

Classes Supérieures (CS)

2ème année- option SIL

Module: ANAD
TP1

<p>TP AFCM</p>

Réalisé par :

- ZIDELMAL Yacine

Table de matières

Partie1: AFCM	3
Etude des statistiques des données	4
Visualisation de la fréquence des catégories des variables	5
Transformation du tableau de données en tableau disjonctif:	6
Faire l'AFCM	8
Représentation du biplot individus-variables	9
Etude du tableau des contributions et donner une signification aux axes.	10
contributions des variables	10
contributions des individus	12
Autres visualisations possibles.	13
Corrélation entre variables	13
Coordonnées des modalités des variables	14
Qualité de représentation des variables (modalités)	15
Qualité de représentation des individus	18
représentations avec groupement des points en ellipses	19
Partie 2 :AFC	22
Croiser 2 questions pour en faire un tableau de contingence.	22
Résultats de l'AFC	23
Interprétation	23

Partie1: AFCM

Nous allons, dans cette partie, réaliser une Analyse Factorielle des Correspondances Multiples sur un ensemble de données qui constitue les réponses au formulaire d'évaluation des enseignements de l'ESI durant l'année universitaire 2018/2019 pour le module Réseau 1 qu'on désignera par RES1.

L'ensemble des données pourra être consulté dans ce [lien](#).

Dans le but de ne pas encombrer l'étude et la rendre plus lisible nous allons faire un pré-traitement à l'ensemble des données

- a) nous allons renommer l'ensemble des variables données dans la feuille excel comme suit

object_			bien_exp						
clairs	prog	pré-requis	osée	bon_encadr	assimil_cours	sup_pedag	vol_hor	control	quant_travail

- b) lire le fichier Excel pour continuer les traitement sur le R
- c) nous allons également remplacer les signes “++”, “+”, “- -”, “-” par “TB”, “B”, “M”, “TM” respectivement. Tel que ces abréviations signifient:
- “TB” : Très bien fait
- “B” : Bien fait
- “M” : Mal fait
- “TM” : Très mal fait

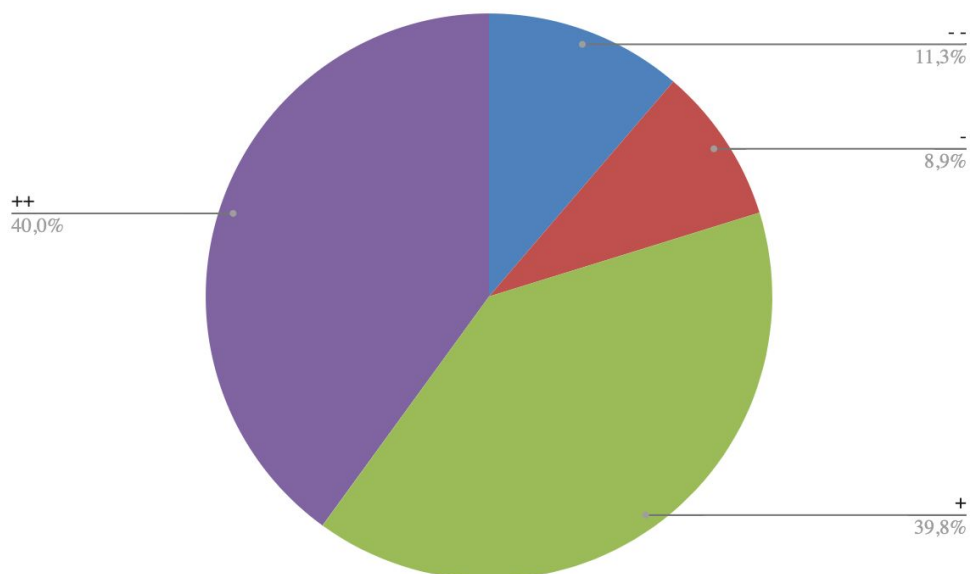
cette série de remplacements est réalisée par le R via le script suivant

```
1 library("readxl")
2 library("FactoMineR")
3 library("factoextra")
4
5 my_data <- read_excel("dataset.xlsx", sheet = "RES1")
6
7 my_data[my_data=="++"]<- "TB"
8 my_data[my_data=="+" ]<- "B"
9 my_data[my_data=="- -"]<- "TM"
10 my_data[my_data=="- " ]<- "TM"
11 my_data[my_data=="- " ]<- "M"
12
13 clean_responses = my_data[1:47,c(2:11)]
```

le résultat que nous allons obtenir à la fin de cette phase de pré-traitement peut être consulté dans ce [lien](#).

Etude des statistiques des données

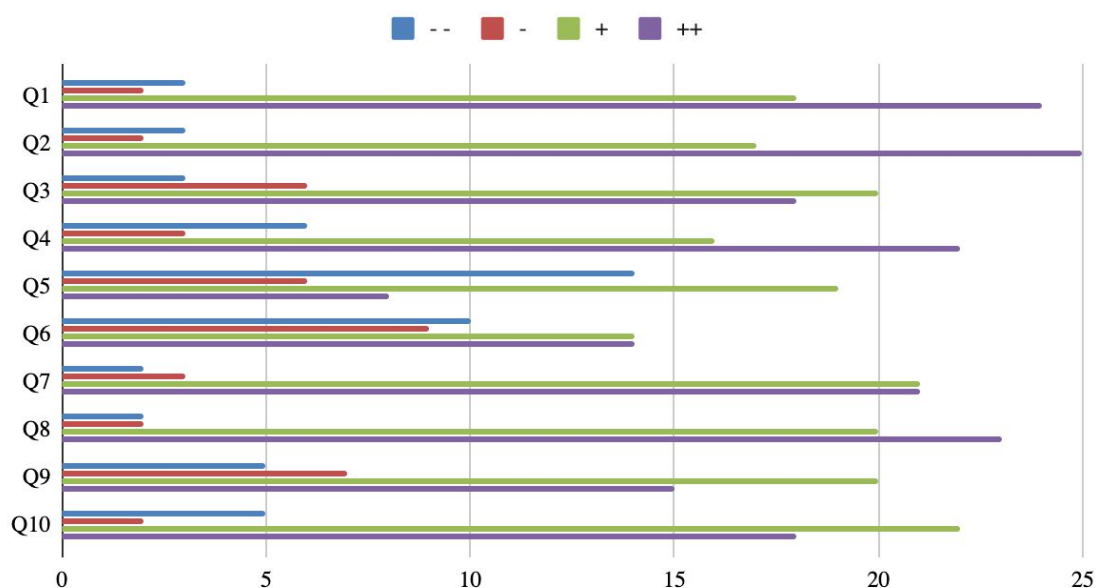
a) Pourcentage general de chacune des réponses



Nous remarquons que 79.8% des réponses des étudiants sont positives ce qui donne une vue globale de satisfaction vis-à-vis du module.

