

## ASSIGNMENT 1:

The data analyzed are based on measurements made on a sample of 31 men during a sporting event. They represent age (in years), weight (WEIGHT in kg), oxygen consumed per kg per minute (OXYGEN), time taken to cover 2.4km (RUNTIME, in minutes), resting heart rate (RESTPULSE), heart rate during the race (RUNPULSE) and the maximum of this rate (MAXPULSE).

The age variable has been classified as follows: C1= between 38 and 44 years old, C2= between 45 and 50 years old, C3= between 51 and 57 years old.

To highlight the links between the different measures, a standardized Principal Component Analysis was performed on these data.

1) By analyzing the first table on page 3 explain why it was decided to carry out a second standard PCA on only 28 individuals.

The following questions concern the second standard PCA. The weight and age class variables are supplementary variables.

2) Justify why a standard PCA is used.

3) Give the inertia of the cloud of points. Justify the answer.

4) Specify the possible alternative to the standard PCA and its effects on the results of the analysis? Give the value of inertia in this case.

5) Recalling the usual criteria, how many components do you select?

6) Interpret (internal) the first two principal components and in particular the position of individuals 2, 4, 9, 21, and 28. 7) Explain how the additional variables weight and age class were added to the PCA results. Give the respective formulas. 8) Interpret (external) the results using supplementary variables. Check the consistency between the internal and external interpretation of the age and age-class variables.

9) A hierarchical clustering by Ward's method has been performed on the PCA results. The results in the appendices (pages 9 to 12) are identical to those of the direct classification on the original variables.

- a. Justify this result.
- b. Justify the choice of partitioning into 4 classes.
- c. What do the level indices represent?
- d. How much is the sum of the level indices? Justify the answer.
- e. Recall what mixed classification is all about. The usefulness of each of its steps should be explained.
- f. Calculate the inertia of cluster 3.
- g. What is the value of the between inertia of the partition? Justify the answers (formulas can be given).
- h. Interpret the clusters.

As for the age variable, it was decided to classify all the other variables: 7 classes for restpulse and maxpulse, 8 classes for runpulse, 2 classes for runtime, 4 classes for oxygen and 3 classes for weight. A multiple correspondence analysis was then carried out on all the qualitative variables.

Answer the following questions:

- 10) Briefly recall the broad outlines of the method.
- 11) Regarding the number of categories of the variables, which preliminary treatments should be carried out?
- 12) How many non-trivial eigenvalues do we have? The first eigenvalue is denoted  $\lambda$ .
- 13) What are the rules for the choice of factorial components to retain?

- 14) The MCA could have been performed on the BURT table. According to give the value of the first eigenvalue that we would have obtained.

