

TABLE DE MATIERE

1.	INTRODUCTION	6
2.	ASPECT THÉORIQUE	7
2.1	DONNÉES, NOTATIONS, HYPOTHÈSE D'INDÉPENDANCE	7
2.2	OBJECTIFS	9
2.3	TRANSFORMATIONS DES DONNÉES EN PROFILS	10
2.4	RESSEMBLANCE ENTRE PROFILS : DISTANCE DU x^2	11
2.5	LES DEUX NUAGES	12
2.5.1	Nuage des profils-lignes	12
2.5.2	Nuage des profils-colonnes	13
2.6	AJUSTEMENT DES DEUX NUAGES	14
2.6.1	Ajustement du nuage des profils-lignes	14
2.6.2	Ajustement du nuage des profils-colonnes	15
2.7	UN ASPECT TECHNIQUE DU CENTRAGE EN AFC	15
2.8	LA DUALITÉ	16
2.9	DUALITE ENTRE LES FACTEURS SUR I ET LES FACTEURS SUR J	17
2.10	REPRESENTATION SIMULTANEE DES LIGNES ET DES COLONNES ; RELATIONS BARYCENTRIQUES	19
2.11	NOMBRE D'AXES ET INERTIE TOTALE	20
2.12	AIDES À L'INTERPRÉTATION ET ÉLÉMENTS SUPPLÉMENTAIRES	20
3.	L'ASPECT PRATIQUE	21
3.1	ORIGINE DE DONNES	21
3.1.1	Objectif de cette étude	21
3.2	LES TECHNOLOGIE UTILISER POUR APPLIQUER AFC	21
3.3	APPLIQUER L'AFC PAR R	22
3.3.1	Packages utilisée pour calculer l'afc	22
3.3.2	Description de donnes	23
3.3.3	Visualiser le tableau de contingence en diagramme a bulle	23
3.3.4	Test de khi2	25
3.3.5	Calculer l'afc	26
3.3.6	Valeurs propres variances	27
3.3.7	Le graphique des valeurs propres avec une droite en pointillee	28
3.3.8	Le biplot des lignes et des colonnes	28

3.4	GRAPHIQUE DES POINTS LIGNES.....	29
3.4.1	Les coordonnées des lignes.....	30
3.4.2	Qualite de representation des lignes	31
3.5	GRAPHIQUE DES POINTS COLONNES	32
	Donner le résultat suivant :.....	33
3.5.1	Les coordonnées des colonnes.....	33
3.5.2	Qualite de representation des colonnes	34
3.6	APPLIQUER L’AFC PAR EXCEL.....	35
3.6.1	L’utilisation de excel	36
3.6.2	Test de khi2	37
3.6.3	Les valeurs propres.....	37
3.6.4	Screeplot.....	38
3.6.5	Resultats pour lignes	38
3.6.6	Resultat pour les colonnes	40
3.6.7	Graphique symétrique.....	42
3.7	APPLIQUER AFC PAR PYTHON	43
3.7.1	Charger et afficher les donnees	43
3.7.2	Test de khi2	44
3.7.3	AFC en python	44
3.7.4	Les valeurs propres.....	45
3.7.5	Résultat des lignes.....	46
3.7.6	Résultat des colonnes.....	48
3.7.7	Représentation simultanée	50
4.	INTERPRETATION.....	51
4.1	PROFILS LIGNES	51
4.2	PROFILS COLONNES.....	51
4.3	RECAPITULATIF	52
4.4	PRESENTATION DES POINTS DE LIGNE	52
4.5	PRESENTATION DES POINTS DE COLONNES.....	53
4.6	GRAPHE DES POINTS DE LIGNES ET DE COLONNES.....	54
5.	CONCLUSION	55

TABLE DE FIGURES

Figure 1 : Tableau des données brutes	7
Figure 2 : Tableau F des fréquences relatives et ses marges	8
Figure 3 : Profil-ligne (à gauche) et profil-colonne (à droite). GI et GJ : profils marginaux	10
Figure 4 : la distance de khi2	11
Figure 5 : Le nuage NI des profils-lignes dans RJ	12
Figure 6 : Le nuage NJ des profils-colonnes dans RI	13
Figure 7 : Ajustement dans RJ du nuage des profils-lignes	14
Figure 8 : Le premier axe factoriel du nuage NI non centré est le facteur trivial OGI orthogonal	16
Figure 9 : Analyse par rapport au barycentre et par rapport à l'origine	16
Figure 10 : La statuesque X2	17
Figure 11 : Inertie total de NI rapport a GI	17
Figure 12 : Les formules de transition	18
Figure 13 : Tableau de donnas	21
Figure 14 : Les technologies utilises	22
Figure 155 Installer packages Facto Miner et facto extra	22
Figure 16 : Charger les packages Facto Miner et facto extra	23
Figure 17 : la fonction summary ()	23
Figure 18 : code le tableau de contingence en diagramme a bulle	24
Figure 19 : tableau de contingence en diagramme a bulle	24
Figure 20 :colorRampPalette()	24
Figure 21 : tableau de contingence en diagramme a bulle selon effectif	25
Figure 22 : Test de khi2	25
Figure 23 :CA ()	26
Figure 24 : Calculer AFCpar CA ()	26
Figure 25 : Resultat de AFC	27
Figure 26:get_eigenvalue()	27
Figure 27 : Valeurs propres variances en R	27
Figure 28 :la fonction fviz_screplot ()	28
Figure 29 : Le graphe des valeurs propres	28
Figure 30 : la fonction fviz_ca_biplot ()	29
Figure 31 : le biplot de lignes et colonnes	29
Figure 32:get_ca_row()	29
Figure 33 :la fonction head()	30
Figure 34 : Les coordonnes des lignes	30
Figure 35 : fviz_ca_row	30
Figure 36 : Representation des lignes	31
Figure 37 : la fonction head()	31
Figure 38 : La qualite des points lignes	31
Figure 39 : La fonction fviz_ca_row	32
Figure 40 : Qualite de representation	32
Figure 41 : get_ca_col()	32
Figure 42 : Les coordonnes des colonnes	33

Figure 43 : la Representation des colonnes	33
Figure 44 : la fonction fviz_ca_col()	33
Figure 45 : la fonction head()	34
Figure 46 : Qualité de représentation des colonnes	34
Figure 47 : La fonction fviz_ca_col()	34
Figure 48 : La representationn des colonnes selon la qualite	35
Figure 49 : L'UTILISATION DE EXCEL	35
Figure 50 : L'UTILISATION DE EXCEL 1	36
Figure 51 : Test de khi2	37
Figure 52 : Les valeurs propres par excel	37
Figure 53 : la Distance et le poids pour les lignes	38
Figure 54 : Les coordonnes des lignes	39
Figure 55 : Contributions des lignes	39
Figure 56 : La distance et le poids des colonnes	40
Figure 57 : Les coordonnes des colonnes	40
Figure 58 : Les contrubutions des colonnes	41
Figure 59 : Graphique symetrique	42
Figure 60 : Charger et afficher les donnes	43
Figure 61 : Les donnes affichees	43
Figure 62 : Test de khi2	44
Figure 63 : Application de AFC	44
Figure 64 : Les resultat de AFC	44
Figure 65 : l'objet afc.eig	45
Figure 66 : Les valeurs propres	45
Figure 67 : la fonction plot_eigenvalues	45
Figure 68 : le graphe des valeurs propres	46
Figure 69 : L'objet row_coord	46
Figure 70 : Les coordonnees des points lignes	46
Figure 71 : L'objet row_contrib	47
Figure 72 : Les Contribution des points lignes	47
Figure 73 : L'objet row.cos2	47
Figure 74 : Les qualites des points lignes	48
Figure 75 : L'objet col_coord	48
Figure 76 : Les coordonnes des points lignes	48
Figure 77 : L'objet col_contrib	49
Figure 78 : Les contributions des points colonnes	49
Figure 79 : Les qualites des points colonnes	49
Figure 80 : Les qualités des points colonnes	50
Figure 81 : la fonction mapping()	50
Figure 82 : Représentation simultanée	50
Figure 83 : Profils lignes	51
Figure 84 : Profils colonnes	51
Figure 85 : Recapitulatif	52
Figure 86 : Presentation des points de ligne	52
Figure 87 : Presentation des points de colonne	53

Figure 88 : Graphe des points de lignes et de colonnes	54
--	----

1. INTRODUCTION

L'analyse des données : outil de connaissance dans les plus divers Depuis une trentaine d'années, les méthodes des données ont largement démontré leur efficacité dans l'étude grandes masses complexes d'informations. Ce sont des méthodes dites multidimensionnelles en opposition aux de la statistique descriptive qui ne traitent qu'une ou variables à la fois. Elles permettent donc la confrontation entre de nombreuses informations, qui est infiniment plus riche que leur examen séparé. Les représentations simplifiées de grands tableaux de données que méthodes permettent d'obtenir s'avèrent un outil de synthèse remarquable. De données trop nombreuses pour être appréhendées directement, elles les tendances les plus marquantes, les hiérarchisent et éliminent effets marginaux ou ponctuels qui perturbent la perception globale faits.

Nées à l'université, elles ont d'abord été connues essentiellement et appliquées aux domaines scientifiques comme l'écologie, la linguistique, l'économie, etc. Ils l'ont rendu accessible à des études nouvelles, plus riches, plus complexes. Mais leur champ d'application a longtemps dépassé ce cadre, d'autant plus que la collecte et le stockage de l'information ont été facilités par le développement de l'informatique. Dans tous les domaines, d'importants fichiers de données s'accumulent. Le premier objectif est de retenir l'information et de pouvoir s'y référer facilement. Mais on s'est vite rendu compte que pour miner tout le contenu dans ces fichiers, les collecter nécessitait souvent les bons outils statistiques. Il existe plusieurs méthodes adaptées à différents types de données : ainsi, pour citer les plus connues, l'analyse en composantes principales traite des tableaux croisant des individus et des variables quantitatives, l'analyse factorielle des correspondances traite des tableaux de fréquence et l'analyse des correspondances multiples s'applique à des tableaux croisant des individus et des variables qualitatives L'analyse factorielle des correspondances vise à rassembler la plupart des informations initiales dans un nombre réduit de dimensions en examinant les correspondances, ou les valeurs relatives, entre les variables plutôt que les valeurs absolues. Cette réduction est plus significative pour une dimensionnalité initiale plus élevée.

Le concept de "réduction" est commun à toutes les méthodes factorielles (c'est-à-dire lorsque des facteurs sont extraits). AFC offre la particularité (contrairement à ACP) de fournir un espace commun de représentation des variables et des individus. L'AFC justifie de réduire le tableau ou la fréquence pour cela.

2. ASPECT THÉORIQUE

2.1 DONNÉES, NOTATIONS, HYPOTHÈSE D'INDÉPENDANCE

À l'origine, l'analyse des correspondances factorielles (FCA) a été développée pour examiner des tableaux communément appelés tableaux de contingence (ou tableaux croisés). Il s'agit de tableaux de fréquences obtenus par croisement des modalités de deux variables qualitatives définies pour une même population de n personnes. Dans l'exemple décrit au chapitre 10, la population est constituée de tous ceux qui ont quitté le système scolaire français en 1972 et ont été embauchés en 1973. Nous connaissons le niveau de qualification et le domaine d'expertise de chacun. La figure 3.1 résume les notations les plus importantes. On parle indifféremment de modalité i (par exemple les bacheliers) ou de classe i , la classe des personnes de modalité i (par exemple les bacheliers). Dans ce chapitre, nous nous limitons à l'étude des tableaux de contingence. Cependant, la plupart des concepts introduits et les résultats présentés

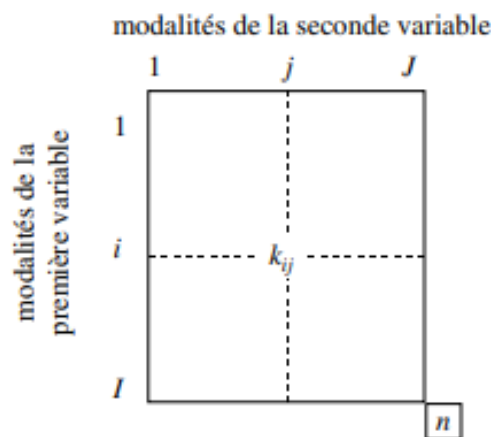


Figure 1 : Tableau des données brutes.

peuvent être généralisés à des tableaux qui ne sont pas strictement de ce

I : ensemble des lignes et nombre de lignes (8 niveaux de diplôme).

J : ensemble des colonnes et nombre de colonnes (9 catégories d'emploi).

K_{ij} : nombre d'individus possédant à la fois la modalité i de la première variable et la modalité j de la seconde (i.e. qui ont le niveau de diplôme i et qui occupent un emploi de la catégorie j). $\sum_i \sum_j k_{ij} = n$ (nombre total d'individus).

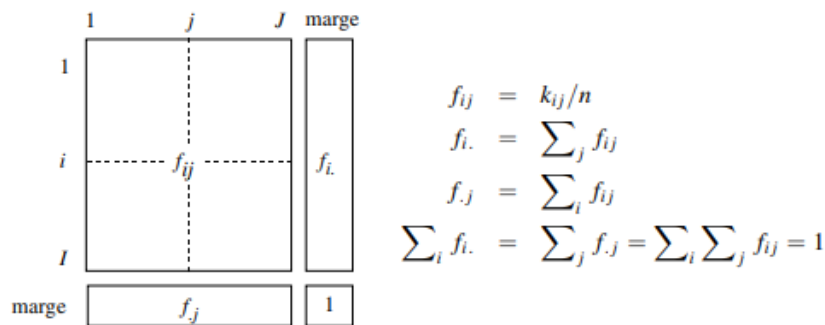


Figure 2 :Tableau F des fréquences relatives et ses marges

Analyse des correspondances multiples. À la fin de ce chapitre, nous fournissons quelques points de référence pour appliquer l'AFC à des tables autres que les tables de contingence. On voit souvent un tableau de fréquences relatives F obtenu en divisant chaque fréquence k_{ij} par la fréquence totale n . Ce nouveau tableau définit la mesure de probabilité pour l'intersection de l'ensemble $I \times J$. Sa marge ou probabilité marginale

Ont pour terme général $f_{i.}$. Pour les marges des colonnes et $f_{.j}$ pour les marges des lignes. Un tableau de contingence montre la relation entre deux attributs. Classiquement, pour une mesure de probabilité, on dit qu'il y a indépendance entre deux variables lorsque, pour tout i et pour tout j , on a l'égalité :

$$f_{ij} = f_{i.} f_{.j}$$

Il existe une relation entre deux variables d'une cellule du tableau f_{ij} autre du produit $f_{i.} f_{.j}$. Si f_{ij} est supérieur à ce produit, alors i et j sont plus liés qu'eux, rendant l'hypothèse indépendante : on dit que i et j A l'inverse, si f_{ij} est inférieur au produit de marge, i et j sont moins associés dans l'hypothèse : on dit qu'il y a répulsion entre ces méthodes. L'indépendance est également exprimée en traitant le tableau comme l'une des lignes. En effet, l'égalité ci-dessus est équivalente à l'égalité :

$$f_{ij} f_{i.} = f_{.j}$$

La quantité $f_{.j}$ représente le pourcentage de la population totale qui possède la modalité j tandis que $f_{ij}/f_{.i}$ Représente le même pourcentage dans la sous-population qui possède i . Lorsqu'elles sont indépendantes, les sous-populations I caractérisées par la forme i de la première variable sont distribuées selon les J j de la deuxième variable avec un pourcentage. Toutes les lignes suivantes sont proportionnelles. L'inverse est vrai : lorsque toutes les courbes sont proportionnelles, elles sont proportionnelles à la frontière $f_{.j}$ et les deux variables sont indépendantes. Ainsi, il y a liaison lorsque toutes ne sont pas proportionnelles à la marge, c'est-à-dire non identiques du point de vue de l'association avec toutes les colonnes. Remarquons enfin que, dans un tableau de contingence, les lignes et les colonnes jouent un rôle absolument symétrique : l'indépendance s'exprime de la même façon sur l'ensemble des colonnes. Les deux équations ci-dessus sont en fait équivalentes comme

Suit :

$$f_{ij} f_{.j} = f_{.i}$$

L'indépendance existe si tous les pourcentages d'une colonne sont égaux à la marge $f_{.i}$. Autrement dit, si les colonnes sont proportionnelles. Même quand ils ne le sont pas, il y a un lien.