a) fréquences des réponses pour chaque question

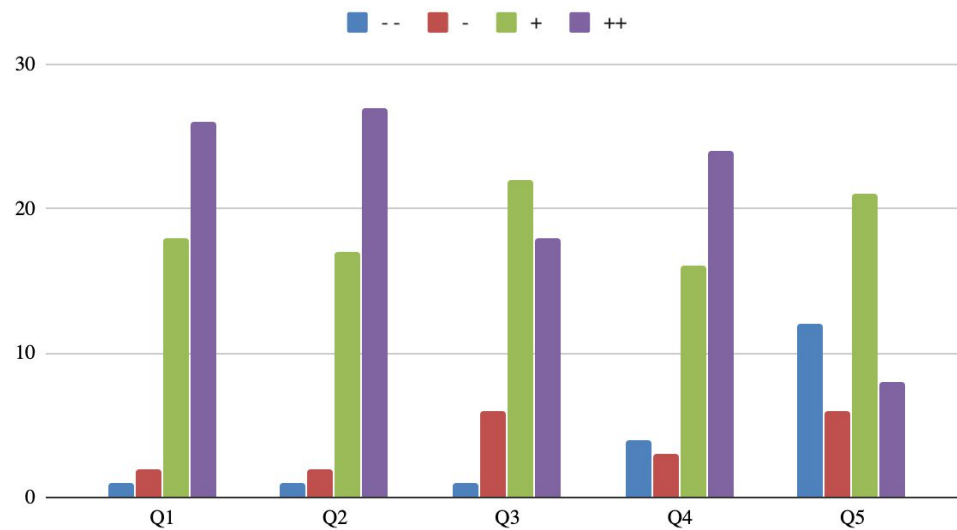
--, -, + et ++



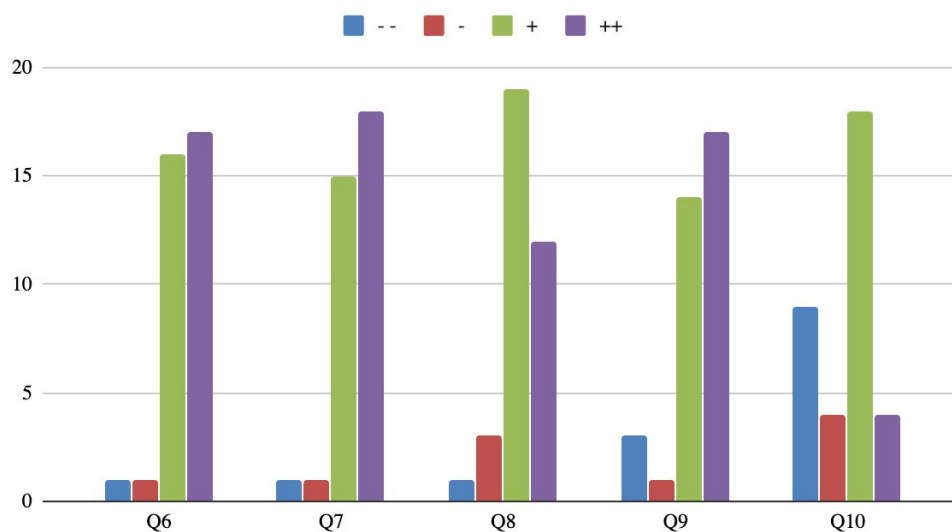
on constate que pour la plupart des questions , on a une opinion positive à l'exception des questions Q5("Je suis bien encadré dans la résolution des exercices") et Q6("Les TD/TP/Projet étaient utiles pour une bonne assimilation du cours"]").

Visualisation de la fréquence des catégories des variables

--, -, + et ++



--, -, + et ++



Transformation du tableau de données en tableau disjonctif:

Pour cela nous allons exécuter la commande suivante sur le R

```
21 "construction de la table disjonctive"
22 library(FactoMineR)
23 disjonctif = tab.disjonctif.prop(clean_responses, seed=NULL, row.w=NULL)
```

La figure suivante représente un extrait de la table disjonctif relativement aux trois premières variables pour les 15 premiers individus.

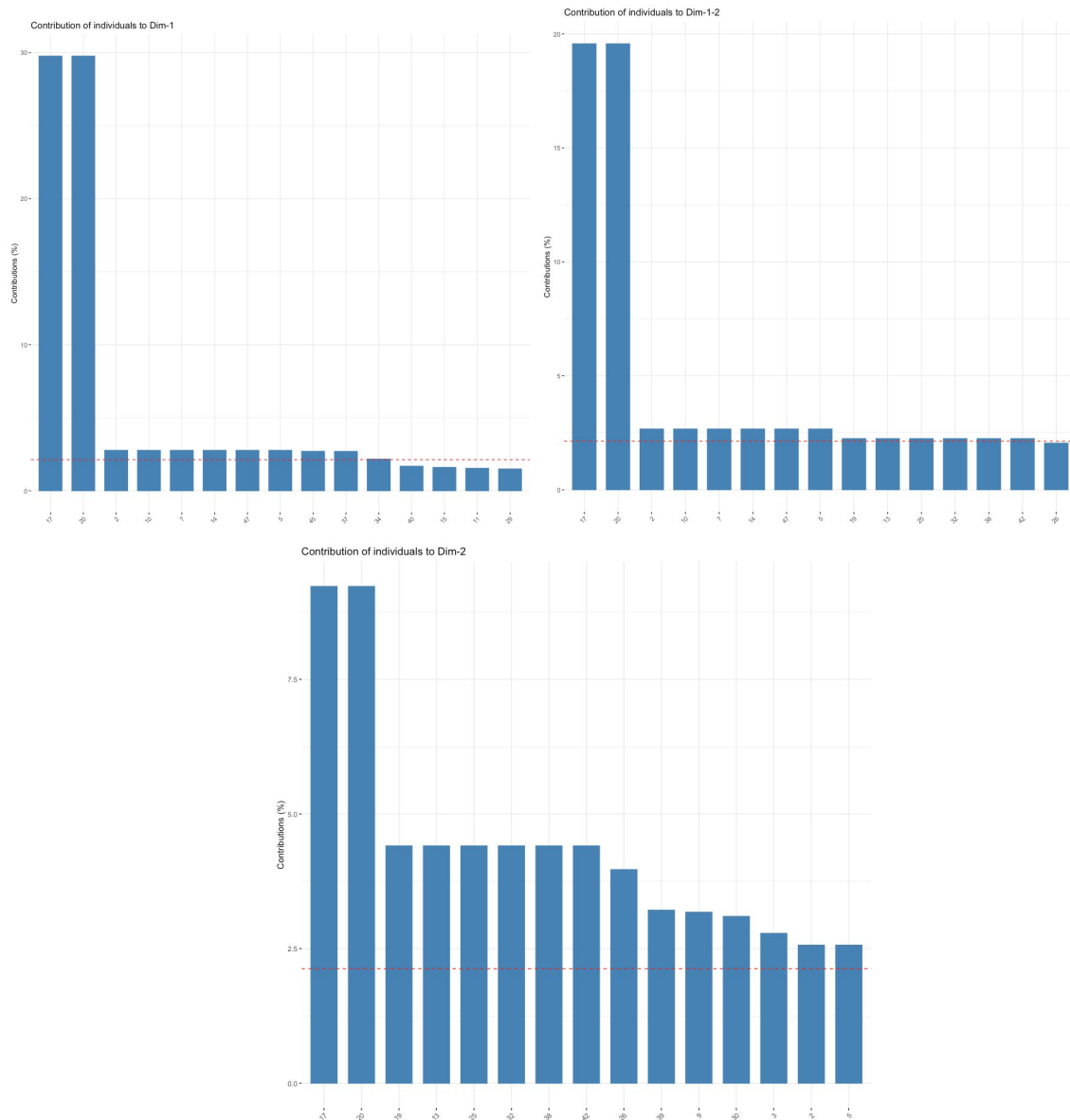
	B	M	TB	TM	B	M	TB	TM	B	M	TB	TM
1	0	0	1	0	0	0	1	0	0	0	1	0
2	0	0	1	0	0	0	1	0	0	0	1	0
3	1	0	0	0	0	1	0	0	0	1	0	0
4	0	0	1	0	0	0	1	0	1	0	0	0
5	0	0	1	0	0	0	1	0	0	0	1	0
6	0	0	1	0	0	0	1	0	1	0	0	0
7	0	0	1	0	0	0	1	0	0	0	1	0
8	0	0	1	0	0	0	1	0	0	0	1	0
9	1	0	0	0	1	0	0	0	0	1	0	0
10	0	0	1	0	0	0	1	0	0	0	1	0
11	0	0	1	0	0	0	1	0	1	0	0	0
12	0	1	0	0	1	0	0	0	0	1	0	0
13	1	0	0	0	1	0	0	0	1	0	0	0
14	0	0	1	0	0	0	1	0	0	0	1	0
15	0	0	1	0	0	0	1	0	0	0	1	0

Remarque:

en effectuant l'analyse pour la première fois , nous sommes rendu compte de l'existence de deux points (individus) aberrants dans le dataset et ils s'agit des individus 17 et 20.

17	16	++	++	+	++	+	+	++	+	+
18	17	--	--	--	--	--	--	--	--	--
19	18	+	+	+	+	+	-	+	+	--
20	19	+	+	+	+	+	+	+	+	+
21	20	--	--	--	--	--	--	--	--	--
22	21	++	++	++	++	+	+	++	+	++
23	22	++	++	++	++	--	+	+	+	+

Ce constat est confirmé en effectuant une AFCM "préliminaire" puis afficher les contributions de chacun des individus sur deux axes puis sur le plan engendré et voici les résultats obtenus



NB:

les commandes qui ont permis d'obtenir ces résultats sont :

```
81 res.mca <- MCA(clean_responses)
82 #***** contributions des individus *****
83 # Contributions of rows to dimension 1
84 fviz_contrib(res.mca, choice = "ind", axes = 1, top = 15)
85 # Contributions of rows to dimension 2
86 fviz_contrib(res.mca, choice = "ind", axes = 2, top = 15)
87 #contribution of rows to plot 1-2
88 fviz_contrib(res.mca, choice = "ind", axes = 1:2, top = 15)
```

Donc ces deux individus (17 et 20) monopolisent une bonne partie de l'information et risquent de nous fausser les résultats.

Alors, pour la suite de toute l'étude, nous allons prendre ces deux individus (17 et 20) comme des individus supplémentaires qui ne participeront pas à la construction des axes factoriels.

