Introduction à R

Méthodes de pondération

Respecter le cadre théorique

Methodes stat (ACP- reduire dim et obtenir axes decorelés, mais 2 var tres corrélés et pas parce que tres corrélé qu’il faut les mettre sur mm dim, donc cadre théorique primordial)

Corrélation ≠ causalité

Si tres corrélé pas parce qu’info en double mais parce que tres forte donc indispensable à l’indice

Données avec ≠ échelles : bien normaliser (données sur la meme échele) et pour sous indicateurs normaliser sur la meme echelle

Pondération

Impact significatif sur valeur indice composite, on peut la base sur méthode stats (ACP) ou jugeemnt expert (questionnaire qui indiquent variables bonne à prendre)

Dans le projet on peut avoir un avis critique sur la construction de l’indicateur avec papier de recherche

Pondérer equal weighting (associer le mm poids à chaque var dans l’indicateur compositie) n variables= 1/n poids à chaque variable (chaque var a le meme impact sur indice composite) donc chaque indicateurs sur la meme echelle, methode de normalisation homogène (Z-score, min-max … mais les normaliser de la mm manière à chaque fois, à discuter et faire sur ensemble des variables). Si variables – dispo pour certains pays on suropondère les précédentes pour faire face aux troués

Corrélation : établit double comptage dans l’indice, donc évaluation de la corrélation (Pearson ou Spearman) permet d’éviter double-comptage puis introduire une regle empirique (si corrélé à +0.5 on les supprime …- règle discutable et dépend de chacun)

Double comptage : voir si intéressant mais s’appuyer sur référence académique car info pertinente (ex : pas parce que une entreprise utilise internet qu’elle a un site, surtout en 2003 car après bulle internet, même si donnée fortement corrélé)`

Element matrice nommatrice[ligne,colonnes]

Nb\_lignes<-nrow(mat)

Nb\_col<-ncol(mat)

Length(mat) donne nb lignes x nb colonnes

*Exercice 1 –faire une fonction qui prenne en argument une matrice de n variables, et qui renvoie un indicateur qui agrège chaque variable avec un poids égal*

function (mat) {

vect\_res<-vector(mode="numeric",length=nrow(mat))

nb\_lignes<-nrow(mat)

nb\_col<-ncol(mat)

for (k in 1:nb\_lignes) {

for (t in 1:nb\_col) {

vect\_res[i]<-vect\_res[i]+mat[k,t]

}

}

vect\_res<-(1/nb\_lignes)\*vect\_res

return (vect\_res)

}

Cependant possible en calculant la moyenne à chaque ligne

function (mat) {

vect\_res<-vector(mode="numeric",length=nrow(mat))

nb\_lignes<-nrow(mat)

nb\_col<-ncol(mat)

for (i in 1 :nb\_lignes) {

Vect\_res[i]<-mean(mat[nb\_lignes,]) permet d’éviter double boucle

}

return(vect\_res)

}

*Exercice 2*

rnorm(nb tirages, moyenne, ecart-type) ici n=100 moyenne= 0 ecart-ytpe=1

Set.seed () on utilize avant de faire tirage aléatoire pour toujours retomber sur les mêmes valeurs

Etapes

1. Créer Matrice vide 100 lignes et 5 colonnes matrix(matrix(data = NA, nrow = 1, ncol = 1, byrow = FALSE, dimnames = NULL) si byrow=TRUE alors a matrice est complété ligne par ligne
2. Set.seed()
3. Remplir matrice avec rnorm()
4. Utiliser finction de l’exercice 1pour créer Indice composite

n<-100

k<-5

mat\_2<-matrix(data = NA, nrow=n, ncol=k,byrow = FALSE, dimnames = NULL)

set.seed(123)

for (i in 1:k) {

mat\_2[,i]<-rnorm(n,0,1) #\*on s’en fout de la ligne, le rnorm remplit pour n tirages et ligne=n

}

head(mat\_2)

results\_V1<-IC(mat\_2)

results\_V2<-IC(mat\_2[,1:3])

results\_V3<-IC(mat\_2[,4:5])

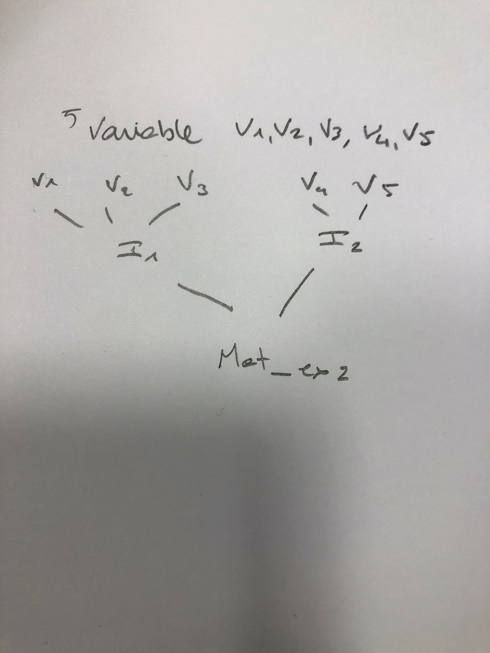
mat\_ex2<-matrix(0,100,2)

mat\_ex2<-c(results\_V2,results\_V3) #on agrège les 2 sous indicateurs ensemble

head(mat\_ex2)

*\*on ne fait qu'une seule boucle car n lignes donc rnorm(n,0,1) remplit la ligne entière on cherche à se déplacer en colonne. On ne cherche pas à remplir 1 par 1 mais lignes par lignes*

Explication



Donc pondération ≠ entre les deux exos

Ex1= 1/5 V1+1/5 V2

Ex2= 1/6 V1+1/6 V2+1/6 V3+1/4 V4+1/4 V5

Dans le cas 2 on accorde – de poids aux variables 1,2 et 3

2. pondération de l’ACP

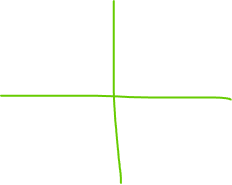
Etapes de la pondération dur l’ACP

Etapes 1 2 & 3 on peut se permettre la flexibilité dans regles de décision et choix des hypothèse

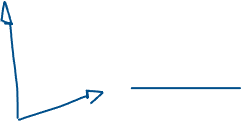
Transformation et une variable n’expliquera jamais 2 axes différents

Dim 1

Dim 2



Dans une autre dimension



Pas du tout corrélé (elles sont dans ≠ dimensions)



Fortement corrélé mais pas dans le meme sens (elles sont dans la meme dimension)

Projet expliquer quel choix on prefere