2.2 OBJECTIFS

Les objectifs de l'AFC peuvent être exprimés de manière similaire à ceux de l'APC, bien que les tableaux explorés soient très différents de ceux explorés par l'APC. J'essaie d'obtenir la typologie des lignes, la typologie des colonnes et de relier ces deux typologies. Cependant, la notion de similarité entre deux lignes ou deux colonnes est différente de celle de l'ACP.

Dans un tableau de contingence, la similarité entre deux lignes d'une part et deux colonnes d'autre part s'exprime de manière totalement symétrique. Si deux lignes correspondent de manière identique à toutes les colonnes, c'est-à-dire si trop (ou trop peu) de lignes correspondent à la même colonne, ces lignes sont considérées comme serrées. Les termes « trop » et « trop peu » font référence à des situations indépendantes. Deux colonnes sont symétriquement adjacentes si elles sont également associées à toutes les lignes. Schématiquement, l'étude de l'ensemble des lignes revient à mettre en évidence une typologie dans laquelle on cherche les lignes dont la répartition s'écarte le plus de celle de l'ensemble de la population, celles qui se ressemblent entre elles (dans le sens précisé ci-dessus) et celles qui

s'opposent. Pour mettre en relation la typologie des lignes avec l'ensemble des colonnes, on caractérise chaque groupe de lignes par les colonnes auxquelles ce groupe s'associe trop ou trop peu. L'étude de l'ensemble des colonnes est absolument analogue. Cette approche, grâce à la notion de ressemblance utilisée, permet d'étudier la liaison entre les deux variables, c'est-à-dire l'écart du tableau à l'hypothèse d'indépendance. Cette analyse de liaison est l'objectif fondamental de l'AFC. Une approche complémentaire de la précédente, impliquant des lignes et des colonnes communes en ne priorisant ni l'une ni l'autre. Prenons l'exemple d'un croisement de types d'emplois et de qualifications. L'ensemble des diplômes est ordonné par la longueur des études tandis que celui des catégories d'emploi l'est par le salaire moyen. La relation entre ces deux ordres explique clairement une partie du lien entre les emplois et les diplômes. Mais ce lien ne se restreint peut-être pas à cet unique aspect ; D'autres phénomènes peuvent exister, comme l'association quasi exclusive de certains diplômes à certains emplois.

2.3 TRANSFORMATIONS DES DONNÉES EN PROFILS

Dans AFC, les tableaux bruts ne sont pas analysés directement. Dans l'étude des lignes, le tableau de données est divisé en divisant le terme f_{ij} de chaque ligne par la marge f_i de cette ligne i . La nouvelle ligne s'appelle profile-line .

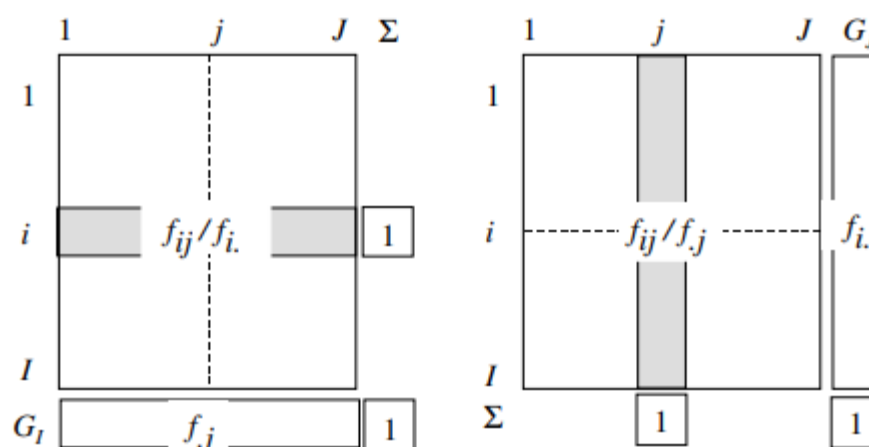


Figure 3: Profil-ligne (à gauche) et profil-colonne (à droite). G_i et G_j : profils marginaux

Cette transformation découle de l'objectif d'examiner la relation entre deux variables par la différence des pourcentages de ligne. Aussi, la comparaison de deux lignes du tableau brut est directement justifiée car elle risque d'être principalement influencée par leurs nombres

de marge. Ainsi, la différence entre les lignes brutes Bac technologie et Bac général du tableau Profession et Diplôme reflète essentiellement la différence de la somme de ces deux diplômes.

Quantité $f_{i.}/f_{i.}$. Dans cet exemple, il représente la probabilité de pourvoir un poste de la catégorie j étant donné que l'on sait que vous avez le niveau de diplôme i . Un profil de ligne i est juste une loi de probabilité conditionnelle définie par i sur toutes les colonnes. Pour analyser les différences d'indépendance, nous comparons ces profils avec le profil marginal du terme général $f_{.j}$ (=déterminé pour l'ensemble de la population et exprimé en GI). En raison du rôle symétrique des lignes dans les colonnes, un argument similaire peut être avancé pour les colonnes. Il conduit à la notion de profil-colonne. Donc, en AFC, selon que l'on s'intéresse aux lignes ou aux colonnes, on ne considère pas le même tableau. Cependant, les deux transformations en profils ont pour eux le même objet d'intérêt. Ces transformations sont intéressantes en elles-mêmes, indépendamment de toute analyse factorielle. Lorsqu'un tableau croisé est commenté, c'est presque toujours sous la forme d'un pourcentage, ligne à colonne, selon l'aspect que l'on souhaite mettre en évidence.

2.4 RESSEMBLANCE ENTRE PROFILS : DISTANCE DU χ^2

En AFC, la similarité entre deux lignes ou deux colonnes est définie par la distance entre les profils appelée distance χ^2 . Défini symétriquement pour les lignes et les colonnes.

Soit :

$$\begin{aligned} d\chi^2(\text{profil-ligne } i, \text{profil-ligne } l) &= \sum_j \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{lj}}{f_{l.}} \right)^2 \\ d\chi^2(\text{profil-colonne } j, \text{profil-colonne } k) &= \sum_i \frac{1}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ik}}{f_{.k}} \right)^2 \end{aligned}$$

Figure 4: la distance de χ^2

Dans ces relations, la distance entre deux lignes dépend essentiellement des différences terme à terme entre les deux profils dont elle fait une somme des carrés pondérés. La pondération $1/f_{.j}$ équilibre l'influence des colonnes sur la distance entre les lignes : elle augmente les termes, a priori plus faibles, concernant les modalités rares ; elle joue, jusqu'à un certain point, un rôle analogue à celui de la division par l'écart-type dans le cas des variables numériques. La distance du χ^2 jouit d'une propriété fondamentale appelée équivalence distributionnelle. Selon cette propriété, si deux colonnes proportionnelles d'un tableau sont

cumulées en une seule, la distance entre les profils-lignes est inchangée. Le cas d'une proportionnalité parfaite entre deux colonnes ne se rencontre guère en pratique mais constitue une situation limite dont on peut être assez proche. La propriété mathématique est alors utilisée sous la forme d'une règle pragmatique : remplacer, par leur somme, deux colonnes ou deux lignes presque proportionnelles ne modifie pas sensiblement les résultats d'une AFC. On se réfère surtout à cette règle lorsque plusieurs ensembles de modalités sont possibles pour définir une même variable. Ainsi, la variable catégorie d'emploi peut être plus ou moins détaillée : par exemple, on peut se demander si les catégories ouvrier qualifié et ouvrier non qualifié peuvent être regroupées en une seule catégorie. Du fait de l'équivalence distributionnelle, si ces deux catégories ont des profils voisins, le choix entre les deux solutions n'est pas fondamental puisque les AFC des deux tableaux aboutissent à des résultats analogues.

2.5 LES DEUX NUAGES

2.5.1 Nuage des profils-lignes

Concernant la modalité de la première variable, nous voyons les données comme des juxtapositions de lignes de profil. Chaque profil de ligne est une séquence de J nombres, qui peuvent être représentés par des points dans l'espace RJ dont les J dimensions sont chacune associées à la modalité de la deuxième variable. La distance x_2 (voir section 3.4) qui définit la similarité entre les lignes de profil a la propriété de distance euclidienne et donne à RJ la structure de l'espace euclidien. Cette distance affectera un poids $1/f_j$ à la j ème dimension de RJ . La somme des coordonnées de chaque profil de ligne est 1. Par conséquent, le nuage N_I appartient à l'hyperplan noté H_I .

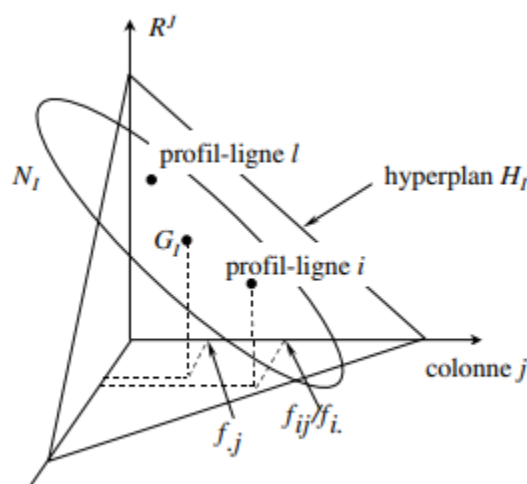


Figure 5: Le nuage N_I des profils-lignes dans R_J

L'AFC applique des pondérations attribuées à chaque point de nuage. Un point i a un poids égal à la fréquence limite $f_{i.}$ (Ce poids est proportionnel à la taille de la classe de personnes représentée par le point i). Notons G_I le barycentre des points de N_I avec ces poids. Sa j ème coordonnée est égale à la fréquence de marginale $f_{.j}$.

$$f_{.j} = \sum_i f_{i.} f_{ij} / f_{i.}$$

Interprété comme profil moyen. Dans l'exemple du tableau qui croise les niveaux de diplôme et les catégories d'emploi, G_I est le profil d'emplois de l'ensemble de la population, tous les diplômes étant cumulés. Il sert constamment de référence dans l'étude des lignes du tableau ; ainsi, étudier dans quelle mesure et de quelle façon une classe d'individus i diffère de l'ensemble de la population revient à étudier l'écart entre le profil de cette classe i et le profil moyen. Étudier la dispersion du nuage

autour de son barycentre revient à étudier l'écart entre les profils des lignes et le profil marginal, et donc la liaison entre les deux variables (cf. section 3.1).

2.5.2 Nuage des profils-colonnes

Compte tenu du rôle symétrique joué par les lignes et les colonnes en AFC, la construction du nuage des profils-colonnes s'effectue selon une démarche strictement identique à celle du nuage des profils-lignes. Il est toutefois utile de la décrire, ne serait-ce que pour fixer les notations. S'intéresser aux modalités de la seconde variable revient à considérer les données comme une juxtaposition de profils-colonnes. Chaque profil-colonne est une suite de I valeurs

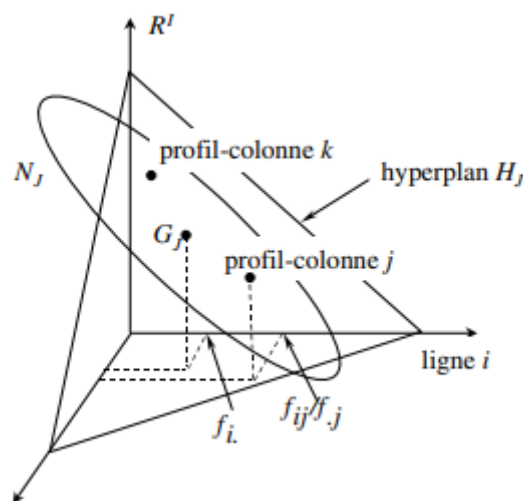


Figure 6: Le nuage N_J des profils-colonnes dans R^I

numériques et peut être représenté par un point de l'espace RI dont chacune des dimensions est associée à une modalité de la première variable. RI est muni d'une structure euclidienne par la distance du x_2 : à la i e dimension on affecte le poids $1/f_i$ (cf. Figure 3.5).

Le point GJ représente la marge $\{ f_{i.} | i = 1, \dots, I \}$; c'est le barycentre de NJ lorsque l'on munit chaque profil-colonne j du poids $f_{.j}$; en tant que profil moyen, il sert constamment de référence dans l'étude de NJ .

2.6 AJUSTEMENT DES DEUX NUAGES

2.6.1 Ajustement du nuage des profils-lignes

L'ajustement dans RJ vise à obtenir une séquence d'images grossièrement planes de nuages NI. Semblable à ACP, AFC recherche un ensemble d'axes orthogonaux sur lesquels les nuages NI sont projetés. Chaque axe a des propriétés qui maximisent l'inertie projetée du nuage NI, contraint d'être orthogonal aux axes déjà trouvés. Une image NI plane doit être telle que les distances entre les points de l'image soient aussi proches que possible des distances entre les points du NI. Cet objectif est très similaire à l'objectif de réglage des nuages de personnes dans ACP.

En pratique, cela signifie que le nuage analysé est centré. H. Leurs barycentres sont choisis comme origines des axes (cf. paragraphe 3.5). Dans le nuage central, la classe définie par la modalité i est représentée par un point dont la coordonnée est sur le jème axe f_{ij} . L'emplacement de ce point représente la différence entre la distribution des individus dans la classe i et la population totale dans toutes les modalités pour la seconde variable. Une recherche de la direction d'inertie maximale du nuage central mettra donc en évidence les classes qui s'écartent le plus du profil de la population totale. Chaque profil reçoit un poids égal à sa fréquence de bord $f_{i.}$. Ce poids affecte principalement les calculs du centroïde du nuage. Cela affecte également l'inertie et donc les critères de positionnement rencontrés par l'axe .

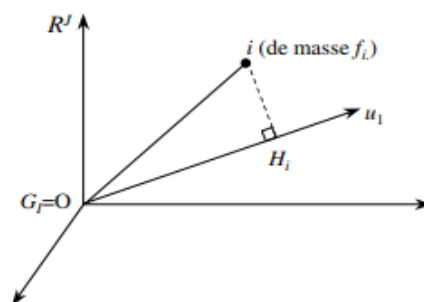


Figure 7: Ajustement dans RJ du nuage des profils-lignes

En introduisant les poids f_i . Dans les critères d'ajustement, chaque modalité a un poids proportionnel à la population qu'elle représente. Les axes de capacité tendent donc à mettre l'accent sur des phénomènes connexes avec la même disparité de profil. la majeure partie de la population totale. D'un autre point de vue, les quelques modalités susceptibles d'avoir des profils moins fiables n'interfèrent pas beaucoup avec la construction des essieux. En résumé, le réglage des nuages NI dans CA est similaire au réglage de nuages individuels dans PCA. Il en diffère sur les trois points suivants :

1. les lignes interviennent au travers de leur profil ;
2. la distance entre les profils est celle du ;
3. chaque ligne i est affectée du poids f_i .