Pour cela nous allons exécuter la commande MCA de FactoMineR

```
81 res.mca <- MCA(clean_responses, ind.sup = c(17, 20))
```

au lieu de la commande

```
81 res.mca <- MCA(clean_responses)
```

Faire l'AFCM

a) Pour cela nous allons exécuter la commande MCA de FactoMineR

```
81 res.mca <- MCA(clean_responses, ind.sup = c(17, 20))
```

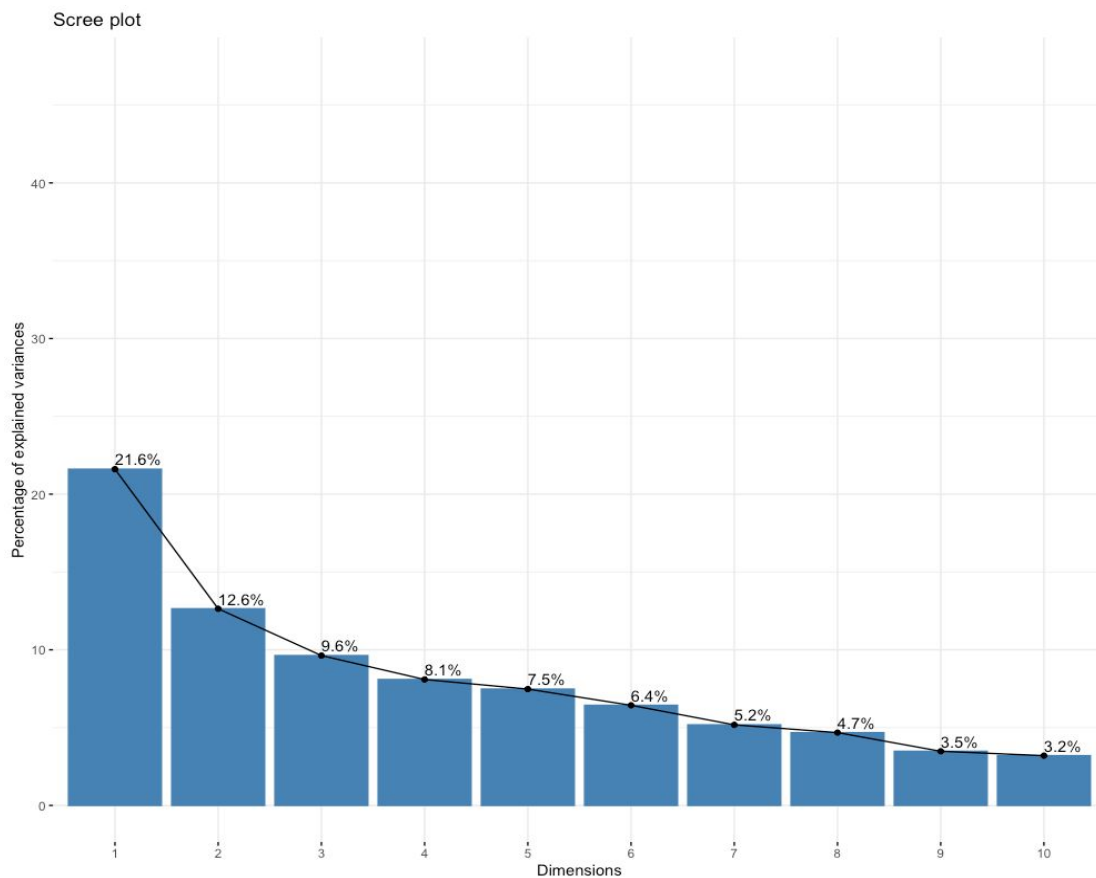
b) Etudier le tableau de valeurs propres:

Exécuter ces commandes pour obtenir un tableau des valeurs propres ainsi qu'une représentation en diagramme à bâtons

```
31 eig.val <- get_eigenvalue(res.mca)
32 View(eig.val)
33 fviz_screplot(res.mca, addlabels = TRUE, ylim = c(0, 47))
```

Voici les résultats obtenus

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	0.6050030634	21.60725226	21.60725
Dim.2	0.3538729687	12.63832031	34.24557
Dim.3	0.2696717948	9.63113553	43.87671
Dim.4	0.2265860027	8.09235724	51.96907
Dim.5	0.2093679997	7.47742856	59.44649
Dim.6	0.1801919599	6.43542714	65.88192
Dim.7	0.1450315986	5.17969995	71.06162
Dim.8	0.1310244752	4.67944554	75.74107
Dim.9	0.0974618628	3.48078081	79.22185
Dim.10	0.0895842414	3.19943719	82.42128
Dim.11	0.0797518440	2.84828014	85.26956
Dim.12	0.0698885970	2.49602132	87.76559
Dim.13	0.0625653211	2.23447575	90.00006



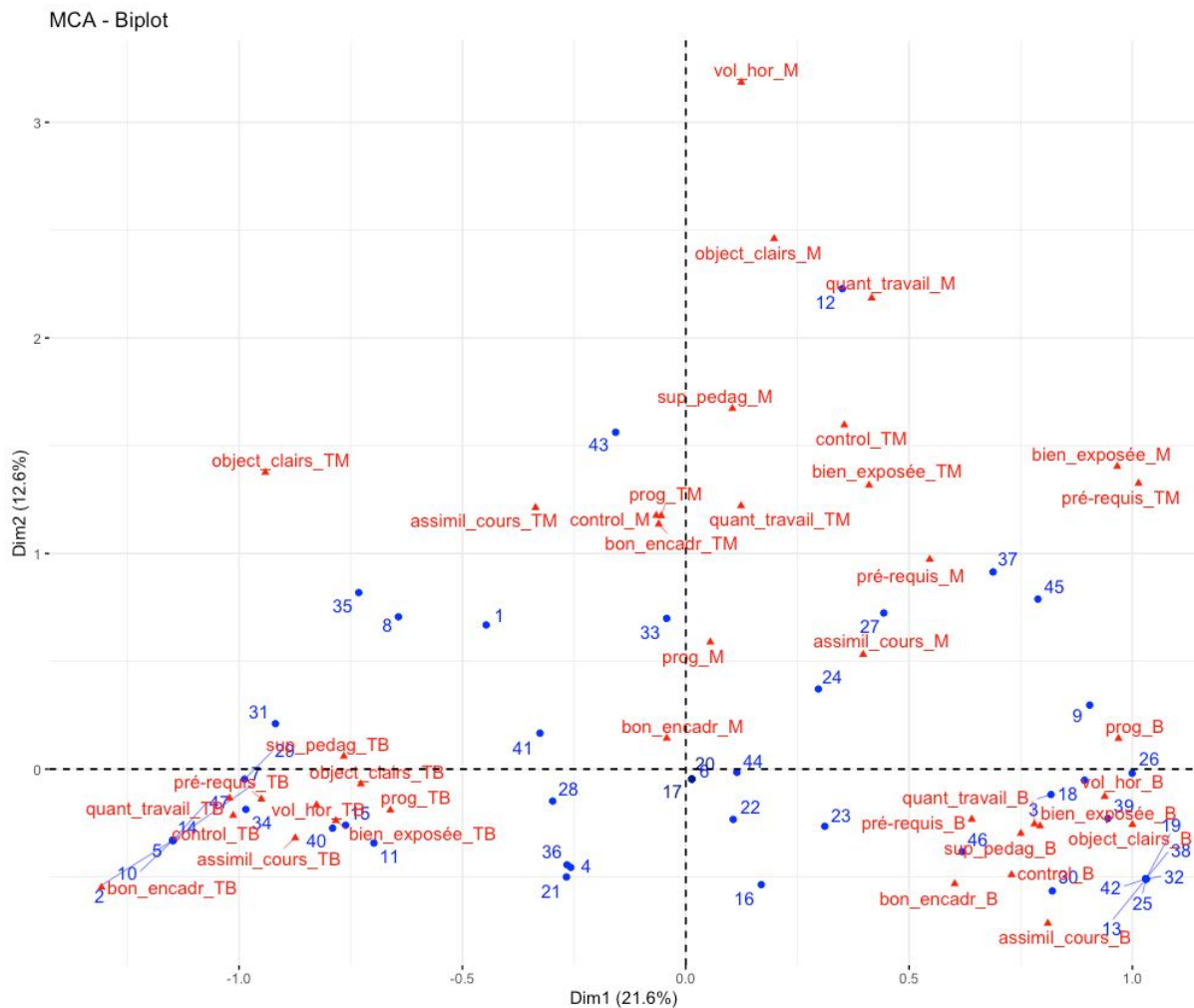
⇒ En appliquant la méthode du coude nous allons prendre les deux plus grandes valeurs propres

Représentation du biplot individus-variables

```

35 fviz_mca_biplot(res.mca,
36                   repel = TRUE, # Avoid text overlapping (slow if many point)
37                   ggtheme = theme_minimal())
38 
```

on obtient le biplot suivant :