INDIVIDUS			COORDONNEES					CONTRIBUTIONS					COSINUS					CARRES				
IDENTIFICATEUR	P.REL	DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5					
01	3.23	6.10	2.10	-0.61	0.83	-0.70	0.35	4.9	0.7	2.3	2.1	1.0	0.73	0.06	0.11	0.08	0.02	02	3.23	7.79	1.90	-1.57
-1.04	-0.36	-0.02	4.0	4.4	3.7	0.5	0.0	0.47	0.32	0.14	0.02	0.00	03	3.23	8.83	-2.22	-1.35	1.36	-0.37	-0.09	5.5	3.2
0.6	0.1	0.56	0.21	0.21	0.02	0.00	04	3.23	14.45	-2.72	-2.33	-0.78	0.68	-0.67	8.2	9.6	2.1	2.0	3.8	0.51	0.38	0.04
0.03	05	3.23	7.99	0.73	-2.47	0.86	-0.56	0.03	0.6	10.7	2.5	1.3	0.0	0.07	0.76	0.09	0.04	0.00	06	3.23	1.67	1.19
0.29	-0.22	-0.18	-0.13	1.6	0.1	0.2	0.1	0.1	0.85	0.05	0.03	0.02	0.01	07	3.23	9.13	2.17	-0.48	-0.93	-1.79	-0.33	5.2
0.4	2.9	13.5	0.9	0.51	0.03	0.09	0.35	0.01	08	3.23	3.92	0.11	-0.20	0.16	-1.95	-0.01	0.0	0.1	0.1	16.0	0.0	0.00
0.01	0.97	0.00	09	3.23	8.25	2.39	0.92	0.25	-1.01	-0.78	6.3	1.5	0.2	4.3	5.1	0.69	0.10	0.01	0.12	0.07	10	3.23
0.01	-0.87	-3.54	0.18	-0.02	-0.08	0.8	22.1	0.1	0.0	0.1	0.05	0.89	0.00	0.00	0.00	11	3.23	2.99	-1.19	-0.59	-0.13	0.29
0.01	1.6	0.6	0.1	0.4	8.1	0.47	0.12	0.01	0.03	0.32	12	3.23	18.66	4.03	0.36	0.89	0.69	-0.84	18.0	0.2	2.7	2.0
0.01	0.04	0.03	0.04	13	3.23	3.04	0.48	0.03	-1.19	0.56	-1.00	0.3	0.0	4.8	1.3	8.4	0.08	0.00	0.46	0.10	0.33	14
2.55	-1.12	0.41	0.73	-0.09	-0.71	1.4	0.3	1.8	0.0	4.2	0.49	0.07	0.21	0.00	0.20	15	3.23	3.28	-1.01	0.35	0.79	0.39
0.01	1.1	0.2	2.1	0.6	6.8	0.31	0.04	0.19	0.05	0.25	16	3.23	6.05	0.04	-1.45	0.35	1.82	0.67	0.0	3.7	0.4	13.9
0.02	0.55	0.07	17	3.23	3.25	0.36	1.39	-0.81	0.04	-0.03	0.1	3.4	2.3	0.0	0.0	0.04	0.60	0.20	0.00	0.00	18	3.23
-1.14	0.61	0.47	0.34	0.07	1.4	0.7	0.8	0.5	0.0	0.61	0.18	0.10	0.06	0.00	19	3.23	5.35	-0.86	0.23	2.07	-0.28	-0.01
0.1	14.7	0.3	0.0	0.14	0.01	0.80	0.02	0.00	20	3.23	4.49	-0.24	0.25	-0.99	-1.48	1.03	0.1	0.1	3.4	9.2	8.8	0.01
0.22	0.49	0.24	21	3.23	8.72	1.46	2.38	-0.47	0.67	0.44	2.4	10.0	0.8	1.9	1.6	0.24	0.65	0.03	0.05	0.02	22	3.23
-0.54	1.81	0.29	-1.20	0.81	0.3	5.8	0.3	6.1	5.5	0.05	0.56	0.01	0.25	0.11	23	3.23	2.94	-1.22	0.78	0.29	0.50	0.22
1.1	0.3	1.0	0.4	0.51	0.21	0.03	0.08	0.02	24	3.23	14.62	-3.70	0.52	-0.07	-0.81	-0.07	15.2	0.5	0.0	2.8	0.0	0.93
0.00	0.05	0.00	25	3.23	2.91	-0.25	0.89	-0.86	1.01	-0.50	0.1	1.4	2.5	4.3	2.1	0.02	0.27	0.25	0.35	0.09	26	3.23
0.89	1.92	2.33	1.14	-0.45	0.9	6.5	18.5	5.4	1.7	0.07	0.32	0.47	0.11	0.02	27	3.23	6.39	1.71	-0.37	-1.12	0.95	1.09
3.2	0.2	4.3	3.8	9.9	0.46	0.02	0.20	0.14	0.18	28	3.23	18.22	-3.39	2.22	-1.32	-0.05	0.02	12.8	8.7	5.9	0.0	0.0
0.27	0.10	0.00	0.00	29	3.23	6.14	1.12	-1.29	-0.82	1.04	1.20	1.4	2.9	2.3	4.5	12.1	0.21	0.27	0.11	0.18	0.24	30
4.47	-0.11	0.49	-1.74	0.47	-0.83	0.0	0.4	10.4	0.9	5.7	0.00	0.05	0.68	0.05	0.15	31	3.23	1.18	-0.12	0.42	0.62	0.25
0.70	0.0	0.3	1.3	0.3	4.1	0.01	0.15	0.33	0.05	0.41												

## ANALYSE EN COMPOSANTES PRINCIPALES STATISTIQUES SOMMAIRES DES VARIABLES CONTINUES

NUM	IDEN	LIBELLE	EFFECTIF	POIDS	MOYENNE	ECART-TYPE	MINIMUM	MAXIMUM	2	V1	AGE	28	28.00
48.04	5.02	38.00	57.00	4	V2	OXYGEN	28	28.00	47.02	4.37	39.20	59.57	

