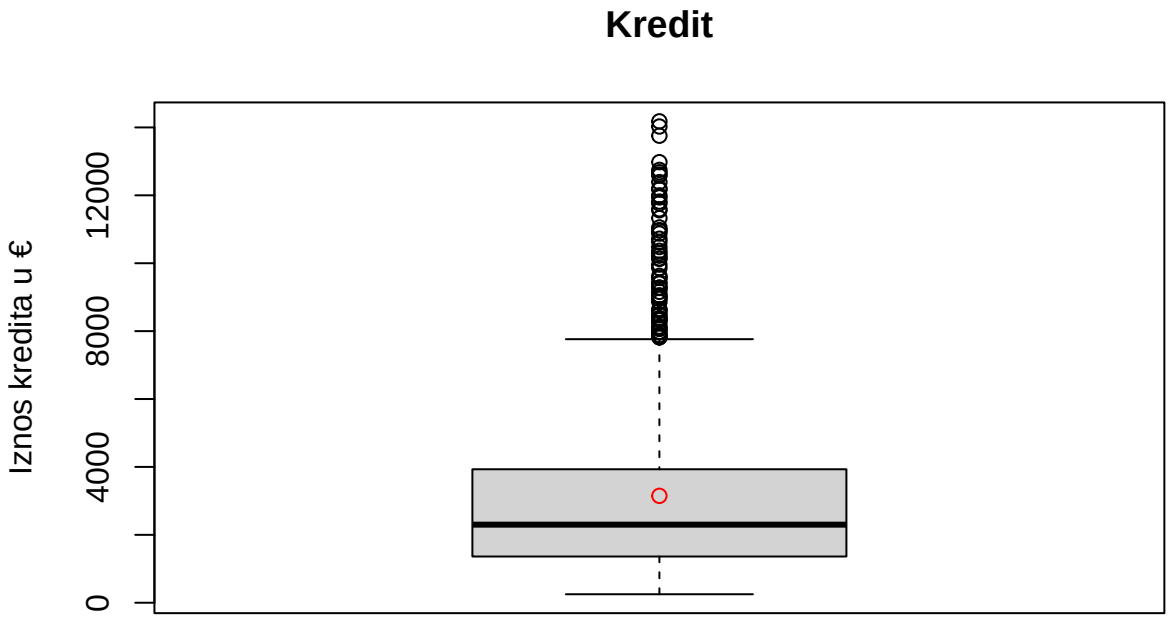
```
##  critical account/ other credits existing (not at this bank):293
##  no credits taken/ all credits paid back duly               : 40
##  existing credits paid back duly till now                   :530
##  all credits at this bank paid back duly                    : 49
##
##
##                 Purpose     CreditAmount                          Account
##  radio/television   :280   Min.   :  250   unknown/ no savings account:183
##  car (new)          :234   1st Qu.: 1366   ... < 100                  :603
##  furniture/equipment:181   Median : 2320   100 <= ... < 500           :103
##  car (used)         :103   Mean   : 3271   500 <= ... < 1000          : 63
##  business           : 97   3rd Qu.: 3972   ... >= 1000                : 48
##  education          : 50   Max.   :18424
##  (Other)            : 55
##           EmploymentSince         PercentOfIncome
##  unemployed        : 62    ... < 20%       :476
##  ... < 1 year      :172    20% <= ... < 25%:157
##  1 <= ... < 4 years:339    25% <= ... < 35%:231
##  4 <= ... < 7 years:174    ... >= 35%      :136
##  ... >= 7 years    :253
##
##
##                              PersonalStatus       OtherDebtors
##  female - divorced/separated/married:310    none        :907
##  male - divorced/separated          : 50    guarantor   : 52
##  male - married/widowed             : 92    co-applicant: 41
##  male - single                      :548
##
##
##
##            ResidenceSince
##  ... < 1 year      :130
##  1 <= ... < 4 years:308
##  4 <= ... < 7 years:149
##  .. >= 7 years     :413
##
##
##
##                                                  Property        Age
##  unknown / no property                              :154   Min.   :19.00
##  building society savings agreement/ life insurance:232   1st Qu.:27.00
##  car or other, not in attribute Account             :332   Median :33.00
##  real estate                                        :282   Mean   :35.55
##                                                            3rd Qu.:42.00
##                                                            Max.   :75.00
##
##  OtherInstallPlans     Housing    NumExistingCredits
##  none  :814        for free:108   1      :633
##  stores: 47        rent    :179   2 or 3 :333
##  bank  :139        own     :713   4 or 5 : 28
##                                   above 6:  6
##
##
##
```

```
##                                                          Job
##  unemployed/ unskilled - non-resident                      : 22
##  unskilled - resident                                      :200
##  management/ self-employed/highly qualified employee/ officer:148
##  skilled employee / official                               :630
##
##
##
##    NumberOfDependents                                 Telephone
##  less than 3:155      none                                 :596
##  3 or more  :845      yes, registered under the customers name:404
##
##
##
##
##
##  ForeignWorker  Default
##  no : 37       FALSE:700
##  yes:963       TRUE :300
##
##
##
##
##
```

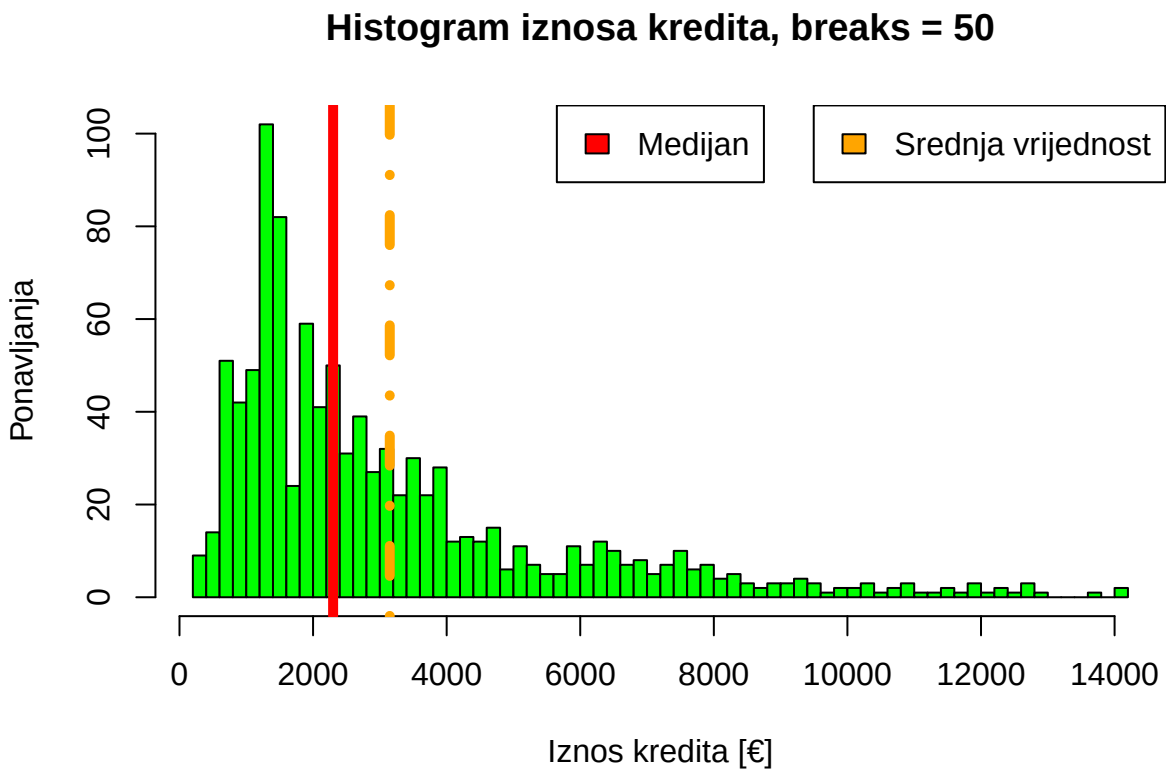
Vidimo da je skup poprilično čist (nema nedostajućih vrijednosti). Iako bi neki stupci moguće bili korisniji da su numerički prije nego kategorički.

Uvodni grafovi

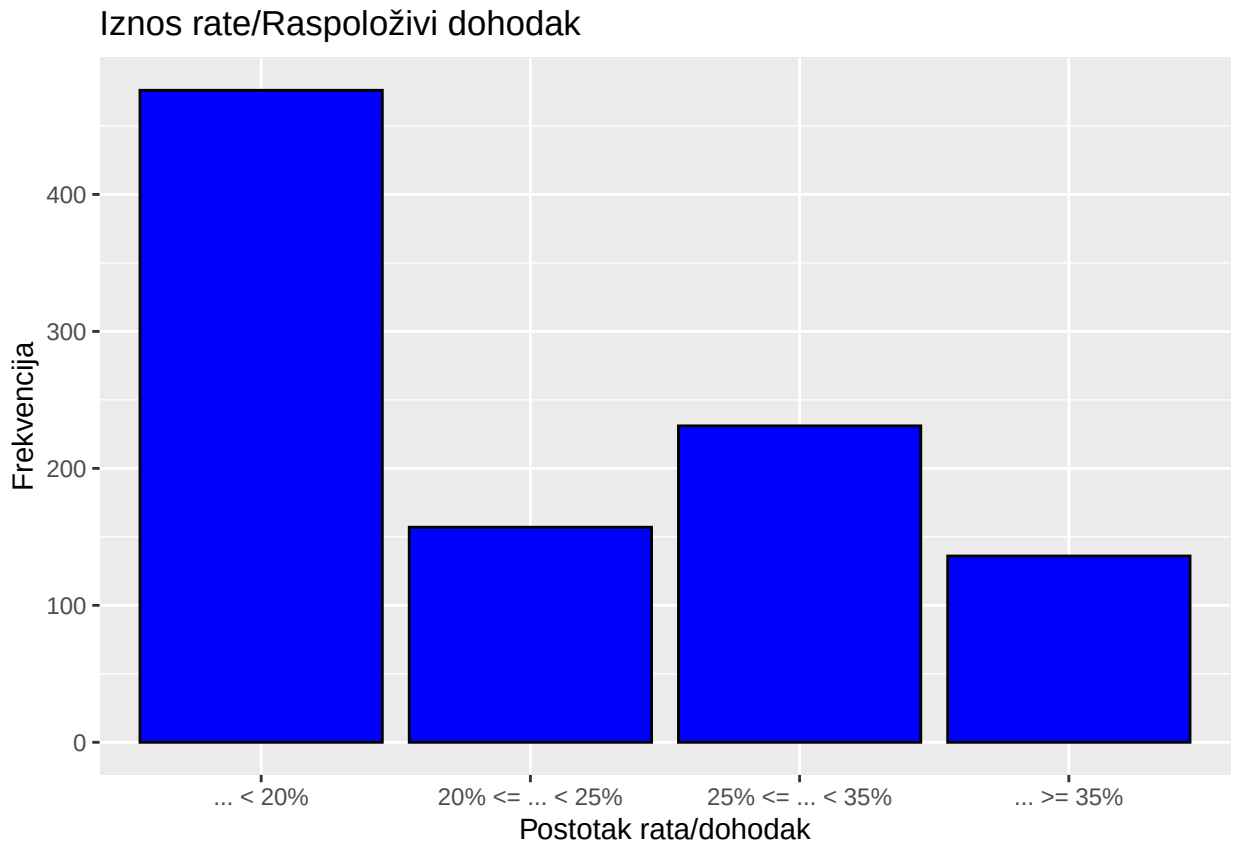
```
threshold <- quantile(data$CreditAmount, 0.99)  # Set threshold at 99th percentile
# Exclude data points above the threshold
filtered_data <- subset(data, CreditAmount <= threshold)
boxplot(filtered_data$CreditAmount, main="Kredit", ylab="Iznos kredita u €")
points(mean(filtered_data$CreditAmount), col = "red")
```




```
threshold <- quantile(data$CreditAmount, 0.99)  # Set threshold at 99th percentile
# Exclude data points above the threshold
filtered_data <- subset(data, CreditAmount <= threshold)
h = hist(filtered_data$CreditAmount,
         breaks = 50,
         main="Histogram iznosa kredita, breaks = 50",
         xlab="Iznos kredita [€]",
         ylab='Ponavljanja',
         col="green"
         )
legend("topright", legend = "Srednja vrijednost", fill = "orange")
legend("top", legend = "Medijan", fill = "red")
abline(v = mean(filtered_data$CreditAmount), col = "orange", lwd=5, lty=10)
abline(v = median(filtered_data$CreditAmount), col= "red", lwd=5)
```




```r
# Create a bar plot
bar_plot <- ggplot(data, aes(x = PercentOfIncome)) +
  geom_bar(fill = "blue", color = "black") +
  labs(title = "Iznos rate/Raspoloživi dohodak",
       x = "Postotak rata/dohodak",
       y = "Frekvencija")

# Print the plot
print(bar_plot)
```



