

Rapport d'étude statistique

La Santé Mentale



Réalisé par :
Oussama ELFEHRI
Youness MERIAF
Amine ZENNAI
Amine MHARZI

Encadré par :
Imane SATAURI

Année universitaire :
2021-2022

Sommaire

1. Datamining.....	5
2. Les méthodes utilisées.....	5
Analyse ACP:.....	8
Eigen values	8
Factor Loadings [Communality Estimates]	8
Factor Score Coefficients	9
Matrice des corrélations brutes et partielles :	10
Matrices	10
Correlations	10
Partial Correlations Controlling all other Variables	10
Original, reproduced and residual correlations.....	11
La qualité de representation:	12
Cercle des corrélations :	13
Analyse par k-means	16
Statistiques descriptives :	16
Centrer et réduire les variables actives :	17
K-Means :	18
Cluster centroids.....	20
Interprétation des groupes :	21
Graphique « Nuage de points » :	23
Projection dans le premier plan factoriel :	24
HAC :	25
Caractérisation des groupes :	26
Note :	27
Régression linière :	28
Régressions logistiques :	31
Rappel théorique	31
Interprétation de résultat :	33
Classification table	35
Variable in the Equation	35
Variable not in the equation.....	36
Les tests historiques des itérations	37

Les tests omnibus des coefficients du modèle	37
Hosmer and Lemeshow test.....	38
Classification table	38
Variable in the Equation	39
Conclusion :	40

Introduction

- La santé mentale est un état de bien-être dans lequel un individu peut réaliser son propre potentiel et faire face aux situations normales de la vie et au stress qu'elles génèrent. Une personne en bonne santé mentale peut notamment contribuer à sa communauté et travailler de façon productive. La santé mentale est donc un état de bien-être physique, mental et social complet et n'est pas simplement l'absence de troubles mentaux
- Les problèmes de santé mentale touchent les deux sexes, de tous âges, quel que soient leur culture, leur scolarité ou leur niveau de revenu. Les répercussions économiques et sociales de la maladie mentale touchent tout le monde dans notre pays, n'épargnant personne.
- Nous avons tous des périodes de vie difficiles où nous sommes bouleversés par nos pensées ou nos émotions, malgré tout, nous arrivons à poursuivre nos activités quotidiennes. Cependant, chez certaines personnes, des émotions intenses et incessantes et des schémas de pensée de nature chaotique et persistante créent des entraves considérables dans leur vie. Sans un traitement pour les aider à rétablir leur assise émotionnelle, elles continuent de se sentir submergées.
- Une maladie de l'esprit est comme une maladie du corps. Personne ne choisit la maladie et, moyennant une bonne évaluation ainsi qu'un traitement et un soutien adéquats, la guérison est possible. Il est temps de cesser de traiter les maladies de l'esprit différemment des maladies du corps. Il est temps de commencer à traiter les gens






Outils et méthodes utilisées

1. Datamining

Extraction d'information intéressante (non triviale, implicite, non connue précédemment et potentiellement utile) ou de patterns. Découverte de connaissance (mining) dans des Bdd, extraction de connaissance, analyse de données/pattern. Propose des résumés d'information (rapports multidimensionnels, résumés statistiques).

2. Les méthodes utilisées

Dans cette partie on va présenter les méthodes d'apprentissage supervisé que nous avons utilisé dans notre travail :

-  Analyse ACP
-  Analyse par K-means
-  HAC
-  Régression linéaire
-  Régressions logistiques

La base de données







</

Partie 1 : ANALYSE FONCTIONNELLE

Analyse ACP:

Eigen values

Matrix trace	6.000000
Average	1.000000

Axis	Eigen value	Difference	Proportion (%)	Histogram	Cumulative (%)
1	2.564134	1.065976	42.74 %		42.74 %
2	1.498158	0.615412	24.97 %		67.70 %
3	0.882746	0.388322	14.71 %		82.42 %
4	0.494424	0.141799	8.24 %		90.66 %
5	0.352625	0.144710	5.88 %		96.53 %
6	0.207914	-	3.47 %		100.00 %
Tot.	6.000000	-	-	-	-

Tanagra fournit le tableau des valeurs propres « Eigen values », avec le pourcentage d'inertie expliquée (individuelle et cumulée) par les axes. Un histogramme permet de situer leur décroissance.

⇒ On a inertie d'axe 1 + axe 2 + axe 3 = 82.4% > 80%, donc av travailler par trois axes.

Factor Loadings [Communality Estimates]

Corrélations variables / axes. La seconde partie des résultats indique la corrélation et le COS² en % et % cumulé des variables avec les axes factoriels.

- L'axe 1 représente la variable Travail parce qu'on a Corr dans l'axe 1 égale 0.87.
- L'axe 2 représente la variable Age parce qu'on a Corr dans l'axe 2 égale 0.80.
- L'axe 3 représente la variable Genre parce qu'on a Corr dans l'axe 3 égale -0.70.

Attribute	Axis_1		Axis_2		Axis_3	
	Corr.	% (Tot. %)	Corr.	% (Tot. %)	Corr.	% (Tot. %)
-						
Travail	0.87245	76 % (76 %)	-0.24838	6 % (82 %)	-0.12909	2 % (84 %)
Stresse	0.81042	66 % (66 %)	0.08847	1 % (66 %)	0.38863	15 % (82 %)
Prb.Sommeil	0.75146	56 % (56 %)	0.28786	8 % (65 %)	-0.01307	0 % (65 %)
Profession	0.69473	48 % (48 %)	-0.50183	25 % (73 %)	-0.36409	13 % (87 %)
age	0.31174	10 % (10 %)	0.80214	64 % (74 %)	0.28604	8 % (82 %)
genre	0.04061	0 % (0 %)	0.67120	45 % (45 %)	-0.70746	50 % (95 %)
Var. Expl.	2.56413	43 % (43 %)	1.49816	25 % (68 %)	0.88275	15 % (82 %)

Factor Score Coefficients

Coordonnées factorielles : projection, COS^2 et contribution. Un troisième tableau propose les paramètres des équations de projections pour chaque axe. Il est bien entendu nécessaire de centrer et réduire les données avec les moyennes et écarts-types affichés avant d'appliquer les coefficients.

Attribute	Mean	Std-dev	Axis_1	Axis_2	Axis_3
Age	1.8333333	0.8198916	0.1946814	0.6553475	0.3044414
Genre	0.2666667	0.4422166	0.0253631	0.5483670	-0.7529760
Profession	0.5666667	0.4955356	0.4338580	-0.4099965	-0.3875221
Prb.Sommeil	1.6000000	0.9865766	0.4692842	0.2351852	-0.0139079
Stresse	1.1000000	0.6506407	0.5061035	0.0722829	0.4136399
Travail	0.4000000	0.4898979	0.5448437	-0.2029268	-0.1373965

Matrice des corrélations brutes et partielles :

Matrices

La matrice des corrélations brutes montre la relation entre les variables.

La matrice des corrélations partielles indique les relations entre les variables en contrôlant l'impact de toutes les autres. Nous pouvons ainsi déceler la nature factice de certaines corrélations. Par exemple, $r[\text{Travail, Stress}] = 0.60654$ laisse à penser que le travail provoque le Stress. Aussi, $r[\text{Stresse, Prb.Sommeil}] = 0.52967$ cette résultat montre que les problème de sommeil peut causer le stress, on se rend compte que la liaison est en réalité déterminée par les autres préférences.

Correlations

	Travail	Stresse	Prb.Sommeil	Profession	age	genre
Travail	1.00000	0.60654	0.46898	0.71401	0.08299	-0.03077
Stresse	0.60654	1.00000	0.52967	0.34117	0.34367	-0.09268
Prb.Sommeil	0.46898	0.52967	1.00000	0.32728	0.32967	0.16809
Profession	0.71401	0.34117	0.32728	1.00000	-0.17776	-0.08113
age	0.08299	0.34367	0.32967	-0.17776	1.00000	0.30645
genre	-0.03077	-0.09268	0.16809	-0.08113	0.30645	1.00000

Partial Correlations Controlling all other Variables

	Travail	Stresse	Prb.Sommeil	Profession	age	genre
Travail	1.00000	0.43298	0.10169	0.64535	0.03524	0.04584
Stresse	0.43298	1.00000	0.30308	-0.07742	0.31198	-0.27688
Prb.Sommeil	0.10169	0.30308	1.00000	0.12862	0.18175	0.18862
Profession	0.64535	-0.07742	0.12862	1.00000	-0.30452	-0.01150
age	0.03524	0.31198	0.18175	-0.30452	1.00000	0.31397
genre	0.04584	-0.27688	0.18862	-0.01150	0.31397	1.00000

Original, reproduced and residual correlations

	Travail	Stresse	Prb.Sommeil	Profession	age	genre
Travail	-	0.6065 0.6349 (-0.0284)	0.4690 0.5858 (-0.1168)	0.7140 0.7778 (-0.0638)	0.0830 0.0358 (0.0472)	-0.0308 -0.0400 (0.0092)
Stresse	0.6065 0.6349 (-0.0284)	-	0.5297 0.6294 (-0.0997)	0.3412 0.3771 (-0.0360)	0.3437 0.4348 (-0.0911)	-0.0927 -0.1826 (0.0900)
Prb.Sommeil	0.4690 0.5858 (-0.1168)	0.5297 0.6294 (-0.0997)	-	0.3273 0.3824 (-0.0551)	0.3297 0.4614 (-0.1318)	0.1681 0.2330 (-0.0649)
Profession	0.7140 0.7778 (-0.0638)	0.3412 0.3771 (-0.0360)	0.3273 0.3824 (-0.0551)	-	-0.1778 -0.2901 (0.1123)	-0.0811 -0.0510 (-0.0301)
age	0.0830 0.0358 (0.0472)	0.3437 0.4348 (-0.0911)	0.3297 0.4614 (-0.1318)	-0.1778 -0.2901 (0.1123)	-	0.3065 0.3487 (-0.0422)
genre	-0.0308 -0.0400 (0.0092)	-0.0927 -0.1826 (0.0900)	0.1681 0.2330 (-0.0649)	-0.0811 -0.0510 (-0.0301)	0.3065 0.3487 (-0.0422)	-

La qualité de representation:

examples	Axe 1	Axe 2	Axe 3	COS 1	COS 2	COS 3
1.00	-0.64	-1.34	-0.14	0.13	0.56	0.01
2.00	-0.17	-1.11	-0.15	0.01	0.41	0.01
3.00	1.49	-0.85	0.59	0.45	0.15	0.07
4.00	-0.74	-0.41	1.28	0.10	0.03	0.29
5.00	-2.77	-0.87	0.02	0.87	0.08	0.00
6.00	0.95	-1.52	-0.43	0.23	0.60	0.05
7.00	1.96	-0.61	0.58	0.81	0.08	0.07
8.00	-1.04	1.08	1.39	0.23	0.25	0.40
9.00	-1.90	-1.69	-0.76	0.43	0.34	0.07
10.00	-1.04	-0.28	0.63	0.30	0.02	0.11
11.00	1.66	-0.48	-0.07	0.58	0.05	0.00
12.00	-1.04	1.08	1.39	0.23	0.25	0.40
13.00	1.96	-0.61	0.58	0.81	0.08	0.07
14.00	-0.03	2.80	-0.34	0.00	0.89	0.01
15.00	-1.46	0.72	-1.06	0.35	0.09	0.18
16.00	2.26	1.43	-0.76	0.56	0.22	0.06
17.00	0.55	-0.07	0.21	0.08	0.00	0.01
18.00	0.95	-1.52	-0.43	0.23	0.60	0.05
19.00	-2.48	1.17	-1.31	0.60	0.13	0.17
20.00	2.73	1.67	-0.77	0.68	0.25	0.05
21.00	-2.77	-0.87	0.02	0.87	0.08	0.00
22.00	0.95	-1.52	-0.43	0.23	0.60	0.05
23.00	1.96	-0.61	0.58	0.81	0.08	0.07
24.00	-1.04	1.08	1.39	0.23	0.25	0.40
25.00	-1.84	-0.45	-2.46	0.32	0.02	0.57
26.00	-1.04	-0.28	0.63	0.30	0.02	0.11
27.00	1.72	0.76	-1.78	0.42	0.08	0.45
28.00	-1.04	1.08	1.39	0.23	0.25	0.40
29.00	1.96	-0.61	0.58	0.81	0.08	0.07
30.00	-0.03	2.80	-0.34	0.00	0.89	0.01