3

5	.	V3 - RUNTIME	28	28.00	10.59	1.18	8.17	13.08	6	.	V4 - RESTPULSE	28	28.00	53.75	7.73	40.00	70.00
7	.	V5 - RUNPULSE	28	28.00	169.89	9.11	148.00	186.00	8	.	V6 - MAXPULSE	28	28.00	173.36	7.75	155.00	188.00
3	.	V7 - WEIGHT	28	28.00	77.16	8.26	59.08	91.63									

# MATRICE DES CORRELATIONS

	V1	V2	V3	V4	V5	V6
V1	1.00					
V2	-0.28	1.00				
V3	0.21	-0.83	1.00			
V4	-0.21	-0.37	0.45	1.00		
V5	-0.35	-0.29	0.13	0.33	1.00	
V6	-0.37	-0.20	0.06	0.33	0.95	1.00

# VALEURS PROPRES

NUMERO	VALEUR	POURCENTAGE	POURCENTAGE	PROPRE	CUMULE
1	2.6121	43.53	43.53	2	1.9187
2	1.9187	31.98	75.51	3	0.8280
3	0.8280	13.80	89.31	4	0.4547
4	0.4547	7.58	96.89	5	0.3008
5	0.3008	0.07	99.96	6	0.0000

# COORDONNEES DES VARIABLES SUR LES AXES 1 A 5

## VARIABLES ACTIVES

VARIABLES	COORDONNEES	CORRELATIONS	VARIABLE-FACTEUR	ANCIENS AXES UNITAIRES
IDEN - LIBELLE COURT	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
V1 - AGE	-0.26 0.71 0.50 0.42 0.05	-0.26 0.71 0.50 0.42 0.05	-0.16 0.51 0.55 0.63 0.12	V2 - OXYGEN
V2 - OXYGEN	-0.66 -0.67 -0.14 0.19 0.26	-0.41 -0.48 -0.15 0.28 0.70	0.59 0.72 -0.12 -0.23 0.25	V3 - RUNTIME
V3 - RUNTIME	0.59 0.72 -0.12 -0.23 0.25	0.36 0.52 -0.13 -0.35 0.68	0.68 0.11 -0.59 0.42 -0.04	V4 - RESTPULSE
V4 - RESTPULSE	0.68 0.11 -0.59 0.42 -0.04	0.42 0.08 -0.65 0.62 -0.10	0.51 -0.31 0.36 0.04 0.02	V5 - RUNPULSE
V5 - RUNPULSE	0.42 0.08 -0.65 0.62 -0.10	0.82 -0.43 0.33 0.03 0.01	0.82 -0.43 0.33 0.03 0.01	V6 - MAXPULSE
V6 - MAXPULSE	0.82 -0.43 0.33 0.03 0.01	0.79 -0.50 0.30 0.08 0.07	0.49 -0.36 0.33 0.12 0.18	

## VARIABLES ILLUSTRATIVES

VARIABLES	COORDONNEES	CORRELATIONS	VARIABLE-FACTEUR	ANCIENS AXES UNITAIRES
-----------	-------------	--------------	------------------	------------------------

IDEN - LIBELLE COURT	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
V7 - WEIGHT	0.12	-0.05	-0.05	-0.20	-0.16	0.12	-0.05	-0.05	-0.20	-0.16					