TESTOVI

1.pitanje: Možemo li temeljem drugih dostupnih varijabli predvidjeti hoće li nastupiti *default* za određenog klijenta? Koje varijable povećavaju tu vjerojatnost?

Primjeren način za odgovoriti na ovo pitanje je razavoj dobrog modela logističke regresije. Kako bi dobili nekakvu okvirnu sliku o međuovisnostima naših regresora valjalo bi dobiti korelacijsku matricu.

```
cor_matrix <- cor(sapply(data, as.numeric))
cor_matrix
```

```
##                   AccountStatus     Duration CreditHistory     Purpose
## AccountStatus       1.0000000000  0.035049995   0.101690276 -0.01110151
## Duration            0.0350499949  1.000000000  -0.054683547 -0.09087113
## CreditHistory       0.1016902756 -0.054683547   1.000000000  0.08011650
## Purpose            -0.0111015069 -0.090871128   0.080116496  1.00000000
## CreditAmount        0.0245612304  0.624984198  -0.074147809 -0.17293218
## Account            -0.0056144454 -0.064525576   0.005164772 -0.01343600
## EmploymentSince    -0.1085355580  0.057381027  -0.116385754  0.01033098
## PercentOfIncome     0.0579415530 -0.074748816   0.006712577 -0.12582399
## PersonalStatus     -0.0494576924  0.099678338  -0.124240433 -0.06125627
## OtherDebtors        0.0419697300  0.006711354   0.078826518  0.08217475
## ResidenceSince     -0.0595551870  0.034067202  -0.049506574 -0.04539649
## Property           -0.0271098988 -0.231634130   0.015900959  0.08933634
## Age                -0.0490579639 -0.036136374  -0.133804189 -0.07557358
## OtherInstallPlans   0.0454707676  0.054883979   0.067624218 -0.05254707
```

```
## Housing            -0.0489837148 -0.137434101  -0.055202980  0.09360424
## NumExistingCredits -0.0930810062 -0.011283597  -0.534080412 -0.06058002
## Job                -0.0958309536  0.127674030  -0.031824008  0.05754028
## NumberOfDependents  0.0408888937  0.023834475   0.037265261  0.10148441
## Telephone          -0.0392092109  0.164718207  -0.070684436 -0.10455471
## ForeignWorker       0.0002049167  0.138196285  -0.009667845  0.07636717
## Default             0.1977876363  0.214926665   0.134122824 -0.08909612
##                    CreditAmount       Account EmploymentSince PercentOfIncome
## AccountStatus       0.024561230 -0.0056144454    -0.108535558    0.0579415530
## Duration            0.624984198 -0.0645255762     0.057381027   -0.0747488163
## CreditHistory      -0.074147809  0.0051647724    -0.116385754    0.0067125769
## Purpose            -0.172932183 -0.0134359995     0.010330980   -0.1258239941
## CreditAmount        1.000000000 -0.1075380478    -0.008366714    0.2713157012
## Account            -0.107538048  1.0000000000     0.014600228    0.0008051318
## EmploymentSince    -0.008366714  0.0146002279     1.000000000   -0.1261613068
## PercentOfIncome     0.271315701  0.0008051318    -0.126161307    1.0000000000
## PersonalStatus      0.114810137 -0.0519757119     0.226806949   -0.1183667992
## OtherDebtors        0.037920776 -0.0475745399    -0.028757855    0.0148351218
## ResidenceSince      0.028926323 -0.0117720666     0.245080745   -0.0493023708
## Property           -0.269289104  0.0542061688    -0.053739349    0.0381293321
## Age                 0.032716417 -0.0179966762     0.256227362   -0.0582656844
## OtherInstallPlans   0.046007723  0.0003668156     0.040153588   -0.0009827830
## Housing            -0.171584858  0.0338270900    -0.035992369   -0.0150288475
## NumExistingCredits  0.020794552 -0.0041759458     0.125790651   -0.0216687426
## Job                 0.032137899  0.0172989396     0.158418453   -0.0789583397
## NumberOfDependents -0.017142154  0.0213015831    -0.097192004   -0.0712069430
## Telephone           0.276995112 -0.0374523403     0.060518081   -0.0144128800
## ForeignWorker       0.050050007 -0.0053175005     0.027231899   -0.0900244286
## Default             0.154738641 -0.0338712656    -0.116002036   -0.0724039373
##                    PersonalStatus OtherDebtors ResidenceSince    Property
## AccountStatus         -0.04945769  0.041969730   -0.059555187 -0.02710990
## Duration               0.09967834  0.006711354    0.034067202 -0.23163413
## CreditHistory         -0.12424043  0.078826518   -0.049506574  0.01590096
## Purpose               -0.06125627  0.082174747   -0.045396490  0.08933634
## CreditAmount           0.11481014  0.037920776    0.028926323 -0.26928910
## Account               -0.05197571 -0.047574540   -0.011772067  0.05420617
## EmploymentSince        0.22680695 -0.028757855    0.245080745 -0.05373935
## PercentOfIncome       -0.11836680  0.014835122   -0.049302371  0.03812933
## PersonalStatus         1.00000000  0.022214073    0.020201062 -0.09729702
## OtherDebtors           0.02221407  1.000000000   -0.012690139  0.02303672
## ResidenceSince         0.02020106 -0.012690139    1.000000000 -0.15909556
## Property              -0.09729702  0.023036719   -0.159095565  1.00000000
## Age                    0.17715269 -0.028294429    0.266419184 -0.12768867
## OtherInstallPlans      0.04793930  0.030109877   -0.002088674 -0.08408118
## Housing                0.00528134  0.006429581   -0.304022693  0.49324037
## NumExistingCredits     0.11243815 -0.017661711    0.089625233  0.01680560
## Job                    0.02034705 -0.006980105    0.007060177 -0.07565059
## NumberOfDependents    -0.25357416  0.010990013   -0.042643426  0.04289844
## Telephone              0.07891891 -0.050995818    0.095359367 -0.15961358
## ForeignWorker         -0.04909927 -0.107639000    0.054097396 -0.06958913
## Default               -0.08953503  0.028440518    0.002967159 -0.14304786
##                             Age OtherInstallPlans      Housing
## AccountStatus      -0.049057964      0.0454707676 -0.048983715
## Duration           -0.036136374      0.0548839789 -0.137434101
```

```
## CreditHistory      -0.133804189      0.0676242178 -0.055202980
## Purpose            -0.075573577     -0.0525470742  0.093604236
## CreditAmount        0.032716417      0.0460077225 -0.171584858
## Account            -0.017996676      0.0003668156  0.033827090
## EmploymentSince     0.256227362      0.0401535884 -0.035992369
## PercentOfIncome    -0.058265684     -0.0009827830 -0.015028848
## PersonalStatus      0.177152693      0.0479392963  0.005281340
## OtherDebtors       -0.028294429      0.0301098774  0.006429581
## ResidenceSince      0.266419184     -0.0020886735 -0.304022693
## Property           -0.127688673     -0.0840811816  0.493240369
## Age                 1.000000000      0.0423457393 -0.112050595
## OtherInstallPlans   0.042345739      1.0000000000 -0.030743983
## Housing            -0.112050595     -0.0307439830  1.000000000
## NumExistingCredits  0.149253582      0.0484422330  0.022506464
## Job                -0.121634589     -0.0670221352 -0.014420854
## NumberOfDependents -0.118200833     -0.0768906418  0.072815132
## Telephone           0.145258701      0.0193606090 -0.076801193
## ForeignWorker       0.006151396      0.0152113951 -0.044100345
## Default            -0.091127409      0.1098440987 -0.127789062
##                    NumExistingCredits           Job NumberOfDependents
## AccountStatus            -0.093081006 -0.0958309536        0.040888894
## Duration                 -0.011283597  0.1276740299        0.023834475
## CreditHistory            -0.534080412 -0.0318240077        0.037265261
## Purpose                  -0.060580022  0.0575402786        0.101484409
## CreditAmount              0.020794552  0.0321378994       -0.017142154
## Account                  -0.004175946  0.0172989396        0.021301583
## EmploymentSince           0.125790651  0.1584184534       -0.097192004
## PercentOfIncome          -0.021668743 -0.0789583397       -0.071206943
## PersonalStatus            0.112438154  0.0203470477       -0.253574164
## OtherDebtors             -0.017661711 -0.0069801051        0.010990013
## ResidenceSince            0.089625233  0.0070601768       -0.042643426
## Property                  0.016805597 -0.0756505881        0.042898435
## Age                       0.149253582 -0.1216345894       -0.118200833
## OtherInstallPlans         0.048442233 -0.0670221352       -0.076890642
## Housing                   0.022506464 -0.0144208535        0.072815132
## NumExistingCredits        1.000000000 -0.0160021029       -0.109666700
## Job                      -0.016002103  1.0000000000        0.122351384
## NumberOfDependents       -0.109666700  0.1223513844        1.000000000
## Telephone                 0.065553213  0.0954110165        0.014753439
## ForeignWorker             0.009716975  0.0802389378        0.077070853
## Default                  -0.045732489  0.0004976867        0.003014853
##                     Telephone ForeignWorker       Default
## AccountStatus      -0.03920921  0.0002049167  0.1977876363
## Duration            0.16471821  0.1381962853  0.2149266654
## CreditHistory      -0.07068444 -0.0096678446  0.1341228244
## Purpose            -0.10455471  0.0763671738 -0.0890961201
## CreditAmount        0.27699511  0.0500500071  0.1547386411
## Account            -0.03745234 -0.0053175005 -0.0338712656
## EmploymentSince     0.06051808  0.0272318995 -0.1160020364
## PercentOfIncome    -0.01441288 -0.0900244286 -0.0724039373
## PersonalStatus      0.07891891 -0.0490992653 -0.0895350264
## OtherDebtors       -0.05099582 -0.1076389995  0.0284405180
## ResidenceSince      0.09535937  0.0540973962  0.0029671588
## Property           -0.15961358 -0.0695891267 -0.1430478614
```

```
## Age                  0.14525870  0.0061513960 -0.0911274093
## OtherInstallPlans    0.01936061  0.0152113951  0.1098440987
## Housing             -0.07680119 -0.0441003451 -0.1277890617
## NumExistingCredits   0.06555321  0.0097169747 -0.0457324893
## Job                  0.09541102  0.0802389378  0.0004976867
## NumberOfDependents   0.01475344  0.0770708531  0.0030148531
## Telephone            1.00000000  0.1074009094 -0.0364661902
## ForeignWorker        0.10740091  1.0000000000  0.0820794988
## Default             -0.03646619  0.0820794988  1.0000000000
```

