

**Projet analyse de données : Les défis de la femme travailleuse**

**Encadré par :**

* **Mr. BARRY Khalidou**
* **Mr. HARTI Mostapha**

**Réalisé par :**

* **CHAOUI Marwa**
* **TAGMOUTI Ghita**
* **BELHAJ Ikram**
* **ABDELLAH Qatre en nada**
* **EL IDRISSI Yassine**
* **LAKHAL Khalid**

**Sommaire :**

1 – Introduction

2 – Enquete et la collecte des données

3 – Langage R

4 – Tanagra

5 – Régression Linéaire

6 – Analyse en composantes principales (ACP)

7 – Analyse factorielle des correspondances (AFC)

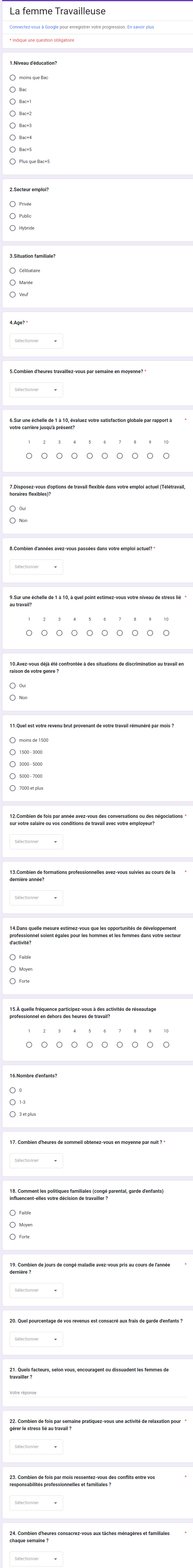
8 – Conclusion

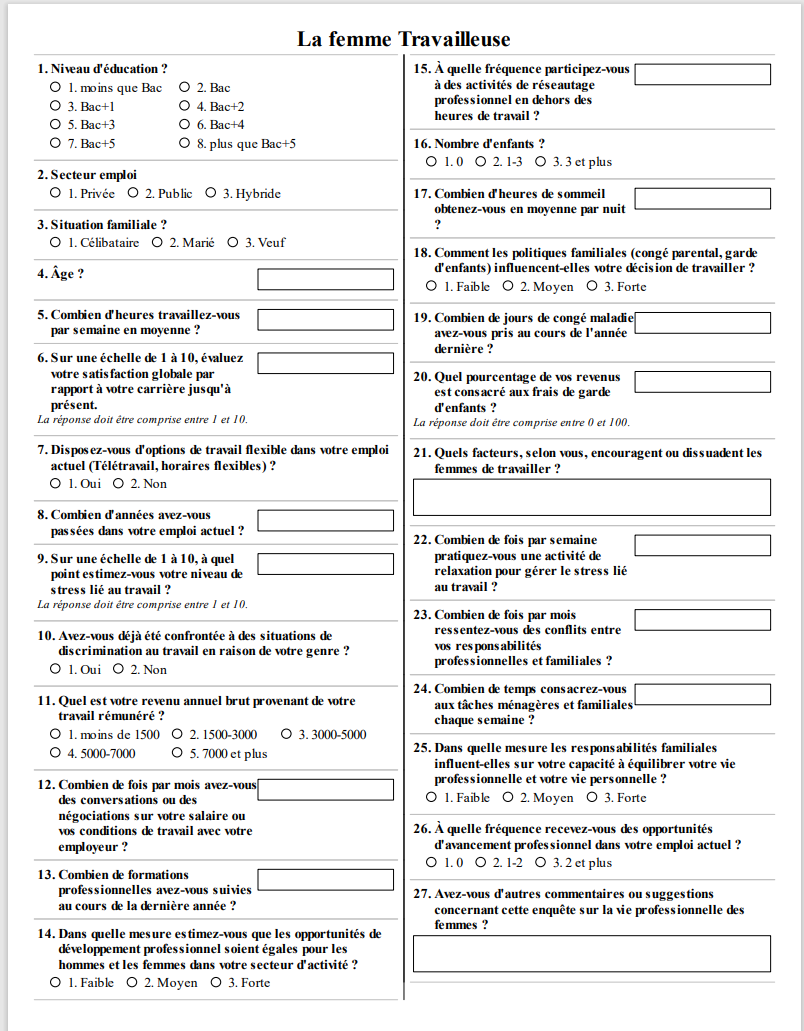
Introduction

Dans le contexte actuel du Maroc, l'équilibre entre vie professionnelle et familiale des femmes est un enjeu crucial. Cette étude examine diverses variables telles que l'âge, le secteur d'emploi et la satisfaction professionnelle pour mieux comprendre les défis et les opportunités auxquels sont confrontées ces femmes. En analysant ces données, l'objectif est de dégager des tendances significatives afin d'éclairer les politiques et les pratiques visant à soutenir leur intégration professionnelle et leur épanouissement personnel, contribuant ainsi à une société plus équitable et inclusive.

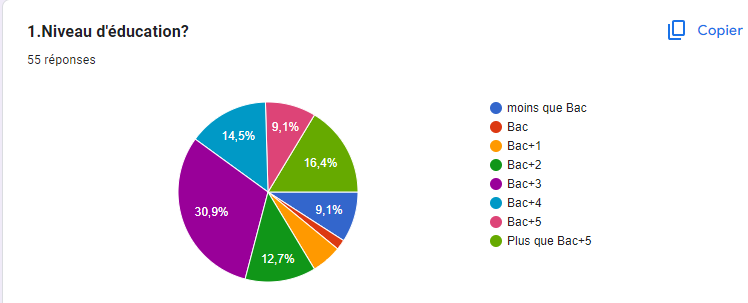
Cette étude vise également à sensibiliser davantage l'opinion publique et les décideurs sur l'importance de créer un environnement favorable pour les femmes sur le marché du travail. En mettant en lumière les défis spécifiques auxquels elles sont confrontées et en identifiant les leviers d'action potentiels, elle aspire à encourager des initiatives politiques et sociales qui favorisent l'égalité des chances et la promotion de la diversité dans tous les secteurs de l'économie. En renforçant la reconnaissance et la valorisation du rôle des femmes dans la société, cette démarche contribue à créer un climat propice à leur épanouissement professionnel et personnel, tout en favorisant un développement durable et inclusif pour l'ensemble de la population marocaine.

**Enquete et la collecte des données**

* **Nous avons entrepris une enquête approfondie portant sur les femmes actives, en utilisant une méthodologie mixte. Pour ce faire, nous avons combiné l'utilisation de deux logiciels : Sphinx, pour la distribution du questionnaire sur support papier, et Google Forms, pour sa diffusion en ligne. Dans le cadre de cette enquête, un total de 27 questions a été posé, couvrant divers aspects liés à la vie professionnelle des femmes. Nous avons reçu un total de 55 réponses, représentant ainsi une base de données substantielle pour notre analyse. Toutefois, pour garantir la qualité des données, nous avons procédé à un nettoyage préliminaire du fichier initial. Cela a impliqué la suppression des valeurs aberrantes et la substitution des valeurs manquantes par la moyenne de chaque question, assurant ainsi la cohérence et la fiabilité de notre ensemble de données.**





**Signification de chaque variable :**

niveau d'education : ne

secteur d'emploi : se

situation familiale : sf

heure de travaille par semaine : hts

satisfction globale : sg

travaille flexible :tf

annees emploi actuel : aea

niveau de stress lie au travail : st

discrimination au travail : disc

revenue mensielle : rm

negociation salaire : negs

formation pro : fp

egalite pro : ep

acctivite de reseautage pro : arp

nombre enfants : enfants

heures sommeils / nuit : sn

impact politique sur l'emploi : ipe

congee maladie annee / jour : cmj

pourcentage de revenue consacre aux frais de garde d'enfants : prfe

acctivites de relaxation par semaine : ars

conflits entre responsab prof et familiale / mois : conf

heures des taches menagers/semaine : hms

frequence d'opportunites pro : foap

Celibataire 0 Marié 1 Divorcé 2 Veuf 3

-secteur d'emploi : public => 0 privé => 1 hybrid => 2

-travail flexible : oui => 0 non => 1

-situation de discrimination : oui => 0 non => 1

-egalite professionnelle : faible => 0 moyen => 1 forte => 2

-l'impact de politique : faible => 0 moyen => 1 forte => 2

**Langage R**

Le langage R est un outil essentiel en analyse de données, offrant des fonctionnalités avancées pour explorer, manipuler et visualiser des données. Dans notre enquête sur la femme travailleuse, R nous permet d'analyser efficacement les données recueillies, en identifiant des tendances, des corrélations et des facteurs influents. Cela nous aide à prendre des décisions éclairées pour promouvoir l'égalité des sexes sur le lieu de travail.



**Tanagra**

Tanagra, un logiciel d'analyse de données, offre un outil précieux pour explorer les tendances des travailleuses au Maroc. En intégrant Tanagra dans notre présentation, on peut extraire des insights significatifs et formuler des recommandations pertinentes pour promouvoir l'égalité des sexes sur le marché du travail marocain.