COORDONNEES ET VALEURS-TEST DES MODALITES AXES 1 A 5

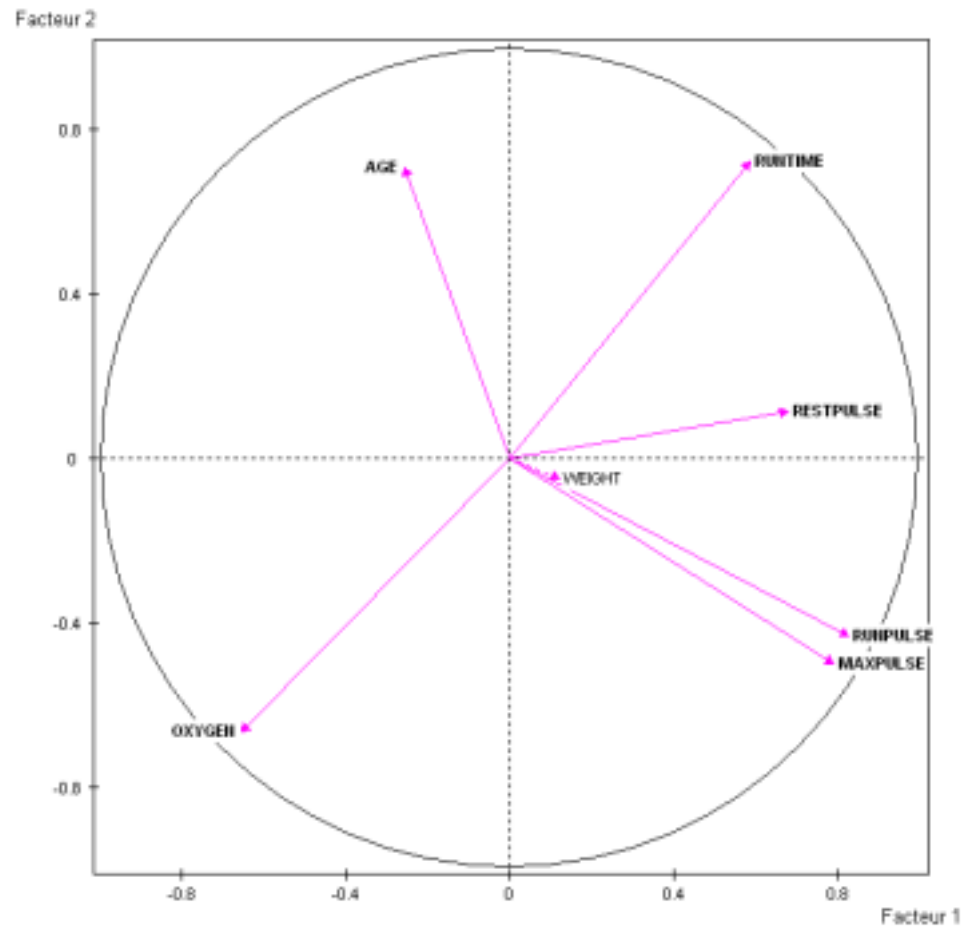
MODALITES					VALEURS-TEST					COORDONNEES										
IDEN	-	LIBELLE	EFF.	P.ABS	1	2	3	4	5	1	2	3	4	5	DISTO.					
1	.	CLASSE_AGE	-	C1 9 9.00	0.7	-2.6	2.9	-1.8	0.3	0.34	-1.01	-0.73	-0.33	0.03	1.79	-	C2 8 8.00	0.2	-0.7	0.7
0.0	0.2	0.10	-0.30	0.20	0.00	0.02	0.15	-	C3 11 11.00	-0.9	3.2	2.1	1.7	-0.5	-0.35	1.05	0.45	0.27	-0.05	1.51