Budući je izlaz u R-u nepregledan predočit ćemo koeficijente varijable čija je apsolutna vrijednost veća od 0.3. Varijable kod kojih dolazi do takvih korelacija su sljedeće:

Duration i CreditAmount

```
cor_matrix["Duration", "CreditAmount"]
```

```
## [1] 0.6249842
```

CreditHistory i NumExistingCredits

```
cor_matrix["CreditHistory", "NumExistingCredits"]
```

```
## [1] -0.5340804
```

ResidenceSince i Housing

```
cor_matrix["ResidenceSince","Housing"]
```

```
## [1] -0.3040227
```

Property i Housing

```
cor_matrix["Property", "Housing"]
```

```
## [1] 0.4932404
```

```
logreg.mdl <- glm(Default ~ AccountStatus + Duration + CreditHistory + Purpose + CreditAmount + Account
summary(logreg.mdl)
```

```
##
## Call:
## glm(formula = Default ~ AccountStatus + Duration + CreditHistory +
##     Purpose + CreditAmount + Account + EmploymentSince + PercentOfIncome +
##     PersonalStatus + OtherDebtors + ResidenceSince + Property +
##     Age + OtherInstallPlans + Housing + NumExistingCredits +
##     Job + NumberOfDependents + Telephone + ForeignWorker, family = binomial(),
##     data = data)
##
## Coefficients:
##                                                                    Estimate
## (Intercept)                                                      -4.588e+00
## AccountStatus... < 0                                              1.780e+00
## AccountStatus0 <= ... < 200                                       1.397e+00
## AccountStatus... >= 200                                           8.061e-01
## Duration                                                          2.801e-02
## CreditHistorycritical account/ other credits existing (not at this bank) -5.461e-01
## CreditHistoryno credits taken/ all credits paid back duly         9.496e-01
## CreditHistoryexisting credits paid back duly till now             3.824e-01
## CreditHistoryall credits at this bank paid back duly              1.119e+00
```

```
## Purposecar (new)                                                        6.888e-01
## Purposecar (used)                                                      -9.717e-01
## Purposedomestic appliances                                              1.779e-01
## Purposeeducation                                                        8.018e-01
## Purposefurniture/equipment                                             -5.930e-02
## Purposeothers                                                          -7.966e-01
## Purposeradio/television                                                -1.855e-01
## Purposerepairs                                                          5.285e-01
## Purposeretraining                                                      -1.242e+00
## CreditAmount                                                            1.233e-04
## Account... < 100                                                        9.732e-01
## Account100 <= ... < 500                                                 6.094e-01
## Account500 <= ... < 1000                                                6.068e-01
## Account... >= 1000                                                     -4.872e-01
## EmploymentSince... < 1 year                                             6.662e-02
## EmploymentSince1 <= ... < 4 years                                      -2.293e-01
## EmploymentSince4 <= ... < 7 years                                      -7.634e-01
## EmploymentSince... >= 7 years                                          -2.213e-01
## PercentOfIncome20% <= ... < 25%                                        -3.109e-01
## PercentOfIncome25% <= ... < 35%                                        -6.727e-01
## PercentOfIncome... >= 35%                                              -9.369e-01
## PersonalStatusmale - divorced/separated                                 2.616e-01
## PersonalStatusmale - married/widowed                                   -1.148e-01
## PersonalStatusmale - single                                            -5.811e-01
## OtherDebtorsguarantor                                                  -9.828e-01
## OtherDebtorsco-applicant                                                4.329e-01
## ResidenceSince1 <= ... < 4 years                                        7.613e-01
## ResidenceSince4 <= ... < 7 years                                        5.246e-01
## ResidenceSince.. >= 7 years                                             3.885e-01
## Propertybuilding society savings agreement/ life insurance             -4.669e-01
## Propertycar or other, not in attribute Account                         -5.760e-01
## Propertyreal estate                                                    -7.367e-01
## Age                                                                    -1.279e-02
## OtherInstallPlansstores                                                 5.587e-01
## OtherInstallPlansbank                                                   6.475e-01
## Housingrent                                                             6.303e-01
## Housingown                                                              1.729e-01
## NumExistingCredits2 or 3                                                4.050e-01
## NumExistingCredits4 or 5                                                2.741e-01
## NumExistingCreditsabove 6                                               4.550e-01
## Jobunskilled - resident                                                 4.416e-01
## Jobmanagement/ self-employed/highly qualified employee/ officer         3.691e-01
## Jobskilled employee / official                                          4.694e-01
## NumberOfDependents3 or more                                            -2.628e-01
## Telephoneyes, registered under the customers name                      -2.848e-01
## ForeignWorkeryes                                                        1.461e+00
##                                                                       Std. Error
## (Intercept)                                                             1.203e+00
## AccountStatus... < 0                                                    2.358e-01
## AccountStatus0 <= ... < 200                                             2.358e-01
## AccountStatus... >= 200                                                 3.852e-01
## Duration                                                                9.448e-03
## CreditHistorycritical account/ other credits existing (not at this bank)  3.392e-01
## CreditHistoryno credits taken/ all credits paid back duly               4.780e-01
```

```
## CreditHistoryexisting credits paid back duly till now                    3.330e-01
## CreditHistoryall credits at this bank paid back duly                     4.840e-01
## Purposecar (new)                                                         3.377e-01
## Purposecar (used)                                                        4.456e-01
## Purposedomestic appliances                                               8.182e-01
## Purposeeducation                                                         4.677e-01
## Purposefurniture/equipment                                               3.570e-01
## Purposeothers                                                            8.137e-01
## Purposeradio/television                                                  3.402e-01
## Purposerepairs                                                           5.945e-01
## Purposeretraining                                                        1.202e+00
## CreditAmount                                                             4.502e-05
## Account... < 100                                                         2.661e-01
## Account100 <= ... < 500                                                  3.557e-01
## Account500 <= ... < 1000                                                 4.520e-01
## Account... >= 1000                                                       5.811e-01
## EmploymentSince... < 1 year                                              4.396e-01
## EmploymentSince1 <= ... < 4 years                                        4.212e-01
## EmploymentSince4 <= ... < 7 years                                        4.596e-01
## EmploymentSince... >= 7 years                                            4.236e-01
## PercentOfIncome20% <= ... < 25%                                          2.551e-01
## PercentOfIncome25% <= ... < 35%                                          2.341e-01
## PercentOfIncome... >= 35%                                                3.047e-01
## PersonalStatusmale - divorced/separated                                  3.885e-01
## PersonalStatusmale - married/widowed                                     3.181e-01
## PersonalStatusmale - single                                              2.133e-01
## OtherDebtorsguarantor                                                    4.264e-01
## OtherDebtorsco-applicant                                                 4.127e-01
## ResidenceSince1 <= ... < 4 years                                         2.994e-01
## ResidenceSince4 <= ... < 7 years                                         3.359e-01
## ResidenceSince.. >= 7 years                                              3.029e-01
## Propertybuilding society savings agreement/ life insurance               4.180e-01
## Propertycar or other, not in attribute Account                           4.079e-01
## Propertyreal estate                                                      4.294e-01
## Age                                                                      9.317e-03
## OtherInstallPlansstores                                                  3.806e-01
## OtherInstallPlansbank                                                    2.403e-01
## Housingrent                                                              4.854e-01
## Housingown                                                               4.607e-01
## NumExistingCredits2 or 3                                                 2.456e-01
## NumExistingCredits4 or 5                                                 6.087e-01
## NumExistingCreditsabove 6                                                1.072e+00
## Jobunskilled - resident                                                  6.867e-01
## Jobmanagement/ self-employed/highly qualified employee/ officer          6.708e-01
## Jobskilled employee / official                                           6.625e-01
## NumberOfDependents3 or more                                              2.518e-01
## Telephoneyes, registered under the customers name                        2.031e-01
## ForeignWorkeryes                                                         6.265e-01
##                                                                         z value
## (Intercept)                                                              -3.815
## AccountStatus... < 0                                                      7.547
## AccountStatus0 <= ... < 200                                               5.923
## AccountStatus... >= 200                                                   2.093
## Duration                                                                  2.965
```

```
## CreditHistorycritical account/ other credits existing (not at this bank)  -1.610
## CreditHistoryno credits taken/ all credits paid back duly                   1.987
## CreditHistoryexisting credits paid back duly till now                       1.148
## CreditHistoryall credits at this bank paid back duly                        2.311
## Purposecar (new)                                                            2.040
## Purposecar (used)                                                          -2.181
## Purposedomestic appliances                                                  0.217
## Purposeeducation                                                            1.714
## Purposefurniture/equipment                                                 -0.166
## Purposeothers                                                              -0.979
## Purposeradio/television                                                    -0.545
## Purposerepairs                                                              0.889
## Purposeretraining                                                          -1.034
## CreditAmount                                                                2.740
## Account... < 100                                                            3.657
## Account100 <= ... < 500                                                     1.713
## Account500 <= ... < 1000                                                    1.343
## Account... >= 1000                                                         -0.838
## EmploymentSince... < 1 year                                                 0.152
## EmploymentSince1 <= ... < 4 years                                          -0.544
## EmploymentSince4 <= ... < 7 years                                          -1.661
## EmploymentSince... >= 7 years                                              -0.523
## PercentOfIncome20% <= ... < 25%                                            -1.219
## PercentOfIncome25% <= ... < 35%                                            -2.874
## PercentOfIncome... >= 35%                                                  -3.075
## PersonalStatusmale - divorced/separated                                     0.673
## PersonalStatusmale - married/widowed                                       -0.361
## PersonalStatusmale - single                                                -2.725
## OtherDebtorsguarantor                                                      -2.305
## OtherDebtorsco-applicant                                                    1.049
## ResidenceSince1 <= ... < 4 years                                            2.543
## ResidenceSince4 <= ... < 7 years                                            1.562
## ResidenceSince.. >= 7 years                                                 1.282
## Propertybuilding society savings agreement/ life insurance                 -1.117
## Propertycar or other, not in attribute Account                             -1.412
## Propertyreal estate                                                        -1.716
## Age                                                                        -1.373
## OtherInstallPlansstores                                                     1.468
## OtherInstallPlansbank                                                       2.694
## Housingrent                                                                 1.299
## Housingown                                                                  0.375
## NumExistingCredits2 or 3                                                    1.649
## NumExistingCredits4 or 5                                                    0.450
## NumExistingCreditsabove 6                                                   0.424
## Jobunskilled - resident                                                     0.643
## Jobmanagement/ self-employed/highly qualified employee/ officer             0.550
## Jobskilled employee / official                                              0.709
## NumberOfDependents3 or more                                                -1.044
## Telephoneyes, registered under the customers name                          -1.402
## ForeignWorkeryes                                                            2.333
##                                                                           Pr(>|z|)
## (Intercept)                                                               0.000136
## AccountStatus... < 0                                                      4.45e-14
## AccountStatus0 <= ... < 200                                               3.15e-09
```

```
## AccountStatus... >= 200                                                    0.036373
## Duration                                                                   0.003028
## CreditHistorycritical account/ other credits existing (not at this bank) 0.107362
## CreditHistoryno credits taken/ all credits paid back duly                  0.046964
## CreditHistoryexisting credits paid back duly till now                      0.250810
## CreditHistoryall credits at this bank paid back duly                       0.020824
## Purposecar (new)                                                           0.041386
## Purposecar (used)                                                          0.029211
## Purposedomestic appliances                                                 0.827872
## Purposeeducation                                                           0.086459
## Purposefurniture/equipment                                                 0.868070
## Purposeothers                                                              0.327552
## Purposeradio/television                                                    0.585437
## Purposerepairs                                                             0.374052
## Purposeretraining                                                          0.301348
## CreditAmount                                                               0.006153
## Account... < 100                                                           0.000255
## Account100 <= ... < 500                                                    0.086644
## Account500 <= ... < 1000                                                   0.179389
## Account... >= 1000                                                         0.401801
## EmploymentSince... < 1 year                                                0.879539
## EmploymentSince1 <= ... < 4 years                                          0.586261
## EmploymentSince4 <= ... < 7 years                                          0.096714
## EmploymentSince... >= 7 years                                              0.601303
## PercentOfIncome20% <= ... < 25%                                            0.222947
## PercentOfIncome25% <= ... < 35%                                            0.004050
## PercentOfIncome... >= 35%                                                  0.002106
## PersonalStatusmale - divorced/separated                                    0.500728
## PersonalStatusmale - married/widowed                                       0.718286
## PersonalStatusmale - single                                                0.006431
## OtherDebtorsguarantor                                                      0.021160
## OtherDebtorsco-applicant                                                   0.294219
## ResidenceSince1 <= ... < 4 years                                           0.010985
## ResidenceSince4 <= ... < 7 years                                           0.118342
## ResidenceSince.. >= 7 years                                                0.199687
## Propertybuilding society savings agreement/ life insurance                 0.264061
## Propertycar or other, not in attribute Account                             0.157924
## Propertyreal estate                                                        0.086230
## Age                                                                        0.169876
## OtherInstallPlansstores                                                    0.142081
## OtherInstallPlansbank                                                      0.007056
## Housingrent                                                                0.194111
## Housingown                                                                 0.707338
## NumExistingCredits2 or 3                                                   0.099170
## NumExistingCredits4 or 5                                                   0.652475
## NumExistingCreditsabove 6                                                  0.671370
## Jobunskilled - resident                                                    0.520167
## Jobmanagement/ self-employed/highly qualified employee/ officer            0.582115
## Jobskilled employee / official                                             0.478594
## NumberOfDependents3 or more                                                0.296625
## Telephoneyes, registered under the customers name                          0.160870
## ForeignWorkeryes                                                           0.019658
##
## (Intercept)                                                                ***
```

```
## AccountStatus... < 0                                                    ***
## AccountStatus0 <= ... < 200                                             ***
## AccountStatus... >= 200                                                 *
## Duration                                                                **
## CreditHistorycritical account/ other credits existing (not at this bank)
## CreditHistoryno credits taken/ all credits paid back duly               *
## CreditHistoryexisting credits paid back duly till now
## CreditHistoryall credits at this bank paid back duly                    *
## Purposecar (new)                                                        *
## Purposecar (used)                                                       *
## Purposedomestic appliances
## Purposeeducation                                                        .
## Purposefurniture/equipment
## Purposeothers
## Purposeradio/television
## Purposerepairs
## Purposeretraining
## CreditAmount                                                            **
## Account... < 100                                                        ***
## Account100 <= ... < 500                                                 .
## Account500 <= ... < 1000
## Account... >= 1000
## EmploymentSince... < 1 year
## EmploymentSince1 <= ... < 4 years
## EmploymentSince4 <= ... < 7 years                                       .
## EmploymentSince... >= 7 years
## PercentOfIncome20% <= ... < 25%
## PercentOfIncome25% <= ... < 35%                                         **
## PercentOfIncome... >= 35%                                               **
## PersonalStatusmale - divorced/separated
## PersonalStatusmale - married/widowed
## PersonalStatusmale - single                                             **
## OtherDebtorsguarantor                                                   *
## OtherDebtorsco-applicant
## ResidenceSince1 <= ... < 4 years                                        *
## ResidenceSince4 <= ... < 7 years
## ResidenceSince.. >= 7 years
## Propertybuilding society savings agreement/ life insurance
## Propertycar or other, not in attribute Account
## Propertyreal estate                                                     .
## Age
## OtherInstallPlansstores
## OtherInstallPlansbank                                                   **
## Housingrent
## Housingown
## NumExistingCredits2 or 3                                                .
## NumExistingCredits4 or 5
## NumExistingCreditsabove 6
## Jobunskilled - resident
## Jobmanagement/ self-employed/highly qualified employee/ officer
## Jobskilled employee / official
## NumberOfDependents3 or more
## Telephoneyes, registered under the customers name
## ForeignWorkeryes                                                        *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1221.7  on 999  degrees of freedom
## Residual deviance:  887.4  on 945  degrees of freedom
## AIC: 997.4
##
## Number of Fisher Scoring iterations: 5
```

