Vin banii odata cu varsta?

Proiect Probabilitati si Statistica

Niculae Stefan
Trinh Xuan Hung

Popovici Georgiana

ian 2015
An 2 Sem I
FMI UB

Introducere

Revista Forbes a publicat lista celor mai instariti 100 de directori executivi (CEO) din 2012. Acestia au fost pusi in clasament in functie de venitul anual (in milioane de dolari).
Pe langa venit, noi ii vom analiza si in functie de varsta, pentru a putea raspunde la intrebarea *vin banii odata cu varsta*? Mai concis; exista o corelatie intre varsta directorilor si venitul lor anual?

intrebarea:

Exista vreo corelatie intre varsta si venitul directorilor executivi?

ipoteze:

Vom mai testa trei ipoteze

1. Datele sunt independete
2. Datele sunt repartizate normal
3. Datele au aceeasi dispersie

Prima ipoteza este verificata prin definitie, odata cu colectarea datelor.

sursa:

date - http://www.forbes.com/lists/2012/12/ceo-compensation-12_rank.html

Importare date

```
> ceos = read.table ("~/Desktop/CEOs.txt",
col.names=c("salary", "age"))
> ceos
    salary age
1   131.19  53
2    66.65  72
3    64.40  55
4    60.94  67
5    55.79  59
6    51.52  57
7    50.18  55
8    48.83  59
9    43.71  61
10   43.19  60
11   42.66  59
12   41.99  56
13   41.48  63
14   41.47  58
15   41.09  58
16   40.74  52
17   39.83  61
18   38.53  65
19   39.27  61
20   36.33  50
21   34.30  65
22   33.82  59
23   33.20  58
24   33.17  66
25   32.75  57
26   32.30  64
27   30.89  58
28   29.67  59
29   29.00  66
30   28.62  56
31   27.18  61
32   27.05  73
33   26.48  57
34   26.00  52
35   25.92  58
36   25.84  60
37   25.62  54
38   25.48  59
39   25.37  67
40   24.84  63
41   24.79  81
42   24.04  56
43   23.84  60
44   23.38  59
45   23.15  61
46   22.19  57
47   21.78  46
48   21.74  57
49   21.61  63
50   21.32  61
51   21.18  63
52   20.87  69
53   20.83  66
54   20.81  61
55   20.78  63
56   20.05  59
57   19.99  57
58   19.87  51
59   19.83  61
60   19.83  52
61   19.79  59
62   19.77  60
63   19.66  62
64   19.34  63
65   19.26  49
66   19.23  60
67   19.20  60
68   19.09  43
69   18.77  52
70   18.59  54
71   18.54  64
72   18.39  59
73   17.85  63
74   17.76  52
75   17.68  49
76   17.68  59
77   17.54  49
78   17.52  56
79   17.45  63
80   17.36  64
81   17.27  61
82   17.17  57
83   17.13  60
84   17.03  52
85   16.84  63
86   16.67  57
87   16.55  62
88   16.54  68
89   16.51  57
90   16.44  66
91   16.38  63
92   16.24  61
93   16.17  66
94   16.07  68
95   15.84  65
96   15.61  65
97   15.57  58
98   15.53  60
99   15.52  61
100  15.44  49
```

Partitionare date

Am ales jumatatea listei ca limita dintre bogati si “saraci” apoi am divizat lista in cele doua categorii: wealthy si poor.

```
> limit = ceos[50, 1]
> limit
[1] 21.32
> wealthy = ceos[ceos[,"salary"] > limit, "age"]
> wealthy
 [1] 53 72 55 67 59 57 55 59 61 60 59 56 63 58 58 52 61 65
[19] 61 50 65 59 58 66 57 64 58 59 66 56 61 73 57 52 58 60
[37] 54 59 67 63 81 56 60 59 61 57 46 57 63
> poor = ceos[ceos[,"salary"] <= limit, "age"]
> poor
 [1] 61 63 69 66 61 63 59 57 51 61 52 59 60 62 63 49 60 60
[19] 43 52 54 64 59 63 52 49 59 49 56 63 64 61 57 60 52 63
[37] 57 62 68 57 66 63 61 66 68 65 65 58 60 61 49 50 53
```