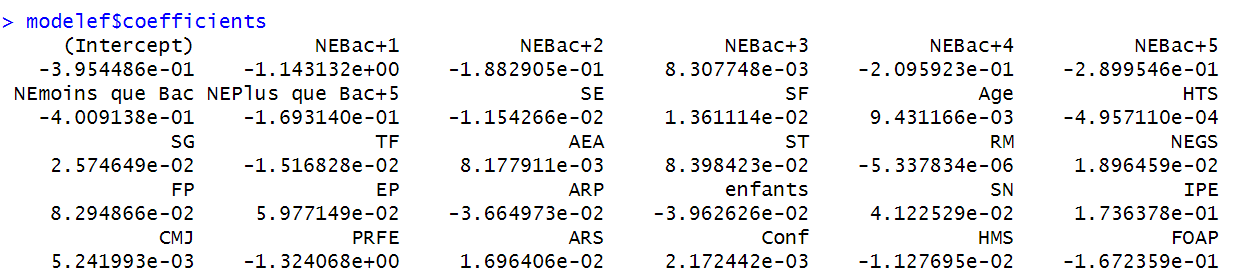
**Régression Linéaire**

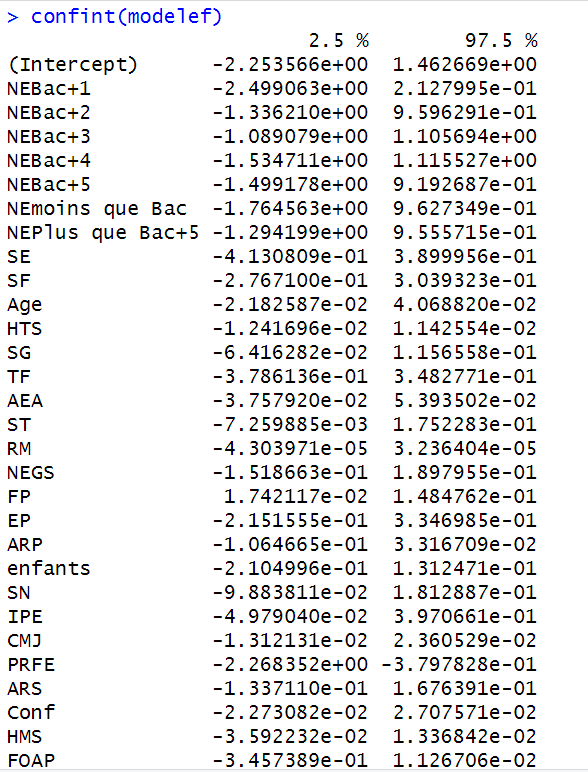
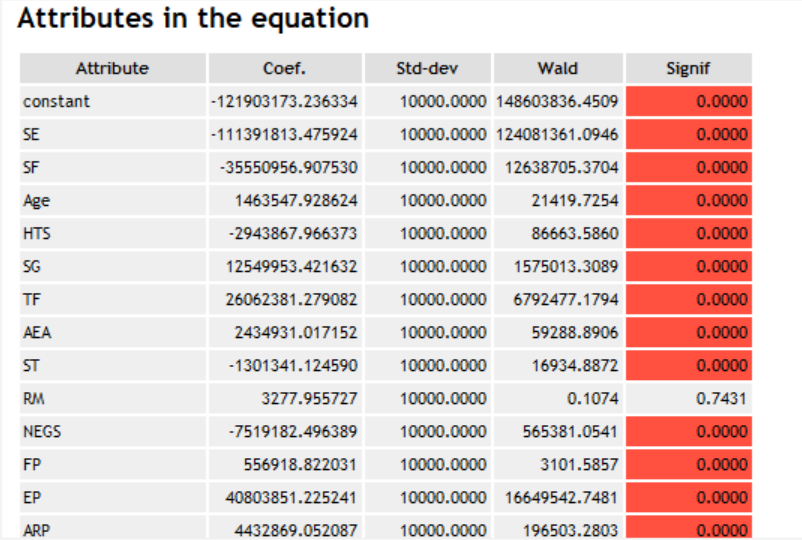
la Régression Linéaire : Une méthode statistique pour modéliser la relation entre une variable dépendante et une ou plusieurs variables indépendantes. En résumé, elle permet de prédire la valeur d'une variable en fonction des valeurs d'autres variables.

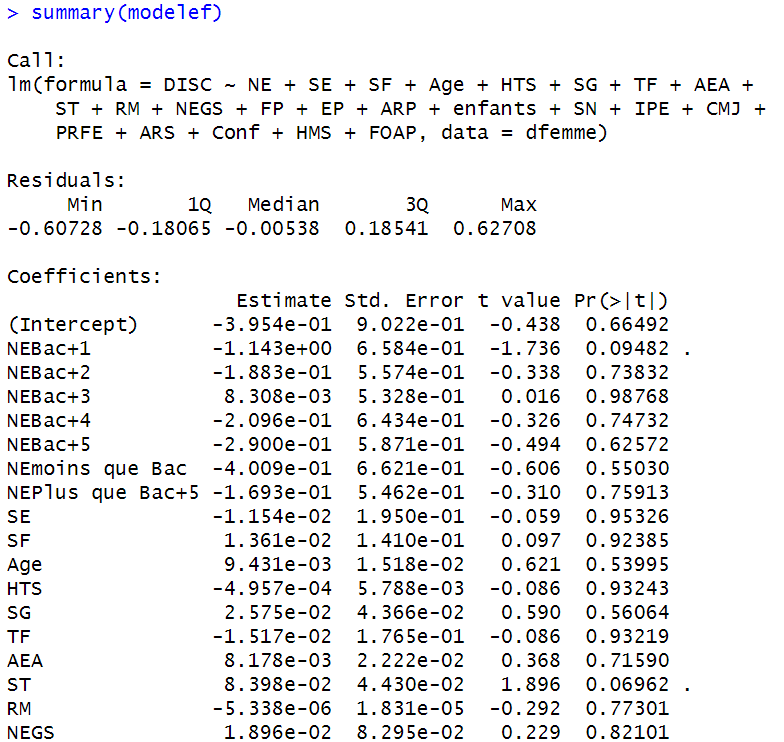
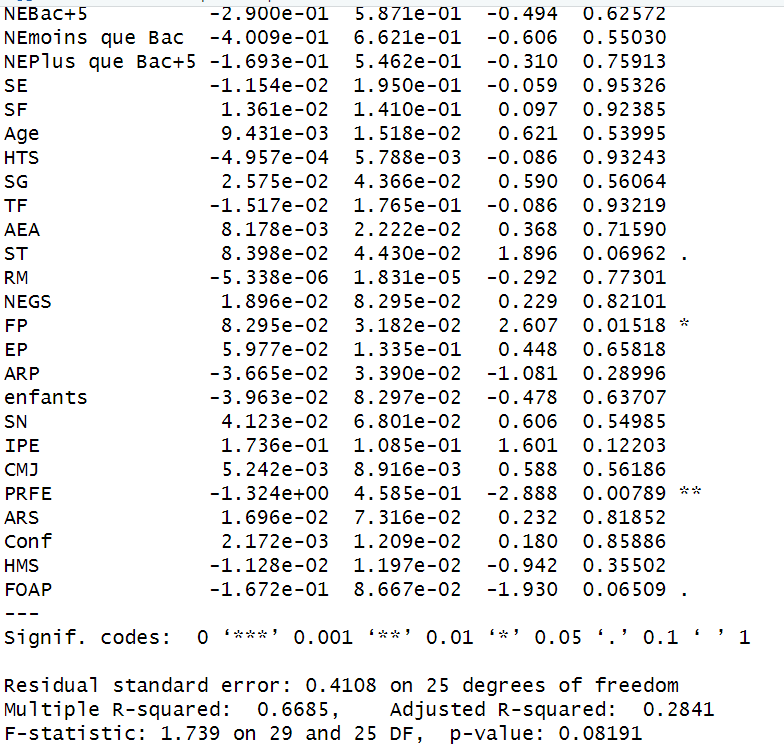
**1) Première étape :**

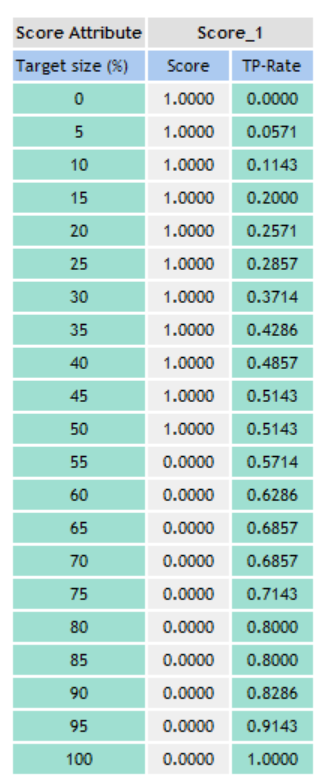
Préparation des données et ajustement du modèle : Dans cette étape, nous chargeons les données nécessaires et ajustons un modèle de régression linéaire à l'aide de la fonction `lm()`. Nous utilisons toutes les variables disponibles dans notre modèle initial, appelé `modelef`.

**2) 2éme étape:**

Évaluation du modèle : Nous évaluons ensuite la performance du modèle complet et modèle incomplet en examinant les coefficients de régression, l'intervalle de confiance et les statistiques de résumé.

**Les coefficients de régression** : Dans un modèle de régression linéaire, chaque variable indépendante est associée à un coefficient de régression. Ce coefficient représente l'effet de cette variable sur la variable dépendante, en supposant que toutes les autres variables restent constantes. Par exemple, si le coefficient pour l'âge est 0.5, cela signifie que pour chaque augmentation d'un an, la variable dépendante augmentera de 0.5, toutes choses étant égales par ailleurs.