Sad ćemo razmotriti neke od mjera kvalitete modela.

```
Ypredicted <- logreg.mdl$fitted.values >= 0.5
tab <- table(data$Default, Ypredicted)

tab
```

```
##        Ypredicted
##         FALSE TRUE
##   FALSE   624   76
##   TRUE    135  165
```

```
accuracy = sum(diag(tab)) / sum(tab)
precision = tab[2,2] / sum(tab[,2])
recall = tab[2,2] / sum(tab[2,])
specificity = tab[1,1] / sum(tab[,1])

accuracy
```

```
## [1] 0.789
```

```
precision
```

```
## [1] 0.6846473
```

```
recall
```

```
## [1] 0.55
```

```
specificity
```

```
## [1] 0.8221344
```

2.pitanje: Jesu li muškarci skloniji neispunjavanju obveza po kreditu od žena?

U ovom odsječku uspoređujemo odnos između dviju kategorijskih varijabli (spol, izvršavanje svojih novčanih obveza). Uspoređivat ćemo je li kod muškaraca i žena jednaka proporcija onih koji nisu izvršili svoje novčane obaveze (default).

Sve statistike provjeravamo na razina značajnosti $\alpha = 0.05$. Ispitujemo jednostranu alternativu (neispunjavanje obveza je češće kod muškaraca).

Statistika nad svim muškarcima i ženama u skupu podataka

H0: Proporcija onih koji nisu ispunili obveza naspram onih koji su ispunili obaveze jednaka je kod muškaraca i žena (ili je manja kod muškaraca).
H1: Proporcija osoba koje nisu ispunile obaveze naspram onih koji su ispunili obaveze veća je kod muškaraca.

```r
female_clients <- data[str_detect(data$PersonalStatus, "female"),]
male_clients <- data[!str_detect(data$PersonalStatus, "female"),]
num_female_default <- nrow(female_clients[female_clients$Default == 1,])
num_male_default <- nrow(male_clients[male_clients$Default == 1,])

proportion_matrix <- matrix(c(nrow(male_clients)-num_male_default,
                              num_male_default,
                            nrow(female_clients)-num_female_default,
                             num_female_default), nrow=2, byrow = T)
colnames(proportion_matrix) <- c("no_default", "default")
rownames(proportion_matrix) <- c("male", "female")
# proportion_matrix
prop.test(proportion_matrix, alternative = "less")
```

```
## Warning in prop.test(proportion_matrix, alternative = "less"): Chi-squared
## approximation may be incorrect
```

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  proportion_matrix
## X-squared = NaN, df = 1, p-value = NA
## alternative hypothesis: less
## 95 percent confidence interval:
##  -1  0
## sample estimates:
## prop 1 prop 2
##      1      1
```

Iz ovoga zaključujemo, na razini značajnosti 0.05, da muškarci ispunjavaju kreditne obveze razmjerno ženama (tj. ne možemo reći da su skloniji neispunjavanju obveza).

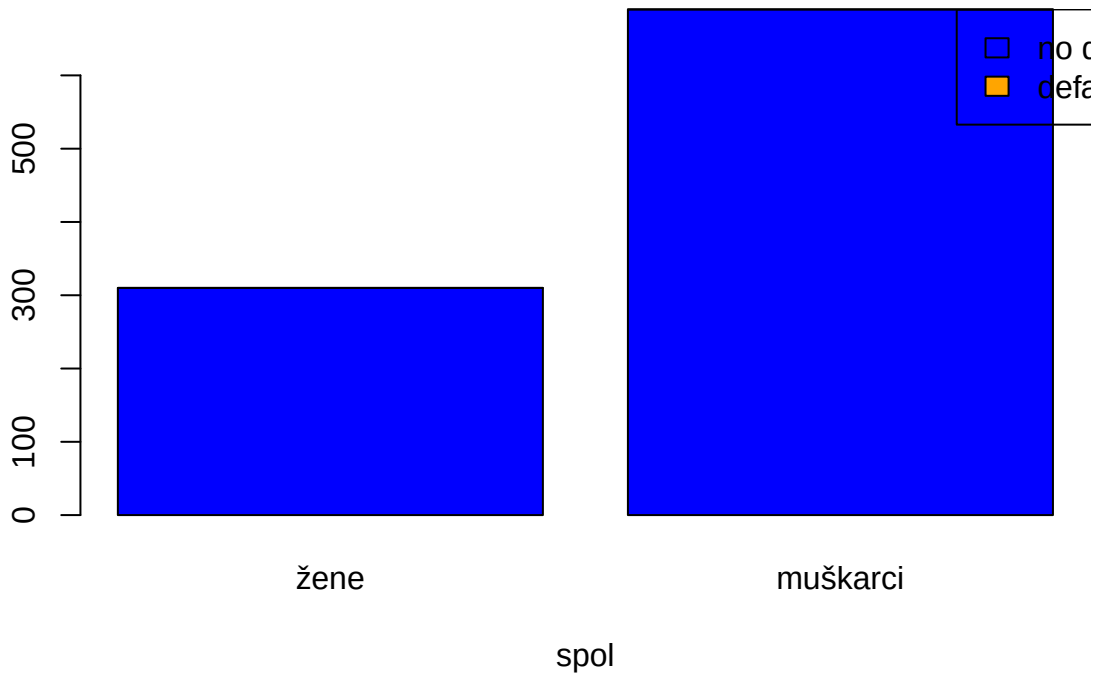
Provodimo Z-test o dvije proporcije s očekivanjem da će nam dati vrlo slične rezultate kao i χ^2-test.