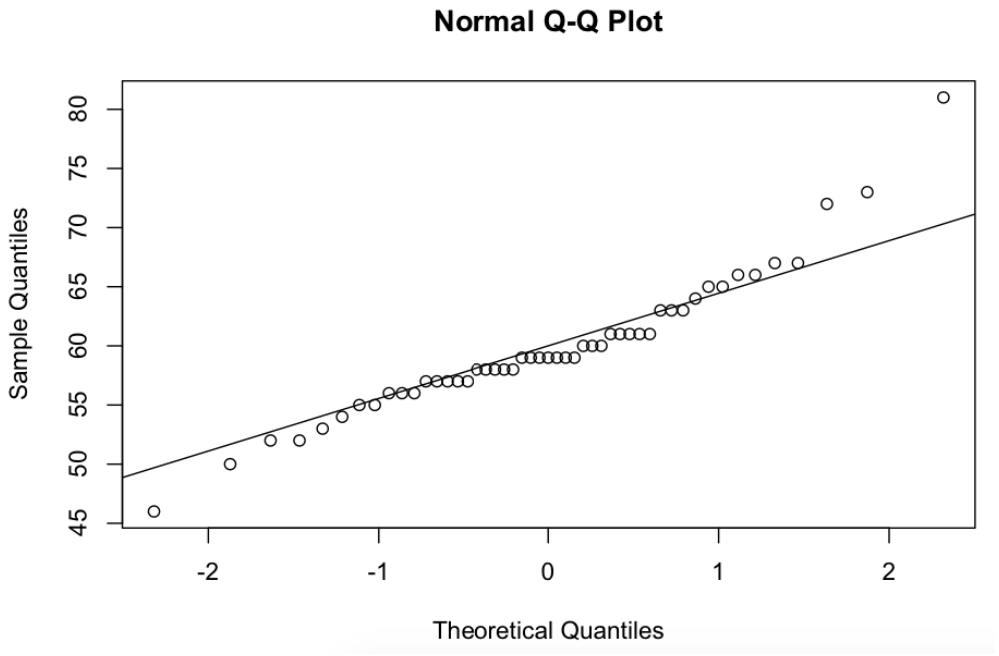
rezumate:

```
> summary(wealthy)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  46.00   57.00   59.00   59.86   63.00   81.00
> summary(poor)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  43.00   56.00   60.00   58.96   63.00   69.00
```

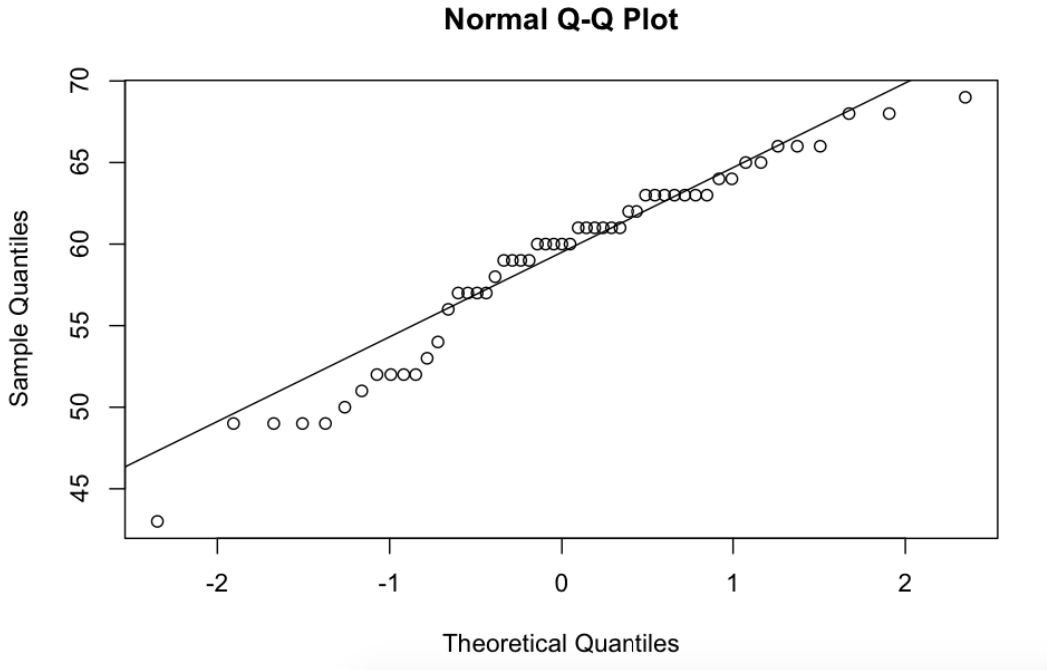
Normalitatea distributiei

wealthy:

qqnorm(wealthy); qqline(wealthy)



poor:

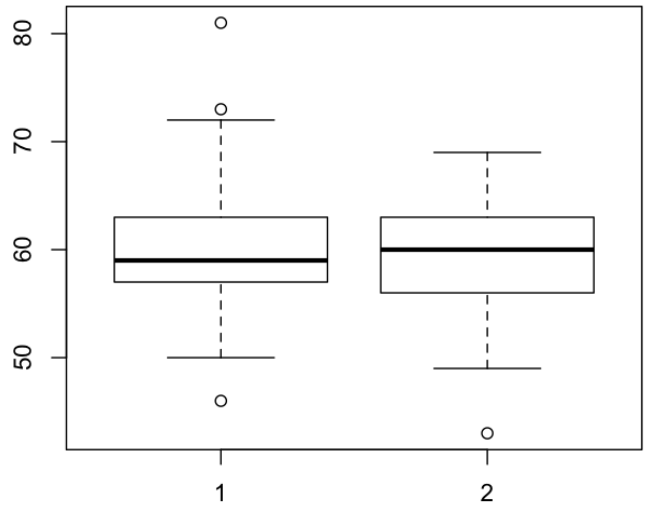


Dispersia si asemanarea

Dispersile de varsta in functie de venit sunt foarte asemanatoare:

```
> var(wealthy)
[1] 35.41667
> var(poor)
[1] 34.11393
```

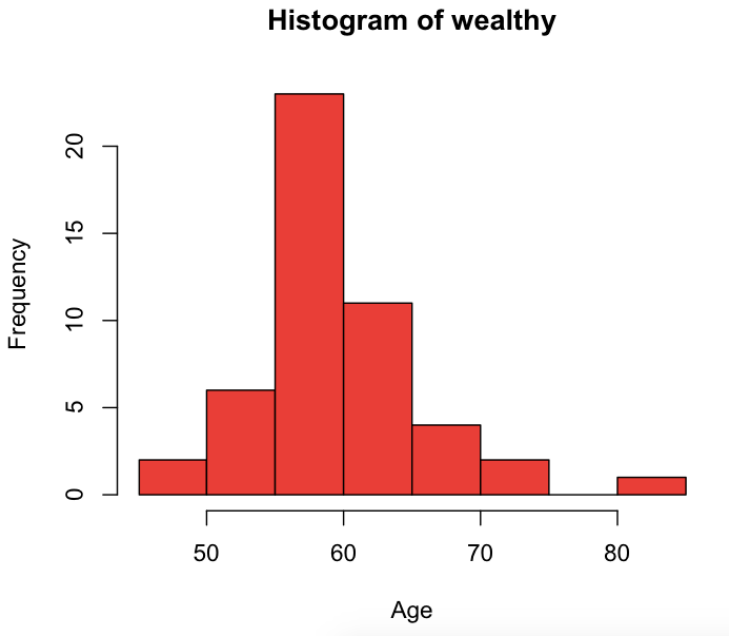
Boxploturile ne confirma acelasi lucru:

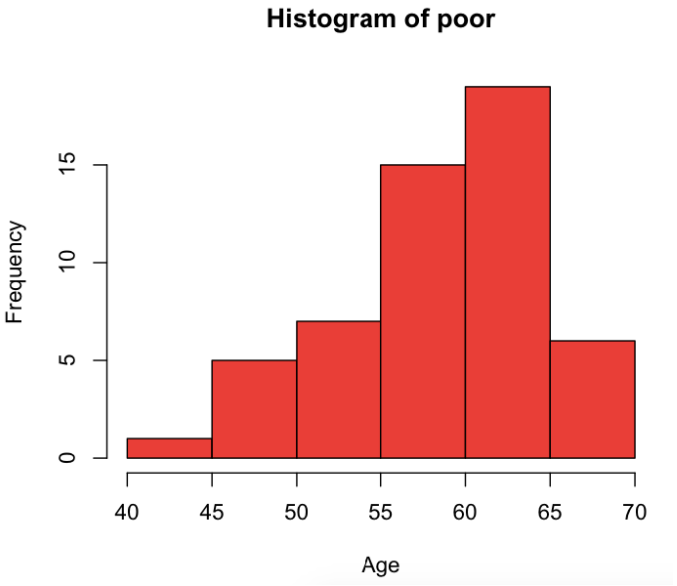


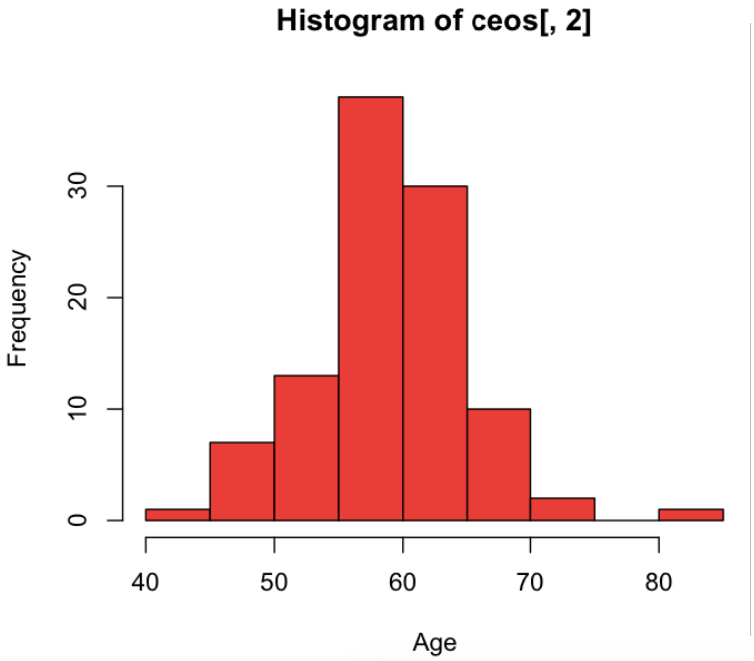
Testele F si T:

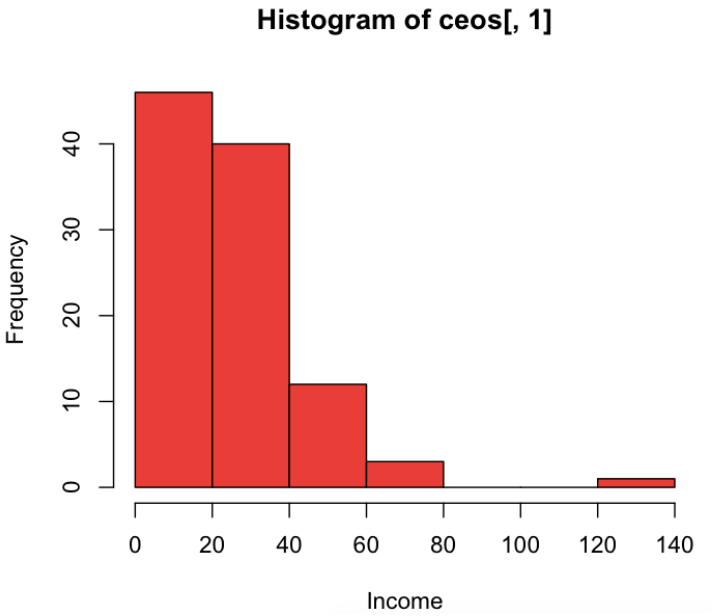
```
> var.test(wealthy, poor)

    F test to compare two variances

data:  wealthy and poor
F = 1.0382, num df = 48, denom df = 52, p-value =
0.8922
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5941998 1.8255053
sample estimates:
ratio of variances
          1.038188

> t.test(wealthy, poor)

    Welch Two Sample t-test

data:  wealthy and poor
t = 0.7655, df = 99.049, p-value = 0.4458
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.424561  3.214318
sample estimates:
mean of x mean of y
 59.85714  58.96226
```

Histograme


Histogram of wealthy
Frequency
0
5
10
15
20
50
60
70
80
Age


Histogram of poor
Frequency
0
5
10
15
40
45
50
55
60
65
70
Age


Histogram of ceos[, 2]
Frequency
0
10
20
30
40
50
60
70
80
Age


Histogram of ceos[, 1]
Frequency
0
10
20
30
40
0
20
40
60
80
100
120
140
Income

Concluzie

ipoteza 2:

Dupa cum se observa din ploturile quarteliene, atat varstele celor bogati cat si ale celor saraci par sa urmeze o distributie normala.

ipoteza 3:

Dispersia este si ea foarte similara in ambele cazuri (~34 in ambele cazuri).

raspunsul:

In concluzie, venitul nu pare sa tina cont de varsta, cel putin asta indica topul celor mai instariti 100 de directori executivi ai Americii. Cu toate ca exista o mica inclinatie spre partea mai tanara, mediile de varsta sunt foarte apropiate:

```
> mean(wealthy)
[1] 59.85714
> mean(poor)
[1] 58.96226
```

Testul de corelatie ne confirma acest lucru:

```
> cor.test(ceos["salary"], ceos["age"])

    Pearson's product-moment correlation

data:  ceos[, "salary"] ard ceos[, "age"]
t = -0.2234, df = 98, p-vclue = 0.8237
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.2180144  0.1746296
sample estimates:
        cor
-0.02256246
```