Les individus qui sont bien représenter dans l'axe 1 :

Axe 1 > 0	Axe 1 < 0
7-13-16-20-23-29	5-9-19-21

Les individus qui sont bien représenter dans l'axe 2 :

Axe 2 > 0	Axe 2 < 0
14-30	6-18-22

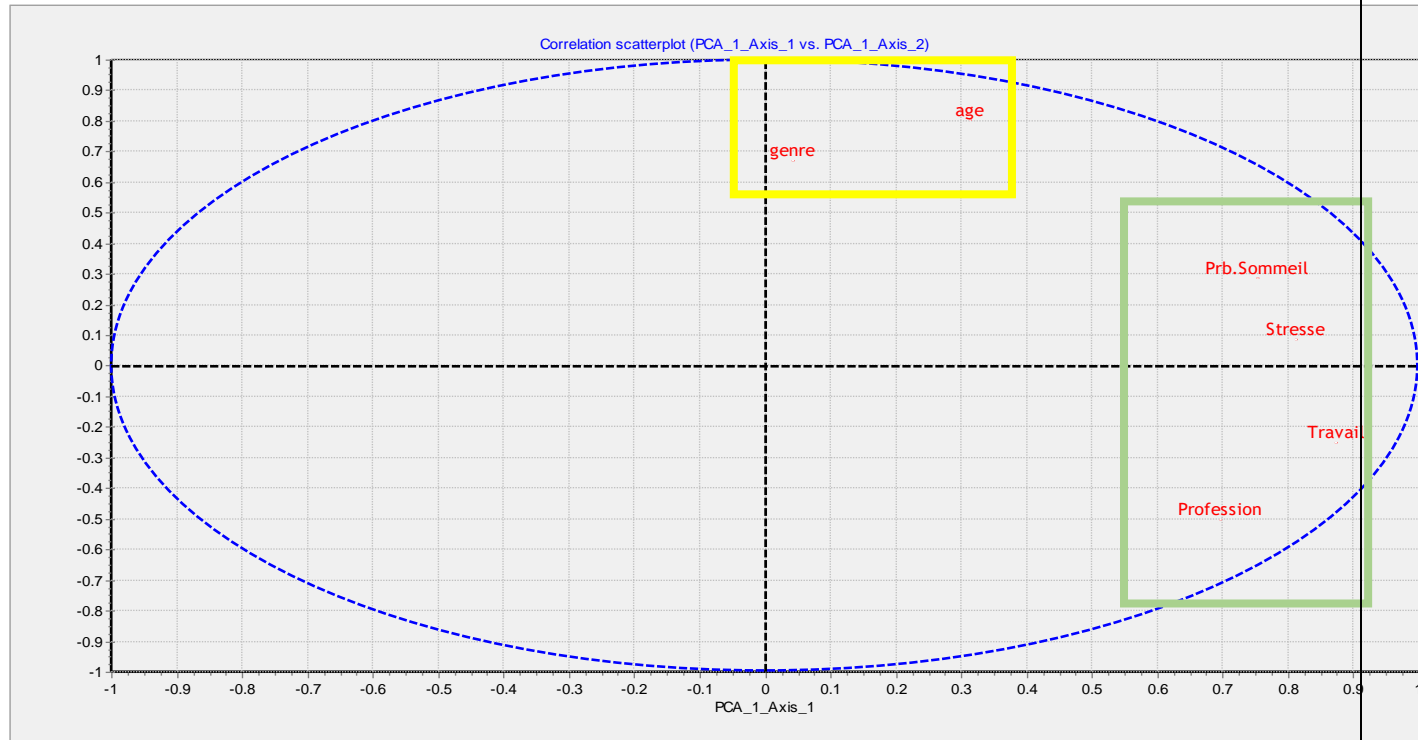
Les individus qui sont bien représenter dans l'axe 3 :

Axe 3 > 0	Axe 3 < 0
12-8-24-25	27-28

Cercle des corrélations :

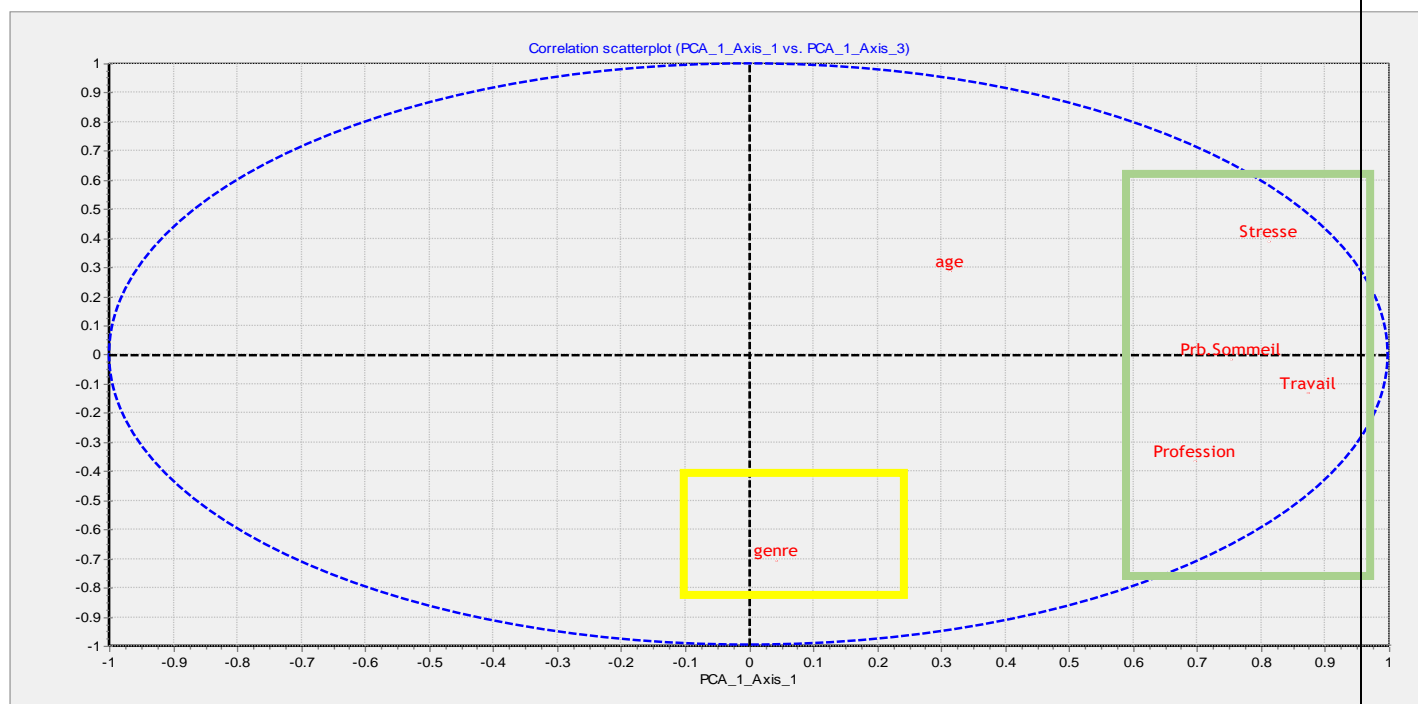
Le cercle de corrélation sert à obtenir une vue synthétique du positionnement des variables par rapport aux axes.

Axe 1 et Axe 2 :



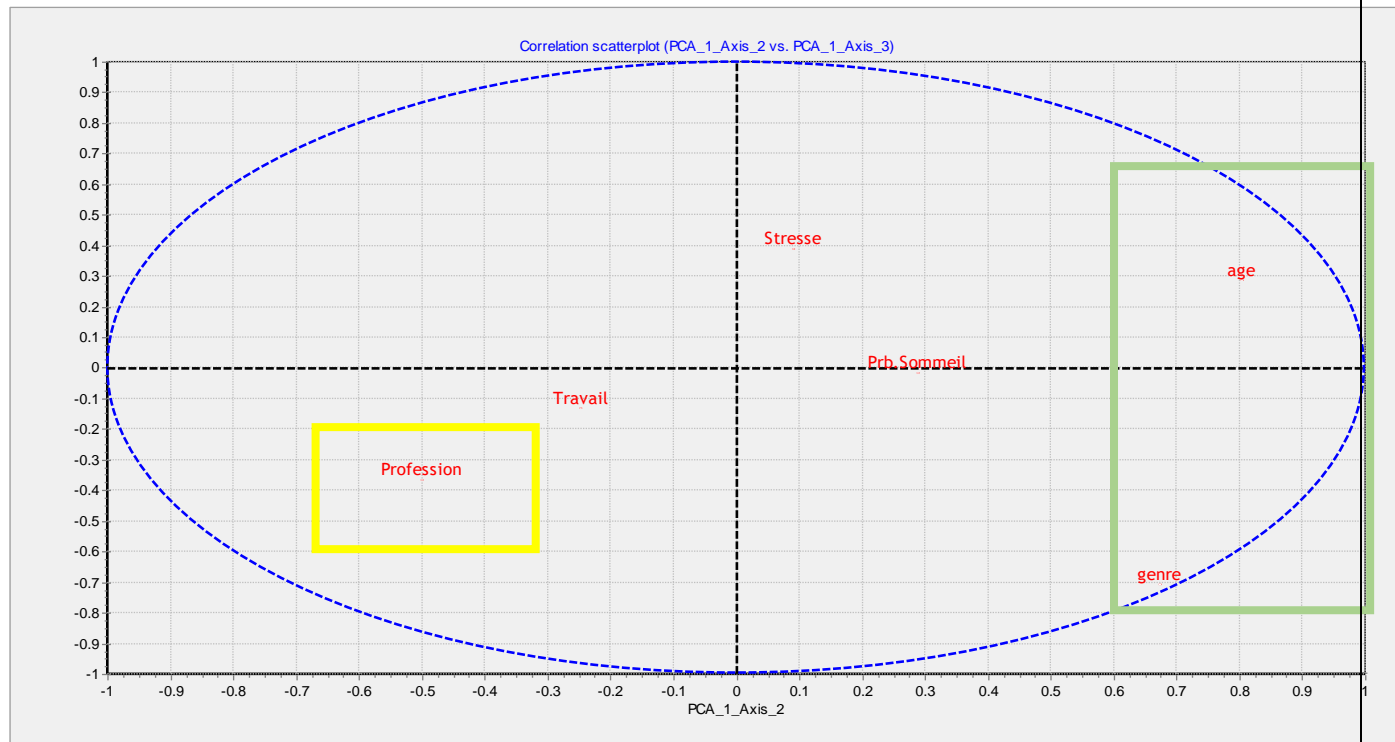
On distingue une corrélation positive entre **Prb.Sommeil**, **Stress**, **Travail** et **Profession** et l'axe 1 qui représente le travail, et on a une corrélation positive entre **genre**, **age** et l'axe 2 qui représente l'âge.