COORDONNEES, CONTRIBUTIONS ET COSINUS CARRES DES INDIVIDUS AXES 1 A 5

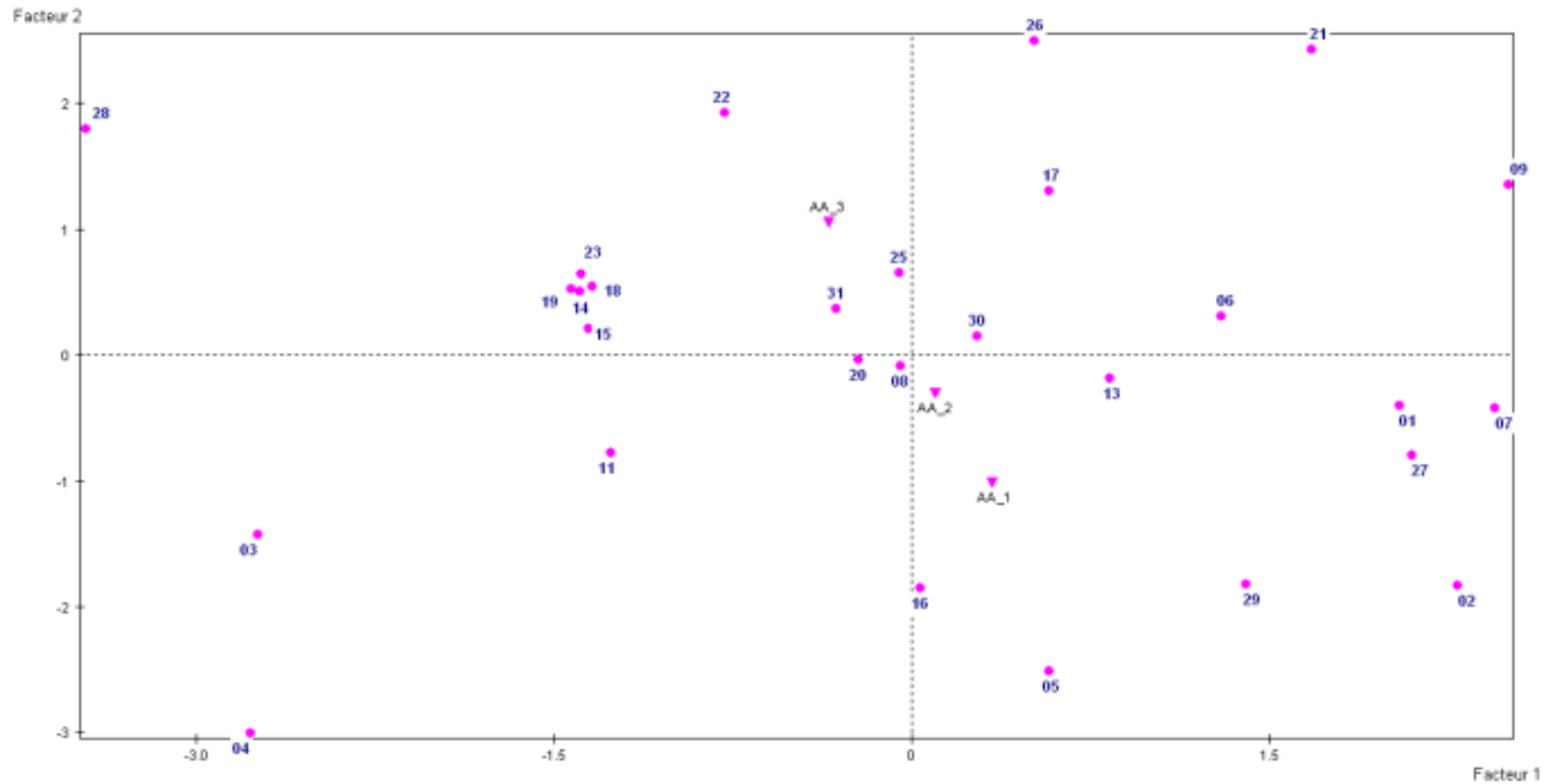
[illegible]

5

```
| 28 3.57 15.92 | -3.46 1.79 -0.32 0.78 -0.01 | 16.4 6.0 0.4 4.8 0.0 | 0.75 0.20 0.01 0.04 0.00 | | 29 3.57 7.98 | 1.40 -1.82 1.25 1.06 -0.05 | 2.7
6.2 6.8 8.8 0.1 | 0.25 0.42 0.20 0.14 0.00 | | 30 3.57 0.80 | 0.27 0.15 0.13 -0.31 0.74 | 0.1 0.0 0.1 0.8 13.9 | 0.09 0.03 0.02 0.12 0.69 | | 31
3.57 0.68 | -0.32 0.37 0.44 0.47 0.09 | 0.1 0.2 0.8 1.8 0.2 | 0.15 0.20 0.28 0.33 0.01 |
+-----+-----+-----+-----+-----+
```



6



7

# CLASSIFICATION HIERARCHIQUE SUR LES 6 AXES FACTORIELS

## DESCRIPTION DES NOEUDS

NUM. AINE BENJ EFF. POIDS INDICE HISTOGRAMME DES INDICES DE NIVEAU

```

29 17 12 2 2.00 0.00280 *
30 21 16 2 2.00 0.00592 *
31 28 22 2 2.00  *
32 11 6 2 2.00 0.02079 *
33 26 24 2 2.00 0.02787 **
34 27 32 3 3.00 0.02913 **
35 7 1 2 2.00  **