2.6.2 Ajustement du nuage des profils-colonnes

Étant donné que les lignes et les colonnes jouent des rôles symétriques dans CFA, l'ajustement de NJ à RI se produit dans les mêmes conditions et a les mêmes propriétés que l'ajustement de NI à RJ. Nous les résumons ci-dessous.

1. Les images planes de NJ doivent être telles que la distance entre les profils projetés soit aussi proche que possible de la distance entre les profils de RI. En conséquence, les nuages NJ doivent être analysés par rapport au centroïde GJ. L'inertie globale de NJ par rapport à GJ provient des différences de profils des différentes classes j et de l'ensemble de la population.

2. Chaque colonne j se voit attribuer un poids égal à sa fréquence marginale $f_{.j}$. En utilisant une notation similaire à la figure 3.6, H_j est la projection du point j associé à la colonne de profil j sur v_1 (le vecteur unitaire du premier axe factoriel de RI), où v_1 maximise la magnitude. $j f_{.j}(\text{OH}_j)$ 2. Cette justification des poids $f_{.j}$ est exactement similaire à celle développée pour le profil de ligne.

2.7 UN ASPECT TECHNIQUE DU CENTRAGE EN AFC

D'un point de vue technique, on peut montrer que le Cloud NI n'a pas besoin d'être explicitement centralisé avant l'analyse (voir section 5.5, page 121). A l'exception du premier facteur, l'analyse des nuages par rapport à O sans centrage conduit aux mêmes facteurs que l'analyse des nuages centrés. Lorsque le nuage NI CA décentré est effectué (c'est-à-dire à l'origine O sans centrage), le premier axe factoriel a les propriétés suivantes :

1. Connectez l'origine O au centroïde GI du nuage NI.

2. Cet axe est orthogonal au sens de la distance (ie distance à x_2) utilisé pour l'hyperplan H_I contenant le nuage NI .
3. L'inertie projetée de NI dans cette direction est 1.

Bien sûr, l'axe lui-même n'est pas intéressant.

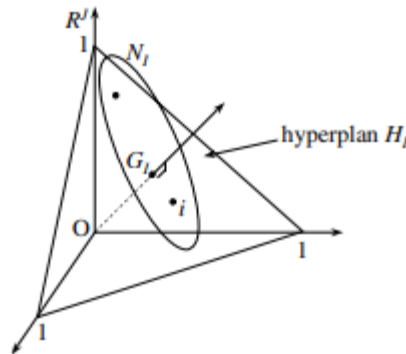


Figure 8: Le premier axe factoriel du nuage NI non centré est le facteur trivial OG_I orthogonal

La projection de chaque point de NI sur OG_I est confondue avec G_I . Cette projection de NI sur l'axe OG_I est appelée le facteur trivial ou constant. L'orthogonalité du premier axe OG_I avec l'hyperplan H_I a des conséquences importantes. L'axe suivant est par définition orthogonal à l' OG_I , de sorte que l'analyse peut être effectuée indépendamment de O ou G_I .

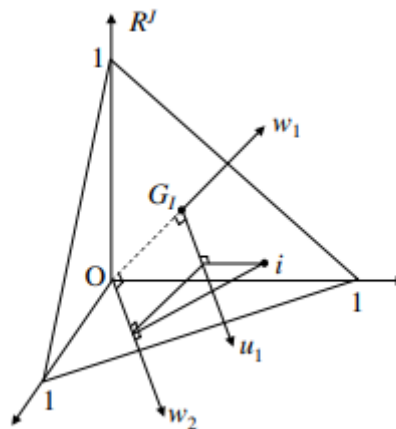


Figure 9: Analyse par rapport au barycentre et par rapport à l'origine

2.8 LA DUALITÉ

Les deux nuages NI et NJ constituent deux représentations d'un même tableau, l'une à travers ses profils-lignes, l'autre à travers ses profils-colonnes. Il s'ensuit que les analyses de ces deux nuages ne sont pas indépendantes : les relations entre ces deux analyses sont

communément regroupées sous le terme de dualité. Cette dualité est plus fondamentale et plus riche en AFC qu'en ACP car les lignes et les colonnes représentent des objets de même nature, ce qui n'est pas le cas en ACP Statistique du x^2 et inertie des deux nuages NI et NJ

Lorsque l'on étudie un tableau de contingence, c'est-à-dire une population de n individus au travers de deux variables qualitatives, il est classique de mesurer la significativité de la liaison entre ces deux variables à l'aide de la statistique x^2 . Appliquée à un tableau d'effectifs, cette statistique mesure l'écart entre les effectifs observés et les effectifs théoriques que l'on obtiendrait en moyenne si les deux variables étaient indépendantes. Elle s'écrit :

$$\chi^2 = \sum_{ij} \frac{(\text{effectif observé} - \text{effectif théorique})^2}{\text{effectif théorique}} = \sum_{ij} \frac{(n f_{ij} - n f_{i.} f_{.j})^2}{n f_{i.} f_{.j}}$$

Figure 10: La statistique X^2

La statistique x^2 est égale, au coefficient n près, l'inertie totale par rapport à leur barycentre de l'un l'autre des nuages NI et NJ . En effet, dans RI , l'inertie totale de NI rapport à GI s'écrit :

$$\text{Inertie}(N_I) = \sum_i \text{Inertie}(i) = \sum_i f_{i.} d^2(i, G_I) = \sum_i f_{i.} \sum_j \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - f_{.j} \right)^2$$

Figure 11: Inertie totale de NI rapport à GI

Cette double équation montre que les inerties totales respectives des deux nuages NI et NJ décrivent la relation entre les deux variables sous deux formes différentes.

Remarque:

La quantité x^2/n notée F^2 mesure la force du lien entre deux variables qualitatives (ce lien est plus fort lorsqu'une modalité est exclusivement associée à l'autre modalité). l'indice x^2 mesure la signification (une forte association peut ne pas être significative si elle est observée chez un très petit nombre d'individus ; une forte association peut ne pas être significative si elle est observée chez un grand nombre d'individus ; , des associations faibles peuvent être significatives).

2.9 DUALITE ENTRE LES FACTEURS SUR I ET LES FACTEURS SUR J

Comme dans l'ACP, un facteur est un ensemble de coordonnées qui projettent un point de nuage sur l'un de ses axes factoriels. Le facteur de ligne est la prédiction NI et le facteur de colonne est la prédiction NJ . Le rang d'un facteur est le rang de l'axe des facteurs correspondant.

En plus d'avoir la même inertie totale, les nuages NI et NJ ont les propriétés surprenantes suivantes :

Leur ajustement conduit à deux séquences de facteurs « doubles ». Plus précisément, au chapitre 5 nous montrons :

1. L'inertie associée à chaque axe de coordonnées du nuage est égale.
2. Les éléments de ligne (égaux) et les éléments de colonne sont liés par des relations dites de transition (permettant des transitions de RI à RJ et vice versa). Les deux paragraphes suivants expliquent cette dualité. Ses résultats les plus importants sont :

Les facteurs I et J de même rang accentuent la même proportion de couplage et doivent être interprétés ensemble. L'un est représenté par des lignes de profil et l'autre par des colonnes de profil.

2.9.1.1 Relations de transition

Les formules de transition précisent les relations entre les points représentant d'une part les lignes et d'autre part les colonnes. Avec les notations suivantes :

1. $F_s(i)$: projection de la ligne i sur l'axe de rang s de NI ,
2. $G_s(j)$: projection de la colonne j sur l'axe de rang s de NJ ,
3. λ_s : valeur commune de l'inertie associée à chacun de ces deux axes, les deux relations de transition s'écrivent :

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_j \frac{f_{ij}}{f_{i.}} G_s(j)$$

$$G_s(j) = \frac{1}{\sqrt{\lambda_s}} \sum_i \frac{f_{ij}}{f_{.j}} F_s(i)$$

Figure 12: Les formules de transition

Ces deux propriétés d'expression des résultats d'une analyse de nuage en fonction de l'autre analyse de nuage entraînent des économies de calcul. Mais surtout, il est logique d'afficher les lignes et les colonnes en même temps.

2.10 REPRESENTATION SIMULTANEE DES LIGNES ET DES COLONNES ; RELATIONS BARYCENTRIQUES

La représentation conjointe est obtenue en superposant les projections respectives des deux nuages NI et NJ sur le plan engendré par les axes de même rang.eux nuages. Sur le graphique ainsi obtenu, la relation de position entre les points de ligne et les points de colonne peut être décrite à partir de la relation de transition comme suit.

La projection de la ligne i sur l'axe des rangs s (en unités RJ) noté $F_s(i)$ est l'axe des rangs s (unités RI), où chaque colonne j se voit attribuer un poids f_{ij}/f_i (cette représentation de l'équation de transition est appelée centroïde). L'élément "lourd" est le centroïde et la colonne j attire la ligne i par la valeur de f_{ij}/f_i est élevé. Au niveau factoriel, les points éloignés de l'origine méritent une attention particulière, car ce sont les profils qui s'écartent le plus du profil moyen. Donc, pour un facteur, trouvez la colonne la plus pertinente j du même côté que la ligne i et la colonne la moins pertinente j du côté opposé. Cela permet d'interpréter les positions des lignes par rapport à toutes les colonnes, justifiant les avantages pratiques de la représentation simultanée. Avec les rôles des lignes et des colonnes inversés, une formulation symétrique tient. D'où le nom de propriété du double barycentre donné à la règle principale d'interprétation du graphe AFC. Cette double caractéristique n'est pas seulement unique à FCA, c'est aussi une caractéristique de FCA.

Nous montrons que les facteurs de CA peuvent être trouvés en essayant de construire une fonction définie par les lignes et les colonnes du tableau de contingence telle que la dualité des barycentres soit vérifiée. La représentation simultanée est largement acceptée en CFA, mais pas en ACP. Deux arguments importants peuvent être avancés à propos de cette superposition.

1. Dans ACP, les lignes et les colonnes représentent des objets de natures très différentes (individus et variables), alors que dans le tableau de contingence FCA, les lignes et les colonnes sont de même nature, la classe des individus. De ce simple point de vue, mettre toutes ces classes dans le même graphe n'est pas un problème.

2. Il existe d'autres représentations de CFA où les classes de personnes qui composent les lignes et les colonnes du tableau de contingence sont dans le même espace.

Dans ce cas, leur affichage simultané est naturel. En résumé, pour un diagramme à courbes jointes ligne-colonne, le diagramme à courbes jointe lignes-colonnes interprète la position relative de deux points dans le même ensemble (ligne ou colonne) comme une distance, et la position d'un point dans l'ensemble comme la distance de tous les points Comparer avec la

position. D'autres ensembles de points sont interprétés comme des centroïdes. Les associations entre les lignes et les colonnes impliquées par la proximité dans le graphique doivent être vérifiées dans le tableau de données.

2.11 NOMBRE D'AXES ET INERTIE TOTALE

Dans l'espace RJ, le nuage NI est contenu dans un sous-espace de dimension $J - 1$; dans cet espace, on peut donc trouver au maximum $J - 1$ dimensions orthogonales d'inertie non nulle. De même, dans l'espace RI, on peut trouver au maximum $I - 1$ dimensions orthogonales d'inertie non nulle. Compte tenu de la dualité (même inertie sur les axes de même rang dans les deux espaces), en AFC on peut trouver au maximum $\min\{I - 1, J - 1\}$ axes d'inertie non nulle. L'inertie associée à un axe étant au maximum égale à 1, l'inertie totale en AFC est donc comprise entre 0 (indépendance) et $\min\{I - 1, J - 1\}$ (liaison d'intensité maximum = association stricte entre les modalités des deux variables mises en correspondances).

2.12 AIDES À L'INTERPRÉTATION ET ÉLÉMENTS SUPPLÉMENTAIRES

Les indices d'aide à l'interprétation définis dans l'ACP (voir section 1.9) (la qualité de représentation d'un élément par son axe ou son plan, et la contribution de l'élément à l'inertie de l'axe) s'appliquent à tout nuage. Elles s'appliquent donc à l'AFC. Notez que si dans ACP tous les poids des éléments sont généralement égaux, dans FCA ce n'est pas le cas. Cependant, ces poids compensent la contribution du point à l'inertie de l'axe. FCA, comme PCA, utilise la technique de l'ajout d'éléments presque systématiquement. Elle consiste à projeter sur des axes factoriels les profils de lignes ou de colonnes qui ne participent pas au calcul de ces axes. Des lignes supplémentaires sont connectées à la colonne active par des formules centroïdes. De même, la formule centroïde connecte des colonnes supplémentaires à la ligne active. Dans de nombreux cas, ces éléments facilitent également l'interprétation. Par exemple, pour les grands tableaux, il est très utile de connaître l'emplacement du centre de gravité et la qualité d'affichage de plusieurs lignes ou de plusieurs colonnes.

3. L'ASPECT PRATIQUE

3.1 ORIGINE DE DONNES

Pour pratiquer l'utilisation de R, Python et Excel, j'ai choisi L'ANCIENNETÉ DE CHÔMAGE par niveau d'études comme données. Ces données font partie d'un jeu de données publié par le Haut-Commissariat aux Comptes Nationaux (HCP). En fait, les comptes nationaux constituent l'un des éléments clés du système d'information statistique d'un pays. Ils fournissent une représentation numérique détaillée de l'économie aux niveaux annuel et trimestriel. Leur importance réside dans le rôle qu'ils peuvent jouer en tant qu'outils statistiques pour l'analyse économique, la prise de décision, la définition et la formulation des politiques économiques et sociales et les comparaisons internationales.

3.1.1 Objectif de cette étude

L'objectif d'une telle étude est donc de savoir s'il existe une relation de dépendance entre l'ancienneté dans le chômage et le niveau d'enseignement selon la version de la formation continue.

Tableau représente le tableau de contingence ou des correspondances :

Table des correspondances									
Niveau d'enseignements	Ancienneté de chômage								
	entre 1 et 1,5	entre 1 et 3 mois	entre 1,5 et 2	entre 2 et 3	entre 3 et 6	entre 6 mois	moins de 1	plus de 3	Marge active
Bac et enseignement supérieur sans diplôme bac+2	124	291	42	49	247	207	63	70	1093
College	153	101	56	98	119	176	19	109	831
Diplome bac+5 et plus	66	142	31	33	151	147	43	37	650
Diplome niveau bac+2	168	326	50	67	274	284	93	101	1363
Diplome niveau licence	134	245	45	41	185	218	45	48	961
Enseignement primaire	150	99	65	106	105	151	17	168	861
lycee	941	1127	319	462	1191	1346	260	531	6177
Niveau bac sans etudes superieures	400	657	124	178	630	635	152	209	2985
Pas d'etudes	26	12	8	29	29	35	6	35	180
Marge active	2162	3000	740	1063	2931	3199	698	1308	15101

Figure 13 : Tableau de donnes

3.2 LES TECHNOLOGIE UTILISER POUR APPLIQUER AFC

Dans cette partie, nous allons essayer d'écrire le code dans chacune des applications suivantes. Nous commencerons par expliquer comment Nous l'avons fait, commenté chaque

code, et c'est parti Nous l'appliquons aux mêmes données sur lesquelles nous avons travaillé Ce projet et une petite comparaison entre les applications



Figure 14: Les technologies utilisées

3.3 APPLIQUER L'AFC PAR R

R est un environnement intégré de manipulation de données, de calcul et de préparation de graphiques. Toutefois, ce n'est pas seulement un « autre » environnement statistique (comme SPSS ou SAS, par exemple), mais aussi un langage de programmation complet et autonome

3.3.1 Packages utilisée pour calculer l'afc

Plusieurs fonctions de différents packages sont disponibles dans le logiciel R pour calculer l'AFC:

- **CA()** [package FactoMineR] ;
- **ca()** [package ca] ;
- **dudi.coa()** [package ade4] ;
- **corresp()** [package MASS] ;
- **epCA()** [package ExPosition] ;

Quelle que soit la fonction que vous choisissiez, vous pouvez facilement extraire et visualiser les résultats FCA à l'aide des fonctions R fournies dans le package factoextra.