Axe 1 et Axe 3 :



On distingue une corrélation positive entre **Prb.Sommeil**, **Stress**, **Travail** et **Profession** et l'axe 1 qui représente le travail, et on a une corrélation négative entre **genre** et l'axe 3 qui représente le genre.

Axe 2 et Axe 3 :

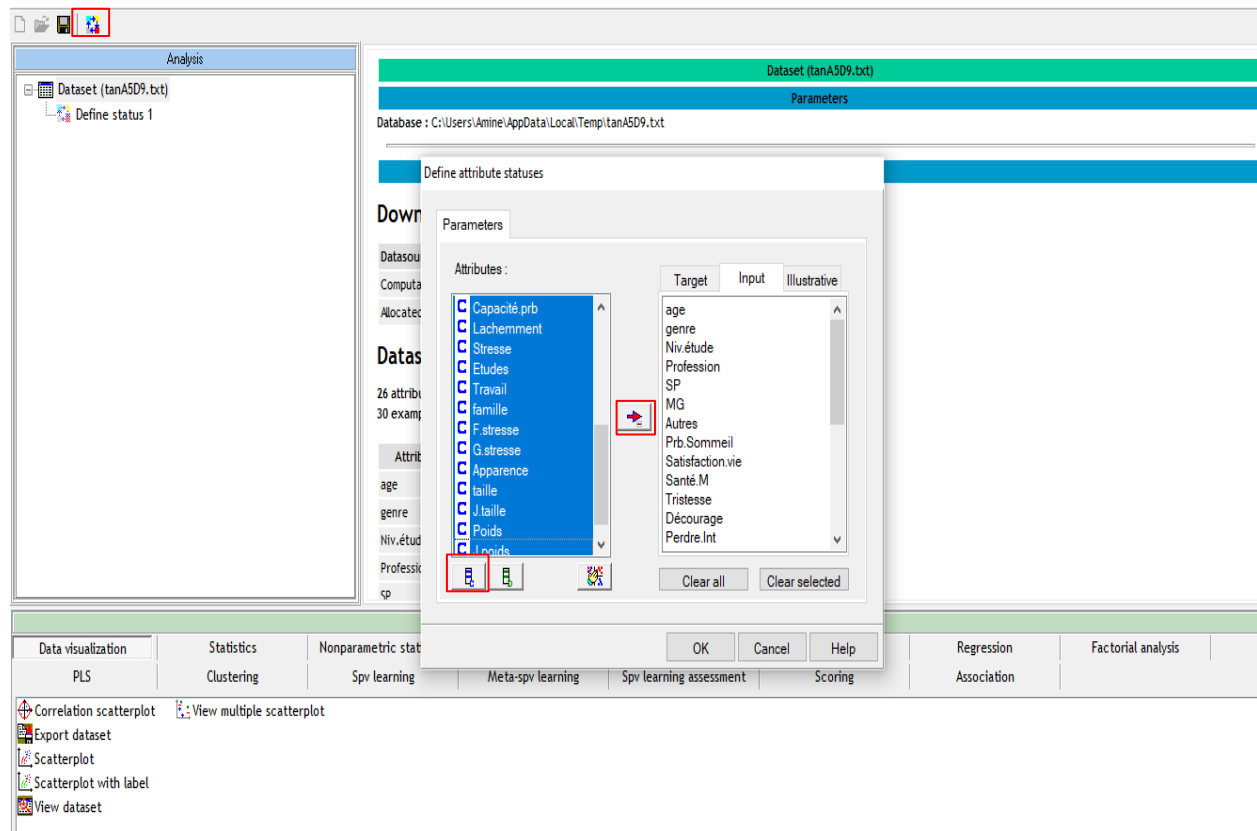


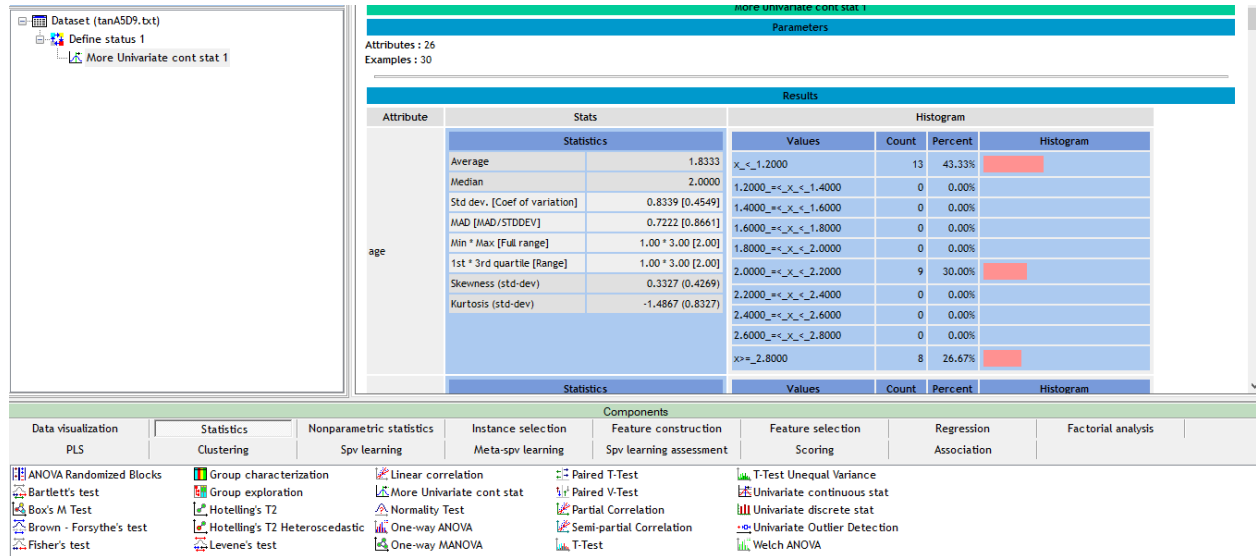
On distingue une corrélation positive entre **age**, **genre** et l'axe 3 qui représente le genre, et on a une corrélation négative entre **genre** et une corrélation positive avec **age** et l'axe 2 qui représente l'âge.

Analyse par k-means

Statistiques descriptives :

Première étape, nous calculons quelques indicateurs de statistique descriptive sur les données. L'idée est de détecter les éventuelles anomalies. Nous insérons le composant DEFINE STATUS via le raccourci dans la barre d'outils. Nous plaçons les variables continues en INPUT. Il s'agit des variables actives de l'analyse.

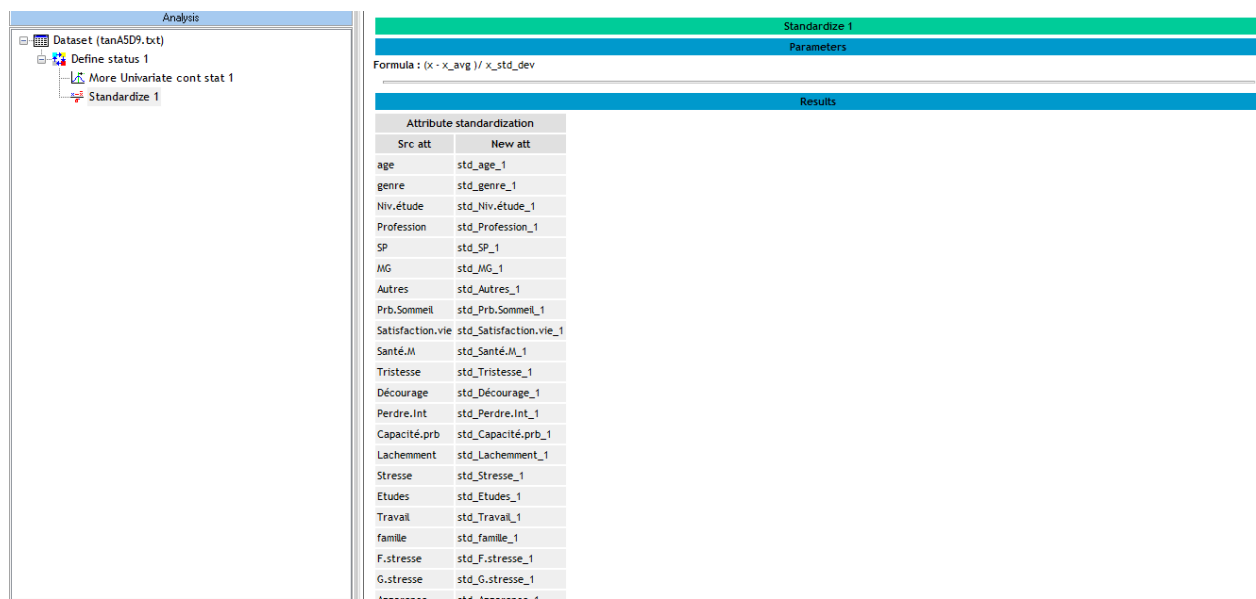




Il n'y a pas de données franchement atypiques ni d'asymétrie marquée, rien en tous les cas qui justifie un pré traitement spécifique.

Centrer et réduire les variables actives :

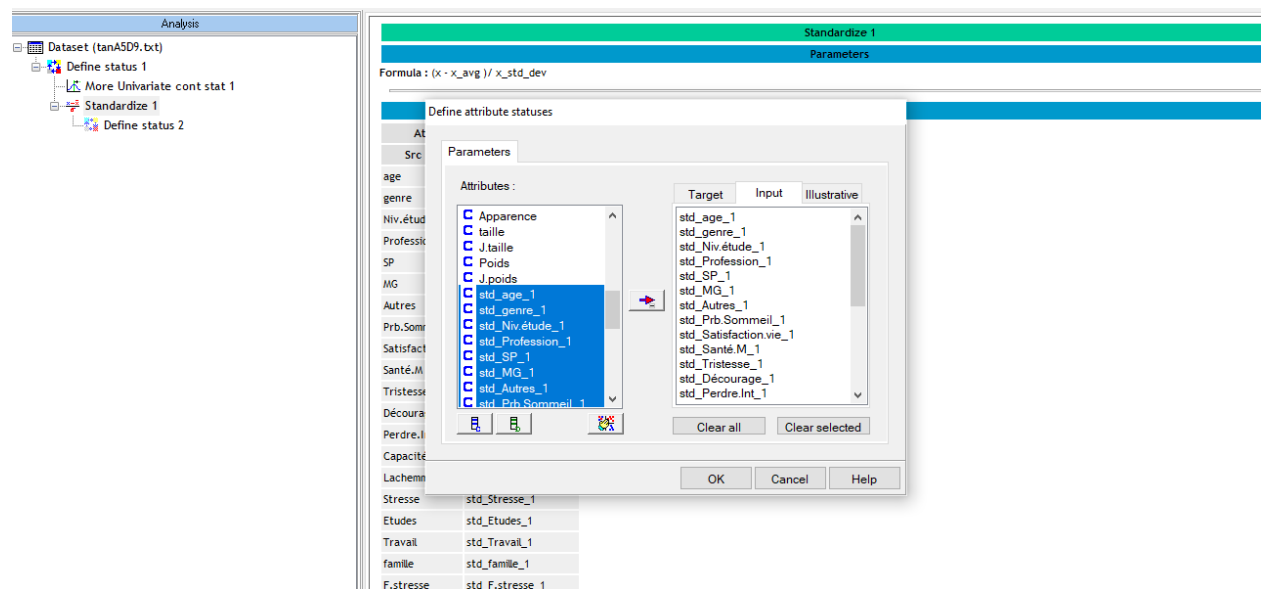
Nous souhaitons centrer et réduire les variables avant de le présenter à la méthode des K-MEANS, ceci afin d'éliminer les disparités d'échelle². Nous introduisons le composant STANDARDIZE (onglet FEATURE CONSTRUCTION) dans le diagramme, à la suite de DEFINE STATUS 1. Nous actionnons le menu contextuel VIEW.