**L'intervalle de confiance** : L'intervalle de confiance est une estimation de la plage dans laquelle la vraie valeur du coefficient de régression se situe. Il est généralement calculé à un niveau de confiance de 95%, ce qui signifie que si l'expérience était répétée de nombreuses fois, 95% des intervalles de confiance calculés contiendraient la vraie valeur du coefficient. Si l'intervalle de confiance pour un coefficient ne comprend pas zéro, cela signifie que l'effet de cette variable sur la variable dépendante est significatif à ce niveau de confiance.

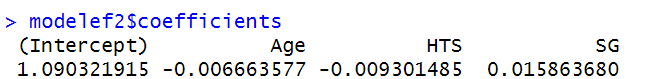
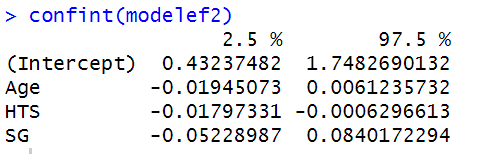
****

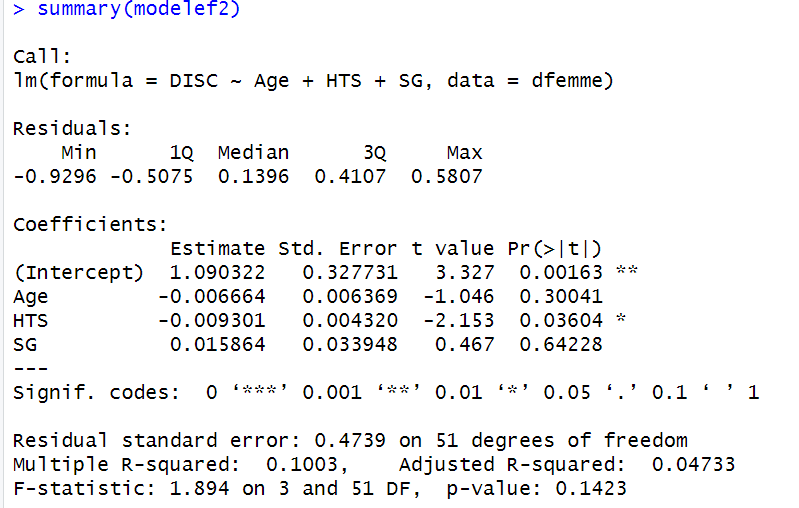
**Les statistiques de résumé :** Les statistiques de résumé fournissent des informations sur la qualité de l'ajustement du modèle de régression. Elles comprennent le R carré, qui est une mesure de la proportion de la variance de la variable dépendante qui est expliquée par les variables indépendantes. Elles comprennent également la statistique F et la p-value associée, qui sont utilisées pour tester l'hypothèse que tous les coefficients de régression sont nuls. Si la p-value est inférieure à 0.05 ( alpha ) => on rejette H0 ,ce qui signifie que le modèle de régression est significatif.

* Dans ce cas on constate que tous les paramètres sont nuls sauf : PRFE,FP

**3) 3éme étape :**

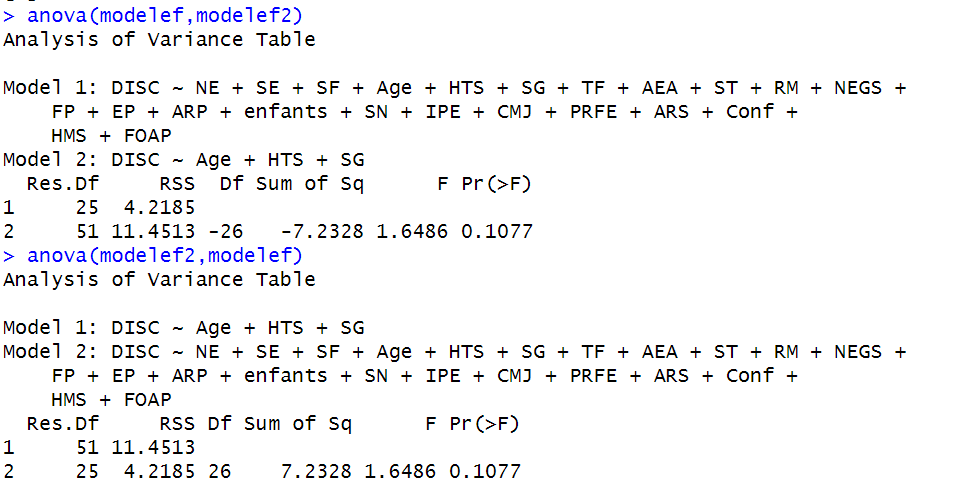
Simplification du modèle : nous simplifions ensuite notre modèle en n'utilisant que les variables les plus significatives. Nous ajustons un nouveau modèle, `modelef2`, en utilisant seulement trois variables : Age, HTS et SG.



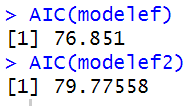
* d’après le calcul des statistiques de résumé du modelf2 ( incomplet ) => on constate que tous les paramètres sont nuls sauf : HTS

**4) Quatrième étape :**

Comparaison des modèles : Nous comparons ensuite les performances de nos deux modèles en utilisant l'analyse de la variance (ANOVA) et les critères d'information d'Akaike (AIC) et de Bayes (BIC).

****

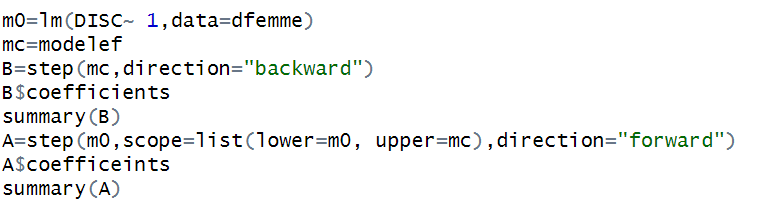
d’après l’utilisation la variance ANOVA pour comparer les deux modèles nous avons trouves que leur p-value sont égaux donc on les recomparent en utilisant les critères AIC et BIC

****

la valeur de AIC(modelef) est plus faible que la valeur de AIC(modelef2) , donc le modèle ‘modelef’ complet est le meilleur modèle.

**5) Cinquième étape** :

Sélection du modèle final : Enfin, nous utilisons la méthode de sélection de modèle pas à pas pour déterminer le modèle final le plus approprié.