Ici, nous utiliserons les deux packages **FactoMineR** (pour l'analyse) et **factoextra** (pour la visualisation basée sur ggplot2). Installez les deux packages comme suit :



Figure 155 Installer packages FactoMineR et factoextra

Chargez-les dans R, en tapant ceci :



Figure 16:Charger les packages FactoMineR et factoextra

3.3.2 Description de données

La fonction `summary()` permet d'avoir la description statistique d'une variable ou d'une table de donnée. Pour une variable donnée, la fonction renvoie 5 valeurs : le minimum (Min.), le premier quartile (1st Qu.), la médiane (Median), la moyenne (Mean), le troisième quartile (3rd Qu.) et le maximum (Max). On a utilisé la fonction `summary()` pour vérifier que les variables ont été bien spécifiées comme quantitatives et qualitatives.

```
> #Statistique Descriptive verifier que les variables ont ete bien specifiees entre quantitatives et qualitatives
> summary(Projet_Fin)
```

moins de 1 mois	entre 1 et 3 mois	entre 3 et 6mois	entre 6 mois et 1 an	entre 1 et 1,5 ans	entre 1,5 et 2ans
Min. : 6.00	Min. : 12.0	Min. : 29.0	Min. : 35.0	Min. : 26.0	Min. : 8.00
1st Qu.: 19.00	1st Qu.: 101.0	1st Qu.: 119.0	1st Qu.: 151.0	1st Qu.:124.0	1st Qu.: 42.00
Median : 45.00	Median : 245.0	Median : 185.0	Median : 207.0	Median :150.0	Median : 50.00
Mean : 77.56	Mean : 333.3	Mean : 325.7	Mean : 355.4	Mean :240.2	Mean : 82.22
3rd Qu.: 93.00	3rd Qu.: 326.0	3rd Qu.: 274.0	3rd Qu.: 284.0	3rd Qu.:168.0	3rd Qu.: 65.00
Max. :260.00	Max. :1127.0	Max. :1191.0	Max. :1346.0	Max. :941.0	Max. :319.00

entre 2 et 3ans	plus de 3ans
Min. : 29.0	Min. : 35.0
1st Qu.: 41.0	1st Qu.: 48.0
Median : 67.0	Median :101.0
Mean :118.1	Mean :145.3
3rd Qu.:106.0	3rd Qu.:168.0
Max. :462.0	Max. :531.0

Figure 17: la fonction `summary()`

3.3.3 Visualiser le tableau de contingence en diagramme à bulle

Le tableau de contingence ci-dessus n'est pas très gros. Par conséquent, il est facile d'inspecter et d'interpréter visuellement les profils des lignes et des colonnes:



Figure 18:code le tableau de contingence en diagramme a bulle

Le résultat de l'exécution

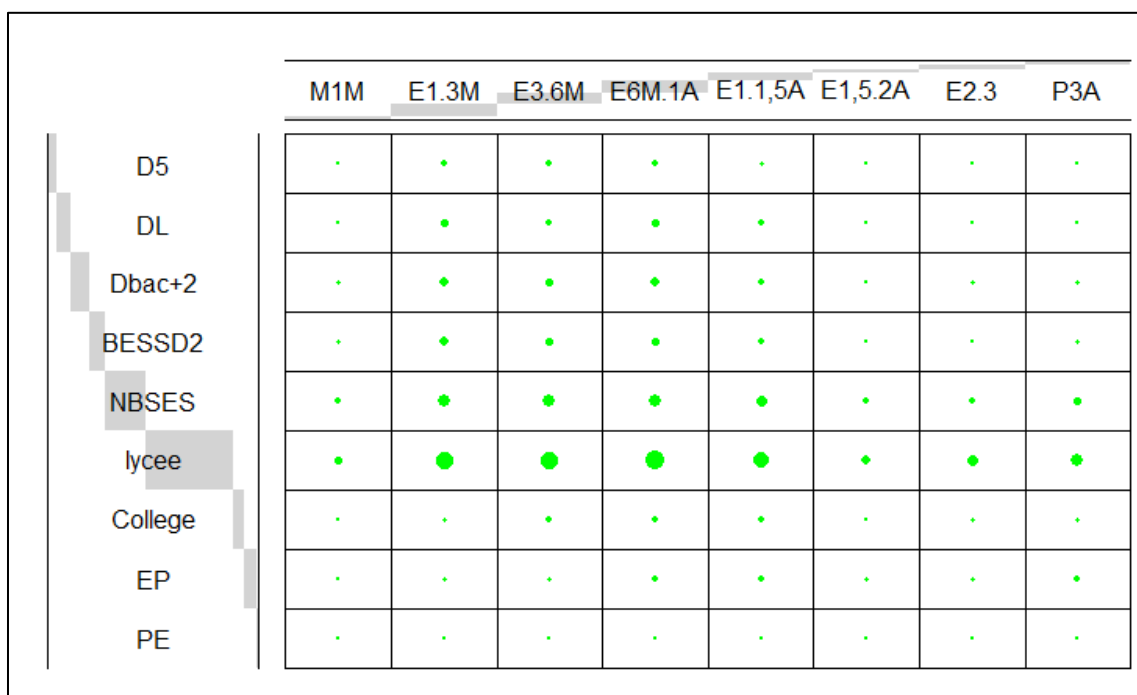


Figure 19: tableau de contingence en diagramme a bulle

Le tableau de contingence peut être visualisé en utilisant les fonctions `balloonplot()` [package `gplots`] et `mosaicplot()` [package `garphics`]:



Figure 20 :colorRampPalette()

Donne le résultat suivant :

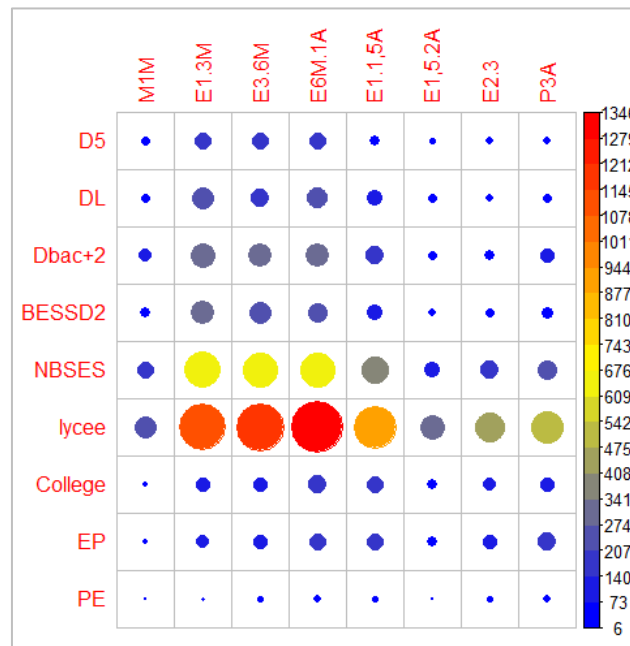


Figure 21:tableau de contingence en diagramme a bulle selon effectif

3.3.4 Test de khi2

Pour interpréter l'analyse des correspondances, la première étape à faire c'est de savoir où il y a une dépendance entre les lignes et les colonnes. On peut utiliser le test de khi-2 : en fait il permet de trouver la probabilité que les deux variables d'un tableau de contingence soient indépendantes, c'est-à-dire qu'il n'y a pas une liaison entre les modalités en ligne les modalités en colonne

Dans notre étude, les variables de ligne et de colonne sont statistiquement significativement associées

```
> chisq <- chisq.test(Projet_Fin)
> chisq

Pearson's Chi-squared test


data:  Projet_Fin
X-squared = 641.05, df = 56, p-value < 2.2e-16
```

Figure 22:Test de khi2

3.3.5 Calculer l'afc

Fonction R: CA() [FactoMiner].

Format simplifié :



```
CA (X, ncp = 5, graph = TRUE)
```

Figure 23:CA()

- X: un data frame (tableau de contingence)
- ncp : nombre de dimensions à conserver dans les résultats finaux.
- graph : une valeur logique. Si TRUE le graphique est affiché.

Pour calculer l'AFC, tapez ceci :



```
library ("FactoMineR")  
res.ca <- CA (Projet_Fin, graph = FALSE)
```

Figure 24:Calculer AFCpar CA()

Le résultat de la fonction CA() est une liste comprenant :

```

**Results of the Correspondence Analysis (CA)**
The row variable has 9 categories; the column variable has 8 categories
The chi square of independence between the two variables is equal to 641.0495 (p-value
= 2.865656e-100 ).
*The results are available in the following objects:

  name          description
1  "$eig"       "eigenvalues"
2  "$col"       "results for the columns"
3  "$col$coord" "coord. for the columns"
4  "$col$cos2"  "cos2 for the columns"
5  "$col$contrib" "contributions of the columns"
6  "$row"       "results for the rows"
7  "$row$coord" "coord. for the rows"
8  "$row$cos2"  "cos2 for the rows"
9  "$row$contrib" "contributions of the rows"
10 "$call"      "summary called parameters"
11 "$call$marge.col" "weights of the columns"
12 "$call$marge.row" "weights of the rows"

```

Figure 25:Resultat de AFC

3.3.6 Valeurs propres variances

L'examen des valeurs propres permet de déterminer le nombre d'axes principaux à considérer. Les valeurs propres correspondent à la quantité d'informations retenue par chaque axe. Elles sont grandes pour le premier axe et petites pour l'axe suivant. Les valeurs propres et la proportion de variances expliquées (capturées) par les différents axes peuvent être extraites à l'aide de la fonction `get_eigenvalue()` [factoextra package].

```

library ("factoextra")
eig.val <- get_eigenvalue (RProjet_Fin)
eig.val

```

Figure 26: get_eigenvalue()

Donne le résultat suivant :

```

> eig.val
  eigenvalue variance.percent cumulative.variance.percent
Dim.1 3.745118e-02      88.2225552             88.22256
Dim.2 2.708713e-03       6.3808287             94.60338
Dim.3 1.406930e-03       3.3142591             97.91764
Dim.4 4.009316e-04       0.9444618             98.86210
Dim.5 3.239543e-04       0.7631289             99.62523
Dim.6 1.093738e-04       0.2576483             99.88288
Dim.7 4.971748e-05       0.1171179             100.00000
> |

```

Figure 27: Valeurs propres variances en R

3.3.7 Le graphique des valeurs propres avec une droite en pointillée

Une autre méthode pour déterminer le nombre de dimensions est de regarder le graphique des valeurs propres (scree plot), ordonnées de la plus grande à la plus petite valeur. Le nombre d'axes est déterminé par le point point, au-delà duquel les valeurs propres restantes sont toutes relativement petites et de tailles comparables. Le graphique des valeurs propres peut être produit à l'aide de la fonction `fviz_eig()` ou `fviz_screplot()` [package `factoextra`].

```
library("factoextra")
fviz_screplot(RProjet_Fin) + geom_hline(
  ( yintercept = 33.33,
    linetype = 2,
    color = "red")
```

Figure 28: la fonction `fviz_screplot()`

Donne le résultat suivant :

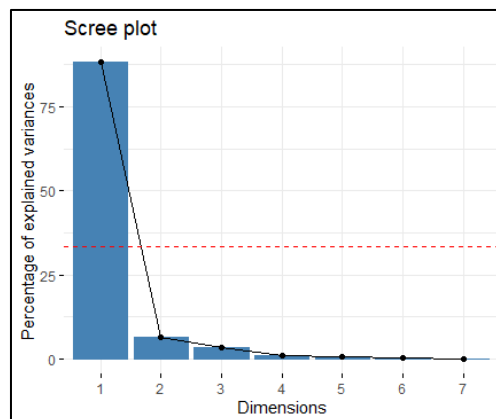


Figure 29: Le graphe des valeurs propres

3.3.8 Le biplot des lignes et des colonnes

La fonction `fviz_ca_biplot()` [package `factoextra`] peut être utilisée pour dessiner le biplot des lignes et des colonnes. Dans le graphique les lignes sont représentées par des points bleus et des colonnes par des triangles rouges.

La distance entre les points lignes ou entre les points colonnes donne une mesure de leur similitude (ou dissemblance). Les points lignes avec un profil similaire sont proches sur le graphique. Il en va de même pour les points colonnes.



Figure 30: la fonction `fviz_ca_biplot()`

Le résultat de l'exécution :

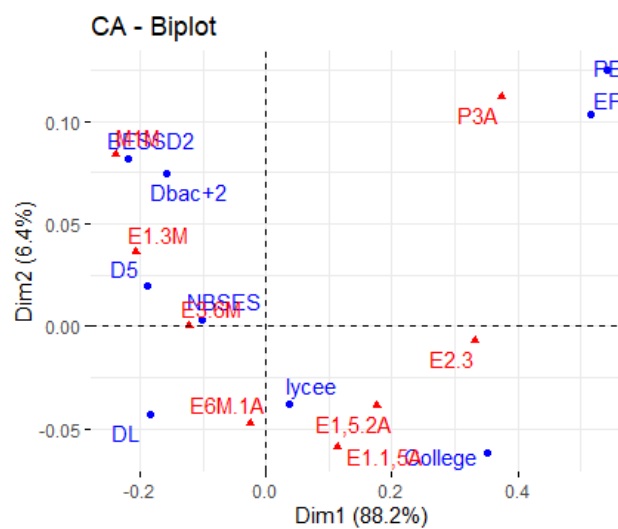


Figure 31: le biplot de lignes et colonnes

3.4 GRAPHIQUE DES POINTS LIGNES

La fonction `get_ca_row()` [factoextra] est utilisée pour extraire les résultats pour les lignes. Cette fonction renvoie une liste contenant les coordonnées, les cos2 et les contributions des lignes :



Figure 32: `get_ca_row()`

Donner le résultat suivant :

```
Correspondence Analysis - Results for rows
=====
  Name      Description
1 "$coord"  "Coordinates for the rows"
2 "$cos2"   "Cos2 for the rows"
3 "$contrib" "contributions of the rows"
4 "$inertia" "Inertia of the rows"
> |
```

3.4.1 Les coordonnées des lignes

Le code R ci-dessous affiche les coordonnées de chaque point ligne pour chacune des dimensions (1, 2, 3, 4 et 5) :



```
head(row$coord)
```

Figure 33: la fonction head()

Donne le résultat suivant :

```
> head(row$coord)
      Dim 1      Dim 2      Dim 3      Dim 4      Dim 5
D5      -0.1492233 -0.03783187  0.049484797  0.027877429  0.0565123225
DL      -0.1801355 -0.04989754 -0.083364140  0.012063906 -0.0206302412
Dbac+2  -0.1321324  0.05482862  0.019826794  0.046280559 -0.0154728970
BESSD2  -0.1876864  0.06145398 -0.017476054 -0.044259819  0.0053524788
NBSES   -0.0838227 -0.01849606  0.016674914 -0.015383199 -0.0008389043
College  0.4001703 -0.09943039 -0.002976133  0.002711569 -0.0056586405
> |
```

Figure 34: Les coordonnées des lignes

Utilisez la fonction fviz_ca_row() [dans factoextra] pour visualiser uniquement les points lignes :



```
fviz_ca_row(RProjet_Fin, repel = TRUE)
```

Figure 35: fviz_ca_row

Donne le résultat suivant :

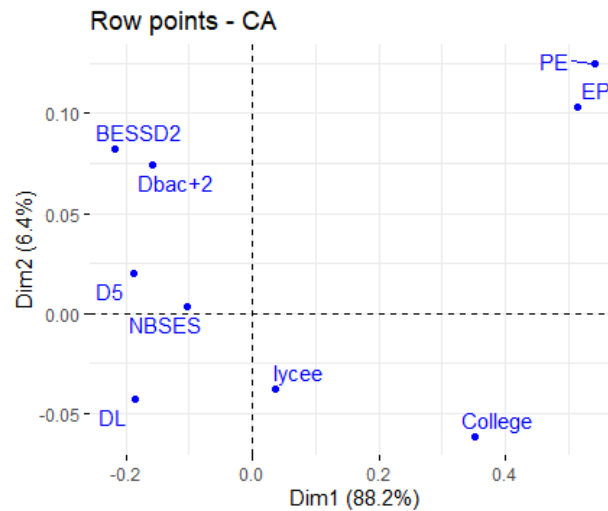


Figure 36:Représentation des lignes

3.4.2 Qualité de représentation des lignes

Le résultat de l'analyse montre que le tableau de contingence est bien représenté dans un espace de faibles dimensions en utilisant l'AFC. Les deux dimensions 1 et 2 sont suffisantes pour conserver 88,6% de l'inertie totale (variation) contenue dans les données. Cependant, tous les points ne sont pas aussi bien représentés dans les deux dimensions.