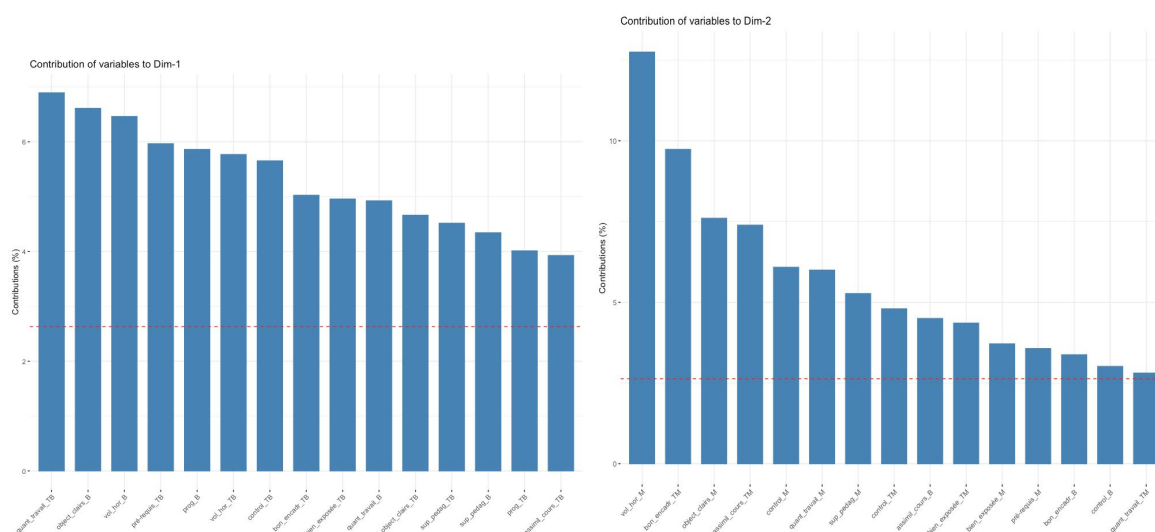
Etude du tableau des contributions et donner une signification aux axes.

a) contributions des variables

⇒ script à exécuter

```
75 #***** contributions des variables *****
76 # Contributions of rows to dimension 1
77 fviz_contrib(res.mca, choice = "var", axes = 1, top = 15)
78 # Contributions of rows to dimension 2
79 fviz_contrib(res.mca, choice = "var", axes = 2, top = 15)
80 #contribution of rows to plot 1-2
81 fviz_contrib(res.mca, choice = "var", axes = 1:2, top = 15)
```

⇒ résultats obtenus



Signification Axe1

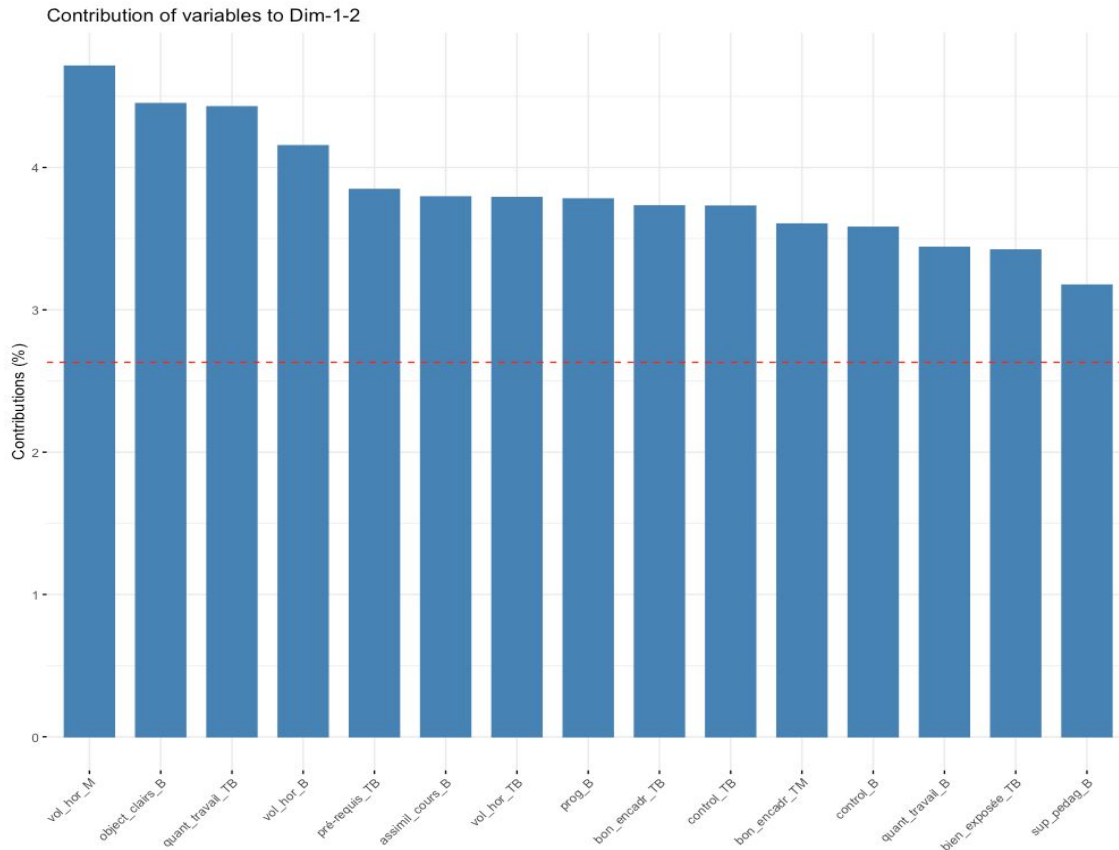
+	-
pre_requis_TB , quant_travail_TB	quant_travail_TB, bon_encadr_TB, control_TB, bien_expose_TB, vol_hor_B, assimil_cours_TB , prog_TB,

→ L'axe 1 est un axe d'opposition qui oppose les modalités TB(Très Bien) de chaque question avec les modalités B(Bien)

Signification Axe2

+	-
bon_encadr_TM,object_clairs_M,..	bon_encadr_B ...

→ l'axe 2 est un axe d'opposition lui aussi , il oppose les modalités TB(Très Bien) , B(Bien) d'un coté et les modalités TM (Très Mauvais) , M (Mauvais) de l'autre côté.

