

# ClustInvest. Classification non supervisée avec données binaires. Application à des données d'accidents de décompression

**Gérard GRÉGOIRE(\*) / Jean-Pierre IMBERT**  
**Laboratoire LJK UGA/ Société Divetech**

**Rencontres R**  
**Grenoble 24-26 Juin 2015**

# Plan de l'exposé

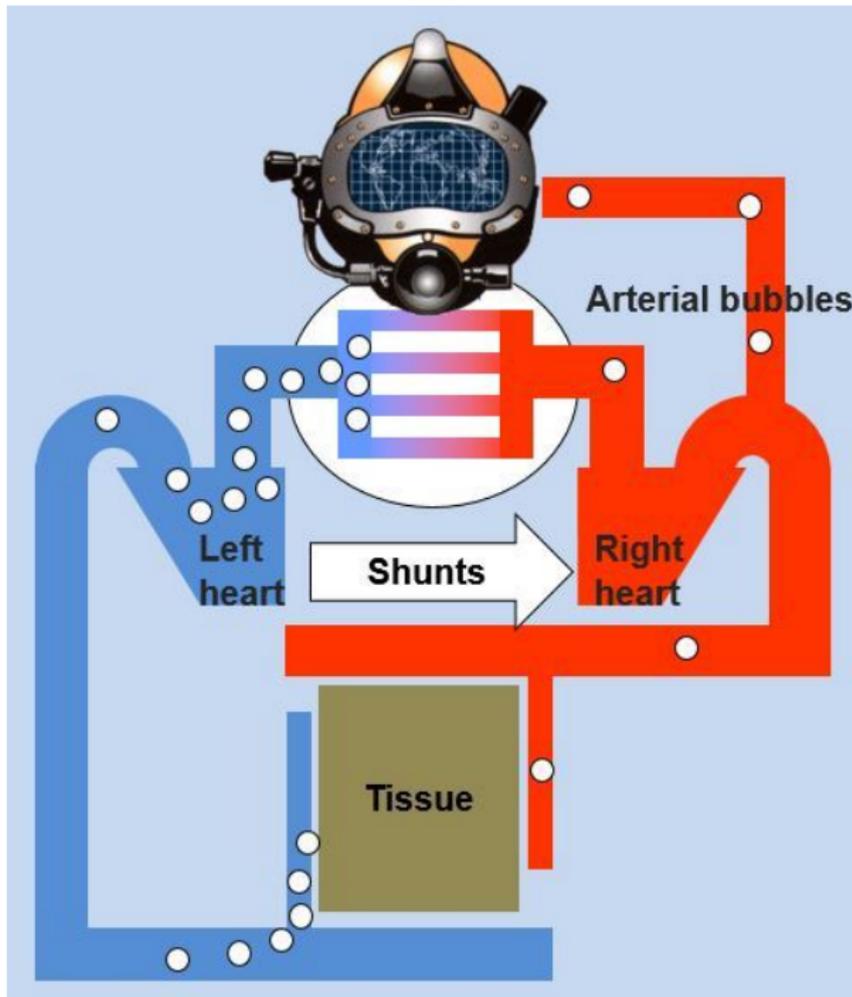
- 1 Introduction
- 2 Fonctions de ClustInvest
- 3 Exemple de mise en oeuvre
  - Informations de base
  - Analyse d'une classification
  - Calcul d'entropie
- 4 Application

## Le problème des accidents de décompression



## Les accidents de décompression en plongée professionnelle:

- Des procédures de décompression de plus en plus sûres (adaptées en fonction de la profondeur de travail, du temps passé au niveau de travail, du mélange gazeux utilisé...)
- Une fréquence d'accidents qui va en diminuant
- On cherche cependant toujours à approfondir la connaissance des mécanismes physiologiques impliqués de manière à améliorer les traitements et la prévention



Les données:

- 785 enregistrements d'accidents (1 accident = 1 individu statistique)
- On observe 17 symptômes pour lesquels on relève la présence ou l'absence
- A chaque symptôme est associée une variable binaire  $X_i$ , ( $X_i = 1$ ) si le symptôme est présent, ( $X_i = 0$ ) sinon.
- Un individu présente au plus 5 symptômes.

Les variables (symptômes) :

- **Joint** : douleurs articulaires ou dans les tissus musculaires environnant
- **Dizz\_Vert\_Bal**: étourdissement, vertige, troubles de l'équilibre
- **Nausea**: présence de nausées, vomissements
- **Paresth**: fourmillements, picotements, sensations d'aiguilles ou d'épingles, parésie, hypoésie (perte ou altération de la sensibilité dans une région du corps)
- **Fatigue**: sensation de fatigue, maux de tête
- **Back**: douleurs dans le dos (cervicales, dorsales, lombaires)
- **Shock**: pâleur, faiblesse, anxiété
- **Eye\_mov**: mouvements des yeux (nystagmus)
- **Vision**: problèmes de vision
- **Diff.StandUp**: faiblesse musculaire, difficulté à se tenir debout, paralysie
- **Skin\_marmor**: rougeurs, tâches marmoréennes sur la peau
- **Dyspnea**: difficultés respiratoires, sensation d'étouffement
- **Hearing\_pb**: présence d'acouphènes, bourdonnements
- **Speech**: problèmes d'élocution, confusion, changements de personnalité
- **Loss\_consc**: perte de conscience
- **Skin\_rash**: irritations de la peau, boursouffures, éruptions, peau squameuse
- **Emphys**: présence d'emphysème

Le fichier se présente comme un data\_frame de dimension  $785 \times 17$ , dont toutes les valeurs sont en 0/1, avec optionnellement des vecteurs apportant des informations complémentaires. L'étude statistique porte sur la matrice  $X$  de dimension  $785 \times 17$ .