26 nouvelles variables intermédiaires sont maintenant disponibles pour les traitements ultérieurs.

K-Means :

Nous devons préciser à TANAGRA qui sont les variables transformées qui seront utilisées pour les calculs. Ensuite Nous insérons un nouveau DEFINE STATUS, nous plaçons en INPUT les variables STD_MPG_1 à STD_ACCELERATION_1.



Nous pouvons introduire le composant K-MEANS (onglet CLUSTERING). Nous actionnons dans un premier temps le menu PARAMETERS pour spécifier les paramètres de traitement.

tan151.txt)
: status 1
standardize 1
Define status 3
K-Means 2

Formula : $(x - x_{avg}) / s$ K-Means parameters

Parameters

Number of clusters : 2
Max iterations : 10
Number of trials : 5

Distance normalization
☒ None
☐ Variance

Average computation
☐ Forgy
☒ Mc Queen

Seed random generator
☐ Random
☒ Standard

OK Cancel Help

Statistics Nonparametric statistics
Clustering Spv learning
Group exploration
Brown - Forsythe's test
Fisher's test
Hotelling's T2
Levene's test
Linear correlation
Normality Test
One-way ANOVA
Regression
Association
Paired T-Test
Paired V-Test
Factorial analysis
Semi-partial
T-Test

Nous demandons une partition en 2 groupes. Il n'est pas nécessaire de normaliser la distance puisque les variables sont déjà réduites. Nous validons, puis nous cliquons sur le menu VIEW.

Dataset (tan59F5.txt)
Define status 1
More Univariate cont stat 1
Standardize 1
Define status 2
K-Means 1

K-Means 1
Parameters

K-Means parameters

Clusters 2
Max Iteration 10
Trials 5
Distance normalization none
Average computation McQueen
Seed random generator Standard

Results

Global evaluation

Within Sum of Squares 607.3804
Total Sum of Squares 754.0000
R-Square 0.1945

Cluster size and WSS

Cluster	Description	Size	WSS
cluster n°1	c_kmeans_1	10	172.0856
cluster n°2	c_kmeans_2	20	435.2949

R-Square for each attempt

TANAGRA nous annonce qu'il y a respectivement 20 et 10 observations dans chaque groupe. La partition explique 19.45 % de l'inertie totale.

Dans la partie basse de la fenêtre de visualisation, Tanagra affiche les moyennes conditionnelles sur les variables ayant participé à la construction de la partition. Elles sont donc calculées sur les données centrées et réduites. Elles permettent de comprendre les différenciations entre les groupes mais, n'étant pas exprimées dans les unités des variables initiales, elles ne sont pas vraiment utilisables pour l'interprétation.

Cluster centroids

Attribute	Cluster n1	Cluster n2
std_age_1	0.079945	-0.039972
std_genre_1	0.074111	-0.037055
std_Niv.ȳde_1	-0.528059	0.264029
std_Profession_1	-0.727503	0.363752
std_SP_1	-0.262769	0.131384
std_MG_1	0.154973	-0.077486
std_Autres_1	0.000000	0.000000
std_Prб.Sommeil_1	-1.096226	0.548113
std_Satisfaction.vie_1	0.807969	-0.403984
std_Santȳ_1	0.871691	-0.435845
std_Tristesse_1	1.292168	-0.646084
std_Dȳurage_1	-0.429100	0.214550
std_Perdre.Int_1	-0.370554	0.185277
std_Capacitȳrb_1	0.234609	-0.117305
std_Lachement_1	-0.200693	0.100347
std_Stresse_1	-0.906668	0.453334
std_Etudes_1	-0.370554	0.185277
std_Travail_1	-0.802773	0.401386
std_famille_1	-0.764744	0.382372
std_F.stresse_1	-0.965475	0.482738
std_G.stresse_1	0.276435	-0.138217
std_Apparence_1	0.446077	-0.223039
std_taille_1	-0.234609	0.117305
std_J.taille_1	0.519391	-0.259696

std_Poids_1	-0.548511	0.274256
std_J.poids_1	0.372700	-0.186350

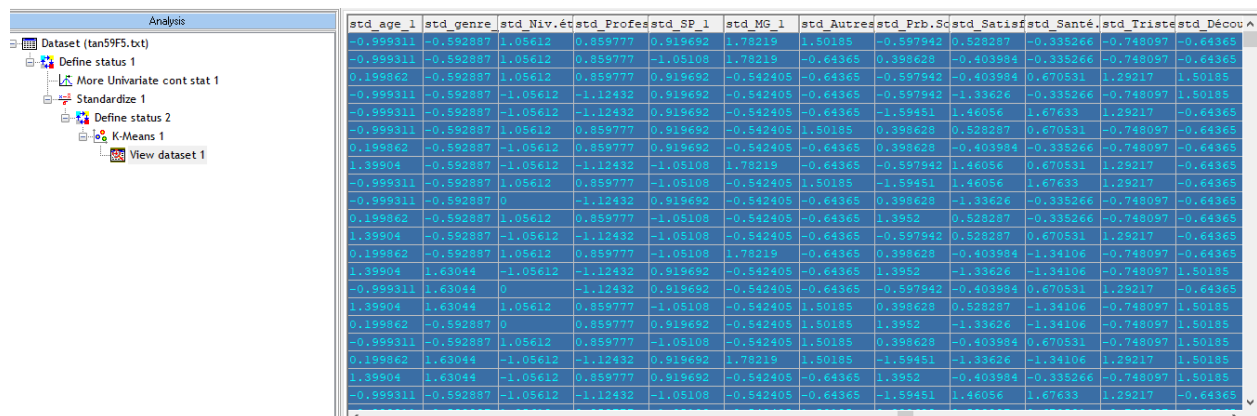
Tout ce qu'on peut dire sur ce résultat est que, pour toutes les variables, les écarts vont inférieure de 1 écart type, sauf (Tristesse).

Interprétation des groupes :

La partie délicate de la typologie : l'interprétation des résultats. Nous abordons ici les différentes pistes qui permettent de comprendre la formation des groupes.

1. Affectation des classes aux individus :

Première approche, revenir aux données en inspectant l'affectation des classes aux individus. Elle est réellement viable si les effectifs sont faibles et les individus identifiés (cela pourrait être la marque et le modèle du véhicule dans notre exemple). TANAGRA produit automatiquement une variable supplémentaire dans la base courante. Elle décrit la classe affectée à chaque individu. Nous pouvons la visualiser avec le composant VIEW DATASET (onglet DATA VISUALIZATION). Elle est positionnée en dernière colonne. L'intérêt de cet outil est très limité dès que la base va au-delà de quelques dizaines d'observations.



std_age_1	std_genre	std_Niv.ét	std_Profes	std_SP_1	std_MG_1	std_Autres	std_Prb.Sc	std_Satisf	std_Santé	std_Tristesse	std Découv
-0.999311	-0.592887	1.05612	0.859777	0.919692	1.78219	1.50185	-0.597942	0.528287	-0.335266	-0.748097	-0.64365
-0.999311	-0.592887	1.05612	0.859777	-1.05108	1.78219	-0.64365	0.398628	-0.403984	-0.335266	-0.748097	-0.64365
0.199862	-0.592887	1.05612	0.859777	0.919692	-0.542405	-0.64365	-0.597942	-0.403984	0.670531	1.29217	1.50185
-0.999311	-0.592887	-1.05612	-1.12432	0.919692	-0.542405	-0.64365	-0.597942	-1.33626	-0.335266	-0.748097	1.50185
-0.999311	-0.592887	-1.05612	-1.12432	0.919692	-0.542405	-0.64365	-1.59451	1.46056	1.67633	1.29217	-0.64365
-0.999311	-0.592887	1.05612	0.859777	0.919692	-0.542405	1.50185	0.398628	0.528287	0.670531	-0.748097	-0.64365
0.199862	-0.592887	1.05612	0.859777	0.919692	-0.542405	-0.64365	0.398628	-0.403984	-0.335266	-0.748097	-0.64365
1.39904	-0.592887	-1.05612	-1.12432	-1.05108	1.78219	-0.64365	-0.597942	1.46056	0.670531	1.29217	-0.64365
-0.999311	-0.592887	1.05612	0.859777	-1.05108	-0.542405	1.50185	-1.59451	1.46056	1.67633	1.29217	-0.64365
-0.999311	-0.592887	0	-1.12432	0.919692	-0.542405	-0.64365	0.398628	-1.33626	-0.335266	-0.748097	-0.64365
0.199862	-0.592887	1.05612	0.859777	-1.05108	-0.542405	-0.64365	1.3952	0.528287	-0.335266	-0.748097	-0.64365
1.39904	-0.592887	-1.05612	-1.12432	-1.05108	-0.542405	-0.64365	-0.597942	0.528287	0.670531	1.29217	-0.64365
0.199862	-0.592887	1.05612	0.859777	-1.05108	1.78219	-0.64365	0.398628	-0.403984	-1.34106	-0.748097	-0.64365
1.39904	1.63044	-1.05612	-1.12432	0.919692	-0.542405	-0.64365	1.3952	-1.33626	-1.34106	-0.748097	1.50185
-0.999311	1.63044	0	-1.12432	0.919692	-0.542405	-0.64365	-0.597942	-0.403984	0.670531	1.29217	-0.64365
1.39904	1.63044	1.05612	0.859777	-1.05108	-0.542405	1.50185	0.398628	0.528287	-1.34106	-0.748097	1.50185
0.199862	-0.592887	0	0.859777	0.919692	-0.542405	1.50185	1.3952	-1.33626	-1.34106	-0.748097	1.50185
-0.999311	-0.592887	1.05612	0.859777	-1.05108	-0.542405	1.50185	0.398628	-0.403984	0.670531	-0.748097	1.50185
0.199862	1.63044	-1.05612	-1.12432	0.919692	1.78219	1.50185	-1.59451	-1.33626	-1.34106	1.29217	1.50185
1.39904	1.63044	-1.05612	0.859777	0.919692	-0.542405	-0.64365	1.3952	-0.403984	-0.335266	-0.748097	1.50185
-0.999311	-0.592887	-1.05612	-1.12432	-1.05108	-0.542405	-0.64365	-1.59451	1.46056	1.67633	1.29217	-0.64365

2.Statistiques descriptives comparatives :

Seconde approche, très simple et pourtant très instructive, nous pouvons calculer les statistiques descriptives conditionnelles sur les variables actives et illustratives. Les oppositions permettent souvent de situer les caractéristiques marquantes des groupes.

