

I Questions de cours (4 points)

1. Quel est le nom du package de R permettant de réaliser une analyse en composantes principales (ACP) ?
2. Quelle est le nom de la fonction permettant de réaliser une ACP sous R ?
3. A quel type de données est adapté une analyse factorielle des correspondances (AFC) ?
4. Quel est le nombre maximal de nouveaux axes que l'on peut construire lorsque l'on réalise une AFC ?

II Commentaires de sorties de logiciels (16 points)

On dispose de données concernant 48 salariés d'une entreprise pour lesquels on a relevé les 6 caractéristiques suivantes : NIVETUD (nombre d'années passées à l'école), SALDEB (salaire annuel à l'embauche en dollars), SALACT (salaire annuel actuel en dollars), TEMPS (ancienneté dans l'entreprise en nombre de mois), EXP (expérience passée en dehors de l'entreprise en nombre de mois) et AGE (en années).

A. On réalise une ACP avec R sur ces données (voir sorties ci-après).

1. Calculer les valeurs A, B et C qui ont été supprimées et préciser de façon succincte à quoi correspondent ces valeurs.

2. On souhaite conserver 85% de l'inertie. Donner, en le justifiant, le nombre de composantes principales à retenir pour cette ACP.

3. Interpréter les composantes principales retenues à l'aide des variables initiales.

4. Interpréter le salarié 2 sur le premier axe et le salarié 1 sur le deuxième axe après avoir vérifié que ces salariés sont bien représentés sur les axes considérés.

B. On réalise une AMM (sur les variables standardisées) en 6 groupes avec SAS des 48 salariés à l'aide des 6 variables précédentes.

1. Pourquoi faut-il standardiser les variables ?

2. Quels sont les 2 groupes les plus proches? Justifier.

3. Justifier le nombre de groupes choisis pour cette classification.

4. Faire la typologie de chaque groupe en utilisant les variables initiales.

Sorties R (ACP)

	eigenvalue	percentage of variance	cumulative percentage of variance	percentage of variance
comp 1	2.3814207		B	39.69034
comp 2	1.8272101	30.453501		C
comp 3	A	16.066724		86.21057
comp 4	0.4892824	8.154706		94.36528

```

comp 5 0.2330477          3.884128          98.24940
comp 6 0.1050358          1.750596          100.00000

```

```

> round(resu_acp_empred$var$cor,2)
      Dim.1 Dim.2 Dim.3 Dim.4 Dim.5
SALDEB  0.92  0.20  0.03 -0.24 -0.06
TEMPS   -0.04 -0.31  0.95 -0.04  0.05
AGE     -0.02  0.89  0.23  0.25 -0.30
SALACT  0.94 -0.06  0.02 -0.24 -0.08
NIVETUD 0.80 -0.17 -0.01  0.55  0.15
EXP      0.01  0.93  0.08 -0.11  0.33

```

```

> resu_acp_empred$ind$coord
      Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
1 -0.52182673  1.57484559  1.54137819 -0.10006110  0.588541336
2  2.23109488 -0.79304252 -1.70208683 -0.13319834 -0.265138343

```

```

> resu_acp_empred$ind$cos2
      Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
1  0.0484827311 4.415810e-01 4.230122e-01 1.782646e-03 6.167205e-02
2  0.5242550498 6.623682e-02 3.051195e-01 1.868547e-03 7.403748e-03

```

Sorties SAS (AMM)

Statistiques pour variables

Variable	E-t	total	Dans E-t	R-carré	RSQ/(1-RSQ)
SALDEB	1.00000	0.42123	0.841440	5.306763	
TEMPS	1.00000	0.97237	0.155074	0.183535	
AGE	1.00000	0.40861	0.850797	5.702258	
SALACT	1.00000	0.35837	0.885237	7.713573	
NIVETUD	1.00000	0.79152	0.440143	0.786169	
EXP	1.00000	0.56270	0.717049	2.534185	
OVER-ALL	1.00000	0.62736	0.648290	1.843250	

Distance Between Cluster Centroids

	1	2	3	4	5	6
1	.	1.742098330	3.714649545	7.196802699	5.831076894	2.518690522
2	1.742098330	.	3.936383180	7.150771263	5.316117967	3.201638689
3	3.714649545	3.936383180	.	4.577649510	4.376541809	2.613043373
4	7.196802699	7.150771263	4.577649510	.	3.648698642	6.632764197
5	5.831076894	5.316117967	4.376541809	3.648698642	.	6.036458239
6	2.518690522	3.201638689	2.613043373	6.632764197	6.036458239	.

Cluster Means

Cluster	SALDEB	TEMPS	AGE	SALACT	NIVETUD	EXP	Effectif
1	4730	81.33	52.17	8250	11.83	10.17	6
2	5870	75.17	48.5	10170	11.33	21.83	6
3	10058	75.57	22.71	22260	16.43	4.57	7
4	18000	96	40	44250	19	10	1
5	21000	83	50	26700	16	22	1
6	5582	83.63	29.22	11494	12.48	3.63	27