```r
n1 <- nrow(male_clients)
n2 <- nrow(female_clients)
k1 <- n1 - num_male_default
k2 <- n2 - num_female_default

Z_stat <- (k1/n1-k2/n2)/sqrt(((k1+k2)/(n1+n2))*(1-(k1+k2)/(n1+n2))*(1/n1+1/n2))
cat("The p-value of the Z statistic is: ", pnorm(Z_stat))
```

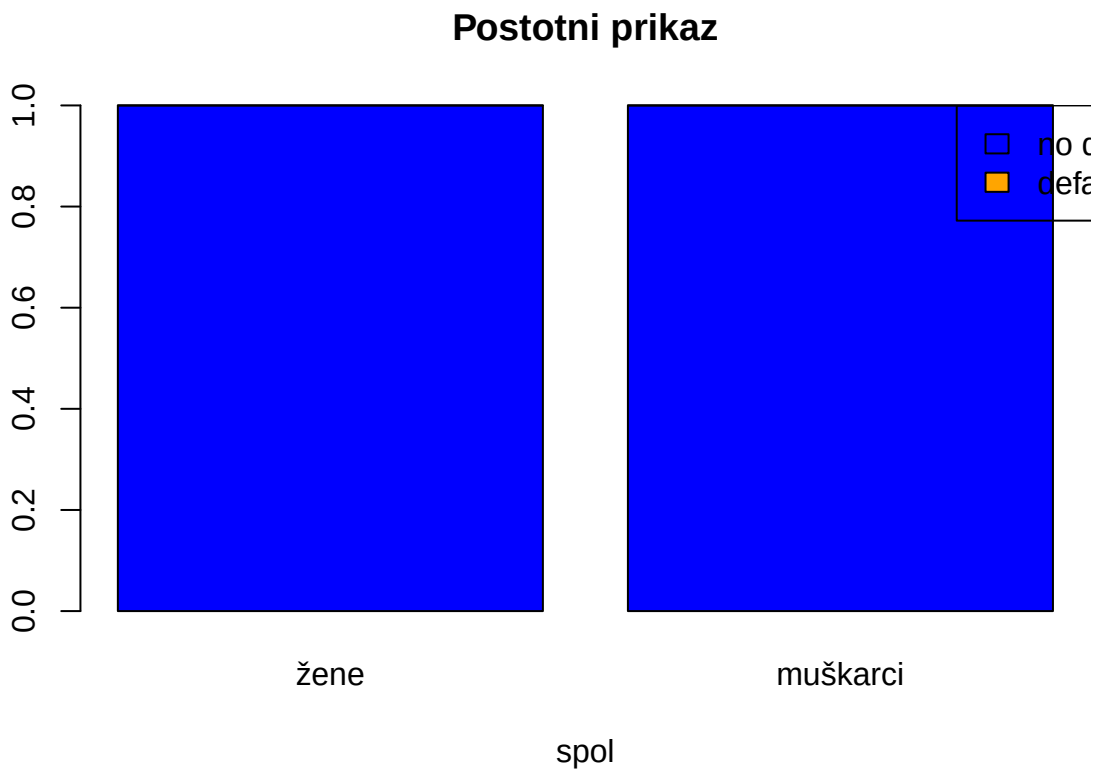
```
## The p-value of the Z statistic is:  NaN
```

Kao što možemo uočiti Z-test nam daje isti zaključak i vrlo sličnu p-vrijednost kao i χ^2-test pa ćemo nadalje koristiti χ^2 jer je on implementiran u R-u.

```r
Values <- matrix(c((nrow(female_clients)-num_female_default),
                   (nrow(male_clients)-num_male_default),
                   num_female_default,
                   num_male_default
                     ), nrow=2, ncol=2, byrow = T)

barplot(Values, main="Kvantitativni prikaz", names.arg=c("žene", "muškarci"), xlab="spol", col = c("blu

legend("topright", inset = c(-0.1, 0), c("no default","default"), fill = c("blue", "orange"))
```




```
Values <- matrix(c((nrow(female_clients)-num_female_default)/nrow(female_clients),
                   (nrow(male_clients)-num_male_default)/nrow(male_clients),
                   num_female_default/nrow(female_clients),
                   num_male_default/nrow(male_clients)
                     ), nrow=2, ncol=2, byrow = T)


barplot(Values, main="Postotni prikaz", names.arg=c("žene", "muškarci"), xlab="spol", col = c("blue", "

legend("topright", inset = c(-0.1, 0), c("no default","default"), fill = c("blue", "orange"))
```



Statistika nad slobodnim muškarcima i ženama u skupu podataka

H0: Proporcija onih koji nisu ispunili obveza naspram onih koji su ispunili obaveze jednaka je kod slobodnih muškaraca i žena (ili je manja kod slobodnih muškaraca).
H1: Proporcija osoba koje nisu ispunile obaveze naspram onih koji su ispunili obaveze veća je kod slobodnih muškaraca.

```
male_single_clients <- male_clients[str_detect(male_clients$PersonalStatus, "single"),]
num_male_single_default <- sum(male_single_clients$Default == 1)

proportion_matrix[1,] <- c(nrow(male_single_clients) - num_male_single_default,
                           num_male_single_default)
# proportion_matrix
prop.test(proportion_matrix, alternative = "less")
```

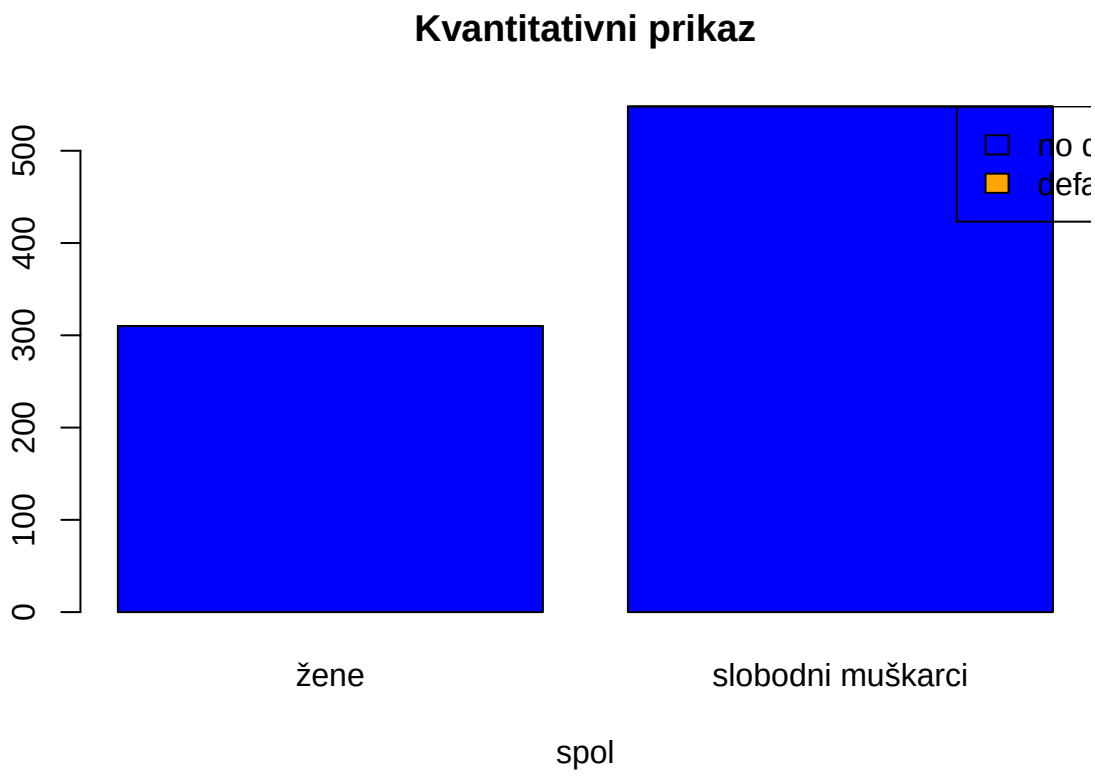
```
## Warning in prop.test(proportion_matrix, alternative = "less"): Chi-squared
## approximation may be incorrect
```

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  proportion_matrix
## X-squared = NaN, df = 1, p-value = NA
## alternative hypothesis: less
## 95 percent confidence interval:
##  -1  0
## sample estimates:
```

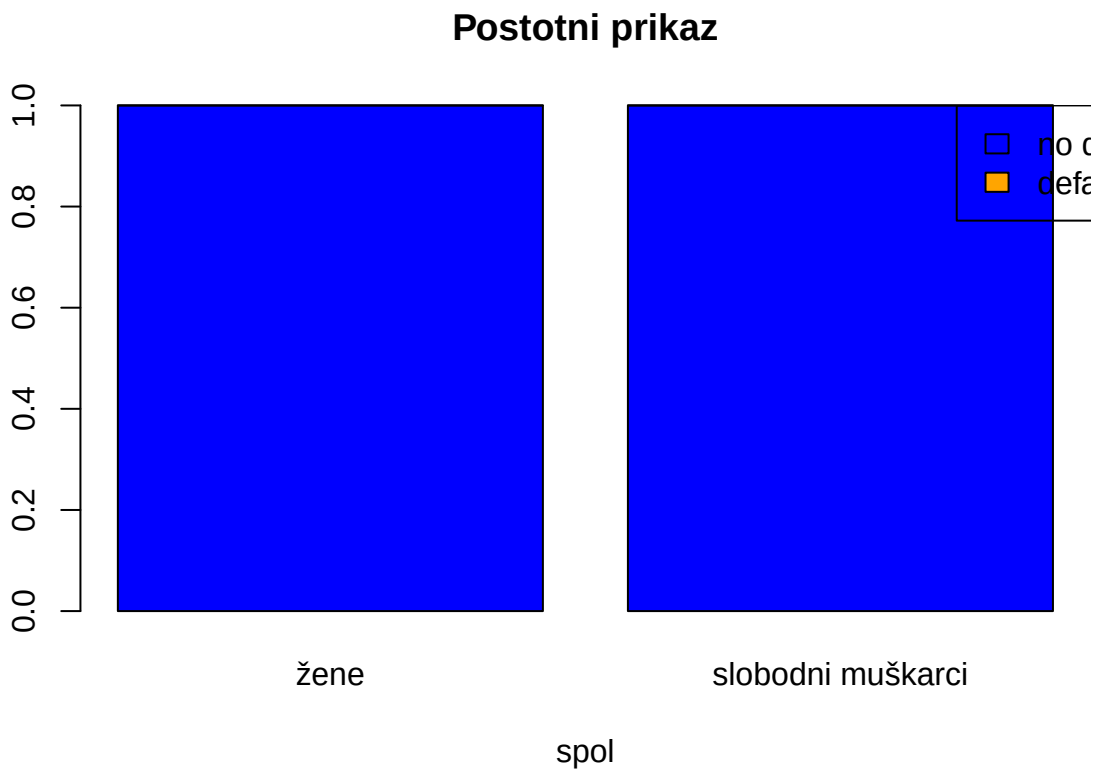
```
## prop 1 prop 2
##      1      1
```