```

```

36 30 29 4 4.00 0.03985 **
37 18 8 2 2.00 0.04627 ***
38 13 31 3 3.00 0.05693 ***
39 5 2 2 2.00 0.06410 ***
40 4 3 2 2.00 0.06551 ***
41 36 10 5 5.00 [REDACTED] ****
42 19 15 2 2.00 0.07617 ****
43 9 35 3 3.00 0.09463 *****
44 23 42 3 3.00 0.09800 *****
45 33 14 3 3.00 [REDACTED] *****
46 20 37 3 3.00 0.12011 *****
47 41 38 8 8.00 0.12532 *****
48 34 43 6 6.00 0.20003 *****
49 45 39 5 5.00 0.23328 *****
50 47 25 9 9.00 0.29144 *****
51 50 46 12 12.00 0.36873 *****
52 44 48 9 9.00 0.42957 *****
53 51 40 14 14.00 0.65906 *****
54 49 52 14 14.00 0.88042 *****
55 53 54 28 28.00 1.82388 *****

```

# DECOMPOSITION DE L'INERTIE

```

+-----+
| | INERTIES | EFFECTIFS | POIDS | DISTANCES |
| INERTIES | AVANT APRES | AVANT APRES | AVANT APRES | AVANT APRES |
+-----+
| | | | | |
| INTER-CLASSES | [REDACTED] [REDACTED] | | | |
| | | | |
| INTRA-CLASSE | | | | |
| | | | |
| CLASSE 1 / 4 | 0.9862 0.9862 | 9 9 | 9.00 9.00 | 2.5666 2.5666 |
| CLASSE 2 / 4 | 0.4338 0.4338 | 5 5 | 5.00 5.00 | 5.4173 5.4173 |
| CLASSE 3 / 4 | 0.0655 0.0655 | 2 2 | 2.00 2.00 | 12.8058 12.8058 |
| CLASSE 4 / 4 | 1.1511 1.1511 | 12 12 | 12.00 12.00 | 1.5314 1.5314 |
| | | | |
+-----+

```

# COORDONNEES ET VALEURS-TEST APRES CONSOLIDATION AXES 1 A 5

CLASSES					VALEURS-TEST					COORDONNEES					
IDEN	-	LIBELLE	EFF.	P.ABS	1	2	3	4	5	1	2	3	4	5	DISTO.



```

| Coupure 'a' de l'arbre en 4 classes | | | aala - CLASSE 1 / 4 9 9.00 | 3.0 2.0 -0.2 -1.8 1.2 | 1.35 0.78 -0.06 -0.34 0.12 |
2.57 | | aa2a - CLASSE 2 / 4 5 5.00 | 1.9 -3.1 1.9 1.3 -1.4 | 1.28 -1.77 0.70 0.36 -0.22 | 5.42 | | aa3a - CLASSE 3 / 4 2 2.00 |
-2.5 -2.3 -0.3 -0.7 1.0 | -2.75 -2.22 -0.21 -0.35 0.26 | 12.81 | | aa4a - CLASSE 4 / 4 12 12.00 | -3.0 1.7 -1.0 1.1 -0.5 | -1.09
0.52 -0.21 0.16 -0.05 | 1.53 |
+-----+-----+-----+-----+-----+

```

# PARANGONS

CLASSE 1/ 4

EFFECTIF: 9

```

-----
|RK | DISTANCE | IDENT. ||RK | DISTANCE | IDENT. ||RK | DISTANCE | IDENT. |
+---+-----+-----+---+-----+-----+---+-----+-----+
| 1| 0.35191|06 || 2| 1.64372|13 || 3| 1.93441|17 |
| 4| 1.99440|30 || 5| 2.11036|01 || 6| 2.77116|09 |
| 7| 4.43489|21 || 8| 5.84695|07 || 9| 6.52595|26 |
+---+-----+-----+---+-----+-----+---+-----+-----+

```

CLASSE 2/ 4

EFFECTIF: 5

```

-----
|RK | DISTANCE | IDENT. ||RK | DISTANCE | IDENT. ||RK | DISTANCE | IDENT. |
+---+-----+-----+---+-----+-----+---+-----+-----+
| 1| 0.84315|29 || 2| 2.43767|02 || 3| 2.44223|27 |
| 4| 3.14665|16 || 5| 3.27622|05 || | | |
+---+-----+-----+---+-----+-----+---+-----+-----+

```

CLASSE 3/ 4

EFFECTIF: 2

```

-----
|RK | DISTANCE | IDENT. ||RK | DISTANCE | IDENT. ||RK | DISTANCE | IDENT. |
+---+-----+-----+---+-----+-----+---+-----+-----+
| 1| 0.91721|04 || 2| 0.91721|03 || | | |
+---+-----+-----+---+-----+-----+---+-----+-----+

