

گز ارش پروژه اول دانیال سعیدی

بخش اول

کد های این بخش در فایل part1.r نوشته شده است. نتیجه اجرای برنامه:

```
Last login: Wed Nov 4 19:13:12 on ttys000

(base) daniel@Daniels-MacBook-Pro - % cd Documents/GitHub
(base) daniel@Daniels-MacBook-Pro GitHub % cd EngineeringProbabilityStatistics
(base) daniel@Daniels-MacBook-Pro EngineeringProbabilityStatistics % ls

CAI
(base) daniel@Daniels-MacBook-Pro EngineeringProbabilityStatistics % cd (AI
(base) daniel@Daniels-MacBook-Pro CAI % ls
Archive.zip (A#1,pdf Rplots,pdf exam_data.csv outcome.csv outcome2.csv part1.r part2.r
(base) daniel@Daniels-MacBook-Pro CAI % Rscript part1.r

During startup - Morning messages:
1: Setting LC_OULATE failed, using "C"
2: Setting LC_ITIME failed, using "C"
3: Setting LC_MONETARY failed, using "C"
4: Setting LC_MONETARY failed, using "C"
5: Setting LC_MONETARY failed, using "C"
11 "Exam 1:"

Max: 100 Madian: 77.5 Mean: 79.25 Max: 99 Variance: 102.3026
[1] "Exam 2:"

Max: 100 Madian: 75 Mean: 80.55 Max: 100 Variance: 160.7868

% 25% 50% 75% 100%

-16.00 -6.25 -5.00 6.75 29.00

(base) daniel@Daniels-MacBook-Pro CAI %

A online of the failed of
```

امتحان اول ویژگی های زیر را دارد:

• Max: 99

Median: 77.5Mean: 79.25

• Max: 99

• Variance: 102.3026

امتحان دوم ویژگی های زیر را دارد:

Max: 100Median: 75Mean: 80.55

• Max: 100

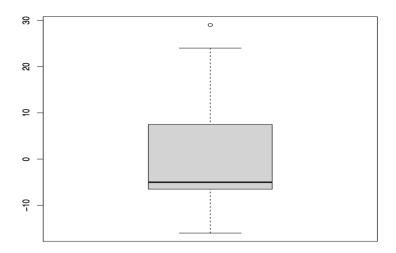
• Variance: 160.7868

یک وکتور به نام diffVector تعریف کردم که اختلاف نمرات افراد را ذخیره می کند. مقدار چارک های diffVector با استفاده از تابع quantile به دست می آید:

0% 25% 50% 75% 100%

-16.00 -6.25 -5.00 6.75 29.00

نمودار جعبه diffVecotr



بخش دوم

کد های این بخش در فایل part1.r نوشته شده است. نتیجه اجرای برنامه:

```
During startup - Warning messages:

1: Setting LC_COLLATE failed, using "C"

2: Setting LC_TIME failed, using "C"

3: Setting LC_MESSAGES failed, using "C"
    Setting LC_MONETARY failed, using "C"
                                                                         "martial.status"
"sex"
 [1] "age"
[5] "occupation"
                                                   "education"
                            "workclass"
                            "relationship"
      "nativecountry" "income"
ge workclass education
                                                         martial.status
    age
                   State-gov
                                 Bachelors
                                                          Never-married
                                                                                    Adm-clerical
         Self-emp-not-inc
Private
                                 Bachelors
                                                    Married-civ-spouse
                                                                                Exec-managerial
                                                                Divorced
                                                                              Handlers-cleaners
                                    HS-grad
                                                    Married-civ-spouse
                                                                              Handlers-cleaners
                      Private
                                  Bachelors
                                                    Married-civ-spouse
                                                                                  Prof-specialty
                      Private
                                    Masters
                                                    Married-civ-spouse
                                                                                 Exec-managerial
                      Private
                                        9th Married-spouse-absent
                                                                                   Other-service
         Self-emp-not-inc
                                    HS-grad
                                                    Married-civ-spouse
                                                                                 Exec-managerial
    31
42
                                                          Never-married
                                  Bachelors
                                                    Married-civ-spouse
                                                                                 Exec-managerial
                                     sex nativecountry income
Male United-States <=50K
                         race
     Not-in-family
                        White
                                    Male
            Husband
            n-family White
Husband Black
                                    Male United-States <=50K
Male United-States <=50K
     Not-in-family
                Wife Black
                                 Female
                                                       Cuba
                        White
                                  Female
                                            United-States
                                   Female Jamaica
Male United-States
     Not-in-family
                        Black
                                                                >50K
>50K
            Husband White
    Not-in-family White
Husband White
                                 Female
Male
                                            United-States
                                            United-States
 Amer-Indian-Eskimo Asian-Pac-Islander
311 1039
                                                                      Black
3124
                                                                                                Other
271
                  White
Warning message:
In yinch(0.1) : y log scale: yinch() is nonsense
(base) daniel@Daniels-MacBook-Pro CA1 %
```