Le cos2 mesure le degré d'association entre les lignes / colonnes et un axe particulier. Le cos2 des points lignes peut être extrait comme suit :



Figure 37:la fonction head()

Donner le Résultat suivant :

```
> head(row$cos2)
      Dim 1      Dim 2      Dim 3      Dim 4      Dim 5
D5      0.7393216 0.04751993 8.130245e-02 2.580272e-02 1.060343e-01
DL      0.7637400 0.05860106 1.635708e-01 3.425493e-03 1.001742e-02
Dbac+2  0.7505067 0.12922617 1.689822e-02 9.207317e-02 1.029151e-02
BESSD2  0.8493245 0.09105592 7.363680e-03 4.723103e-02 6.907455e-04
NBSES   0.8842606 0.04305417 3.499324e-02 2.978175e-02 8.856902e-05
College 0.9396932 0.05801426 5.197579e-05 4.314573e-05 1.878973e-04
> |
```

Figure 38:La qualité des points lignes

Les valeurs de \cos^2 sont comprises entre 0 et 1. La somme des \cos^2 pour les lignes sur toutes les dimensions de l'AFC est égale à 1.

```

# Colorer en fonction du cos2:
fviz_ca_row (RProjet_Fin, col.row = "cos2",
             gradient.cols = c ("#00AFBB", "#E7B800",
                                "#FC4E07"),
             repel = TRUE)

```

Figure 39: La fonction `fviz_ca_row`

Donne le résultat suivant :

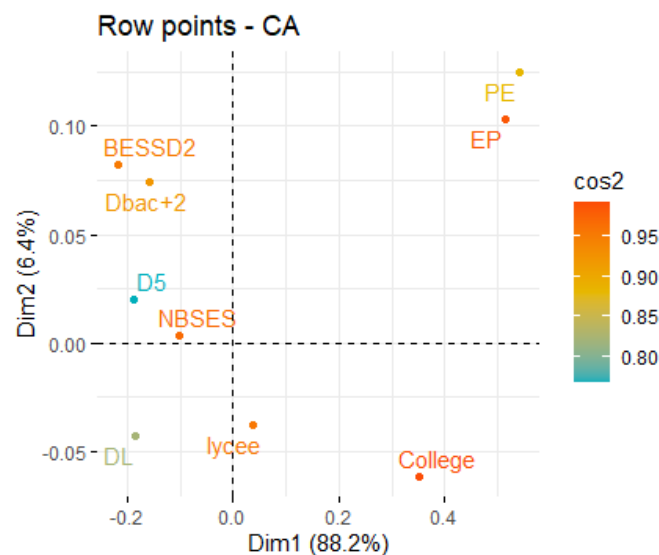


Figure 40: Qualité de représentation

3.5 GRAPHIQUE DES POINTS COLONNES

La fonction `get_ca_col()` [factoextra] est utilisée pour extraire les résultats pour les colonnes. Cette fonction renvoie une liste contenant les coordonnées, les \cos^2 et les contributions des colonnes:

```

col <- get_ca_col(RProjet_Fin)
col

```

Figure 41: `get_ca_col()`

Donner le résultat suivant :

```
Correspondence Analysis - Results for columns
=====
Name      Description
1 "$coord" "Coordinates for the columns"
2 "$cos2"  "Cos2 for the columns"
3 "$contrib" "contributions of the columns"
4 "$inertia" "Inertia of the columns"
> |
```

3.5.1 Les coordonnées des colonnes

Le code R ci-dessous affiche les coordonnées de chaque point colonnes pour chacune des dimensions (1, 2, 3, 4 et 5):

```
> head(col$coord)
      Dim 1      Dim 2      Dim 3      Dim 4      Dim 5
M1M      -0.25846316  0.075146619  0.0910444936  0.07341205  0.01159412
E1.3M    -0.22472940  0.032438565 -0.0494115671 -0.01032150 -0.01541923
E3.6M    -0.13664691 -0.001929715  0.0418316488 -0.03091713  0.01980947
E6M.1A   -0.01685988 -0.061258605 -0.0001533463  0.02340386 -0.01110147
E1,5.2A  0.25790282 -0.054491721 -0.1009377996  0.01911431  0.06022099
E2.3     0.47592045 -0.048266825  0.0502545448 -0.02538525 -0.02567556
> |
```

Figure 42: Les coordonnées des colonnes

Utilisez la fonction `fviz_ca_col()` [dans `factoextra`] pour visualiser uniquement les points lignes



```
fviz_ca_col(RProjet_Fin, repel = TRUE)
```

Figure 44: la fonction `fviz_ca_col()`

Donne le résultat suivant :

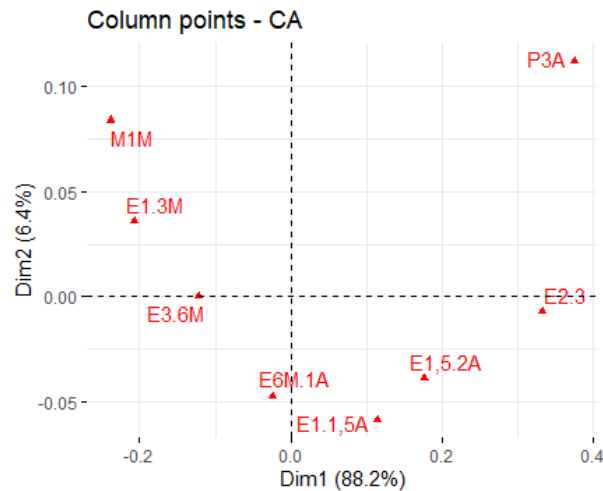


Figure 43: la Représentation des colonnes

3.5.2 Qualité de représentation des colonnes

Le résultat de l'analyse montre que le tableau de contingence est bien représenté dans un espace de faibles dimensions en utilisant l'AFC. Les deux dimensions 1 et 2 sont suffisantes pour conserver 88,6% de l'inertie totale (variation) contenue dans les données. Cependant, tous



Figure 45: la fonction head()

les points ne sont pas aussi bien représentés dans les deux dimensions. Le cos2 mesure le degré d'association entre les lignes / colonnes et un axe particulier. Le cos2 des points colonnes peut être extrait comme suit :

Donner le Résultat suivant :

```
> head(col$cos2)
      Dim 1      Dim 2      Dim 3      Dim 4      Dim 5
M1M      0.77252584 0.0653032212 9.585683e-02 0.062323302 0.001554502
E1.3M     0.92904371 0.0193570703 4.491316e-02 0.001959757 0.004373624
E3.6M     0.85686214 0.0001708825 8.030106e-02 0.043864142 0.018007649
E6M.1A     0.05950669 0.7855814189 4.922702e-06 0.114665287 0.025799898
E1,5.2A    0.79349063 0.0354234348 1.215451e-01 0.004358599 0.043263877
E2.3      0.97198677 0.0099974465 1.083783e-02 0.002765379 0.002828990
```

Figure 46 : Qualité de représentation des colonnes

Les valeurs de \cos^2 sont comprises entre 0 et 1. La somme des \cos^2 pour les lignes sur toutes les dimensions de l'AFC est égale à 1.

```
# Colorer en fonction du cos2:
fviz_ca_col (RProjet_Fin
  , col.row = "cos2"
  , gradient.cols = c ("#00AFBB", "#E7B800",
    "#FC4E07")
  , repel = TRUE)
```

Figure 47: La fonction `fviz_ca_col()`

Donne le résultat suivant :

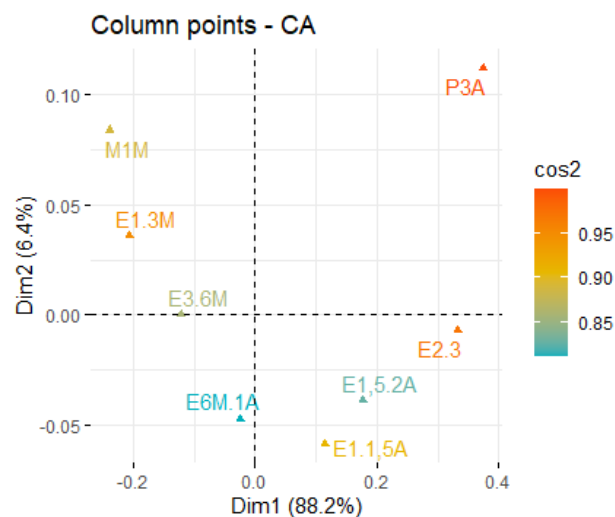


Figure 48: La representation des colonnes selon la qualite

3.6 APPLIQUER L'AFC PAR EXCEL

Cette partie vous aidera à mettre en place et à interpréter une Analyse des Correspondances (AC) dans Excel à l'aide du logiciel XLSTAT. XLSTAT est un logiciel s'appuyant sur le tableur Office Excel. Vous ne savez pas s'il s'agit du bon outil d'analyse de données multivariées dont vous avez besoin ?

3.6.1 L'utilisation de excel

Une fois XLSTAT ouvert, sélectionnez la fonctionnalité XLSTAT / Analyse des données / Analyse des correspondances

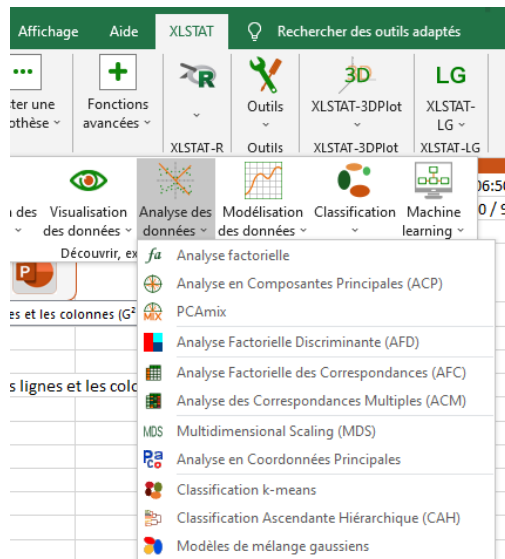


Figure 49:L'UTILISATION DE EXCEL

- Une fois que vous avez cliqué sur le bouton, la boîte de dialogue **Analyse des correspondances** apparaît.

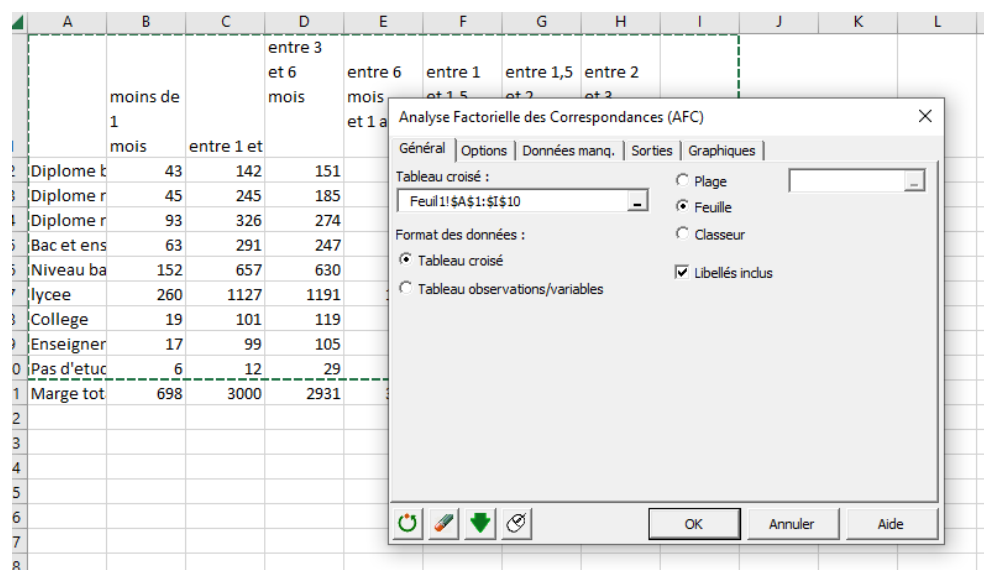


Figure 50:L'UTILISATION DE EXCEL 1

- Dans l'onglet **Général**, sélectionnez le tableau entier sur la feuille Excel. Si vos données se trouvent dans un tableau de contingence, comme dans cet exemple, sélectionnez le format **de tableau à double entrée**. Si vos données se trouvent dans un tableau Individus/Variables, sélectionnez l'option Tableau **Observations/Variables**.
- Si les noms des catégories pour chaque variable sont présents pour les lignes et les colonnes, assurez-vous que la case **Libellés inclus** est cochée.

- Avant de commencer l'interprétation introduisons la **notion de profil**. L'Analyse des Correspondances est basée sur l'analyse des profils. Un profil est un ensemble de fréquences divisées par leur total, c'est-à-dire les fréquences relatives. En d'autres termes, un profil reflète la façon dont la catégorie d'une variable change en fonction des catégories de l'autre variable.

3.6.2 Test de khi2

Test d'indépendance entre les lignes et les colonnes :	
Khi ² (Valeur)	641,050
Khi ² (Valeur)	74,468
DDL	56
p-value	<0,0001
alpha	0,05
Interprétation du test :	
H0 : Les lignes et les colonnes du tableau sont indépendantes.	
Ha : Il existe un lien entre les lignes et les colonnes du tableau.	
Etant donné que la p-value calculée est inférieure au niveau de signification alpha=0,05, on doit rejeter l'hypothèse nulle H0, et retenir l'hypothèse alternative Ha.	