Un extrait du fichier :

	Joint	Dizz_Vert_Bal	Nausea	Paresth	Fatigue	Back	Shock	Eye_mov	Vision	Diff.StandUp
8	0	0	0	1	0	1	0	0	0	1
13	0	0	0	0	0	0	1	0	1	0
14	1	0	0	0	0	1	0	0	0	0
24	0	1	1	0	1	0	0	0	0	0
25	0	0	0	1	0	0	0	0	1	0
29	0	1	1	0	0	0	1	0	0	0

	Skin_marmor	Dyspnea	Hearing_pb	Speech	Loss_consc	Skin_rash	Emphys
8	0	0	0	0	0	0	0
13	0	0	0	1	1	0	0
14	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0

>

La problématique.

Se basant sur les outils de classification non supervisée, on cherche à détecter des groupes d'individus dont les caractéristiques communes pourraient suggérer une interprétation physiologique des symptômes exprimés.

On souhaite aussi confronter les partitions proposées par les méthodes de classification à celles fournies par les "experts", médecins spécialistes du domaine de la plongée.

La classification sur données binaires ne peut être basée sur des méthodes utilisant la distance euclidienne (en particulier, le plus souvent les 0 et les 1 ne jouent pas un rôle symétrique).

Des méthodes où on substitue à la distance euclidienne des dissimilarités entre individus sont proposées dans la littérature. Elles méritent qu'on examine de manière approfondie les partitions obtenues.

La problématique.

Se basant sur les outils de classification non supervisée, on cherche à détecter des groupes d'individus dont les caractéristiques communes pourraient suggérer une interprétation physiologique des symptômes exprimés.

On souhaite aussi confronter les partitions proposées par les méthodes de classification à celles fournies par les " experts", médecins spécialistes du domaine de la plongée.

La classification sur données binaires ne peut être basée sur des méthodes utilisant la distance euclidienne (en particulier, le plus souvent les 0 et les 1 ne jouent pas un rôle symétrique).

Des méthodes où on substitue à la distance euclidienne des dissimilarités entre individus sont proposées dans la littérature. Elles méritent qu'on examine de manière approfondie les partitions obtenues.

- **Analyse du fichier de données**

- a) *Informations de base*

- Nombre d'individus, nombre de variables

- Nombre total de symptômes exprimés

- Nombre maximal de symptômes pour un individu

- Nombre total de patterns différents

- Vecteur du nombre de manifestations de chaque symptôme

- Entropie de l'ensemble des variables

- b) *Multiplicités*

- Tables de multiplicités

- Tables des couples

- Table des triples

- **Analyse d'une classification**

- a) Table des effectifs, tables des pourcentages

- b) Exploration des groupes

- c) Calcul de l'entropie conditionnelle. Calcul de l'indice entropique de la classification. Calcul d'indices d'entropie pour les variables.

- **Application**

- Comparaison de résultats de méthodes différentes.

- Retour sur le problème des accidents de décompression en plongée professionnelle.

- **Analyse du fichier de données**

- a) *Informations de base*

- Nombre d'individus, nombre de variables

- Nombre total de symptômes exprimés

- Nombre maximal de symptômes pour un individu

- Nombre total de patterns différents

- Vecteur du nombre de manifestations de chaque symptôme

- Entropie de l'ensemble des variables

- b) *Multiplicités*

- Tables de multiplicités

- Tables des couples

- Table des triples

- **Analyse d'une classification**

- a) Table des effectifs, tables des pourcentages

- b) Exploration des groupes

- c) Calcul de l'entropie conditionnelle. Calcul de l'indice entropique de la classification. Calcul d'indices d'entropie pour les variables.

- **Application**

- Comparaison de résultats de méthodes différentes.

- Retour sur le problème des accidents de décompression en plongée professionnelle.

- **Analyse du fichier de données**

- a) *Informations de base*

- Nombre d'individus, nombre de variables

- Nombre total de symptômes exprimés

- Nombre maximal de symptômes pour un individu

- Nombre total de patterns différents

- Vecteur du nombre de manifestations de chaque symptôme

- Entropie de l'ensemble des variables

- b) *Multiplicités*

- Tables de multiplicités

- Tables des couples

- Table des triples

- **Analyse d'une classification**

- a) Table des effectifs, tables des pourcentages

- b) Exploration des groupes

- c) Calcul de l'entropie conditionnelle. Calcul de l'indice entropique de la classification. Calcul d'indices d'entropie pour les variables.

- **Application**

- Comparaison de résultats de méthodes différentes.

- Retour sur le problème des accidents de décompression en plongée professionnelle.

- **Analyse du fichier de données**

- a) *Informations de base*

- Nombre d'individus, nombre de variables

- Nombre total de symptômes exprimés

- Nombre maximal de symptômes pour un individu

- Nombre total de patterns différents

- Vecteur du nombre de manifestations de chaque symptôme

- Entropie de l'ensemble des variables

- b) *Multiplicités*

- Tables de multiplicités

- Tables des couples

- Table des triples

- **Analyse d'une classification**

- a) Table des effectifs, tables des pourcentages

- b) Exploration des groupes

- c) Calcul de l'entropie conditionnelle. Calcul de l'indice entropique de la classification. Calcul d'indices d'entropie pour les variables.

- **Application**

- Comparaison de résultats de méthodes différentes.

- Retour sur le problème des accidents de décompression en plongée professionnelle.

