**1)Identificați regiunile care obțin scoruri peste Q75% pentru prima componentă principală.**

O imagine care conține text, Font, linie, captură de ecran

Descriere generată automat

Pentru setul de date ales de mine, regiunile care au scoruri peste Q75% pentru prima componentă sunt cele din figura de mai sus, și anume: EL30, ES30, ES51, ES52, ES61, FRJ1, FRK2 și așa mai departe. Cum scorurile reprezintă coordonatele în noul spațiu, iar aceste regiuni înregistrează valori peste Q75% pentru acestea, înseamnă că ele vor avea valori mari pentru prima componentă principală. Deoarece prima componentă principală am denumit-o nivelul de ocupație al populației, putem afirma că aceste regiuni sunt cele mai dezvoltate din punct de vedere al gradului de ocupare al populației.

**2)Care sunt variabilele X cu care această componentă este puternic corelată.**

O imagine care conține text, captură de ecran, Font, număr

Descriere generată automat Dimensiunea arată corelația dintre variabilă și componenta principală. Se poate observa că prima componentă, nivelul de ocupare al populației, va prelua informație din variabilele X6,X9 și X3 deoarece prezintă corelații puternice (0.877, 0.806, -0.771). Aceste variabile reprezintă rata de angajare, rata de angajare a persoanelor cu o vârstă de peste 15 ani și respectiv, populația activă din punct de vedere economic care are un nivel al educației mai mic decât nivelul primar și secundar inferior.

**3)Reprezentați cercul corelațiilor și extrageți două concluzii.**

O imagine care conține text, diagramă, linie, Interval

Descriere generată automat

Așa cum reiese din figura de mai sus, variabila inițială care este cea mai bine reprezentată din punct de vedere al celor 2 componente principale, nivelul de ocupare al populației și respectiv, populația ocupată cu un nivel educațional secundar superior și peste, este X1, care reprezintă populația totală. Variabila X1 se află în al doilea cadran al cercului, cu o corelație directă puternică de 1 cu cea de-a doua componentă, ceea ce arată că va avea un comportament asemănător cu populația ocupată cu un nivel educațional secundar superior și peste (deoarece avem date standardizate). De asemenea, aceasta prezintă o corelație inversă puternică de aproape 0.75 față de prima componentă, ceea ce arată că va evolua opus față de nivelul de ocupare al populației.

O altă concluzie care poate fi extrasă din cercul corelațiilor este faptul că, pentru grupurile de variabile X1,X3,X4,X5 și respectiv X2,X6,X7,X8,X9 se observă un coeficient de corelație mare.

**4)Alegeți 2 variabile X corelate negativ. Transformați-le in variabile categoriale cu 4 nivele și aplicați analiza corespondențelor. Identificați categoriile cu cele mai mari contribuții la inerția totală.**

## Legendă:

L1, ML1,MH1,H1-categorii ale ratei de angajare(X6)

L2, ML2,MH2,H2-categorii al numărul mediu de ore săptămânale obișnuite de muncă la locul de muncă principal, a persoanelor cu vârsta de peste 15 ani(X10)

**O imagine care conține text, captură de ecran, Font, număr

Descriere generată automat**

Din outputul de mai sus se remarcă o valoare a inerției totale de 0.386. Coloana qlt măsoară calitatea reprezentării punctelor respective în harta corespondențelor. În cazul nostru, cele mai bine reprezentate puncte sunt L1 și ML1 deoarece valorile lor se apropie foarte mult de 1 în raport cu prima dimensiune, pe când pentru a doua dimensiune sunt punctele MH2 și L2.

Din punct de vedere al contribuției fiecărei categorii la inerția totala, cea mai mare contribuție o prezintă L2, cu 48,2%, respectiv MH1 cu 33,3%.

**O imagine care conține text, linie, captură de ecran, diagramă

Descriere generată automat**

Categoria L2 este singura poziționată pe partea stângă, în contrast cu celelalte 4. Din punct de vedere al primei axe, distanța dintre L2 și ML2 este mult mai mare față de cea dintre H2 și MH2. Prima dimensiune pune în evidență contrastul dintre categoria corespunzătoare nivelului foarte scăzut al numărul mediu de ore săptămânale de muncă (L2) și toate celelalte.

În ceea ce privește cealaltă variabilă, care reprezintă rata de angajare, observă că L1 și ML1 se detașează de celelalte 2 categorii, fiind situate în extremitatea dreapta.

Se observă o corespondență între nivelul mic al ratei angajării(ML1) cu un nivel ridicat al numărului mediu de ore săptămânale de muncă(MH2), remarcându-se contrastul dintre acestea și restul categoriilor, fiind situate în partea dreaptă, sus. De asemenea, se mai remarcă o corespondență între nivelul extrem de căzut al numărului mediu de ore săptămânale de muncă(L2) și nivelul ridicat al ratei de angajare(MH1).

**5)Efectuați analiza factorială pentru 50 de regiuni extrase aleator dintre cele din data frame-ul X.**

**O imagine care conține text, Font, captură de ecran

Descriere generată automat**

Indicele KMO indică proporția din varianța variabilelor care poate fi cauzată de factorii latenți. Cum acest indice are o valoare mai mică decât 0,5, ceea ce reprezintă un nivel inacceptabil pentru realizarea analizei factoriale, ceea ce înseamnă că nu se pot extrage factori comuni.

O imagine care conține text, captură de ecran, Font, număr

Descriere generată automat

Conform testului Bartlett, cum p-value<0.05, se respinge H0 și se acceptă H1, ceea ce înseamnă că există cel puțin un factor comun.