Figure 51: Test de khi2

Le premier résultat affiché est le test d'indépendance entre les lignes et les colonnes, basé sur une statistique du chi carré. Si la valeur du chi carré observée est supérieure à la valeur critique, et si la valeur de p est inférieure au niveau alpha choisi, alors on peut conclure que les lignes et les colonnes du tableau sont significativement associées. Dans cette étude, il est fort probable que de réelles différences existent entre les tranches d'âge en termes et leurs profils d'appréciation.

3.6.3 Les valeurs propres

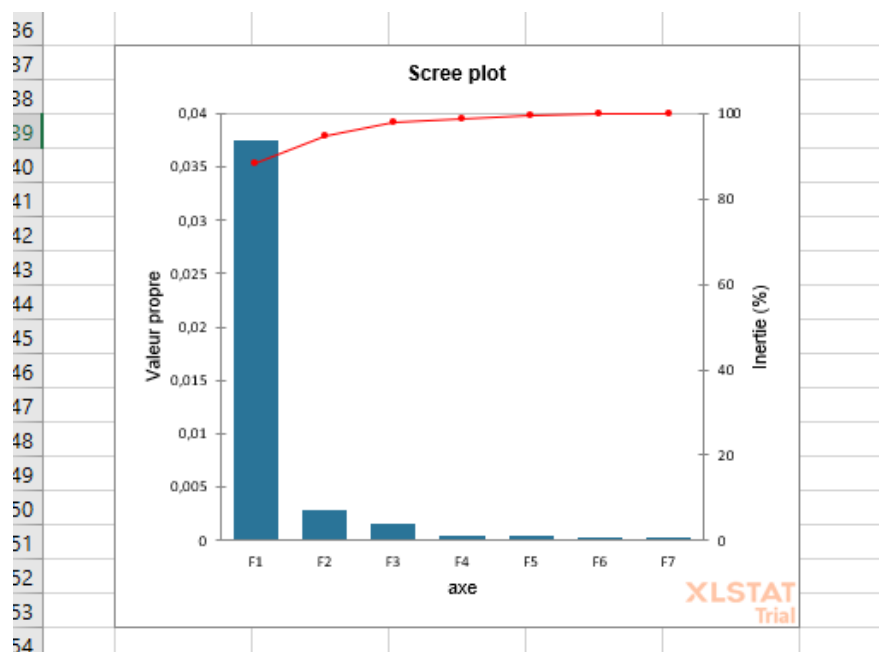
Inertie totale	0,042						
Valeurs propres et pourcentages d'inertie :							
	F1	F2	F3	F4	F5	F6	F7
Valeur propre	0,037	0,003	0,001	0,000	0,000	0,000	0,000
Inertie %	88,223	6,381	3,314	0,944	0,763	0,258	0,117
% cumulé	88,223	94,603	97,918	98,862	99,625	99,883	100,000

Figure 52: Les valeurs propres par Excel

Les valeurs propres correspondent à la variance extraite par chaque facteur (dimension). La qualité de l'analyse peut être évaluée en consultant le tableau des valeurs propres ou le scree plot correspondant. Si la somme des deux (ou quelques) premières valeurs

propres est proche du total représenté, alors la qualité de l'analyse est très élevée. L'analyse des correspondances dans cette étude est de bonne qualité puisque la somme des deux premières valeurs propres totalise 94 % de l'inertie totale.

3.6.4 Screeplot



3.6.5 Resultats pour lignes

3.6.5.1 Poids, distances et distances quadratiques a l'origine, inerties et inerties relatives

Résultats pour les lignes :					
Poids, distances et distances quadratiques à l'origine, inerties et inerties relatives (lignes) :					
	Poids (relatif)	Distance	Distance ²	Inertie	Inertie relative
Diplome bac	0,043	0,216	0,047	0,002	0,048
Diplome niv	0,064	0,209	0,044	0,003	0,065
Diplome niv	0,090	0,182	0,033	0,003	0,070
Bac et enseig	0,072	0,239	0,057	0,004	0,097
Niveau bac s	0,198	0,104	0,011	0,002	0,050
lycee	0,409	0,054	0,003	0,001	0,028
College	0,055	0,359	0,129	0,007	0,167
Enseigneme	0,057	0,529	0,280	0,016	0,376
Pas d'etudes	0,012	0,592	0,350	0,004	0,098

Figure 53: la Distance et le poids pour les lignes

Le tableau montre les poids, les distances et les distances au carré à l'origine, les inerties et les inerties relatives des lignes (respectivement et des colonnes). Les poids sont

des proportions marginales utilisées pour pondérer les profils de points lors du calcul des distances. Plus la distance à l'origine est grande, plus il y a dissemblance entre le profil de catégorie et le profil moyen (plus la catégorie participe à la dépendance entre les deux variables). Les groupes de diplôme lycée, Diplôme niveau bac+2 et plus et Niveau bac sans études supérieures ont la distance la plus courte à l'origine, indiquant que ces profils de groupe sont proches du profil moyen.

3.6.5.2 Les coordonnées des lignes

Coordonnées principales (lignes) :							
	F1	F2	F3	F4	F5	F6	F7
Diplome bac	-0,189	0,020	-0,077	0,020	0,066	0,016	0,002
Diplome niv	-0,185	-0,043	0,083	0,013	-0,002	0,024	-0,010
Diplome niv	-0,158	0,074	-0,003	0,046	-0,020	-0,006	0,004
Bac et enseig	-0,218	0,082	0,026	-0,043	0,004	0,002	0,011
Niveau bac s	-0,102	0,003	-0,008	-0,013	-0,009	-0,001	-0,004
lycee	0,037	-0,038	-0,009	-0,001	0,002	-0,007	-0,001
College	0,352	-0,062	0,010	0,006	-0,012	0,016	0,020
Enseigneme	0,515	0,103	0,058	0,003	0,023	-0,005	-0,007
Pas d'etudes	0,541	0,125	-0,188	-0,025	-0,061	0,044	-0,019
Coordonnées standard (lignes) :							
	F1	F2	F3	F4	F5	F6	F7
Diplome bac	-0,974	0,379	-2,058	0,999	3,642	1,507	0,335
Diplome niv	-0,954	-0,824	2,219	0,628	-0,101	2,322	-1,396
Diplome niv	-0,817	1,428	-0,072	2,317	-1,132	-0,534	0,633
Bac et enseig	-1,128	1,576	0,691	-2,152	0,202	0,163	1,539
Niveau bac s	-0,529	0,065	-0,225	-0,667	-0,506	-0,096	-0,539
lycee	0,189	-0,724	-0,231	-0,032	0,096	-0,645	-0,149
College	1,819	-1,188	0,261	0,277	-0,651	1,576	2,895
Enseigneme	2,662	1,987	1,544	0,140	1,298	-0,519	-0,952
Pas d'etudes	2,795	2,400	-4,999	-1,235	-3,375	4,168	-2,656

Figure 54: Les coordonnées des lignes

Les coordonnées principales et les coordonnées standard des lignes (colonnes respectivement) sont également affichées. Les coordonnées standard sont les coordonnées principales divisées par la racine carrée de la valeur propre du facteur correspondant. La somme des carrés pondérée des coordonnées standard est égale à 1 pour chaque facteur.

3.6.5.3 Les contributions des lignes

Contributions (lignes) :								
	Poids (relatif)	F1	F2	F3	F4	F5	F6	F7
Diplome bac	0,043	0,041	0,006	0,182	0,043	0,571	0,098	0,005
Diplome niv	0,064	0,058	0,043	0,313	0,025	0,001	0,343	0,124
Diplome niv	0,090	0,060	0,184	0,000	0,485	0,116	0,026	0,036
Bac et ensei	0,072	0,092	0,180	0,035	0,335	0,003	0,002	0,171
Niveau bac s	0,198	0,055	0,001	0,010	0,088	0,051	0,002	0,057
lycee	0,409	0,015	0,215	0,022	0,000	0,004	0,170	0,009
College	0,055	0,182	0,078	0,004	0,004	0,023	0,137	0,461
Enseignement	0,057	0,404	0,225	0,136	0,001	0,096	0,015	0,052
Pas d'etudes	0,012	0,093	0,069	0,298	0,018	0,136	0,207	0,084

Figure 55:Contributions des lignes

Voici un tableau des **contributions des lignes** (colonnes respectivement). Les contributions correspondent à l'importance de chaque catégorie pour chaque facteur (dimension). La somme des contributions est égale à 1 pour chaque facteur. En règle générale, si la contribution est supérieure à $1/I$, I étant le nombre de lignes (respectivement $1/J$ avec J le nombre de colonnes), la catégorie est importante pour le facteur donné. Dans notre étude, le groupe Enseignement primaire est important pour le facteur F1, les groupes Enseignement primaire et lycée sont importants pour le facteur F2.

3.6.6 Resultat pour les colonnes

3.6.6.1 Poids, distances et distances quadratiques à l'origine, inerties et inerties relatives

Résultats pour les colonnes :					
Poids, distances et distances quadratiques à l'origine, inerties et inerties relatives (colonnes) :					
	Poids (relatif)	Distance	Distance ²	Inertie	Inertie relative
moins de 1m	0,046	0,268	0,072	0,003	0,078
entre 1 et 3 r	0,199	0,215	0,046	0,009	0,217
entre 3 et 6m	0,194	0,131	0,017	0,003	0,079
entre 6 mois	0,212	0,059	0,003	0,001	0,017
entre 1 et 1,5	0,143	0,135	0,018	0,003	0,061
entre 1,5 et 2	0,049	0,199	0,040	0,002	0,046
entre 2 et 3a	0,070	0,338	0,114	0,008	0,189
plus de 3ans	0,087	0,391	0,153	0,013	0,312

Figure 56:La distance et le poids des colonnes

3.6.6.2 Les coordonnées des colonnes

Coordonnées principales (colonnes) :							
	F1	F2	F3	F4	F5	F6	F7
moins de 1m	-0,238	0,084	-0,064	0,060	-0,004	-0,009	0,014
entre 1 et 3 r	-0,207	0,036	0,047	-0,007	-0,007	0,008	0,001
entre 3 et 6m	-0,122	0,000	-0,040	-0,026	0,009	-0,008	-0,001
entre 6 mois	-0,024	-0,047	-0,014	0,018	0,002	0,008	-0,008
entre 1 et 1,5	0,114	-0,059	0,032	0,003	-0,020	-0,015	0,003
entre 1,5 et 2	0,177	-0,039	0,047	0,009	0,066	0,002	0,011
entre 2 et 3a	0,332	-0,007	-0,047	-0,018	-0,019	0,021	0,012
plus de 3ans	0,375	0,112	0,011	0,004	0,003	-0,006	-0,009
Coordonnées standard (colonnes) :							
	F1	F2	F3	F4	F5	F6	F7
moins de 1m	-1,232	1,611	-1,712	2,985	-0,237	-0,882	1,961
entre 1 et 3 r	-1,068	0,693	1,246	-0,369	-0,367	0,744	0,190
entre 3 et 6m	-0,629	0,004	-1,072	-1,293	0,522	-0,796	-0,172
entre 6 mois	-0,126	-0,905	-0,382	0,885	0,086	0,741	-1,184
entre 1 et 1,5	0,591	-1,127	0,855	0,163	-1,105	-1,465	0,493
entre 1,5 et 2	0,912	-0,749	1,256	0,444	3,683	0,179	1,626
entre 2 et 3a	1,718	-0,136	-1,264	-0,880	-1,029	1,989	1,687
plus de 3ans	1,935	2,151	0,298	0,180	0,167	-0,563	-1,307

Figure 57: Les coordonnées des colonnes

Les coordonnées principales et les coordonnées standard des colonnes (lignes respectivement) sont également affichées. Les coordonnées standard sont les coordonnées principales divisées par la racine carrée de la valeur propre du facteur correspondant. La somme des carrés pondérée des coordonnées standard est égale à 1 pour chaque facteur.

3.6.6.3 Les contributions des colonnes

Contributions (colonnes) :								
	Poids (relatif)	F1	F2	F3	F4	F5	F6	F7
moins de 1m	0,046	0,070	0,120	0,136	0,412	0,003	0,036	0,178
entre 1 et 3 r	0,199	0,227	0,095	0,308	0,027	0,027	0,110	0,007
entre 3 et 6m	0,194	0,077	0,000	0,223	0,324	0,053	0,123	0,006
entre 6 mois	0,212	0,003	0,173	0,031	0,166	0,002	0,116	0,297
entre 1 et 1,5	0,143	0,050	0,182	0,105	0,004	0,175	0,307	0,035
entre 1,5 et 2	0,049	0,041	0,027	0,077	0,010	0,665	0,002	0,130
entre 2 et 3a	0,070	0,208	0,001	0,113	0,054	0,075	0,279	0,200
plus de 3ans	0,087	0,324	0,401	0,008	0,003	0,002	0,027	0,148

Figure 58: Les contributions des colonnes

Voici un tableau des contributions des colonnes (lignes respectivement). Les contributions correspondent à l'importance de chaque catégorie pour chaque facteur (dimension). La somme des contributions est égale à 1 pour chaque facteur. En règle générale, si la contribution est supérieure à $1/I$, I étant le nombre de lignes (respectivement $1/J$ avec J le nombre de colonnes), la catégorie est importante pour le facteur donné. Dans notre étude, le

groupe entre 6 mois et 1 année et entre 1.5 et 2 ans est important pour le facteur F1, les groupes entre 3 et 6 mois et entre 2 et 3 ans sont importants pour le facteur F2.

3.6.7 Graphique symétrique

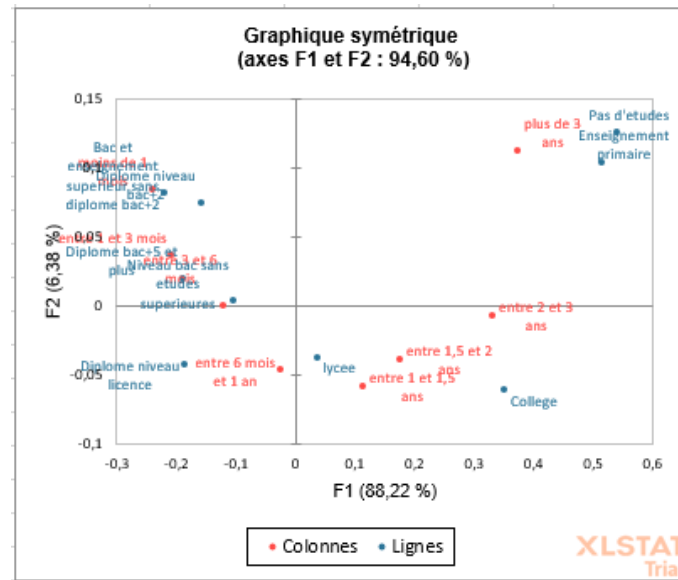


Figure 59: Graphique symétrique

Les différentes cartes sont alors affichées en commençant par le **tracé symétrique** des lignes et des colonnes ou tracé français qui est le plus couramment utilisé. Les profils de lignes et les profils de colonnes sont superposés dans un affichage conjoint (tous deux en coordonnées principales). Cet affichage est très pratique car les points de ligne et de colonne sont également répartis. La distance entre les points de lignes (respectivement les points de colonnes) se rapproche de la distance du chi carré interlignes (respectivement inter-colonnes). Les tranches Bac et enseignement supérieur sans diplôme bac+2 et Diplôme niveau bac+2 se superposent presque sur la carte symétrique, montrant des profils très similaires.