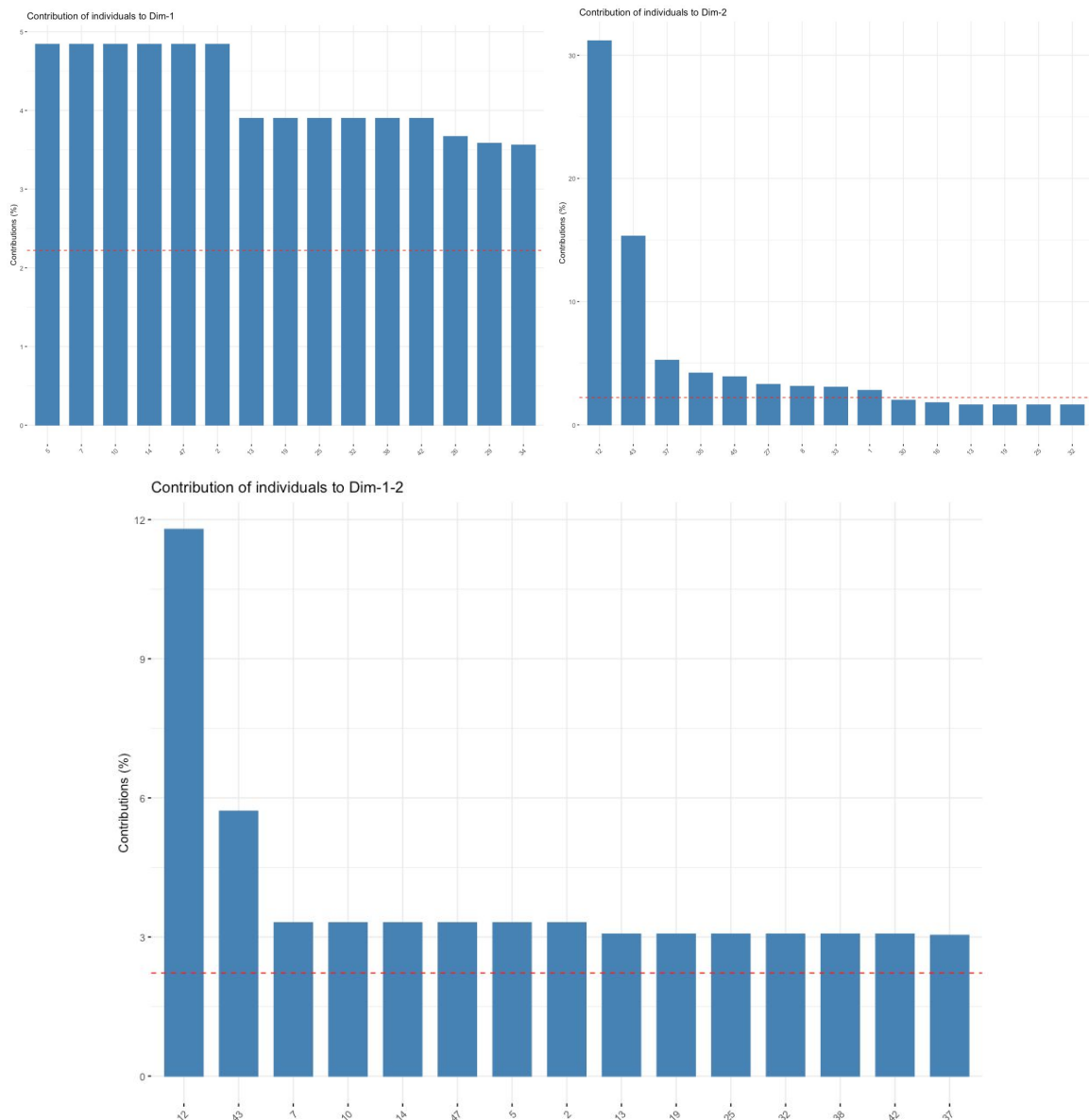


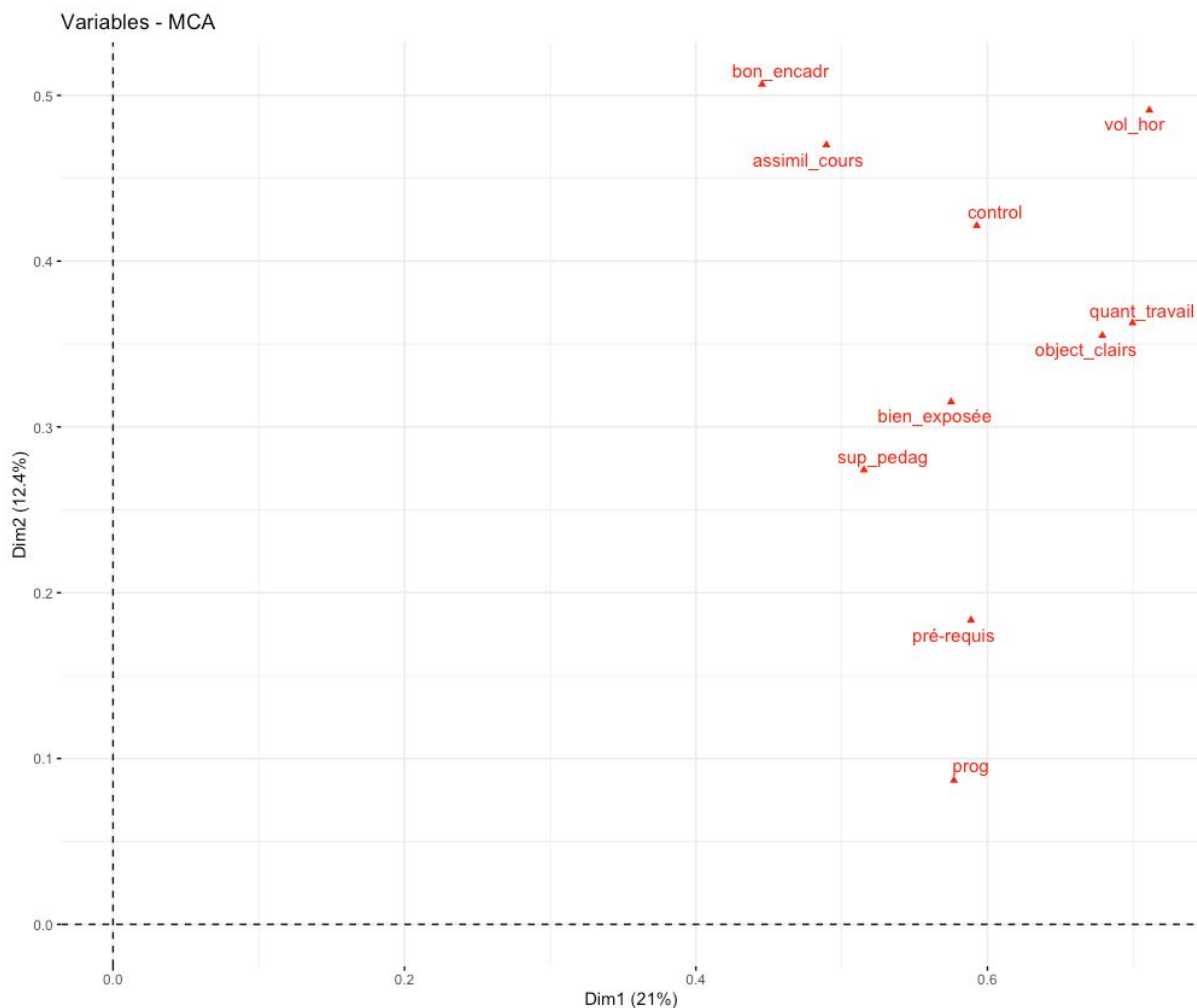
b) contributions des individus

⇒ script à exécuter

```
83 ***** contributions des individus *****
84 # Contributions of rows to dimension 1
85 fviz_contrib(res.mca, choice = "ind", axes = 1, top = 15)
86 # Contributions of rows to dimension 2
87 fviz_contrib(res.mca, choice = "ind", axes = 2, top = 15)
88 #contribution of rows to plot 1-2
89 fviz_contrib(res.mca, choice = "ind", axes = 1:2, top = 15)
```

⇒ résultats obtenus





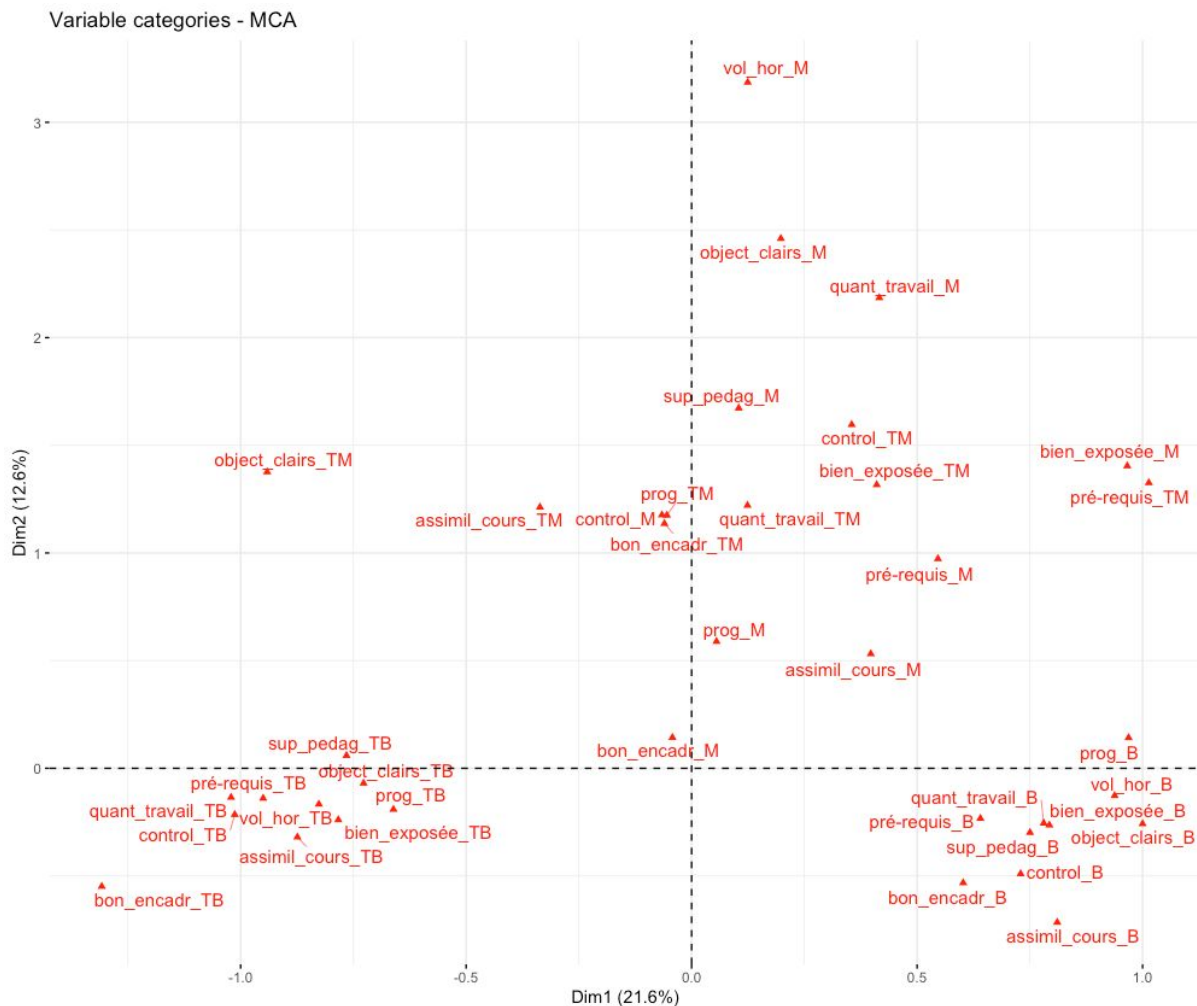
→ On remarque que la corrélation entre la quantité du travail et les objectifs clairs est élevée, ceci sous-entend que plus les objectifs sont clairs les étudiants fournissent plus d'effort.
→ la variable bien_exposé est corrélée avec support bien fourni ce qui est logique car un cours bien exposé est souvent accompagné d'un support de qualité.