مطابق زیر فایل outcome.csv را می خوانیم و در یک dataframe ذخیره می کنیم.

```
elements <- read.csv("./outcome.csv", head = TRUE,",")

df <- data.frame(elements)

df[df==""] <- NA</pre>
```

ليست نام ستون ها با تابع colnames و به كمك تابع ۱۰ head اول رو را چاپ كردم.

```
colnames(df)
## [1] "age"
                       "workclass"
                                       "education"
                                                       "martial.status"
## [5] "occupation"
                     "relationship"
                                       "race"
                                                       "sex"
## [9] "nativecountry" "income"
head(df, 10)
                workclass education
                                           martial.status
     age
occupation
## 1 39
                State-gov
                           Bachelors
                                            Never-married
                                                               Adm-
clerical
## 2
      50 Self-emp-not-inc Bachelors
                                        Married-civ-spouse
                                                            Exec-
managerial
## 3
      38
                  Private HS-grad
                                                 Divorced Handlers-
cleaners
## 4
      53
                  Private
                               11th
                                        Married-civ-spouse Handlers-
cleaners
## 5
                  Private Bachelors
                                        Married-civ-spouse
                                                             Prof-
specialty
                  Private
                            Masters
                                        Married-civ-spouse
## 6 37
                                                             Exec-
managerial
## 7 49
                  Private
                                9th Married-spouse-absent
                                                              Other-
service
      52 Self-emp-not-inc HS-grad
                                        Married-civ-spouse
## 8
                                                             Exec-
managerial
                                            Never-married
                                                             Prof-
## 9
      31
                  Private
                            Masters
specialty
## 10 42
                  Private Bachelors
                                        Married-civ-spouse
                                                             Exec-
managerial
       relationship
                    race
                              sex nativecountry income
                             Male United-States <=50K
## 1
      Not-in-family White
            Husband White
## 2
                            Male United-States <=50K
## 3
      Not-in-family White
                             Male United-States <=50K
           Husband Black
                             Male United-States <=50K
## 4
## 5
              Wife Black Female
                                           Cuba <=50K
              Wife White Female United-States <=50K
## 6
## 7
      Not-in-family Black Female
                                        Jamaica <=50K
## 8
           Husband White Male United-States >50K
```

```
## 9 Not-in-family White Female United-States >50K
## 10 Husband White Male United-States >50K
```

به کمک تابع table مقدار تکرار متغیر های race را چاپ کردم:

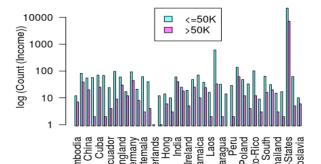
```
table(df$race)
##
## Amer-Indian-Eskimo Asian-Pac-Islander Black
Other
## 311 1039 3124
271
## White
## 27816
```

grouped bar plot رسم نمودار

```
counts <- table(df$income, df$nativecountry)</pre>
counts[counts==0] <- NA # because Log(0) is not defined
barplot(counts, main="Income vs Native Country",
        ylab="log (Count (Income))",
                                              # set column colors
        col = cm.colors(2),
        legend = rownames(counts),
                                              # legend variables
        args.legend = list(x = "top"),
                                              # Legend position
        las = 2,
                                              # Rotate x Labels name 90 degree
        beside=TRUE,
                                              # columns of a row come beside
                                              # y axis in Logarithm
        log='y',
## Warning in yinch(0.1): y log scale: yinch() is nonsense
```

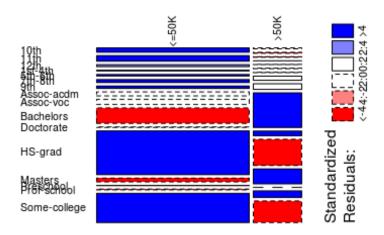
محور x اسم nativecountry و محور y لگاریتم تعداد تکرار متغیر ها می باشد. با آرگومان las = 2 اسم ها را ۹۰ درجه می چرخانیم تا بشود در یک صفحه جا داد!

Income vs Native Country



نمودار موزاییکی متغیرهای income,education

Income vs Education



نمودار تجمعي فراواني افراد

```
plot(ecdf(df$age),
    main= "cumulative distribution of Age",
    xlab="age",
    ylab="cumulative distribution function",
    col = cm.colors(1),
    )
```

cumulative distribution of Age