Analysis

- Dataset (tan59F5.txt)
 - Define status 1
 - More Univariate cont stat 1
 - Standardize 1
 - Define status 2
 - K-Means 1
 - View dataset 1
 - Define status 3

Define attribute statuses

Parameters

Attributes:

- age
- genre
- Niv.étude
- Profession
- SP
- MG
- Autres
- Prb.Sommeil
- Satisfaction.vie
- Santé.M
- Tristesse
- Décourage
- Perdre.Int

Target: Cluster_KMeans_1

Input:

Illustrative:

Clear all Clear selected

OK Cancel Help

Data visualization: PLS, Correlation scatterplot, Export dataset, Scatterplot

Statistics: Clustering, View dataset, View multiple scatterplot

Nonparametric stat: Spv learning, Meta-spv learning, Spv learning assessment, Scoring

Regression: Association

Factorial analysis

Nous introduisons le composant DEFINE STATUS dans le diagramme. Nous plaçons en TARGET la variable désignant les classes CLUSTER_KMEANS_1, en INPUT les variables actives originelles. Puis, nous insérons le composant GROUP CHARACTERIZATION (onglet STATISTICS).

Results

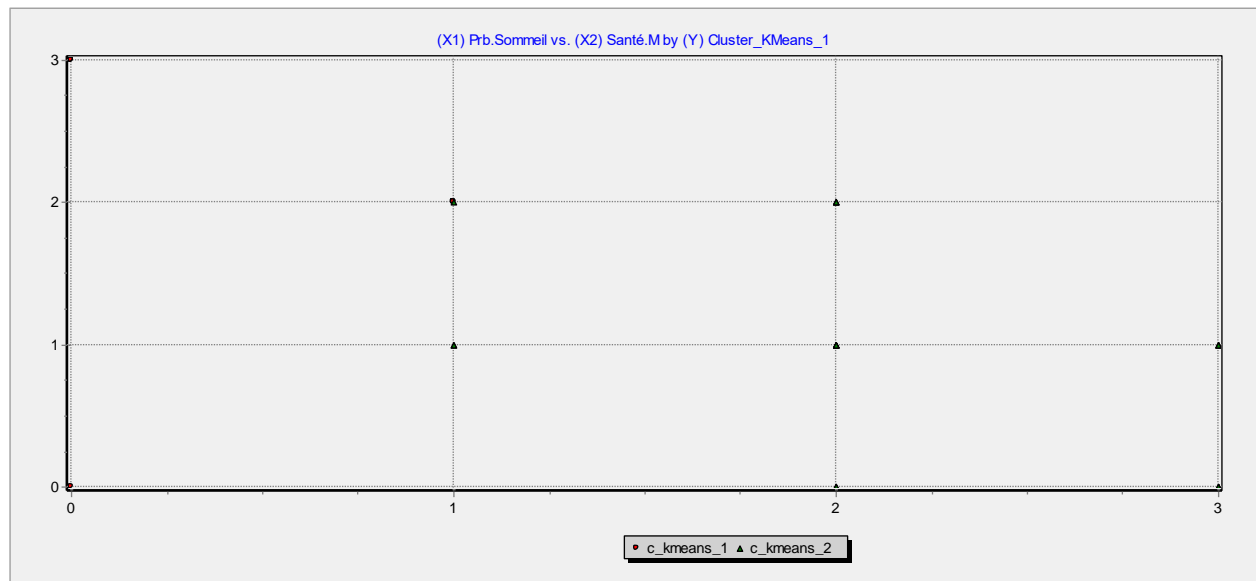
Description of "Cluster_KMeans_1"

Cluster_KMeans_1=c_kmeans_1 [33.3 %] 10					Cluster_KMeans_1=c_kmeans_2 [66.7 %] 20				
Examples	Att - Desc	Test value	Group	Overall	Examples	Att - Desc	Test value	Group	Overall
Continuous attributes : Mean (StdDev)					Continuous attributes : Mean (StdDev)				
	Tristesse	4.92	1.00 (0.00)	0.37 (0.49)		Prb.Sommeil	4.17	2.15 (0.67)	1.60 (1.00)
	Santé.M	3.32	2.20 (0.92)	1.33 (0.99)		F.stresse	3.68	1.85 (0.75)	1.40 (0.93)
	Satisfaction.vie	3.08	2.30 (1.06)	1.43 (1.07)		Stresse	3.45	1.40 (0.50)	1.10 (0.66)
	J.taille	1.98	1.80 (0.42)	1.40 (0.77)		Travail	3.06	0.60 (0.50)	0.40 (0.50)
	Apparence	1.70	2.20 (0.92)	1.77 (0.97)		famille	2.91	0.85 (0.37)	0.67 (0.48)
	J.poids	1.42	2.00 (0.00)	1.77 (0.63)		Profession	2.77	0.75 (0.44)	0.57 (0.50)
	G.stresse	1.05	1.50 (1.35)	1.17 (1.21)		Poids	2.09	1.40 (1.19)	1.10 (1.09)
	Capacité.prb	0.89	1.80 (1.32)	1.53 (1.14)		Niv.étude	2.01	1.25 (0.91)	1.00 (0.95)
	MG	0.59	0.30 (0.48)	0.23 (0.43)		Décourage	1.63	0.40 (0.50)	0.30 (0.47)
	age	0.30	1.90 (0.99)	1.83 (0.83)		Perdre.Int	1.41	0.35 (0.49)	0.27 (0.45)
	genre	0.28	0.30 (0.48)	0.27 (0.45)		Etudes	1.41	0.35 (0.49)	0.27 (0.45)
	Autres	0.00	0.30 (0.48)	0.30 (0.47)		SP	1.00	0.60 (0.50)	0.53 (0.51)
	Lachement	-0.76	0.50 (0.53)	0.60 (0.50)		taille	0.89	1.30 (0.57)	1.23 (0.57)
	taille	-0.89	1.10 (0.57)	1.23 (0.57)		Lachement	0.76	0.65 (0.49)	0.60 (0.50)
	SP	-1.00	0.40 (0.52)	0.53 (0.51)		Autres	0.00	0.30 (0.47)	0.30 (0.47)
	Etudes	-1.41	0.10 (0.32)	0.27 (0.45)		genre	-0.28	0.25 (0.44)	0.27 (0.45)
	Perdre.Int	-1.41	0.10 (0.32)	0.27 (0.45)		age	-0.30	1.80 (0.77)	1.83 (0.83)
	Décourage	-1.63	0.10 (0.32)	0.30 (0.47)		MG	-0.59	0.20 (0.41)	0.23 (0.43)
	Niv.étude	-2.01	0.50 (0.85)	1.00 (0.95)		Capacité.prb	-0.89	1.40 (1.05)	1.53 (1.14)

Le principal intérêt de GROUP CHARACTERIZATION est qu'il permet d'introduire à la fois les variables actives et illustratives, qu'elles soient quantitatives ou qualitatives. Dans notre étude, la variable dans le premier group permet de mieux comprendre les classes c'est Tristesse et pour le deuxième c'est prb.sommeil.

Graphique « Nuage de points » :

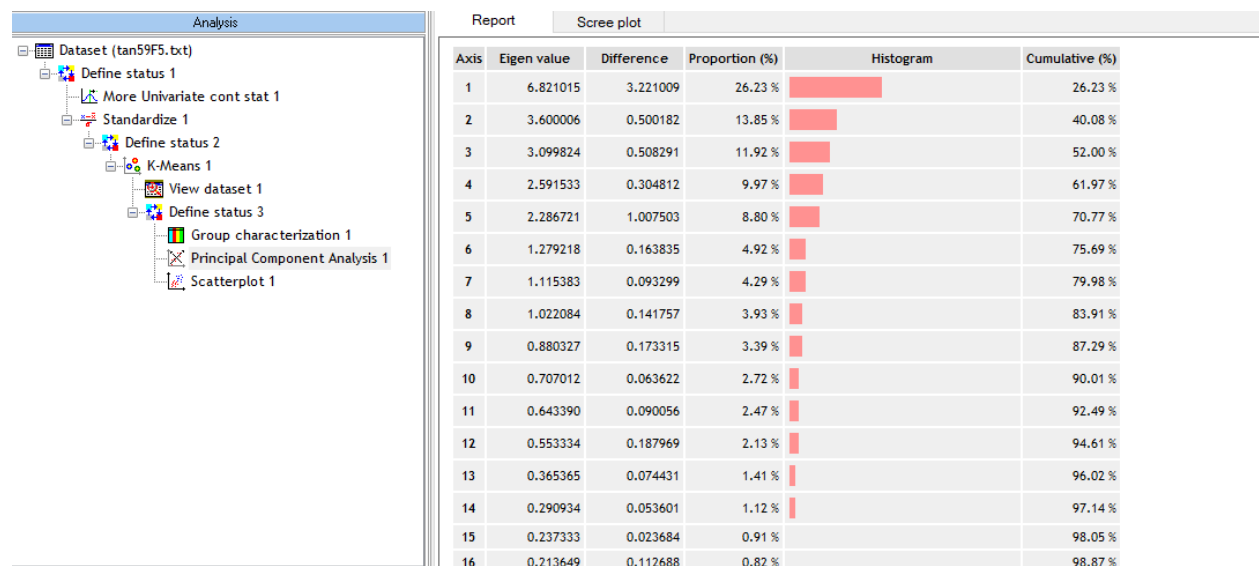
Une autre manière d'interpréter les résultats est de positionner les groupes dans l'espace des couples de variables. On peut ainsi analyser l'action conjointe de deux variables. Le graphique « nuage de points » est un outil privilégié pour cela.



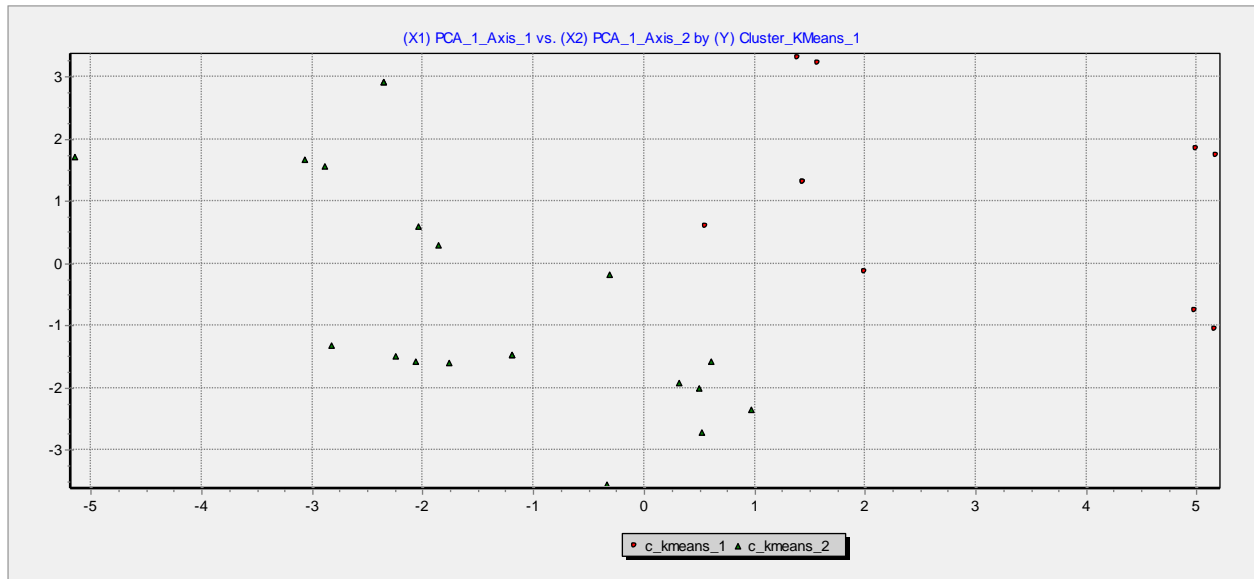
Dans TANAGRA, il suffit d'introduire l'outil SCATTERPLOT (onglet DATA VISUALIZATION). Nous cliquons sur VIEW. Dans l'outil de visualisation, nous mettons en abscisse la sente mentale et en ordonnée la problème de sommeil (prb.sommeil). Il nous reste à coloriser les données selon leur classe d'appartenance.