### Informations de base

```
> print(fun_base_nb(data_classif_x)$nb_indiv)
[1] 785
> print(fun_base_nb(data_classif_x)$nb_var)
[1] 17
> print(fun_base_nb(data_classif_x)$nb_tot_sympt)
[1] 926
> print(fun_base_nb(data_classif_x)$nb_max_sympt)
[1] 5

> print(fun_patterns(data_classif_x)$nb_patterns)
[1] 64
> print(fun_Entrop(data_classif_x)$Entropie)
[1] 1.289778
```

### Informations de base

```
> print(fun_base_nb(data_classif_x)$nb_indiv)
[1] 785
> print(fun_base_nb(data_classif_x)$nb_var)
[1] 17
> print(fun_base_nb(data_classif_x)$nb_tot_sympt)
[1] 926
> print(fun_base_nb(data_classif_x)$nb_max_sympt)
[1] 5

> print(fun_patterns(data_classif_x)$nb_patterns)
[1] 64
> print(fun_Entrop(data_classif_x)$Entropie)
[1] 1.289778
```

### Informations de base

```
> print(fun_base_nb(data_classif_x)$nb_indiv)
[1] 785
> print(fun_base_nb(data_classif_x)$nb_var)
[1] 17
> print(fun_base_nb(data_classif_x)$nb_tot_sympt)
[1] 926
> print(fun_base_nb(data_classif_x)$nb_max_sympt)
[1] 5

> print(fun_patterns(data_classif_x)$nb_patterns)
[1] 64
> print(fun_Entrop(data_classif_x)$Entropie)
[1] 1.289778
```

```

> print( fun_nb_by_sympt(data_classif_x)$nb)
  Joint Dizz_Vert_Bal Nausea Paresth Fatigue Back Shock Eye_mov
    712           36     36     35     17  16   13     12

Vision Diff.StandUp Skin_marmor Dyspnea Hearing_pb Speech
    12           8     8     6     6     4

Loss_consc Skin_rash Emphys
    2         2     1

> print(fun_mult(data_classif_x)$vec_mult)

Mult_order   1    2    3    4    5
Obs_Nbr     700  46  26   9   4

```

```

> print( fun_nb_by_sympt(data_classif_x)$nb)
  Joint Dizz_Vert_Bal Nausea Paresth Fatigue Back Shock Eye_mov
    712           36     36     35     17  16   13     12

Vision Diff.StandUp Skin_marmor Dyspnea Hearing_pb Speech
    12           8     8     6     6     4

Loss_consc Skin_rash Emphys
    2         2     1

> print(fun_mult(data_classif_x)$vec_mult)

Mult_order   1    2    3    4    5
Obs_Nbr      700  46  26   9   4

```

Occurrences multiples par symptôme.

```
> print(fun_var_multi(data_classif_x)$matr_mult)
```

	Joint	Dizz_Vert_Bal	Nausea	Paresth	Fatigue	Back	Shock	Eye_mov	Vision	Diff.StandUp	Skin_marmor	Dyspnea
1	680	1	2	7	1	1	1	1	1	0	2	0
2	23	12	11	18	6	6	2	3	3	1	4	1
3	5	16	17	7	3	7	6	6	4	5	0	0
4	2	5	5	2	4	0	3	2	3	0	1	3
5	2	2	1	1	3	2	1	0	1	2	1	2
6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0
18	712	36	36	35	17	16	13	12	12	8	8	6

	Hearing_pb	Speech	Loss_consc	Skin_rash	Emphys	tot_co_occur
1	0	0	0	2	1	700
2	2	0	0	0	0	46
3	1	1	0	0	0	26
4	3	2	1	0	0	9
5	0	1	1	0	0	4
6	0	0	0	0	0	0
7	0	0	0	0	0	0
8	0	0	0	0	0	0
9	0	0	0	0	0	0
10	0	0	0	0	0	0
11	0	0	0	0	0	0
12	0	0	0	0	0	0
13	0	0	0	0	0	0
14	0	0	0	0	0	0
15	0	0	0	0	0	0
16	0	0	0	0	0	0
17	0	0	0	0	0	0
18	6	4	2	2	1	NA

>

```
> print(fun_couples(x)$mat_couples)
```

	Joint	Dizz_Vert_Bal	Nausea	Paresth	Fatigue	Back	Shock	Eye_mov	Vision	Diff.StandUp
Joint	0	4	5	17	5	5	1	0	0	1
Dizz_Vert_Bal	4	0	24	4	5	2	5	10	2	2
Nausea	5	24	0	3	5	4	4	8	1	2
Paresth	17	4	3	0	2	4	5	0	3	2
Fatigue	5	5	5	2	0	2	3	1	4	2
Back	5	2	4	4	2	0	2	1	1	3
Shock	1	5	4	5	3	2	0	0	3	1
Eye_mov	0	10	8	0	1	1	0	0	0	0
Vision	0	2	1	3	4	1	3	0	0	4
Diff.StandUp	1	2	2	2	2	3	1	0	4	0
Skin_marmor	3	0	0	0	2	2	1	0	1	1
Dyspnea	4	3	3	1	2	2	0	0	0	0
Hearing_pb	0	4	4	0	1	0	0	1	1	0
Speech	1	2	0	1	1	0	1	0	3	1
Loss_consc	1	0	1	0	1	0	1	0	1	0
Skin_rash	0	0	0	0	0	0	0	0	0	0
Emphys	0	0	0	0	0	0	0	0	0	0