b) Coordonnées des modalités des variables

⇒ script à exécuter

```
41 "correlation entre les variables"
42 fviz_mca_var(res.mca, choice = "mca.cor",
43             repel = TRUE, # Avoid text overlapping (slow)
44             ggtheme = theme_minimal())
```

⇒ résultats obtenus



→ on remarque que les modalités B (Bien) et TB (Très Bien) sont partitionnées est regroupés dans deux cotés.

c) Qualité de représentation des variables (modalités)

⇒ Script à exécuter

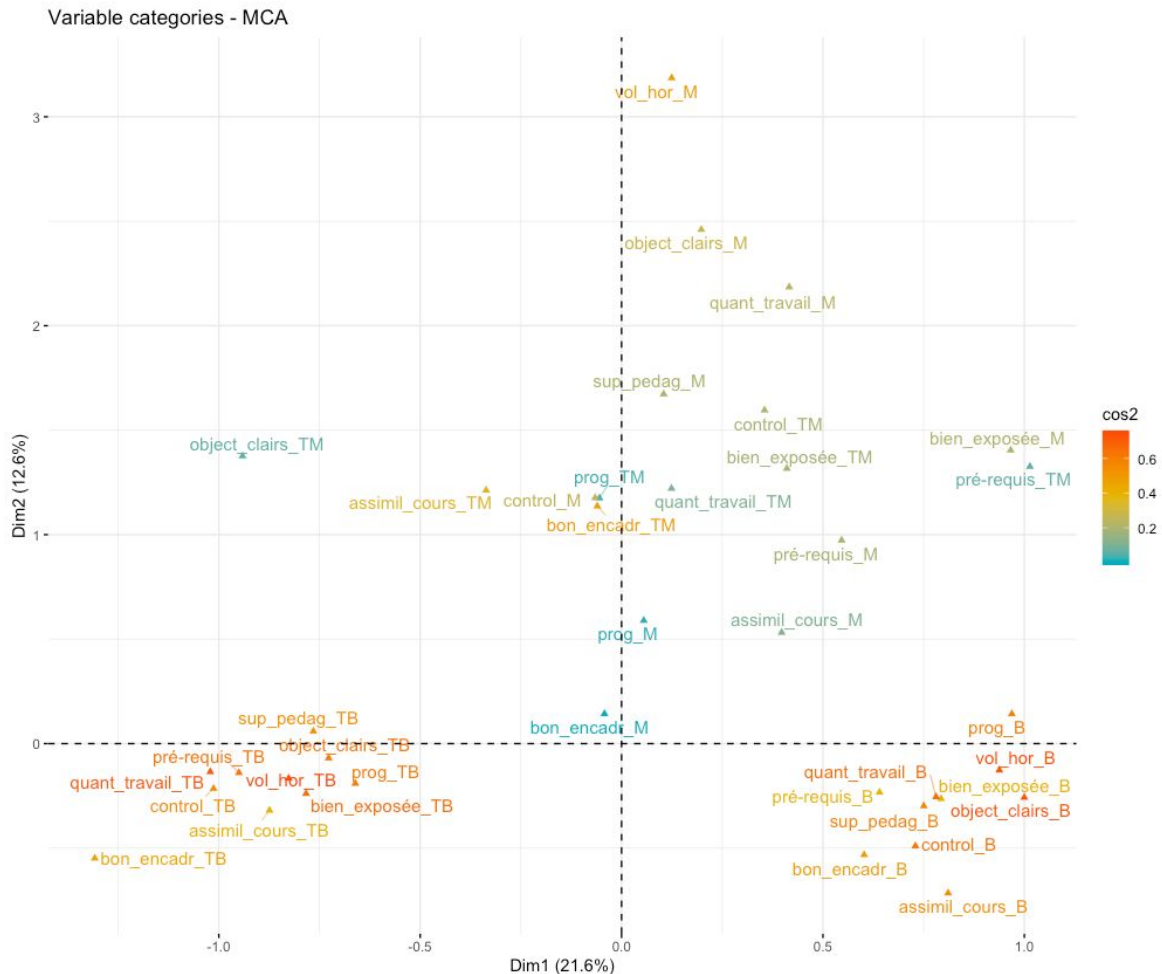
```
54 "qualité de représentation"
55 var <- get_mca_var(res.mca)
56 var$cos2
57
58 "visualiser les variables colorés selon les qualités de représentations"
59 fviz_mca_var(res.mca, col.var = "cos2",
60               gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
61               repel = TRUE, # Avoid text overlapping
62               ggtheme = theme_minimal())
```

⇒ Résultats obtenu

Voici un extrait de la table des qualités de représentations des modalités

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
object_clairs_B	6.667312e-01	0.044245865	3.618577e-03	7.691081e-03	0.0005887958
object_clairs_M	1.821535e-03	0.281529873	4.659565e-01	7.483318e-04	0.0383032207
object_clairs_TB	6.045444e-01	0.005465970	1.136205e-02	9.516235e-03	0.0455284993
object_clairs_TM	2.013664e-02	0.043058428	1.549542e-01	2.014224e-07	0.2800359998
prog_B	5.701040e-01	0.012269365	8.295998e-04	3.273220e-02	0.0400087742
prog_M	1.413644e-04	0.016140004	8.267387e-02	3.891712e-05	0.4912779484
prog_TB	5.464006e-01	0.045501007	9.067363e-03	2.053248e-03	0.0409912797
prog_TM	6.821227e-05	0.031374814	3.946891e-01	2.034349e-01	0.1299488795
pré-requis_B	3.280888e-01	0.043411194	4.201027e-02	1.464885e-01	0.0982695225
pré-requis_M	4.590955e-02	0.145678912	2.711567e-02	1.157927e-01	0.0982100911
pré-requis_TB	6.016293e-01	0.012917027	5.451269e-02	9.284417e-03	0.0081829477
pré-requis_TM	2.336170e-02	0.039954093	8.685763e-02	3.431436e-02	0.0011153769
bien_exposée_B	3.473035e-01	0.038409838	4.503316e-02	7.037431e-02	0.0114899869
bien_exposée_M	6.666925e-02	0.140828723	1.005946e-01	3.493458e-01	0.0009018016
bien_exposée_TB	5.870186e-01	0.054642921	1.179346e-02	1.203560e-03	0.0116655561
bien_exposée_TM	1.643654e-02	0.169185031	1.973104e-01	1.763481e-02	0.0002857973

Le graphe suivant représente les modalités projetées sur le nouveau plan factoriel colorés selon les qualités de représentation de chacune d'elle



- on remarque que les modalités B et TB des questions sont très bien représentés, contrairement ai modalités M et TM de la plupart des questions.
- ainsi on constate que les questions les mieux représentés sont:
 - quantité du travail
 - et la conformité des contrôles au programme
 - objectifs clairs

- **Les variables les mieux représentées:**

⇒ script à exécuter

```
66 # visualiser la qualité de representation des variables
67 View(res.mca$var$eta2)
```

⇒ résultats obtenus

Le tableau suivant illustre les qualités de représentation des variables selon les axes. les donnés sont ordonnées par ordre décroissant, donc la variable "objectif_clair" est la meilleure en terme de qualité de représentation.

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
object_clairs	0.6788203	0.35512068	0.56240939	0.02149555	0.348819423
prog	0.5770111	0.08687337	0.47492283	0.25134359	0.557865639
pré-requis	0.5887267	0.18346010	0.14878095	0.09187835	0.261131831
bien_exposée	0.5750680	0.31512563	0.30827995	0.41676441	0.038283907
bon_encadr	0.4453231	0.50666821	0.36186435	0.09703203	0.302272318
assimil_cours	0.4895135	0.47009924	0.32928622	0.51071108	0.006906022
sup_pedag	0.5152712	0.27407452	0.08576230	0.38618989	0.063791291
vol_hor	0.7110938	0.49109166	0.03638774	0.07181819	0.098827502
control	0.5926850	0.42145206	0.24545942	0.12134787	0.191570239
quant_travail	0.6995796	0.36273696	0.13226065	0.28366962	0.215105164