3.7 APPLIQUER AFC PAR PYTHON

3.7.1 Charger et afficher les données



Figure 60:Charger et afficher les données

On importe la librairie pandas pour charger les données, ainsi que la classe CA du package fanalysis. Les données sont transformées en matrice de type numpy.ndarray. Les données doivent se présenter sous forme de tableau croisé, avec des fréquences absolues (effectifs de chaque cellule)

	moins de 1\nmois	...	plus de 3\nans
Diplome bac+5 et plus	43	...	37
Diplome niveau licence	45	...	48
Diplome niveau bac+2	93	...	101
Bac et enseignement superieur sans diplome bac+2	63	...	70
Niveau bac sans etudes superieures	152	...	209
lycee	260	...	531
College	19	...	109
Enseignement primaire	17	...	168
Pas d'etudes	6	...	35

Figure 61:Les données affichées

Les données doivent se présenter sous forme de tableau croisé, avec des fréquences absolues (effectifs de chaque cellule)

3.7.2 Test de khi2

```
from scipy.stats import chi2_contingency
chi2,pvalue,degrees,expected = chi2_contingency(D)
print("chi2=",chi2," degrees=",degrees," pvalues=",pvalue)
```

Figure 62:Test de khi2

Donne le résultat suivant :

```
chi2= 641.0495356536996  degrees= 56  pvalues= 2.8656558768419315e-100
```

3.7.3 AFC en python

```
#importation de la librairie
from fanalysis.ca import CA
#lancer les calculs
afc = CA(row_labels=D.index,col_labels=D.columns)
afc.fit(D.values)
print(dir(afc))
```

Figure 63:Application de AFC

Pour lancer les calculs, nous importons la classe « CA » de « fanalysis », nous l’instancions et nous passons le tableau à traiter à la méthode fit(). Nous affichons les propriétés de l’objet.

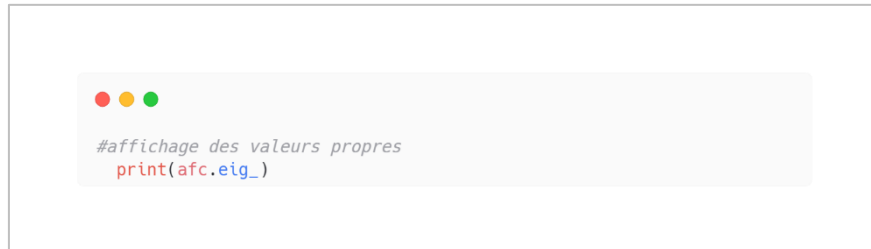
Donne le résultat suivant :

```
['__class__', '__delattr__', '__dict__', '__dir__', '__doc__', '__eq__', '__format__', '__ge__', '__getattr__', '__getstate__', '__gt__', '__hash__', '__init__', '__init_subclass__', '__le__', '__lt__', '__module__', '__ne__', '__new__', '__reduce__', '__reduce_ex__', '__repr__', '__setattr__', '__setstate__', '__sizeof__', '__str__', '__subclasshook__', '__weakref__', '_check_feature_names', '_check_n_features', '_compute_stats', '_compute_svd', '_get_param_names', '_get_tags', '_more_tags', '_repr_html_', '_repr_html_inner', '_repr_mimebundle_', '_validate_data', 'c_', 'col_contrib_', 'col_coord_', 'col_cos2_', 'col_labels', 'col_labels_', 'col_labels_short_', 'col_top_andas', 'eig_', 'fit', 'fit_transform', 'get_params', 'mapping', 'mapping_col', 'mapping_row', 'model_', 'n_', 'n_components', 'n_components_', 'plot_col_contrib', 'plot_col_cos2_', 'plot_eigenvalues', 'plot_row_contrib', 'plot_row_cos2_', 'r_', 'row_contrib_', 'row_coord_', 'row_cos2_', 'row_labels', 'row_labels_', 'row_topandas', 'set_params', 'stats', 'transform']
```

Figure 64:Les résultat de AFC

3.7.4 Les valeurs propres

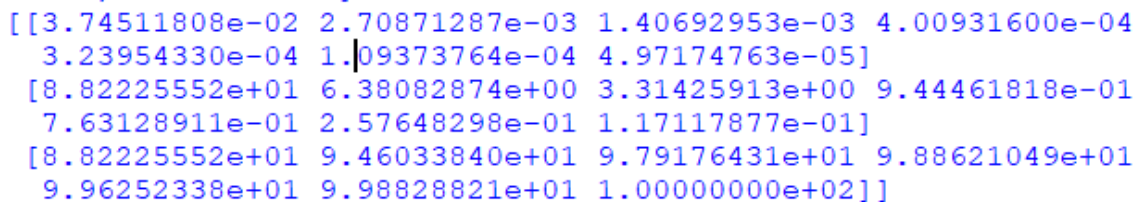
L'évaluation des valeurs propre permet de déterminer le nombre d'axes à interpréter, les valeurs propres représente la quantité d'information résumée par chaque axe (dim) Les valeurs propres et la repartitions de variances sur les différents axes peuvent être extraits à l'aide de `afc.eig_`:



```
#affichage des valeurs propres
print(afc.eig_)
```

Figure 65:l'objet `afc.eig_`

Donne le résultat suivant :



```
[[3.74511808e-02 2.70871287e-03 1.40692953e-03 4.00931600e-04
 3.23954330e-04 1.09373764e-04 4.97174763e-05]
 [8.82225552e+01 6.38082874e+00 3.31425913e+00 9.44461818e-01
 7.63128911e-01 2.57648298e-01 1.17117877e-01]
 [8.82225552e+01 9.46033840e+01 9.79176431e+01 9.88621049e+01
 9.96252338e+01 9.98828821e+01 1.00000000e+02]]
```

Figure 66:Les valeurs propres

Les valeurs propres sont fournies de 3 manières par « fanalysis » : les valeurs brutes, en pourcentages, et en pourcentages cumulés et pour afficher le graphique des valeurs propres en utilise la fonction `plot_eigenvalues`



```
#affichage graphique des v.p.
afc.plot_eigenvalues()
```

Figure 67 : la fonction `plot_eigenvalues`

Donne les résultat suivant :

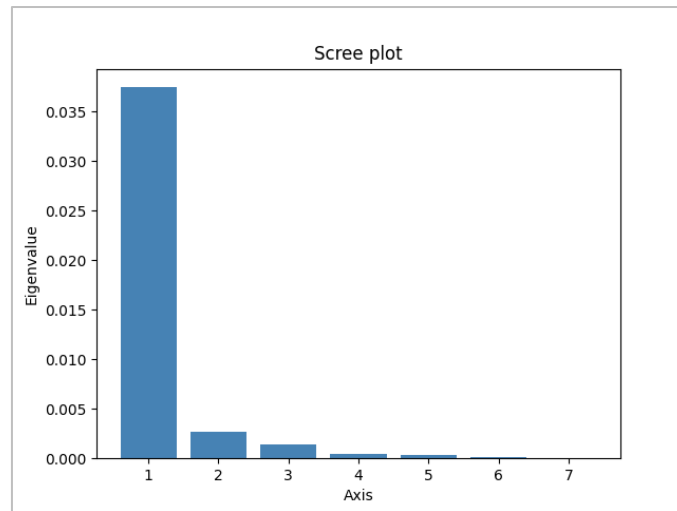


Figure 68:le graphe des valeurs propres

3.7.5 Résultat des lignes

3.7.5.1 Les Coordonnées des points lignes

```
#Coordonnees des points lignes
print(afc.row_coord)
```

Figure 69:L'objet row_coord

Résultat de l'exécution :

```
[[-0.1885493  0.01974418 -0.07717991 -0.02000662 -0.06555938  0.01576229
  0.0023655 ]
 [-0.18453703 -0.04287067  0.08323733 -0.01257079  0.00182032  0.02428614
 -0.00984518]
 [-0.15803873  0.07434254 -0.00270014 -0.04640239  0.02037397 -0.00558889
  0.00446541]
 [-0.21830511  0.08200586  0.02590797  0.04309871 -0.00363356  0.0017063
  0.01084843]
 [-0.10231228  0.00337058 -0.00842462  0.013357  0.00910966 -0.00100743
 -0.00379837]
 [ 0.03657518 -0.03769205 -0.00867071  0.00063508 -0.00172995 -0.00675002
 -0.00105159]
 [ 0.35205149 -0.06180637  0.00979276 -0.00554406  0.01172471  0.01648302
  0.02041522]
 [ 0.51506258  0.10341753  0.05790927 -0.00280238 -0.02336714 -0.00542737
 -0.0067131 ]
 [ 0.54091675  0.124916  -0.18751358  0.02472669  0.06075257  0.04358796
 -0.01872958]]
```

Figure 70:Les coordonnees des points lignes

3.7.5.2 Les Contribution des points lignes



Figure 71:L'objet row_contrib_

Résultat de l'exécution :

```
[[4.08593982e+00 6.19473914e-01 1.82240079e+01 4.29719211e+00
 5.71075488e+01 9.77762075e+00 4.84444979e-01]
[5.78654377e+00 4.31792737e+00 3.13387460e+01 2.50825902e+00
 6.50924291e-02 3.43179511e+01 1.24066827e+01]
[6.01937915e+00 1.84162829e+01 4.67723796e-02 4.84730705e+01
 1.15653215e+01 2.57767555e+00 3.61996201e+00]
[9.21036305e+00 1.79697188e+01 3.45309728e+00 3.35330293e+01
 2.94981853e-01 1.92668138e-01 1.71332415e+01]
[5.52495333e+00 8.29060378e-02 9.97163934e-01 8.79602067e+00
 5.06359319e+00 1.83425155e-01 5.73619497e+00]
[1.46109777e+00 2.14540031e+01 2.18579332e+00 4.11491562e-02
 3.77880347e-01 1.70399566e+01 9.09818156e-01]
[1.82113512e+01 7.76066376e+00 3.75087855e-01 4.21871622e-01
 2.33515684e+00 1.36696043e+01 4.61311597e+01]
[4.03879607e+01 2.25124526e+01 1.35900571e+01 1.11681560e-01
 9.61003169e+00 1.53554855e+00 5.16815018e+00]
[9.31241126e+00 6.86657148e+00 2.97892742e+01 1.81772606e+00
 1.35803933e+01 2.07055498e+01 8.41034579e+00]]
```

Figure 72:Les Contribution des points lignes

3.7.5.3 Les qualités des points lignes

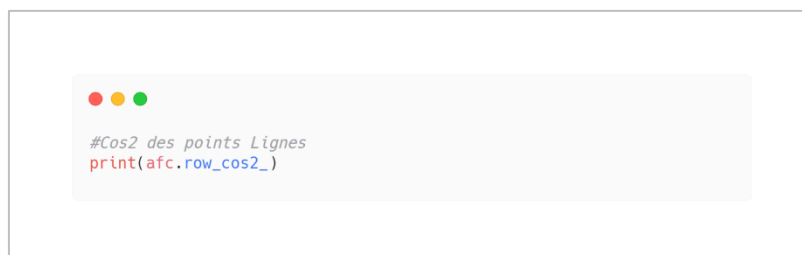


Figure 73:L'objet row.cos2_

Résultat de l'exécution :

```
[ [7.58826589e-01 8.32091438e-03 1.27145589e-01 8.54358445e-03
  9.17407673e-02 5.30311942e-03 1.19436919e-04]
[7.79830689e-01 4.20875717e-02 1.58660758e-01 3.61874713e-03
  7.58804386e-05 1.35067256e-02 2.21962862e-03]
[7.53890457e-01 1.66823021e-01 2.20065945e-04 6.49922916e-02
  1.25294651e-02 9.42827938e-04 6.01871971e-04]
[8.35436048e-01 1.17889520e-01 1.17666302e-02 3.25622212e-02
  2.31446190e-04 5.10380421e-05 2.06309658e-03]
[9.66826113e-01 1.04930894e-03 6.55530872e-03 1.64782380e-02
  7.66472963e-03 9.37402911e-05 1.33256138e-03]
[4.63901589e-01 4.92665756e-01 2.60712891e-02 1.39866043e-04
  1.03781313e-03 1.58002046e-02 3.83481786e-04]
[9.62920543e-01 2.96786817e-02 7.45054625e-04 2.38799476e-04
  1.06802794e-03 2.11082326e-03 3.23806987e-03]
[9.47575632e-01 3.82016594e-02 1.19781589e-02 2.80509906e-05
  1.95031631e-03 1.05213909e-04 1.60968407e-04]
[8.36190846e-01 4.45944347e-02 1.00487031e-01 1.74733617e-03
  1.05480914e-02 5.42972300e-03 1.00253761e-03]]
```

Figure 74: Les qualites des points lignes

3.7.6 Résultat des colonnes

3.7.6.1 Les Coordonnées des points colonnes

```
#Coordonnees des points colonnes
print(afc.col_coord_)
```

Figure 75: L'objet col_coord_

Résultat de l'exécution :

```
[ [-2.38482385e-01 8.38402127e-02 -6.42234788e-02 -5.97732179e-02
  4.27051872e-03 -9.22079998e-03 1.38250470e-02]
[-2.06680832e-01 3.60593317e-02 4.67231964e-02 7.39781900e-03
  6.59987561e-03 7.78591298e-03 1.34103460e-03]
[-1.21766078e-01 2.24783551e-04 -4.02244656e-02 2.58809476e-02
  -9.39254740e-03 -8.32406811e-03 -1.21541051e-03]
[-2.43724215e-02 -4.70765020e-02 -1.43273468e-02 -1.77226227e-02
  -1.54164739e-03 7.74773875e-03 -8.34600616e-03]
[ 1.14453229e-01 -5.86480415e-02 3.20518967e-02 -3.27035546e-03
  1.98808337e-02 -1.53182780e-02 3.47443181e-03]
[ 1.76542243e-01 -3.89794756e-02 4.71240739e-02 -8.89595086e-03
  -6.62849228e-02 1.87665237e-03 1.14630423e-02]
[ 3.32482860e-01 -7.10115495e-03 -4.74240336e-02 1.76142236e-02
  1.85231855e-02 2.08053742e-02 1.18963022e-02]
[ 3.74501934e-01 1.11950177e-01 1.11874789e-02 -3.59674750e-03
  -3.01286578e-03 -5.88341871e-03 -9.21398072e-03]]
```

Figure 76: Les coordonnées des points lignes

3.7.6.2 Les Contributions des points colonnes



```
#Contribution des points colonnes
print(afc.col_contrib_)
```

Figure 77: L'objet `col_contrib_`

Résultat de l'exécution :

```
[[7.01934861e+00 1.19947579e+01 1.35508027e+01 4.11900862e+01
 2.60211919e-01 3.59313685e+00 1.77694453e+01]
 [2.26595060e+01 9.53647605e+00 3.08253698e+01 2.71176881e+00
 2.67118064e+00 1.10108660e+01 7.18598717e-01]
 [7.68417468e+00 3.62056373e-04 2.23212408e+01 3.24265070e+01
 5.28558562e+00 1.22961231e+01 5.76696098e-01]
 [3.36000851e-01 1.73322096e+01 3.09077743e+00 1.65956485e+01
 1.55415684e-01 1.16264013e+01 2.96795191e+01]
 [5.00772609e+00 1.81800067e+01 1.04540628e+01 3.81917547e-01
 1.74676861e+01 3.07154381e+01 3.47622991e+00]
 [4.07809913e+00 2.74874862e+00 7.73462610e+00 9.67253869e-01
 6.64617439e+01 1.57790378e-01 1.29514003e+01]
 [2.07778480e+01 1.31045637e-01 1.12525822e+01 5.44733745e+00
 7.45547163e+00 2.78589958e+01 2.00374504e+01]
 [3.24372966e+01 4.00763935e+01 7.70538163e-01 2.79480589e-01
 2.42704494e-01 2.74124840e+00 1.47906602e+01]]
```