	Skin_marmor	Dyspnea	Hearing_pb	Speech	Loss_consc	Skin_rash	Emphys
Joint	3	4	0	1	1	0	0
Dizz_Vert_Bal	0	3	4	2	0	0	0
Nausea	0	3	4	0	1	0	0
Paresth	0	1	0	1	0	0	0
Fatigue	2	2	1	1	1	0	0
Back	2	2	0	0	0	0	0
Shock	1	0	0	1	1	0	0
Eye_mov	0	0	1	0	0	0	0
Vision	1	0	1	3	1	0	0
Diff.StandUp	1	0	0	1	0	0	0
Skin_marmor	0	0	1	0	0	0	0
Dyspnea	0	0	1	1	1	0	0
Hearing_pb	1	1	0	0	0	0	0
Speech	0	1	0	0	1	0	0
Loss_consc	0	1	0	1	0	0	0
Skin_rash	0	0	0	0	0	0	0
Emphys	0	0	0	0	0	0	0

```
> print(fun_couples_exact(data_classif_x)$mat_couples_exact)
```

	Joint	Dizz_Vert_Bal	Nausea	Paresth	Fatigue	Back	Shock	Eye_mov	Vision	Diff.StandUp	
Joint	0	0	0	15	2	3	0	0	0	0	0
Dizz_Vert_Bal	0	0	8	0	1	0	0	2	0	0	0
Nausea	0	8	0	0	0	1	0	1	0	0	0
Paresth	15	0	0	0	1	0	1	0	1	0	0
Fatigue	2	1	0	1	0	0	1	0	1	0	0
Back	3	0	1	0	0	0	0	0	0	0	0
Shock	0	0	0	1	1	0	0	0	0	0	0
Eye_mov	0	2	1	0	0	0	0	0	0	0	0
Vision	0	0	0	1	1	0	0	0	0	0	1
Diff.StandUp	0	0	0	0	0	0	0	0	1	0	0
Skin_marmor	3	0	0	0	0	1	0	0	0	0	0
Dyspnea	0	0	0	0	0	1	0	0	0	0	0
Hearing_pb	0	1	1	0	0	0	0	0	0	0	0
Speech	0	0	0	0	0	0	0	0	0	0	0
Loss_consc	0	0	0	0	0	0	0	0	0	0	0
Skin_rash	0	0	0	0	0	0	0	0	0	0	0
Emphys	0	0	0	0	0	0	0	0	0	0	0

	Skin_marmor	Dyspnea	Hearing_pb	Speech	Loss_consc	Skin_rash	Emphys
Joint	3	0	0	0	0	0	0
Dizz_Vert_Bal	0	0	1	0	0	0	0
Nausea	0	0	1	0	0	0	0
Paresth	0	0	0	0	0	0	0
Fatigue	0	0	0	0	0	0	0
Back	1	1	0	0	0	0	0
Shock	0	0	0	0	0	0	0
Eye_mov	0	0	0	0	0	0	0
Vision	0	0	0	0	0	0	0
Diff.StandUp	0	0	0	0	0	0	0
Skin_marmor	0	0	0	0	0	0	0
Dyspnea	0	0	0	0	0	0	0
Hearing_pb	0	0	0	0	0	0	0
Speech	0	0	0	0	0	0	0
Loss_consc	0	0	0	0	0	0	0
Skin_rash	0	0	0	0	0	0	0
Emphys	0	0	0	0	0	0	0

```
> print(fun_list_couples(x)$names_tab)
```

	sympt_i	sympt_j	freq
1	Back	Dyspnea	1
2	Back	Skin_marmor	1
3	Dizz_Vert_Bal	Eye_mov	2
4	Dizz_Vert_Bal	Fatigue	1
5	Dizz_Vert_Bal	Hearing_pb	1
6	Dizz_Vert_Bal	Nausea	8
7	Fatigue	Shock	1
8	Fatigue	Vision	1
9	Joint	Back	3
10	Joint	Fatigue	2
11	Joint	Paresth	15
12	Joint	Skin_marmor	3
13	Nausea	Back	1
14	Nausea	Eye_mov	1
15	Nausea	Hearing_pb	1
16	Paresth	Fatigue	1
17	Paresth	Shock	1
18	Paresth	Vision	1
19	Vision	Diff.StandUp	1

```
> print(fun_list_triples(x)$names_tab)
```

	sympt_i	sympt_j	sympt_k	freq
1	Back	Vision	Diff.StandUp	1
2	Dizz_Vert_Bal	Back	Eye_mov	1
3	Dizz_Vert_Bal	Nausea	Eye_mov	5
4	Dizz_Vert_Bal	Nausea	Fatigue	1
5	Dizz_Vert_Bal	Nausea	Hearing_pb	1
6	Dizz_Vert_Bal	Nausea	Paresth	1
7	Dizz_Vert_Bal	Nausea	Shock	3
8	Dizz_Vert_Bal	Paresth	Diff.StandUp	1
9	Dizz_Vert_Bal	Shock	Vision	1
10	Joint	Dizz_Vert_Bal	Fatigue	1
11	Joint	Dizz_Vert_Bal	Nausea	1
12	Joint	Nausea	Back	1
13	Joint	Nausea	Diff.StandUp	1
14	Joint	Paresth	Shock	1
15	Nausea	Fatigue	Back	1
16	Nausea	Paresth	Back	1
17	Nausea	Vision	Diff.StandUp	1
18	Paresth	Back	Diff.StandUp	1
19	Paresth	Back	Shock	1
20	Paresth	Vision	Speech	1

```

> cl<-kmeans(data_classif_x,nb_k,nstart=100) # Appel kmeans
> CLUSTER_IG<- cl$cluster
> print(fun_clusters(data_classif_x,vec_clust_of_indiv=CLUSTER_IG)