d) Qualité de représentation des individus

⇒ script à exécuter

```

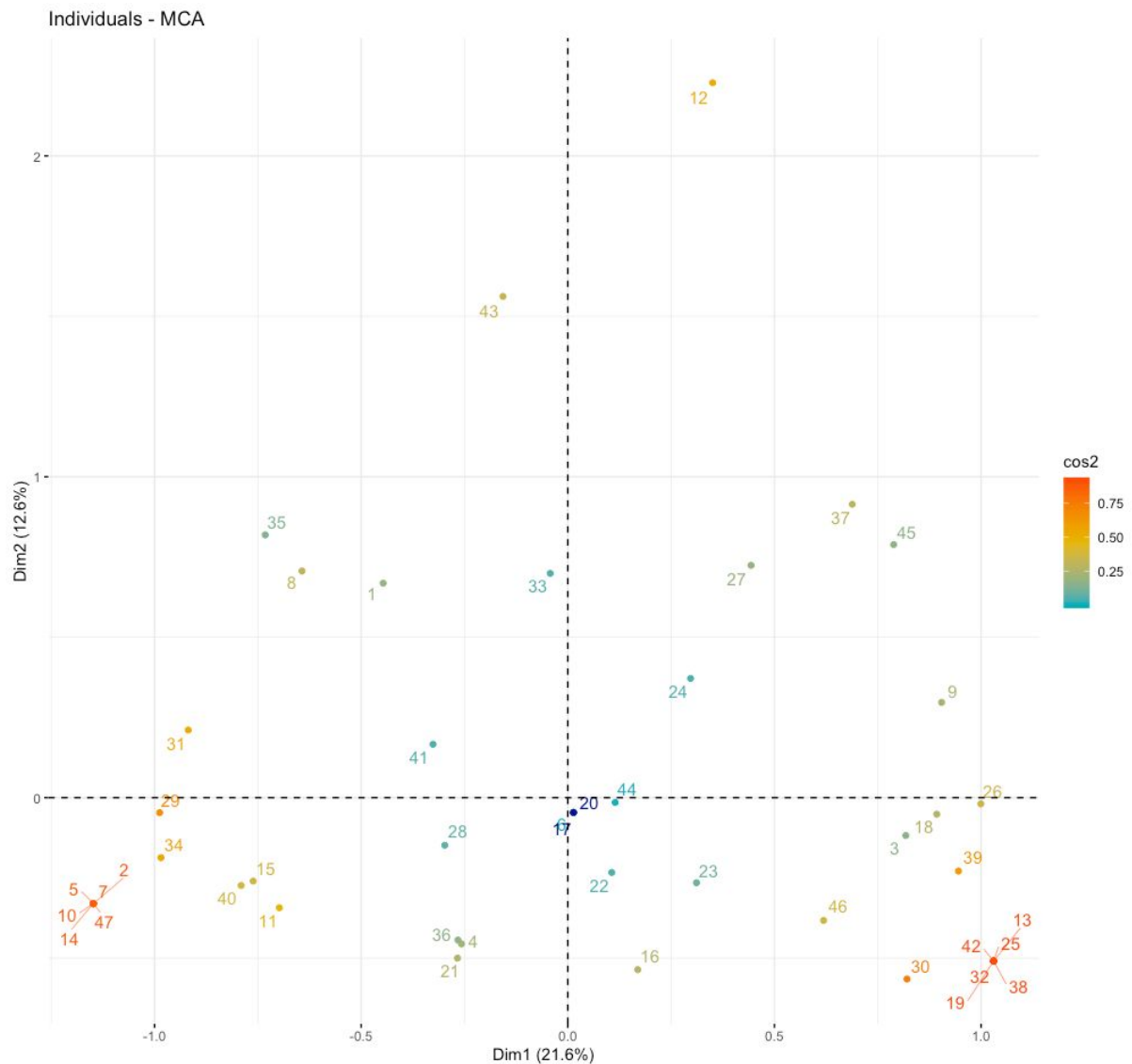
64 # ***** qualité de representation des individus *****
65 var1 <- get_mca_ind(res.mca)
66 var1$cos2
67
68 "visualiser les variables colorés selon les qualités de representations"
69 fviz_mca_ind(res.mca, col.ind = "cos2",
70             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
71             repel = TRUE, # Avoid text overlapping
72             ggtheme = theme_minimal())

```

⇒ résultats obtenus

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
1	0.0588999967	0.1318538612	1.336350e-01	3.944528e-02	2.004296e-02
2	0.7912089392	0.0653891831	1.594097e-02	1.092134e-02	9.416758e-03
3	0.1496088113	0.0031114554	1.845778e-04	1.525915e-04	4.185380e-01
4	0.0534083861	0.1675126125	6.859549e-04	6.373734e-03	3.233732e-02
5	0.7912089392	0.0653891831	1.594097e-02	1.092134e-02	9.416758e-03
6	0.0001058743	0.0011955724	3.238175e-02	7.235805e-02	1.712034e-02
7	0.7912089392	0.0653891831	1.594097e-02	1.092134e-02	9.416758e-03
8	0.1387147637	0.1672194958	2.622707e-01	5.642974e-03	1.121787e-03
9	0.2106445445	0.0227068165	5.889034e-02	2.125752e-01	7.468460e-02
10	0.7912089392	0.0653891831	1.594097e-02	1.092134e-02	9.416758e-03
11	0.3704566941	0.0894495536	3.546128e-02	7.698785e-04	3.980602e-02
12	0.0122866398	0.4966184819	1.262045e-01	2.387490e-01	6.953890e-03
13	0.7332010120	0.1791192606	1.240546e-05	1.261305e-02	2.287982e-03
14	0.7912089392	0.0653891831	1.594097e-02	1.092134e-02	9.416758e-03
15	0.3259331055	0.0380124477	2.722738e-02	1.588847e-04	3.930332e-02
16	0.0239521208	0.2405003474	1.394105e-02	8.074866e-03	3.017261e-02

Le graphe suivant représente les individus projetées sur le nouveau plan factoriel colorés selon les qualités de représentation de chacune d'elle



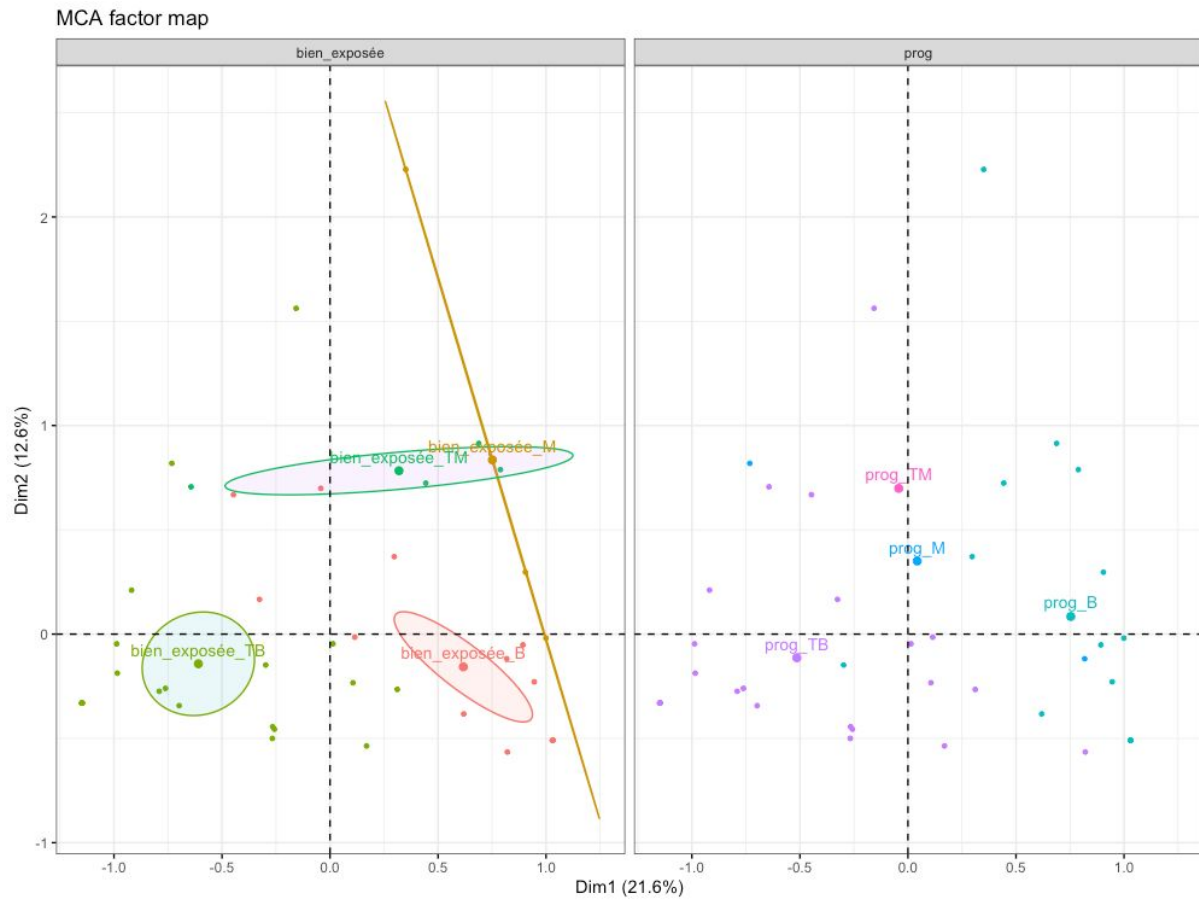
e) représentations avec groupement des points en ellipses

