TP - Analyse des données n^o 4

Analyse en composantes principales

Splus -e	Lancer Splus (avec l'option diteur de ligne).
motif()	Ouverture de la fentre graphique.
par(mfrow=c(1,2))	Dcoupage de la fentre graphique.
help.start()	Ouvrir la fentre d'aide (l'iconifier).
?crbg	Information sur le jeu de donnes "**Cr**ues du **Bag**et".
crbg	Visualisation des donnes numriques.
pairs(crbg)	Visualisation graphique des liaisons deux deux.
res_ acp(crbg,reduc=F)	Lancement de l'acp non rduite.
summary(res)	Contrler les contributions des variables l'inertie. Notez celles qui sont le plus influentes.
contri(res, ind=F)	Contributions et qualits Quelle est la variable mal reprsente sur le plan principal? (cf. tableau `carrs des corrlations multiples ...`).
pairs(cbind(res$cmpr[,1],crbg))	Diagramme de dispersion entre la 1 re composante principale (`res$cmpr[,1]`) et les variables initiales Ne pas tenir compte des `warnings` qui ne sont pas des erreurs
round(cor(cbind(res$cmpr[,1],crbg)),2)	`cor()` calcule la matrice de corrlation
plot(res)	Commentaires?
biplot2(res)	Choisir "isomtrique colonne" : commentaires?
cercor(res)	Des variables de faible influence peuvent tre bien reprsentes.
par(mfrow=c(1,2))	Dcoupage de la fentre graphique.
res_acp(crbg)	Acp rduite :
summary(res)	Quid des contributions?
contri(res, ind=F)	Contributions et qualits : qu'est-ce qui "fait" le troisime facteur? (cf. tableau `Contribution des variables aux facteurs`).
contri(res,ind=T)	Contributions (un facteur s'interprte en fonction. des variables qui y ont la plus forte contribution. ce facteur).
biplot2(res)	Choisir "isomtrique colonne" : commentaires?
par(mfrow=c(2,2))	Dcoupage de la fentre graphique.
biplot2(res,lisinv=list(1:34),marg=F)	Apperu de la succession des relevs. Interprtez les deux premiers facteurs. Copiez le graphique (menu `graph` dans la fentre graphique).
biplot2(res,1,3, lisinv=list(1:34),marg=F)	Visualisation sur les axes 1 et 3. Essayez de caractriser des priodes de la crue.
rescah_cahd(dist(scale(crbg)))	Classification hirarchique : critre de Ward (variables rduites par la fonction `scale`).
plclust(rescah)	Visualisation de l'arbre hirarchique.
pp3_cutree(rescah,3)	Coupure de l'arbre en 3 classes.
par(mfrow=c(2,2))	
biplot2(res,pltcl=pp3)	Visualisation des 3 classes sur 1-2.
biplot2(res,1,3,pltcl=pp3)	Visualisation des 3 classes sur 1-3. Copier ce graphique (menu `graph`).
pp4_cutree(rescah,4)	Coupure de l'arbre en 4 classes.
biplot2(res,pltcl=pp4)	Visualisation des 4 classes sur 1-2.
biplot2(res,1,3,pltcl=pp4)	Visualisation des 4 classes sur 1-3. En conclusion, dites en combien de priodes on peut diviser cette crue et faite une interprtation rapide de ces priodes.

pp4 _factor(pp4)	Transformation en facteur des diffrents numros de classe.
res_cdgsup(res,pp4)	Introduction de la variable numro de classe.
biplot2(res,isup="a")	Visualisation des barycentres de classe “seuls” (alone).
biplot2(res,isup="b", pltcl=pp4)	Visualisation des barycentres de classe avec les individus (both).
dimnames(crbg)[[2]]	Nom des variables de l'acp.
sel_c(F,F,T,T,F,F,T,T,F)	Dfinition d'un “masque” pour slectionner des variables.
dimnames(crbg)[[2]][sel]	Variables slectionnes.
dimnames(crbg)[[2]][!sel]	Variables “masques”.
par(mfrow=c(1,1))	Un graphique pleine page.
biplot2(res,selv=sel,lisinv=list(1:34), callib=T,sep=F,nmark=3,nomi=NULL)	Biplot gradu.
round(reconacp(res,2),2)	Valeurs reconstitues l'ordre 2.
crbg	Vraies valeurs.
	Comparer les valeurs reconstitues avec les valeurs lues sur le biplot gradu.
round(crbg-reconacp(res,2),2)	rsidus : comparer avec le tableau
contri(res, ind=F)	donnant la qualit de la reprsentation des variables.
	Interprtation d'un nouveau jeu de donnes
q()	Sortir de Splus
cp ∼carlier/.Data/.data1.2/perou .Data	Recherche d'un nouveau jeu de données.
	On étudie la tension artérielle de 54 habitants qui ont migré des hauts-plateaux ruraux du Pérou vers les villes situées beaucoup plus bas (une étude précédente a suggéré que la tension artérielle augmentait puis tendait à revenir à la normale).

On effectuera l'acp, en rflchissant aux options utiliser (on pourra en comparer plusieurs). On interprtera les facteurs qui semblent pertinents en expliquant leur nombre. On regardera s'il y lieu de faire des groupements d'individus et on caractérisera les individus 3, 5, 38 et 39.

Les variables initiales sont:

age:	age en années	ep.avbr:	épaisseur du pli de la peau de l'avant-bras en m
ans:	nombre d'années depuis la migration	ep.moll:	épaisseur du pli de la peau du mollet en mms
poids:	poids en kgs	pouls:	pouls en pulsations par mn
taille:	taille en mms	tas:	tension artérielle systolique
ep.ment:	épaisseur du pli de la peau du menton en mms	tad:	tension artérielle diastolique.