Figure 78: Les contributions des points colonnes

3.7.6.3 Les qualités des points colonnes



```
#Cos2 des points colonnes
print(afc.col_cos2_)
```

Figure 79: Les qualités des points colonnes

Résultat de l'exécution :

```
[ [7.91069137e-01 9.77702157e-02 5.73706120e-02 4.96952756e-02
  2.53666484e-04 1.18260315e-03 2.65849018e-03]
 [9.21398652e-01 2.80467482e-02 4.70882185e-02 1.18046896e-03
  9.39547345e-04 1.30757395e-03 3.87906569e-05]
 [8.58348348e-01 2.92509516e-06 9.36680566e-02 3.87767395e-02
  5.10713966e-03 4.01127276e-03 8.55178789e-05]
 [1.71599470e-01 6.40216634e-01 5.92993767e-02 9.07349424e-02
  6.86576107e-04 1.73407757e-02 2.01222243e-02]
 [7.18999588e-01 1.88790254e-01 5.63871327e-02 5.87032925e-04
  2.16941046e-02 1.28793052e-02 6.62582755e-04]
 [7.88742723e-01 3.84512151e-02 5.61983728e-02 2.00273199e-03
  1.11190477e-01 8.91262736e-05 3.32535392e-03]
 [9.69079020e-01 4.42057382e-04 1.97159473e-02 2.71986707e-03
  3.00782124e-03 3.79465003e-03 1.24063676e-03]
 [9.16370956e-01 8.18865835e-02 8.17763357e-04 8.45245927e-05
  5.93092941e-05 2.26163596e-04 5.54699512e-04]]
```

Figure 80: Les qualités des points colonnes

3.7.7 Représentation simultanée

```
#représentation simultanée
afc.mapping(num_x_axis=1,num_y_axis=2,figsize=(7,7))
```

Figure 81: la fonction mapping()

Résultat de l'exécution :

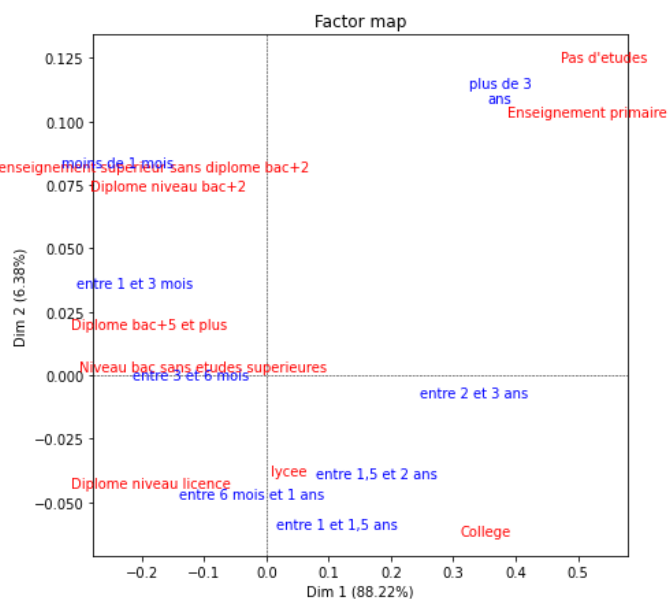


Figure 82: Représentation simultanée

4. INTERPRETATION

4.1 PROFILS LIGNES

Profils lignes									
V1	V2								Marge active
	entre 1 et 1,5 ans	entre 1 et 3 mois	entre 1,5 et 2 ans	entre 2 et 3 ans	entre 3 et 6 mois	entre 6 mois et 1 an	moins de 1 mois	plus de 3 ans	
Bac et enseignement supérieur sans diplôme bac+2	,113	,266	,038	,045	,226	,189	,058	,064	1,000
College	,184	,122	,067	,118	,143	,212	,023	,131	1,000
Diplome bac+5 et plus	,102	,218	,048	,051	,232	,226	,066	,057	1,000
Diplome niveau bac+2	,123	,239	,037	,049	,201	,208	,068	,074	1,000
Diplome niveau licence	,139	,255	,047	,043	,193	,227	,047	,050	1,000
Enseignement primaire	,174	,115	,075	,123	,122	,175	,020	,195	1,000
lycee	,152	,182	,052	,075	,193	,218	,042	,086	1,000
Niveau bac sans etudes superieures	,134	,220	,042	,060	,211	,213	,051	,070	1,000
Pas d'etudes	,144	,067	,044	,161	,161	,194	,033	,194	1,000
Masse	,143	,199	,049	,070	,194	,212	,046	,087	

Figure 83: Profils lignes

Ce tableau se lit ligne par ligne (chaque niveau d'enseignement correspond à une durée d'ancienneté de chômage). Par exemple, pour les personnes qui ont le diplôme Bac+5 et plus, 6,6% ont une ancienneté dans le chômage de moins d'un mois

4.2 PROFILS COLONNES

Profils colonnes									
V1	V2								Masse
	entre 1 et 1,5 ans	entre 1 et 3 mois	entre 1,5 et 2 ans	entre 2 et 3 ans	entre 3 et 6 mois	entre 6 mois et 1 an	moins de 1 mois	plus de 3 ans	
Bac et enseignement supérieur sans diplôme bac+2	,057	,097	,057	,046	,084	,065	,090	,054	,072
College	,071	,034	,076	,092	,041	,055	,027	,083	,055
Diplome bac+5 et plus	,031	,047	,042	,031	,052	,046	,062	,028	,043
Diplome niveau bac+2	,078	,109	,068	,063	,093	,089	,133	,077	,090
Diplome niveau licence	,062	,082	,061	,039	,063	,068	,064	,037	,064
Enseignement primaire	,069	,033	,088	,100	,036	,047	,024	,128	,057
lycee	,435	,376	,431	,435	,406	,421	,372	,406	,409
Niveau bac sans etudes superieures	,185	,219	,168	,167	,215	,198	,218	,160	,198
Pas d'etudes	,012	,004	,011	,027	,010	,011	,009	,027	,012
Marge active	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	

Figure 84: Profils colonnes

Les personnes qui ont une ancienneté dans le chômage de moins d'un mois, 6,5% ont le diplôme Bac+5 et plus et 13% ont le niveau diplôme Bac+2

4.3 RECAPITULATIF

Récapitulatif								
Dimension	Valeur singulière	Inertie	Khi-deux	Sig.	Proportion d'inertie		Valeur singulière de confiance	
					Représentation	Cumulé	Ecart type	Corrélation 2
1	,194	,037			,882	,882	,008	,150
2	,052	,003			,064	,946	,009	
3	,038	,001			,033	,979		
4	,020	,000			,009	,989		
5	,018	,000			,008	,996		
6	,010	,000			,003	,999		
7	,007	,000			,001	1,000		
Total		,042	641,050	<,001 ^a	1,000	1,000		
a. 56 degrés de liberté								

Figure 85:Récapitulatif

Ce tableau présente les contributions. Il permet de savoir comment les modalités de la variable ligne (ici niveau d'enseignement) expliquent les axes. Il permet également de savoir la façon dont ces modalités sont expliquées par les axes. • Les scores de la dimension permettent de placer les modalités sur les axes retenus, ils représentent, en quelques sortes, les coordonnées des modalités lignes.

Les scores : • Le premier axe oppose les modalités dont les coordonnées sont négatives (les diplômes supérieurs au Bac) aux modalités dont les coordonnées sont positives (les diplômes inférieurs au Bac).

4.4 PRESENTATION DES POINTS DE LIGNE

Présentation des points de ligne ^a									
V1	Masse	Score de la dimension		Inertie	Contribution				
		1	2		Du point vers l'inertie de la dimension		De la dimension vers l'inertie du point		
					1	2	1	2	Total
Bac et enseignement supérieur sans diplôme bac+2	,072	-,496	,359	,004	,092	,180	,835	,118	,953
College	,055	,800	-,271	,007	,182	,078	,963	,030	,993
Diplome bac+5 et plus	,043	-,429	,087	,002	,041	,006	,759	,008	,767
Diplome niveau bac+2	,090	-,359	,326	,003	,060	,184	,754	,167	,921
Diplome niveau licence	,064	-,419	-,188	,003	,058	,043	,780	,042	,822
Enseignement primaire	,057	1,171	,453	,016	,404	,225	,948	,038	,986
lycee	,409	,083	-,165	,001	,015	,215	,464	,493	,957
Niveau bac sans etudes superieures	,198	-,233	,015	,002	,055	,001	,967	,001	,968
Pas d'etudes	,012	1,230	,548	,004	,093	,069	,836	,045	,881
Total actif	1,000			,042	1,000	1,000			
a. Normalisation symétrique									

a. Normalisation symétrique

Figure 86:Présentation des points de ligne

Pour la variable Niveau d'enseignement, la contribution la plus forte à la dimension 1

est celle de la tranche Enseignement primaire (40,4%). La dimension 2 quant à elle, résulte de la contribution de la tranche Enseignement primaire (22%). Les tranches d'Enseignement primaire et collège sont les mieux représentées sur la dimension 1 (respectivement 44% et 19%), tandis que la dimension 2 représente mieux la tranche d'Enseignement primaire (22%)

4.5 PRESENTATION DES POINTS DE COLONNES

Présentation des points de colonne ^a									
V2	Masse	Score de la dimension		Inertie	Contribution				
		1	2		Du point vers l'inertie de la dimension		De la dimension vers l'inertie du point		
					1	2	1	2	Total
entre 1 et 1,5 ans	,143	,260	-,257	,003	,050	,182	,719	,189	,908
entre 1 et 3 mois	,199	-,470	,158	,009	,227	,095	,921	,028	,949
entre 1,5 et 2 ans	,049	,401	-,171	,002	,041	,027	,789	,038	,827
entre 2 et 3 ans	,070	,756	-,031	,008	,208	,001	,969	,000	,970
entre 3 et 6 mois	,194	-,277	,001	,003	,077	,000	,858	,000	,858
entre 6 mois et 1 ans	,212	-,055	-,206	,001	,003	,173	,172	,640	,812
moins de 1 mois	,046	-,542	,368	,003	,070	,120	,791	,098	,889
plus de 3 ans	,087	,851	,491	,013	,324	,401	,916	,082	,998
Total actif	1,000			,042	1,000	1,000			
a. Normalisation symétrique									

Figure 87: Présentation des points de colonne

Pour la variable ancienneté du chômage, la contribution la plus forte à la dimension 1 est celle de plus de 3 ans (32,4%). La dimension 2 quant à elle, résulte de la contribution de la modalité plus de 3 ans (40%). La modalité entre 2 et 3 ans est la mieux représentée sur la dimension 1 (96,9%), tandis que la dimension 2 représente mieux la modalité entre 6 mois et 1 ans (64%). On peut donc affirmer que sur la dimension 1, il y a une certaine correspondance entre la tranche plus de 3 ans et plus et la modalité enseignement primaire, alors que la dimension 2 met en correspondance la tranche d'plus de 3 ans et la modalité d'enseignement primaire.

4.6 GRAPHE DES POINTS DE LIGNES ET DE COLONNES

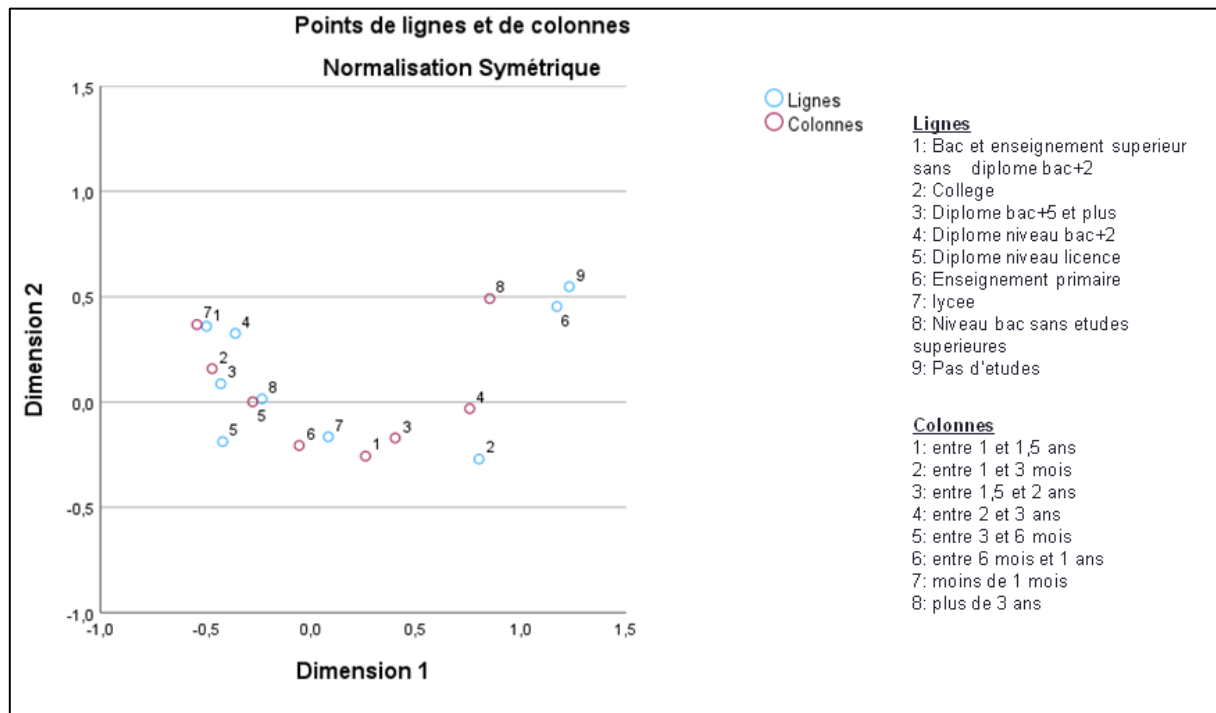


Figure 88: Graphe des points de lignes et de colonnes

Ce diagramme confirme le résultat des tableaux des contributions, en effet, on peut voir sur l'axe horizontal une certaine correspondance entre la tranche plus de 3 ans et plus et la modalité enseignement primaire, par contre sur l'axe vertical, on peut voir une nette opposition entre la tranche d'plus de 3 ans et la modalité d'enseignement primaire.

D'après les résultats obtenus, on peut conclure que : La projection sur le premier axe factoriel montre que les personnes ayant des diplômes inférieures au Bac ont plus du mal à trouver un travail que ceux ayant des diplômes supérieures au Bac. Ainsi, on note que la durée d'ancienneté de chômage dépasse 2 ans ("entre 2 et 3 ans" et "plus de 3 ans") pour les niveaux "Enseignement primaire", "collège" et Par ailleurs, pour les niveaux "Bac" et "Bac+2", la durée d'ancienneté de chômage ne dépasse pas 3 ans.

5. CONCLUSION

L'exécution de la méthode AFC avec Python et Excel donne des résultats similaires à R. avec une préférence pour ce dernier. L'analyse des données avec R est plus simple et plus rapide que celle avec Python.