```

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	tot_in_pop
Clusters_size	32	694	26	33	785
lign1	-----	-----	-----	-----	-----
Joint	17	694	0	1	712
Dizz_Vert_Bal	2	2	3	29	36
Nausea	1	4	3	28	36
Paresth	32	0	1	2	35
Fatigue	1	5	9	2	17
Back	4	4	7	1	16
Shock	3	0	6	4	13
Eye_mov	0	0	1	11	12
Vision	2	0	10	0	12
Diff_StandUp	2	1	5	0	8
Skin_marmor	0	3	5	0	8
Dyspnea	1	3	1	1	6
Hearing_pb	0	0	1	5	6
Speech	1	1	2	0	4
Loss_consc	0	1	1	0	2
Skin_rash	0	0	2	0	2
Emphys	0	0	1	0	1
chain_white1					
Sympt_tot_nb	66	718	58	84	926
patterns_nb	12	10	24	18	64
mean_sympt_nbr_by_indiv	2.062	1.035	2.231	2.545	1.18

\$matr\_influence\_v

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	tot_in_pop	perc_by_nbindiv
Cluster_size	32	694	26	33	785	<NA>
lign	-----	-----	-----	-----	-----	-----
Joint	0.531	1	0	0.03	712	0.907
Dizz_Vert_Bal	0.062	0.003	0.115	0.879	36	0.0459
Nausea	0.031	0.006	0.115	0.848	36	0.0459
Paresth	1	0	0.038	0.061	35	0.0446
Fatigue	0.031	0.007	0.346	0.061	17	0.0217
Back	0.125	0.006	0.269	0.03	16	0.0204
Shock	0.094	0	0.231	0.121	13	0.0166
Eye_mov	0	0	0.038	0.333	12	0.0153
Vision	0.062	0	0.385	0	12	0.0153
Diff.StandUp	0.062	0.001	0.192	0	8	0.0102
Skin_marmor	0	0.004	0.192	0	8	0.0102
Dyspnea	0.031	0.004	0.038	0.03	6	0.0076
Hearing_pb	0	0	0.038	0.152	6	0.0076
Speech	0.031	0.001	0.077	0	4	0.0051
Loss_consc	0	0.001	0.038	0	2	0.0025
Skin_rash	0	0	0.077	0	2	0.0025
Emphys	0	0	0.038	0	1	0.0013

\$matr\_influence\_h

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	tot_in_pop	perc_by_nbdiv
Cluster_size	32	694	26	33	785	<NA>
lign	-----	-----	-----	-----	-----	-----
Joint	0.024	0.975	0	0.001	712	0.907
Dizz_Vert_Bal	0.056	0.056	0.083	0.806	36	0.0459
Nausea	0.028	0.111	0.083	0.778	36	0.0459
Paresth	0.914	0	0.029	0.057	35	0.0446
Fatigue	0.059	0.294	0.529	0.118	17	0.0217
Back	0.25	0.25	0.438	0.062	16	0.0204
Shock	0.231	0	0.462	0.308	13	0.0166
Eye_mov	0	0	0.083	0.917	12	0.0153
Vision	0.167	0	0.833	0	12	0.0153
Diff.StandUp	0.25	0.125	0.625	0	8	0.0102
Skin_marmor	0	0.375	0.625	0	8	0.0102
Dyspnea	0.167	0.5	0.167	0.167	6	0.0076
Hearing_pb	0	0	0.167	0.833	6	0.0076
Speech	0.25	0.25	0.5	0	4	0.0051
Loss_consc	0	0.5	0.5	0	2	0.0025
Skin_rash	0	0	1	0	2	0.0025
Emphys	0	0	1	0	1	0.0013

\$matr\_influence\_symp

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	tot_in_pop	perc_by_nbsympt
Cluster_size	32	694	26	33	785	<NA>
lign	-----	-----	-----	-----	-----	-----
Joint	0.258	0.967	0	0.012	712	0.7689
Dizz_Vert_Bal	0.03	0.003	0.052	0.345	36	0.0389
Nausea	0.015	0.006	0.052	0.333	36	0.0389
Paresth	0.485	0	0.017	0.024	35	0.0378
Fatigue	0.015	0.007	0.155	0.024	17	0.0184
Back	0.061	0.006	0.121	0.012	16	0.0173
Shock	0.045	0	0.103	0.048	13	0.014
Eye_mov	0	0	0.017	0.131	12	0.013
Vision	0.03	0	0.172	0	12	0.013
Diff_StandUp	0.03	0.001	0.086	0	8	0.0086
Skin_marmor	0	0.004	0.086	0	8	0.0086
Dyspnea	0.015	0.004	0.017	0.012	6	0.0065
Hearing_pb	0	0	0.017	0.06	6	0.0065
Speech	0.015	0.001	0.034	0	4	0.0043
Loss_consc	0	0.001	0.017	0	2	0.0022
Skin_rash	0	0	0.034	0	2	0.0022
Emphys	0	0	0.017	0	1	0.0011
chain_white						
Sympt_tot_nb	66	718	58	84	926	<NA>
mean_sympnt_nbr_by_indiv	2.062	1.035	2.231	2.545	1.18	<NA>

```
> data_classif_xJ<- data_classif_x[CLUSTER_IG==3,]

> print(fun_base_nb(data_classif_xJ)$nb_indiv)
[1] 26
> print(fun_base_nb(data_classif_xJ)$nb_var)
[1] 17
> print(fun_base_nb(data_classif_xJ)$nb_tot_sympt)
[1] 58
> print(fun_patterns(data_classif_xJ)$nb_patterns)
[1] 24
> print(fun_base_nb(data_classif_xJ)$nb_max_sympt)
[1] 5

> print(fun_Entrop(data_classif_xJ)$Entropie)
[1] 4.546594
```