Iz ovoga zaključujemo, na razini značajnosti 0.05, da slobodni muškarci ispunjavaju kreditne obveze razmjerno ženama (tj. ne možemo reći da su skloniji neispunjavanju obveza).

```
Values <- matrix(c((nrow(female_clients)-num_female_default),
                   (nrow(male_single_clients)-num_male_single_default),
                   num_female_default,
                   num_male_single_default
                     ), nrow=2, ncol=2, byrow = T)

barplot(Values, main="Kvantitativni prikaz", names.arg=c("žene", "slobodni muškarci"), xlab="spol", col

legend("topright", inset = c(-0.1, 0), c("no default","default"), fill = c("blue", "orange"))
```




```
Values <- matrix(c((nrow(female_clients)-num_female_default)/nrow(female_clients),
                   (nrow(male_single_clients)-num_male_single_default)/nrow(male_single_clients),
                   num_female_default/nrow(female_clients),
                   num_male_single_default/nrow(male_single_clients)
                     ), nrow=2, ncol=2, byrow = T)

barplot(Values, main="Postotni prikaz", names.arg=c("žene", "slobodni muškarci"), xlab="spol", col = c(

legend("topright", inset = c(-0.1, 0), c("no default", "default"), fill = c("blue", "orange"))
```



Statistika nad rastavljenim muškarcima i ženama u skupu podataka

H0: Proporcija onih koji nisu ispunili obveza naspram onih koji su ispunili obaveze jednaka je kod rastavljenih muškaraca i žena (ili je manja kod rastavljenih muškaraca).
H1: Proporcija osoba koje nisu ispunile obaveze naspram onih koji su ispunili obaveze veća je kod rastavljenih muškaraca.

```
male_divor_clients <- male_clients[str_detect(male_clients$PersonalStatus, "divorced"),]
num_male_divor_default <- sum(male_divor_clients$Default == 1)

proportion_matrix[1,] <- c(nrow(male_divor_clients) - num_male_divor_default,
                           num_male_divor_default)
# proportion_matrix
prop.test(proportion_matrix, alternative = "less")
```

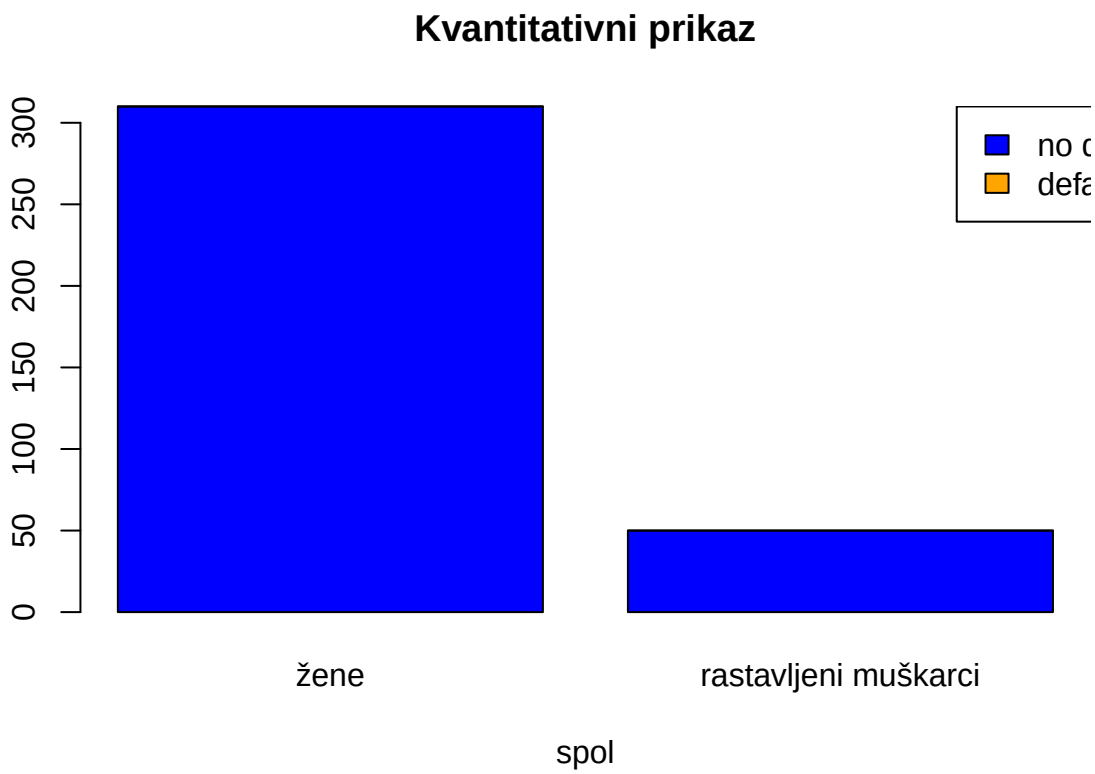
```
## Warning in prop.test(proportion_matrix, alternative = "less"): Chi-squared
## approximation may be incorrect
```

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  proportion_matrix
## X-squared = NaN, df = 1, p-value = NA
## alternative hypothesis: less
## 95 percent confidence interval:
##  -1  0
## sample estimates:
```

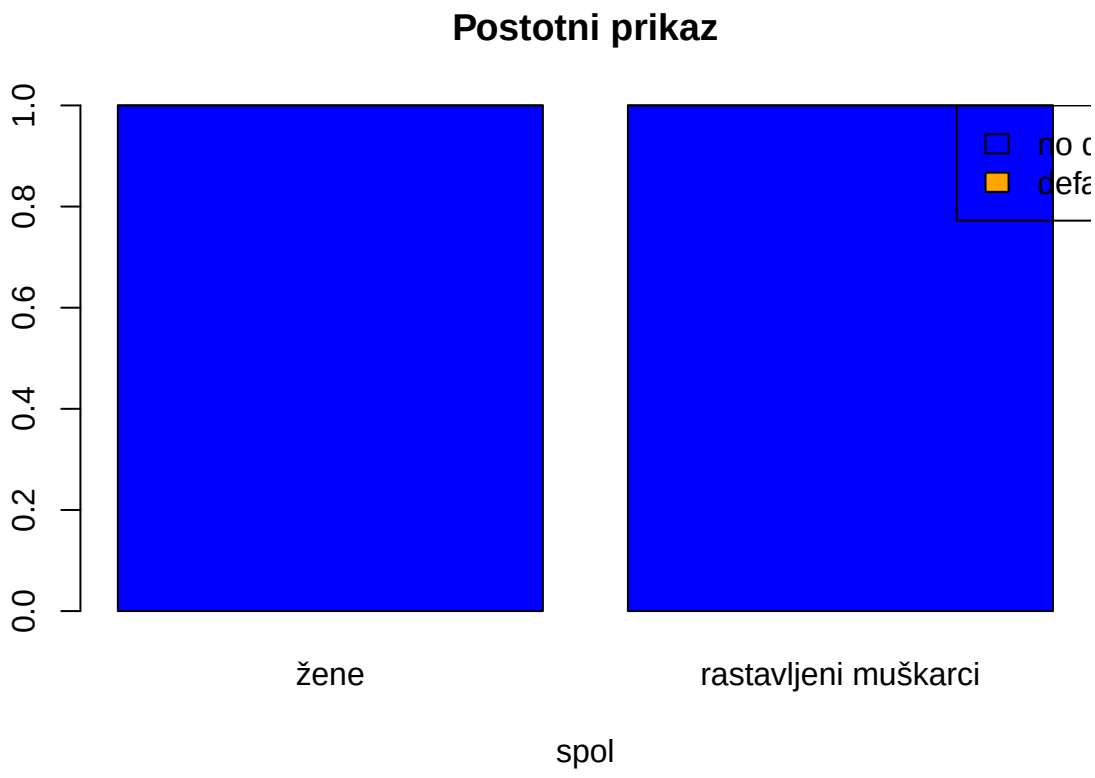
```
## prop 1 prop 2
##      1      1
```

Iz ovoga zaključujemo, na razini značajnosti 0.05, da rastavljeni muškarci ispunjavaju kreditne obveze razmjerno ženama (tj. ne možemo reći da su skloniji neispunjavanju obveza).

```
Values <- matrix(c((nrow(female_clients)-num_female_default),
                   (nrow(male_divor_clients)-num_male_divor_default),
                   num_female_default,
                   num_male_divor_default
                     ), nrow=2, ncol=2, byrow = T)

barplot(Values, main="Kvantitativni prikaz", names.arg=c("žene", "rastavljeni muškarci"), xlab="spol", c

legend("topright", inset = c(-0.1, 0), c("no default", "default"), fill = c("blue", "orange"))
```




```
Values <- matrix(c((nrow(female_clients)-num_female_default)/nrow(female_clients),
                   (nrow(male_divor_clients)-num_male_divor_default)/nrow(male_divor_clients),
                   num_female_default/nrow(female_clients),
                   num_male_divor_default/nrow(male_divor_clients)
                     ), nrow=2, ncol=2, byrow = T)

barplot(Values, main="Postotni prikaz", names.arg=c("žene", "rastavljeni muškarci"), xlab="spol", col =

legend("topright", inset = c(-0.1, 0), c("no default", "default"), fill = c("blue", "orange"))
```



Statistika nad oženjenim muškarcima i ženama u skupu podataka

H0: Proporcija onih koji nisu ispunili obveza naspram onih koji su ispunili obaveze jednaka je kod oženjenih muškaraca i žena (ili je manja kod oženjenih muškaraca).
H1: Proporcija osoba koje nisu ispunile obaveze naspram onih koji su ispunili obaveze veća je kod oženjenih muškaraca.