Projection dans le premier plan factoriel :

Pour tenir compte du rôle simultané des variables actives, nous pouvons projeter les données dans le premier plan factoriel de l'ACP. S'il traduit une part suffisamment importante de l'information contenue dans les données, le positionnement des groupes dans cet espace devrait être assez fidèle du mécanisme de formation des groupes. Nous insérons le composant PRINCIPAL COMPONENT ANALYSIS (onglet FACTORIAL ANALYSIS) à la suite du K-MEANS 1. Il utilisera de fait les mêmes variables actives. Nous cliquons sur VIEW. Le premier plan factoriel traduit 90.01 % de l'information disponible.



Lorsque nous construisons le graphique à l'aide du composant SCATTERPLOT, en abscisse le premier axe PCA_1_AXIS_1, en ordonnée le second PCA_1_AXIS_2. Les groupes se démarquent très nettement.



HAC :

Nous choisissons la CAH pour la construction de la typologie. Elle affiche les différents niveaux d'agrégation et donne des indications sur le nombre de classes à retenir.

Enfin, les effectifs étant faibles pour cet exemple, nous pouvons prendre comme racine de la hiérarchie les observations du fichier de données. Si l'effectif est élevé, plusieurs milliers d'observations, il est plus judicieux de s'appuyer sur le principe de la classification mixte en définissant un premier partitionnement assez fin (une cinquantaine de classes) avec une méthode pouvant gérer de gros volumes de données (un K-MEANS ou un SOM par exemple), puis de les utiliser comme racine du dendrogramme. Le temps de calcul est nettement amélioré tout en préservant la qualité des résultats.

Voici le diagramme de traitement correspondant, nous paramétrons la méthode de manière à normaliser les données. En effet, étant exprimées dans des unités différentes, il est préférable de ramener les variables dans le même référentiel.

Analysis

- Dataset (tan59F5.txt)
 - Define status 1
 - More Univariate cont stat 1
 - Standardize 1
 - Define status 2
 - K-Means 1
 - View dataset 1
 - Define status 3
 - Group characterization 1
 - Principal Component Analysis 1
 - Scatterplot 2
 - Scatterplot 1
 - HAC 1

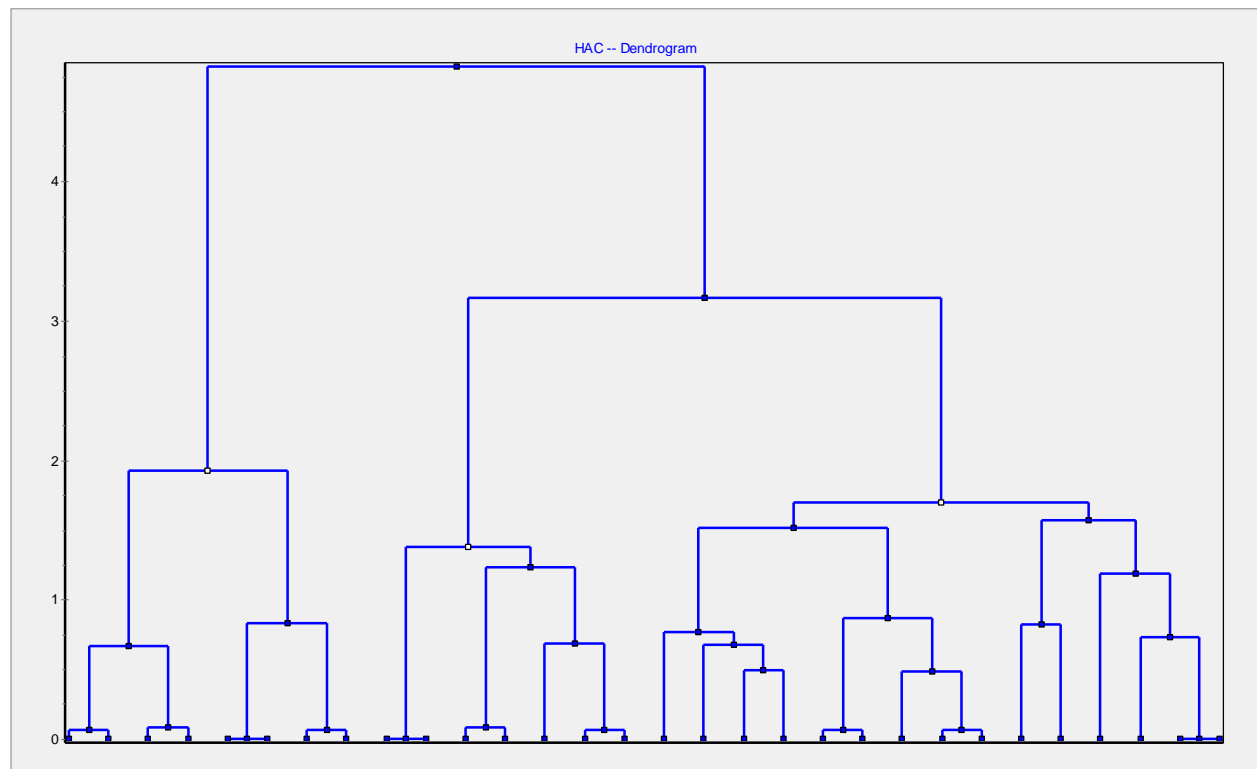
Report **Dendrogram**

Best cluster selection

Clusters	BSS ratio	Gap
1	0.0000	0.0000
2	0.1855	1.6602
3	0.3072	1.2402
4	0.3812	0.2288
5	0.4464	0.1280
6	0.5066	0.0485
7	0.5650	0.1435
8	0.6179	0.1409
9	0.6654	0.0476
10	0.7110	0.3244

Cluster centroids

Attribute	Cluster n°1	Cluster n°2	Cluster n°3
std_age_1	0.199862	0.713794	-0.439697
std_genre_1	-0.314971	0.677585	-0.148222
std_Niv.étude_1	-0.528059	-0.452622	0.492855
std_Profession_1	-0.628298	0.292891	0.198410
std_SP_1	-0.558385	0.075077	0.262769



Caractérisation des groupes :

