ACADEMIA DE STUDII ECONOMICE

Facultatea: Cibernetică, Statistică și Informatică Economică

Specializarea: Cibernetică Economică



**Proiect Analiza Datelor**

Studenți: Ivasișin Andrei, Murgeanu Andrei

Grupa: 1062

Profesor îndrumător: Matei Monica Mihaela

1. Definirea variabilelor utilizate

Pentru analiza corespondetelor am folosit datele din proiectul analizei componentelor principale pe care le-am extins la un numar de 33 de inregistrari, referitoare la PIB-ul si numarul imigrantilor din tarile europene. Dorim sa urmarim daca exista corespondenta perfecta intre PIB si numarul de imigranti, in sensul ca o tara cu un PIB mare are si un numar mare de imigranti.

Datele folosite au fost preluate de pe ec.europa.eu/eurostat

1. Analiza distributiilor variabilelor

-PIB

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Min | 1st Qu. | Median | Mean | 3rd Qu. | Max | Abatare std. |
| 5386 | 27211 | 172614 | 405697 | 335282 | 2393250 | 642070.6 |

Observam ca s-au inregistrat valori intre 5386 si 2393250 milioane euro.

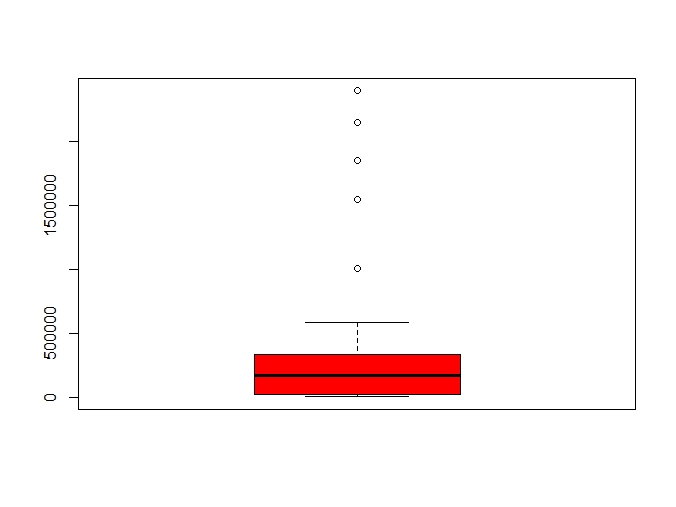
25% dintre inregistrari sunt situate sub valoarea de 27211 mil. Euro.

50% dintre inregistrari sunt situate sub valoarea de 172614 mil. Euro.

25% dintre inregistrari sunt situate peste valoarea de 335282 mil. Euro.

Tarile au inregistrat in medie valoarea de 405697 mil. Euro.

Valorile s-au abatut de la medie cu 642070.6 mil. Euro.



Observam 5 tari (Germania, Franta, Spania, etc) care au un PIB foarte mare in comparative cu restul tarilor.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Min | 1st Qu. | Median | Mean | 3rd Qu. | Max | Abatare std. |
| 1320 | 12546 | 29661 | 135767 | 122415 | 958266 | 223078.5 |

Observam ca s-au inregistrat valori intre 1320 si 958266 de imigranti.

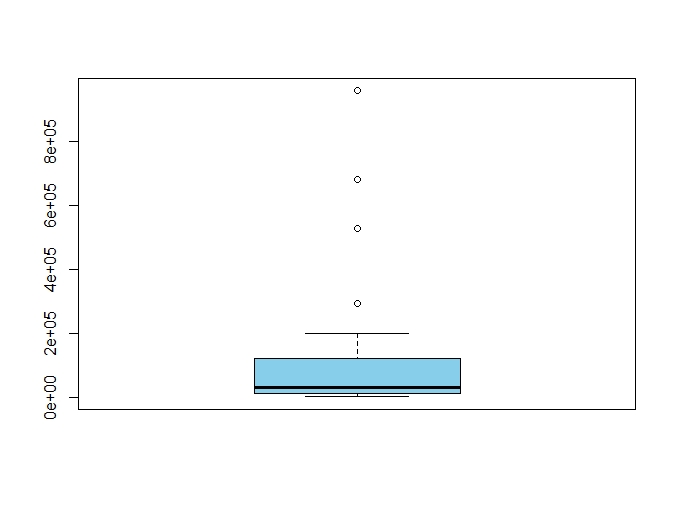
25% dintre inregistrari sunt situate sub valoarea de 12546 de imigranti.