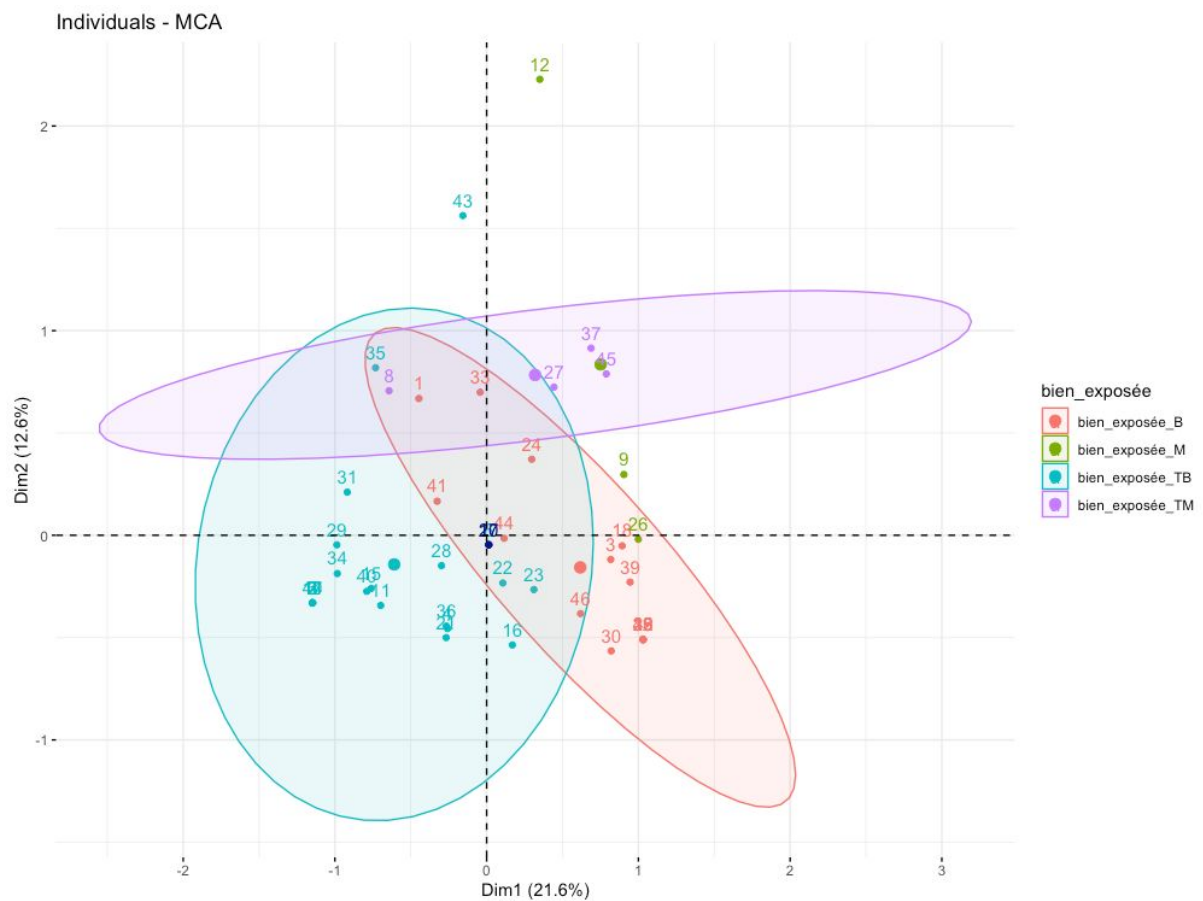
⇒ script exécuté

```
104 # ***** representarions avec ellipse *****
105 #une seule variable
106 fviz_mca_ind(res.mca, habillage = 4, addEllipses = TRUE)
107
108 #deux variables
109 fviz_ellipses(res.mca, c("prog", "bien_exposée"), geom = "point")
```

⇒ résultats obtenus

dans cette section, nous avons représenté les deux variables “bien_exposé” et “prog” l’une à côté de l’autre





La figure précédente représentation des tendances regroupés dans des ellipses illustrant les régions pour chacune des modalités d'une question donnée ("contenu bien exposé" dans notre cas à titre d'exemple)

Partie 2 :AFC

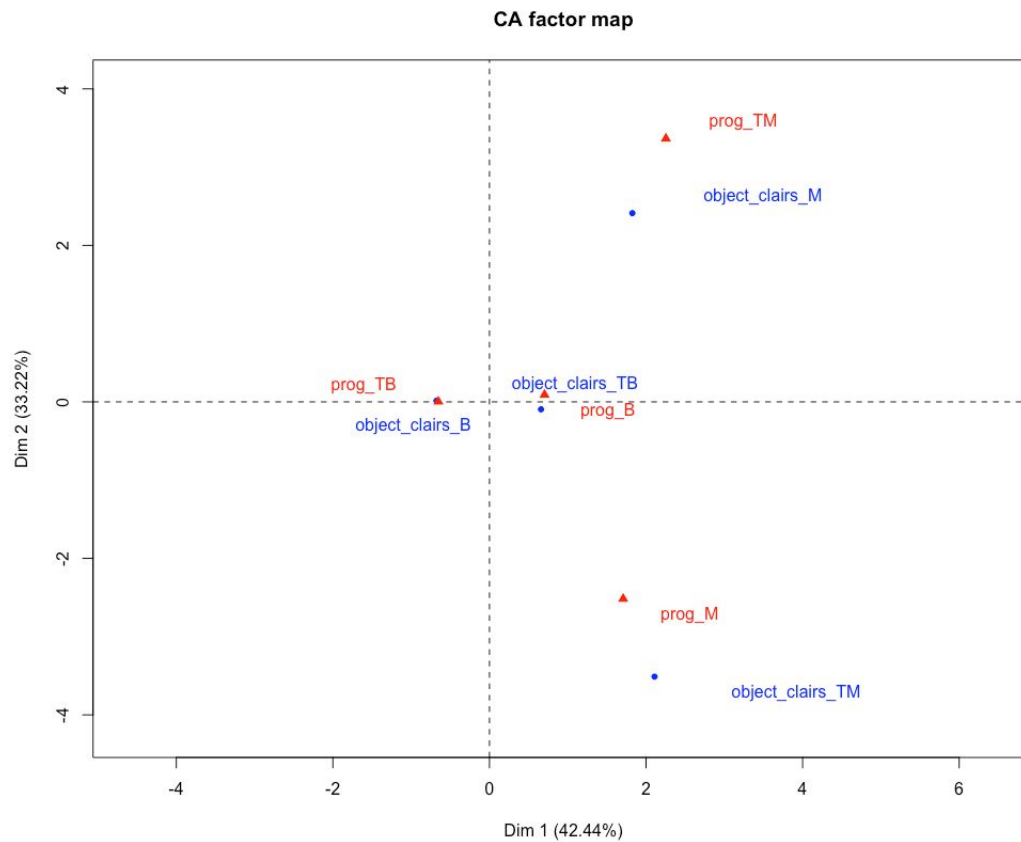
Croiser 2 questions pour en faire un tableau de contingence.

Nous allons croiser les questions 'objectifs clairs' et 'prog' :

⇒ script à exécuter

```
121 #croiser deux questions
122 x <- disjonctif[,c(1:8)]
123
124 #renommer les colonnes pour une meilleure representation
125 x<- matrix(data = x, nrow = 47, ncol = 8, dimnames = list(NULL,c(
126 'object_clairs_B','object_clairs_M','object_clairs_TB',
127 'object_clairs_TM','prog_B','prog_M','prog_TB','prog_TM'))))
128 |
129 burt = t(x)%*%x
130 burt.useful = burt[1:4 , c(5:8)]
131 burt.ca =CA(burt.useful)
```

Résultats de l'AFC



Interprétation

⇒ script à exécuter

```
134 # cotributions & comparaison avec le poids des lignes
135 View(burt.ca$row$contrib)
136 View(burt.ca$call$marge.row *100)
137
138 # cotributions & comparaison avec le poids des colonne
139 View(burt.ca$col$contrib)
140 View(burt.ca$call$marge.col *100)
```

⇒ résultats

- lignes

	Dim 1	Dim 2	Dim 3		V1
object_clairs_B	25.29762	0.67500609	35.729502	object_clairs_B	38.297872
object_clairs_M	21.60789	48.19965751	25.937130	object_clairs_M	4.255319
object_clairs_TB	38.67475	0.02559689	5.980509	object_clairs_TB	55.319149
object_clairs_TM	14.41974	51.09973951	32.352859	object_clairs_TM	2.127660

- colonnes

	Dim 1	Dim 2	Dim 3		V1
prog_B	27.28430	0.585464937	35.960018	prog_B	36.170213
prog_M	18.92150	52.484029045	24.339150	prog_M	4.255319
prog_TB	37.32720	0.002135845	5.223856	prog_TB	57.446809
prog_TM	16.46699	46.928370173	34.476977	prog_TM	2.127660

Axe 1:

+	-
object_clair_TM , prog_M, prog_TM	

⇒ l'axe 1 mesure la médiocrité de la variable prog selon l'avis des étudiants

Axe 2:

+	-
object_clair_M, prog_M,	object_clair_TM , prog_TM

⇒ l'axe 2 oppose la modalité prog_M (qui à été choisi par les mêmes personnes qui ont choisi la modalité avec prog_TM