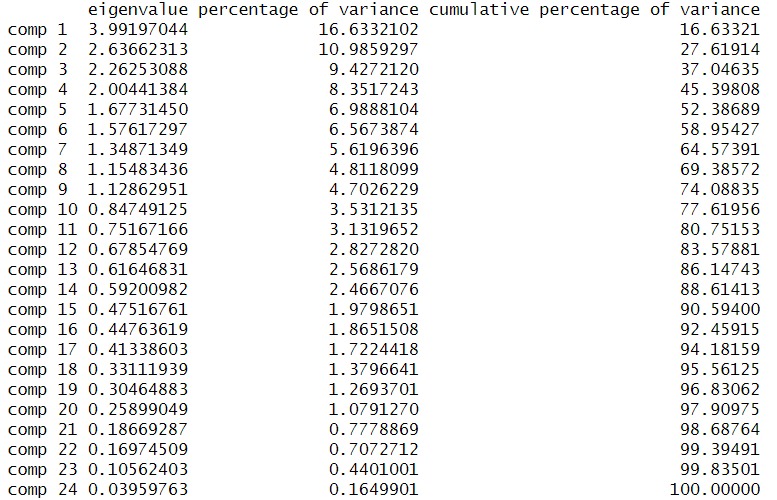


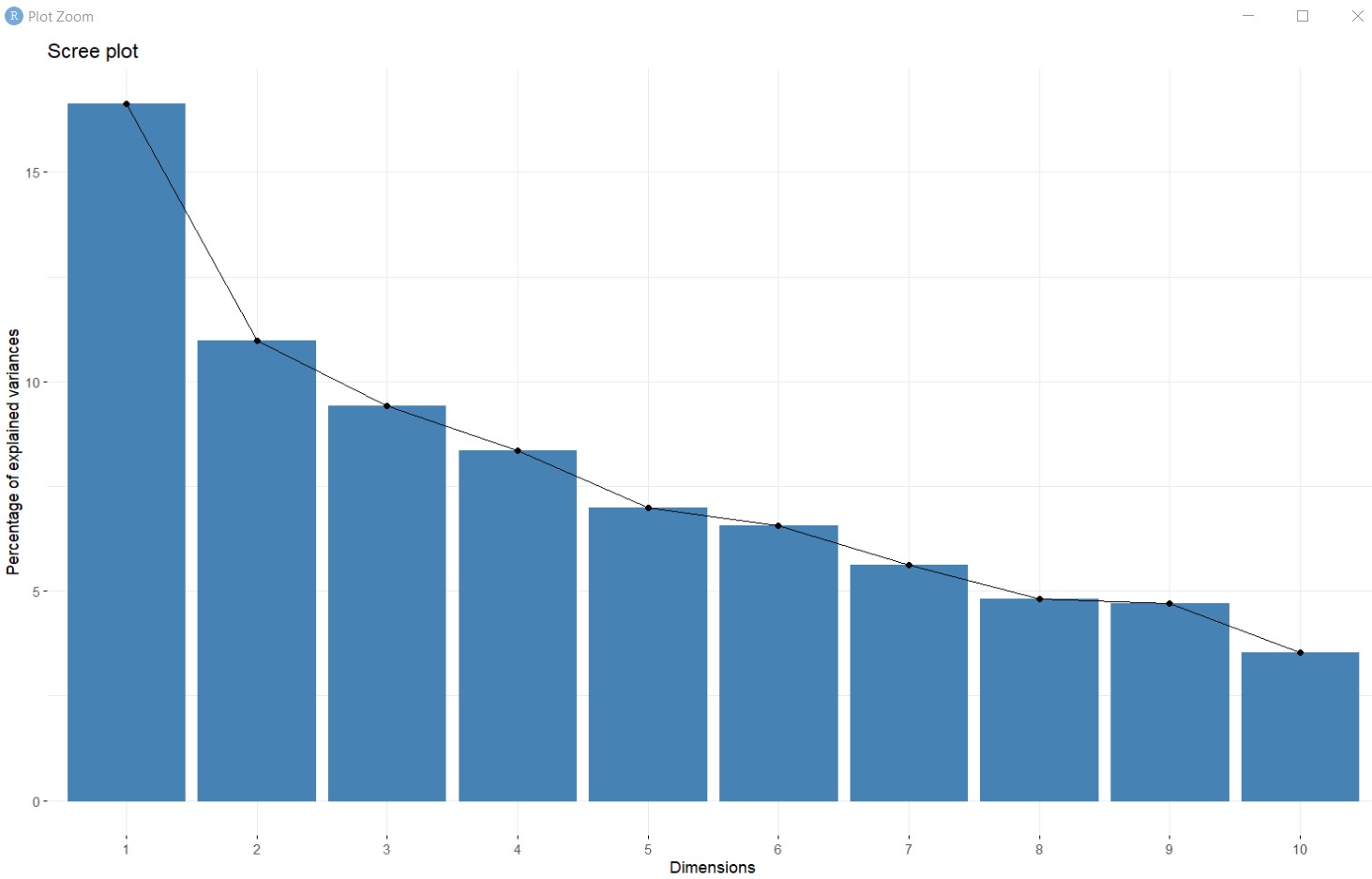
ces démarches sont utilisées pour trouver le modèle le plus simple qui prédit encore bien la variable dépendante. Cela nous aide à éviter le surajustement, où un modèle complexe s'adapte très bien aux données d'entraînement mais ne généralise pas bien aux nouvelles données.

**Analyse en composantes principales (ACP)**

L'Analyse en Composantes Principales (ACP) est une méthode statistique pour réduire la dimensionnalité des données tout en préservant leur structure essentielle. En résumé, elle permet de visualiser et de comprendre les relations entre les variables en les projetant dans un espace de dimensions réduites appelées composantes principales.

1. **Première Etape :**

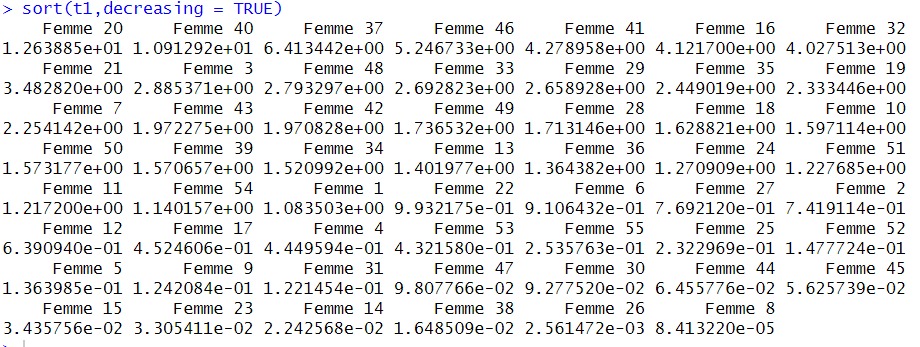
* **le critère de kaiser :** En choisissant de conserver neuf axes principaux, nous nous concentrons sur les aspects les plus importants de nos données, simplifiant ainsi notre analyse tout en préservant les principales tendances et variations.
* **critère du coude**: nous constatons un changement notable dans la pente de la courbe de l'inertie donc on va admettre 9 axes.



**=>** Nous pouvons donc admettre qu'à ce point, d’après le deux critère qu’on va admettre 9 axes.

1. **2ème Etape :**

**2.1) 1er Axe :**

* **Pour les individus :** On à 55 individus dons le seuil est : 1,81%

Donc on constate que tous les contributions qui sont supérieur a 1,81% :

-Femme 20 : 12.63885

-Femme 40 : 10.91292

-Femme 37 : 6.413442

-Femme 46 : 5.246733

-Femme 41 : 4.278958

-Femme 16 : 4.1217

-Femme 32 : 4.027513

-Femme 21 : 3.48282

-Femme 3 : 2.885371

-Femme 48 : 2.793297

-Femme 33 : 2.692823

-Femme 29 : 2.658928

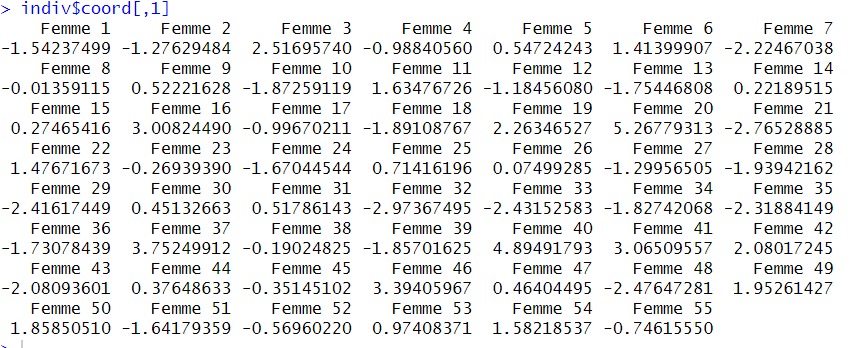
-Femme 35 : 2.449019

-Femme 19 : 2.333446

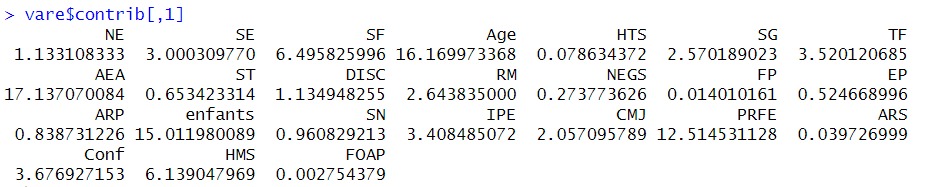
-Femme 7 : 2.254142

-Femme 43 : 1.972275

-Femme 42 : 1.970828



* **Pour les variables :** On a 24 variables donc : 1/24=4.16



**Selon la contribution** les variables dont les contributions sur le premier axe de l'ACP sont supérieures à 4,16 sont considérées comme ayant une influence significative sur la variation observée le long de cet axe. Voici les variables correspondantes :

- Âge (Age) : 16,17

- Satisfaction globale (SG) : 6,49

- Années d'emploi actuel (AEA) : 17,14

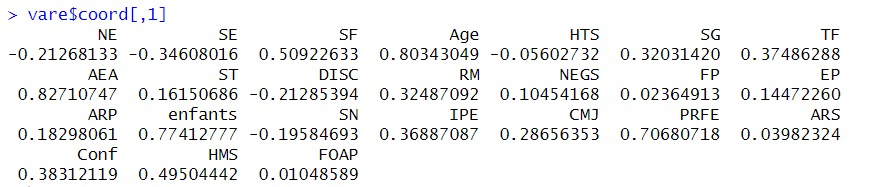
- Impact politique sur l'emploi (IPE) : 4,41

-Pourcentage de revenu consacré aux frais de garde d'enfants (PRFE) : 12,51

- Conflits entre responsabilités professionnelles et familiales par mois (Conf) : 3,68

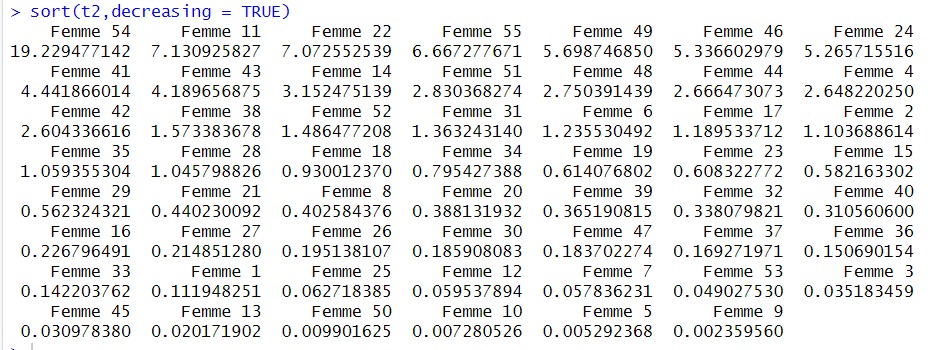
- Heures des tâches ménagères par semaine (HMS) : 6,14

**Selon les coordonnes**

****

**=> Interprétation :** Dans notre analyse, nous constatons que certaines femmes, comme Femme 3, sont fortement influencées par des variables telles que la satisfaction envers la famille, l'âge avancé et les aspects économiques et administratifs. Par exemple, Femme 3 pourrait être une personne âgée qui accorde une grande importance à sa famille et qui est financièrement stable. En revanche, d'autres femmes, comme Femme 16, semblent être influencées par des caractéristiques opposées, telles qu'une satisfaction moindre envers la famille ou des aspects économiques moins importants. Par exemple, Femme 16 pourrait être une jeune professionnelle concentrée sur sa carrière et moins préoccupée par sa situation familiale. Ces exemples illustrent comment les variables influencent différemment les individus dans notre analyse.