50% dintre inregistrari sunt situate sub valoarea de 29661 de imigranti.

25% dintre inregistrari sunt situate peste valoarea de 122415 de imigranti.

Tarile au inregistrat in medie valoarea de 122415 de imigranti.

Valorile s-au abatut de la medie cu 223078.5 unitati.



Observam 4 tari care au inregistrat un numar foarte mare de imigranti in anul respectiv.

3.Analiza corespondentelor

Am grupat valorile PIB-ului in trei categorii:PIB-High, PIB-Medium, PIB-Low.

|  |  |  |
| --- | --- | --- |
| PIB-High | PIB-Medium | PIB-Low |
| 7 | 13 | 13 |

Am folosit functia “qnorm” pentru a putea imparti esantionul in trei intervale egale.

In primul interval se regasesc tarile care au PIB-ul mai mic sau egal cu valoarea de 58615.27mil. Euro.

In al doilea interval se regasesc tarile a caror PIB este cuprins intre valorile 58615.27 si 564929.

Iar in ultimul interval sunt situate tarile cu un PIB mai mare decat 564929 mil. Euro.

Au reiesit 7 tari cu PIB-ul mare, 13 cu PIB-ul mediu si alte 13 cu PIB-ul redus.

La fel ca mai sus, am grupat numarul imigrantilor:

|  |  |  |
| --- | --- | --- |
| Imigranti-High | Imigranti-Medium | Imigranti-Low |
| 7 | 10 | 16 |

In primul interval se regasesc tarile care au avut un numar de imigranti mai mic decat 13485.

In al doilea interval sunt situate tarile cu un numar de imigranti cuprins intre 13485 si 189256.

In al treilea interval sunt situate tarile cu un numar de imigranti mai mare decat 189256.

In urma gruparii au reiesit 7 tari care au un numar mare de imigranti, 10 care au un numar relative mediu, iar 16 tari care au un numar mic de imigranti.

**Matricea de contingenta**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Imigranti-High | Imigranti-Medium | Imigranti-Low |
| PIB-High | 5 | 1 | 1 |
| PIB-Medium | 1 | 8 | 4 |
| PIB-Low | 1 | 1 | 11 |

In matrice putem observa ca:

-5 tari al caror PIB este mare, iar numarul de imigranti este mare

-O tara a carei PIB este mare, iar numarul de imigranti este mediu

-O tara a carei PIB este mare, iar numarul de imigranti este mic

-O tara a carei PIB este mediu, iar numarul de imigranti este mare

-8 tari al caror PIB este mediu, iar numarul de imigranti este mediu

-4 tari al caror PIB este mediu, iar numarul de imigranti este mic

-O tara a carei PIB este mic, iar numarul de imigranti este mare

-O tara a carei PIB este mic, iar numarul de imigranti este mediu

-11 tari al caror PIB este mic, iar numarul de imigranti este mic

**Testul Chi-square**

|  |  |  |
| --- | --- | --- |
| X-squared=23.564 | Df=4 | P-value= 9.768e-05 |

P-value este egal cu 0.0000976, aproximativ 0. Sunt sanse mici sa gresim daca daca ignoram ipoteza nula, deci acceptam ipoteza alternative. Exista asociere intre cele doua variabile categoriale. Spre exemplu, tarile cu un PIB mare au si un numar mare de imigranti. La fel si in celelalte cazuri.