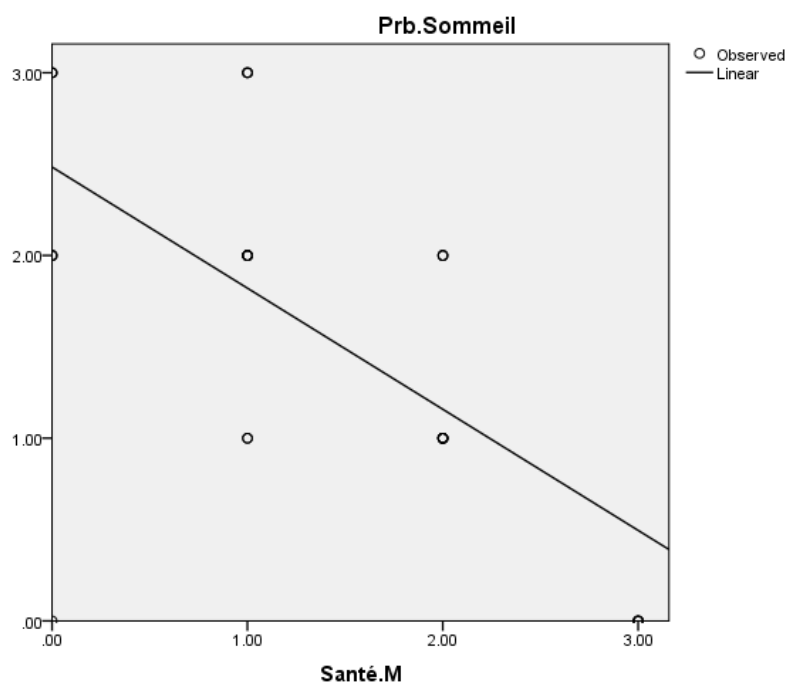
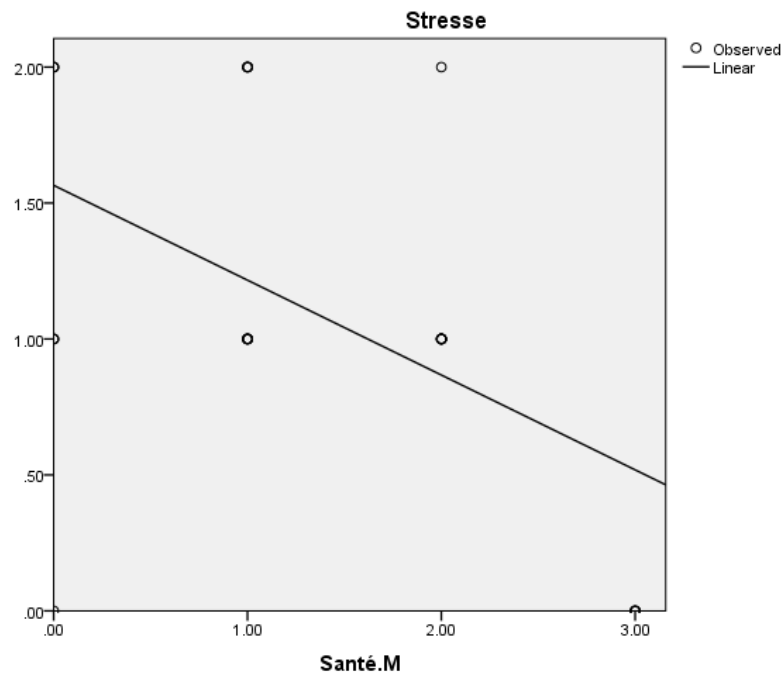
Analysis					Cluster_HAC_1=c_hac_1					Cluster_HAC_1=c_hac_2					Cluster_HAC_1=c_hac_3				
Dataset (tan59FS.txt)					Examples [26.7 %] 8					Examples [23.3 %] 7					Examples [50.0 %] 15				
Define status 1					Att - Desc Test value Group Overall					Att - Desc Test value Group Overall					Att - Desc Test value Group Overall				
More Univariate cont stat 1					Continuous attributes : Mean (StdDev)					Continuous attributes : Mean (StdDev)					Continuous attributes : Mean (StdDev)				
Standardize 1					std_Tristesse_1	4.20	1.29 (0.00)	0.00 (1.00)		std_Poids_1	4.38	1.48 (0.45)	0.00 (1.00)		std_Etudes_1	3.19	0.59 (1.15)	0.00 (1.00)	
Define status 2					std_Satisfaction.vie_1	3.99	1.23 (0.43)	0.00 (1.00)		std_taille_1	4.01	1.35 (0.00)	0.00 (1.00)		std_Niv.étude_1	2.65	0.49 (0.78)	0.00 (1.00)	
K-Means 1					std_Santé.M_1	3.81	1.17 (0.54)	0.00 (1.00)		std_Prb.Sommeil_1	3.30	1.11 (0.49)	0.00 (1.00)		std_Perdre.Int_1	2.39	0.44 (1.15)	0.00 (1.00)	
View dataset 1					std_Apparence_1	2.45	0.75 (0.55)	0.00 (1.00)		std_F.stresse_1	3.28	1.10 (0.57)	0.00 (1.00)		std_famille_1	2.25	0.42 (0.73)	0.00 (1.00)	
Define status 3					std_G.stresse_1	1.57	0.48 (1.15)	0.00 (1.00)		std_age_1	2.12	0.71 (0.64)	0.00 (1.00)		std_Lachement_1	2.16	0.40 (0.83)	0.00 (1.00)	
Group characterization 1					std_J.taille_1	1.48	0.45 (0.60)	0.00 (1.00)		std_genre_1	2.01	0.68 (1.19)	0.00 (1.00)		std_Autres_1	1.93	0.36 (1.11)	0.00 (1.00)	
Principal Component Analysis 1					std_Capacité.prb_1	1.33	0.41 (1.15)	0.00 (1.00)		std_Travail_1	1.87	0.63 (0.98)	0.00 (1.00)		std_SP_1	1.42	0.26 (0.96)	0.00 (1.00)	
Scatterplot 2					std_J.poids_1	1.21	0.37 (0.00)	0.00 (1.00)		std_Stresse_1	1.48	0.50 (0.81)	0.00 (1.00)		std_Stresse_1	1.36	0.25 (0.90)	0.00 (1.00)	
Scatterplot 1					std_age_1	0.65	0.20 (1.28)	0.00 (1.00)		std_Profession_1	0.87	0.29 (0.97)	0.00 (1.00)		std_MG_1	1.25	0.23 (1.13)	0.00 (1.00)	
HAC 1					std_MG_1	0.13	0.04 (1.08)	0.00 (1.00)		std_Décourage_1	0.82	0.28 (1.15)	0.00 (1.00)		std_Décourage_1	1.16	0.21 (1.09)	0.00 (1.00)	
Define status 4					std_taille_1	0.10	0.03 (0.81)	0.00 (1.00)		std_Capacité.prb_1	0.47	0.16 (0.43)	0.00 (1.00)		std_Profession_1	1.07	0.20 (0.97)	0.00 (1.00)	
Group characterization 2					std_Autres_1	-0.35	-0.11 (0.99)	0.00 (1.00)		std_famille_1	0.30	0.10 (1.02)	0.00 (1.00)		std_Travail_1	0.72	0.13 (1.04)	0.00 (1.00)	
					std_Lachement_1	-0.65	-0.20 (1.07)	0.00 (1.00)		std_SP_1	0.22	0.08 (1.05)	0.00 (1.00)		std_Prb.Sommeil_1	0.36	0.07 (0.72)	0.00 (1.00)	
					std_genre_1	-1.02	-0.31 (0.79)	0.00 (1.00)		std_J.poids_1	-0.25	-0.08 (1.21)	0.00 (1.00)		std_Apparence_1	0.18	0.03 (0.89)	0.00 (1.00)	
					std_Niv.étude_1	-1.71	-0.53 (0.98)	0.00 (1.00)		std_G.stresse_1	-0.76	-0.26 (0.31)	0.00 (1.00)		std_F.stresse_1	0.00	0.00 (0.79)	0.00 (1.00)	
					std_Poids_1	-1.78	-0.55 (0.49)	0.00 (1.00)		std_Perdre.Int_1	-0.82	-0.28 (0.84)	0.00 (1.00)		std_J.taille_1	-0.47	-0.09 (0.94)	0.00 (1.00)	
					std_SP_1	-1.81	-0.56 (0.91)	0.00 (1.00)		std_J.taille_1	-0.99	-0.33 (1.39)	0.00 (1.00)		std_G.stresse_1	-0.74	-0.14 (1.09)	0.00 (1.00)	
					std_Etudes_1	-1.93	-0.59 (0.00)	0.00 (1.00)		std_Satisfaction.vie_1	-1.20	-0.40 (0.76)	0.00 (1.00)		std_genre_1	-0.80	-0.15 (0.92)	0.00 (1.00)	
					std_Perdre.Int_1	-1.93	-0.59 (0.00)	0.00 (1.00)		std_Niv.étude_1	-1.34	-0.45 (1.03)	0.00 (1.00)		std_J.poids_1	-0.86	-0.16 (1.16)	0.00 (1.00)	
					std_Profession_1	-2.04	-0.63 (0.92)	0.00 (1.00)		std_MG_1	-1.61	-0.54 (0.00)	0.00 (1.00)		std_Capacité.prb_1	-1.58	-0.29 (1.06)	0.00 (1.00)	
					std_Décourage_1	-2.09	-0.64 (0.00)	0.00 (1.00)		std_Etudes_1	-1.76	-0.59 (0.00)	0.00 (1.00)		std_Santé.M_1	-1.81	-0.34 (0.85)	0.00 (1.00)	
					std_Travail_1	-2.61	-0.80 (0.00)	0.00 (1.00)		std_Santé.M_1	-1.85	-0.62 (0.49)	0.00 (1.00)		std_Tristesse_1	-1.83	-0.34 (0.84)	0.00 (1.00)	
					std_famille_1	-2.82	-0.87 (0.97)	0.00 (1.00)		std_Lachement_1	-1.87	-0.63 (0.98)	0.00 (1.00)		std_Poids_1	-2.13	-0.40 (0.66)	0.00 (1.00)	
					std_Stresse_1	-2.94	-0.91 (0.81)	0.00 (1.00)		std_Autres_1	-1.91	-0.64 (0.00)	0.00 (1.00)		std_age_1	-2.37	-0.44 (0.77)	0.00 (1.00)	
					std_Tristesse_1	-3.14	-0.93 (0.53)	0.00 (1.00)		std_Tristesse_1	-3.33	-0.95 (0.00)	0.00 (1.00)		std_Etudes_1	-3.54	-0.43 (0.74)	0.00 (1.00)	

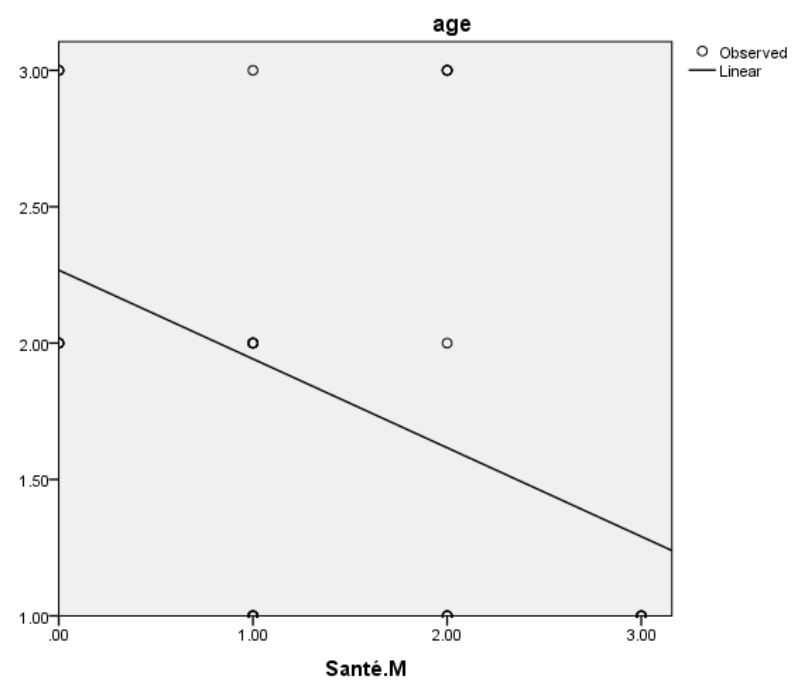
Note :

On constate que la tristesse contribue mieux dans le premier hac_cluster , dans le deuxième on a le poids et pour le dernier l'études, mais on a le troisième cluster représente la moitié donc on peut conclure que L'étude provoque principalement.

Régression linière :

Après l'application de la méthode de régression on peut conclure les bonnes corrélations entre la santé mentale avec :





Partie 2 :

Prévision et Prédictive

Régressions logistiques :

Rappel théorique

La régression logistique propose de tester un modèle de régression dont la variable dépendante est dichotomique (codée 0-1) et dont les variables indépendantes peuvent être continue ou catégorielles.

Un modèle de régression logistique permet aussi de prédire la probabilité qu'un événement arrive (valeur de 1) ou non (valeur de 0) à partir de l'optimisation des coefficients de régression. Ce résultat varie toujours entre 0 et 1. Lorsque la valeur prédite est supérieure à 0,5, l'événement est susceptible de se produire, alors que lorsque cette valeur est inférieure à 0,5, il ne l'est pas.

Pour réaliser une régression logistique, on a cliqué sur **Analyse, Régression, Logistique binaire**.

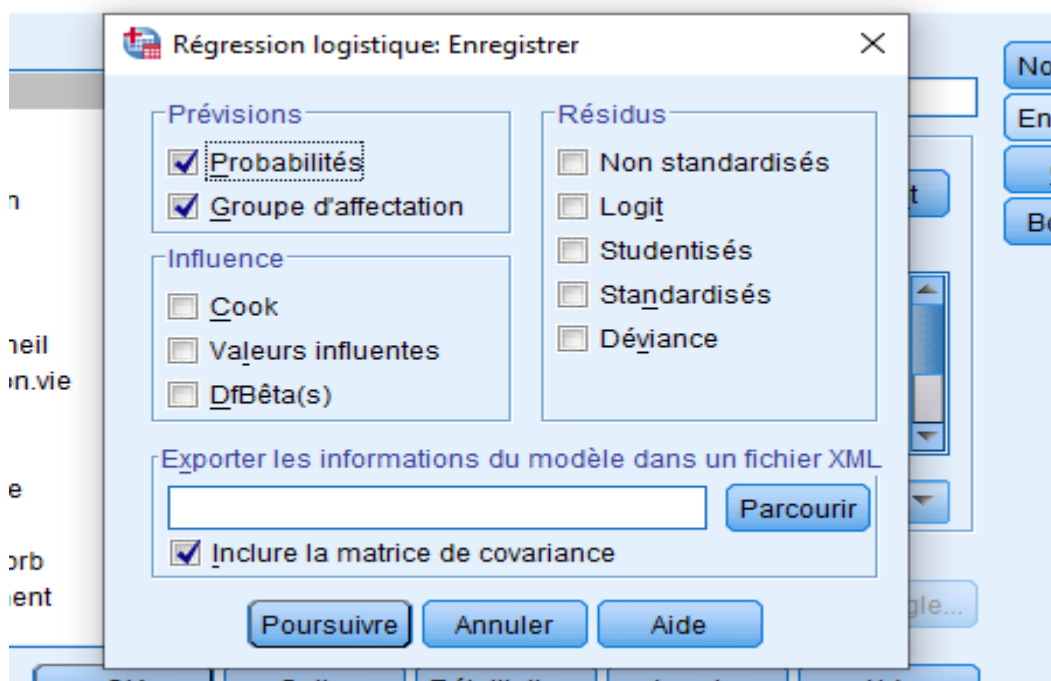
Puis cette boîte de dialogue s'affiche, pour insérer la variable dépendante dans la boîte **Dépendant** cette variable doit être une vraie variable dichotomique et non une variable continue recodée en 2 groupes, ce qui serait associé à une importante perte d'information, et les variables prédictrices dans la boîte **Covariables**.