## Exploration du groupe 3

```
> print( fun_nb_by_symp( data_classif_xJ )$nb)
```

Joint	Dizz_Vert_Bal	Nausea	Paresth	Fatigue	Back	Shock	Eye_mov
0	3	3	1	9	7	6	1
Vision	Diff.StandUp	Skin_marmor	Dyspnea	Hearing_pb	Speech		
10	5	5	1	1	2		
Loss_consc	Skin_rash	Emphys					
1	2	1					

```
> print( fun_mult( data_classif_xJ )$vec_mult)
```

Mult_order	1	2	3	4	5
Obs_Nbr	10	7	4	3	2

```
> print(fun_var_multi(data_classif_xJ)$matr_mult)
```

```
> print(fun_var_multi(data_classif_xJ)$matr_mult)
```

	Joint	Dizz_Vert_Bal	Nausea	Paresth	Fatigue	Back	Shock	Eye_mov	Vision	Diff.StandUp	Skin_marmor
1	0	0	0	0	0	1	1	1	1	0	2
2	0	0	1	1	0	3	3	1	0	2	1
3	0	0	1	2	0	1	2	1	0	3	0
4	0	0	0	0	1	2	0	2	0	3	1
5	0	0	1	0	0	2	1	1	0	1	1
6	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0
	0	3	3	1	9	7	6	1	10	5	5

	Dyspnea	Hearing_pb	Speech	Loss_consc	Skin_rash	Emphys	tot_co_occur
1	0	0	0	0	2	1	10
2	1	0	0	0	0	0	7
3	0	0	0	0	0	0	4
4	0	1	1	1	0	0	3
5	0	0	1	0	0	0	2
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0
	1	1	2	1	2	1	NA

```
> print(fun_clusters(data_classif_x, CLUSTER_IG)$groupes[[3]])
```

```
[[1]]
```

```
[1] "CLUSTER DESCRIPTION : 3"
```

```
[[2]]
```

```
[1] "Cluster size : 26"
```

```
$Joint_onsets
```

	vec_clust_of_indiv	V1	V2	V3	V4	V5
13	3	Shock	Vision	Speech	Loss_consc	
77	3	Nausea	Fatigue	Back		
128	3	Back	Dyspnea			
190	3	Skin_marmor				
200	3	Vision				
207	3	Back				
233	3	Vision	Diff.StandUp			
237	3	Fatigue				
262	3	Dizz_Vert_Bal	Shock	Vision		
279	3	Shock				
290	3	Dizz_Vert_Bal	Fatigue	Vision	Diff.StandUp	Speech

```
> data_classif_xJ<- data_classif_x[CLUSTER_IG==3,]  
> print(fun_list_couples(data_classif_xJ)$names_tab)
```

	sympt_i	sympt_j	freq
1	Back	Dyspnea	1
2	Back	Skin_marmor	1
3	Dizz_Vert_Bal	Fatigue	1
4	Fatigue	Shock	1
5	Fatigue	Vision	1
6	Nausea	Back	1
7	Vision	Diff.StandUp	1

```
> print(fun_list_triples(data_classif_xJ)$names_tab)
```

	sympt_i	sympt_j	sympt_k	freq
1	Back	Vision	Diff.StandUp	1
2	Dizz_Vert_Bal	Shock	Vision	1
3	Nausea	Fatigue	Back	1
4	Nausea	Vision	Diff.StandUp	1

```

$stable_des_pourcentages
      nbr perc_by_totsympt perc_by_ind_nbr perc_in_totv
Joint      0 0.00000000 0.00000000 0.00000000
Dizz_Vert_Bal 3 0.05172414 0.11538462 0.08333333
Nausea      3 0.05172414 0.11538462 0.08333333
Paresth     1 0.01724138 0.03846154 0.02857143
Fatigue     9 0.15517241 0.34615385 0.52941176
Back        7 0.12068966 0.26923077 0.43750000
Shock       6 0.10344828 0.23076923 0.46153846
Eye_mov     1 0.01724138 0.03846154 0.08333333
Vision     10 0.17241379 0.38461538 0.83333333
Diff.StandUp 5 0.08620690 0.19230769 0.62500000
Skin_marmor 5 0.08620690 0.19230769 0.62500000
Dyspnea     1 0.01724138 0.03846154 0.16666667
Hearing_pb  1 0.01724138 0.03846154 0.16666667
Speech      2 0.03448276 0.07692308 0.50000000
Loss_consc  1 0.01724138 0.03846154 0.50000000
Skin_rash   2 0.03448276 0.07692308 1.00000000
Emphys     1 0.01724138 0.03846154 1.00000000
lign_sum    58 1.00000000                NA                NA

```

```
[[6]]
```

```
[1] "Mean symptom number by individual = 2.231"
```

Entropie globale, entropie conditionnelle à la classification, entropie de chacune des classes, indice entropique de la classification

```
> print(fun_Entrop(data_classif_x)$Entropie)
```

```
[1] 1.289778
```

```
> print(fun_Entr_cond(data_classif_x,CLUSTER_IG)$Entrp_cd)
```

```
[1] 0.5894051
```

```
> print(fun_Entr_cond(data_classif_x,CLUSTER_IG)$Entrp_cl)
```

```
[1] 2.554536 0.202631 4.546594 3.700014
```

```
> print(fun_Entr_cond(data_classif_x,CLUSTER_IG)$Entrp_ind)
```

```
[1] 0.5430182
```