```
male_married_clients <- male_clients[str_detect(male_clients$PersonalStatus, "married"),]
num_male_married_default <- sum(male_married_clients$Default == 1)

proportion_matrix[1,] <- c(nrow(male_married_clients) - num_male_married_default,
                           num_male_married_default)
# proportion_matrix
prop.test(proportion_matrix, alternative = "less")
```

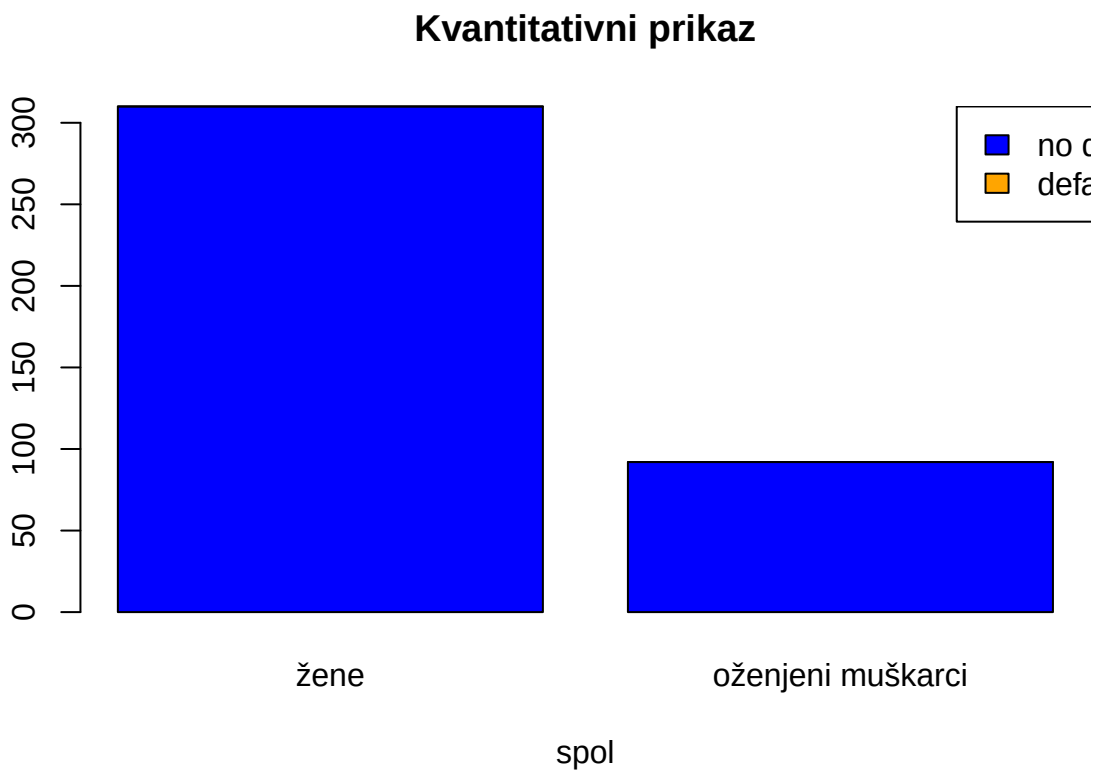
```
## Warning in prop.test(proportion_matrix, alternative = "less"): Chi-squared
## approximation may be incorrect
```

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  proportion_matrix
## X-squared = NaN, df = 1, p-value = NA
## alternative hypothesis: less
## 95 percent confidence interval:
##  -1  0
## sample estimates:
```

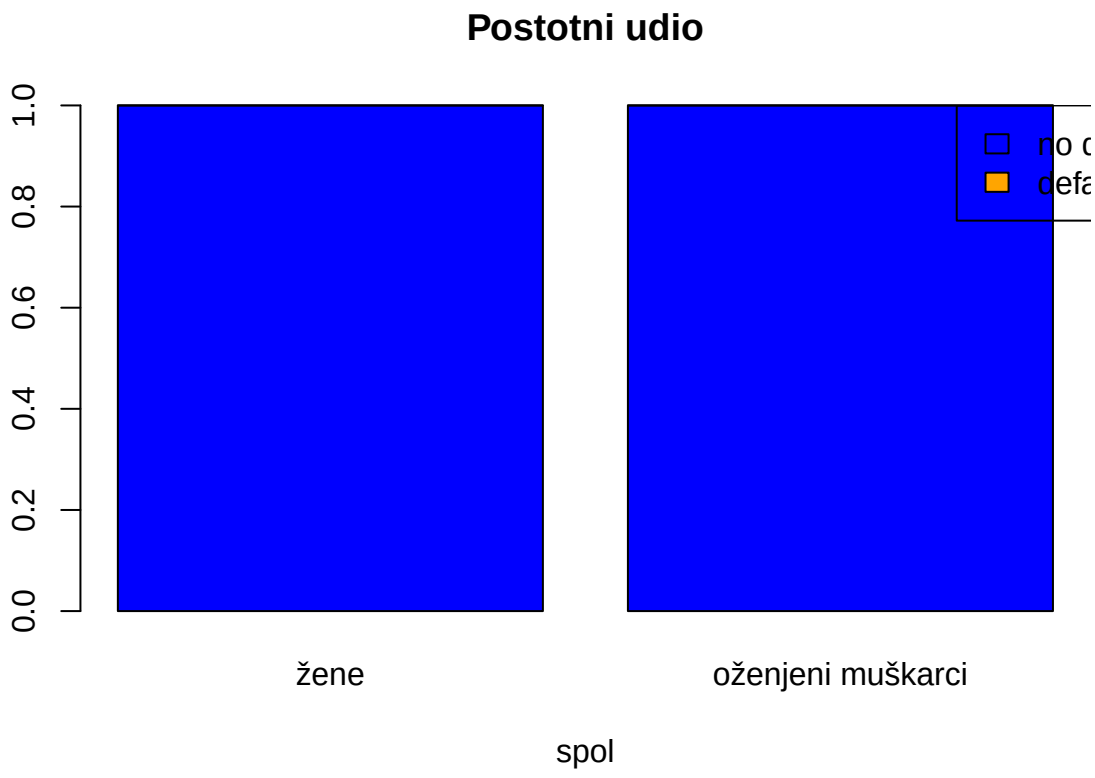
```
## prop 1 prop 2
##      1      1
```

Iz ovoga zaključujemo, na razini značajnosti 0.05, da oženjeni muškarci ispunjavaju kreditne obveze razmjerno ženama (tj. ne možemo reći da su skloniji neispunjavanju obveza).

```
Values <- matrix(c((nrow(female_clients)-num_female_default),
                   (nrow(male_married_clients)-num_male_married_default),
                   num_female_default,
                   num_male_married_default
                     ), nrow=2, ncol=2, byrow = T)

barplot(Values, main="Kvantitativni prikaz", names.arg=c("žene", "oženjeni muškarci"), xlab="spol", col

legend("topright", inset = c(-0.1, 0), c("no default", "default"), fill = c("blue", "orange"))
```




```
Values <- matrix(c((nrow(female_clients)-num_female_default)/nrow(female_clients),
                   (nrow(male_married_clients)-num_male_married_default)/nrow(male_married_clients),
                   num_female_default/nrow(female_clients),
                   num_male_married_default/nrow(male_married_clients)
                     ), nrow=2, ncol=2, byrow = T)

barplot(Values, main="Postotni udio", names.arg=c("žene", "oženjeni muškarci"), xlab="spol", col = c("bl

legend("topright", inset = c(-0.1, 0), c("no default", "default"), fill = c("blue", "orange"))
```



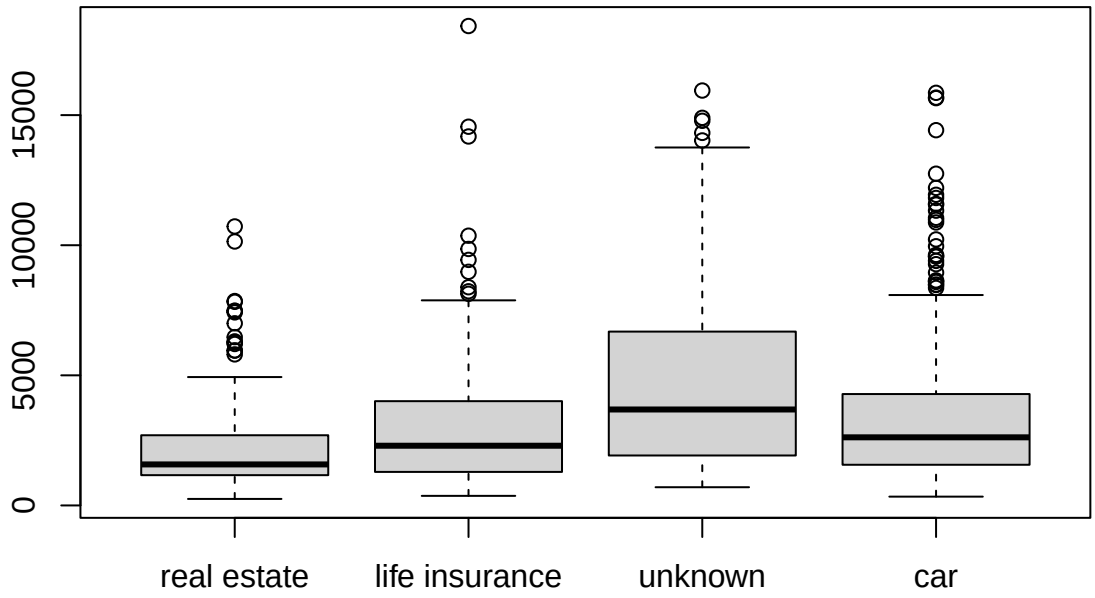
3. pitanje: Postoje li razlike u traženom iznosu kredita prema imovini klijenta?

```
c("real estate", "building society savings agreement/ life insurance",
  "unknown / no property", "car or other, not in attribute Account") %>%
  sapply(function(x) {
    filter(data, Property==x) %>% pull(CreditAmount) -> numbers
    str_c(x, " n: ", length(numbers), "\n") %>% cat()
    print(summary(numbers))
    str_c(x, " standard deviation: ", sd(numbers), "\n") %>% cat()
    cat("--------------------------------------------\n")
    numbers
  }) -> Prop_category
```

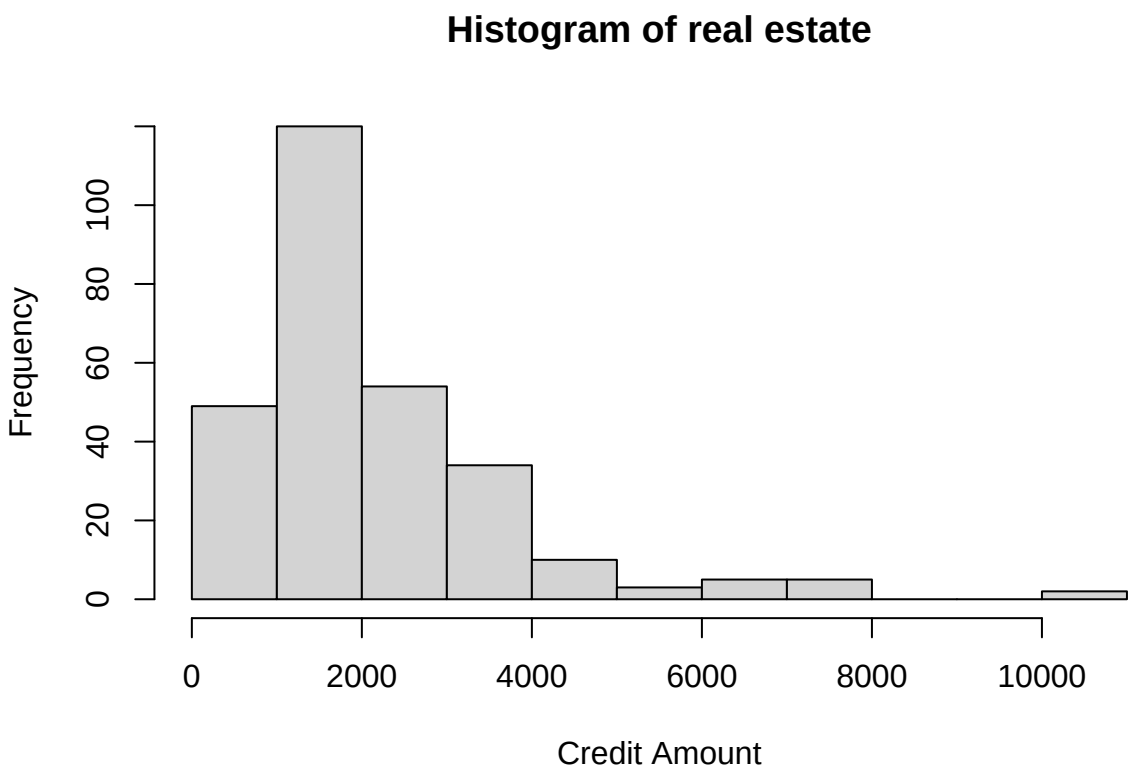
```
## real estate n: 282
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     250    1164    1576    2153    2694   10722
## real estate standard deviation: 1606.27879330167
## --------------------------------------------
## building society savings agreement/ life insurance n: 232
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     368    1288    2294    3104    3990   18424
## building society savings agreement/ life insurance standard deviation: 2602.53168475544
## --------------------------------------------
## unknown / no property n: 154
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     700    1923    3687    4917    6664   15945
```