```

CLASSE 4/ 4

EFFECTIF: 12

```

-----
|RK | DISTANCE | IDENT. ||RK | DISTANCE | IDENT. ||RK | DISTANCE | IDENT. |
+---+-----+-----+---+-----+-----+---+-----+-----+
| 1| 0.60790|18 || 2| 0.77253|19 || 3| 1.11533|14 |
| 4| 1.15094|31 || 5| 1.24812|23 || 6| 1.67260|15 |
| 7| 2.52033|25 || 8| 3.07113|11 || 9| 3.48443|22 |
| 10| 4.23489|20 || | | | | |
+---+-----+-----+---+-----+-----+---+-----+-----+

```

DESCRIPTION DE PARTITION(S) DESCRIPTION DE LA Coupure 'a' de l'arbre en 4 classes

CARACTERISATION DES CLASSES PAR LES MODALITES

CARACTERISATION DES CLASSES PAR LES CONTINUES

CARACTERISATION PAR LES MODALITES DES CLASSES OU MODALITES DE Coupure 'a' de l'arbre en 4 classes

CLASSE 1 / 4

V.TEST PROBA ---- POURCENTAGES ---- MODALITES IDEN POIDS CLA/MOD MOD/CLA GLOBAL CARACTERISTIQUES DES VARIABLES

CLASSE 2 / 4

V.TEST PROBA ---- POURCENTAGES ---- MODALITES IDEN POIDS CLA/MOD MOD/CLA GLOBAL CARACTERISTIQUES DES VARIABLES

CLASSE 3 / 4

V.TEST PROBA ---- POURCENTAGES ---- MODALITES IDEN POIDS CLA/MOD MOD/CLA GLOBAL CARACTERISTIQUES DES VARIABLES

CLASSE 4 / 4

V.TEST PROBA ---- POURCENTAGES ---- MODALITES IDEN POIDS CLA/MOD MOD/CLA GLOBAL CARACTERISTIQUES DES VARIABLES

CARACTERISATION PAR LES CONTINUES DES CLASSES OU MODALITES DE Coupure 'a' de l'arbre en 4 classes

CLASSE 1 / 4

V.TEST	PROBA	MOYENNES	ECARTS TYPES	VARIABLES CARACTERISTIQUES	CLASSE GENERALE	CLASSE GENERAL	LIBELLE IDEN
CLASSE 1 / 4 ( POIDS = 9.00 EFFECTIF = 9 )	aa1a	3.95	0.000	11.90 10.59	0.74 1.18	RUNTIME	-3.32
0.000		42.96	47.02	3.07	4.37	OXYGEN	

10

CLASSE 2 / 4

V.TEST	PROBA	MOYENNES	ECARTS TYPES	VARIABLES CARACTERISTIQUES	CLASSE GENERALE	CLASSE GENERAL	LIBELLE IDEN
CLASSE 2 / 4 ( POIDS = 5.00 EFFECTIF = 5 )	aa2a	3.70	0.000	185.20 173.36	2.93 7.75	MAXPULSE	3.49
183.00 169.89		3.35 9.11	RUNPULSE				0.000

CLASSE 3 / 4

V.TEST	PROBA	MOYENNES	ECARTS TYPES	VARIABLES CARACTERISTIQUES	CLASSE GENERALE	CLASSE GENERAL	LIBELLE IDEN
CLASSE 3 / 4 ( POIDS = 2.00 EFFECTIF = 2 )	aa3a	3.27	0.001	56.93 47.02	2.64 4.37	OXYGEN	-2.66

```

0.004 | 8.41 10.59 | 0.24 1.18 | RUNTIME |
+-----+-----+-----+-----+-----+-----+-----+-----+
CLASSE 4 / 4
+-----+-----+-----+-----+-----+-----+-----+-----+
| V.TEST | PROBA | MOYENNES | ECARTS TYPES | VARIABLES CARACTERISTIQUES | | | CLASSE GENERALE | CLASSE GENERAL | LIBELLE IDEN |
+-----+-----+-----+-----+-----+-----+-----+-----+
| CLASSE 4 / 4 ( POIDS = 12.00 EFFECTIF = 12 ) aa4a | | | | | | | -3.24 | 0.001 | 163.33 169.89 | 6.24 9.11 | RUNPULSE | |
-3.79 | 0.000 | 166.83 173.36 | 4.45 7.75 | MAXPULSE |
+-----+-----+-----+-----+-----+-----+-----+-----+

```