Entropie globale, entropie conditionnelle à la classification, entropie de chacune des classes, indice entropique de la classification

```
> print(fun_Entrop(data_classif_x)$Entropie)
[1] 1.289778
> print(fun_Entr_cond(data_classif_x,CLUSTER_IG)$Entrp_cd)
[1] 0.5894051

> print(fun_Entr_cond(data_classif_x,CLUSTER_IG)$Entrp_cl)
[1] 2.554536 0.202631 4.546594 3.700014
> print(fun_Entr_cond(data_classif_x,CLUSTER_IG)$Entrp_ind)
[1] 0.5430182
```

Entrpie des variables et indices.

```
> print(fun_tab_info_v(data_classif_x, CLUSTER_IG)$table_info)
```

	n0	Entrp_0	n1	Entrp_1	Entrp_cond	Indice0	Indice1
Joint	73	1.5211288	712	0.1779688	0.3028741	0.4893216	0.9544714
Dizz_Vert_Bal	749	0.4860564	36	1.0133601	0.5102385	0.8700413	0.7023674
Nausea	749	0.5017673	36	1.0765687	0.5281276	0.8655151	0.6794966
Paresth	750	0.4571494	35	0.5007116	0.4590917	0.8783082	0.8658200
Fatigue	768	0.6360208	17	1.6087002	0.6570852	0.8258266	0.4423233
Back	769	0.6371274	16	1.7717822	0.6602541	0.8254915	0.3378001
Shock	772	0.6303806	13	1.5262349	0.6452164	0.8275323	0.4867058
Eye_mov	773	0.6360769	12	0.4138169	0.6326793	0.8258096	0.8905569
Vision	773	0.6315907	12	0.6500224	0.6318725	0.8271666	0.8215770
Diff.StandUp	777	0.6628371	8	1.2987949	0.6693182	0.8176683	0.5921170
Skin_marmor	777	0.6743704	8	0.9544340	0.6772245	0.8141344	0.7230373
Dyspnea	779	0.6869135	6	1.7924813	0.6953637	0.8102736	0.3221170
Hearing_pb	779	0.6693770	6	0.6500224	0.6692290	0.8156663	0.8215770
Speech	781	0.6850747	4	1.5000000	0.6892271	0.8108407	0.5000000
Loss_consc	783	0.6956239	2	1.0000000	0.6963994	0.8075816	0.7071068
Skin_rash	783	0.6894641	2	0.0000000	0.6877075	0.8094862	1.0000000
Emphys	784	0.6949612	1	0.0000000	0.6940759	0.8077867	1.0000000

Méthode	Nbclasses = 4	Nbclasses = 5
kmeans	0.5430	0.5882
CAH euclidienne	0.5876	0.6424
CAH diss. meth. 1	0.5989	0.6185
CAH diss. meth. 3	0.5831	0.6139
CAH diss. meth. 5	0.5935	0.6459
CAH diss. meth. 7	0.5942	0.6465
pam euclidien	0.5366	0.5713
pam diss. meth. 1	0.5366	0.5760
pam diss. meth. 3	0.5366	0.5766
pam diss. meth. 5	0.5366	0.5710
pam diss. meth. 7	0.5366	0.5722
mélange binaire	0.5575(*)	0.5872
mélange gaussien	0.5413	0.5752

(\*) de 0.48 à 0.58

CAH avec dissymétrie Jaccard ( $a/(a+b+c)$ ) et algorithme de Ward

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	tot_in_pop
Clusters_size	680	42	41	22	785
lign1	-----	-----	-----	-----	-----
Joint	680	10	7	15	712
Dizz_Vert_Bal	0	3	33	0	36
Nausea	0	5	31	0	36
Paresth	0	10	3	22	35
Fatigue	0	11	6	0	17
Back	0	14	2	0	16
Shock	0	9	4	0	13
Eye_mov	0	0	12	0	12
Vision	0	12	0	0	12
Diff.StandUp	0	7	1	0	8
Skin_marmor	0	8	0	0	8
Dyspnea	0	1	5	0	6
Hearing_pb	0	1	5	0	6
Speech	0	3	1	0	4
Loss_consc	0	1	1	0	2
Skin_rash	0	2	0	0	2
Emphys	0	1	0	0	1
chain_white1					
Sympt_tot_nb	680	98	111	37	926
patterns_nb	1	35	26	2	64
mean_sympt_nbr_by_indiv	1	2.333	2.707	1.682	1.18
[1]	0.5988889				

# La classification du Dr Fructus (Comex)

```
> print(fun_clusters(data_classif_x,Code_Comex)$matr_effect_cl)
```

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	tot_in_pop
Clusters_size	681	52	44	5	2	1	785
lign1	-----	-----	-----	-----	-----	-----	-----
Joint	681	7	21	3	0	0	712
Dizz_Vert_Bal	0	34	2	0	0	0	36
Nausea	0	31	5	0	0	0	36
Paresth	0	4	31	0	0	0	35
Fatigue	1	12	4	0	0	0	17
Back	0	3	13	0	0	0	16
Shock	0	9	4	0	0	0	13
Eye_mov	0	12	0	0	0	0	12
Vision	0	7	5	0	0	0	12
Diff.StandUp	0	3	5	0	0	0	8
Skin_marmor	0	1	2	5	0	0	8
Dyspnea	0	3	3	0	0	0	6
Hearing_pb	0	6	0	0	0	0	6
Speech	0	3	1	0	0	0	4
Loss_consc	0	2	0	0	0	0	2
Skin_rash	0	0	0	0	2	0	2
Emphys	0	0	0	0	0	1	1
chain_white1							
Sympt_tot_nb	682	137	96	8	2	1	926
patterns_nb	2	37	22	2	1	1	64
mean_sympt_nbr_by_indiv	1.001	2.635	2.182	1.6	1	1	1.18

```
> print(fun_Entr_cond(data_classif_x,CLUSTER_IG)$Entrp_ind)
```

```
[1] 0.5802651]
```

```
> tab<- table(Code_Comex,CLUSTER_IG)
```

```
> tab
```