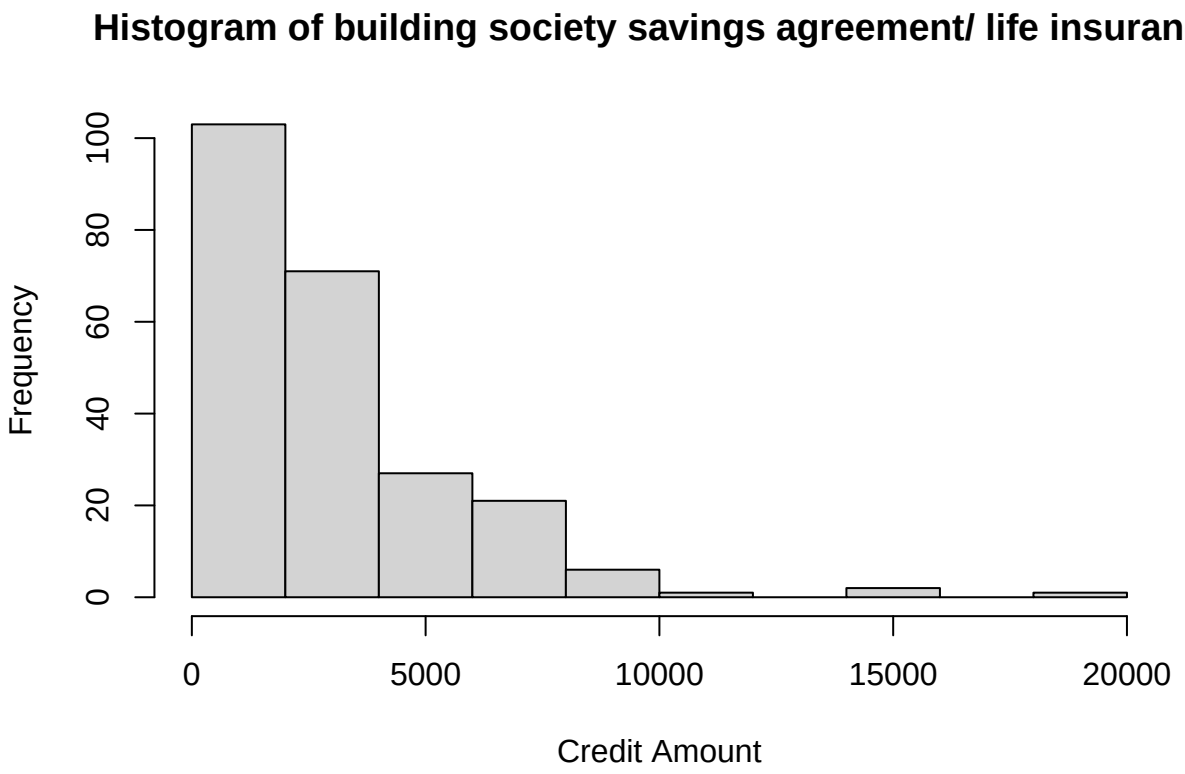
```
## unknown / no property standard deviation: 3725.2304734243
## ------------------------------------------
## car or other, not in attribute Account n: 332
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     338    1565    2618    3574    4280   15857
## car or other, not in attribute Account standard deviation: 2877.33655331269
## ------------------------------------------
```

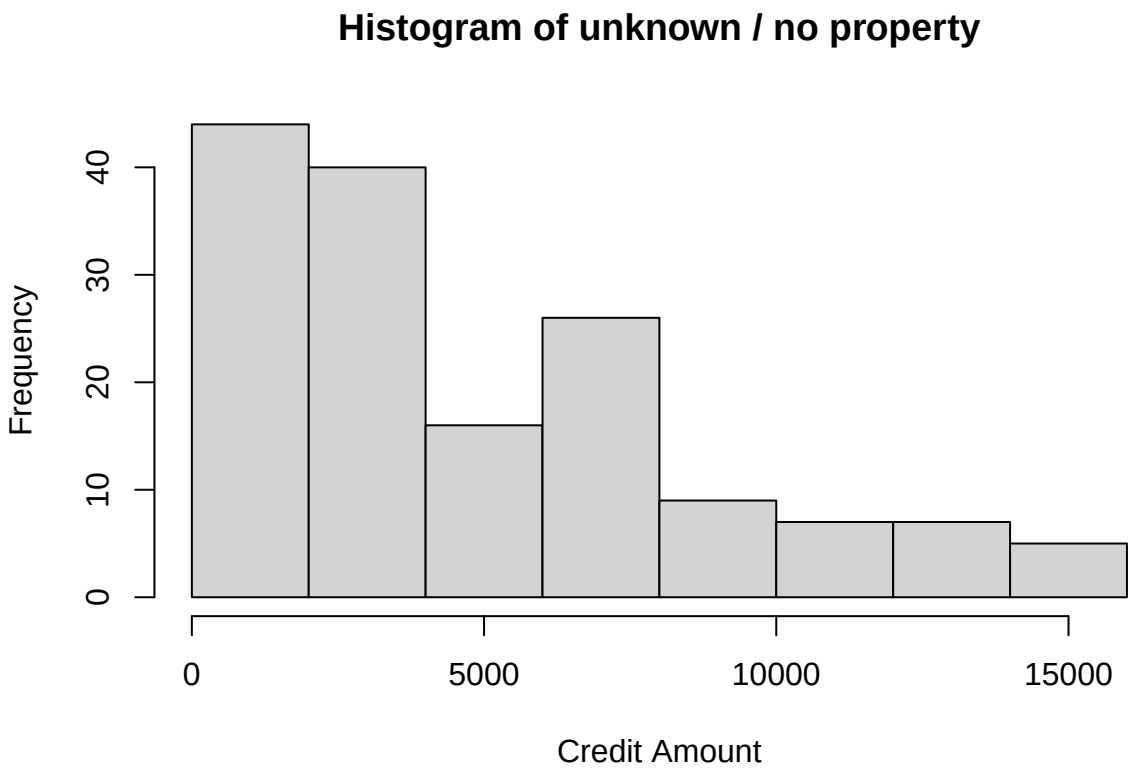
```
boxplot(Prop_category, names=c("real estate", "life insurance", "unknown", "car"))
```

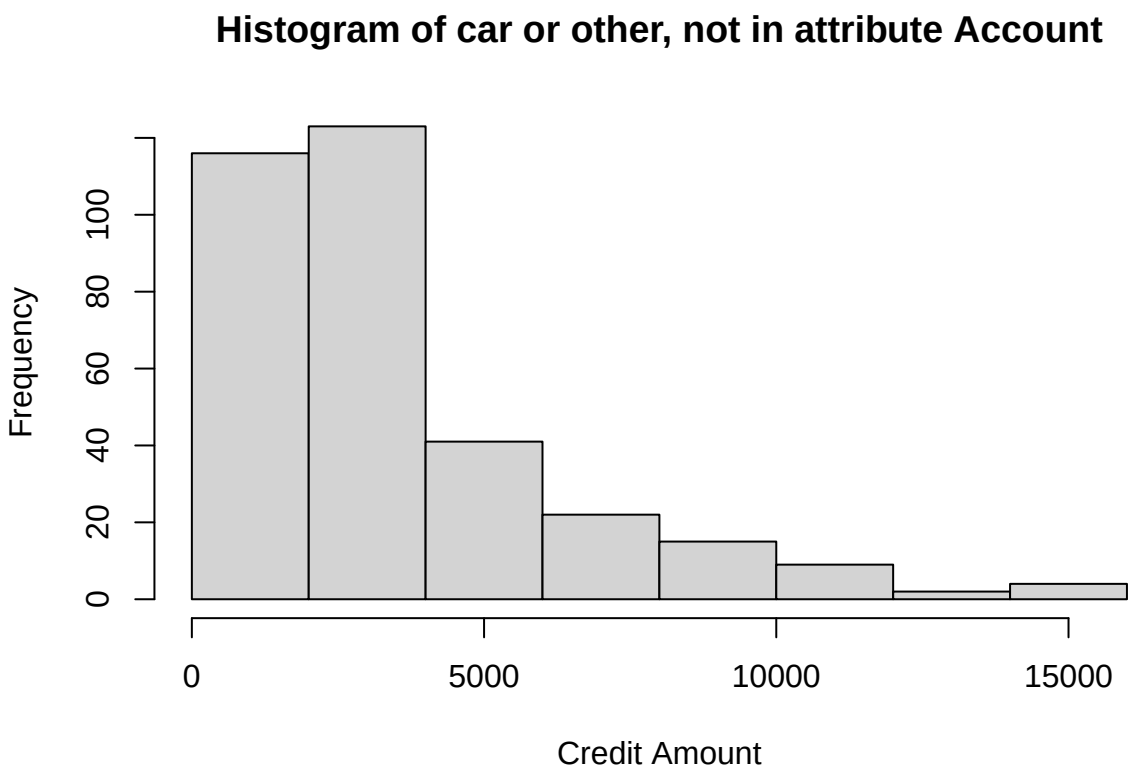



```
for(x in 1:length(Prop_category)) {
  hist(Prop_category[[x]], main = str_c("Histogram of ", names(Prop_category)[x]), xlab="Credit Amount"
}
```


Histogram of real estate
Frequency
0
20
40
60
80
100
0
2000
4000
6000
8000
10000
Credit Amount


Histogram of building society savings agreement/ life insurance
Frequency
0
20
40
60
80
100
0
5000
10000
15000
20000
Credit Amount


Histogram of unknown / no property
Frequency
0
10
20
30
40
0
5000
10000
15000
Credit Amount


Histogram of car or other, not in attribute Account
Frequency
0
20
40
60
80
100
0
5000
10000
15000
Credit Amount