**Statisticile descriptive-analiza corespondetelor**

Principal inertias (eigenvalues):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dim | Value | % | Cum% | Screen plot |
| 1 | 0.412330 | 57.7 | 57.7 | \*\*\*\*\*\*\*\*\*\*\*\*\*\* |
| 2 | 0.301720 | 42.3 | 100 | \*\*\*\*\*\*\*\*\*\*\* |
|  | ------------ | ------- |  |  |
| Total | 0.714050 | 100 |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | name | mass | qlt | inr | K=1 | cor | ctr | K=2 | Cor | ctr |
| 1 | PIBH | 212 | 1000 | 450 | -1222 | 970 | 756 | 214 | 30 | 32 |
| 2 | PIBL | 394 | 1000 | 261 | 180 | 68 | 31 | -664 | 932 | 575 |
| 3 | PIBM | 394 | 1000 | 289 | 473 | 426 | 213 | 548 | 574 | 393 |

Columns:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | name | mass | qlt | inr | K=1 | cor | ctr | K=2 | Cor | ctr |
| 1 | ImgH | 212 | 1000 | 448 | -1203 | 956 | 744 | 249 | 41 | 43 |
| 2 | ImgL | 303 | 1000 | 296 | 109 | 17 | 9 | -828 | 983 | 688 |
| 3 | ImgM | 485 | 1000 | 256 | 458 | 557 | 247 | 409 | 443 | 268 |

Inertia totala este egala cu 0.71.

Prima dimensiune explica din aceasta inertie 57.7%, iar a doua dimensiune explica 42.3%.

O proportie de 0.41 din informatie este explicate din prima dimensiune, iar0.3 din a doua.

Calitatea reprezentarii pentru PIBH,PIBL,PIBM si ImgH,ImgL,ImgM este 1.

In ambele cazuri, atat PIBH cat si ImgH contribuie mai mult la inertia totala (PIBH-450 si ImgH-448), insa celelalte contribuie aproape in mod egal (PIBL:261-PIBM:289 si ImgL:296-ImgM:256).

PIBH se coreleaza cu dimensiunea 1 (0.97, in comparative cu 0.03 din dimensiunea 2). Acesta are coordonatele -1.22 in prima dimensiune si 0.21 in a doua. Contributia reprezentarea axei dimensiunii 1 este egala cu 0.75 iar pentru dimnesiunea 2 este egala cu 0.03.

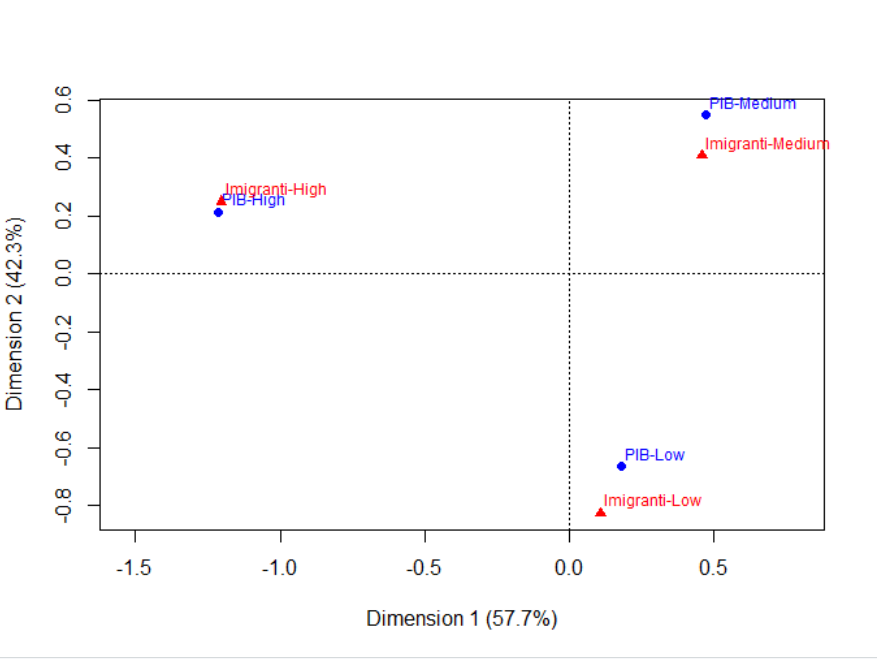
PIBL se coreleaza cu dimensiunea 2 (0.932, in comparative cu 0.068 din dimensiunea 1). Acesta are coordonatele 0.18 in prima dimensiune si -0.66 in a doua. Contributia pentru reprezentarea axei dimensiunii 1 este egala cu 0.03 iar pentru dimnesiunea 2 este egala cu 0.57.