**2.2) 1er Axe :**

* **Pour les individus :** On à 55 individus dons le seuil est : 1,81%

Donc on constate que tous les contributions qui sont supérieur a 1,81% :

- Femme 54 : 19,229477142

- Femme 11 : 7,130925827

- Femme 22 : 7,072552539

- Femme 55 : 6,667277671

- Femme 49 : 5,698746850

- Femme 46 : 5,336602979

- Femme 24 : 5,265715516

- Femme 41 : 4,441866014

- Femme 43 : 4,189656875

- Femme 14 : 3,152475139

-Femme 51 : 2,830368274

- Femme 48 : 2,750391439

- Femme 44 : 2,666473073

- Femme 4 : 2,648220250

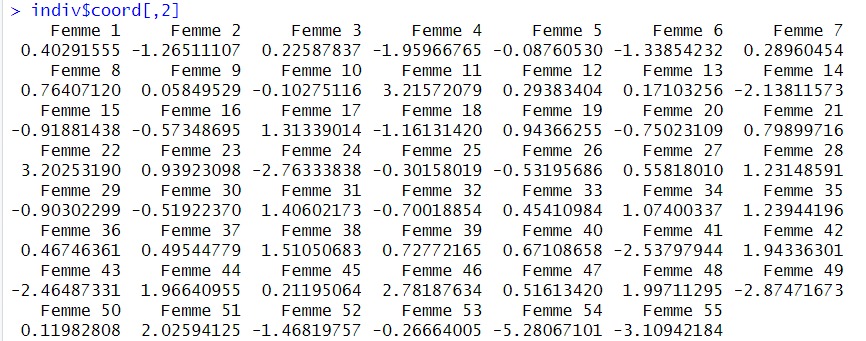
- Femme 42 : 2,604336616

- Femme 38 : 1,573383678

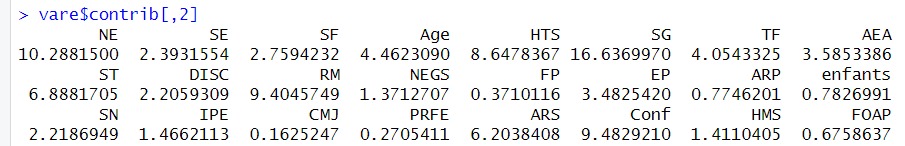
- Femme 52 : 1,486477208

- Femme 31 : 1,363243140

- Femme 6 : 1,235530492



* **Pour les variables :** On a 24 variables donc : 1/24=4.16



**Selon la contribution** Les variables dont les contributions sur le premier axe de l'ACP sont supérieures à 4,16 sont considérées comme ayant une influence significative sur la variation observée le long de cet axe. Voici les variables correspondantes :

-NE : 10,2881500

-HTS : 8,6478367

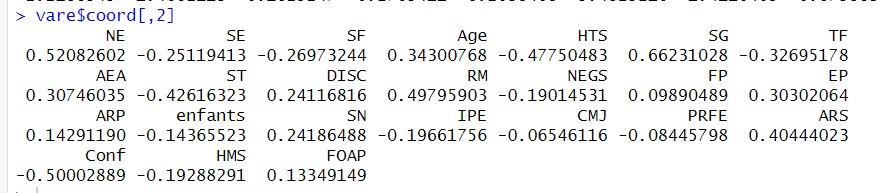
-SG : 16,6369970

-ST : 6,8881705

-RM : 9,4045749

-ARS : 6,2038408

-Conf : 9,4829210

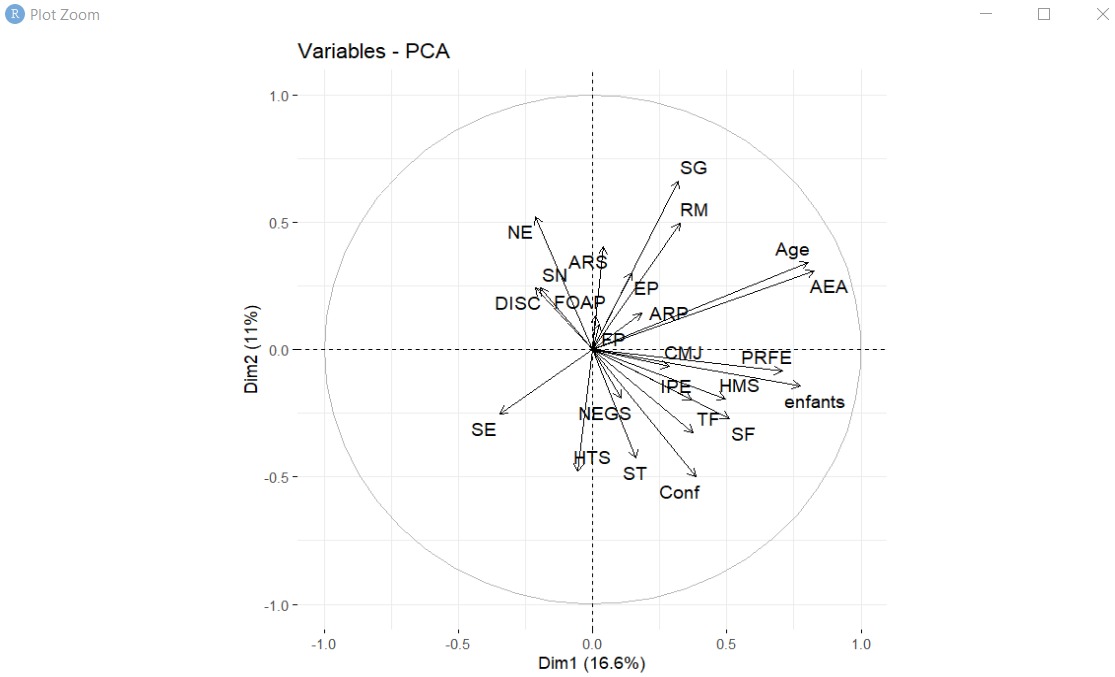
**Selon les coordonnes**

**=>Interprétation :** Parmi les individus dont les contributions sont élevées, on observe deux groupes distincts en fonction des variables qui les influencent le plus. Dans un groupe, des individus comme Femme 54, Femme 11 et Femme 22 sont fortement influencés par des facteurs tels que le secteur d'emploi, la satisfaction globale au travail et l'impact des politiques. Cela suggère qu'ils peuvent être particulièrement sensibles à leur environnement professionnel, leur satisfaction au travail et les politiques gouvernementales en matière d'emploi. Dans l'autre groupe, des individus comme Femme 46, Femme 24 et Femme 41 sont moins influencés par ces facteurs, mais peuvent être davantage touchés par des aspects tels que la flexibilité du travail, la discrimination au travail et le niveau de stress lié au travail.

1. **3eme et 4eme Etape :**

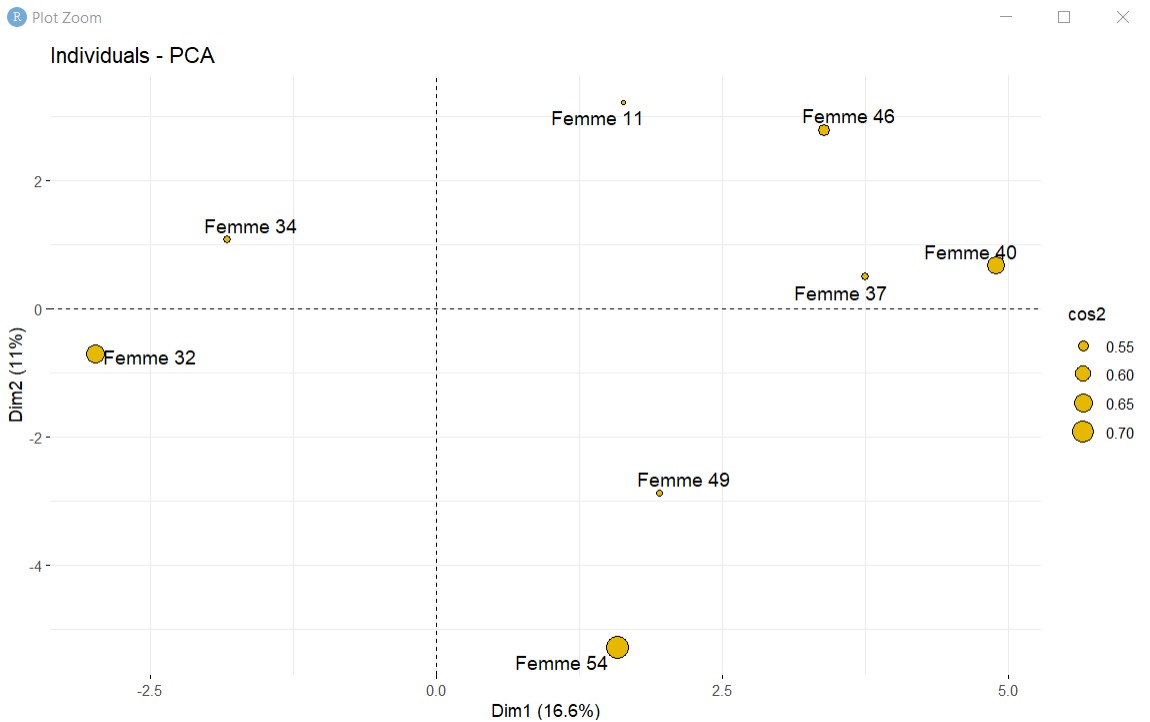
**Pour les variables :**

Parmi les variables qui dépassent le seuil de 0,5 en termes de qualité de représentation, nous identifions l'AEA, l'Age, les enfants, le SG, le PRFE, le Conf, le RM et le SF. Ces variables jouent un rôle crucial dans la caractérisation de la composante principale, ce qui signifie qu'elles contribuent de manière significative à la structure des données observées. Leur forte qualité de représentation indique qu'elles capturent efficacement la variation présente dans l'ensemble des données, fournissant ainsi des informations essentielles pour l'interprétation et la compréhension des tendances sous-jacentes.