	CLUSTER_IG			
Code_Comex	1	2	3	4
1	680	1	0	0
2	0	13	39	0
3	0	20	2	22
4	0	5	0	0
5	0	2	0	0
6	0	1	0	0

```
> randIndex(tab)
```

```
ARI
```

```
0.9789895
```

```
>
```

On observe dans la plupart des classifications :

- Un groupe "problèmes dans les articulations" contenant tous les individus pour lesquels "Joint" est le seul symptôme, plus éventuellement des individus qui ont des symptômes additionnels. Gravité faible de l'événement. (Groupe 1)
- Une classe "vestibulaire (troubles de l'oreille interne)-cérébral". Les manifestations sont principalement Troubles de l'équilibre/Nausée-vomissements/Mouvements des yeux (nystagmus)/troubles de l'audition/fatigue auxquels s'ajoutent éventuellement état de choc/dyspnée/douleurs dans les articulations. Gravité élevée. (groupe 2)
- Une classe "médullaire-neurologique" caractérisée par Fatigue/troubles de la vision/problèmes de dos/Difficulté à se tenir debout(paralyse)/Skin marmorata/état de choc/parésie. Gravité élevée. (Groupe 3)
- Une classe constituée essentiellement d'individus présentant les 2 symptômes "douleurs dans les articulations" et "parésie". Ces symptômes semblent attribuables à des atteintes nerveuses locales ou médullaires légères. Gravité modérée. (Groupe 4)

On observe dans la plupart des classifications :

- Un groupe "problèmes dans les articulations" contenant tous les individus pour lesquels "Joint" est le seul symptôme, plus éventuellement des individus qui ont des symptômes additionnels. Gravité faible de l'événement. (Groupe 1)
- Une classe "vestibulaire (troubles de l'oreille interne)-cérébral". Les manifestations sont principalement Troubles de l'équilibre/Nausée-vomissements/Mouvements des yeux (nystagmus)/troubles de l'audition/fatigue auxquels s'ajoutent éventuellement état de choc/dyspnée/douleurs dans les articulations. Gravité élevée. (groupe 2)
- Une classe "médullaire-neurologique" caractérisée par Fatigue/troubles de la vision/problèmes de dos/Difficulté à se tenir debout(paralyse)/Skin marmorata/état de choc/parésie. Gravité élevée. (Groupe 3)
- Une classe constituée essentiellement d'individus présentant les 2 symptômes "douleurs dans les articulations" et "parésie". Ces symptômes semblent attribuables à des atteintes nerveuses locales ou médullaires légères. Gravité modérée. (Groupe 4)

On observe dans la plupart des classifications :

- Un groupe "problèmes dans les articulations" contenant tous les individus pour lesquels "Joint" est le seul symptôme, plus éventuellement des individus qui ont des symptômes additionnels. Gravité faible de l'événement. (Groupe 1)
- Une classe "vestibulaire (troubles de l'oreille interne)-cérébral". Les manifestations sont principalement Troubles de l'équilibre/Nausée-vomissements/Mouvements des yeux (nystagmus)/troubles de l'audition/fatigue auxquels s'ajoutent éventuellement état de choc/dyspnée/douleurs dans les articulations. Gravité élevée. (groupe 2)
- Une classe "médullaire-neurologique" caractérisée par Fatigue/troubles de la vision/problèmes de dos/Difficulté à se tenir debout(paralyse)/Skin marmorata/état de choc/parésie. Gravité élevée. (Groupe 3)
- Une classe constituée essentiellement d'individus présentant les 2 symptômes "douleurs dans les articulations" et "parésie". Ces symptômes semblent attribuables à des atteintes nerveuses locales ou médullaires légères. Gravité modérée. (Groupe 4)

On observe dans la plupart des classifications :

- Un groupe "problèmes dans les articulations" contenant tous les individus pour lesquels "Joint" est le seul symptôme, plus éventuellement des individus qui ont des symptômes additionnels. Gravité faible de l'événement. (Groupe 1)
- Une classe "vestibulaire (troubles de l'oreille interne)-cérébral". Les manifestations sont principalement Troubles de l'équilibre/Nausée-vomissements/Mouvements des yeux (nystagmus)/troubles de l'audition/fatigue auxquels s'ajoutent éventuellement état de choc/dyspnée/douleurs dans les articulations. Gravité élevée. (groupe 2)
- Une classe "médullaire-neurologique" caractérisée par Fatigue/troubles de la vision/problèmes de dos/Difficulté à se tenir debout(paralyse)/Skin marmorata/état de choc/parèsthésie. Gravité élevée. (Groupe 3)
- Une classe constituée essentiellement d'individus présentant les 2 symptômes "douleurs dans les articulations" et "parèsthésie". Ces symptômes semblent attribuables à des atteintes nerveuses locales ou médullaires légères. Gravité modérée. (Groupe 4)