O imagine care conține text, captură de ecran, linie, diagramă

Descriere generată automat

Primul factor latent ML1 imprimă variație asupra lui X6, X9,X8, și X7, care reprezintă, în ordine: Rata de angajare, Rata de angajare a persoanelor cu o vârstă de peste 15 ani, Rata de angajare a persoanelor cu un nivel educațional terțiar, și respectiv, Rata de angajare a persoanelor cu un nivel educațional mai mic decât nivelul primar și secundar inferior. Astfel, putem denumi primul factor comun nivelul de utilizare al forței de muncă. Se poate observa o corelare negativă a variabilei X10, care reprezintă Numărul mediu de ore săptămânale obișnuite de muncă la locul de muncă principal, a persoanelor cu vârsta de peste 15 ani cu acest factor.

Cel de-al doilea factor latent ML2 imprimă variație asupra lui X1,X5,X4 și X3, care reprezintă, în ordine: Populația totală, Populația activă din punct de vedere economic care au un nivel al educației terțiar, Populația activă din punct de vedere economic care au un nivel al educației secundar superior, post-superior și non-terțiar și respectiv, Populația activă din punct de vedere economic care are un nivel al educației mai mic decât nivelul primar și secundar inferior. Astfel, putem denumi cel de-al doilea factor comun nivelul populației active.

Se remarcă faptul că varianța variabilei X2, care reprezintă Vârsta mediană a populației, nu este surprinsă de nici unul dintre factorii aleși.

O imagine care conține text, captură de ecran, linie, diagramă

Descriere generată automat

Așa cum reiese din outputul de mai sus, în analiza factorială este recomandat să extragem 3 factori latenți.

O imagine care conține text, linie, captură de ecran, diagramă

Descriere generată automat

Așa cum reiese din diagrama de mai sus, reiese că vor fi 2 factori comuni, nivelul de utilizare al forței de muncă și nivelul populației active. Cel de-al treilea factor este unul de tip unic, care va influența în mod unilateral o singură variabilă, și anume X2(Vârsta mediană a populației).

# CODUL ÎN R

date<-read.table(file="BD2.txt",row.names=1,sep="\t",header=TRUE,dec=".")

date

summary(date)

##1. Identificați regiunile care obțin scoruri peste Q75% pentru prima componentă principală.

X\_f<-scale(date,center=T,scale=T)

X\_f

library(corrplot)

R<-cor(date)

corrplot(R,method="number",type="upper")

R\_f<-cor(X\_f)

corrplot(R\_f,method="number",type="upper")

C<-cov(X\_f)

C

library(moments)

library(FactoMineR)

library(factoextra)

library(nFactors)

library(dplyr)

acp<-PCA(X\_f,scale.unit=TRUE)

acp

summary(acp)

fviz\_eig(acp)

fviz\_pca\_var(acp,col.var="contrib")

acp2<-princomp(X\_f,cor=TRUE,scores=TRUE)

acp2$scores

scoruri<-acp2$scores

q<-quantile(scoruri[,1],probs=c(0.25,0.5,0.75))

q

r<-which(acp2$scores[,1]>1.5165200)

r

View(acp2$scores)

##2. Care sunt variabilele X cu care această componentă este puternic corelată.

MF<-cor(X\_f,scoruri)

corrplot(MF,method="number",type="upper")

##3. Reprezentați cercul corelațiilor și extrageți două concluzii.

fviz\_pca\_var(acp,col.var="contrib")

##4. Alegeți 2 variabile X corelate negativ. Transformați-le in variabile categoriale cu 4 nivele și aplicați

##analiza corespondențelor. Identificați categoriile cu cele mai mari contribuții la inerția totală.

q1<-quantile(date$X6,probs=c(0.25,0.5,0.75))

q1

dim(date)

X6\_cat<-rep(0,224)

X6\_cat[date$X6<=q1[1]]<-"L1"

X6\_cat[date$X6>q1[1]& date$X6<=q1[2]]<-"ML1"

X6\_cat[date$X6>q1[2]& date$X6<=q1[3]]<-"MH1"

X6\_cat[date$X6>q1[3]]<-"H1"

table(X6\_cat)

q2<-quantile(date$X10,probs=c(0.25,0.5,0.75))

q2

X10\_cat<-rep(0,224)

X10\_cat[date$X10<=q2[1]]<-"L2"

X10\_cat[date$X10>q2[1]& date$X10<=q2[2]]<-"ML2"

X10\_cat[date$X10>q2[2]& date$X10<=q2[3]]<-"MH2"

X10\_cat[date$X10>q2[3]]<-"H2"

table(X10\_cat)

contingenta<-table(X6\_cat,X10\_cat)

contingenta

library(ca)

ac<-ca(contingenta)

summary(ac)

plot(ac)

##5. Efectuați analiza factorială pentru 50 de regiuni extrase aleator dintre cele din data frame-ul X.

dim(date)

set.seed(100)

indici<-sample(1:166,50,replace=FALSE)

date2<-date[indici,]

dim(date2)

library(psych)

d2\_f<-scale(date2,center=TRUE,scale=TRUE)

R2<-cor(d2\_f)

KMO(R2)

cortest.bartlett(R2)

af3<-fa(d2\_f,nfactors=2,n.obs=50,rotate="varimax",fm="ml")

fa.diagram(af3)

paralel<-fa.parallel(d2\_f)

paralel$nfact

af3<-fa(d2\_f,nfactors=3,n.obs=50,rotate="varimax",fm="ml")

fa.diagram(af3)