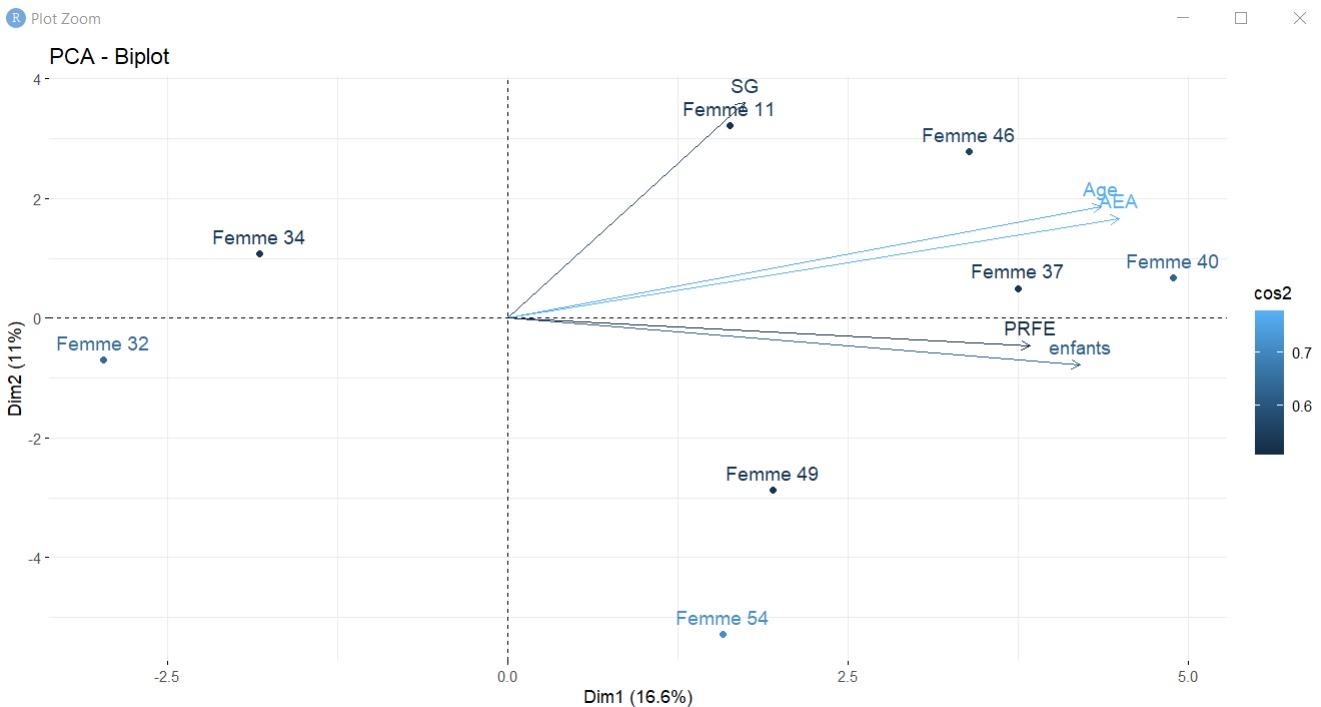
****

**Pour les individus :**

Parmi les individus dont la qualité de représentation dépasse le seuil de 0,5, nous n’observons que les femmes telles que Femme 54, Femme 32, Femme 40, Femme 46, Femme 34 et Femme 37 sont bien représentées dans la composante principale. Cela signifie que ces femmes contribuent de manière significative à la structure des données observées et capturent efficacement la variation présente dans l'ensemble des données.



1. **Synthèse :**

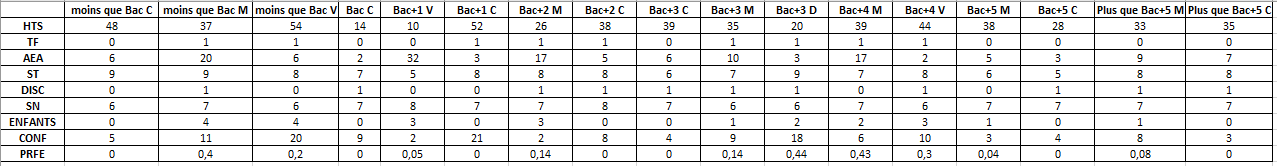


Lors de la cinquième étape de l'analyse, une conclusion importante a émergé : un seul groupe de femmes de 37, 40 et 46. Ce groupe est caractérisé par une influence significative de l'âge, du nombre d'années d'emploi actuel, du pourcentage de revenu consacré aux frais de garde d'enfants et du nombre d'enfants. Ces variables semblent jouer un rôle majeur dans la situation de ce groupe spécifique de femmes, suggérant qu'elles sont confrontées à des défis ou des dynamiques similaires, indépendamment de leur âge précis. Cette observation souligne l'importance de considérer non seulement l'âge, mais également d'autres facteurs tels que la situation professionnelle et les responsabilités familiales lors de l'analyse des données démographiques et socio-économiques.

**Analyse factorielle des correspondances (AFC)**

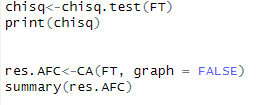
1. **Introduction**

L'Analyse Factorielle des Correspondances (AFC) est une méthode pour analyser des données catégorielles, en révélant les associations entre les variables et en les représentant dans un espace de dimensions réduites. Elle aide à comprendre les relations entre différentes catégories de variables. Dans votre projet sur les femmes travailleuses, l'AFC pourrait éclairer les caractéristiques influençant leur situation professionnelle.

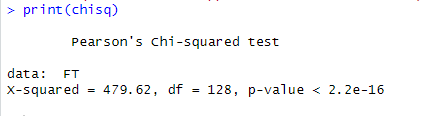


1. **Etape 1 : Le test Khi 2 et lancement du AFC**

**Code :**

****

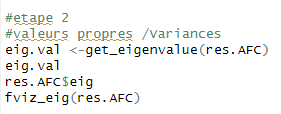
**Résultats :**

****

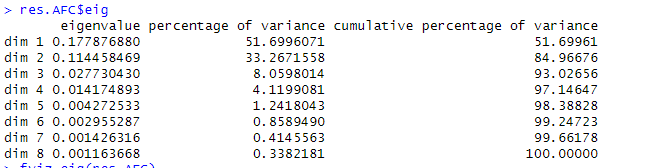
* Le test de khi² sert à connaitre l’indépendance des variables.
* On a alpha = 0,05 > p-value = 2,2e-16 Donc on rejette l’hypothèse 0 et les variables ne sont pas indépendantes.