La variable que nous voulons expliquer, c'est le Santém alors on la considère comme variable

dépendant, on a choisi cette variable pour savoir si les différentes caractéristiques des gens qu'on a étudié ont une relation avec leur maladie et surtout et la santé mentale (SantéM) qui'ils dépendent.

Et cette variable va s'expliquer par l'âge du client, son revenu, son sexe, la fréquence d'achat par mois, alors on les considère comme des variables indépendantes et on les met sur la boîte **Covariables**



On a coché sur les deux cases **Probabilités** et **Groupe d'affectations**

Probabilités : Enregistre pour chaque observation la probabilité d'occurrence prévue pour l'événement. Dans les résultats, un tableau affiche le nom et le contenu de toutes les nouvelles variables. L'événement est la catégorie de la variable dépendante qui a la plus haute valeur, par exemple, si la valeur dépendante prend la valeur 0 et 1, la probabilité prédite de la catégorie 1 est enregistrée.

Groupes d'affectation. : Le groupe qui possède la probabilité a posteriori la plus élevée, basé sur les écarts discriminants. Le groupe prévu par le modèle est celui auquel appartient l'observation.

Dans Options :

On a coché **Qualité d'ajustement de Hosmer_lemeshow**

Cette statistique de qualité d'ajustement utilisée pour la régression logistique, particulièrement pour les modèles ayant des covariables continues et les études d'échantillons de petite taille. Elle est basée sur le regroupement des observations en déciles de risque et la comparaison de la probabilité observée avec la probabilité théorique à l'intérieur de chaque décile.

Après on clique sur OK

Interprétation de résultat :

« Case Processing Summary » et « Dependent Variable Encoding »

Récapitulatif de traitement des observations

Observations non pondérées ^a		N	Pourcentage
Observations sélectionnées	Inclus dans l'analyse	30	100,0
	Observations manquantes	0	,0
	Total	30	100,0
Observations exclues		0	,0
Total		30	100,0

a. Si la pondération est activée, reportez-vous au tableau de classement pour connaître le nombre total d'observations.

Codage de variables dépendantes

Valeur d'origine	Valeur interne
Mauvaise	0
Excellente	1

Pour ces deux premiers tableaux

La première table. Elles affichent un résumé de traitement de cas, c'est-à-dire qu'elles indiquent les valeurs incluses et les valeurs (manquantes) non incluses dans les données incluse sur cette analyse et ici on a 30 individus sont inclus et 0 non inclus.

La deuxième table est indiquée tous simplement que SPSS a conservé les mêmes valeurs que celles utilisées pour coder la variable dépendante budget, soit 0 pour les individus qui possèdent mauvaise « santé mentale » et 1 qui possèdent bonne.

Bloc 0 : bloc de départ

Historique des itérations^{a,b,c}

Itération		-2log-vraisemblance	Coefficients
			Constante
Etape 0	1	41,455	-,133
	2	41,455	-,134

- a. La constante est incluse dans le modèle.
- b. -2log-vraisemblance initiale : 41,455
- c. L'estimation a été interrompue au numéro d'itération 2 parce que les estimations de paramètres ont changé de moins de ,001.

Dans ce tableau ce qui nous intéresse c'est la dernière valeur **-2log-vraisemblance initial** =41,455 parce ce qu'on va l'utiliser pour faire un comparaison.

Bloc 0 : bloc de départ

Tableau de classement^{a,b}

Observations			Prévisions		
			SantéM		Pourcentage correct
			Mauvaise	Excellente	
Etape 0	SantéM	Mauvaise	16	0	100,0
		Excellente	14	0	,0
Pourcentage global					53,3

a. La constante est incluse dans le modèle.

b. La valeur de césure est ,500

Classification table

Le tableau de classification « Classification Table » montre les résultats pratiques de l'utilisation du modèle de régression logistique

Cette table décrit un modèle des prédictions de base qui sont faites uniquement sur la catégorie la plus fréquente de nos données. Ici, le modèle suppose toujours «0» car la plupart des participant possèdent des maladies « mauvaise » (16 contre 14 selon la première colonne). La ligne de pourcentage globale nous indique que cette approche de la prédiction est correcte 53,30% des gens maladies.

Variable in the Equation

Variables dans l'équation

	A	E.S.	Wald	ddl	Sig.	Exp(B)
Etape 0 Constante	-,134	,366	,133	1	,715	,875

Cette table Les variables dans l'équation nous montrent le coefficient de la constante (B0), dans notre cas, il est de $-0,134$.

Selon ce tableau, le modèle statistiquement n'est pas significatif car « Sig = $0,715 > 0,05$ ».

Mais les résultats de cette table ne peuvent pas pleinement nous confirmer que nos variables ne sont pas significatives.

Variable not in the equation

Variables hors de l'équation

	Score	ddl	Sig.
Etape 0 Variables age	1,417	1	,234
genre	2,058	1	,151
Niv.étude	,000	1	1,000
Travail	1,094	1	,296
Stresse	2,139	1	,144
Tristesse	2,625	1	,105
Prb.Sommeil	,930	1	,335
Profession	,621	1	,431
Statistiques globales	8,314	8	,403

Ce tableau « *Variable hors de l'équation* » montre les valeurs de la statistique Score pour chaque variable prédictrice hors de l'équation qui s'apparente aux valeurs de corrélation partielle dans la régression multiple.

Et depuis la valeur de « Sig » on peut déduire que toutes les variables sont statistiquement non significatives alors ils peuvent ne pas tous contribuer à l'amélioration du modèle.

Les tests historiques des itérations

Block 1 : Méthode = Entrée

Historique des itérations^{a,b,c,d}

Itération	-2log-vraisemblance	Coefficients								
		Constante	age	genre	Niv.étude	Travail	Stresse	Tristesse	Prb.Sommeil	Profession
Etape 1	32,428	,681	-1,291	-,657	-,845	,120	,874	,504	,639	,628
2	31,956	,760	-1,628	-,931	-1,146	,116	1,173	,740	,852	,838
3	31,941	,778	-1,690	-1,000	-1,214	,119	1,237	,787	,895	,872
4	31,941	,779	-1,693	-1,003	-1,217	,120	1,240	,789	,897	,873
5	31,941	,779	-1,693	-1,003	-1,217	,120	1,240	,789	,897	,873

a. Méthode : Entrée

b. La constante est incluse dans le modèle.

c. -2log-vraisemblance initiale : 41,455

d. L'estimation a été interrompue au numéro d'itération 5 parce que les estimations de paramètres ont changé de moins de ,001.

Dans ce tableau ce qui nous intéresse c'est la dernière valeur -2log-vraisemblance = 31,941 (bloc 1) parce ce qu'on va l'utiliser pour faire une comparaison.

Maintenant revenant au résultat du tableau d'historique des itérations du bloc 0 nous avons gardé une valeur qui égale à -2log-vraisemblance initial = 41,455 La comparaison des deux valeurs nous donne est ce que le modèle significatif ou non, si la valeur du bloc 1 < à la valeur bloc 0 donc le modèle peut bien prédire. Dans notre cas on trouve que : -2log-vraisemblance (bloc 0) = 41,455 > -2log-vraisemblance (bloc 1) = 31,941 Donc le modèle est significatif

Les tests omnibus des coefficients du modèle

Nous passons maintenant au modèle de régression qui inclut nos variables explicatives. Le prochain ensemble de tableaux commence par l'en-tête du « block 1: Methode = Entrée ».

Block 1 : Méthode = Entrée

Tests de spécification du modèle

		Khi-Chi-deux	ddl	Sig.
Etape 1	Etape	9,514	8	,301
	Bloc	9,514	8	,301
	Modèle	9,514	8	,301

Ces tests sont utilisés pour vérifier que le nouveau modèle (avec les variables

explicatives incluses) est une amélioration par rapport au modèle de référence. Il utilise des tests du « *Khi-chi-deux* » pour voir s'il existe une différence significative entre les log-vraisemblances (en particulier les -2LL) du modèle de base et du nouveau modèle. Si le nouveau modèle a un -2LL significativement réduit par rapport à la ligne de base, cela suggère que le nouveau modèle explique davantage la variance du résultat et constitue une amélioration.

Ici, le Khi-chi-deux = 9,514 ddl = 8, sig = 0,301 > 0,05, donc notre nouveau modèle est nettement significatives .

Hosmer and Lemeshow test

Test de Hosmer-Lemeshow

Etape	Khi-Chi-deux	ddl	Sig.
1	2,807	6	,833

Ce tableau aide à déterminer si le modèle décrit correctement les données. Ici le test Hosmer-Lemeshow indique un mauvais ajustement car la valeur de signification est inférieure à 0,05.

Classification table

Tableau de classement^a

Observations			Prévisions		
			SantéM		Pourcentage correct
			Mauvaise	Excellente	
Etape 1	SantéM	Mauvaise	11	5	68,8
		Excellente	5	9	64,3
Pourcentage global					66,7

a. La valeur de césure est ,500

Il est maintenant possible d'examiner si le modèle permet de bien classer les gens dans leur groupe d'appartenance à partir de l'équation finale. Nous nous rappelons que le hasard permettait de classer correctement 53,3 % des participants. Nous voyons que le pourcentage correct de classification passe de 66,7 % et selon les résultats précédents, cette amélioration est significative.

Variable in the Equation

Variables dans l'équation

		A	E.S.	Wald	ddl	Sig.	Exp(B)
Etape 1 ^a	age	-1,693	,973	3,026	1	,082	,184
	genre	-1,003	1,198	,700	1	,403	,367
	Niv.étude	-1,217	,915	1,766	1	,184	,296
	Travail	,120	1,465	,007	1	,935	1,127
	Stresse	1,240	1,090	1,295	1	,255	3,455
	Tristesse	,789	2,240	,124	1	,725	2,201
	Prb.Sommeil	,897	1,151	,607	1	,436	2,452
	Profession	,873	1,787	,239	1	,625	2,394
	Constante	,779	2,322	,113	1	,737	2,179

a. Variable(s) entrées à l'étape 1 : age, genre, Niv.étude, Travail, Stresse, Tristesse, Prb. Sommeil, Profession.

Ce tableau fournit le coefficient de régression (A), la statistique de Wald (pour tester la signification statistique) et le rapport de cotes (Exp (B)) le plus important pour chaque catégorie de variable.

En regardant d'abord les résultats pour la variable, il y a un effet global très significatif (Wald = 3,026, ddl = 1, Sig >0 ,05). Les coefficients A significatifs et positifs pour la variable Travail, Stresse , Tristesse et Profession, ce qui indique qu' ils influencent dans la santéM. Comparativement le coefficient des variables age , genre , Niv_étude est négatives ce qui indique qu'ils ne sont pas des facteurs essentiel pour influence sur la santé mentale.

Conclusion :

A travers ce projet, nous avons pu en effet consolider les connaissances acquises et d'enrichir notre expérience en matière d'analyse. A travers le sujet choisi « Santé Mentale », on constate que le datamining peut être appliqué dans tous les domaines et s'avère utile pour tous les décideurs.

Néanmoins, ce projet ne s'est pas terminé sans difficulté. La base de données ayant fait l'objet d'autre projet, malgré notre volonté de faire une étude totalement différente de la première, la limitation des données nous faisait énormément dans le choix des variables lors de l'application des méthodes. Ce qui a entraîné un certain retard dans la réalisation de ce projet.

Donc on constate d'après la réalisation de ce projet on constate que la santé mentale La santé mentale fait partie intégrante de la santé et du bien-être. D'une autre façon on peut dire que la santé mentale aide d'une manière énorme le développement de la santé d'un être-humains.