PIBM se coreleaza cu dimensiunea 1 (0.574) si cu dimensiunea 2 (0.426). Acesta are coordonatele -0.47 in prima dimensiune si 0.54 in a doua. Contributia pentru reprezentarea axei dimensiunii 1 este egala cu 0.21 iar pentru dimnesiunea 2 este egala cu 0.39.

ImgH se coreleaza cu dimensiunea 1 (0.956, in comparative cu 0.044 din dimensiunea 2). Acesta are coordonatele -1.20 in prima dimensiune si 0.24 in a doua. Contributia pentru reprezentarea axei dimensiunii 1 este egala cu 0.74 iar pentru dimnesiunea 2 este egala cu 0.04.

ImgL se coreleaza cu dimensiunea 2 (0.983, in comparative cu 0.017 din dimensiunea 1). Acesta are coordonatele 0.1 in prima dimensiune si -0.82 in a doua. Contributia pentru reprezentarea axei dimensiunii 1 este egala cu 0.009 iar pentru dimnesiunea 2 este egala cu 0.68.

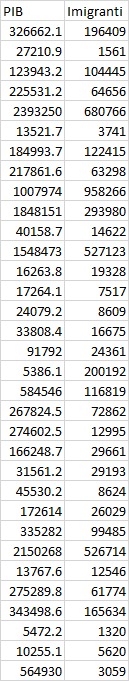
ImgM se coreleaza cu dimensiunea 1 (0.557) si cu dimensiunea 2 (0.443).Acesta are coordonatele -0.45 in prima dimensiune si 0.4 in a doua. Contributia pentru reprezentarea axei dimensiunii 1 este egala cu 0.247 iar pentru dimnesiunea 2 este egala cu 0.268.



Harta ne arata cele doua dimensiuni si corelatiile dintre acestea si categorii.

Observam ca PIB-High si Imigranti-High sunt foarte puternic asociate. Pib-Low si Imigranti-Low sunt asociate, respective PIB-Medium si Imigranti-Medium. Din acestea reiese ca o tara cu PIB-ul mare are si un numar mare de imigranti/ o tara cu PIB mediu are un numar relative mediu de imigranti/ iar unei tari cu PIB mic ii corespunde un numar mic de imigranti.

**ANEXA**



path<-"C:\\Users\\Flo\\Desktop"

date<-read.table(file.path(path,'Date\_2.3.txt'),sep="\t",dec=".",header=TRUE)

date<-as.data.frame(date)

summary(date$PIB)

sd(date$PIB)

qnorm(0.3,400000,651000) #58615.27

qnorm(0.6,400000,651000) #564929

date$PIBc<-rep(0,33)

date$PIBc[which(date$PIB<=58615.27)]<-"PIB-Low"

date$PIBc[which(date$PIB>58615.27 & date$PIB<=564929)]<-"PIB-Medium"

date$PIBc[which(date$PIB>564929)]<-"PIB-High"

table(date$PIBc)

summary(date$Imigranti)

sd(date$Imigranti)

qnorm(0.3,132000,226000)

qnorm(0.6,132000,226000)

date$IMIGRANTIc<-rep(0,33)

date$IMIGRANTIc[which(date$Imigranti<=13485.48)]<-"Imigranti-Low"

date$IMIGRANTIc[which(date$Imigranti>13485.48 & date$Imigranti<=189256.4)]<-"Imigranti-Medium"

date$IMIGRANTIc[which(date$Imigranti>189256.4)]<-"Imigranti-High"

table(date$IMIGRANTIc)

TabelCont<-table(date$PIBc,date$IMIGRANTIc)

TabelCont

chisq.test(TabelCont)

library(ca)

ac<-ca(TabelCont)

summary(ac)

plot(ac)

boxplot(date$PIB,col="red")

boxplot(date$Imigranti,col="skyblue")