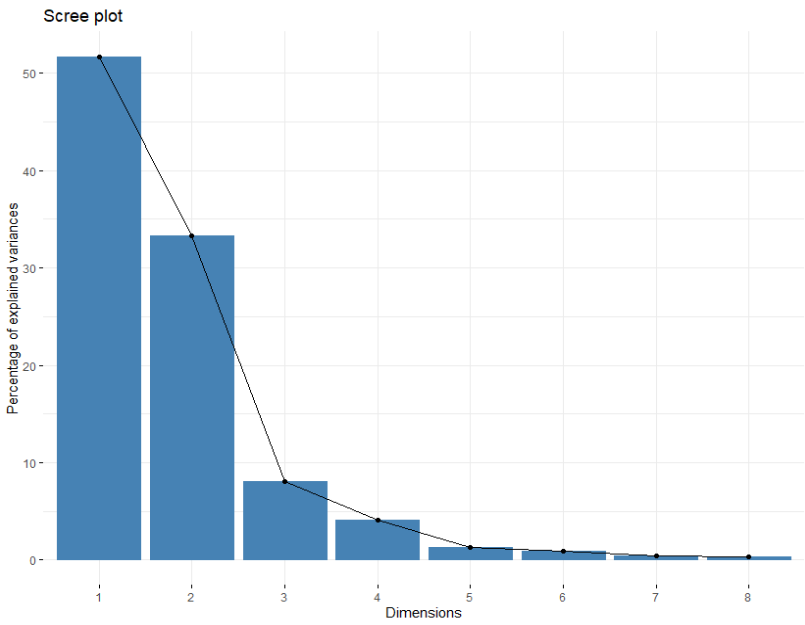
1. **Etape 2 : Choix du nombre d’axes**

**Code :**

****

**Résultats :**

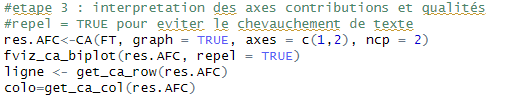
****

****

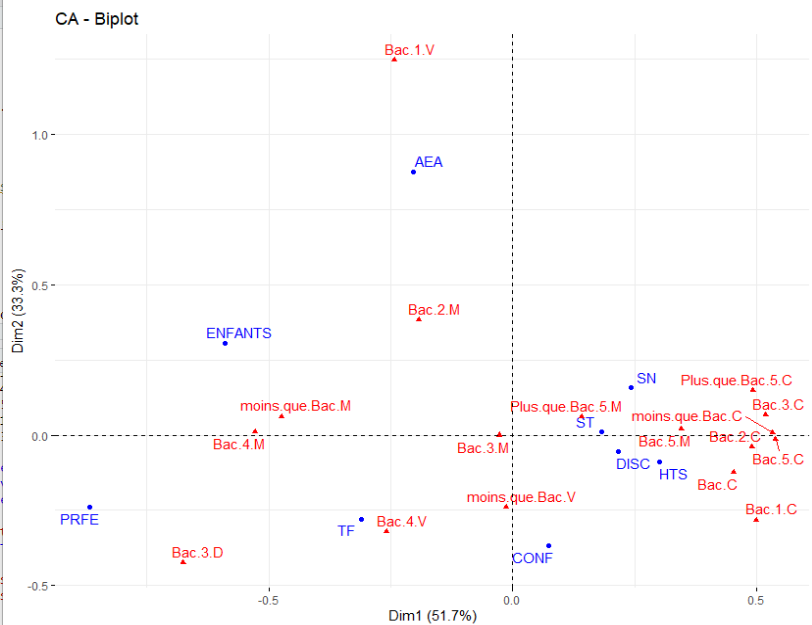
* D’après le cumule et le graphe, on va prendre 2 axes avec un taux d’inertie total de 86 ,96%.

1. **Etape 3 : Interprétation des axes, des contributions et des qualités**

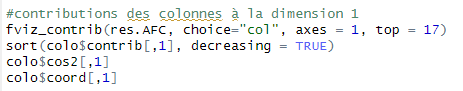
* **On relance l’AFC sur les deux axes et on isole les lignes et les colonnes :**

****

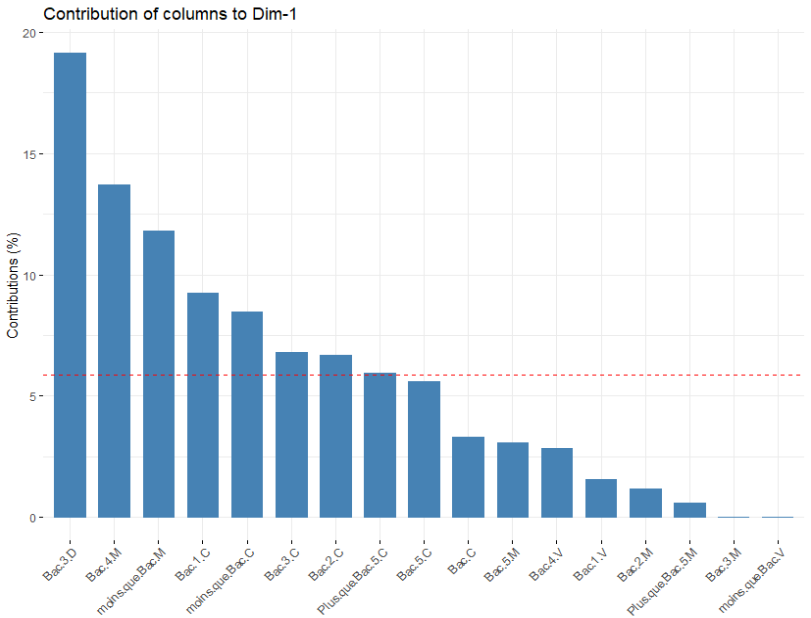
**Resultats :**

****

* **Puis on interprète l’AXE 1 :**
* Pour les colonnes :

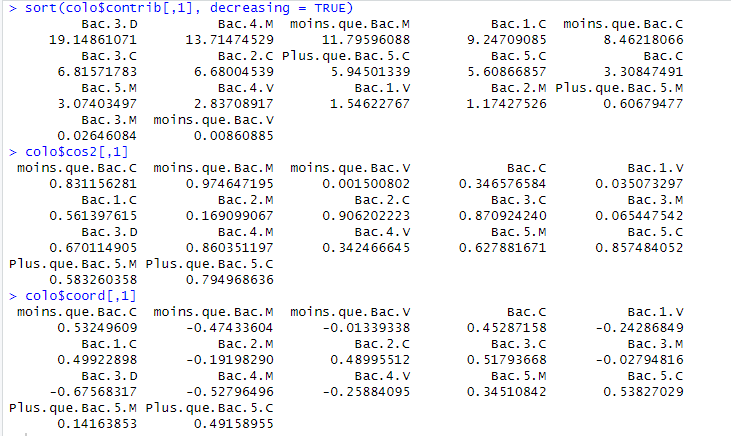
****

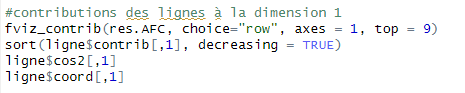
**Résultats :**

****

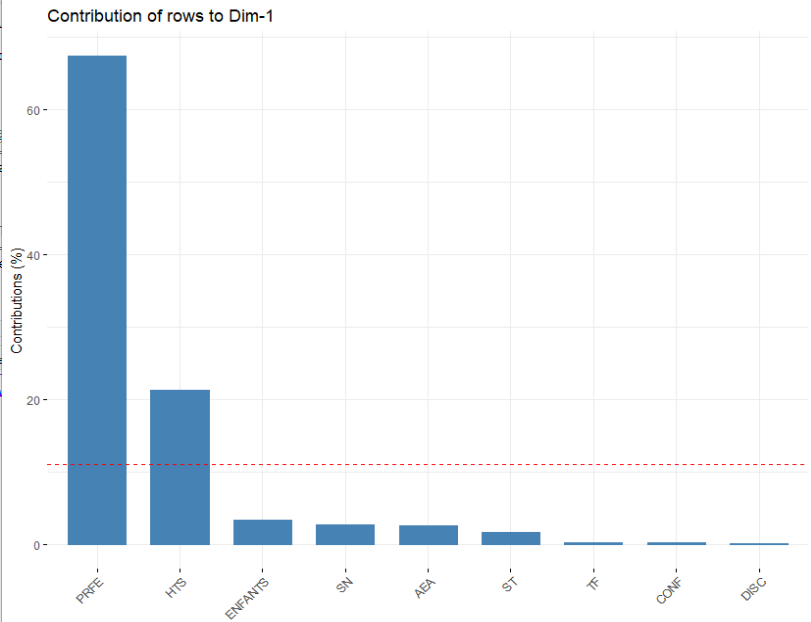
* **On a pris une moyenne de 6,68**
* **Les contributions qui sont supérieure à cette moyenne :**

**Bac+3D, Bac+4M, moins\_que\_BacM, Bac+1C, moins\_que\_BacC, Bac+3C et Bac+2C**

* **Toutes ces variables ont une bonne qualité c-à-d supérieure à 55%**
* Pour les lignes :



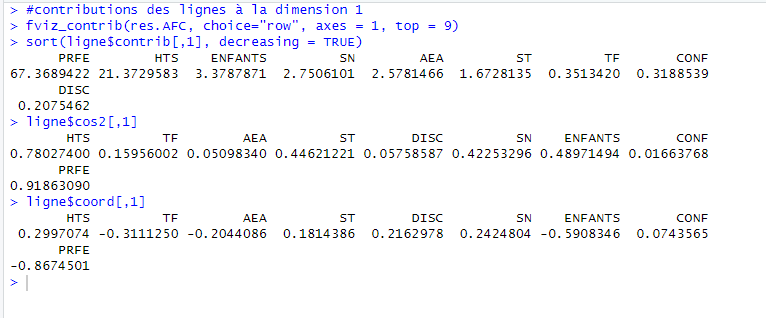
**Résultats :**



* **On a pris une moyenne de 21,37**
* **Les contributions qui sont supérieure à cette moyenne :**

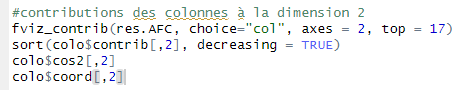
**PRFE et HTS**

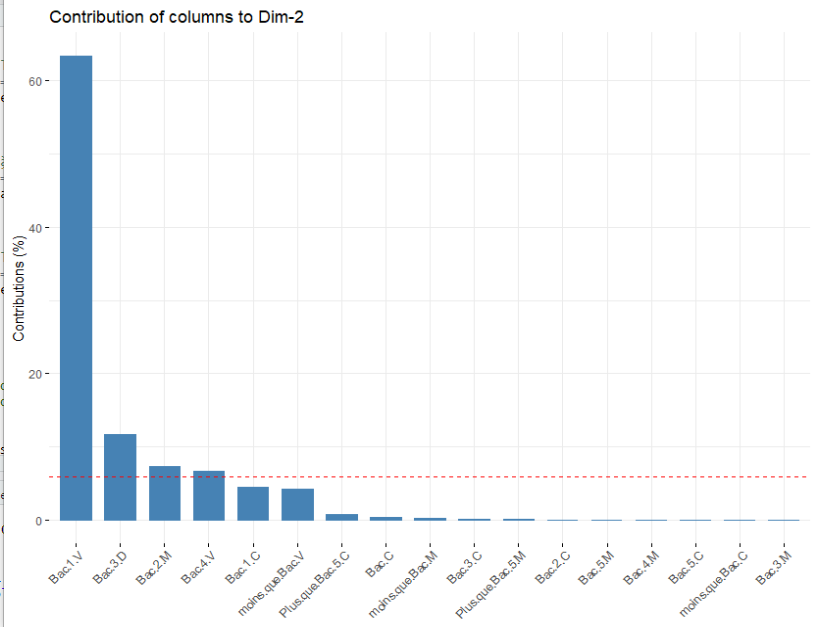
* **Toutes ces variables ont une bonne qualité c-à-d supérieure à 55%**

****

**Résultats Finale :**

* **Les femmes qui ont Bac+3 Divorcée, Bac+4 et moins que Bac Mariée sont ceux qui dépensent le plus grand pourcentage de leurs salaires pour la garde de leurs enfants.**
* **Les femmes qui ont Bac+1, moins que Bac, Bac+3 et Bac+2 tous célibataires sont ceux qui travaillent le plus d’heures par semaine.**
* **Puis on interprète l’AXE 2 :**
* Pour les colonnes :

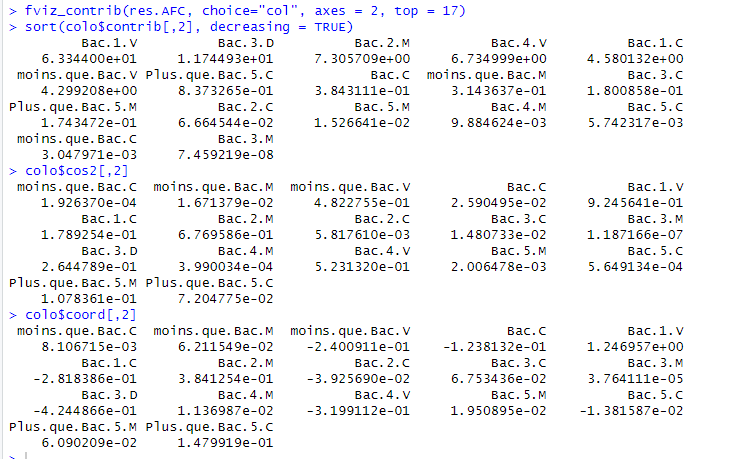
****

**Résultats :** 

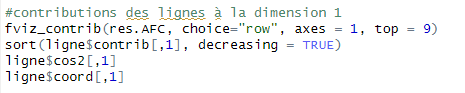
* **On a pris une moyenne de 6,73**
* **Les contributions qui sont supérieure à cette moyenne :**

**Bac+3D, Bac+4V, Bac+1V et Bac+2M**

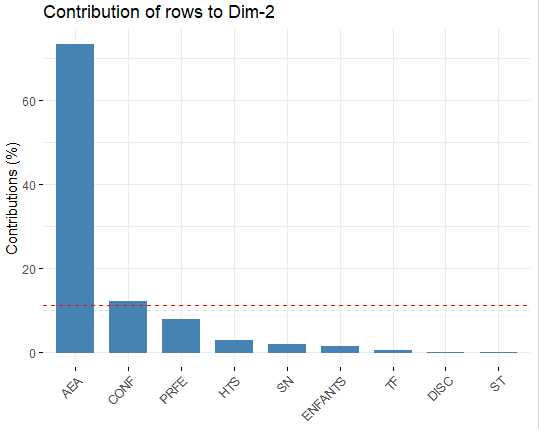
* **Toutes ces variables ont une bonne qualité c-à-d supérieure à 55% sauf Bac+3D et BAC+4V**

****

* Pour les lignes :



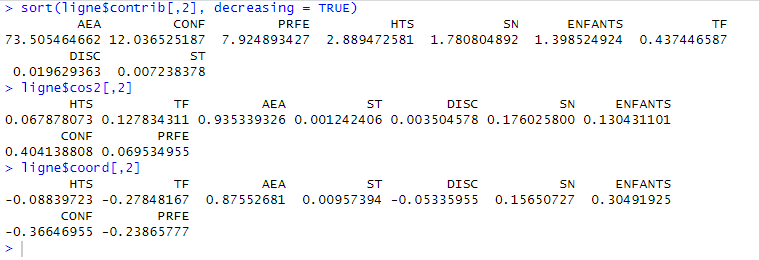
**Résultats :**



* **On a pris une moyenne de 12,03**
* **Les contributions qui sont supérieure à cette moyenne :**

**AEA et CONF**

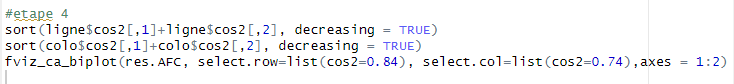
* **Toutes ces variables ont une bonne qualité c-à-d supérieure à 55% sauf CONF**

****

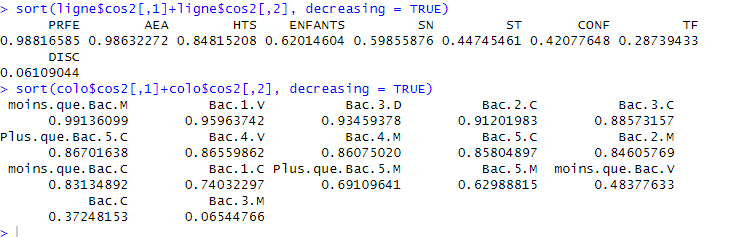
**Résultats Finale :**

* **Les femmes qui ont Bac+1 Veuf, Bac+2 Mariée ont passé le plus d’années dans leurs emploi actuel**

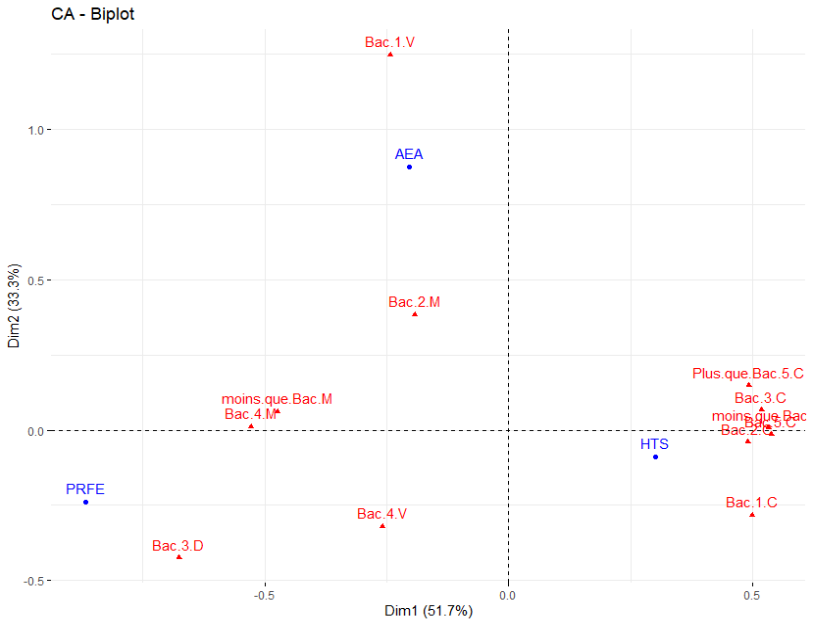
1. **Etape 4 : Formulation des groupes**

**Code : **

* **La 1ère ligne dans le code c’est pour connaitre la qualité des lignes pour la représentation du dernier graphe**
* **La 2ème ligne dans le code c’est pour connaitre la qualité des colonnes pour la représentation du dernier graphe**

****

* **Pour les lignes on a pris la qualité supérieure ou égale à 0,7 c’est : 0,84**
* **Pour les colonnes on a pris la qualité supérieure ou égale à 0,7 c’est : 0,74**

****

**Résultats Finale :**

* **On a trouvé 3 groupes :** 
  + **Groupe 1 :** Bac+1 Veuf, Bac+2 Mariée avec (AEA)
  + **Groupe 2 :** moins que Bac,Bac+1, Bac+2, Bac+3, Bac+4 et Plus que Bac plus 5 qui ont Célibataire avec (HTS)
  + **Groupe 3 :** moins que Bac, Bac+4 qui ont Mariée et Bac+3 Divorcée avec (PRFE)

**Conclusion**

Au nom de notre équipe, nous tenons à exprimer notre sincère gratitude envers nos professeurs, Monsieur Khalidou et Monsieur Harti, pour l'opportunité et l'encadrement qu'ils nous ont offerts lors de cette expérience d'enquête dans la vie réelle. Cette expérience s'est avérée extrêmement enrichissante pour notre parcours académique. Leur guidance et leurs conseils ont été précieux tout au long du processus, nous permettant de développer nos compétences et notre compréhension dans la conduite d'une enquête de manière professionnelle et rigoureuse. Leur dévouement et leur soutien ont grandement contribué à notre réussite, et nous sommes reconnaissants de l'opportunité qui nous a été offerte de bénéficier de leur expertise.